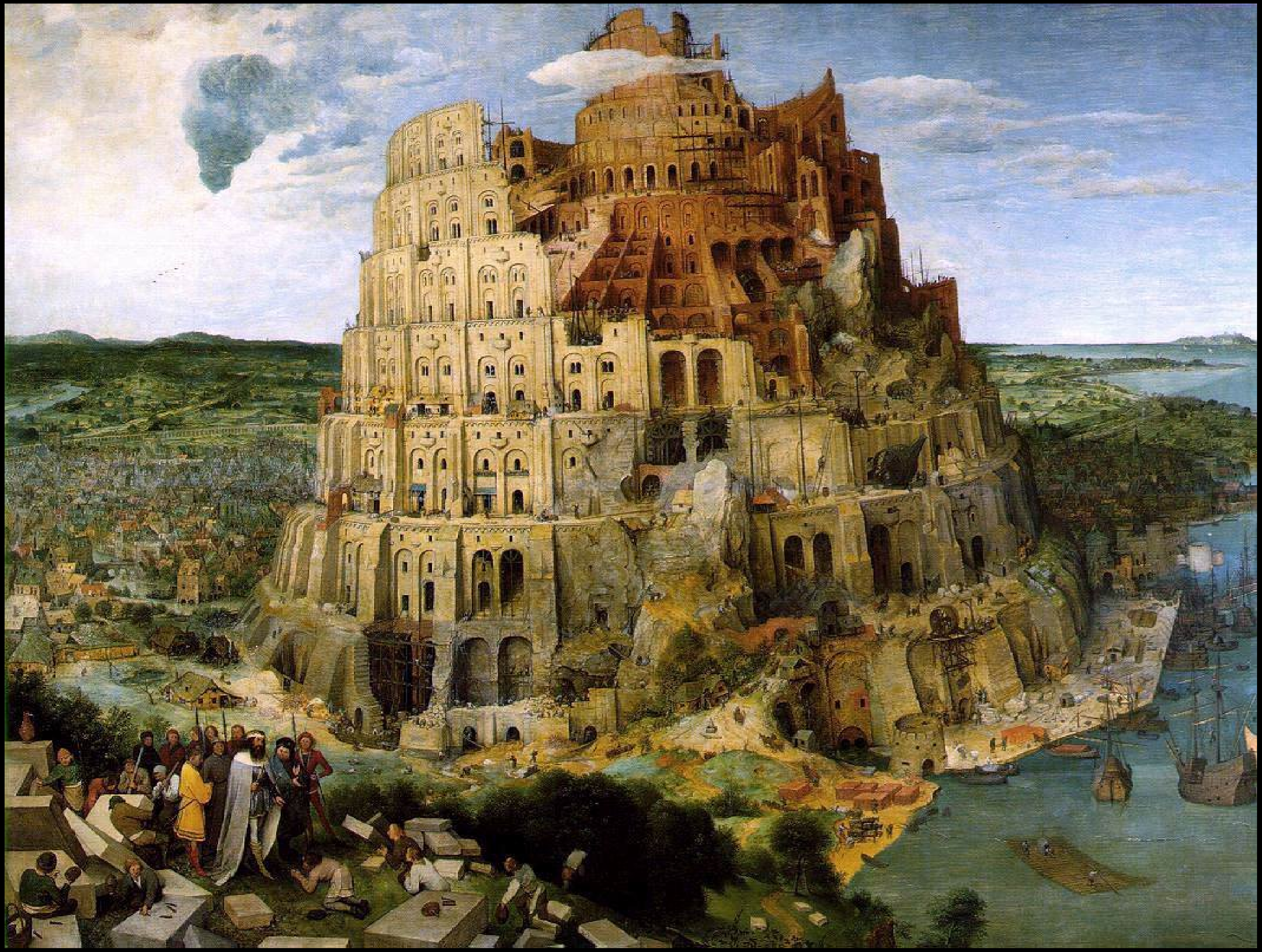
Machine Translation

3 December 2012
Natural Language Processing 600.465
Guest Lecturer: Matt Post
Slides amalgamated from mt-class.org

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

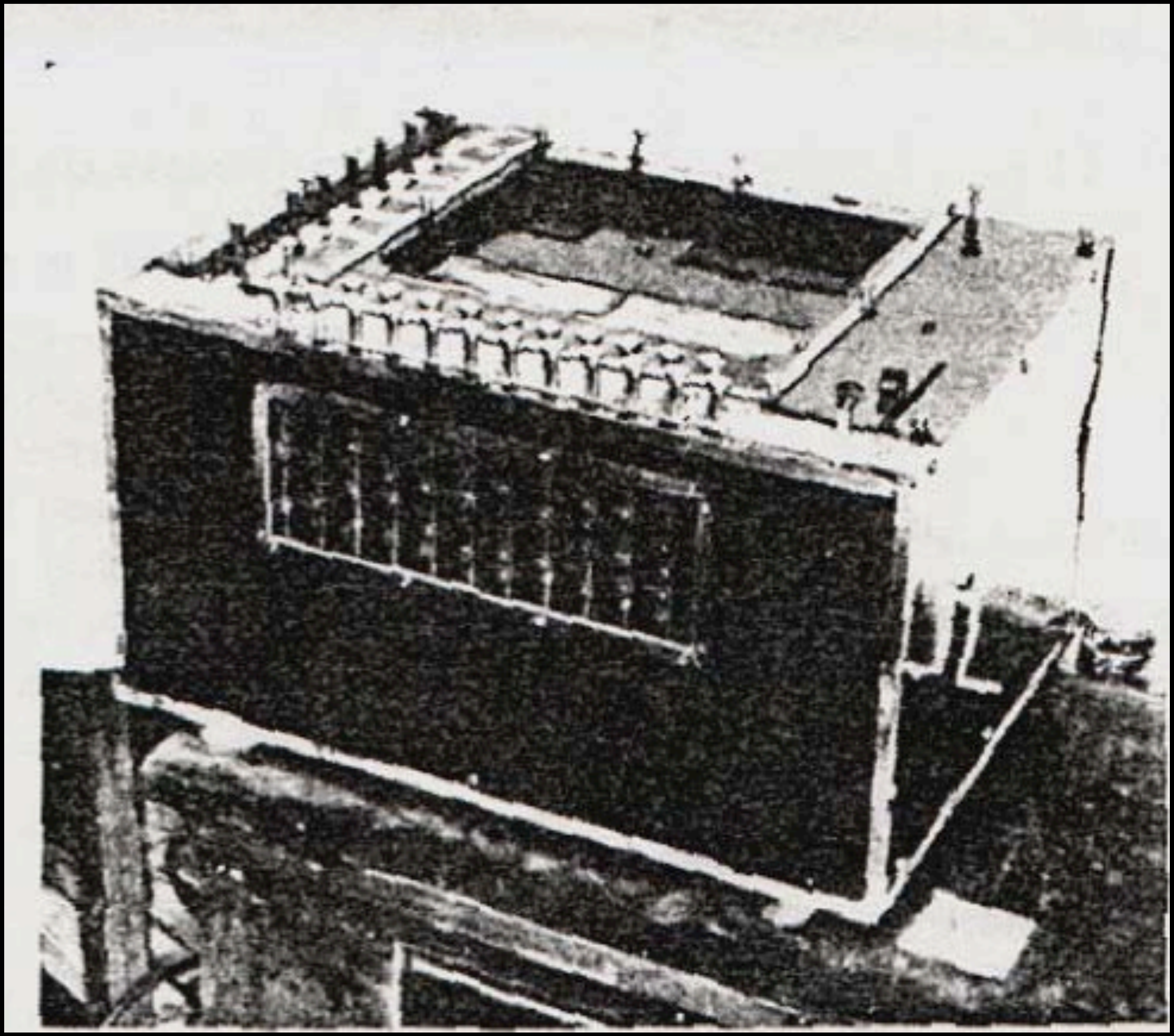
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

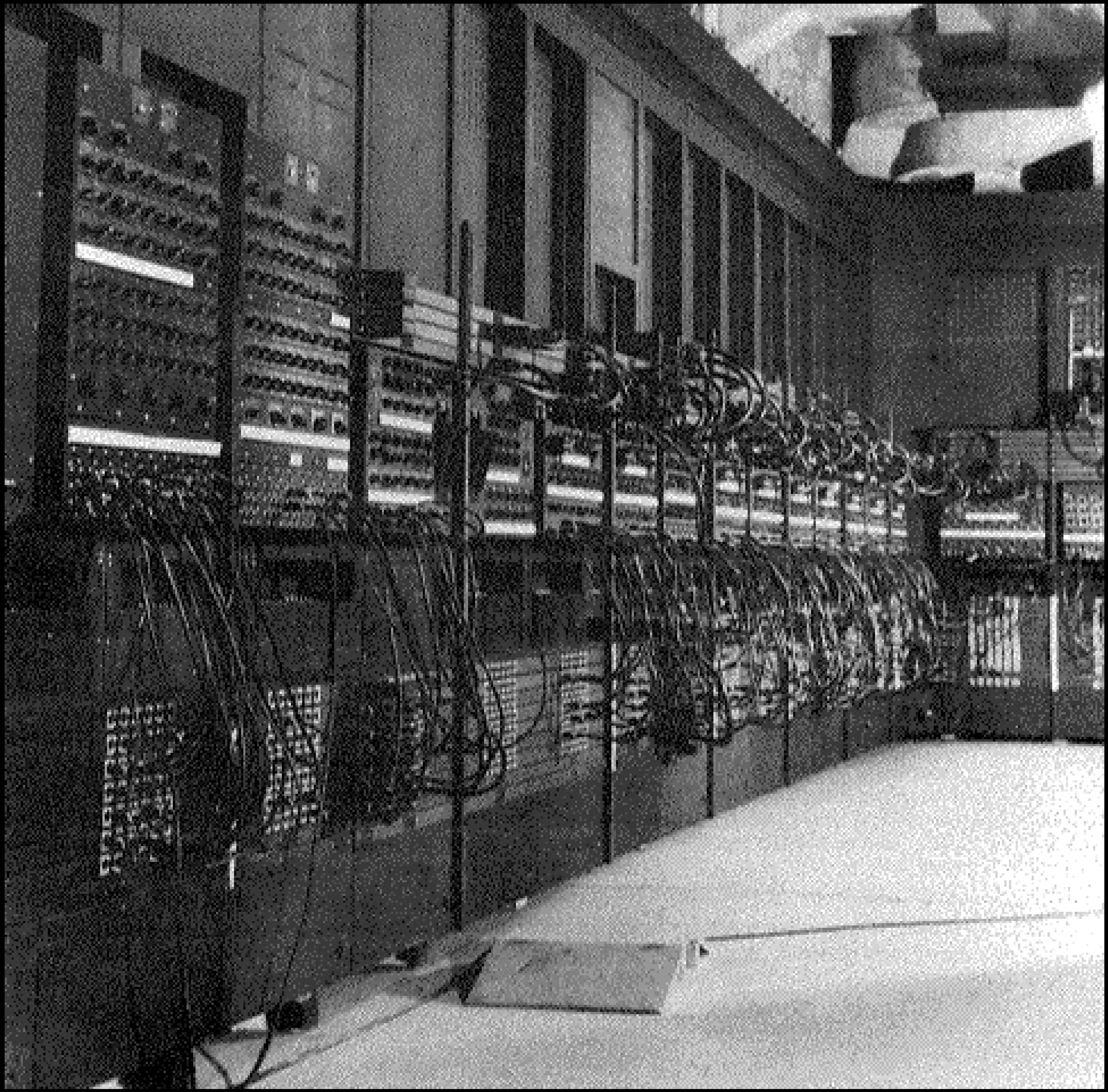


The Tower of Babel

Pieter Brueghel the Elder (1563)



Georges Artsrouni's "mechanical brain", patented 1933 (France)



ENIAC (1946)

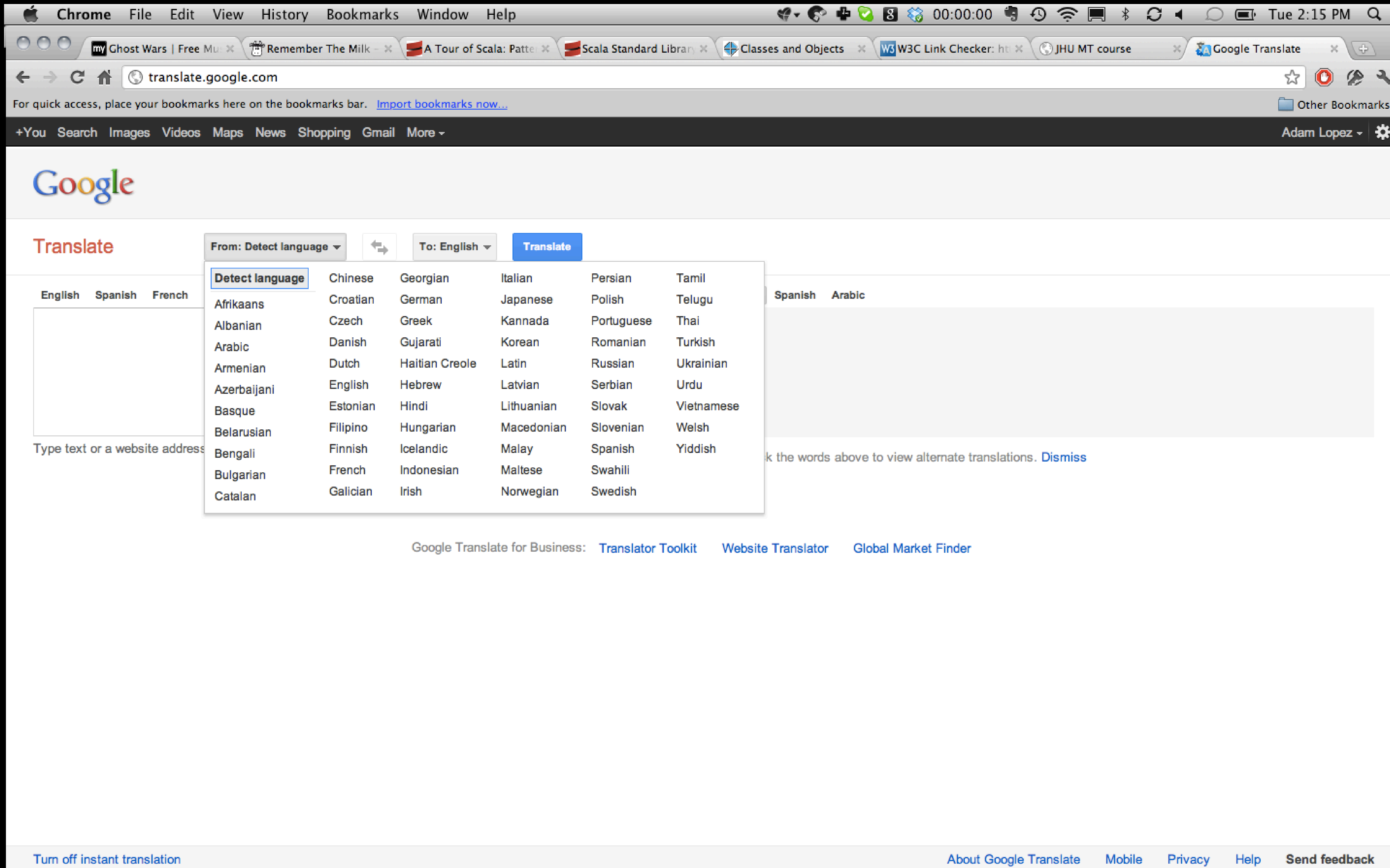
When I look at an article in Russian, I say: "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

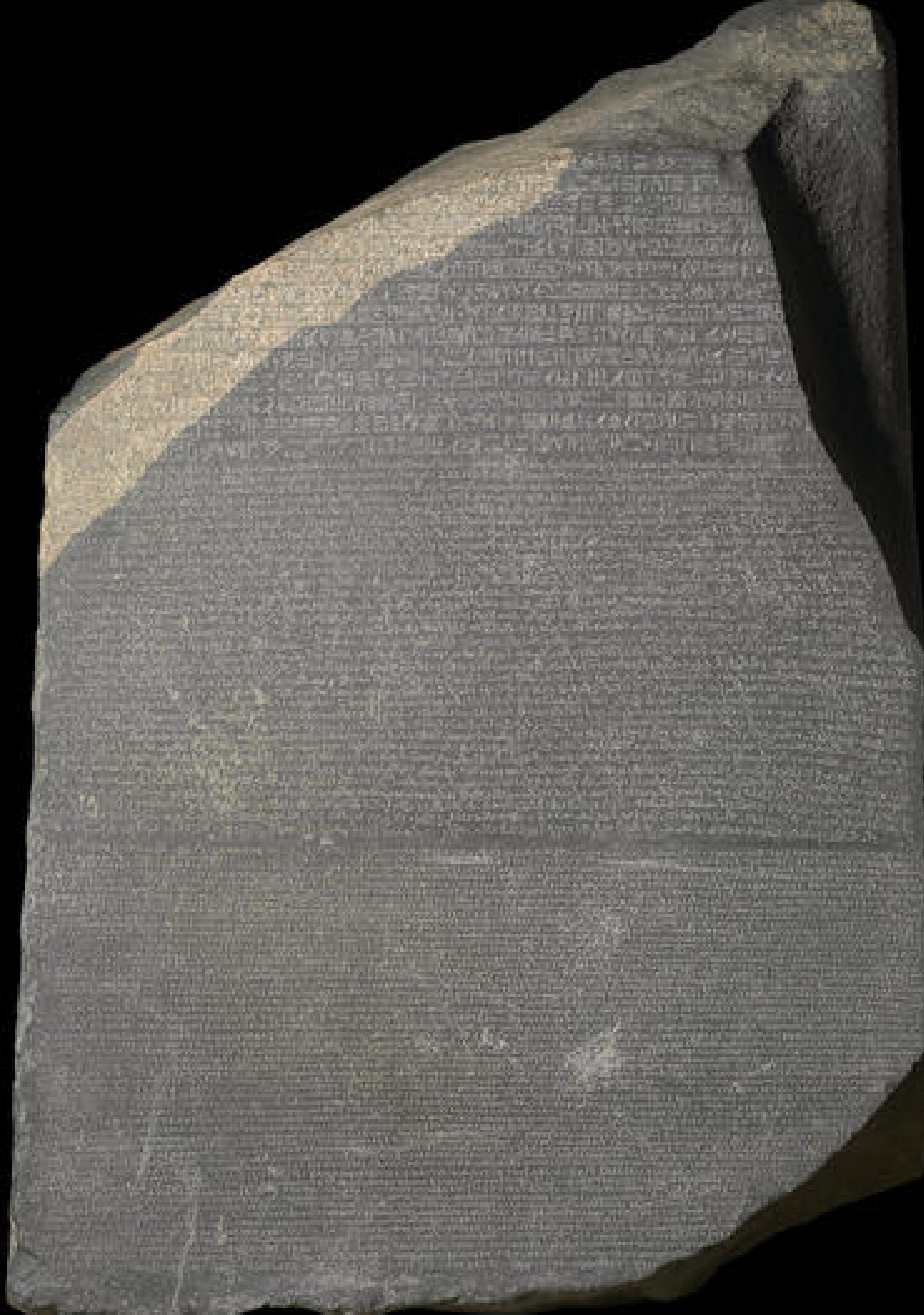
Warren Weaver (1949)

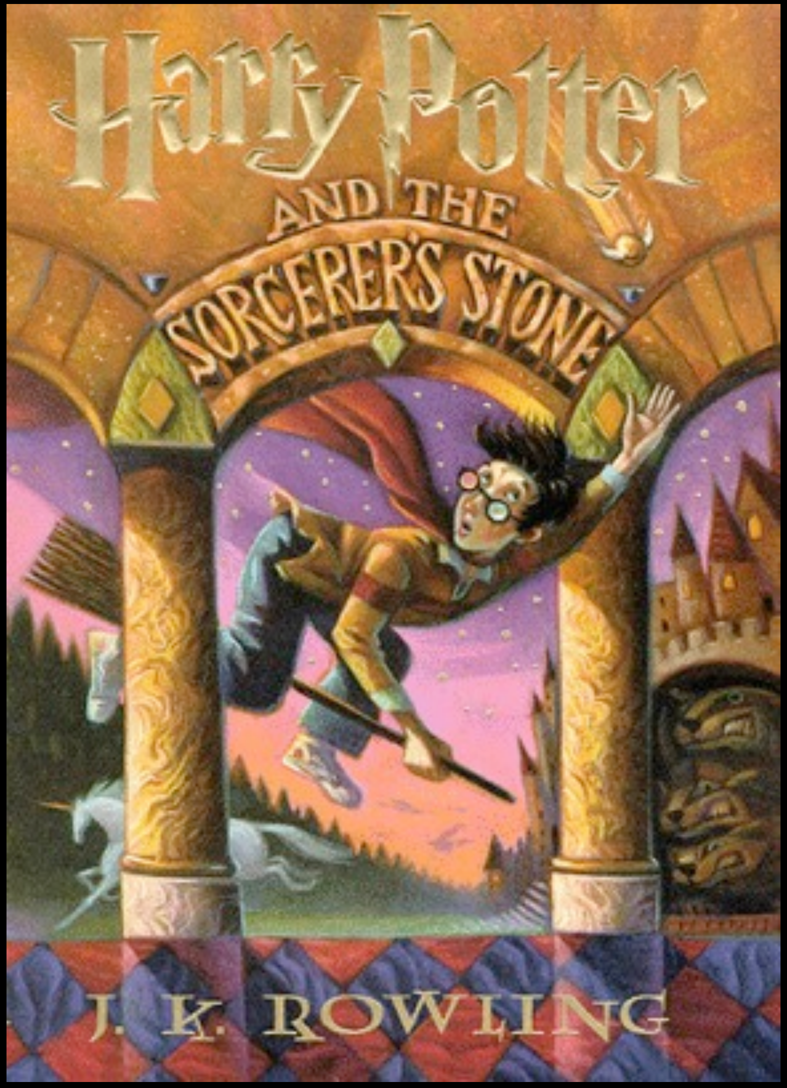
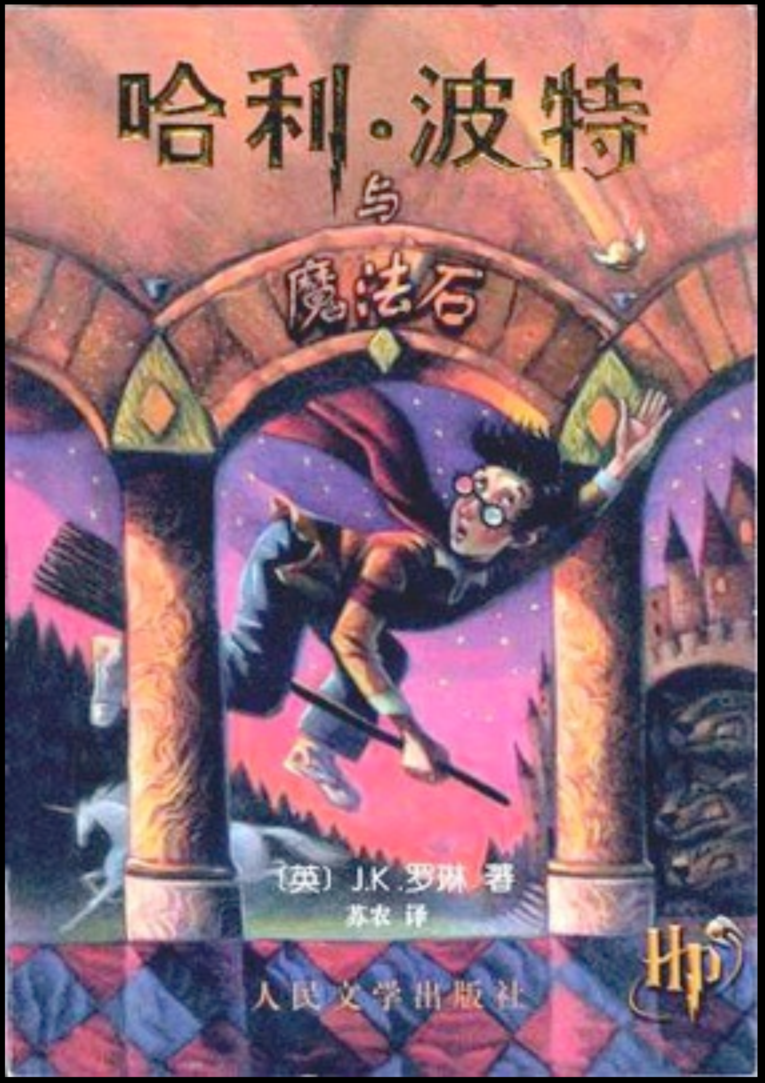
Statistical Machine Translation Live

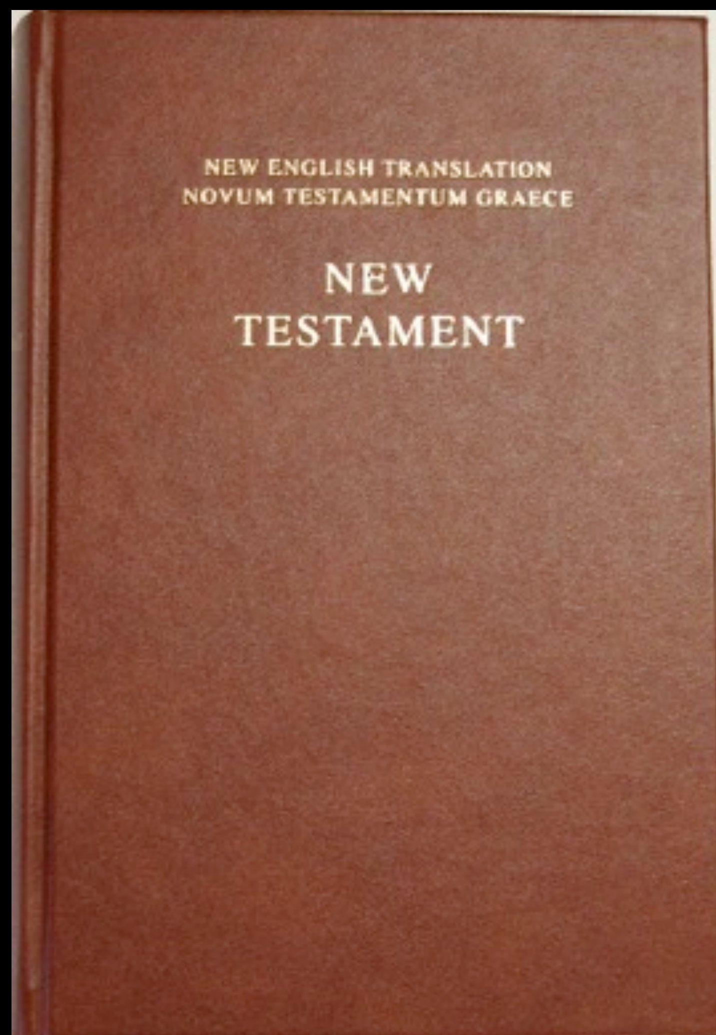
4/28/2006 03:40:00 PM
Franz Och

Because we want to provide everyone with access to all the world's information, including information written in every language, one of the exciting projects at Google Research is machine translation... Now you can see the results for yourself. We recently launched an online version of our system for Arabic-English and English-Arabic. Try it out!


Chrome File Edit View History Bookmarks Window Help
00:00:00 Tue 2:15 PM
Ghost Wars | Free Mu
Remember The Milk -
A Tour of Scala: Patte
Scala Standard Librar
Classes and Objects
W3C Link Checker: ht
JHU MT course
Google Translate
translate.google.com
For quick access, place your bookmarks here on the bookmarks bar. Import bookmarks now...
Other Bookmarks
+You Search Images Videos Maps News Shopping Gmail More
Adam Lopez
Google
Translate
From: Detect language
To: English
Translate
English Spanish French
Detect language
Afrikaans
Albanian
Arabic
Armenian
Azerbaijani
Basque
Belarusian
Bengali
Bulgarian
Catalan
Chinese
Croatian
Czech
Danish
Dutch
English
Estonian
Filipino
Finnish
French
Galician
Georgian
German
Greek
Gujarati
Haitian Creole
Hebrew
Hindi
Hungarian
Icelandic
Indonesian
Irish
Italian
Japanese
Kannada
Korean
Latin
Latvian
Lithuanian
Macedonian
Malay
Maltese
Norwegian
Persian
Polish
Portuguese
Romanian
Russian
Serbian
Slovak
Slovenian
Spanish
Swahili
Swedish
Tamil
Telugu
Thai
Turkish
Ukrainian
Urdu
Vietnamese
Welsh
Yiddish
Spanish Arabic
Type text or a website address
k the words above to view alternate translations. Dismiss
Google Translate for Business: Translator Toolkit Website Translator Global Market Finder
Turn off instant translation
About Google Translate Mobile Privacy Help Send feedback



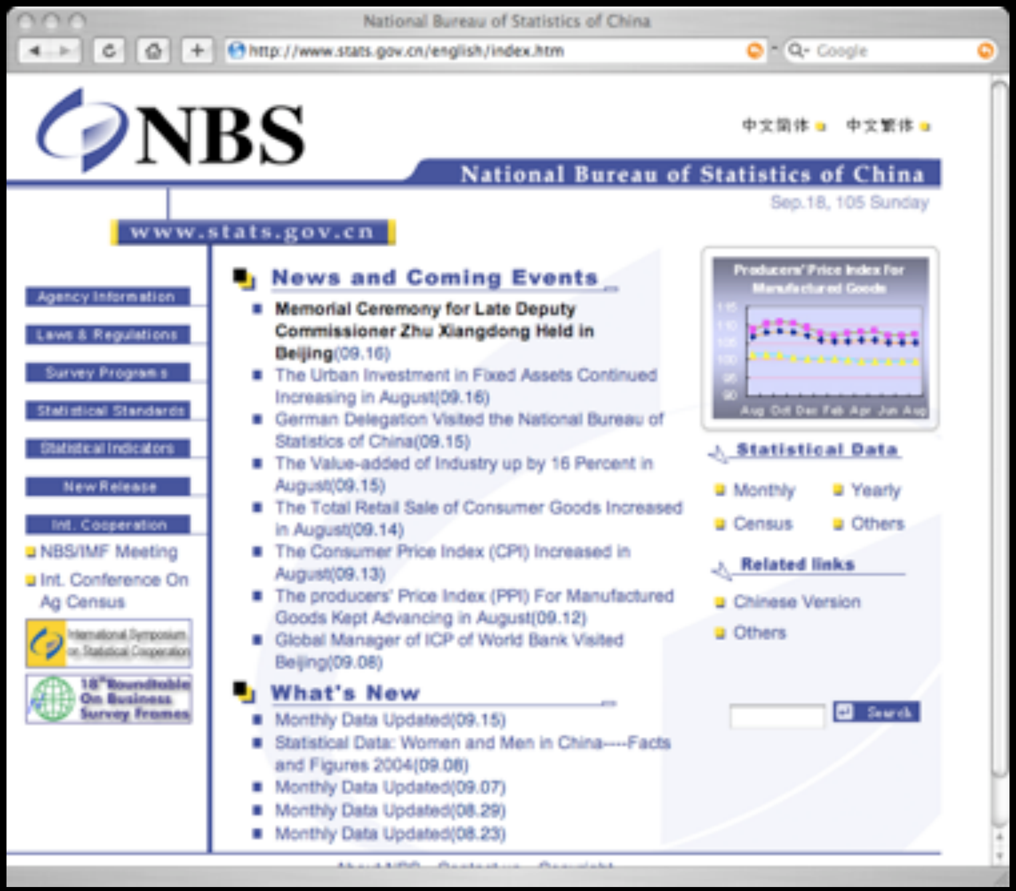

Harry Potter
AND THE
SORCERER'S STONE
J. K. ROWLING

哈利·波特
与
魔法石
〔英〕J.K.罗琳 著
苏农 译
人民文学出版社
HP


NEW ENGLISH TRANSLATION
NOVUM TESTAMENTUM GRAECE
NEW
TESTAMENT


聖經
尋道本


www.un.org
http://www.un.org/english/
Google
We the peoples
Daily Briefing | Radio, TV, Photo | Documents, Maps | Publications, Stamps, Databases | UN Works | Search
Peace & Security | Economic & Social Development | Human Rights | Humanitarian Affairs | International Law
Welcome to the
United Nations
UN Millennium Development Goals
United Nations News Centre
About the United Nations
Main Bodies
Conferences & Events
Member States
General Assembly President
8 September 2005 >>
Secretary-General
Situation in Iraq
Mideast Roadmap
Renewing the UN
UN Action against Terrorism
Issues on the UN Agenda
Civil Society / Business
UN Webcast
CyberSchoolBus
Home | Recent Additions | Employment | UN Procurement | Comments | Q & A | UN System Sites | Index
عربي | 中文 | English | Français | Русский | Español
Copyright, United Nations, 2000-2005 | Use of UN60 Logo | Terms of Use | Privacy Notice | Help
[Text version]
Live and On-Demand Webcasts, 24 Hours a Day: Click on UN Webcas


联合国主页
http://www.un.org/chinese/
Google
我们人民
每日简报 | 多媒体 | 文件与地图 | 出版物 邮票 数据库 | 服务全球 | 网址搜索
和平与安全 | 经济与社会发展 | 人权 | 人道主义事务 | 国际法
欢 迎 来 到
联 合 国
联合国千年发展目标
联合国新闻
联合国概况
联合国主要机关
会议与活动
联合国会员国
联合国大会主席
联大第60届会议一般性辩论
联合国秘书长
伊拉克局势
中东路线图
更新联合国
反恐怖主义
联合国日常议题
民间团体/商业
联合国网络直播
空中校车
新增内容 | 工作机会 | 联合国采购 | 建议 | 问题与解答 | 其他网址 | 网址索引
عربي | 中文 | English | Français | Русский | Español
联合国2000-2005年版权|联合国60周年徽标使用准则|使用条件|隐私通告|帮助
[纯文字版]


National Bureau of Statistics of China
http://www.stats.gov.cn/english/index.htm
NBS
中文简体 中文繁体
National Bureau of Statistics of China
Sep.18, 105 Sunday
www.stats.gov.cn
Agency Information
Laws & Regulations
Survey Programs
Statistical Standards
Statistical Indicators
New Release
Int. Cooperation
NBS/IMF Meeting
Int. Conference On Ag Census
News and Coming Events
Memorial Ceremony for Late Deputy Commissioner Zhu Xiangdong Held in Beijing(09.16)
The Urban Investment in Fixed Assets Continued Increasing in August(09.16)
German Delegation Visited the National Bureau of Statistics of China(09.15)
The Value-added of Industry up by 16 Percent in August(09.15)
The Total Retail Sale of Consumer Goods Increased in August(09.14)
The Consumer Price Index (CPI) Increased in August(09.13)
The producers' Price Index (PPI) For Manufactured Goods Kept Advancing in August(09.12)
Global Manager of ICP of World Bank Visited Beijing(09.08)
What's New
Monthly Data Updated(09.15)
Statistical Data: Women and Men in China---Facts and Figures 2004(09.08)
Monthly Data Updated(09.07)
Monthly Data Updated(08.29)
Monthly Data Updated(08.23)
Producers' Price Index For Manufactured Goods
Statistical Data
Monthly
Yearly
Census
Others
Related links
Chinese Version
Others
Search

中华人民共和国国家统计局
http://www.stats.gov.cn/
中华人民共和国国家统计局
National Bureau of Statistics of China
中文简体
中文繁体
ENGLISH
www.stats.gov.cn
统计公报
统计数据
统计分析
统计出版
统计法规
统计管理
数据直报
统计链接
最新统计信息
最新统计动态
工业品出厂价格指数
重要公告
统计机构
统计动态
统计标准
统计制度
统计知识


中文简体
NBS
www.stats.gov.cn
Agency Information
Laws & Regulations
Survey Programs
Statistical Standards
Statistical Indicators
New Release
Int. Cooperation
NBS/IMF Meeting
Int. Conference On Ag Census
News and Coming
Memorial Ceremony for Late Deputy Commissioner Zhu Xiangdong Held in Beijing(09.16)
The Urban Investment in Fixed Assets Continued Increasing in August(09.16)
German Delegation Visited the National Bureau of Statistics of China(09.15)
The Value-added of Industry up by 16 Percent in August(09.15)
The Total Retail Sale of Consumer Goods Increased in August(09.14)
The Consumer Price Index (CPI) Increased in August(09.13)
The producers' Price Index (PPI) For Manufactured Goods Kept Advancing in August(09.12)
Global Manager of ICP of World Bank Visited Beijing(09.08)
What's New
Monthly Data Updated(09.15)
Statistical Data: Women and Men in China----Facts and Figures 2004(09.08)
Monthly Data Updated(09.07)
Monthly Data Updated(08.29)
Monthly Data Updated(08.23)
Statistical Data
Monthly
Yearly
Census
Others
Related links
Chinese Version
Others
Search


中文繁体
ENGLISH
中华人民共和国国家
National Bureau of Statistics
统计公报
统计数据
统计分析
统计出版
统计法规
统计管理
数据直报
统计链接
最新统计信息
最新统计动态
重要公告
统计机构
统计动态
统计标准
统计制度
统计知识

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

Although north wind howls , but sky still very clear .

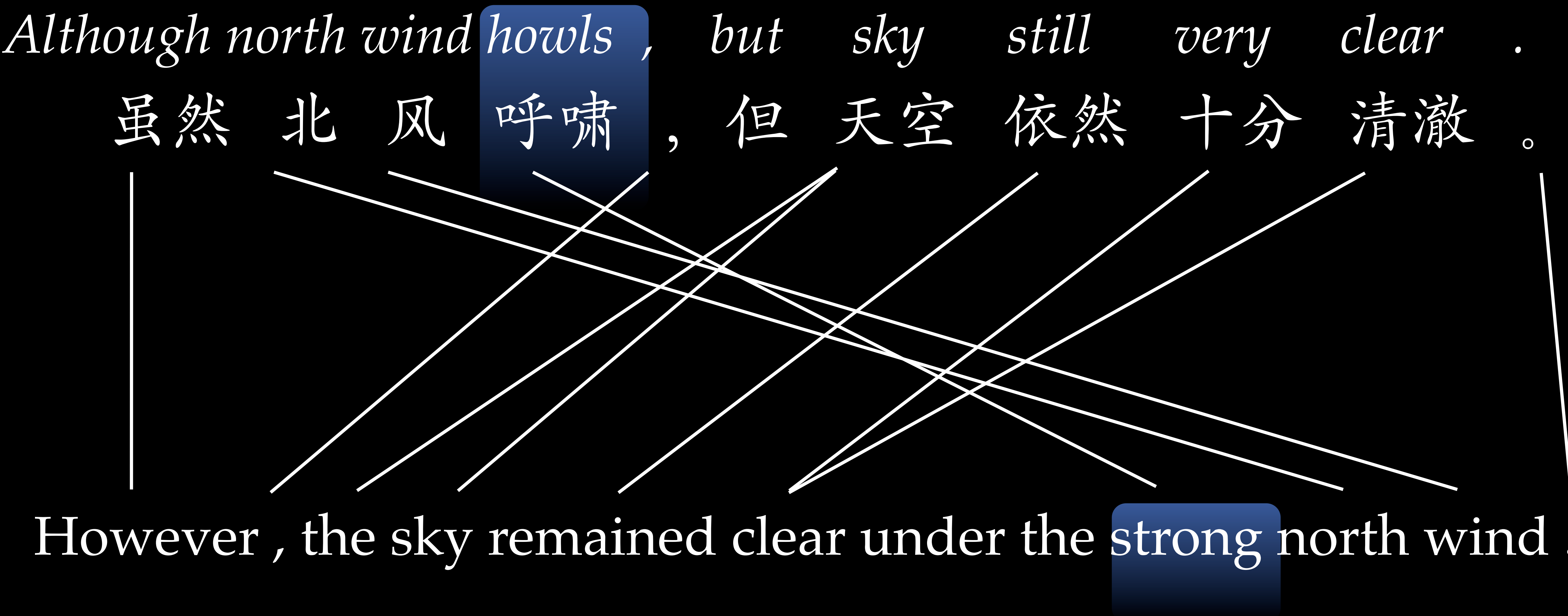
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

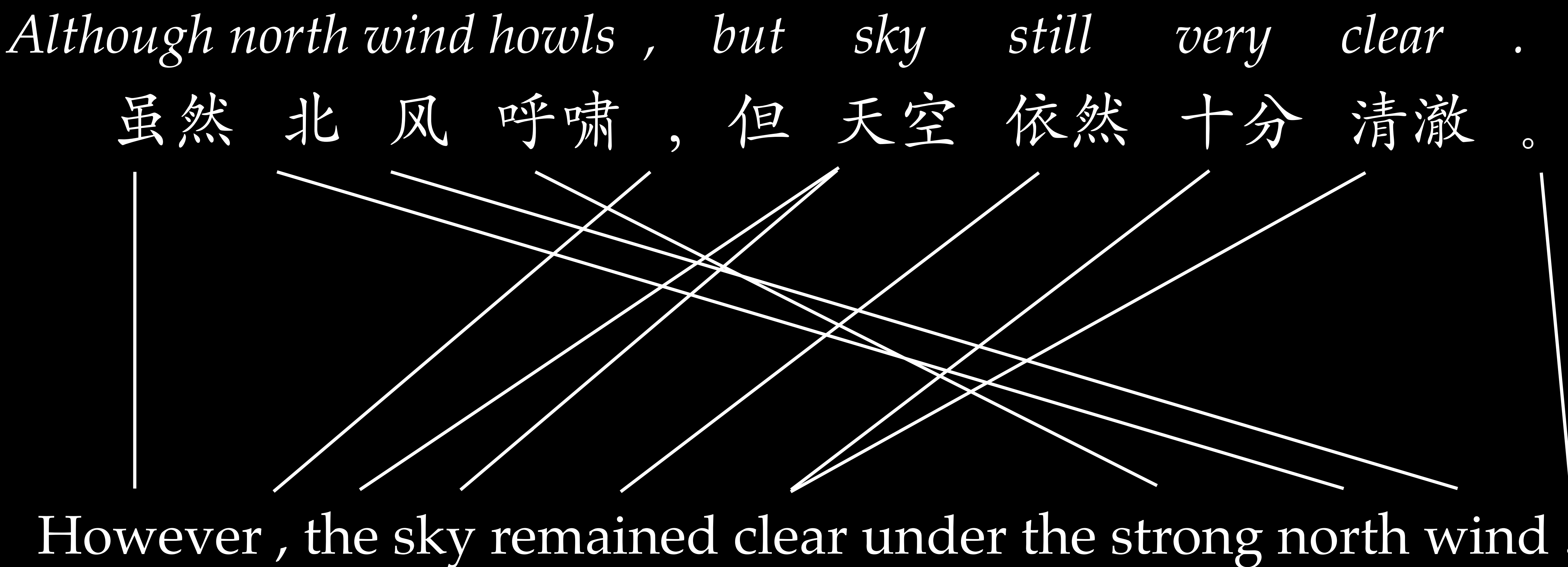
However , the sky remained clear under the strong north wind .

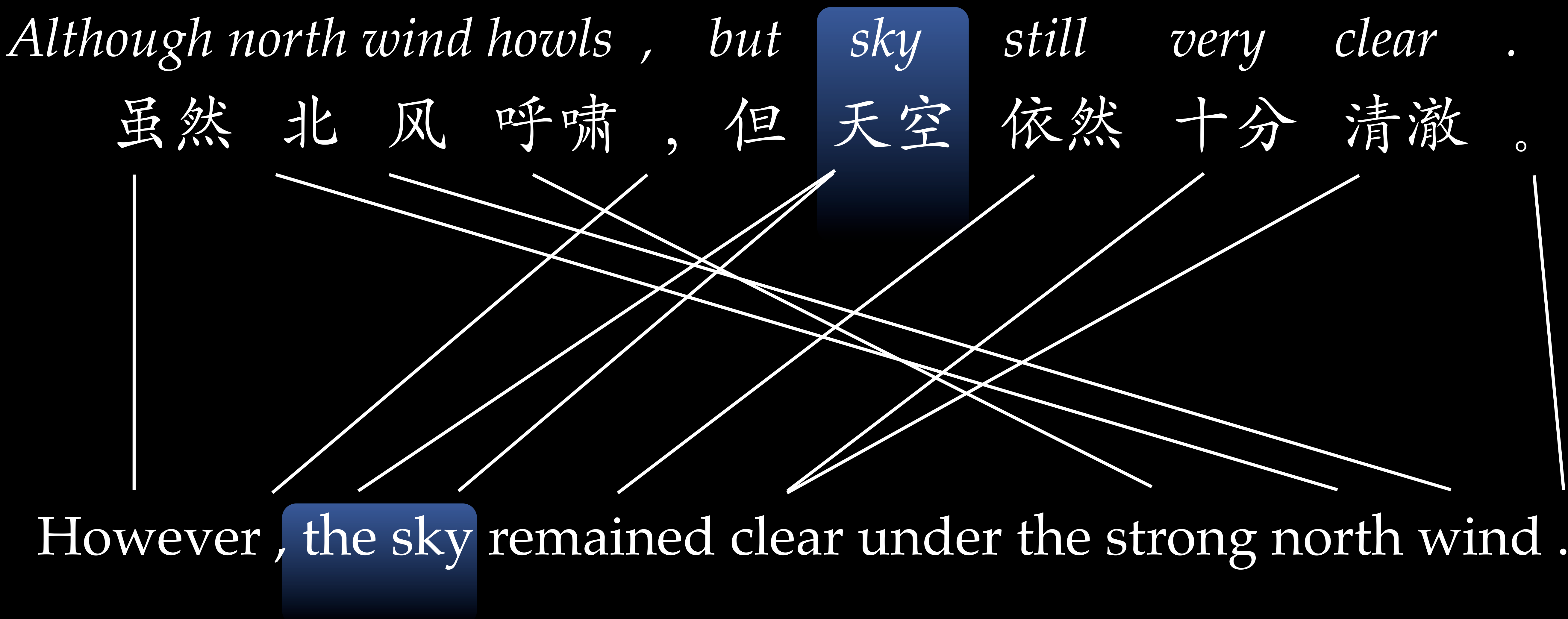
Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。
However , the sky remained clear under the strong north wind .


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。
However , the sky remained clear under the strong north wind .


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。
However , the sky remained clear under the strong north wind .

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

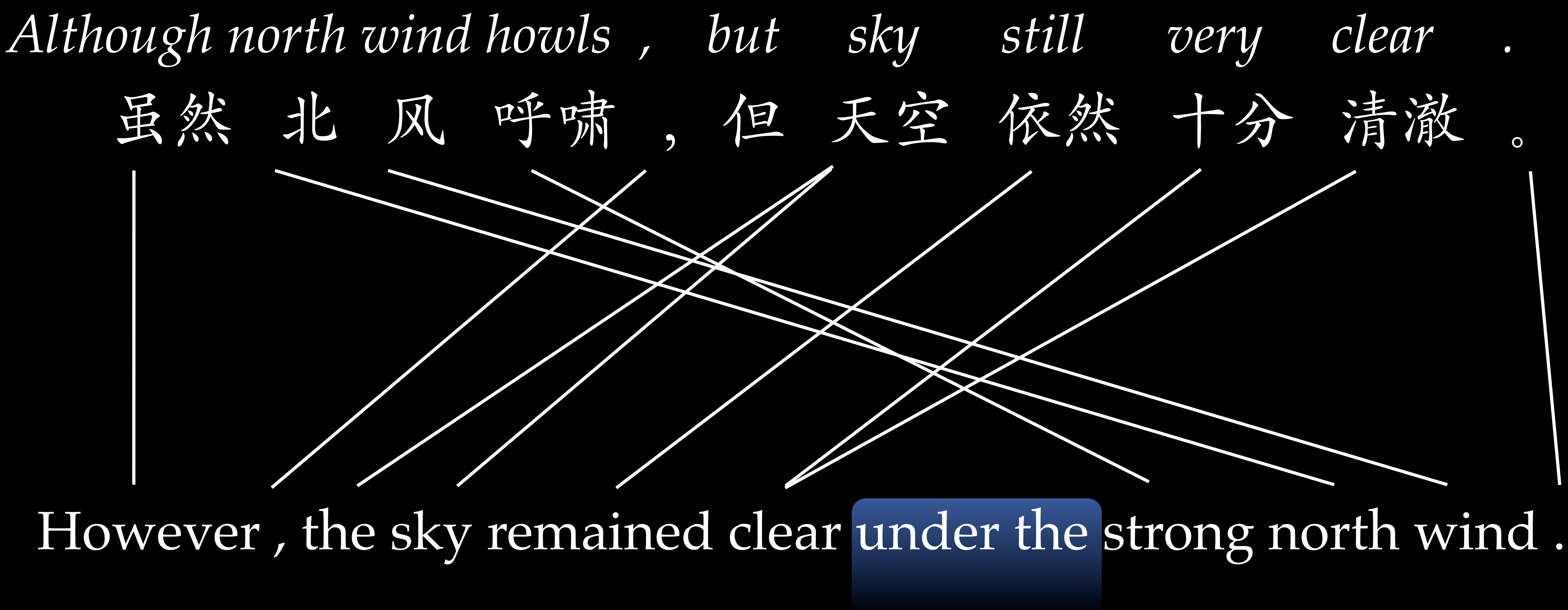
However , the sky remained clear under the strong north wind .


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。
However , the sky remained clear under the strong north wind .

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。
However , the sky remained clear under the strong north wind .

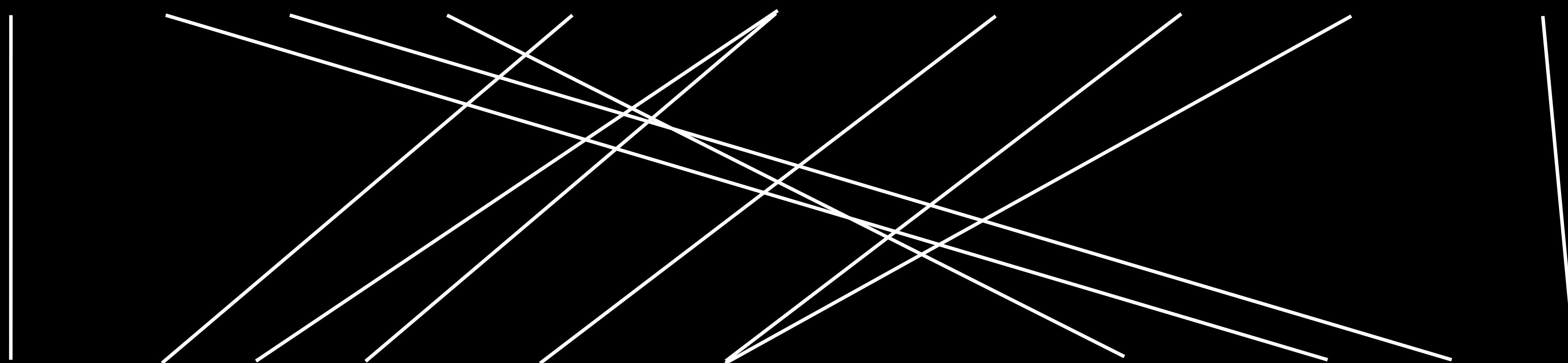
Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。



However , the sky remained clear under the strong north wind .

Despite the strong northerly winds , the sky remains very clear .

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

Despite the strong northerly winds , the sky remains very clear .

The sky was still crystal clear , though the north wind was howling .

Although a north wind was howling , the sky remained clear and blue .

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

Despite the strong northerly winds , the sky remains very clear .

The sky was still crystal clear , though the north wind was howling .

Although a north wind was howling , the sky remained clear and blue .

Questions

- How can we formalize knowledge acquisition from text as an algorithm?
- How can we formalize decoding as an algorithm?
- How well does it work?

Statistical Machine Translation

Develop a statistical ***model*** of translation that can be ***learned*** from ***data*** and used to ***predict*** the correct English translation of new Chinese sentences.

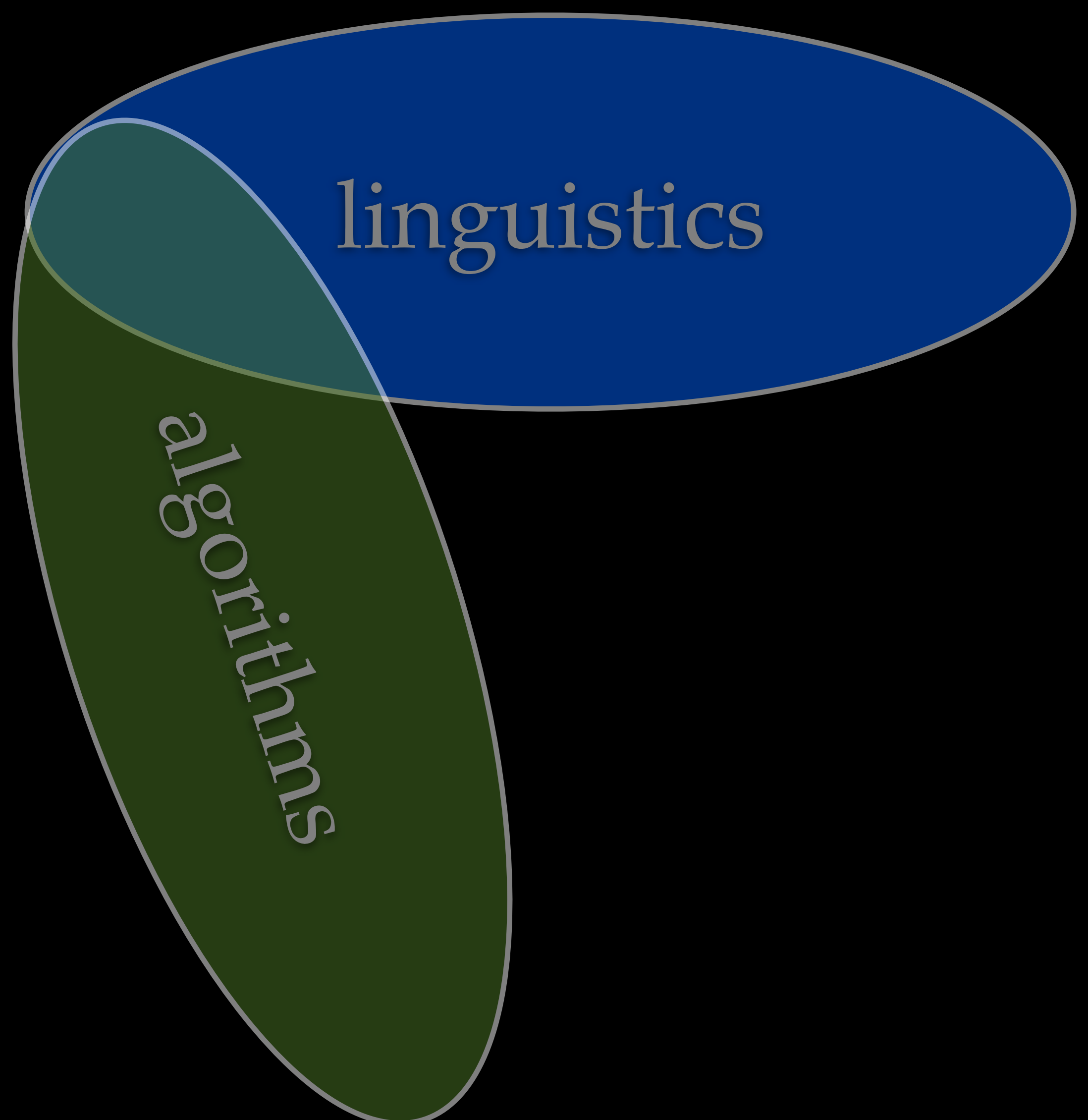
Statistical Machine Translation

Develop a statistical ***model*** of translation that can be ***learned*** from ***data*** and used to ***predict*** the correct English translation of new Chinese sentences.

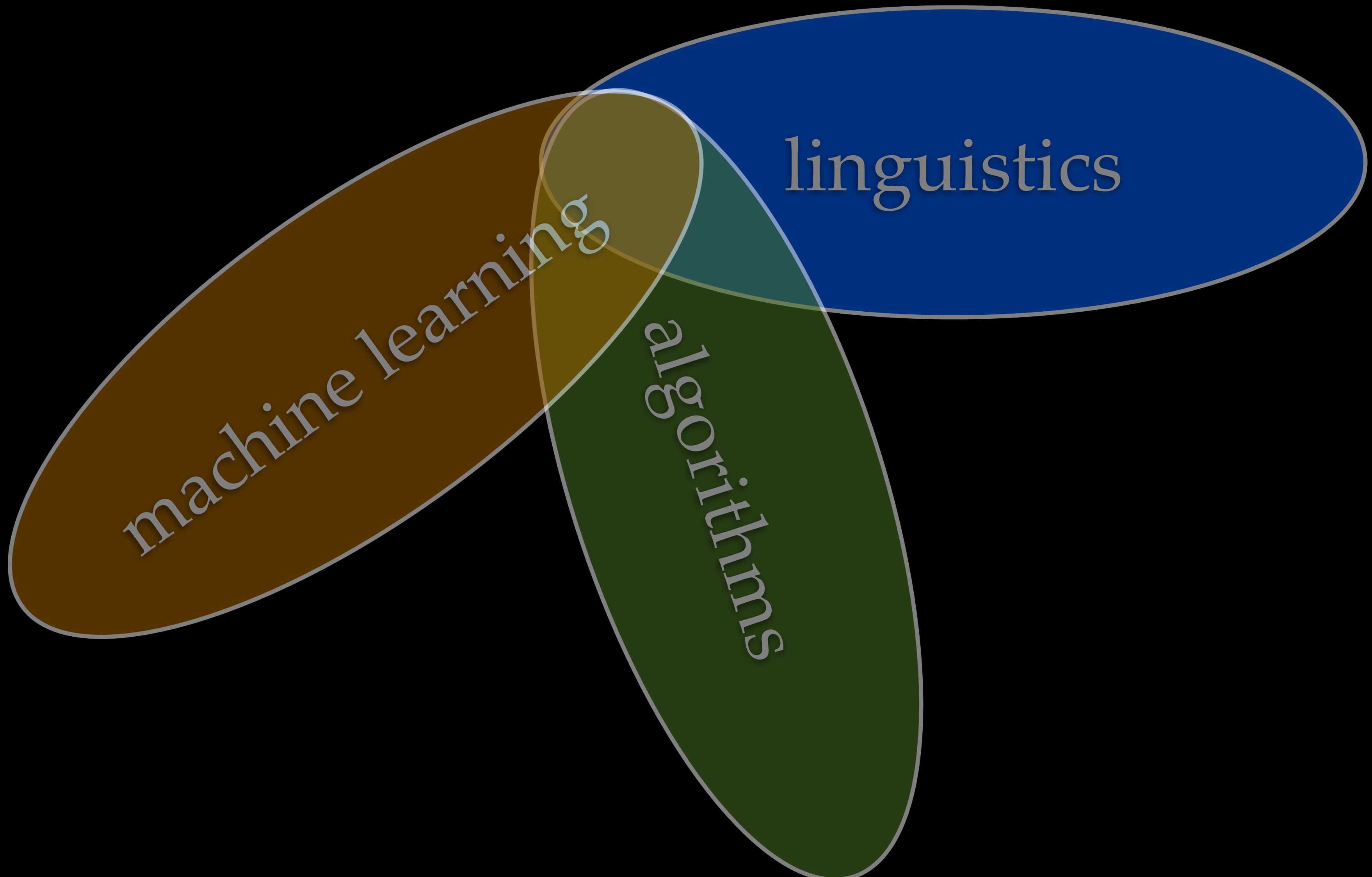
Statistical Machine Translation

Statistical Machine Translation

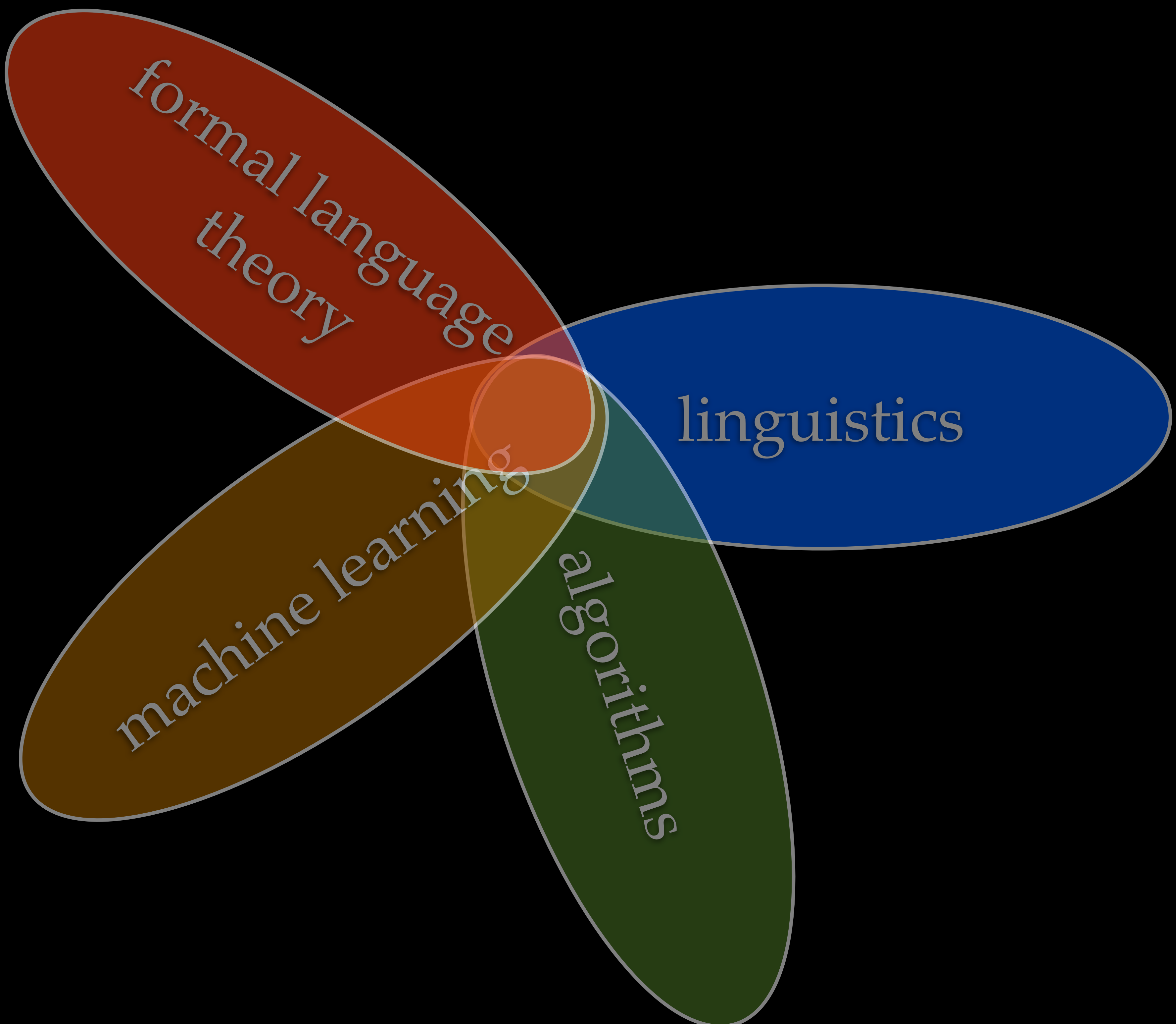
Statistical Machine Translation



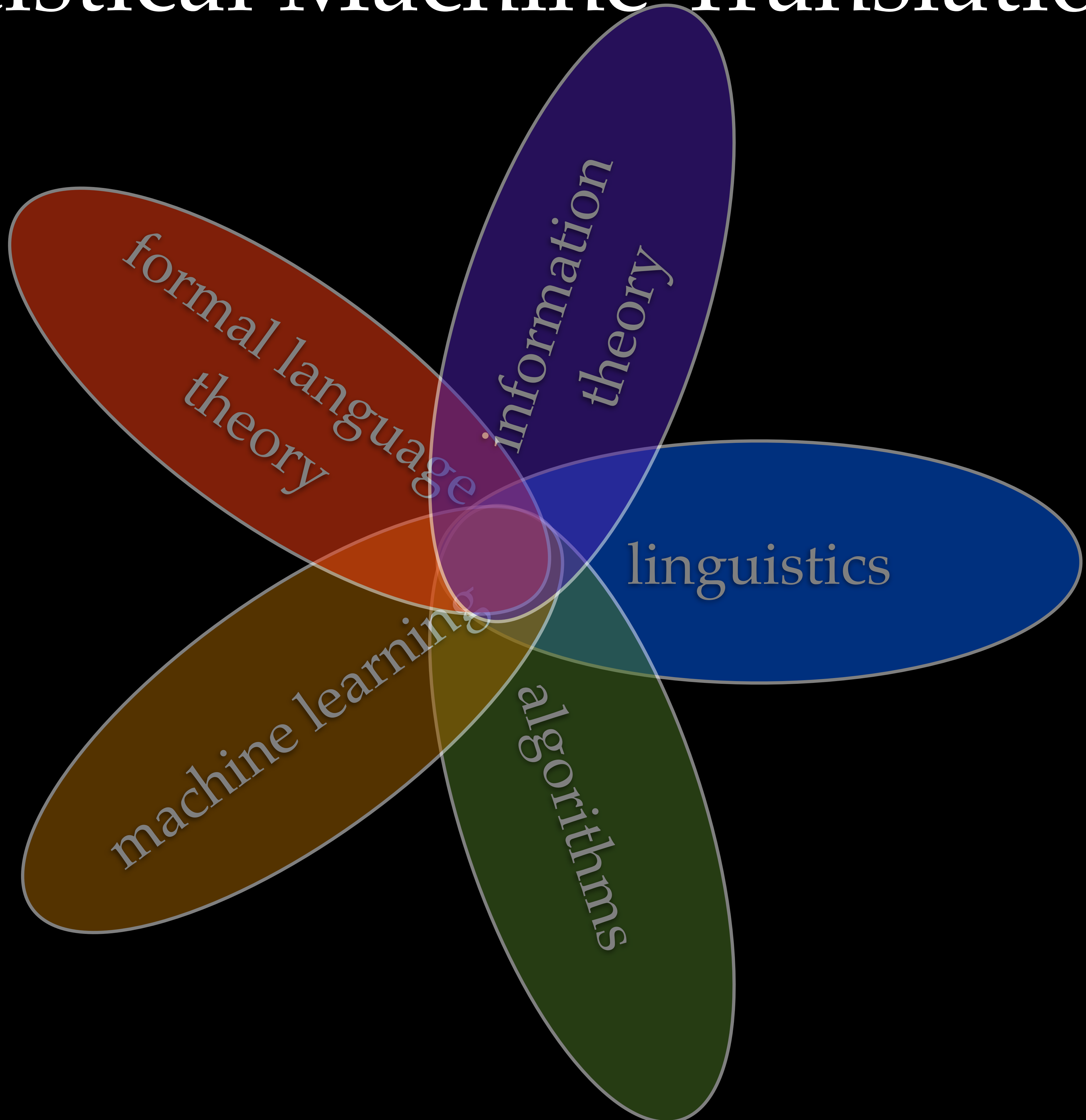
Statistical Machine Translation



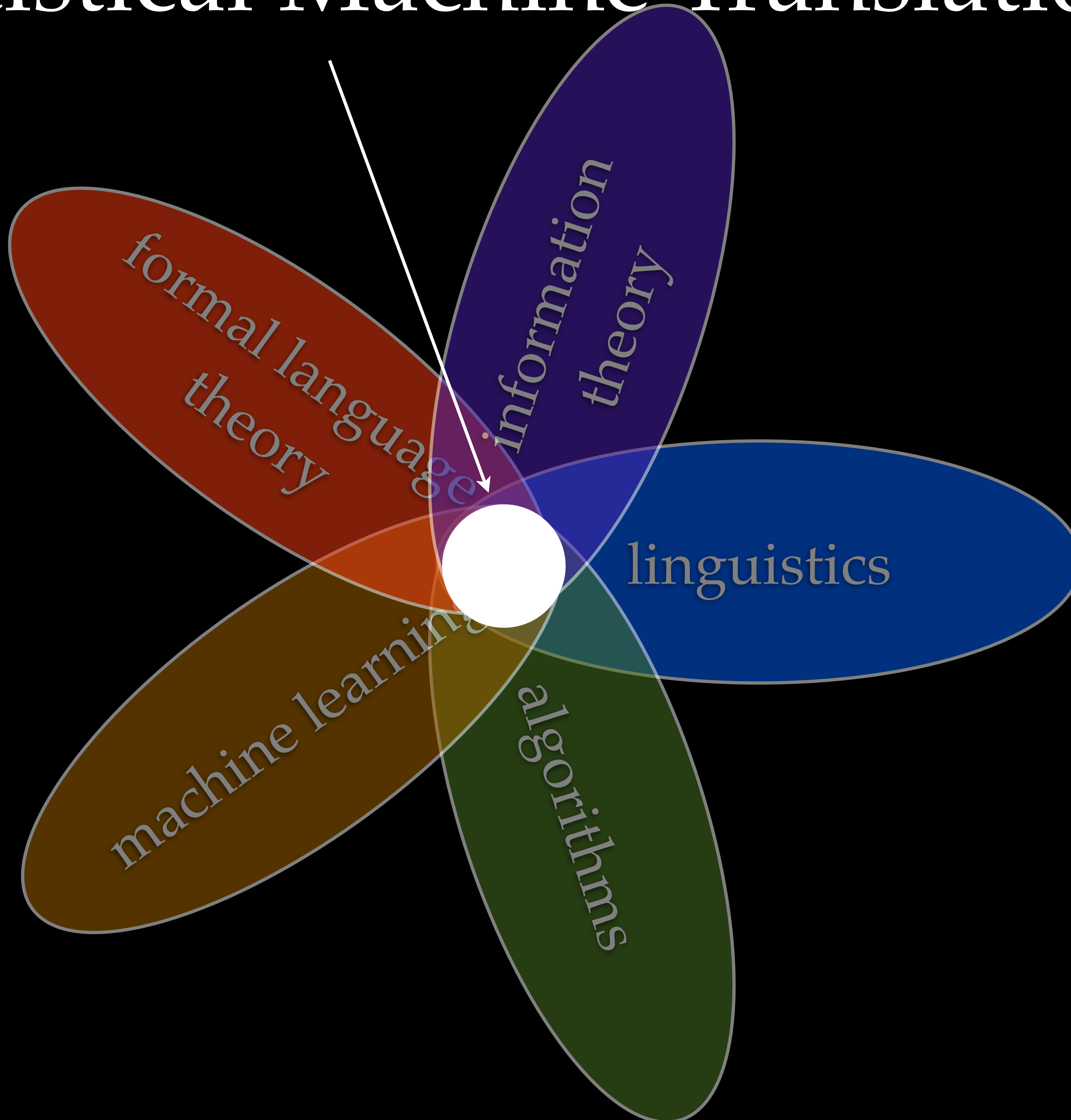
Statistical Machine Translation



Statistical Machine Translation



Statistical Machine Translation



OVERVIEW

OVERVIEW

- MT is a huge topic in NLP

OVERVIEW

- MT is a huge topic in NLP

- Today and Wednesday we’ll be talking about two parts:

OVERVIEW

- MT is a huge topic in NLP

- Today and Wednesday we'll be talking about two parts:

    - **Alignment** (with IBM Model 1): how do we automatically learn probabilistic translation dictionaries?

OVERVIEW

- MT is a huge topic in NLP

- Today and Wednesday we'll be talking about two parts:

  - **Alignment** (with IBM Model 1): how do we automatically learn probabilistic translation dictionaries?

  - **Decoding**: given a model, how can we efficiently search over the huge space of possible translations?

OVERVIEW

- MT is a huge topic in NLP

- Today and Wednesday we'll be talking about two parts:

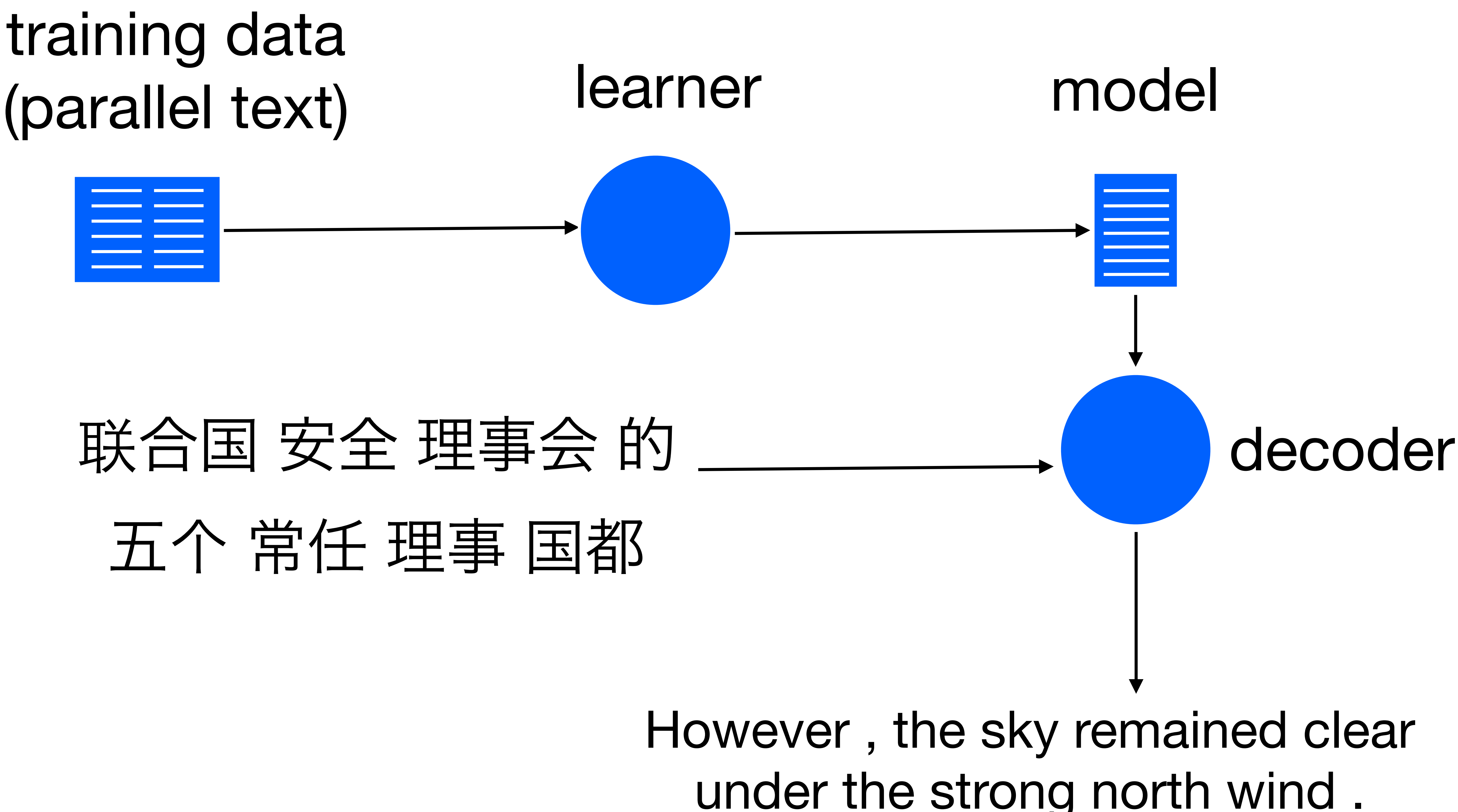
    - **Alignment** (with IBM Model 1): how do we automatically learn probabilistic translation dictionaries?

    - **Decoding**: given a model, how can we efficiently search over the huge space of possible translations?

