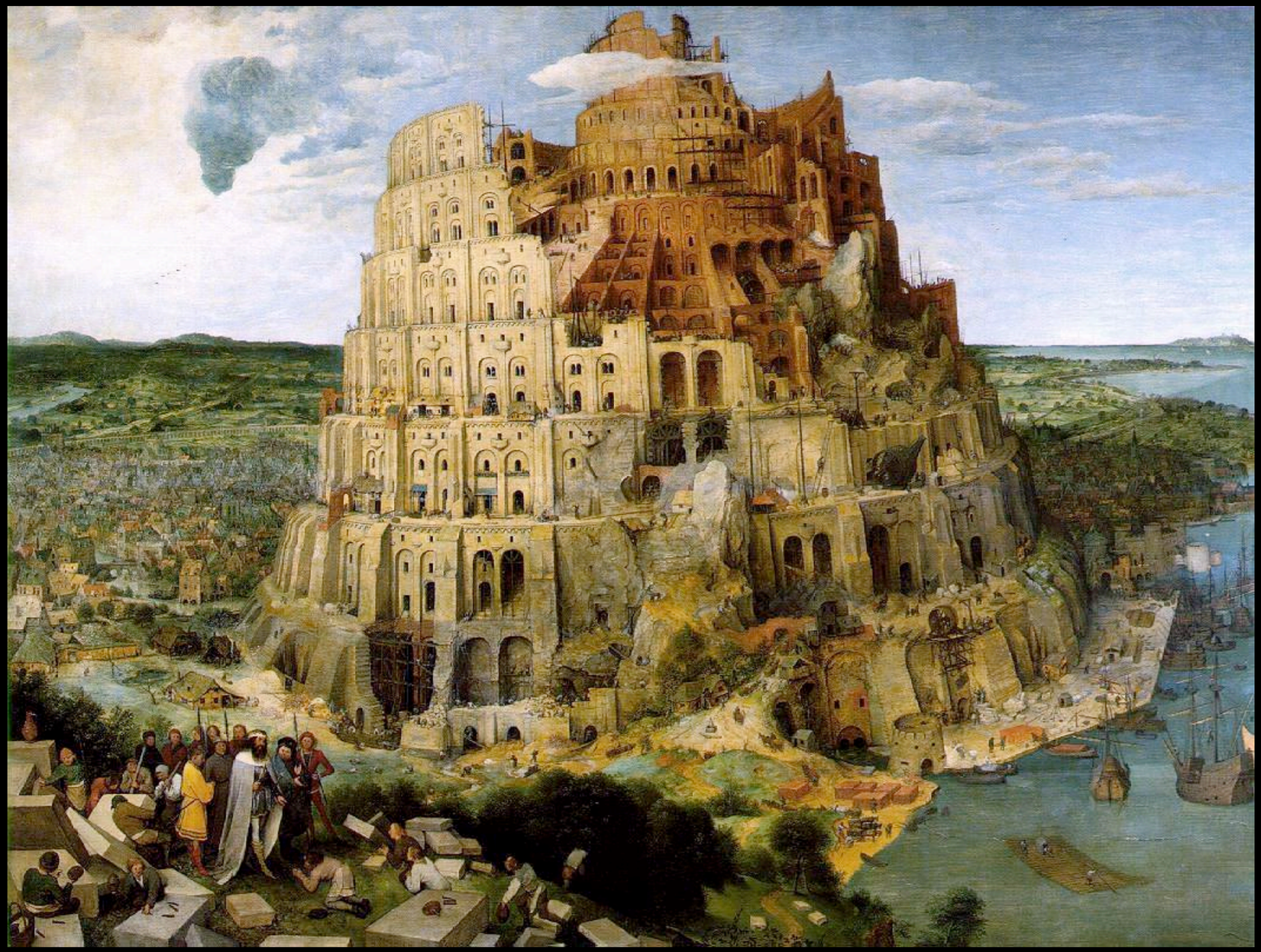
Statistical Machine Translation

Adam Lopez
University of Edinburgh → Johns Hopkins University

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

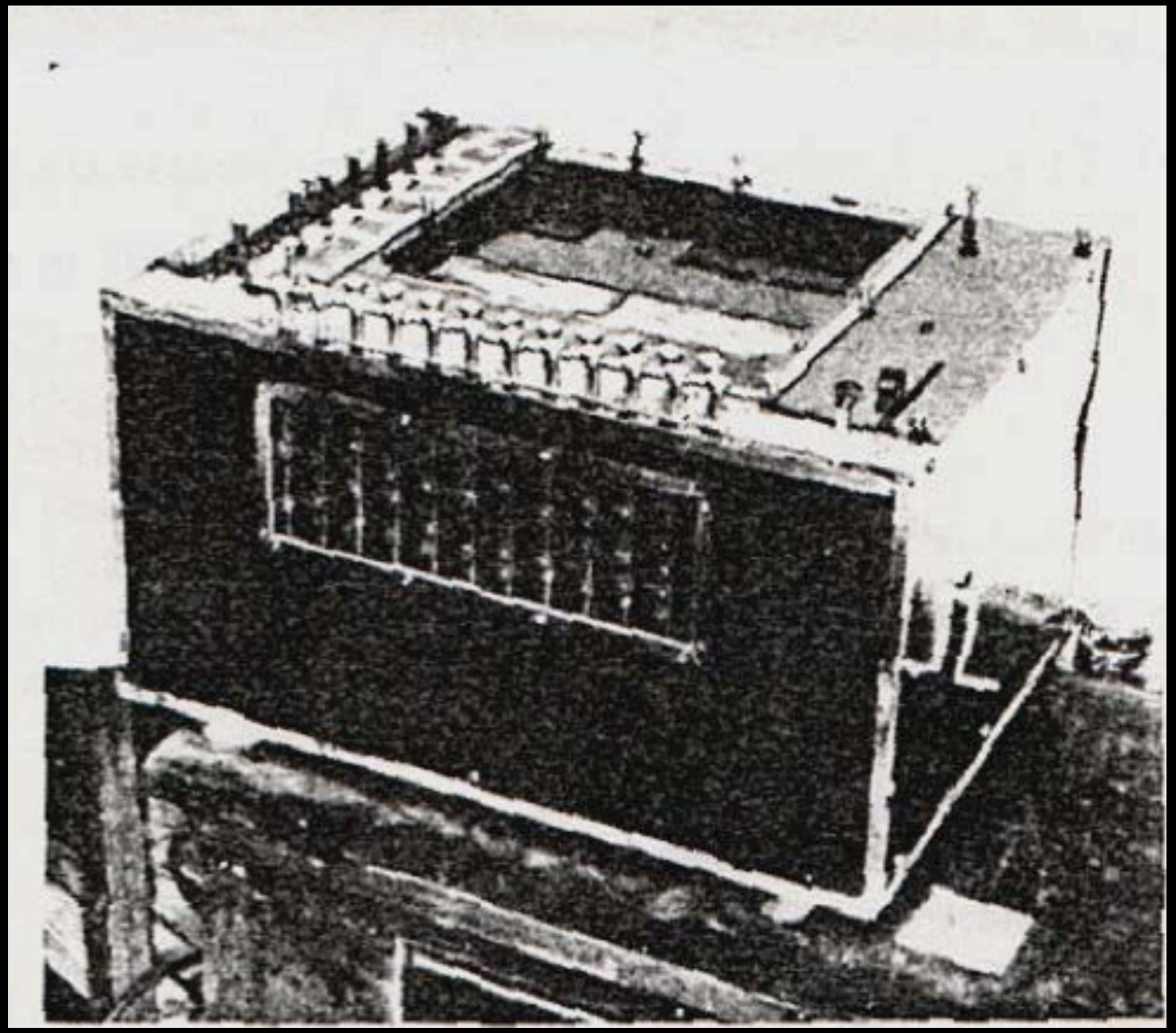
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

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

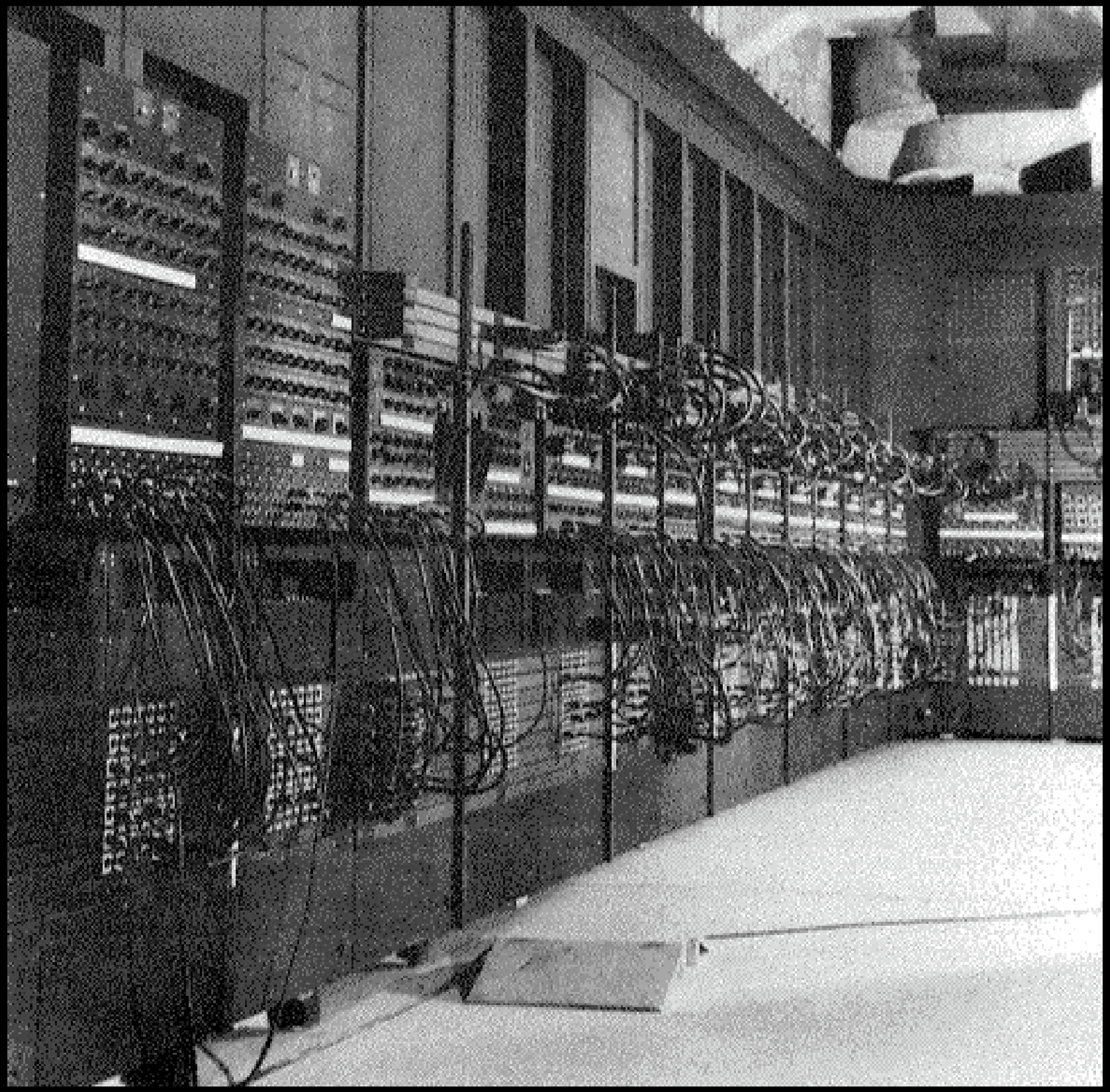


The Tower of Babel

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!

Google **Language Tools**

Translated search

Type a search phrase in your language. Google will find results in other languages and translate them for you to read.

Search for: [ ] Translate and Search

Search pages written in:
- (•) **Automatically selected languages**
- ( ) Specific languages

My language: English ▾

Example: 1. Search for Bern tourist information.
2. We translate your query into French and German, and find French and German results.
3. Finally, we translate the French and German results back into your language.

Translate text

Bienvenue à Le Mans

French » English Translate

Google **Language Tools**

Translated search

Type a search phrase in your language. Google will find results in other languages and translate them for you to read.

Search for: Translate and Search

pages written in:
omatically selected languages
cific languages

My language:
English ▾

ch for Bern tourist information.
ranslate your query into French and German, and find French and German results.
ly, we translate the French and German results back into your language.

Mans

English
Estonian
Filipino
Finnish
French
Galician
German
Greek
Haitian Creole
Hebrew
Hindi
Hungarian
Icelandic
Indonesian
Irish
Italian
Japanese
Korean
Latvian
Lithuanian

French » English Translate

Google **Language Tools**

Translated search

Type a search phrase in your language. Google will find results in other languages and translate them for you to read.

Search for: Translate and Search

pages written in:
omatically selected languages
cific languages

My language:
English ▾

ch for Bern tourist information.
ranslate your query into French and German, and find French and German results.
ly, we translate the French and German results back into your language.

Mans

English
Estonian
Filipino
Finnish
French
Galician
German
Greek
Haitian Creole
Hebrew
Hindi
Hungarian
Icelandic
Indonesian
Irish
Italian
Japanese
Korean
Latvian
Lithuanian

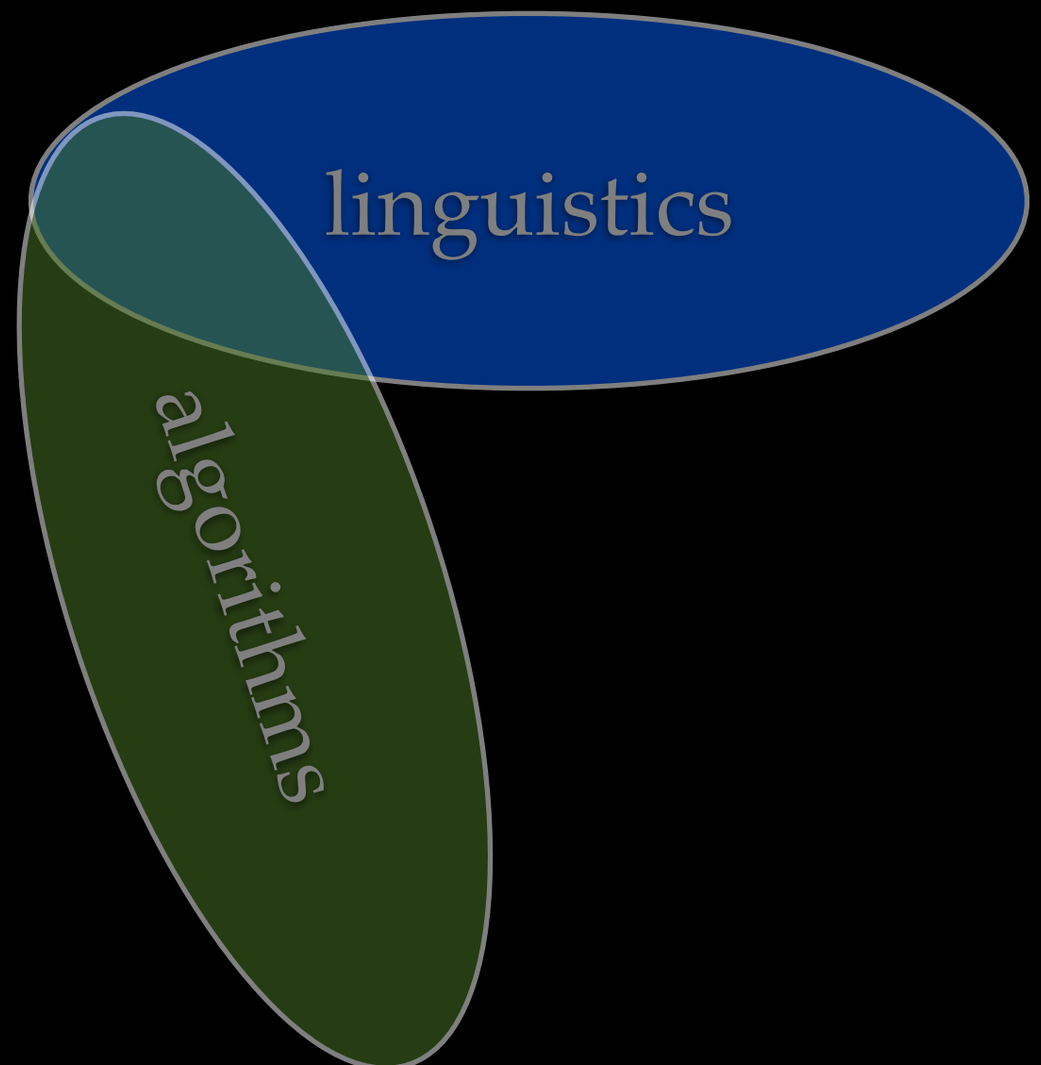
French » English Translate

2756 language pairs!

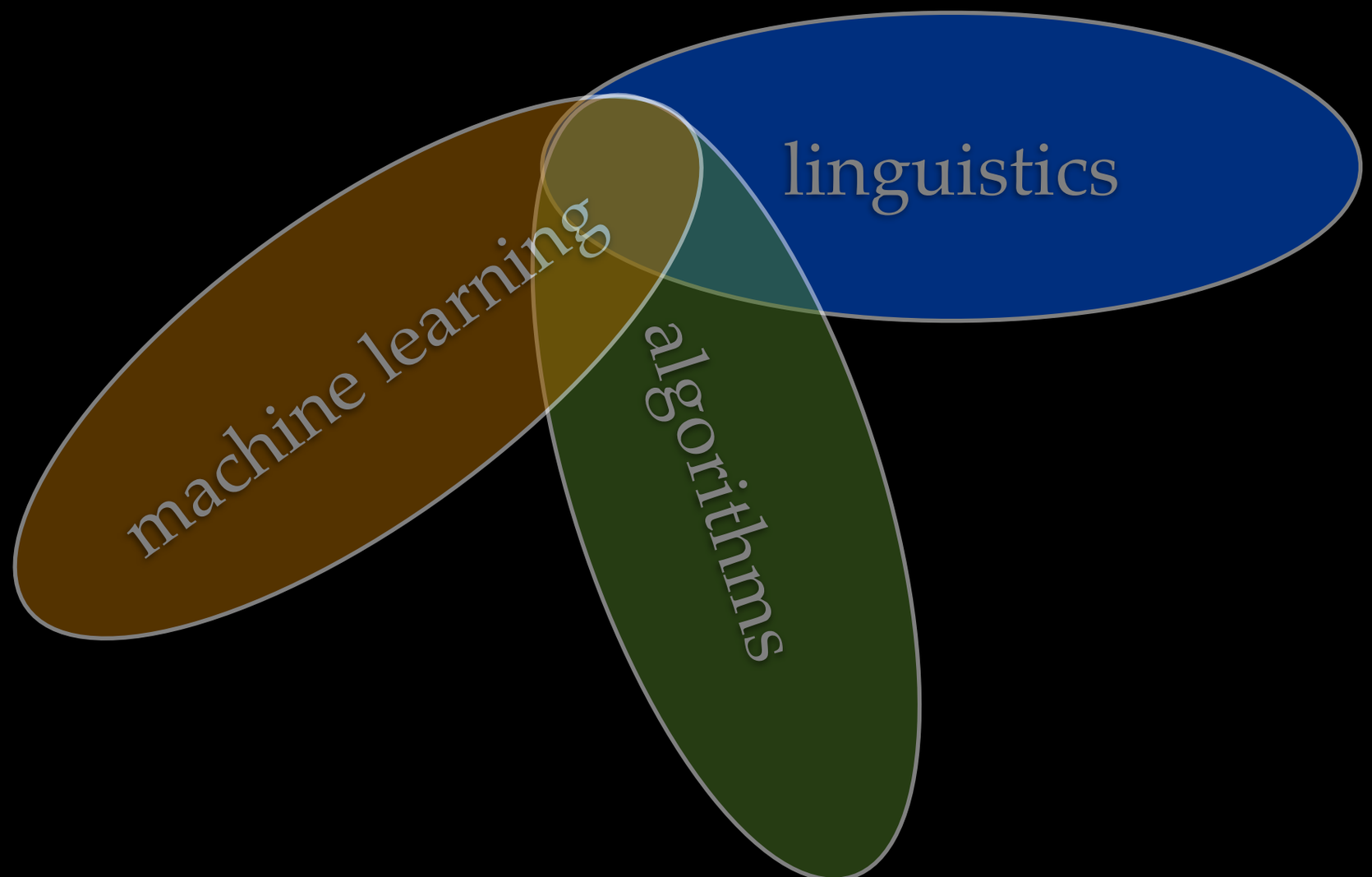
Statistical Machine Translation

Statistical Machine Translation

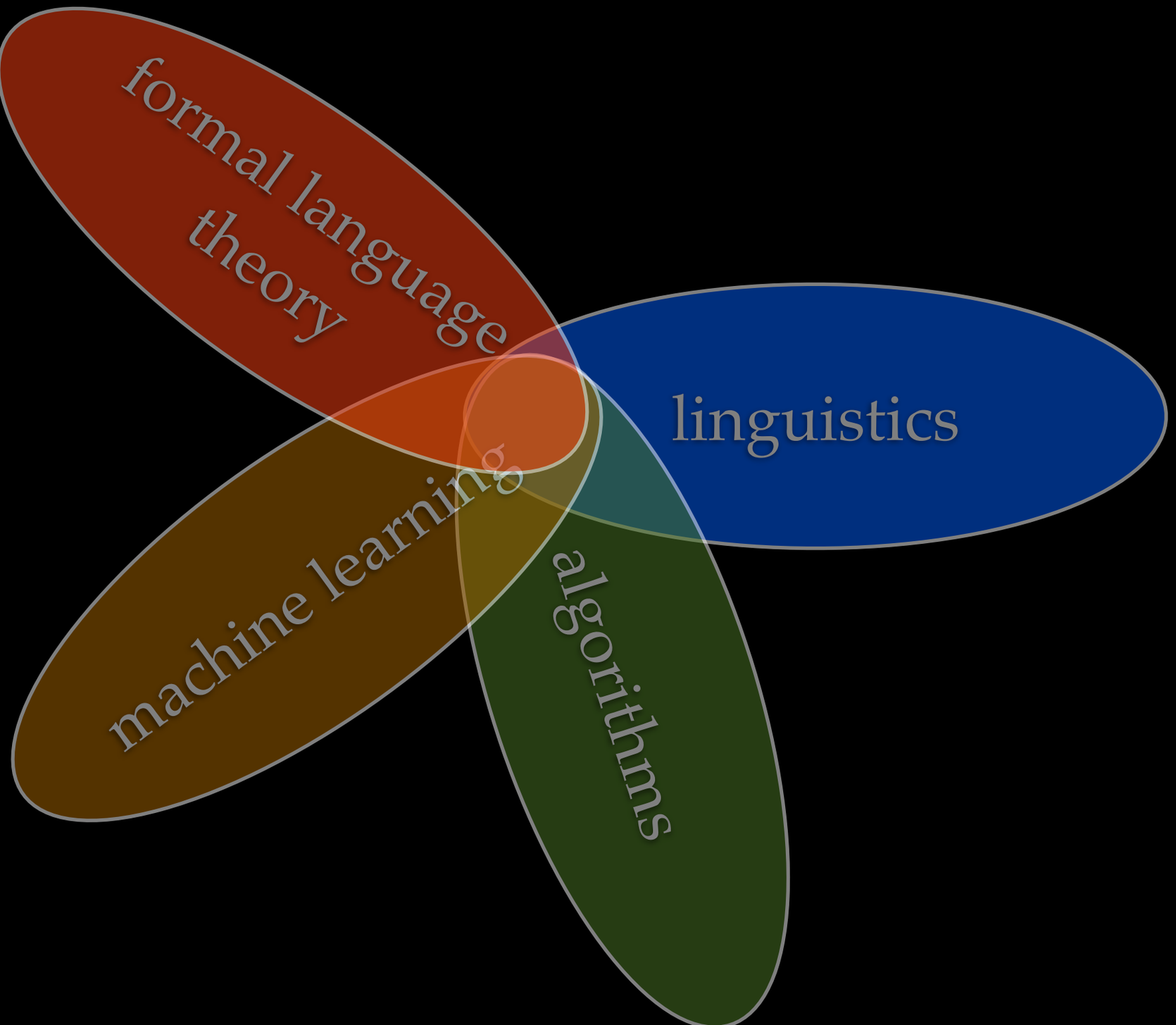
Statistical Machine Translation



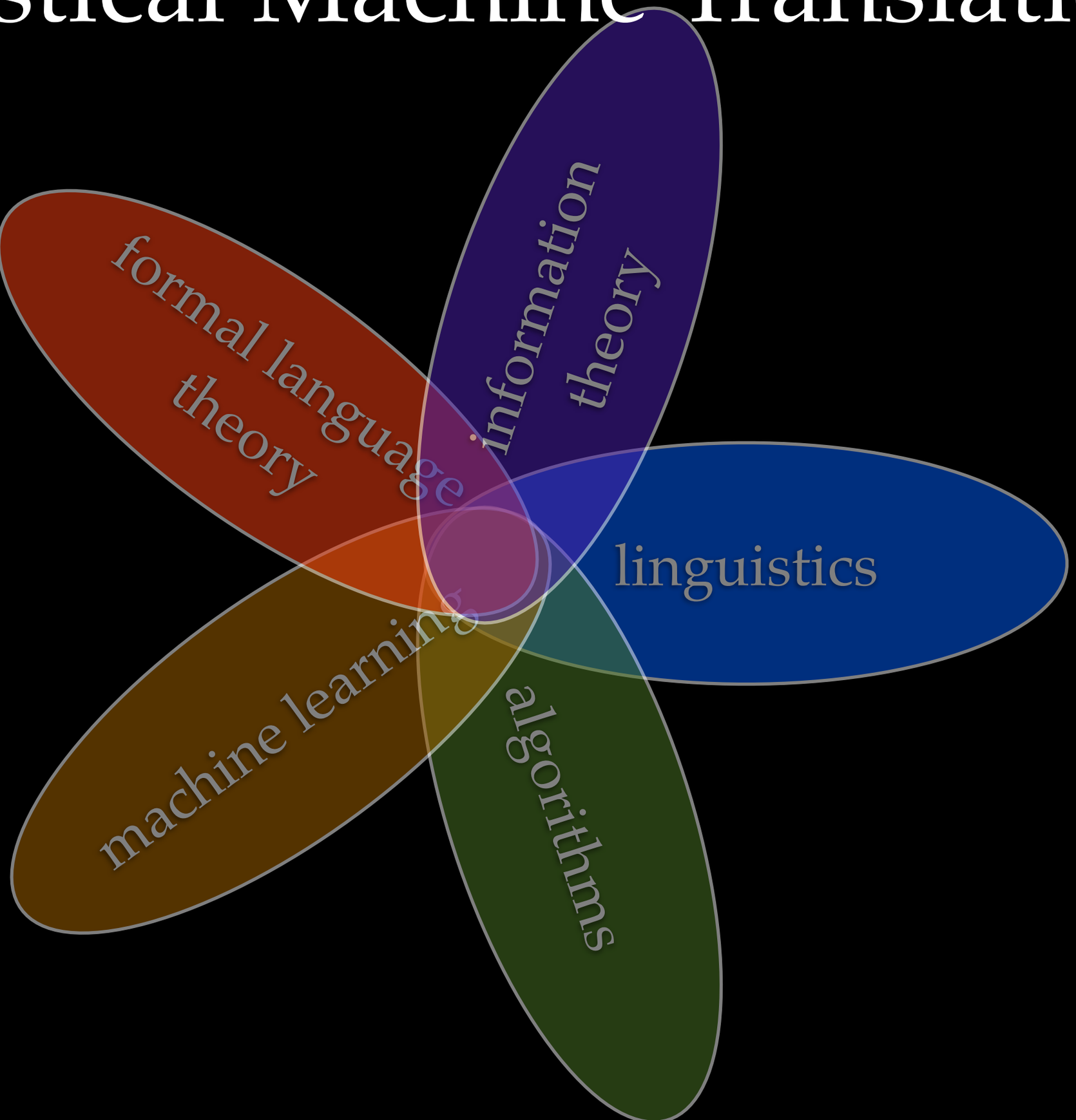
Statistical Machine Translation



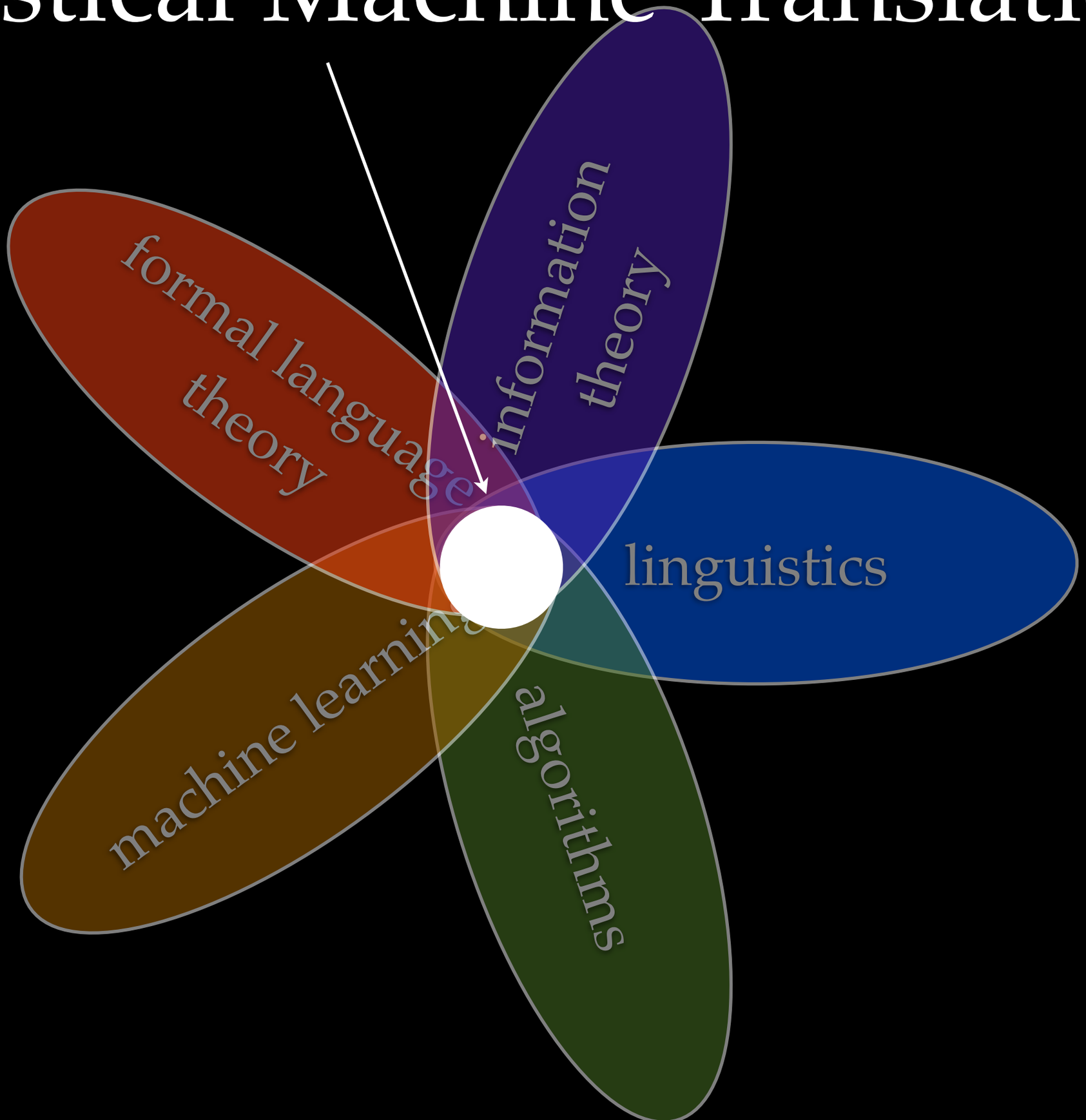
Statistical Machine Translation



Statistical Machine Translation



Statistical Machine Translation



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 .

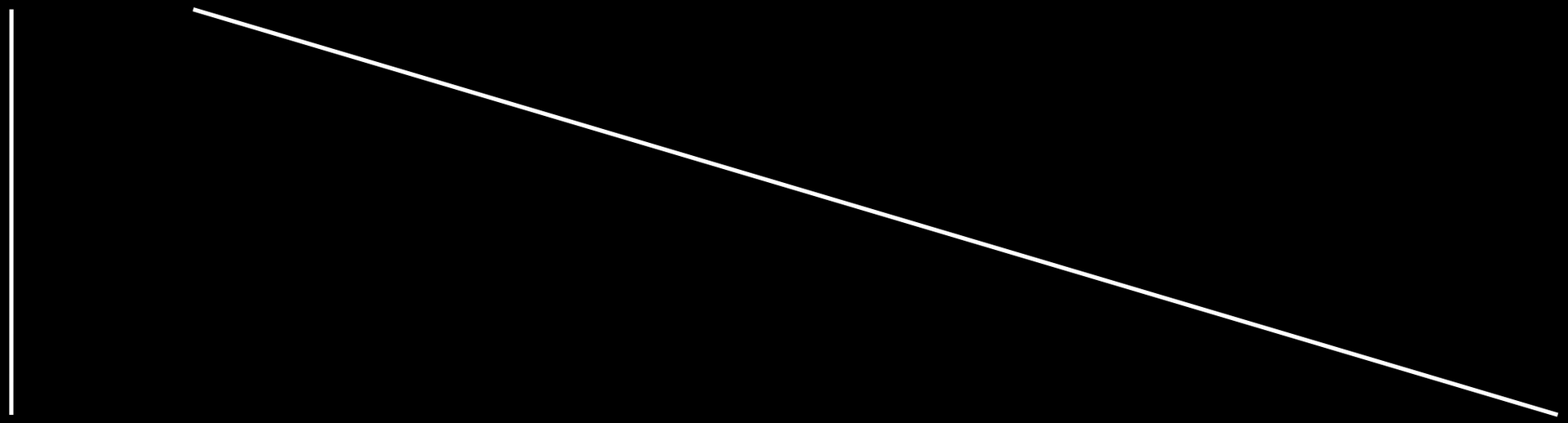
Although north wind howls , but sky still very clear .

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 .

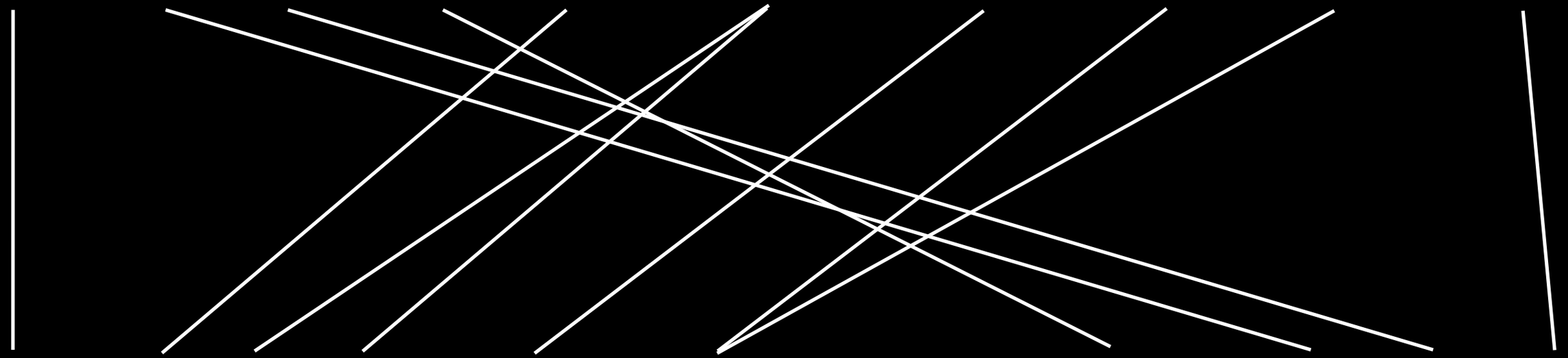
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



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

Although north wind howls , but sky still very clear .

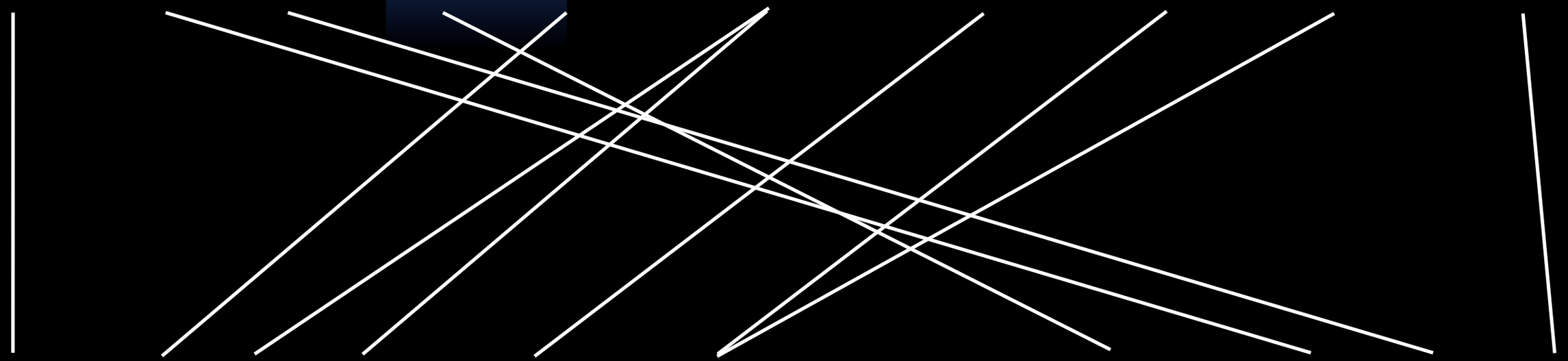
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



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

Although north wind howls , but sky still very clear .

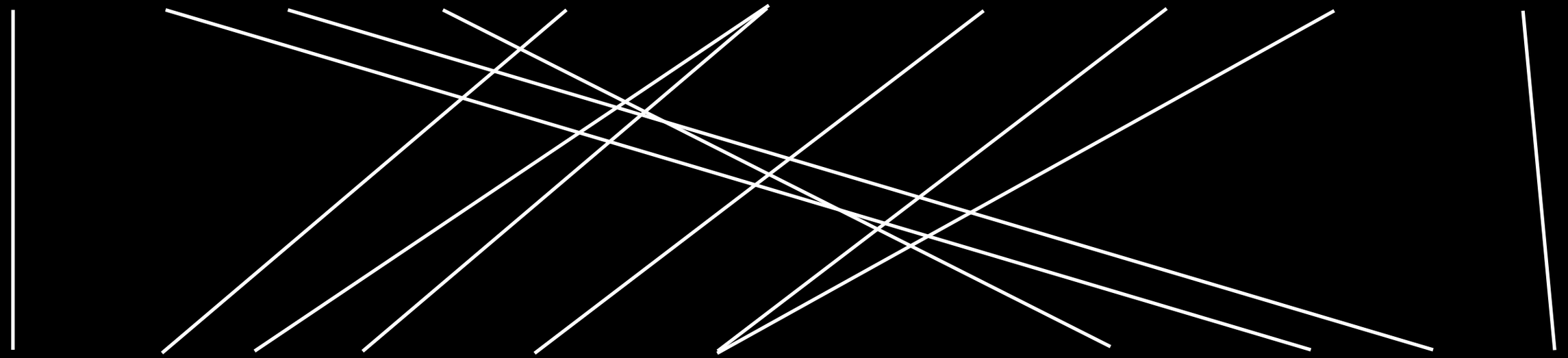
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



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

Although north wind howls , but sky still very clear .

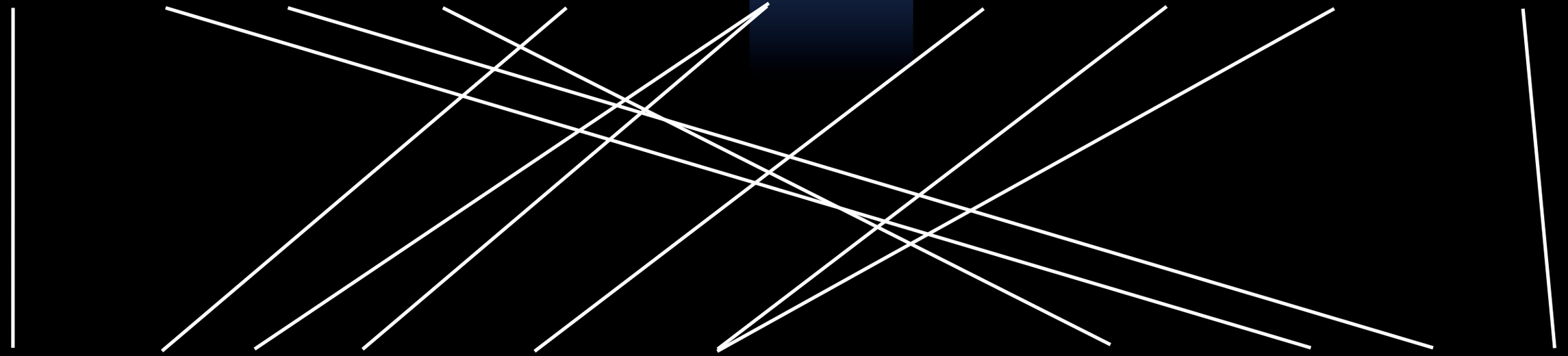
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



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

Although north wind howls , but sky still very clear .

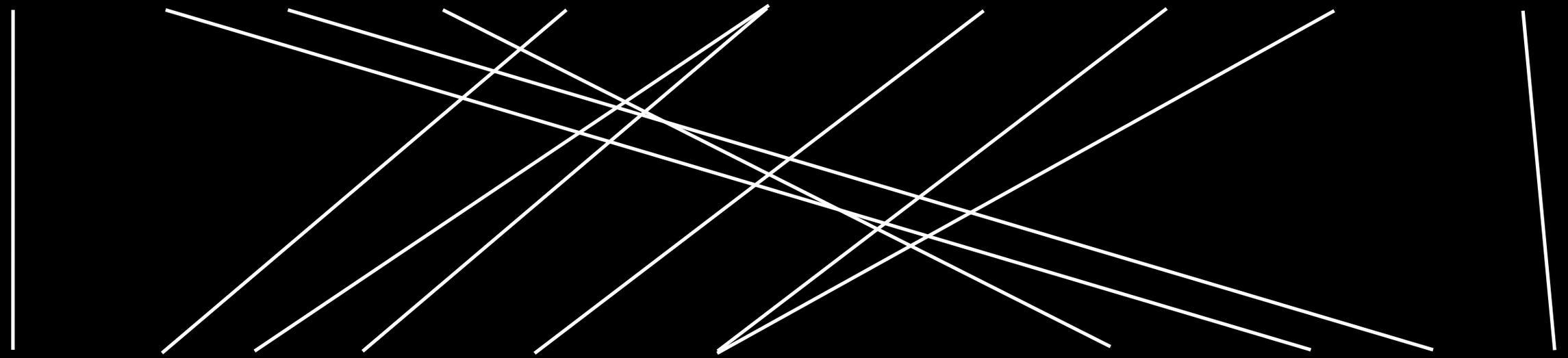
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



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

Although north wind howls , but sky still very clear .

