

Introduction to Statistical Machine Translation

Adam Lopez

Including content and jokes borrowed from / inspired
by: Chris Callison-Burch, Philipp Koehn, Kevin Knight,
and Martin Kay

Introduction to Statistical Machine Translation

Adam Lopez

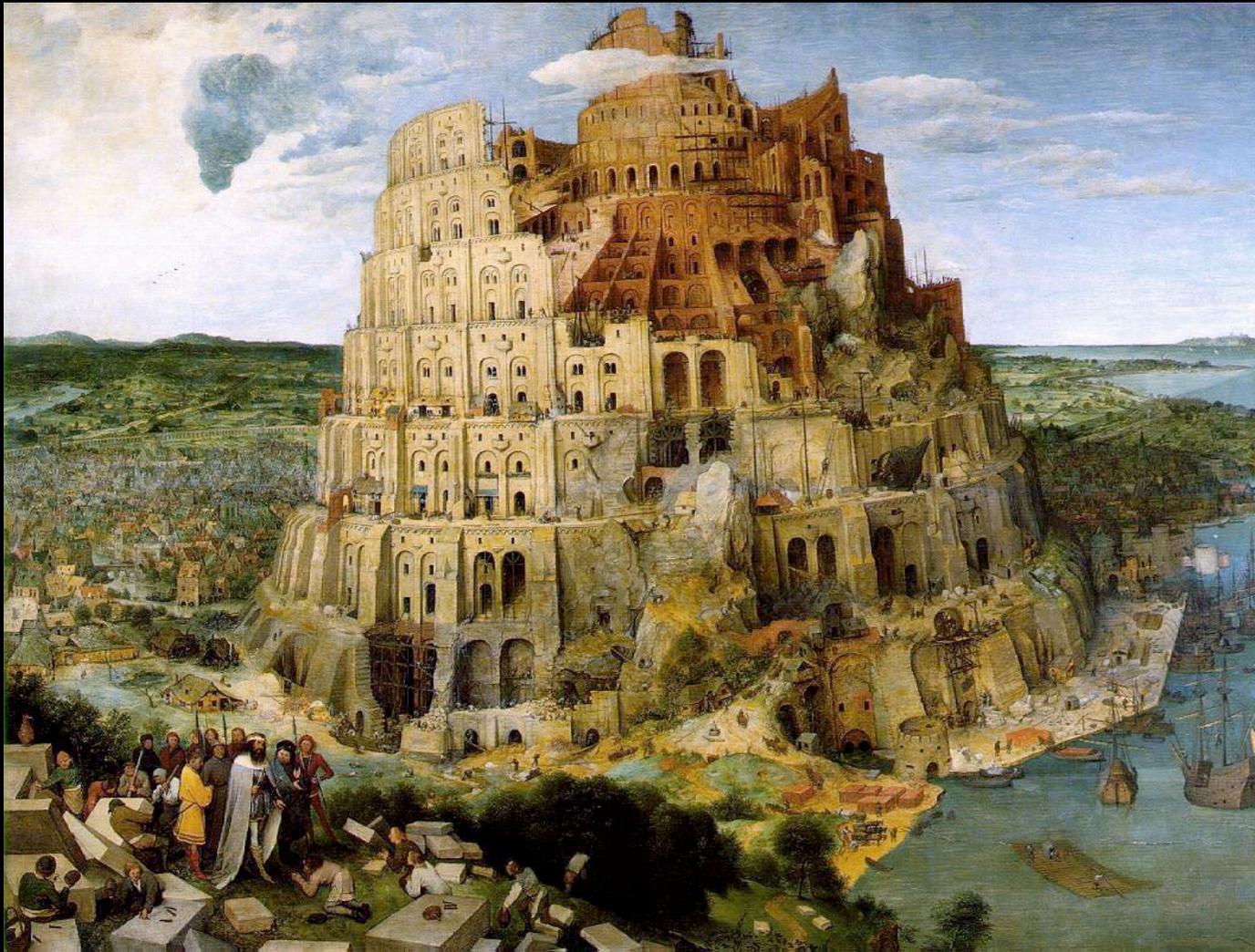
Including content and jokes borrowed from / inspired
by: Chris Callison-Burch, Philipp Koehn, Kevin Knight,
and Martin Kay

Errors and bad timing are all mine, though.

虽然北风呼啸，但天空依然十分清澈。

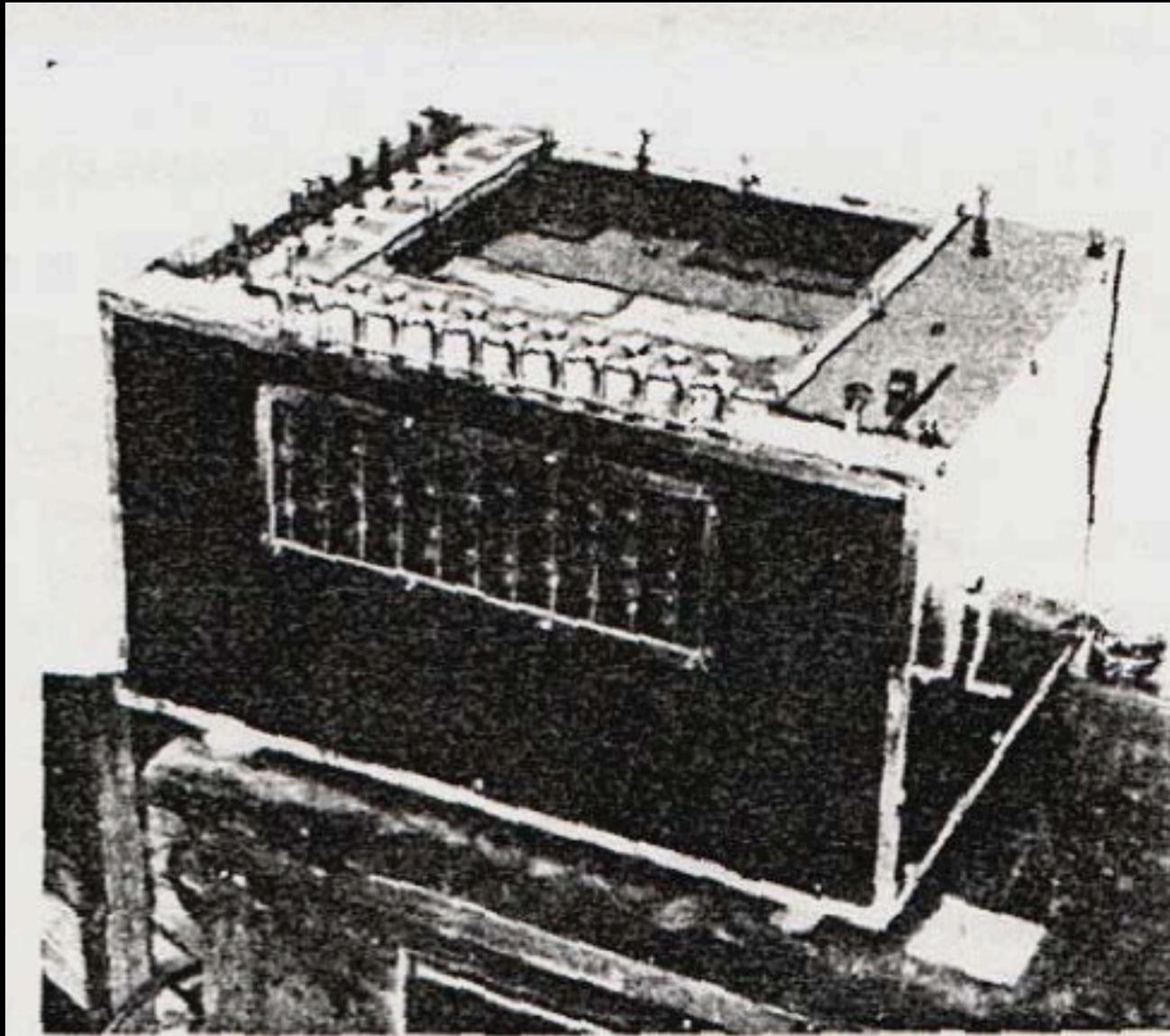
虽然北风呼啸，但天空依然十分清澈。

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



The Tower of Babel

Pieter Bruegel the Elder (1563)



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

COMIT
METAL
Systran
TITUS
SPANAM
ENGSPAN
CUKT
Ariane
GETA
SUSY
AMPAR
NERPA
FRAP
Logos

LUTE
KANT
Candide
SDL
Eurotra
CATALYST
UNITRAN
Rewrite
Pharoah
Moses
Hiero
SAMT
Joshua
Google Translate

Related fields:

artificial intelligence

cognitive science

computer science

linguistics

logic

machine learning

mathematics

philosophy

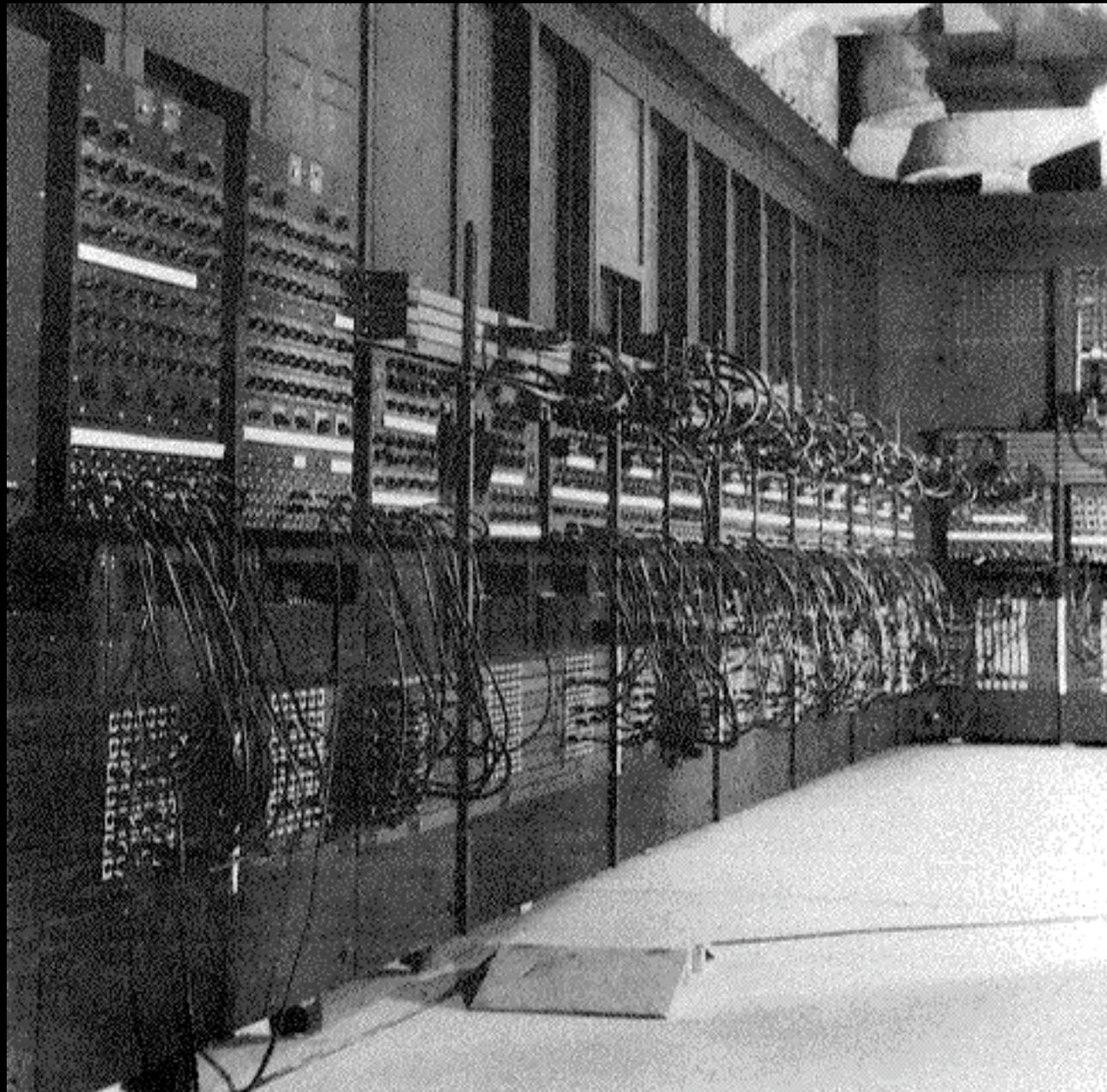
signal processing

statistics

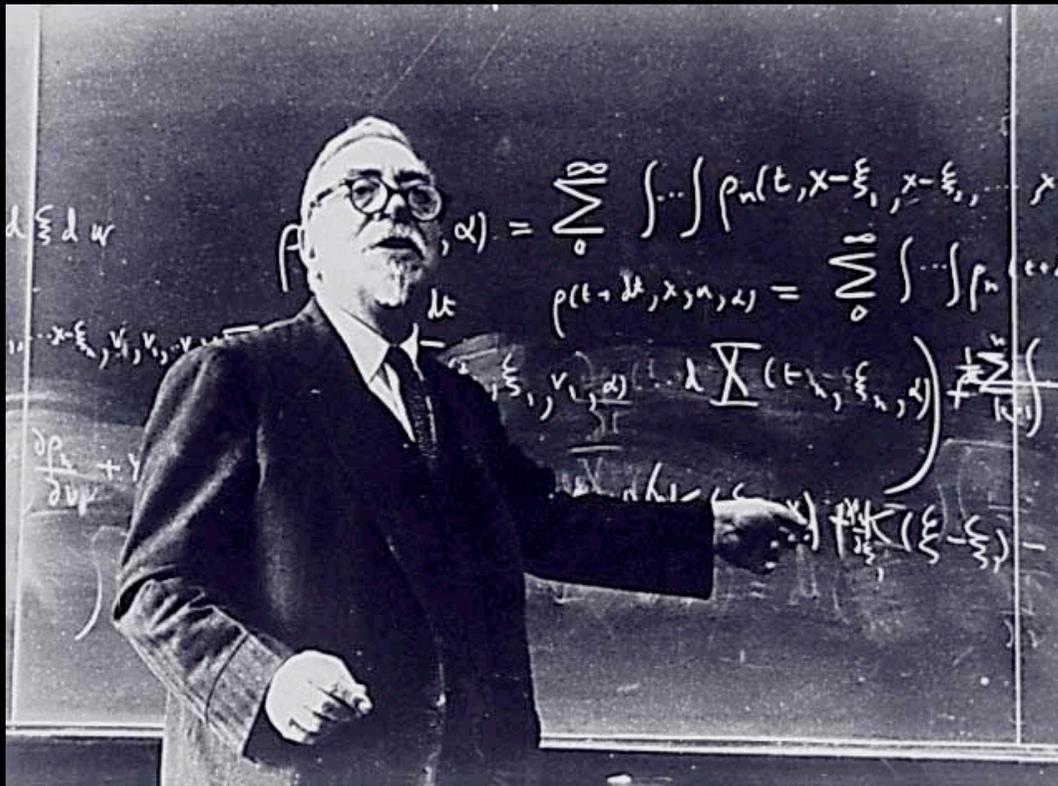
Interlingua-Based MT
Constraint-Based MT
Knowledge-Based MT
Lexical-Based MT
Rule-Based MT
Principle-Based MT
Shake-and-Bake MT
Statistical MT
Example-Based MT
Memory-Based MT
Dialogue-Based MT
Neural-Network-Based MT
Context-Based MT