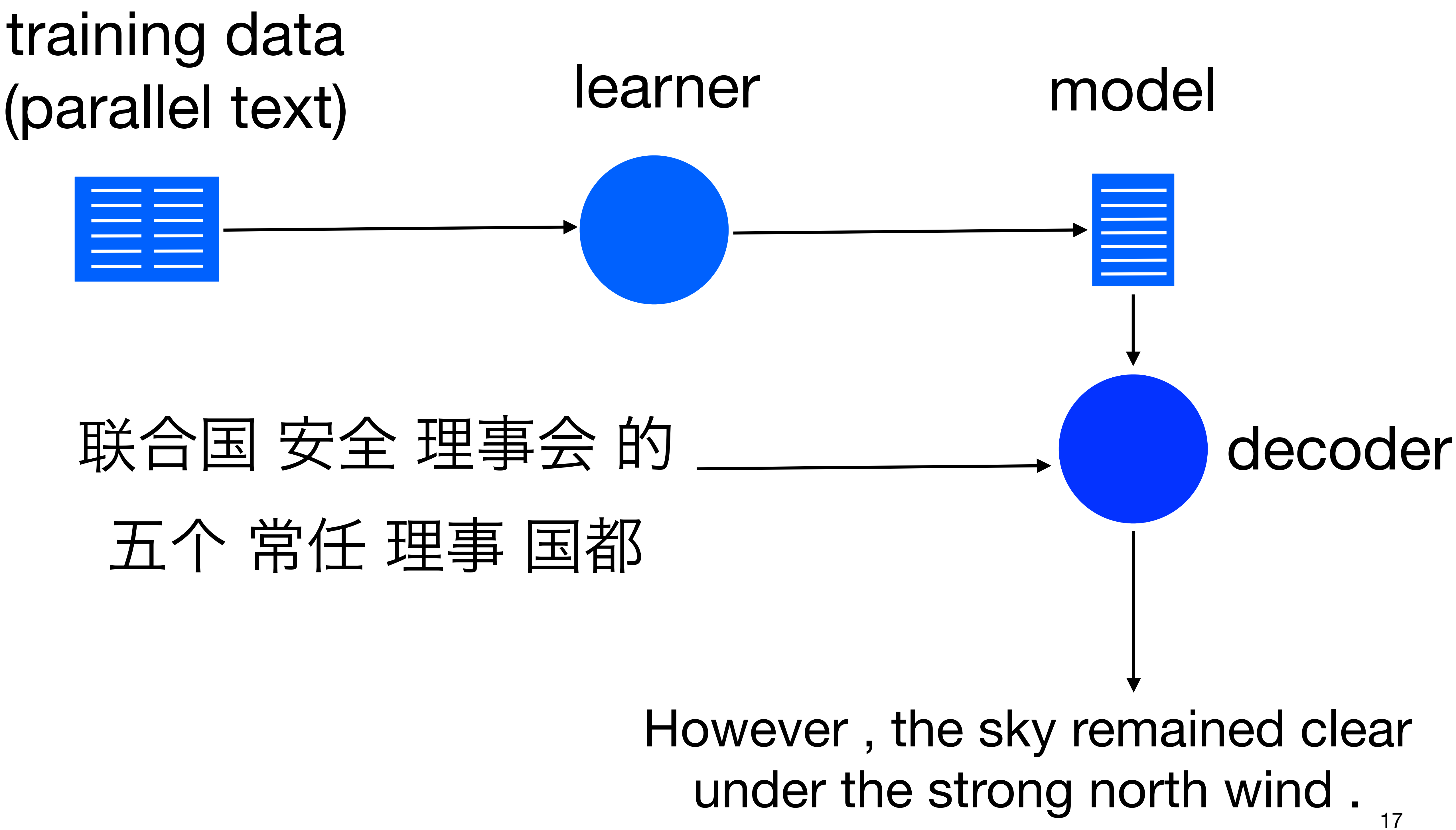
- Brings together many things you have learned about (n-gram language models, unsupervised learning, structured prediction, dynamic programming, EM, noisy channel models)

Alignment

MT PIPELINE OVERVIEW



MT Pipeline Overview



Intuitions

CLASSIC SOUPS

			Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

CLASSIC SOUPS

	No.	Item	Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

CLASSIC SOUPS

	No.	Item	Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲呑	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲呑湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

CLASSIC SOUPS

			Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

CLASSIC SOUPS

			Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

CLASSIC SOUPS

	No.		Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

CLASSIC SOUPS

	No.	Item	Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

CLASSIC SOUPS

			Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

CLASSIC SOUPS

			Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

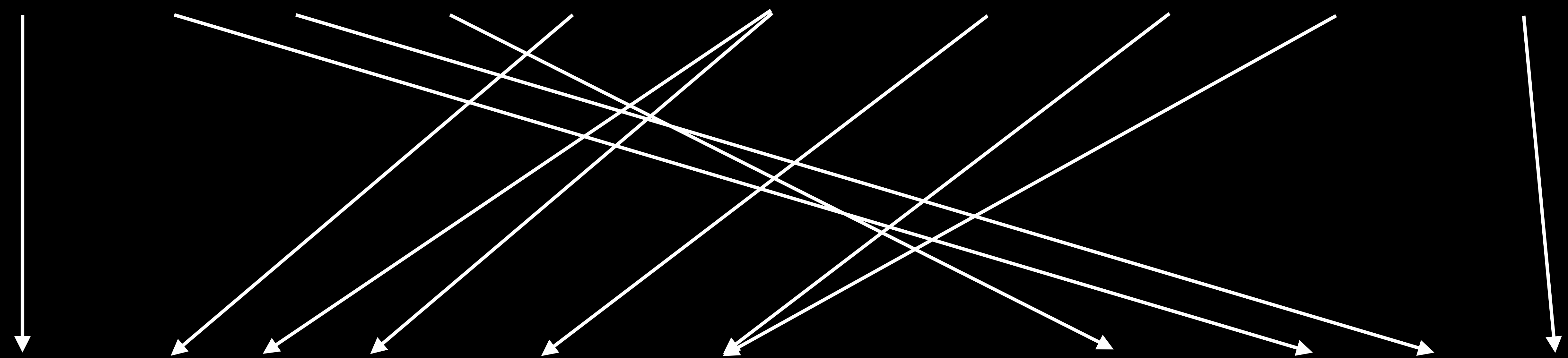
CLASSIC SOUPS

			Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

Translation Models

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

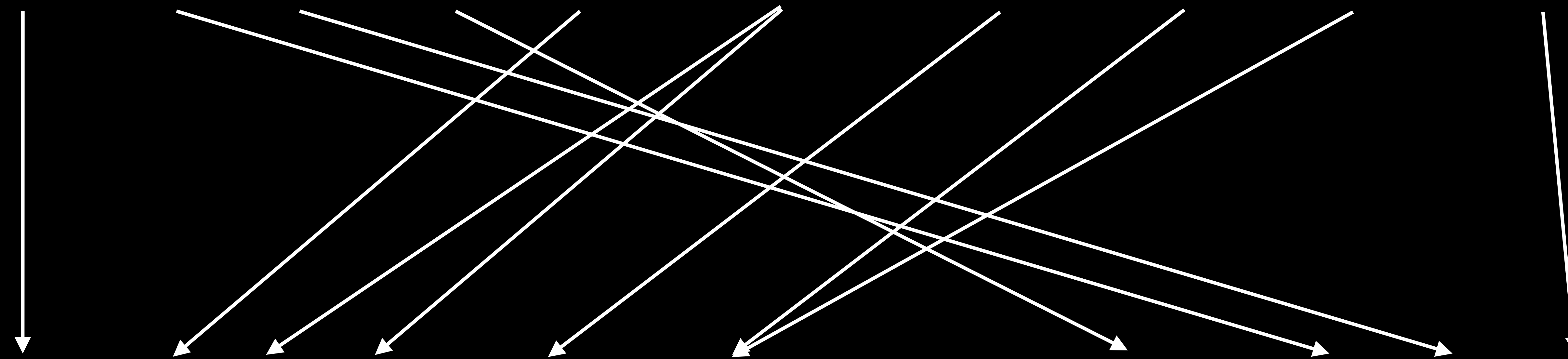


However , the sky remained clear under the strong north wind .

Translation Models

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



However , the sky remained clear under the strong north wind .

$$p(English|Chinese)?$$

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。 ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

$$p(English\ length | Chinese\ length)$$

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

$$p(English\ length|Chinese\ length)$$

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

$$p(Chinese\ word\ position)$$

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However

$$p(English\ word|Chinese\ word)$$

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。 ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

However ,

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However ,

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。 ε

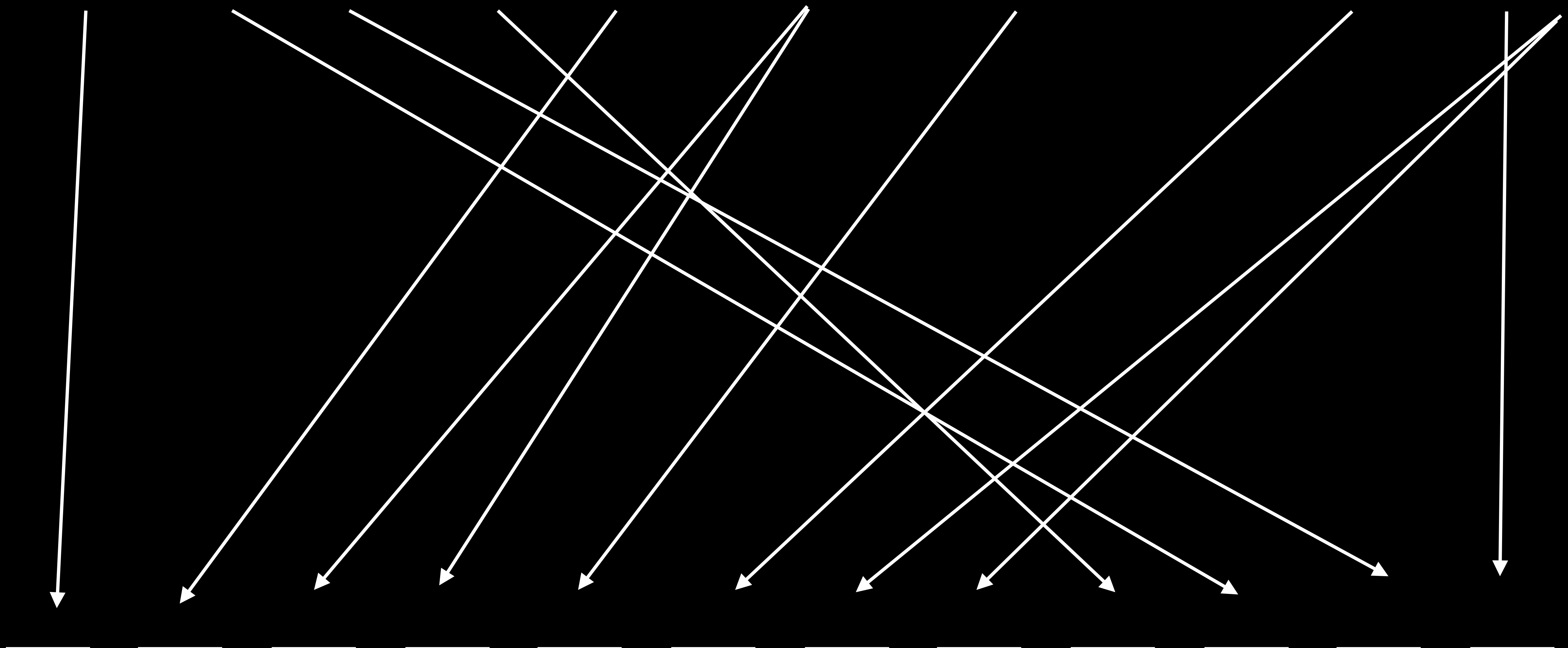
___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

However , the

IBM Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε



However , the sky remained clear under the strong north wind .

IBM Model 1

IBM Model 1

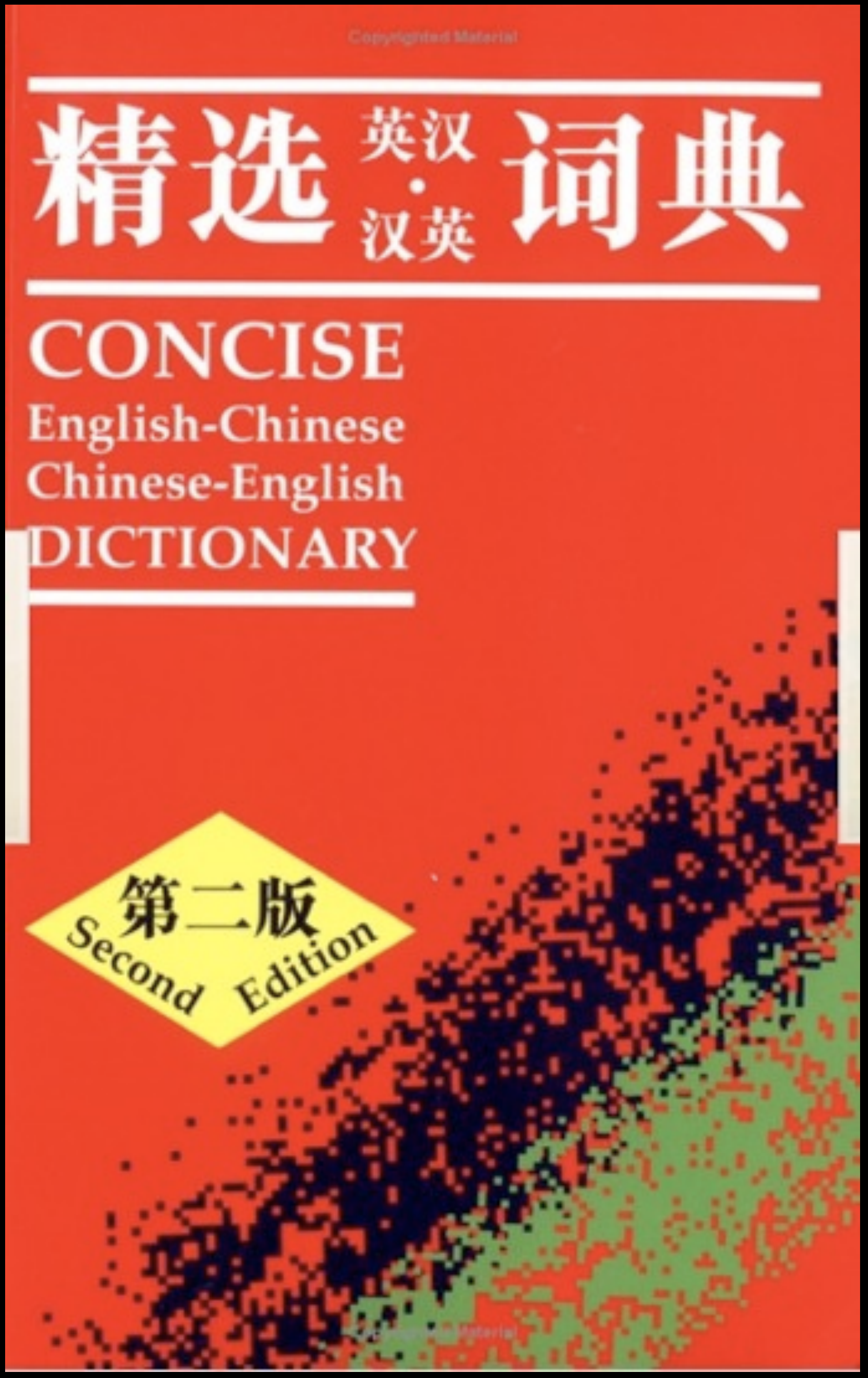
- Word translation probabilities.

IBM Model 1

- Word translation probabilities.
- No real ordering model.
  - This is left to the LM.

IBM Model 1

- Word translation probabilities.
- No real ordering model.
  - This is left to the LM.



IBM Model 1

- Word translation probabilities.
- No real ordering model.
  - This is left to the LM.

IBM Model 1

$$p(despite|\text{虽然})$$

$$p(however|\text{虽然})$$

$$p(although|\text{虽然})$$

$$p(northern|\text{北})$$

$$p(north|\text{北})$$

IBM Model 1

$p(despite|\text{虽然})$???

$p(however|\text{虽然})$???

$p(although|\text{虽然})$???

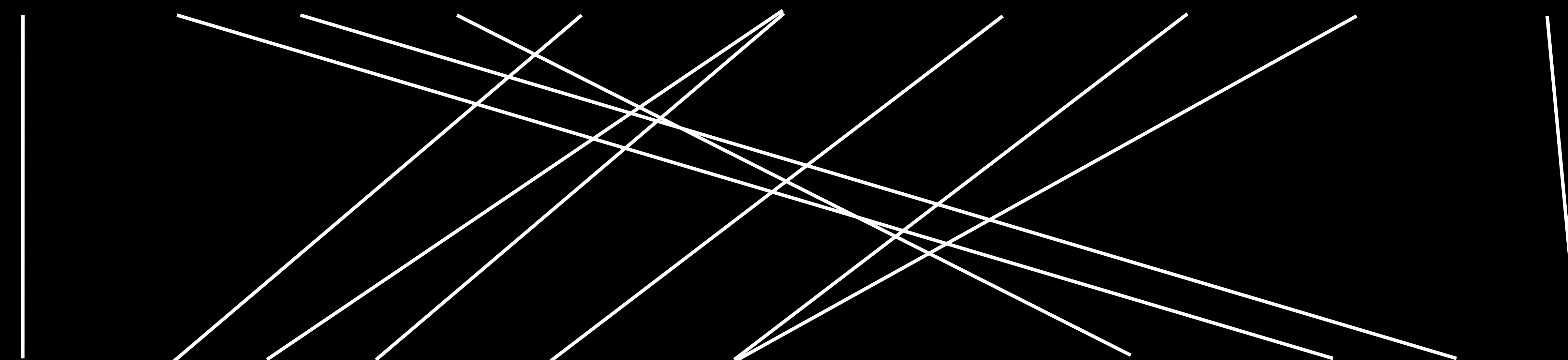
$p(northern|\text{北})$???

$p(north|\text{北})$???

Translation Models

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。



However , the sky remained clear under the strong north wind .

$$p(however \mid \text{虽然}) = \frac{\text{\# of times 虽然 aligns to However}}{\text{\# of times 虽然 occurs}}$$

Translation Models

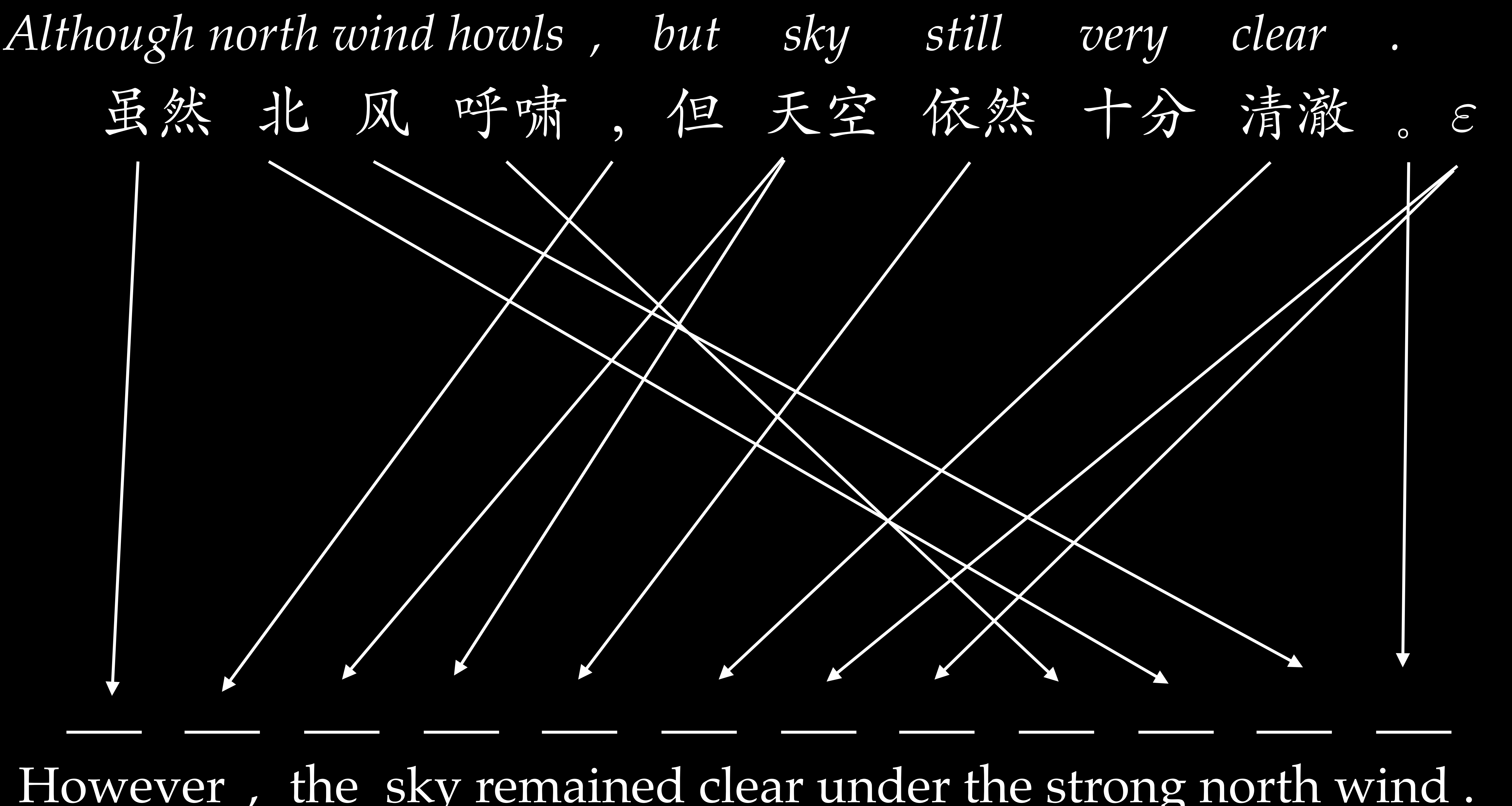
Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

$$p(however|\text{虽然}) = \frac{\text{\# of times 虽然 aligns to However}}{\text{\# of times 虽然 occurs}}$$

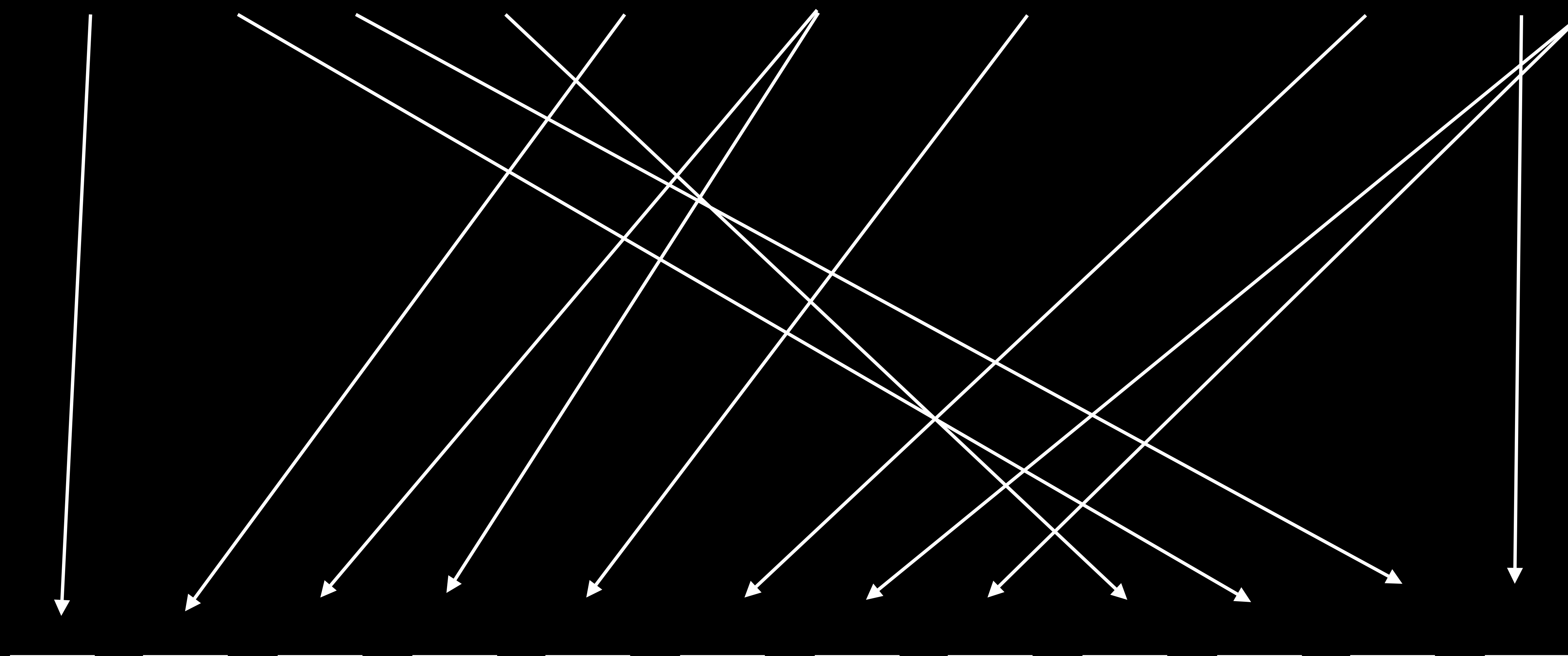
Likelihood Estimation for Model 1



Likelihood Estimation for Model 1

Although north wind howls , but sky still very clear .

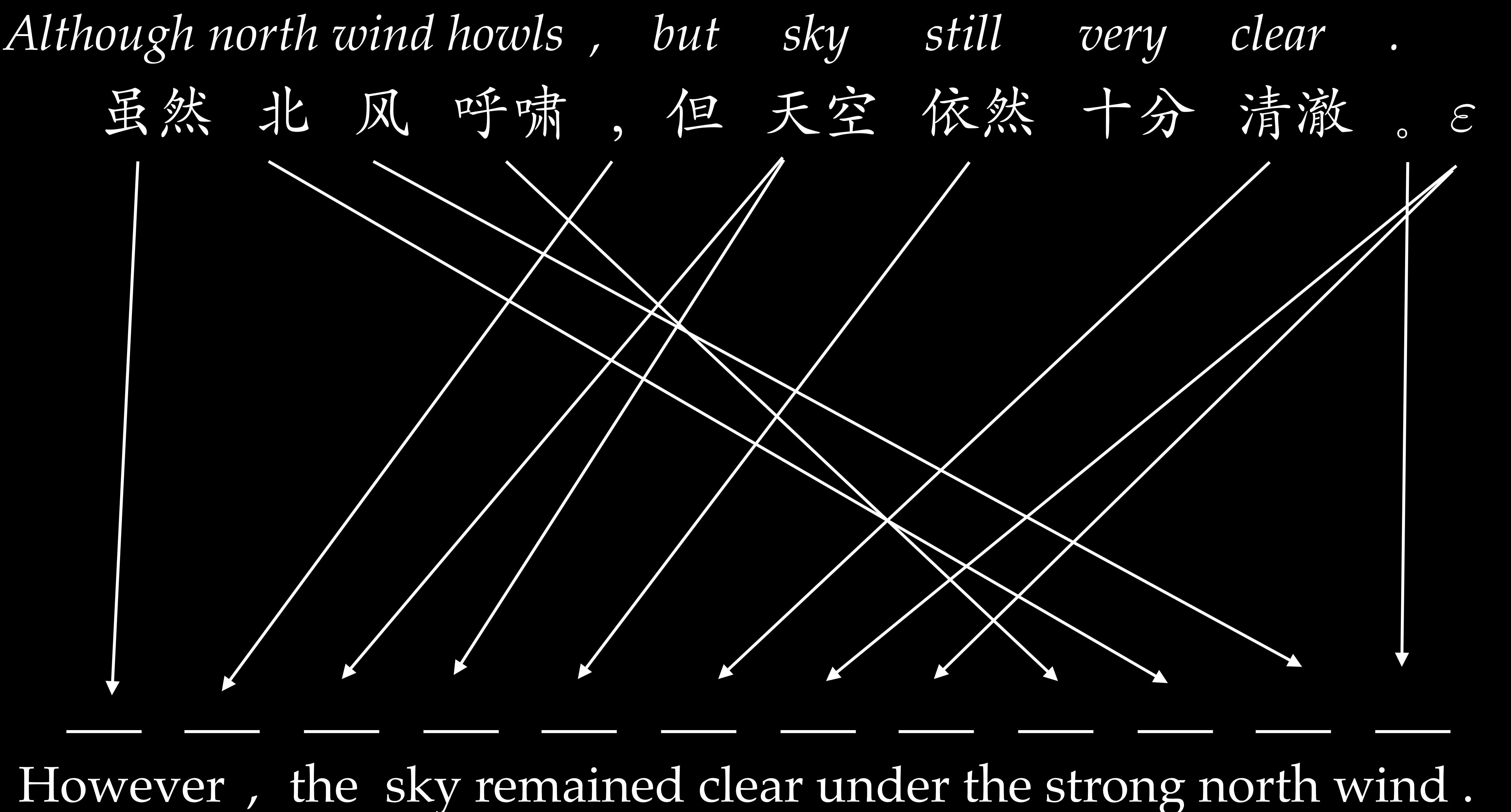
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε



However , the sky remained clear under the strong north wind .

$p(English\ length|Chinese\ length)$ observed

Likelihood Estimation for Model 1



$p(Chinese\ word\ position)$ uniform, no need to estimate

Likelihood Estimation for Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

However , the sky remained clear under the strong north wind .

$p(English\ word|Chinese\ word)$ unobserved!

Likelihood Estimation for Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

the missing alignment is a ***latent variable***

However , the sky remained clear under the strong north wind .

$p(English\ word|Chinese\ word)$ unobserved!

Likelihood Estimation for Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

Parameters and alignments are both unknown.

However , the sky remained clear under the strong north wind .

$p(English\ word|Chinese\ word)$ unobserved!

Likelihood Estimation for Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

Parameters and alignments are both unknown.

If we knew the alignments, we could calculate the values of the parameters.

However , the sky remained clear under the strong north wind .

$p(English\ word|Chinese\ word)$ unobserved!

Likelihood Estimation for Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

Parameters and alignments are both unknown.

If we knew the alignments, we could calculate the values of the parameters.

If we knew the parameters, we could calculate the likelihood of the data.

However , the sky remained clear under the strong north wind .

$p(English\ word|Chinese\ word)$ unobserved!

Likelihood Estimation for Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

Parameters and alignments are both unknown.

If we knew the alignments, we could calculate the values of the parameters.

If we knew the parameters, we could calculate the likelihood of the data.

However , the sky remained clear under the strong north wind .

$p(English\ word|Chinese\ word)$ unobserved!

The Plan: Bootstrapping

- Arbitrarily select a set of parameters (say, uniform).
- Calculate ***expected counts*** of the unseen events.
- Choose new parameters to maximize likelihood, using expected counts as proxy for observed counts.
- Iterate.
- Guarantee: likelihood will be monotonically nondecreasing.

The Plan: Bootstrapping

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

However , the sky remained clear under the strong north wind .

The Plan: Bootstrapping

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。ε

if we had observed the alignment, this line would either be here (count 1) or it wouldn't (count 0).

However , the sky remained clear under the strong north wind .

The Plan: Bootstrapping

Although north wind howls , but sky still very clear .

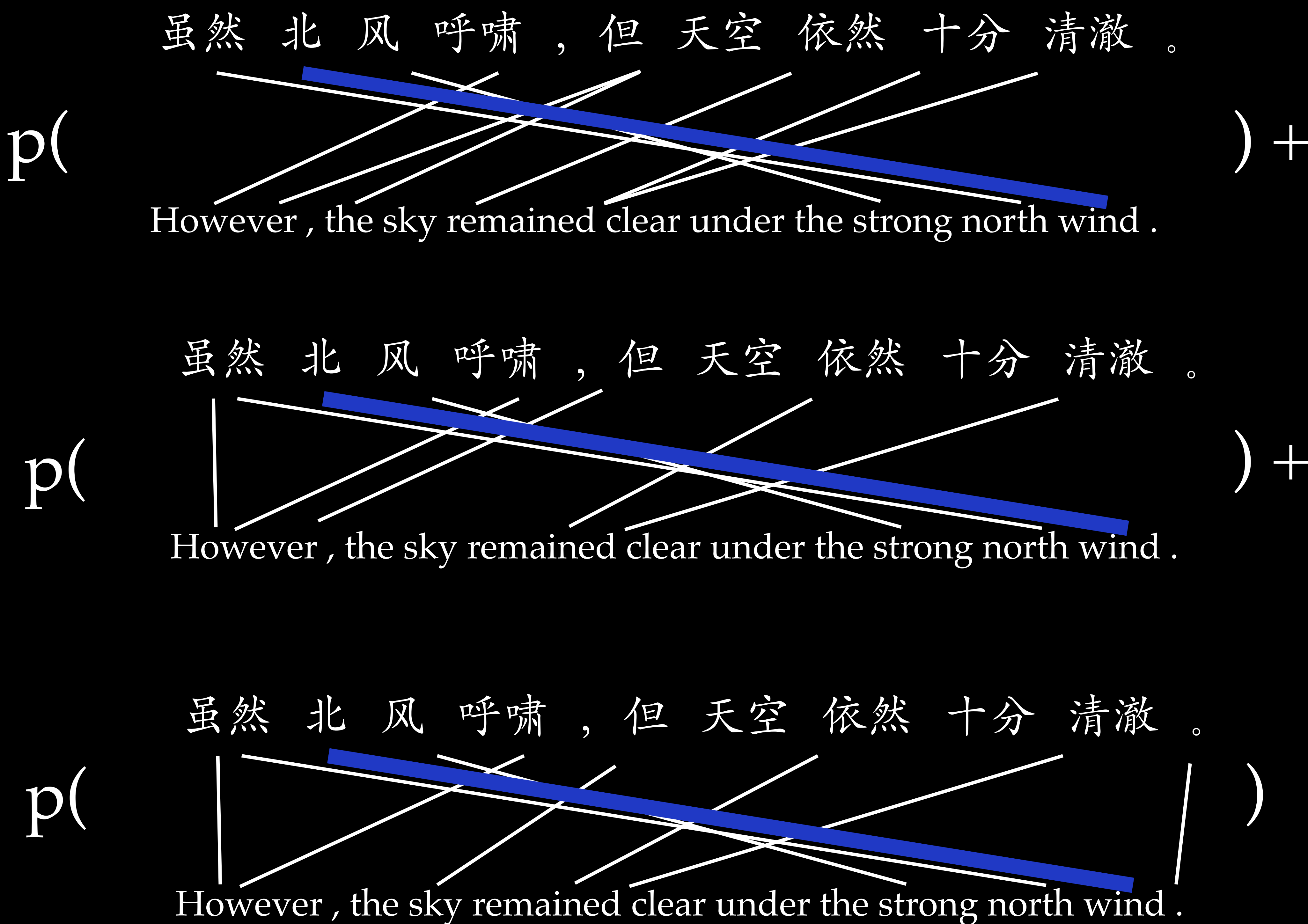
虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。 ε

if we had observed the alignment, this line would either be here (count 1) or it wouldn't (count 0).

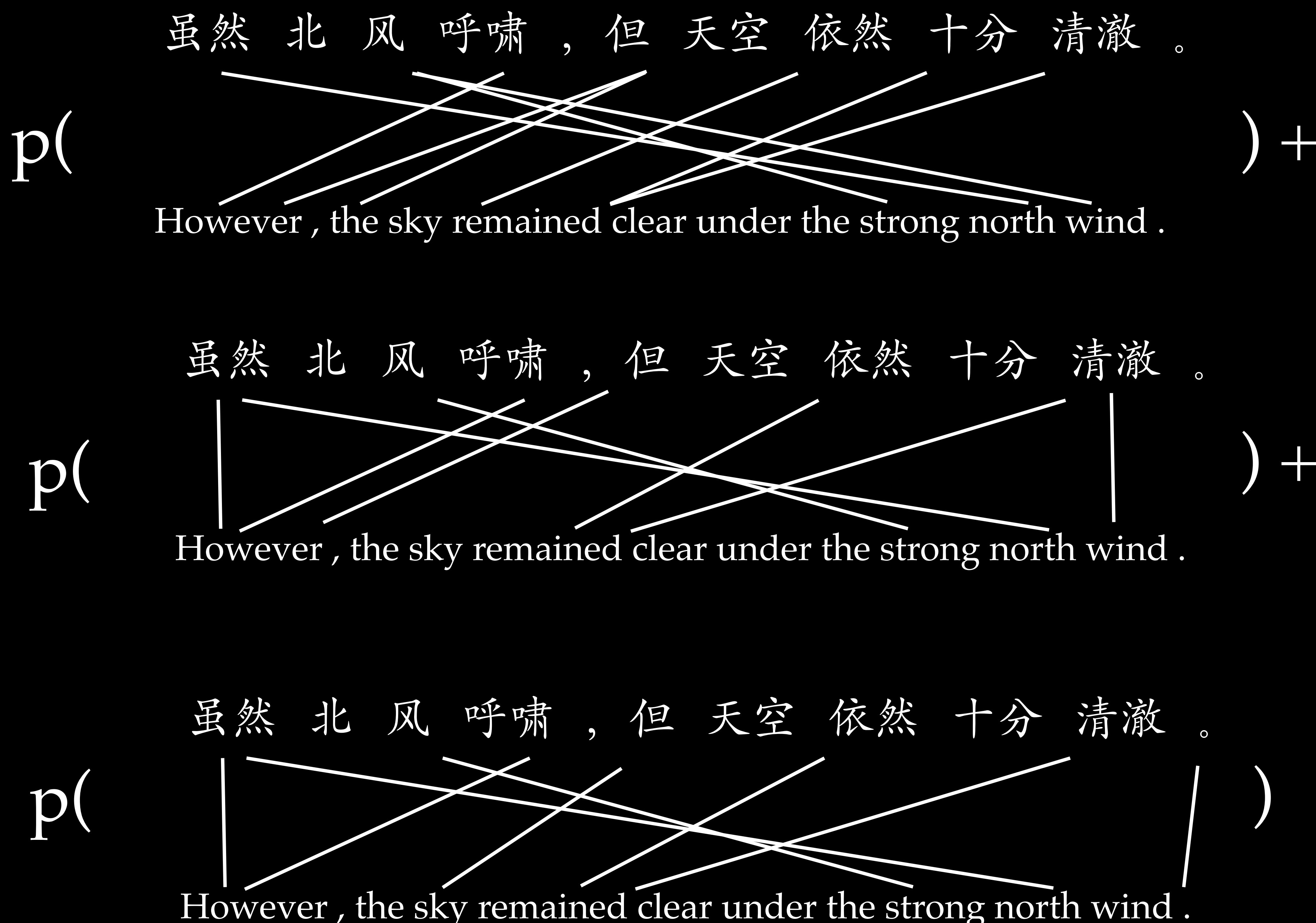
since we didn't observe the alignment, we calculate the probability that it's there.

However , the sky remained clear under the strong north wind .

Marginalize: sum all alignments containing the link



Divide by sum of all *possible* alignments



Divide by sum of all *possible* alignments

p(虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .) +

p(虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .) +

p(虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .)

Is this hard? How many alignments are there?

Expectation Maximization

probability of an alignment.

$$p(F, A|E) = p(I|J) \prod_{a_i} p(a_i = j) p(f_i|e_j)$$

Expectation Maximization

probability of an alignment.

$$p(F, A|E) = p(I|J) \prod_{a_i} p(a_i = j) p(f_i|e_j)$$

observed

uniform

Expectation Maximization

probability of an alignment.

factors across words.

$$p(F, A|E) = p(I|J) \prod_{a_i} p(a_i = j) \boxed{p(f_i|e_j)}$$

observed

uniform

Expectation Maximization

marginal probability of
alignments containing link

$$\sum_{a \in A : 北 \leftrightarrow north} p(north | 北) \cdot p(rest\ of\ a)$$

Expectation Maximization

marginal probability of
alignments containing link

$$p(north|北) \sum_{a \in A: 北 \leftrightarrow north} p(rest\ of\ a)$$

Expectation Maximization

marginal probability of
alignments containing link

$$\frac{p(north|北)\displaystyle\sum_{a\in A:北\leftrightarrow north} p(rest\ of\ a)}{\displaystyle\sum_{c\in Chinese\ words} p(north|c)\sum_{a\in A:\,c\,\leftrightarrow north} p(rest\ of\ a)}$$

marginal probability of all
alignments

Expectation Maximization

marginal probability of
alignments containing link

$$\frac{p(north|北)\displaystyle\sum_{a\in A:北\leftrightarrow north} p(rest\ of\ a)}{\displaystyle\sum_{c\in Chinese\ words} p(north|c)\sum_{a\in A:\,c\leftrightarrow north} p(rest\ of\ a)}$$

identical!

marginal probability of all
alignments

Expectation Maximization

$$\frac{p(north|\text{北})}{\sum_{c \in Chinese\ words} p(north|c)}$$

Expectation Maximization

marginal probability (expected count) of an alignment containing the link

$$\frac{p(north|\text{北})}{\sum_{c \in Chinese\ words} p(north|c)}$$

Expectation Maximization

marginal probability (expected count) of an alignment containing the link

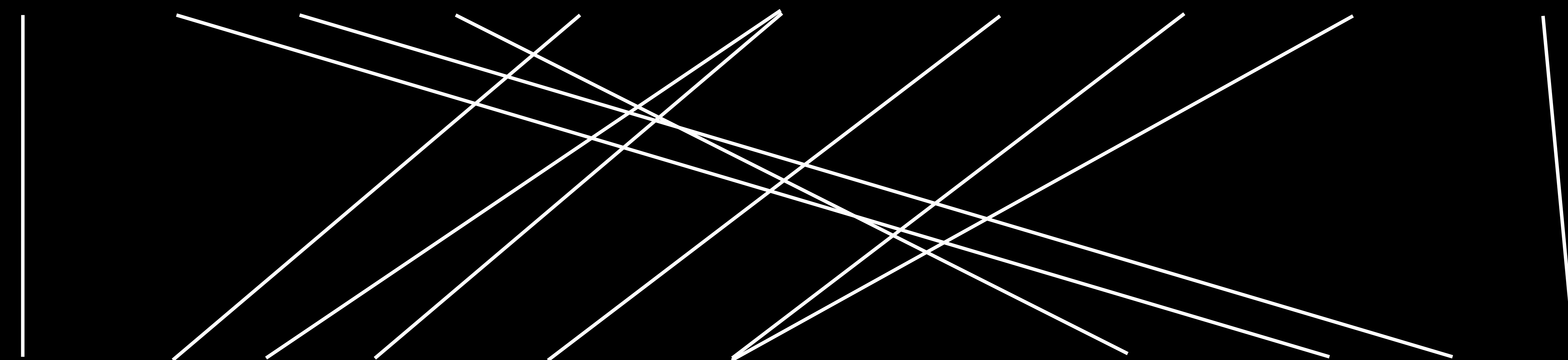
$$\frac{p(north|\text{北})}{\sum_{c \in Chinese\ words} p(north|c)}$$

For each sentence, use this quantity instead of 0 or 1

Supervised Case

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。



However , the sky remained clear under the strong north wind .

$$p(however|\text{虽然}) = \frac{\text{\# of times 虽然 aligns to However}}{\text{\# of times 虽然 occurs}}$$

Unsupervised Case

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

$$p(however|\text{虽然}) = \frac{\textit{Expected}\ \#\ \text{of times 虽然 aligns to However}}{\#\ \text{of times 虽然 occurs}}$$

Expectation Maximization

Why does this even work?

$$\frac{p(north|\text{北})}{\sum_{c \in Chinese\ words} p(north|c)}$$

Expectation Maximization

Observation 1: We are still solving a maximum likelihood estimation problem.

Expectation Maximization

Observation 1: We are still solving a maximum likelihood estimation problem.

$$p(Chinese|English) = \sum_{alignments} p(Chinese, alignment|English)$$

Expectation Maximization

Observation 1: We are still solving a maximum likelihood estimation problem.

$$p(Chinese|English) = \sum_{alignments} p(Chinese, alignment|English)$$

MLE: choose parameters that maximize this expression.

Expectation Maximization

Observation 1: We are still solving a maximum likelihood estimation problem.

$$p(Chinese|English) = \sum_{alignments} p(Chinese, alignment|English)$$

MLE: choose parameters that maximize this expression.

Minor problem: there is no analytic solution.

Expectation Maximization

Observation 1: We are still solving a maximum likelihood estimation problem.

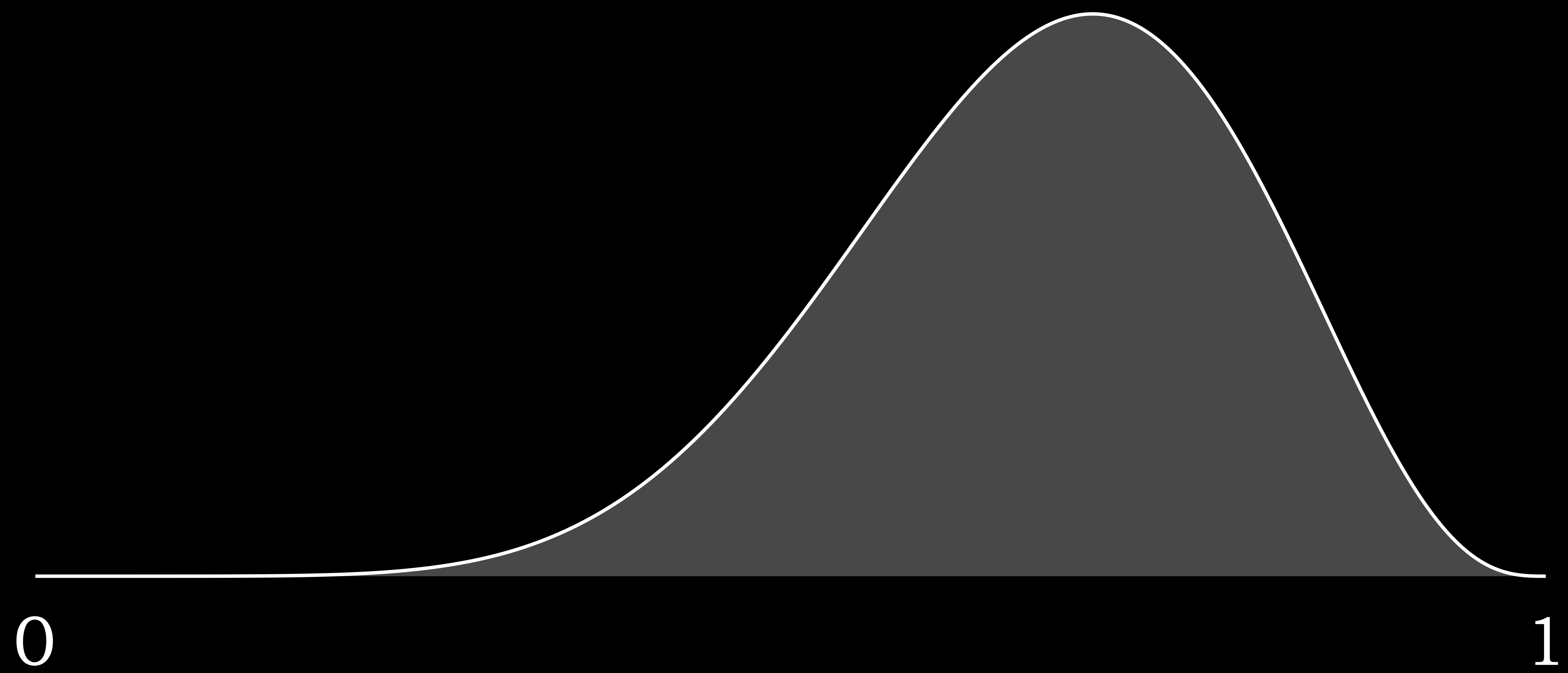
$$p(Chinese|English) = \sum_{alignments} p(Chinese, alignment|English)$$

MLE: choose parameters that maximize this expression.

Minor problem: there is no analytic solution.

Remember: likelihood is monotonically non-decreasing!

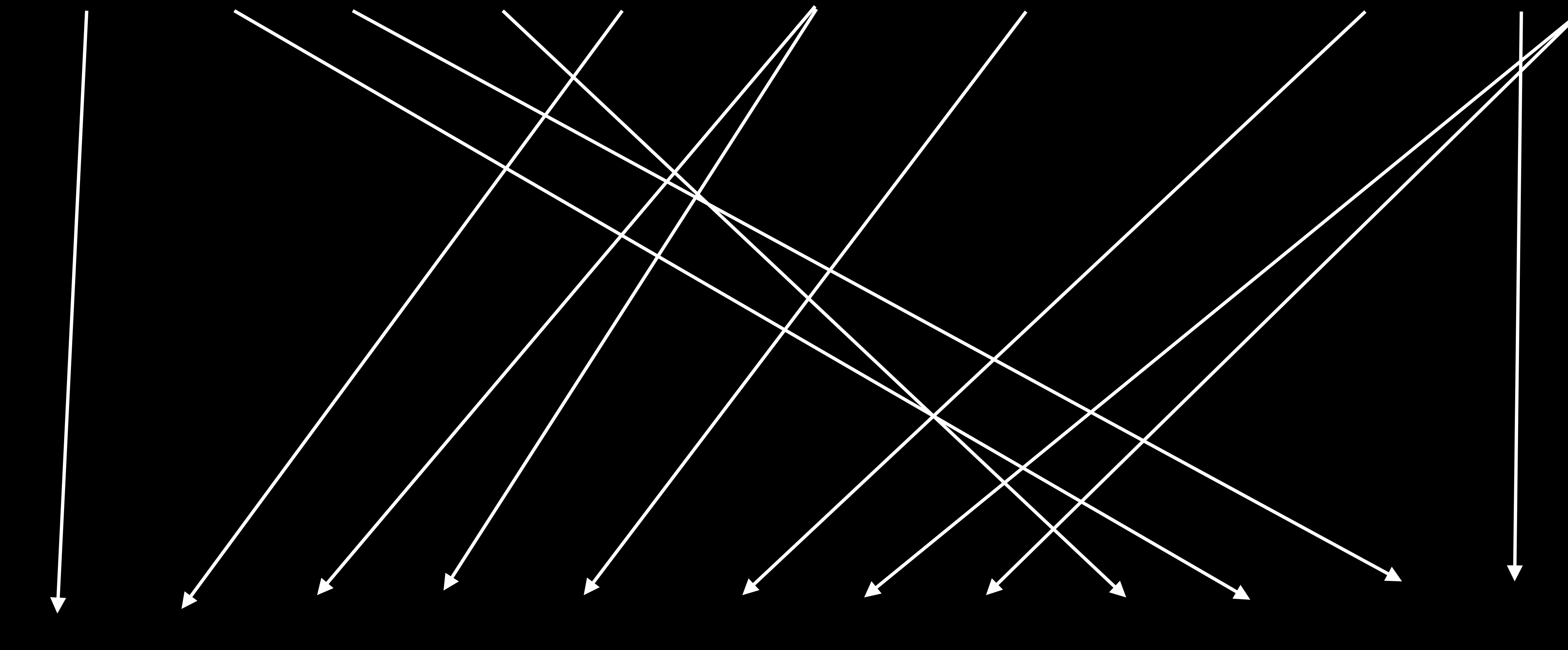
... and, likelihood is ***convex*** for this model:



Likelihood Estimation for Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε



However , the sky remained clear under the strong north wind .

What are some things this model doesn't account for?

NOT DISCUSSED

NOT DISCUSSED

- What's wrong with Model 1?

NOT DISCUSSED

- What’s wrong with Model 1?

- Higher IBM Models

NOT DISCUSSED

- What's wrong with Model 1?

- Higher IBM Models

    - Model 2: includes a term for the probability of alignment positions (spot i linked with spot j)

NOT DISCUSSED

- What's wrong with Model 1?
- Higher IBM Models
  - Model 2: includes a term for the probability of alignment positions (spot i linked with spot j)
  - Model 3: models **fertility**, i.e., how many foreign words were produced by each English word (but independently!)

NOT DISCUSSED

- What's wrong with Model 1?

- Higher IBM Models

    - Model 2: includes a term for the probability of alignment positions (spot i linked with spot j)

    - Model 3: models **fertility**, i.e., how many foreign words were produced by each English word (but independently!)

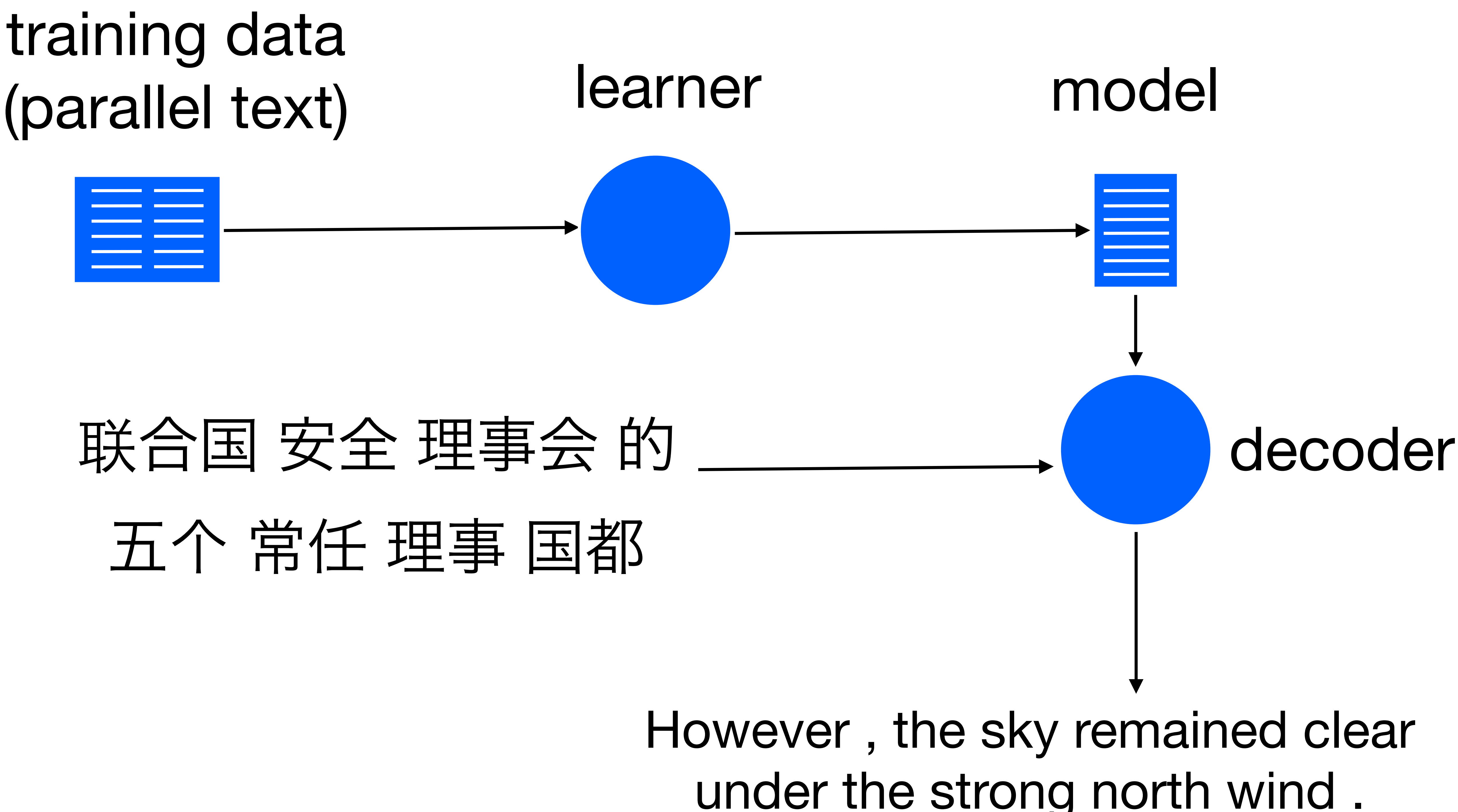
- For decoding, these models are superseded by more general phrase-based models (but are still used for alignment)

Summary

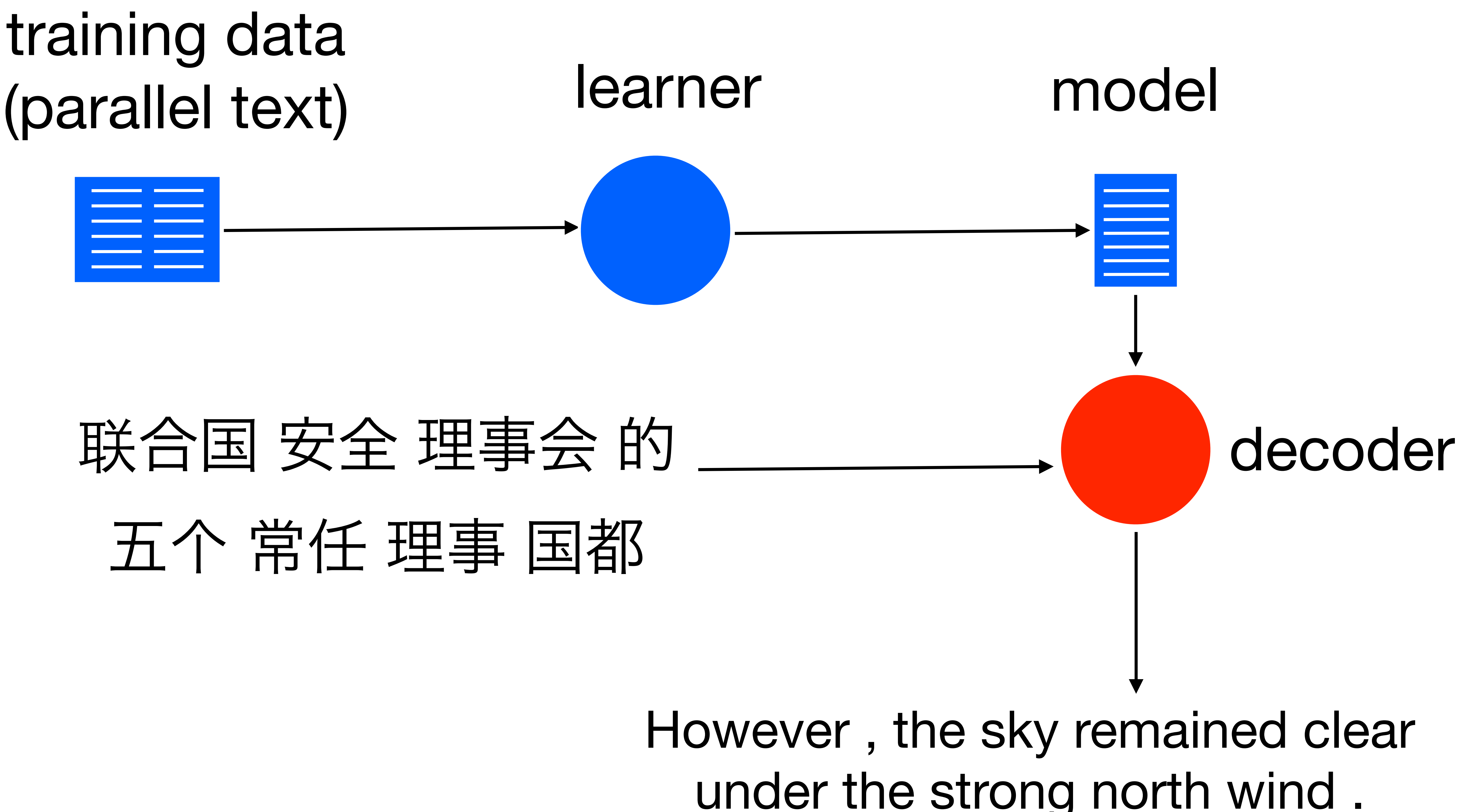
- We can formulate learning as an optimization problem: choose parameters that optimize some function, such as likelihood.
- Supervised: maximum likelihood.
  - Beware of overfitting.
- Unsupervised: expectation maximization.
- Many, many, many other algorithms.

Decoding

THE STORY SO FAR...



THE STORY SO FAR...



DECODING

DECODING

- The process of producing a translation of a sentence

DECODING

- The process of producing a translation of a sentence
  - *input*: foreign-language sentence and a model

DECODING

- The process of producing a translation of a sentence
  - *input*: foreign-language sentence and a model
  - *task:* find the model's favorite translation

DECODING

- The process of producing a translation of a sentence
    - *input*: foreign-language sentence and a model
    - *task:* find the model's favorite translation
- Two main problems:

DECODING

- The process of producing a translation of a sentence
  - *input*: foreign-language sentence and a model
  - *task:* find the model's favorite translation
- Two main problems:
  - **modeling** How do we score translations?

DECODING

- The process of producing a translation of a sentence
  - *input*: foreign-language sentence and a model
  - *task:* find the model's favorite translation
- Two main problems:
  - **modeling** How do we score translations?
  - **search** How do we find the model's preferred translation?

PROBLEM 1: MODELING

- The process of producing a translation of a sentence

PROBLEM 1: MODELING

- The process of producing a translation of a sentence
- Two main problems:

PROBLEM 1: MODELING

- The process of producing a translation of a sentence

- Two main problems:

  - **modeling** – given a pair of sentences, how do we assign a probability to them?

PROBLEM 1: MODELING

- The process of producing a translation of a sentence

- Two main problems:

  - **modeling** – given a pair of sentences, how do we assign a probability to them?

$$P_{(C \rightarrow E)}\left(\begin{array}{c}\text{他们还缺乏国际比赛的经验.} \\ \text{They still lack experience in international competitions}\end{array}\right) = \text{high}$$

PROBLEM 1: MODELING

- The process of producing a translation of a sentence

- Two main problems:

  - **modeling** – given a pair of sentences, how do we assign a probability to them?

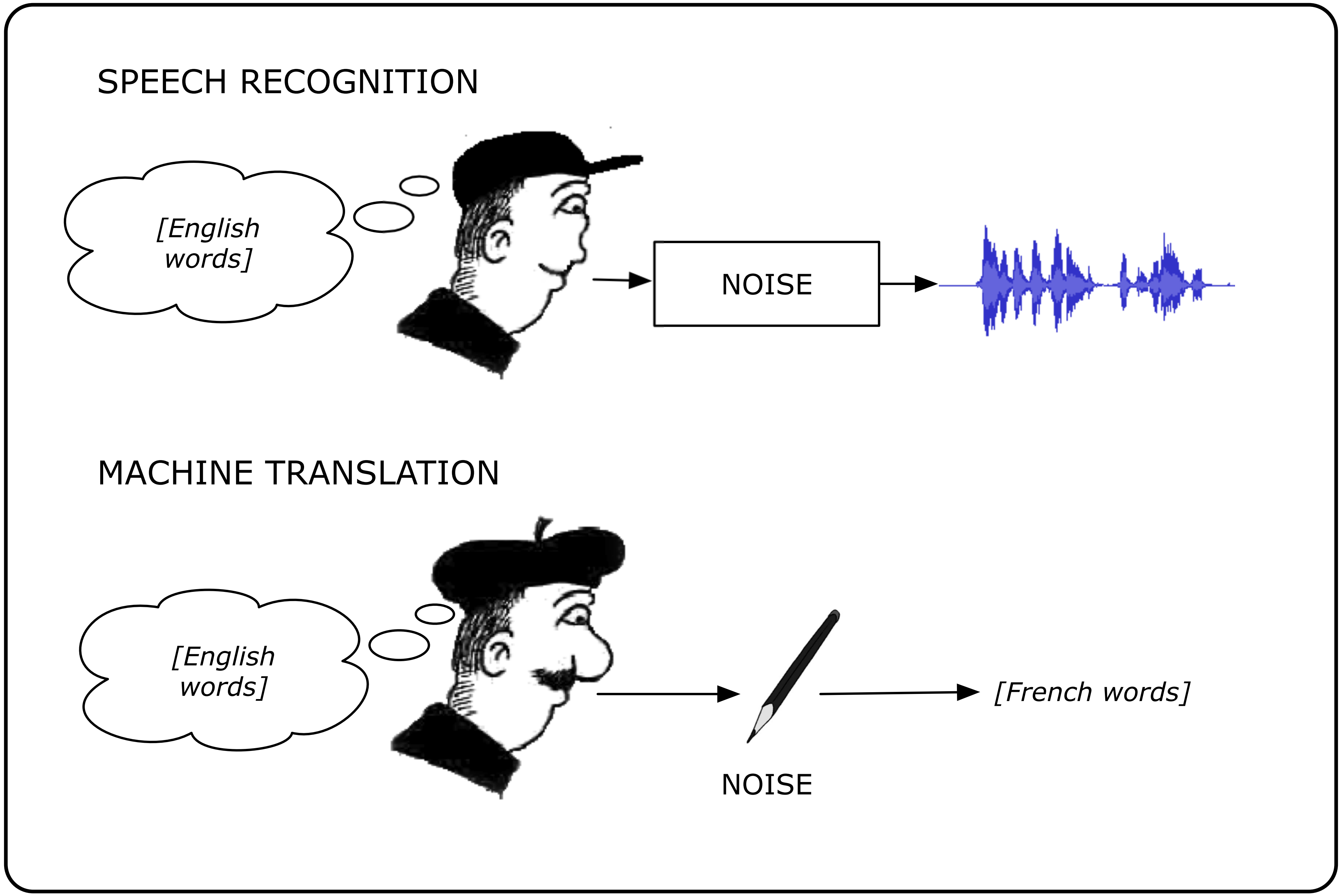
$P_{(C \rightarrow E)}\Big($ 他们还缺乏国际比赛的经验. This is not a good translation of the above sentence. $\Big)$ = low

MODEL

- Noisy Channel model $P(e \mid f) \quad \propto \quad P(f \mid e)P(e)$

MODEL

- Noisy Channel model $P(e \mid f) \propto P(f \mid e)P(e)$



Model Transforms

- Add weights

$$
\begin{aligned}
P(e \mid f) &\propto P(f \mid e)P(e) \\
&\boxed{\propto P(f \mid e)^{\lambda_1} P(e)^{\lambda_2}}
\end{aligned}
$$

WEIGHTS

WEIGHTS

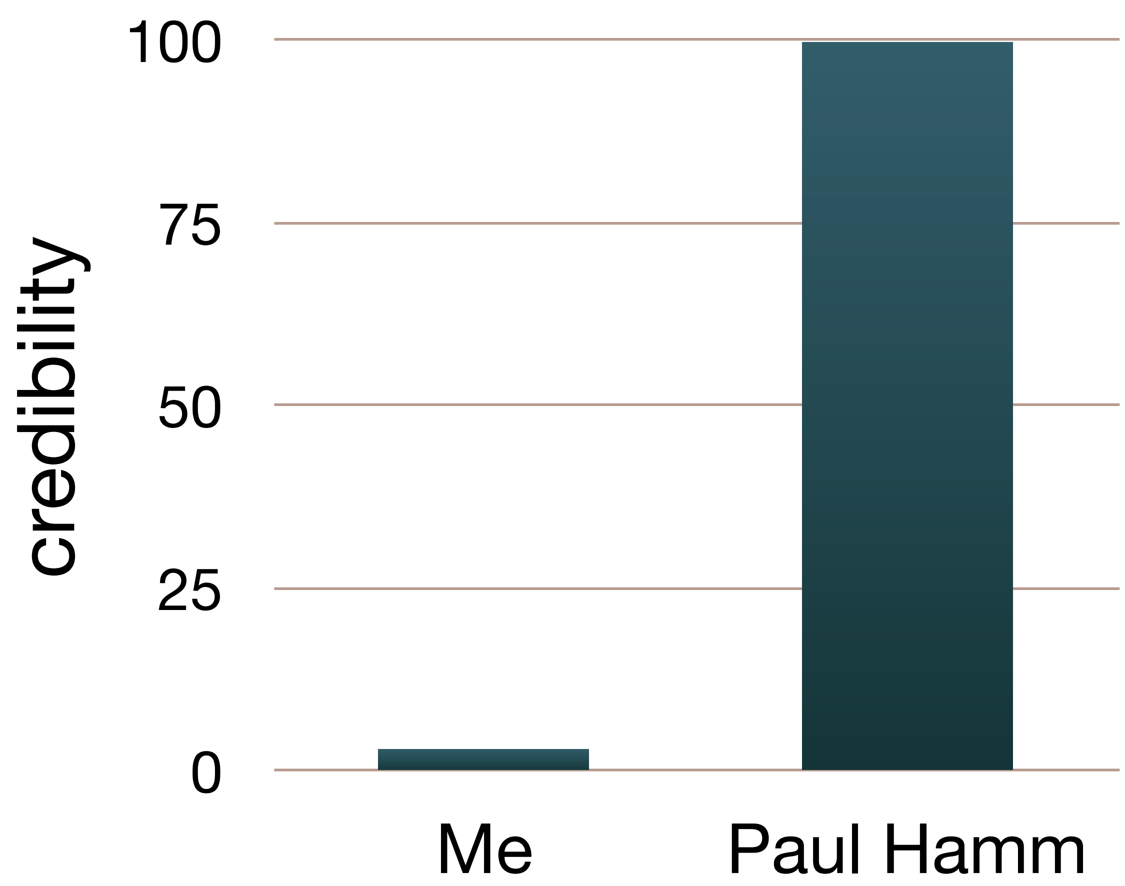
- Why?

WEIGHTS

- Why?
    - Just like in real life, where we trust people's claims differently, we will want to learn how to trust different models

Weights

- Why?
  - Just like in real life, where we trust people's claims differently, we will want to learn how to trust different models

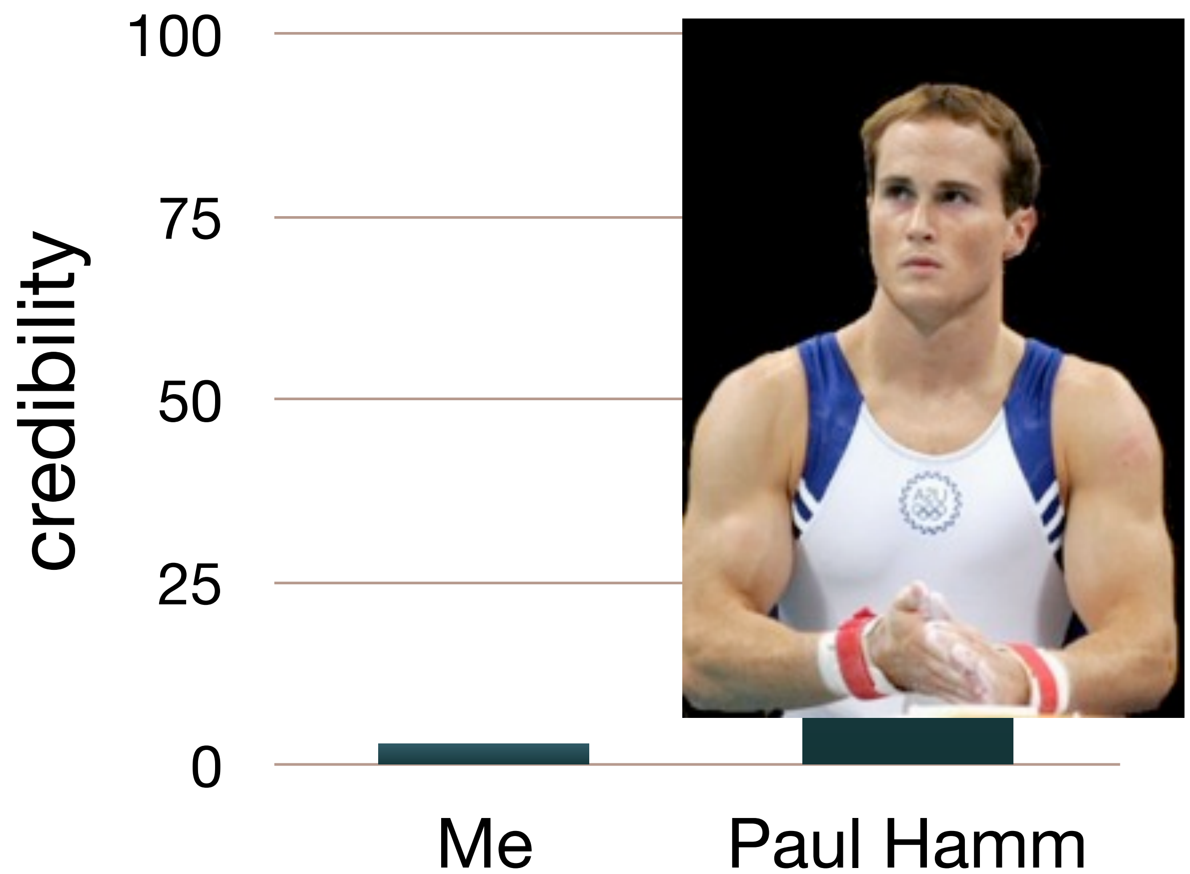


"I can do a backflip off this pommel horse"

Weights

- Why?

    - Just like in real life, where we trust people's claims differently, we will want to learn how to trust different models



"I can do a backflip off this pommel horse"

MODEL TRANSFORMS

- Log space transform

$$
\begin{aligned}
P(e \mid f) &\propto P(f \mid e)P(e) \\
&\propto P(f \mid e)^{\lambda_1} P(e)^{\lambda_2} \\
&\boxed{= \lambda_1 \log P(f \mid e) + \lambda_2 \log P(e)}
\end{aligned}
$$

Model Transforms

- Log space transform

$$\begin{aligned} P(e \mid f) &\propto P(f \mid e)P(e) \\ &\propto P(f \mid e)^{\lambda_1} P(e)^{\lambda_2} \\ &\boxed{= \lambda_1 \log P(f \mid e) + \lambda_2 \log P(e)} \end{aligned}$$

- Because:

    0.0001 * 0.0001 * 0.0001 = **0.000000000001**

    log(0.0001) + log(0.0001) + log(0.0001) = **-12**

MODEL TRANSFORMS

- Generalization (linearly-weighted models)

$$
\begin{aligned}
P(e \mid f) &\propto P(f \mid e)P(e) \\
&\propto P(f \mid e)^{\lambda_1} P(e)^{\lambda_2} \\
&= \lambda_1 \log P(f \mid e) + \lambda_2 \log P(e) \\
&\boxed{= \lambda_1 \phi_1(f,e) + \lambda_2 \phi_2(f,e) \\
&= \sum_i \lambda_i \phi_i(f,e)}
\end{aligned}
$$

PROBLEM 2: SEARCH

- The process of producing a translation of a sentence

PROBLEM 2: SEARCH

- The process of producing a translation of a sentence

- The second problem is **search** – given a model and a source sentence, how do we find the sentence that the model likes best?

PROBLEM 2: SEARCH

- The process of producing a translation of a sentence

- The second problem is **search** – given a model and a source sentence, how do we find the sentence that the model likes best?

  - *impractical*: enumerate all sentences, score them

PROBLEM 2: SEARCH

- The process of producing a translation of a sentence

- The second problem is **search** – given a model and a source sentence, how do we find the sentence that the model likes best?

