

# Statistical Machine Translation

May 13th, 2014



Josef van Genabith  
DFKI GmbH

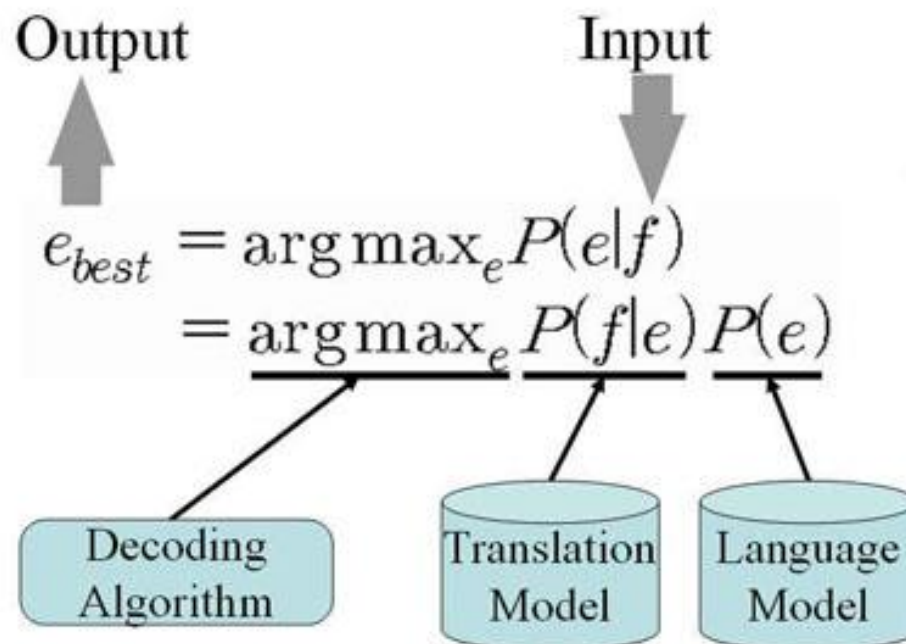
*Josef.van\_Genabith@dfki.de*

**Language Technology II**

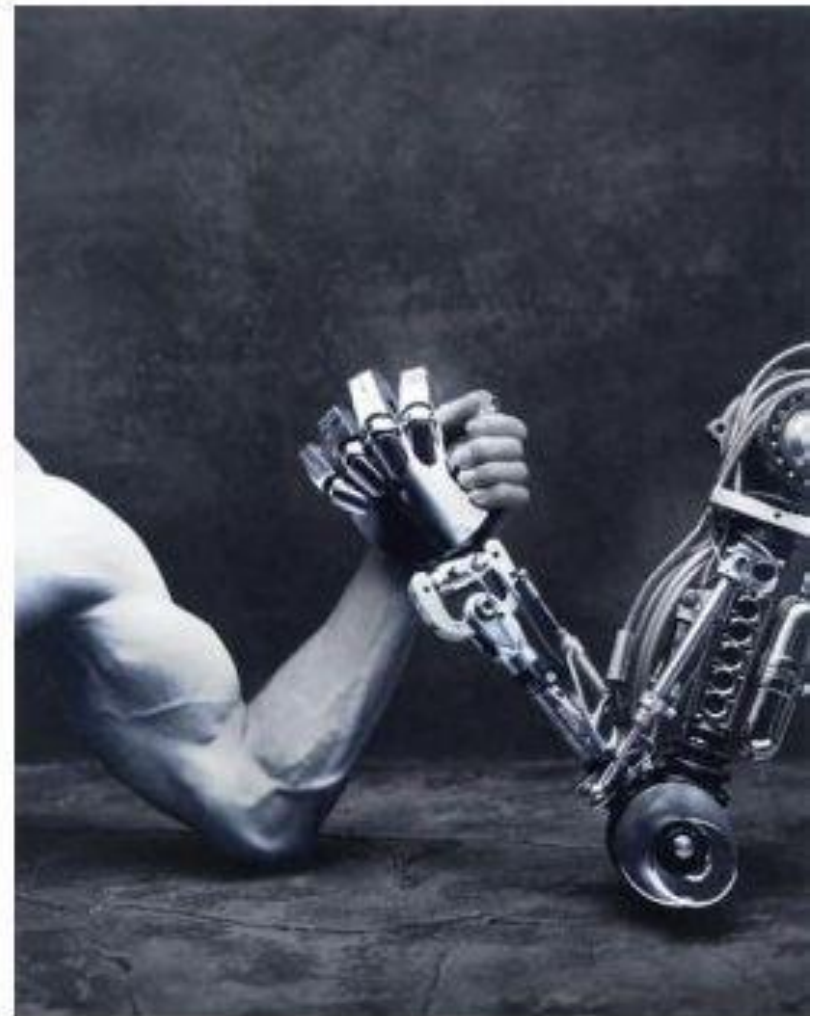
**SS 2014**

With additional slides from Sabine Hunsiker LT-II SS2012

- Introduction: the basic idea
- IBM models
- Phrase-Based SMT



- Human languages are extremely:
  - Elegant
  - Efficient
  - Flexible
- Turns out that this is exactly what makes it difficult for machines to understand human language, translate between them etc.



What's the matter with Human Language:

- Used to communicate
- record, store, manipulate, share, transmit information ...
- Highly ambiguous (a single string can mean many things)
- Flexible (we can say the same thing in many ways)
- Context dependent (meaning changes in context)
- Literal and non-literal meaning, metaphor ....

## ...The President's Research Award ...

Ambiguity

- The award the president won?
- The award the president is awarding?

Structural

- Who is this President anyway?

Context

- the *Áras an Uachtaráin* [...] ... The President's Research Award
- ..... the *Helix* in DCU ..... [...] ... The President's Research Award

- ... The President's Research Award ....
- ... Prof. Brian MacCraith's Research Award ...
- ... Brian's Research Award ...

Many ways  
of saying  
same thing

- Human Languages:
  - Highly ambiguous (a single string can mean many things)
  - Flexible (we can say the same thing in many ways)
  - Context dependent (meaning changes in context)
  - Literal and non-literal, non-compositional ...
- Many-to-many mappings between **form**  $\Leftrightarrow$  **meaning**
- Human Languages  $\neq$  Formal Languages (Math, Logic, ...)

**Unstructured representations of information!**

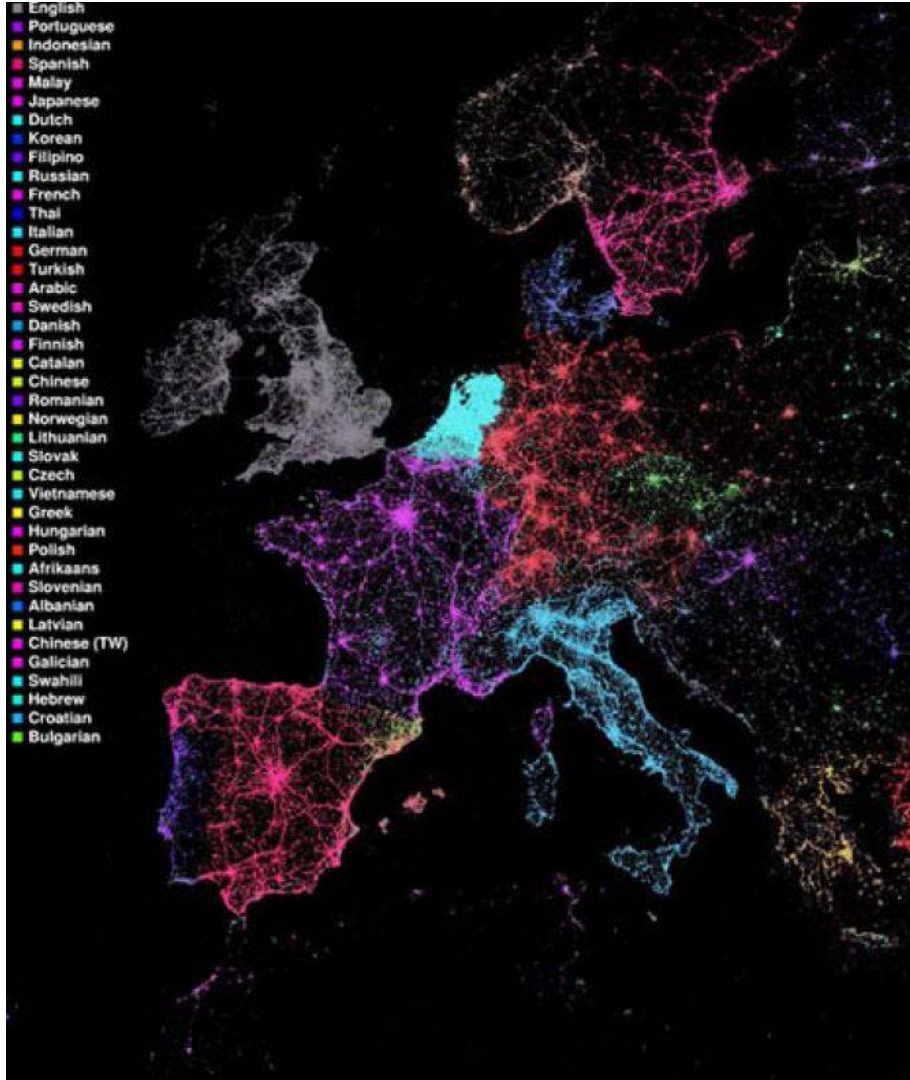
- Difficult for machines! Humans: context, world knowledge, ...

