Machine Translation

Prof. Sameer Singh

CS 295: STATISTICAL NLP

WINTER 2017

February 28, 2017

Based on slides from Jason Eisenstein, Chris Dyer, Alan Ritter, Yejin Choi, and everyone else they copied from.

Upcoming...

Summaries

- Paper summaries: **February 28**, March 14
- Summary 1 graded

Project

- Status report due in 1 weeks: **March 7, 2017**
- Instructions coming today!
- Almost final report, only 5 pages

Homework

- Homework 4 is due on March 13
- Write-up and data releasing soon.

Outline

Machine Translation

Introduction to Statistical MT

IBM Translation Models

Outline

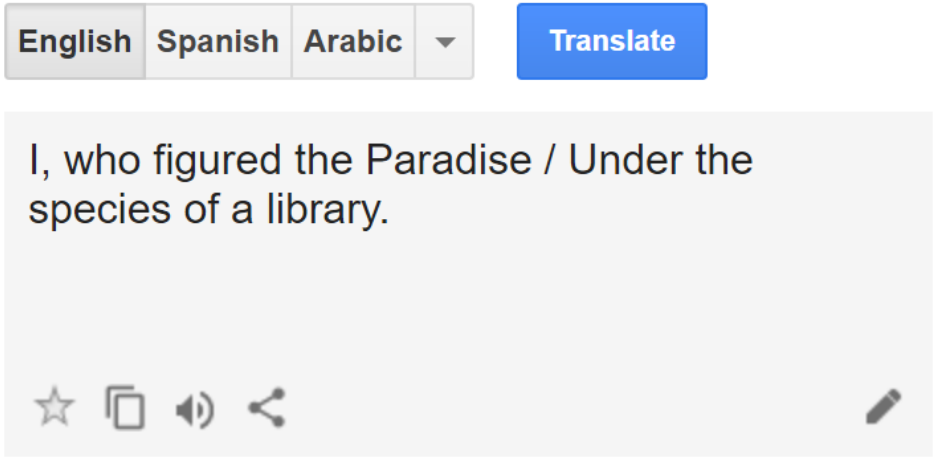
Machine Translation

Introduction to Statistical MT

IBM Translation Models

Machine Translation

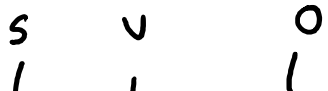
Yo, que me figuraba el Paraíso / Bajo la especie de una biblioteca.



I have always imagined Paradise as a kind of library.

Challenges: Word Order

SVO vs SOV



English: IBM bought Lotus

Japanese: IBM Lotus bought

Even for SVO

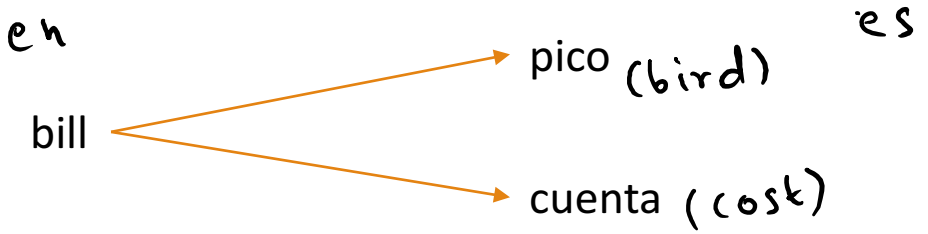
English: I will buy it

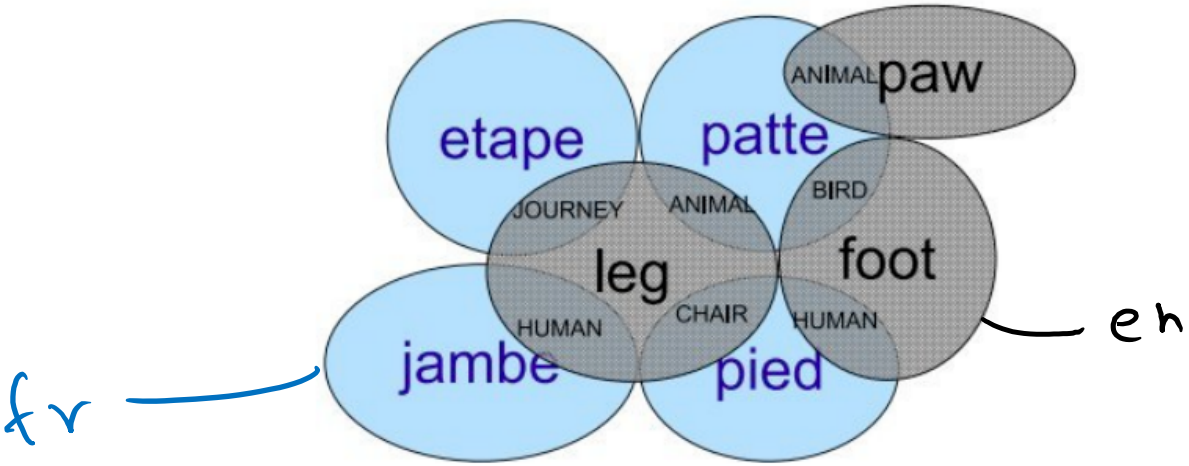
French: Je vais l'acheter (I will it buy)