  - *impractical*: enumerate all sentences, score them

  - *stack decoding*: use factored models and assemble the translations piece by piece

SEARCH

SEARCH

- The solution: factorized models and dynamic programming

SEARCH

- The solution: factorized models and dynamic programming

- We employ models that factorize across pieces of the translation process

SEARCH

- The solution: factorized models and dynamic programming

- We employ models that factorize across pieces of the translation process

  - *n-gram language model*: gives us probabilities for pairs of words

SEARCH

- The solution: factorized models and dynamic programming
- We employ models that factorize across pieces of the translation process
  - *n-gram language model*: gives us probabilities for pairs of words
  - *Model 1 translation model:* gives us probabilities for translation pairs

MODEL: TRANSLATION MODEL

$$\phi_1(e, a, c) = \log P(e, a \mid c)$$

MODEL: TRANSLATION MODEL

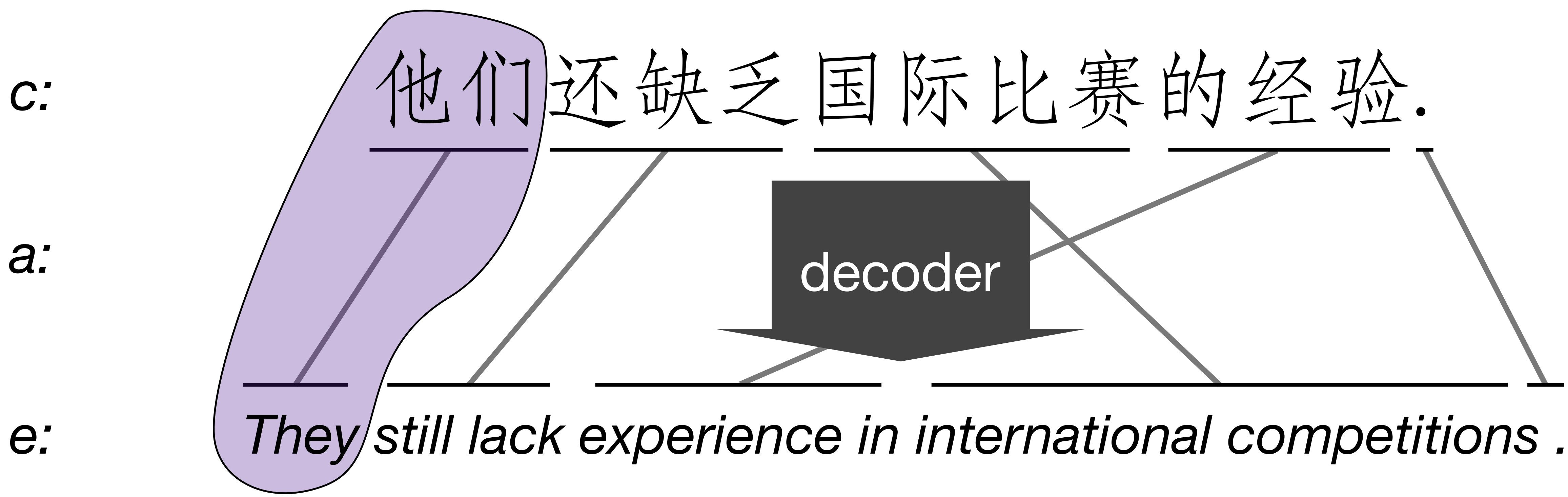
$$\phi_1(e, a, c) = \log P(e, a \mid c)$$

translation model

MODEL: TRANSLATION MODEL

$$\phi_1(e, a, c) = \log P(e, a \mid c)$$

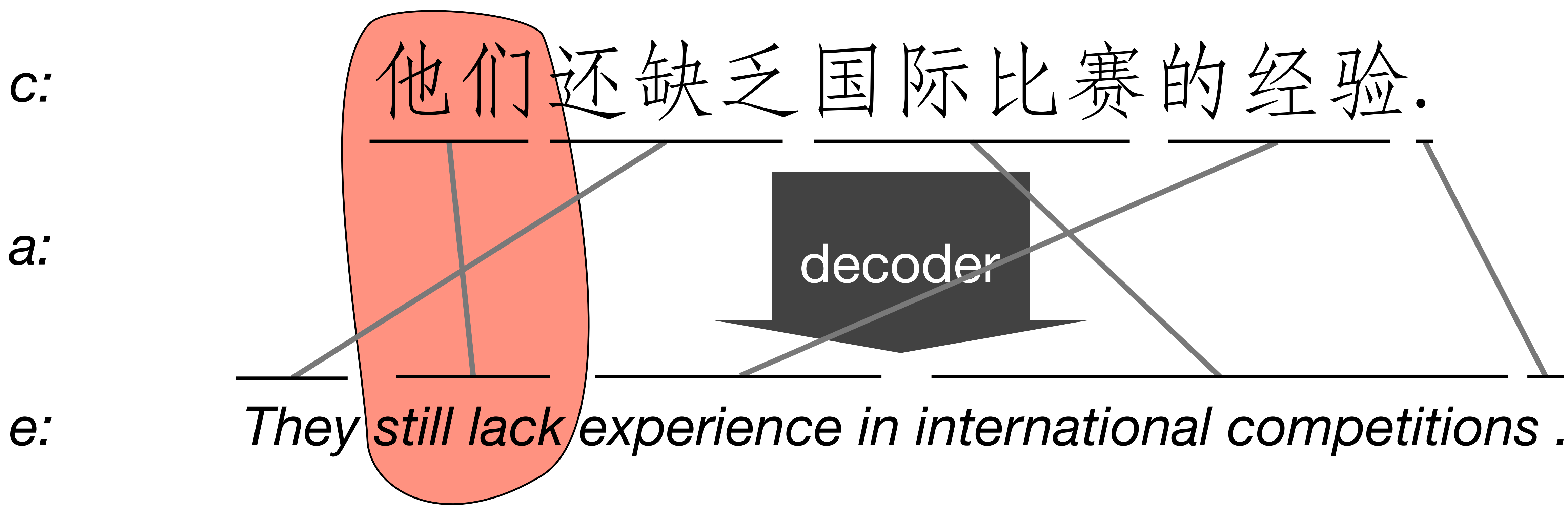
translation model



good translation

MODEL: TRANSLATION MODEL

$$\phi_1(e, a, c) = \log P(e, a \mid c)$$

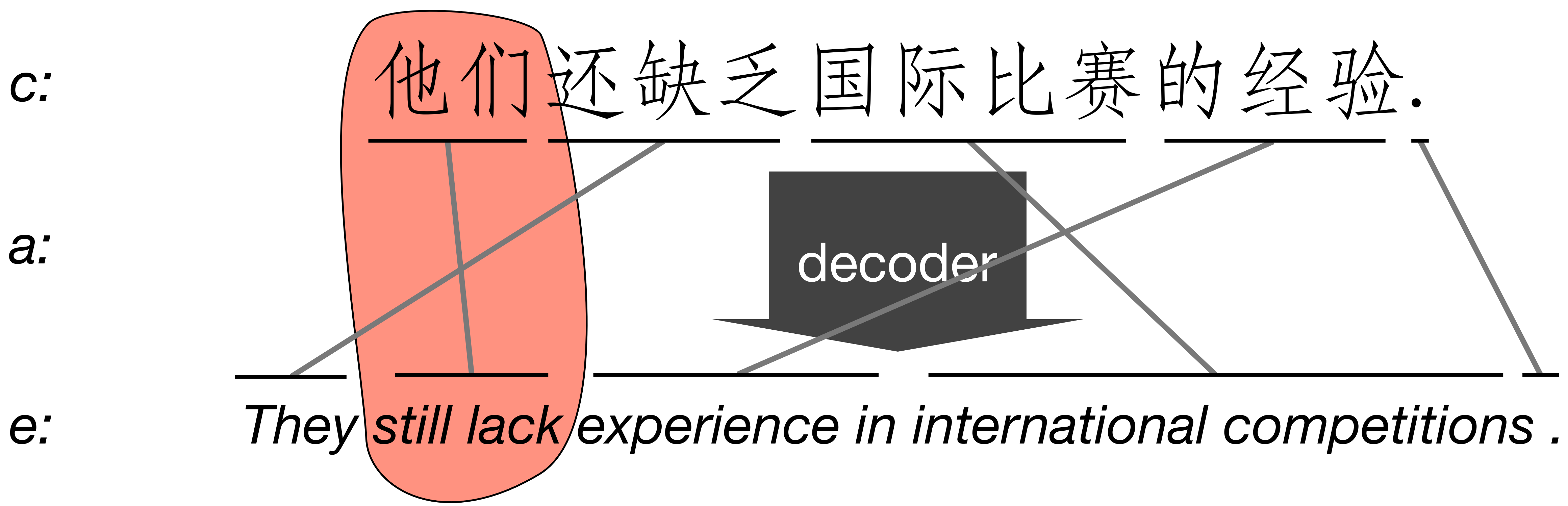


lower translation probability

Model: Translation model

$$\phi_1(e, a, c) = \log P(e, a \mid c)$$

translation model



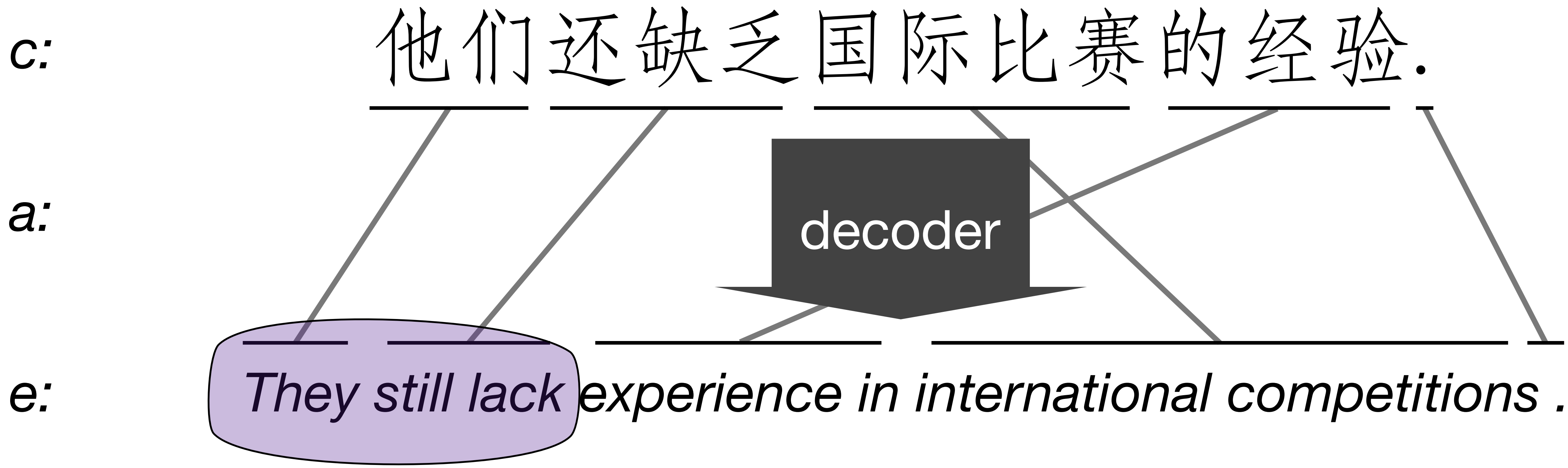
lower translation probability

Model: Language model

$$\phi_2(e, a, c) = \log P(e)$$

MODEL: LANGUAGE MODEL

$$\phi_2(e, a, c) = \log P(e)$$

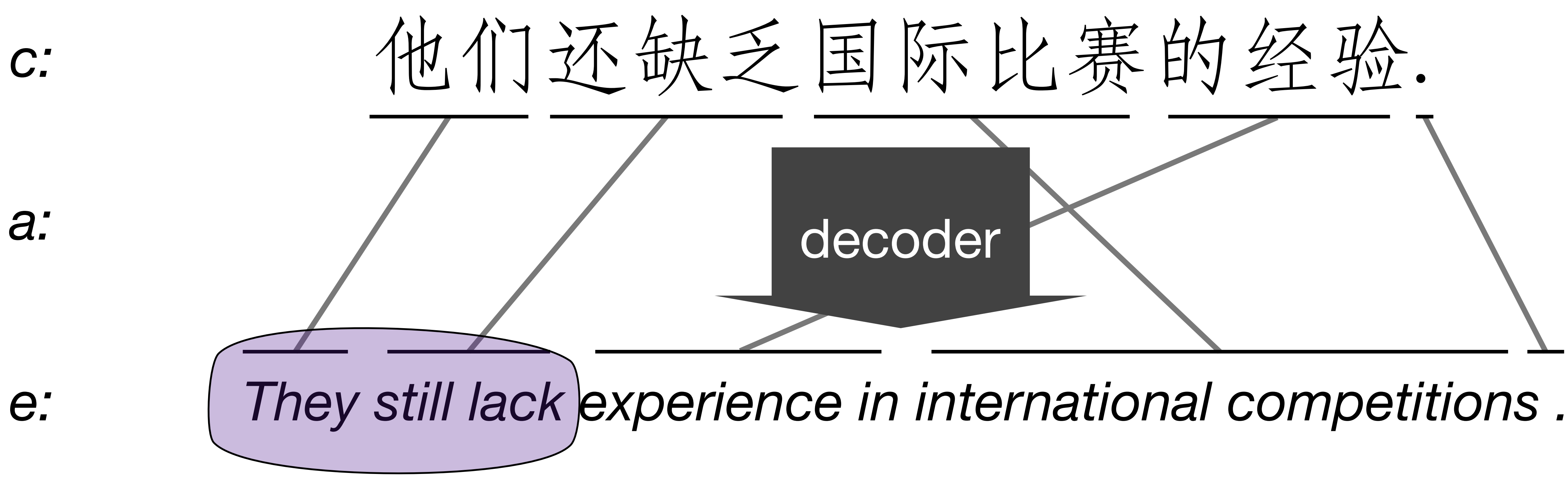


fluent English

MODEL: LANGUAGE MODEL

$$\phi_2(e, a, c) = \log P(e)$$

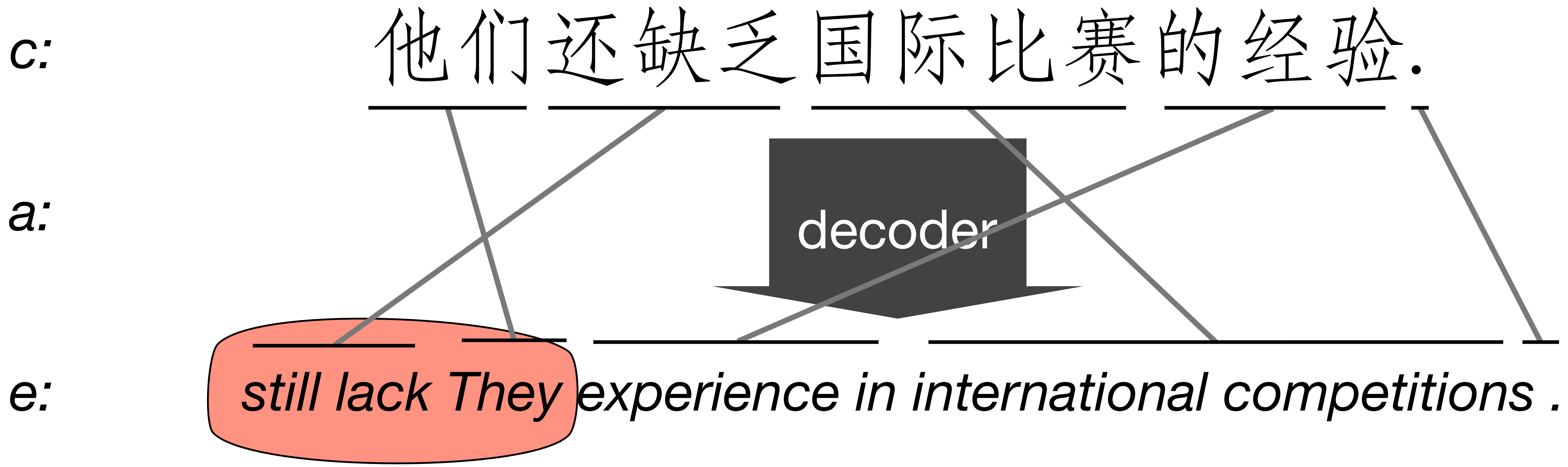
language model



fluent English

Model: Language model

$$\phi_2(e, a, c) = \log P(e)$$

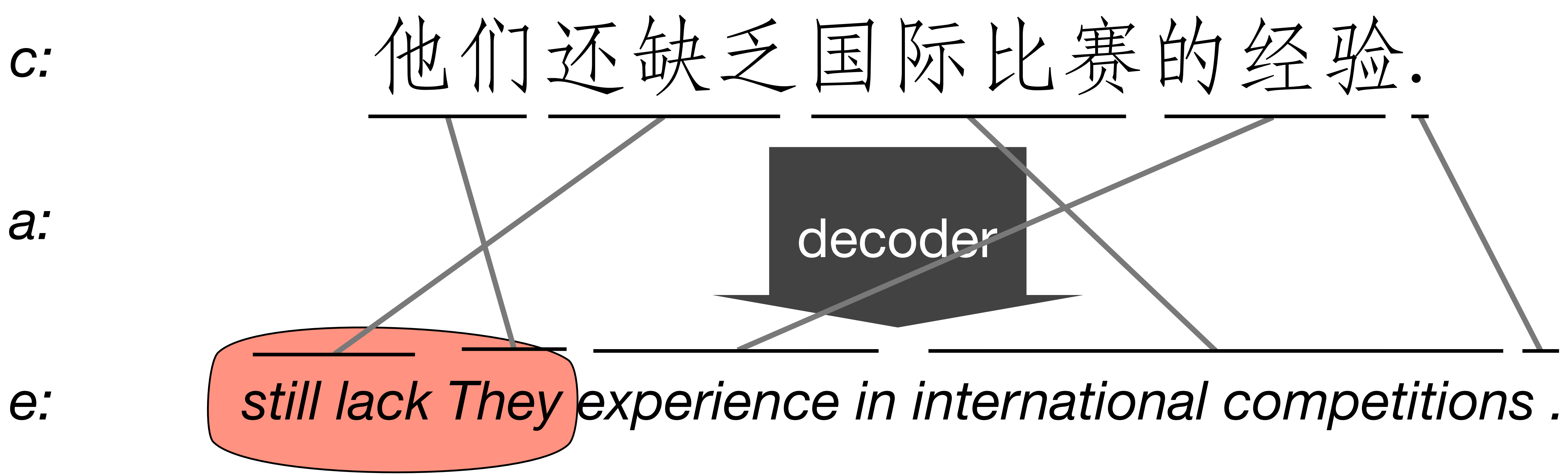


lower language model probability

Model: Language model

$$\phi_2(e, a, c) = \log P(e)$$

language model



lower language model probability

AN EXAMPLE

AN EXAMPLE

- Translate the sentence

  Yo tengo hambre

  from Spanish to English.

An Example

- Translate the sentence

  Yo tengo hambre

  from Spanish to English.

- We'll use an algorithm called *stack decoding* to efficiently explore the space of translations

AN EXAMPLE

- Translate the sentence

  Yo tengo hambre

  from Spanish to English.

- We'll use an algorithm called *stack decoding* to efficiently explore the space of translations

  - (Note: this has nothing to do with stacks; a better name would be *priority queue decoding*)

FACTORING MODELS

- Translating *Yo tengo hambre*

+

tengo
→am

=

FACTORING MODELS

- Translating *Yo tengo hambre*

- “Stack” decoding works by extending hypotheses word by word

FACTORING MODELS

- Translating *Yo tengo hambre*

- “Stack” decoding works by extending hypotheses word by word

- At each step, we take an existing hypothesis, grab an untranslated word, translate it, and extend the hypothesis

FACTORING MODELS

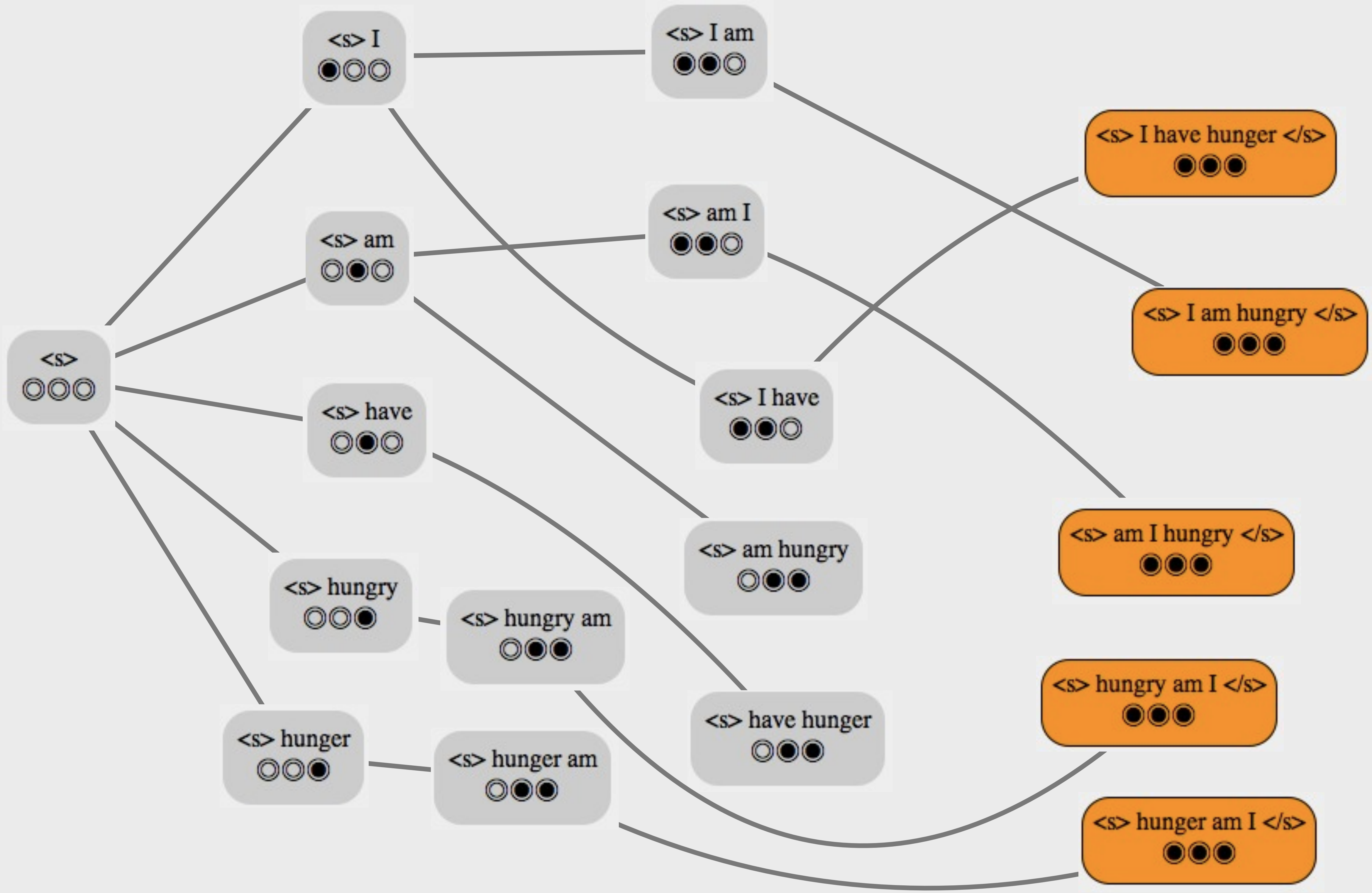
- Translating *Yo tengo hambre*

- “Stack” decoding works by extending hypotheses word by word

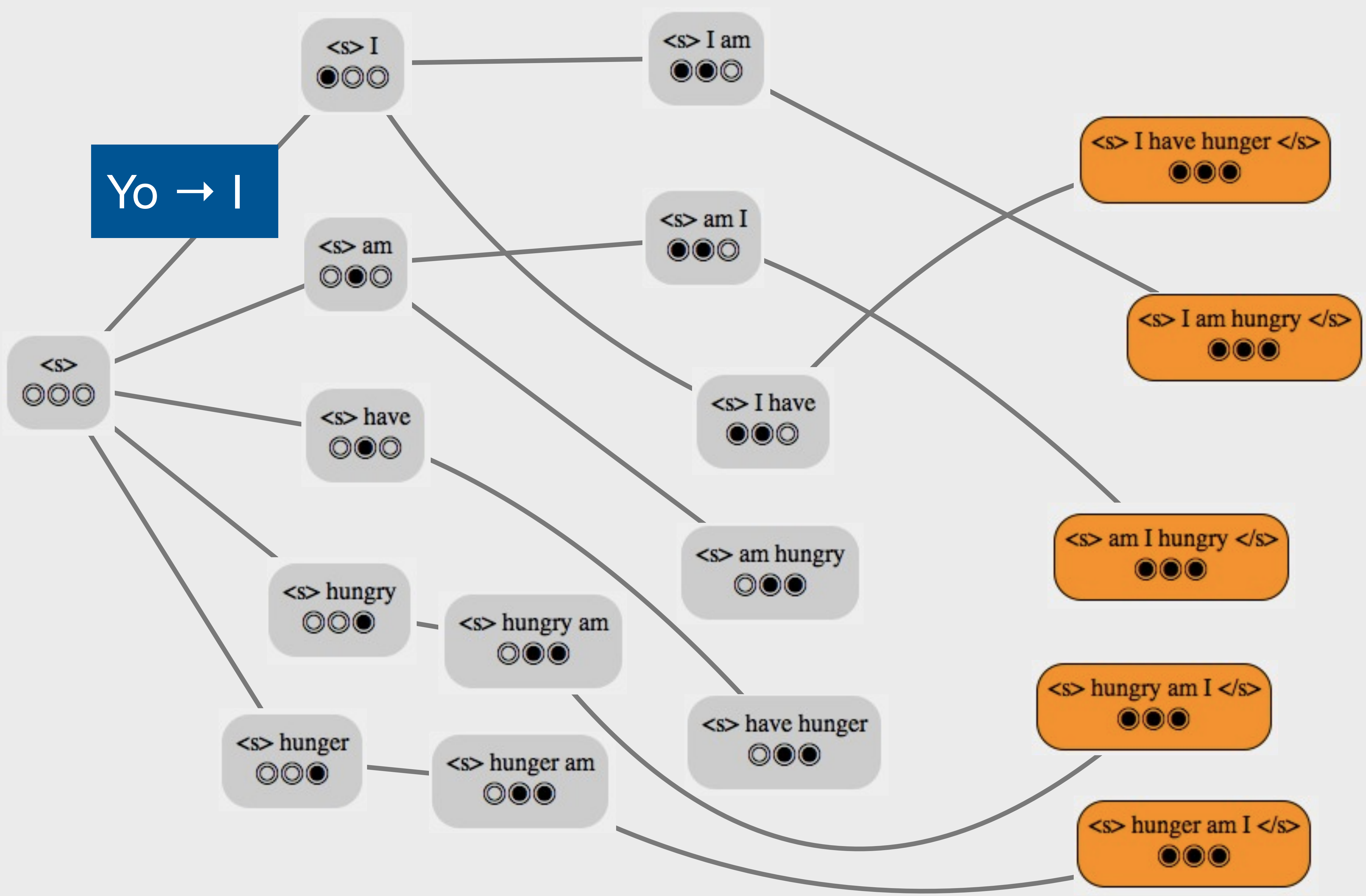
- At each step, we take an existing hypothesis, grab an untranslated word, translate it, and extend the hypothesis

- These can be arranged into a *search graph* representing the space we search

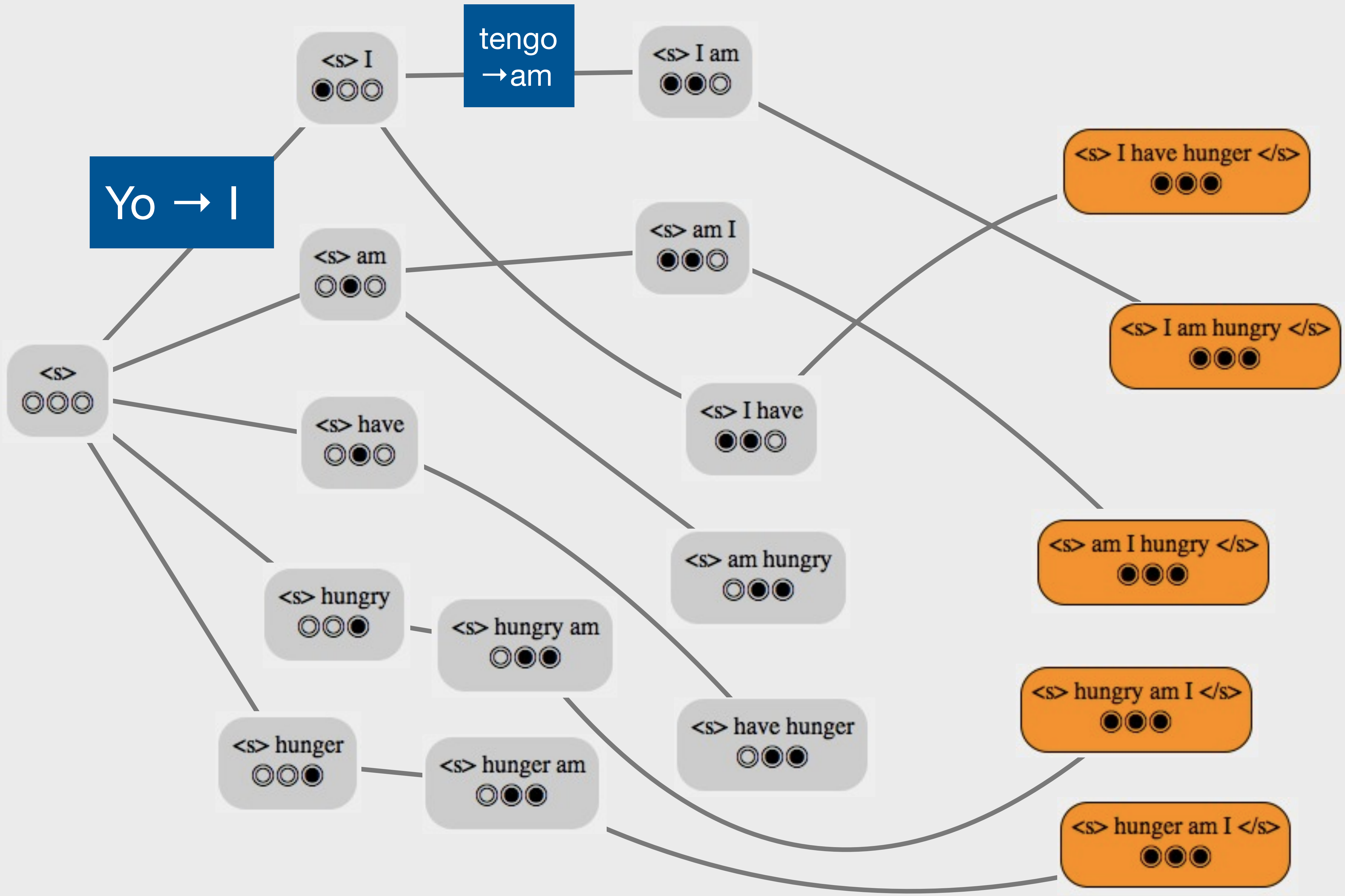
Factoring Models



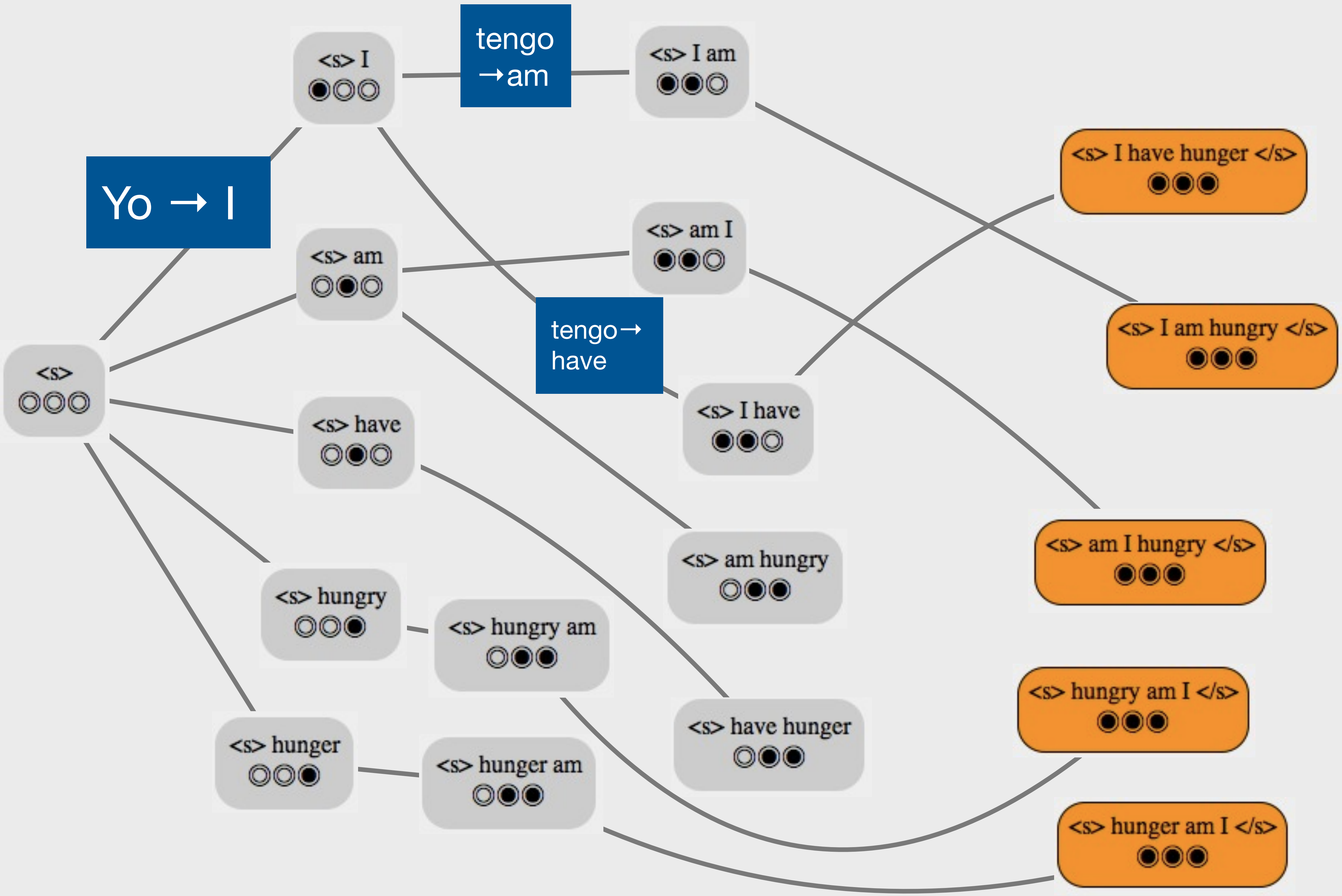
FACTORING MODELS



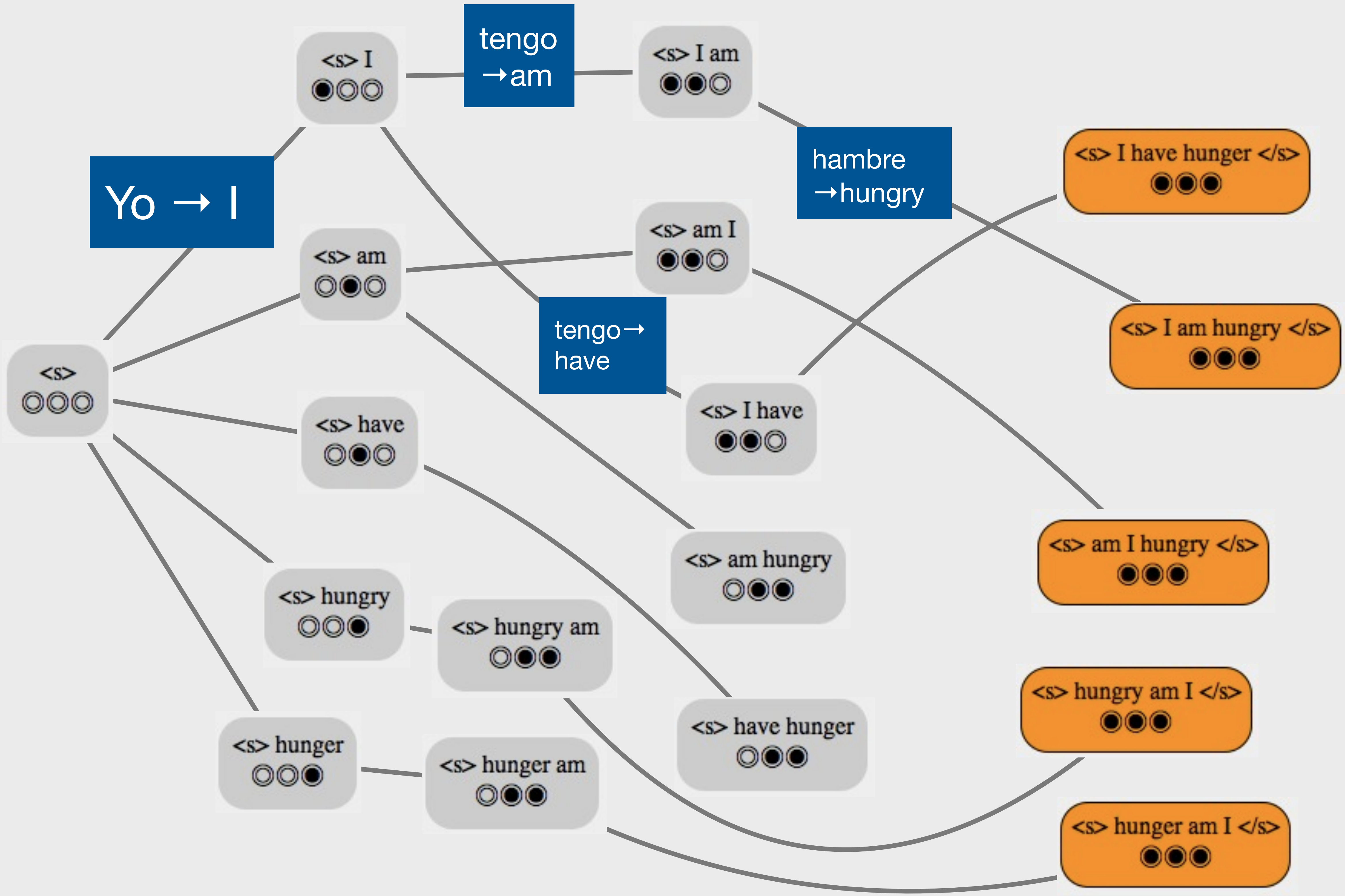
FACTORING MODELS



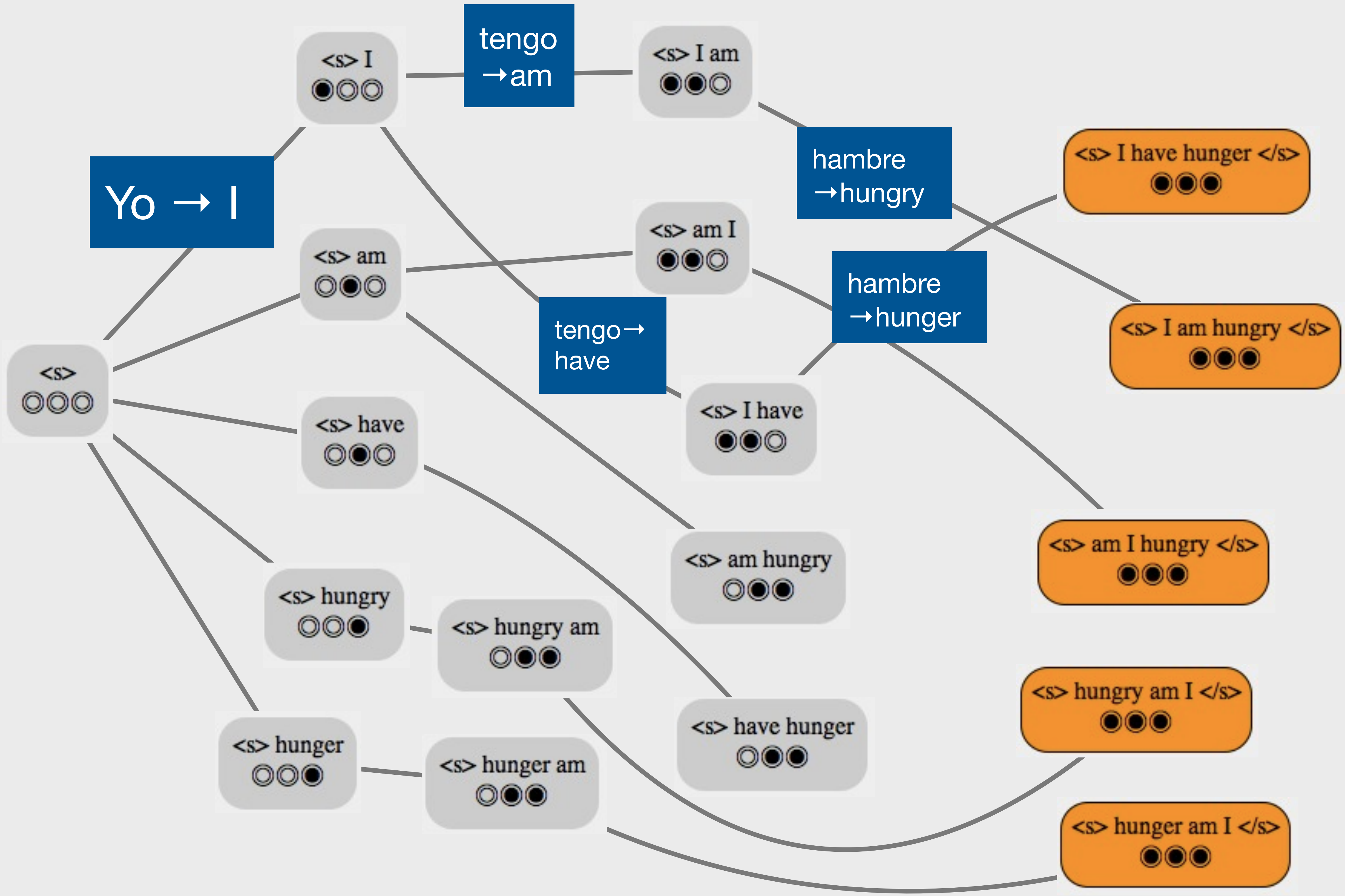
Factoring Models



FACTORING MODELS



FACTORING MODELS



FACTORING MODELS

- Stack decoding works by extending hypotheses word by word

FACTORING MODELS

- Stack decoding works by extending hypotheses word by word

- These can be arranged into a *search graph* representing the space we search

FACTORING MODELS

- Stack decoding works by extending hypotheses word by word

- These can be arranged into a *search graph* representing the space we search

- The component models we use need to *factorize* over this graph, and we accumulate the score as we go

FACTORING MODELS

- Example hypothesis creation:

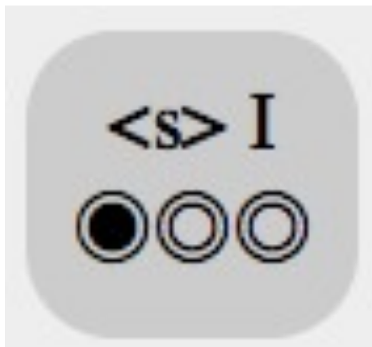
old hypothesis

add word

new hypothesis

FACTORING MODELS

- Example hypothesis creation:



\+ tengo →am =

old hypothesis | add word | new hypothesis

- **translation model**: trivial case, since all the words are translated independently

$$\text{hypothesis.score} \mathrel{+}= \log(P_{TM}(\text{am} \mid \text{tengo}))$$

FACTORING MODELS

- Example hypothesis creation:

old hypothesis | add word | new hypothesis

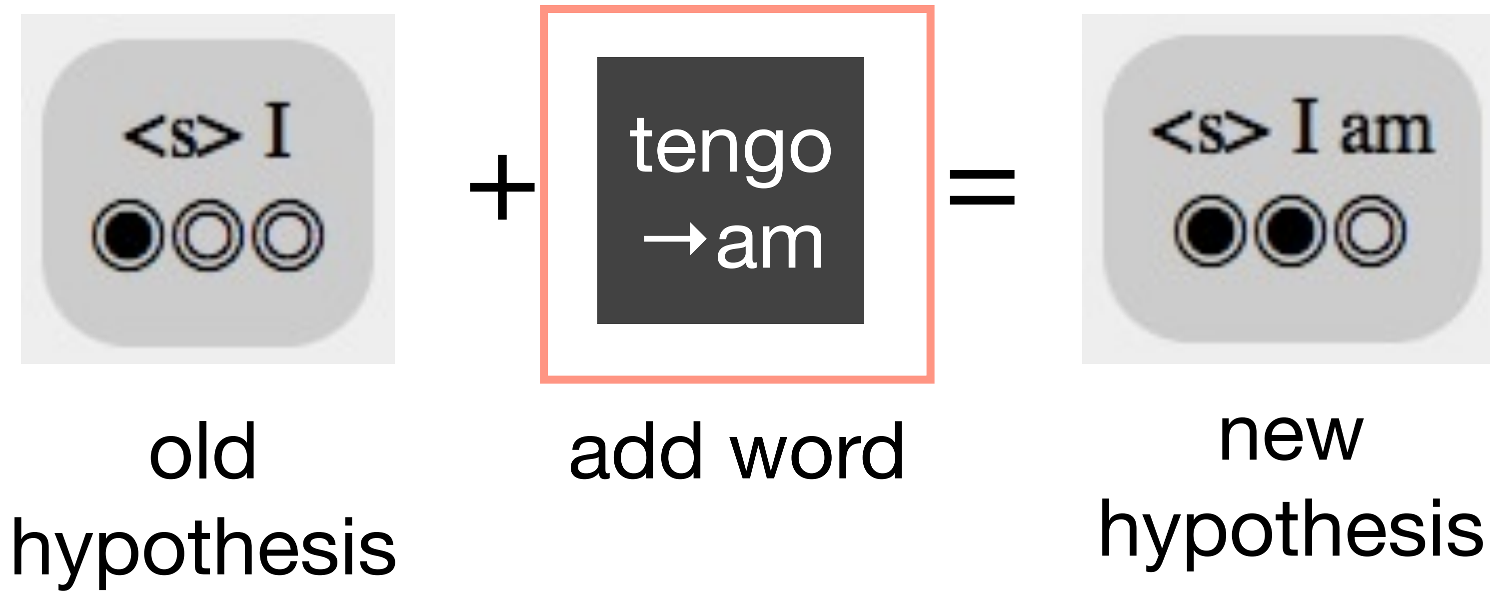
- **translation model**: trivial case, since all the words are translated independently

$$\text{hypothesis.score} \mathrel{+}= \log(P_{TM}(\text{am} \mid \text{tengo}))$$

- a function of just the word that is added

Factoring Models

- Example hypothesis creation:

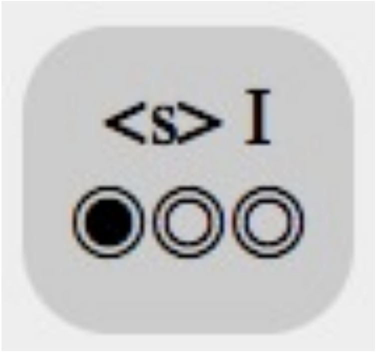


- **translation model**: trivial case, since all the words are translated independently

$$\text{hypothesis.score} \mathrel{+}= \log(P_{TM}(\text{am} \mid \text{tengo}))$$

- a function of just the word that is added

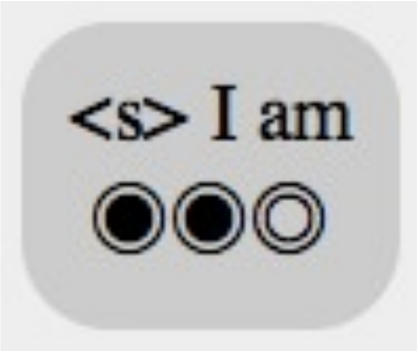
FACTORING MODELS



+

tengo
→am

=



old hypothesis

add word

new hypothesis

FACTORING MODELS

- Example hypothesis creation:

old hypothesis

add word

new hypothesis

FACTORING MODELS

- Example hypothesis creation:

old hypothesis — add word — new hypothesis

- **language model**: still easy, since (bigram) language models depend only on the previous word

$$\text{hypothesis.score} \mathrel{+}= \log(P_{LM}(\text{am} \mid \text{I}))$$

FACTORING MODELS

- Example hypothesis creation:

old hypothesis | add word | new hypothesis

- **language model**: still easy, since (bigram) language models depend only on the previous word

$$\text{hypothesis.score} \mathrel{+}= \log(P_{LM}(\text{am} \mid \text{I}))$$

- a function of the old hyp. and the new word translation

FACTORING MODELS

- Example hypothesis creation:

old hypothesis | add word | new hypothesis

- **language model**: still easy, since (bigram) language models depend only on the previous word

$$\text{hypothesis.score} \mathrel{+}= \log(P_{LM}(\text{am} \mid \text{I}))$$

- a function of the old hyp. and the new word translation

STACK DECODING

- Start with a list of hypotheses, containing only the empty hypothesis

STACK DECODING

- Start with a list of hypotheses, containing only the empty hypothesis

- For each stack

STACK DECODING

- Start with a list of hypotheses, containing only the empty hypothesis
- For each stack
  - For each hypothesis

STACK DECODING

- Start with a list of hypotheses, containing only the empty hypothesis
- For each stack
  - For each hypothesis
    - For each applicable word

Stack decoding

- Start with a list of hypotheses, containing only the empty hypothesis
- For each stack
  - For each hypothesis
    - For each applicable word
      - Extend the hypothesis with the word

Stack decoding

- Start with a list of hypotheses, containing only the empty hypothesis
- For each stack
  - For each hypothesis
    - For each applicable word
      - Extend the hypothesis with the word
      - Place the new hypothesis on the right stack

DYNAMIC PROGRAMMING

DYNAMIC PROGRAMMING

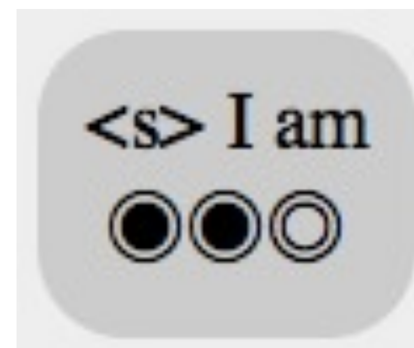
- We saw Tuesday how huge the search space could get

Dynamic Programming

- We saw Tuesday how huge the search space could get

- Notice anything here?

\+ tengo →am =



score +=
P_{TM}(am | tengo)
\+ P_{LM}(am | I)

old hypothesis | add word | new hypothesis

DYNAMIC PROGRAMMING

- We saw Tuesday how huge the search space could get

- Notice anything here?

+

tengo
→am

=

score +=
P_{TM}(am | tengo)
+ P_{LM}(am | I)

old hypothesis

add word

new hypothesis

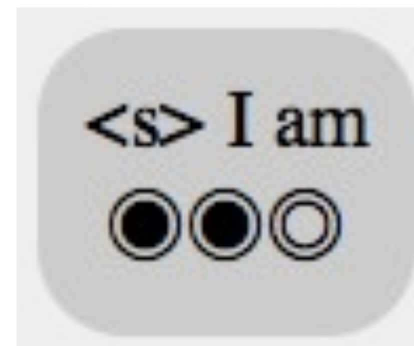
- (1) <s> *is never used* in computing the scores AND
  (2) <s> is implicit in the graph structure

DYNAMIC PROGRAMMING

- We saw Tuesday how huge the search space could get

- Notice anything here?

+ tengo →am =



score += P_{TM}(am | tengo) + P_{LM}(am | I)

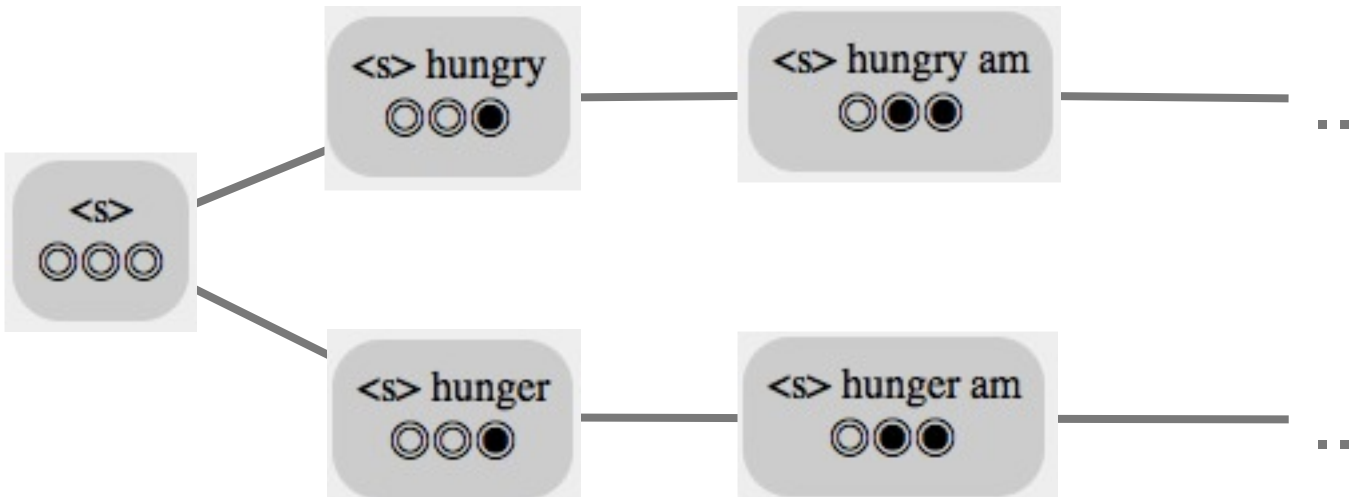
old hypothesis | add word | new hypothesis

- (1) <s> *is never used* in computing the scores AND
  (2) <s> is implicit in the graph structure

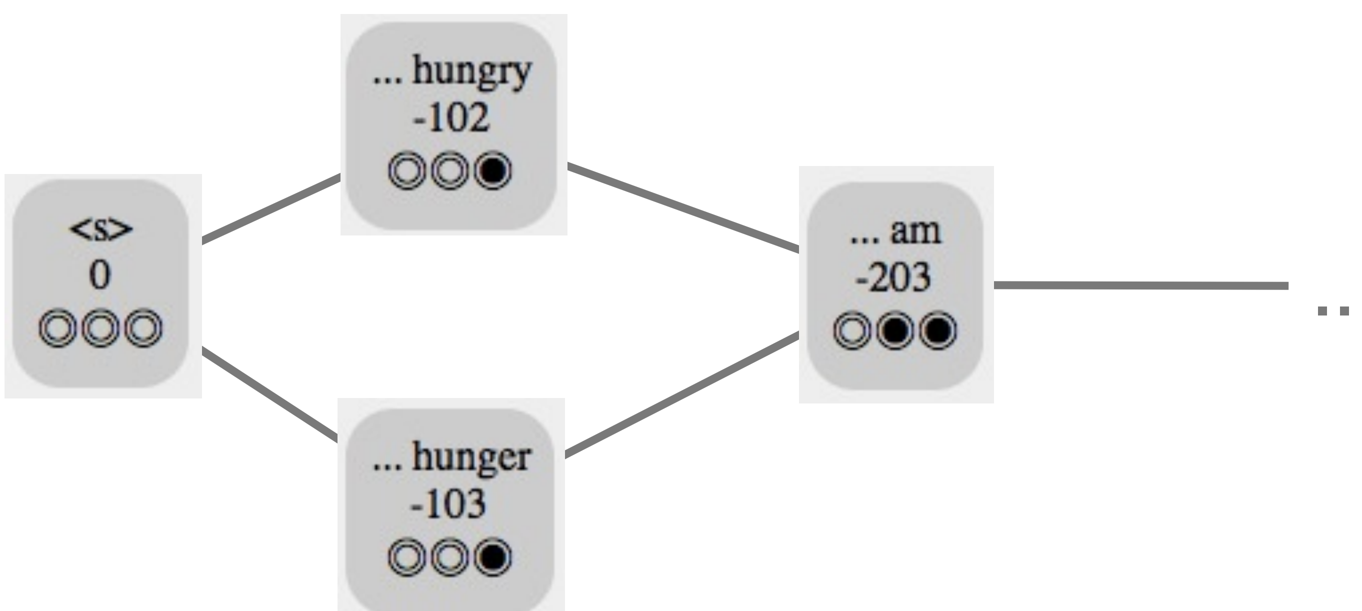
- let's get rid of the extra state!

DYNAMIC PROGRAMMING

- Before



- After



The score of the new hypothesis is the maximum way to compute it

STACK DECODING (WITH DP)

- Start with a list of hypotheses, containing only the empty hypothesis
- For each stack
  - For each hypothesis
    - For each applicable word
      - Extend the hypothesis with the word
      - Place the new hypothesis on the right stack

STACK DECODING (WITH DP)

- Start with a list of hypotheses, containing only the empty hypothesis
- For each stack
  - For each hypothesis
    - For each applicable word
      - Extend the hypothesis with the word
      - Place t

IF either (1) no equivalent hypothesis exists or (2) this hypothesis has a higher score.

OTHER ISSUES IN DECODING

OTHER ISSUES IN DECODING

- Pruning

OTHER ISSUES IN DECODING

- Pruning
- Search heuristics (e.g., distortion limits)

OTHER ISSUES IN DECODING

- Pruning

- Search heuristics (e.g., distortion limits)

- Other useful feature functions

OTHER ISSUES IN DECODING

- Pruning
- Search heuristics (e.g., distortion limits)
- Other useful feature functions
- Tuning of the weights

SUMMARY

SUMMARY

- We've barely touched on issues in decoding, much less machine translation

SUMMARY

- We’ve barely touched on issues in decoding, much less machine translation

- The recommended readings have more information; you can also see our class page at mt-class.org (which we’ll likely teach again in SP/2014)

SUMMARY

- We've barely touched on issues in decoding, much less machine translation

- The recommended readings have more information; you can also see our class page at mt-class.org (which we'll likely teach again in SP/2014)

- On Wednesday, we'll talk about learning and using **hierarchical models** for decoding

Syntax-based Statistical Machine Translation

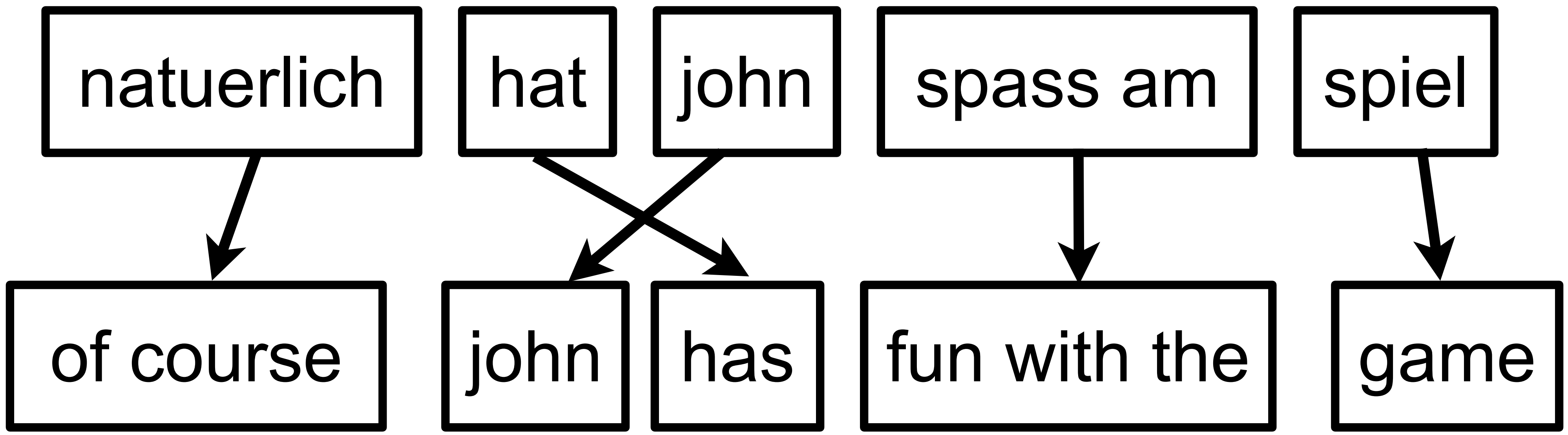
5 December 2012
Natural Language Processing 600.465
Guest Lecturer: Matt Post
Slides amalgamated from mt-class.org

Goals

- Understand why syntax is important for reordering models
  - Review non-syntactic reordering models for phrase-based machine translation
  - Review the "Clause Restructuring" approach of Collins, Koehn, and Kucerova, its advantages and limitations
- Learn about Synchronous Context Free Grammars
  - Introduce notation, and basic algorithm
- Understand how we learn SCFGs from bitexts
- Get a sense of the different flavors of SCFGs

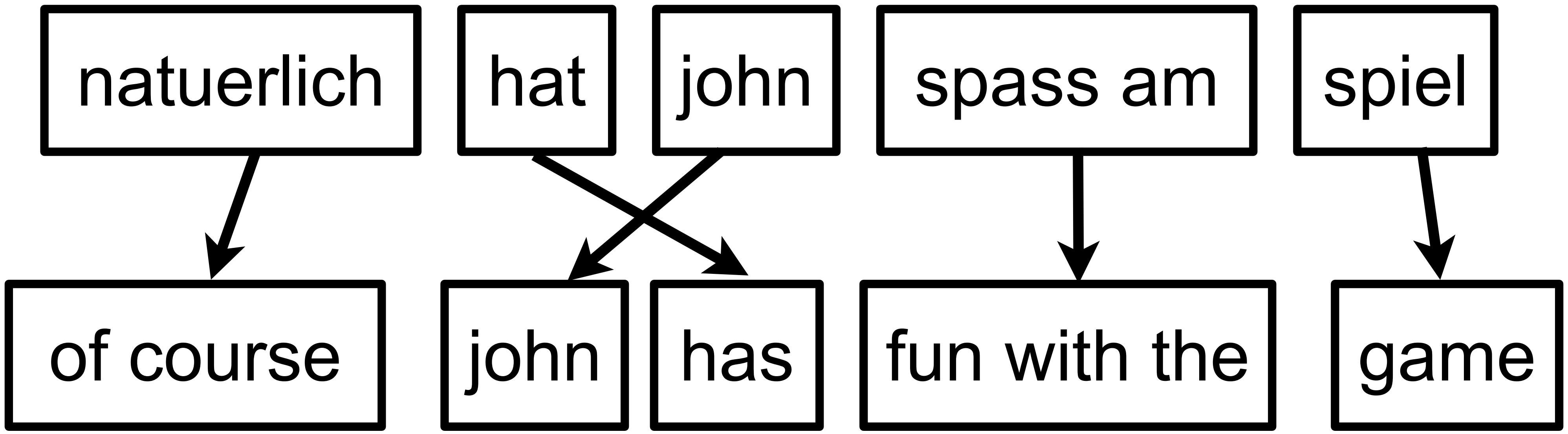
Why syntax matters

Phrase-based model

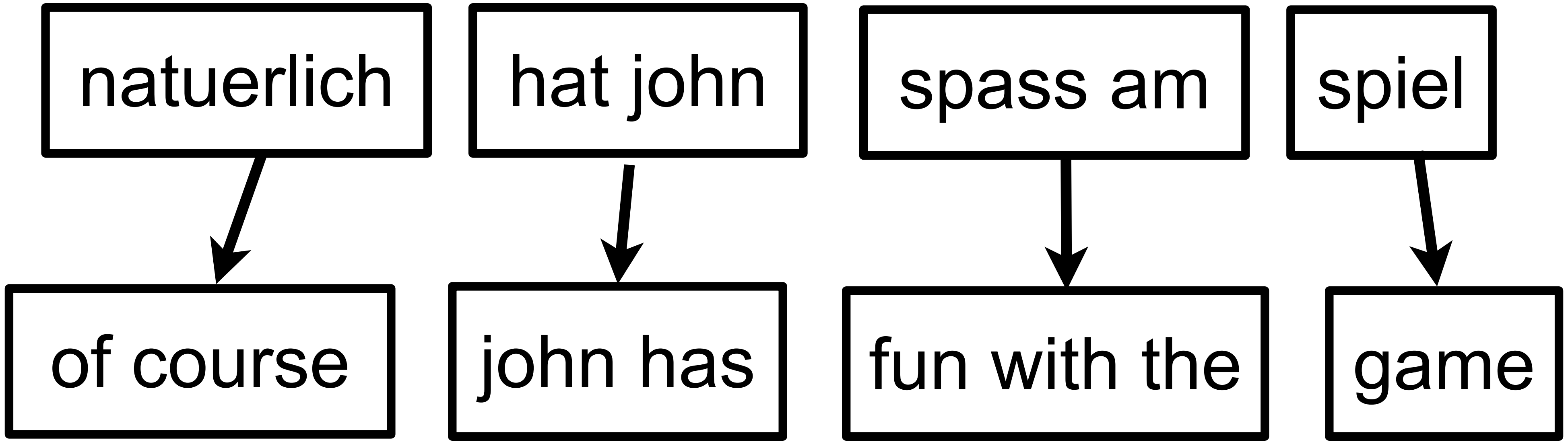


- Foreign input is segmented in phrases
- Each phrase is translated into English
- Phrases are reordered

Some Reordering Already Captured



- Local reordering can be captured within phrases



Phrase translation table

- Main knowledge source: table with phrase translations and their probabilities
- Example: phrase translations for natuerlich

Source	**Translation**	**Probability** $\phi(e\|f)$
natuerlich	of course	0.5
natuerlich	naturally	0.3
natuerlich	of course ,	0.15
natuerlich	, of course ,	0.05

Probabilistic Model

- Bayes rule
  - $\mathbf{e}_{\text{best}} = \text{argmax}_{\mathbf{e}}\ p(\mathbf{e}|\mathbf{f})$
    $= \text{argmax}_{\mathbf{e}}\ p(\mathbf{f}|\mathbf{e})\ p_{\text{lm}}(\mathbf{e})$
  - translation model $p(\mathbf{e}|\mathbf{f})$
  - language model $p_{\text{lm}}(\mathbf{e})$
- Reordering score can be incorporated in the TM

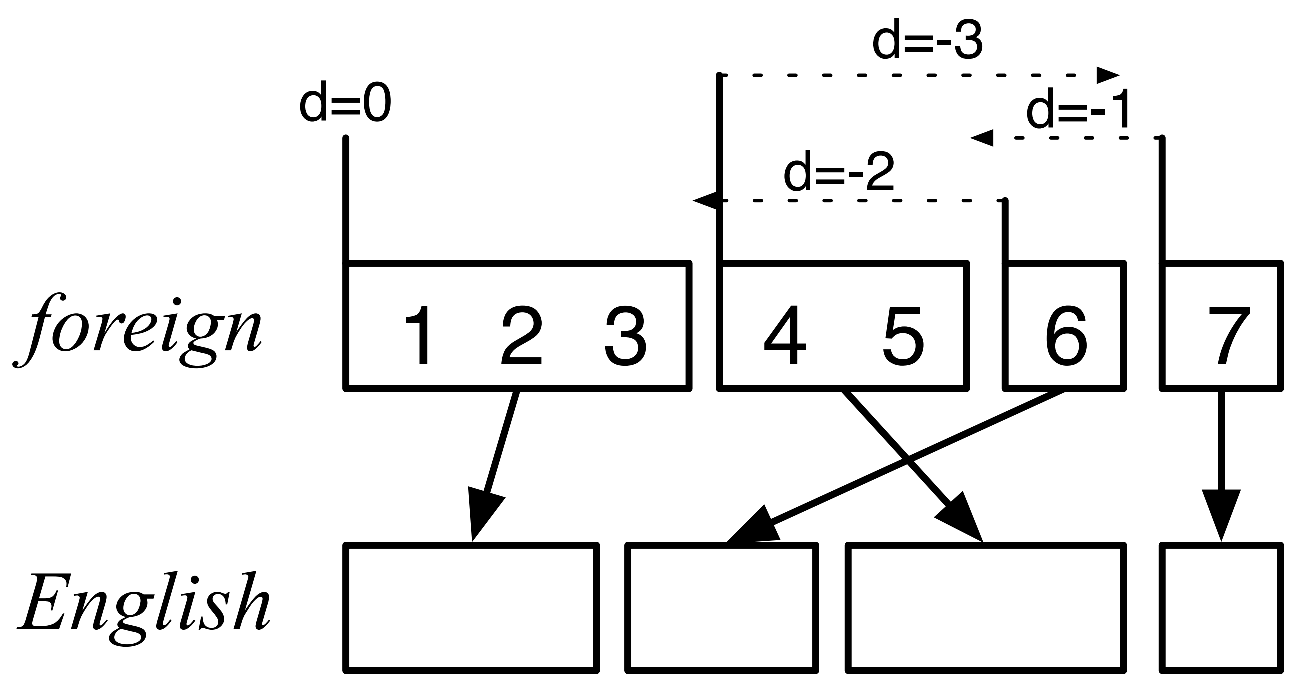
$$p(\bar{f}_1^I|\bar{e}_1^I) = \prod_{i=1}^{I} \phi(\bar{f}_i|\bar{e}_i)\ d(start_i - end_{i-1} - 1)$$

  - phrase translation probability ϕ
  - reordering probability d

Log-linear model

$$
\begin{aligned}
p(e,a|f) = \mathsf{exp}(\lambda_\phi \sum_{i=1}^{I} \mathsf{log}\ \phi(\bar{f}_i|\bar{e}_i) + \\
\lambda_d \sum_{i=1}^{I} \mathsf{log}\ d(a_i - b_{i-1} - 1) + \\
\lambda_{LM} \sum_{i=1}^{|\mathbf{e}|} \mathsf{log}\ p_{LM}(e_i|e_1 \ldots e_{i-1}))
\end{aligned}
$$

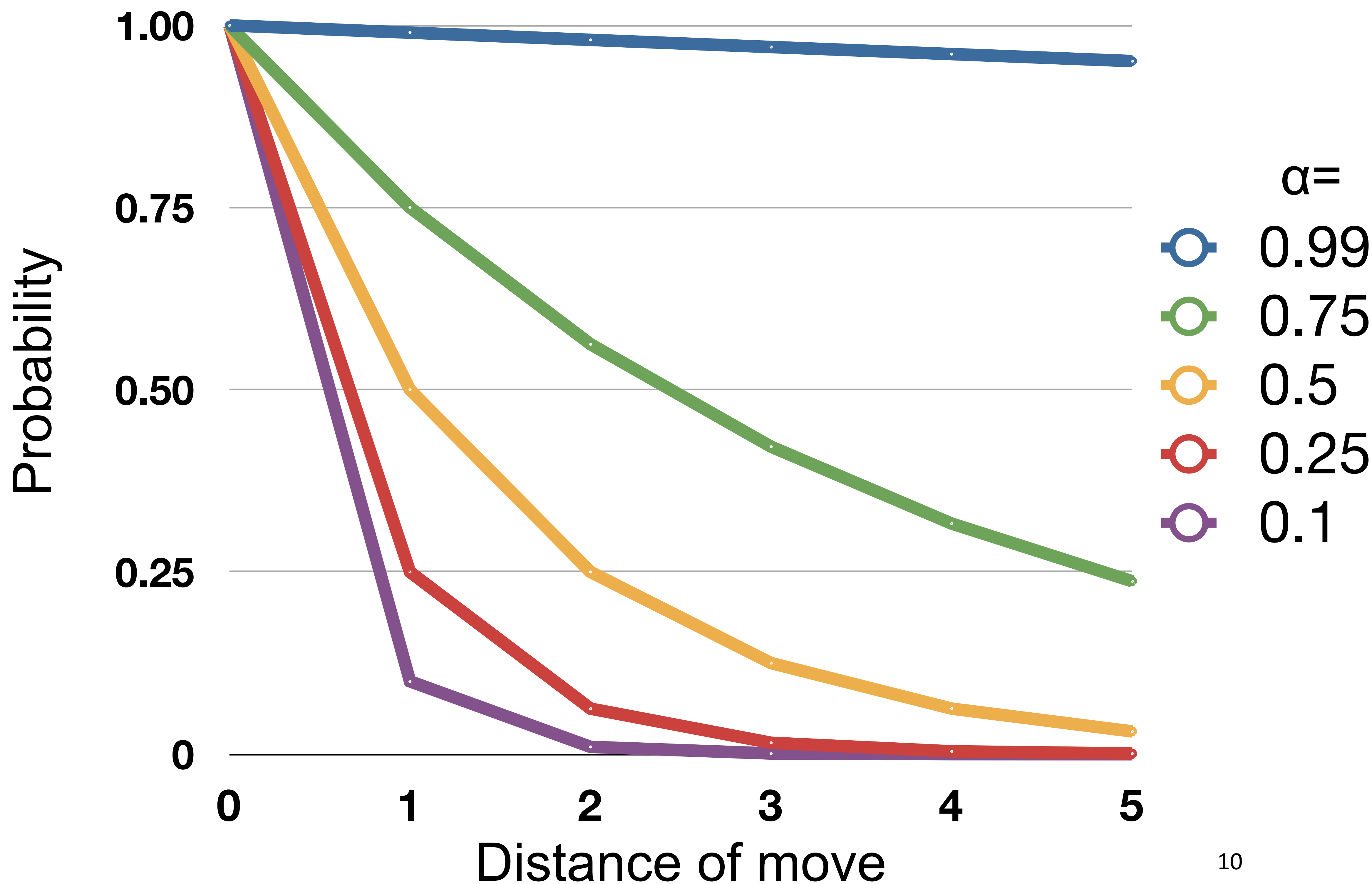
Distance-based Reordering



phrase	translates	movement	distance
1	1–3	start at beginning	0
2	6	skip over 4–5	+2
3	4–5	move back over 4–6	-3
4	7	skip over 6	+1

Scoring function: $d(x) = \alpha^{|x|}$ – exponential with distance

Values of α



1.00
0.75
0.50
0.25
0
Probability
0
1
2
3
4
5
Distance of move
α=
0.99
0.75
0.5
0.25
0.1

Discussion: Distance-based reordering

- What do you think of it?
- Is it a good model for how reordering works across languages?
- What is it missing?

(Discuss with your neighbor)

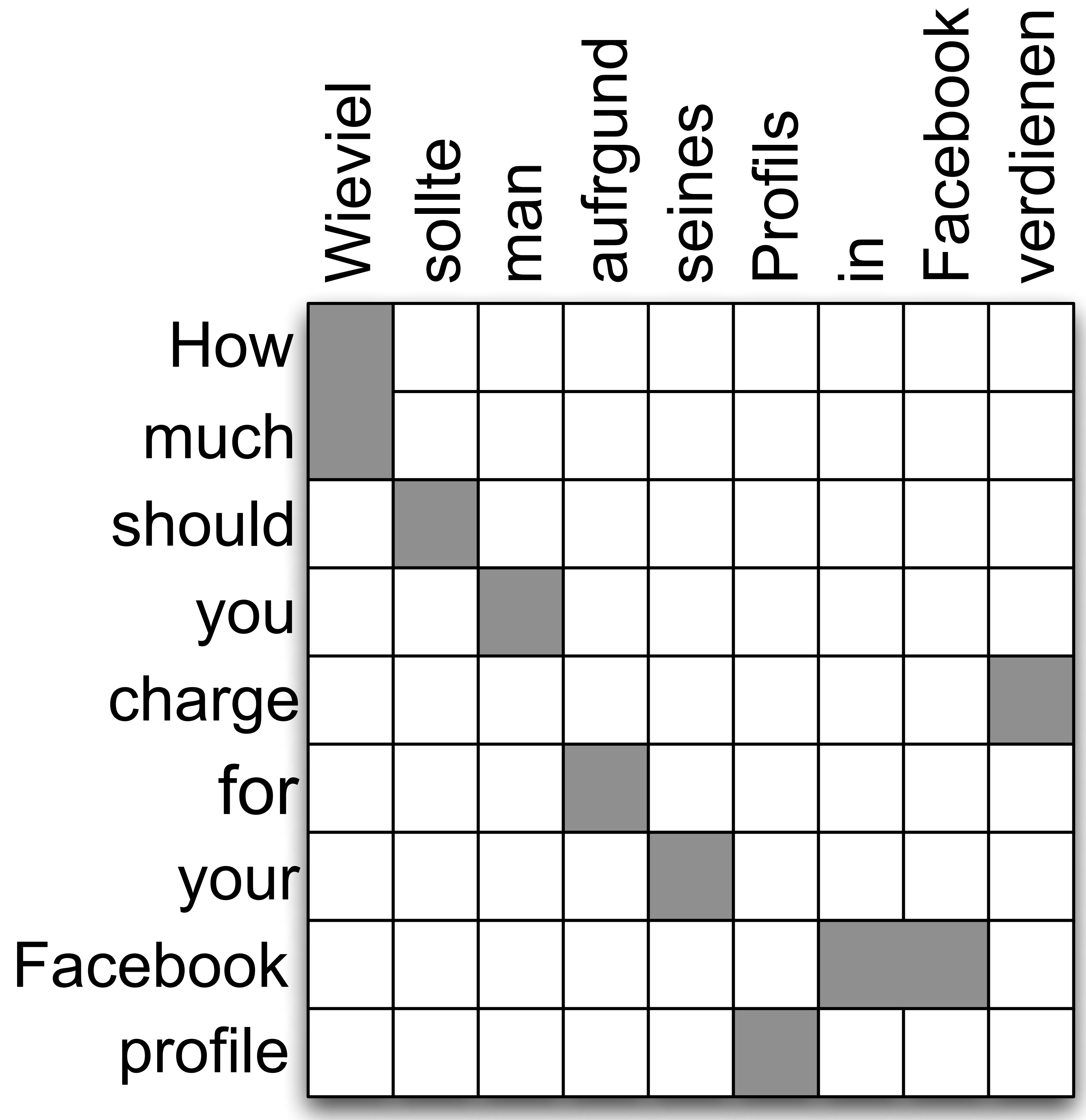
Distance-based reordering

- Small values of α, severely discourage reordering
  - Limit reordering to monotonic or a narrow window
  - OK for languages with very similar word orders
  - Bad for languages with different word orders
- The distance-based penalty applies uniformly to all words and all word types
  - Doesn't know that adjectives and nouns should swap when translating from French to English
- Puts most responsibility on the language model

How else could we model reordering?