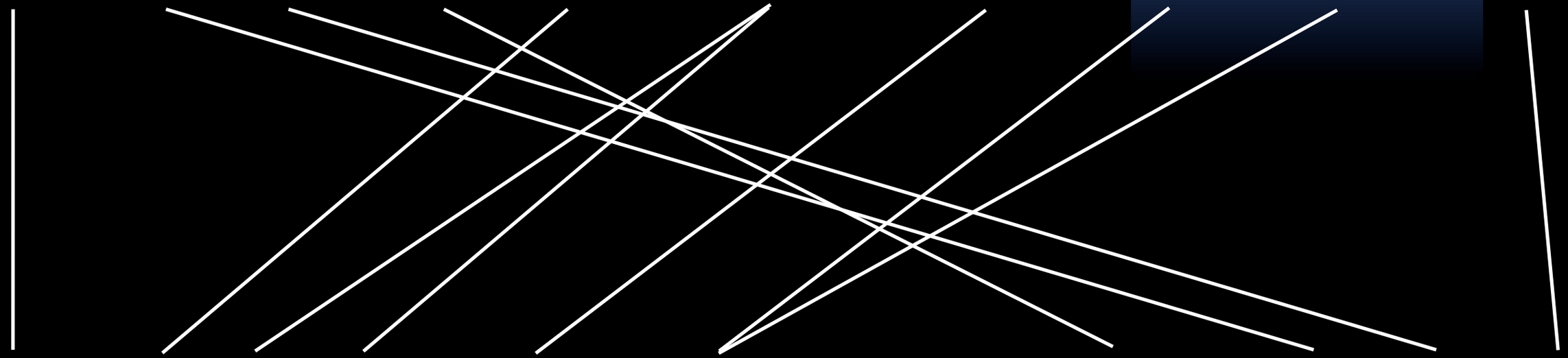
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



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

Although north wind howls , but sky still very clear .

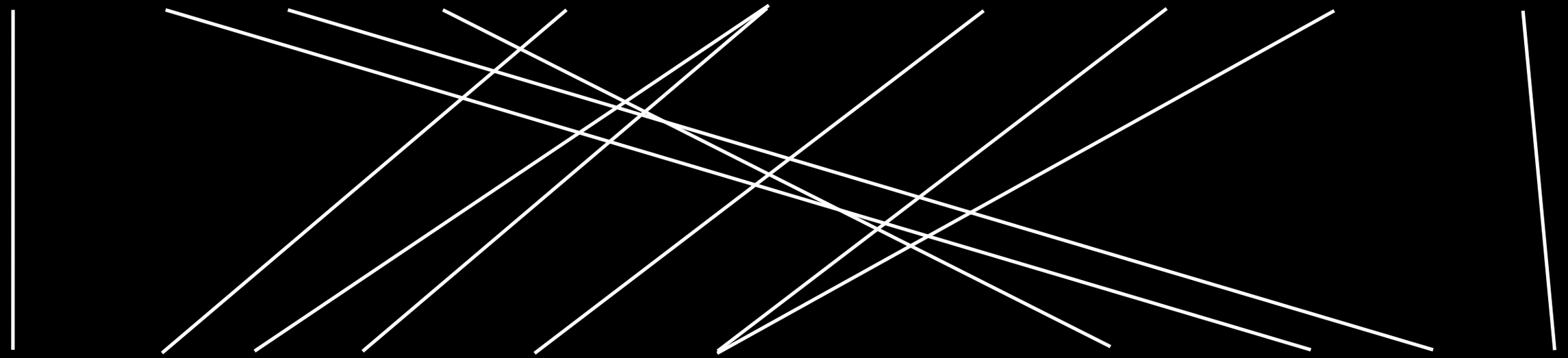
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



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

Although north wind howls , but sky still very clear .

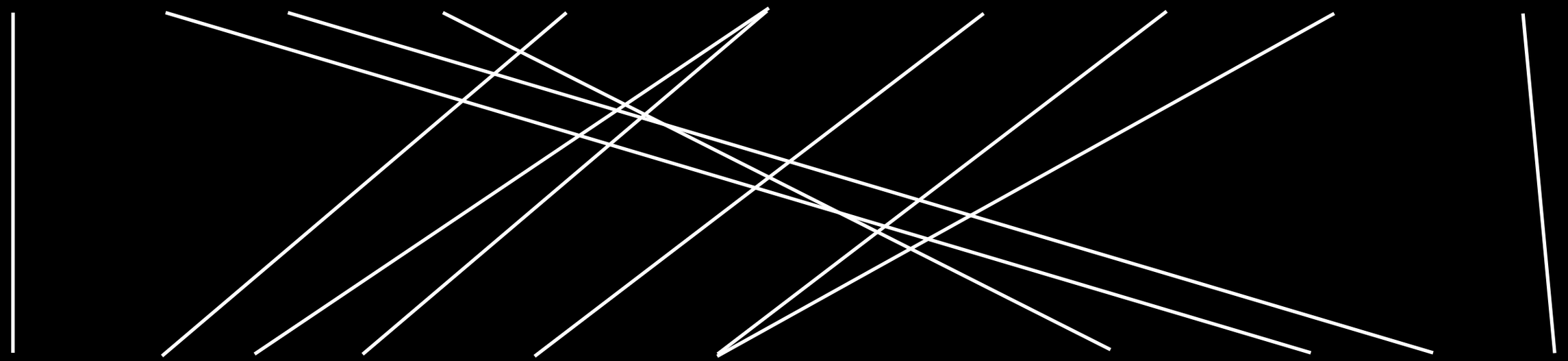
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



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

Although north wind howls , but sky still very clear .

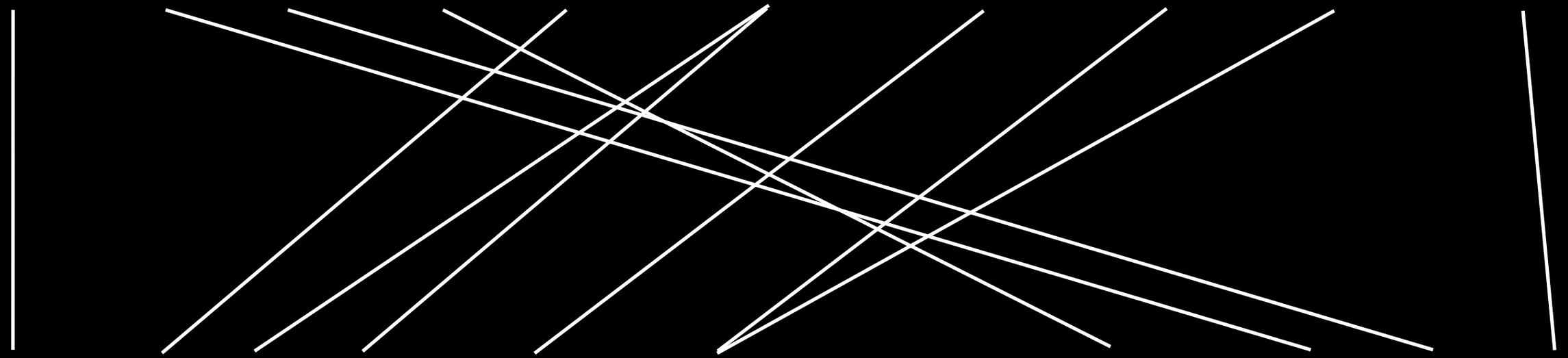
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



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

Although north wind howls , but sky still very clear .

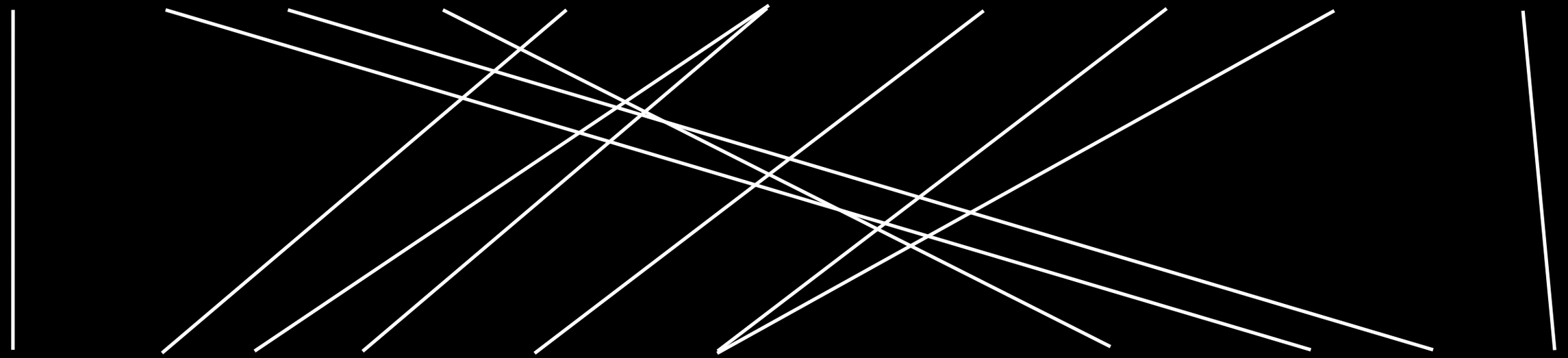
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



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 。

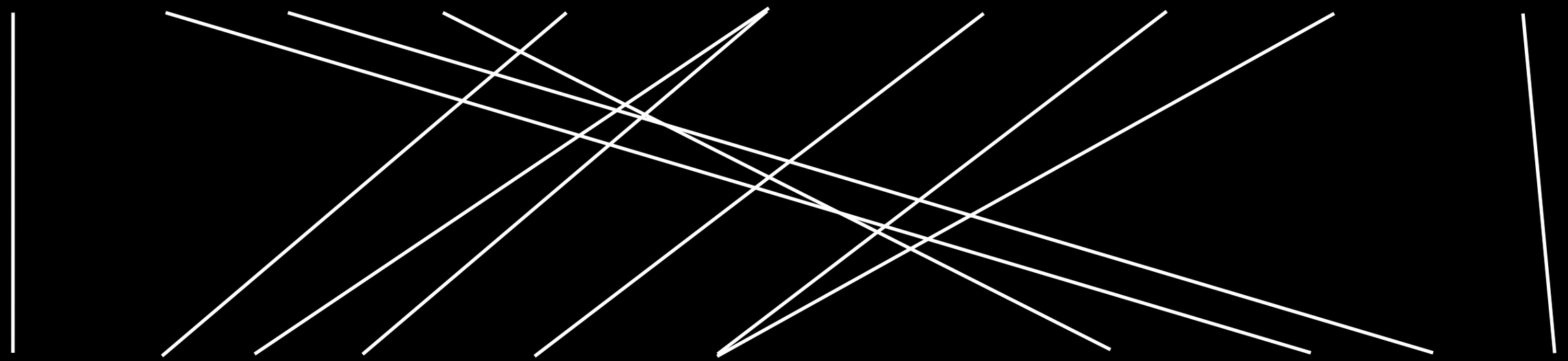


However , the sky remained clear under the strong north wind .

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 .

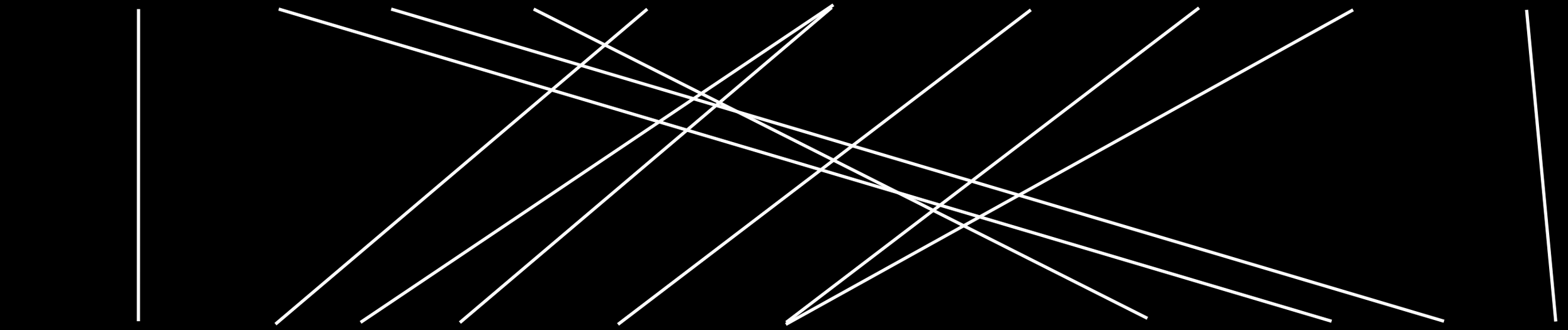
Despite the strong northerly winds , the sky remains very clear .

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 。

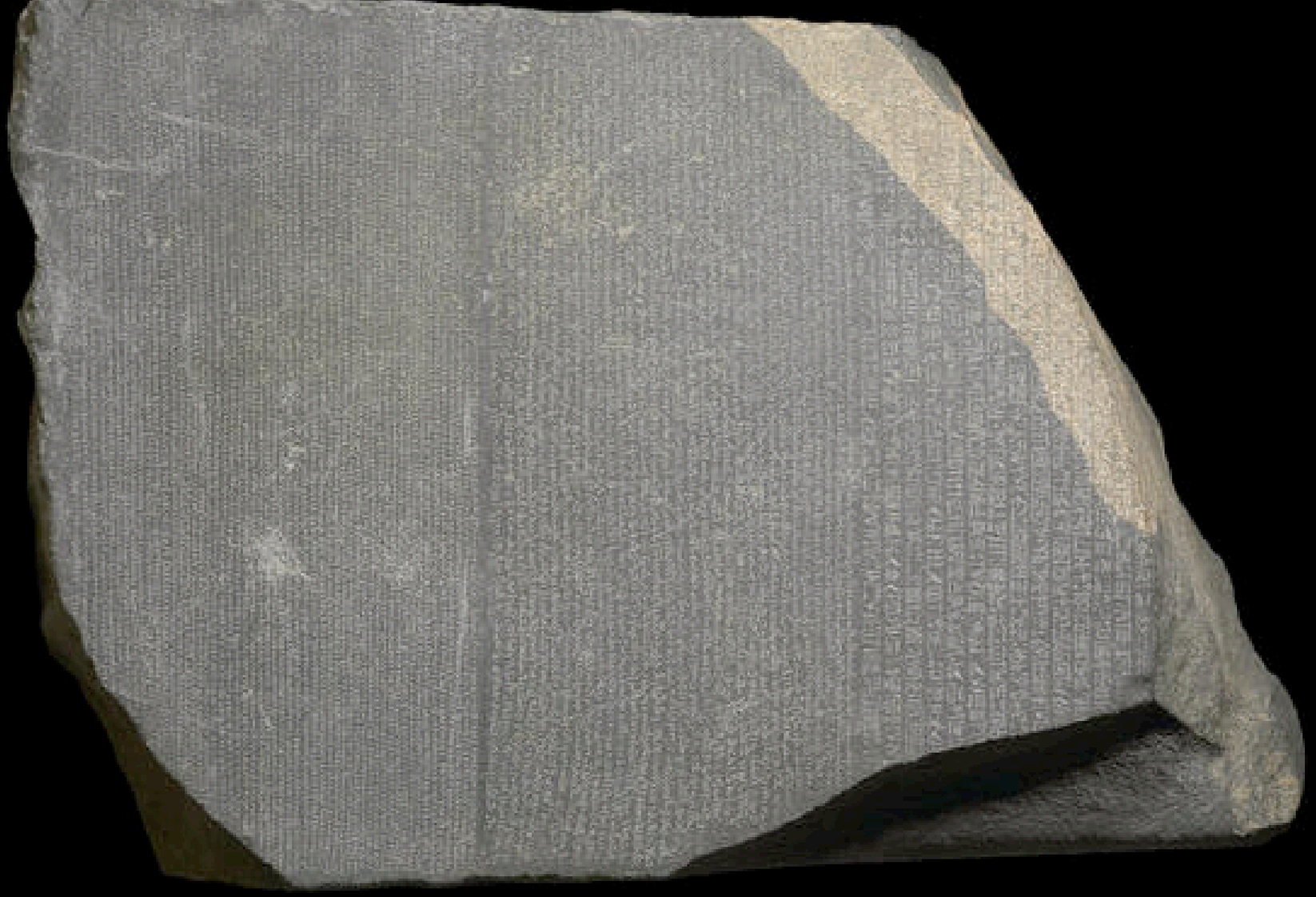


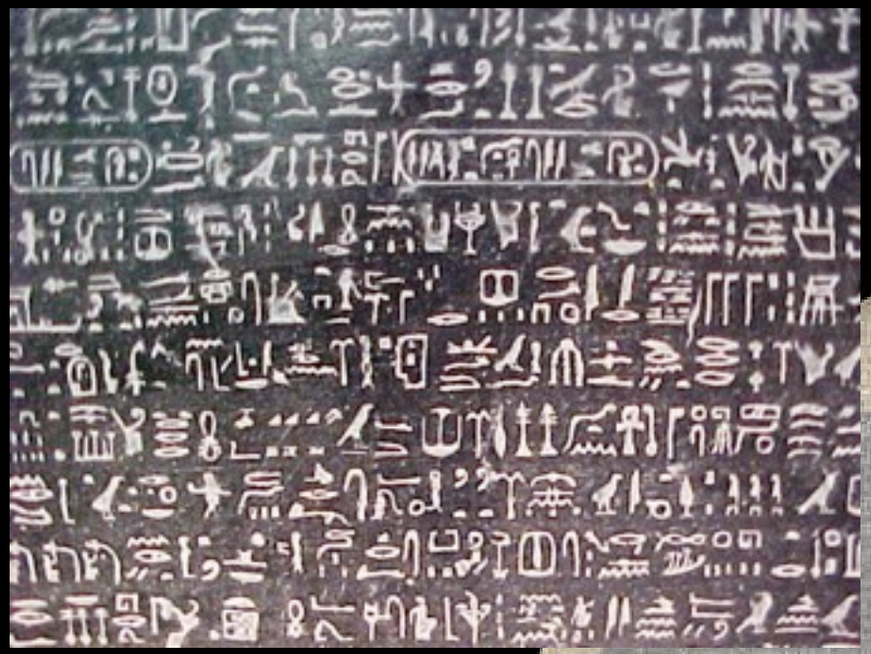
However , the sky remained clear under the strong north wind .

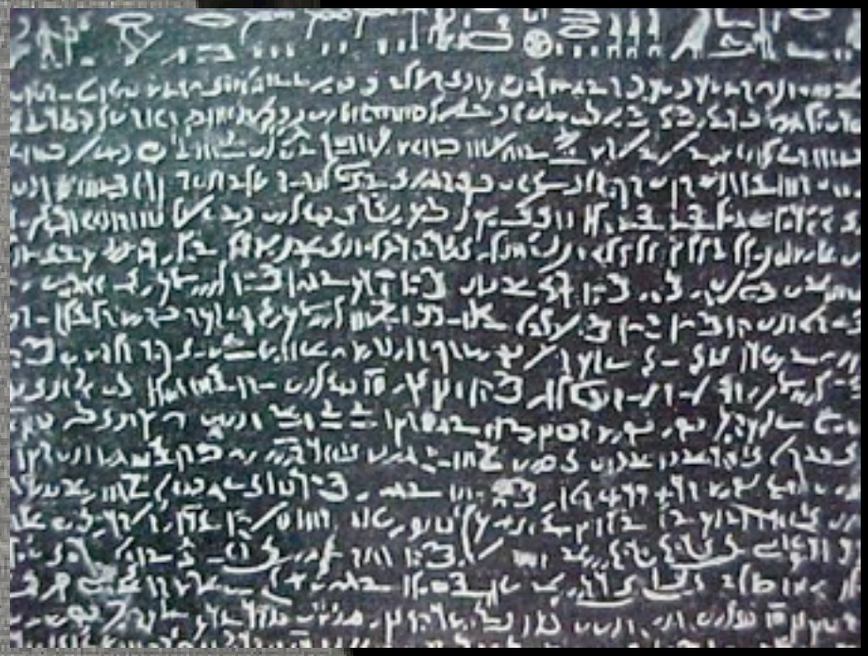
Despite the strong northerly winds , the sky remains very clear .

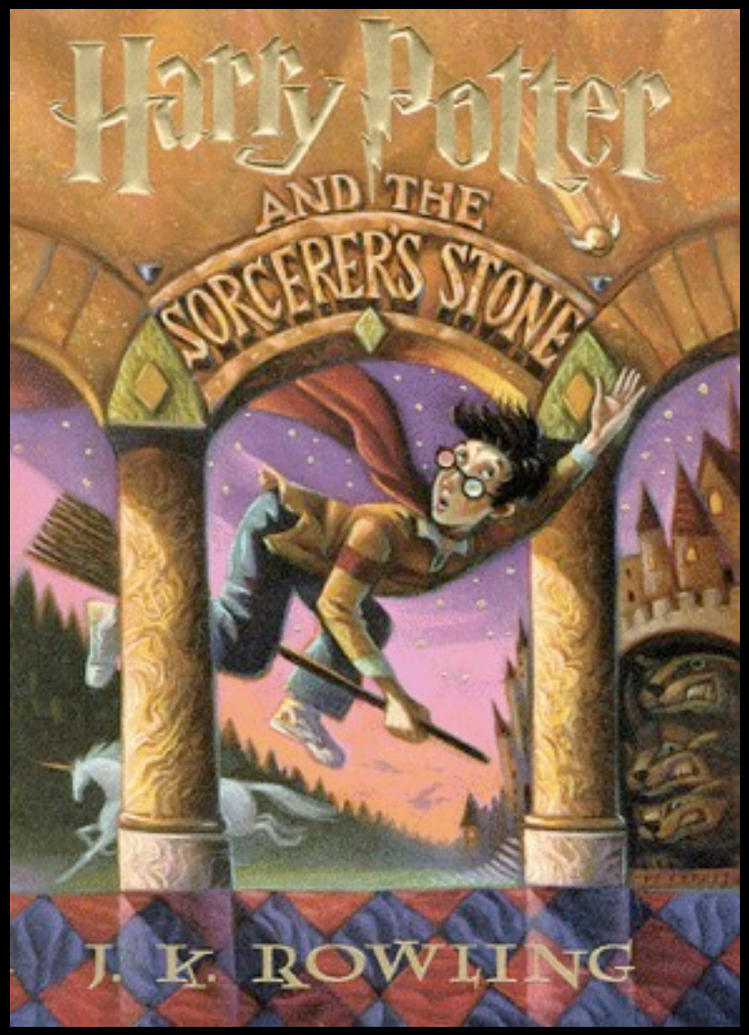
The sky was still crystal clear , though the north wind was howling .

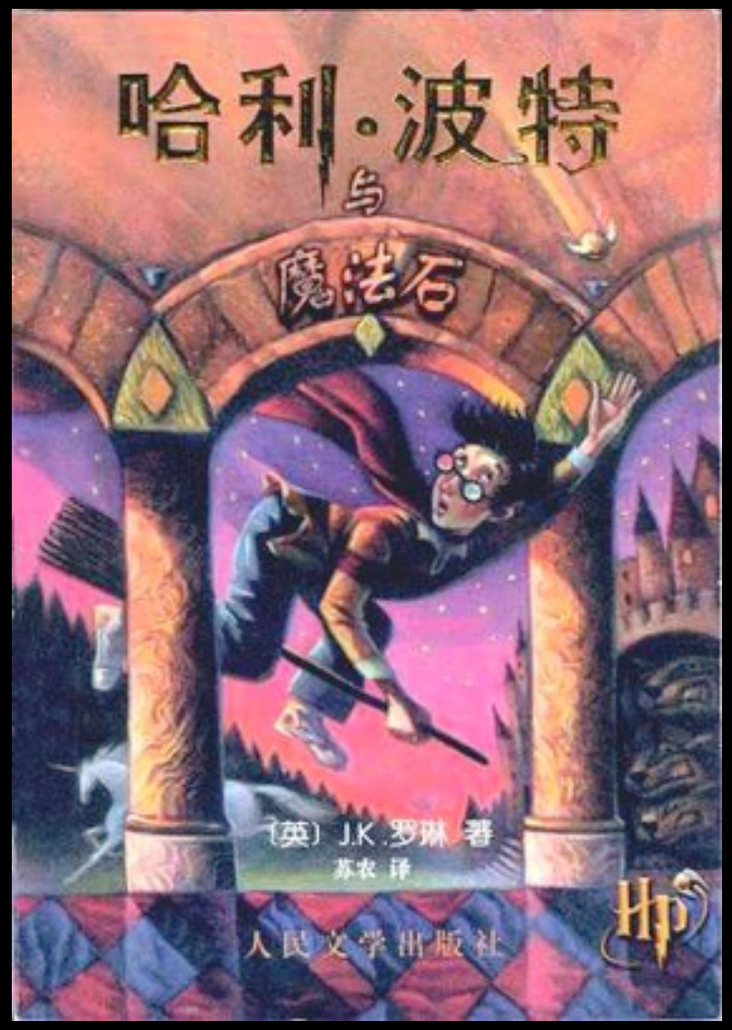
Although a north wind was howling , the sky remained clear and blue .

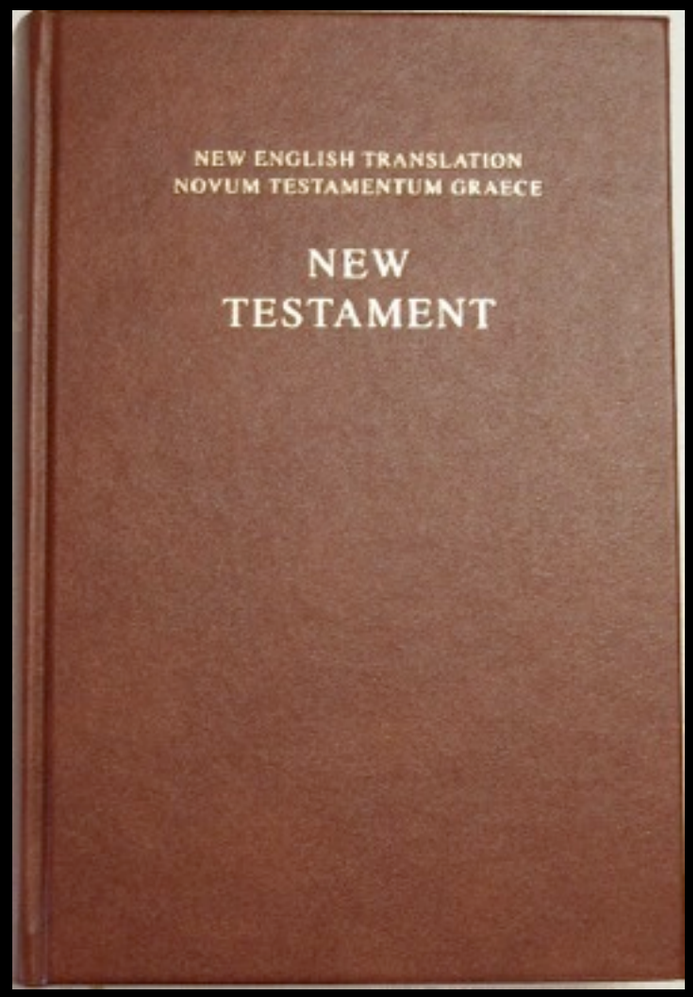







Harry Potter
AND THE
SORCERER'S STONE
J. K. ROWLING


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
■ 帶道本

www.un.org

http://www.un.org/english/

Google

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

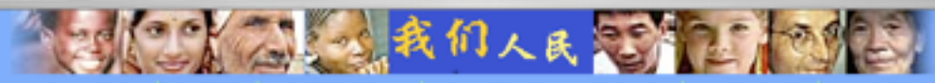
[Text version]

Live and On-Demand Webcasts, 24 Hours a Day: Click on UN Webcas

联合国主页

http://www.un.org/chinese/

Google



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

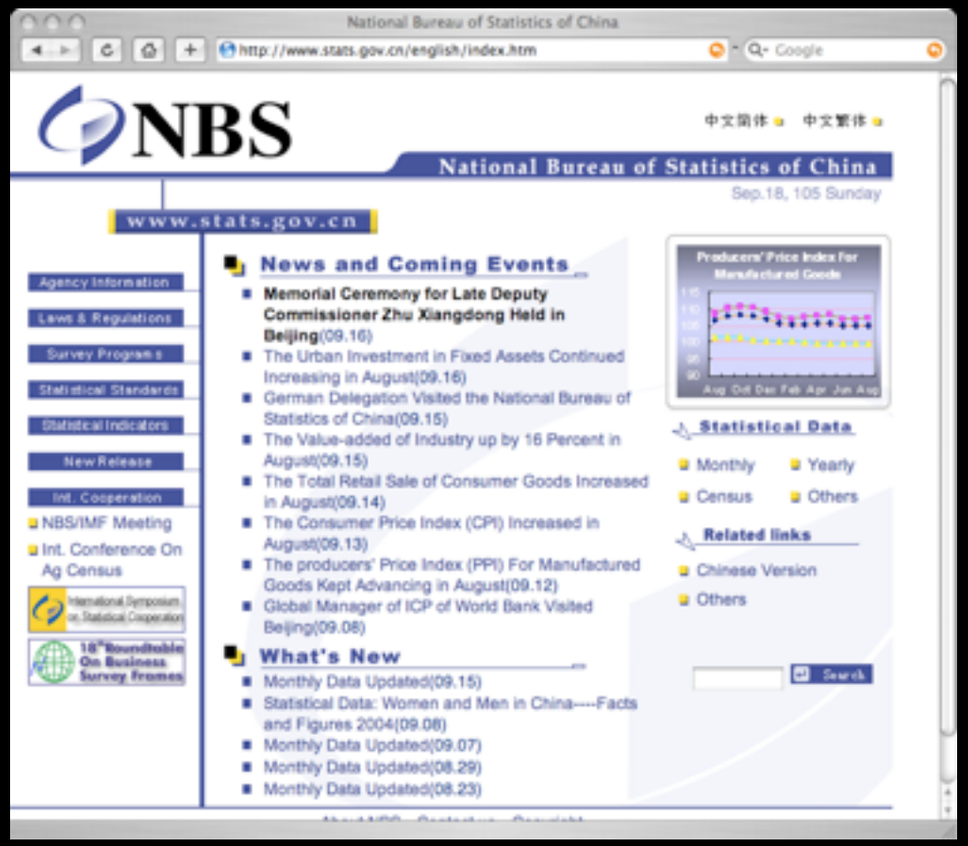
新增内容 | 工作机会 | 联合国采购 | 建议 | 问题与解答 | 其他网址 | 网址索引

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]

联合国网络直播


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

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
蕃茄雪湯	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
雪花湯	64.	Egg Drop Soup	1.10	2.10
雲雪湯	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

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

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

2002
éIRE

EURO

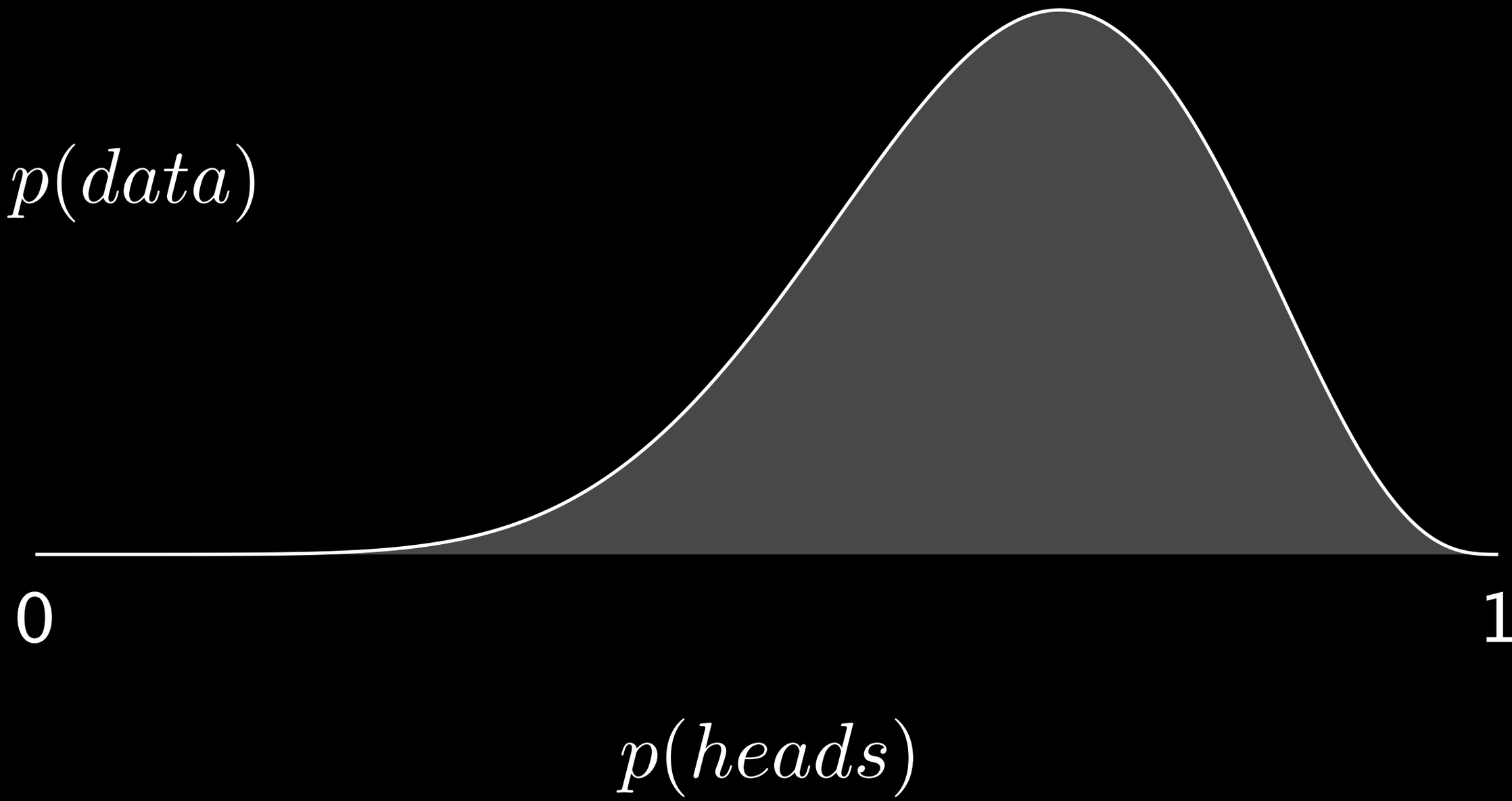
$p(heads)$ $\quad$ $1 - p(heads)$

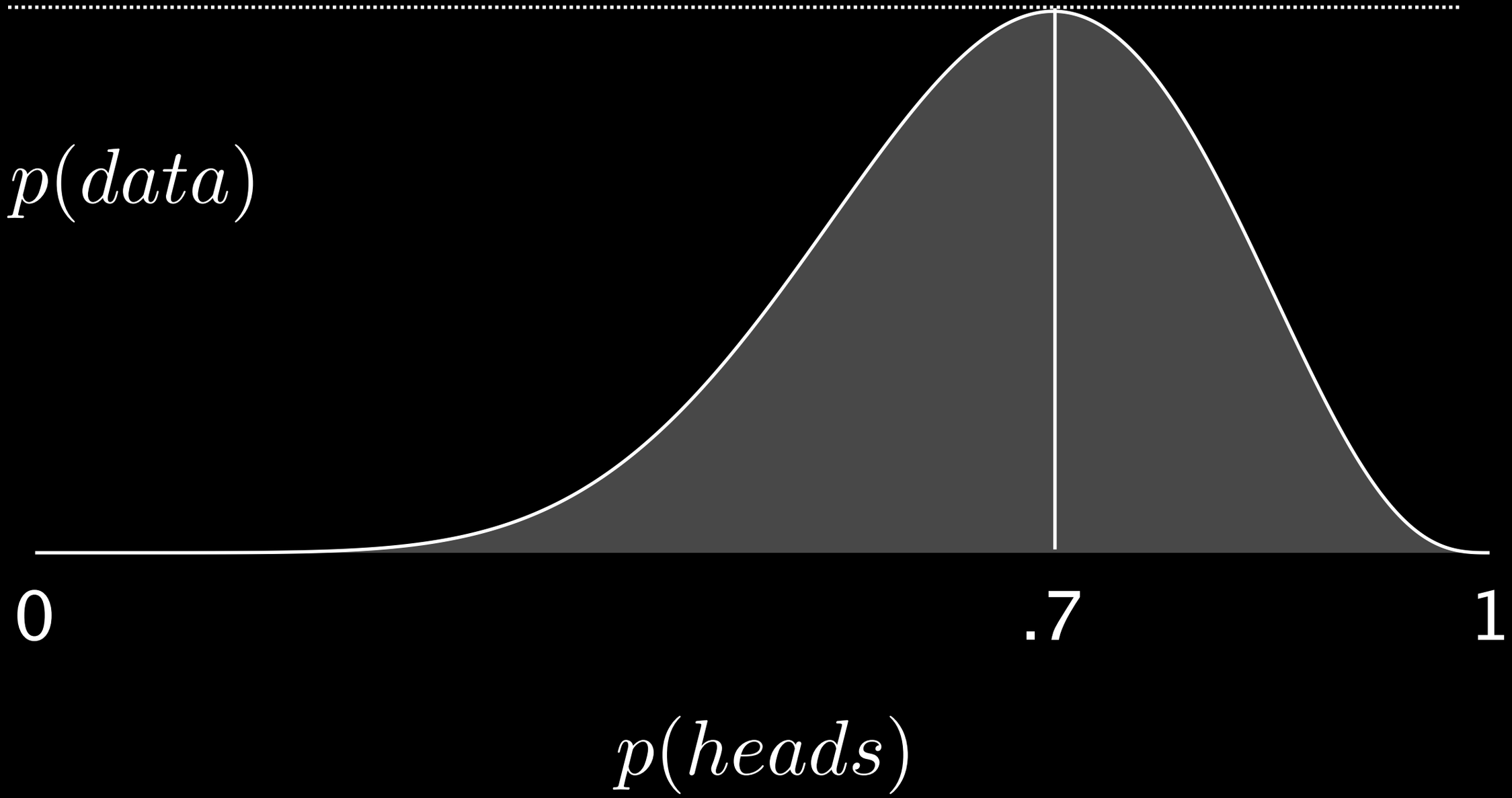
1
EURO
2002
éIRe

$$p(heads)\,?$$

$$p(data) = p(heads)^7 \times p(tails)^3$$

$$p(data) = p(heads)^7 \times [1 - p(heads)]^3$$


$p(data)$
0
1
$p(heads)$


$p(data)$
0
.7
1
$p(heads)$

THE WORD

- Optimization

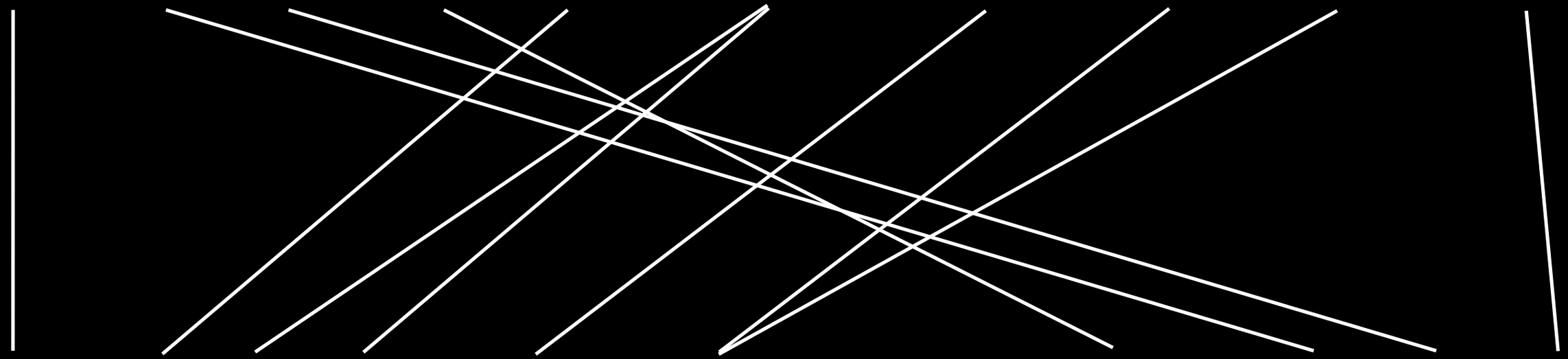
1
EURO

$p(heads)$?

$p(heads)$?

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 .

This is a *latent variable* problem

More on learning from parallel data:

Today's lab

Tomorrow's lecture

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.