## Human Languages:

- Unstructured representation of information
- Many-to-many mappings between **form**  $\leftrightarrow$  **meaning**

- There are **many** of them .....





... of scientific  
 ... nce was very gr  
 ... rked on board, whil  
 ... t sail. I was oblig  
 ... shortly afterwards  
 ... which could not  
 ... considerable per  
 ... y different, as, w  
 ... very different, a  
 ... s the principal t  
 ... e captain is the  
 ... too, is so confin  
 ... is. Besides this  
 ... vibration, and  
 ... it certainly is  
 ... here a complete  
 ... on board a stea  
 ... es, hard dampin  
 ... id do well to ta  
 ... a southern clim  
 ... its for making s  
 ... have children  
 ... not be forgotte  
 ... cheaply in any  
 ... d by a sailor,  
 ... s are employe  
 ... nces he allow  
 ... e glowing heat  
 ... ry is made good  
 ... are being repa  
 ... in some degree  
 ... these matters  
 ... hardly know  
 ... t I myself suff  
 ... and clothes. A  
 ... before daybre  
 ... and compared  
 ... t admiral, on t  
 ... radeans. The  
 ... and thick app  
 ... n the captain  
 ... were eight pa  
 ... ner in the Braz  
 ... s. 44.). In the  
 ... veage at all su  
 ... our company,  
 ... the dolphins,  
 ... we had