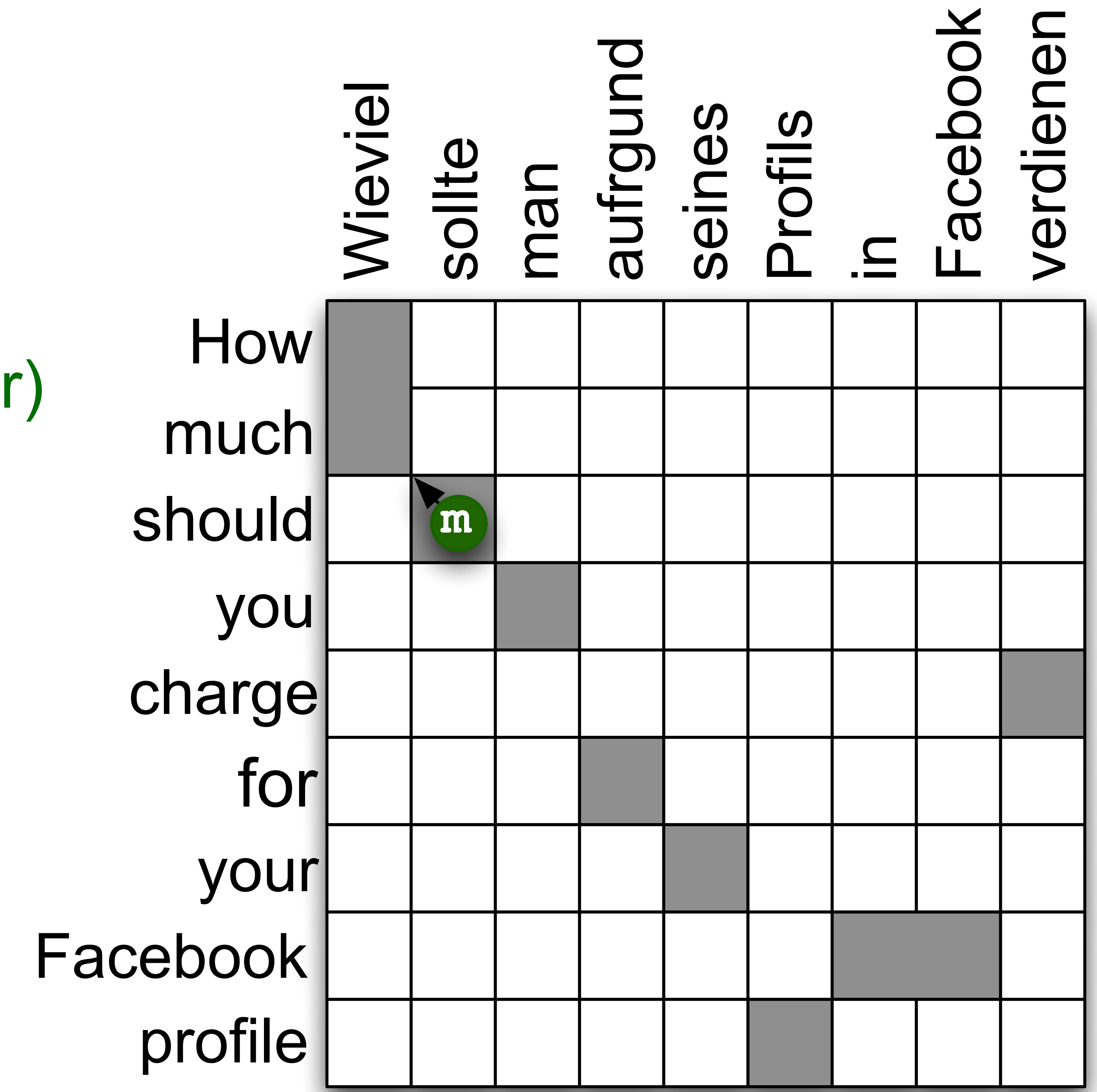
- Why not assign a distinct reordering probability to each word/phrase in the phrase table?
  - p(reorder | f, e)
- This is known as lexicalized reordering
- How can we estimate that probability?

Lexicalized Reordering model



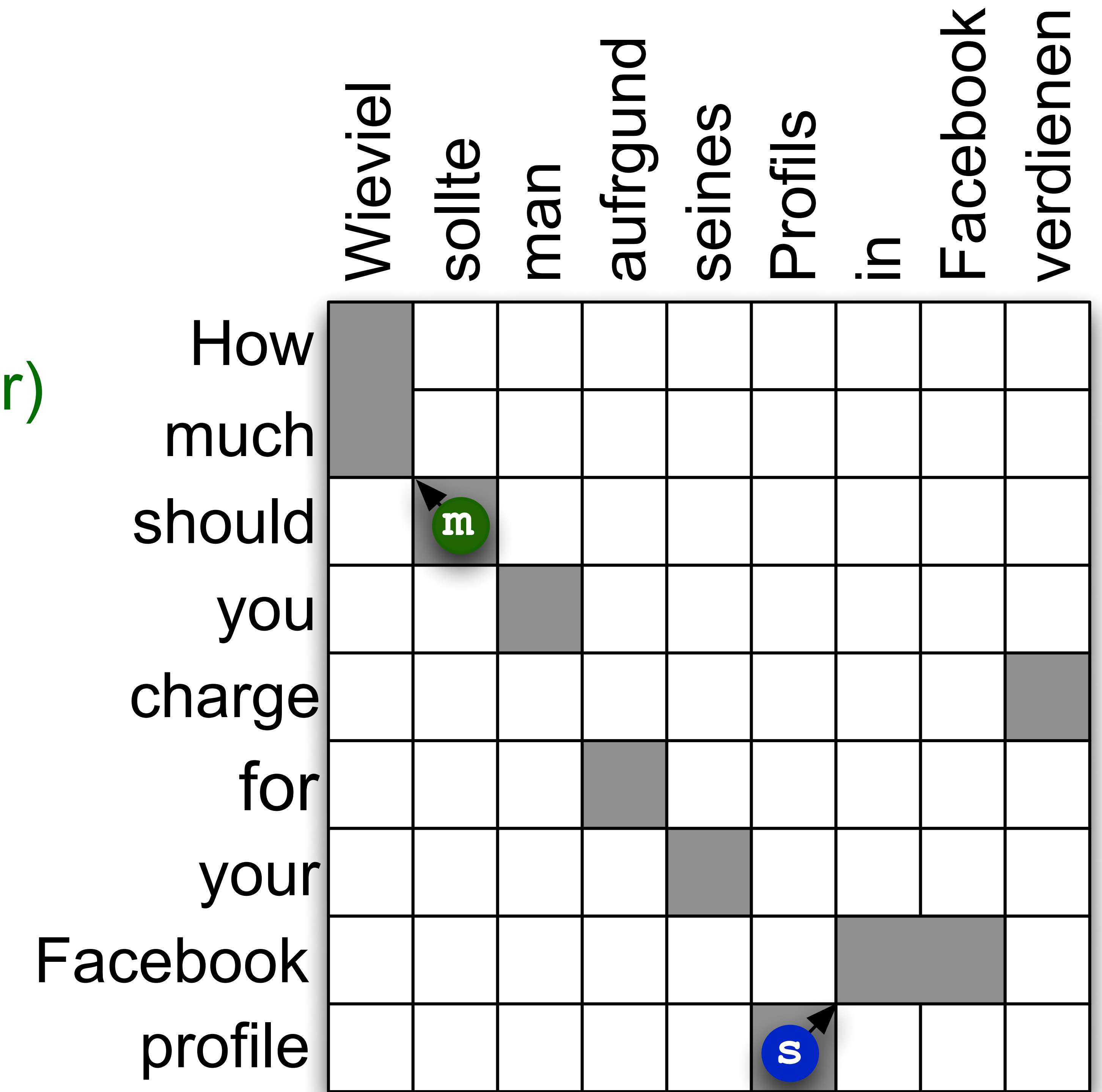
Lexicalized Reordering model

m: monotone (keep order)



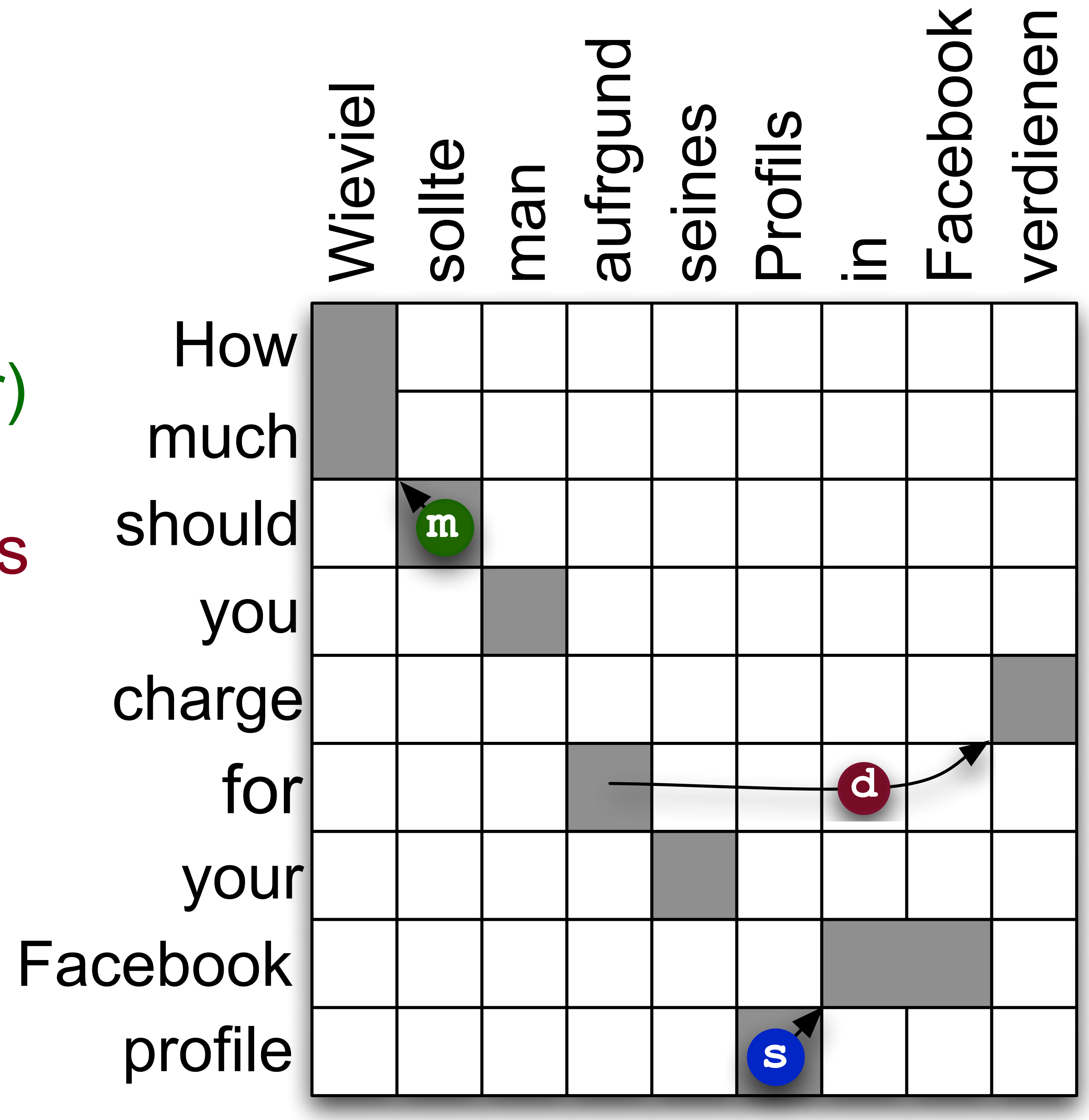
Lexicalized Reordering model

m: monotone (keep order)
s: swap order



Lexicalized Reordering model

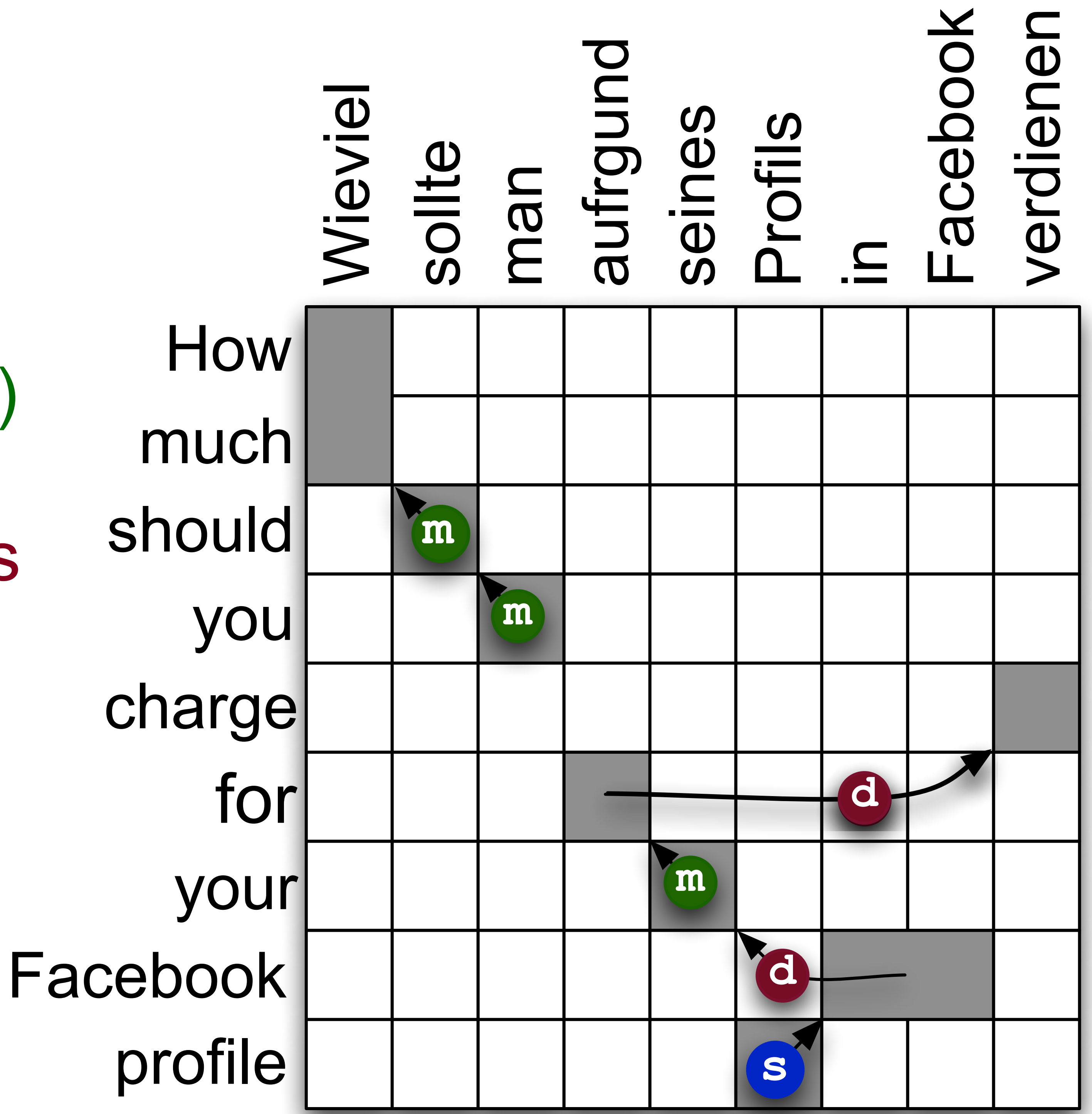
m: monotone (keep order)
s: swap order
d: become discontinuous



Lexicalized Reordering model

m: monotone (keep order)
s: swap order
d: become discontinuous

Reordering features are probability estimates of s, d, and m



Lexicalized Reordering table

- Identical phrase pairs <f,e> as in the phrase translation table
- Contains values for p(monotone|e,f), p(swap|e,f), p(discontinuous|e,f)

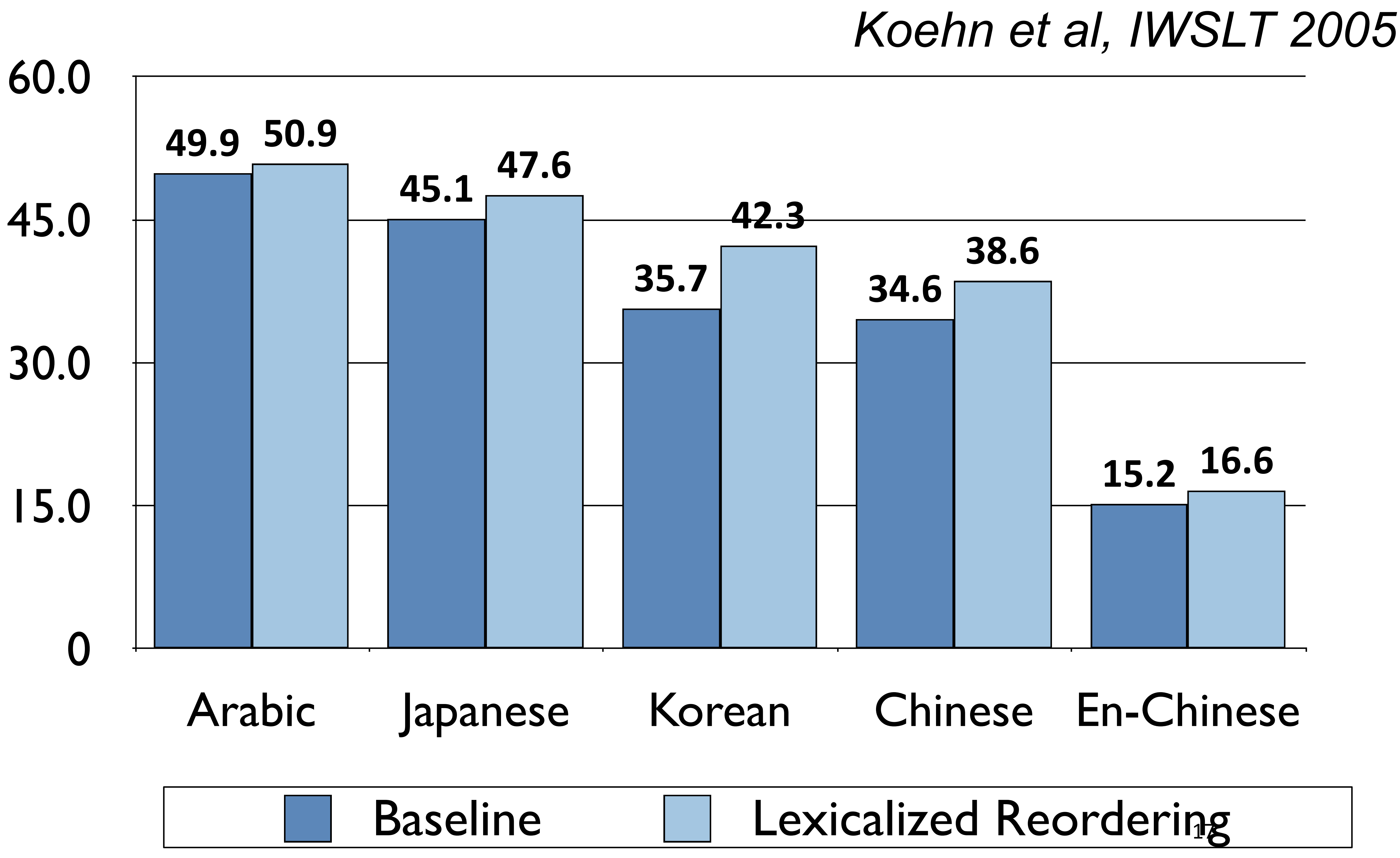
Source	**Translation**	p(m|e,f)	p(s|e,f)	p(d|e,f)
natuerlich	of course	0.52	0.08	0.40
natuerlich	naturally	0.42	0.10	0.48
natuerlich	of course ,	0.50	0.001	0.499
natuerlich	, of course	0.27	0.17	0.56

Discussion: Is this better?

- Do you think that this is a more sensible reordering model than the distance-based one?
- How could you determine if it is better or not?
- What do you think that it still lacks?

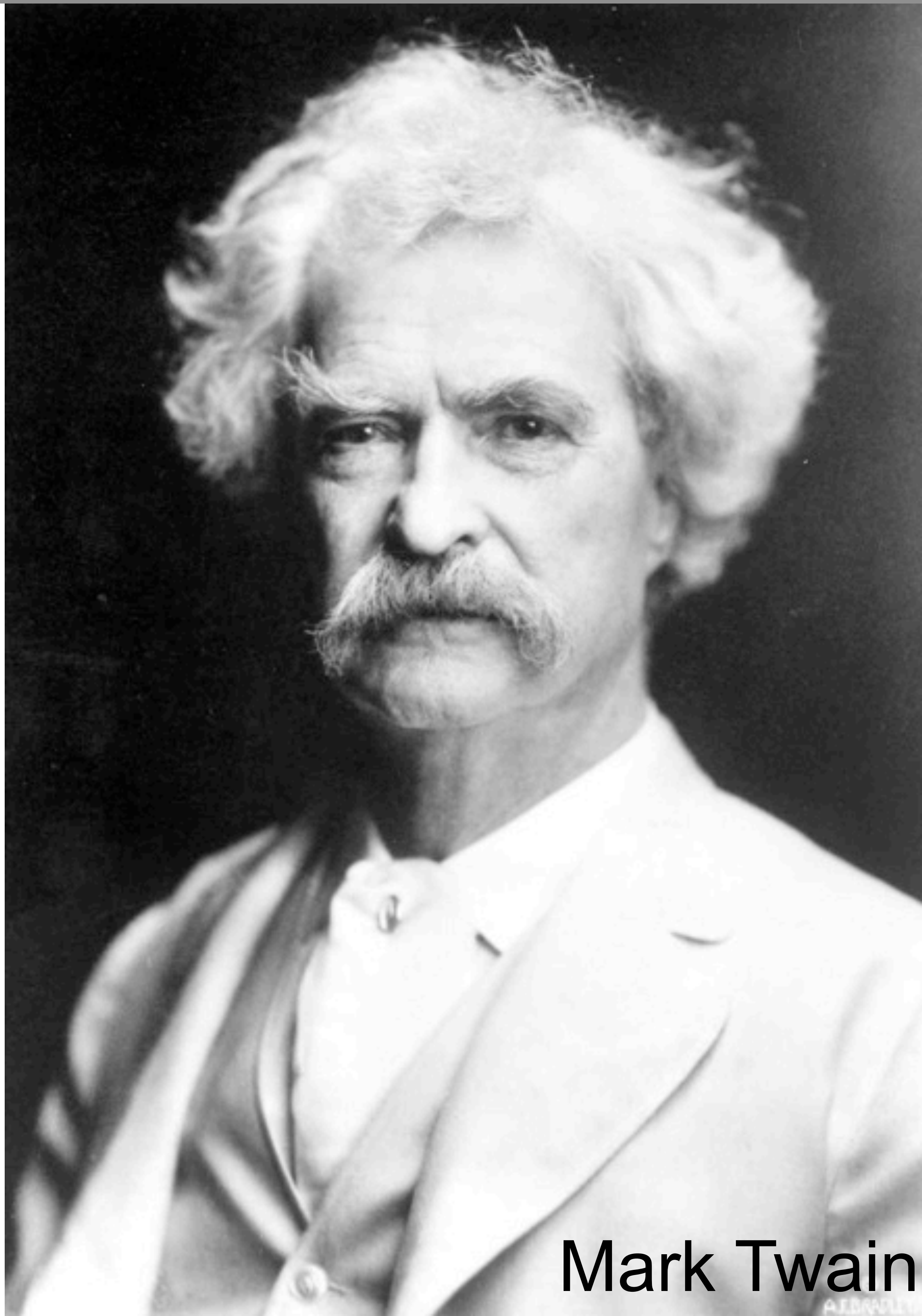
(Discuss with your neighbor)

Empirically, yes!



The Awful German Language

“The Germans have another kind of parenthesis, which they make by splitting a verb in two and putting half of it at the beginning of an exciting chapter and the OTHER HALF at the end of it. Can any one conceive of anything more confusing than that? These things are called ‘separable verbs.’ The wider the two portions of one of them are spread apart, the better the author of the crime is pleased with his performance.”



Mark Twain

German verbs

Ich	werde	Ihnen	den	Report	aushaendigen	.
I	will	to_you	the	report	pass_on	.

German verbs

Ich	werde	Ihnen	den	Report	aushaendigen	.
I	will	to_you	the	report	pass_on	.

Ich	werde	Ihnen	die	entsprechenden	Anmerkungen	aushaendigen	.
I	will	to_you	the	corresponding	comments	pass_on	.

German verbs

Ich werde Ihnen den Report aushaendigen .
I will to_you the report pass_on .

Ich werde Ihnen die entsprechenden Anmerkungen aushaendigen .
I will to_you the corresponding comments pass_on .

Ich werde Ihnen die entsprechenden Anmerkungen am Dienstag aushaendigen
I will to_you the corresponding comments on Tuesday pass_on

German free word order

The finite verb always appears in 2nd position, but ***Any constituent*** (not just the subject) can appear in the 1st position

I will to_you the report pass_on

To_you will I the report pass_on

The report will I to_you pass_on

German verbs

Main clause

Ich	werde	Ihnen	den	Report	aushaendigen	,
I	will	to_you	the	report	pass_on	,

German verbs

Main clause

Ich	werde	Ihnen	den	Report	aushaendigen	,
I	will	to_you	the	report	pass_on	,

Subordinate clause

damit	Sie	den	eventuell	uebernehmen	koennen	.
so_that	you	it	perhaps	adopt	can	.

Collins' Motivation

Phrase-based models have an overly simplistic way of handling different word orders.

We can describe the linguistic differences between different languages.

Collins defines a set of 6 simple, linguistically motivated rules, and demonstrates that they result in significant translation improvements.

Collins’ Pre-ordering Model

Step 1: Reorder the source language

Ich werde Ihnen den Report aushaendigen ,
damit Sie den eventuell uebernehmen koennen .

Collins' Pre-ordering Model

Step 1: Reorder the source language

Ich werde Ihnen den Report aushaendigen ,
damit Sie den eventuell uebernehmen koennen .

⇨ Ich werde aushaendigen Ihnen den Report ,
damit Sie koennen uebernehmen den eventuell .

Collins' Pre-ordering Model

Step 1: Reorder the source language

Ich werde Ihnen den Report aushaendigen ,
damit Sie den eventuell uebernehmen koennen .

Ich werde aushaendigen Ihnen den Report ,
damit Sie koennen uebernehmen den eventuell .

(I will pass_on to_you the report, so_that you can adopt it perhaps .)

Collins' Pre-ordering Model

Step 1: Reorder the source language

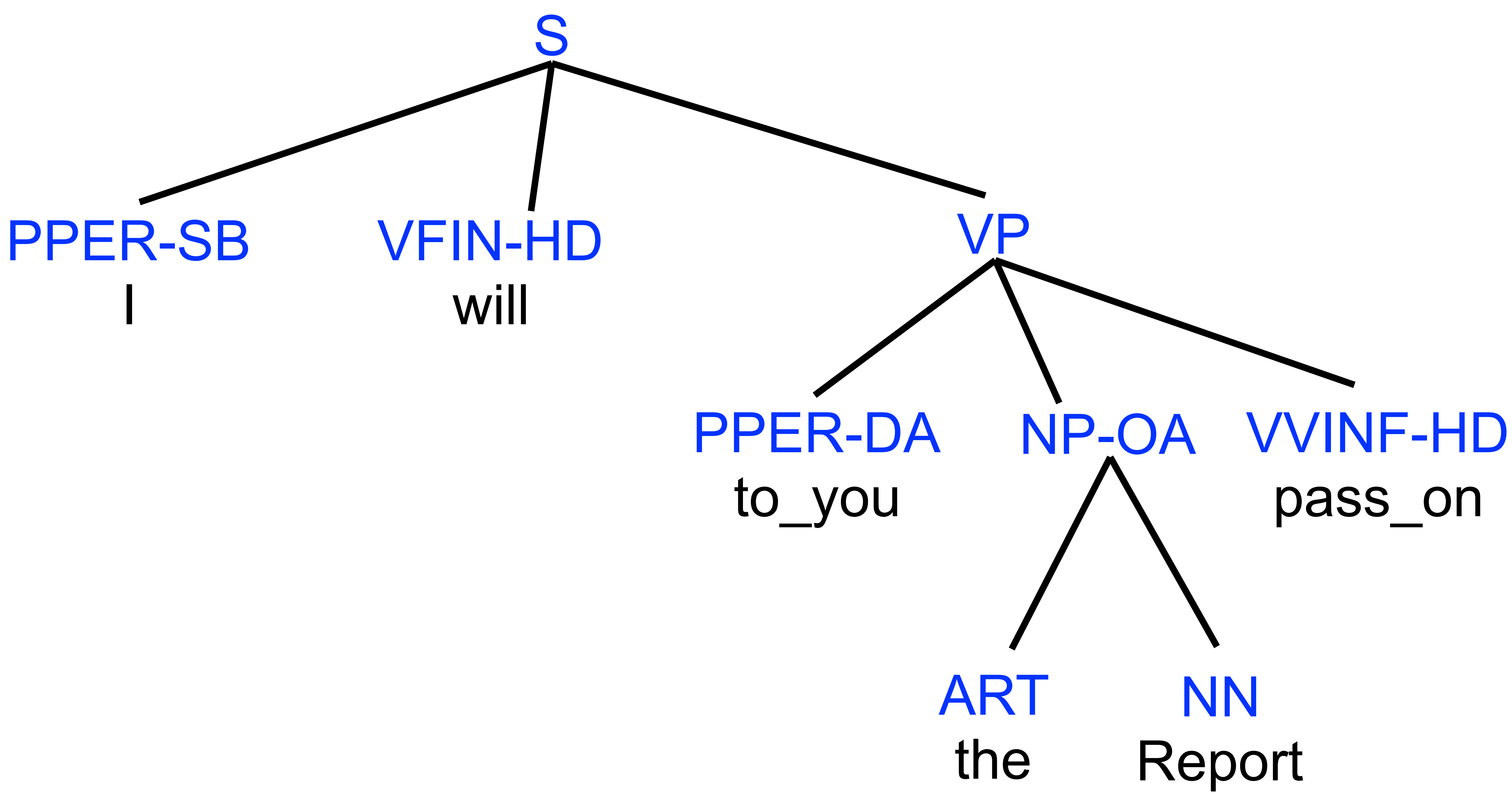
Ich werde Ihnen den Report aushaendigen ,
damit Sie den eventuell uebernehmen koennen .

Ich werde aushaendigen Ihnen den Report ,
damit Sie koennen uebernehmen den eventuell .

(I will pass_on to_you the report, so_that you can adopt it perhaps .)

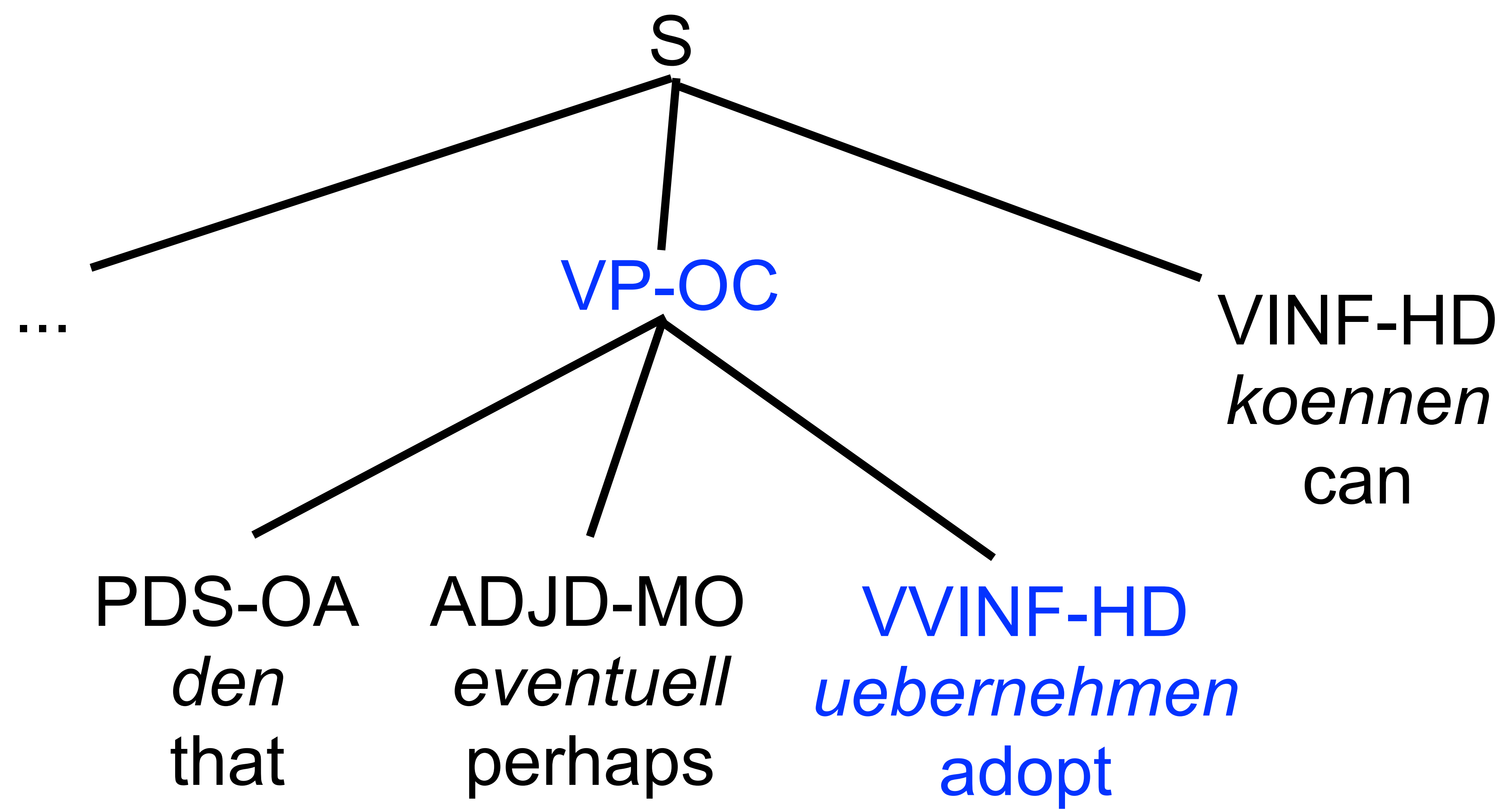
Step 2: Apply the phrase-based machine translation pipeline to the reordered input.

Example Parse Tree



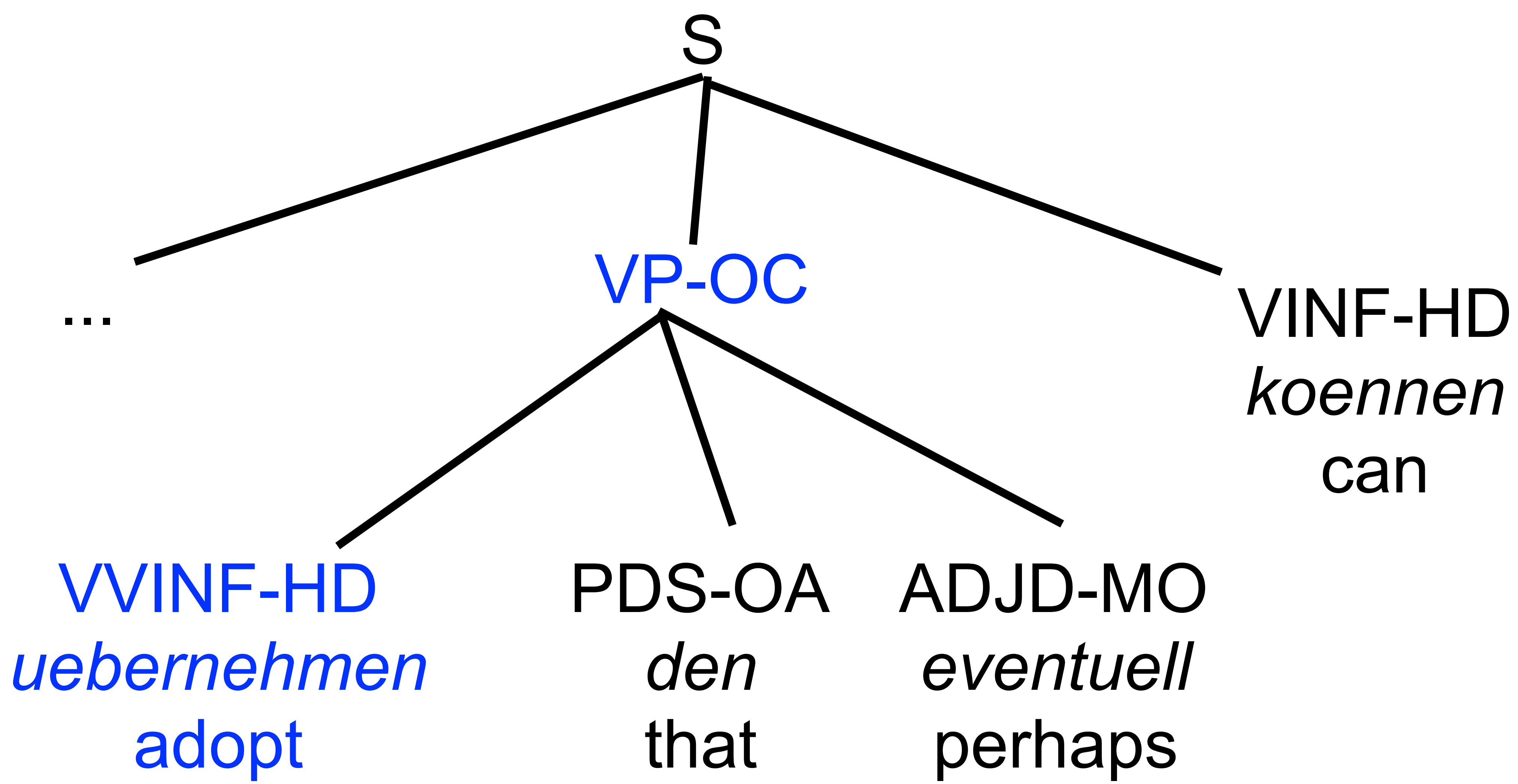
Clause Restructuring

Rule 1: **Verbs are initial in VPs**
Within a VP, move the head to the initial position



Clause Restructuring

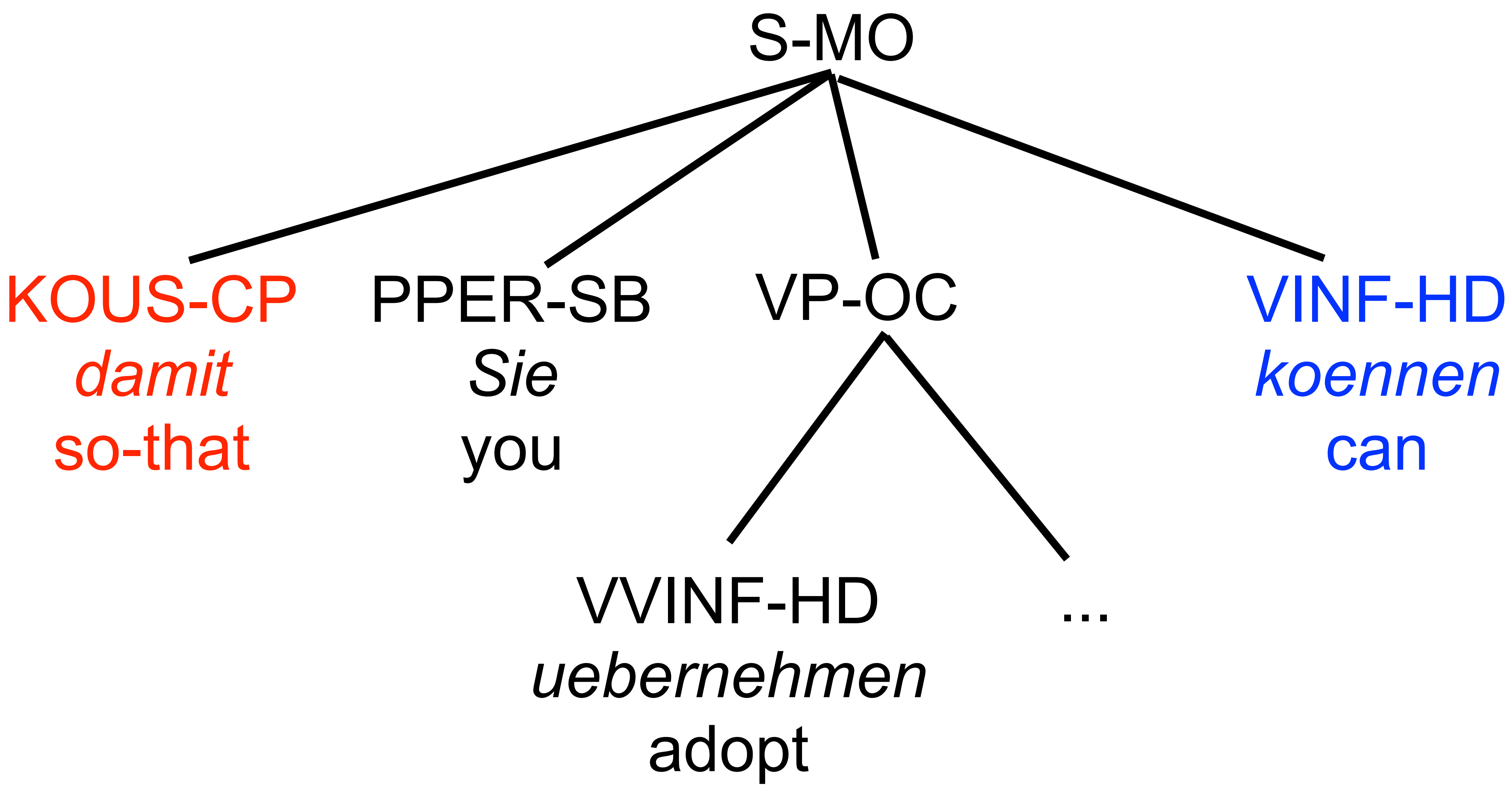
Rule 1: **Verbs are initial in VPs**
Within a VP, move the head to the initial position



Clause Restructuring

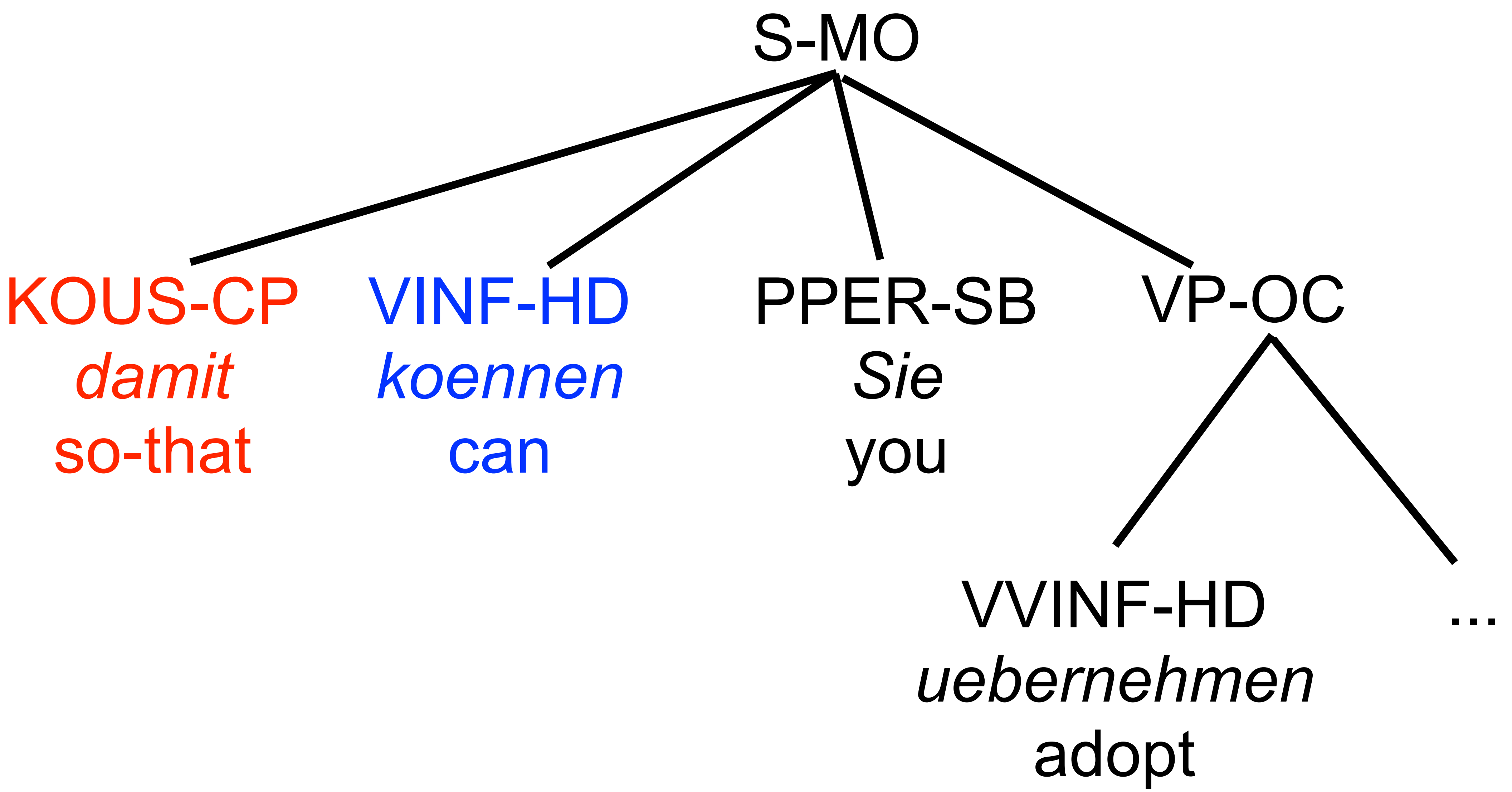
Rule 2: **Verbs follow complementizers**

In a subordinated clause mote the head of the clause to follow the complementizer



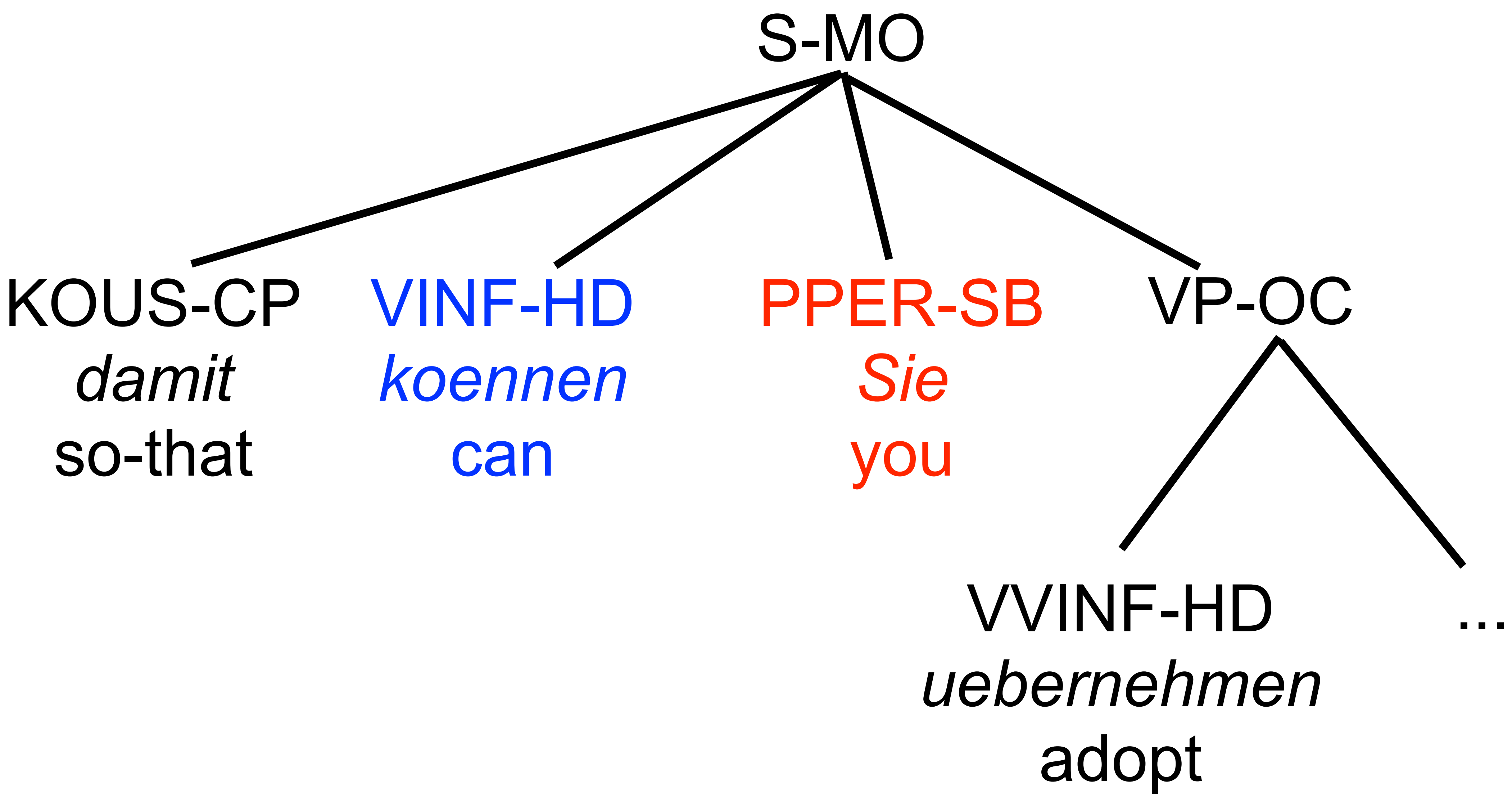
Clause Restructuring

Rule 2: **Verbs follow complementizers**
In a subordinated clause mote the head of the clause to follow the complementizer



Clause Restructuring

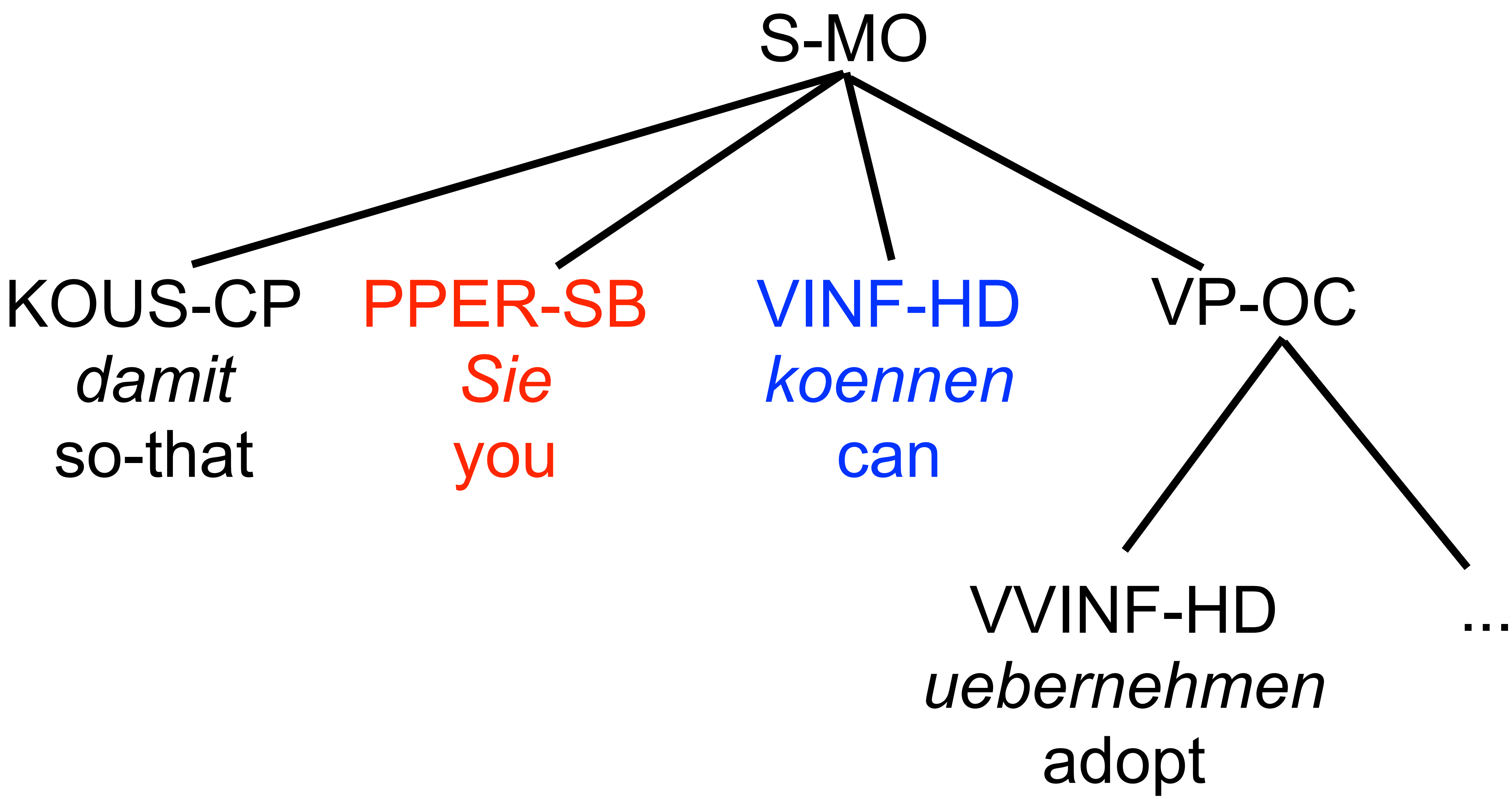
Rule 3: **Move subject**
The subject is moved to directly precede the head of the clause



Clause Restructuring

Rule 3: **Move subject**

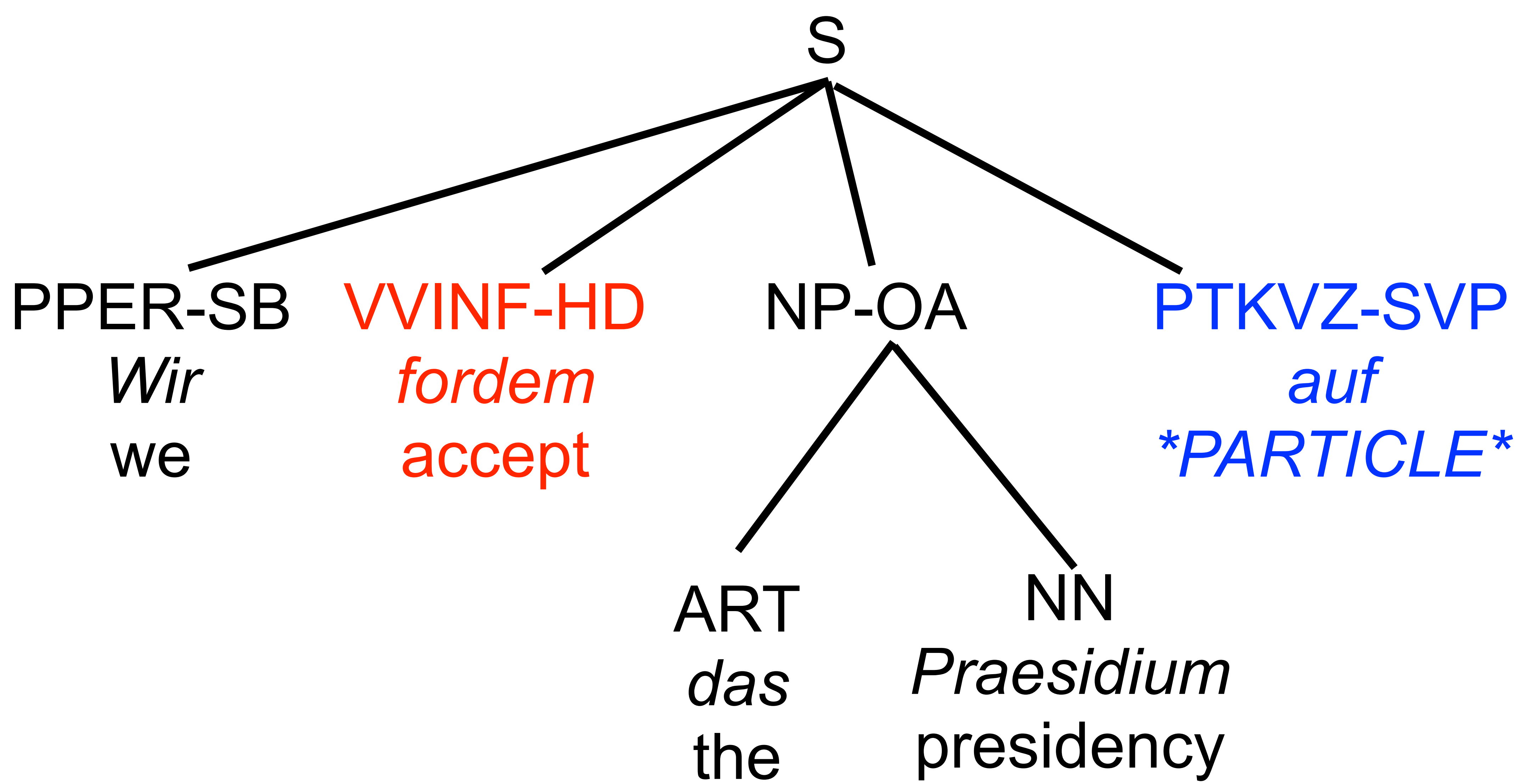
The subject is moved to directly precede the head of the clause



Clause Restructuring

Rule 4: **Particles**

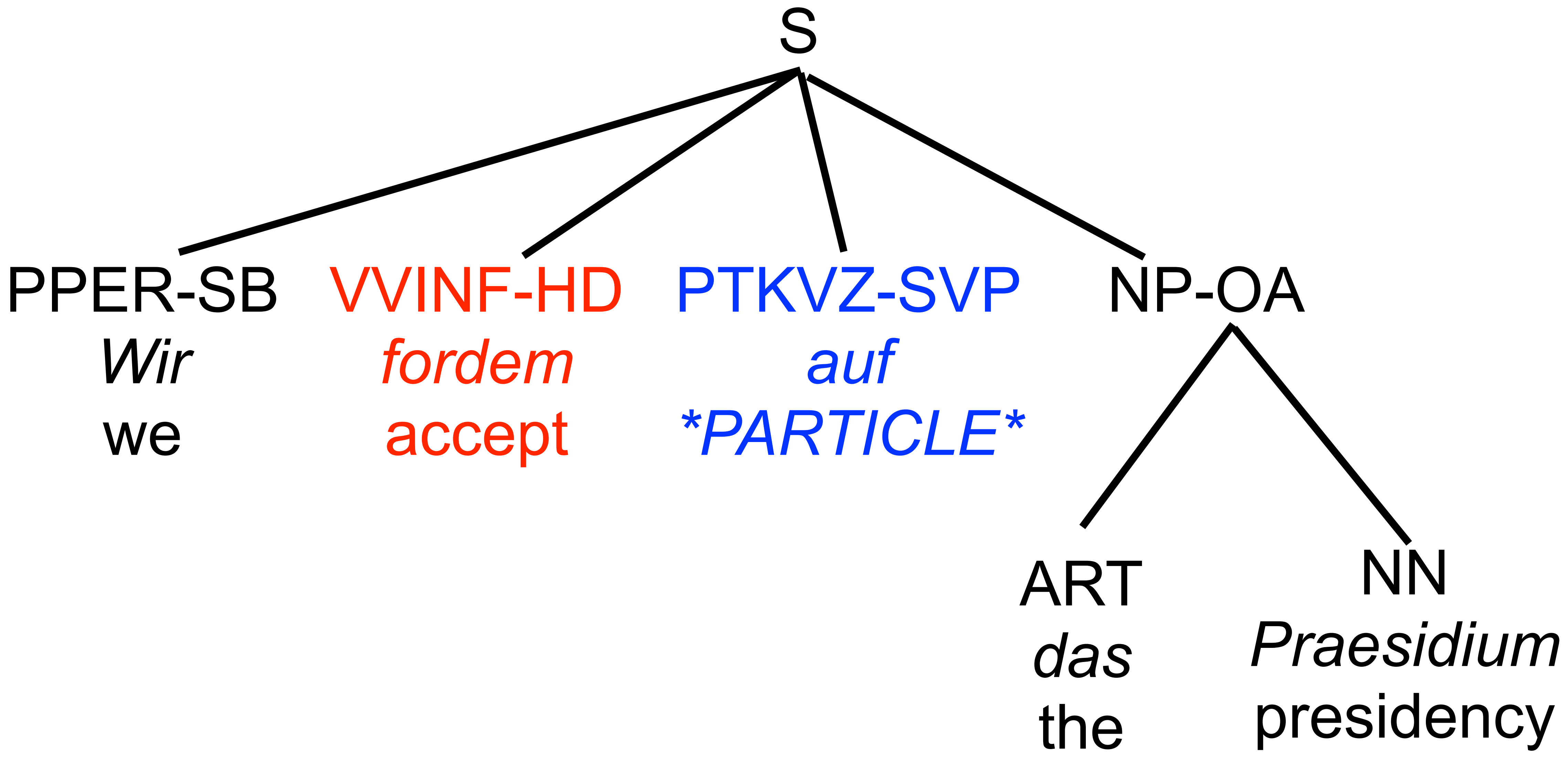
In verb particle constructions, the particle is moved to precede the finite verb



Clause Restructuring

Rule 4: **Particles**

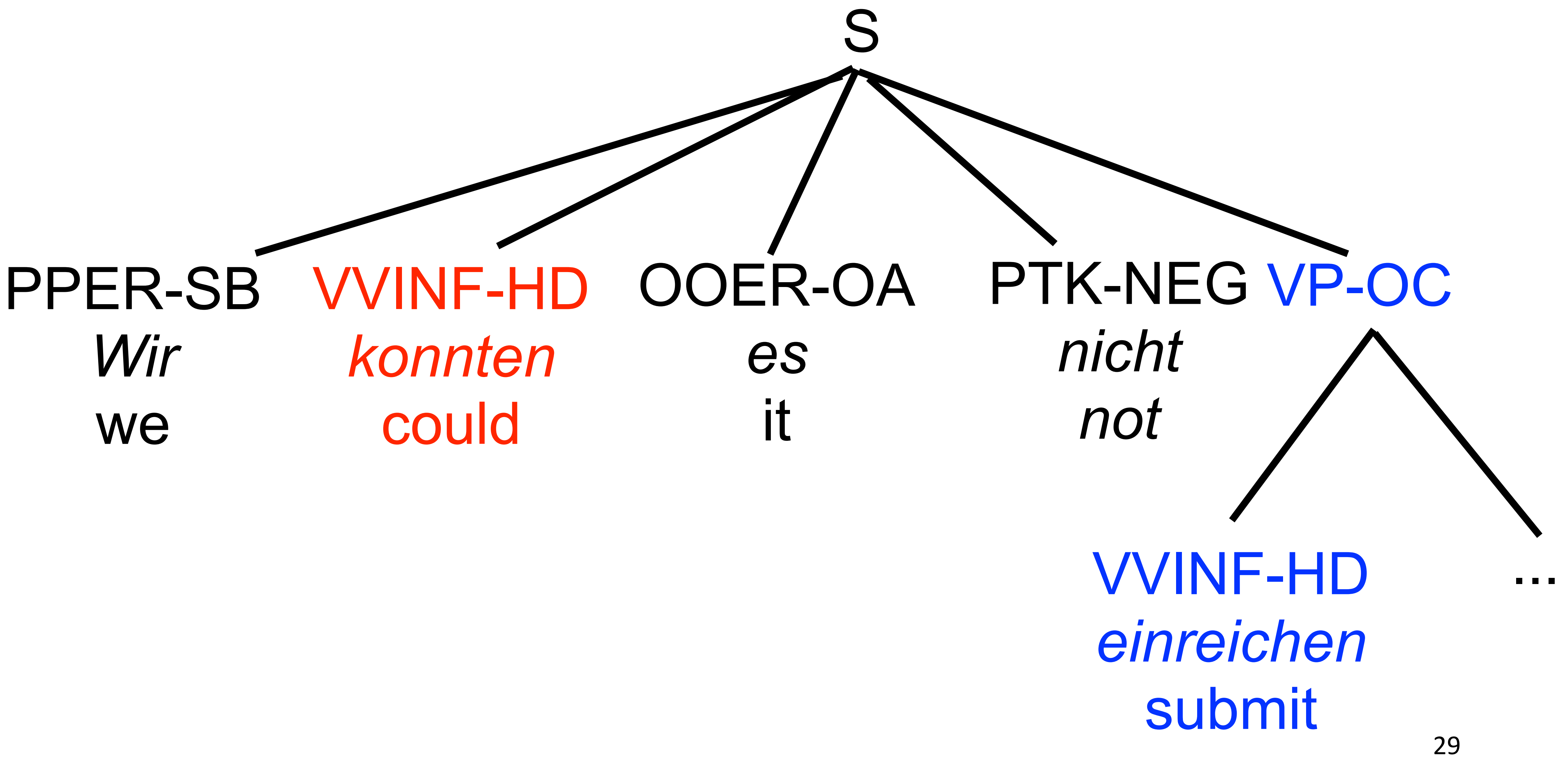
In verb particle constructions, the particle is moved to precede the finite verb



Clause Restructuring

Rule 5: **Infinitives**

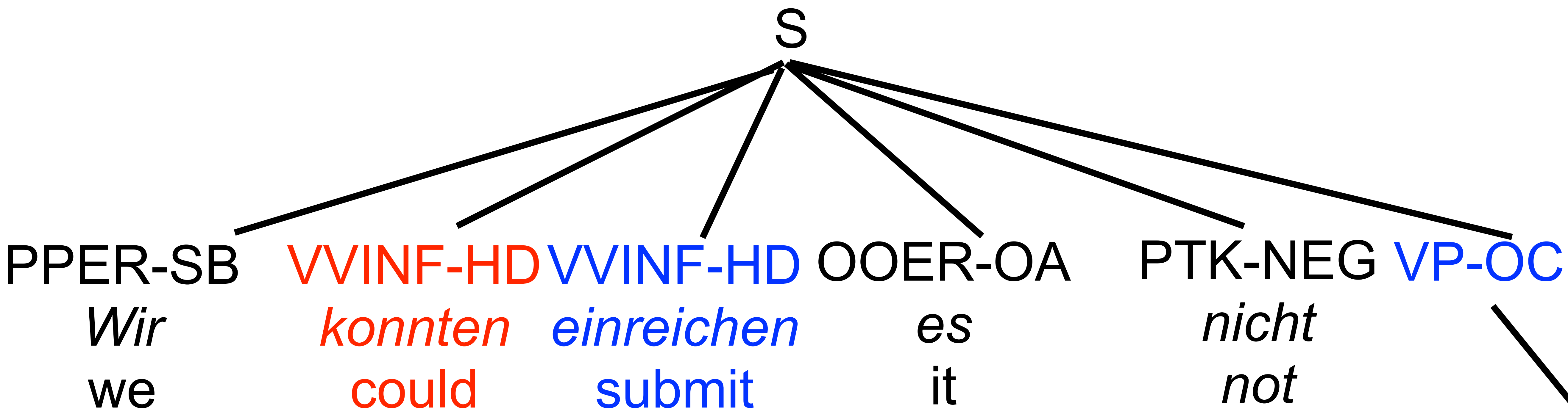
Infinitives are moved to directly follow the finite verb within a clause



Clause Restructuring

Rule 5: **Infinitives**

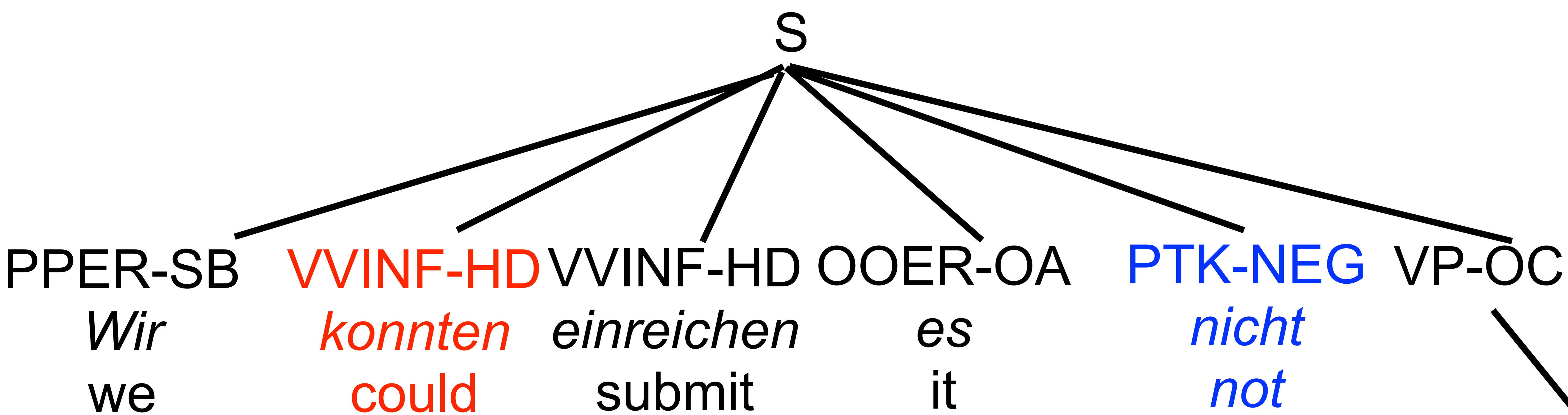
Infinitives are moved to directly follow the finite verb within a clause



Clause Restructuring

Rule 6: **Negation**

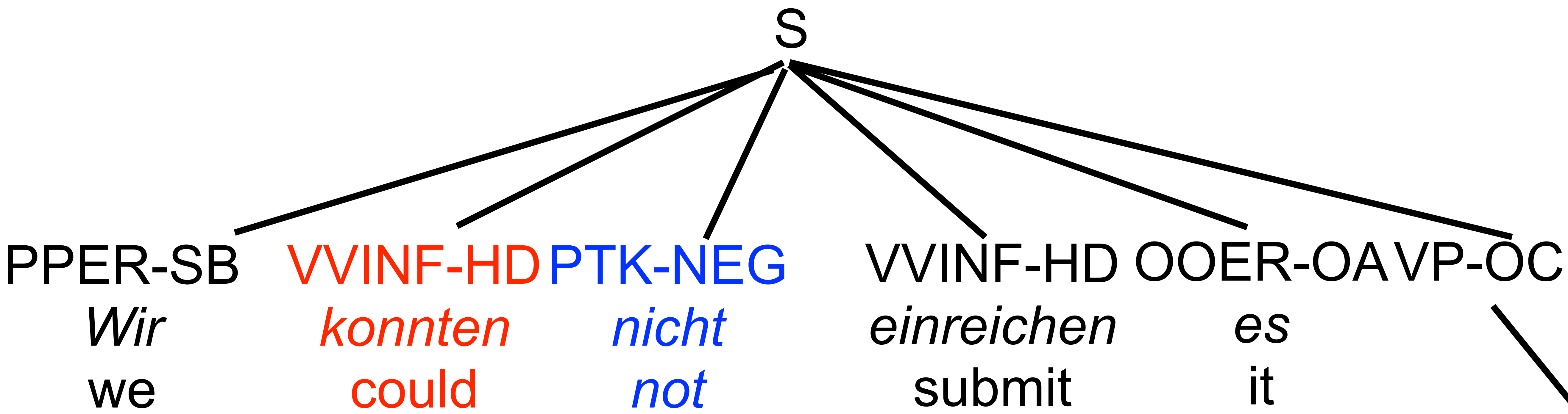
Negative particle is moved to directly follow the finite verb



Clause Restructuring

Rule 6: **Negation**

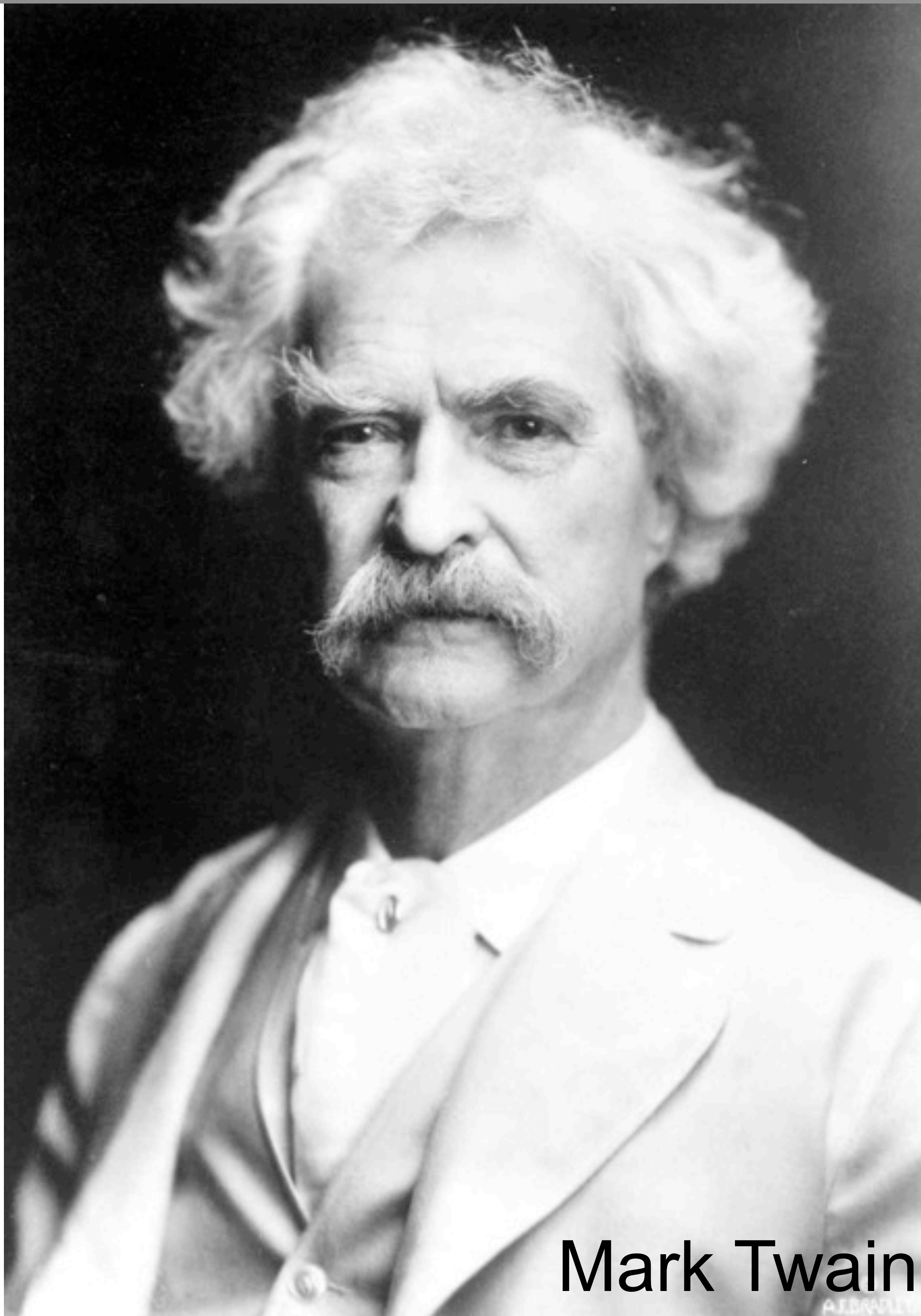
Negative particle is moved to directly follow the finite verb



A Less Awful German Language

Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen koennen.