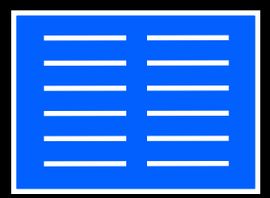
Statistical Machine Translation

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

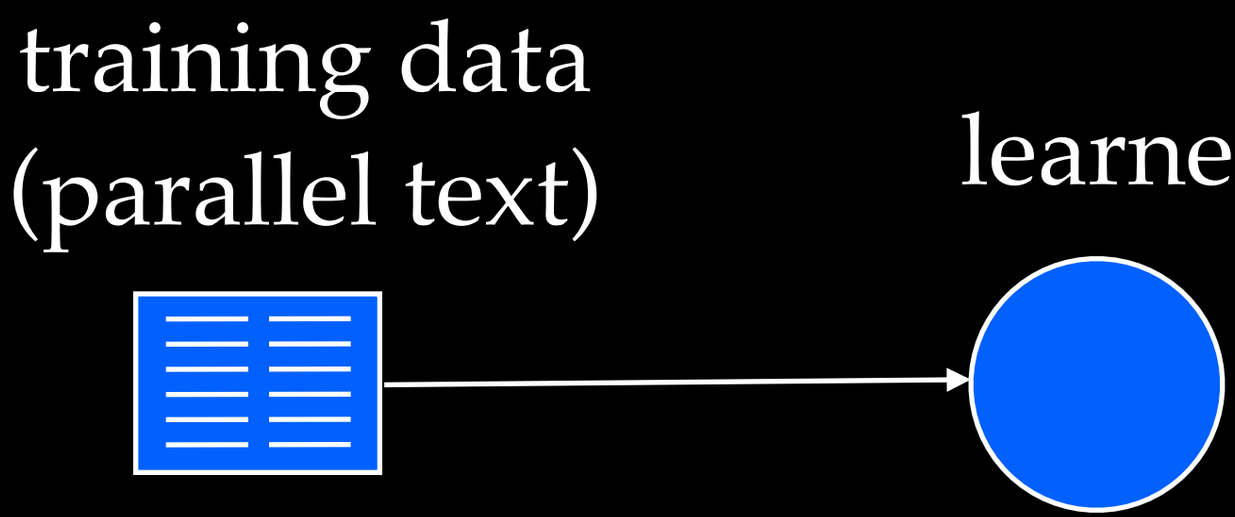
Overview

Overview

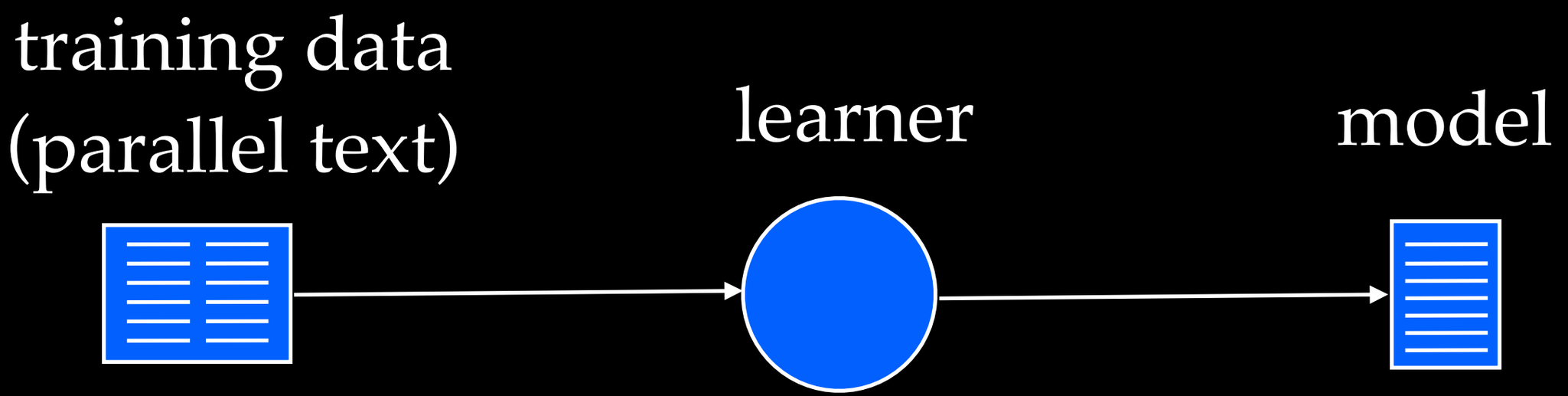
training data
(parallel text)



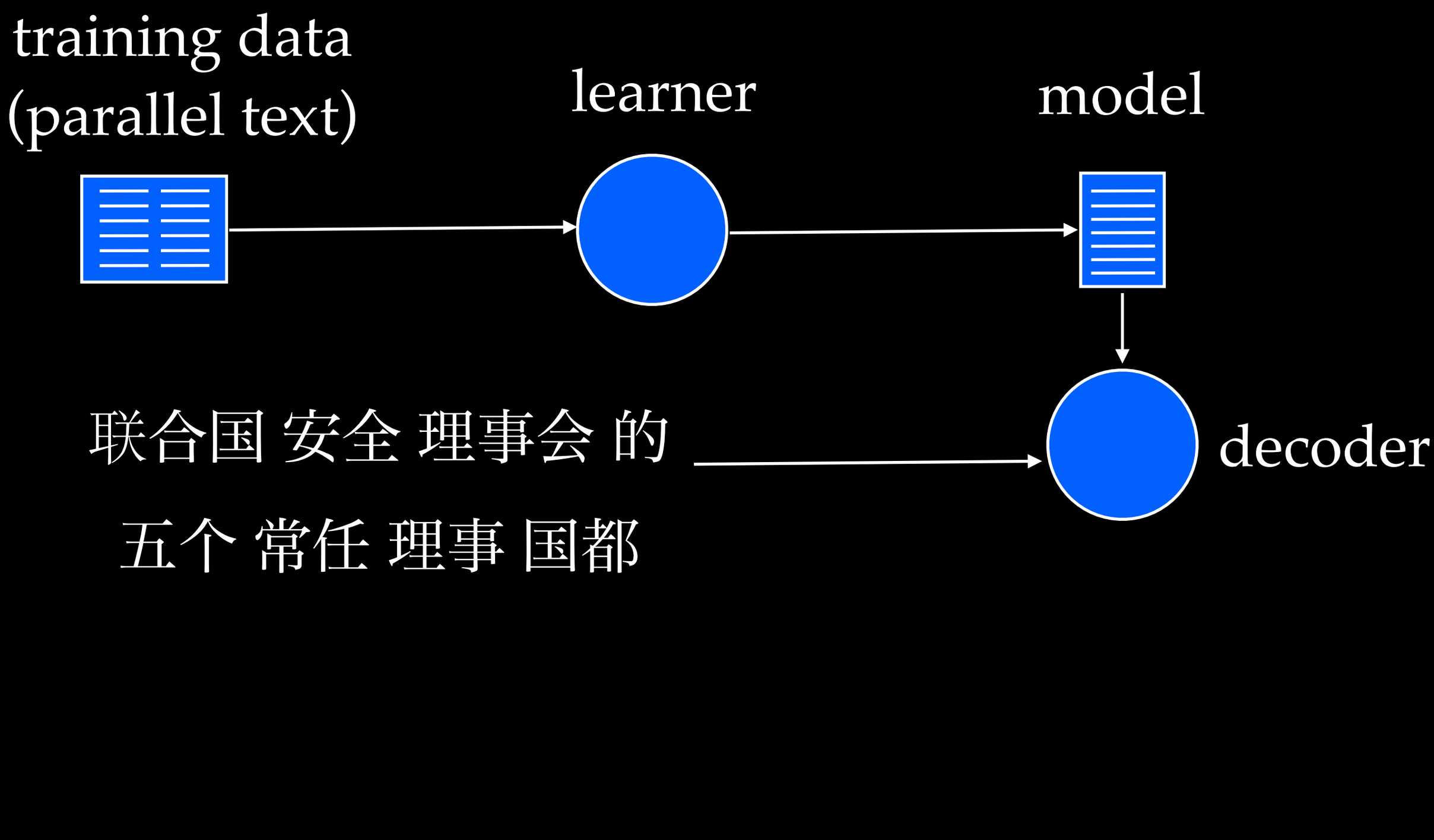
Overview



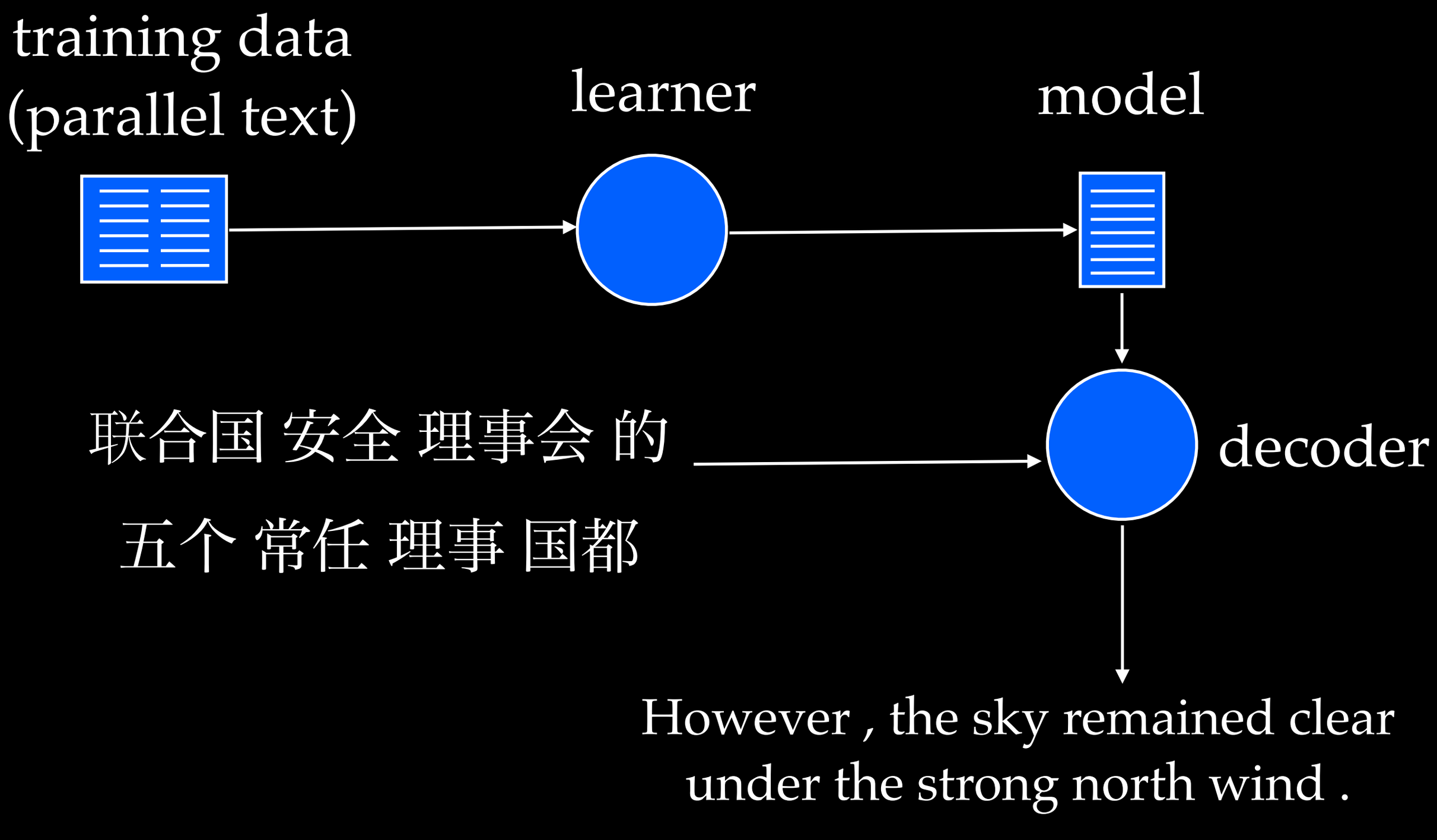
Overview



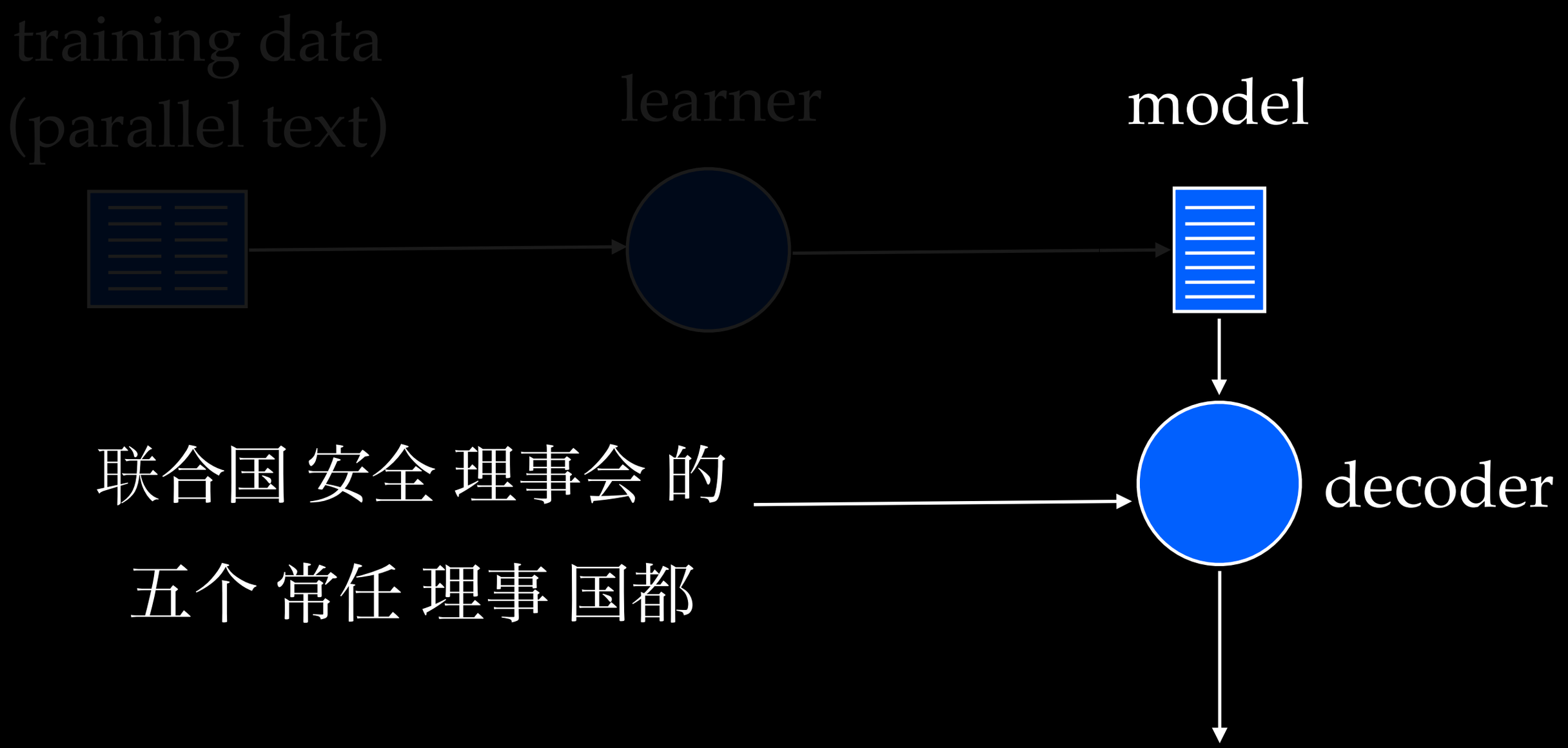
Overview


training data
(parallel text)
learner
model
联合国 安全 理事会 的
五个 常任 理事 国都
decoder

Overview


training data
(parallel text)
learner
model
联合国 安全 理事会 的
五个 常任 理事 国都
decoder
However , the sky remained clear
under the strong north wind .

Overview


training data
(parallel text)
learner
model
联合国 安全 理事会 的
五个 常任 理事 国都
decoder
However , the sky remained clear
under the strong north wind .

What's a model?

What's a model?

For our purposes, a model will be

a probability distribution over sentence pairs.

Why Probability?

Why Probability?

- Access to techniques developed and proven over hundreds of years that work on many problems.

Why Probability?

- Access to techniques developed and proven over hundreds of years that work on many problems.
- In particular, techniques for *learning* and *prediction*.

Why Probability?

- Access to techniques developed and proven over hundreds of years that work on many problems.
- In particular, techniques for *learning* and *prediction*.
- Allows us to answer questions:

Why Probability?

- Access to techniques developed and proven over hundreds of years that work on many problems.
- In particular, techniques for *learning* and *prediction*.
- Allows us to answer questions:
  - What is the best explanation of observed data?

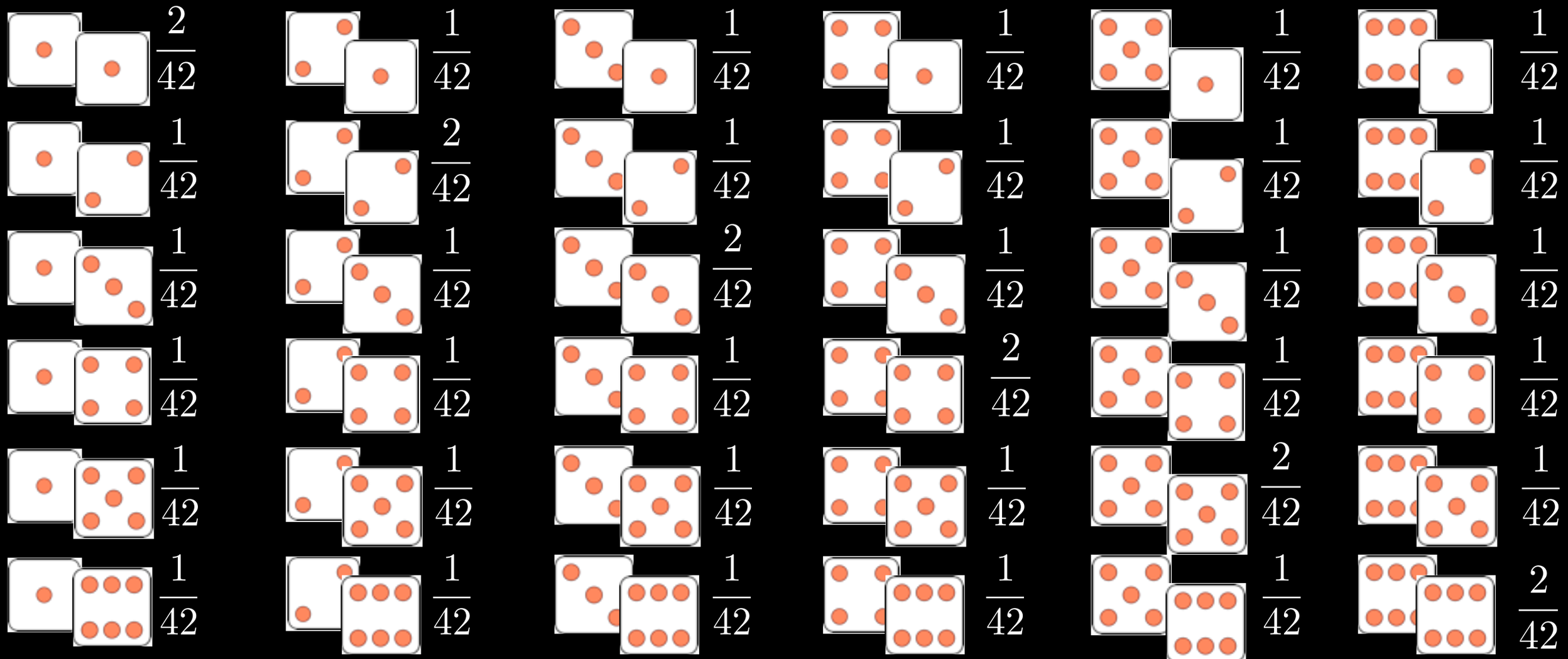
Why Probability?

- Access to techniques developed and proven over hundreds of years that work on many problems.
- In particular, techniques for *learning* and *prediction*.
- Allows us to answer questions:
  - What is the best explanation of observed data?
  - Given some partially observed data (e.g. an input sentence), what is the most likely complete data (e.g. a sentence pair)?

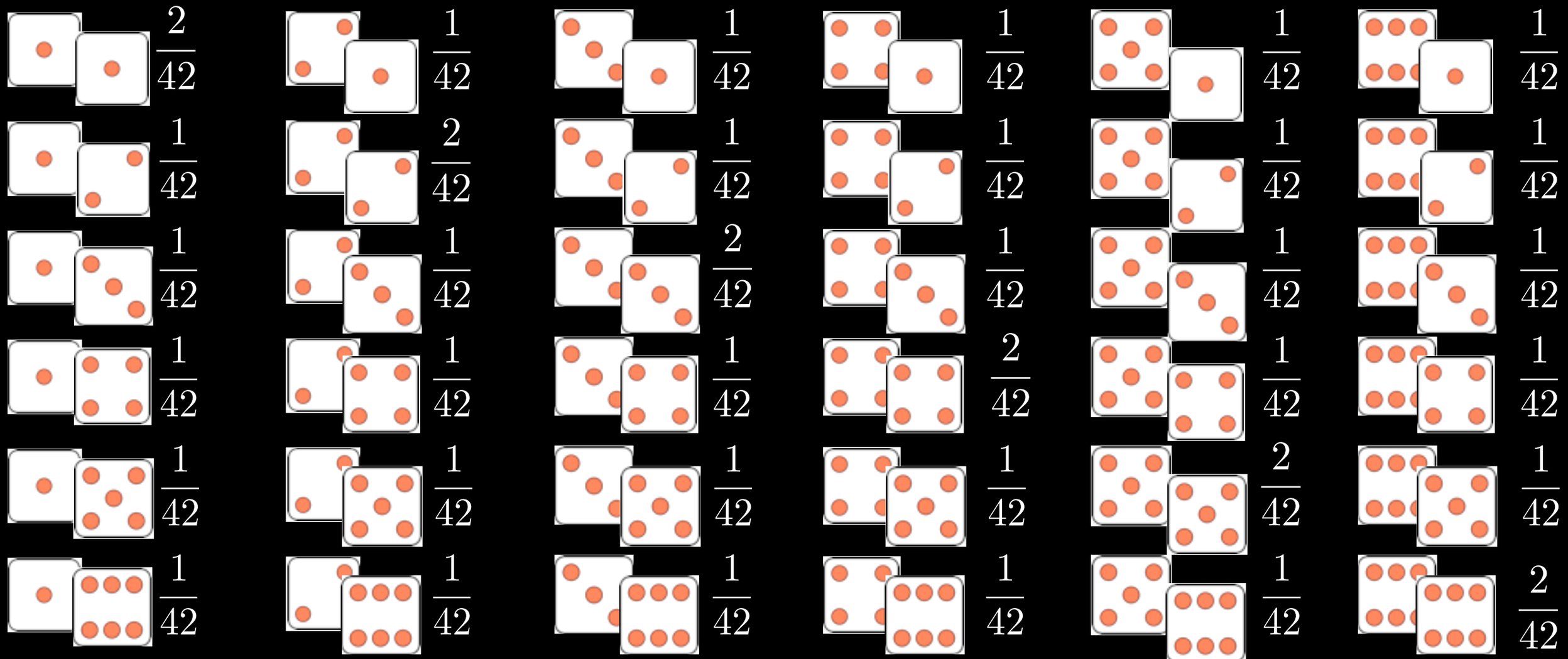
Why Probability?

- Access to techniques developed and proven over hundreds of years that work on many problems.
- In particular, techniques for *learning* and *prediction*.
- Allows us to answer questions:
  - What is the best explanation of observed data?
  - Given some partially observed data (e.g. an input sentence), what is the most likely complete data (e.g. a sentence pair)?
- Common sense in mathematical form!

Probabilistic Primer

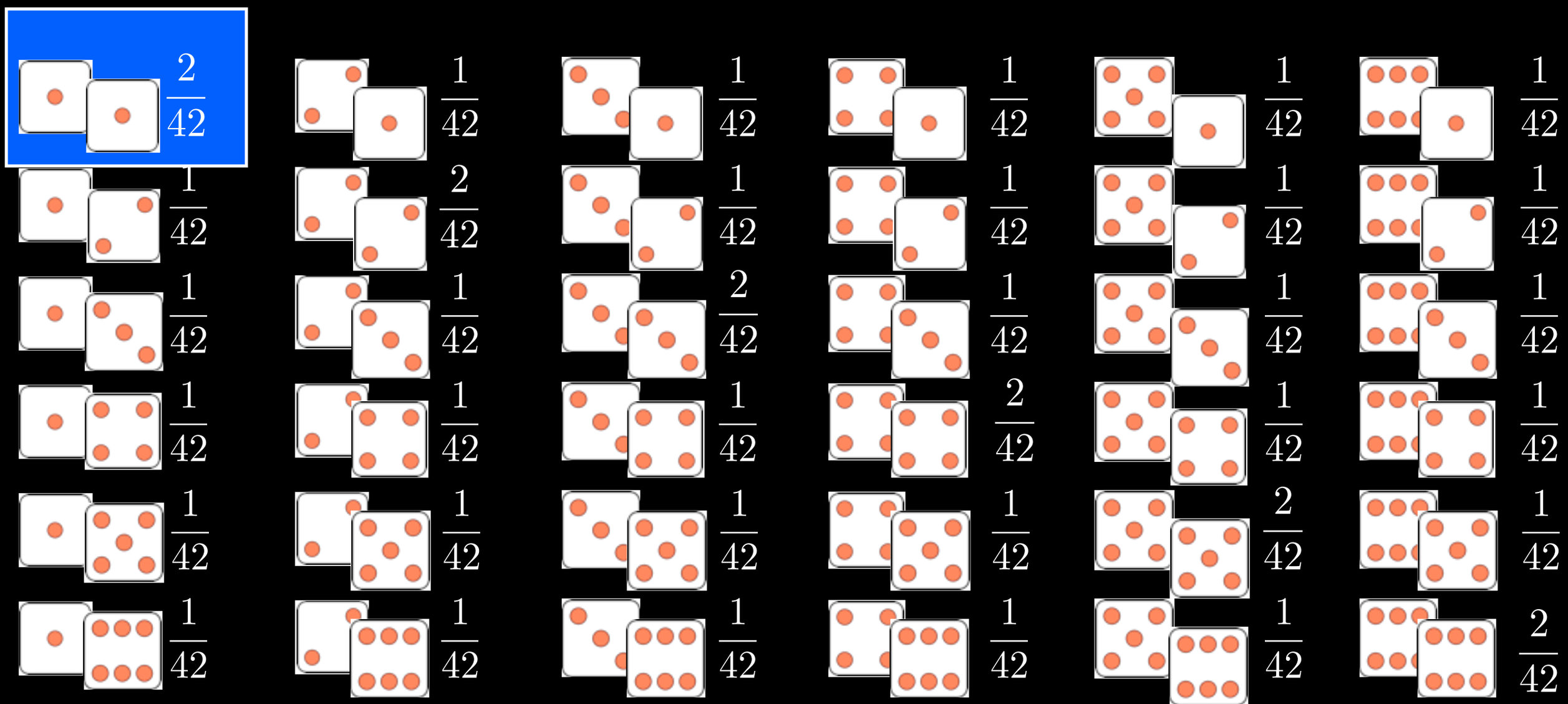

2/42
1/42
1/42
1/42
1/42
1/42
1/42
2/42
1/42
1/42
1/42
1/42
1/42
1/42
2/42
1/42
1/42
1/42
1/42
1/42
1/42
1/42
2/42
1/42
1/42
1/42
1/42
1/42
1/42
1/42
2/42
1/42
1/42
1/42
1/42
1/42
1/42
1

Probabilistic Primer



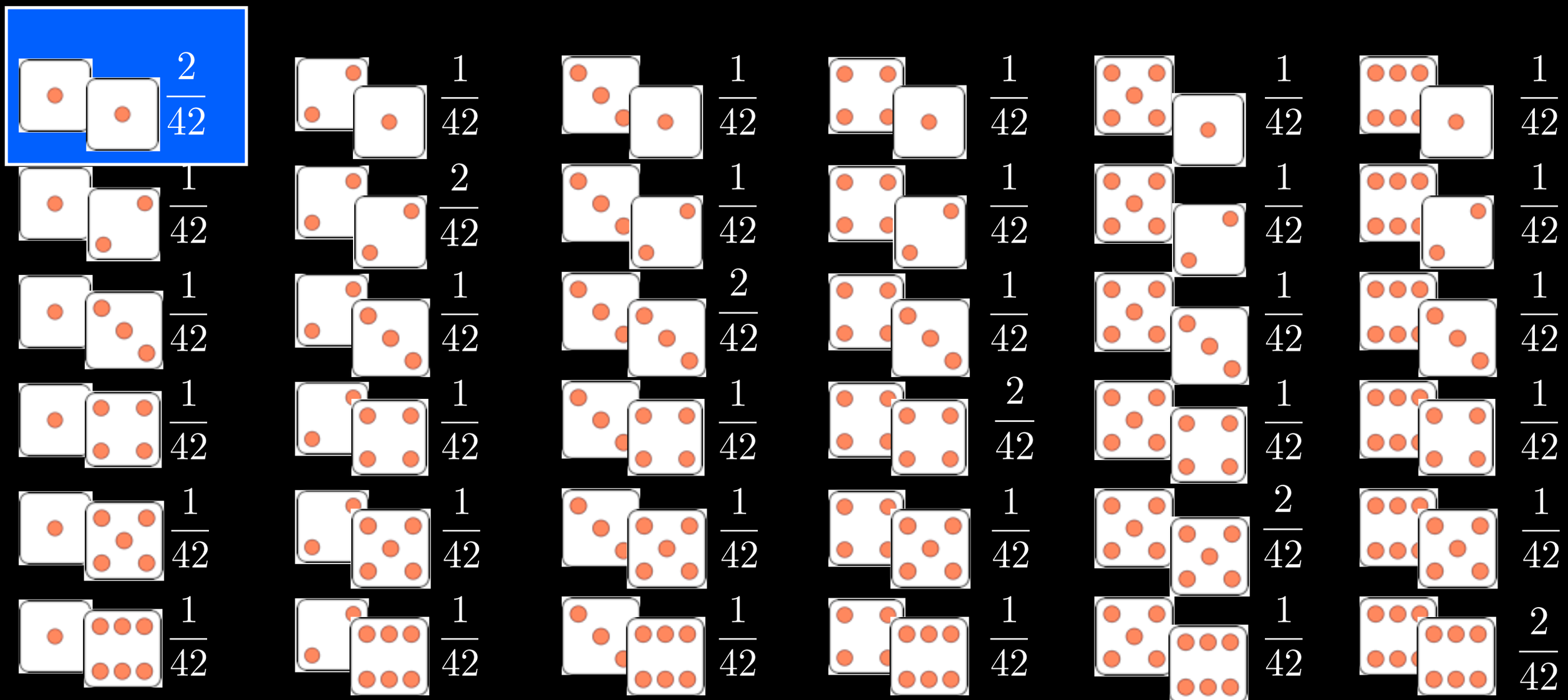
When an event consists of observations about more than one variable, it is a ***joint probability***.

Probabilistic Primer



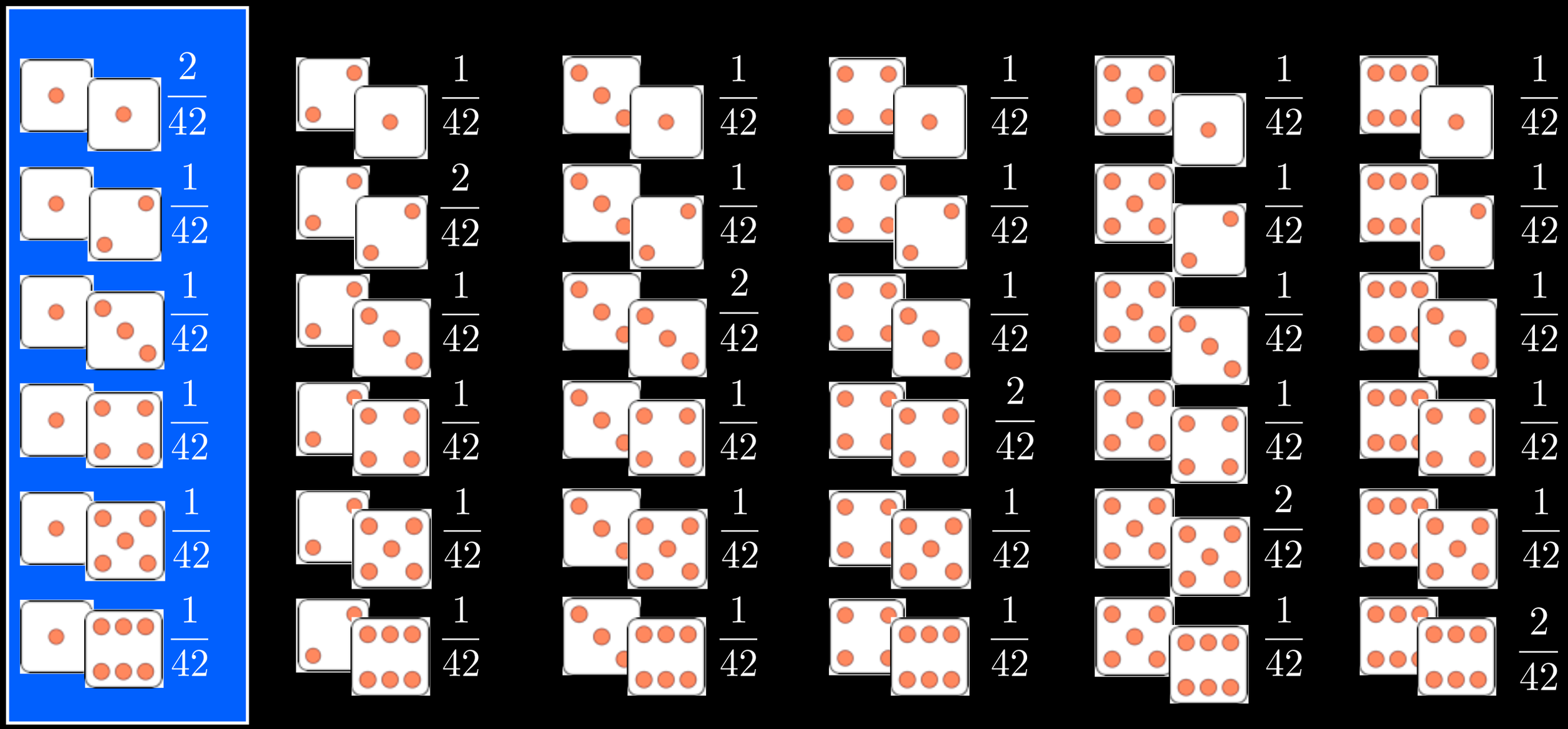
$$p(A = 1, B = 1) = \frac{2}{42}$$

Probabilistic Primer



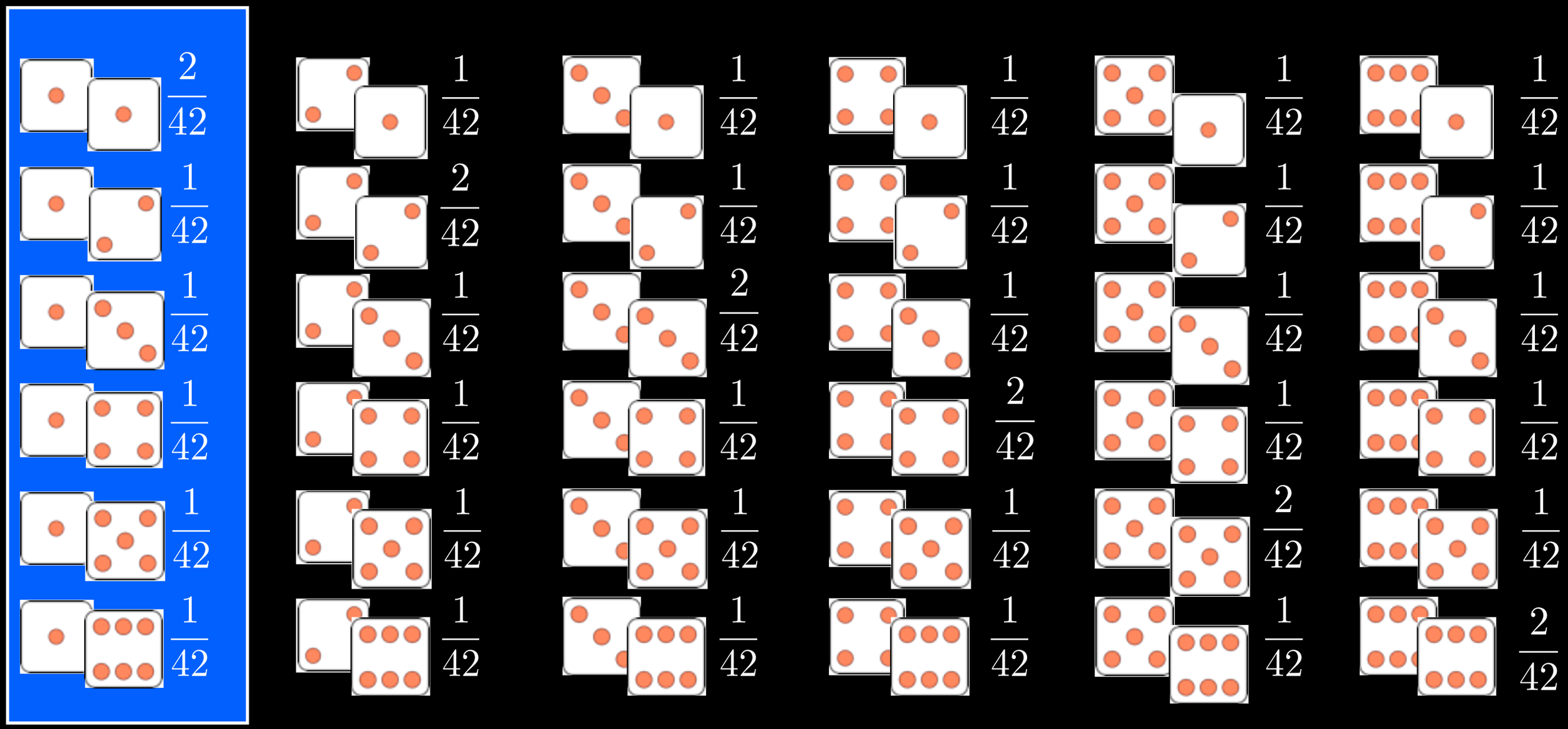
$$p(1,1) = \frac{2}{42}$$

Probabilistic Primer



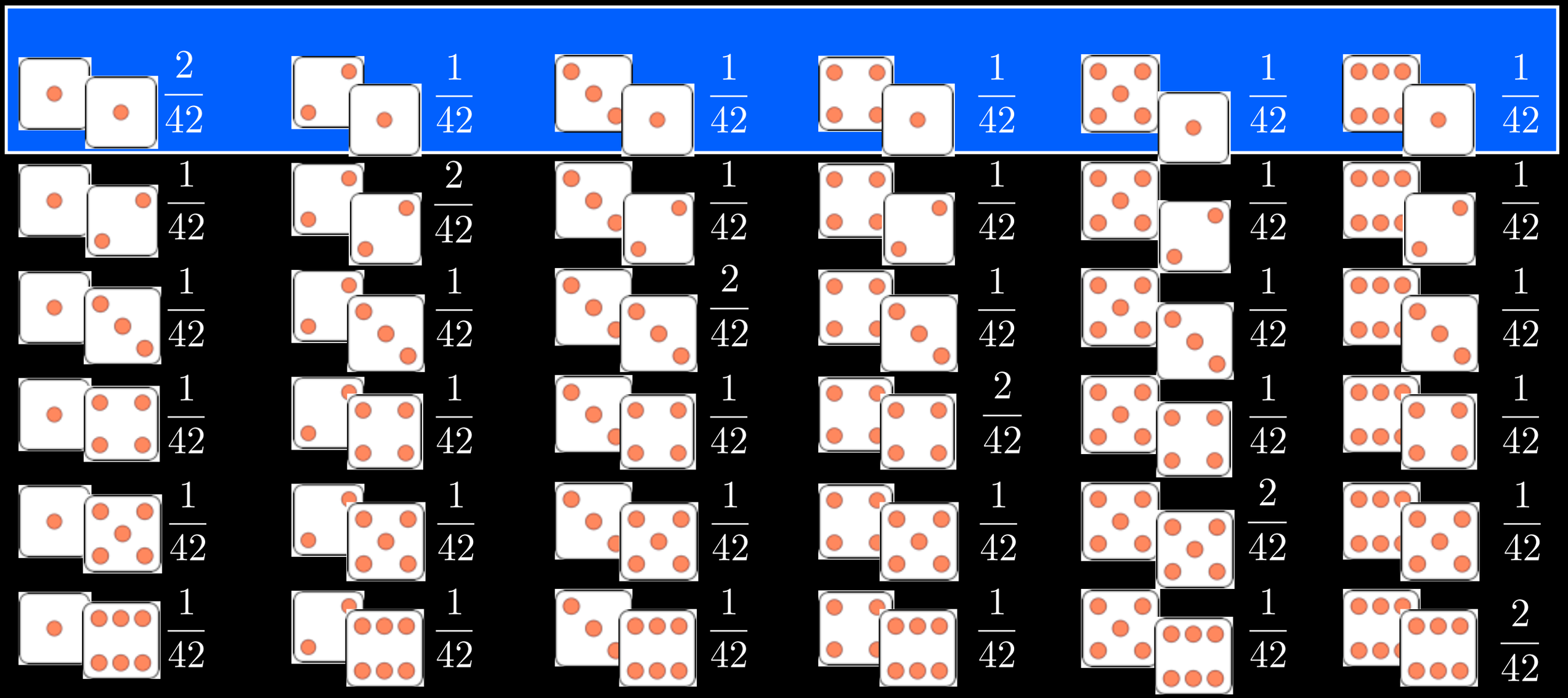
A probability distribution over a subset of variables is a *marginal probability*.