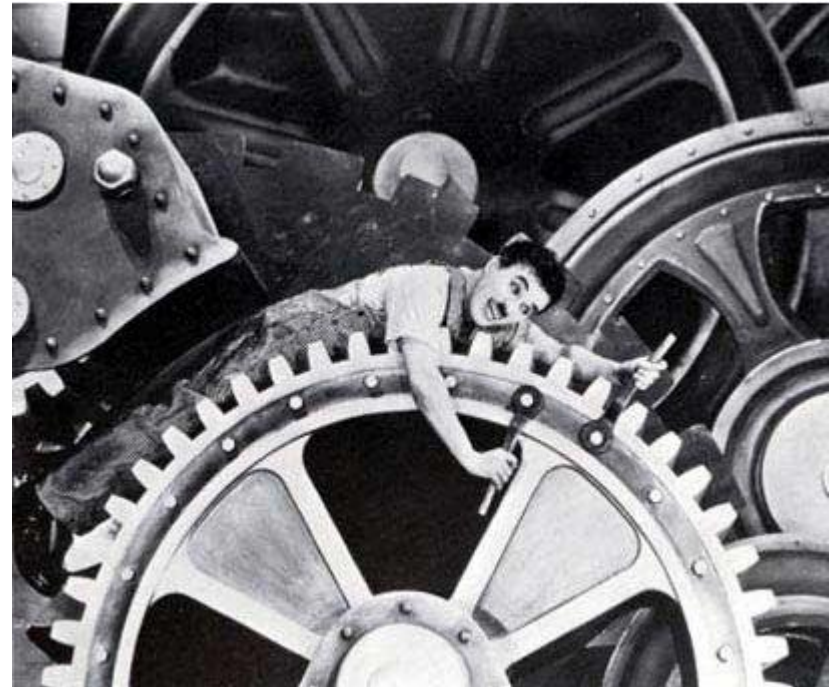
## ■ Translation

- Automatic
- Fast
- Scalable
- Accurate

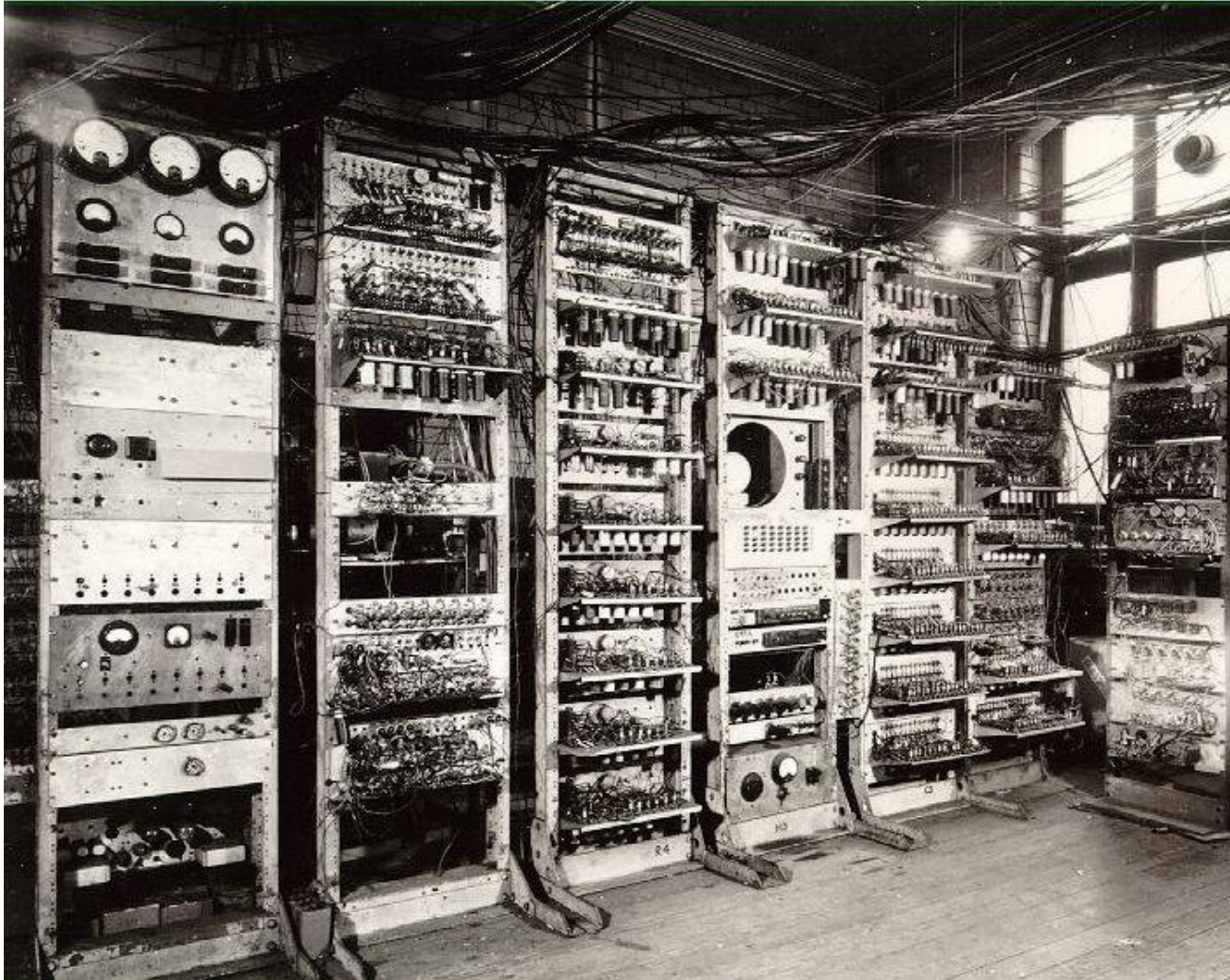
## ■ Machine Translation MT

### ■ How does MT work?

- Rule-based MT
- Statistical MT

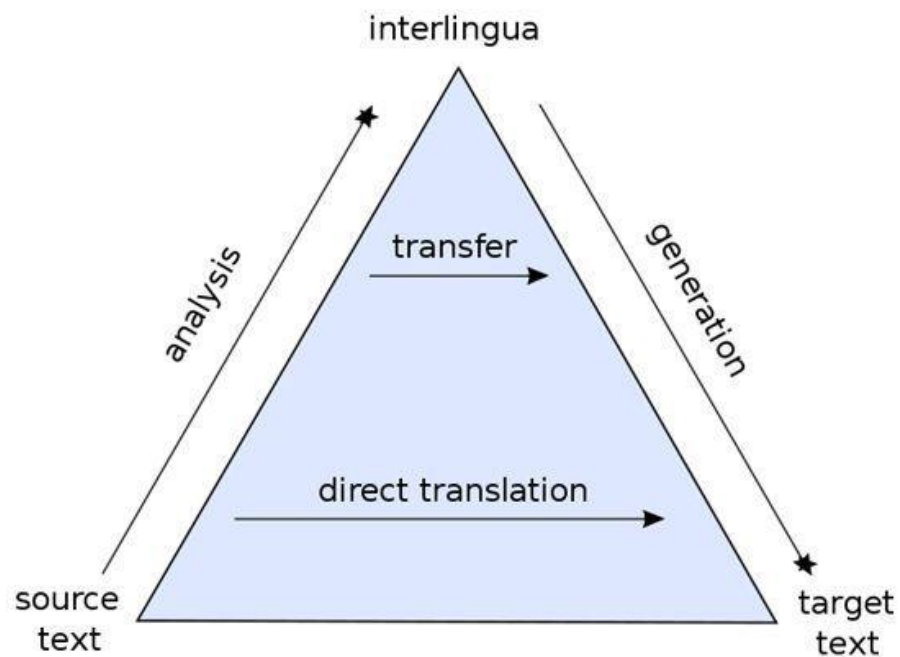


# Machine Translation



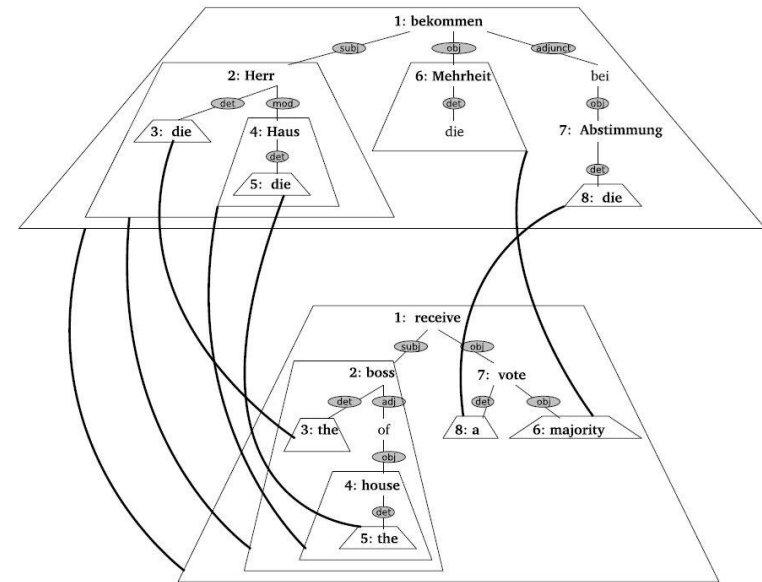
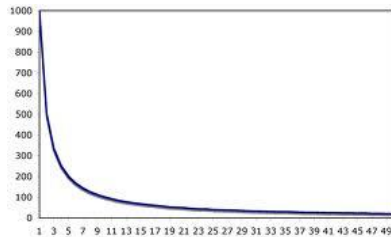
- Rule-Based Machine Translation (RBMT)
- Dominant Paradigm 1950s-1990s

- Words: Dictionaries
- Rules: Regularities
- Hand-crafted
- Highly skillful knowledge engineer



## ■ Problems with RBMT:

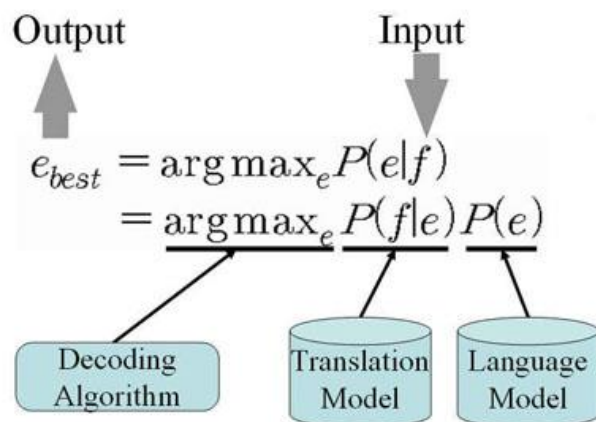
- ❑ Languages are complicated
- ❑ Many words
- ❑ Words are ambiguous
- ❑ Lots of rules
- ❑ Lots of exceptions to the rules



- ❑ Very labour intensive to scale
- ❑ 20-100 ++ person years required for each language pair
- ❑ 27 EU languages: > 700 pairs
- ❑ 100 languages: > 9900 pairs

**Difficult to scale!**

- Is there a simpler way of doing MT?
- Can machines learn **automatically** how to translate just from **data** (= already translated text)?
- Data-Driven MT (machine-learning based)
  - Statistical Machine Translation (SMT)



I love the boy.  
J'aime le garçon.

I love the dog.  
J'aime le chien.

They love the dog.  
Ils aiment le chien.

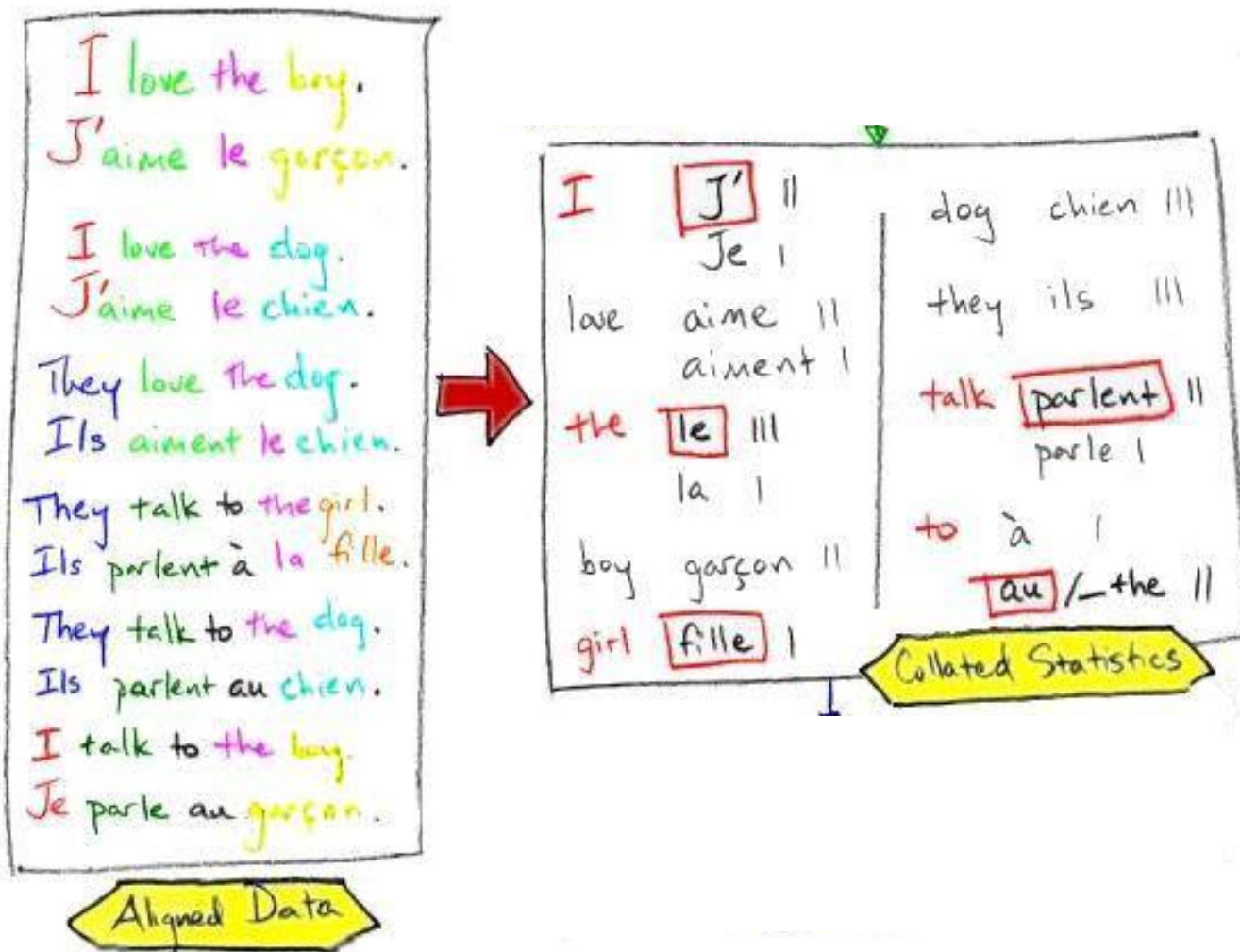
They talk to the girl.  
Ils parlent à la fille.

They talk to the dog.  
Ils parlent au chien.