I will to_you the report pass_on, so_that you it perhaps adopt can.



Mark Twain

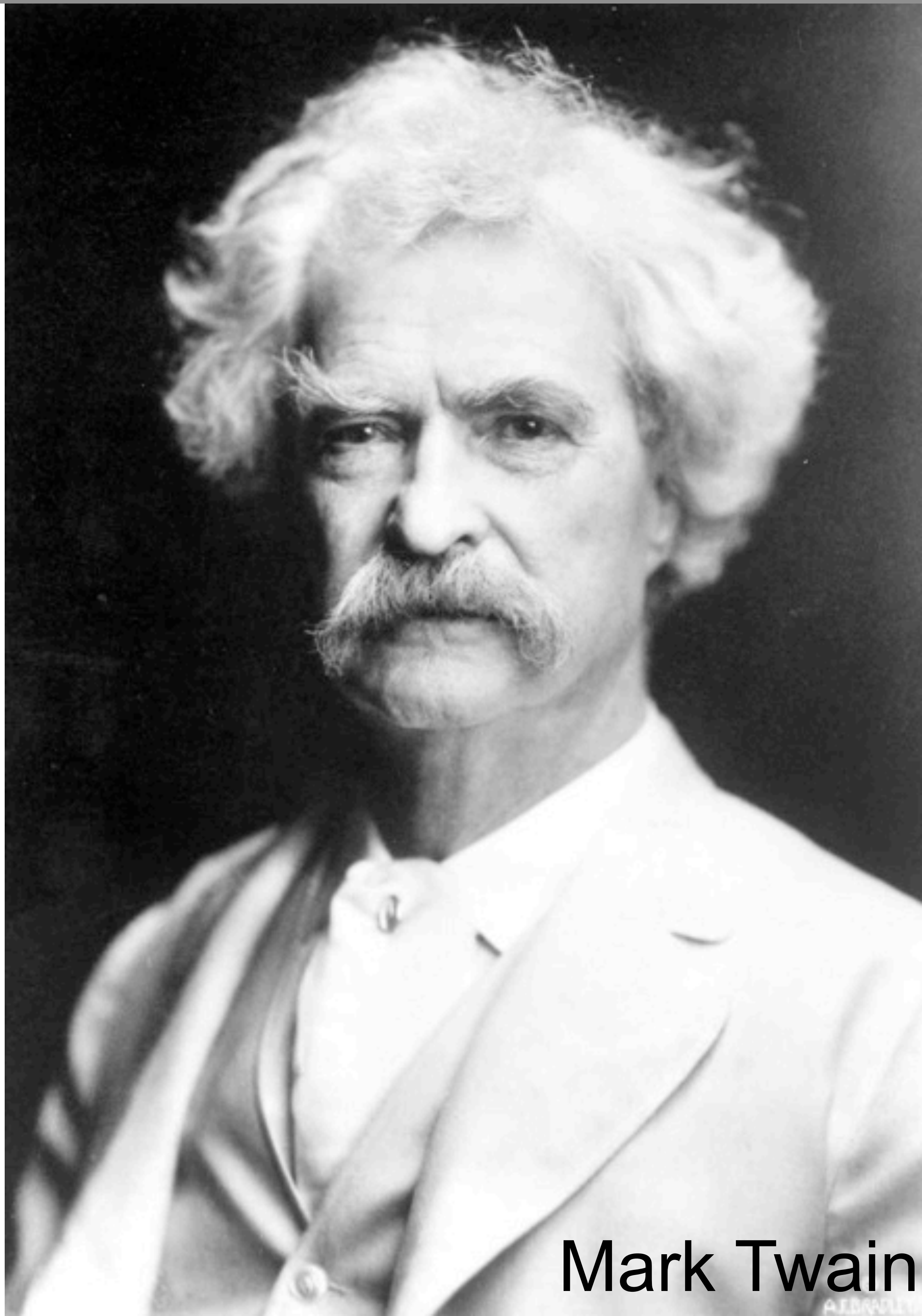
A Less Awful German Language

Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen koennen.

Ich werde aushaendigen Ihnen den Report, damit Sie koennen uebernehmen den eventuell.

I will to_you the report pass_on, so_that you it perhaps adopt can.

I will pass_on to_you the report, so_that you can adopt it perhaps .



A Less Awful German Language

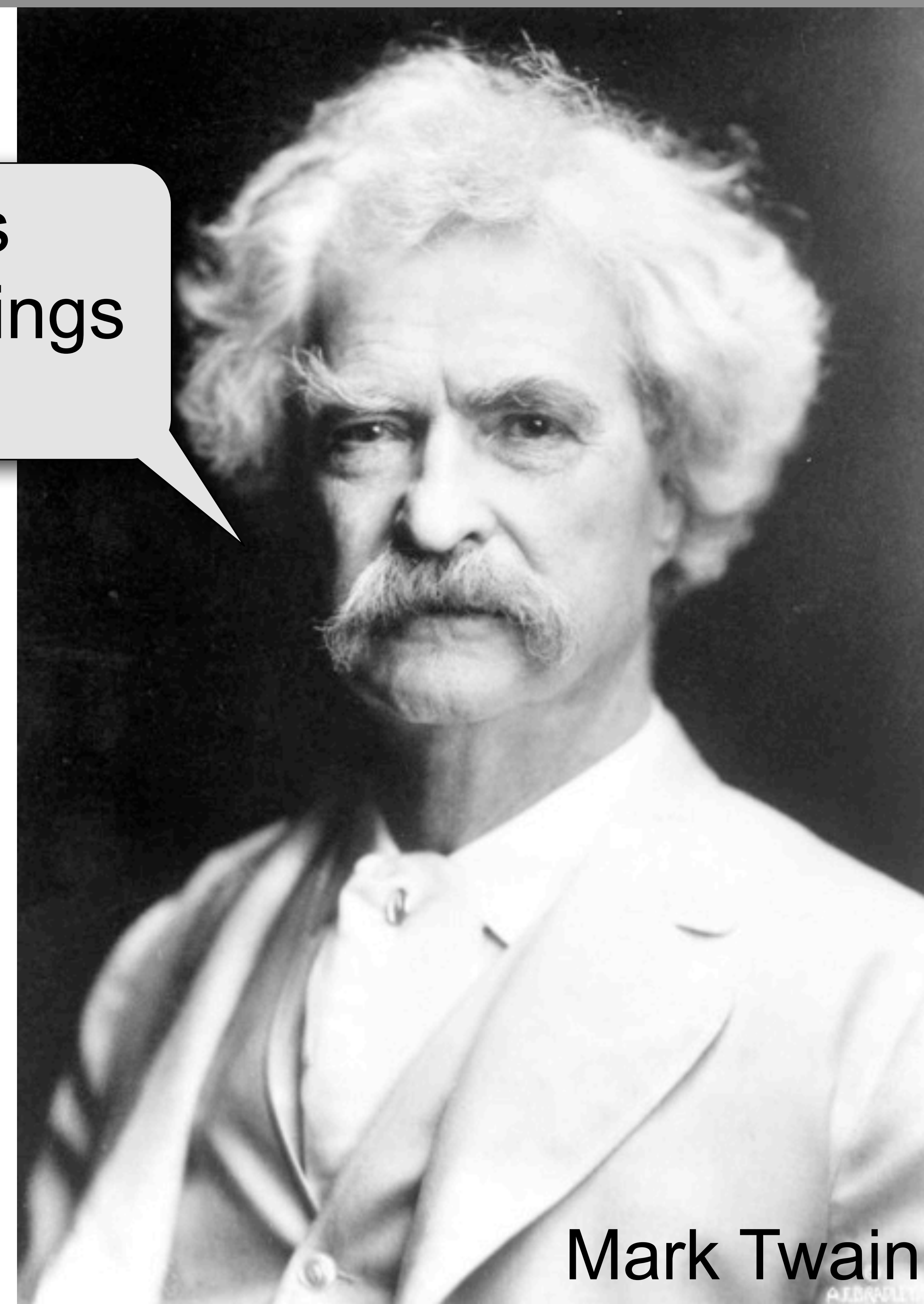
Ich werde Ihnen den Report
aushaendigen, dam
eventuell ueberneh

Now that seems less like the ravings of a madman.

Ich werde aushaendigen Ihnen
den Report, damit Sie koennen
uebernehmen den eventuell.

I will to_you the report pass_on,
so_that you it perhaps adopt can.

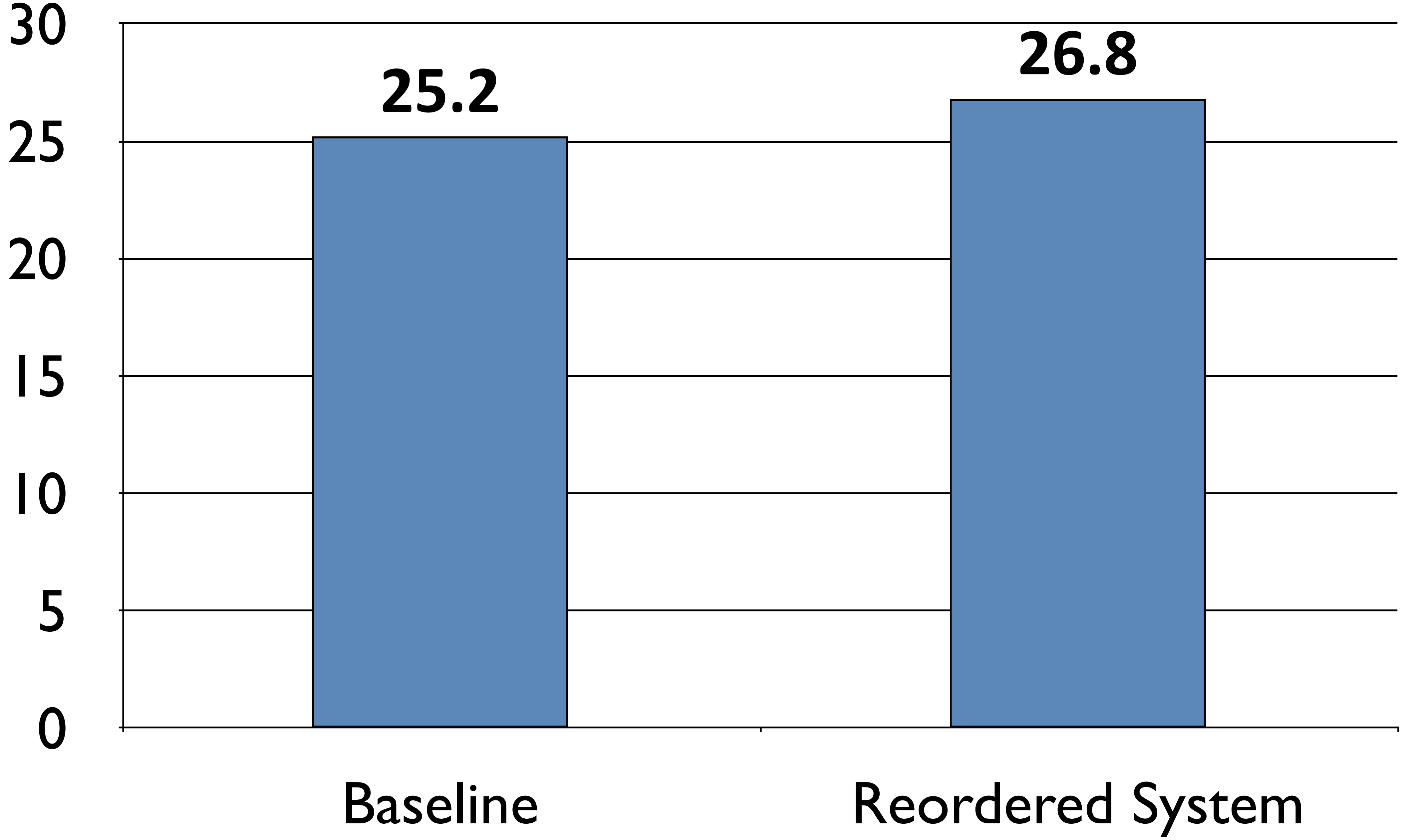
I will pass_on to_you the report,
so_that you can adopt it perhaps .



Experiments

- Parallel training data: Europarl corpus (751k sentence pairs, 15M German words, 16M English)
- Parsed German training sentences
- Reordered the German training sentences with their 6 clause reordering rules
- Trained a phrase-based model
- Parsed and reordered the German test sentences
- Translated them
- Compared against the standard phrase-based model without parsing/reordering

Bleu score increase



Significant improvement at $p<0.01$ using the sign test

Human Translation Judgments

- 100 sentences (10-20 words in length)
- Two annotators
- Judged two different versions
  - Baseline system's translation
  - Reordering system's translation
- Judgments: Worse, better or equal
- Sentences were chosen at random, systems' translations were presented in random order

Human Translation Judgments

	+	=	–
Annotator 1	40%	40%	20%
Annotator 2	44%	37%	19%

+ = reordered translation better
– = baseline better
= = equal

Examples

Reference	I think it is wrong in principle to have such measures in the European Union

I believe that it is wrong in principle to take such measures in the European Union

I believe that it is wrong in principle, such measure in the European Union to take.

Examples

Reference	I think it is wrong in principle to have such measures in the European Union
Reordered	I believe that it is wrong in principle to take such measures in the European Union
Baseline	I believe that it is wrong in principle, such measure in the European Union to take.

Examples

Reference	The current difficulties should encourage us to redouble our efforts to promote coorperation in the Euro-Mediterranean framework.

The current problems should spur us, our efforts to promote coorperation within the framework of the e-prozesses to be intensified.

The current problems should spur us to intensify our efforts to promote cooperation within the framework of the e-prozesses.

Examples

Reference	The current difficulties should encourage us to redouble our efforts to promote coorperation in the Euro-Mediterranean framework.
Baseline	The current problems should spur us, our efforts to promote coorperation within the framework of the e-prozesses to be intensified.
Reordered	The current problems should spur us to intensify our efforts to promote cooperation within the framework of the e-prozesses.

Examples

Reference	To go on subsidizing tobacco cultivation at the same time is a downright contradiction.

At the same time, continue to subsidize tobacco growing, it is quite schizophrenic.

At the same time, to continue to subsidize tobacco growing is schizophrenic.

Examples

Reference	To go on subsidizing tobacco cultivation at the same time is a downright contradiction.
Baseline	At the same time, continue to subsidize tobacco growing, it is quite schizophrenic.
Reordered	At the same time, to continue to subsidize tobacco growing is schizophrenic.

Examples

Reference	We have voted against the report by Mrs. Lalumiere for reasons that include the following:

We have voted, amongst other things, for the following reasons against the report by Mrs. Lalumiere:

We have, among other things, for the following reasons against the report by Mrs. Lalumiere voted:

Examples

Reference	We have voted against the report by Mrs. Lalumiere for reasons that include the following:
Reordered	We have voted, amongst other things, for the following reasons against the report by Mrs. Lalumiere:
Baseline	We have, among other things, for the following reasons against the report by Mrs. Lalumiere voted:

Discussion: Clause Restructuring

- Are you convinced that German-English translation has improved?
- Do you think that this is a good fit for phrase-based machine translation?
- What limitations does this method have?

(Discuss with your neighbor.)

Limitations

- Requires a parser for the source language
  - We have parsers for only a small number of languages
  - Penalizes "low resource languages"
  - Fine for translating from English into other languages
- Involves hand crafted rules
- Removes the nice language-independent qualities of statistical machine translation

Synchronous context-free grammars (SCFGs)

The Syntax Bet

- Longstanding debate about whether linguistic information can help statistical translation
- Two camps

The Syntax Bet

- Longstanding debate about whether linguistic information can help statistical translation
- Two camps

Syntax will improve translation

The Syntax Bet

- Longstanding debate about whether linguistic information can help statistical translation
- Two camps

Syntax will improve translation

Simpler data-driven models will always win

“Every time I fire a linguist my performance goes up”

- Longstanding debate about wheth
information can help statistical tr
- Two camps

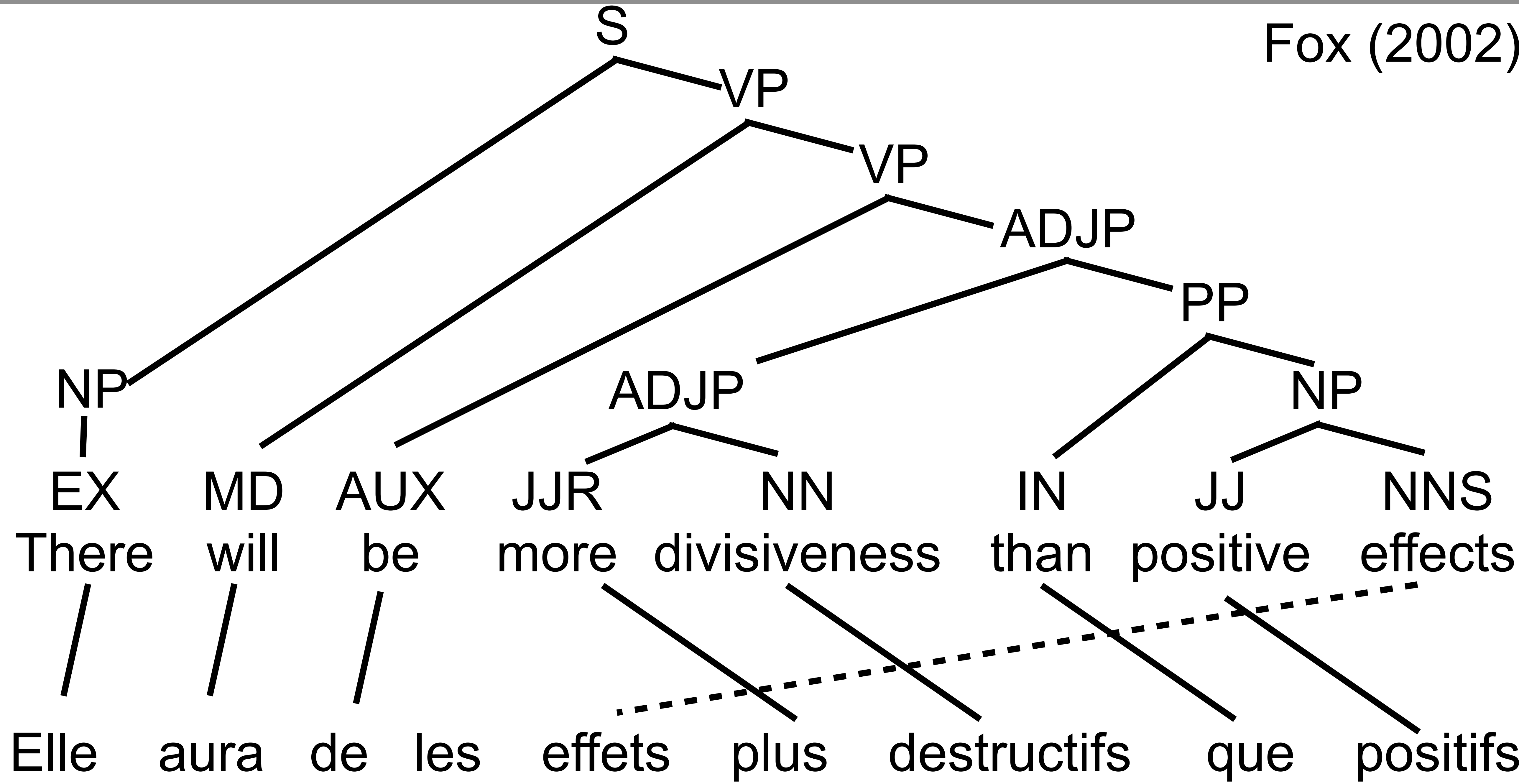
Syntax will improve translation

Simpler data-driven models will always win

Syntax is bad for translation

- The IBM Models were the dominant approach to SMT from the `90s until mid 2000s
  – Eschewed linguistic information
- A number of studies cast doubt on whether linguistic info could help SMT
  – Fox (2002) showed that "phrasal cohesion" was less common than assumed across even related languages
  – Koehn et al (2003) empirically demonstrated that syntactically motivated phrases made PBMT worse

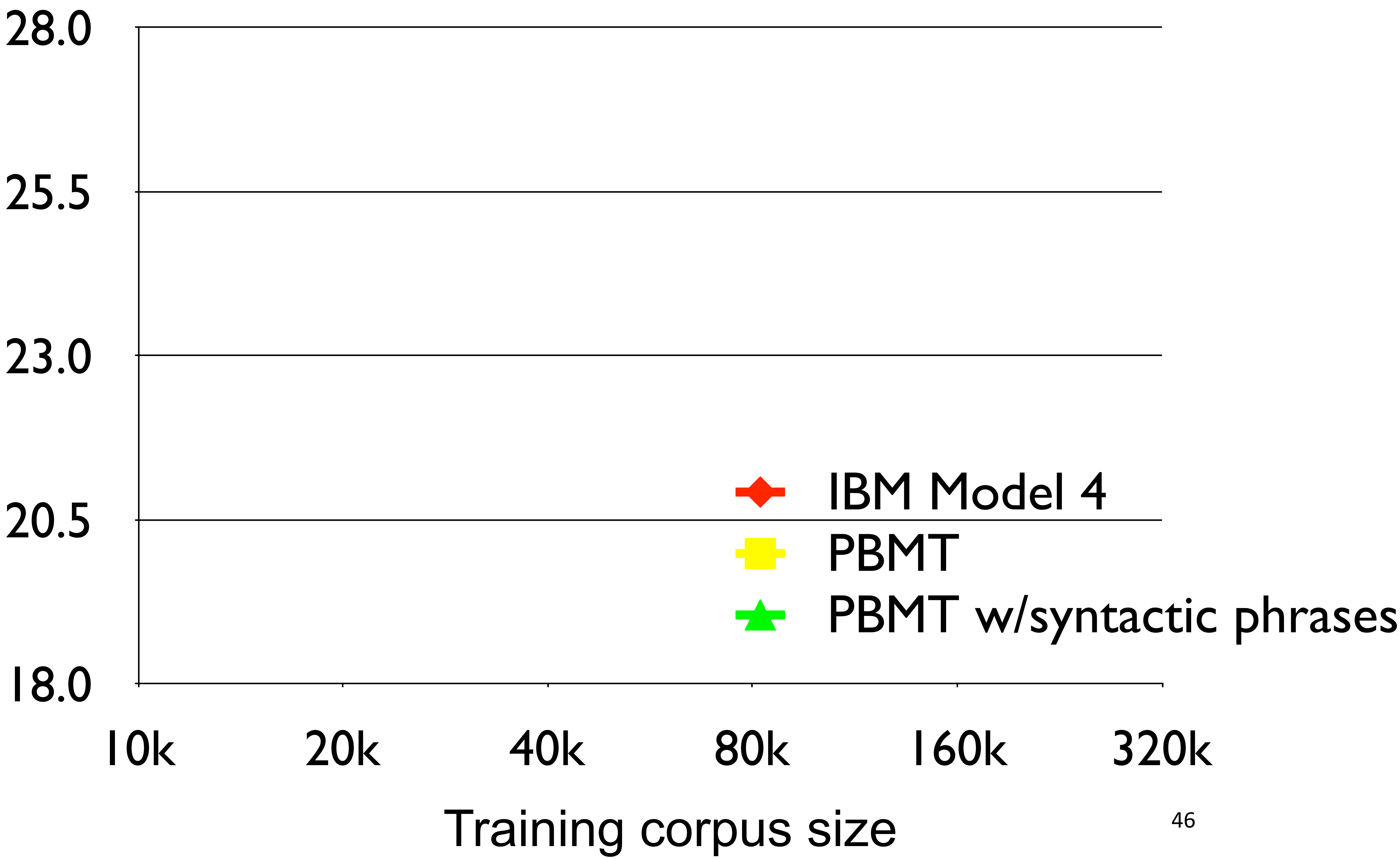
Phrases aren't coherent in bitexts



Gloss: *It will have effects more destructive than positive*

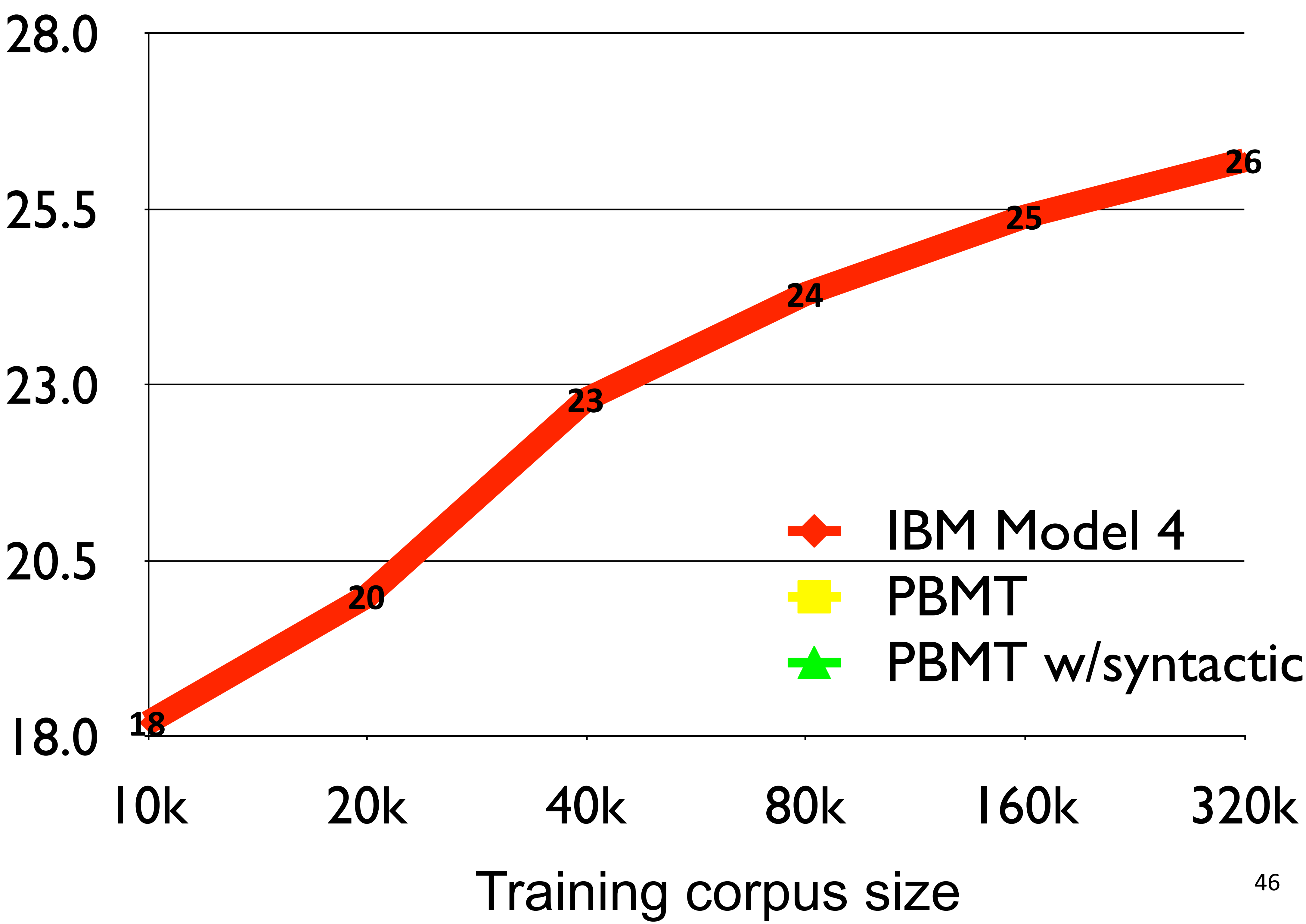
Ouch! Syntax hurts!

Koehn et al (2003)



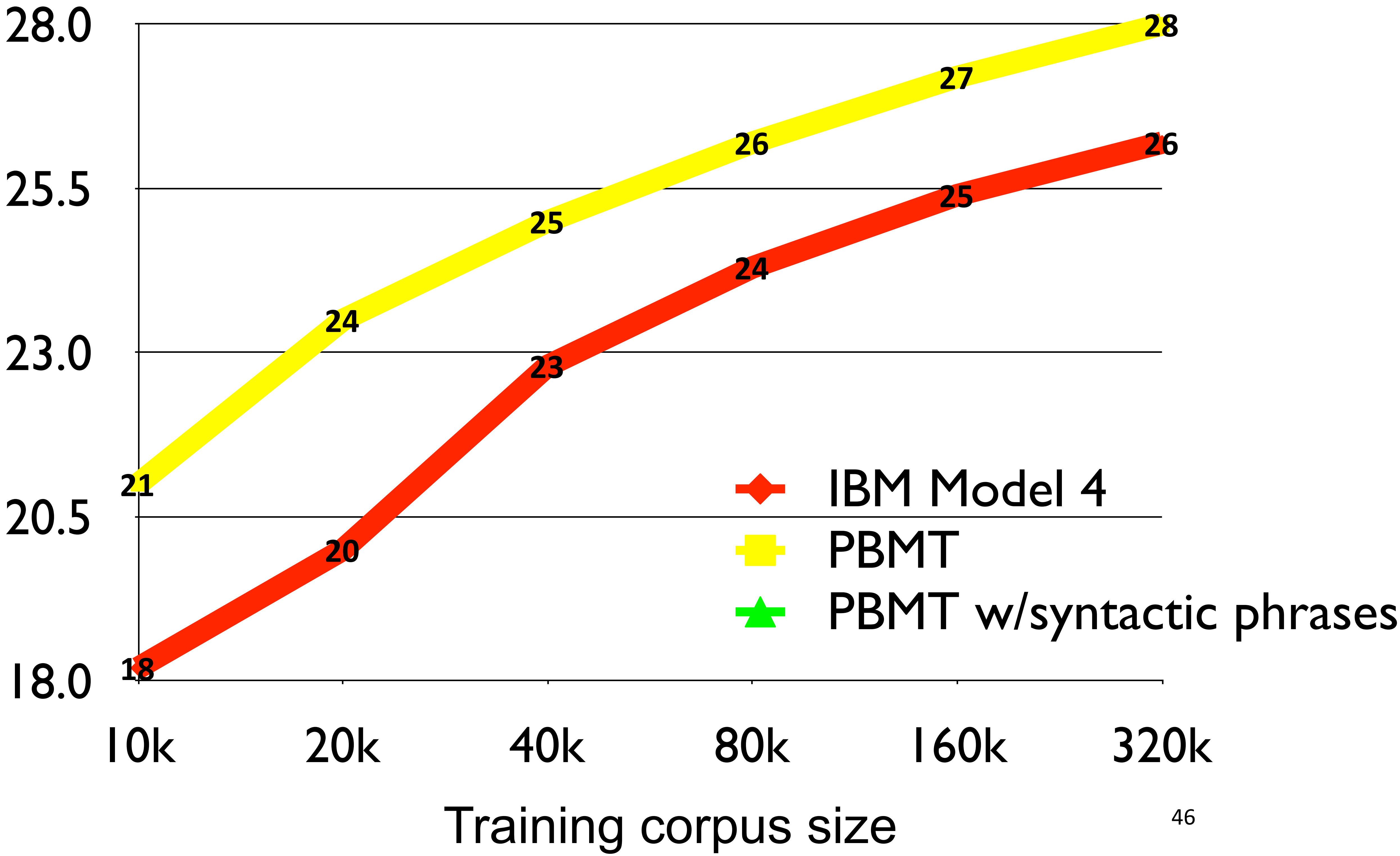
Ouch! Syntax hurts!

Koehn et al (2003)



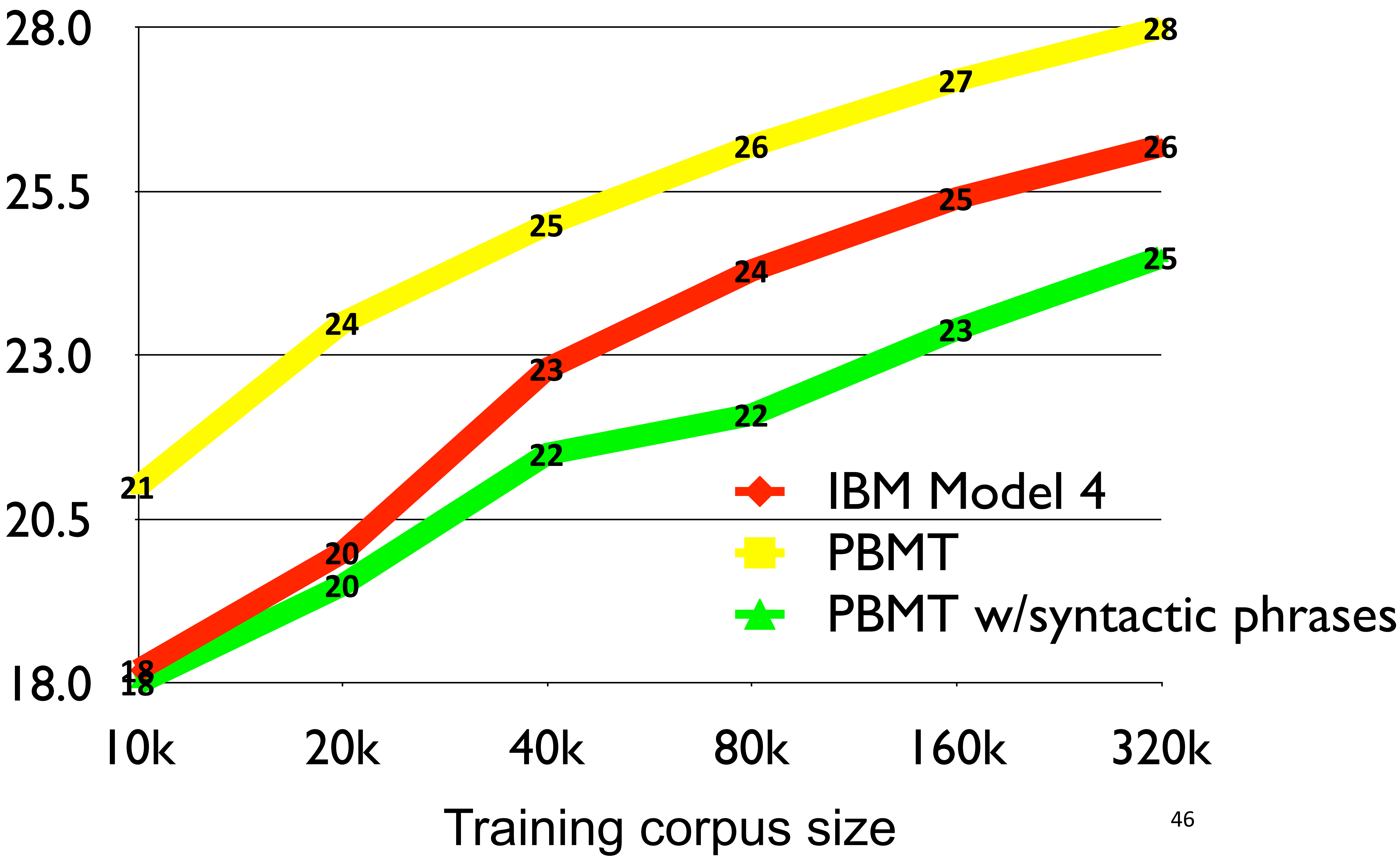
Ouch! Syntax hurts!

Koehn et al (2003)

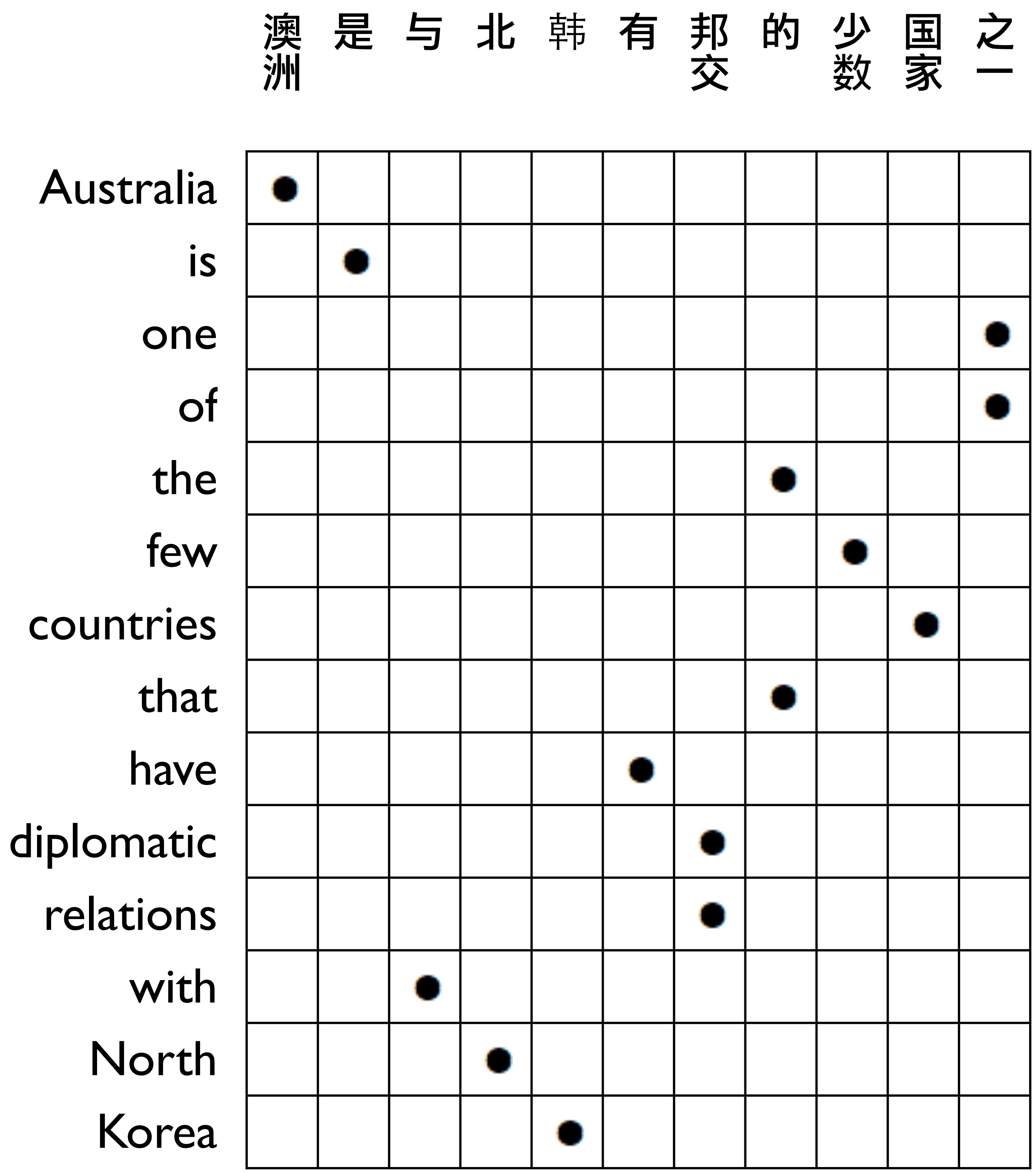


Ouch! Syntax hurts!

Koehn et al (2003)

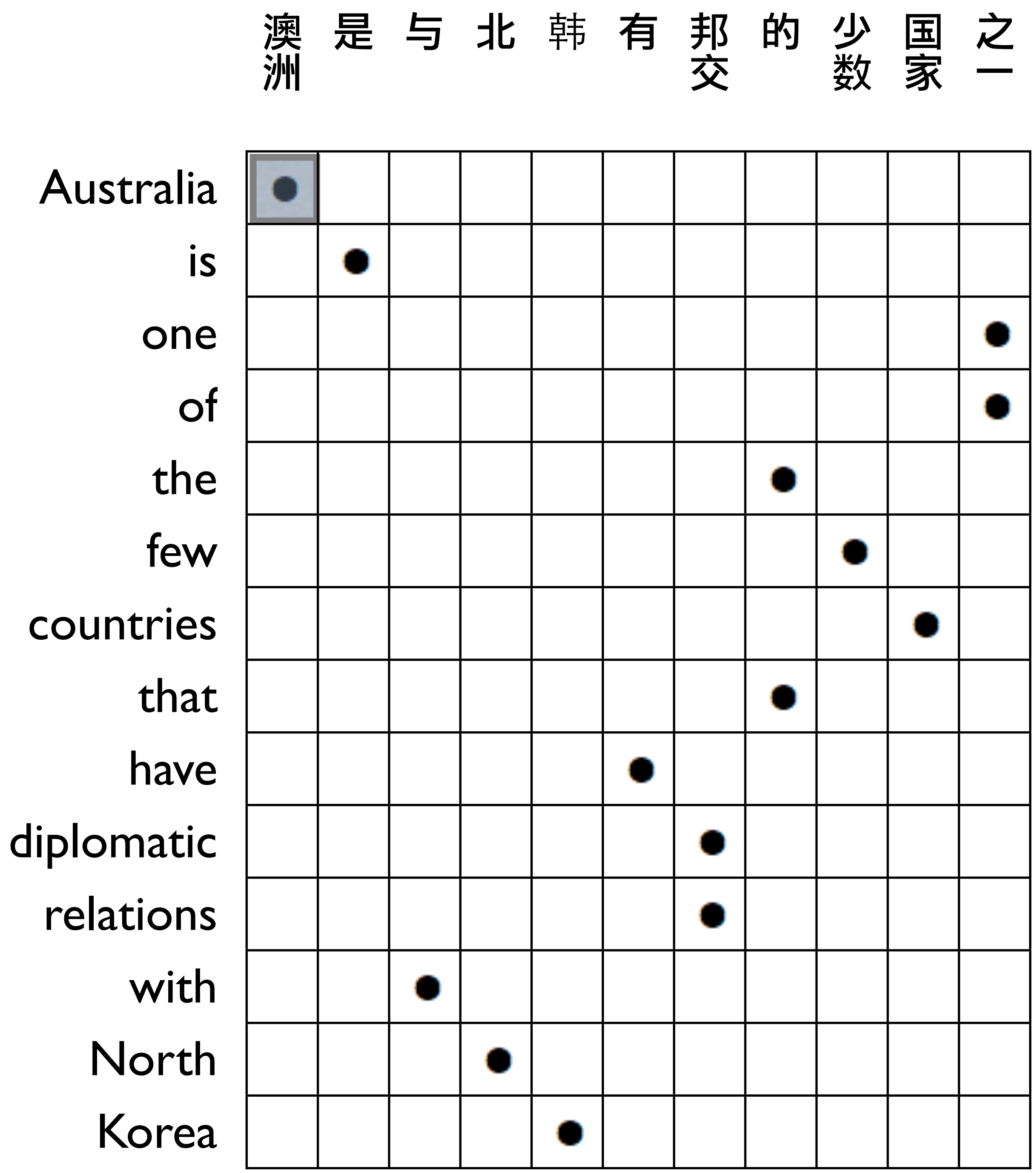


Extracting phrase pairs

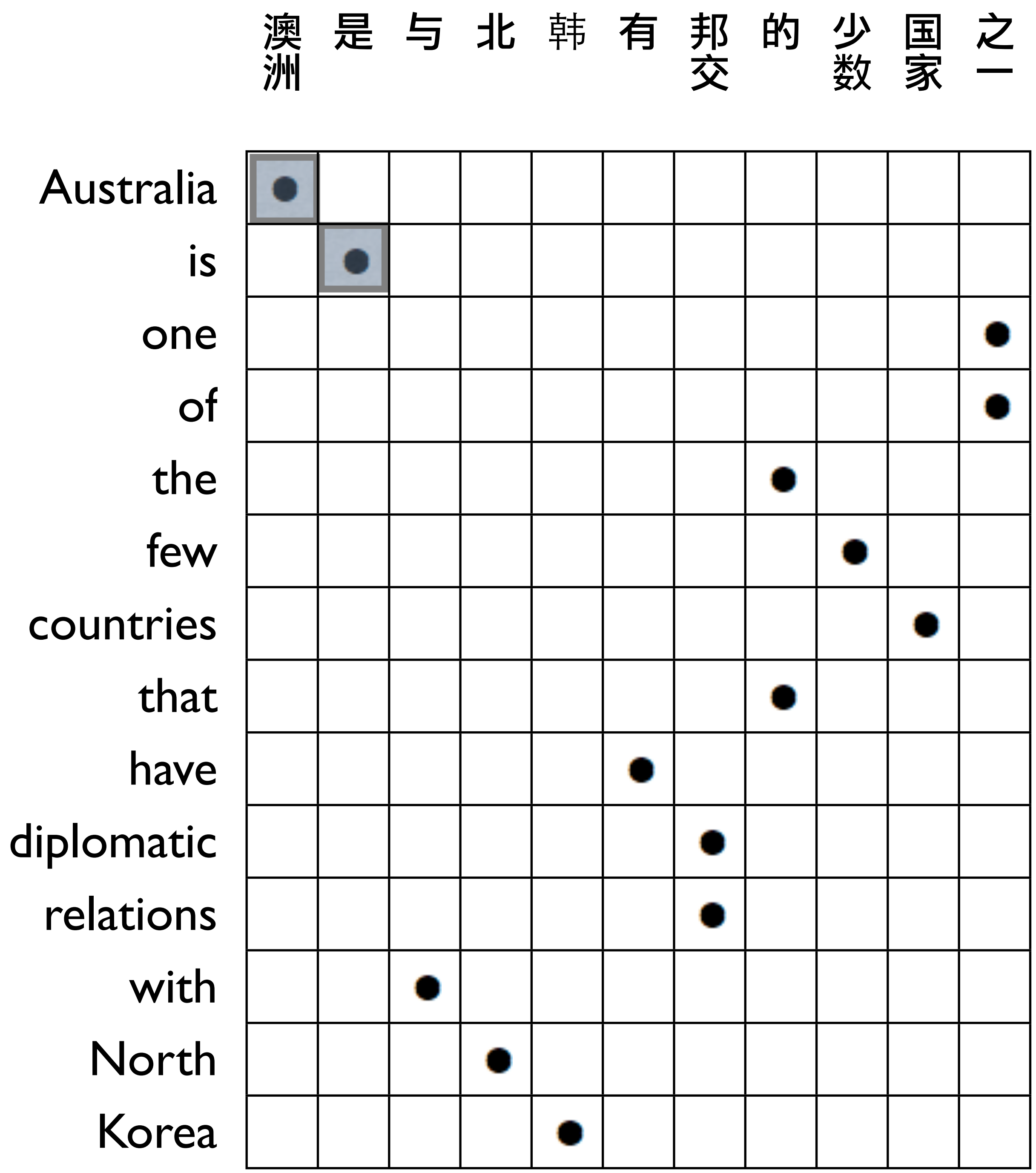


Extracting phrase pairs

澳洲, Australia



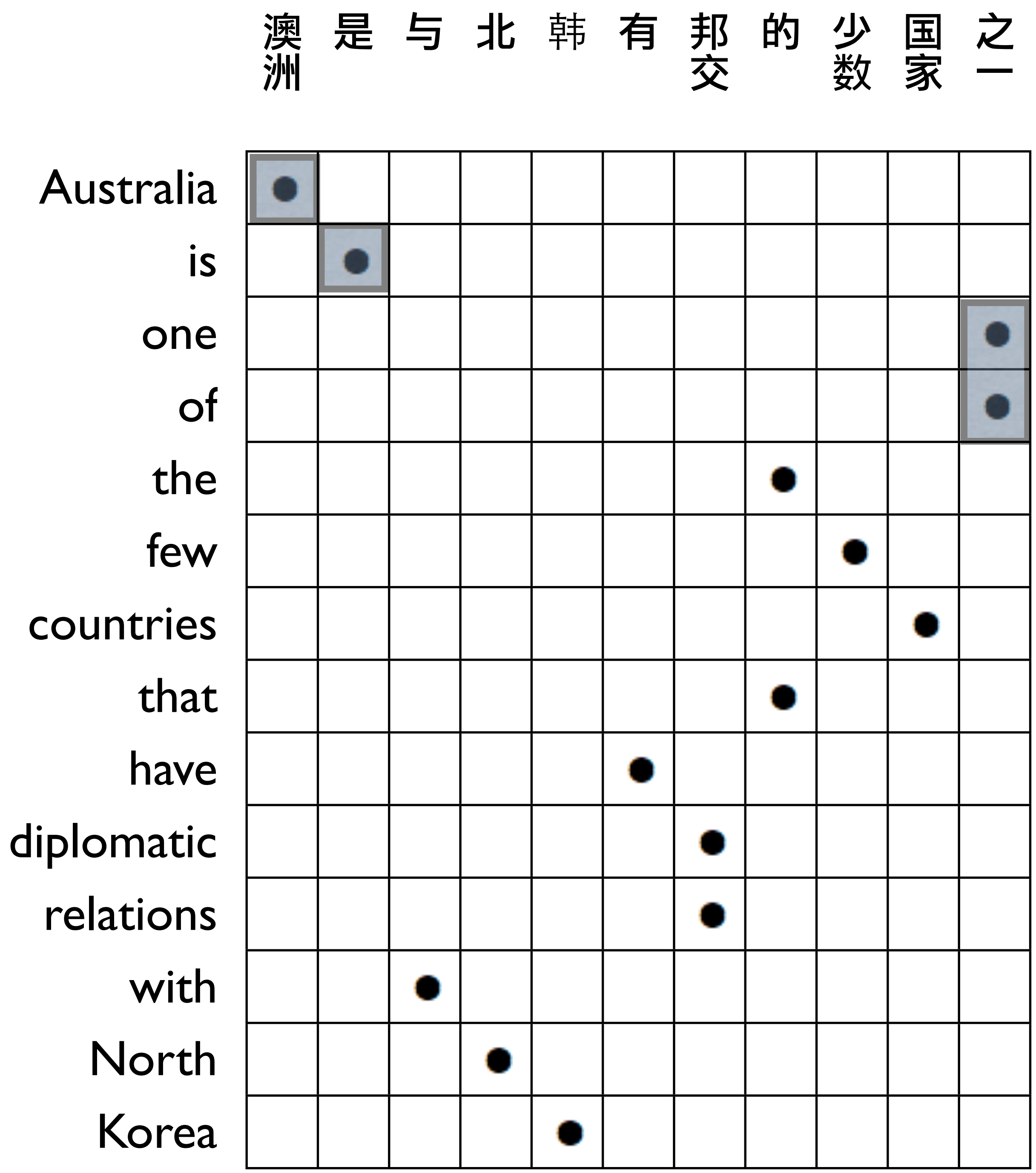
Extracting phrase pairs



澳洲, Australia

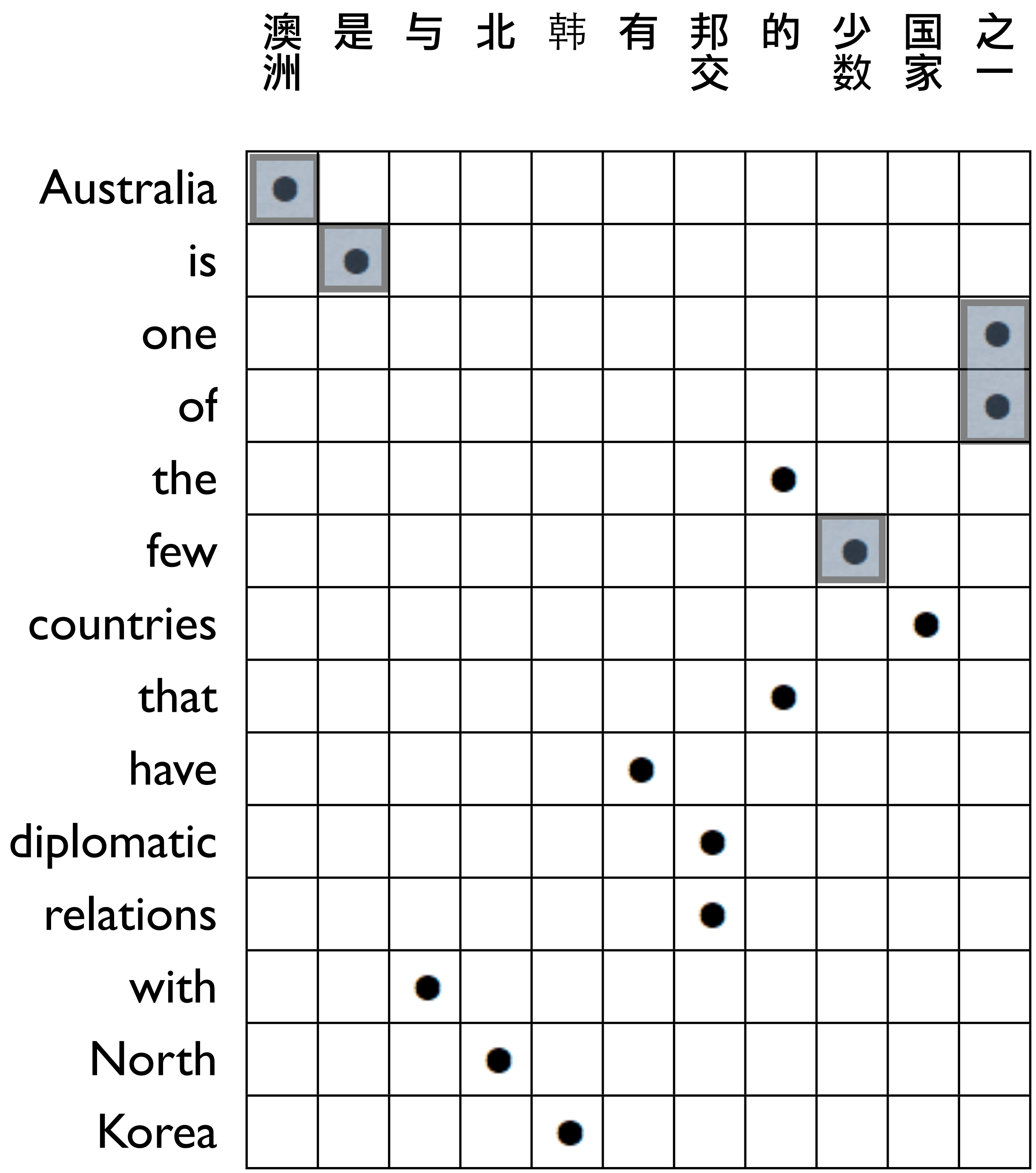
是, is

Extracting phrase pairs



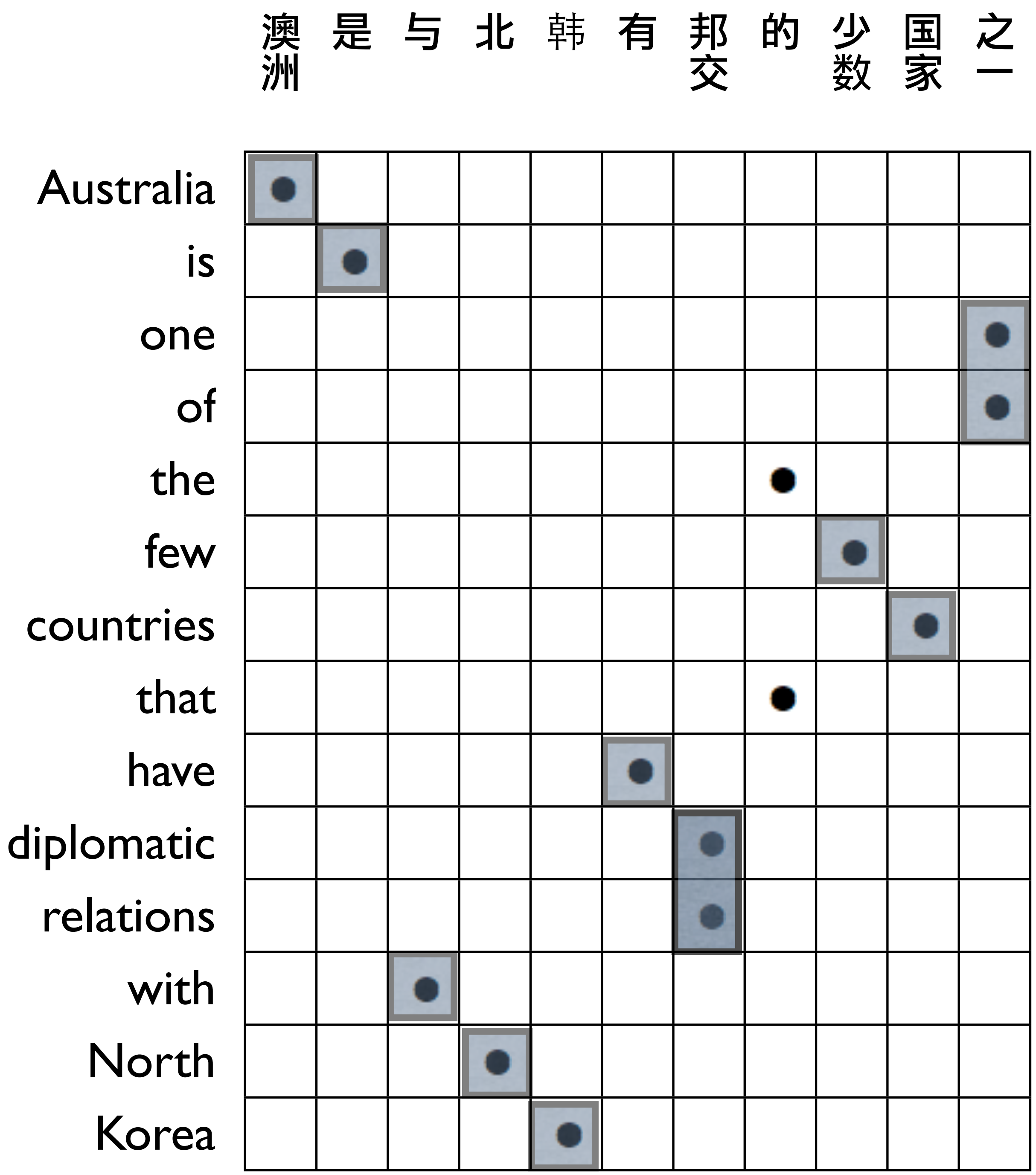
澳洲, Australia
是, is
之一, one of

Extracting phrase pairs



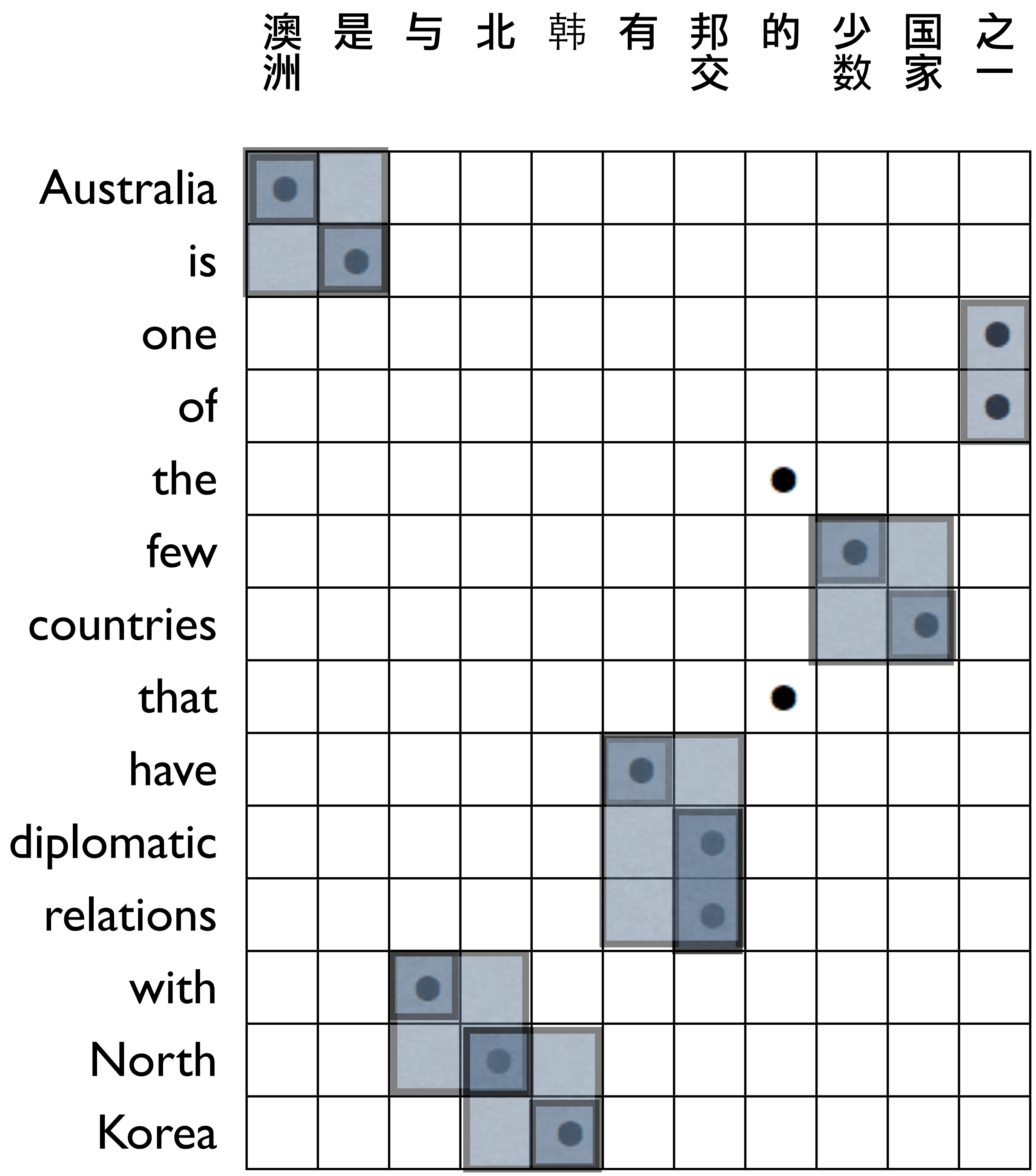
澳洲, Australia
是, is
之一, one of
少数, few

Extracting phrase pairs



澳洲, Australia
是, is
之一, one of
少数, few
国家, countries
有, have
邦交, diplomatic relations
与, with
北, North
韩, Korea

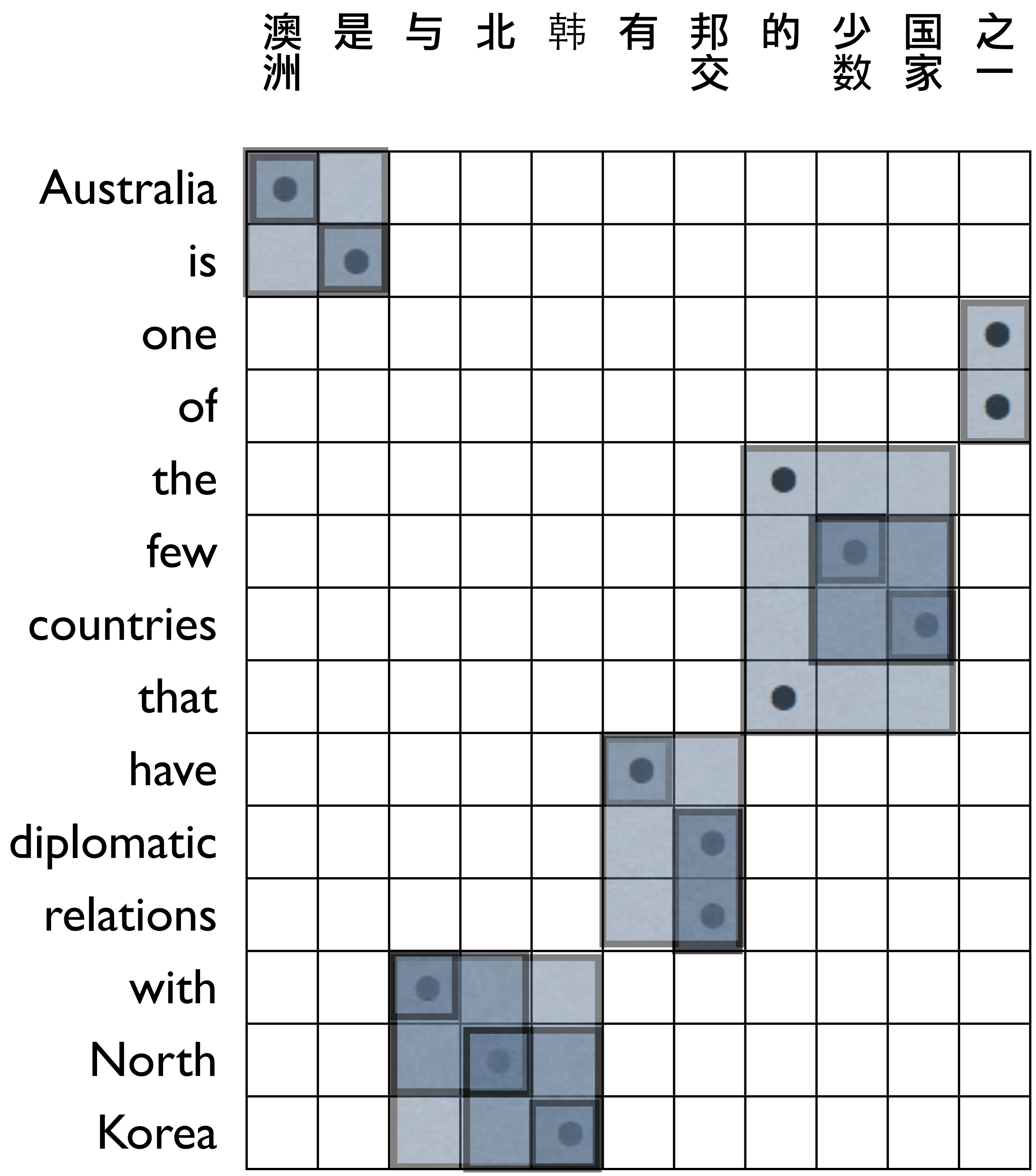
Extracting phrase pairs



澳洲, Australia
是, is
之一, one of
少数, few
国家, countries
有, have
邦交, diplomatic relations
与, with
北, North
韩, Korea

澳洲是, Australia is
少数 国家, few countries
有邦交, have diplomatic relations
与北, with North
北韩, North Korea

Extracting phrase pairs

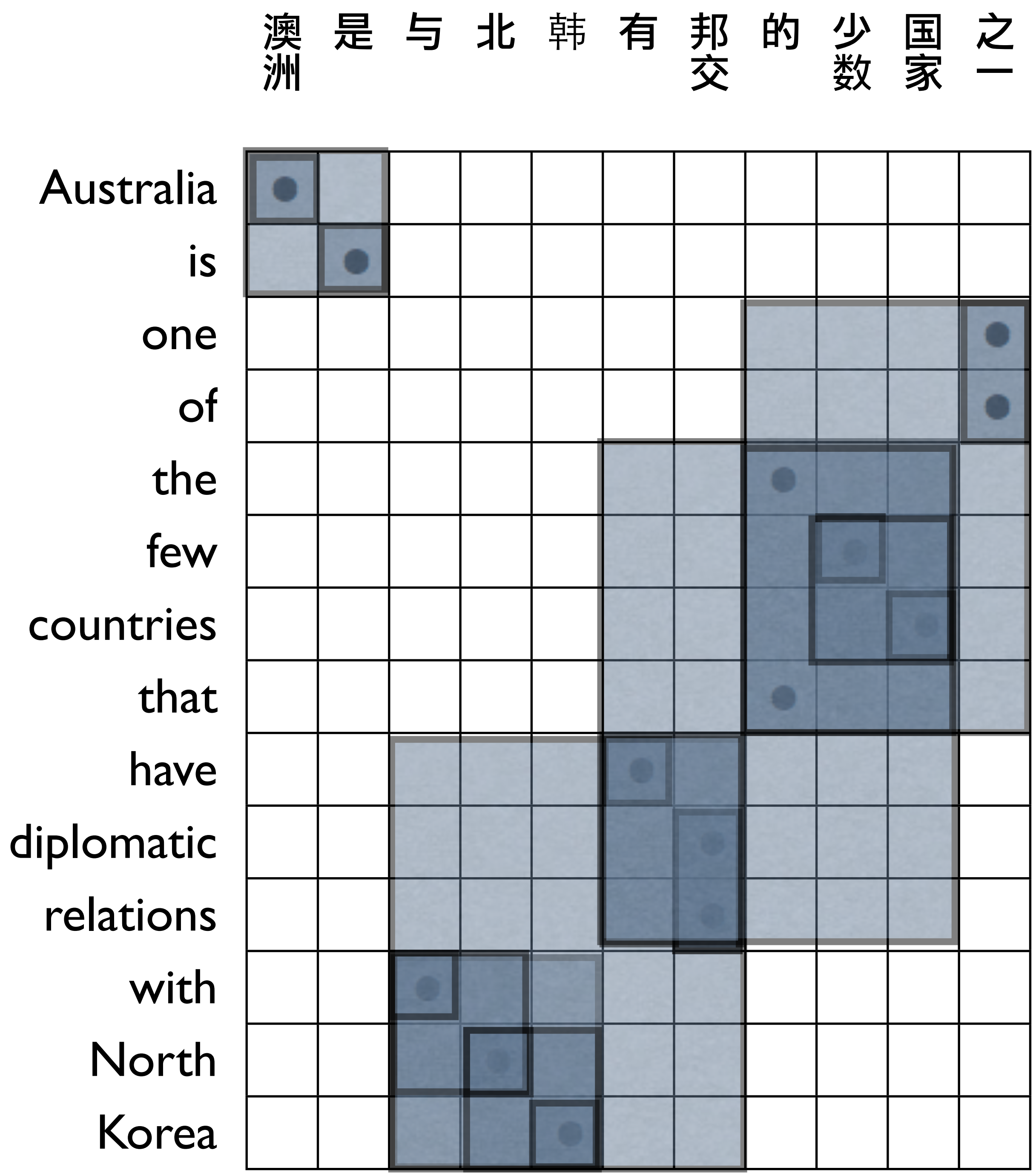


澳洲, Australia
是, is
之一, one of
少数, few
国家, countries
有, have
邦交, diplomatic relations
与, with
北, North
韩, Korea

澳洲是, Australia is
少数 国家, few countries
有邦交, have diplomatic relations
与北, with North
北韩, North Korea

的少数 国家, the few countries that
与北韩, with North Korea

Extracting phrase pairs



澳洲, Australia
是, is
之一, one of
少数, few
国家, countries
有, have
邦交, diplomatic relations
与, with
北, North
韩, Korea

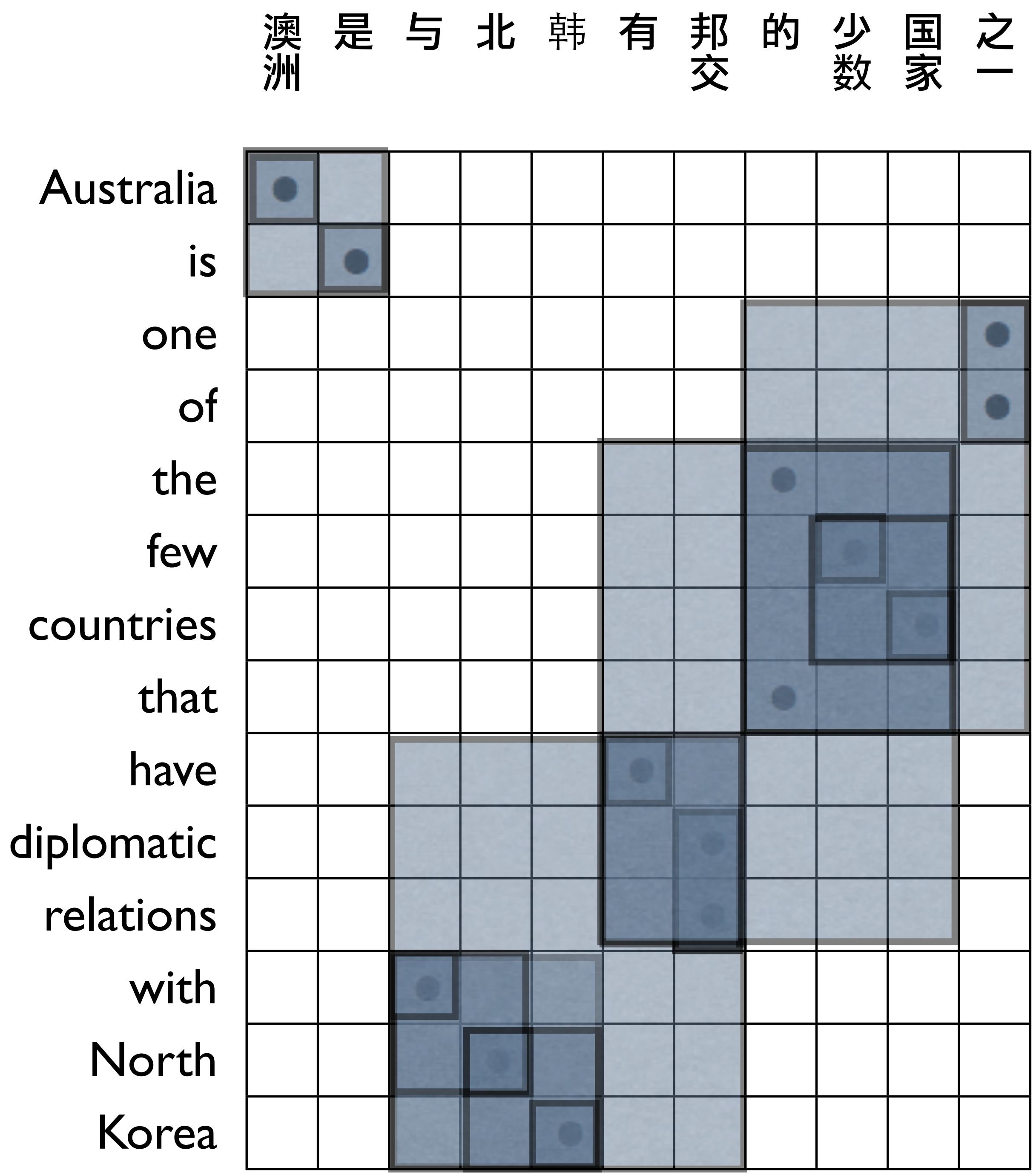
澳洲是, Australia is
少数 国家, few countries
有邦交, have diplomatic relations
与北, with North
北韩, North Korea

的少数 国家, the few countries that
与北韩, with North Korea

之一的少数 国家, one of the the few countries that
与北韩 有邦交, have diplomatic relations with North Korea

有邦交 的少数 国家, the few countries that have diplomatic relations

Extracting phrase pairs



澳洲, Australia
是, is

少数, few
国家, countries
有, have
邦交, diplomatic relations
与, with
北, North
韩, Korea

少数 国家, few countries

北韩, North Korea

与北韩, with North Korea

与北韩 有邦交, have diplomatic relations with North Korea

Why does it hurt to limit to constituents?

- Massively reduces the inventory of phrases that can be used as translation units
- Eliminates non-constituent phrases, many of which are quite useful
  - *there are*
  - *note that*
  - *according to*

So, what should we do?