Probabilistic Primer



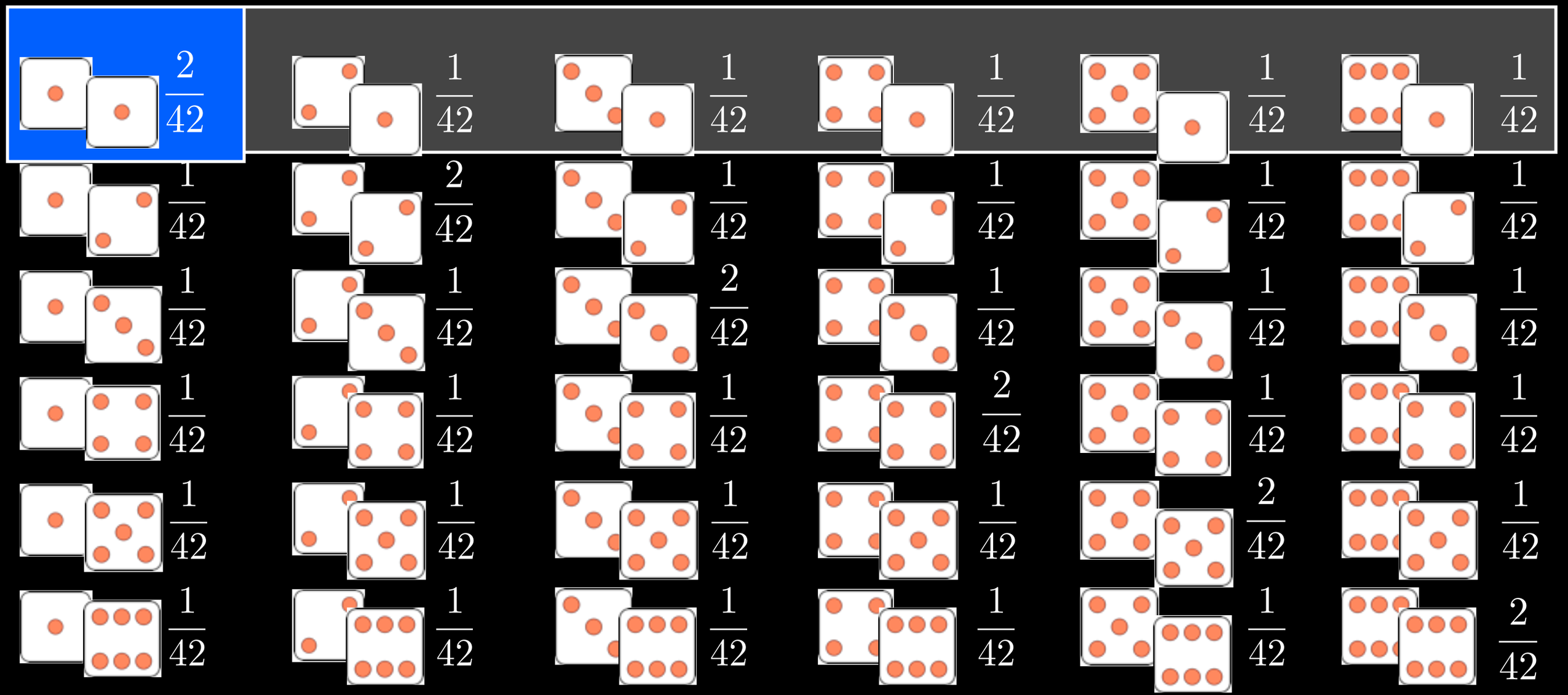
$$p(B = 1) = p(\cdot, 1) = \sum_{a \in A} p(A = a, B = 1) = \frac{1}{6}$$

Probabilistic Primer



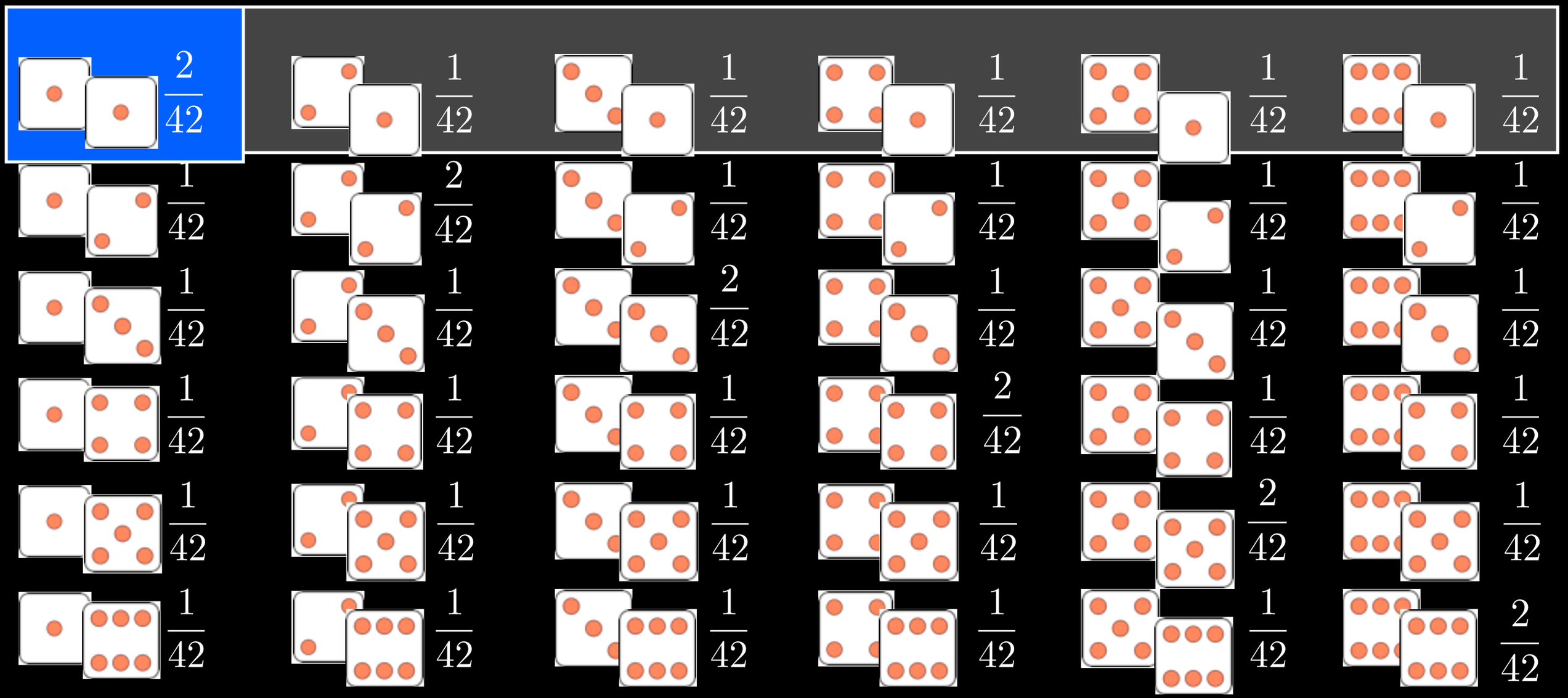
$$p(A = 1) = p(1, \cdot) = \sum_{b \in B} p(A = 1, B = b) = \frac{1}{6}$$

Probabilistic Primer



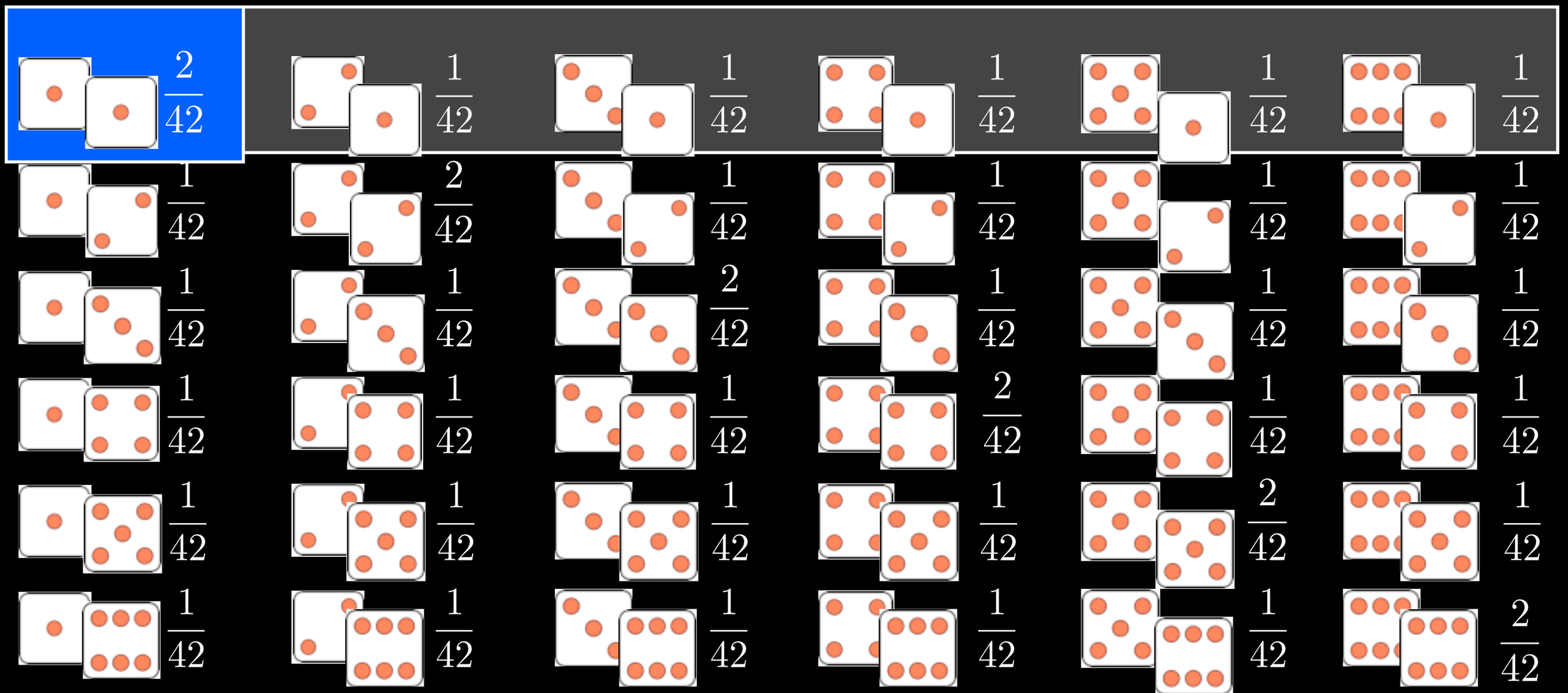
The probability of a variable under the condition that the other variables are fixed is the ***conditional probability***.

Probabilistic Primer



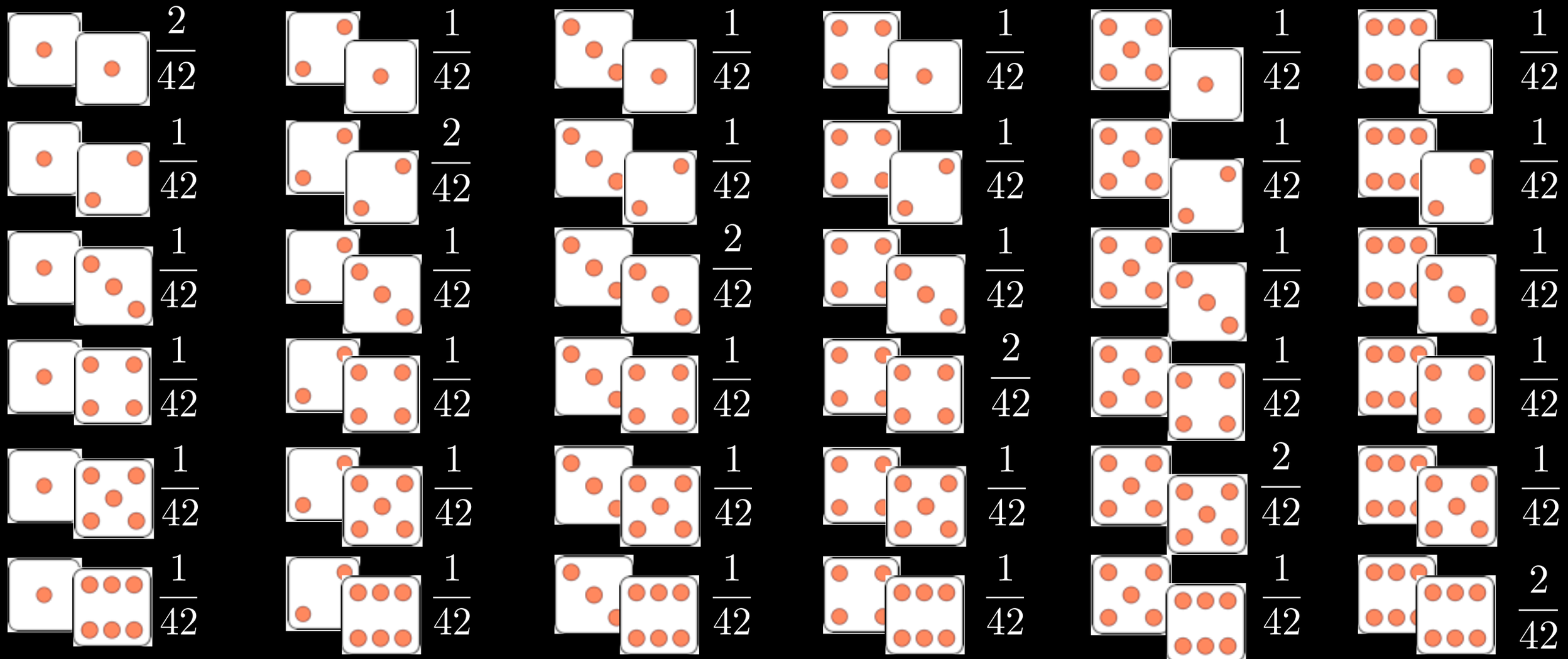
$$p(B = 1|A = 1) = \frac{p(A = 1, B = 1)}{\sum_{b \in B} p(A = 1, B = b)} = \frac{2}{7}$$

Probabilistic Primer



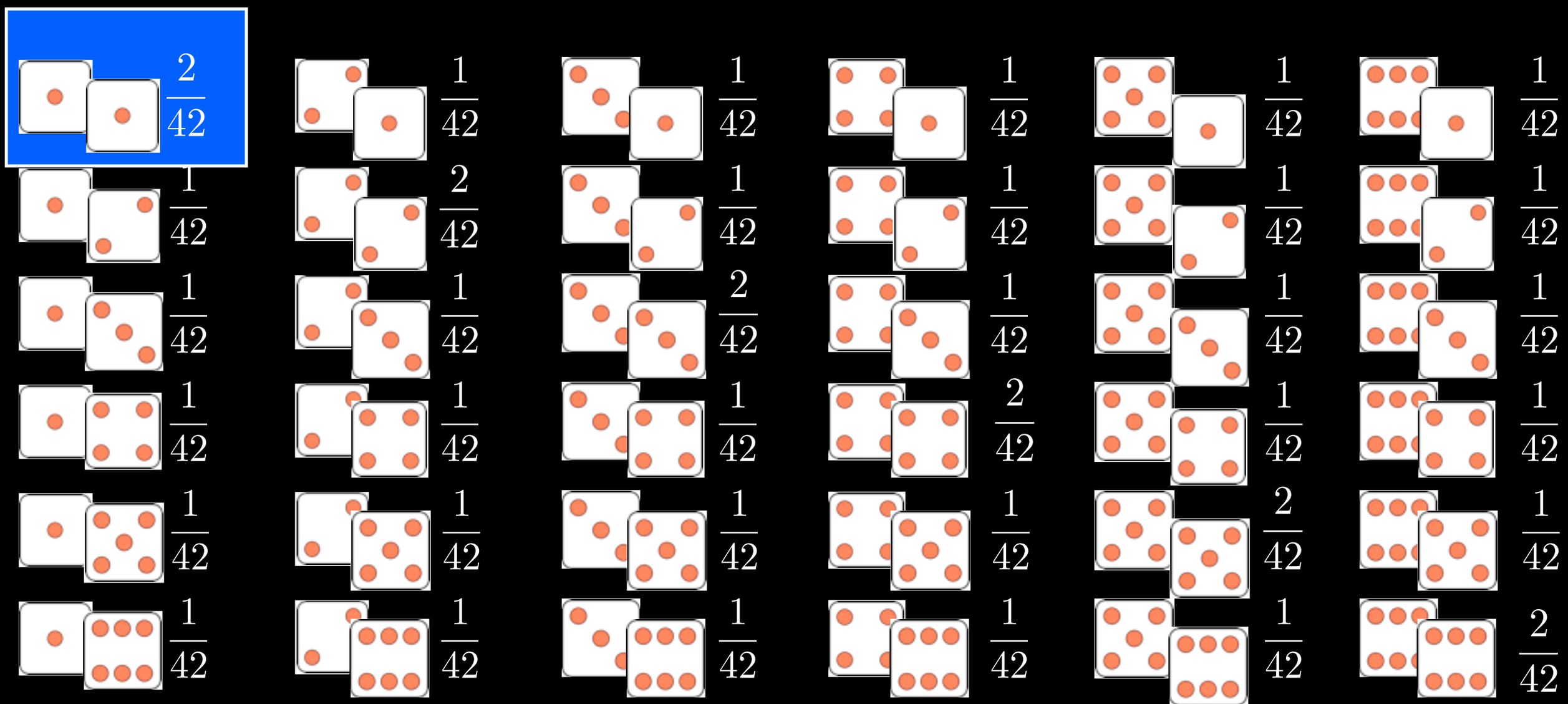
$$p(B = 1|A = 1) = \frac{p(A = 1, B = 1)}{\sum_{b \in B} p(A = 1, B = b)} = \frac{2}{7} \qquad \frac{\text{joint}}{\text{marginal}}$$

Probabilistic Primer



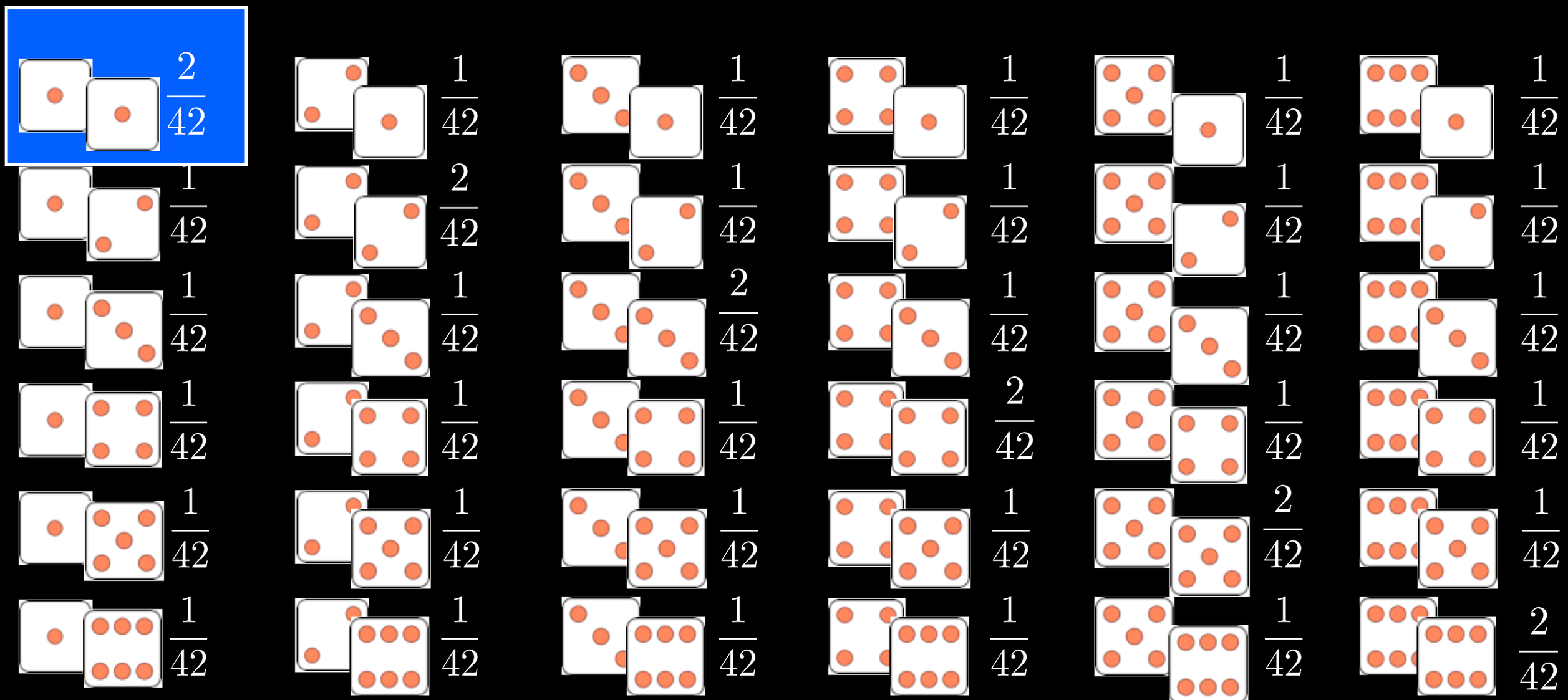
We can still represent the joint distribution as a product of other distributions.

Probabilistic Primer



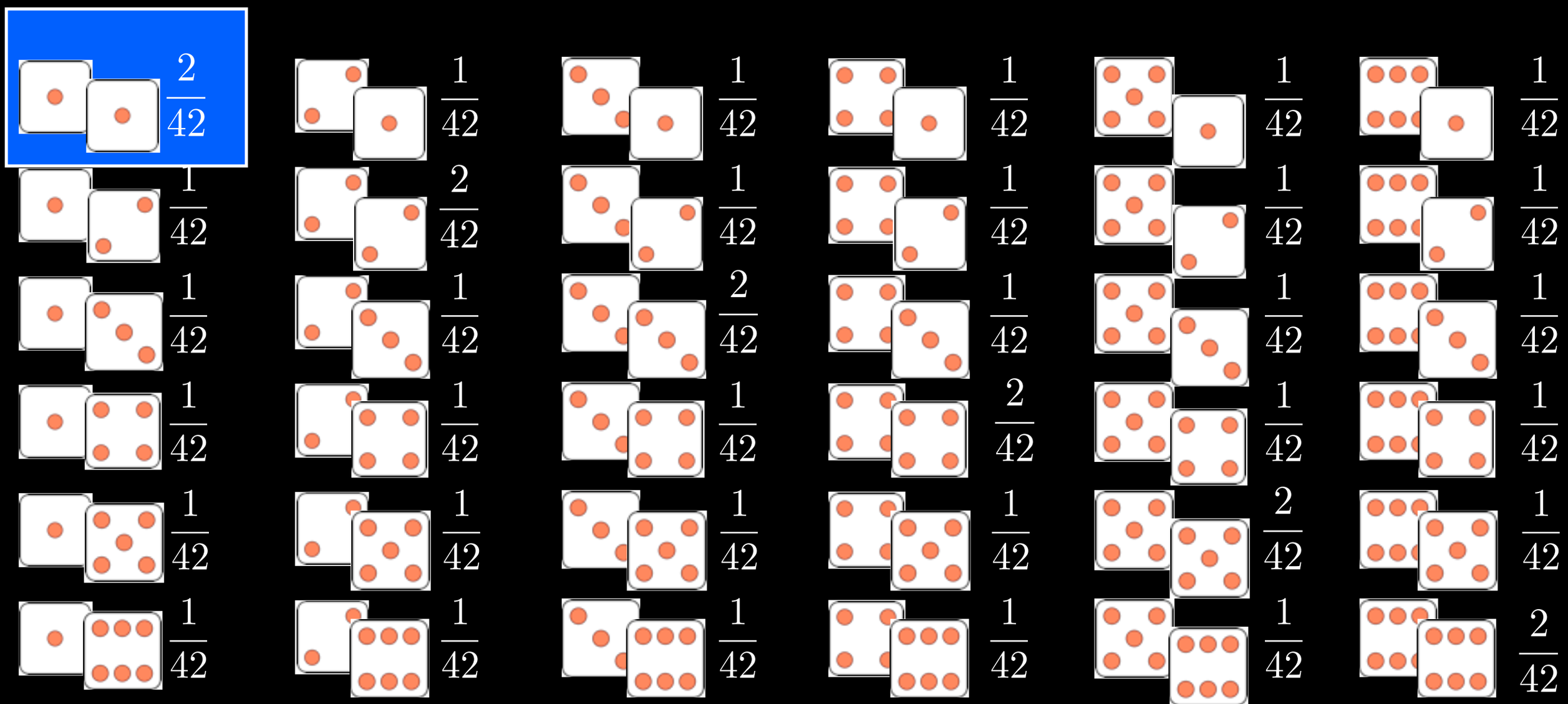
$$p(A = 1, B = 1) = p(A = 1, B = 1)$$

Probabilistic Primer



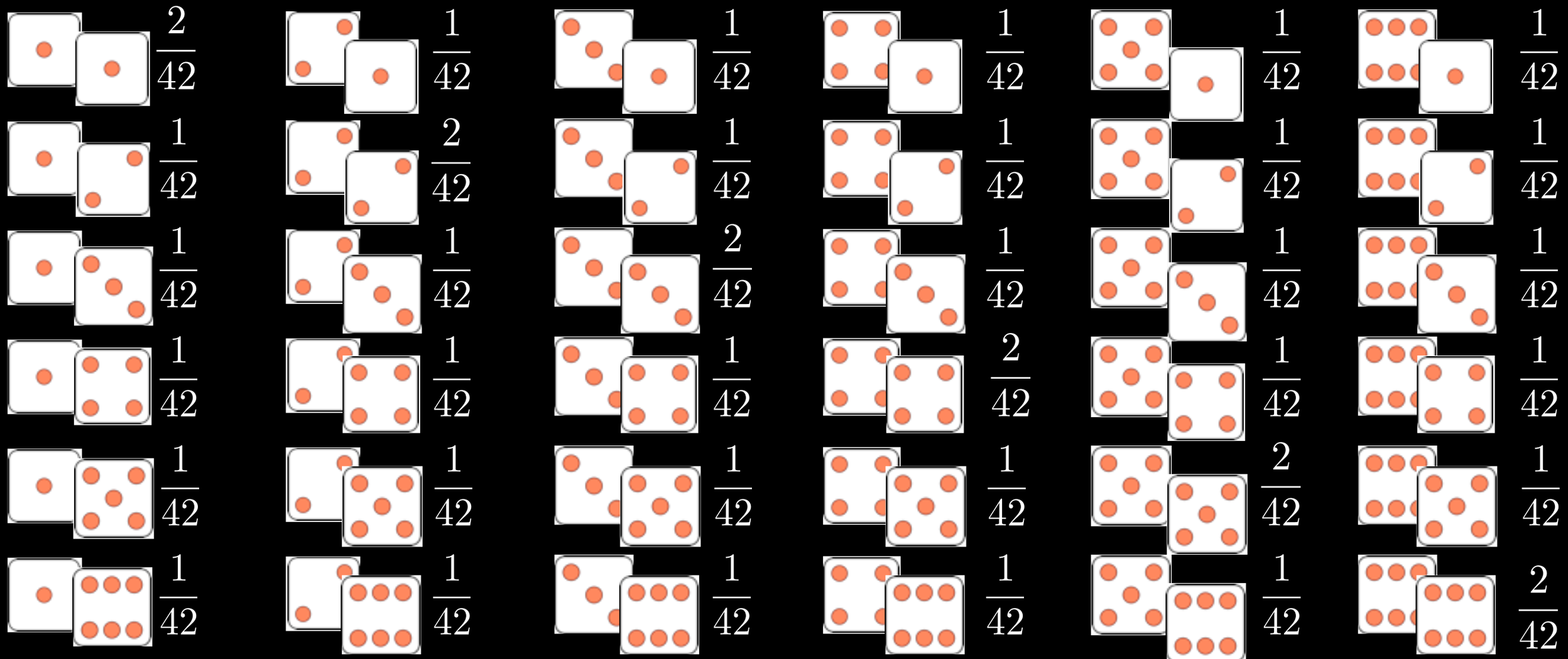
$$p(A=1,B=1) = \sum_{b \in B} p(A=1,B=b) \frac{p(A=1,B=1)}{\sum_{b \in B} p(A=1,B=b)}$$

Probabilistic Primer



$$p(A = 1, B = 1) = p(A = 1) \cdot p(B = 1|A = 1)$$

Probabilistic Primer



$$p(A, B) = p(A) \cdot p(B|A)$$

Probabilistic Primer

$$p(A, B) = p(A) \cdot p(B|A)$$

Probabilistic Primer

$$p(A, B) = p(A) \cdot p(B|A) = p(B) \cdot p(A|B)$$

Probabilistic Primer

$$p(A) \cdot p(B|A) = p(B) \cdot p(A|B)$$

Probabilistic Primer

$$p(B|A) = \frac{p(B) \cdot p(A|B)}{p(A)}$$

Probabilistic Primer

Bayes' Rule

$$p(B|A) = \frac{p(B) \cdot p(A|B)}{p(A)}$$

Probabilistic Primer

Bayes' Rule

posterior prior likelihood

$$p(B|A) = \frac{p(B) \cdot p(A|B)}{p(A)}$$

...But the probability that an event has happened is the same as the probability I have to guess right if I guess it has happened. Wherefore the following proposition is evident: If there be two subsequent events, the probability of the 2d b/N and the probability both together P/N, and it being 1st discovered that the 2d event has also happened, the probability I am right is P/b.

Thomas Bayes

Bayes' Rule

$p(English)$

Bayes' Rule

$p(English)$

configuration

Bayes' Rule

$p(English)$

configuration

$p(image|English)$

Bayes' Rule

$p(English)$

configuration

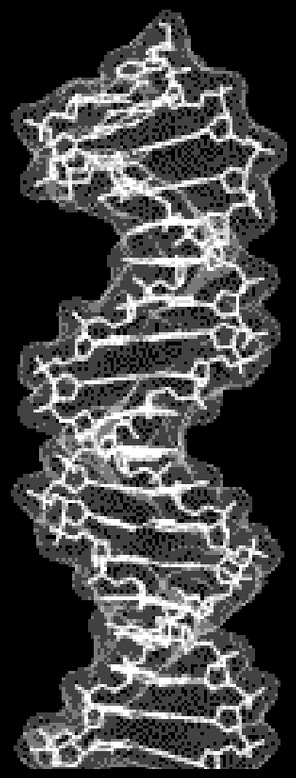
$p(image|English)$

Bayes' Rule

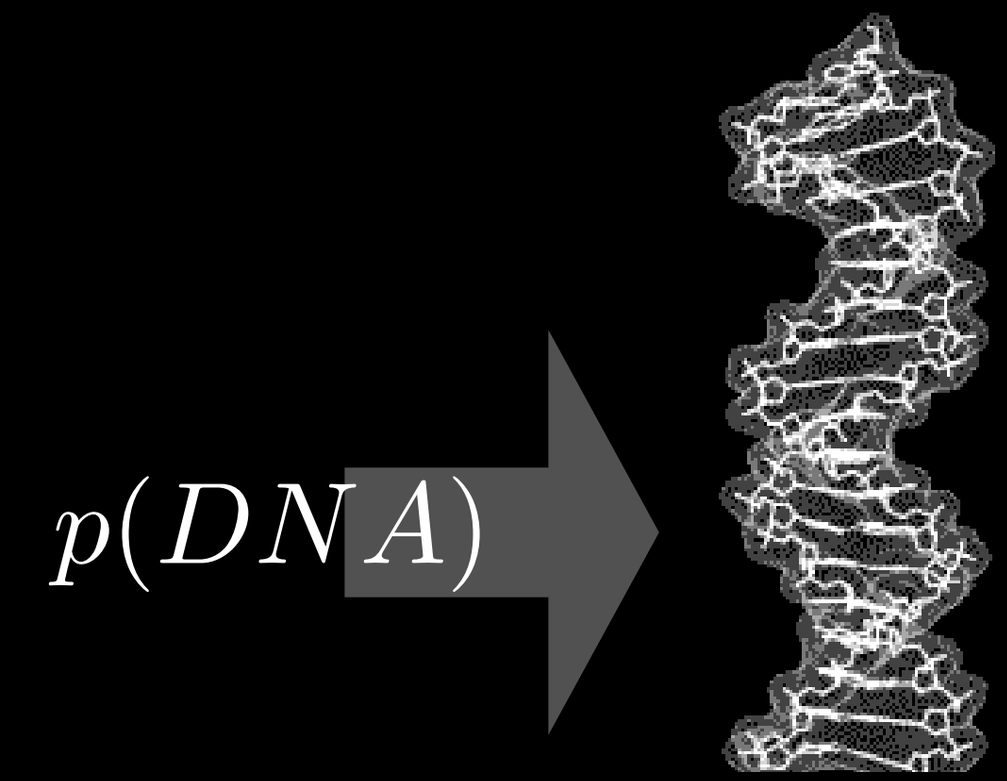
$p(DNA)$

Bayes' Rule

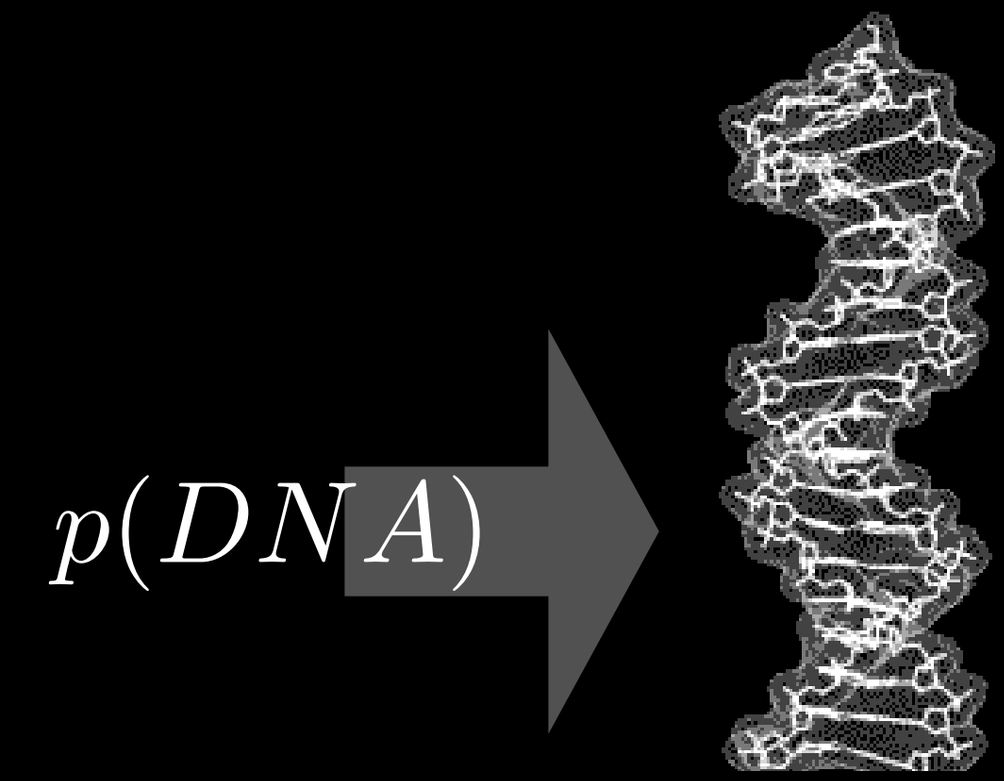
$p(DNA)$

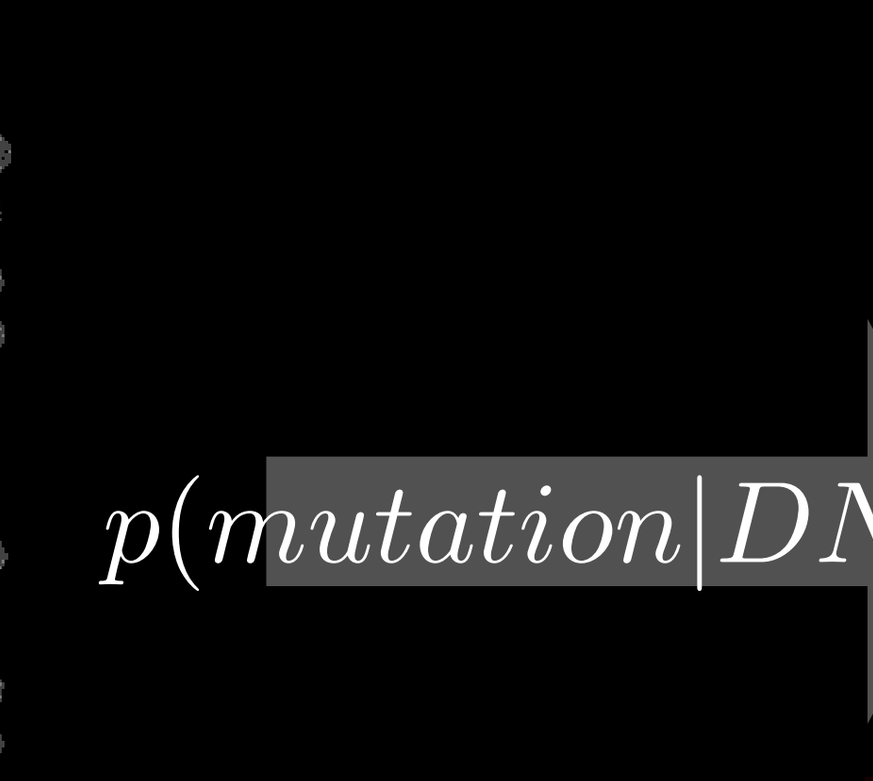


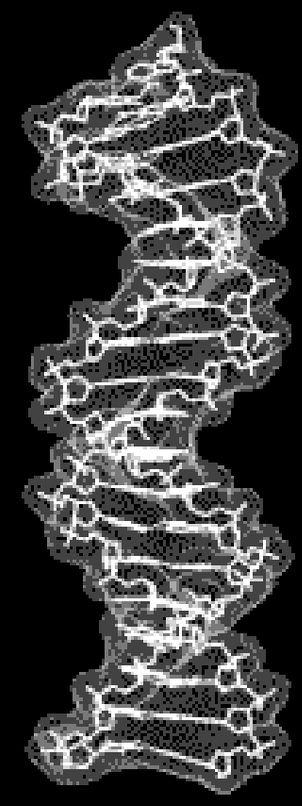
Bayes' Rule

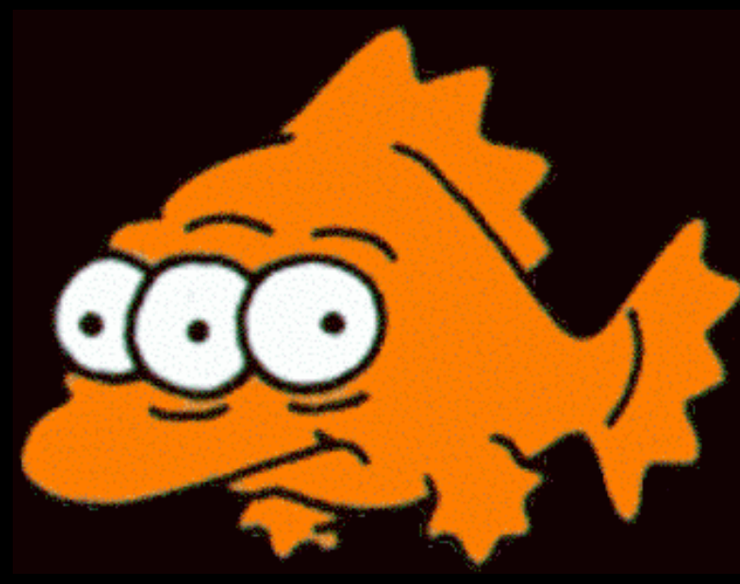


Bayes' Rule









Bayes' Rule

$p(English)$

Bayes' Rule

$p(English)$

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

Bayes' Rule

$$p(English)$$

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

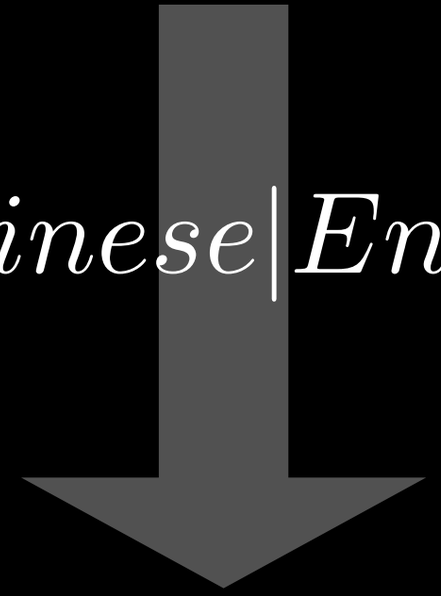
$$p(Chinese|English)$$

Bayes' Rule

$$p(English)$$

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

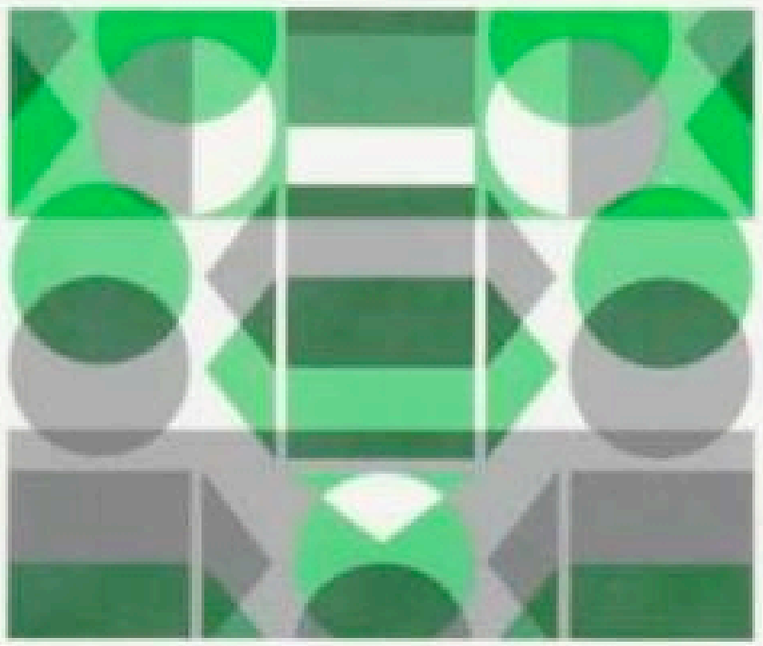
$$p(Chinese|English)$$



虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

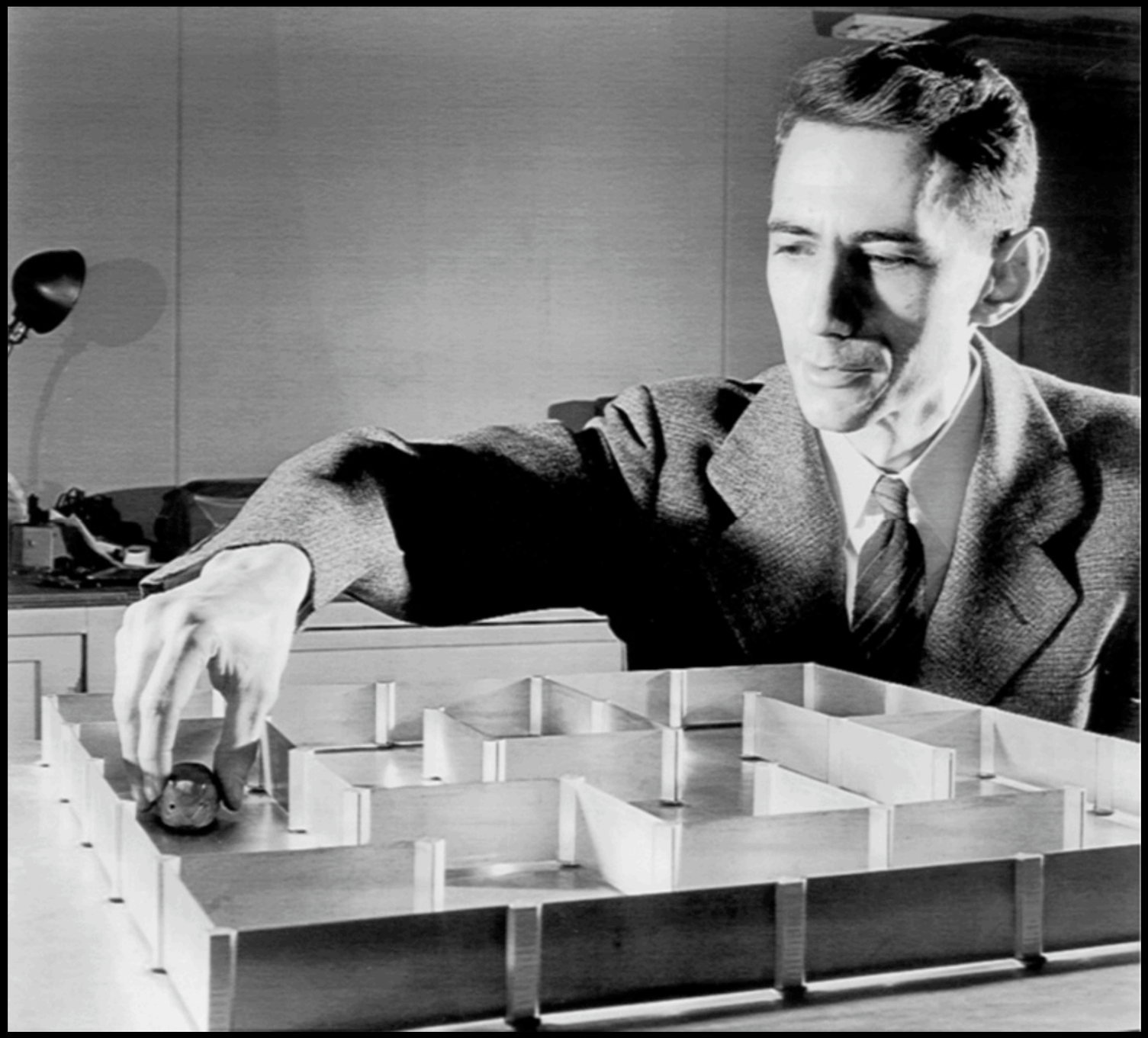
When I look at an article in Russian, I say: "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

Warren Weaver (1949)



THE MATHEMATICAL THEORY OF COMMUNICATION

by Claude E. Shannon and Warren Weaver



Claude Shannon

Bayes' Rule

$$p(English|Chinese) = \frac{p(English) \times p(Chinese|English)}{p(Chinese)}$$

prior

likelihood

normalization term (ensures we're working with valid probabilities).

Noisy Channel

$$p(English|Chinese) = \frac{p(English) \times p(Chinese|English)}{p(Chinese)}$$

signal model

channel model

normalization term (ensures we're working with valid probabilities).

Machine Translation

$$p(English|Chinese) =$$

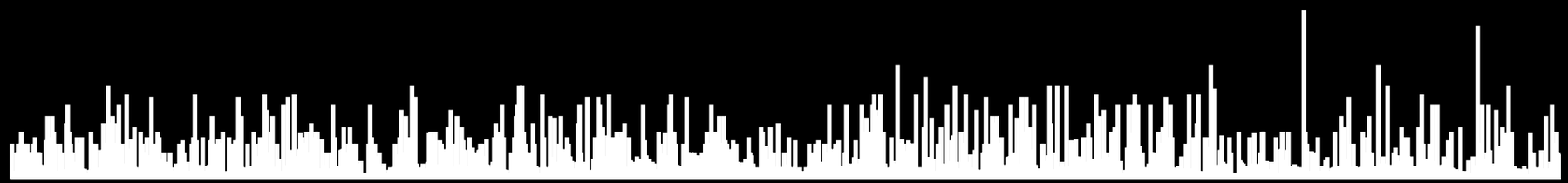
$$\frac{p(English) \times p(Chinese|English)}{p(Chinese)}$$

language model

translation model

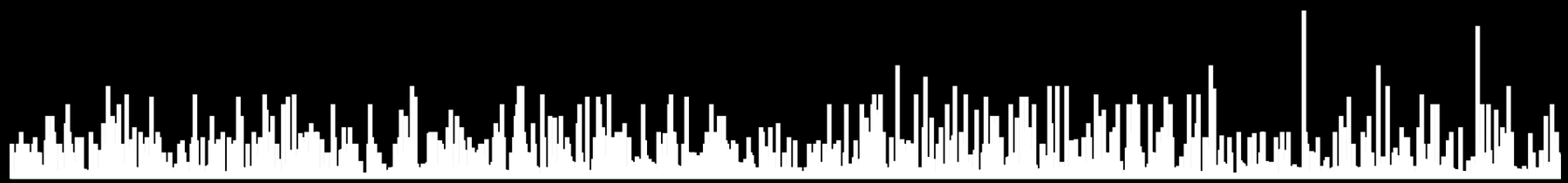
normalization term (ensures we're working with valid probabilities).