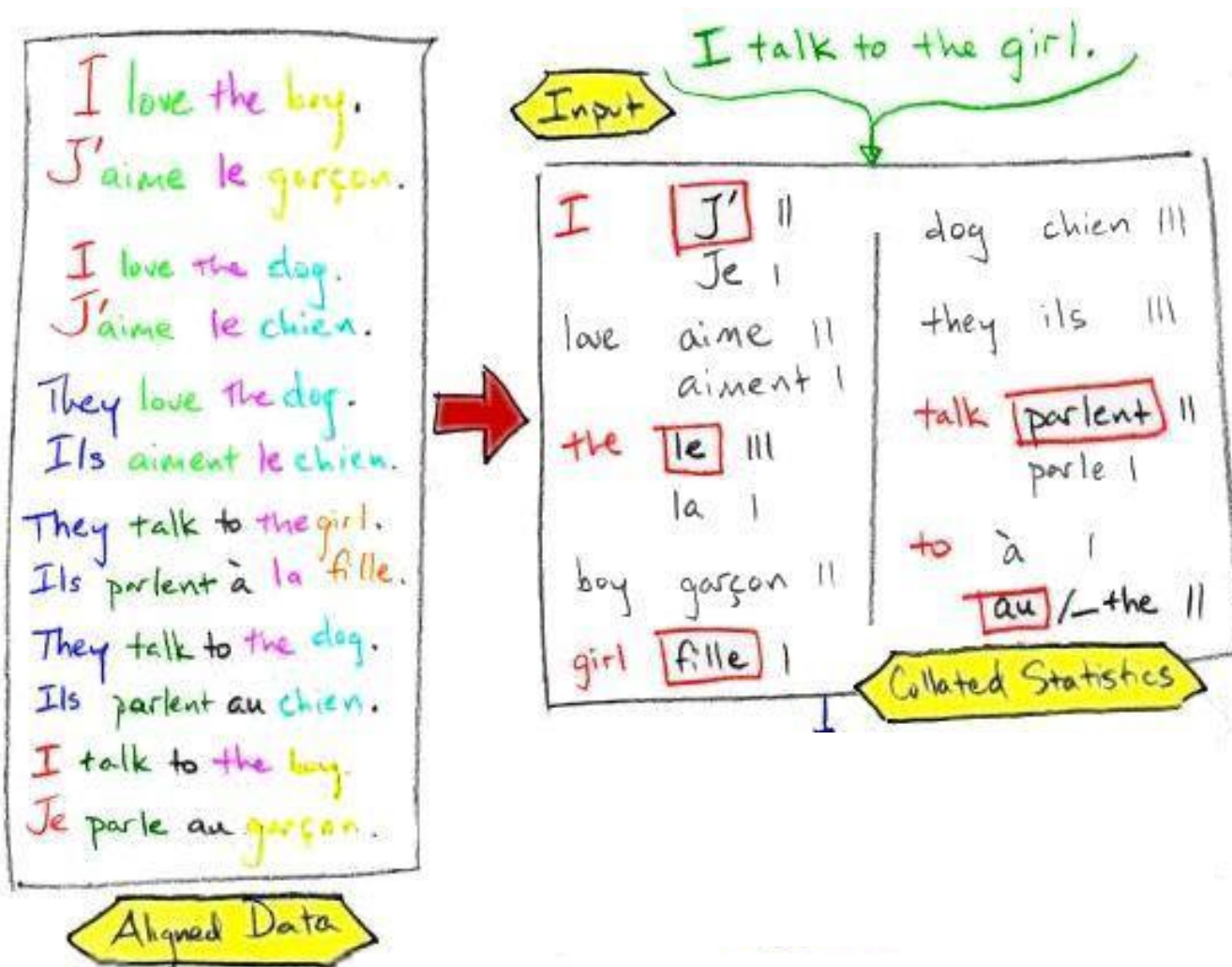
I talk to the boy.  
Je parle au garçon.

Aligned Data

# Statistical Machine Translation

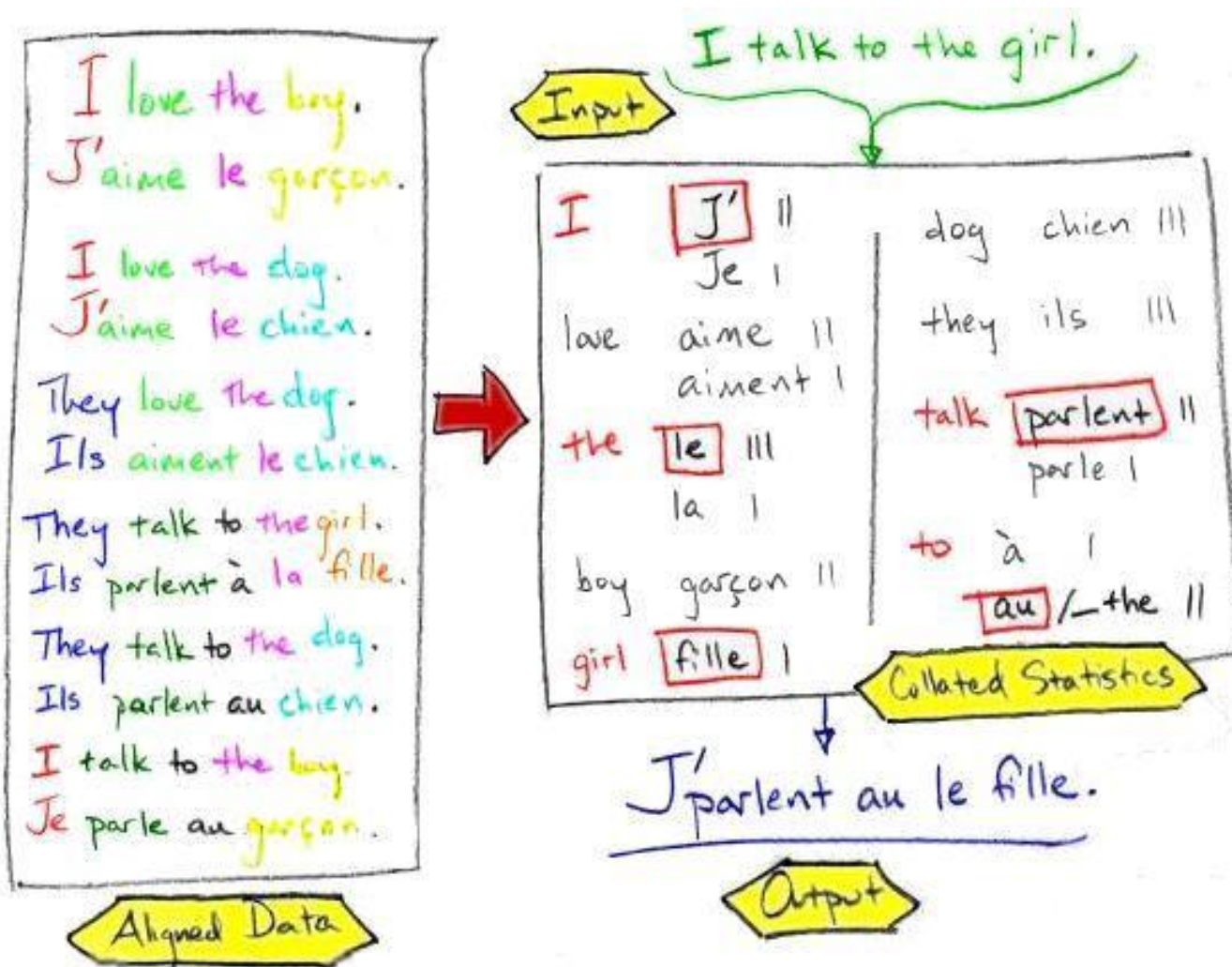


# Statistical Machine Translation





# Statistical Machine Translation



# How the computer learns to align/translate words:

		<b>CLASSIC SOUPS</b>		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		

# How the computer learns to align/translate words:

		<b>CLASSIC SOUPS</b>		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 <b>Egg Drop</b> 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		

# How the computer learns to align/translate words:

		<b>CLASSIC SOUPS</b>		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 <b>Egg Drop</b> 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.	<b>Egg Drop</b> Soup.....	1.10	2.10		
雲	蛋	湯	65.	<b>Egg Drop</b> 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		

# How the computer learns to align/translate words:

		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 <b>Egg Drop</b> 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.	<b>Egg Drop</b> Soup.....	1.10	2.10		
雲	吞	湯	65.	<b>Egg Drop</b> 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		

# How the computer learns to align/translate words:

		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 <b>Egg Drop</b> 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	<b>花</b>	湯	64.	<b>Egg Drop</b> Soup.....	1.10	2.10		
雲	蛋	湯	65.	<b>Egg Drop</b> 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		

# How the computer learns to align/translate words:

		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 <b>Egg Drop</b> 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	蛋	湯	64.	<b>Egg Drop</b> Soup.....	1.10	2.10		
雲	吞	湯	65.	<b>Egg Drop</b> 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		

# How the computer learns to align/translate words:

		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 <b>Egg Drop</b> 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	蛋	湯	64.	<b>Egg Drop</b> Soup.....	1.10	2.10		
雲	吞	湯	65.	<b>Egg Drop</b> 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		



# How the computer learns to align/translate words:

		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			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	68.	Crab Meat Corn Cream Soup.....	NA	3.50
海	鮮		湯	69.	Seafood Soup.....	NA	3.50

# How the computer learns to align/translate words:

		CLASSIC SOUPS		Sm.	Lg.	
清	燉	雞	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot) .....	1.50	2.75
雞	飯		58.	Chicken Rice Soup .....	1.85	3.25
雞	麵		59.	Chicken Noodle Soup .....	1.85	3.25
廣	東	雲吞	60.	Cantonese Wonton Soup.....	1.50	2.75
蕃	茄	蛋	61.	Tomato Clear Egg Drop Soup .....	1.65	2.95
雲	吞		62.	Regular Wonton Soup .....	1.10	2.10
酸	辣		63.	Hot & Sour Soup .....	1.10	2.10
蛋			64.	Egg Drop Soup.....	1.10	2.10
雲	吞		65.	Egg Drop Wonton Mix.....	1.10	2.10
豆	腐	菜	66.	Tofu Vegetable Soup .....	NA	3.50
雞	玉	米	67.	Chicken Corn Cream Soup .....	NA	3.50
蟹	肉	玉	68.	Crab Meat Corn Cream Soup.....	NA	3.50
海	鮮		69.	Seafood Soup.....	NA	3.50