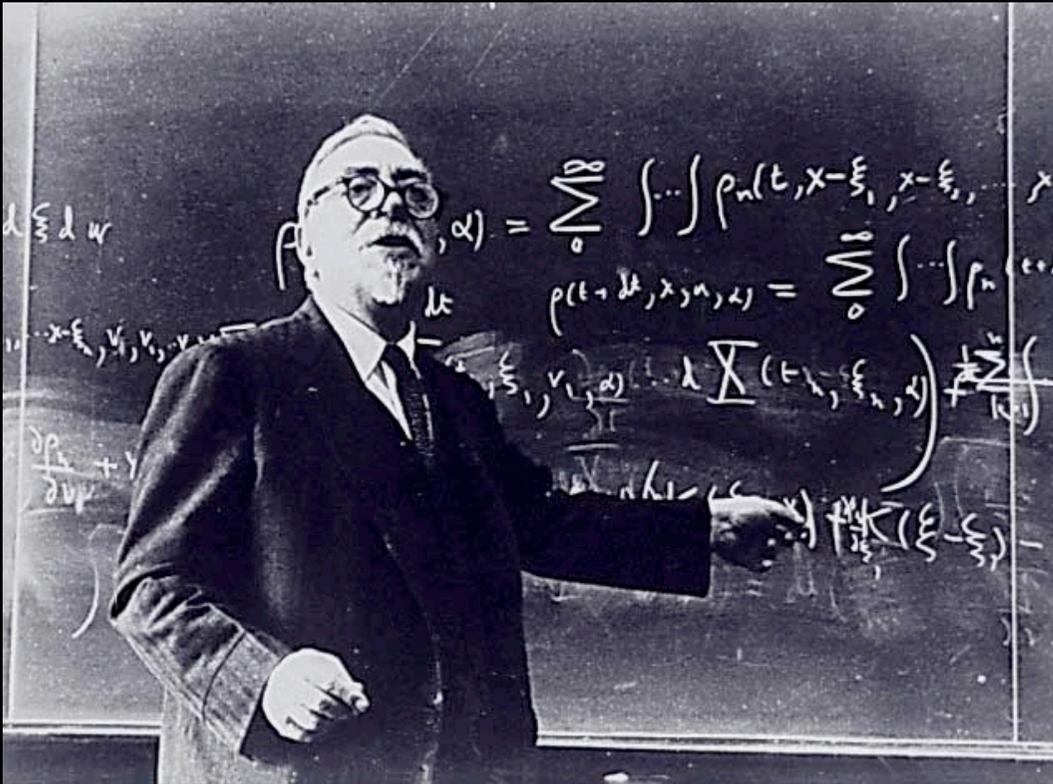
Interlingua-Based MT
Constraint-Based MT
Knowledge-Based MT
Lexical-Based MT
Rule-Based MT
Principle-Based MT
Shake-and-Bake MT
Statistical MT
Example-Based MT
Memory-Based MT
Dialogue-Based MT
Neural-Network-Based MT
Context-Based MT



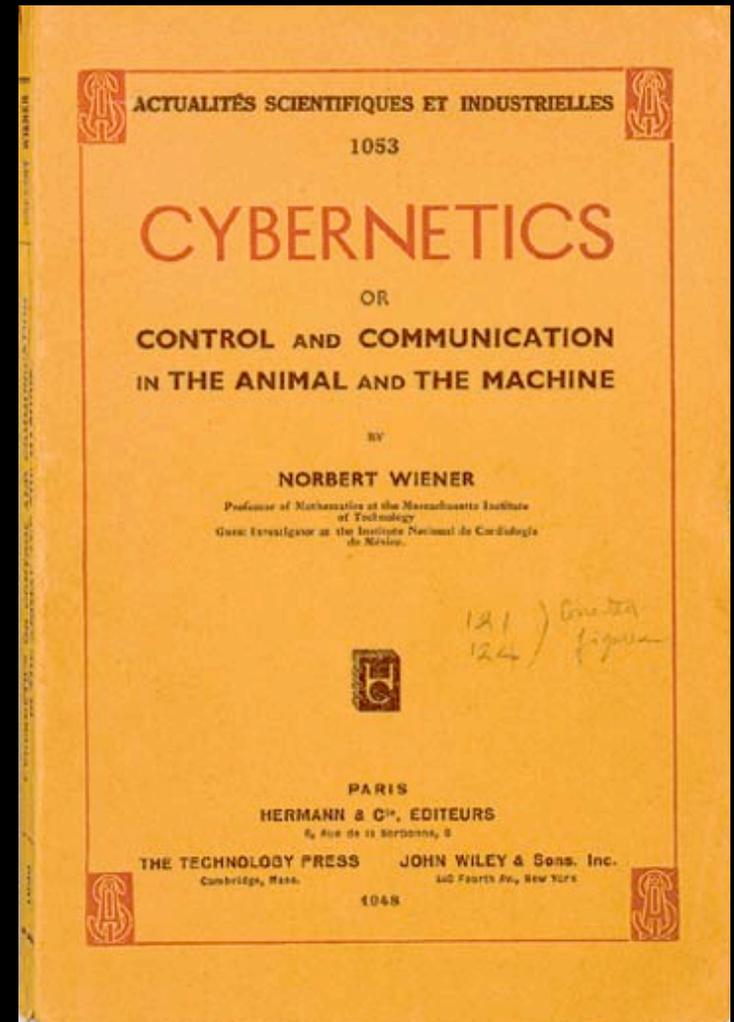
ENIAC (1946)