$p(Chinese|English)$

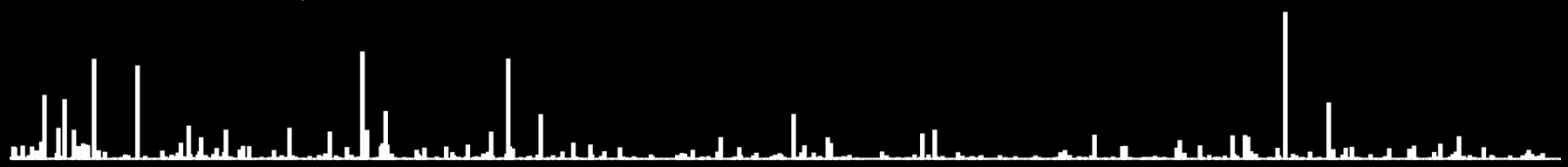


$English$

$p(Chinese|English)$

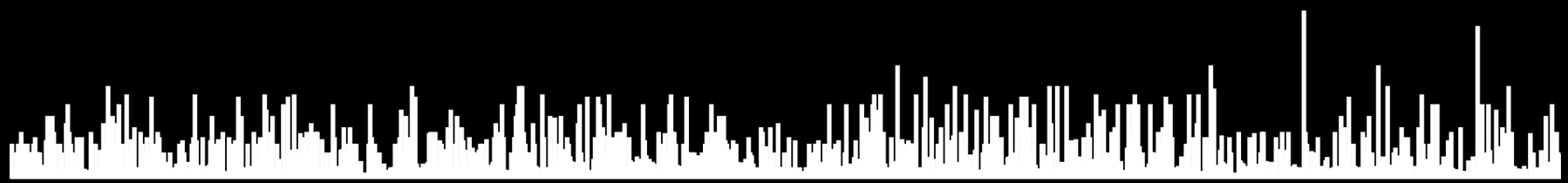


$\times\ p(English)$

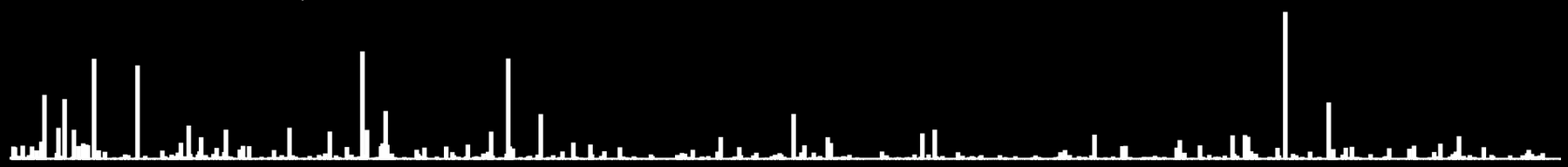


English

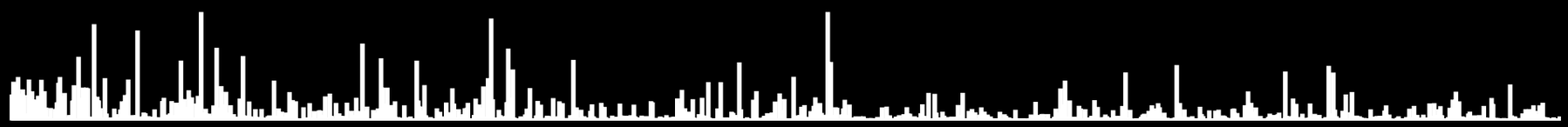
$p(Chinese|English)$



$\times\ p(English)$



$\sim p(English|Chinese)$



$English$

Machine Translation

$$p(English|Chinese) =$$

$$\frac{p(English) \times p(Chinese|English)}{p(Chinese)}$$

language model

translation model

normalization term (ensures we're working with valid probabilities).

Machine Translation

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

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

Machine Translation

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

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

What is the probability of an English sentence?

Machine Translation

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

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

What is the probability of an English sentence?

What is the probability of a Chinese sentence, given a particular English sentence?

Language Models

We can think of our probabilistic model as a story that explains every single word in the sentence pair.

Language Models

$$p(However|START)$$

Language Models

However

$$p(However|START)$$

Language Models

However ,

$$p(, | However)$$

Language Models

However , the

$$p(the|,)$$

Language Models

However , the sky

$$p(sky|the)$$

Language Models

However , the sky remained

$$p(remained|sky)$$

Language Models

However , the sky remained clear

$$p(clear|remained)$$

Language Models

$$length(English)$$

$$p(English) = \prod_{i=1} p(word_i|word_{i-1})$$

Language Models

$$p(English) = \prod_{i=1}^{length(English)} p(word_i|word_{i-1})$$

Note: the prior probability that $word_0$=START is 1.

Language Models

$$p(English) = \prod_{i=1}^{length(English)} p(word_i|word_{i-1})$$

Note: the prior probability that $word_0$=START is 1.

This model explains every word in the English sentence.

Language Models

$$p(English) = \prod_{i=1}^{length(English)} p(word_i | word_{i-1})$$

Note: the prior probability that $word_0$=START is 1.

This model explains every word in the English sentence.

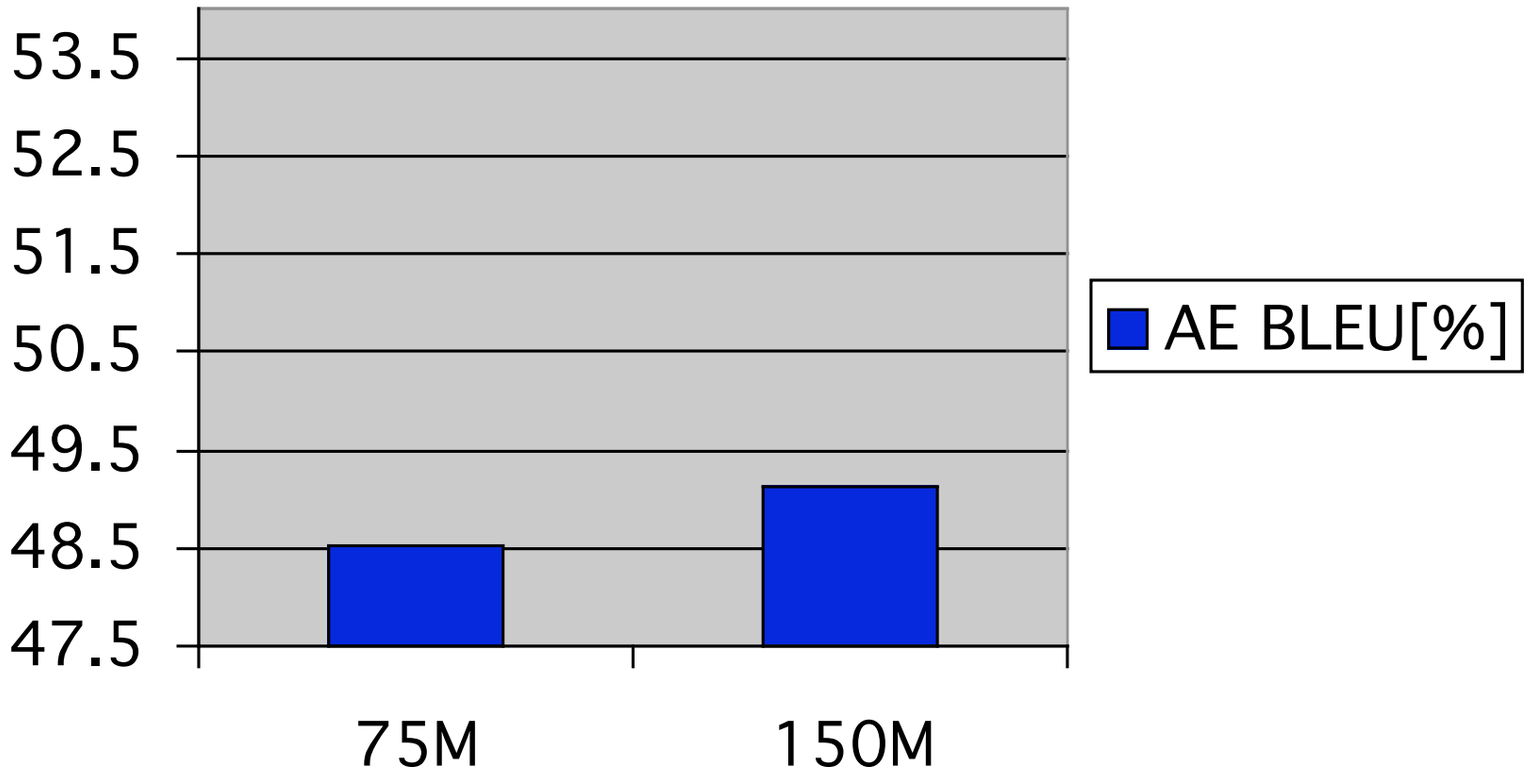
But it makes very strong conditional independence assumptions!

Language Models

- The language model does not depend in any way on parallel data.
- How much English data should we train it on?

Language Models

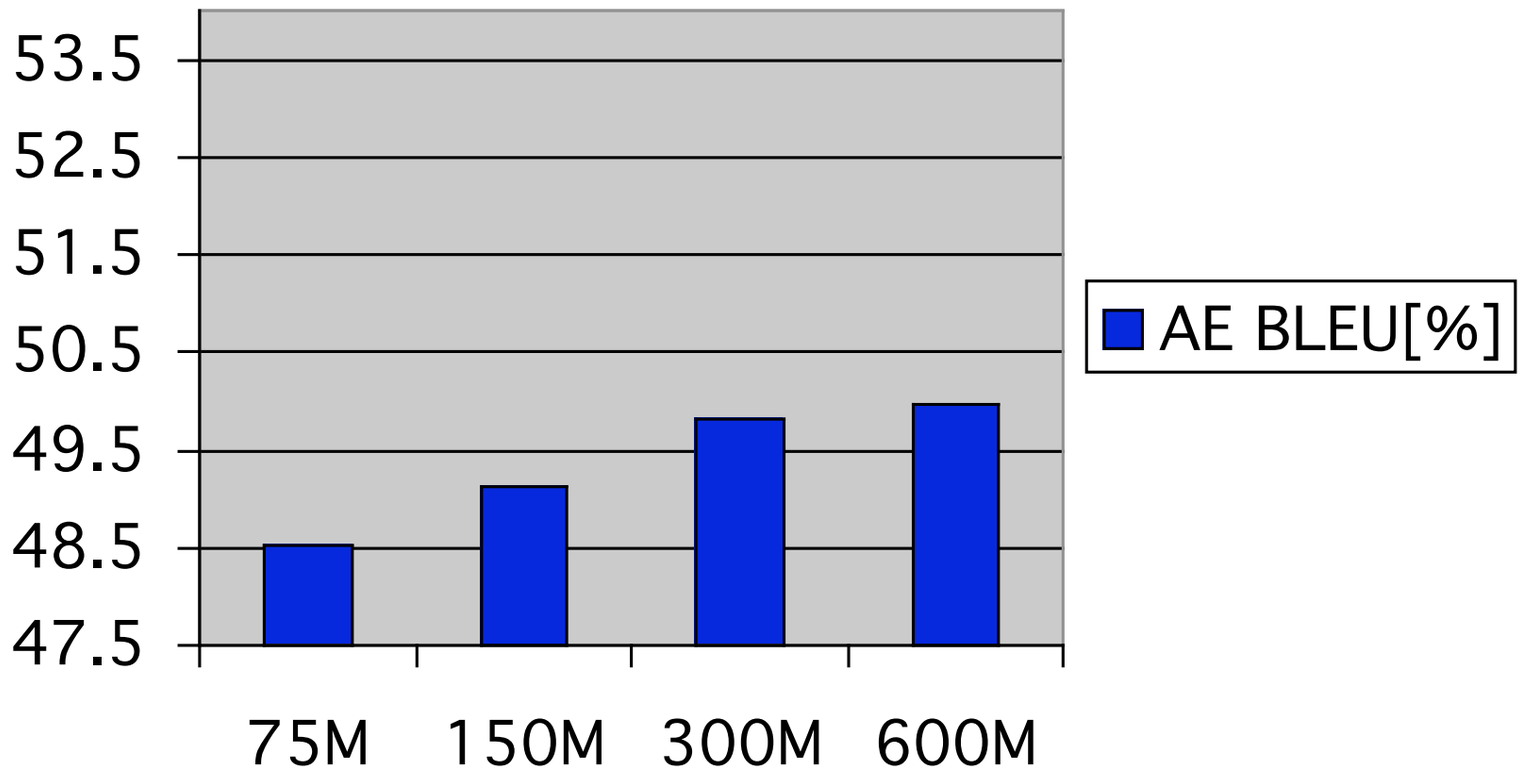
Impact on size of language model training data (in words) on quality of Arabic-English statistical machine translation system (NIST test data)



Google

Language Models

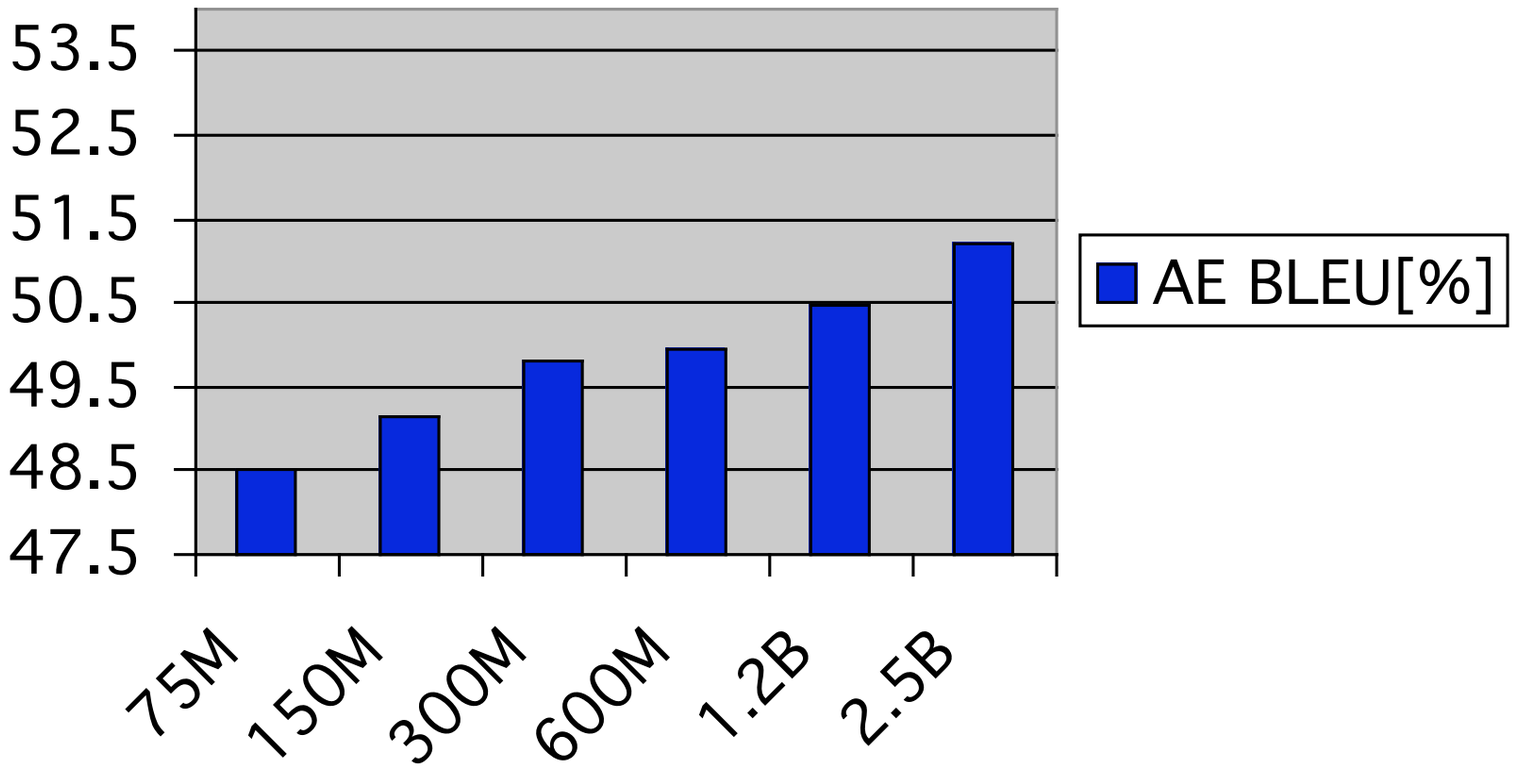
Impact on size of language model training data (in words) on quality of Arabic-English statistical machine translation system



Google

Language Models

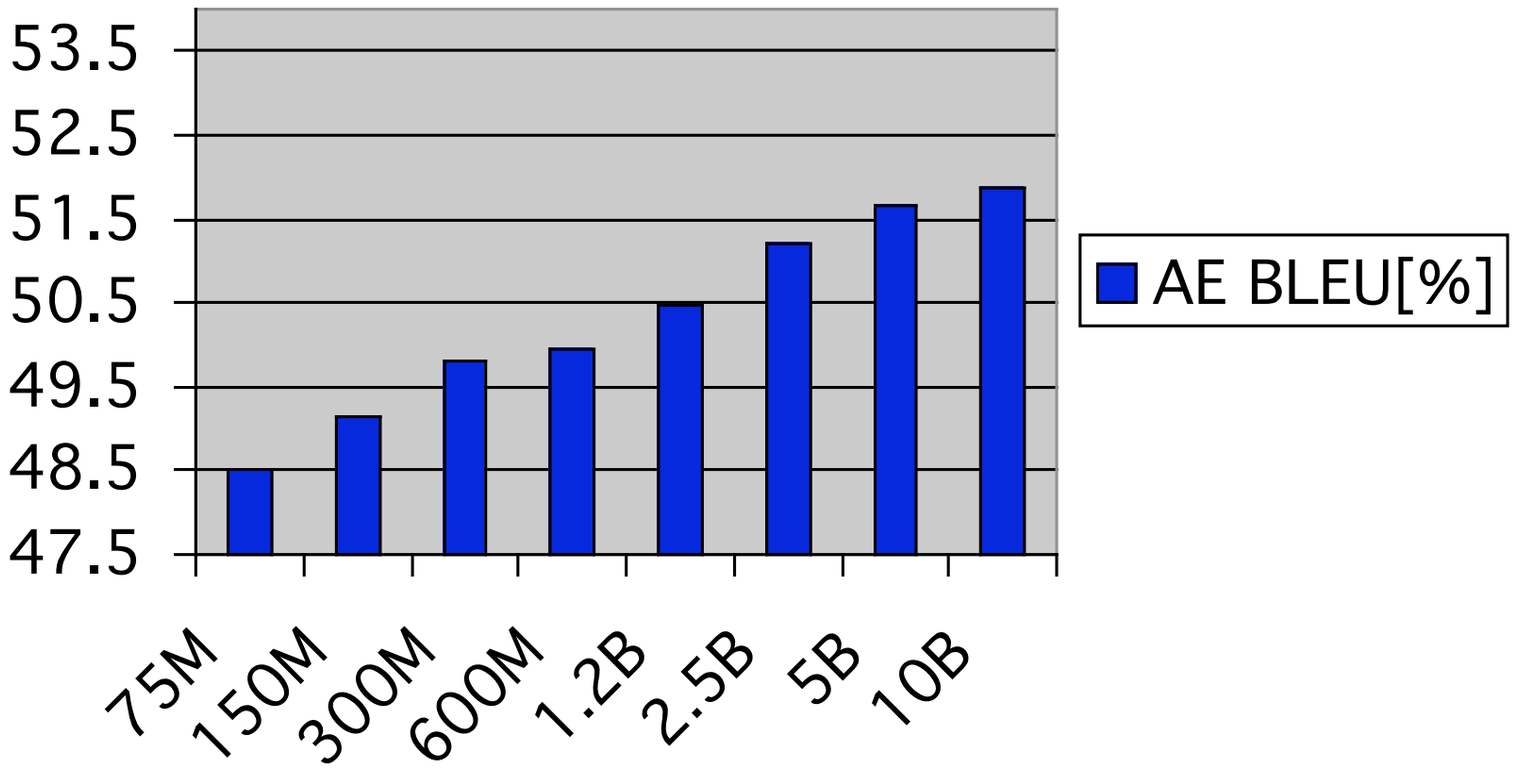
Impact on size of language model training data (in words) on quality of Arabic-English statistical machine translation system



Google

Language Models

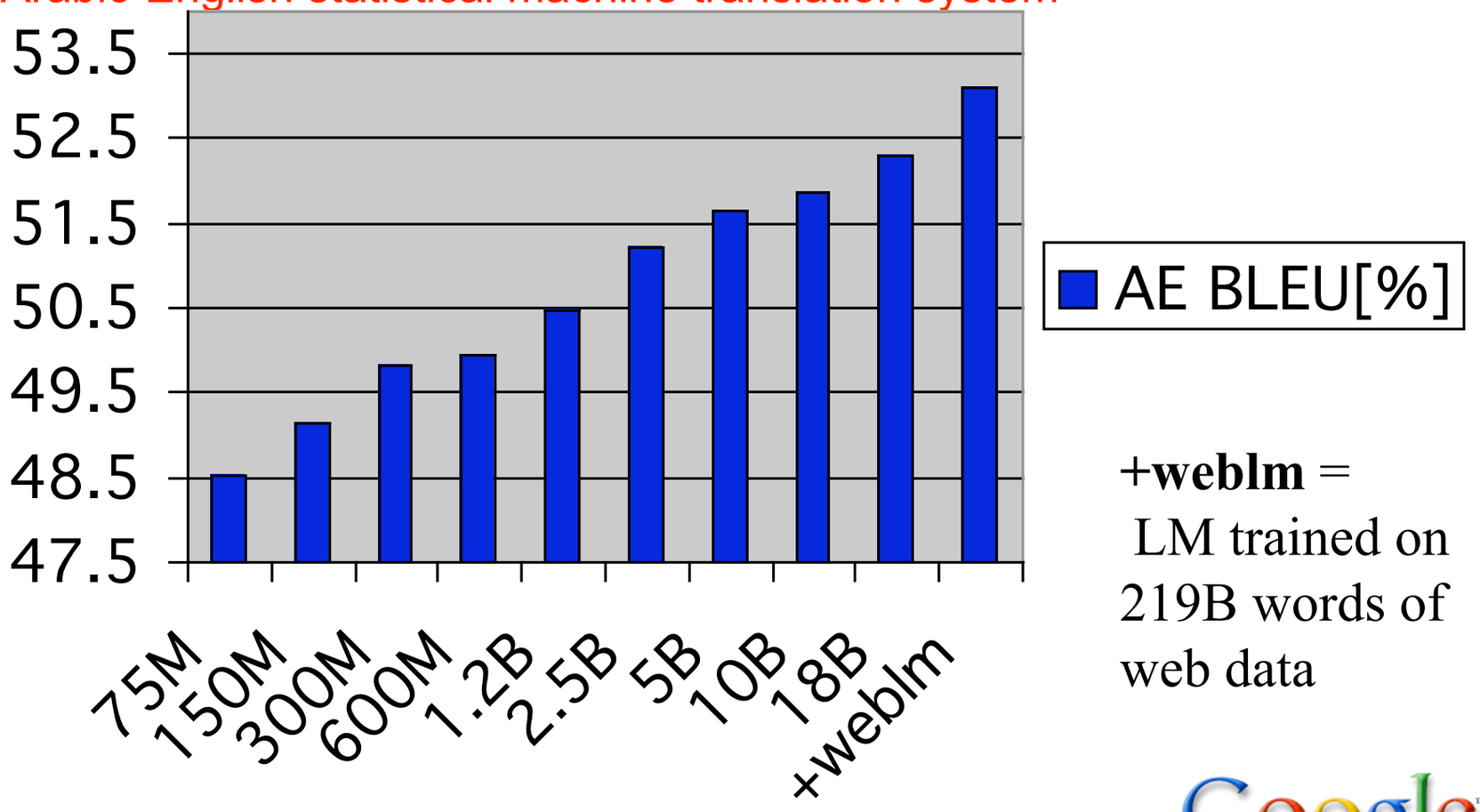
Impact on size of language model training data (in words) on quality of Arabic-English statistical machine translation system



Google

Language Models

Impact on size of language model training data (in words) on quality of Arabic-English statistical machine translation system



+weblm = LM trained on 219B words of web data

Google

Language Models

- There's no data like more data.
- Language models serve a similar function in speech recognition, optical character recognition, and other probabilistic models of text data.
- You'll learn a lot more about them from Nicola Bertoldi on Tuesday and Wednesday

Translation Models

What is a good story about how a Chinese sentence came into being, given that we already have an English sentence?

Translation Models

What is a good story about how a Chinese sentence came into being, given that we already have an English sentence?

Note: in this example I'll show you an English sentence, conditioned on a Chinese sentence. Note that we can apply the same technique in either direction.

Translation Models

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

$$p(English|Chinese)$$

Translation Models

Although north wind howls , but sky still very clear .

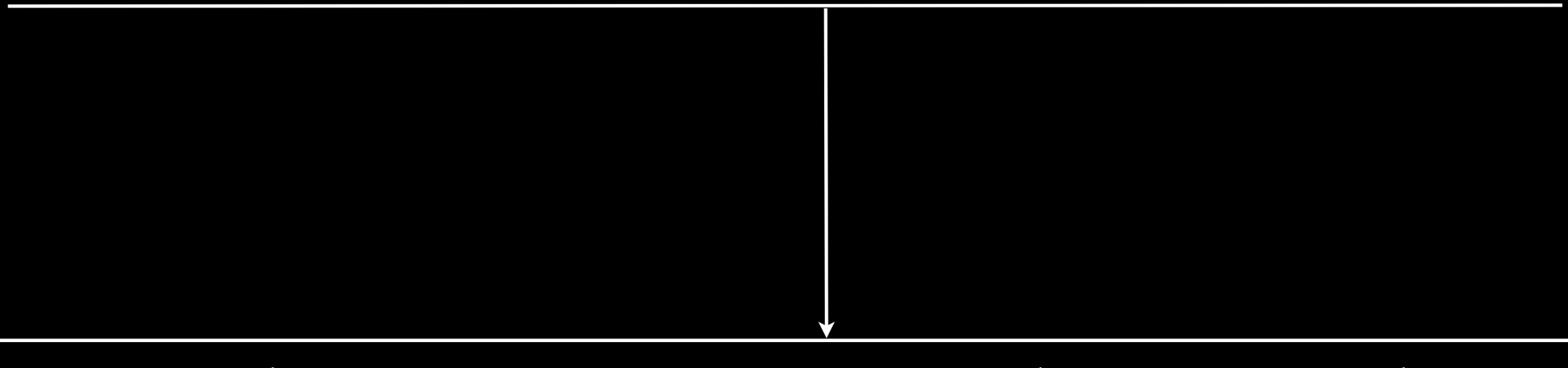
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

$$p(English|Chinese)$$

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 。



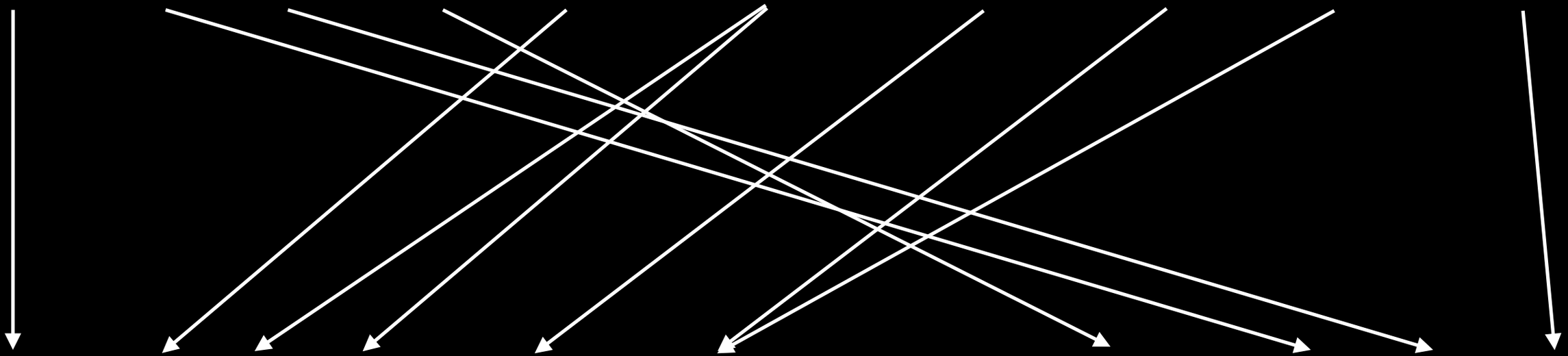
However , the sky remained clear under the strong north wind .

$$p(English|Chinese)$$

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 。

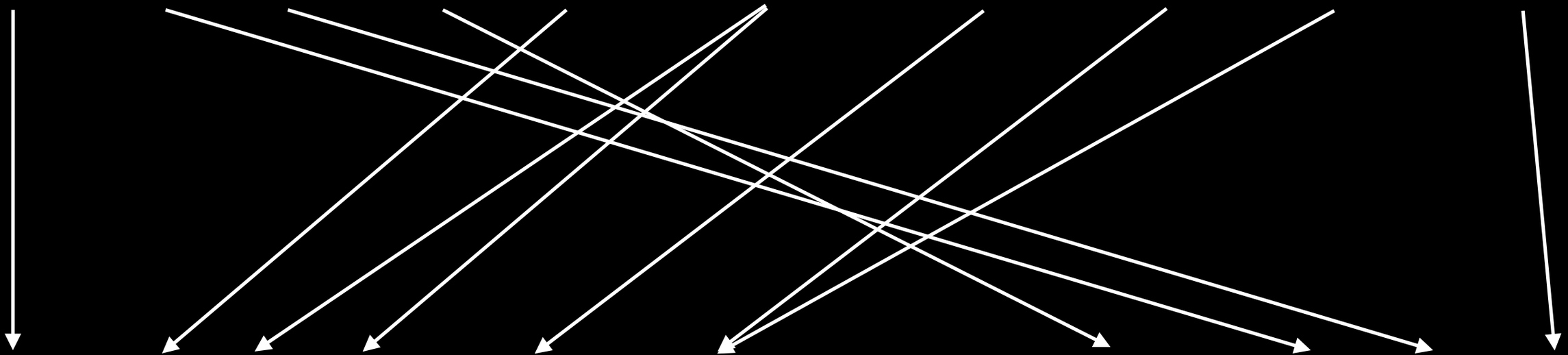


However , the sky remained clear under the strong north wind .

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 4

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 4

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

↓

虽然

IBM Model 4

Although north wind howls , but sky still very clear .

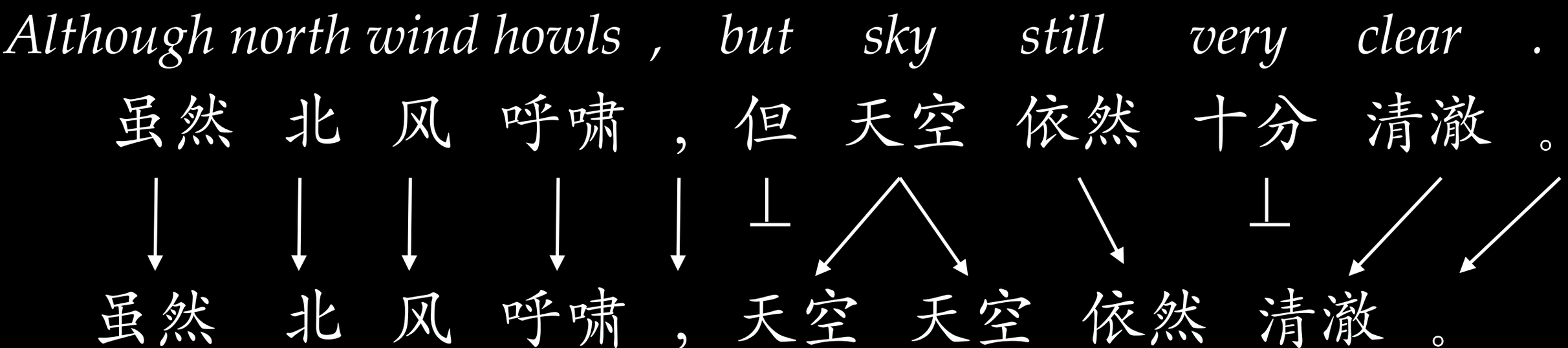
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

↓

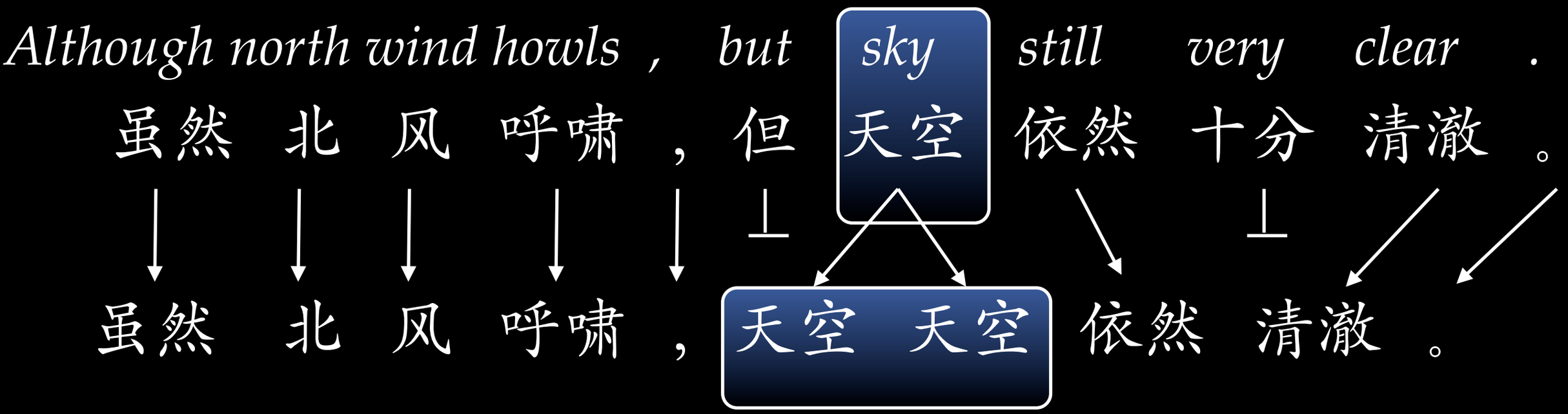
虽然

$$p_f(1|\text{虽然})$$

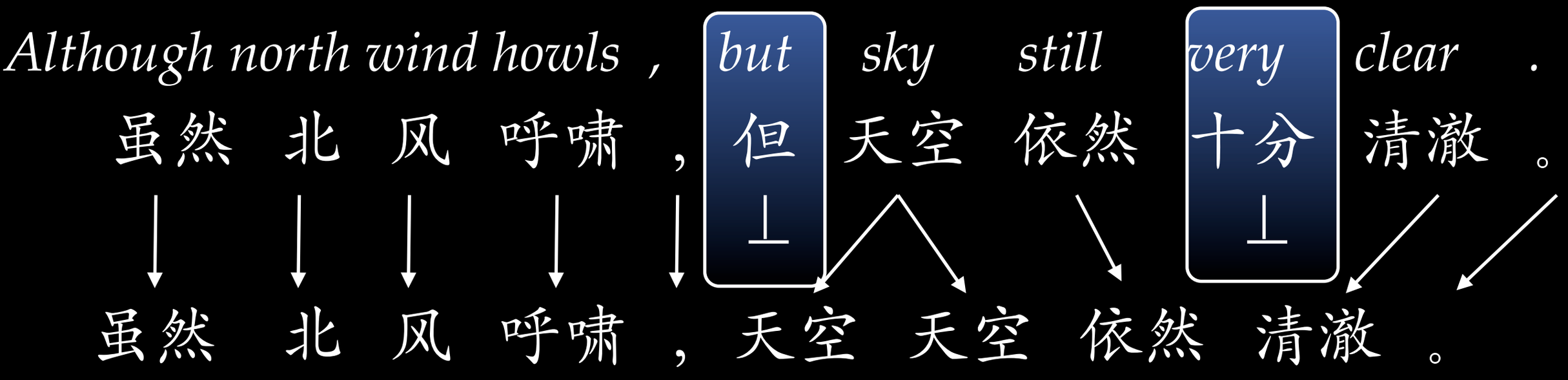
IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。
⊥
⊥
虽然 北 风 呼啸 ，天空 天空 依然 清澈 。

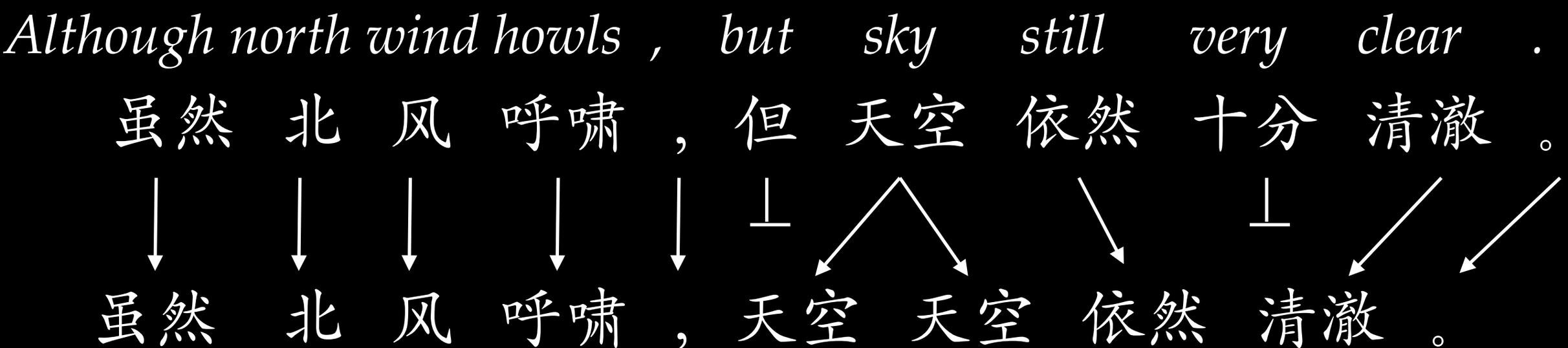
IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。
⊥
⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。

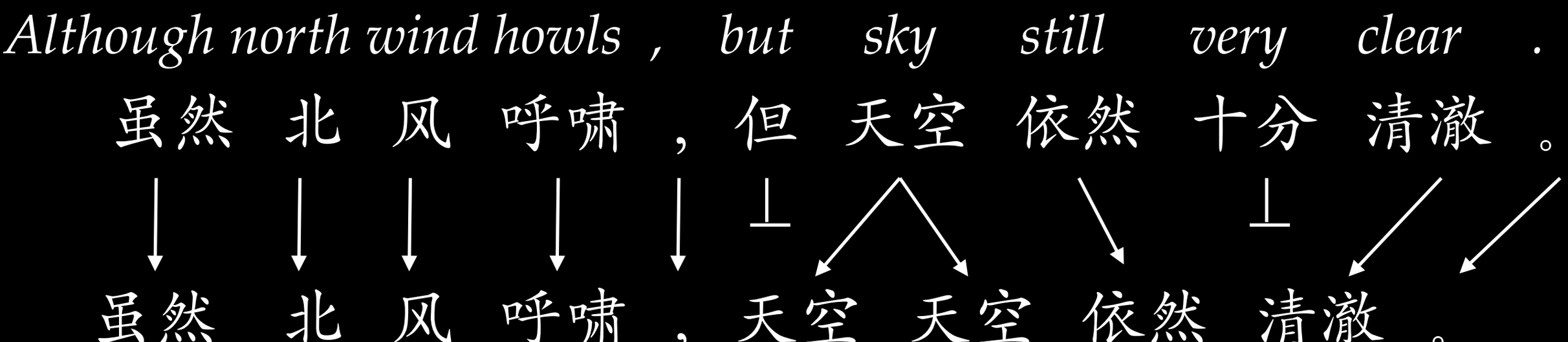
IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。
⊥
⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。

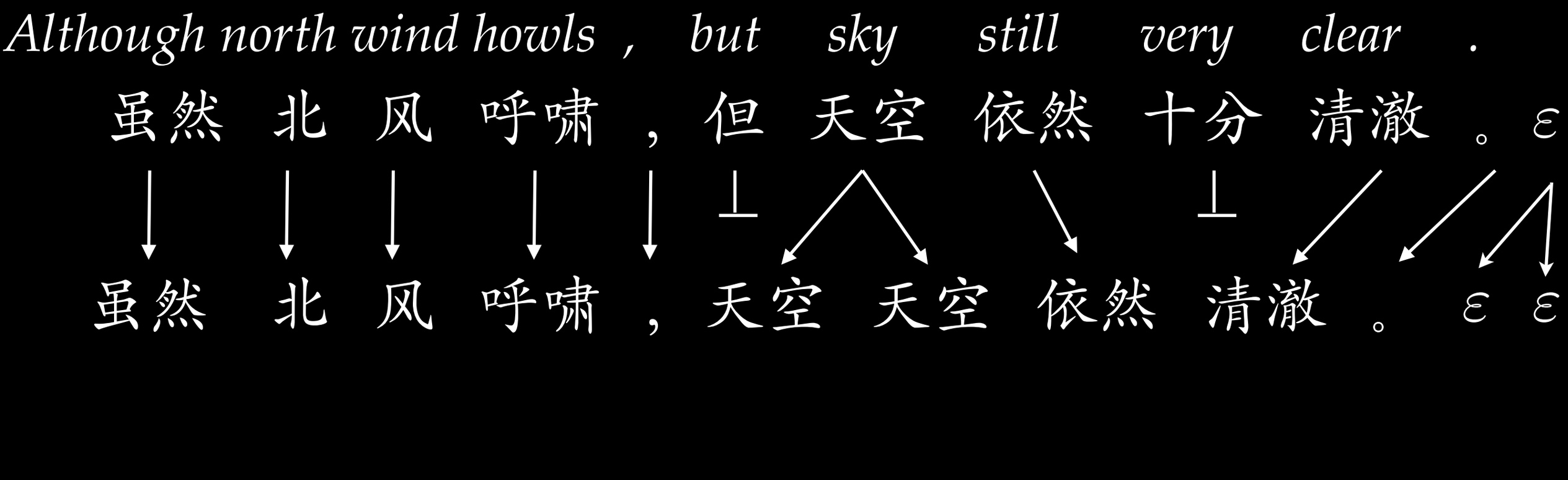
IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。
⊥
⊥
虽然 北 风 呼啸 ，天空 天空 依然 清澈 。

IBM Model 4


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
⊥
⊥
虽然 北 风 呼啸 ，天空 天空 依然 清澈 。

IBM Model 4


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
⊥ ⊥
虽然 北 风 呼啸 ，天空 天空 依然 清澈 。 ε ε

IBM Model 4

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

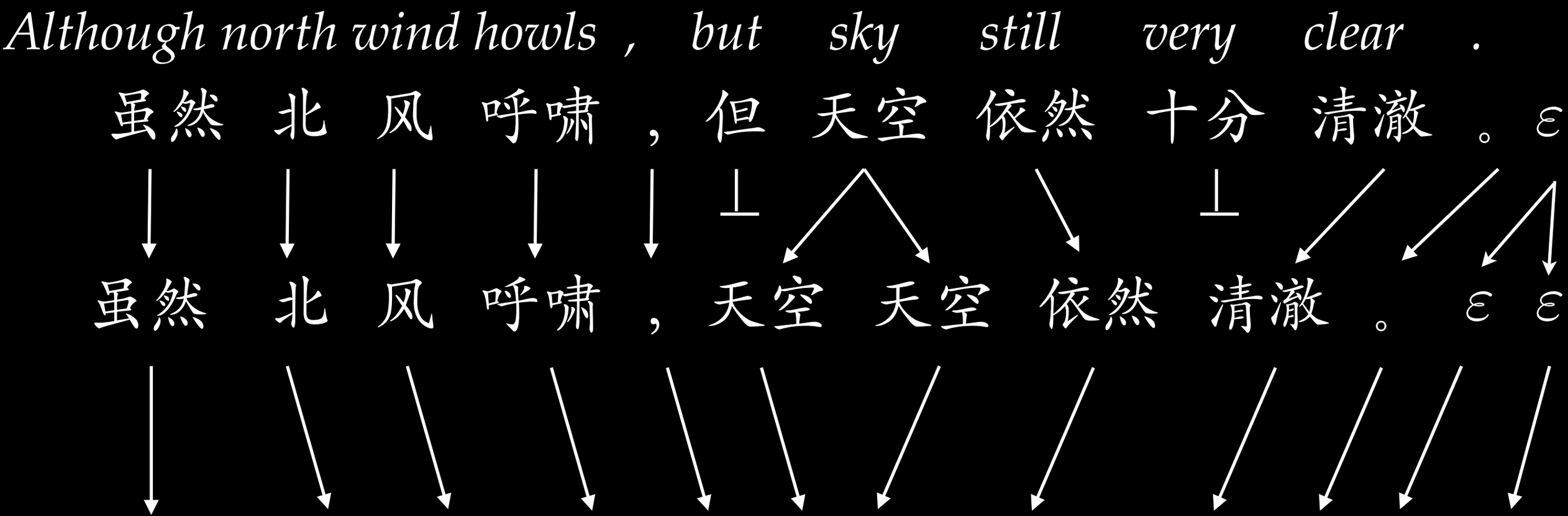
虽然 北 风 呼啸 ， 天空 天空 依然 清澈 。 ε ε

However

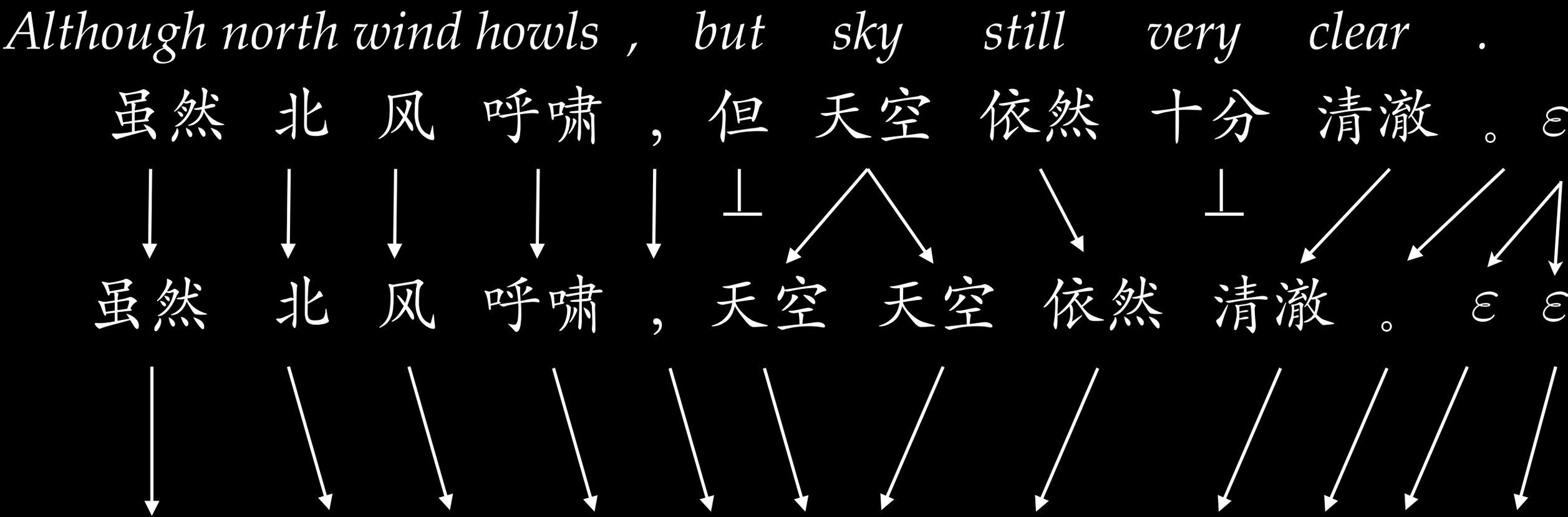
IBM Model 4

$$p_t(However|\text{虽然})$$

IBM Model 4


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
⊥ ⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the

IBM Model 4


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
⊥ ⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the

IBM Model 4

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
⊥ ⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the

IBM Model 4

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
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the
$p_d(0|However)$

IBM Model 4

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

虽然 北 风 呼啸 ，天空 天空 依然 清澈 。 ε ε

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

IBM Model 4

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

虽然 北 风 呼啸 ，天空 天空 依然 清澈 。 ε ε

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

$p_d(8|north)$

IBM Model 4

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
⊥
⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the

IBM Model 4

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
⊥
⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the
However , the sky remained clear under the strong north wind .