# How the computer learns to align/translate words:

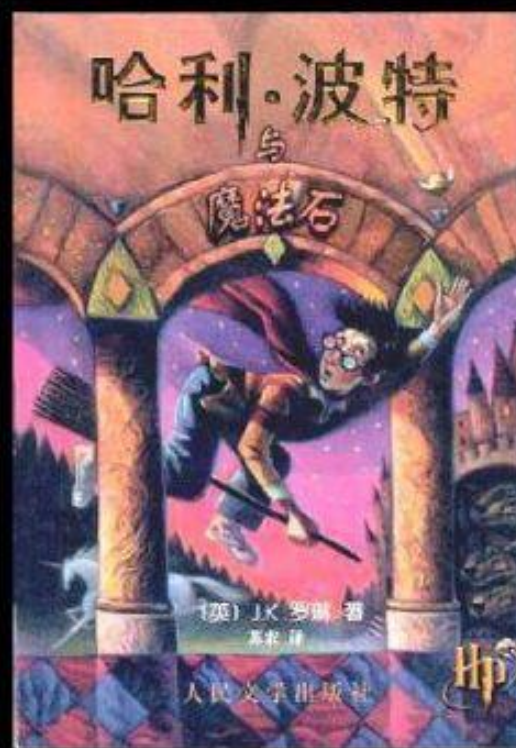
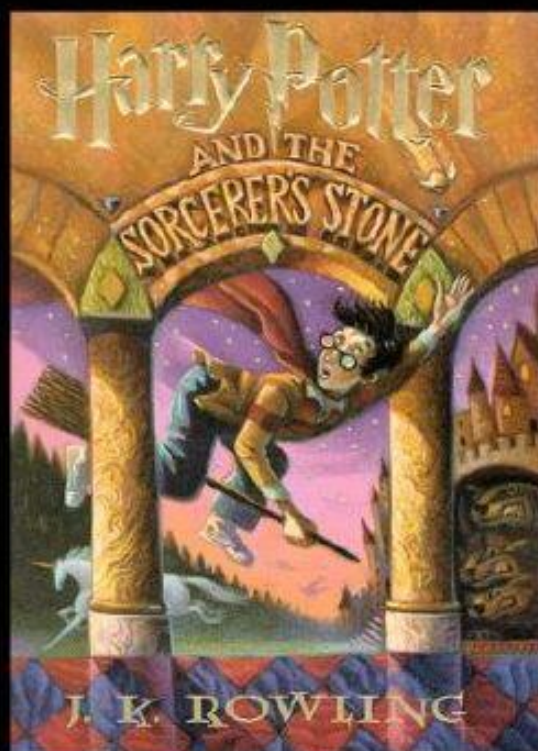
		<b>CLASSIC SOUPS</b>		Sm.	Lg.	
清	燉	雞	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot) .....	1.50	2.75
雞	飯		58.	Chicken Rice Soup .....	1.85	3.25
雞	麵		59.	Chicken Noodle Soup .....	1.85	3.25
廣	東	雲吞	60.	Cantonese Wonton Soup.....	1.50	2.75
蕃	茄	蛋	61.	Tomato Clear <b>Egg Drop</b> Soup .....	1.65	2.95
雲	吞		62.	Regular Wonton Soup .....	1.10	2.10
酸	辣		63.	Hot & Sour Soup .....	1.10	2.10
蛋			64.	<b>Egg Drop</b> Soup.....	1.10	2.10
雲	吞		65.	<b>Egg Drop</b> Wonton Mix.....	1.10	2.10
豆	腐	菜	66.	Tofu Vegetable Soup .....	NA	3.50
雞	玉	米	67.	Chicken Corn Cream Soup .....	NA	3.50
蟹	肉	玉	68.	Crab Meat Corn Cream Soup.....	NA	3.50
海	鮮		69.	Seafood Soup.....	NA	3.50

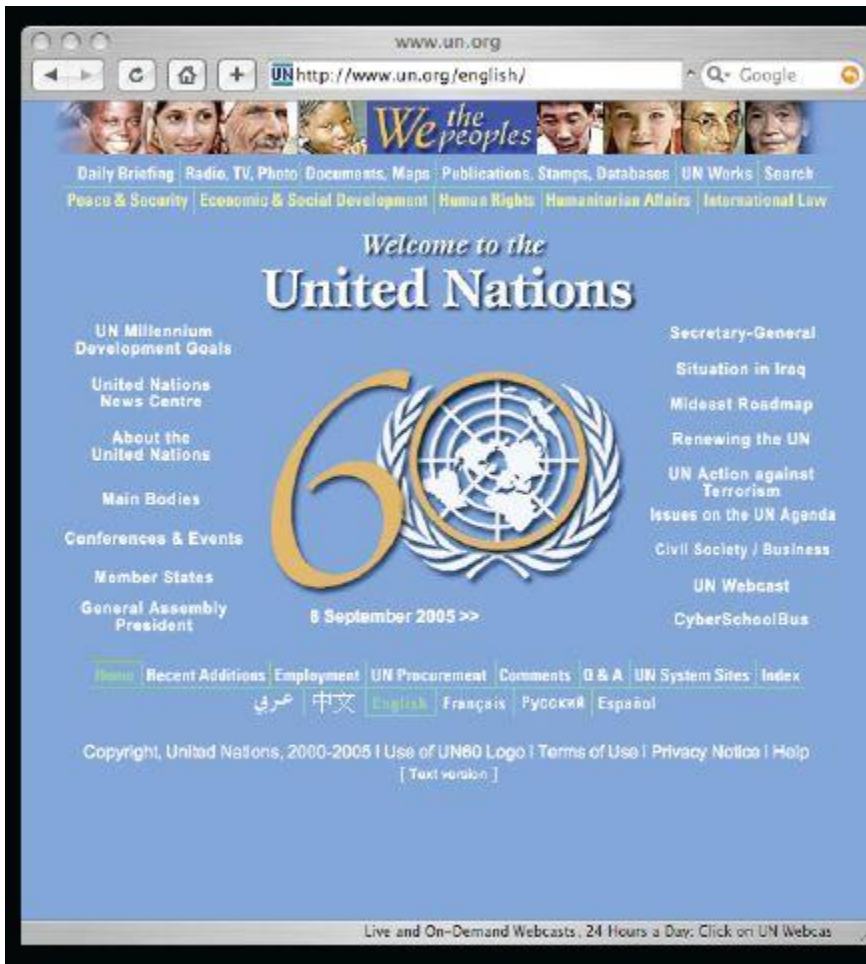
# How the computer learns to align/translate words:

		<b>CLASSIC SOUPS</b>		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 <b>Wonton</b> Soup.....	1.50	2.75	
蕃	茄	蛋	湯	61.	Tomato Clear Egg Drop Soup .....	1.65	2.95	
雲	吞	湯	62.	Regular <b>Wonton</b> 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 <b>Wonton</b> 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		

# How the computer learns to align/translate words:

		<b>CLASSIC SOUPS</b>		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		





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

*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  
Middle East 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/

*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

日本 English Other countries ヘルプ & コンタクト

There's no better way to fly.



フライト予約 お得な料金 Info & Service Miles & More マイアカウント

Booking Check-in

Flight Hotel Car

往復 片道

出発地 → 都市の検索 往路搭乗日  
 2008.05.16

目的地 → 都市の検索 復路搭乗日  
 2008.05.24

料金を優先  
日程を優先

大人 (12才以上) 小児 (2~12才) 幼児 (2才未満)  
1 0 0

クラス

→ 条件を絞り込んで検索する  
→ 複雑な旅程の予約  
→ アワード予約

ビジネスクラスなら「WEB得ピゾ」

ヨーロッパ往復  
¥ 510,000 ~

betterFly - ヨーロッパ内往復スーパー割引  
99 € ~ \*

\*税・手数料込み

お食事とお飲み物  
今月のスターシェフのご紹介をはじめ、ルフートハンザ自慢の機内食とワインについてのご案内です。

お客様へのお知らせ  
ご出発前に必ずご一読ください。  
「機内におけるシートベルト着用義務に関する新規定」「燃油追加料金徴収について」「機内持ち込み手荷物に関する新規定」

「WEB得20」 - 春夏のヨーロッパへ、ルフートハンザで。

- フランクフルト ¥79,000 ~
- パリ ¥79,000 ~
- ニース ¥84,000 ~
- ローマ ¥79,000 ~
- ボルドー ¥84,000 ~
- フィレンツェ ¥95,000 ~
- ベニス ¥84,000 ~
- ロンドン ¥79,000 ~
- その他の都市

クイックリンク

- タイムテーブル
- 発着案内
- オンラインチェックイン
- 自動チェックイン
- マイルの確認について
- 日本のお客様へ
- ルフトハンザモバイル
- メールニュース申し込み
- お問い合わせ
- 企業の皆様へのご案内
- コーポレート予約
- プレスリリース
- ルフトハンザについて

Deutschland English Andere Länder Hilfe & Kontakt

There's no better way to fly.