Norbert Wiener

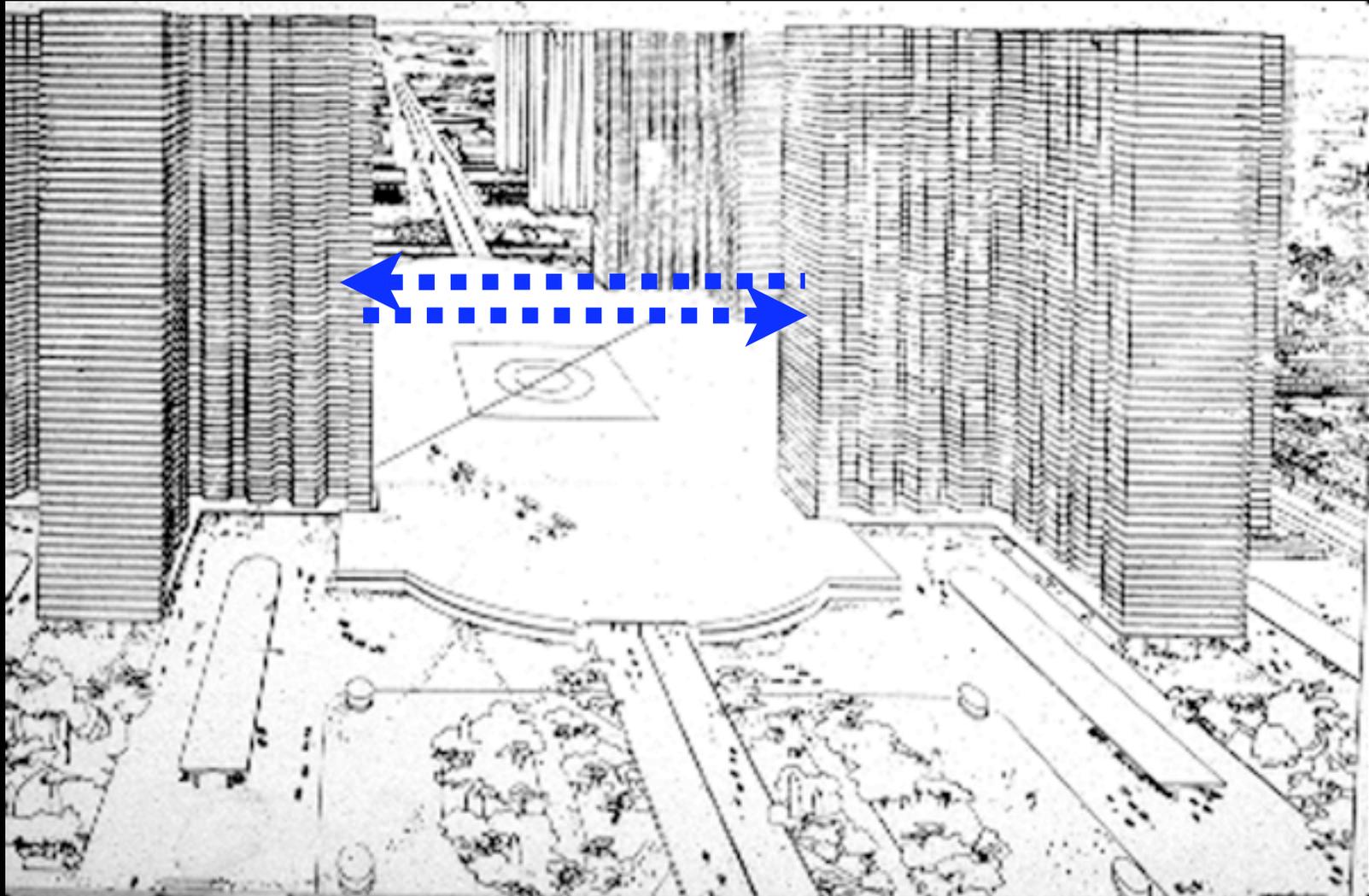


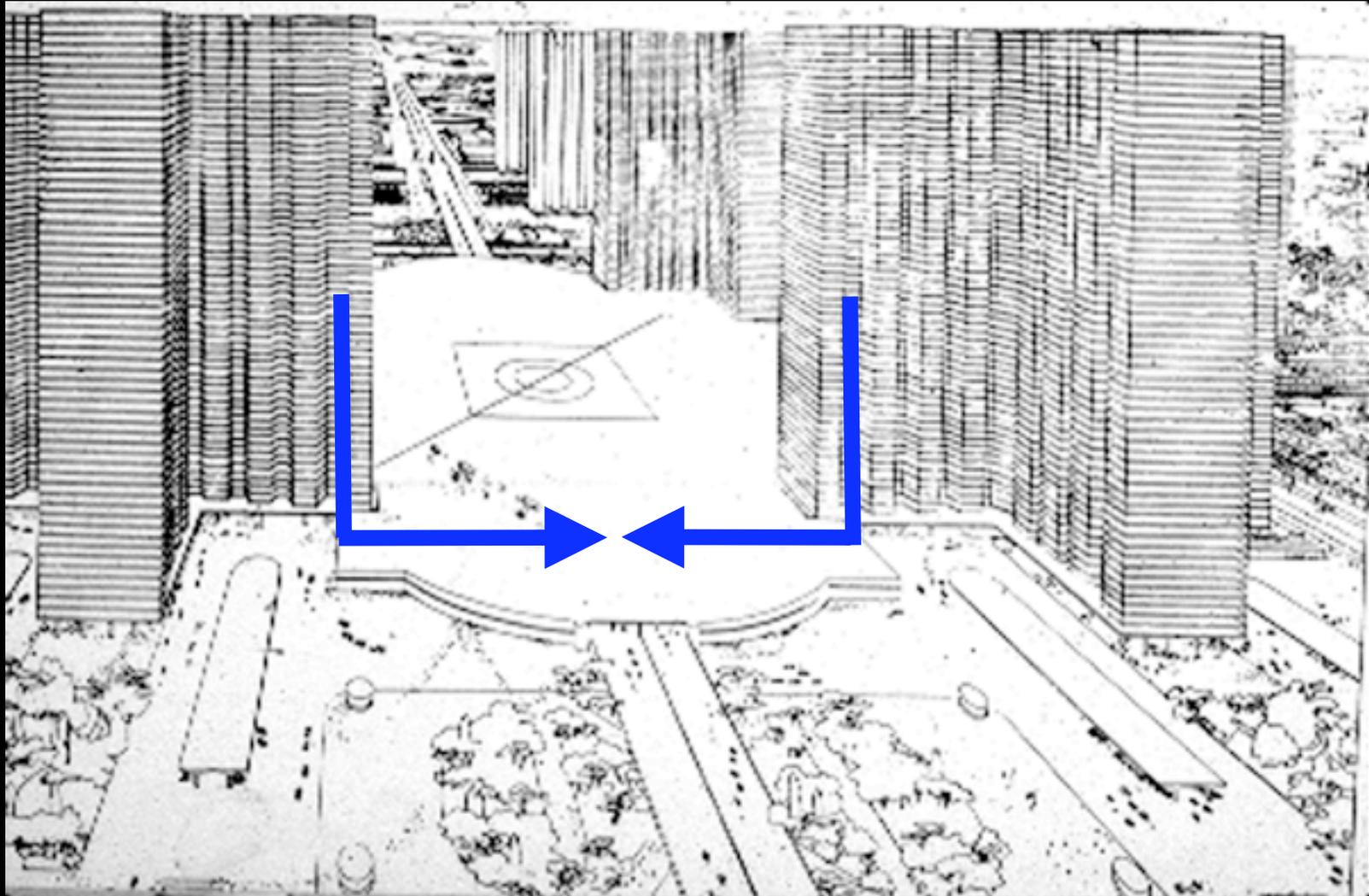
Norbert Wiener





(C) FLC



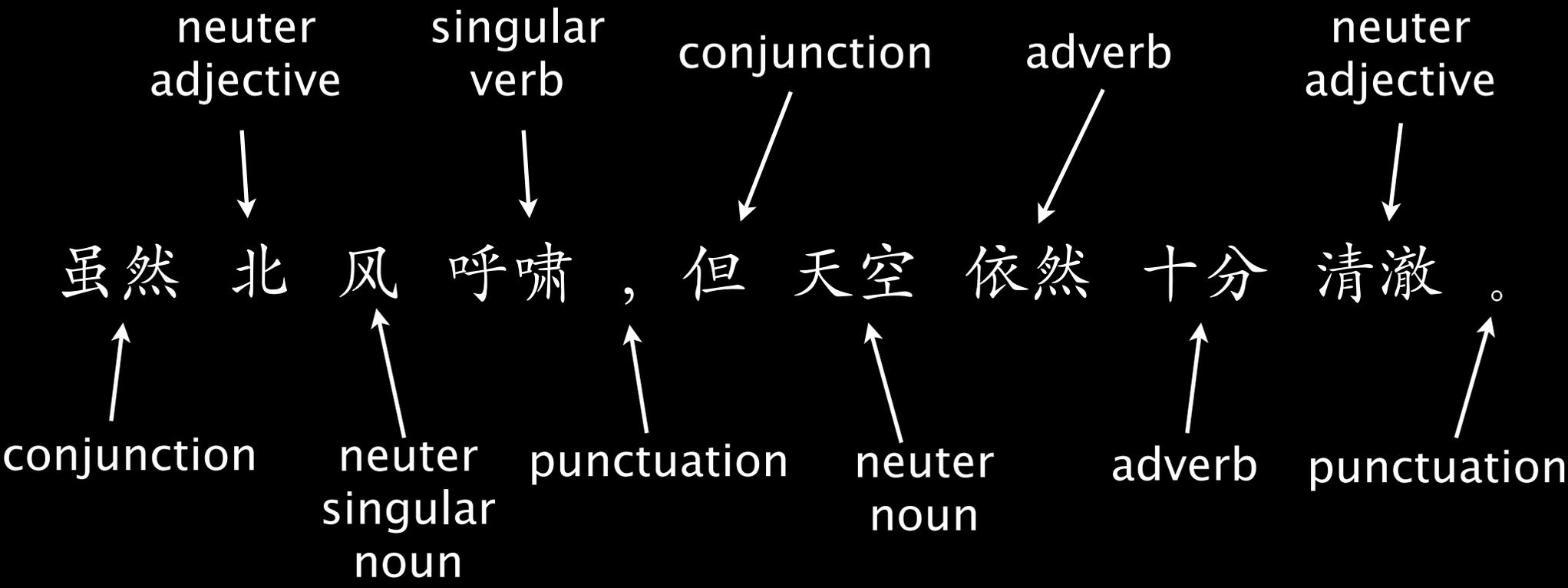


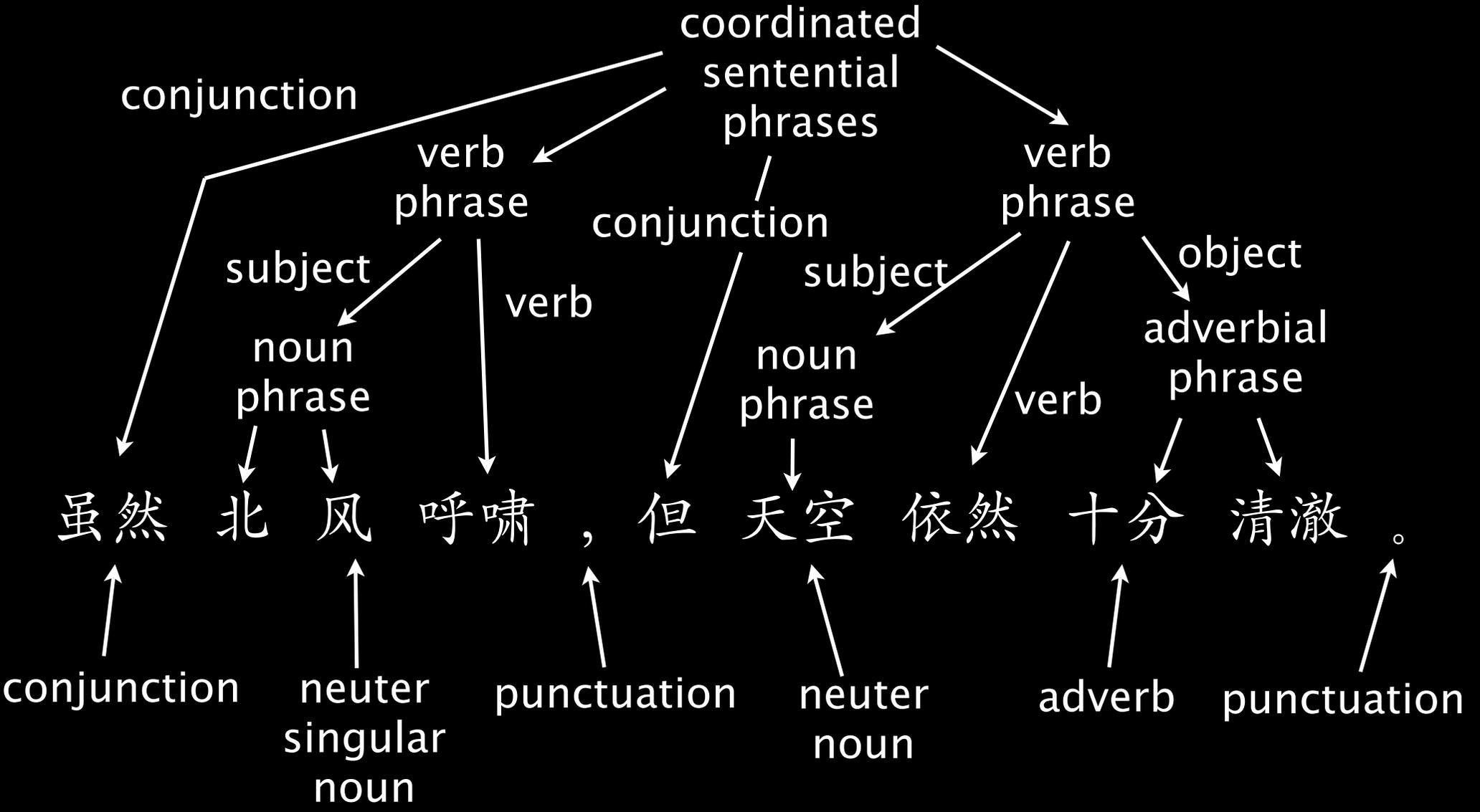
Thus it may be true that the way to translate from Chinese to Arabic, or from Russian to Portuguese, is not to attempt the direct route, shouting from tower to tower. Perhaps the way is to descend, from each language, down to the common base of human communication — the real but as yet undiscovered universal language — and — then re-emerge by whatever particular route is convenient.

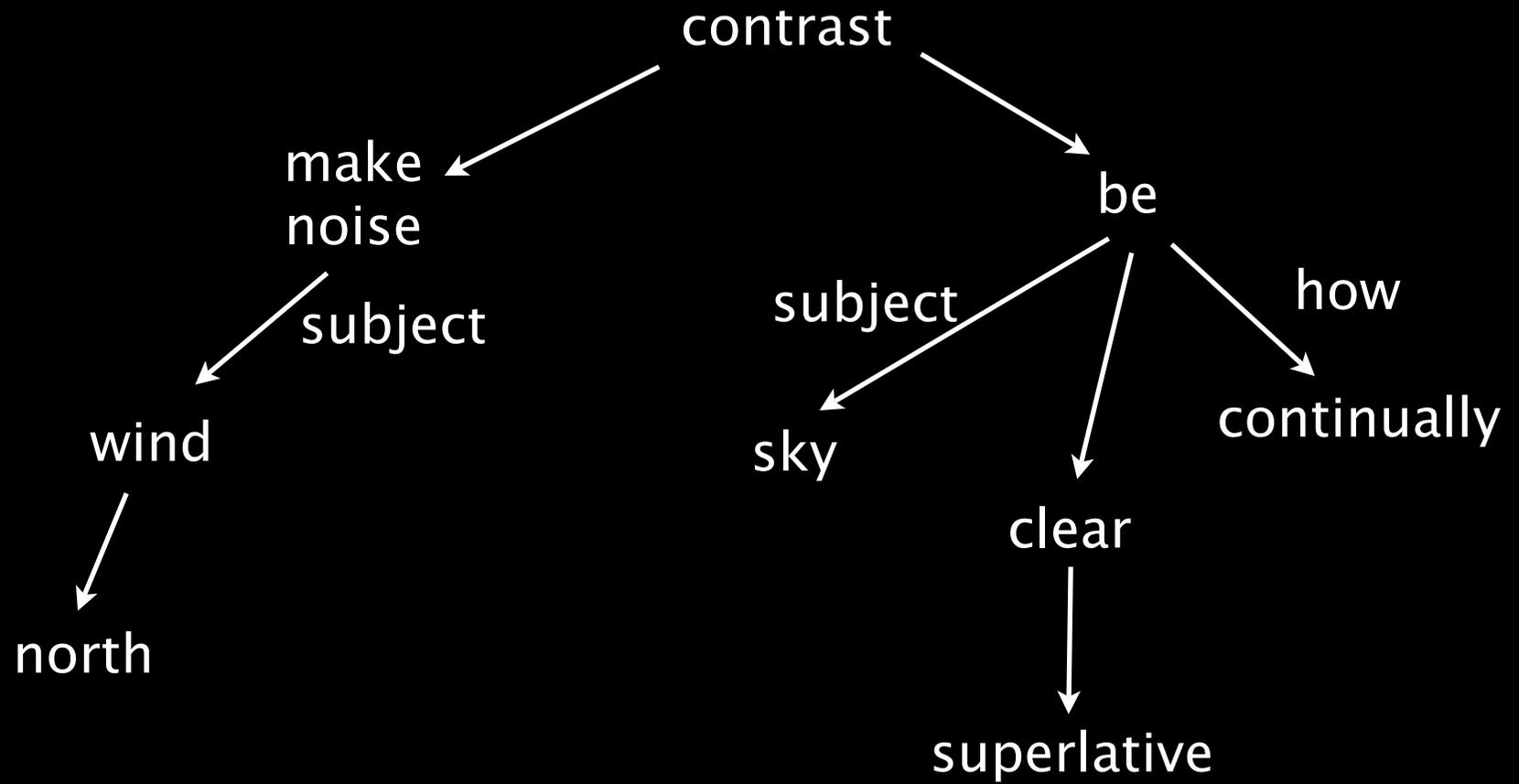
Warren Weaver

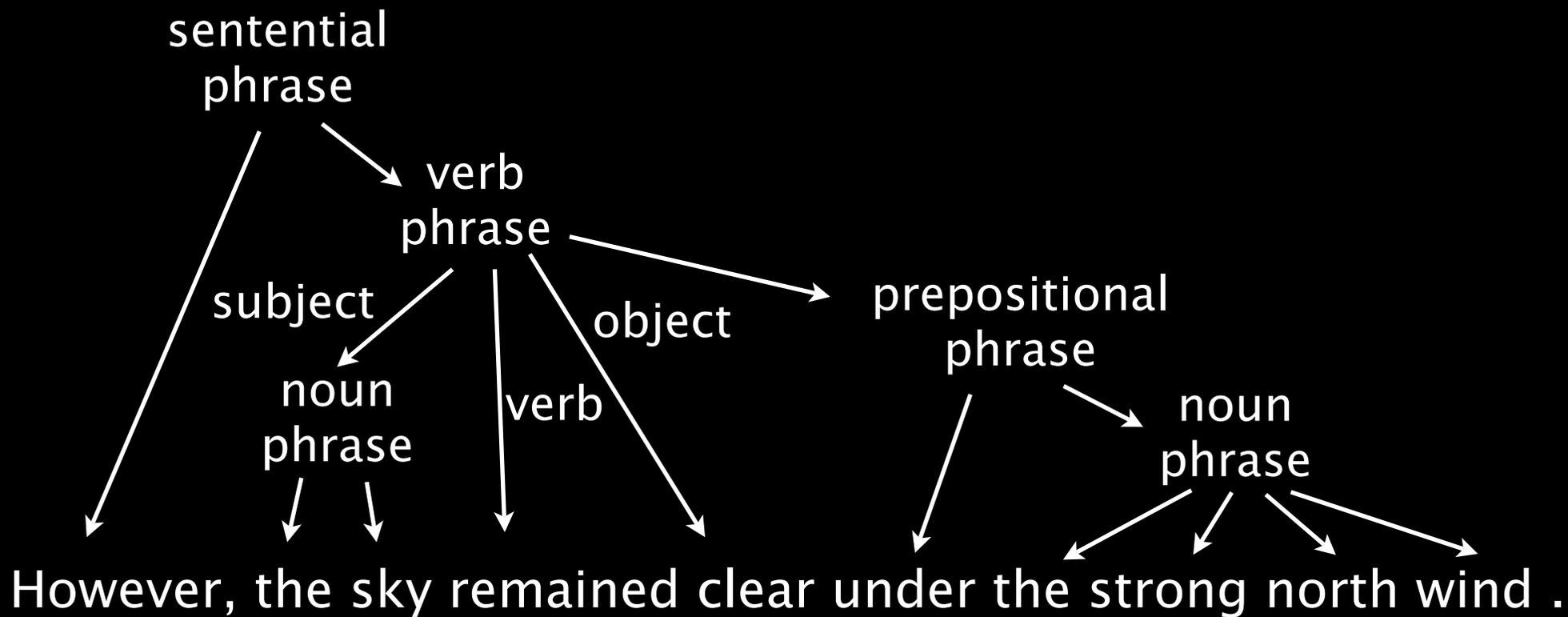


虽然北风呼啸，但天空依然十分清澈。









NINGSUKOL ang duha ka mga tulisan hinungdan nga napusilan hinuon sila ug napatay sa mirespon-ding kapulisan sa milabayng gabii.

Ang kalayo nga niga-was gikan sa sombohan sa armas sa suspetsado maoy ngihimong guide sa kapulisan sa ilang pagba-wos hangtud nga ilang napatay ang usa sa suspetsado alas says sa milabayng gabi sa Ybañez Compound, sitio Alaska Barangay Mambaling, Siyudad sa Sugbo.



Language Tools

Search across languages

Type a search phrase in your own language to easily find pages in another language. We'll translate the results for you to

Search for:

My language:

Search pages written in:

Tip: Use [advanced search](#) to restrict your search by language and country without translating your search phrase.

Translate text





Language Tools

Search across languages

Type a search phrase in your own language to easily find pages in another language. We'll translate the results for you to

Search for:

- Arabic
- Bulgarian
- Catalan
- Chinese
- Croatian
- Czech
- Danish
- Dutch
- English
- Filipino**
- Finnish
- French
- German
- Greek
- Hebrew
- Hindi
- Indonesian
- Italian
- Japanese
- Korean

English

Search pages written in:

Spanish

Translate and Search

[Advanced search](#) to restrict your search by language and country without translating your search phrase.