IBM Model 4

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

虽然 北 风 呼啸 ，天空 天空 依然 清澈 。 ε ε

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

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

$$p(English, alignment|Chinese) = \prod_{p_f} \prod_{p_t} \prod_{p_d}$$

IBM Model 4

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 .

$$p(English, alignment|Chinese) = \prod_{p_f} \prod_{p_t} \prod_{p_d}$$

IBM Model 4

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 .

$$p(English|Chinese) = \sum_{alignments} \prod_{p_f} \prod_{p_t} \prod_{p_d}$$

The IBM Models

The IBM Models

- Fertility probabilities.

The IBM Models

- Fertility probabilities.
- Word translation probabilities.

The IBM Models

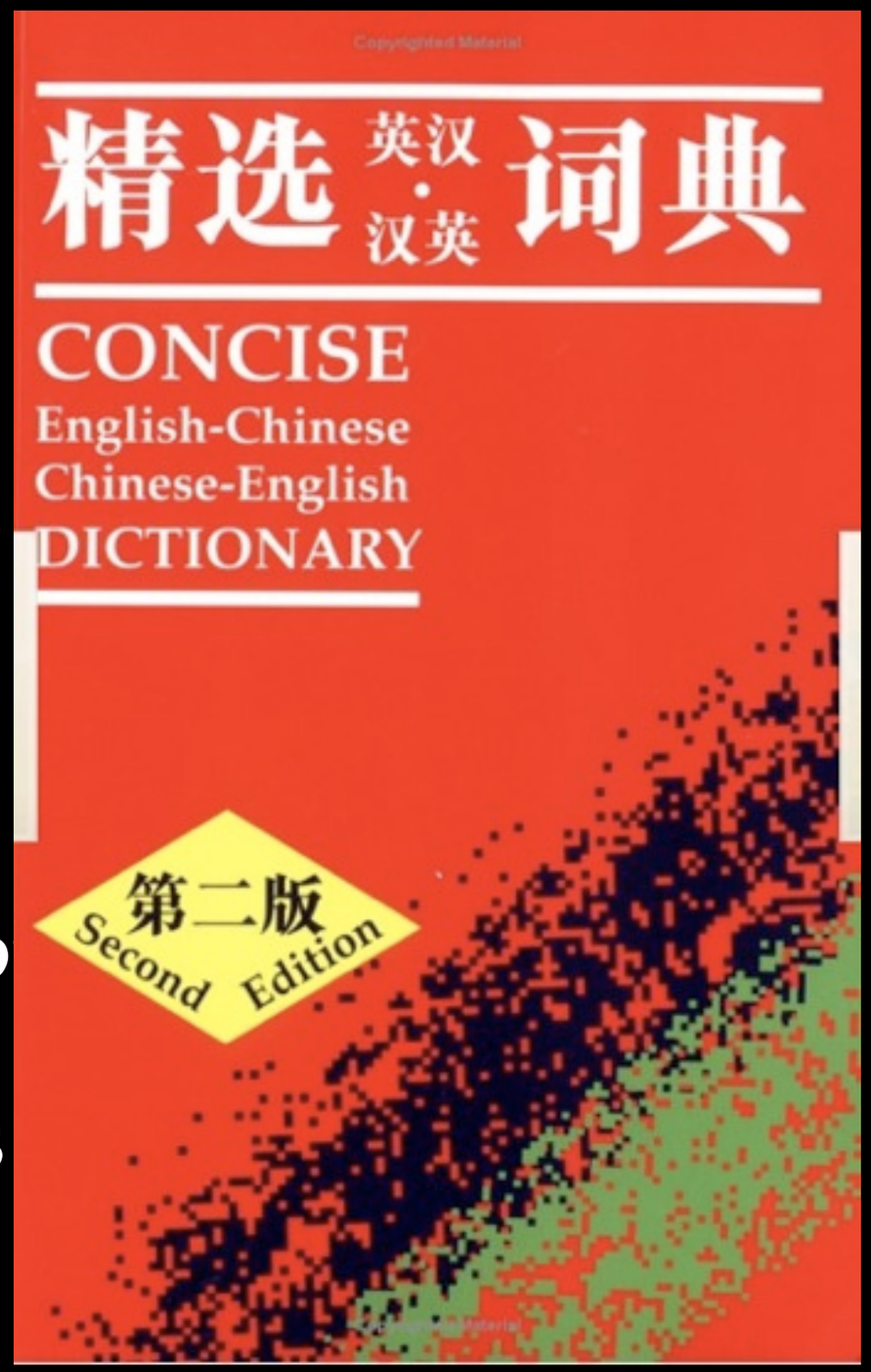
- Fertility probabilities.
- Word translation probabilities.
- Distortion probabilities.

The IBM Models

- Fertility probabilities.
- Word translation probabilities.
- Distortion probabilities.
- Some problems:
  - Weak reordering model -- output is not fluent.
  - Many decisions -- many things can go wrong.

The IBM Models

- Fertility probabilities.
- Word translation probabilities.
- Distortion probabilities.
- Some problems:
  - Weak reordering model -- outp
  - Many decisions -- many things



The IBM Models

- Fertility probabilities.
- Word translation probabilities.
- Distortion probabilities.
- Some problems:
  - Weak reordering model -- output is not fluent.
  - Many decisions -- many things can go wrong.

IBM Model 4 Again

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

虽然 北 风 呼啸 ，天空 天空 依然 清澈 。 ε ε

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

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

IBM Model 4 Again

Phrase-based 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 。

Phrase-based 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 | 。

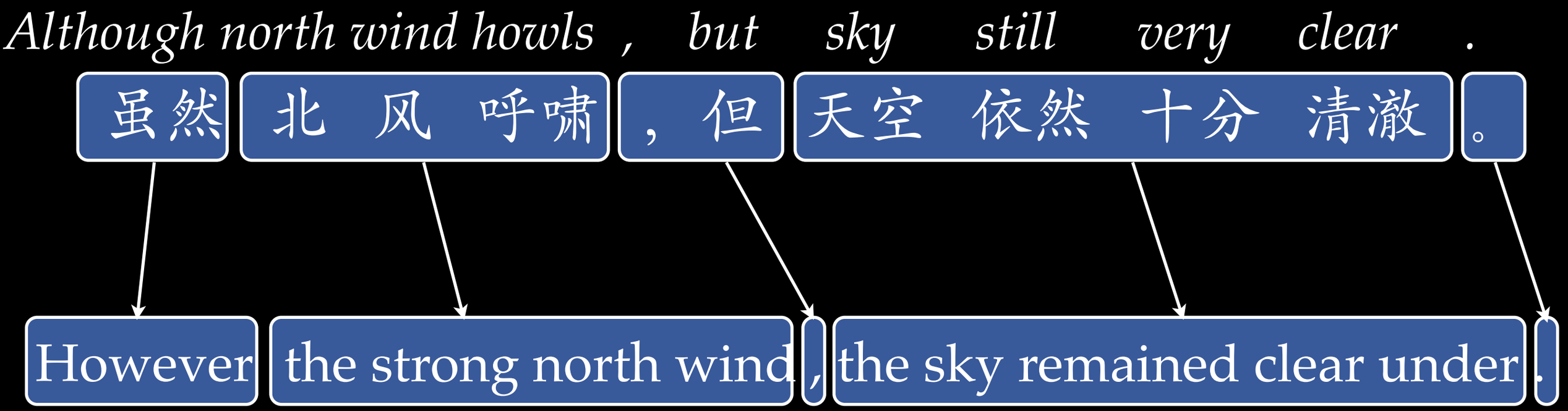
Phrase-based Models

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

Phrase-based Models



Phrase-based 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 strong north wind , the sky remained clear under .

However

Phrase-based 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 strong north wind , the sky remained clear under .

However ,

Phrase-based Models

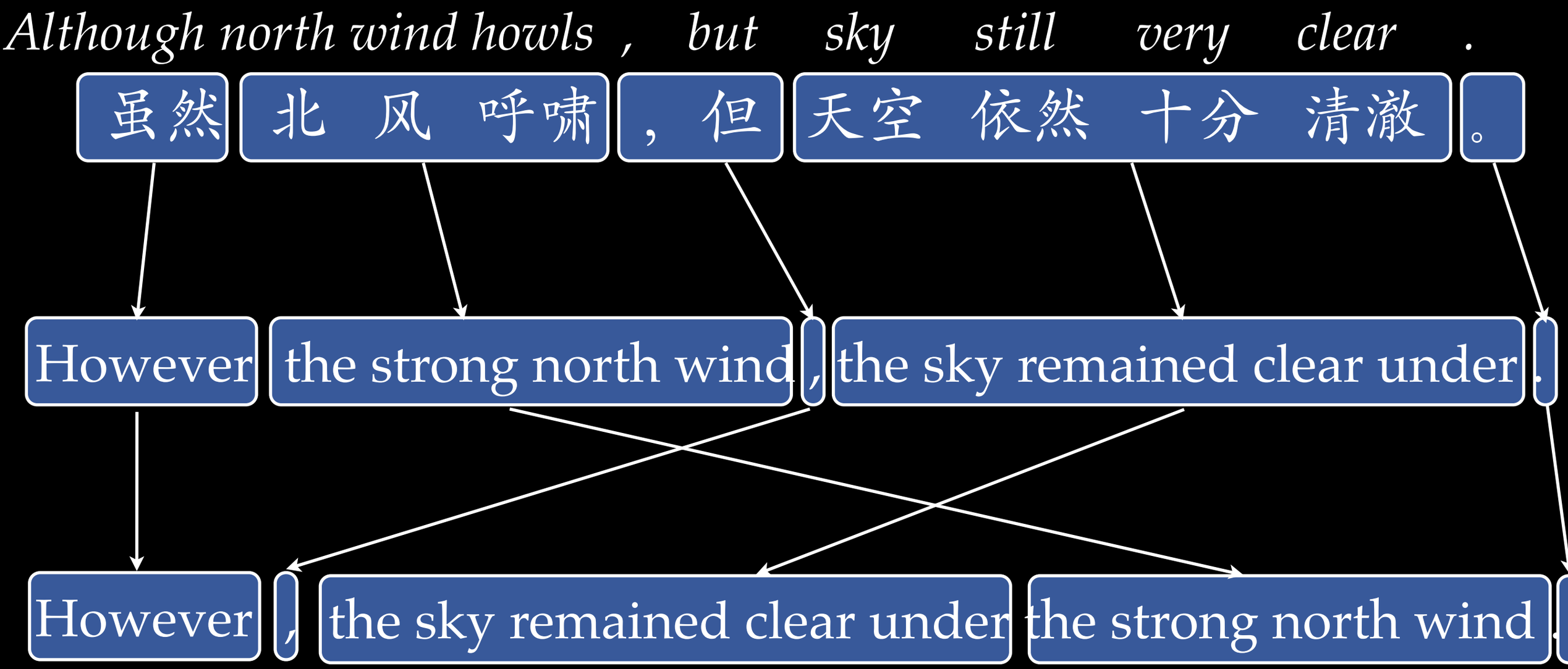
Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

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

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

Phrase-based Models



$$p(English, alignment|Chinese) =$$
$$p(segmentation) \cdot p(translations) \cdot p(reorderings)$$

Phrase-based Models

Phrase-based Models

- Segmentation probabilities.

Phrase-based Models

- Segmentation probabilities.
- Phrase translation probabilities.

Phrase-based Models

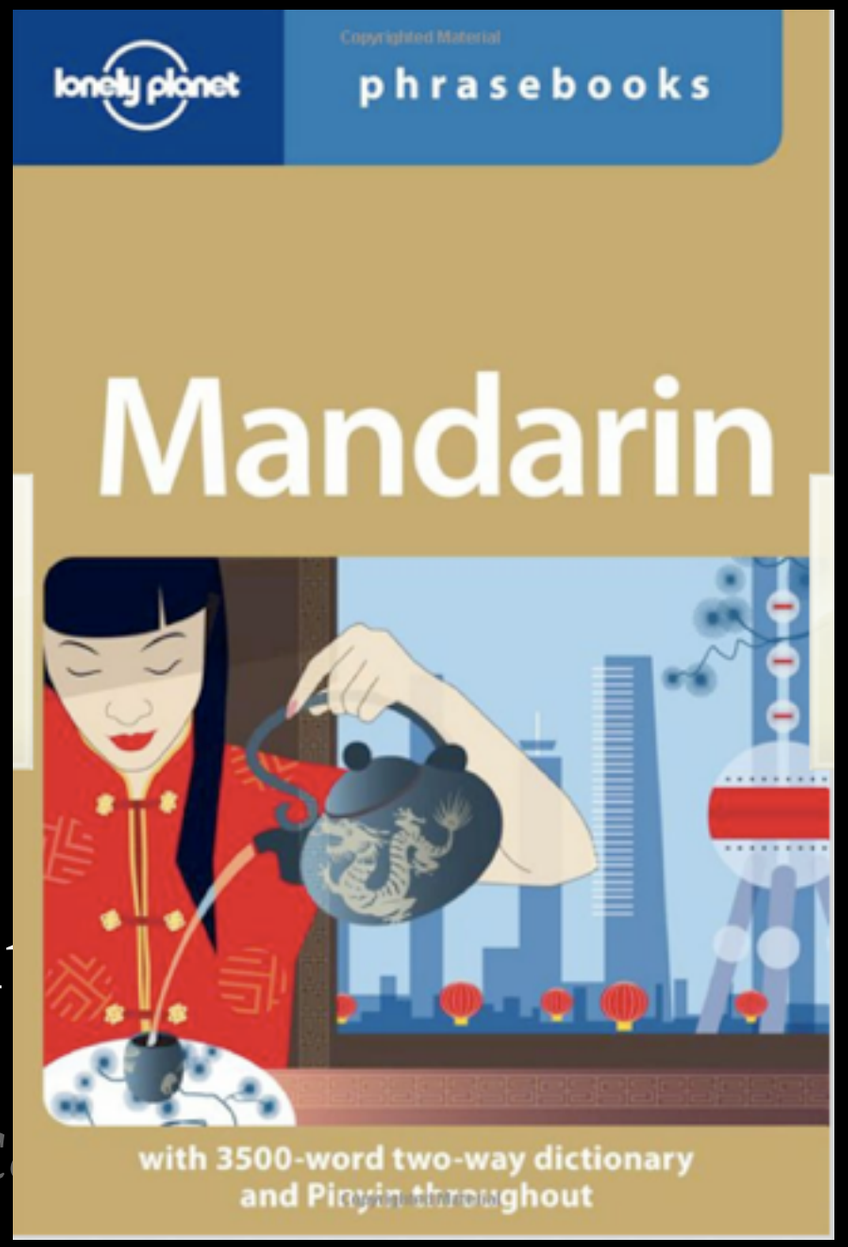
- Segmentation probabilities.
- Phrase translation probabilities.
- Distortion probabilities.

Phrase-based Models

- Segmentation probabilities.
- Phrase translation probabilities.
- Distortion probabilities.
- Some problems:
  - Weak reordering model -- output is not fluent.
  - Many decisions -- many things can go wrong.

Phrase-based Models

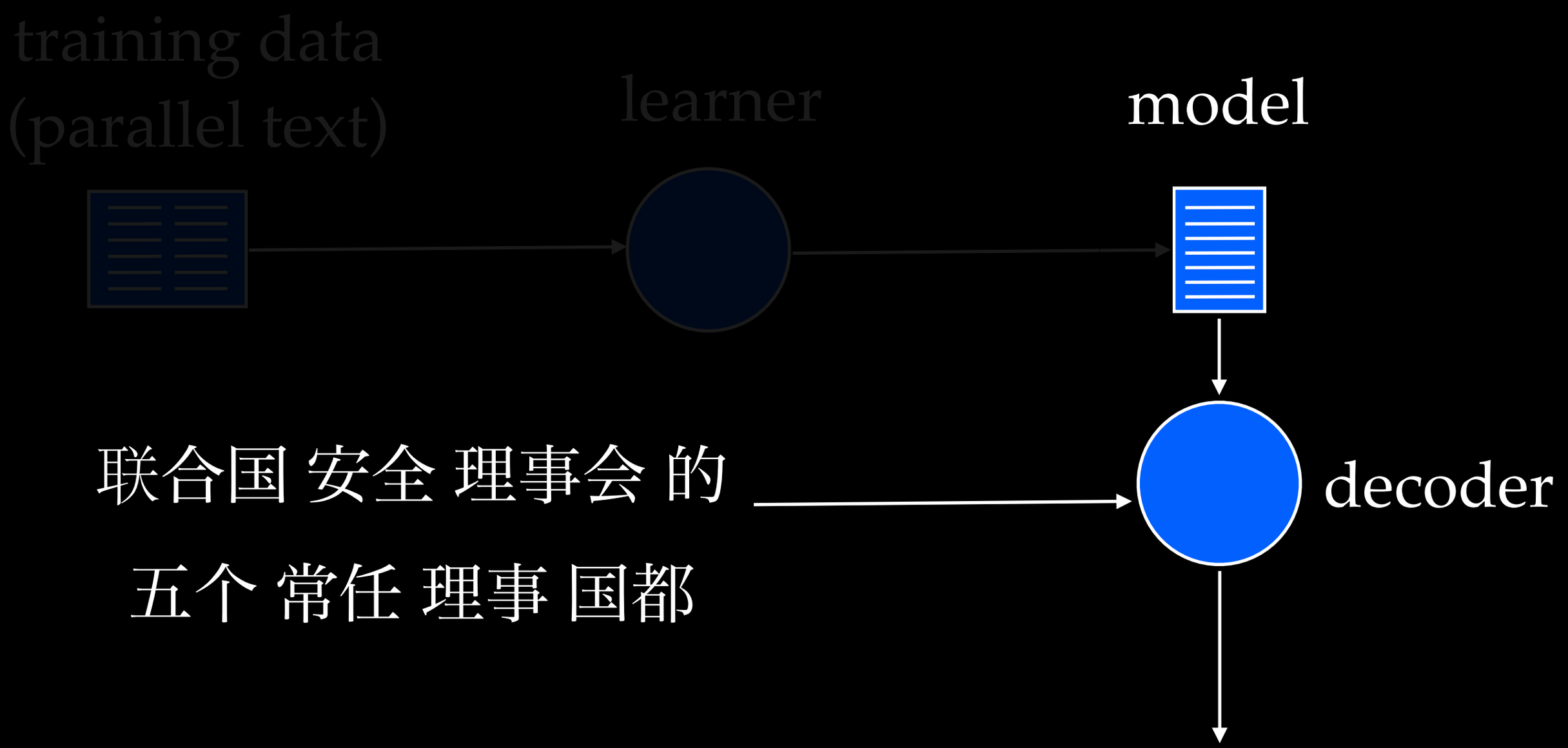
- Segmentation probabilities.
- Phrase translation probabilities.
- Distortion probabilities.
- Some problems:
  - Weak reordering model -- outpu
  - Many decisions -- many things c



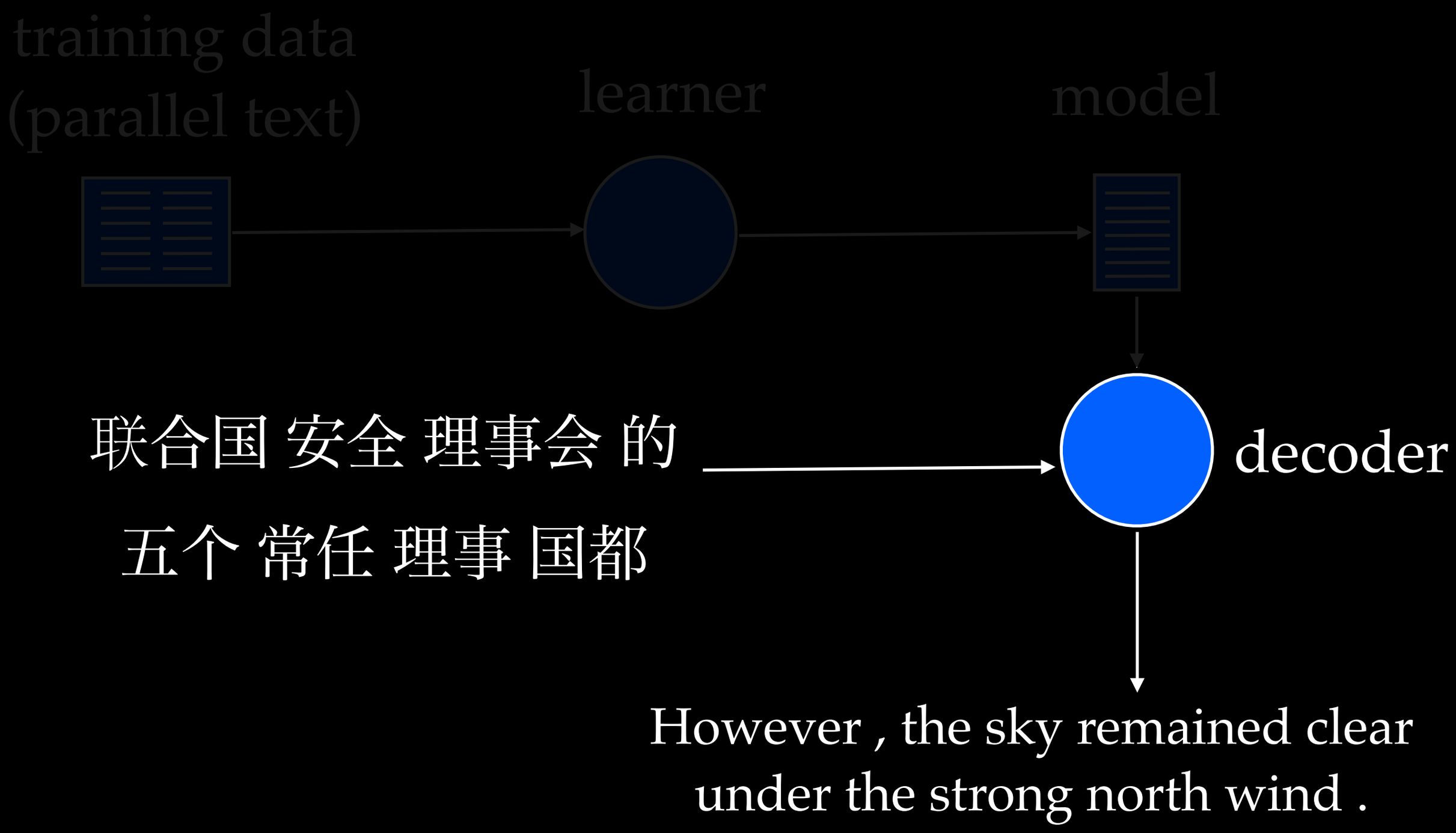
Phrase-based Models

- Segmentation probabilities.
- Phrase translation probabilities.
- Distortion probabilities.
- Some problems:
  - Weak reordering model -- output is not fluent.
  - Many decisions -- many things can go wrong.

Overview


training data
(parallel text)
learner
model
联合国 安全 理事会 的
五个 常任 理事 国都
decoder
However , the sky remained clear
under the strong north wind .

Overview


training data
(parallel text)
learner
model
联合国 安全 理事会 的
五个 常任 理事 国都
decoder
However , the sky remained clear
under the strong north wind .

Decoding

Probability models allow us to make predictions:
Given a particular Chinese sentence, what is the most probable English sentence corresponding to it?

Decoding

Probability models allow us to make predictions:
Given a particular Chinese sentence, what is the most probable English sentence corresponding to it?

In math:

$$\text{argmax}_{English}\, p(English|Chinese)$$

Decoding

Probability models allow us to make predictions:
Given a particular Chinese sentence, what is the most probable English sentence corresponding to it?

In math:

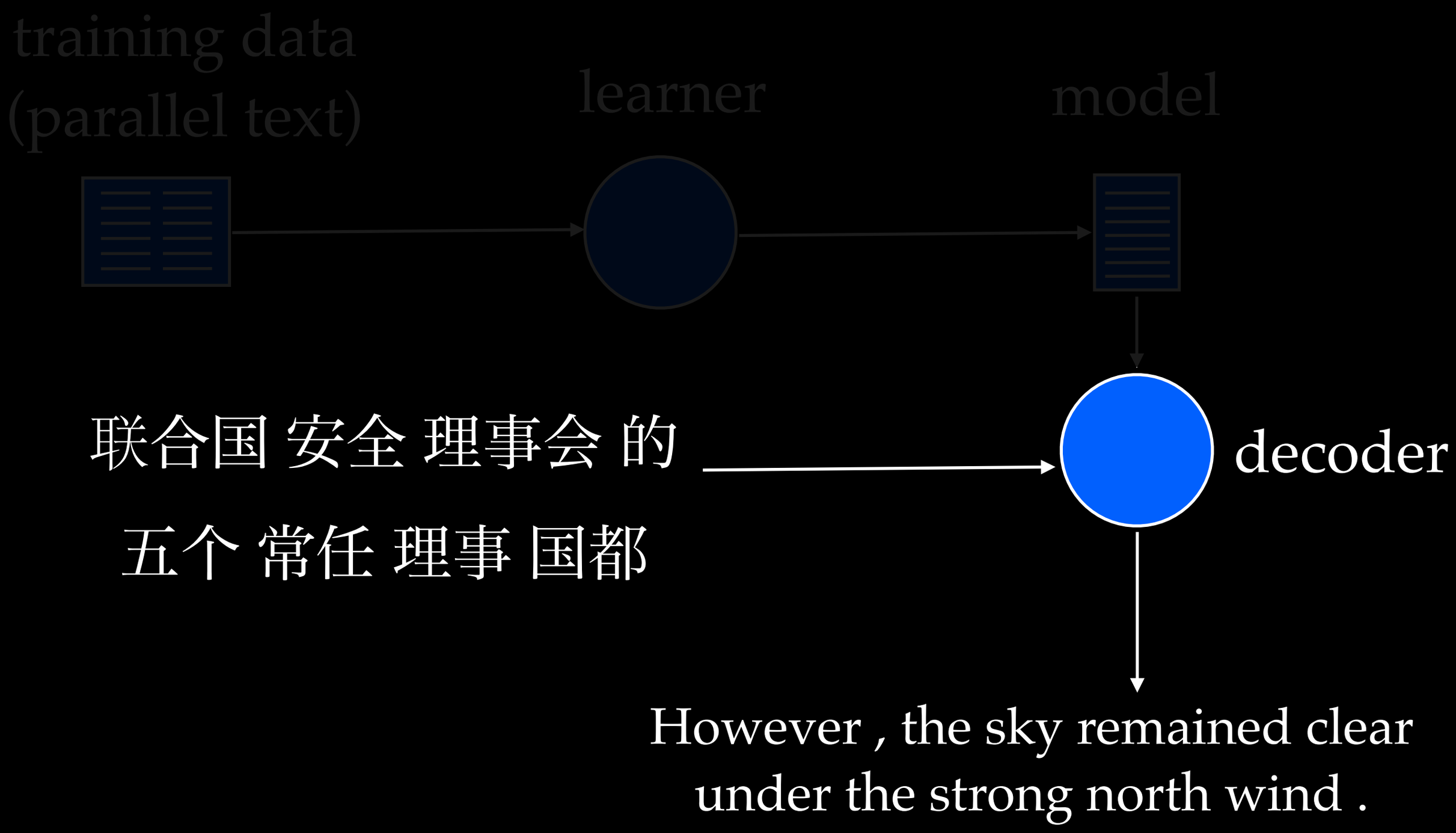
$$\text{argmax}_{English}\, p(English|Chinese)$$

problem: there are a lot of English sentences to choose from!

THE WORD

- Optimization

Overview


training data
(parallel text)
learner
model
联合国 安全 理事会 的
五个 常任 理事 国都
decoder
However , the sky remained clear
under the strong north wind .

北　风　呼啸　。

北　风　呼啸　。

segmentations

substitutions

permutations

北 风 呼啸 。

segmentations $O(2^n)$

substitutions

permutations

北 风 呼啸 。

segmentations	$O(2^n)$
substitutions	$O(5^n)$
permutations	

北 风 呼啸 。

segmentations	$O(2^n)$
substitutions	$O(5^n)$
permutations	$O(n!)$

北 风 呼啸 。

segmentations	$O(2^n)$
substitutions	$O(5^n)$
permutations	$O(n!)$

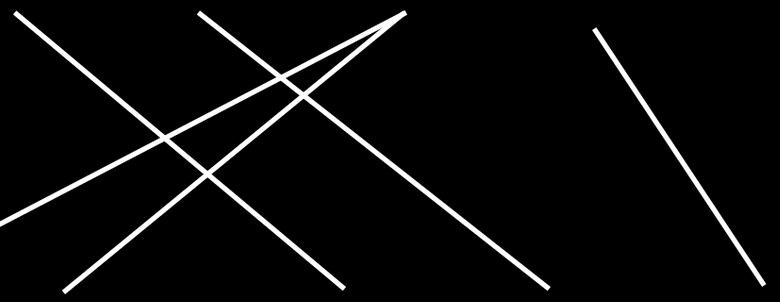
240,000 possibilities!

北　风　呼啸　。

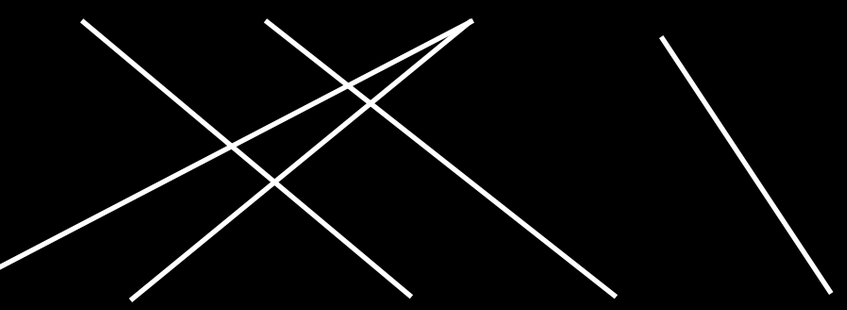
北 风 呼啸 。

the strong north wind .

北 风 呼啸 。



the strong north wind .

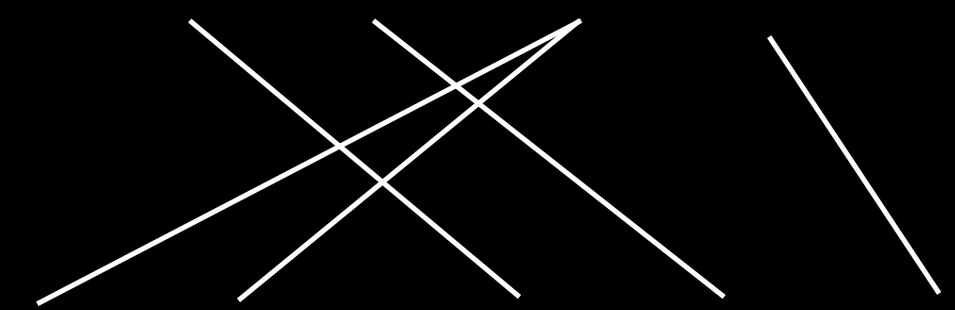


the strong north wind .

Given a sentence pair and an alignment, we can easily calculate

$p(English, alignment|Chinese)$

北 风 呼啸 。



the strong north wind .

Given a sentence pair and an
alignment, we can easily calculate
$p(English, alignment|Chinese)$

Can we do this without enumerating $O(10^n n!)$ pairs?

Key Idea

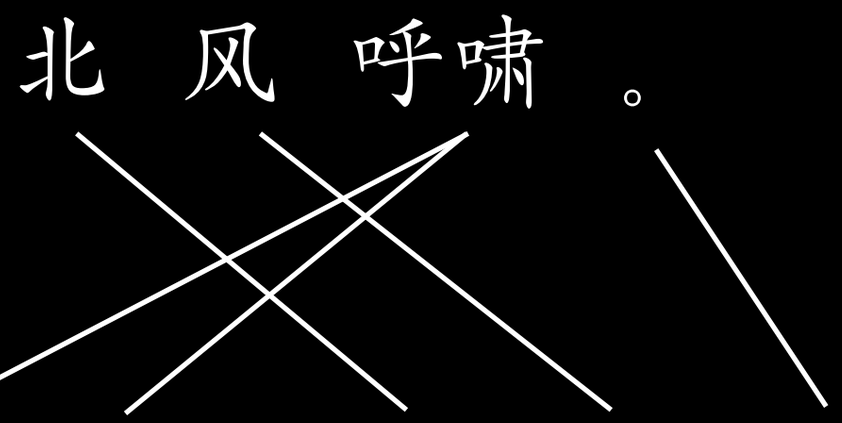
北　风　呼啸　。

Key Idea

北 风 呼啸 。

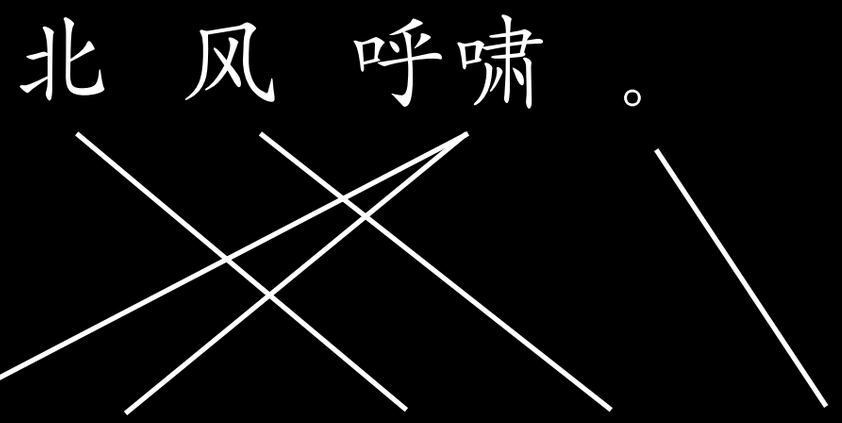
the strong north wind .

Key Idea



the strong north wind .

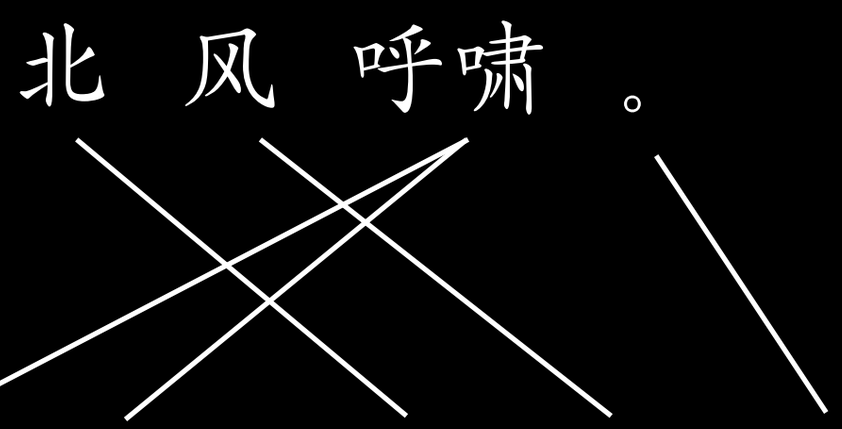
Key Idea



the strong north wind .

There are $O(10^n n!)$ target sentences.

Key Idea



the strong north wind .

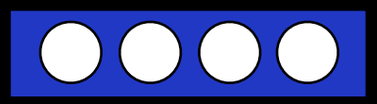
There are $O(10^n n!)$ target sentences.

But there are only $O(5n^2)$ ways to start them.