Buchen Angebote Info & Service Miles & More Urlaubsangebote Mein Konto

Buchen Check-in

Flug Hotel Mietwagen

Hin- und Rückflug Einfacher Flug

Von → Airports Hinflug am  
 16.05.2008

Nach → Airports Rückflug am  
 24.05.2008

Ich bin zeitlich flexibel  
Ich muss am ausgewählten Tag fliegen

Erwachsene (ab 12 Jahre) Kinder (2-11 Jahre) Babys (bis 2 Jahre)  
1 0 0

Klasse

→ Erweiterte Suchoptionen  
→ Mehrere Streckenabschnitte/Gabelflüge  
→ Prämienbuchung

Philadelphia ab 469 €\*  
und über 40 weitere Ziele für Frühbucher  
\*inkl. 15 € Lufthansa Ticket Service Charge

Grenzenlose Möglichkeiten: Urlaub buchen auf [lufthansa.com](http://lufthansa.com)  
Lufthansa bietet in Kooperation mit DERTOUR jetzt auch komplette Urlaubsangebote, z.B. an die Westküste der USA.

Ready to Fly?  
Sind Sie spontan? Hier finden Sie unsere attraktiven Angebote für Kurzentschlossene.  
→ Breslau/Wroclaw ab 99 €\*  
→ Genf ab 99 €\*  
→ Paris ab 99 €\*  
→ Stavanger ab 99 €\*  
→ Venedig ab 99 €\*  
\*inkl. 10 € Lufthansa Ticket Service Charge

Buchen Sie Flüge für Ihre Familie und Freunde  
Sie können über unsere Online Flugbuchung auch dann Flüge buchen, wenn Sie selbst gar nicht mitfliegen.

Fluginformationen  
Wir informieren Sie zum aktuellen Flugbetrieb.

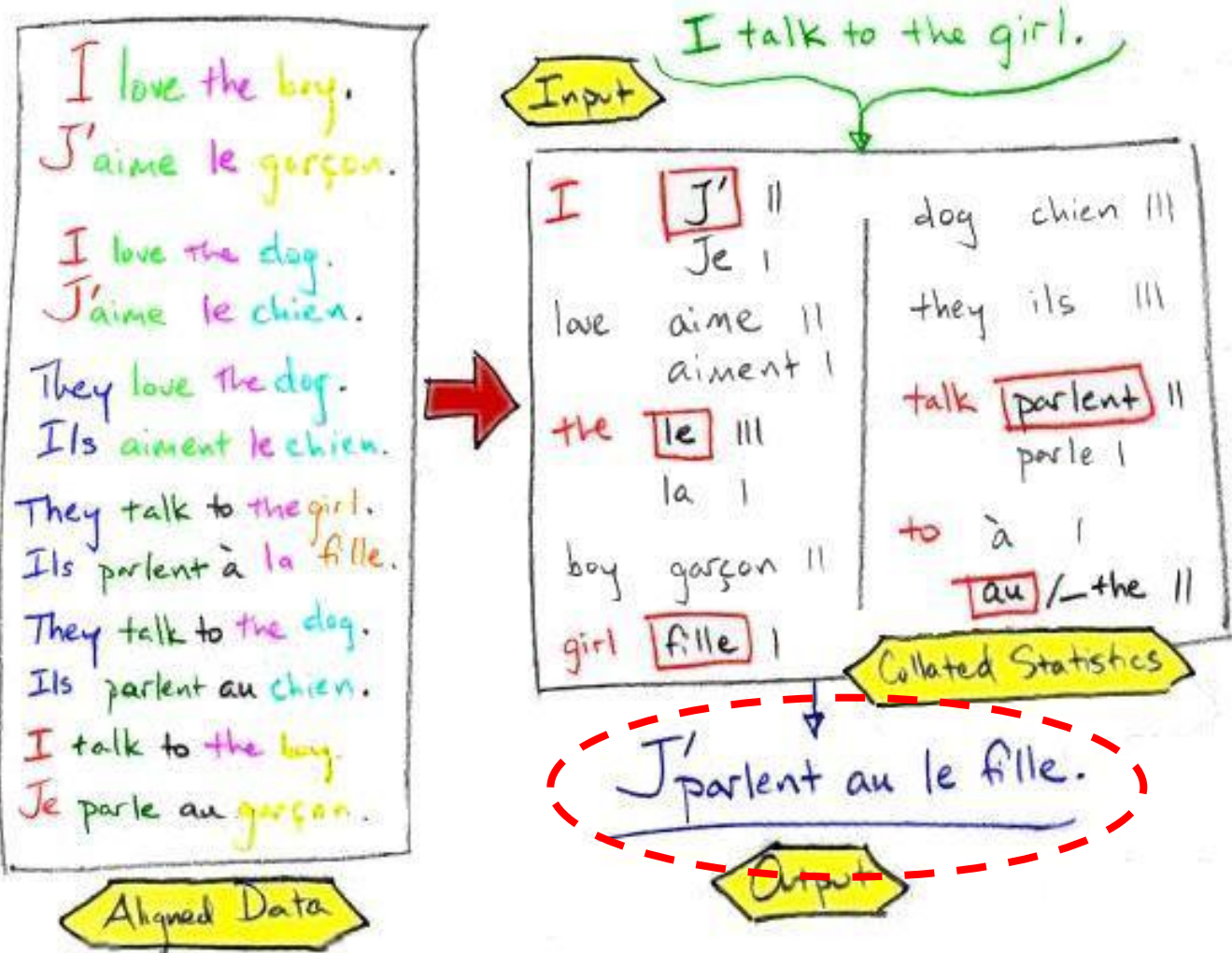
Quicklinks

- Flugplan
- Ankunft und Abflug
- Online Check-in
- Mein Meilenkonto
- Meine Buchungen
- Passenger Receipt
- Gepäck Ratgeber
- Gruppenbuchungen
- Hier starten mit SWISS
- Hier starten mit JetBlue
- Lufthansa Fanflug
- Newsletter
- Klimaschutz-Spende



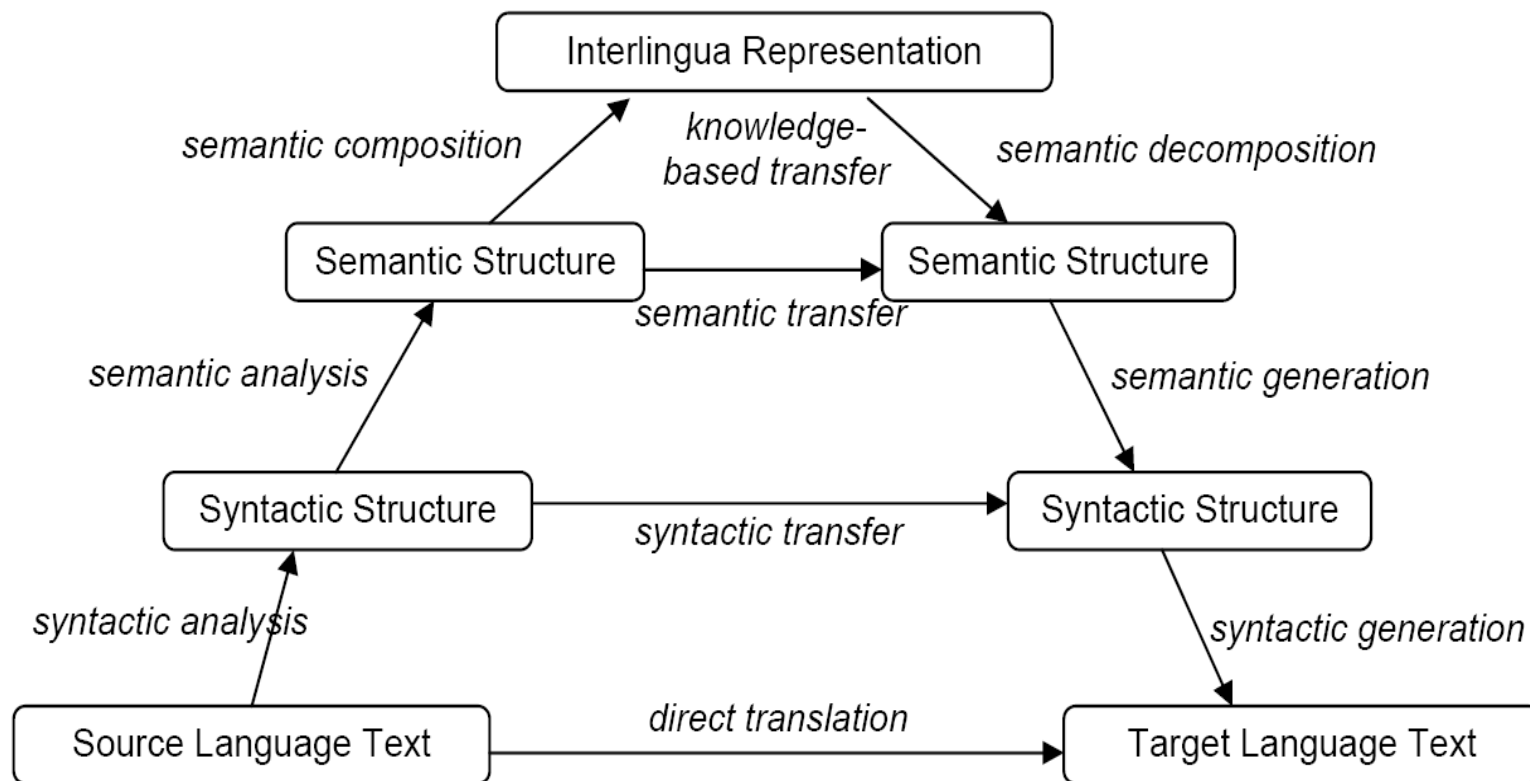
- Data
  - Translations (electronic form)
  - Translation Memories
  - WWW
  