Filipino



English

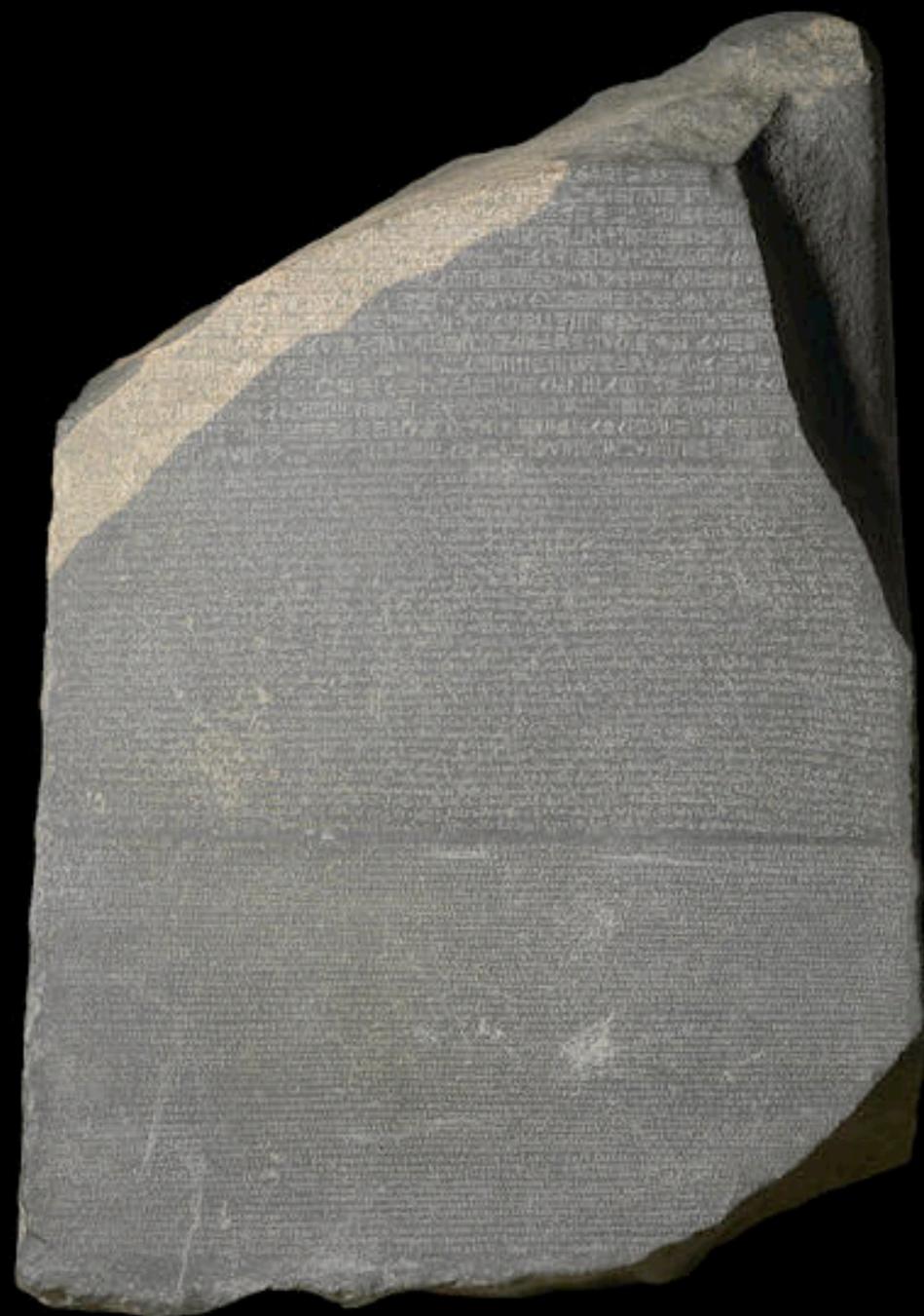
Translate



Statistical Machine Translation Live

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

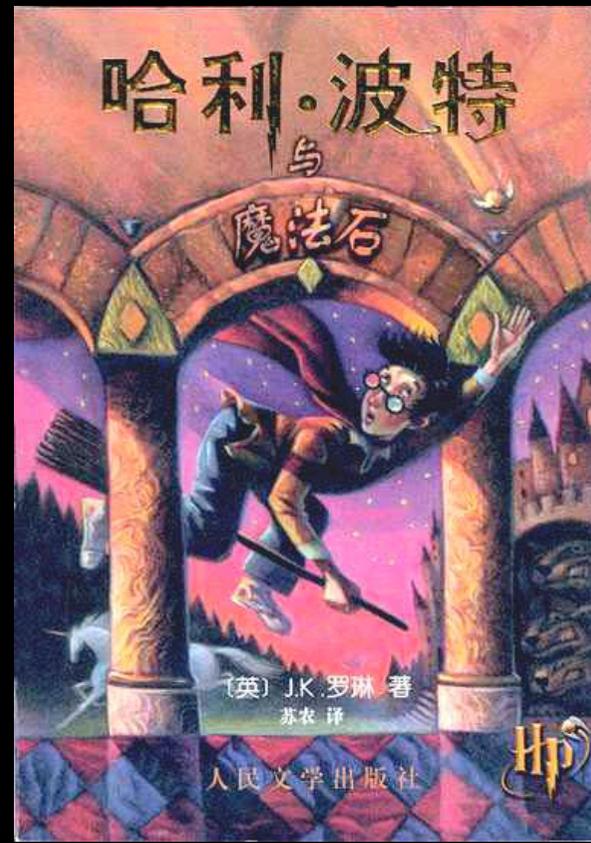
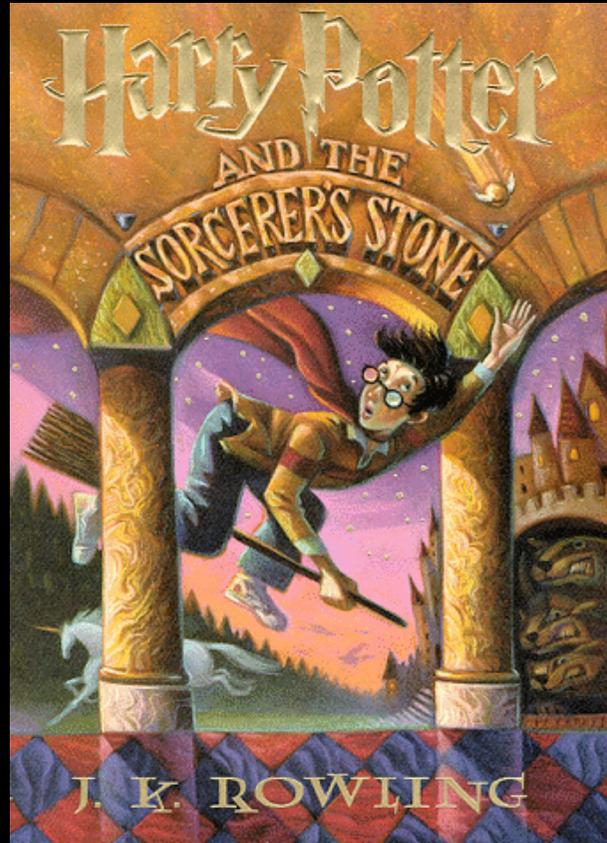
Because we want to provide everyone with access to all the world's information, including information written in every language, one of the exciting projects at Google Research is machine translation... Now you can see the results for yourself. We recently launched an online version of our system for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

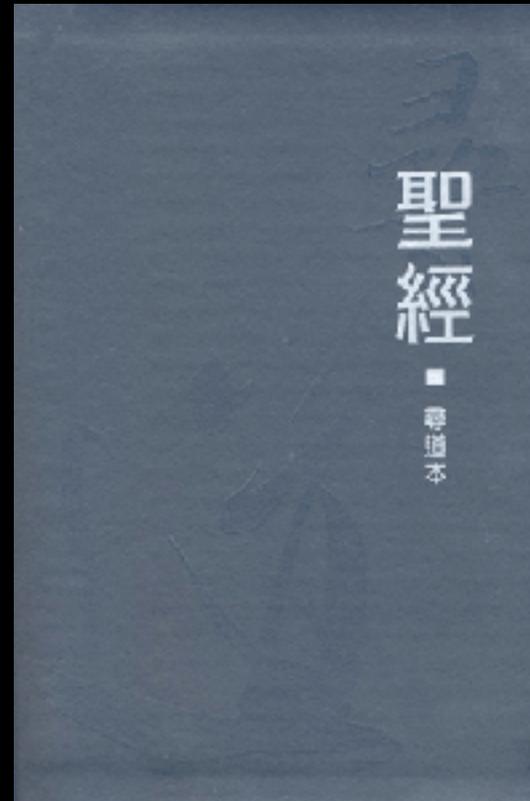
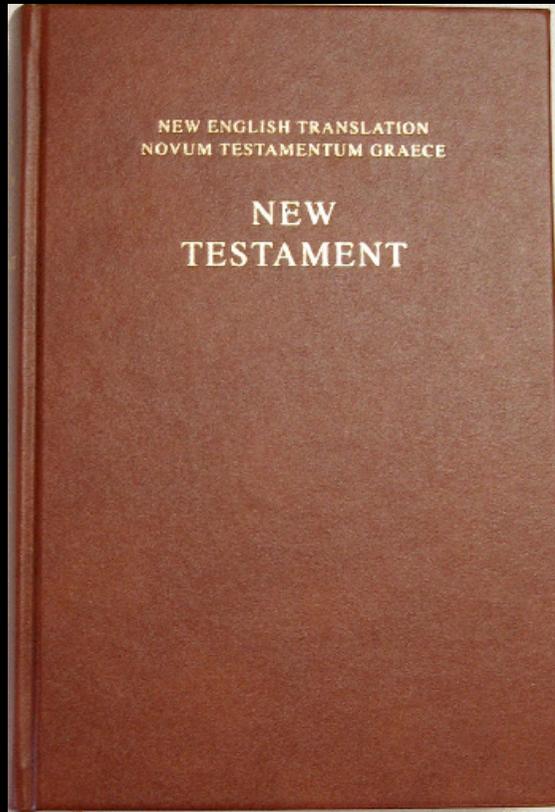


Επιγραφή σε ελληνικό αλφάβητο, γραμμένη σε πολλαπλές στήλες. Το κείμενο είναι αρχαίο ελληνικό, με διάφορα διακριτικά στοιχεία όπως ορθογραφία και σημειώσεις.

Επιγραφή σε ελληνικό αλφάβητο, γραμμένη σε πολλαπλές στήλες. Το κείμενο είναι αρχαίο ελληνικό, με διάφορα διακριτικά στοιχεία όπως ορθογραφία και σημειώσεις.

Επιγραφή σε ελληνικό αλφάβητο, γραμμένη σε πολλαπλές στήλες. Το κείμενο είναι αρχαίο ελληνικό, με διάφορα διακριτικά στοιχεία όπως ορθογραφία και σημειώσεις.





www.un.org

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

We the peoples

Daily Briefing | Radio, TV, Photo Documents, Maps Publications, Stamps, Databases UN Works Search
 Peace & Security | Economic & Social Development Human Rights Humanitarian Affairs International Law

Welcome to the United Nations

UN Millennium Development Goals

United Nations News Centre

About the United Nations

Main Bodies

Conferences & Events

Member States

General Assembly President

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



8 September 2005 >>

Home Recent Additions Employment UN Procurement Comments Q & A UN System Sites Index
 عربي 中文 English Français Русский Español

Copyright, United Nations, 2000-2005 | Use of UN60 Logo | Terms of Use | Privacy Notice | Help [Text version]

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

联合国主页

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

我们人民

每日简报 | 多媒体 文件与地图 出版物 邮票 数据库 服务全球 网址搜索
 和平与安全 | 经济与社会发展 人权 人道主义事务 国际法

欢迎来到联合国

联合国千年发展目标

联合国新闻

联合国概况

联合国主要机关

会议与活动

联合国会员国

联合国大会主席

联合国秘书长

伊拉克局势

中东路线图

更新联合国

反恐主义

联合国日常议题

民间团体/商业

联合国网络直播

空中校车



联大第60届会议一般性辩论

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> Google

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 Censu
 International Symposium on Statistical Cooperation
 18th Roundtable On Business Survey Frames

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)

Producers' Price Index For Manufactured Goods
 Aug Oct Dec Feb Apr Jun Aug

Statistical Data
 Monthly Yearly
 Census Others

Related links
 Chinese Version
 Others

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)

中华人民共和国国家统计局
<http://www.stats.gov.cn/> Google

中文简体 中文繁体 ENGLISH
www.stats.gov.cn

105年9月18日 星期日 本网中文实名: 统计 中国统计 国家统计局

统计公报
统计数据
 月度数据 年度数据
 专题数据 部门数据
统计分析
 月度信息 分析报告
 专题分析 统计图表
 地方信息 企业景气
统计出版
统计法规
统计管理
 部门统计调查项目
 涉外调查管理
数据直报
统计链接
 国内统计网站 政府网站
 国外统计网站 国际网站

最新统计信息

- 2005年全国早稻总产量比上年减产43万吨 (09.16)
- 8月份“国房景气指数”为101.86 同比下降3.10点 (09.16)
- 1-8月湖南城镇居民人均可支配收入同比增长10.4% (09.16)
- 1-8月甘肃固定资产投资增长17.64% 增幅回落3.41% (09.16)
- 株洲: 商品房预售制度对房地产市场的影响浅析 (09.16)
- 经济全球化对江西国民经济发展产生六大影响 (09.16)
- 统计数据: 8月份工业产品产量 各地区产品销售率 (09.15)
- 统计数据: 8月份工业增加值 各地区工业增加值 (09.15)
- 1-8月份全国城镇固定资产投资同比增长27.4% (09.15)
- 加快云南人口城市化进程需解决四大关键问题 (09.15)
- 丹江口: 遏止教育乱收费“一费制”深入人心 (09.15)
- 1-8月浙江规模以上固定资产投资同比增长18.4% (09.15)
- 8月份广西消费品零售额与去年同期相比增长13.6% (09.15)
- 8月份我国工业实现增加值5968亿 同比增长16% (09.14)
- 调查显示: 广东省企业流动资金短缺问题日益突出 (09.14)
- 实施品牌战略 推动吉林省经济快速可持续发展 (09.14)
- 无锡: 城乡居民收入剪刀差十年扩大0.46倍 (09.14)
- 8月份甘肃工业品价格呈现四特点 波动频率有所加快 (09.14)

工业品出厂价格指数
 10月 12月 2月 4月 6月 8月

重要公告

公告
 关于申报2005年度全国统计科研计划项目的通知
 印发《关于统计上对公有和非公有控股经济的分类办法》的通知

统计机构
统计动态
 专项统计工作 业务资料下载
统计标准
 区划代码 行业分类 企业类型
统计制度
统计知识
 统计研究

最新统计动态

- 河南省县级政府领导干部统计知识培训班举行 (09.16)
- 白银市副市长火统元勉励统计工作者当好“参谋” (09.16)
- 吉林省统计机构建设加强 成立实体统计工作站 (09.16)
- 《新疆五十年》出版发行 自治区领导为该书题词 (09.16)
- 国家统计局局长李德水撰文谈心得: 抢桑巨变话地球 (09.15)
- 湖南省副省长肖捷强调统计要实事求是 (09.15)
- 山东省统计局实行行政许可可办理“窗口”式办公 (09.15)
- 湖南省委副书记谭益民在长沙调研指导统计工作 (09.15)

虽然北风呼啸，但天空依然十分清澈。

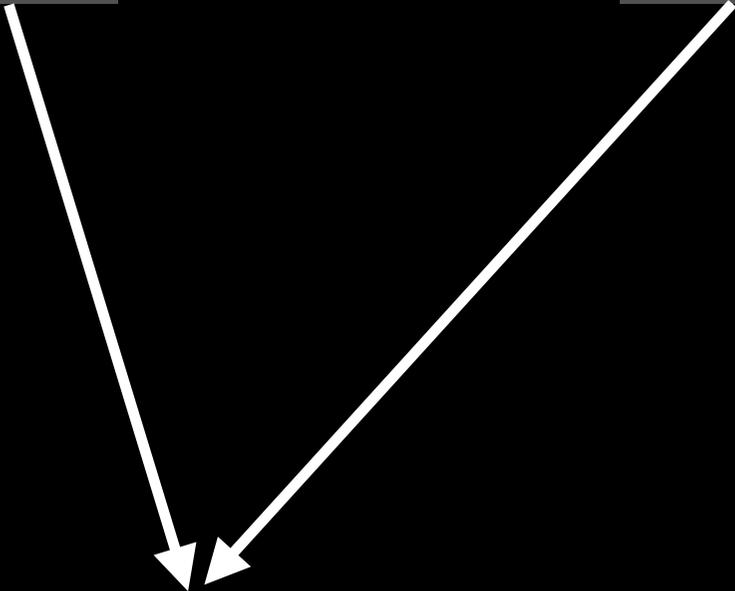
虽然北风呼啸，但天空依然十分清澈。



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

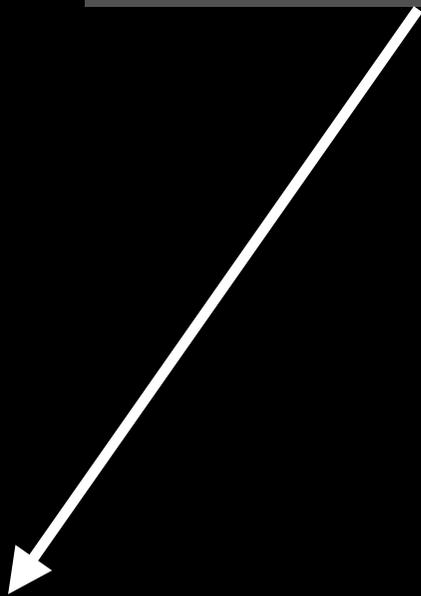
虽然北风呼啸，但天空依然十分清澈。

虽然 北 风 呼 啸 ， 但 天 空 依 然 十 分 清 澈 。



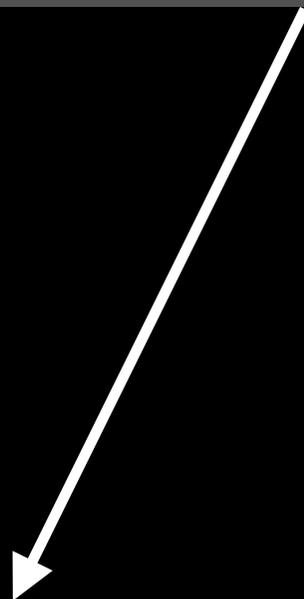
However,

虽然 北风呼啸，但天空 依然 十分 清澈。



However, the sky

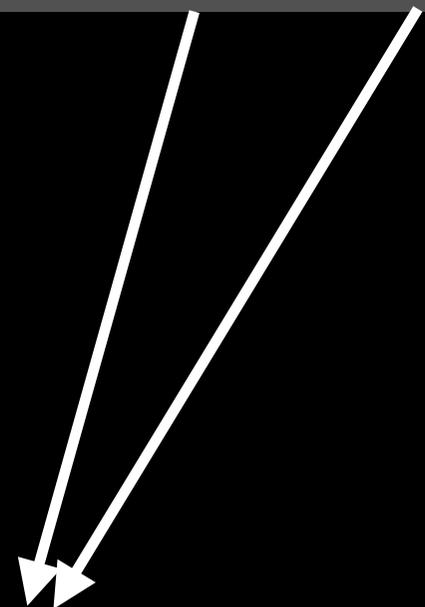
虽然 北风呼啸，但天空依然十分清澈。



However, the sky remained

虽然北风呼啸，但天空依然十分清澈。

However, the sky remained 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 .