- Drop syntax from statistical machine translation, since syntax is a bad fit for the data
- Abandon conventional English syntax and move towards more robust grammars that adapt to the parallel training corpus
- Maintain English syntax but design different syntactic models

Synchronous Context Free Grammars

- A common way of representing syntax in NLP is through context free grammars
- Synchronous context free grammars generate pairs of corresponding strings
- Can be used to describe translation and re-ordering between languages
- SCFGs translate sentences by parsing them

Example SCFG for Urdu

	Urdu	English
S →	NP① VP②	NP① VP②
VP→	PP① VP②	VP② PP①
VP→	V① AUX②	AUX② V①
PP →	NP① P②	P② NP①
NP →	*hamd ansary*	*Hamid Ansari*
NP →	*na}b sdr*	*Vice President*
V →	*namzd*	*nominated*
P →	*kylye*	*for*
AUX →	*taa*	*was*

hamd ansary *na}b sdr* *kylye* *namzd* *taa*

NP❶

hamd ansary *na}b sdr* *kylye* *namzd* *taa*

NP❶

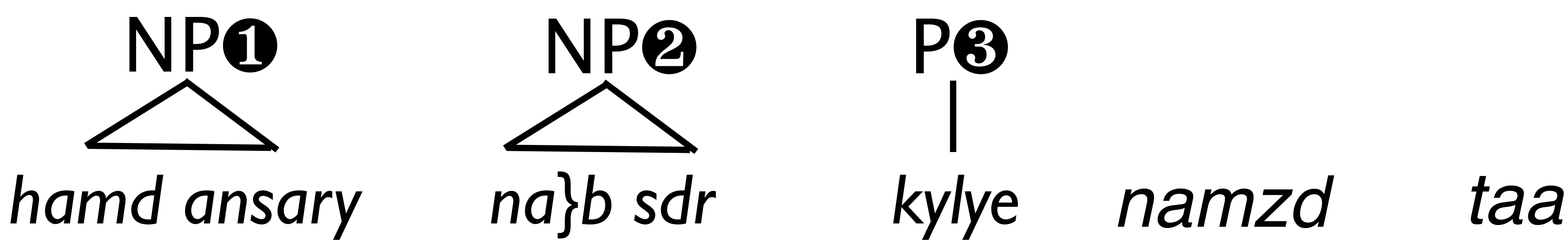
Hamid Ansari

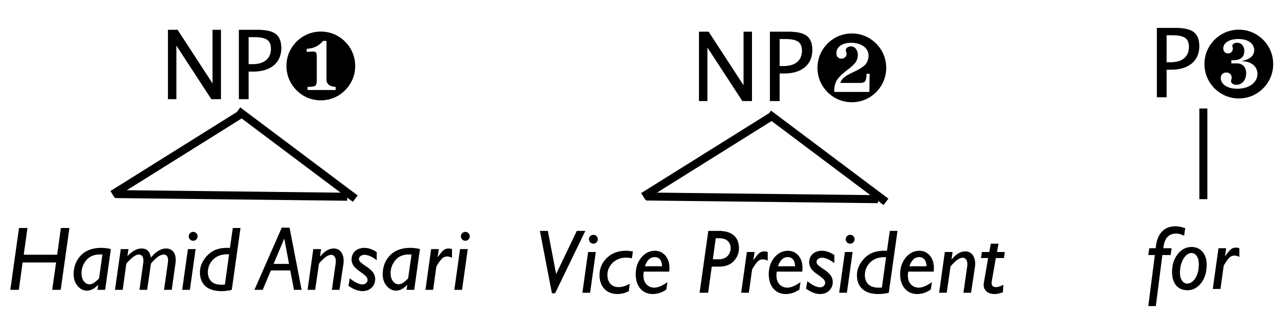
NP❶ NP❷

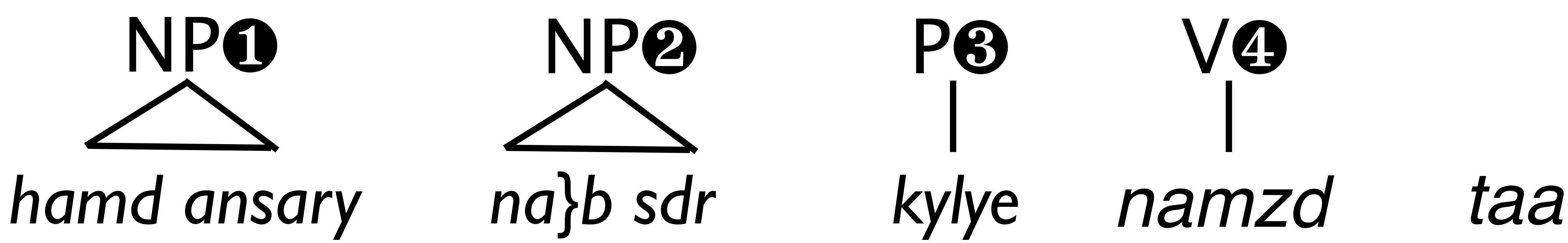
hamd ansary *na}b sdr* *kylye* *namzd* *taa*

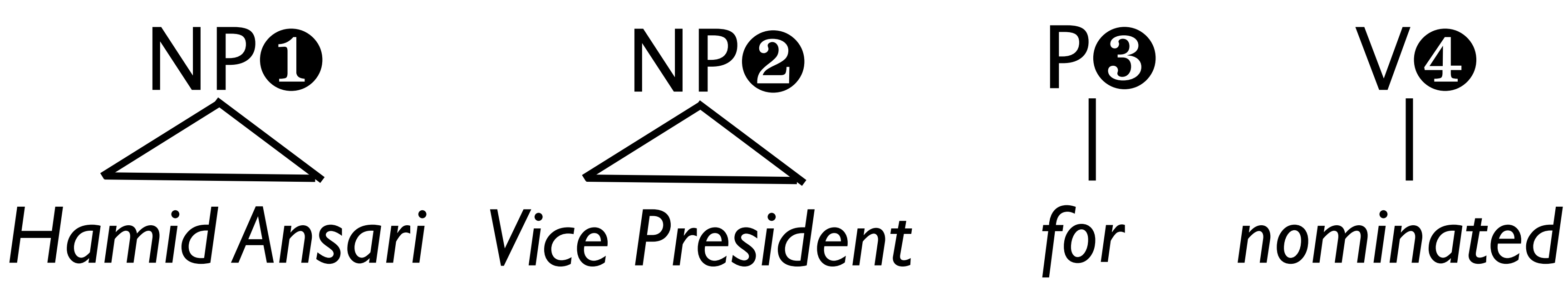
NP❶ NP❷

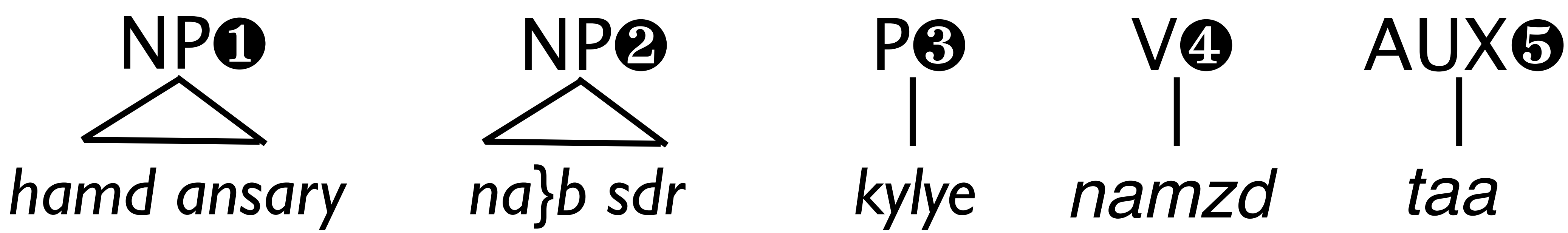
Hamid Ansari *Vice President*

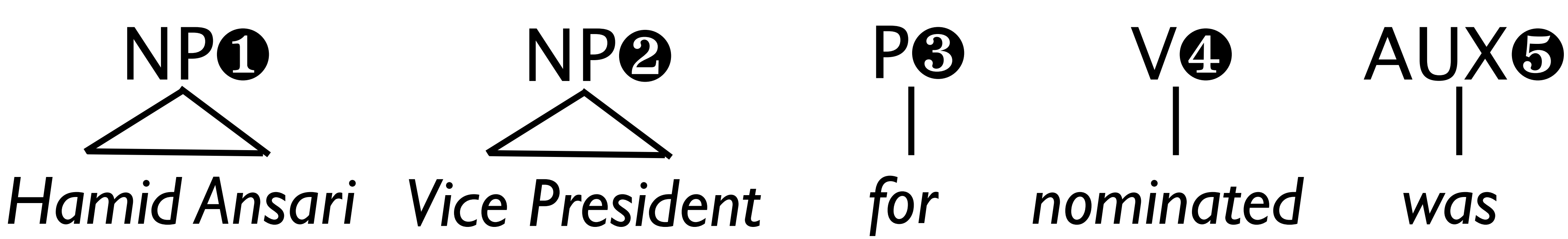

NP❶
NP❷
P❸
hamd ansary
na}b sdr
kylye
namzd
taa

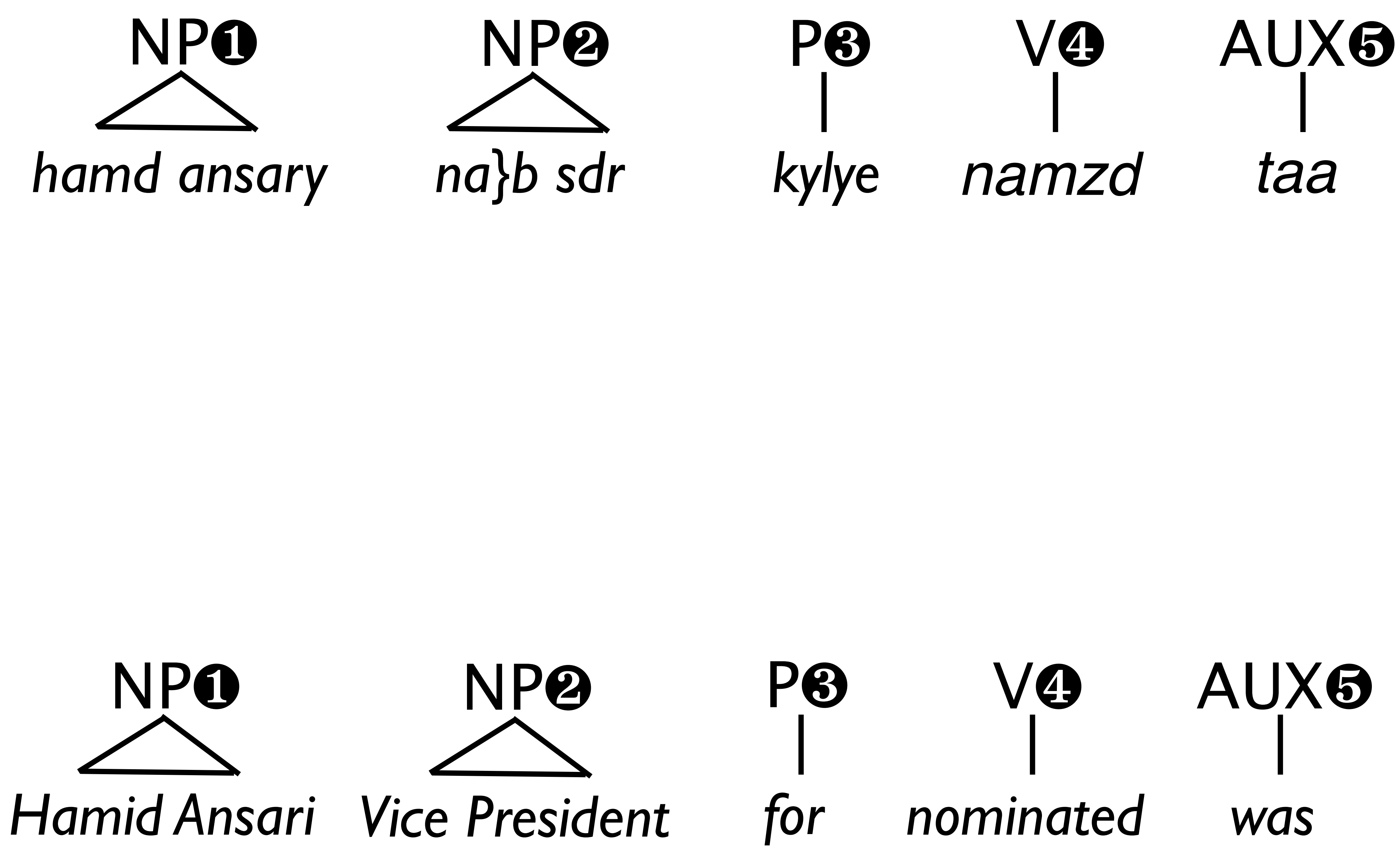

NP❶
NP❷
P❸
Hamid Ansari
Vice President
for

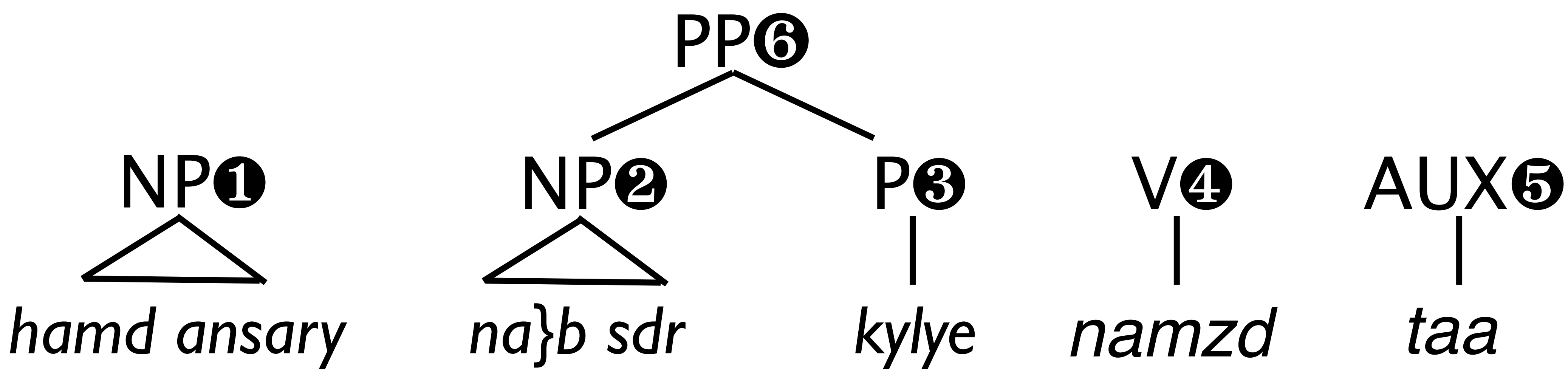

NP❶
NP❷
P❸
V❹
hamd ansary
na}b sdr
kylye
namzd
taa

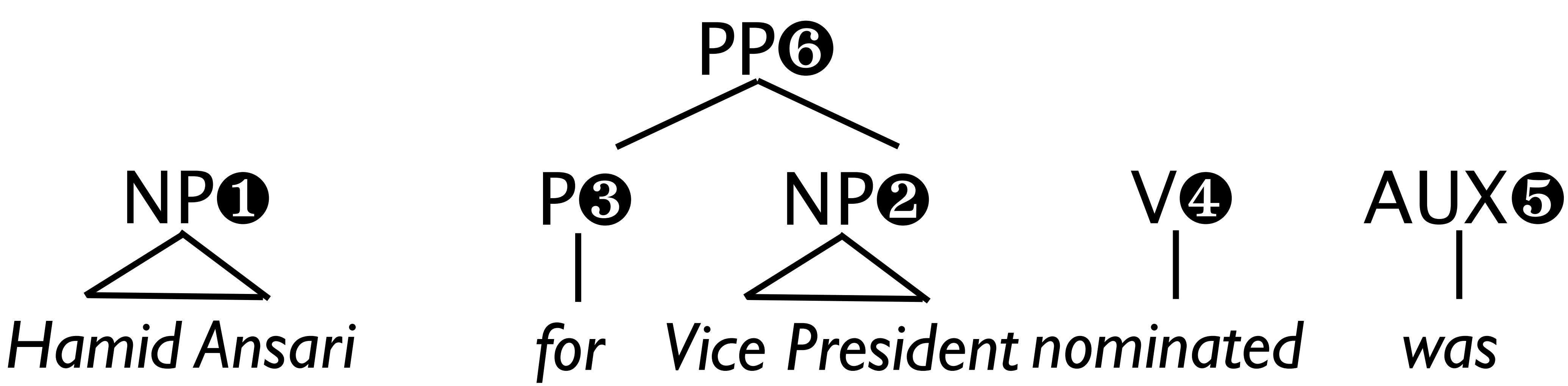

NP❶
NP❷
P❸
V❹
Hamid Ansari
Vice President
for
nominated

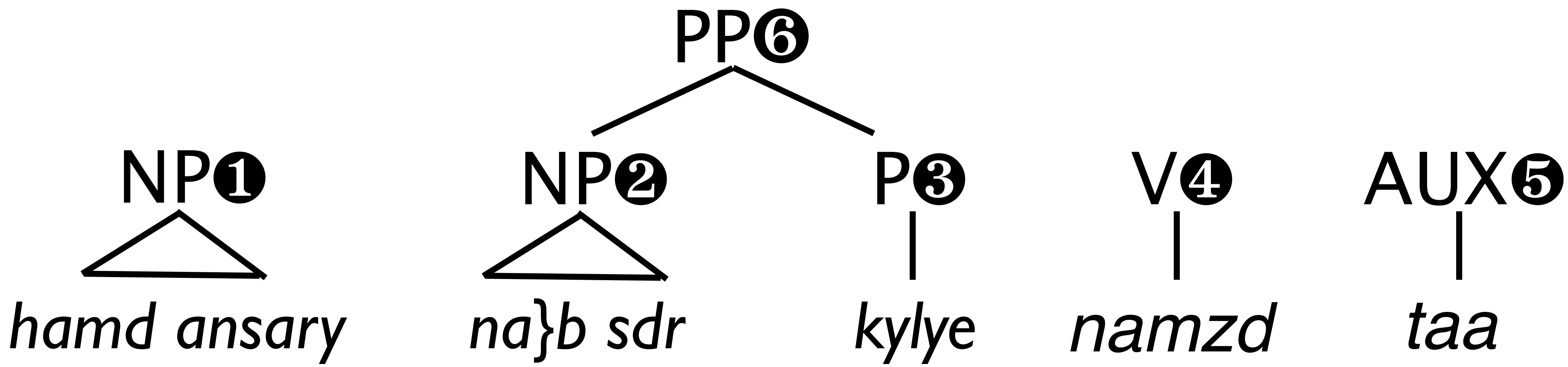

NP❶
NP❷
P❸
V❹
AUX❺
hamd ansary
na}b sdr
kylye
namzd
taa

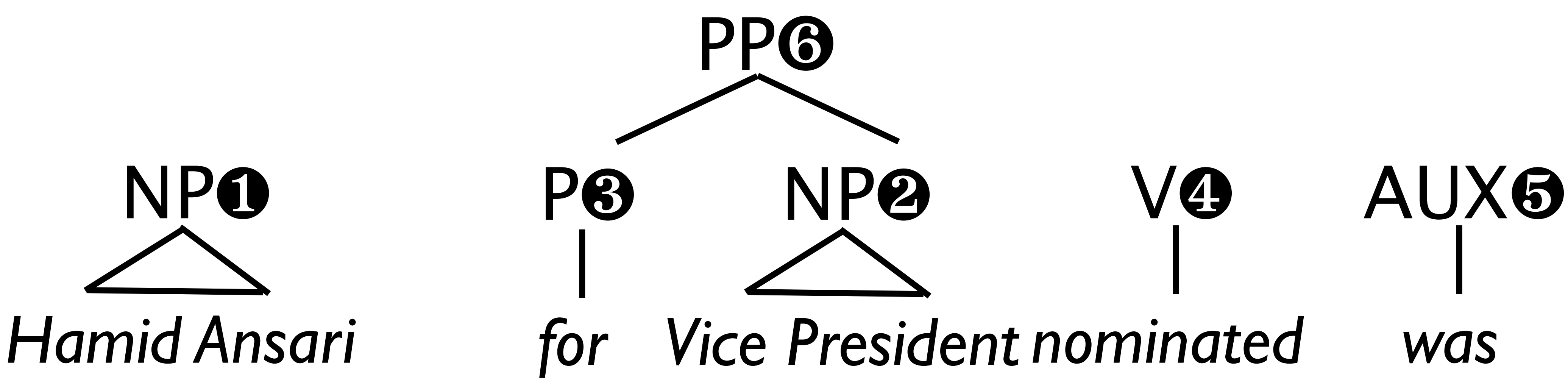

NP❶
NP❷
P❸
V❹
AUX❺
Hamid Ansari
Vice President
for
nominated
was

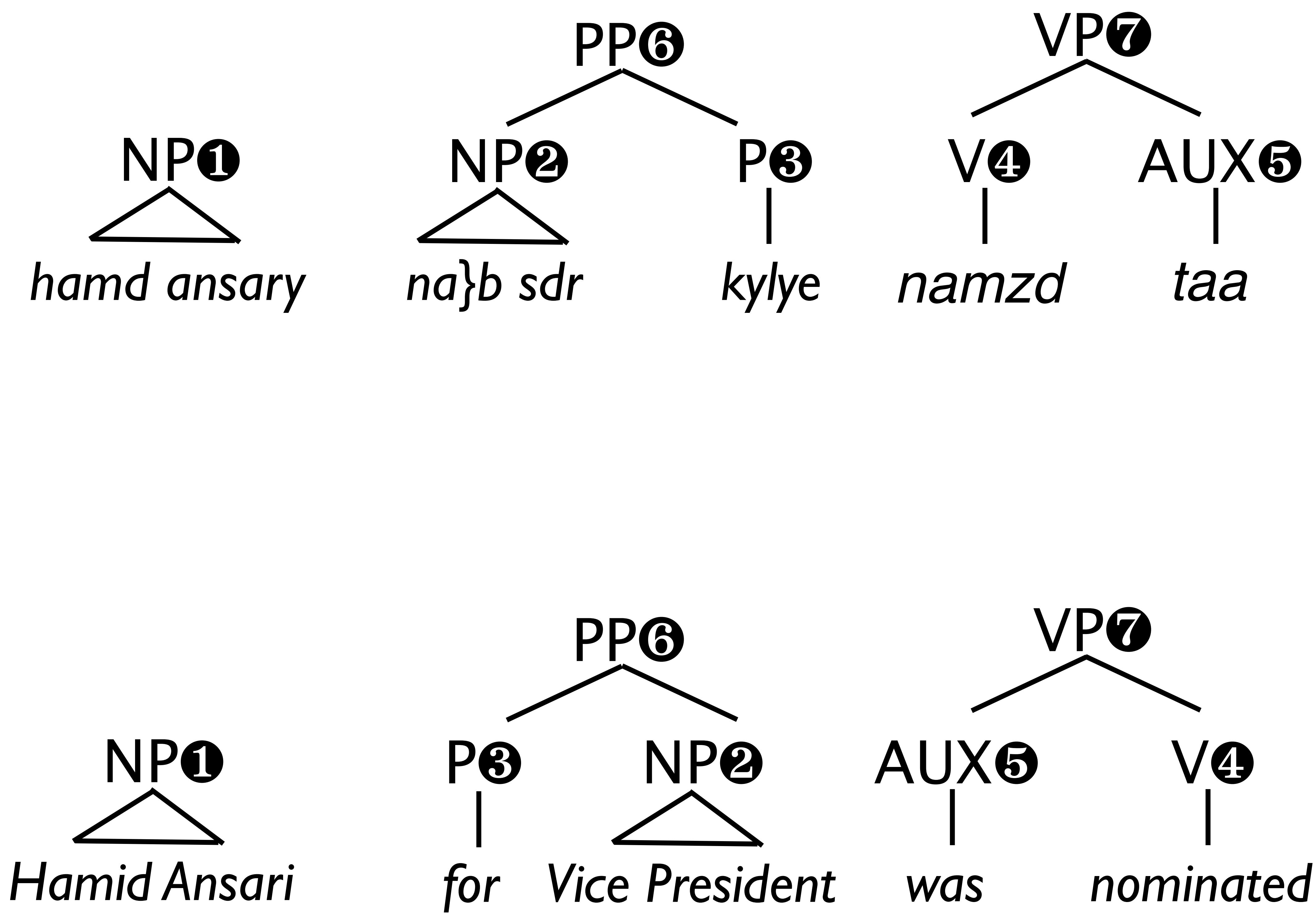

NP❶
hamd ansary
NP❷
na}b sdr
P❸
kylye
V❹
namzd
AUX❺
taa
NP❶
Hamid Ansari
NP❷
Vice President
P❸
for
V❹
nominated
AUX❺
was

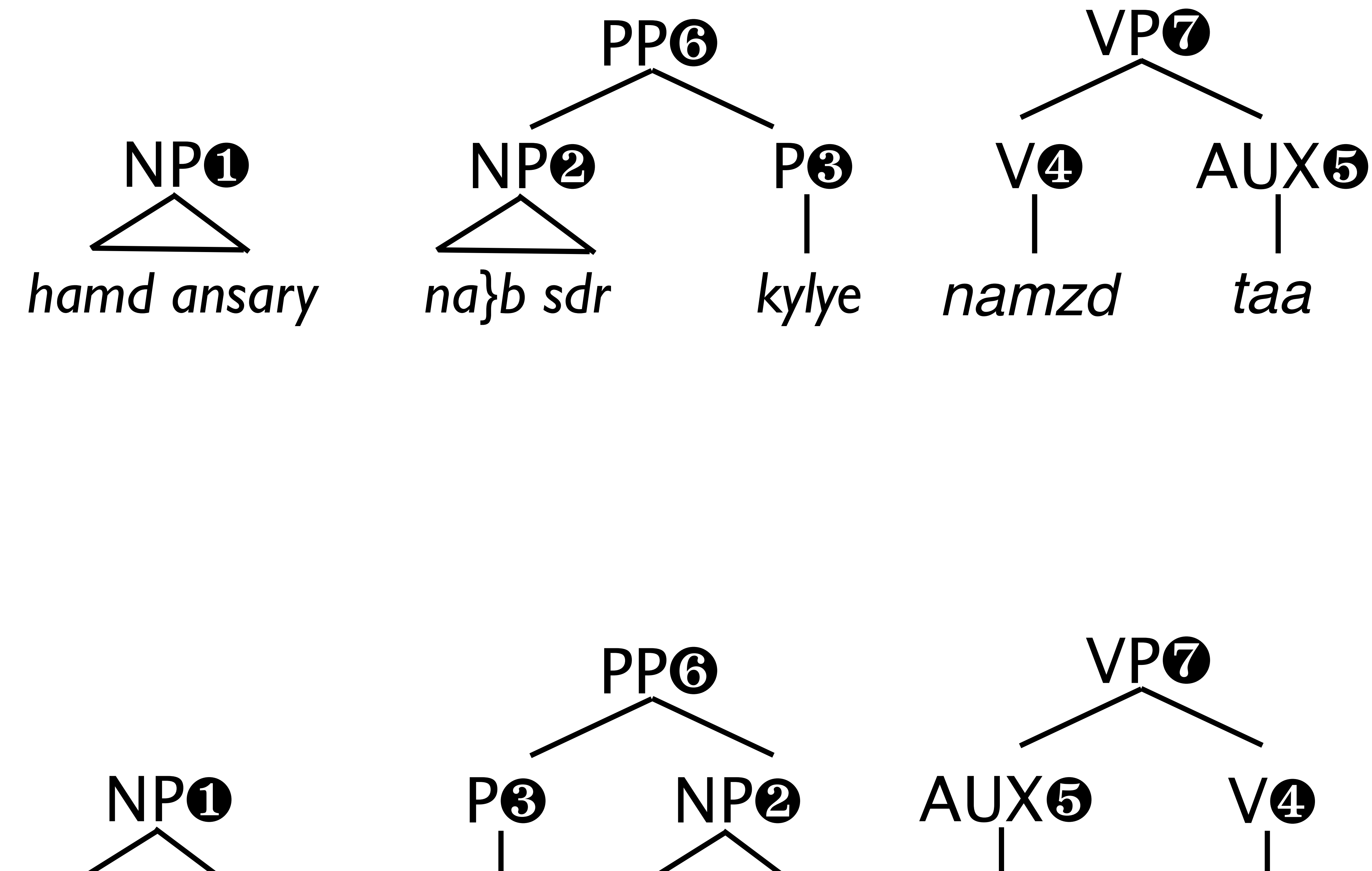

PP❻
NP❶
NP❷
P❸
V❹
AUX❺
hamd ansary
na}b sdr
kylye
namzd
taa

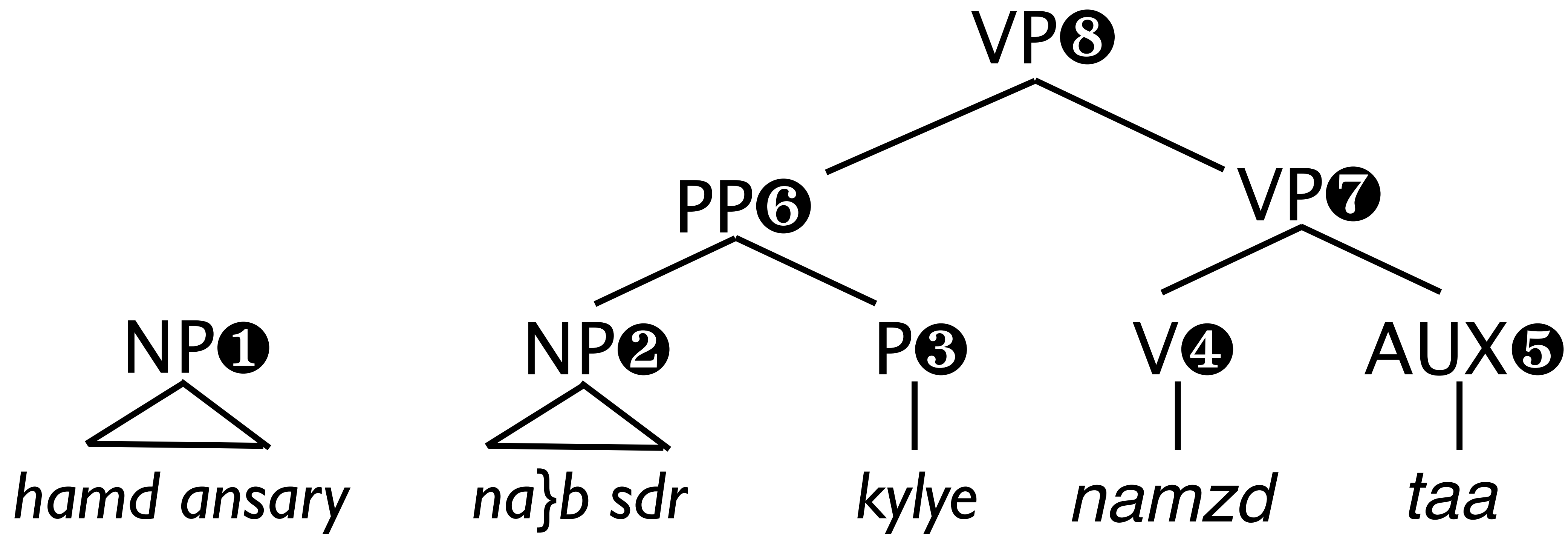

PP❻
NP❶
P❸
NP❷
V❹
AUX❺
Hamid Ansari
for
Vice President
nominated
was

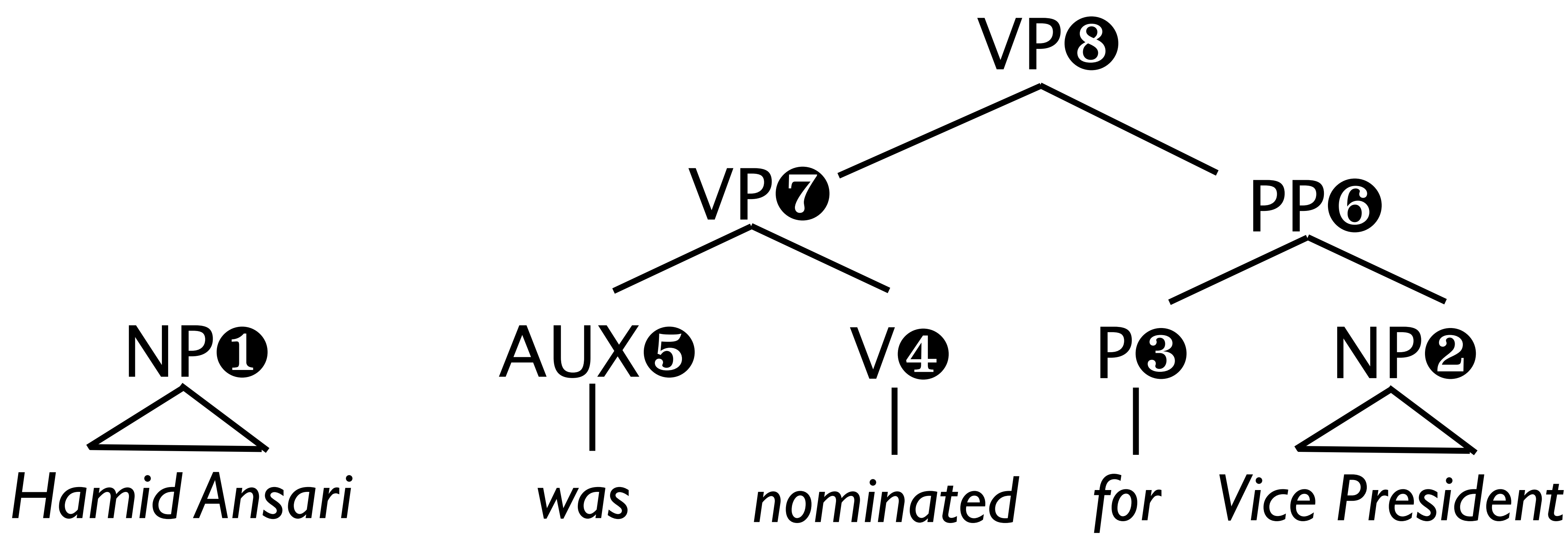

PP❻
NP❶
NP❷
P❸
V❹
AUX❺
hamd ansary
na}b sdr
kylye
namzd
taa

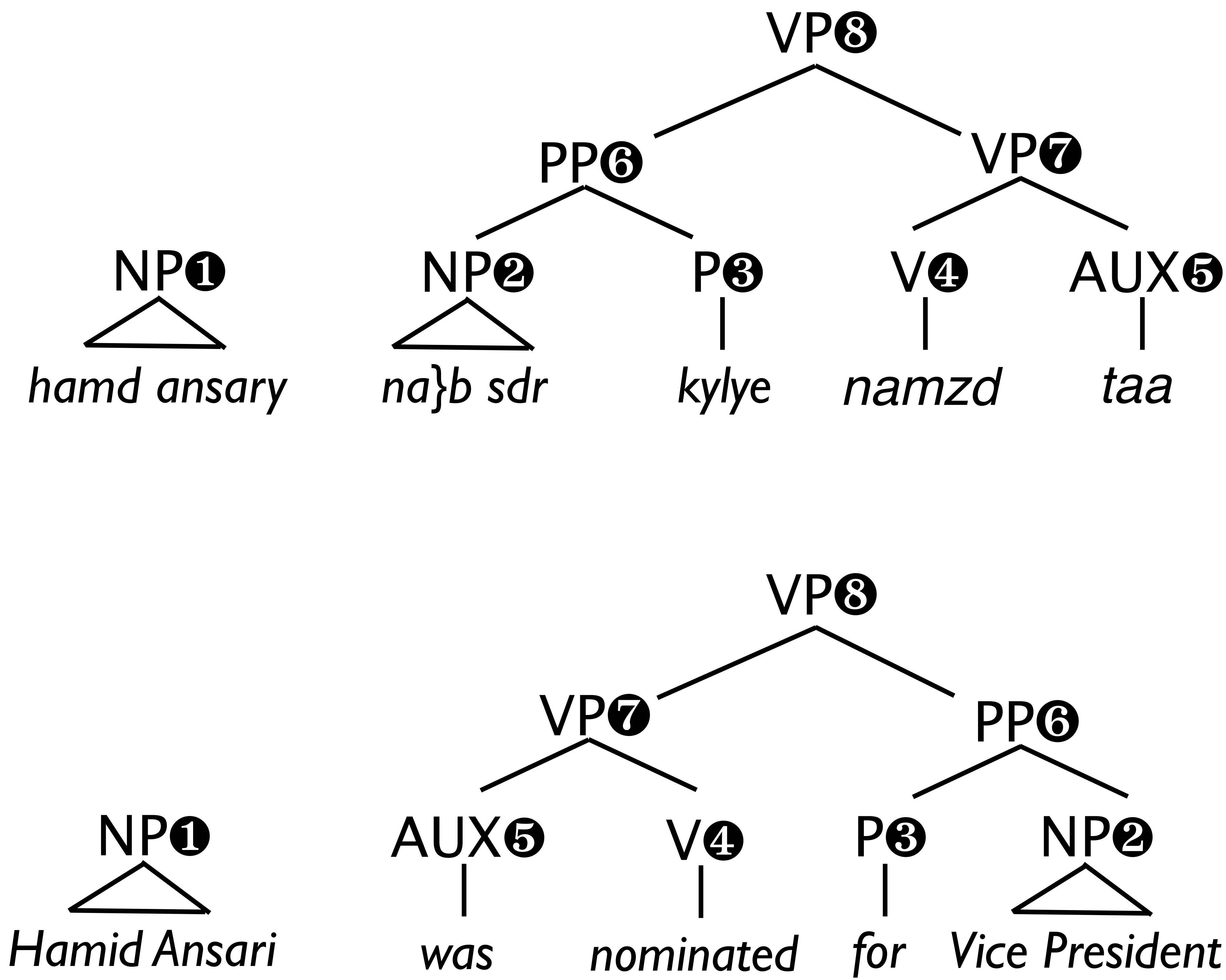

PP❻
NP❶
P❸
NP❷
V❹
AUX❺
Hamid Ansari
for
Vice President
nominated
was


PP❻
VP❼
NP❶
NP❷
P❸
V❹
AUX❺
hamd ansary
na}b sdr
kylye
namzd
taa
PP❻
VP❼
NP❶
P❸
NP❷
AUX❺
V❹
Hamid Ansari
for
Vice President
was
nominated

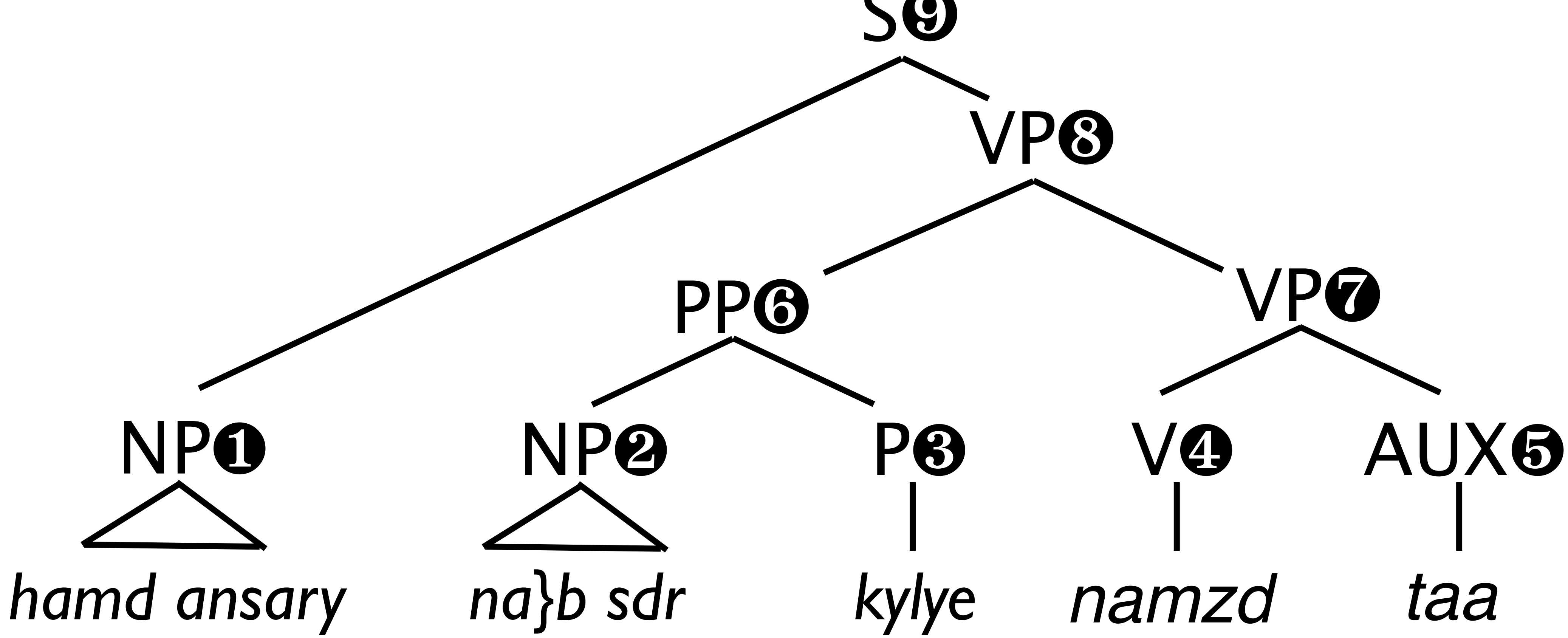

PP❻
VP❼
NP❶
NP❷
P❸
V❹
AUX❺
hamd ansary
na}b sdr
kylye
namzd
taa
PP❻
VP❼
NP❶
P❸
NP❷
AUX❺
V❹
Hamid Ansari
for
Vice President
was
nominated

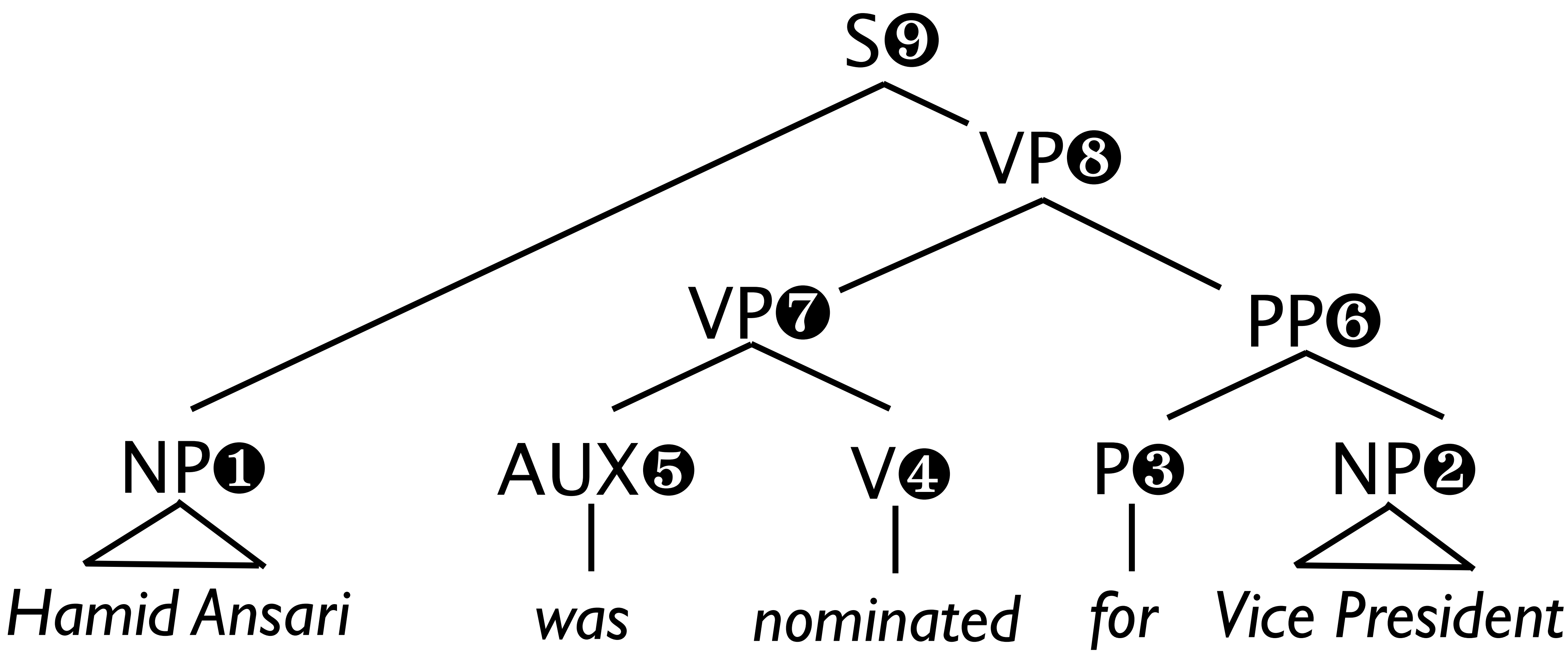

VP❽
PP❻
VP❼
NP❶
NP❷
P❸
V❹
AUX❺
hamd ansary
na}b sdr
kylye
namzd
taa


VP❽
VP❼
PP❻
NP❶
AUX❺
V❹
P❸
NP❷
Hamid Ansari
was
nominated
for
Vice President



VP❽
PP❻
VP❼
NP❶
NP❷
P❸
V❹
AUX❺
hamd ansary
na}b sdr
kylye
namzd
taa
VP❽
VP❼
PP❻
NP❶
AUX❺
V❹
P❸
NP❷
Hamid Ansari
was
nominated
for
Vice President


S❾
VP❽
PP❻
VP❼
NP❶
NP❷
P❸
V❹
AUX❺
hamd ansary
na}b sdr
kylye
namzd
taa

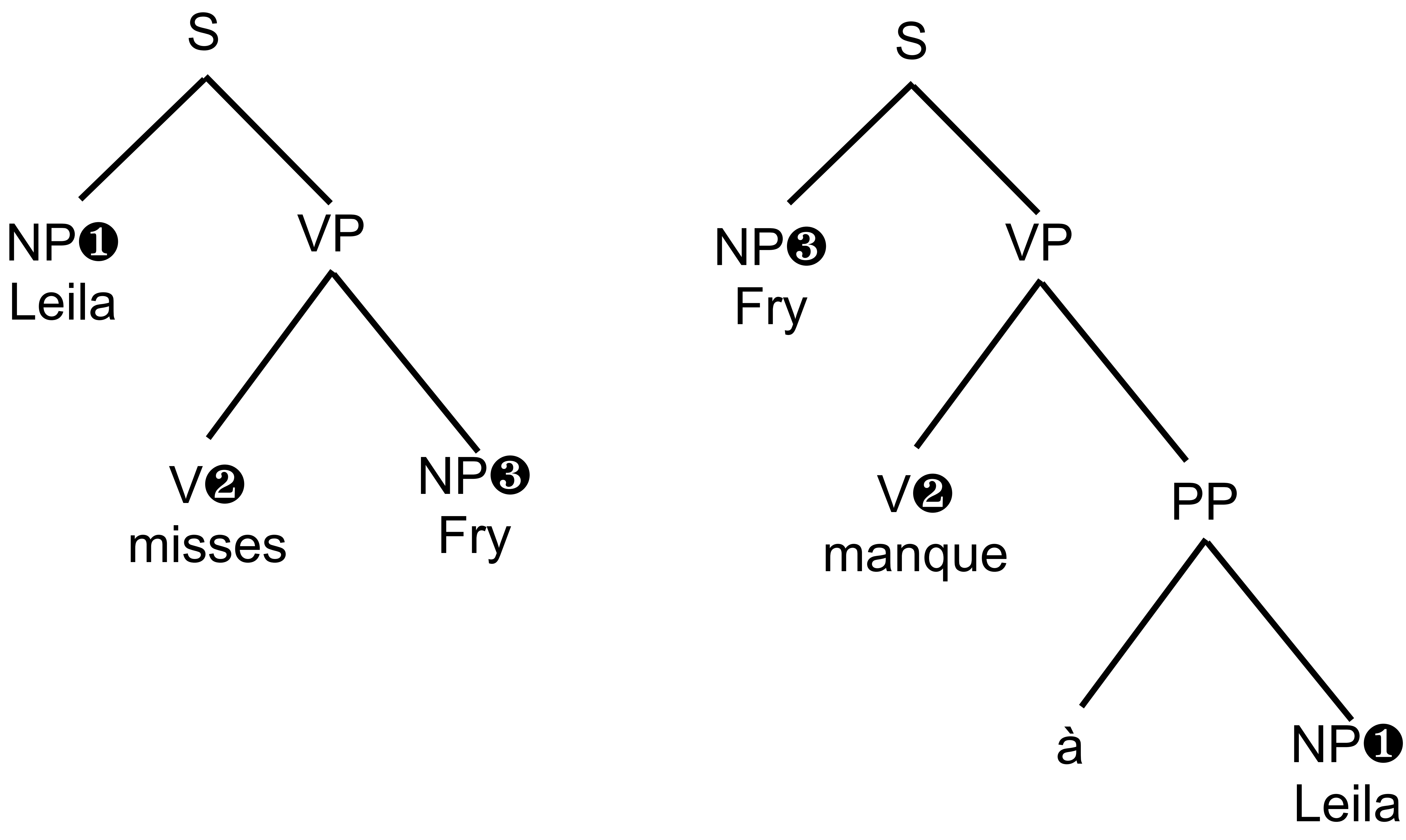

S❾
VP❽
VP❼
PP❻
NP❶
AUX❺
V❹
P❸
NP❷
Hamid Ansari
was
nominated
for
Vice President

Discussion: Do you like SCFG?

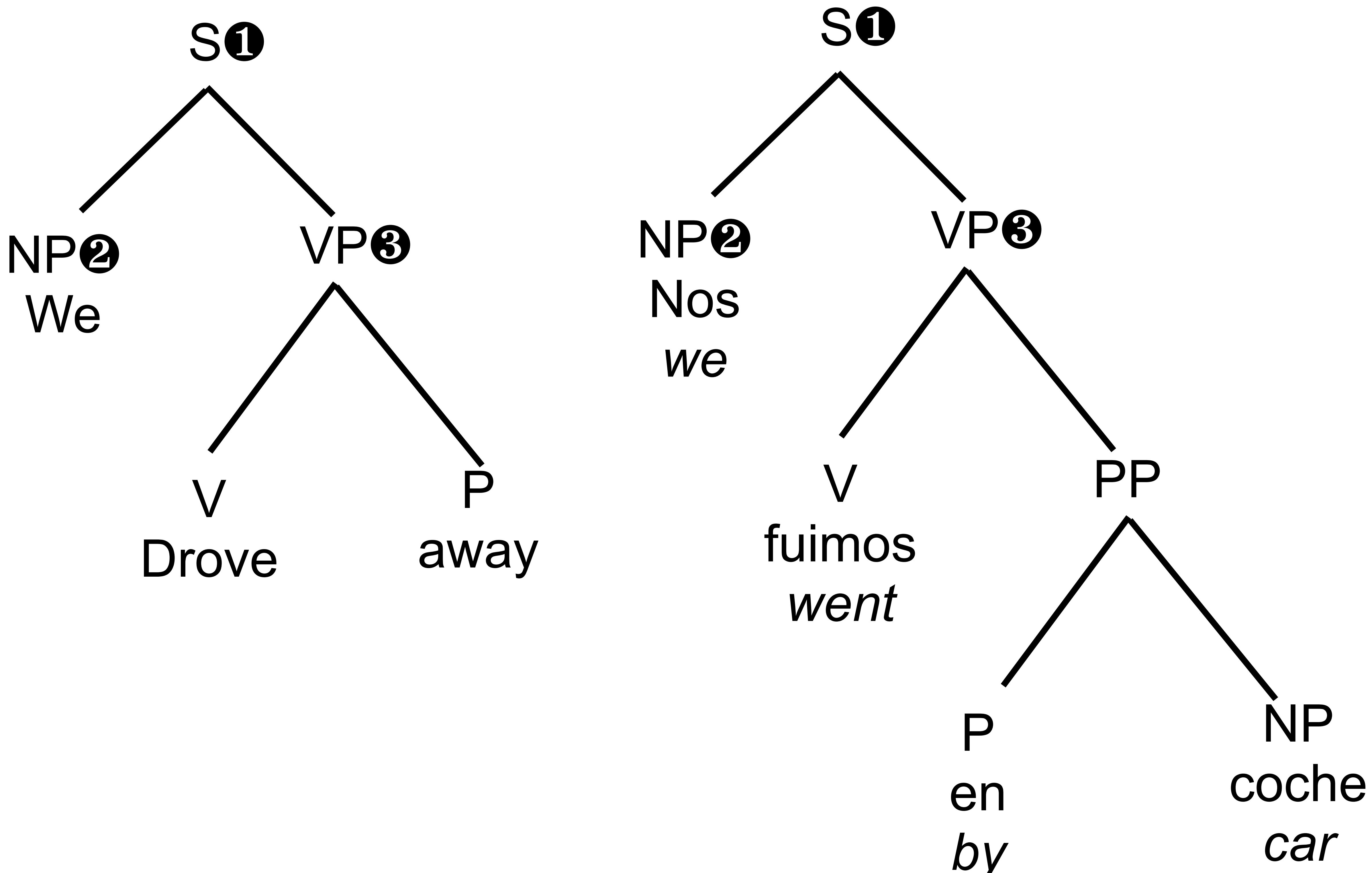
- In what ways are SCFGs better for describing reordering than what we saw before?
- Is this a good model of how languages relate?
- What do you think of the synchronous requirement?

(Discuss with your neighbor)

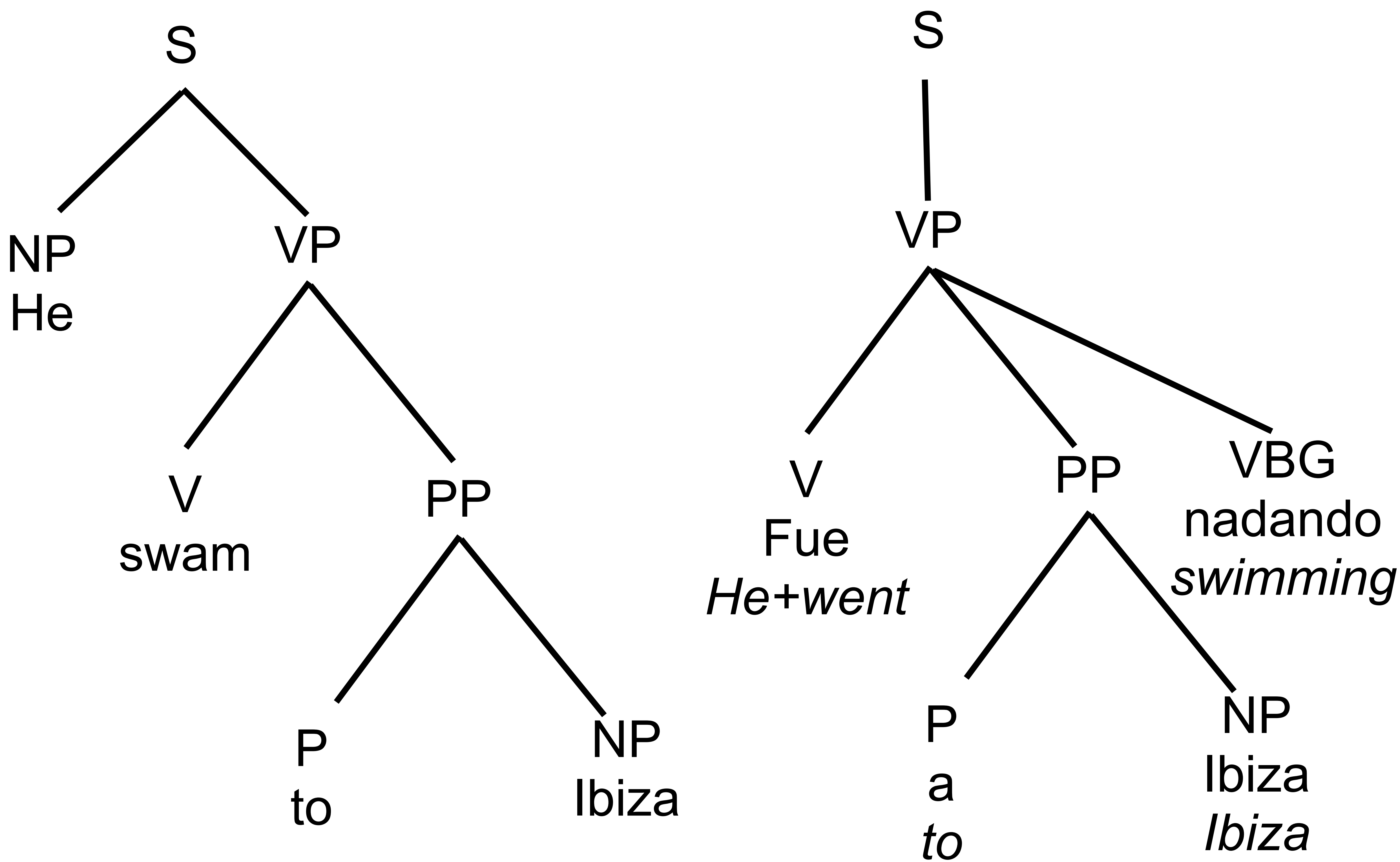
Sometimes languages are mismatched



Spanish motion verb



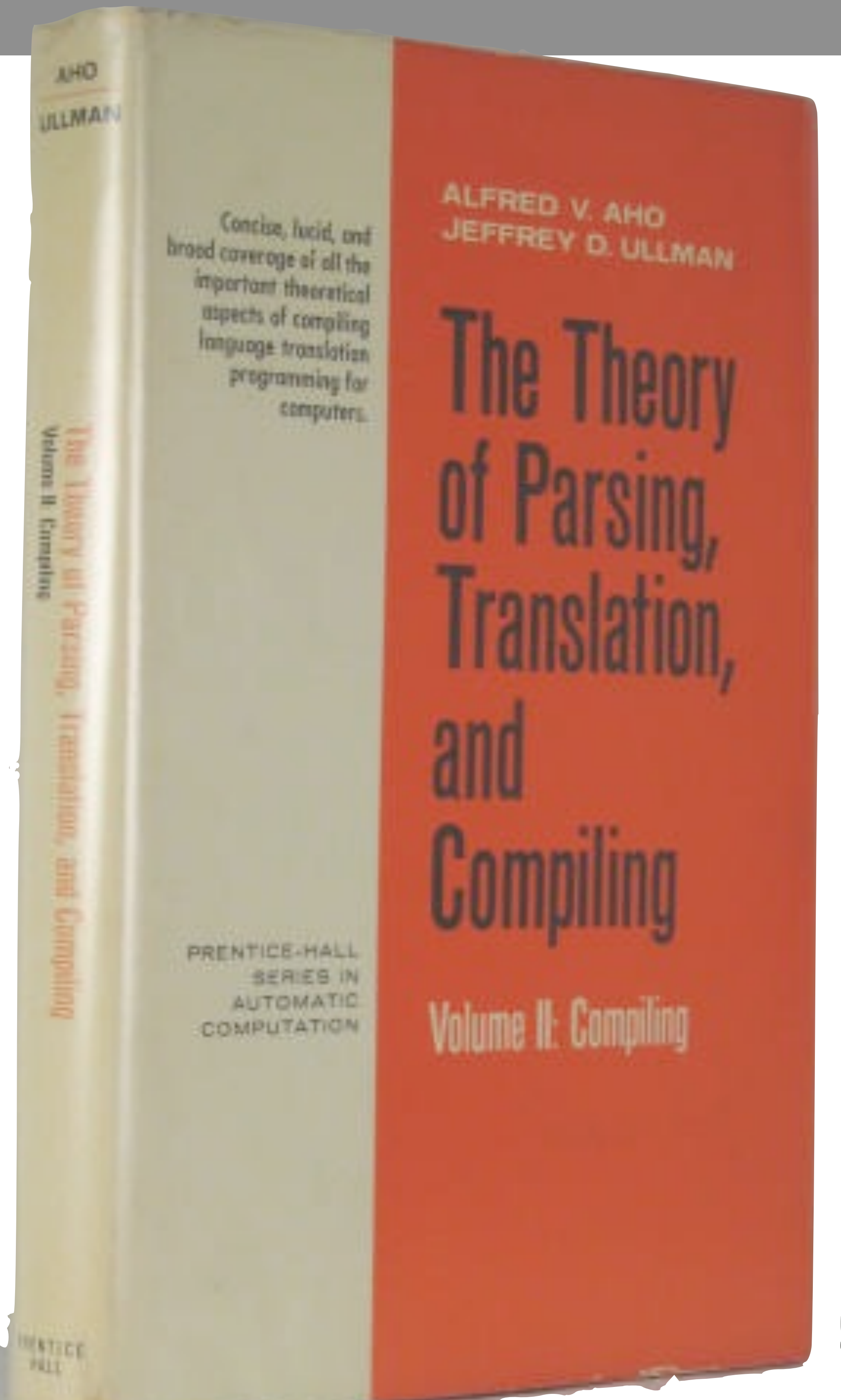
Spanish motion verb, pro-drop



We are going to use them anyway

- SCFGs are mismatched with some linguistic phenomena
- But they have nice formal properties and well-defined algorithms

Formal definition of SCFGs



- Aho and Ullman worked all of this out in the `60s and `70s
- Compiler theory

Formal definition of SCFGs

- A synchronous context free grammar is formally defined by a tuple

$$G = \langle N, T_S, T_T, R, S \rangle$$

- Where

Formal definition of SCFGs

S, NP, VP, PP, P, V, AUX

- A synchronous context free grammar is defined by a tuple

$$G = \langle N, T_S, T_T, R, S \rangle$$

- Where
  - N is a shared set of non-terminal symbols

Formal definition of SCFGs

hamd ansary, na}b sdr, namzd, kylye, taa

S, NP, VP, PP, P, V, AUX

- A synchronous context free grammar is defined by a tuple

$$G = \langle N, T_S, T_T, R, S \rangle$$

- Where
  - –N is a shared set of non-terminal symbols
  - –T_S is the set of source language terminals

Formal definition of SCFGs

hamd ansary, na}b sdr, namzd, kylye, taa

for, Hamid Ansari, nominated, Vice President, was

S, NP, VP, PP, P, V, AUX

- A synchronous context free grammar is defined by a tuple

$$G = <N, T_S, T_T, R, S>$$

- Where
  - –N is a shared set of non-terminal symbols
  - –T_S is the set of source language terminals
  - –T_T is the set of target language terminals

Formal definition of SCFGs

hamd ansary, na}b sdr, namzd, kylye, taa

for, Hamid Ansari, nominated, Vice President, was

S, NP, VP, PP, P, V, AUX

- A synchronous context free grammar is defined by a tuple

$$G = \langle N, T_S, T_T, R, S \rangle$$

- Where
  - –N is a shared set of non-terminal symbols
  - –T_S is the set of source language terminals
  - –T_T is the set of target language terminals
  - –R is a set of production rules

Formal definition of SCFGs

hamd ansary, na}b sdr, namzd, kylye, taa

for, Hamid Ansari, nominated, Vice President, was

S, NP, VP, PP, P, V, AUX

S

- A synchronous context free grammar is defined by a tuple

$$G = \langle N, T_S, T_T, R, S \rangle$$

- Where
  - N is a shared set of non-terminal symbols
  - T_S is the set of source language terminals
  - T_T is the set of target language terminals
  - R is a set of production rules
  - $S \in N$, designated as the goal state

Formal definition of SCFGs

- Each production rule has the form

$$X \rightarrow \langle \alpha, \beta, \sim, w \rangle$$

- Where
  - $X \in N$
  - $\alpha \in (N \cup T_S)^*$
  - $\beta \in (N \cup T_T)^*$
  - ~ is a one-to-one correspondence between the non terminals in α and β
  - w is a weight assigned to the rule

Algorithms for SCFGs

- Translation with SCFGs is done via parsing
- How do we write an algorithm for parsing?
- One way to do it is as a deductive proof system

The CKY Parsing Algorithm

Axioms	$\dfrac{}{A \rightarrow \alpha}$	for all $(A \rightarrow \alpha) \in R$
Inference rules	$\dfrac{A \rightarrow w_{i+1}}{[A, i, i+1]}$ $\dfrac{[B, i, j] \quad [C, j, k] \quad A \rightarrow BC}{[A, i, k]}$	
Goal	$[S, 0, n]$	

Axioms		**Inference rule used**	**Goal**
S →	NP VP		
VP →	PP VP		[S, 0, 5]
VP →	V AUX		
PP →	NP P		
NP →	*hamd ansary*		
NP →	*na}b sdr*		
V →	*namzd*		
P →	*kylye*		
AUX →	*taa*		

Axioms **Inference rule used** **Goal**

S → NP VP
VP→ PP VP
VP→ V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

[S, 0, 5]

$_0$ *hamd ansary* $_1$ *na}b sdr* $_2$ *kylye* $_3$ *namzd* $_4$ *taa* $_5$

Axioms

S	→	NP VP
VP	→	PP VP
VP	→	V AUX
PP	→	NP P
NP	→	*hamd ansary*
NP	→	*na}b sdr*
V	→	*namzd*
P	→	*kylye*
AUX	→	*taa*

Inference rule used

$$\frac{\text{NP} \rightarrow \text{hamd ansary}_1}{[\text{NP}, \mathit{0}, \mathit{1}]}$$

Goal

[S, 0, 5]

$_{0}$ *hamd ansary* $_{1}$ *na}b sdr* $_{2}$ *kylye* $_{3}$ *namzd* $_{4}$ *taa* $_{5}$

Axioms

$S \rightarrow$ NP VP
$VP \rightarrow$ PP VP
$VP \rightarrow$ V AUX
$PP \rightarrow$ NP P
$NP \rightarrow$ *hamd ansary*
$NP \rightarrow$ *na}b sdr*
$V \rightarrow$ *namzd*
$P \rightarrow$ *kylye*
$AUX \rightarrow$ *taa*

Inference rule used

$$\frac{NP \rightarrow \text{hamd ansary}_1}{[NP, \mathit{0}, \mathit{1}]}$$

Goal

[S, 0, 5]

$_{0}$ *hamd ansary* $_{1}$ *na}b sdr* $_{2}$ *kylye* $_{3}$ *namzd* $_{4}$ *taa* $_{5}$

$\boxed{[NP, \mathit{0}, \mathit{1}]}$

Axioms

S	→	NP VP
VP	→	PP VP
VP	→	V AUX
PP	→	NP P
NP	→	*hamd ansary*
NP	→	*na}b sdr*
V	→	*namzd*
P	→	*kylye*
AUX	→	*taa*

Inference rule used

Goal

[S, 0, 5]

₀ hamd ansary ₁ na}b sdr ₂ kylye ₃ namzd ₄ taa ₅

[NP, *0, 1*]

Axioms

S →	NP VP
VP→	PP VP
VP→	V AUX
PP →	NP P
NP →	*hamd ansary*
NP →	*na}b sdr*
V →	*namzd*
P →	*kylye*
AUX →	*taa*

Inference rule used

$$\frac{\text{NP} \rightarrow \text{na\}b sdr}_2}{[\text{NP}, \mathit{1}, \mathit{1}]}$$

Goal

[S, 0, 5]

$_{0}$ *hamd ansary* $_{1}$ *na}b sdr* $_{2}$ *kylye* $_{3}$ *namzd* $_{4}$ *taa* $_{5}$

$\boxed{[\text{NP}, \mathit{0}, \mathit{1}]}$

Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

$$\frac{\mathrm{NP} \rightarrow \mathrm{na\}b\ sdr}_2}{[\mathrm{NP}, \mathit{1}, \mathit{1}]}$$

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]

Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]

Axioms

S	→	NP VP
VP	→	PP VP
VP	→	V AUX
PP	→	NP P
NP	→	*hamd ansary*
NP	→	*na}b sdr*
V	→	*namzd*
P	→	*kylye*
AUX	→	*taa*

Inference rule used

$$\frac{P \rightarrow kylye_3}{[P, 2, 3]}$$

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]

Axioms		
S →	NP VP	
VP→	PP VP	
VP→	V AUX	
PP →	NP P	
NP →	*hamd ansary*	
NP →	*na}b sdr*	
V →	*namzd*	
P →	*kylye*	
AUX →	*taa*	

Inference rule used

$$\frac{P \rightarrow \text{kylye}_3}{[P, \mathit{2}, \mathit{3}]}$$

Goal

[S, 0, 5]

$_{0}$ *hamd ansary* $_{1}$ *na}b sdr* $_{2}$ *kylye* $_{3}$ *namzd* $_{4}$ *taa* $_{5}$

[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]

Axioms

S	→	NP VP
VP	→	PP VP
VP	→	V AUX
PP	→	NP P
NP	→	*hamd ansary*
NP	→	*na}b sdr*
V	→	*namzd*
P	→	*kylye*
AUX	→	*taa*

Inference rule used

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]

Axioms

$S \rightarrow NP\ VP$
$VP \rightarrow PP\ VP$
$VP \rightarrow V\ AUX$
$PP \rightarrow NP\ P$
$NP \rightarrow$ *hamd ansary*
$NP \rightarrow$ *na}b sdr*
$V \rightarrow$ *namzd*
$P \rightarrow$ *kylye*
$AUX \rightarrow$ *taa*

Inference rule used

$$\frac{V \rightarrow \text{namzd}_4}{[V, 3, 4]}$$

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]

Axioms

S	→	NP VP
VP	→	PP VP
VP	→	V AUX
PP	→	NP P
NP	→	*hamd ansary*
NP	→	*na}b sdr*
V	→	*namzd*
P	→	*kylye*
AUX	→	*taa*

Inference rule used

$$\frac{V \rightarrow \text{namzd}_4}{[V, \mathit{3}, \mathit{4}]}$$

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]	[V, *3, 4*]

Axioms

S	→	NP VP
VP	→	PP VP
VP	→	V AUX
PP	→	NP P
NP	→	*hamd ansary*
NP	→	*na}b sdr*
V	→	*namzd*
P	→	*kylye*
AUX	→	*taa*

Inference rule used

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]	[V, *3, 4*]

Axioms

S	→	NP VP
VP	→	PP VP
VP	→	V AUX
PP	→	NP P
NP	→	*hamd ansary*
NP	→	*na}b sdr*
V	→	*namzd*
P	→	*kylye*
AUX	→	*taa*

Inference rule used

$$\frac{\mathrm{AUX} \rightarrow \mathrm{taa}_5}{[\mathrm{AUX}, \mathit{4}, \mathit{5}]}$$

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]	[V, *3, 4*]

Axioms

S	→	NP VP
VP	→	PP VP
VP	→	V AUX
PP	→	NP P
NP	→	*hamd ansary*
NP	→	*na}b sdr*
V	→	*namzd*
P	→	*kylye*
AUX	→	*taa*

Inference rule used

$$\frac{\text{AUX} \rightarrow \text{taa}_5}{[\text{AUX}, \mathit{4}, \mathit{5}]}$$

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]	[V, *3, 4*]	[AUX,4,5]

Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

Goal

[S, 0, 5]

0 *hamd ansary* *1*	*na}b sdr* *2*	*kylye* *3*	*namzd* *4*	*taa* *5*
[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]	[V, *3, 4*]	[AUX,4,5]

Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]	[V, *3, 4*]	[AUX,4,5]

Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

$$\frac{[NP, \mathit{1}, \mathit{2}] \quad [P, \mathit{2}, \mathit{3}] \quad PP \rightarrow NP\ P}{[PP, \mathit{1}, \mathit{3}]}$$

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]	[V, *3, 4*]	[AUX,4,5]

Axioms

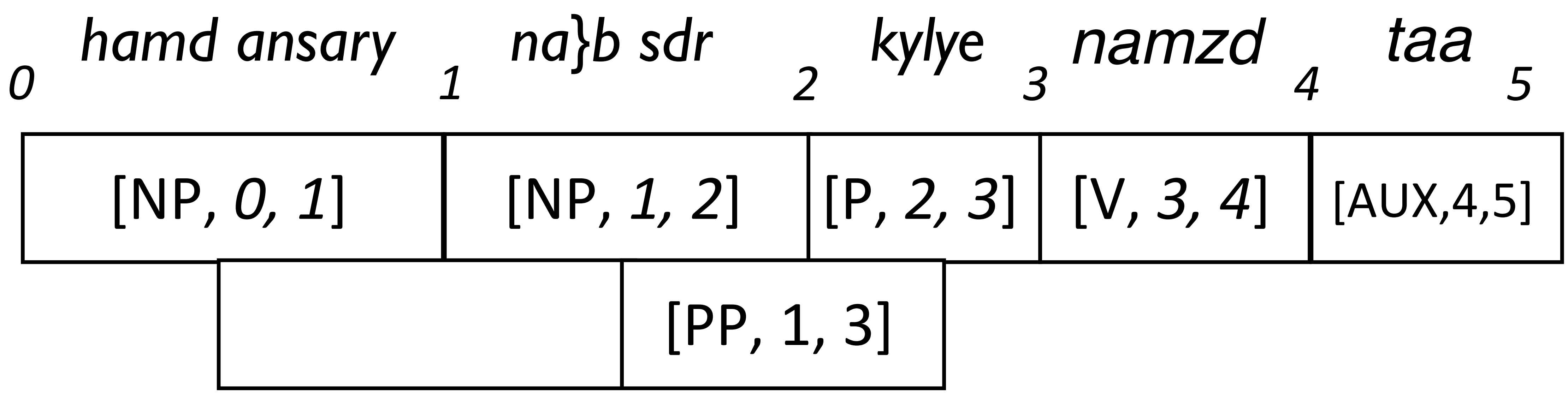
S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

$$\frac{[\mathrm{NP}, \mathit{1}, \mathit{2}] \quad [\mathrm{P}, \mathit{2}, \mathit{3}] \quad \mathrm{PP} \rightarrow \mathrm{NP}\ \mathrm{P}}{[\mathrm{PP}, \mathit{1}, \mathit{3}]}$$

Goal

[S, 0, 5]



Axioms

S → NP VP
VP→ PP VP
VP→ V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]	[V, *3, 4*]	[AUX,4,5]

	[PP, 1, 3]

Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*] [NP, *1, 2*] [P, *2, 3*] [V, *3, 4*] [AUX,4,5]

[PP, 1, 3]

Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

$$\frac{[\mathrm{V}, \mathit{3}, \mathit{4}] \quad [\mathrm{AUX}, \mathit{4}, \mathit{5}] \quad \mathrm{VP} \rightarrow \mathrm{V\ AUX}}{[\mathrm{VP}, \mathit{3}, \mathit{5}]}$$

Goal

[S, 0, 5]

0 *hamd ansary* *1* *na}b sdr* *2* *kylye* *3* *namzd* *4* *taa* *5*

[NP, *0, 1*]	[NP, *1, 2*]	[P, *2, 3*]	[V, *3, 4*]	[AUX,4,5]

[PP, 1, 3]

Axioms

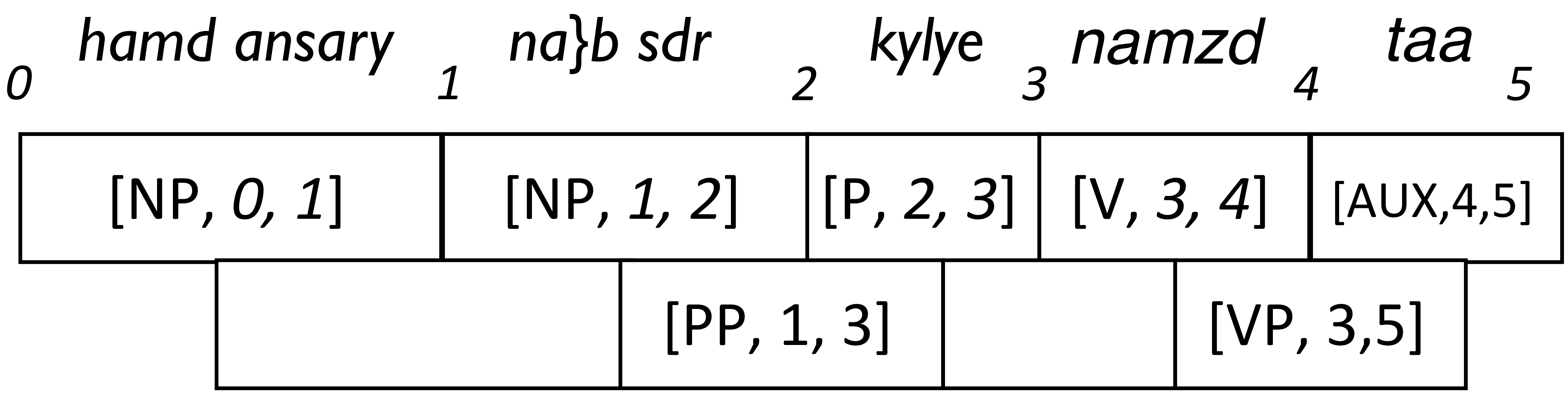
S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

$$\frac{[\mathrm{V}, 3, 4] \quad [\mathrm{AUX}, 4, 5] \quad \mathrm{VP} \rightarrow \mathrm{V\ AUX}}{[\mathrm{VP}, 3, 5]}$$

Goal

[S, 0, 5]