- Europarl: translations of European Parliamentary Debates
  
- French-English corpus: 100M words

# Statistical Machine Translation



# Possible (rule-based) MT architectures

## The „Vauquois Triangle“



# What do we need to model?

---

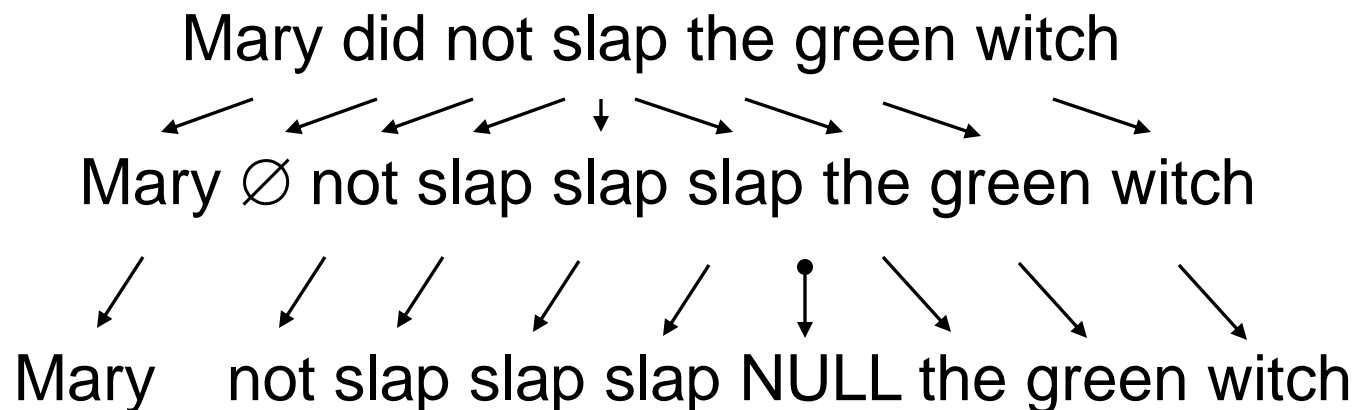
Mary did not slap the green witch

Maria no daba una bofetada a la bruja verde

# What do we need to model?

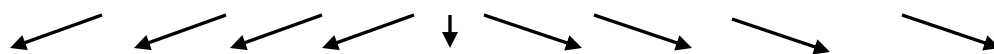
Mary did not slap the green witch  
← ← ← ← ↓ → → → →  
Mary ∅ not slap slap slap the green witch

Maria no daba una bofetada a la bruja verde



Maria no daba una bofetada a la bruja verde

Mary did not slap the green witch



Mary  $\emptyset$  not slap slap slap the green witch



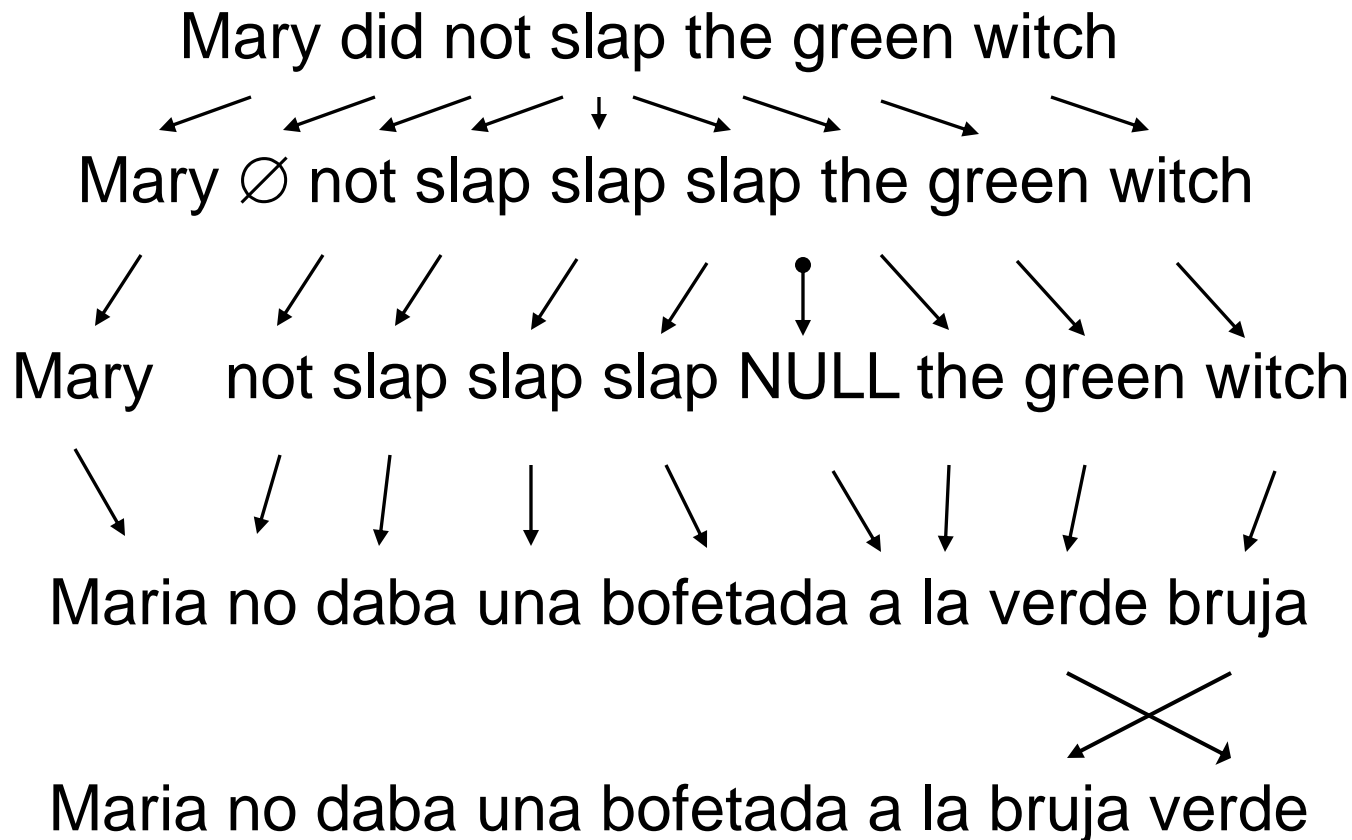
Mary not slap slap slap NULL the green witch