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 .

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	吞	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 吞	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	吞	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		





$p(\textit{heads}) ?$

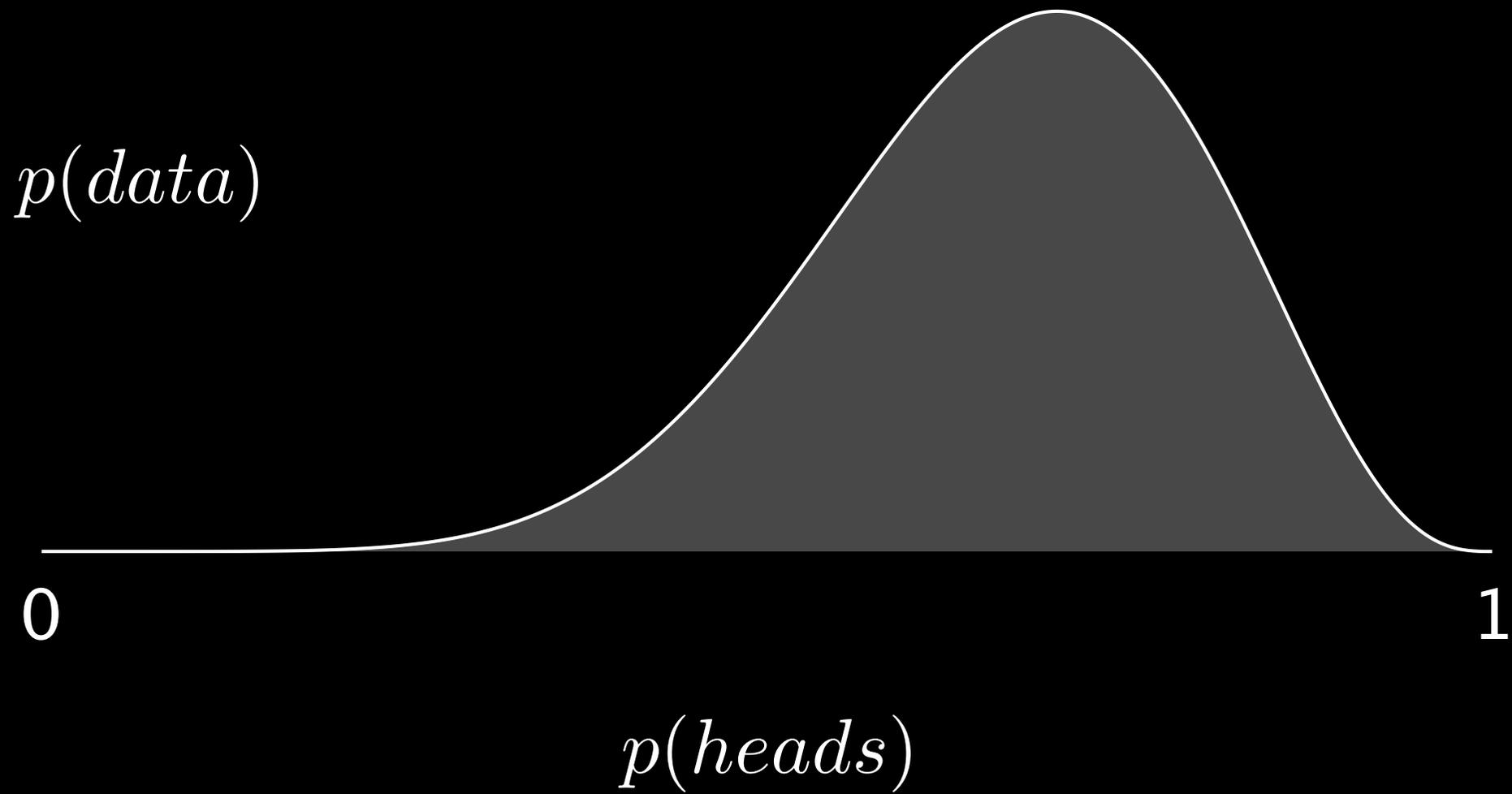


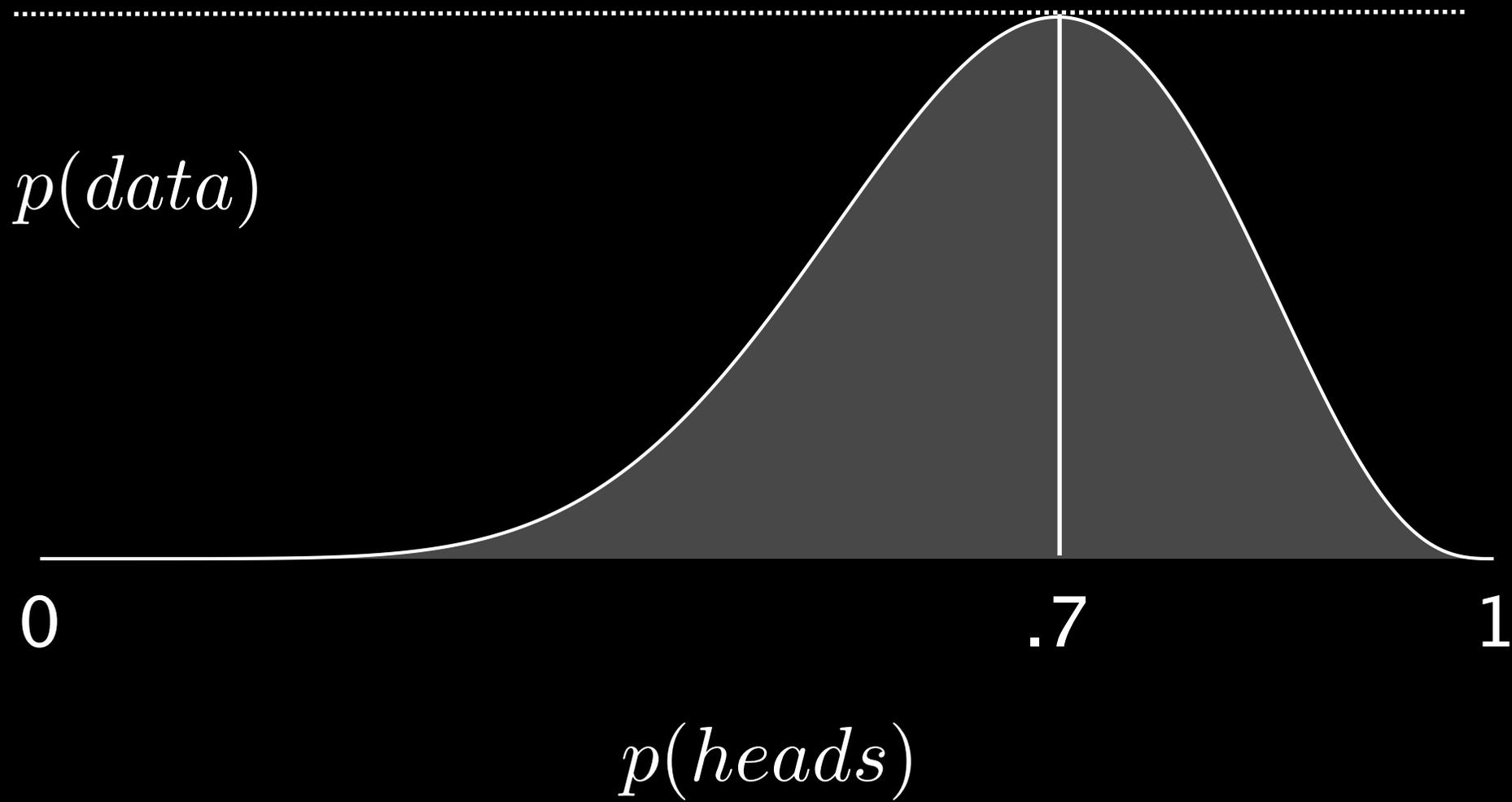


$$p(\text{data}) = p(\text{heads})^7 \times p(\text{tails})^3$$

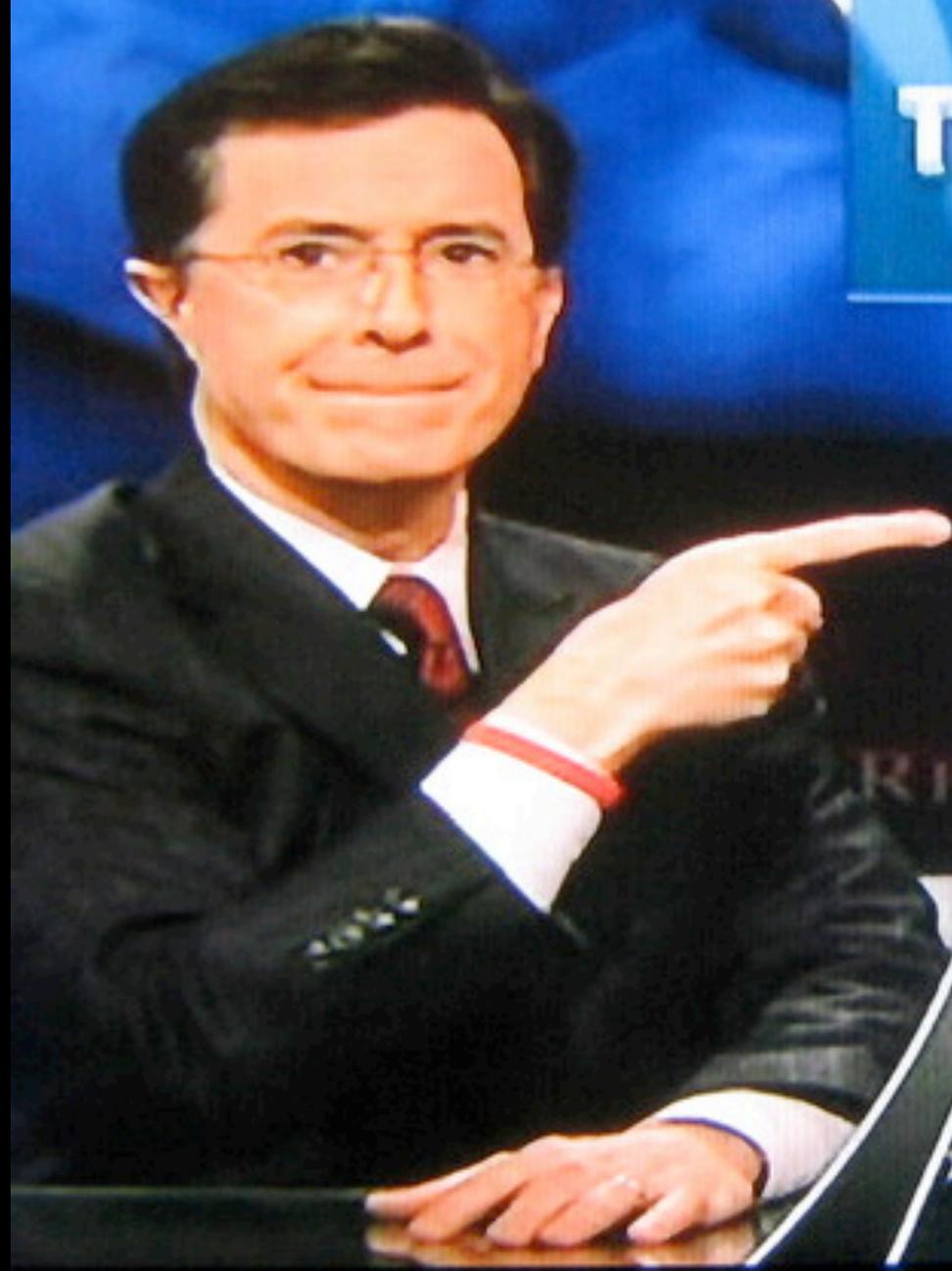


$$p(\text{data}) = p(\text{heads})^7 \times [1 - p(\text{heads})]^3$$





THE ~~W~~ORD



COM
CENT



THE WORD

Optimization



$p(\textit{heads}) ?$

$p(\textit{heads}) ?$

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 .

虽然 北 风 呼 啸 ， 但 天 空 依 然 十 分 清 澈 。

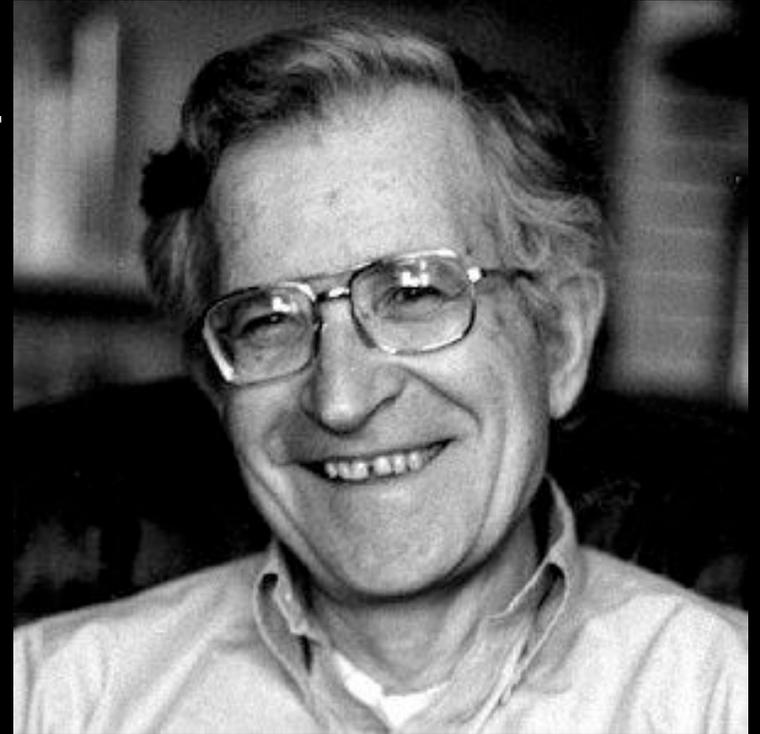
However, the sky remained clear under
the strong north wind .

$$p(\textit{English}|\textit{Chinese}) = \prod_{\textit{rule}} p(\textit{rule})$$

$$\operatorname{argmax}_{\text{English}} p(\text{English}|\text{Chinese}) = \prod_{\text{rule}} p(\text{rule})$$

It must be recognized that the notion of a probability of a sentence is an entirely useless one, under any interpretation of this term.