Key Idea

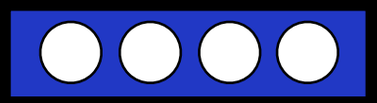
北 风 呼啸 。

Key Idea



北　风　呼啸　。

Key Idea



coverage vector

北　风　呼啸　。

Key Idea

coverage vector

北　风　呼啸　。

Key Idea

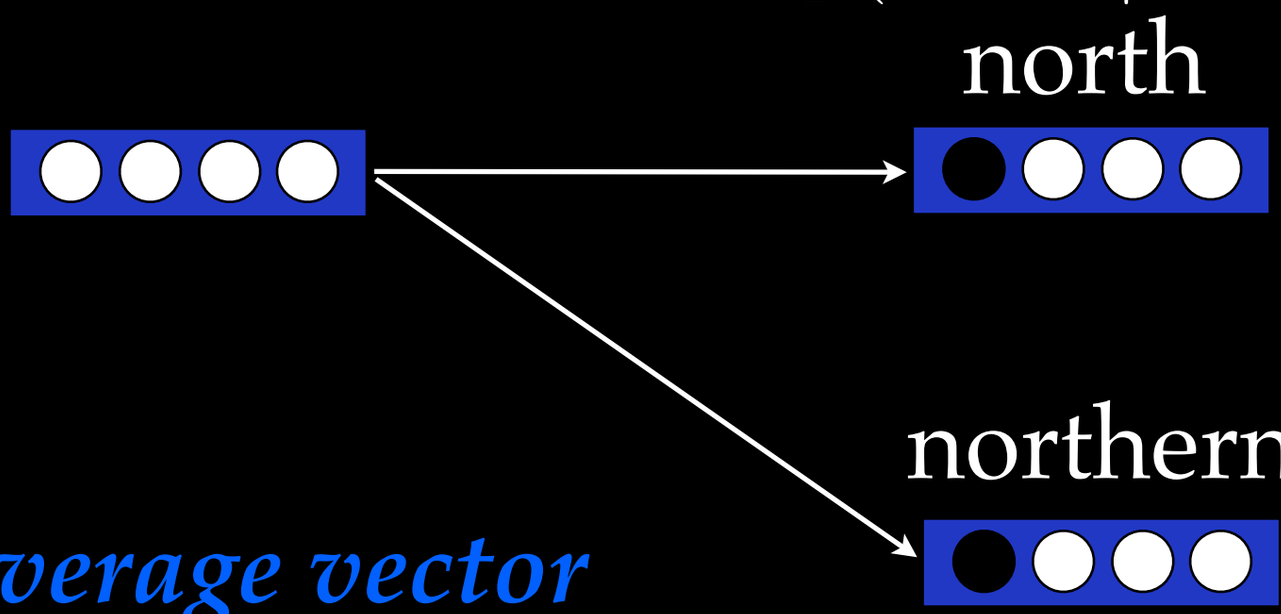
$p(north|START) \cdot p(北|north)$

coverage vector

北 风 呼啸 。

Key Idea

$p(north|START) \cdot p(北|north)$

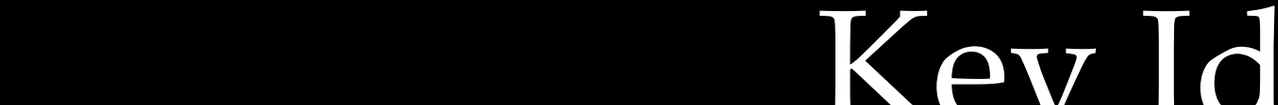


北　风　呼啸　。

Key Idea

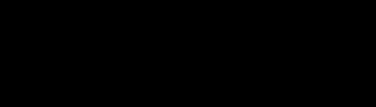
$p(north|START) \cdot p(北|north)$

north



$p(northern|START) \cdot p(北|northern)$

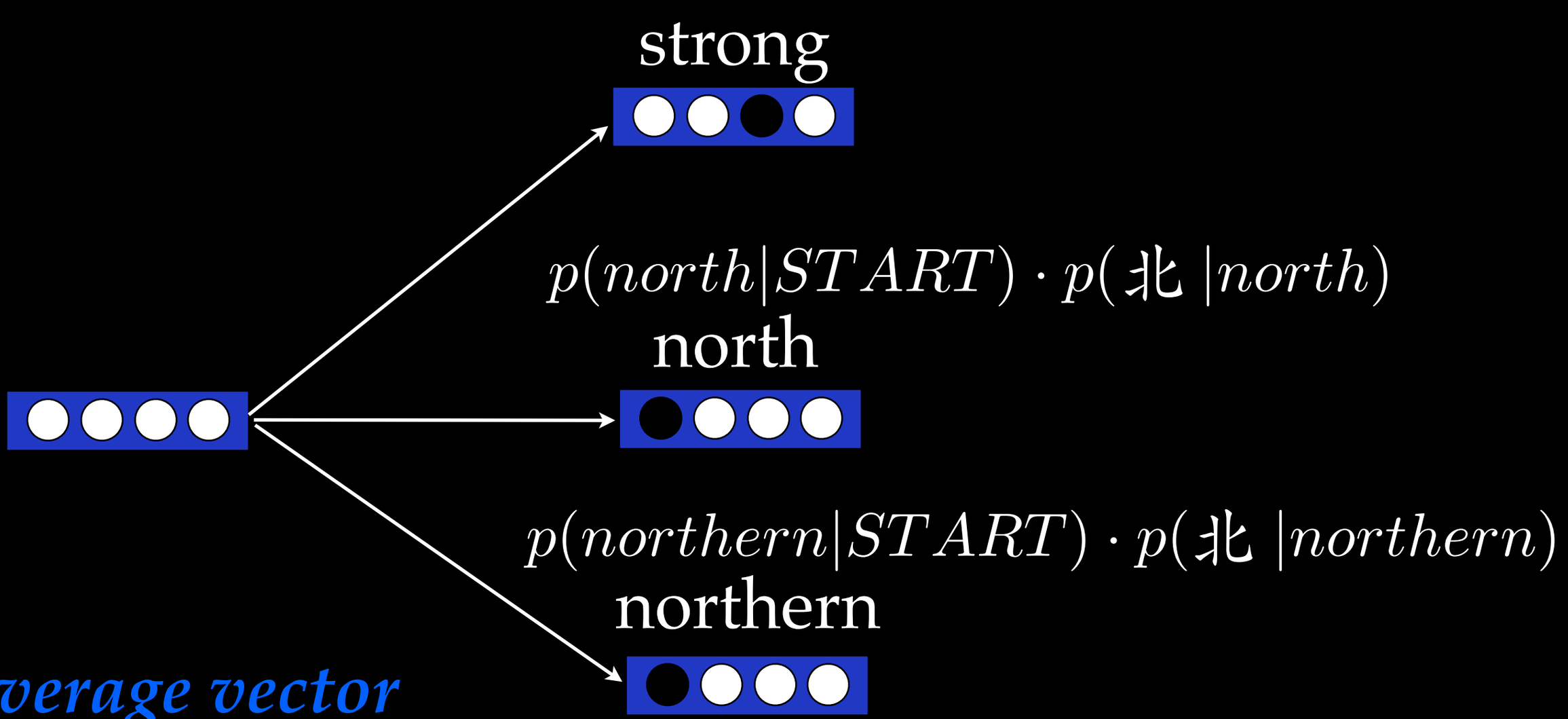
northern



coverage vector

北　风　呼啸　。

Key Idea

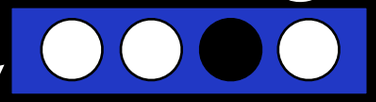


北　风　呼啸　。

Key Idea

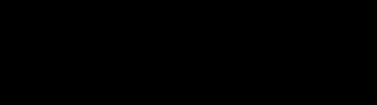
$p(strong|START) \cdot p(\text{呼啸}|strong)$

strong



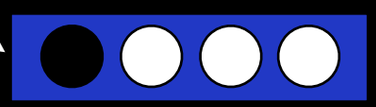
$p(north|START) \cdot p(\text{北}|north)$

north



$p(northern|START) \cdot p(\text{北}|northern)$

northern



coverage vector

北　风　呼啸　。

Key Idea

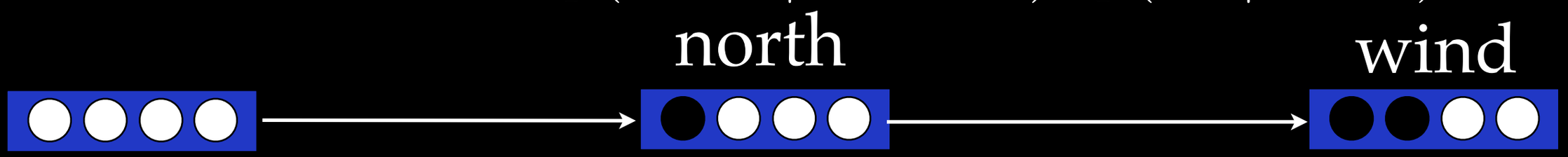
$p(north|START) \cdot p(北|north)$

coverage vector

北 风 呼啸 。

Key Idea

$p(north|START) \cdot p(北|north)$

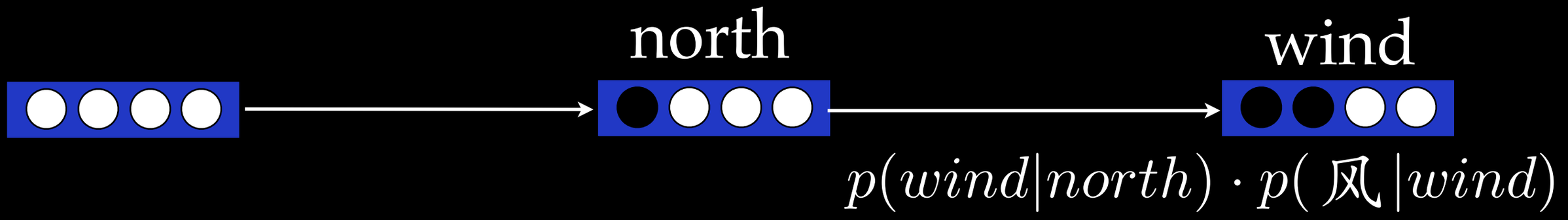


coverage vector

北 风 呼啸 。

Key Idea

$p(north|START) \cdot p(北|north)$

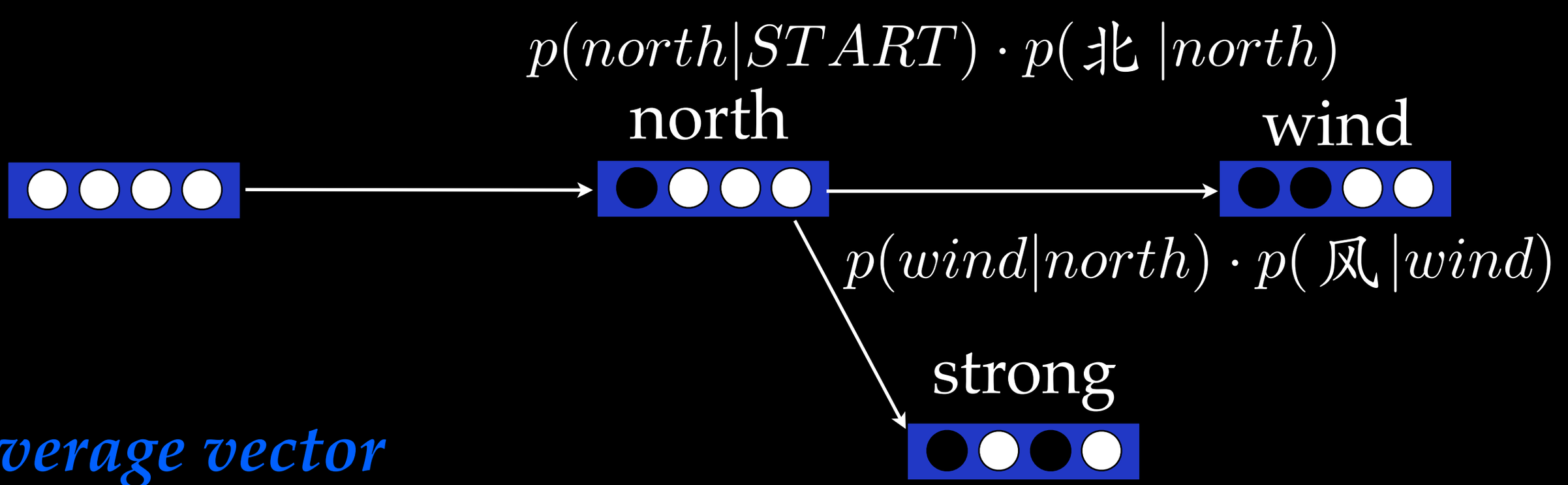


$p(wind|north) \cdot p(风|wind)$

coverage vector

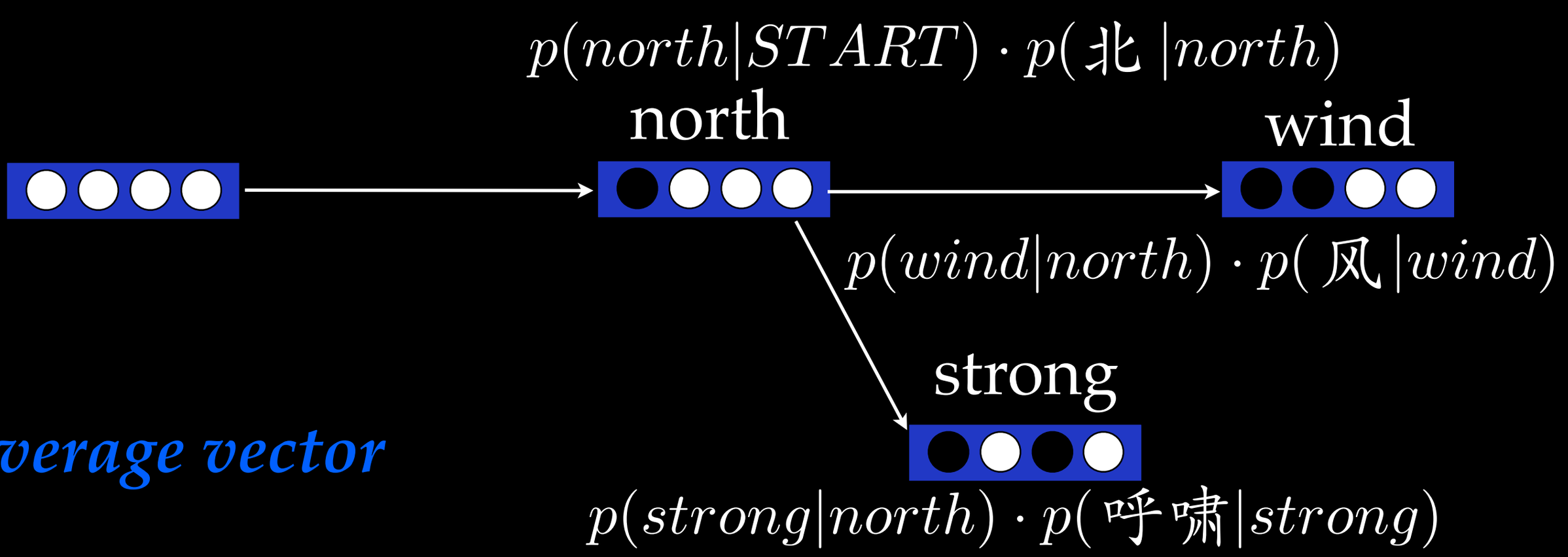
北 风 呼啸 。

Key Idea



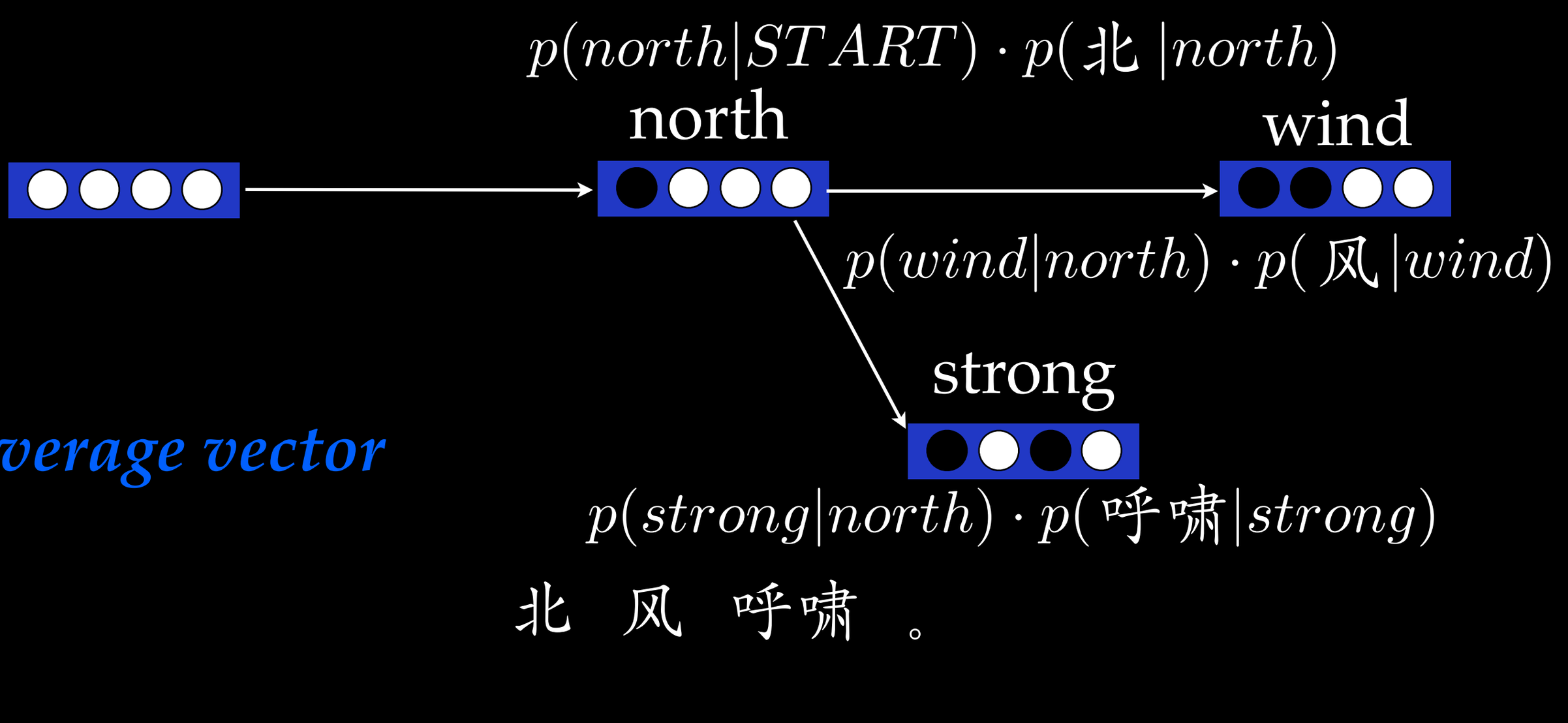
北　风　呼啸　。

Key Idea

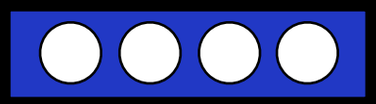

$p(north|START) \cdot p(北|north)$
north
wind
$p(wind|north) \cdot p(风|wind)$
strong
$p(strong|north) \cdot p(呼啸|strong)$
coverage vector
北 风 呼啸 。

Key Idea

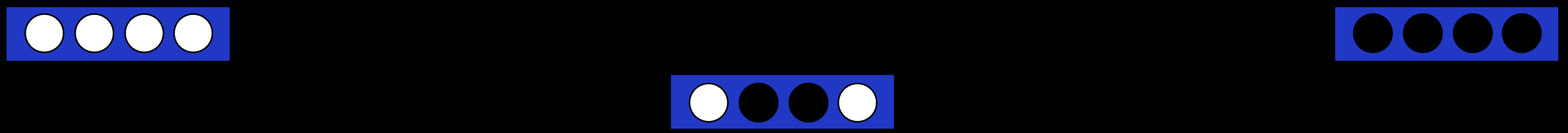
Work done at sentence beginnings is shared across many possible output sentences!



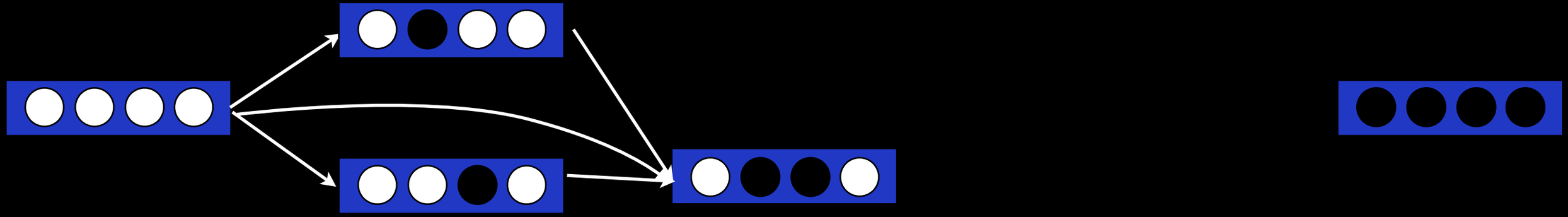
Key Idea



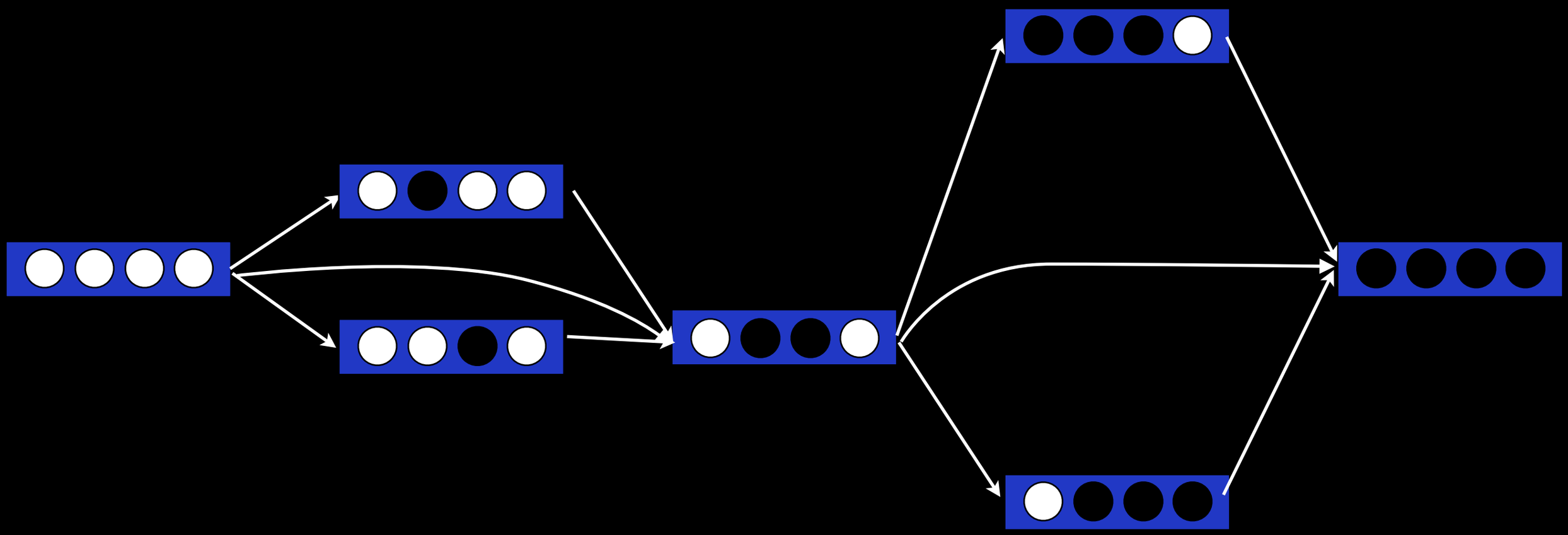
Key Idea



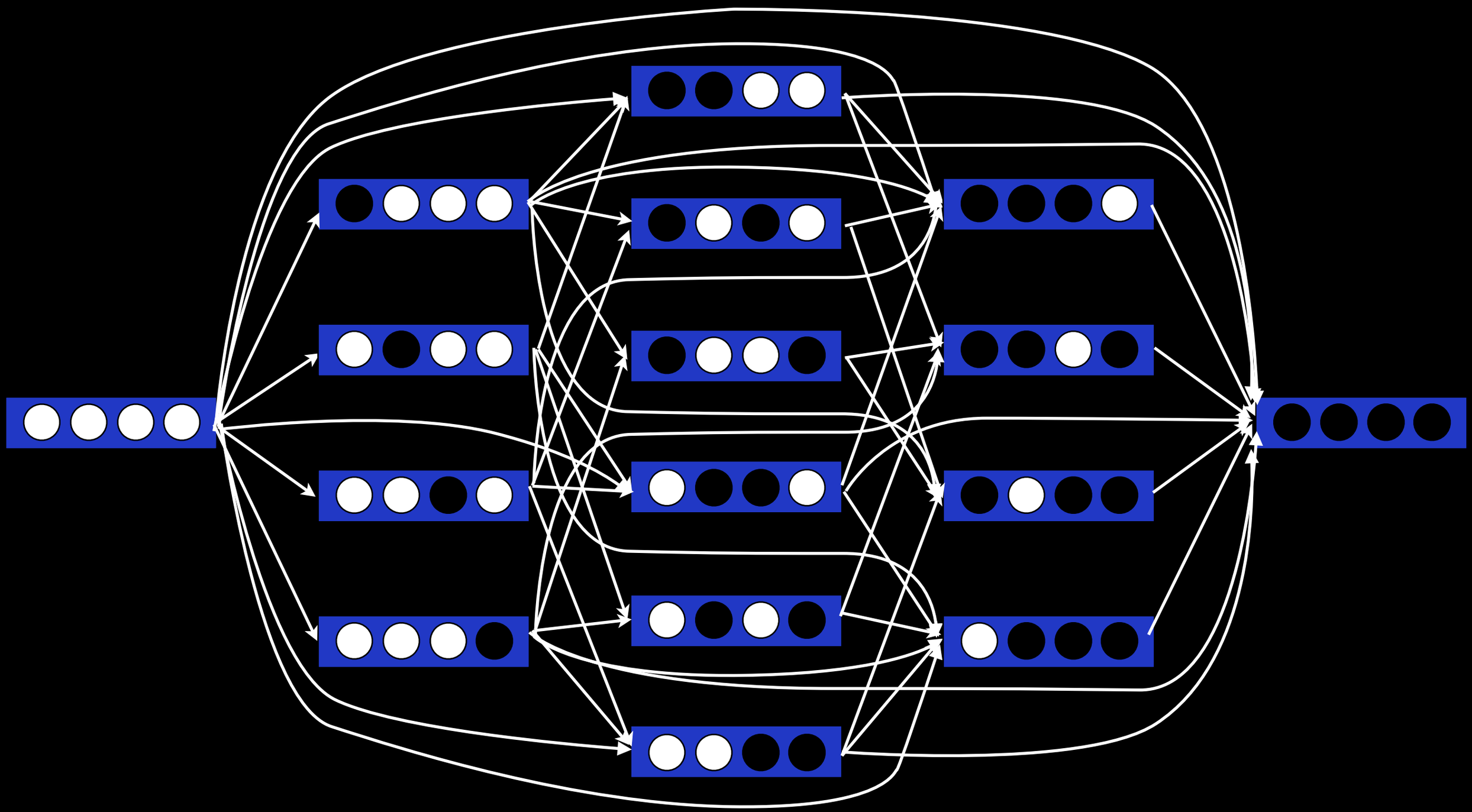
Key Idea



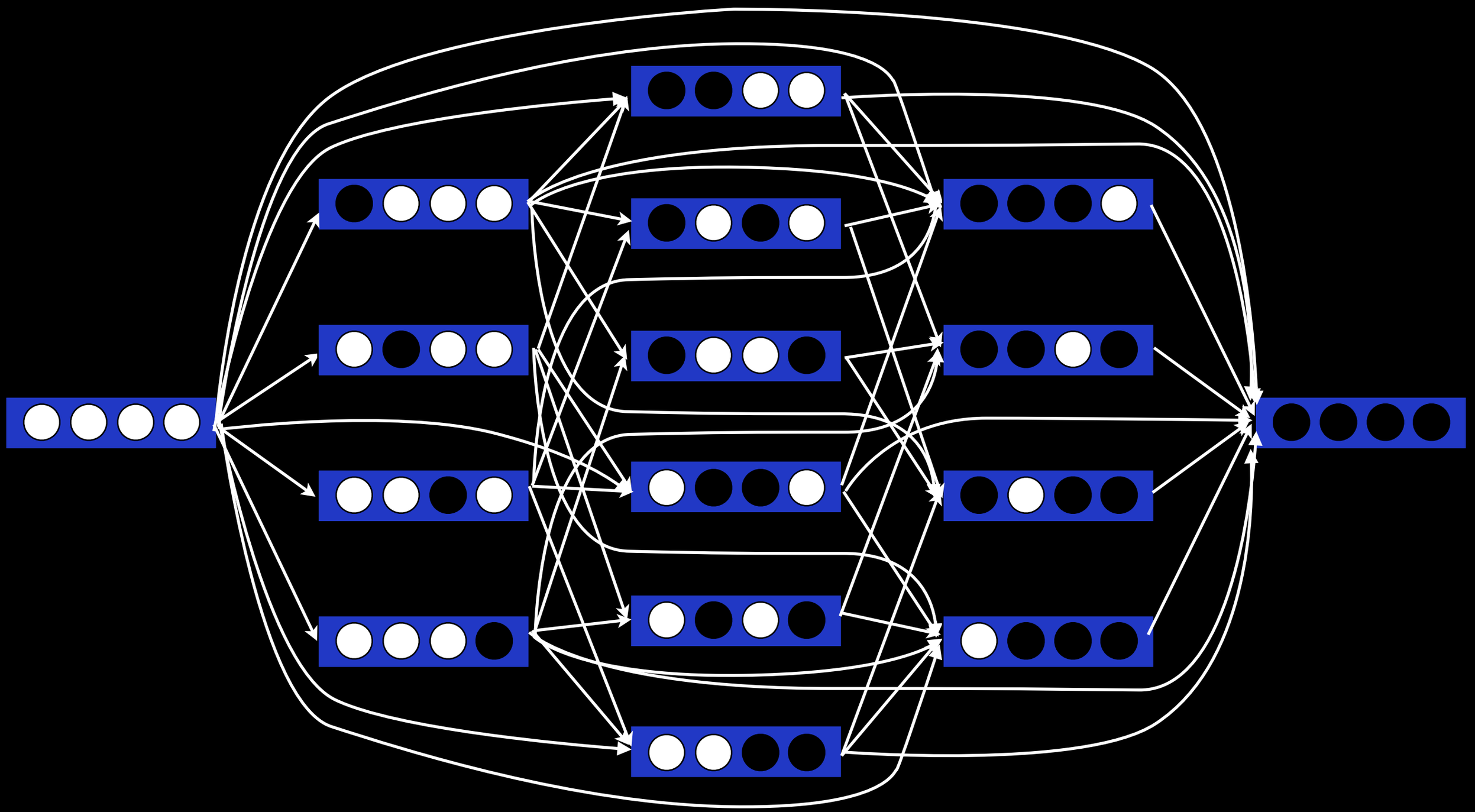
Key Idea



Key Idea



Key Idea

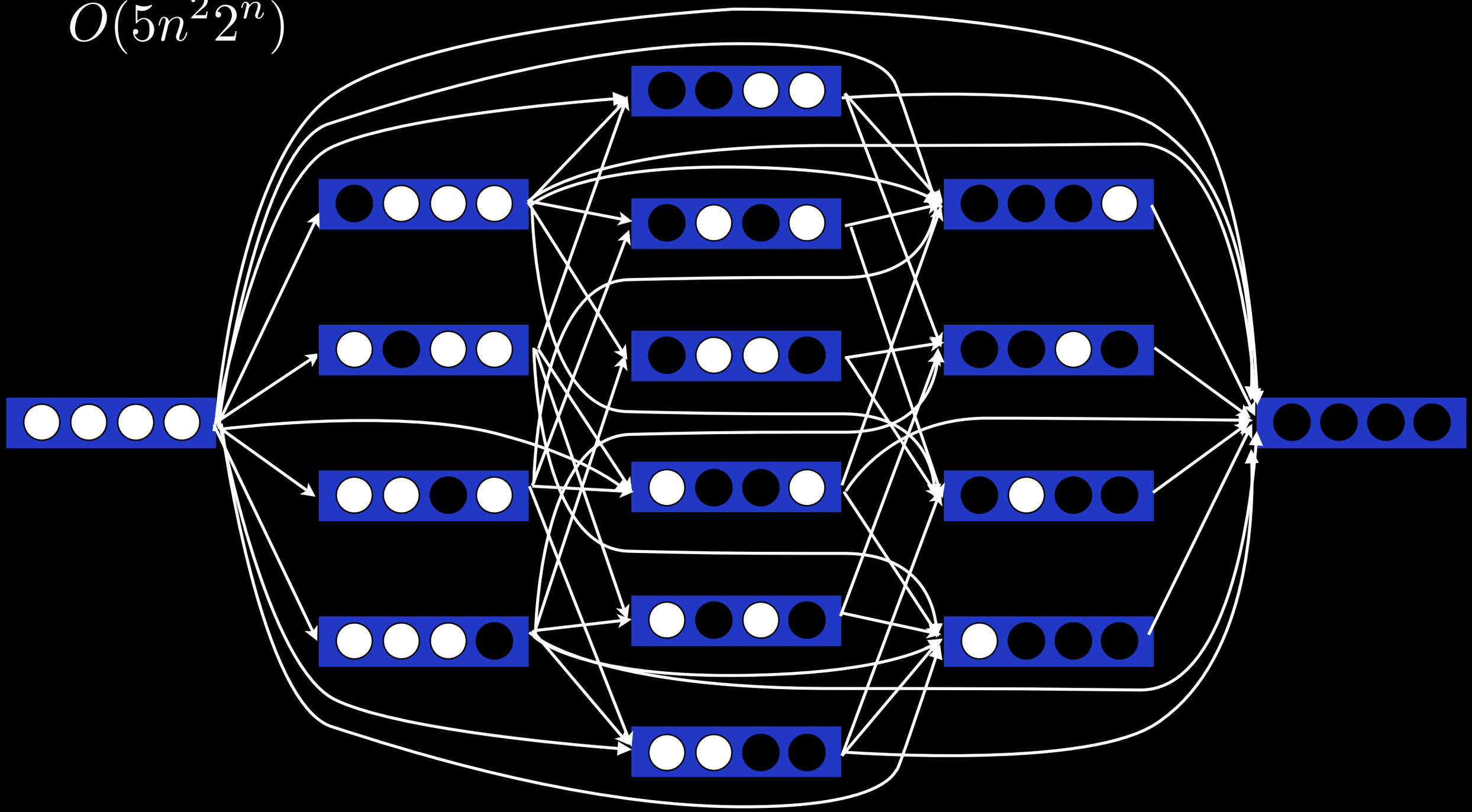


Dynamic Programming

Key Idea

amount of work:

$O(5n^2 2^n)$



Dynamic Programming

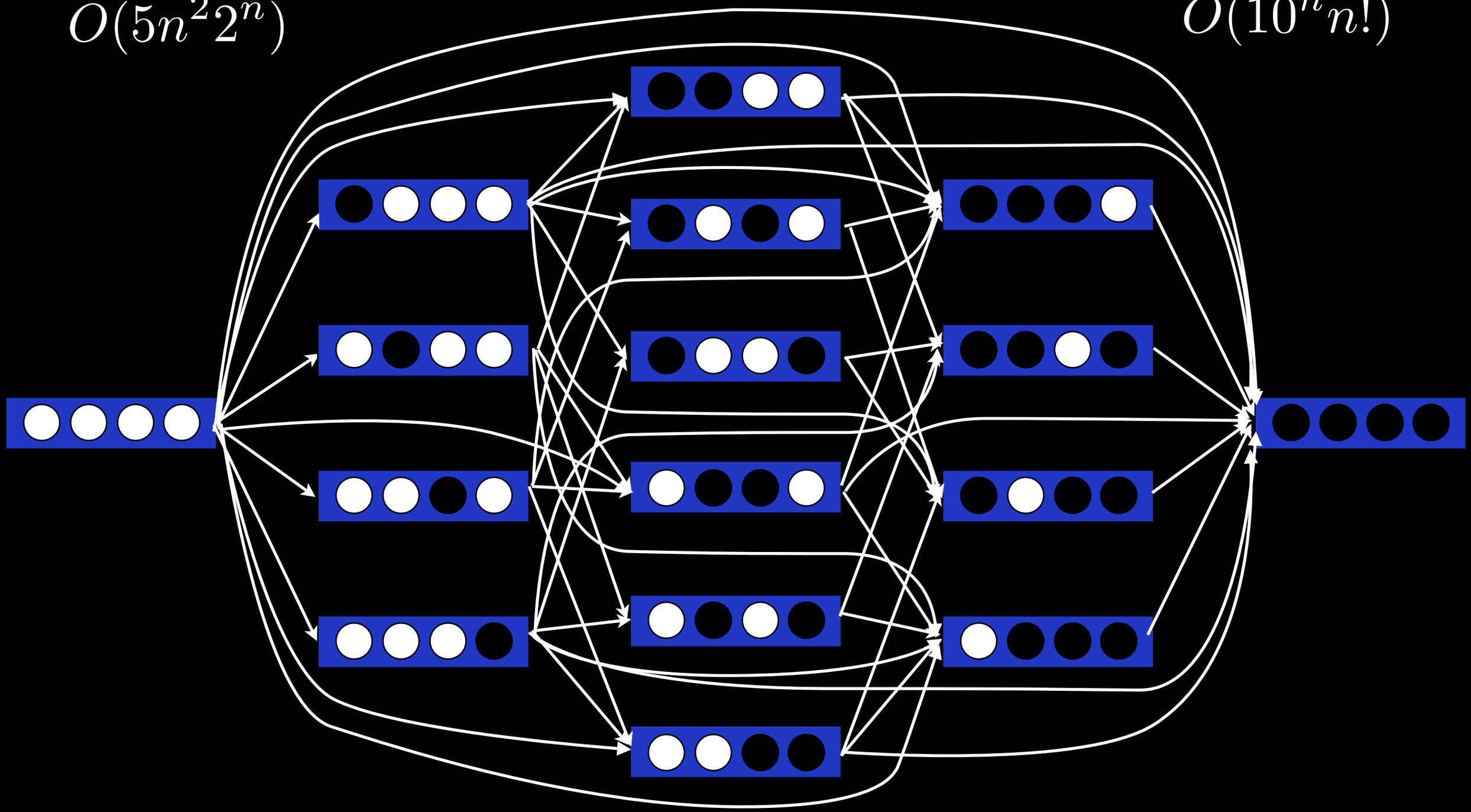
Key Idea

amount of work:

$O(5n^2 2^n)$

bad, but much better than

$O(10^n n!)$



Dynamic Programming

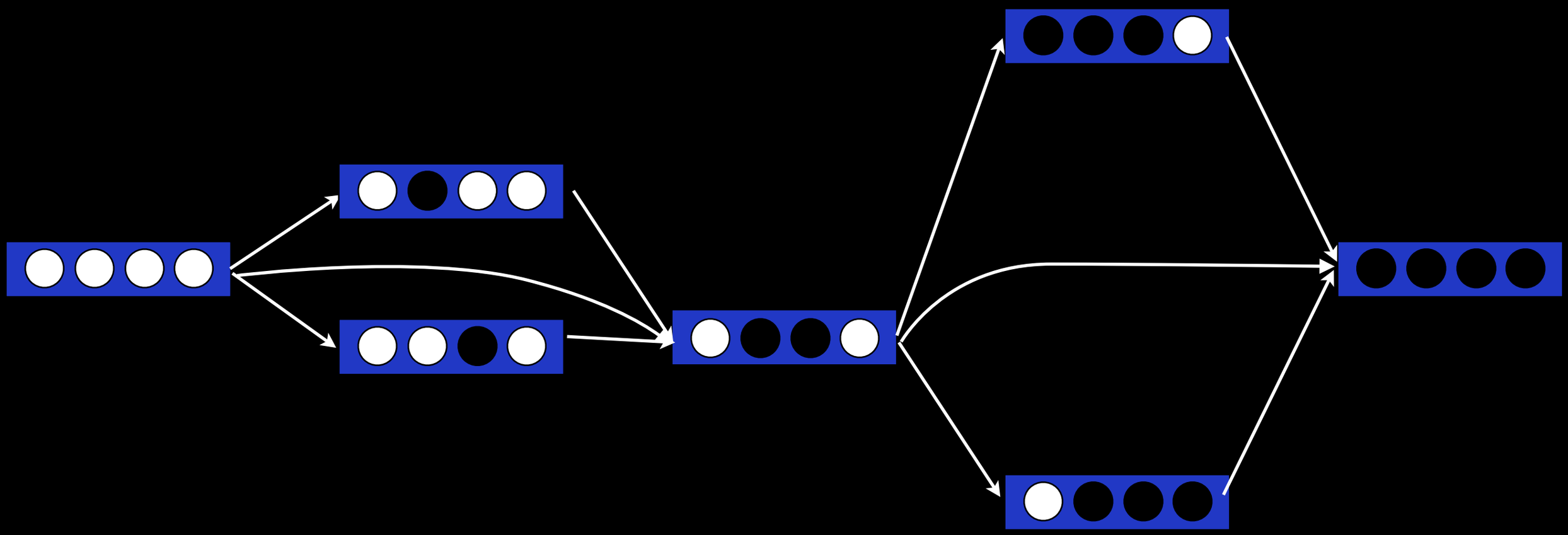
amount of work:

$O(5n^2 2^n)$

Key Idea

bad, but much better than

$O(10^n n!)$



Dynamic Programming

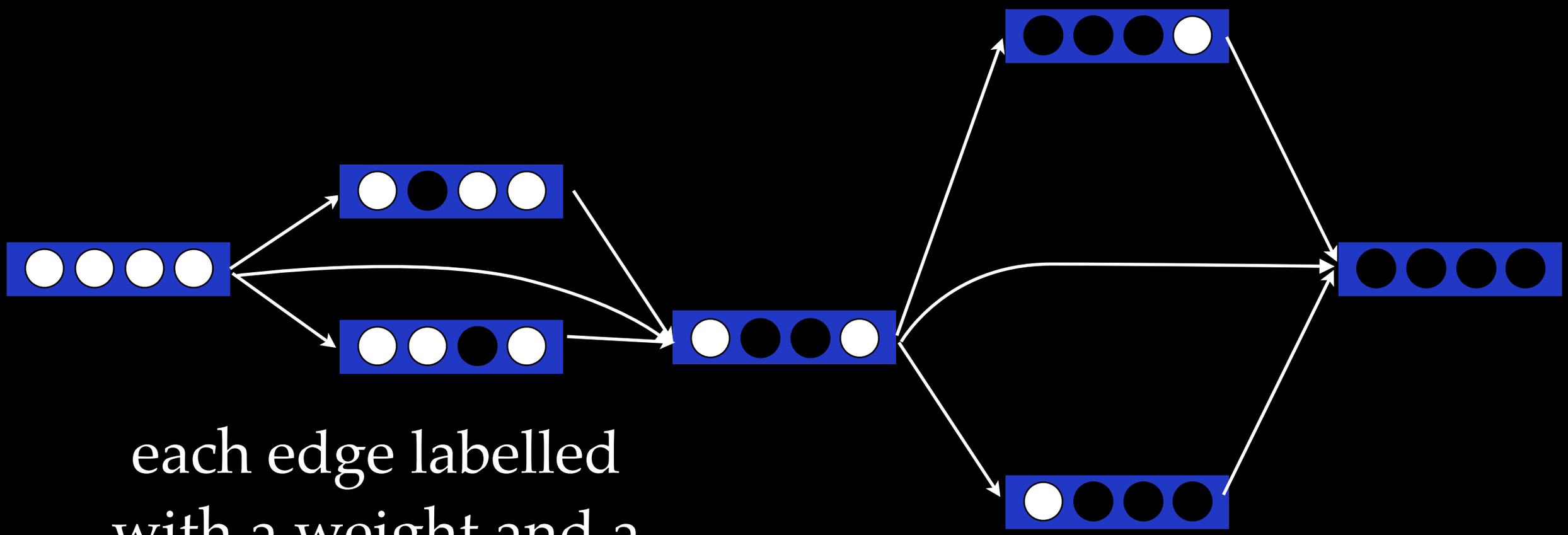
Key Idea

amount of work:

$O(5n^2 2^n)$

bad, but much better than

$O(10^n n!)$



each edge labelled with a weight and a word (or words)

Dynamic Programming

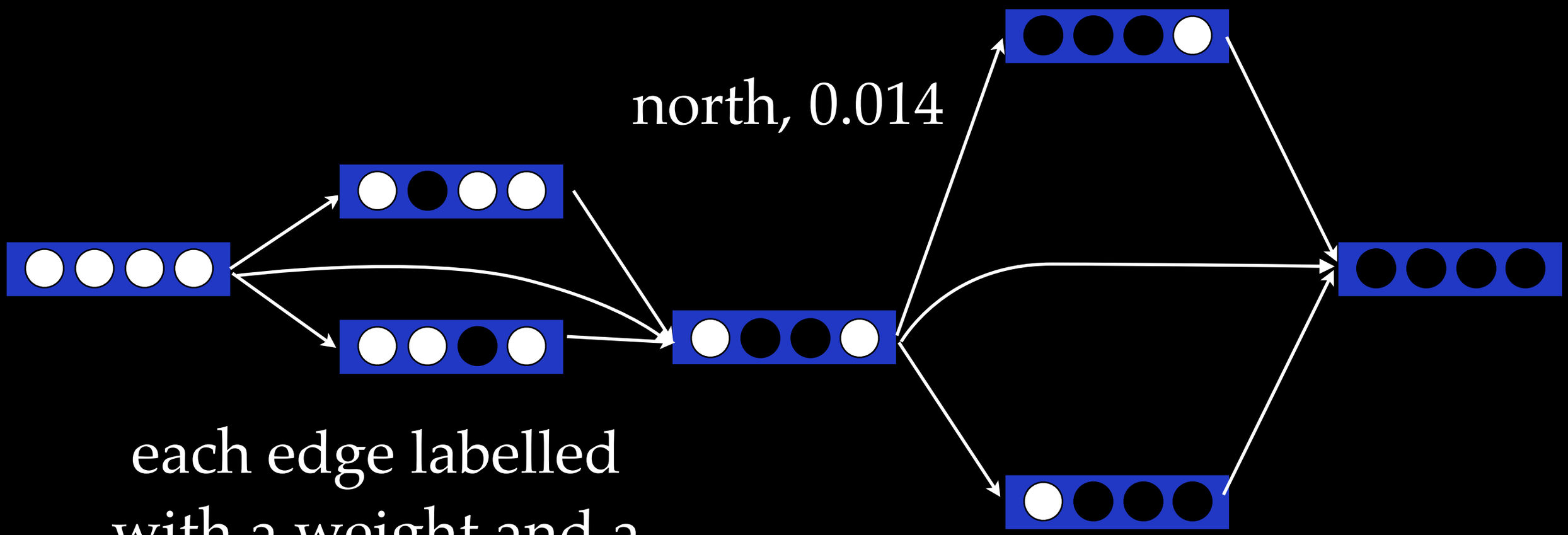
Key Idea

amount of work:

$O(5n^2 2^n)$

bad, but much better than

$O(10^n n!)$



each edge labelled with a weight and a word (or words)

Dynamic Programming

Key Idea

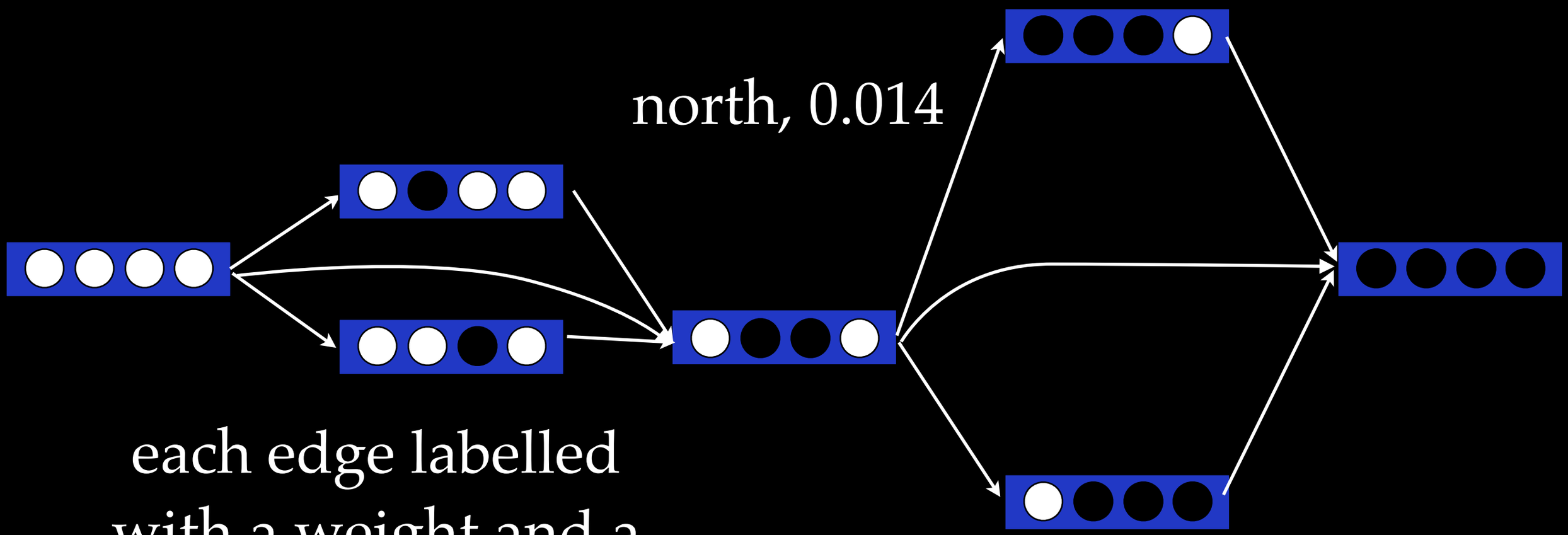
amount of work:

$O(5n^2 2^n)$

bad, but much better than

$O(10^n n!)$

weighted finite-state automata



each edge labelled with a weight and a word (or words)

Dynamic Programming

Weighted languages

- The lattice describing the set of all possible translations is a ***weighted finite state automaton***.
- So is the language model.
- Since regular languages are closed under intersection, we can intersect the devices and run shortest path graph algorithms.
- Taking their intersection is equivalent to computing the probability under Bayes' rule.