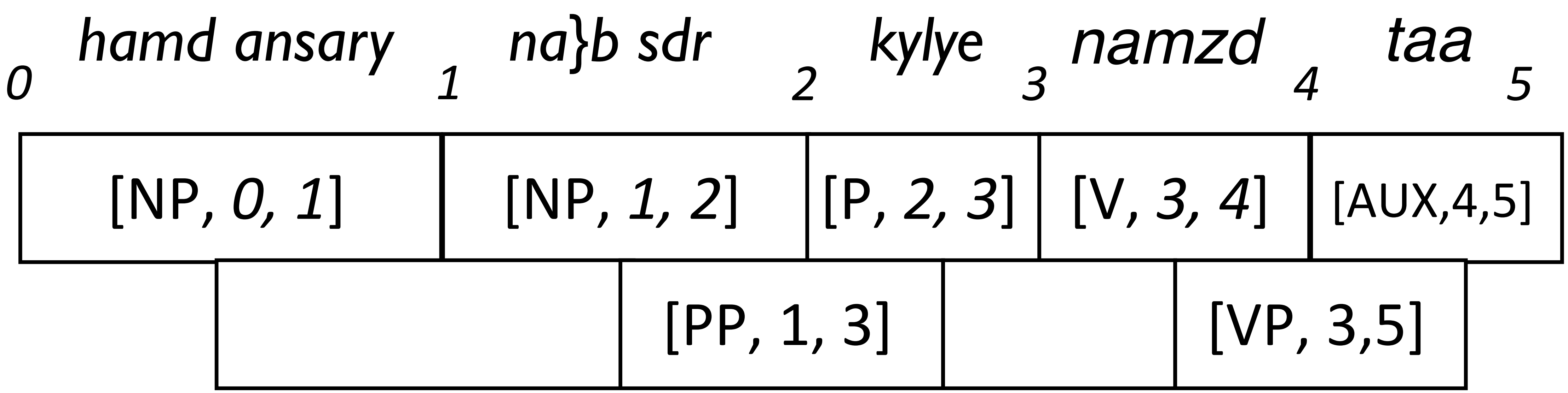
Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

Goal

[S, 0, 5]



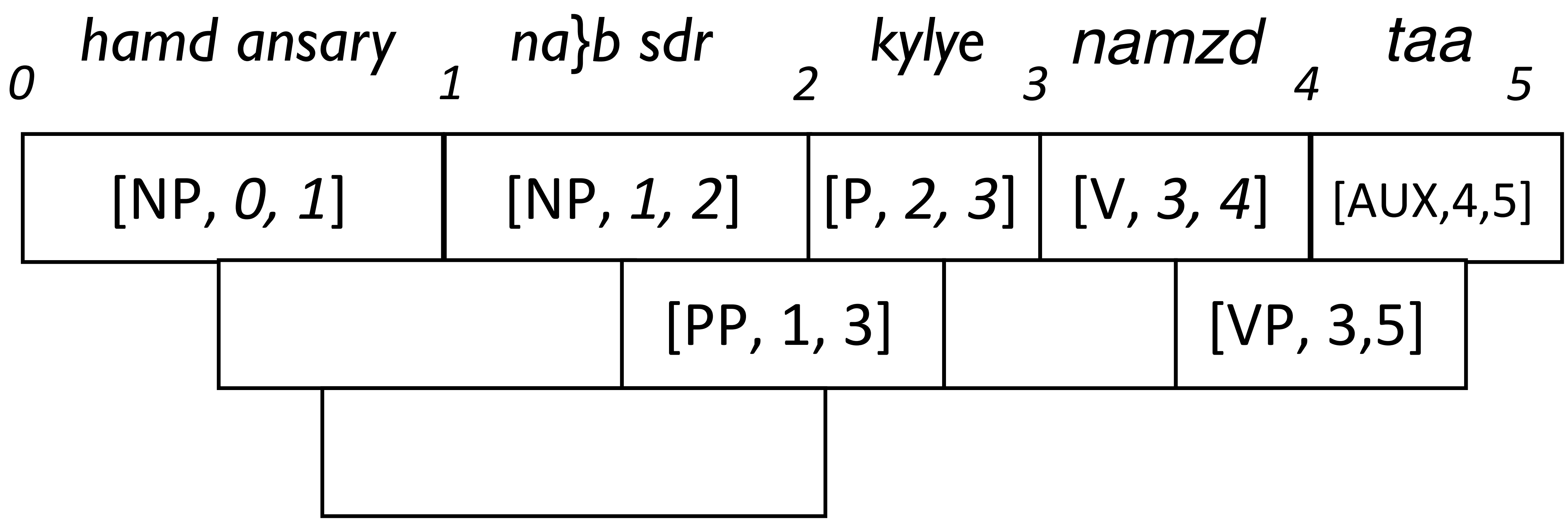
Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

Goal

[S, 0, 5]



Axioms

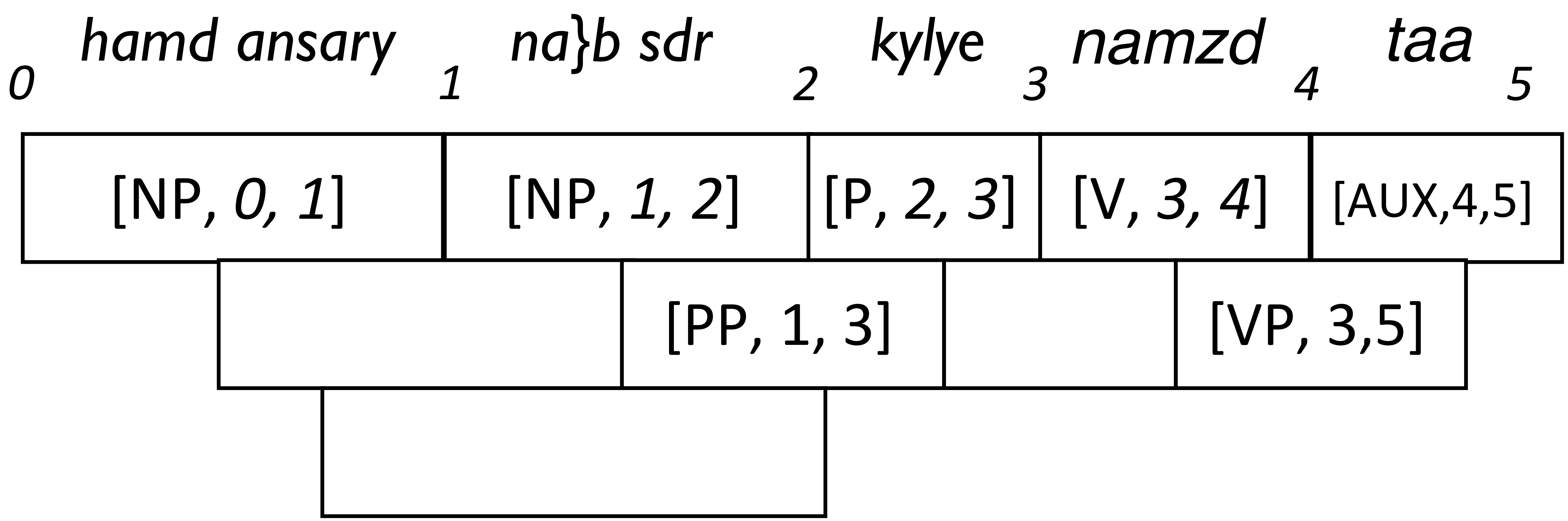
S	→	NP VP
VP	→	PP VP
VP	→	V AUX
PP	→	NP P
NP	→	*hamd ansary*
NP	→	*na}b sdr*
V	→	*namzd*
P	→	*kylye*
AUX	→	*taa*

Inference rule used

$$\frac{[\mathrm{PP}, \mathit{1}, \mathit{3}] \quad [\mathrm{VP}, \mathit{3}, \mathit{5}] \quad \mathrm{VP} \to \mathrm{PP\ CP}}{[\mathrm{VP}, \mathit{1}, \mathit{5}]}$$

Goal

[S, 0, 5]



Axioms

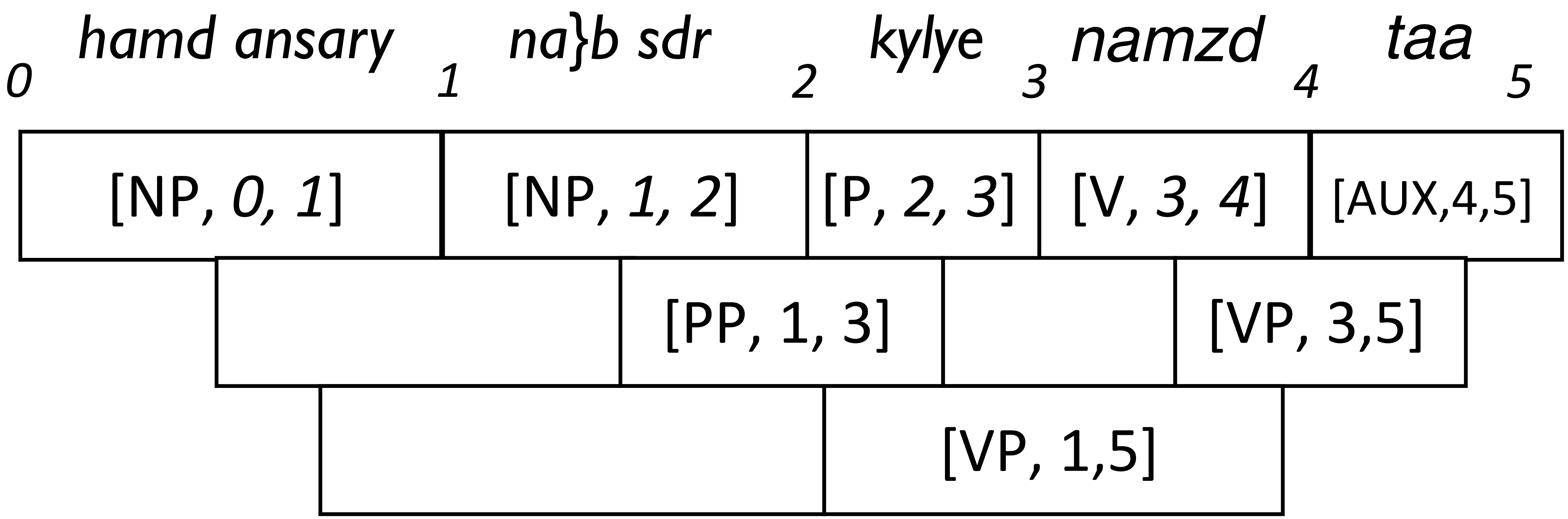
S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

$$\frac{[\mathrm{PP}, \mathit{1}, \mathit{3}]\quad [\mathrm{VP}, \mathit{3}, \mathit{5}]\quad \mathrm{VP} \rightarrow \mathrm{PP\ CP}}{[\mathrm{VP}, \mathit{1}, \mathit{5}]}$$

Goal

[S, 0, 5]



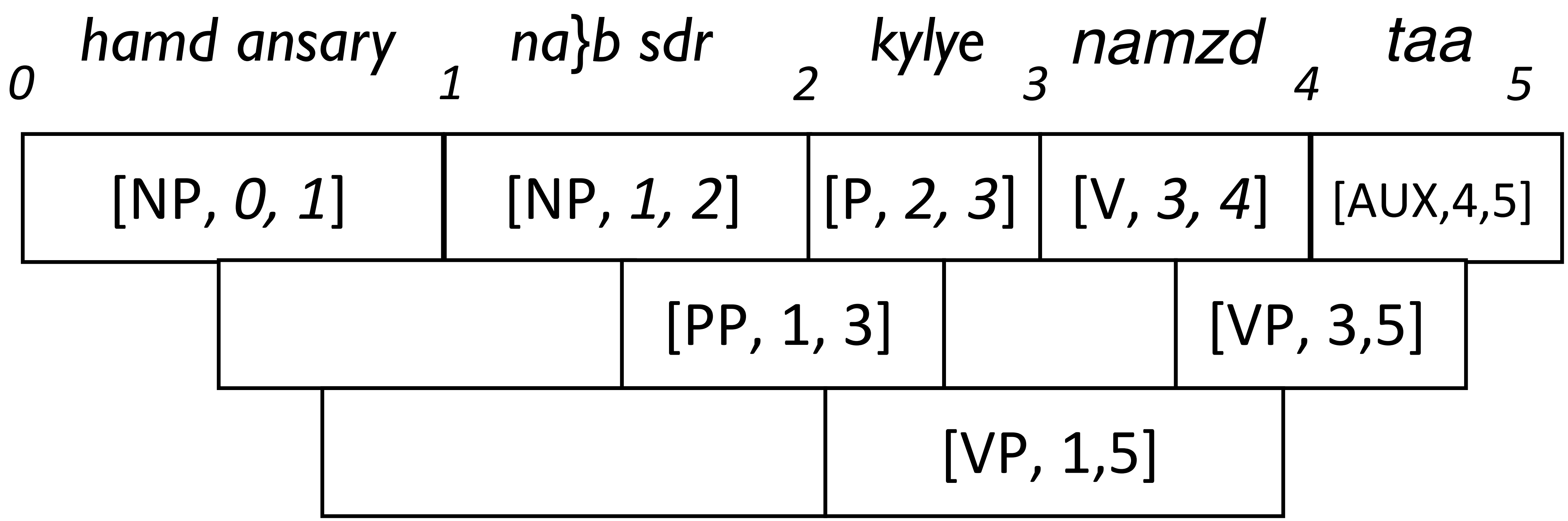
Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

Goal

[S, 0, 5]



Axioms

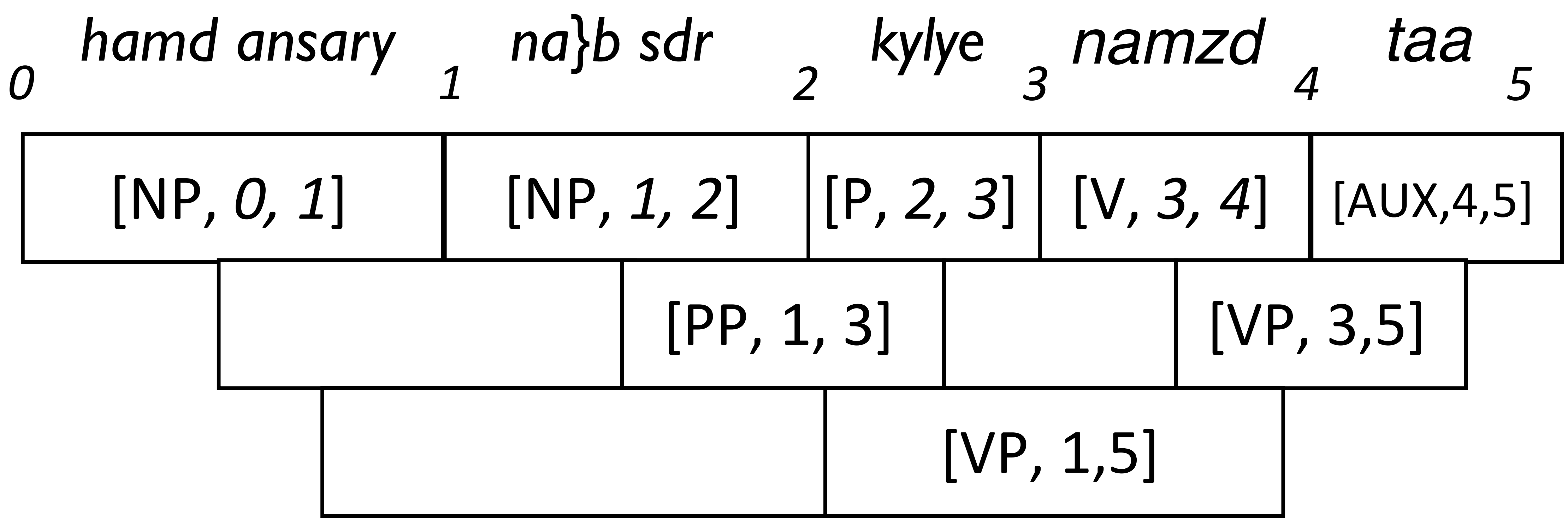
S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

$$\frac{[NP, 0, 1]\quad [VP, 1, 5]\quad S \rightarrow NP\ VP}{[S, 0, 5]}$$

Goal

[S, 0, 5]



Axioms

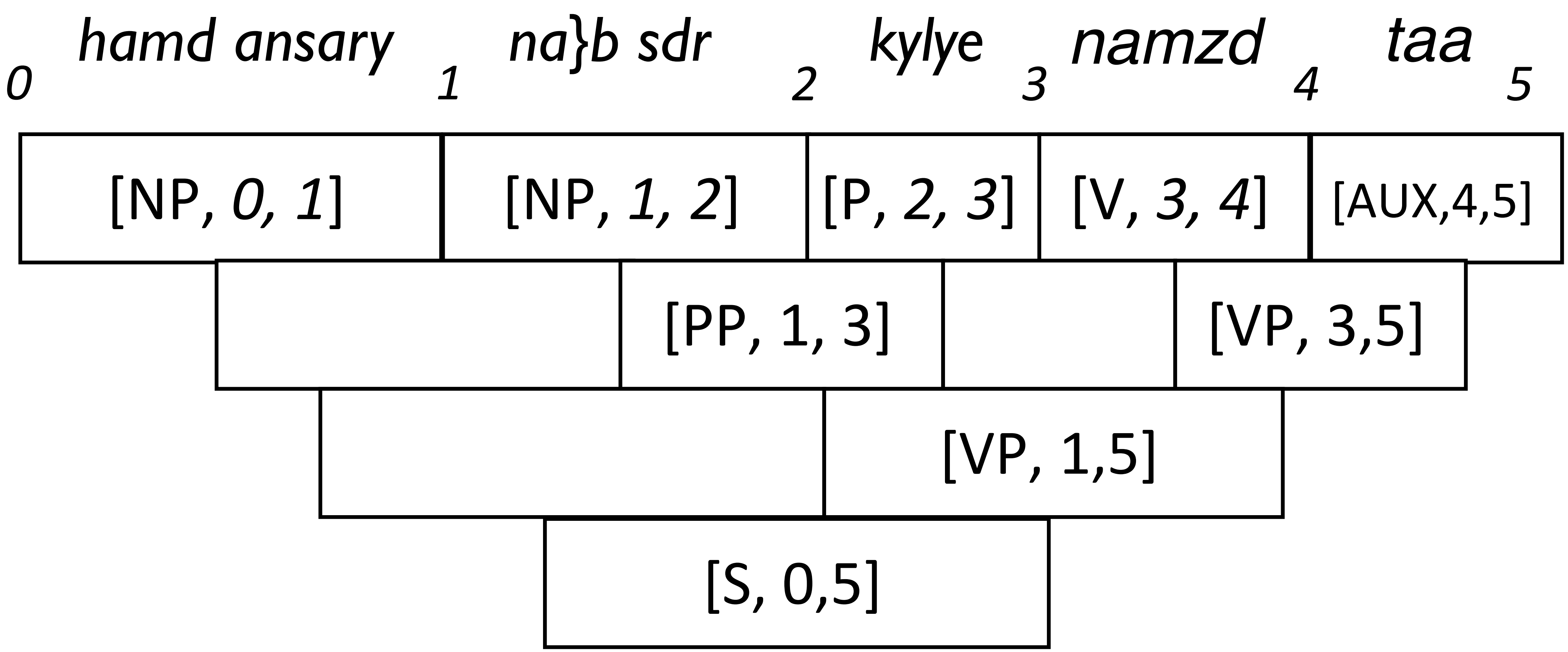
S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

$$\frac{[NP, \mathit{0}, \mathit{1}] \quad [VP, \mathit{1}, \mathit{5}] \quad S \rightarrow NP\ VP}{[S, \mathit{0}, \mathit{5}]}$$

Goal

[S, 0, 5]



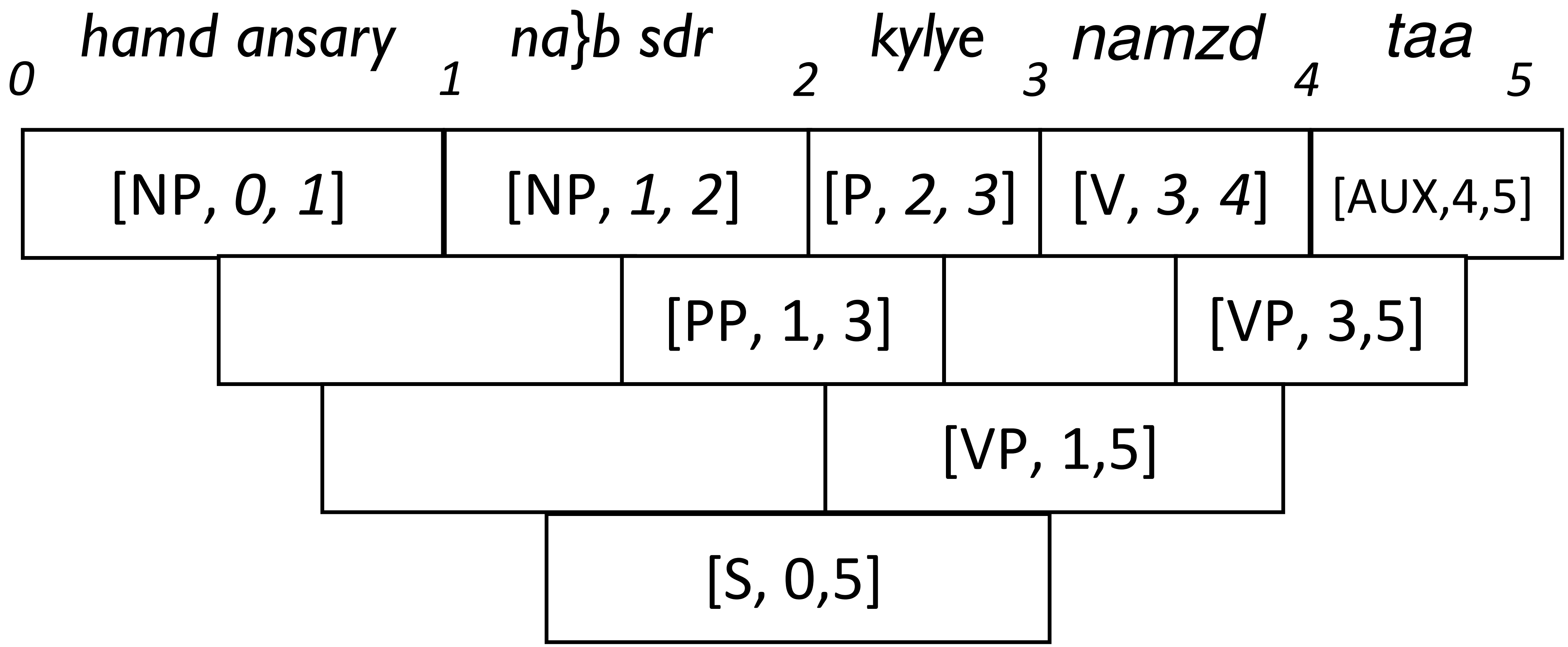
Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → *hamd ansary*
NP → *na}b sdr*
V → *namzd*
P → *kylye*
AUX → *taa*

Inference rule used

Goal

[S, 0, 5]



The CKY Parsing Algorithm

Axioms	$\dfrac{}{A \rightarrow \alpha}$	for all $(A \rightarrow \alpha) \in R$
Inference rules	$\dfrac{A \rightarrow w_{i+1}}{[A, i, i+1]}$ $\dfrac{[B, i, j] \quad [C, j, k] \quad A \rightarrow BC}{[A, i, k]}$	
Goal	$[S, 0, n]$	

The CKY Translation Algorithm

Axioms	$$\frac{}{\mathbf{A \rightarrow \alpha, \beta}}$$	for all $(\mathbf{A \rightarrow \alpha, \beta}) \in R$
Inference rules	$$\frac{A \rightarrow w_{i+1}}{[A, i, i+1]}$$ $$\frac{[B, i, j] \quad [C, j, k] \quad A \rightarrow BC}{[A, i, k]}$$	
Goal	$[S, 0, n]$	

Learning SCFGs from parallel text

Where do grammars come from?

- Great! We now have
  - a formalism for describing the relationship between two languages,
  - an algorithm for producing translations
- All we need now is a synchronous grammar

Where do grammars come from?

- Great! We now have
  - a formalism for describing the relationship between two languages,
  - an algorithm for producing translations
- All we need now is a synchronous grammar
- Where do grammars come from?
- Well, when two languages love each other very much...

Data-driven grammar extraction

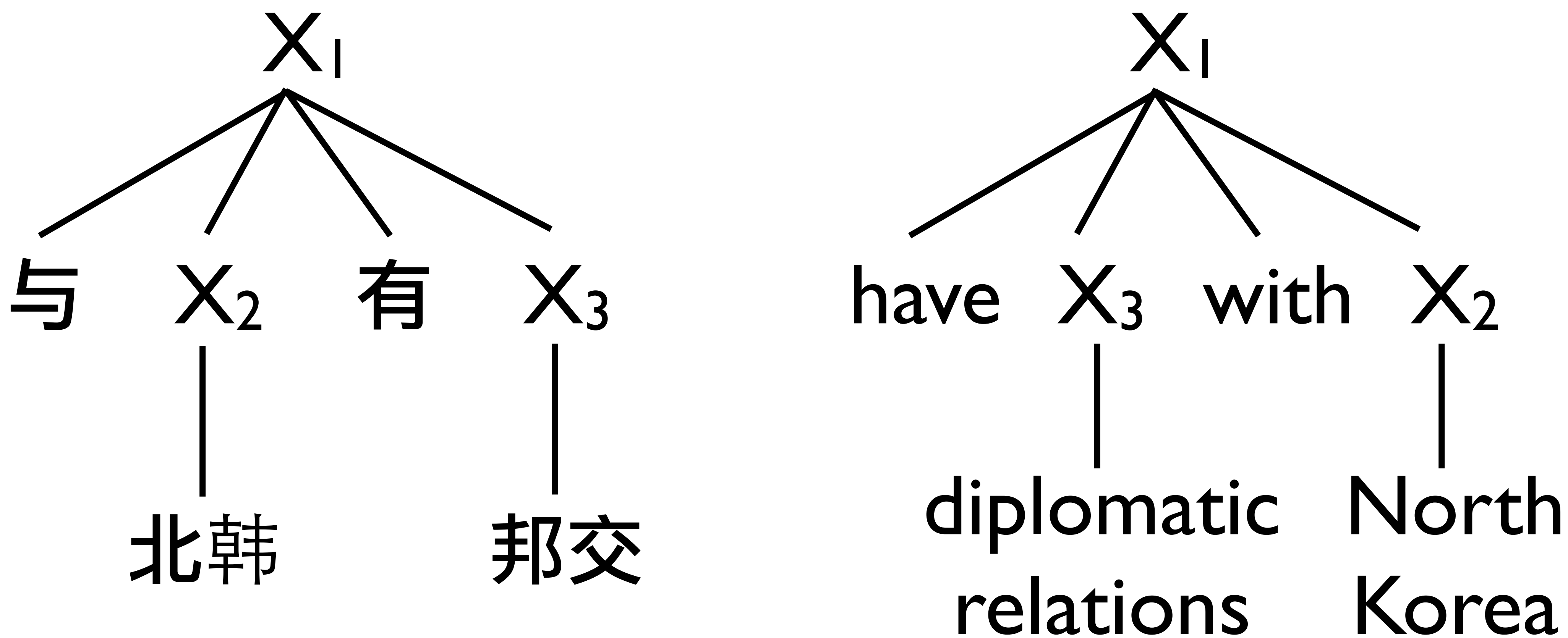
- Grammar rules are not written by hand, they are extracted from bilingual parallel corpora

Arabic	English
فالتعذيب لا يزال يمارس على نطاق واسع	Torture is still being practised on a wide scale.
وتتم عمليات الاعتقال والاحتجاز دون سبب بصورة روتينية	Arrest and detention without cause take place routinely.
وحان وقت التحلى بالبصيرة والشجاعة السياسية .	This is a time for vision and political courage
. . .	. . .

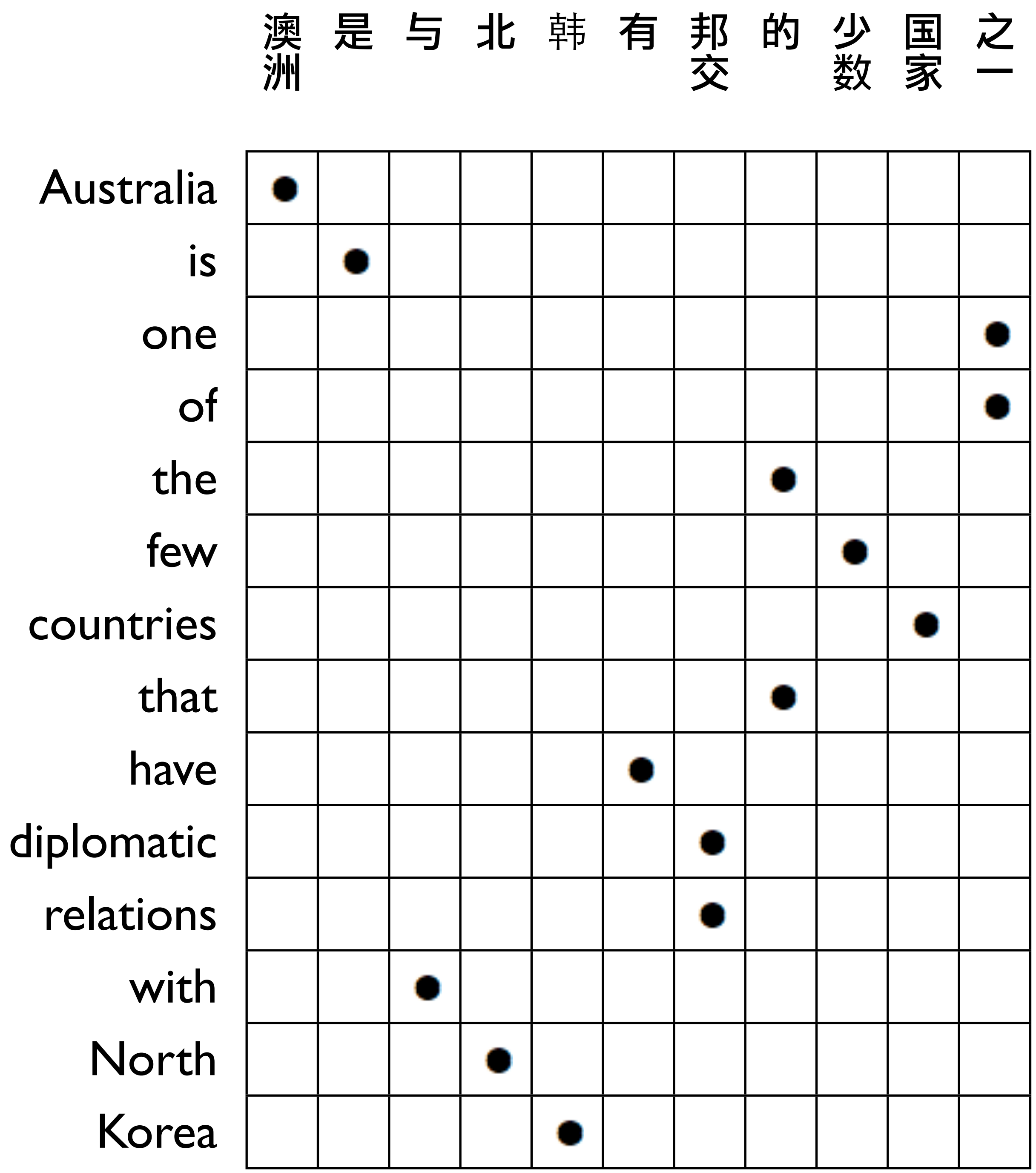
Chinese	English
我国 能源 原材料 工业 生产 大幅度 增长 .	China's energy and raw materials production up.
非国大 要求 阻止 更 多 被 拘留 人员 死亡 .	ANC calls for steps to prevent deaths in police custody .
. . .	. . .

Hiero-style SCFG rules

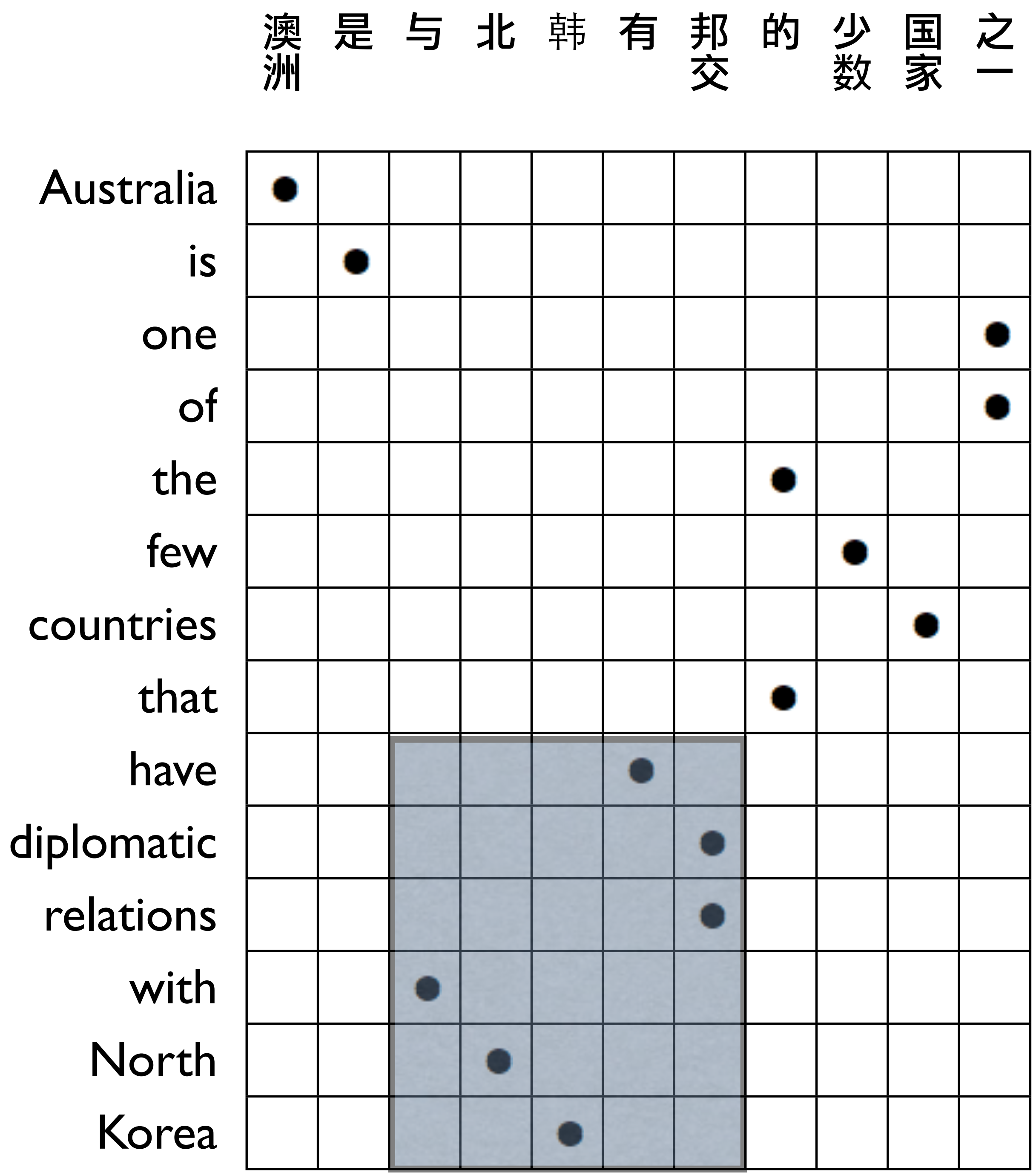
- Most common type of SCFG in SMT is Hiero which has rules w/one non-terminal symbol
- Not as nice as linguistically motivated rules, does not capture the reordering in Urdu



Extracting Hiero rules

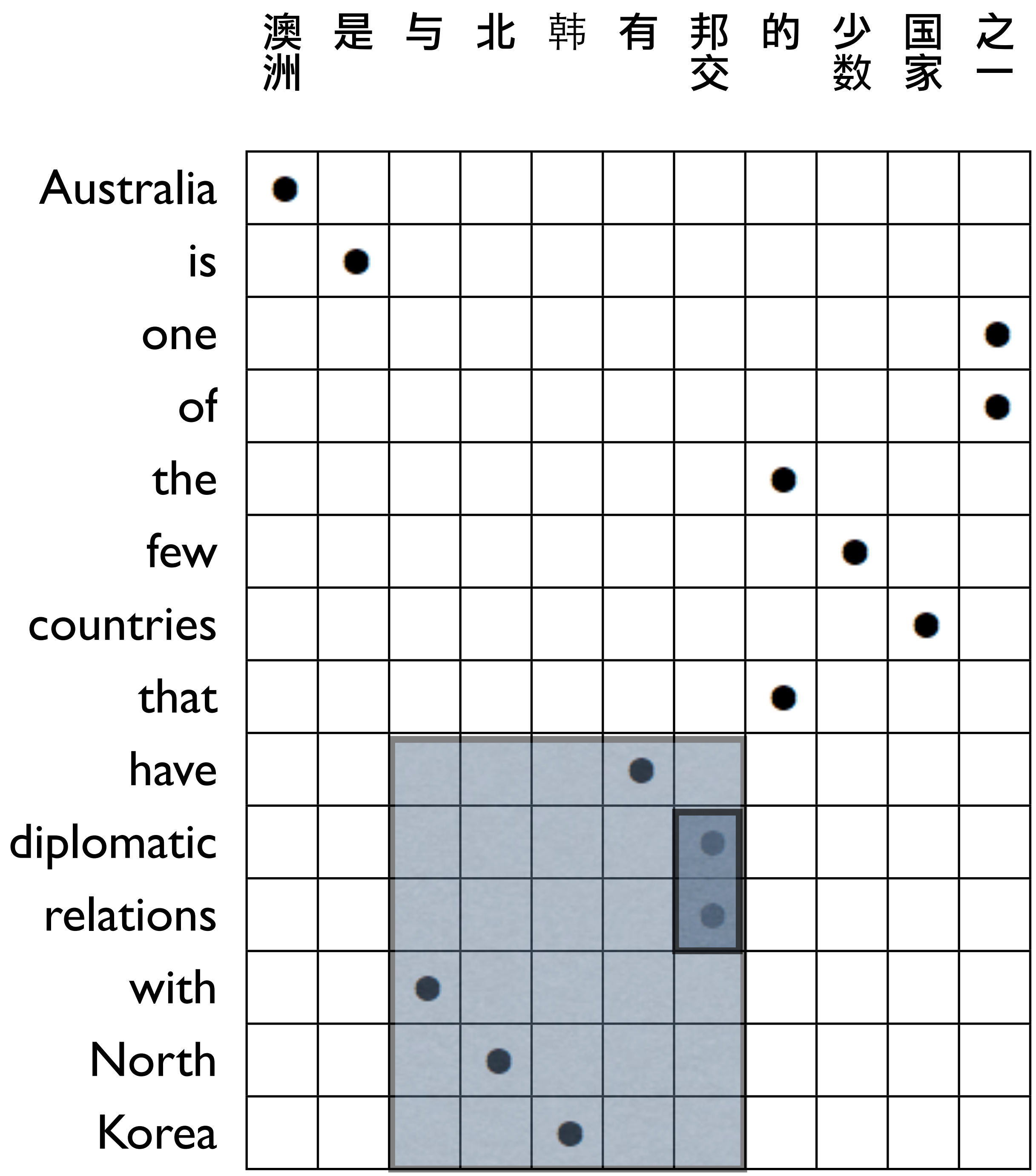


Extracting Hiero rules



X → 与 北 韩 有 邦交,
have diplomatic relations
with North Korea

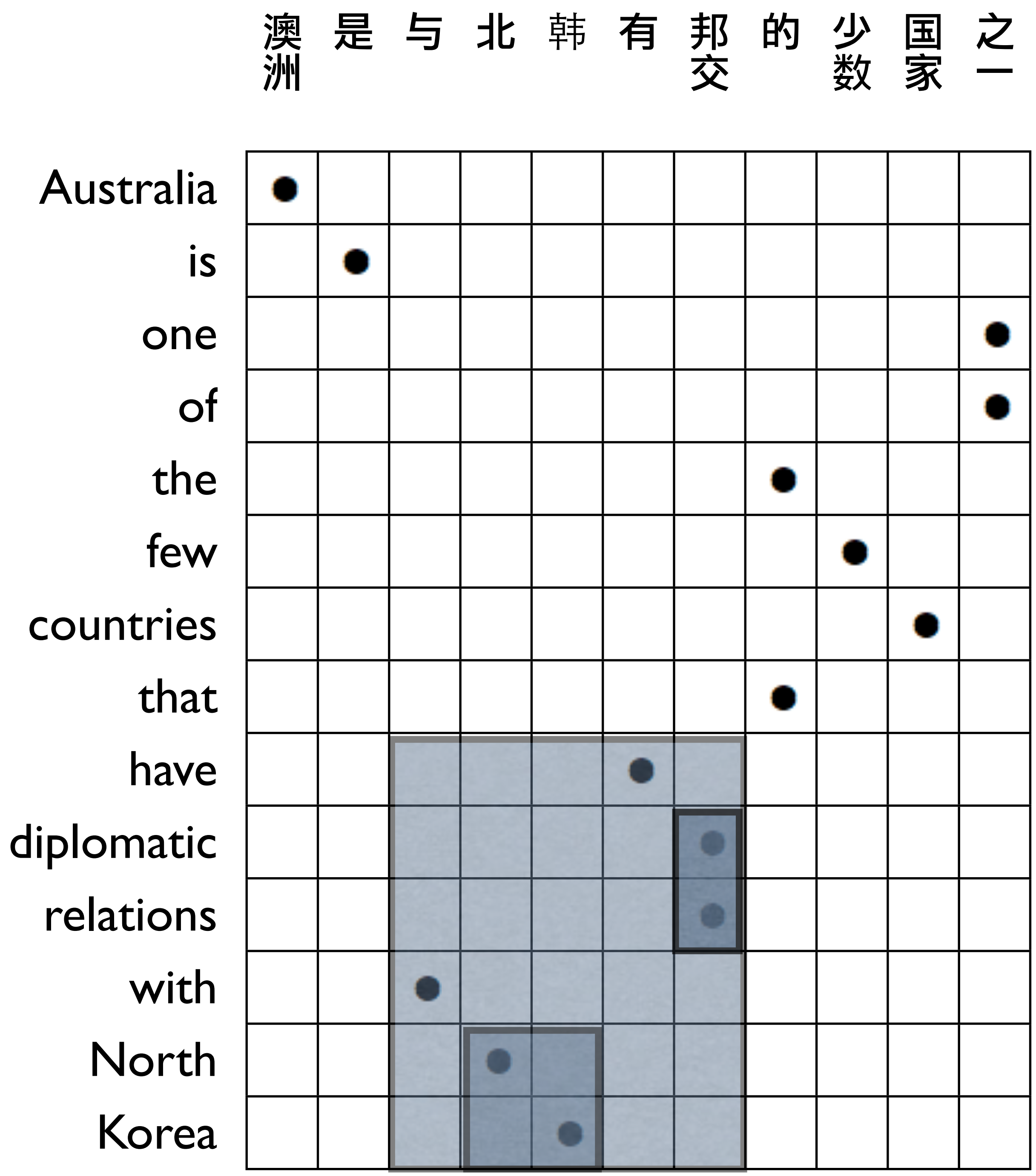
Extracting Hiero rules



X → 与 北 韩 有 邦交, have diplomatic relations with North Korea

X → 邦交, diplomatic relations

Extracting Hiero rules

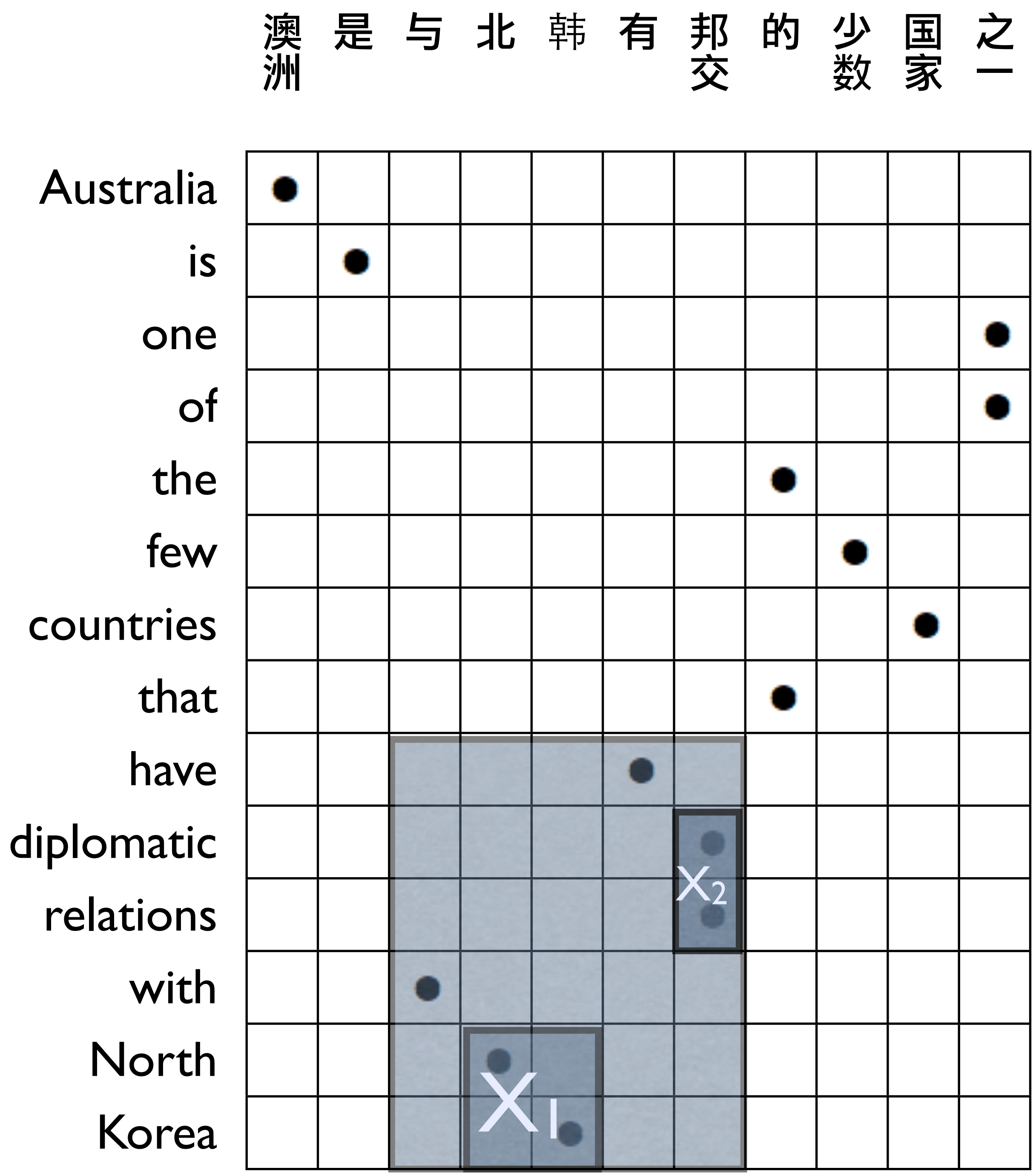


X → 与 北 韩 有 邦交, have diplomatic relations with North Korea

X → 邦交, diplomatic relations

X → 北 韩, North Korea

Extracting Hiero rules

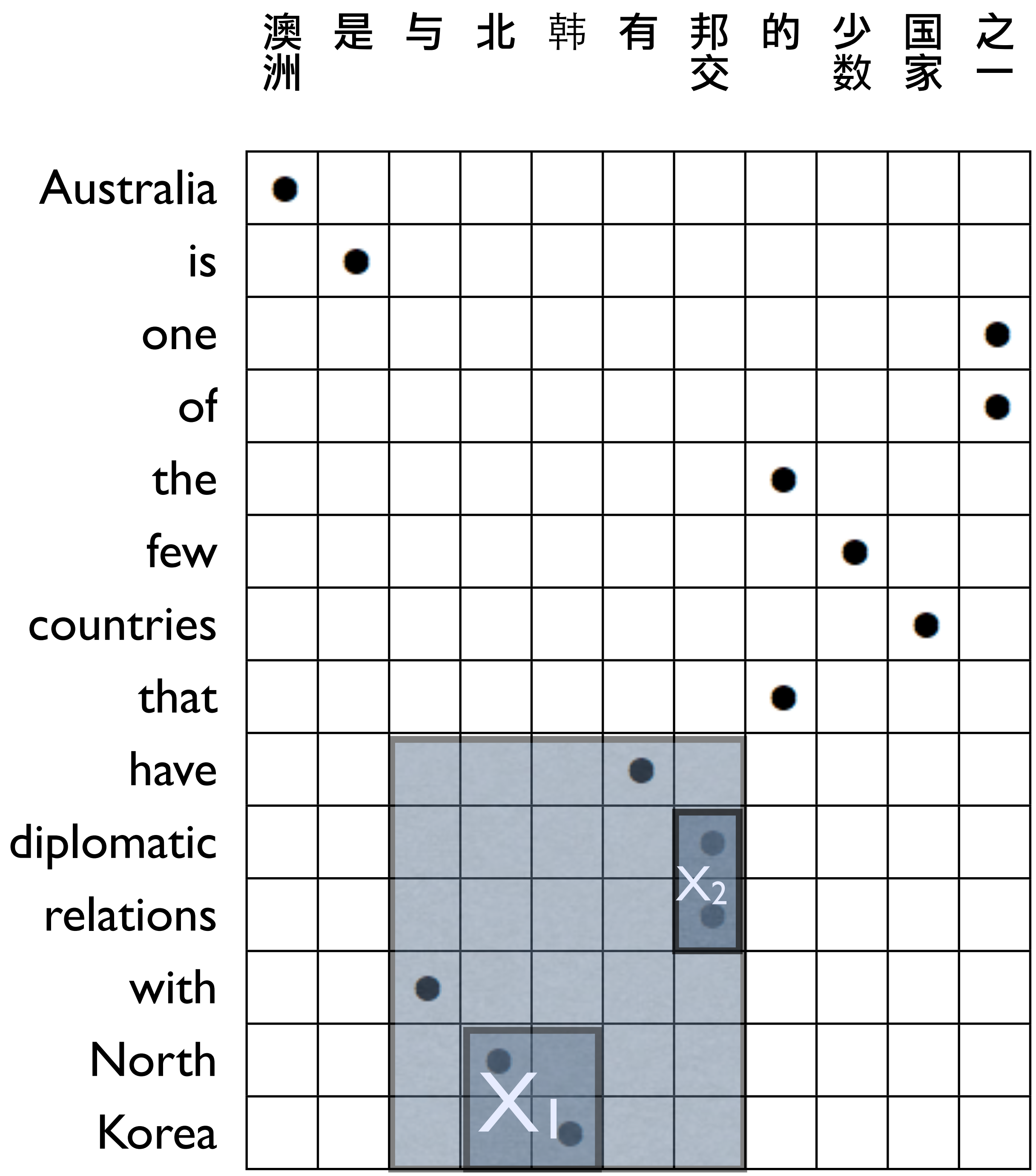


X → 与 北 韩 有 邦交, have diplomatic relations with North Korea

X → 邦交, diplomatic relations

X → 北 韩, North Korea

Extracting Hiero rules



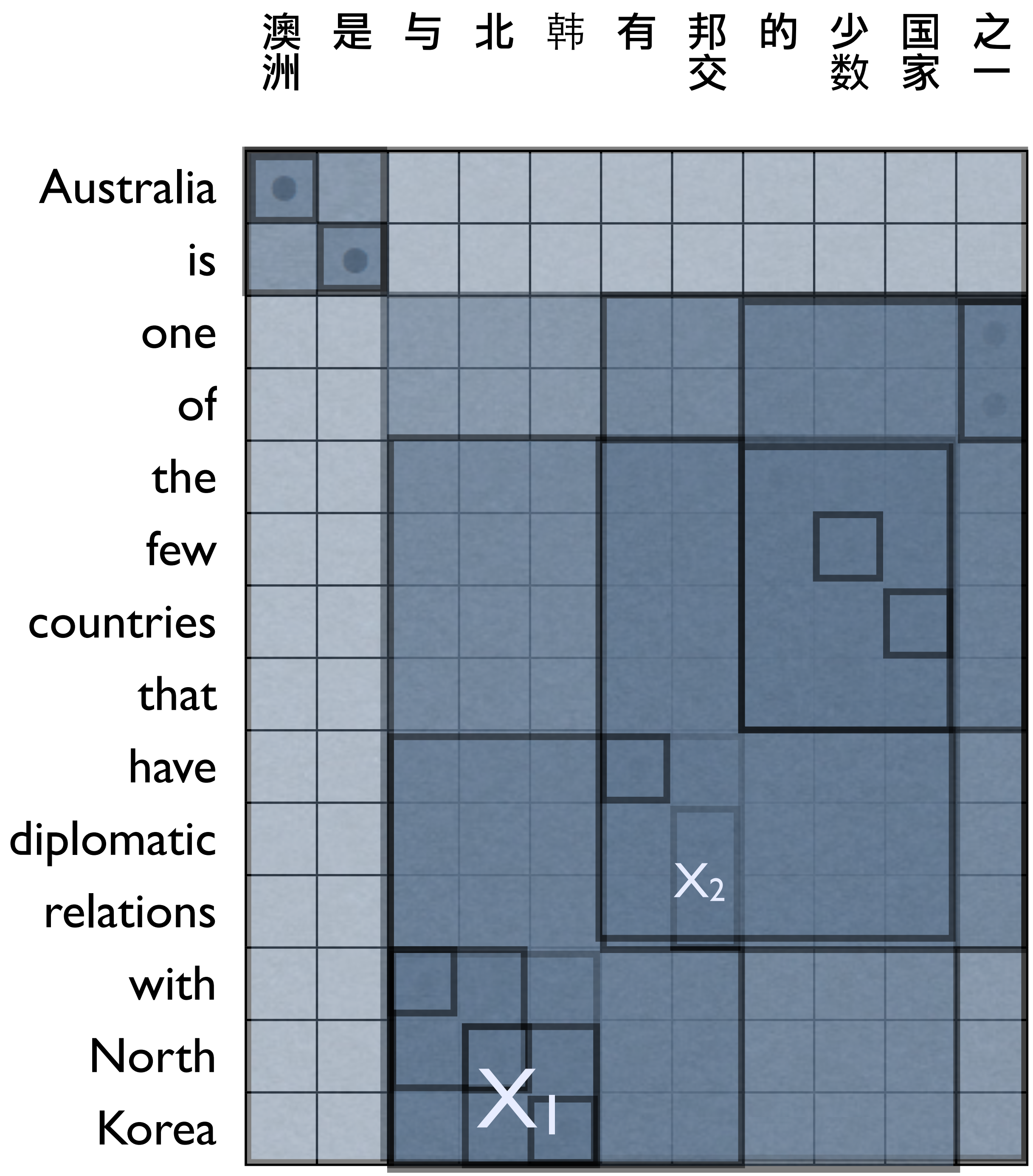
X → 与 北 韩 有 邦交, have diplomatic relations with North Korea

X → 邦交, diplomatic relations

X → 北 韩, North Korea

X → 与 X_1 有 X_2, have X_2 with X_1

Extracting Hiero rules



X → 与 北 韩 有 邦交, have diplomatic relations with North Korea

X → 邦交, diplomatic relations

X → 北 韩, North Korea

X → 与 X_1 有 X_2, have X_2 with X_1

Discussion: what do you think of Hiero?

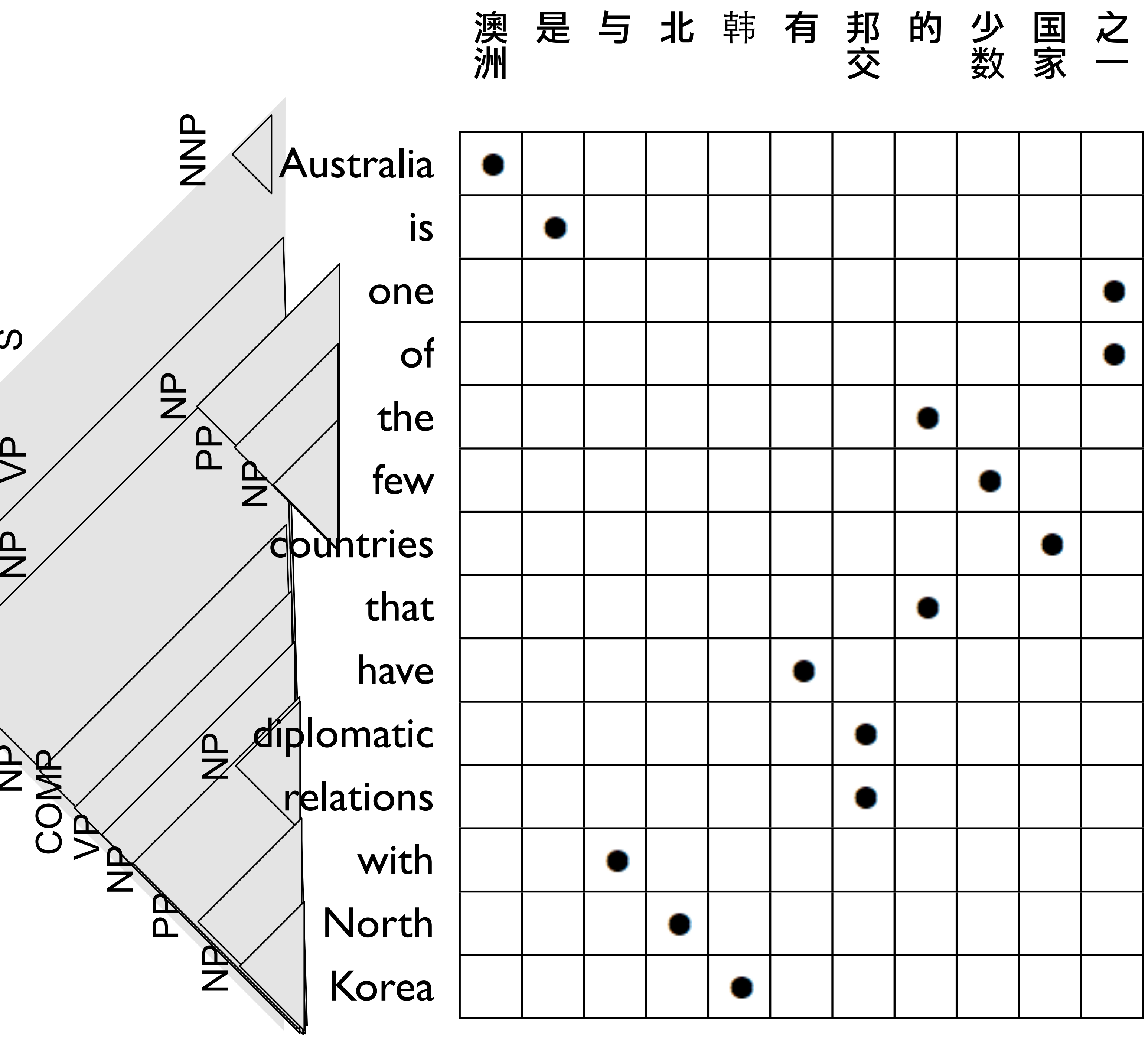
- So, we now have a way of extracting SCFGs from bitexts. Great! So what?
- Is this any better than the phrase based model?
- How?
- Do you feel that it is lacking anything?

(Discuss with your neighbor)

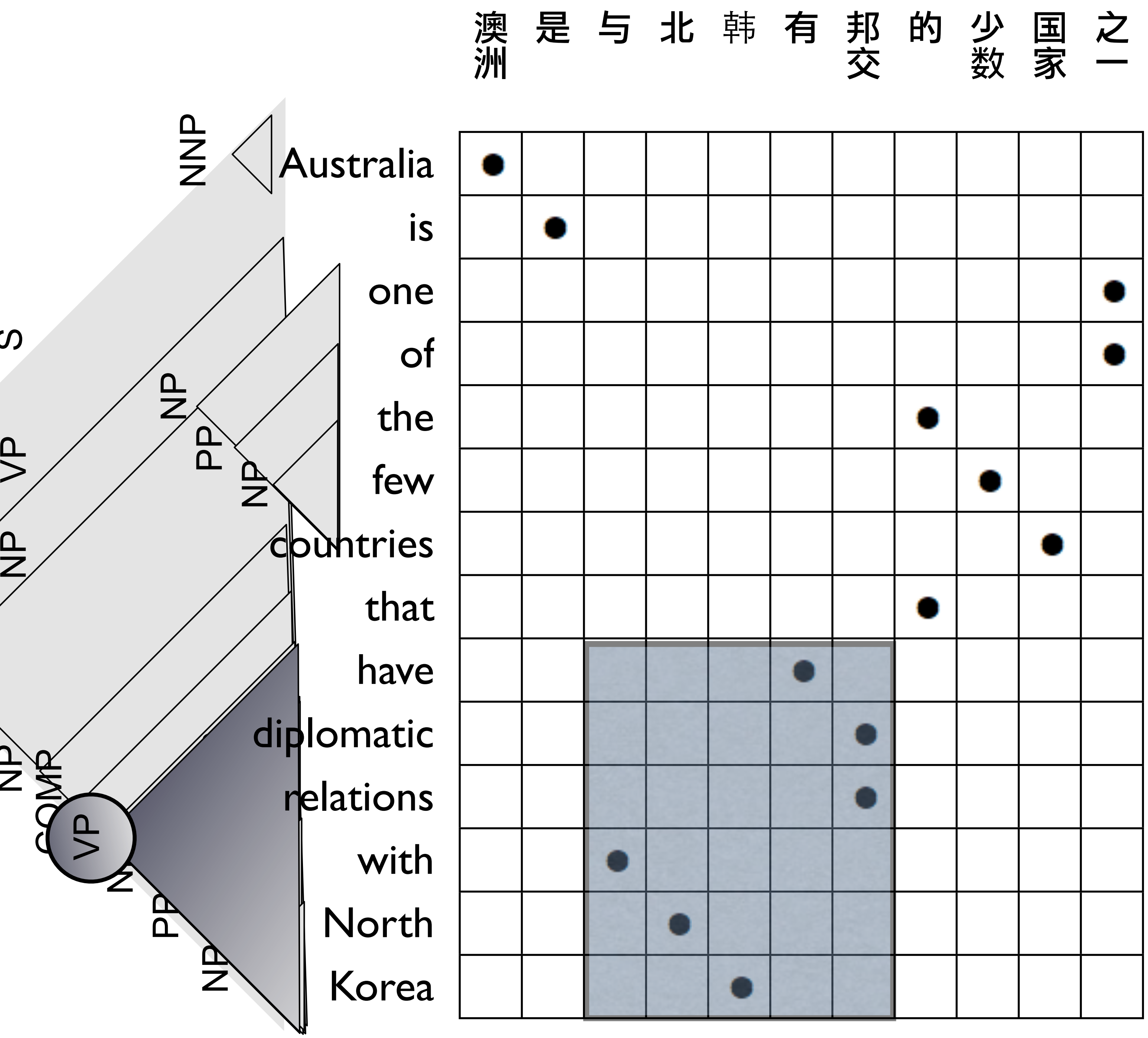
Extracting Syntactic Rules

	澳洲	是	与	北	韩	有	邦交	的	少数	国家	之一
Australia	●										
is		●									
one											●
of											●
the								●			
few									●		
countries										●	
that								●			
have						●					
diplomatic							●				
relations							●				
with			●								
North				●							
Korea					●						

Extracting Syntactic Rules

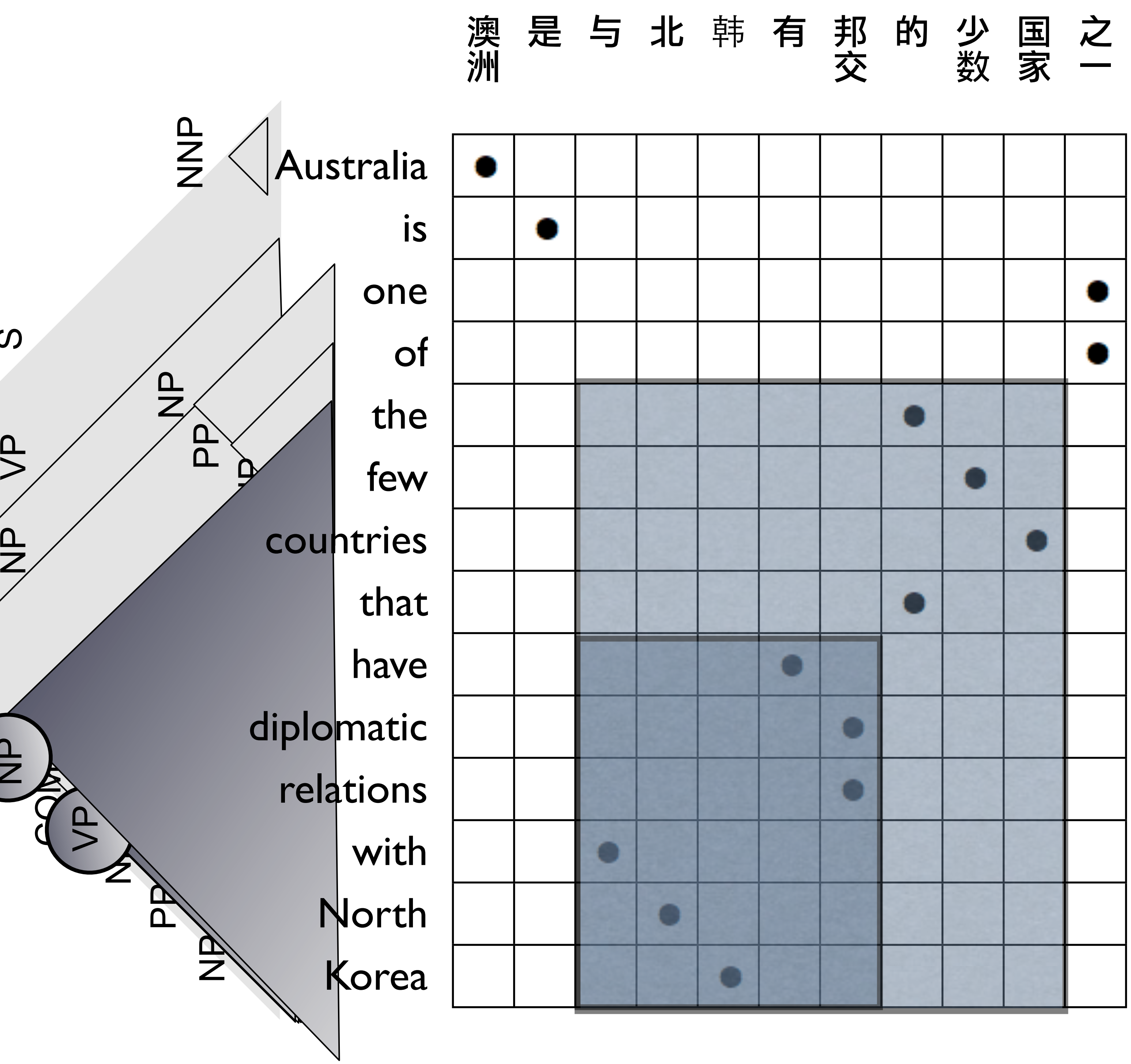


Extracting Syntactic Rules



VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

Extracting Syntactic Rules

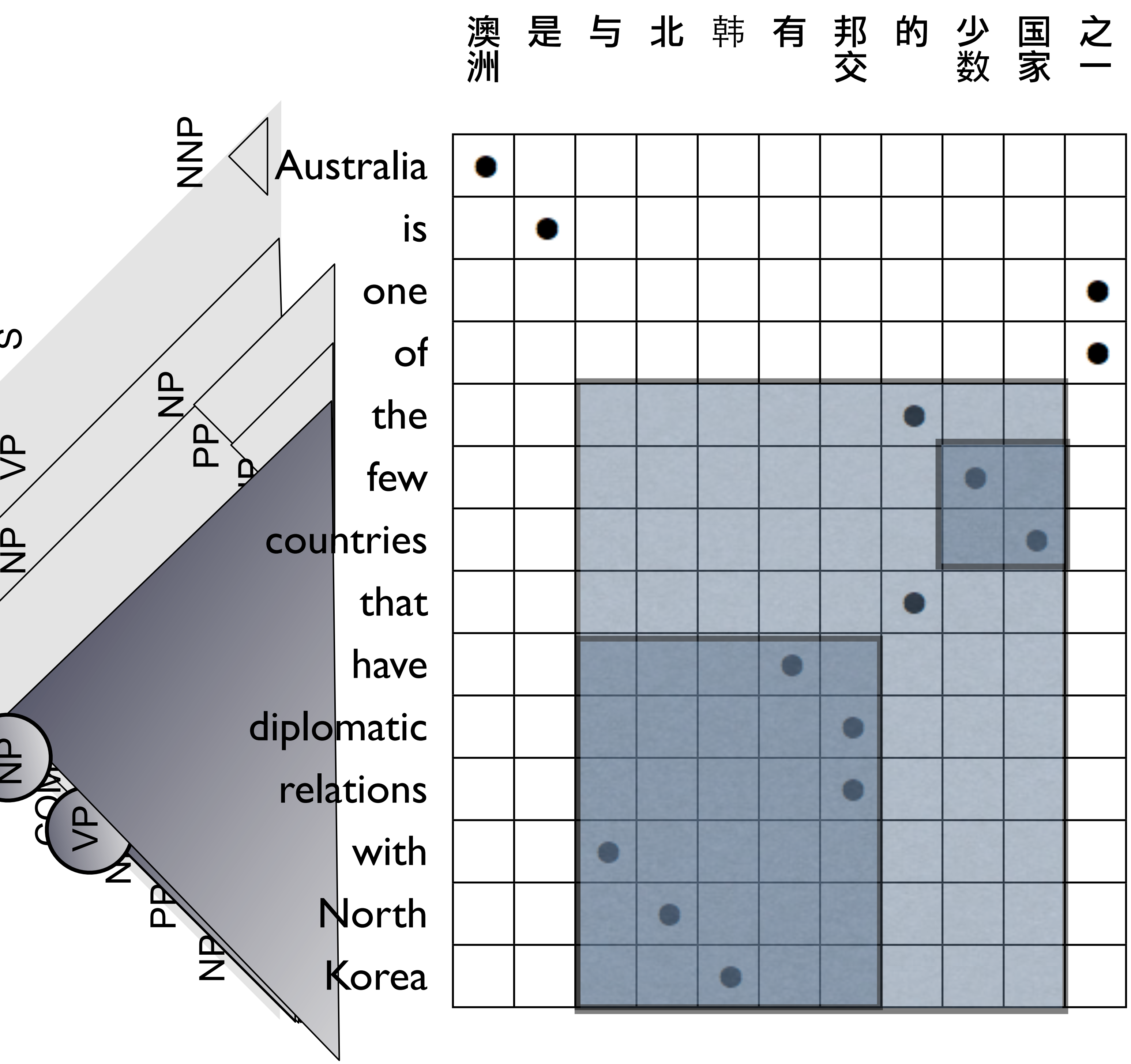


VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

NP → 与 北 韩 有 邦交 的 少数 国家, the few countries that have diplomatic relations with North Korea

NP → VP 的 少数 国家, the few countries that VP

Extracting Syntactic Rules



VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

NP → 与 北 韩 有 邦交 的 少数 国家, the few countries that have diplomatic relations with North Korea

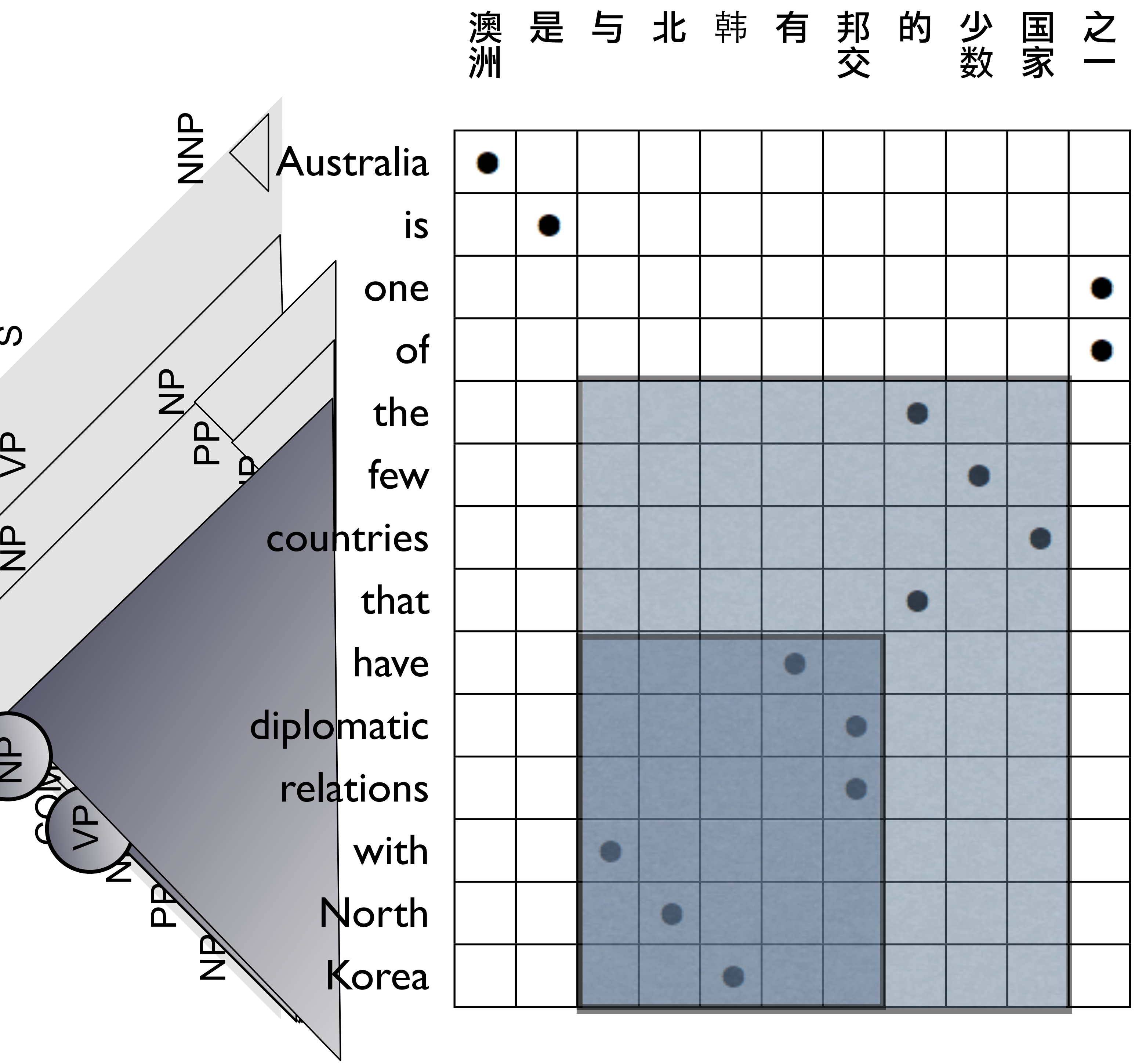
NP → VP 的 少数 国家, the few countries that VP

NP → VP 的 NP, the NP that VP

Wait a minute...

- Didn’t we see this earlier in Koehn’s paper?
- Aren’t we giving up a ton of rules that you said were valuable?
- Something about a reduced inventory because we got rid of non-constituent phrases?