Noam Chomsky

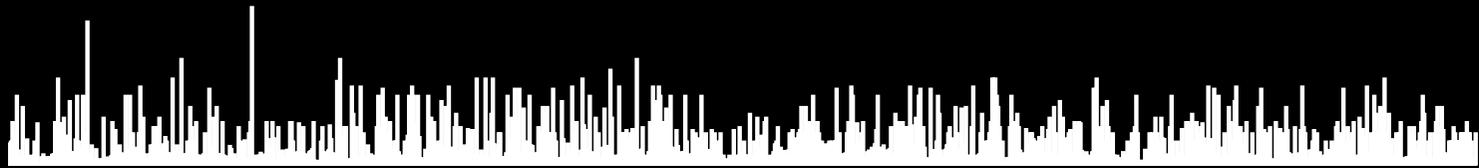


Every time I fire a linguist,
the performance of the
recognizer goes up.

Fred Jelinek (1988)



$p(\textit{English}|\textit{Chinese})$

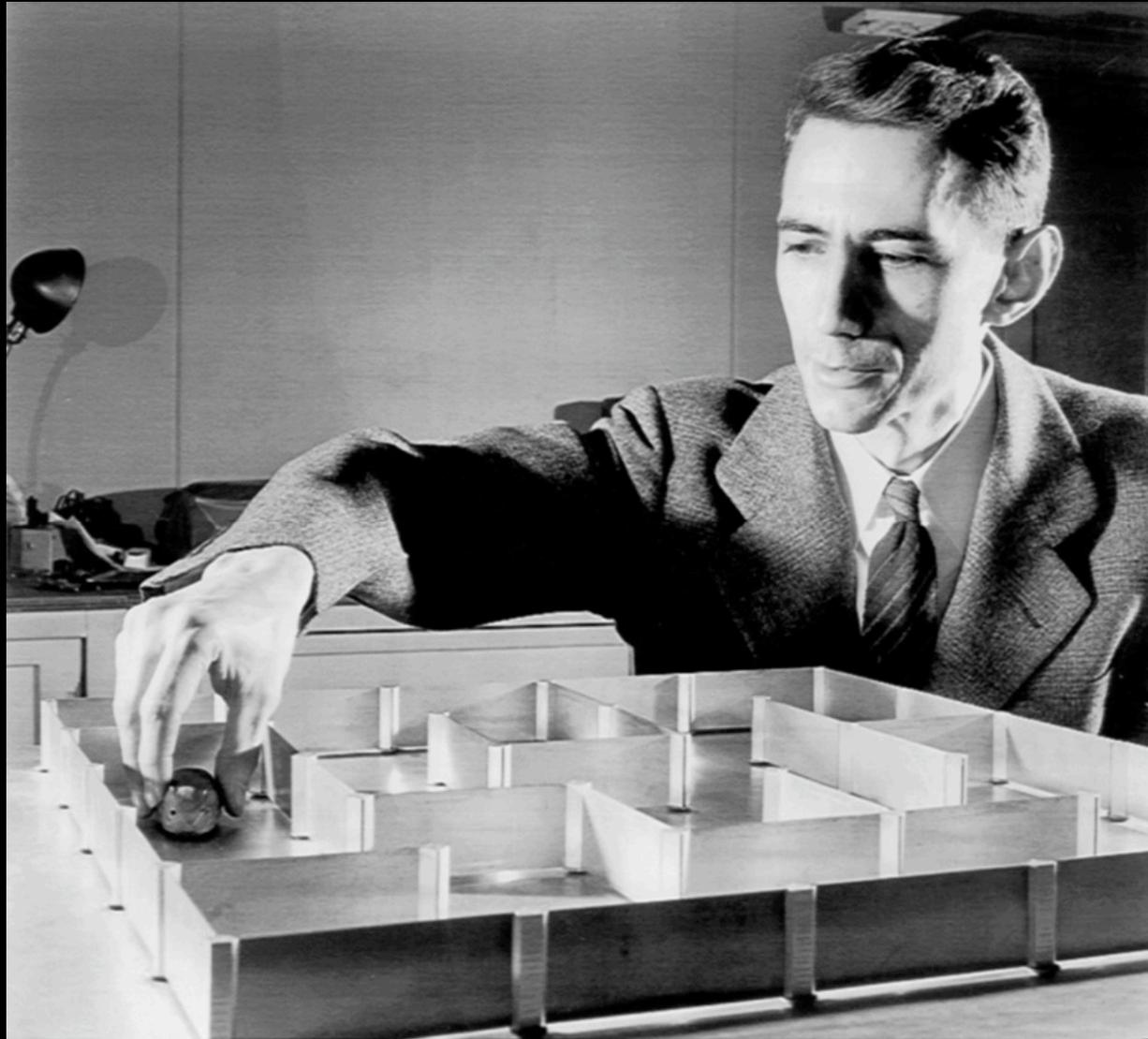


English



THE MATHEMATICAL THEORY OF COMMUNICATION

by Claude E. Shannon and Warren Weaver



Claude Shannon



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

...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 .



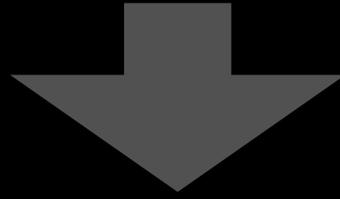
Thomas Bayes

$$p(\textit{English}|\textit{Chinese}) =$$

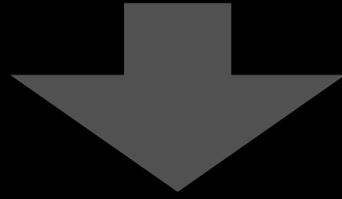
$$\frac{p(\textit{English}) \times p(\textit{Chinese}|\textit{English})}{p(\textit{Chinese})}$$

$$p(\textit{English}|\textit{Chinese}) \sim$$

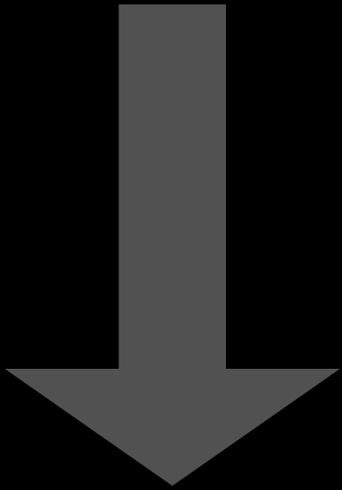
$$p(\textit{English}) \times p(\textit{Chinese}|\textit{English})$$



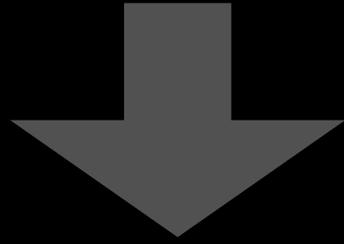
However, the sky remained clear under
the strong north wind .



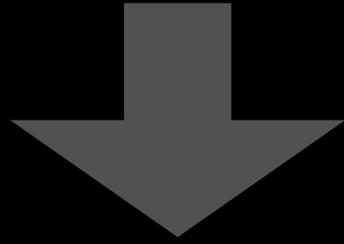
However, the sky remained clear under
the strong north wind .



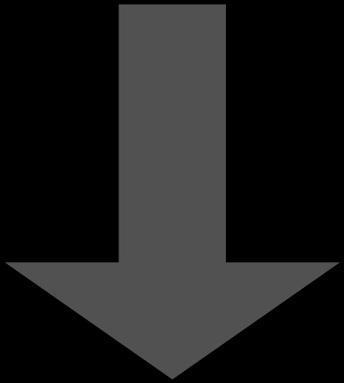
虽然 北 风 呼 啸 ， 但 天 空 依 然 十 分 清 澈 。



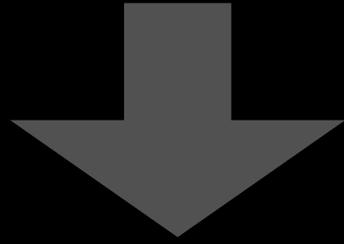
configuration



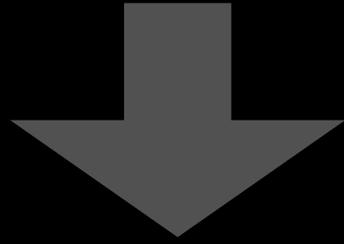
configuration



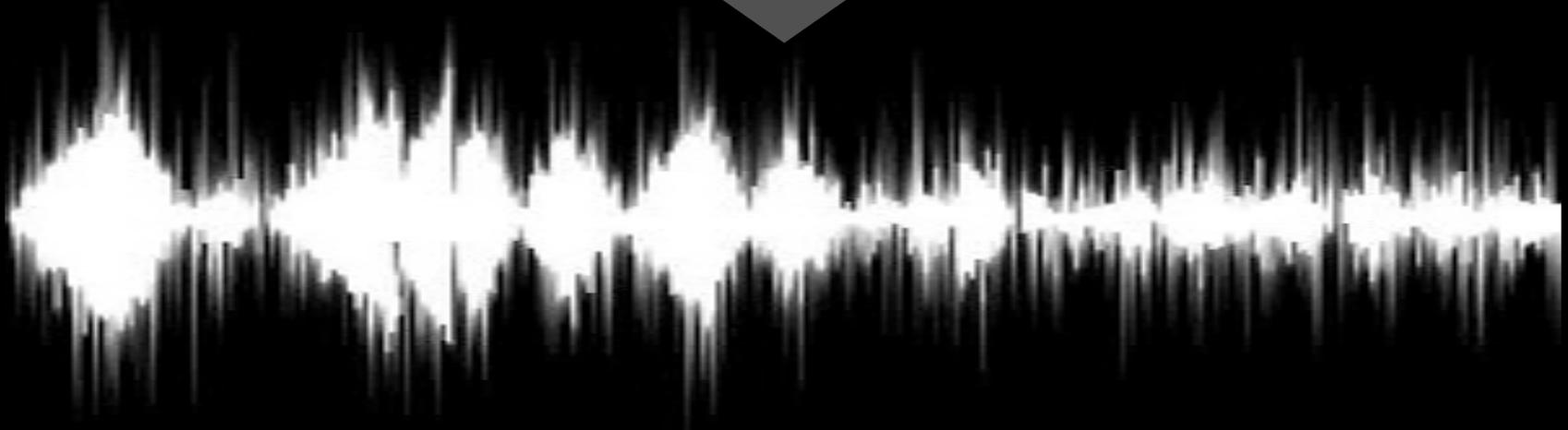
configuration

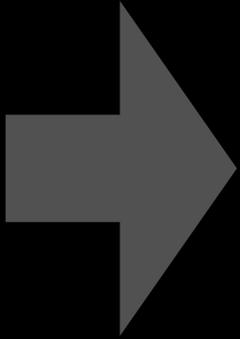


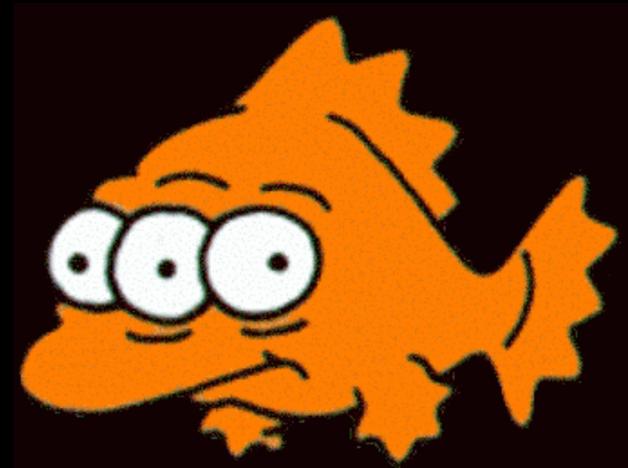
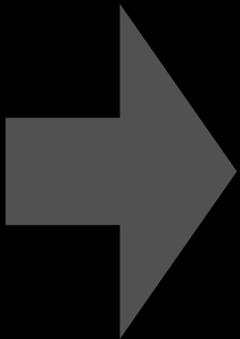
Our enemies are innovative and resourceful, and so are we. They never stop thinking about new ways to harm our country and our people, and neither do we.



Our enemies are innovative and resourceful, and so are we. They never stop thinking about new ways to harm our country and our people, and neither do we.







However

$$p(\textit{However} | \textit{START})$$

However ,

$p(, | \textit{However})$

However , the

$$p(\textit{the} | ,)$$

However , the sky

$$p(\textit{sky}|\textit{the})$$

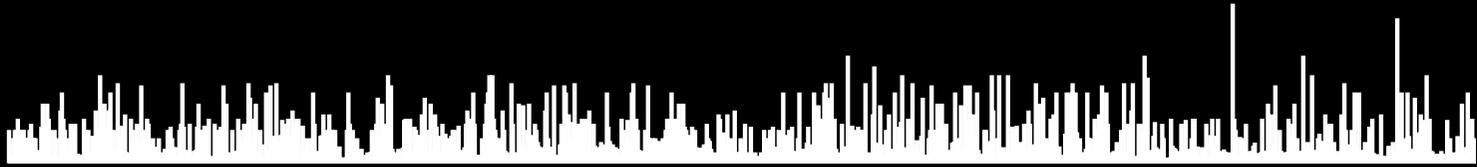
$$p(\textit{English}) = \prod_{i=1}^{\textit{length}(\textit{English})} p(\textit{word}_i | \textit{word}_{i-1})$$