Practical Issues

Practical Issues

$O(5n^2 2^n)$ is still far too much work.

Practical Issues

$O(5n^2 2^n)$ is still far too much work.

Can we do better?

Practical Issues

$O(5n^2 2^n)$ is still far too much work.

Can we do better?

NO! Knight (1999) shows that this is NP-Complete.

Practical Issues

$O(5n^2 2^n)$ is still far too much work.

Can we do better?

NO! Knight (1999) shows that this is NP-Complete.

Barry Haddow will tell you how we can still
make it work on Thursday.

Recap

Recap

- Probability theory enables us to learn from data.

Recap

- Probability theory enables us to learn from data.
- Very simple models get us pretty far!

Recap

- Probability theory enables us to learn from data.
- Very simple models get us pretty far!
- There's no data like more data.

Recap

- Probability theory enables us to learn from data.
- Very simple models get us pretty far!
- There's no data like more data.
- Word-based models follow intuitions, but not all.

Recap

- Probability theory enables us to learn from data.
- Very simple models get us pretty far!
- There's no data like more data.
- Word-based models follow intuitions, but not all.
- Phrase-based models are similar, but more effective.

Recap

- Probability theory enables us to learn from data.
- Very simple models get us pretty far!
- There's no data like more data.
- Word-based models follow intuitions, but not all.
- Phrase-based models are similar, but more effective.
- *All* of these models are weighted regular languages.

Recap

- Probability theory enables us to learn from data.
- Very simple models get us pretty far!
- There's no data like more data.
- Word-based models follow intuitions, but not all.
- Phrase-based models are similar, but more effective.
- *All* of these models are weighted regular languages.
- Need dynamic programming with approximations.

Recap

- Probability theory enables us to learn from data.
- Very simple models get us pretty far!
- There's no data like more data.
- Word-based models follow intuitions, but not all.
- Phrase-based models are similar, but more effective.
- *All* of these models are weighted regular languages.
- Need dynamic programming with approximations.
- Is this the best we can do? Stay tuned on Thursday.

ارتفاع عجز الميزان التجاري الأردني

قيمة الواردات الأردنية بلغت 7.39 مليارات دولار في النصف الأول من العام (الجزيرة نت)

أفادت بيانات رسمية بأن العجز في الميزان التجاري الأردني ارتفع في النصف الأول من العام الحالي بنسبة 18.1% نتيجة زيادة حجم الواردات مقابل الصادرات، كما تراجع حجم القروض الائتمانية التي قدمتها البنوك الأردنية بنسبة 11% خلال الفترة نفسها.

وقالت بيانات لدائرة الإحصاءات العامة نشرت اليوم الاثنين إن قيمة العجز في الميزان التجاري بلغت 2.79 مليار دينار أردني (3.94 مليار دولار أميركي).

وأشارت البيانات إلى ارتفاع حجم الصادرات خلال النصف الأول من العام بنسبة 16.6%, حيث بلغت 2.05 مليار دينار (2.91 مليار دولار)، كما ارتفعت قيمة الواردات بنسبة 11.7% لتبلغ 5.23 مليارات دينار (7.39 مليارات دولار).

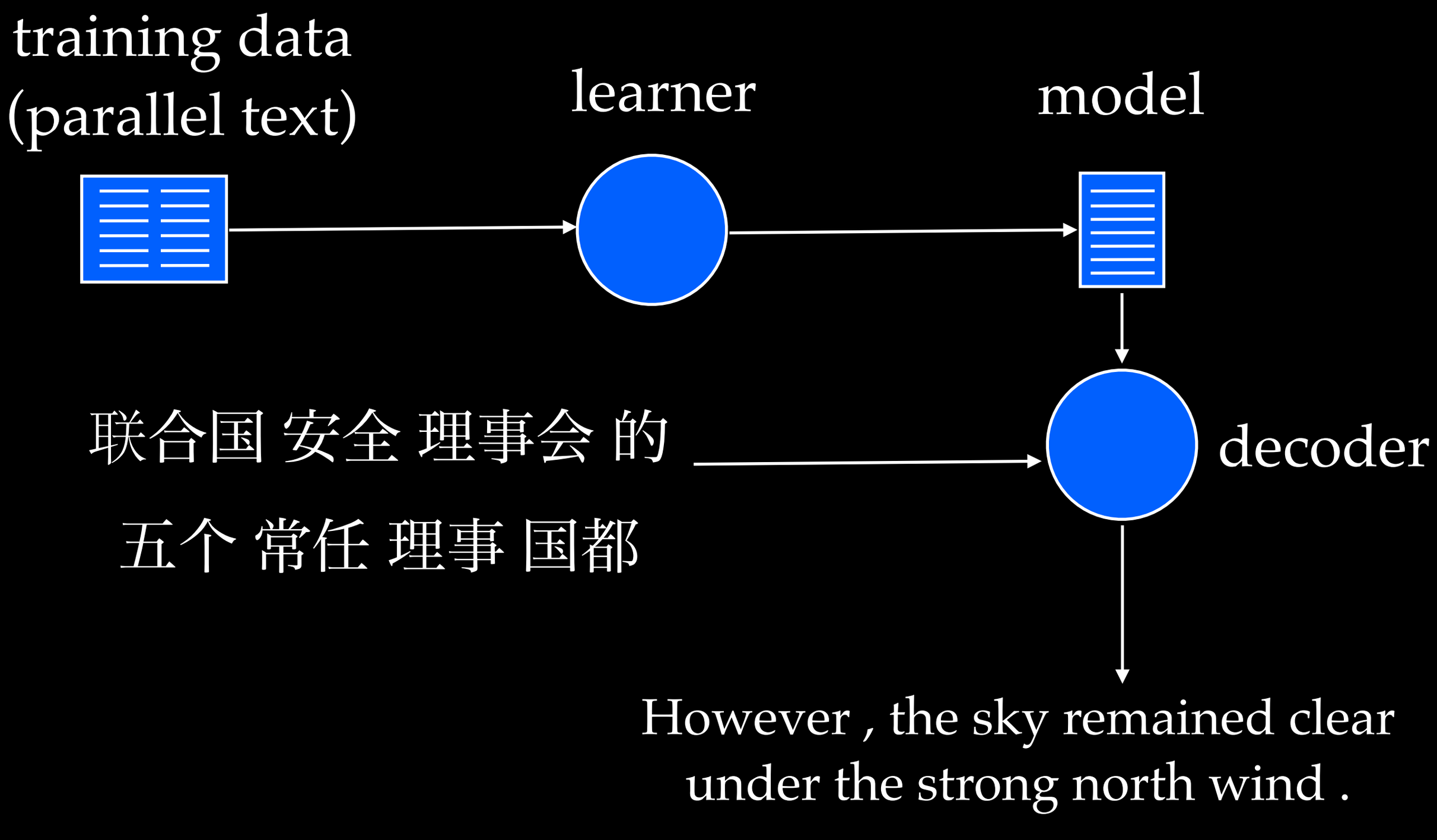
Value of Jordanian imports amounted to 7.39 billion dollars in the first half of the year (island Net)

According to official statements that the trade balance deficit rose Jordan in the first half of this year by 18.1% due to increased volume of imports versus exports, as the decline in the volume of credit provided by banks of Jordan by 11% during the same period.

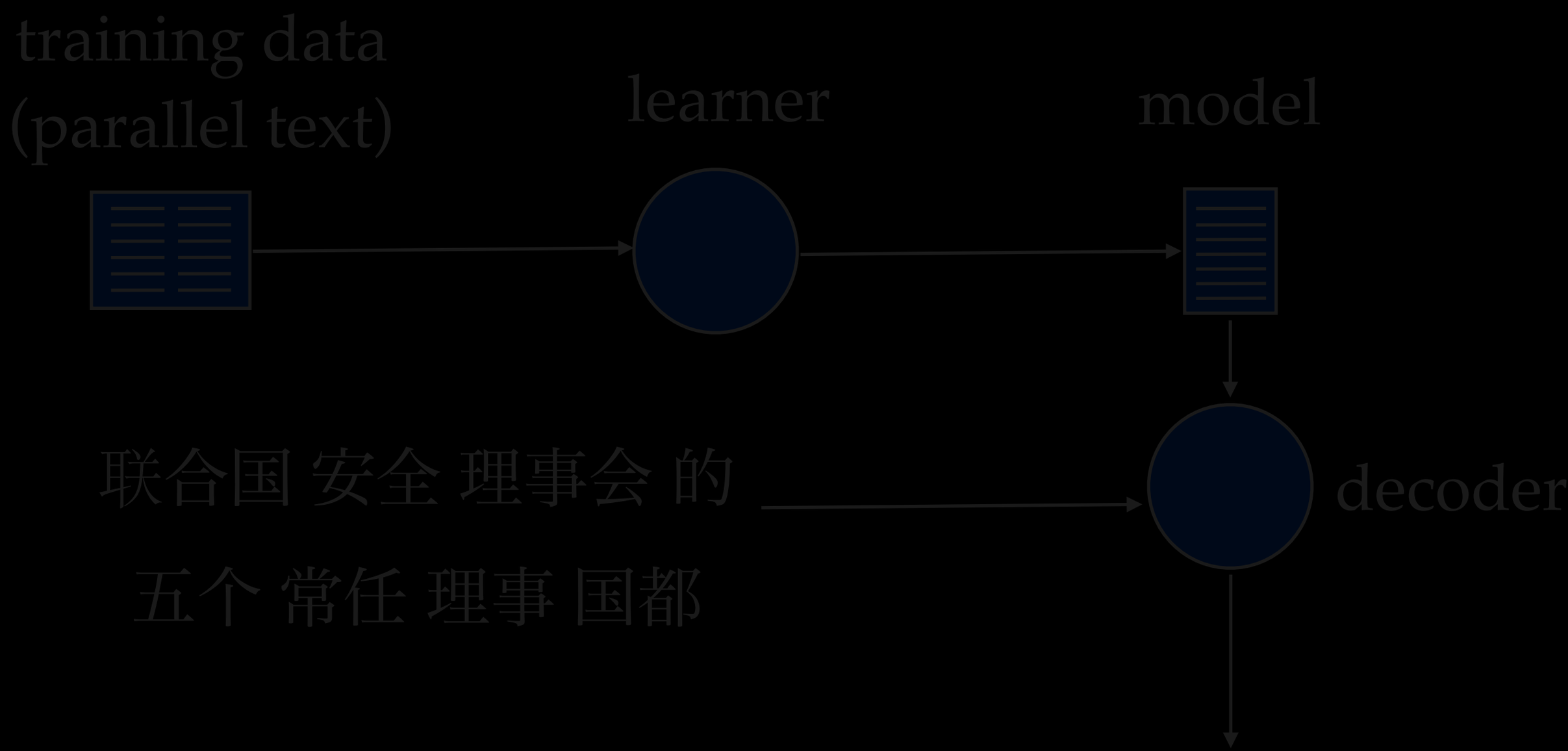
The data for the Department of Statistics published on Monday that the value of the trade balance deficit amounted to 2.79 billion Jordanian dinars (3.94 billion U.S. dollars). The data indicated the high volume of exports during the first half of the year by 16.6%, reaching 2.05 billion dinars (2.91 billion dollars), as imports rose by 11.7% to 5.23 billion dinars (7.39 billion dollars).

The export value increased significantly for the countries of the Greater Arab Free Trade, including Saudi Arabia, and the countries of the free trade agreement for North America, including the United States, as well as non-Arab Asian

Overview


training data
(parallel text)
learner
model
联合国 安全 理事会 的
五个 常任 理事 国都
decoder
However , the sky remained clear
under the strong north wind .

Overview


training data
(parallel text)
learner
model
联合国 安全 理事会 的
五个 常任 理事 国都
decoder
However , the sky remained clear
under the strong north wind .

Overview

Evaluation

More has been written about machine translation evaluation than about machine translation itself.

Yorick Wilks

- Why evaluate?
  - Rank systems
  - Evaluate incremental changes
  - Assess new ideas objectively

美国愿和北韩谈判但拒绝再付出报酬

US willing to negotiate with North Korea but not to pay more compensation.

美国愿和北韩谈判但拒绝再付出报酬

US willing to negotiate with North Korea but not to pay more compensation.

The United States is willing to hold talks with North Korea but refused to pay remuneration.

“奋进”号因机械手故障推迟到升空

Launch of “Endeavour” delayed by robotic arm problems.

“奋进”号因机械手故障推迟到升空

Launch of “Endeavour” delayed by robotic arm problems.

“Progress” postponed because of mechanical hand into the sky.

Chinese people in the traditional Spring Festival is approaching, the CPC Central Committee this afternoon in Zhongnanhai on the 22nd non-Party personages to convene a forum in Spring Festival, invited the central committees of democratic parties, the leadership of the National Federation of Industry and Commerce and personages without party affiliation on behalf of comrades gathered together State yes, talked in length about the friendship, to greet the Chinese New Year. CPC Central Committee General Secretary and State President and Central Military Commission Chairman Hu Jintao on behalf of the CPC Central Committee, the State Council, to the central committees of democratic parties, leaders of the National Federation of Industry and Commerce and personages without party affiliation, to members of the united front, to extend my New Year's blessing.

Although the northern wind shrieked across the sky , it was still very clear .

Although the northern wind shrieked across the sky , it was still very clear .

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

Although the northern wind shrieked across the sky , it was still very clear .

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

Although the northern wind shrieked across the sky , it was still very clear .

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

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

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

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

Although the northern wind shrieked across the sky , it was still very clear .

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

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

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

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

Although the northern wind shrieked across the sky , it was still very clear .

unigram: 11/15 bigram: 4/14 trigram: 1/13

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

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

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

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

very clear .

However , the sky remained clear under the strong north wind .
Although a north wind was howling , the sky remained clear and blue .
The sky was still crystal clear , though the north wind was howling .
Despite the strong northerly winds , the sky remains very clear .

very clear .

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

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

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

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

very clear .

unigram: 3/3 bigram: 2/2 trigram: 1/1

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

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

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

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

very clear .

brevity penalty: 3/12

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

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

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

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

BLEU

very clear .

brevity penalty: 3/12

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

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

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

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

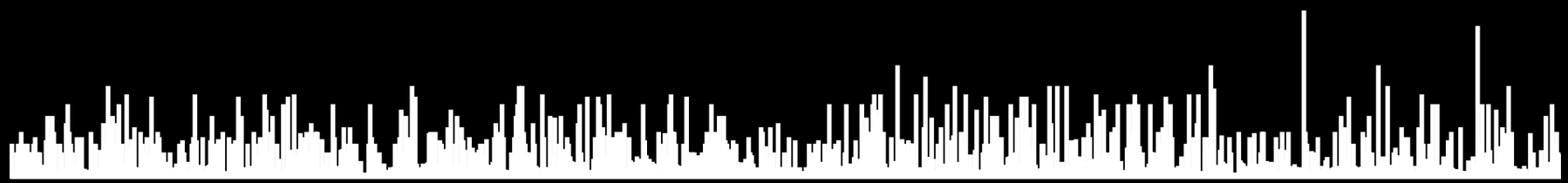
BLEU-1
BLEU-4
BLEU-v11b
BLEU-v12
METEOR-v0.6
NIST-v11b
TER-v0.7.254-GRR
ATEC1
ATEC2
ATEC3
ATEC4
Meteor-v0.7

BEwT-E
Badger
BadgerLite
Bleu-sbp
BleuSP
CDer
DP-Or
DP-Orp
DR-Or
EDPM
LET
METEOR-ranking
MaxSim

RTE
RTE-MT
SEPIA1
SEPIA2
SNR
SR-Or
SVM-Rank
TERp
ULCh
ULCopt
invWer
mBLEU
mTER

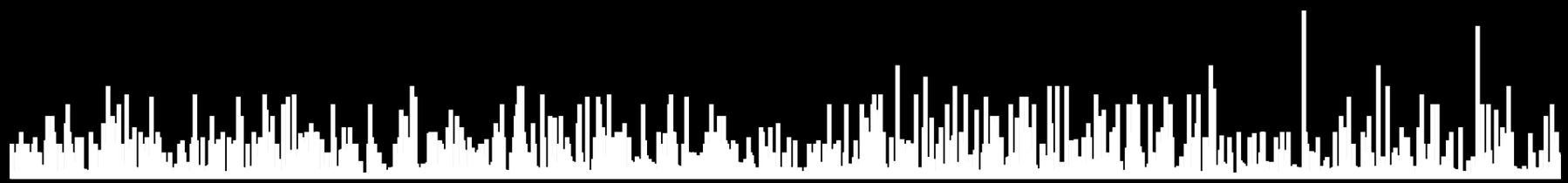
A bit more on learning...

$p(Chinese|English)$

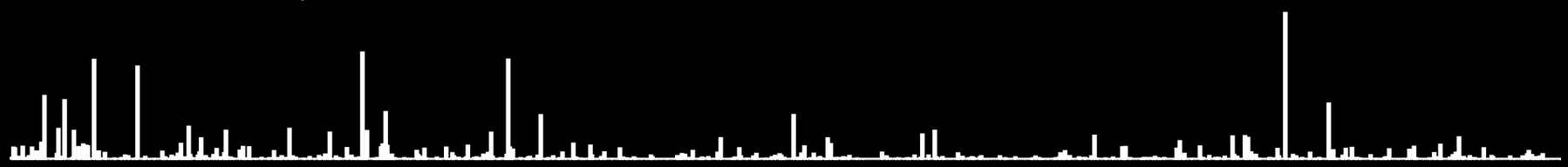


$English$

$p(Chinese|English)$

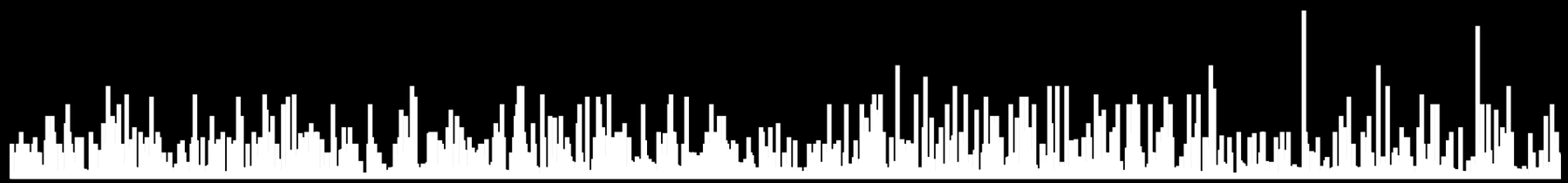


$\times\ p(English)$

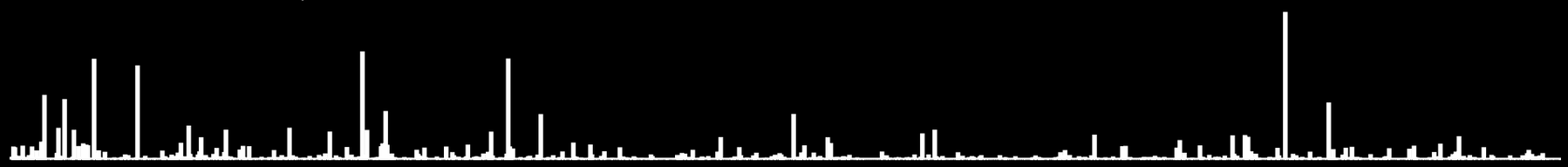


English

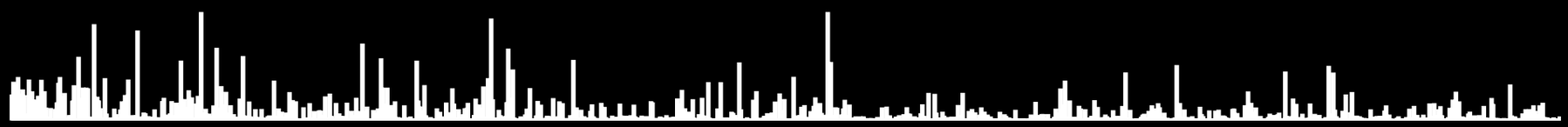
$p(Chinese|English)$



$\times\ p(English)$

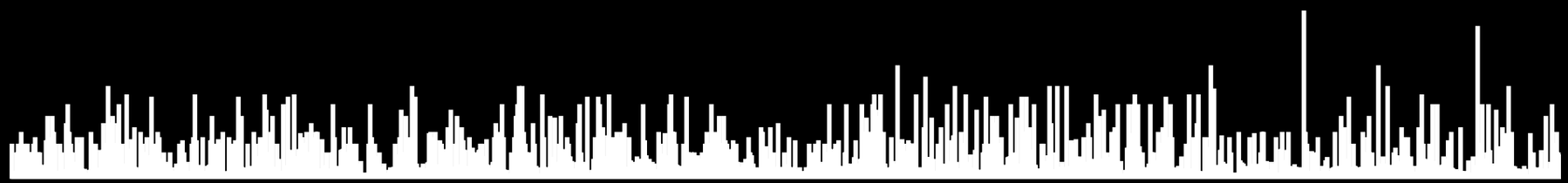


$\sim p(English|Chinese)$

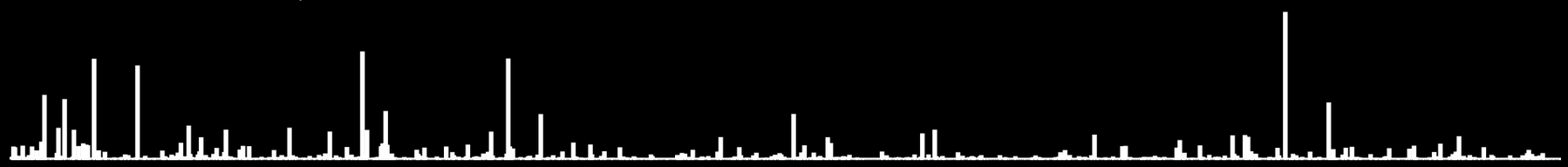


$English$

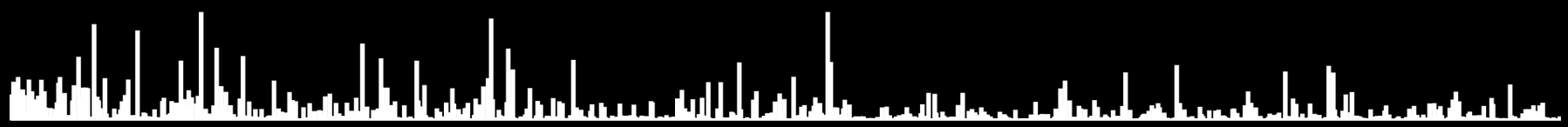
$$p(Chinese|English)^1$$



$$\times\ p(English)^1$$

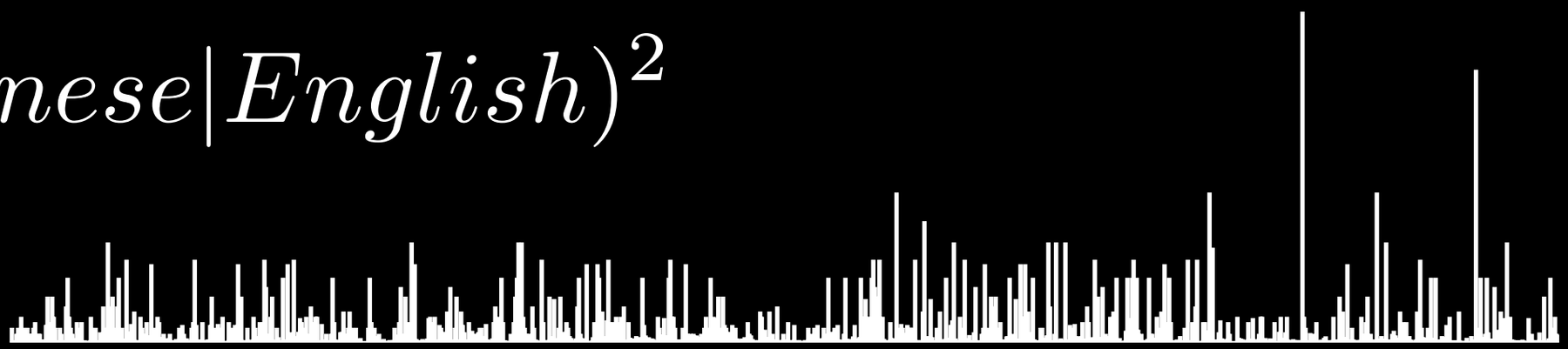


$$\sim p(English|Chinese)$$

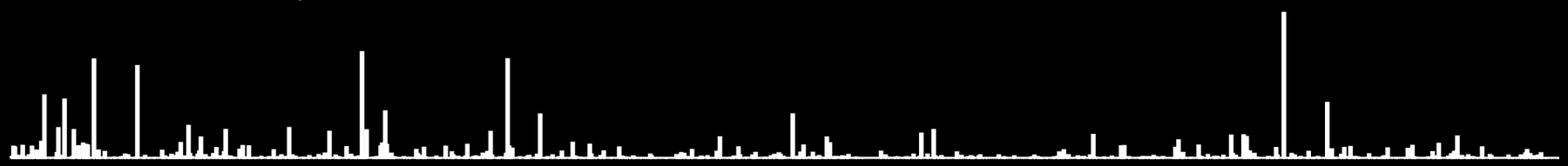


$$English$$

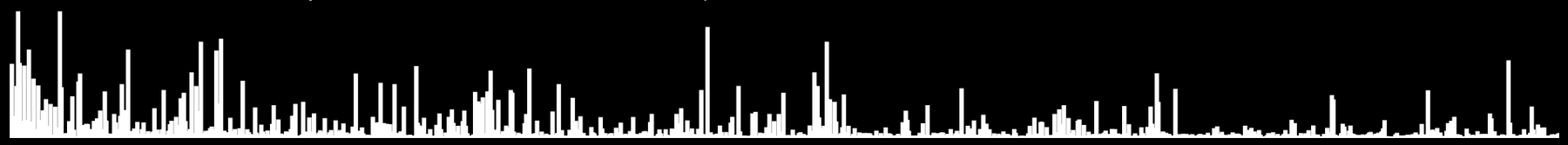
$$p(Chinese|English)^2$$



$$\times\ p(English)^1$$



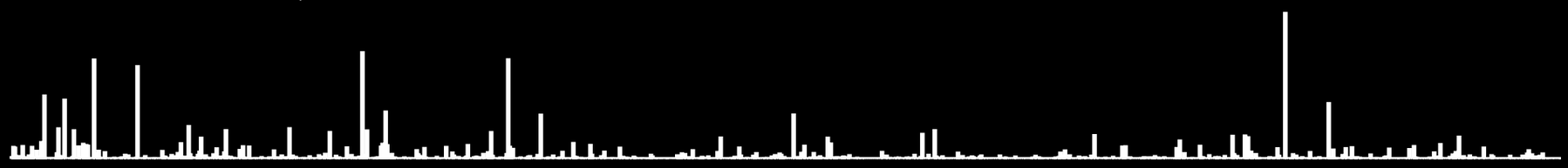
$$\sim p(English|Chinese)$$



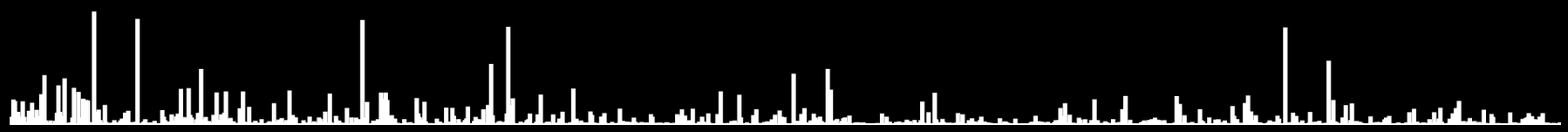
$$English$$

$$p(Chinese|English)^{1/2}$$

$$\times\ p(English)^{1}$$



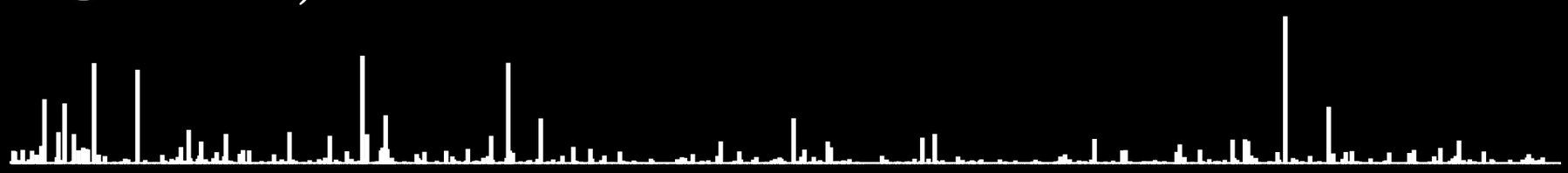
$$\sim p(English|Chinese)$$



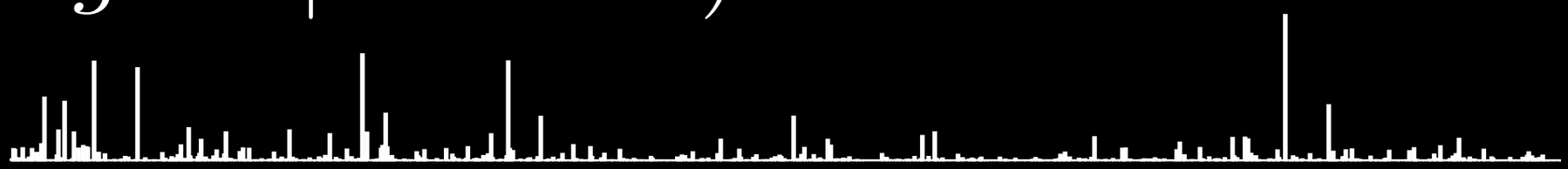
$$English$$

$$p(Chinese|English)^0$$

$$\times\ p(English)^1$$



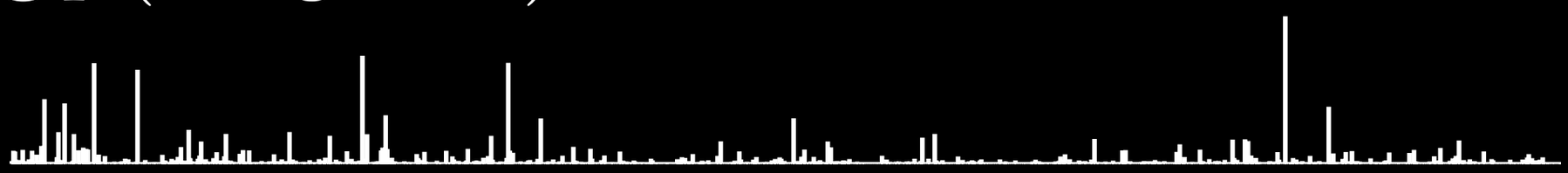
$$\sim p(English|Chinese)$$



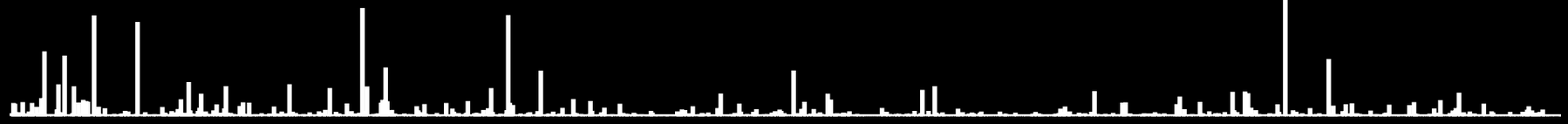
$$English$$

$$0 \cdot \log p(Chinese|English)$$

$$+1 \cdot \log p(English)$$



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



English

$$p(English|Chinese) =$$

$$\frac{1}{Z}\left[\lambda_1 \log p(Chinese|English) + \lambda_2 \log p(English)\right]$$

$$p(English|Chinese) =$$

$$\frac{1}{Z}\left[\sum_i \lambda_i h_i(Chinese, English)\right]$$

$$\underset{English}{\text{argmax}}\ p(English|Chinese) =$$

$$\frac{1}{Z}\left[\sum_i \lambda_i h_i(Chinese, English)\right]$$

log-linear model
conditional random field
case-factor diagram
undirected model

$$\underset{English}{\text{argmax}}\ p(English|Chinese) =$$

$$\frac{1}{Z}\left[\sum_i \lambda_i h_i(Chinese, English)\right]$$

$$\underset{English}{\text{argmax}}\ p(English|Chinese) =$$

$$\frac{1}{Z}\left[\sum_i \lambda_i h_i(Chinese, English)\right]$$

$$\underset{English}{\text{argmax}}\ p(English|Chinese) = \underset{English}{\text{argmax}}\left[\sum_i \lambda_i h_i(Chinese, English)\right]$$

$$\underset{English}{\text{argmax}}\ p(English|Chinese) =$$

$$\left[\sum_i \lambda_i h_i(Chinese, English)\right]$$

linear model

Linear Models

Linear Models

- We could optimize lambdas for likelihood (this would be a log-linear model).

Linear Models

- We could optimize lambdas for likelihood (this would be a log-linear model).
- Good news: optimization is convex.

Linear Models

- We could optimize lambdas for likelihood (this would be a log-linear model).
- Good news: optimization is convex.
- Bad news: computing Z is intractable.

Linear Models

- We could optimize lambdas for likelihood (this would be a log-linear model).
- Good news: optimization is convex.
- Bad news: computing Z is intractable.
- Question: why should we bother with likelihood?

$$BLEU(MT\ output)$$

$$BLEU(\text{argmax}\, score(English|Chinese))$$

$$English$$

$$\sum_{Chinese \in Test} BLEU(\underset{English}{\operatorname{argmax}}\, score(English|Chinese))$$

- Ôptimization

Max Blues

Things to Remember

- Probability gives us a well-founded framework in which to explore a wide variety of models.
- We get many tools for learning and prediction.
- We can express many models in terms of weighted languages.

- Optimization

More Information

Related ESSLLI course web page:
http://homepages.inf.ed.ac.uk/alopez/esslli2010.html

Statistical Machine Translation

ADAM LOPEZ
University of Edinburgh

8

Statistical machine translation (SMT) treats the translation of natural language as a machine learning problem. By examining many samples of human-produced translation, SMT algorithms automatically learn how to translate. SMT has made tremendous strides in less than two decades, and new ideas are constantly introduced. This survey presents a tutorial overview of the state of the art. We describe the context of the current research and then move to a formal problem description and an overview of the main subproblems: translation modeling, parameter estimation, and decoding. Along the way, we present a taxonomy of some different approaches within these areas. We conclude with an overview of evaluation and a discussion of future directions.

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]: Statistical computing; I.2.6 [**Artificial Intelligence**]: Learning—*Parameter learning*; I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Machine translation*; I.5.1 [**Pattern Recognition**]: Models—*Statistical*

General Terms: Algorithms

Additional Key Words and Phrases: Natural language processing, machine translation

ACM Reference Format:
Lopez, A. 2008. Statistical machine translation. ACM Comput. Surv., 40, 3, Article 8 (August 2008), 49 pages DOI = 10.1145/1380584.1380586 http://doi.acm.org/10.1145/1380584.1380586

1. INTRODUCTION

Machine translation (MT) is the automatic translation from one natural language into another using computers. Interest in MT is nearly as old as the electronic computer—popular accounts trace its modern origins to a letter written by Warren Weaver in 1949, only a few years after ENIAC came online.[1] It has since remained a key application in the field of natural language processing (NLP). A good historical overview is given by Hutchins [2007], and a comprehensive general survey is given by Dorr, Jordan, and Benoit [1999].

[1]This letter is reproduced as Weaver [1955].

This research was supported in part by ONR MURI Contract FCPO.810548265, the GALE program of the Defense Advanced Research Projects Agency, Contract No. HR0011-06-02-001, and the EuroMatrix Project funded by the European Commission (6th Framework Programme). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the author and do not necessarily reflect the view of DARPA. This research was carried out while the author was at the University of Maryland.

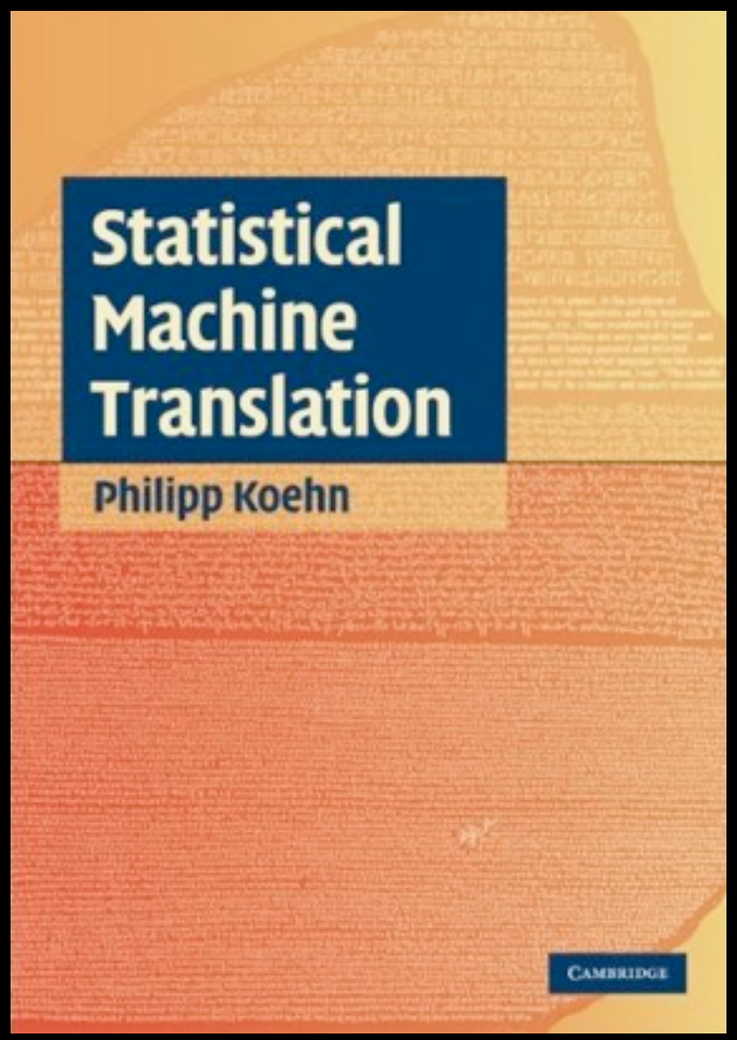
Author's address: A. Lopez, Department of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom; email: alopez@inf.ed.ac.uk.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
©2008 ACM 0360-0300/2008/08-ART8 $5.00. DOI 10.1145/1380584.1380586 http://doi.acm.org/10.1145/1380584.1380586

Statistical Machine Translation.
Adam Lopez. 2008.
In ACM Computing Surveys 40(3).

More Information

Related ESSLLI course web page:
http://homepages.inf.ed.ac.uk/alopez/esslli2010.html



Statistical Machine Translation.
Philipp Koehn. 2010.

THANKS

Necip Fazil Ayan
Phil Blunsom
Ondřej Bojar
Chris Callison-Burch
David Chiang
Trevor Cohn
Bonnie Dorr
Chris Dyer
Martin Kay
Kevin Knight
Philipp Koehn
Nitin Madnani
Christof Monz
Philip Resnik
Michael Subotin