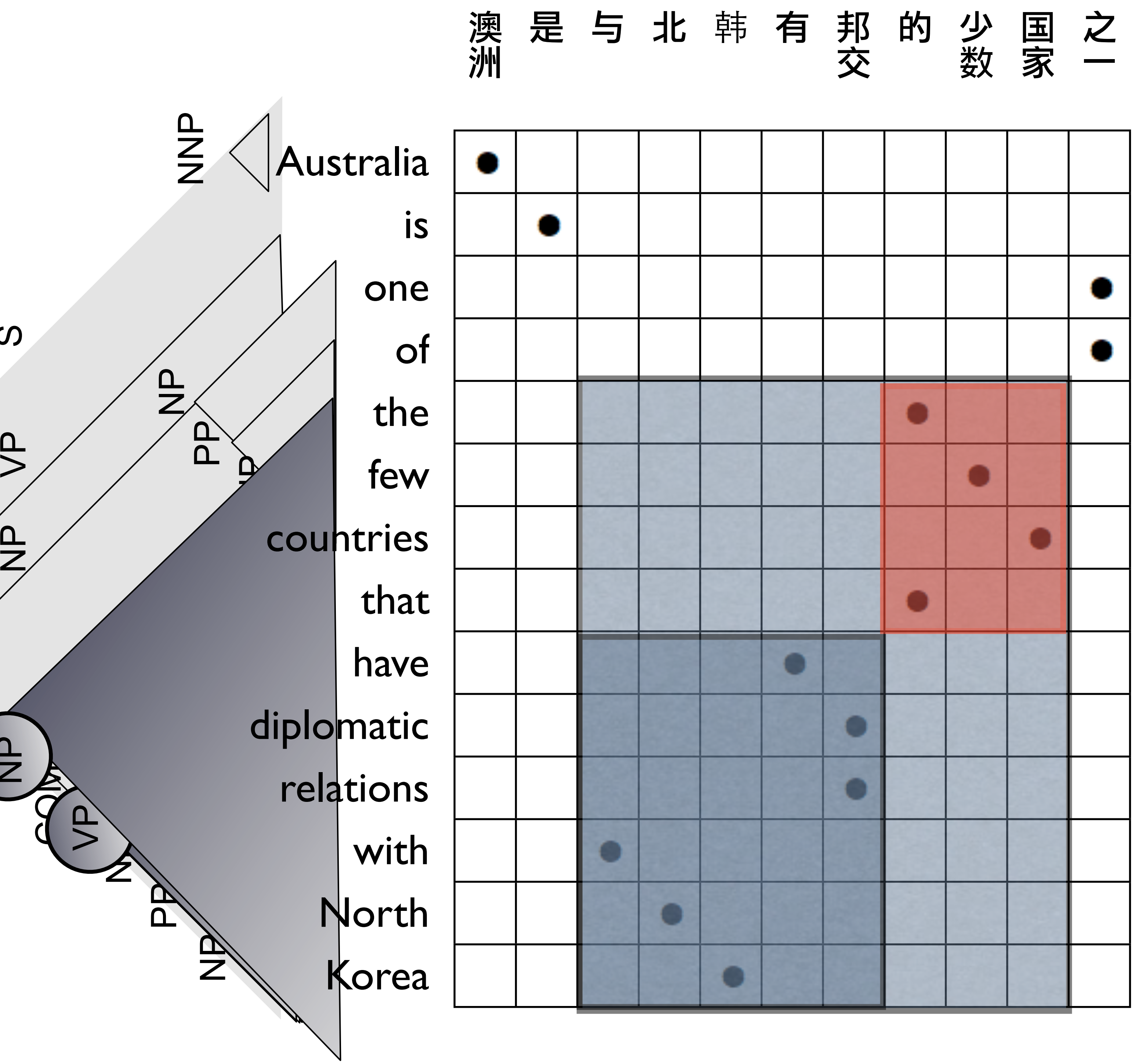
Extracting Syntactic Rules



VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

NP → 与 北 韩 有 邦交 的 少数 国家, the few countries that have diplomatic relations with North Korea

Extracting Syntactic Rules

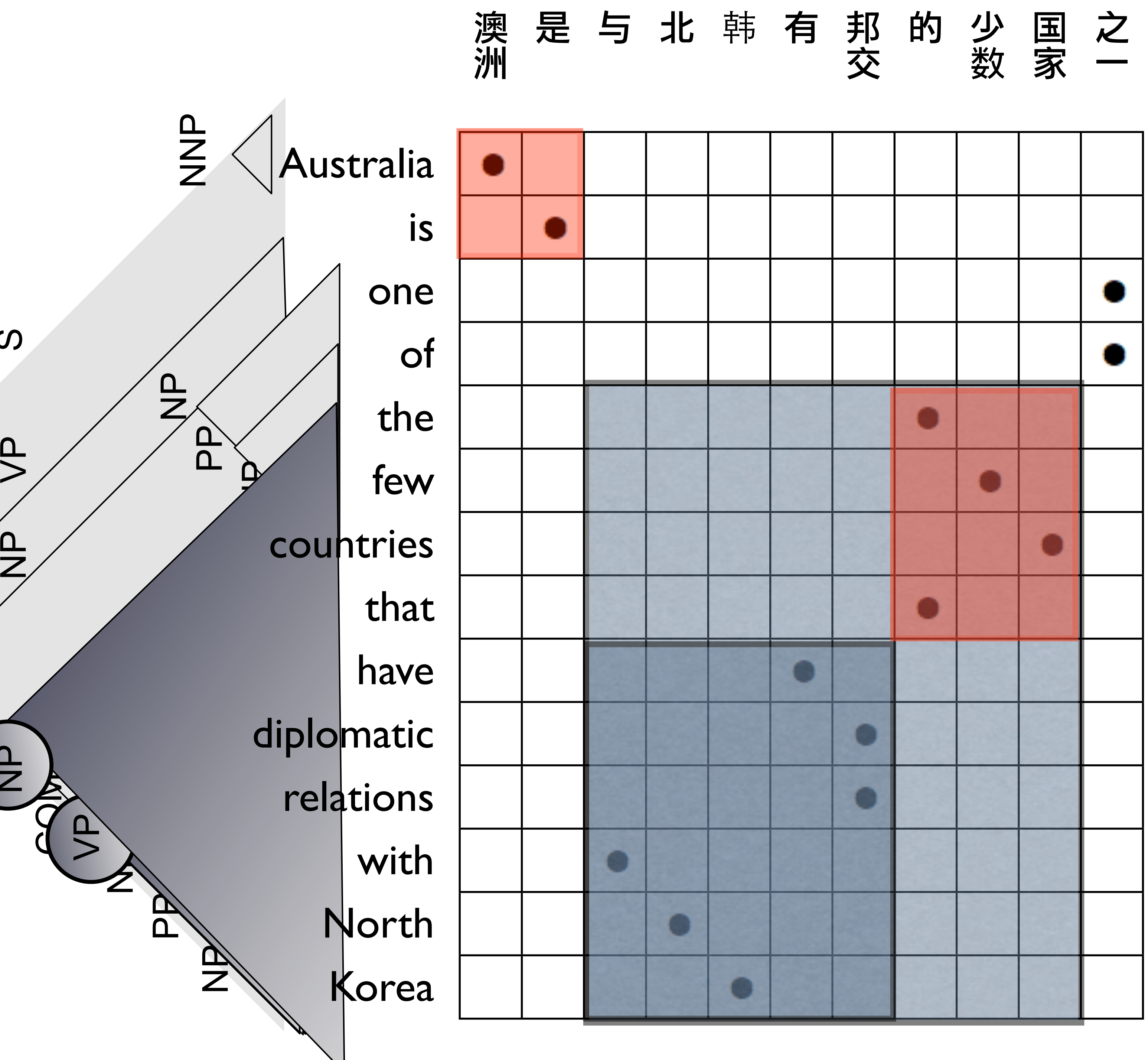


VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

NP → 与 北 韩 有 邦交 的 少数 国家, the few countries that have diplomatic relations with North Korea

??? → 的 少数 国家, the few countries that

Extracting Syntactic Rules



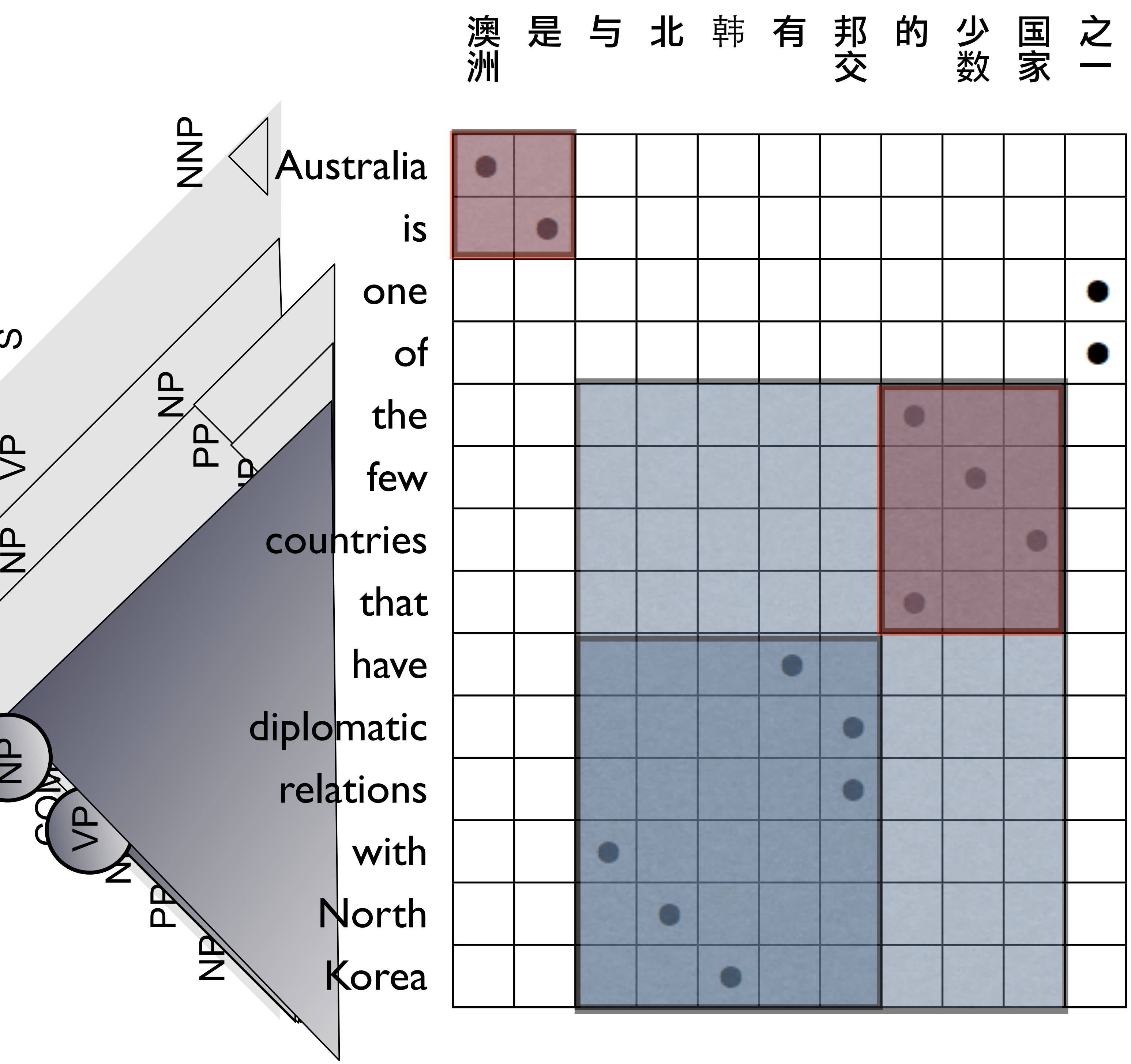
VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

NP → 与 北 韩 有 邦交 的 少数 国家, the few countries that have diplomatic relations with North Korea

??? → 的 少数 国家, the few countries that

??? → 澳洲 是, Australia is

Extracting Syntactic Rules



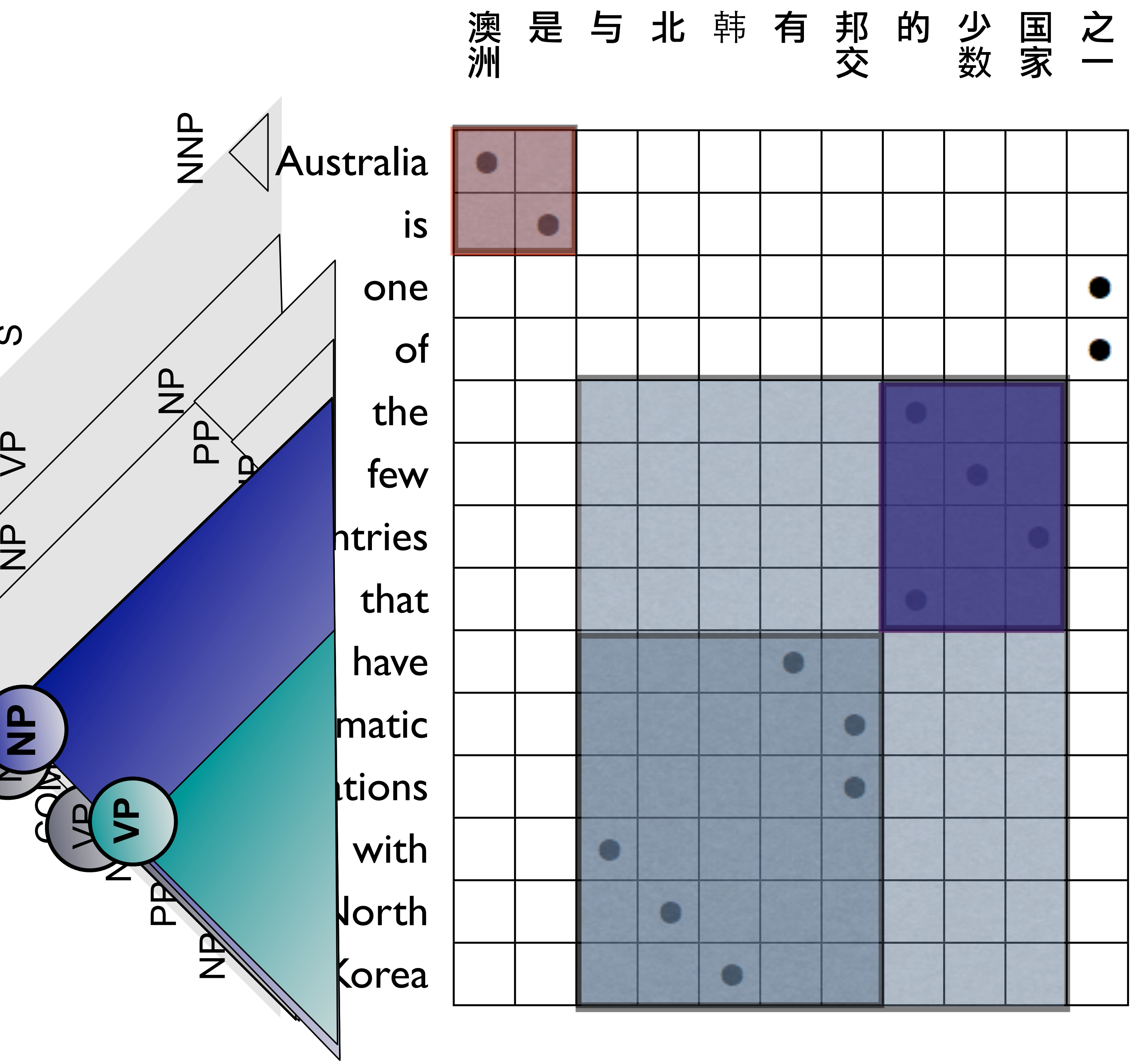
VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

NP → 与 北 韩 有 邦交 的 少数 国家, the few countries that have diplomatic relations with North Korea

??? → 的 少数 国家, the few countries that

??? → 澳洲 是, Australia is

Extracting Syntactic Rules



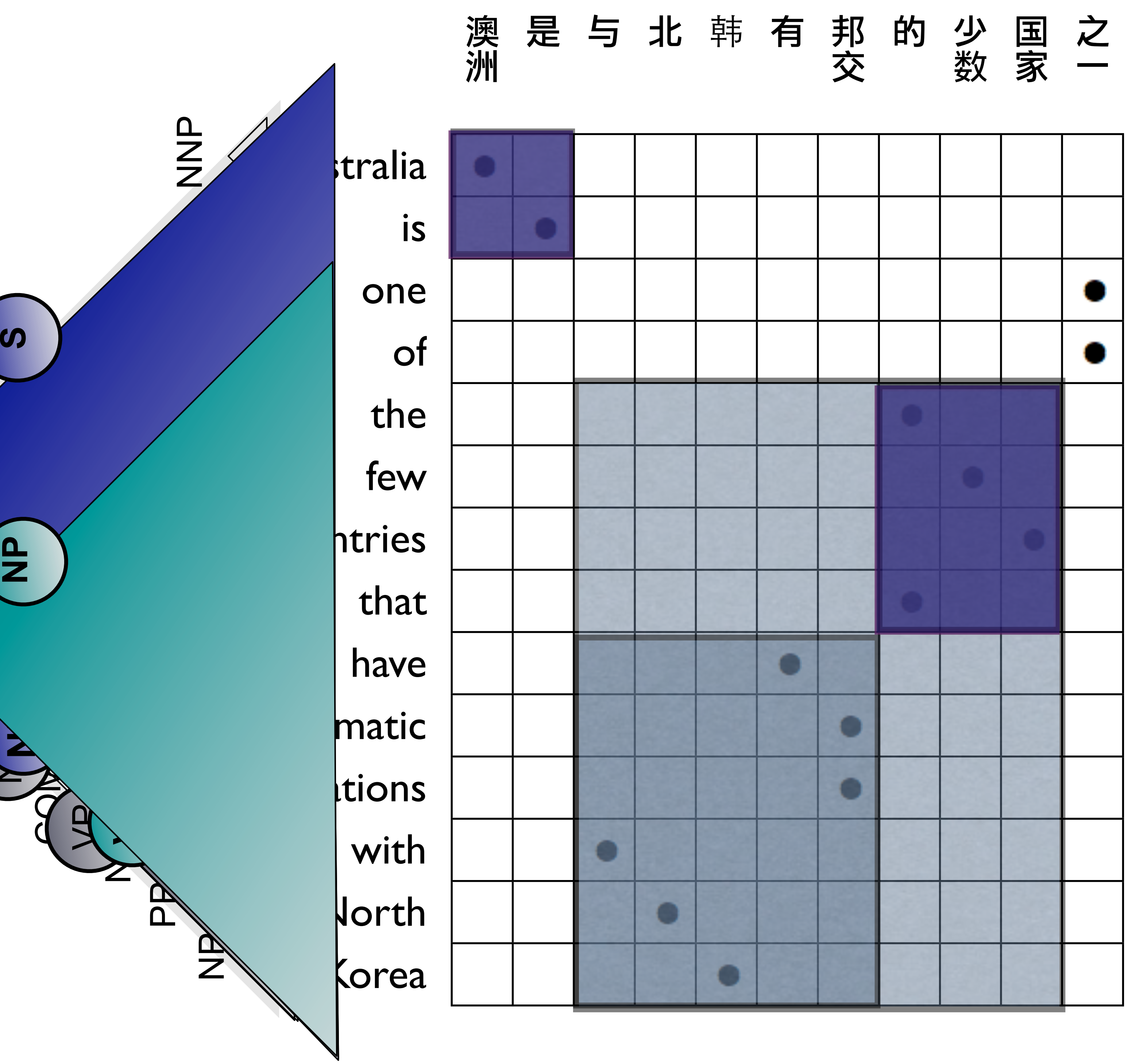
VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

NP → 与 北 韩 有 邦交 的 少数 国家, the few countries that have diplomatic relations with North Korea

NP/ VP → 的 少数 国家, the few countries that

??? → 澳洲 是, Australia is

Extracting Syntactic Rules



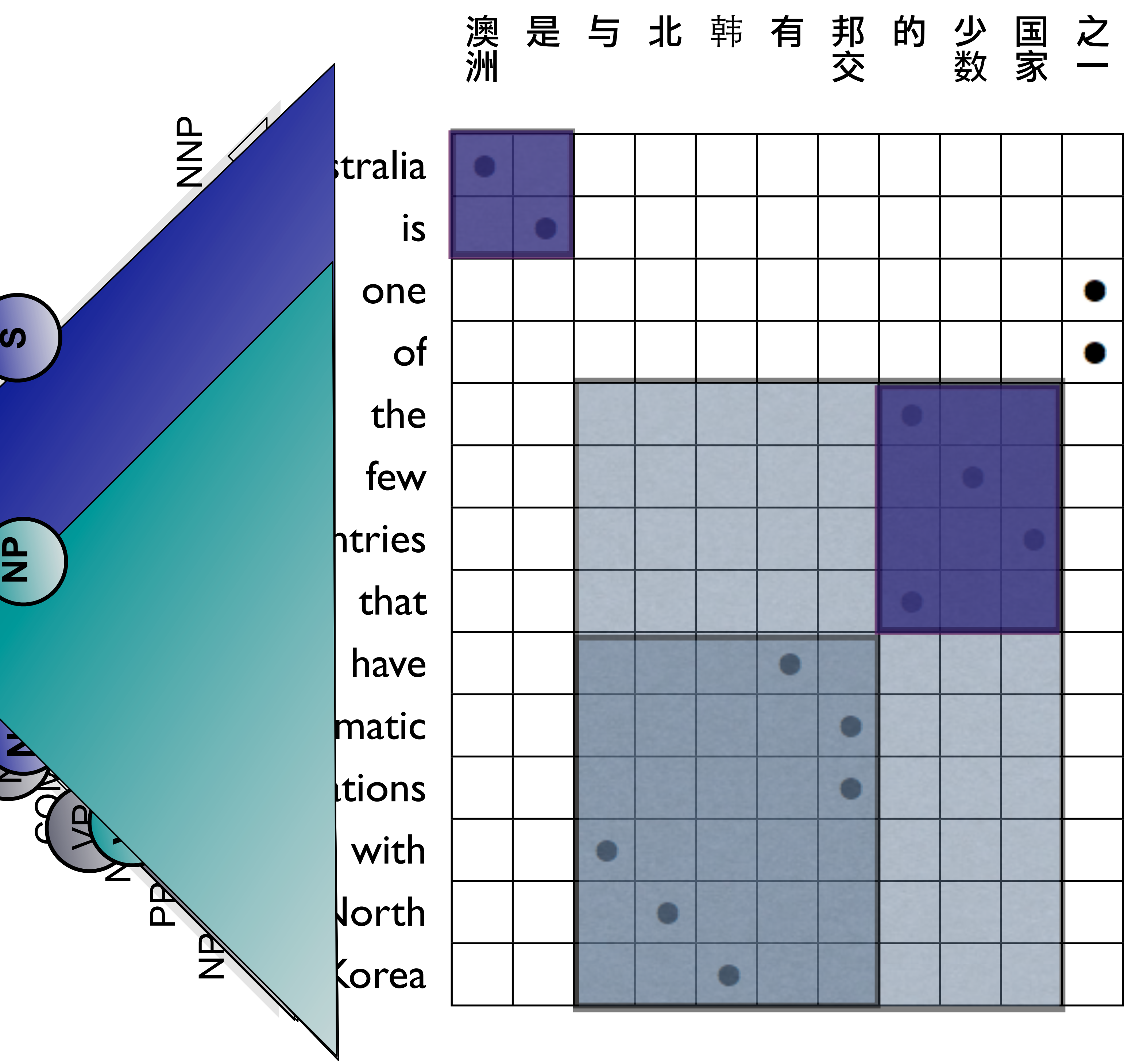
VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

NP → 与 北 韩 有 邦交 的 少数 国家, the few countries that have diplomatic relations with North Korea

NP/ VP → 的 少数 国家, the few countries that

S/ NP → 澳洲 是, Australia is

Extracting Syntactic Rules



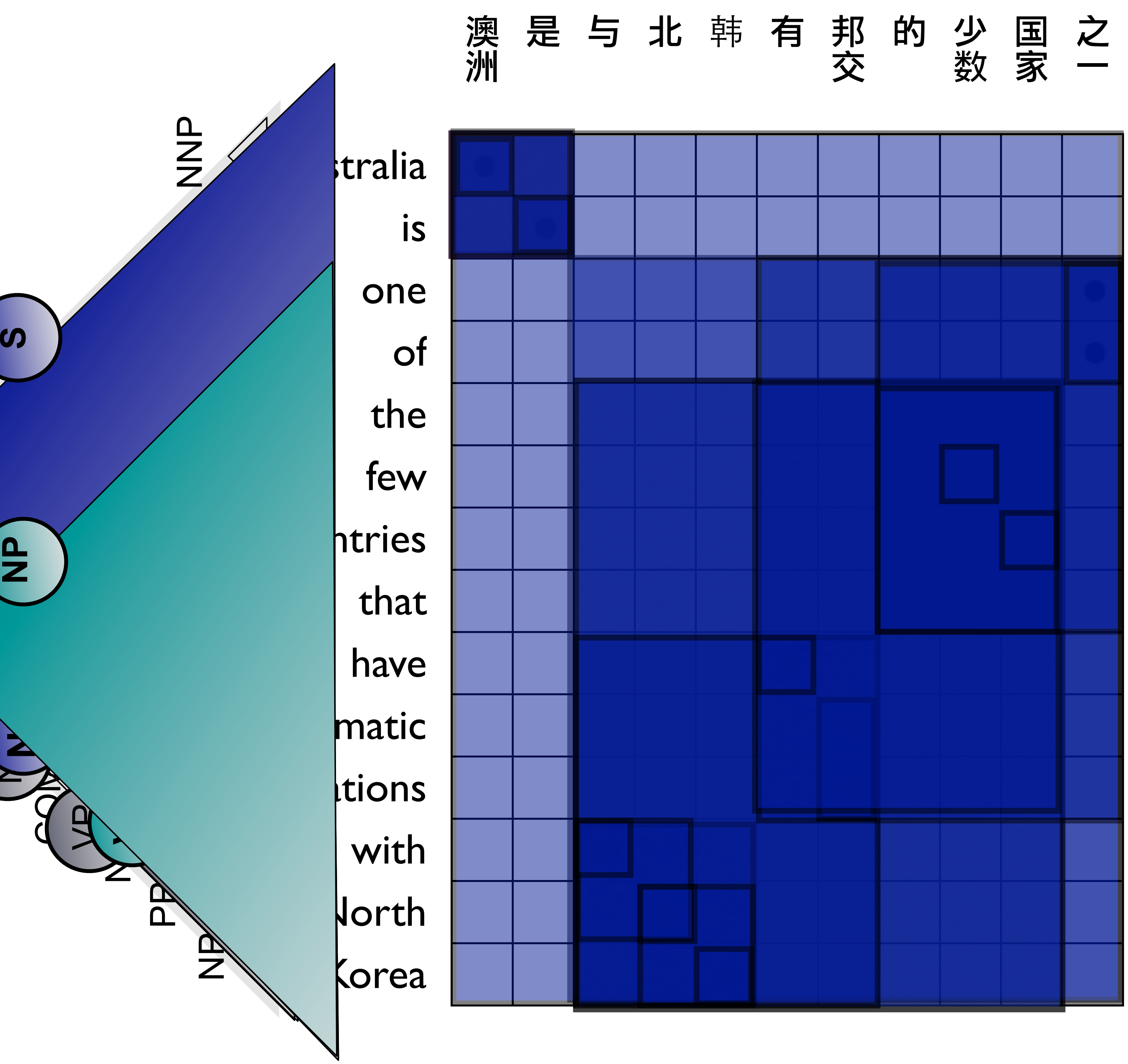
VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

NP → 与 北 韩 有 邦交 的 少数 国家, the few countries that have diplomatic relations with North Korea

NP/ VP → 的 少数 国家, the few countries that

S/ NP → 澳洲 是, Australia is

Extracting Syntactic Rules



VP → 与 北 韩 有 邦交, have diplomatic relations with North Korea

NP → 与 北 韩 有 邦交 的 少数 国家, the few countries that have diplomatic relations with North Korea

NP/ VP → 的 少数 国家, the few countries that

S/ NP → 澳洲 是, Australia is

Discussion: Is this better?

- What do you think of this flavor of SCFGs?
- What are its limitations?
- Do you think that it is better or worse than Hiero?
- How would you prove it?

(Discuss with your neighbors)

New training paradigm

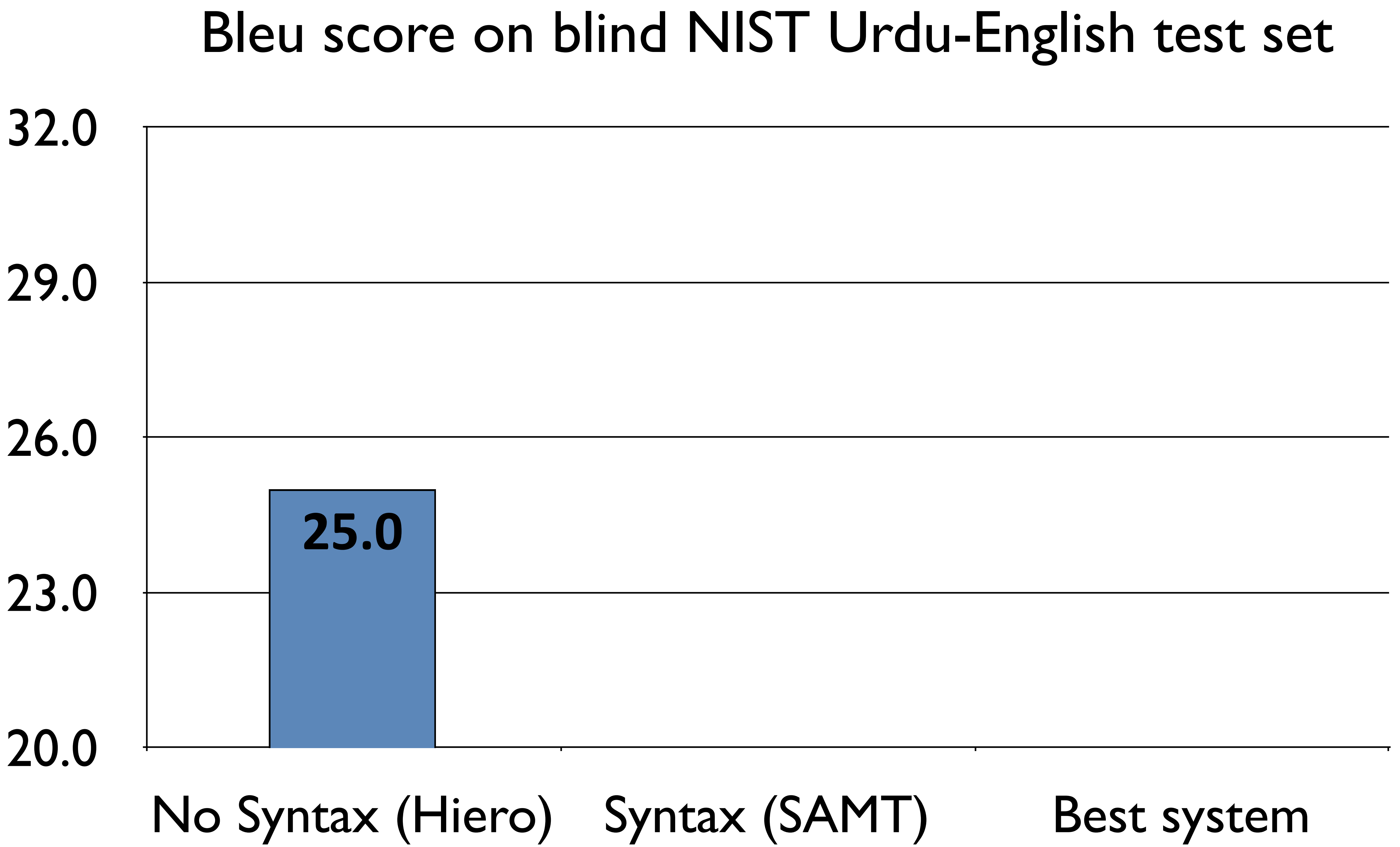
- Training data: word-aligned bilingual parallel corpus, with parse trees
  - No need to parse the Urdu, just parse the English
  - Method is therefore transferable to other resource poor languages
- Extract SCFG rules with syntactic nonterminals
- For non-constituent phrases use CCG-style nonterminals
- Same coverage as Hiero model

Does it work?

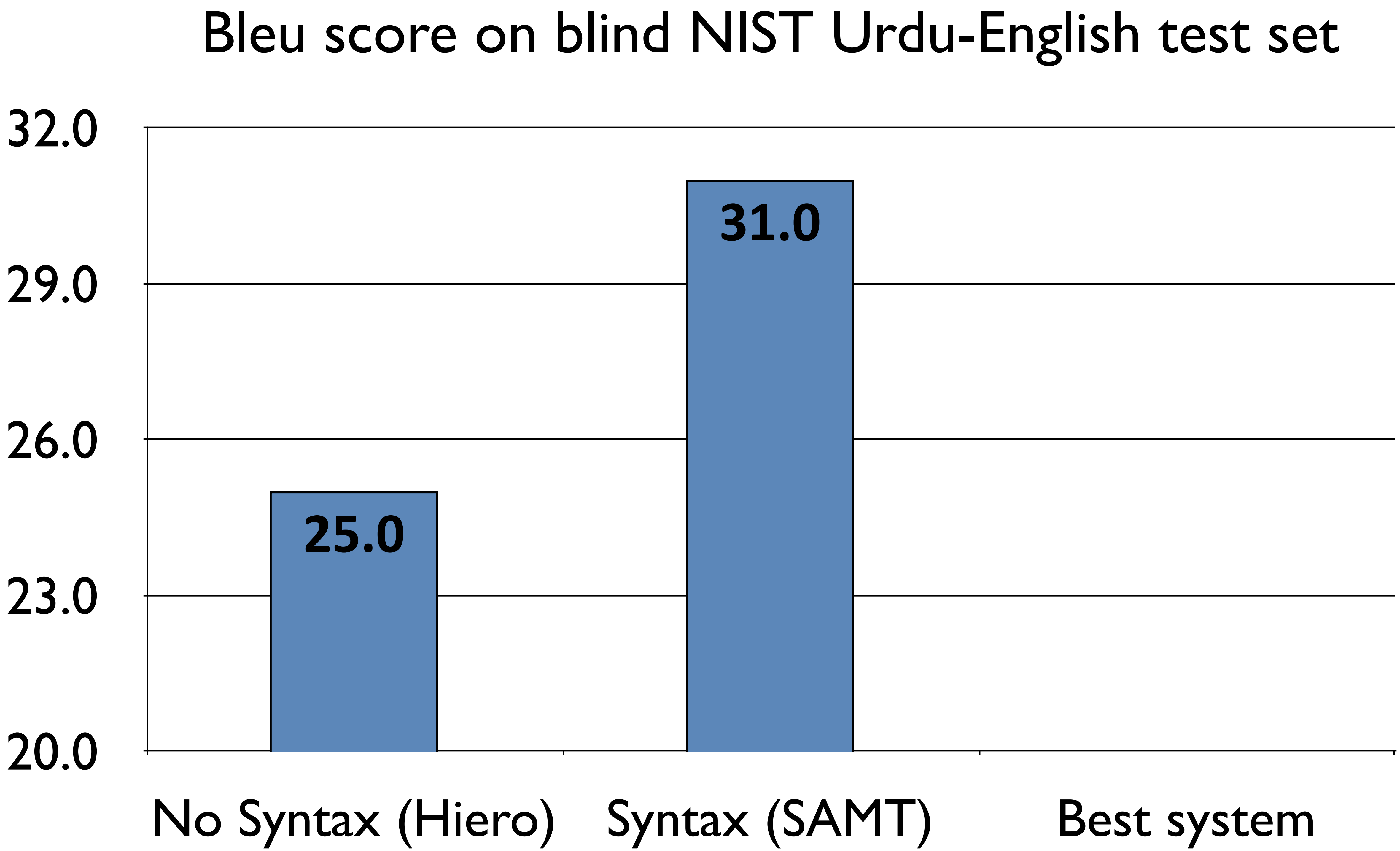
- Tested for Urdu-English MT
- 1.5 Million word parallel corpus
- Two contrastive systems, with different grammar extraction mechanism
  - Hiero
  - Syntax-augmented grammars
- Used same decoder in both cases
- Tested results in a blind test set administered by the National Institute for Standards in Technology

Syntax v. no Syntax

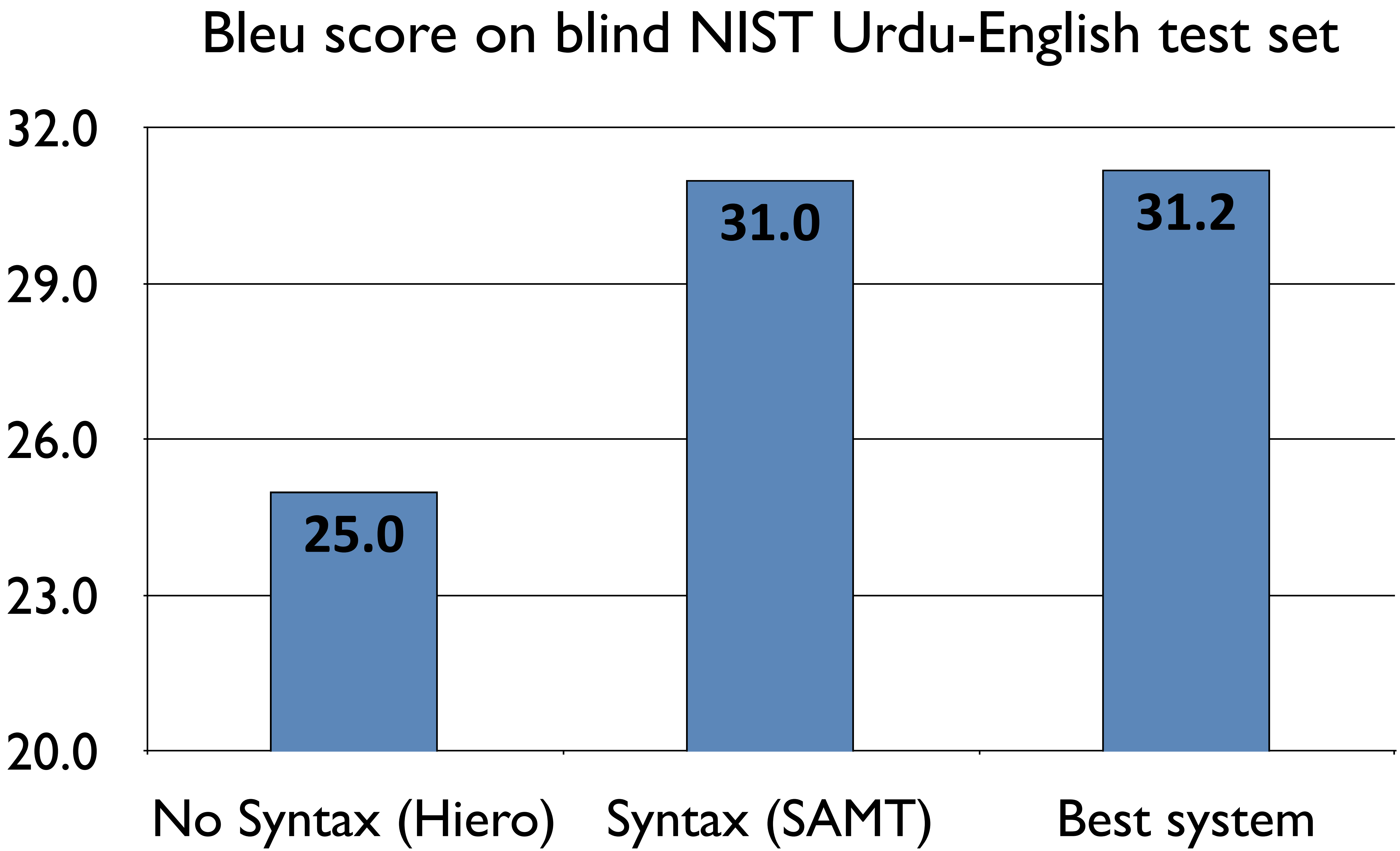
Syntax v. no Syntax



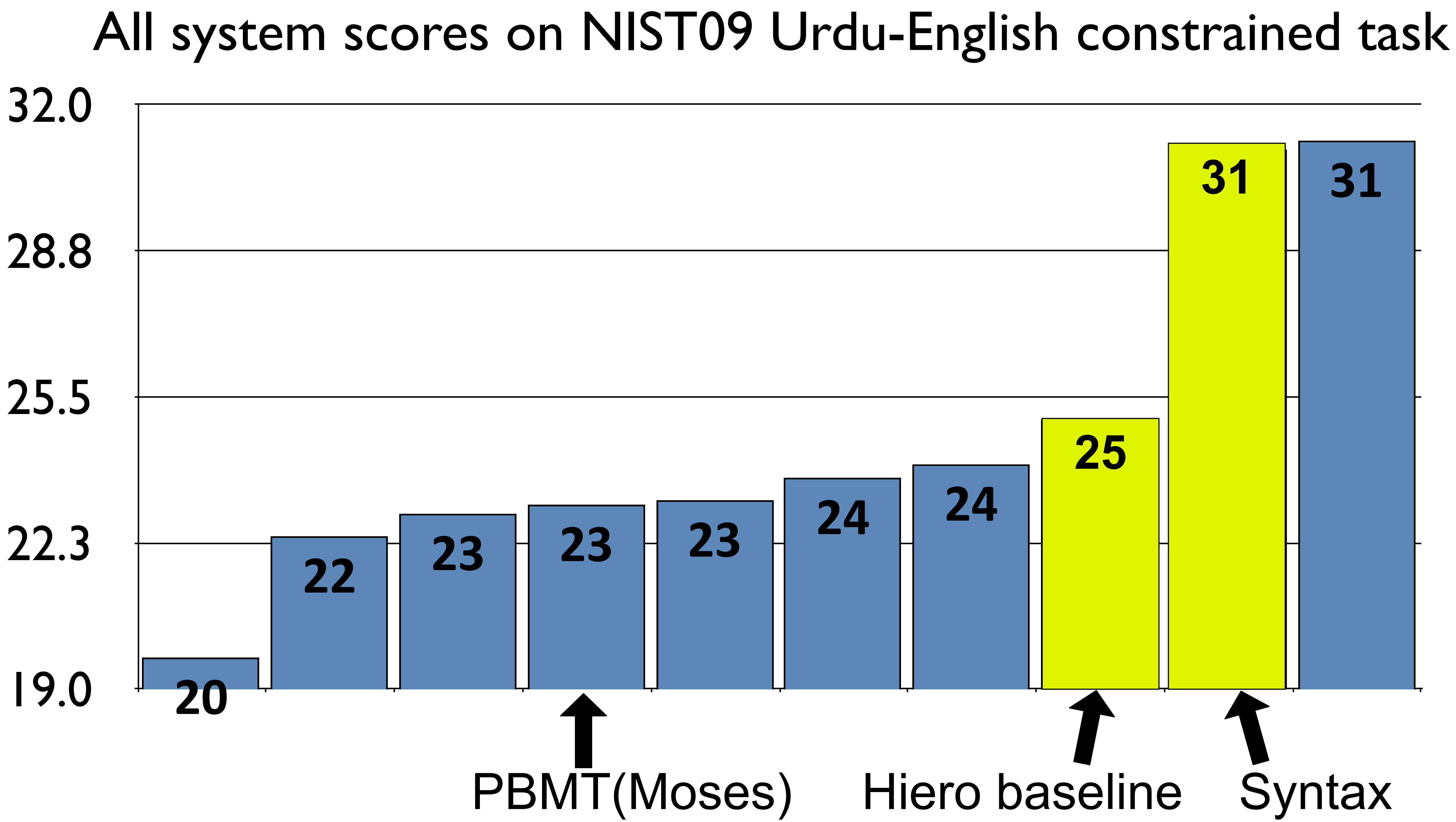
Syntax v. no Syntax



Syntax v. no Syntax



State of the Art Urdu Results



Translation improvements

'first nuclear experiment in 1990 was'
Thomas red Unilever National Laboratory of the United States in ویپن designer, are already working on the book of Los ایلموس National Laboratory ڈینی, former director of the technical انٹیلیجنس written with the cooperation of سٹلمین.

This book 'nuclear express: political history and the expansion of bomb' has been written, and the two writers have also claimed that the country has made nuclear bomb is he or any other country's nuclear secrets to or that of any other nuclear چرائے power cooperation is achieved.

The First Nuclear Test Was in 1990.

Thomas red of the United States, the National Laboratory in designer are already working on the book of Los Alamos National Laboratory, former director of the technical intelligence, with the cooperation of Diana steelman wrote.

This book under the title of the spread of nuclear expressway: the political history of the bomb and this has been written and the two writers have claimed that the country also has made nuclear bomb or any other country, Korea nuclear secrets, or any of the other nuclear power cooperation.

Who did what to whom?

Baseline

He said that China, North Korea, Iran, Syria, Pakistan, through Egypt, Libya and Yemen is to provide nuclear technology.

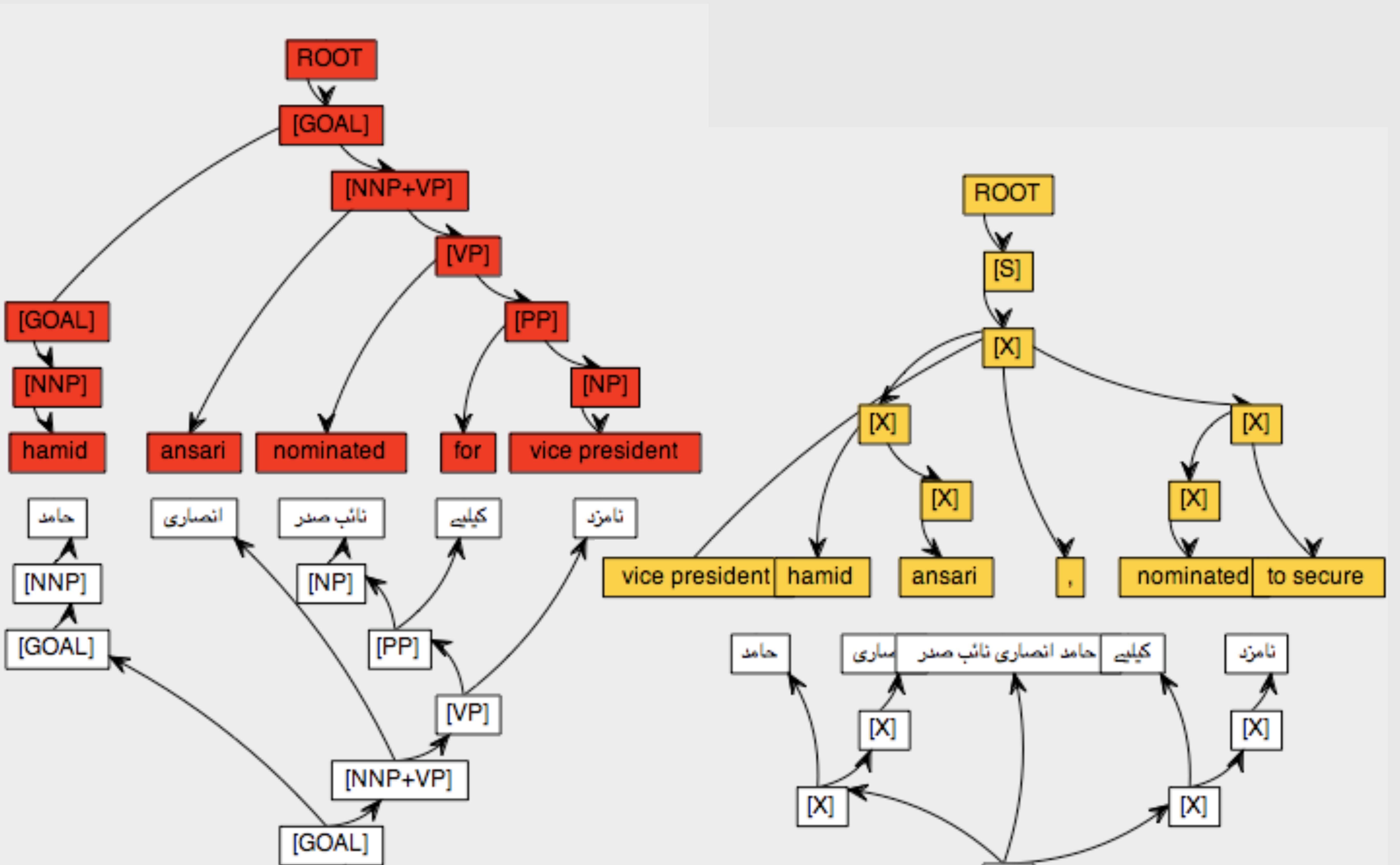
Thomas was red when this question why China has provided the nuclear technology to Pakistan, In response, He said as China and India was joint enemy of Pakistan.

SCALE final system

He said that China would provide nuclear technology to North Korea, Iran, Syria, Pakistan, Egypt, Libya and Yemen.

Thomas red when was this question why China has provided to Pakistan nuclear technology, he said in response to China, Pakistan and India as a common enemy.

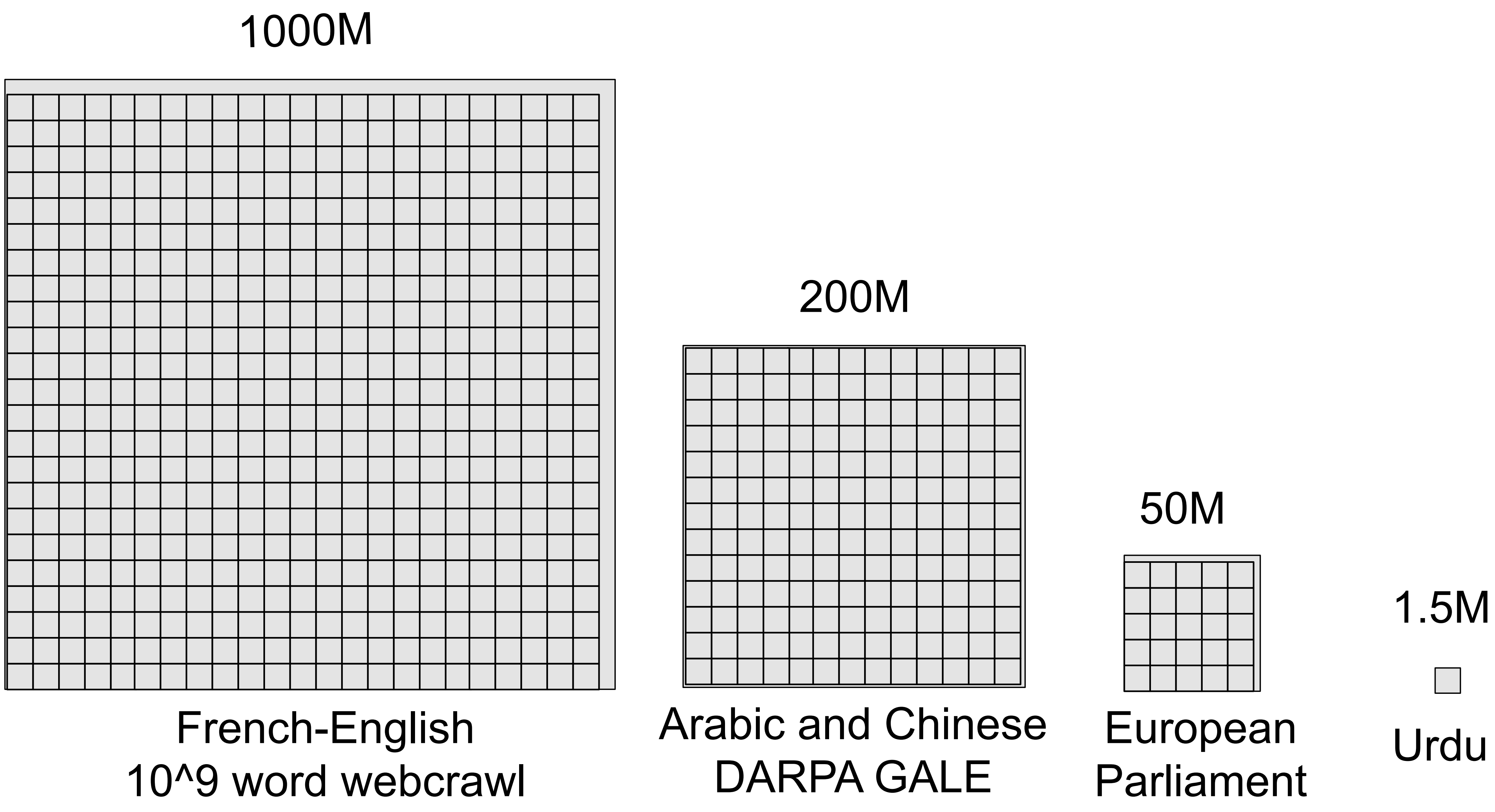
Syntax captures Urdu reordering



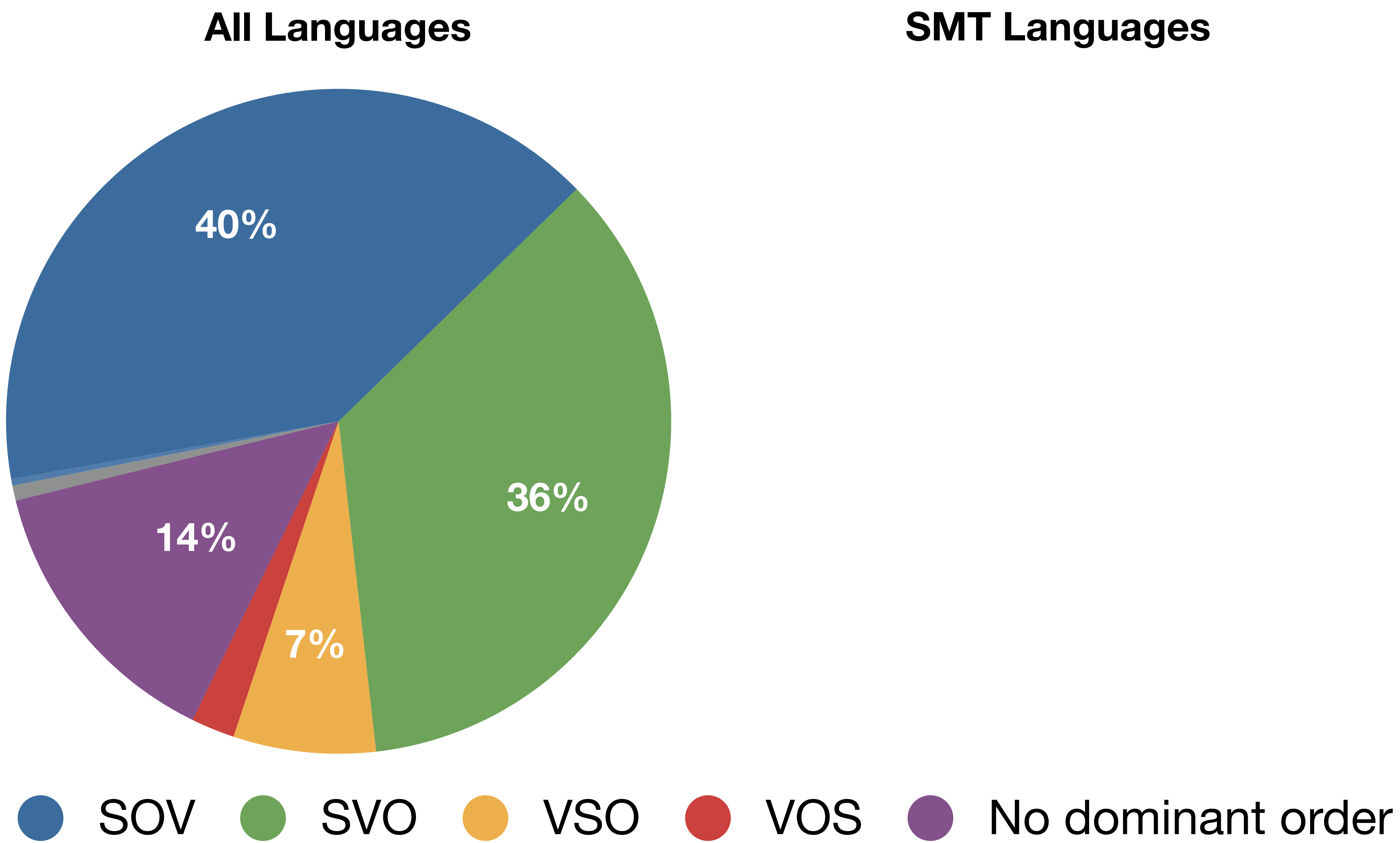
Why did this work?

- Using syntax-based translation models resulted in huge improvements in quality
- Previous work on syntax did not shown significant gains, so why did it work here?
- Urdu is an ideal language to show off the advantages of syntax
  - Very small amount of training data
  - Very different word order than English
- Can't simply memorize translations of phrases
- Must generalize

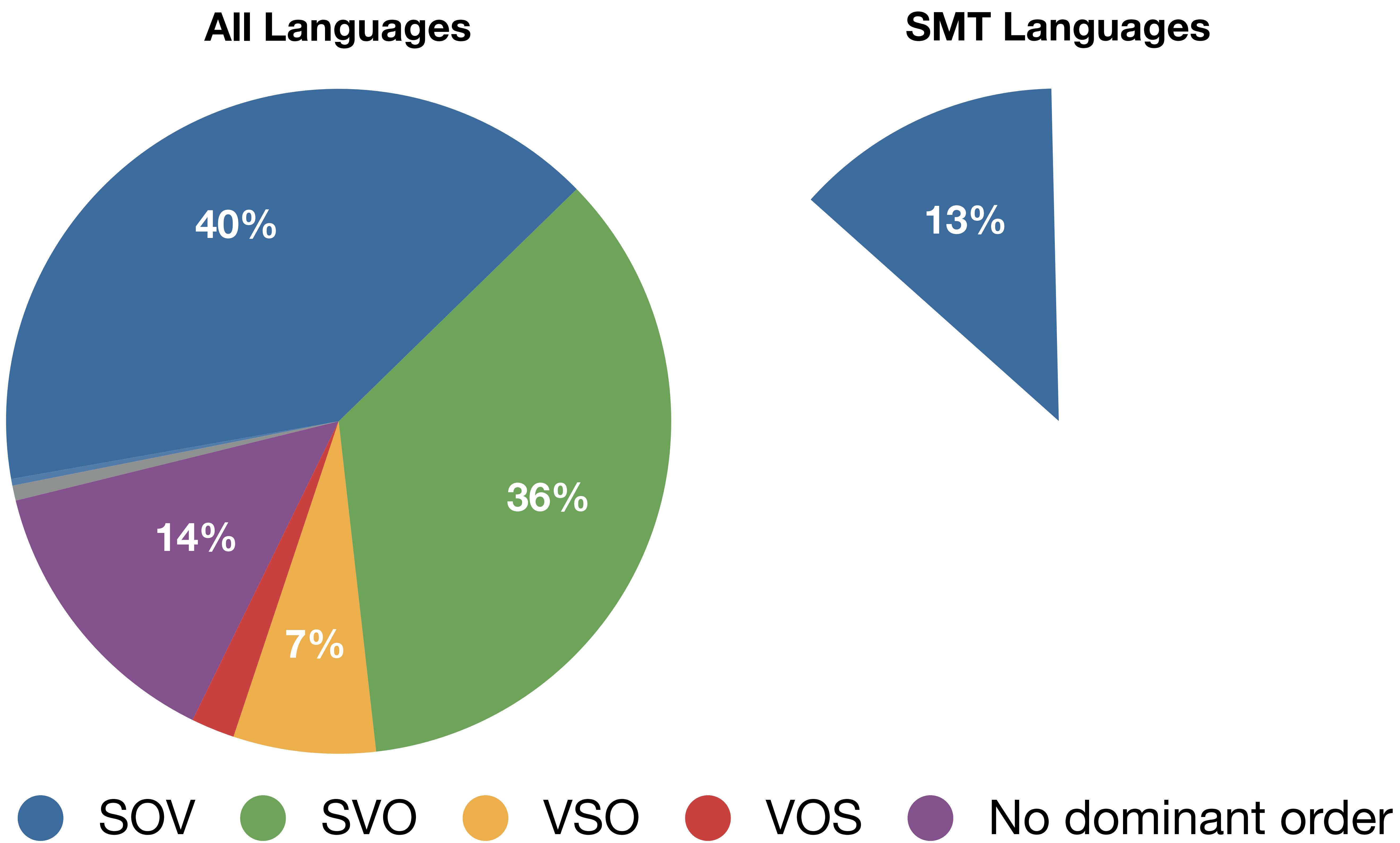
Training data for MT Research



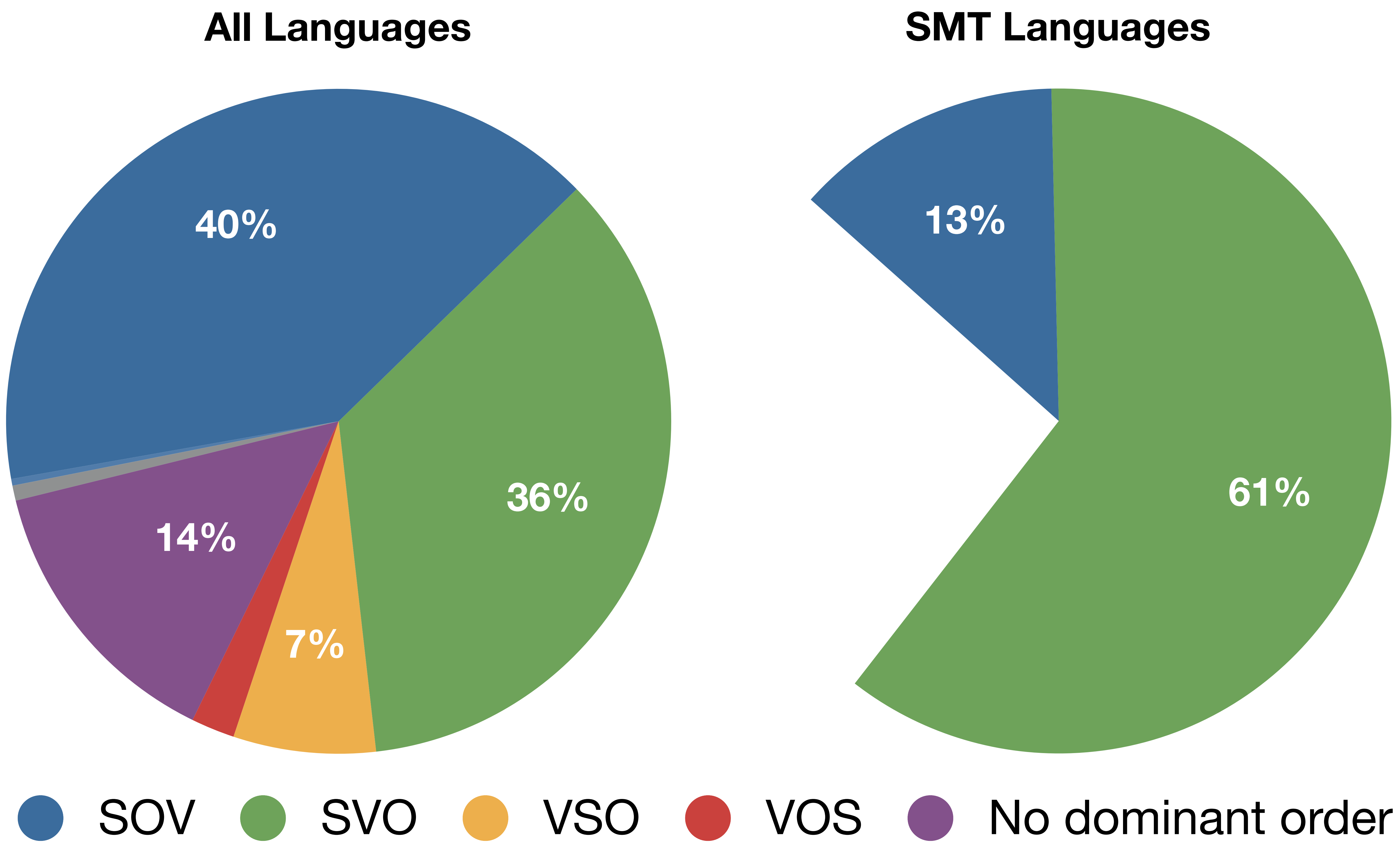
Distribution of Word Orders



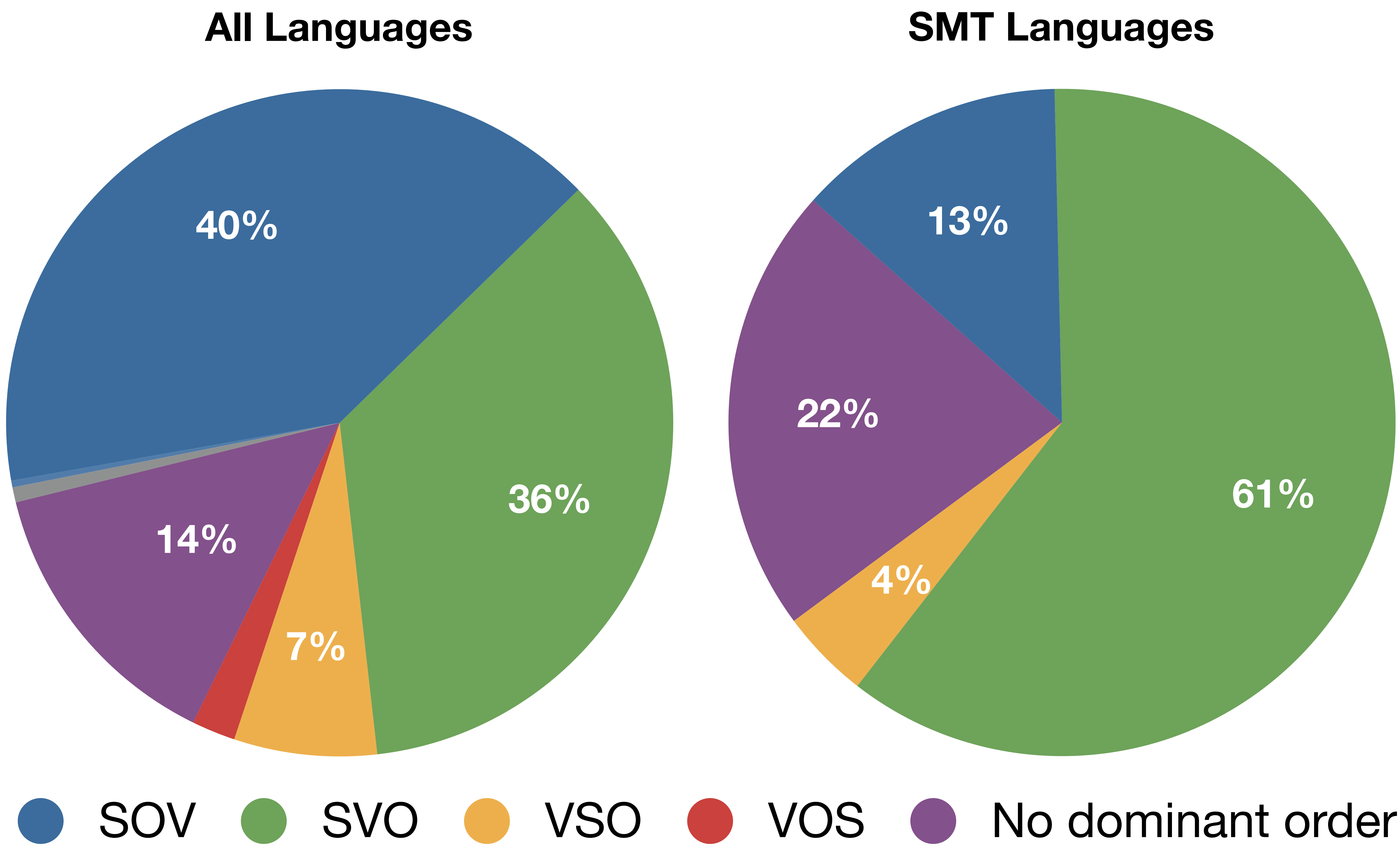
Distribution of Word Orders



Distribution of Word Orders



Distribution of Word Orders



Joshua Decoder

- An open source decoder
- Uses synchronous context free grammars to translate
- Implements all algorithms needed for translating with SCFGs
  - grammar extraction (Thrax!)
  - chart-parsing
  - n-gram language model integration
  - pruning, and k-best extraction

Joshua Decoder

- Download it from
  - http://joshua-decoder.org

Joshua Decoder

- Download it from
  - http://joshua-decoder.org
- Brownie points if you use it in your final projects

Joshua Decoder

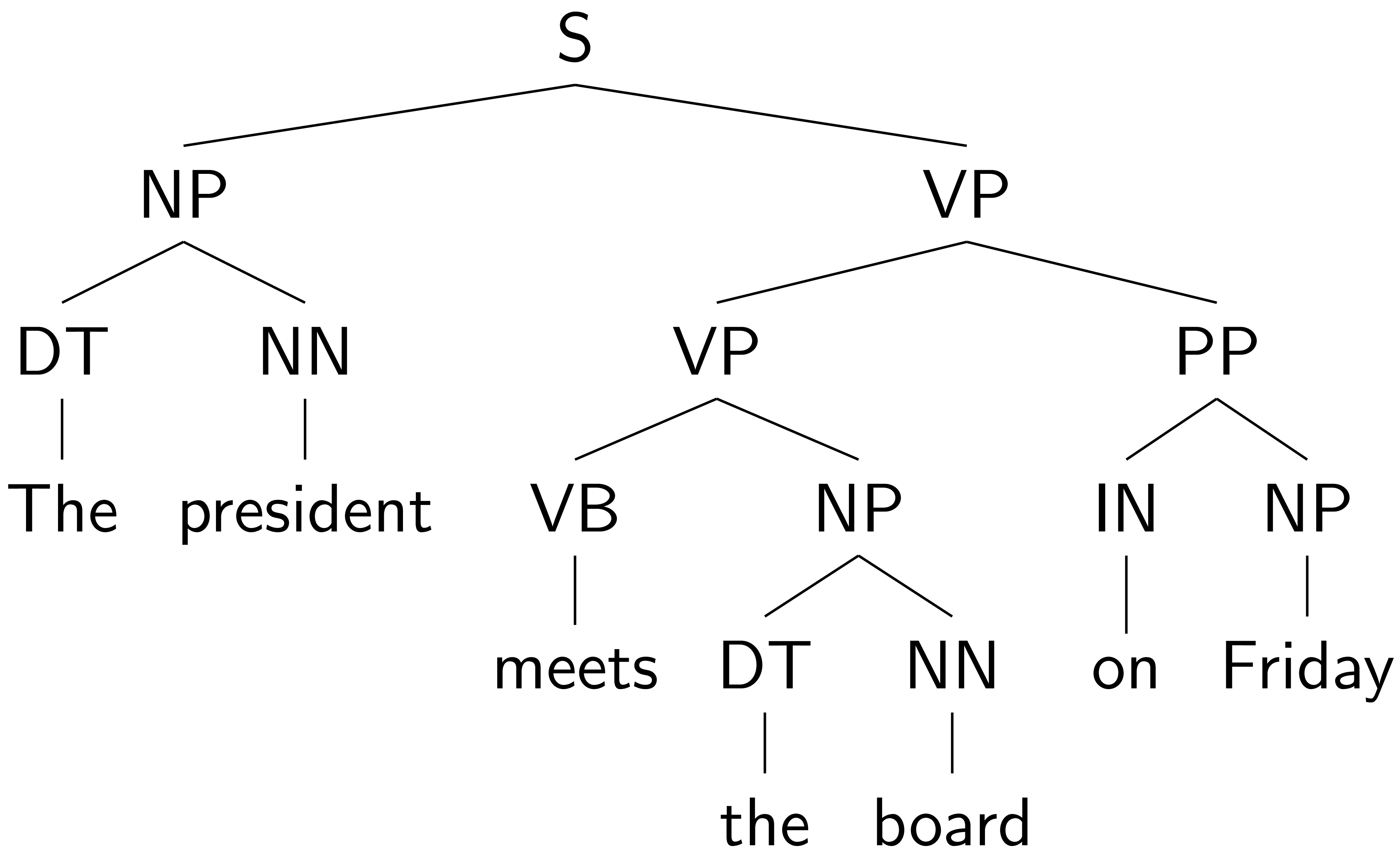
- Download it from
  - http://joshua-decoder.org
- Brownie points if you use it in your final projects
- Use Jonny's Thrax grammar extractor to test different kinds of SFCGs for your problems

More syntax

Syntactic LMs

- Our goal is reorder the translated phrases so that they are grammatical English
- Isn't the language model probability supposed to do that already?
- Instead of an n-gram model, could we augment the LM with syntactic information?

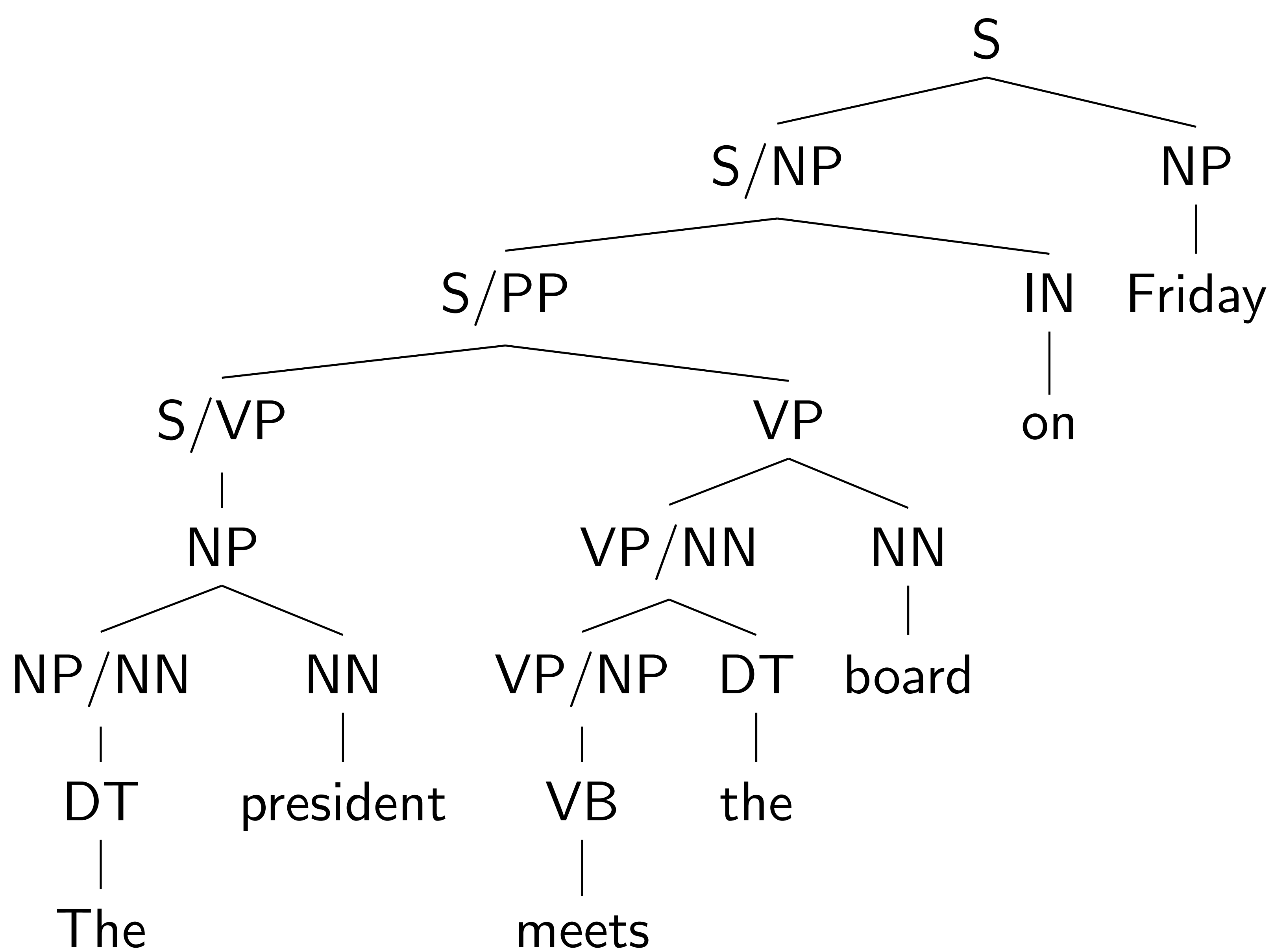
Statistical parsing



Problem: bottom up parsing requires whole sentence

We need the LM to be able to score partial translations

One possibility: Incremental parsing



Questions?