$p(\textit{Chinese}|\textit{English})$

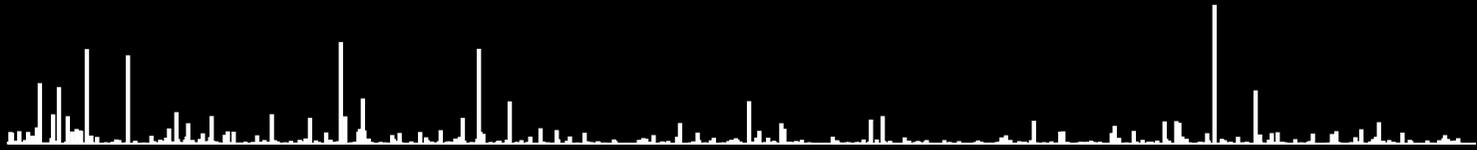


English

$p(\textit{Chinese}|\textit{English})$

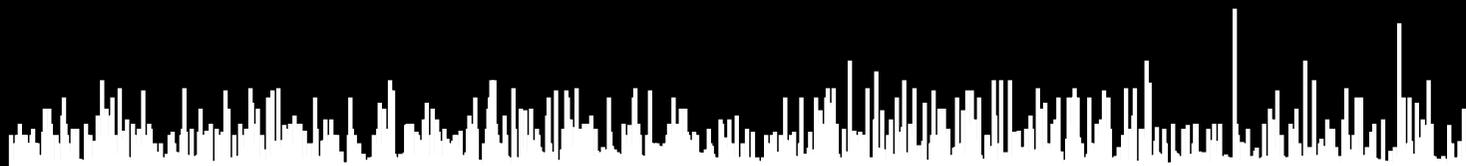


$\times p(\textit{English})$

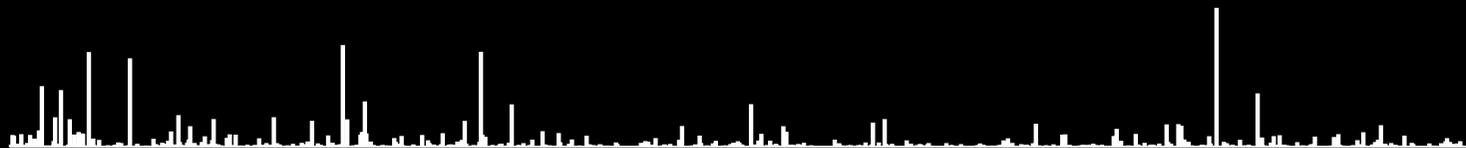


English

$p(\textit{Chinese}|\textit{English})$



$\times p(\textit{English})$

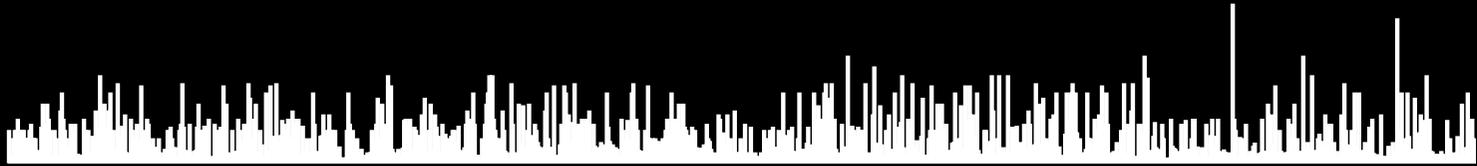


$= p(\textit{English}|\textit{Chinese})$

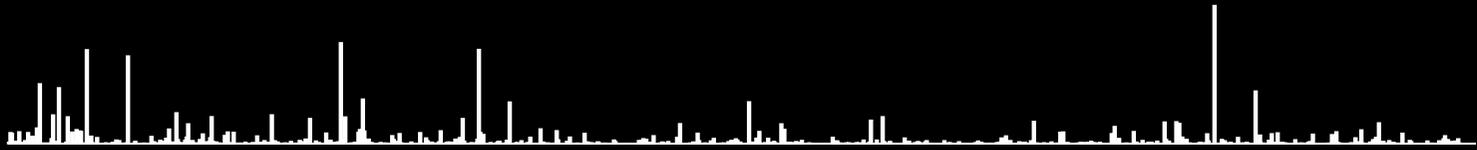


English

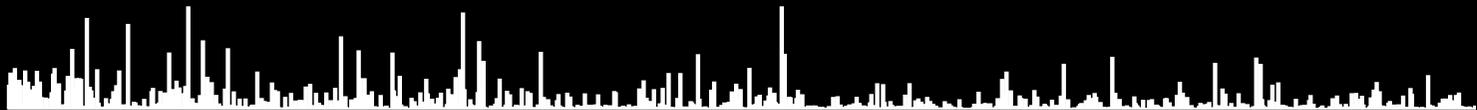
$$p(\textit{Chinese}|\textit{English})^1$$



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

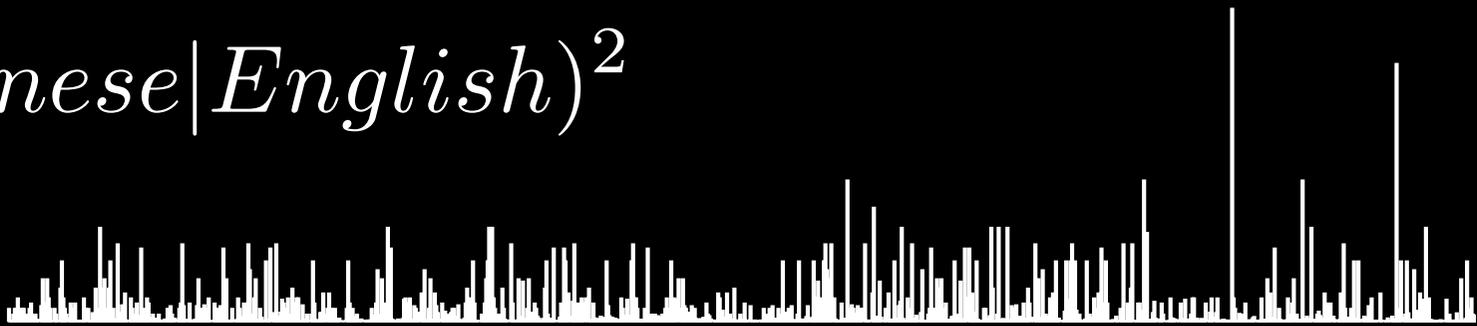


$$= p(\textit{English}|\textit{Chinese})$$

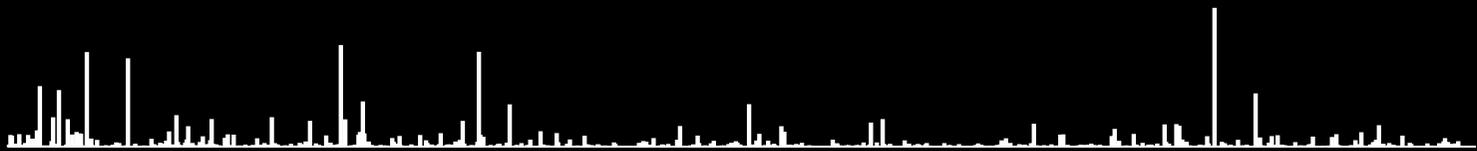


English

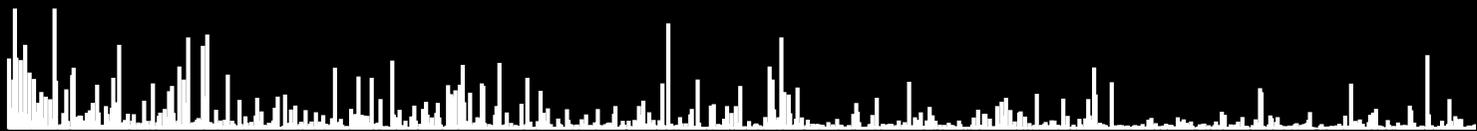
$$p(\textit{Chinese}|\textit{English})^2$$



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



$$= \textit{score}(\textit{English}|\textit{Chinese})$$

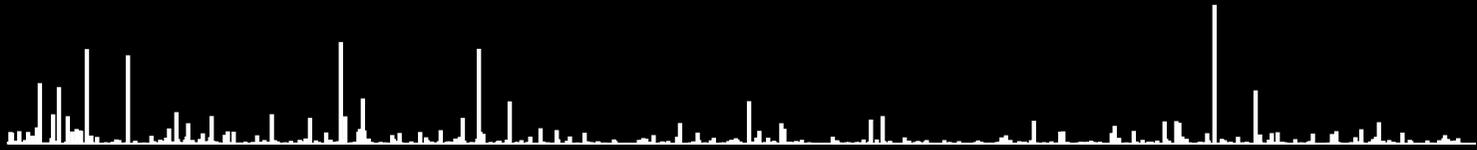


English

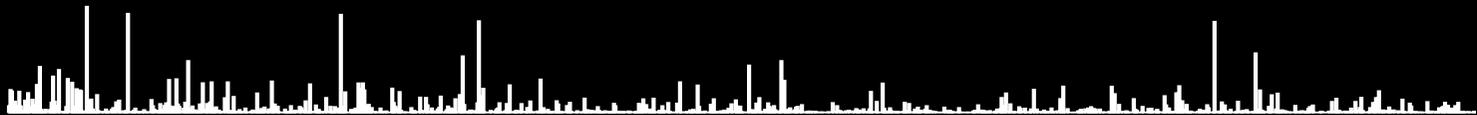
$$p(\textit{Chinese}|\textit{English})^{1/2}$$



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



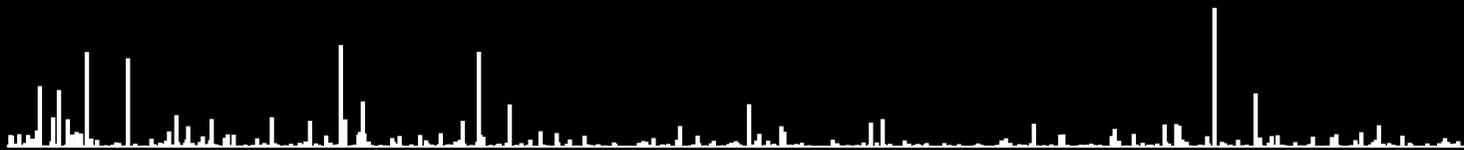
$$= \textit{score}(\textit{English}|\textit{Chinese})$$



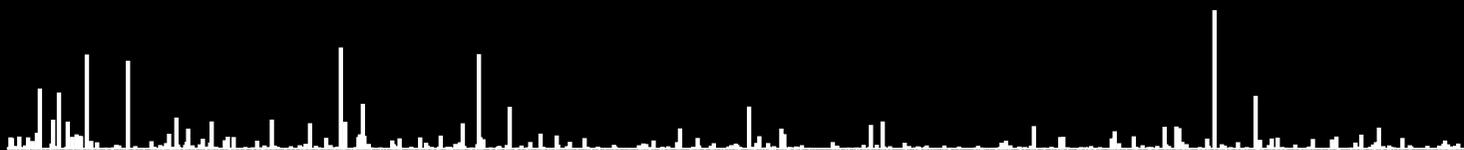
English

$$p(\textit{Chinese}|\textit{English})^0$$

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



$$= \textit{score}(\textit{English}|\textit{Chinese})$$

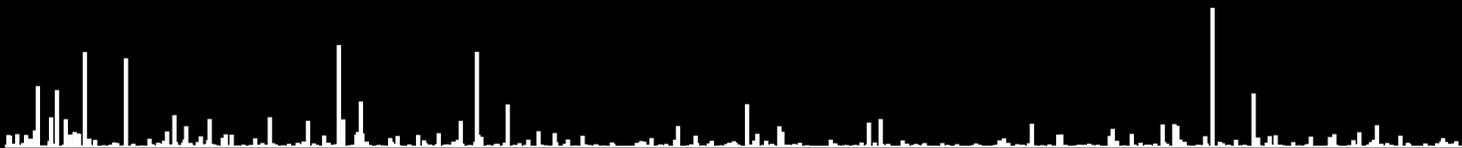


English

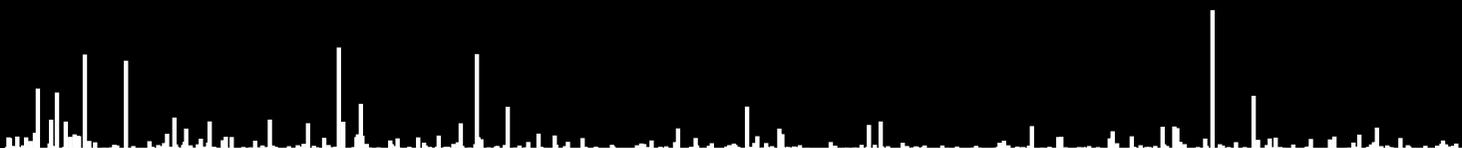
$$0 \cdot \log p(\textit{Chinese}|\textit{English})$$



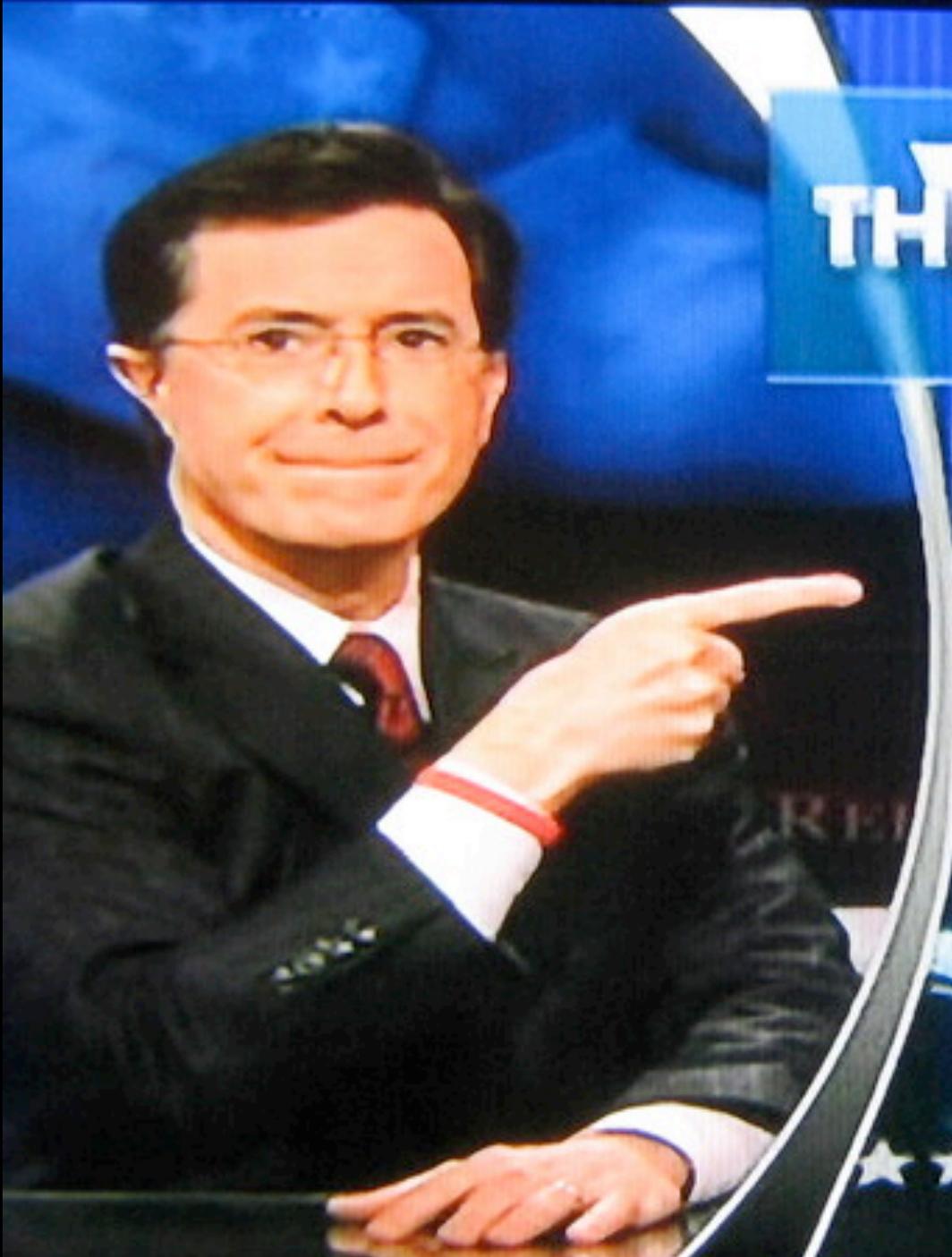
$$+1 \cdot \log p(\textit{English})$$



$$= \textit{score}(\textit{English}|\textit{Chinese})$$



English

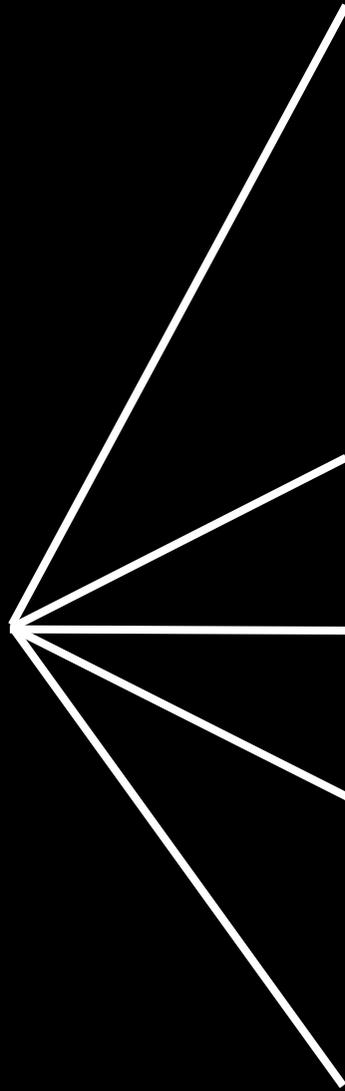
A man in a dark suit, white shirt, and red tie, wearing glasses, is pointing his right index finger towards the right. He is positioned on the left side of the image. The background is a blue-themed graphic with a world map and a row of stars at the bottom. The text 'THE WORD' is prominently displayed in the upper right quadrant.