Maria no daba una bofetada a la verde bruja

Maria no daba una bofetada a la bruja verde

# What do we need to model?

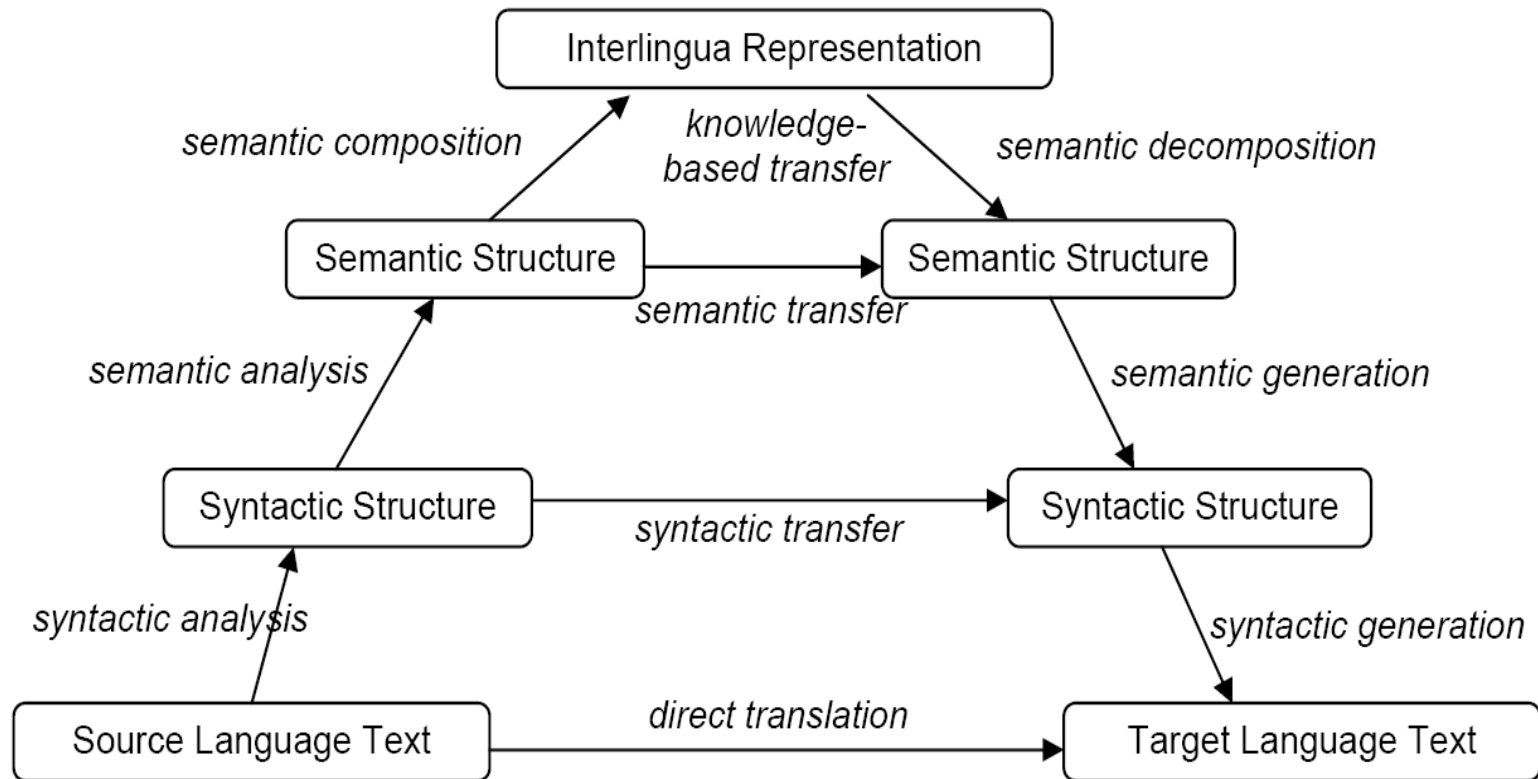




# Possible (rule-based) MT architectures

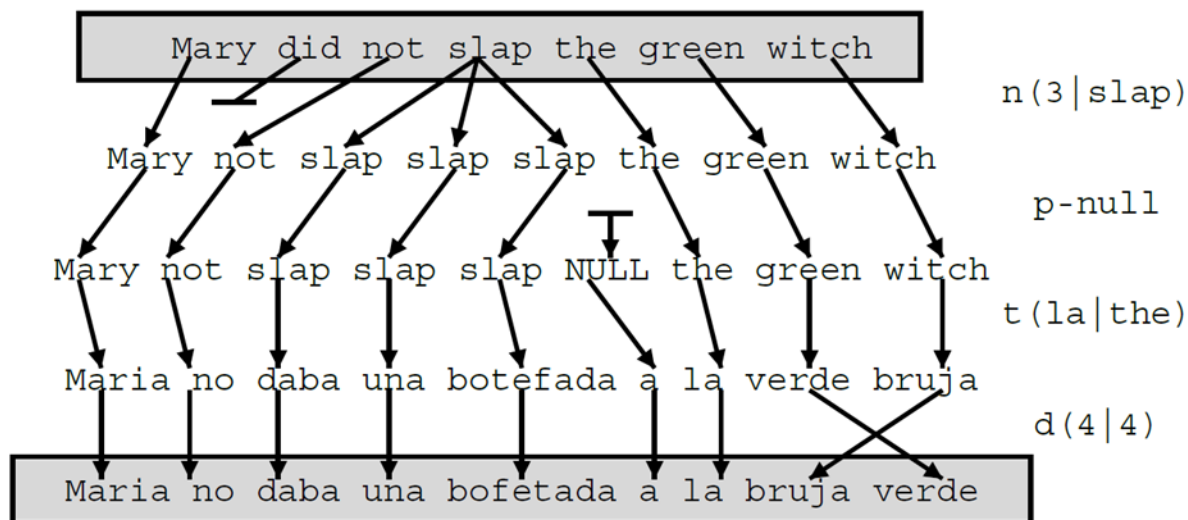
## The „Vauquois Triangle“

Slide by Sabine Hunsiker



# What do we need to model?

## IBM Model 4



# More Slides by Sabine Hunsiker SS2012

---

Slides from the SS2012 version of the course

## ■ MT in general, history:

- <http://www.MT-Archive.info>: Electronic repository and bibliography of articles, books and papers on topics in machine translation and computer-based translation tools, regularly updated, contains over 3300 items
- Hutchins & Somers: An introduction to machine translation. Academic Press, 1992, available under <http://www.hutchinsweb.me.uk/IntroMT-TOC.htm>

## ■ MT systems:

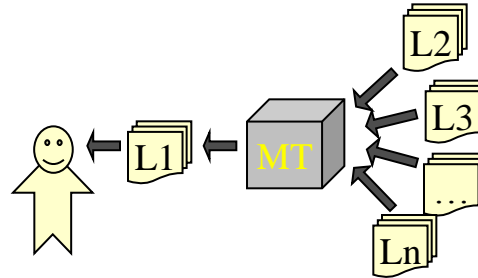
Compendium of Translation Software, see <http://www.hutchinsweb.me.uk/Compendium.htm>

## ■ Statistical Machine Translation:

See [www.statmt.org](http://www.statmt.org)

Book by Philipp Koehn is available in the coli-bib

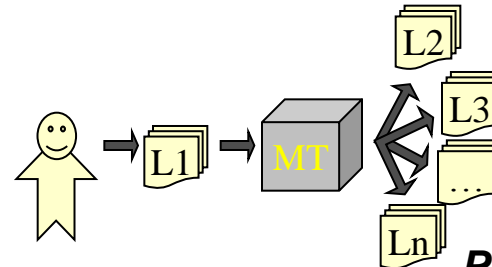
a) MT for assimilation  
„inbound“



**Robustness**  
**Coverage**

*Daily throughput of  
online-MT-Systems  
> 500 M Words*

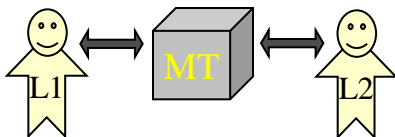
b) MT for dissemination  
„outbound“



**Textual quality**

*Publishable quality can only be  
authored by humans;  
Translation Memories & CAT-  
Tools mandatory for  
professional translators*

c) MT for direct communication



**Speech recognition, context dependence**

*Topic of many running and completed research projects  
(VerbMobil, TC Star, TransTac, ...)  
US-Military uses systems for spoken MT*

## *Some recent examples*



**'I am not in the office at the moment. Please send any work to be translated'**

- Good translation requires knowledge and decisions on many levels
  - syntactic disambiguation (POS, attachments)
  - semantic disambiguation (collocations, scope, word sense)
  - reference resolution
  - lexical choice in target language
  - application-specific terminology, register, connotations, good style ...
- Rule-based models of all these levels are very expensive to build, maintain, and adapt to new domains
- Statistical approaches have been quite successful in many areas of NLP, once data has been annotated
- Learning from existing translation will focus on distinctions that matter (not on the linguist's favorite subject)
- Translation corpora are available in rapidly growing amounts
- SMT *can* integrate rule-based modules (morphologies, lexicons)
- SMT *can* use feed-back for on-line adaptation to domain and user preferences

- 1949: Warren Weaver: *the translation problem can be largely solved by “statistical semantic studies”*
- 1950s..1970s: Predominance of rule-based approaches
- 1966: ALPAC report: general discouragement for MT (in the US)
- 1980s: example-based MT proposed in Japan (Nagao), statistical approaches to speech recognition (Jelinek et al. at IBM)
- Late 80s: Statistical POS taggers, SMT models at IBM, work on translation alignment at Xerox (M. Kay)
- Early 90s: many statistical approaches to NLP in general, IBM ‘s Candide claimed to be as good as Systran
- Late 90s: Statistical MT successful as a fallback approach within Verbmobil System (Ney, Och). Wide distribution of translation memory technology (Trados) indicates big commercial potential of SMT
- 1999 Johns Hopkins workshop: open source re-implementation of IBM’ s SMT methods (GIZA)



- Since 2001: DARPA/NIST evaluation campaign (XYZ → English), uses BLEU score for automatic evaluation
- Various companies start marketing/exploring SMT:  
language weaver, aixplain GmbH, Linear B Ltd., esteam, Google Labs
- 2002: Philipp Koehn (ISI) makes EuroParl corpus available
- 2003: Koehn, Och & Marcu propose *Statistical Phrase-Based MT*
- 2004: ISI publishes Philipp Koehn's SMT decoder *Pharaoh*
- 2005: First SMT workshop with shared task
- 2006: Johns Hopkins workshop on OS factored SMT decoder Moses, Start of EuroMatrix project for MT between all EU languages, Acquis Communautaire (EU laws in 20+ languages) made available
- 2007: Google abandons Systran and switches to own SMT technology
- 2009: Start of EuroMatrixPlus “*bringing MT to the user*”
- 2010: Start of many additional MT-related EU projects (Let's MT, ACCURAT, ...)