THE WORD

Optimization

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

虽然北风呼啸，但天空依然十分清澈。

Although

虽然北风呼啸，但天空依然十分清澈。

The sky was still crystal clear

虽然北风呼啸，但天空依然十分清澈。

However

虽然北风呼啸，但天空依然十分清澈。

Although

虽然北风呼啸，但天空依然十分清澈。

The sky was still crystal clear

虽然北风呼啸，但天空依然十分清澈。

Although the north wind screamed

虽然北风呼啸，但天空依然十分清澈。

Although the northern wind shrieked

虽然北风呼啸，但天空依然十分清澈。

Although a north wind

虽然北风呼啸，但天空依然十分清澈。

Although the north wind screamed

虽然北风呼啸，但天空依然十分清澈。

Although the northern wind shrieked

虽然北风呼啸，但天空依然十分清澈。

Although a north wind

虽然北风呼啸，但天空依然十分清澈。

Although the northern wind shrieked,

虽然北风呼啸，但天空依然十分清澈。

Although the northern wind shrieked across the sky

虽然北风呼啸，但天空依然十分清澈。

Although the northern wind shrieked, yet

虽然北风呼啸，但天空依然十分清澈。

Although the northern wind shrieked,

虽然北风呼啸，但天空依然十分清澈。

Although the northern wind shrieked across the sky

虽然北风呼啸，但天空依然十分清澈。

Although the northern wind shrieked, yet

虽然北风呼啸，但天空依然十分清澈。

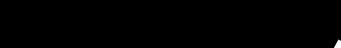
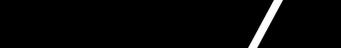
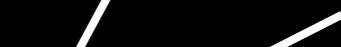
Although the northern wind shrieked across the sky,

→ 虽然北风呼啸，但天空依然十分清澈。



Although the northern wind shrieked across the sky,

虽然北风呼啸，但天空依然十分清澈。



Although the northern wind shrieked
across the sky, still quite

虽然北 风呼啸，但天空依然十分清澈。

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

虽然北 风呼啸，但天空依然十分清澈。

Although the northern wind shrieked
across the sky, it remained crystal clear.

虽然北 风呼啸，但天空依然十分清澈。

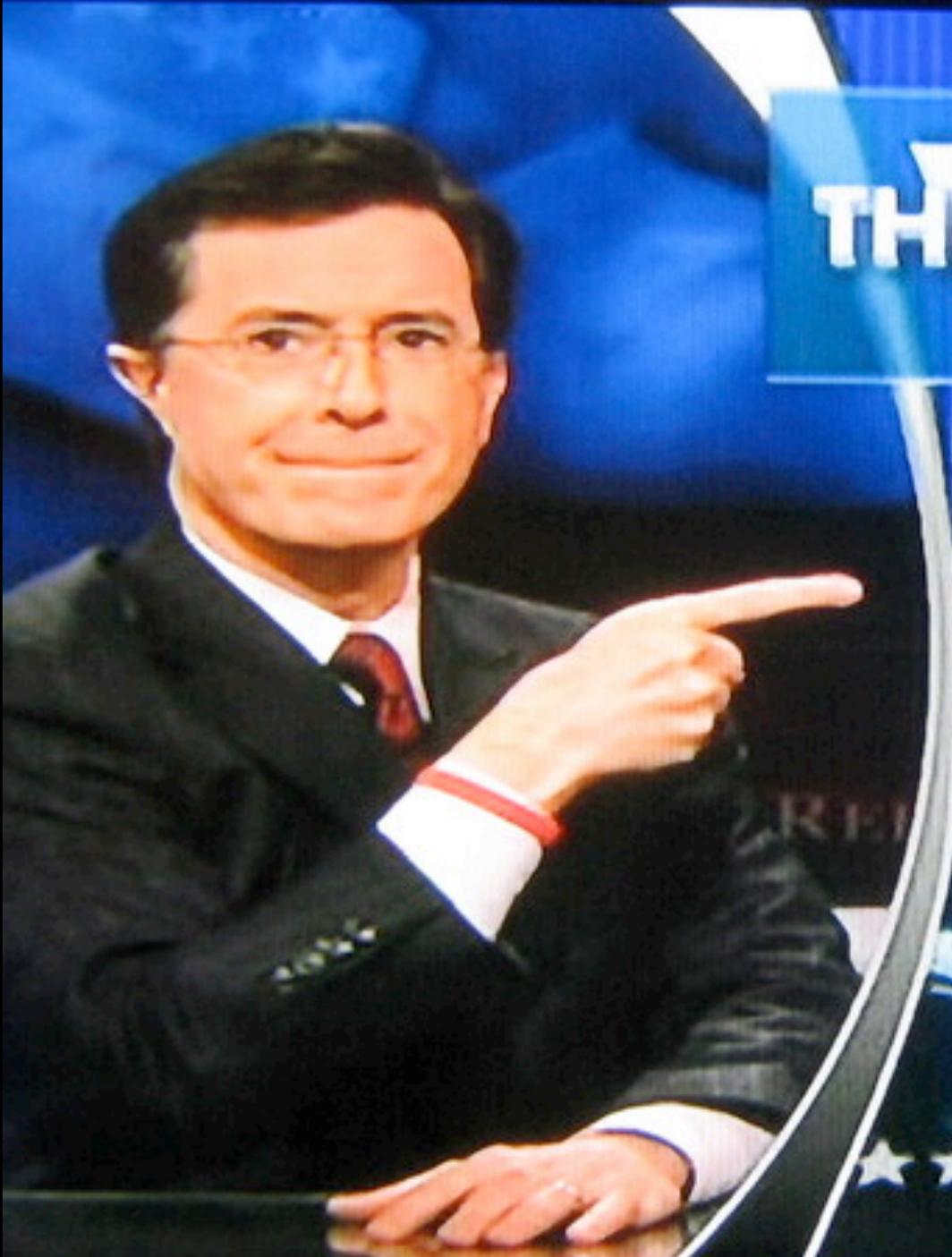
Although the northern wind shrieked
across the sky, but was still very clear.



虽然北风呼啸，但天空依然十分清澈。

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

虽然北风呼啸，但天空依然十分清澈。

A man in a dark suit, white shirt, and red tie, wearing glasses, is pointing his right index finger towards the right. He is positioned on the left side of the image. The background is a blue studio set with a world map and a row of stars at the bottom. The text 'THE WORD' is prominently displayed in the upper right.

THE WORD

Optimization



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.

与此同时,欧陆各地和以外地区的民防单位都表示,降雪已使道路封闭、火车误点和机场瘫痪。

In the meantime, civil defense services across the European continent and beyond all indicated that snowfalls have blocked roads, delayed trains, and paralyzed airports.

与此同时,欧陆各地和以外地区的民防单位都表示,降雪已使道路封闭、火车误点和机场瘫痪。

In the meantime, civil defense services across the European continent and beyond all indicated that snowfalls have blocked roads, delayed trains, and paralyzed airports.

Meanwhile, land and other regional civil defense units have all said to down the road closure and train mistakenly points and the airport.

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 .

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 .



BiLingual Evaluation Understudy

BLEU-1	BEwT-E	RTE
BLEU-4	Badger	RTE-MT
BLEU-v11b	BadgerLite	SEPIA1
BLEU-v12	Bleu-sbp	SEPIA2
METEOR-v0.6	BleuSP	SNR
NIST-v11b	CDer	SR-Or
TER-v0.7.254-GRR	DP-Or	SVM-Rank
ATEC1	DP-Orp	TERp
ATEC2	DR-Or	ULCh
ATEC3	EDPM	ULCopt
ATEC4	LET	invWer
Meteor-v0.7	METEOR-ranking	mBLEU
	MaxSim	mTER

BLEU(MT output)

$BLEU(\operatorname{argmax}_{English} \operatorname{score}(English|Chinese))$

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

A man in a dark suit, white shirt, and red tie, wearing glasses, is pointing his right index finger towards the right. He is positioned on the left side of the image. The background is a blue studio set with a world map and a row of stars at the bottom. The text 'THE WORD' is prominently displayed in the upper right.

THE WORD

Optimization



Max Blues Training

Evaluation drives MT research

- Metrics can drive the research for the topics that they evaluate
- NIST MT Eval / DARPA Sponsorship
- Bleu has lead to a focus on phrase-based translation
- Maximum BLEU training
- Other metrics may similarly change the community's focus

When writing a paper

- If you're writing a paper that claims that
 - one approach to machine translation is better than another, or that
 - some modification you've made to a system has improved translation quality
- Then you need to back up that claim
- Evaluation metrics can help, but good experimental design is also critical

Experimental Design

- Separating training / test / development sets
- Standardized data sets
- Standardized evaluation metric
- Error analysis
- Statistical significance



idea



idea



implementation



idea



implementation



32.6

experiment



?

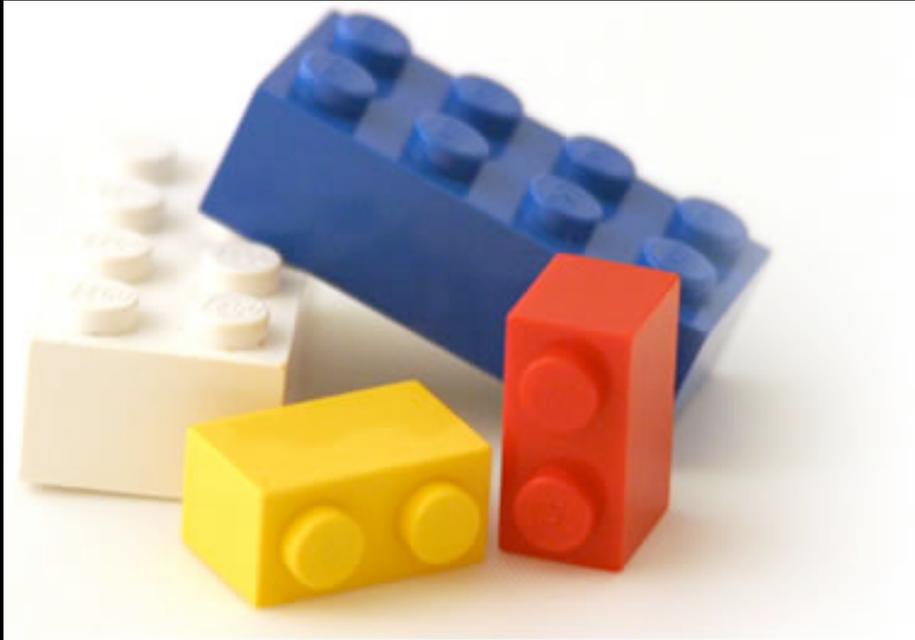


32.6

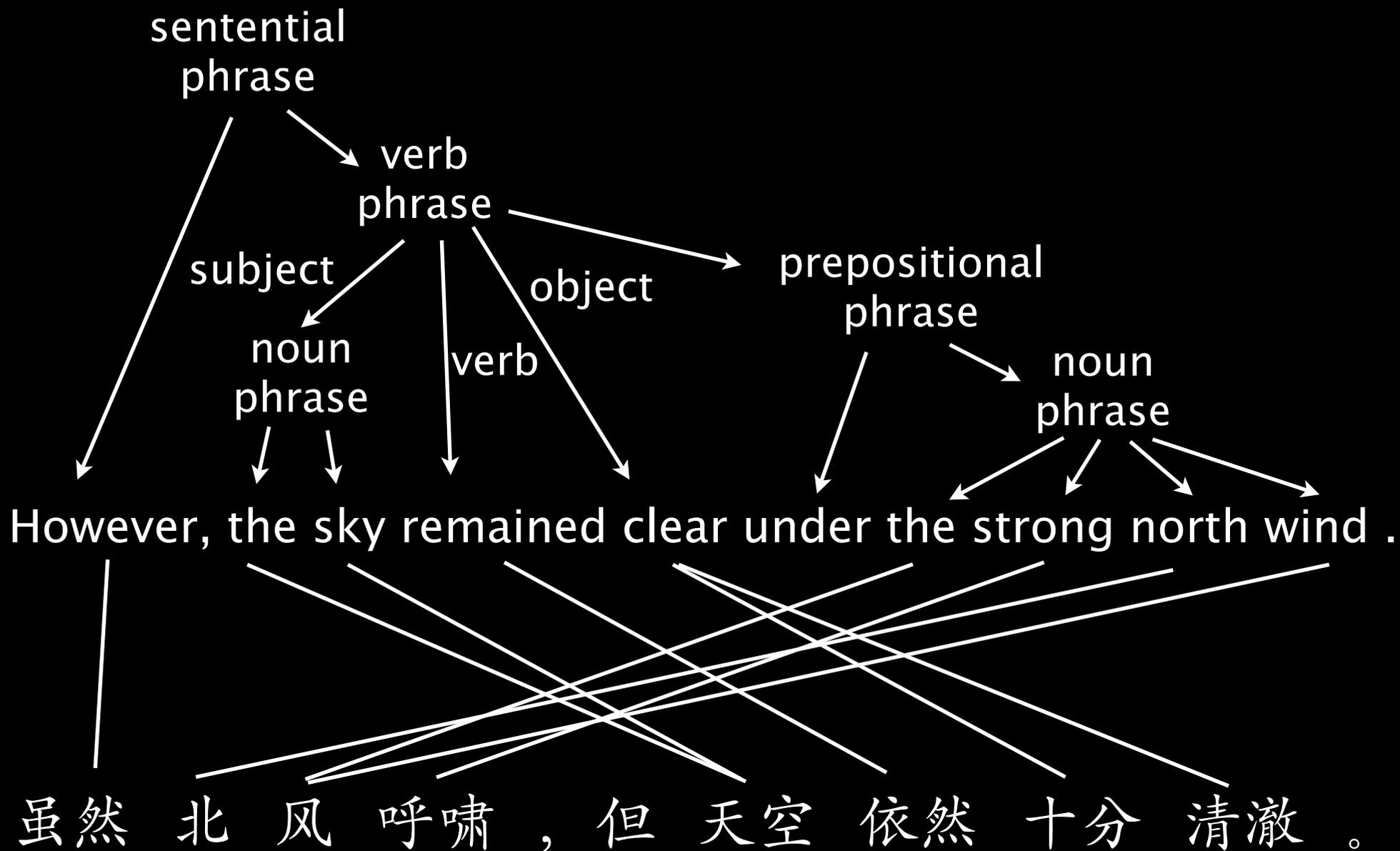
idea

implementation

experiment



- www.statmt.org/europarl/
- www.fjoch.com/GIZA++.html
- www.statmt.org/moses/
- Open source tools from this week.



Some of my best
friends are linguists.

Fred Jelinek (2004)



Open Problems

- What do we do when there's not enough (or any) parallel data?
- What do we do for out-of-domain data?
- What is the best learning algorithm?
- What is the best model?
- What is the best evaluation function?

- <http://www.statmt.org> - papers, tutorials, etc.
- Statistical Machine Translation. In *ACM Computing Surveys* 40(3), Aug 2008.

At <http://homepages.inf.ed.ac.uk/alopez>

BibTeX at <http://github.com/alopez/smtbib>