English: I bought it

French: Je l'ai achet´ e (I it have bought)

Challenges: Lexical Ambiguity





Challenges: Pronouns

Different Pronouns

English possessive pronouns take the gender of the owner:

Marie rides **her** bike

French possessive pronouns take the gender of the object:

Marie monte sur **son** vélo

Dropping Pronouns

In Spanish, you can recover the pronoun from verb inflection:

Viv**imos** en Atlanta → **We** live in Atlanta

I Again, discourse context is often crucial:

Viv**e** en Atlanta → **She/he/it** lives in Atlanta

这块蛋糕很美味。谁烤的?

Zhè kuài dàngāo hěn měiwèi. Shéi kāo de?

This piece cake very beautiful taste. Who bake?

"This cake is very tasty. Who baked **it**?"

Challenges: Tenses

The preterite tense is for events with a definite time, e.g.

I biked to work this morning

The imperfect is for events with indefinite times, e.g.

I biked to work all last summer

To translate English to Spanish, we must pick the right tense.

Challenges: Idioms

As Cool As a Cucumber

Why in the world

Blue in the Face

Lend me your ears

Hold Your Horses

Dead As A Doornail

Kick the bucket

Head In The Clouds

Bob's Your Uncle

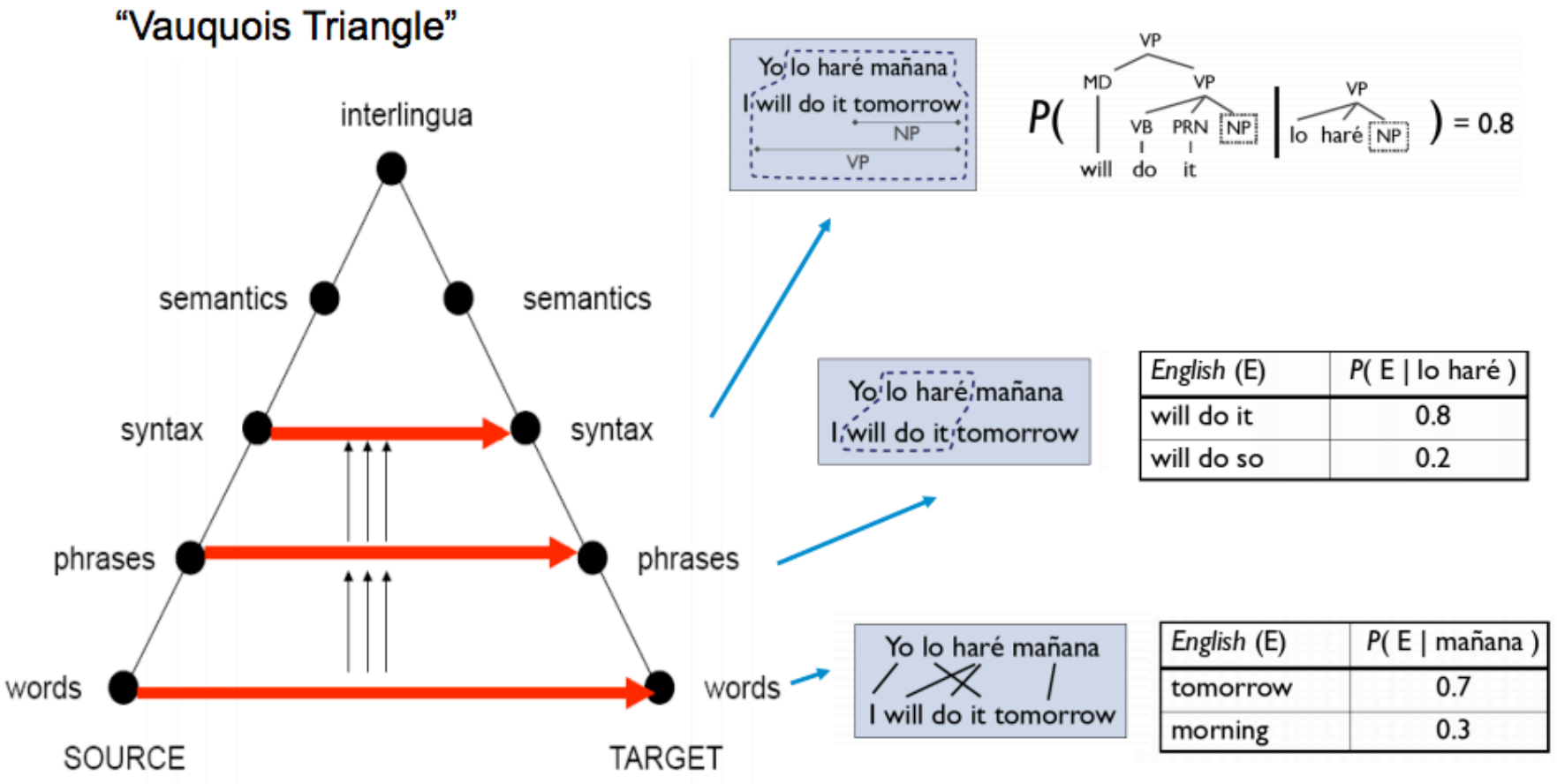
Storm in a Teacup

Rules for Machine Translation

Rules for translating much or many into Russian:

if preceding word is how **return** *skol'ko*
else if preceding word is as **return** *stol'ko zhe*
else if word is much
    if preceding word is very **return** nil
    else if following word is a noun **return** *mnogo*
else (word is many)
    if preceding word is a preposition and following word is noun return *mnogii*
    else return *mnogo*

The Vauquios Triangle



Outline

Machine Translation

Introduction to Statistical MT

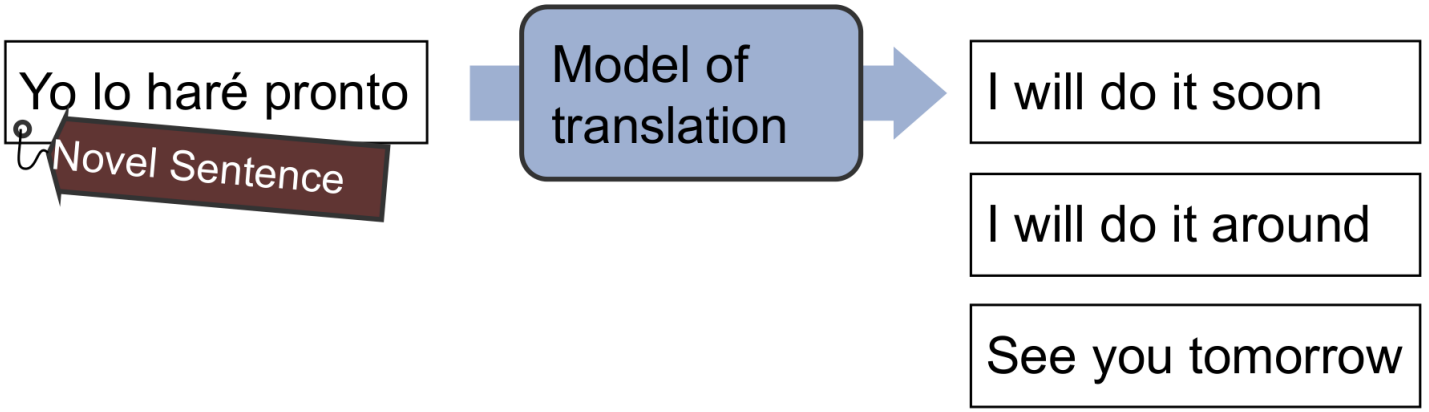
IBM Translation Models

Statistical Machine Translation

Sentence-aligned parallel corpus:

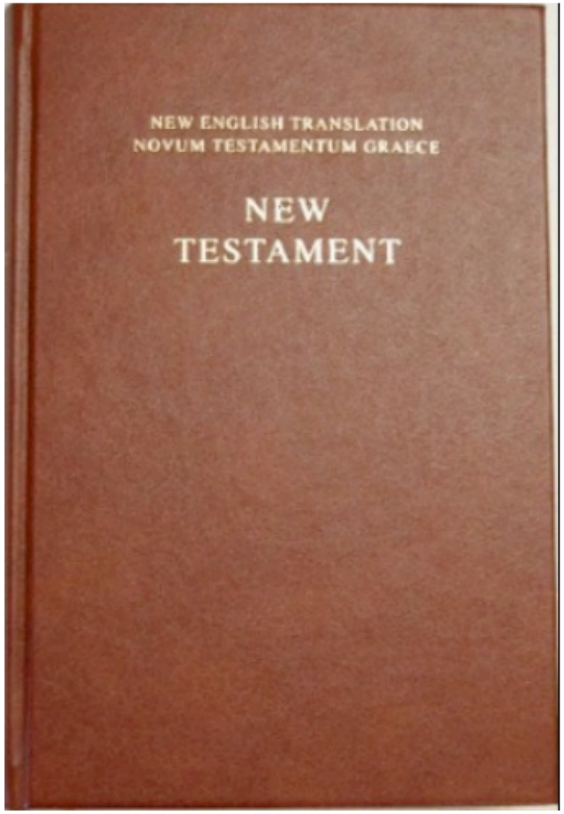
Yo lo haré mañana I will do it tomorrow	Hasta pronto See you soon	Hasta pronto See you around

Machine translation system:

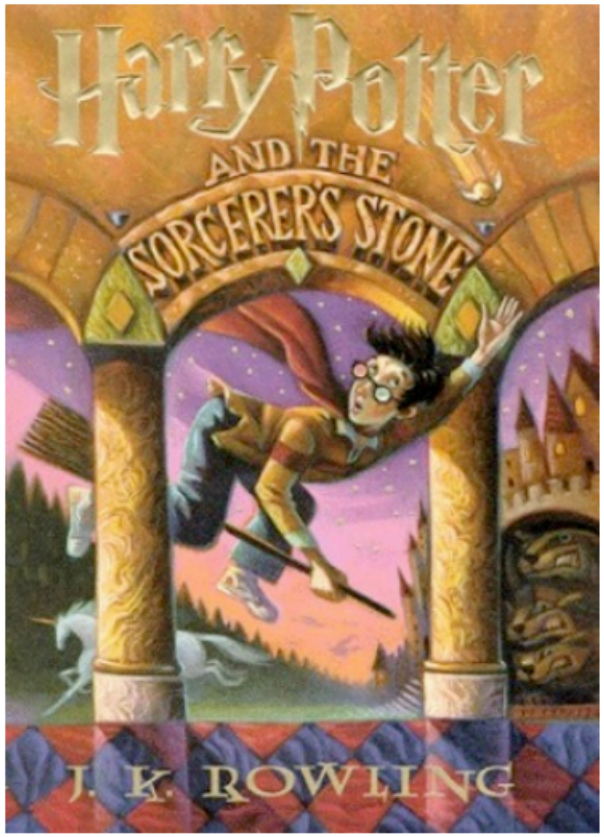


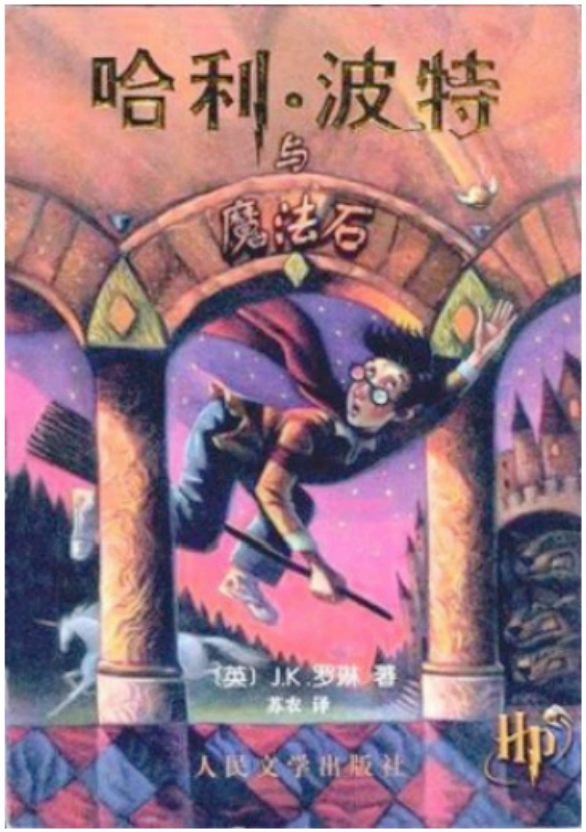
Parallel Corpus: Examples

Parallel Corpus: Examples



Parallel Corpus: Examples





Parallel Corpus: Examples

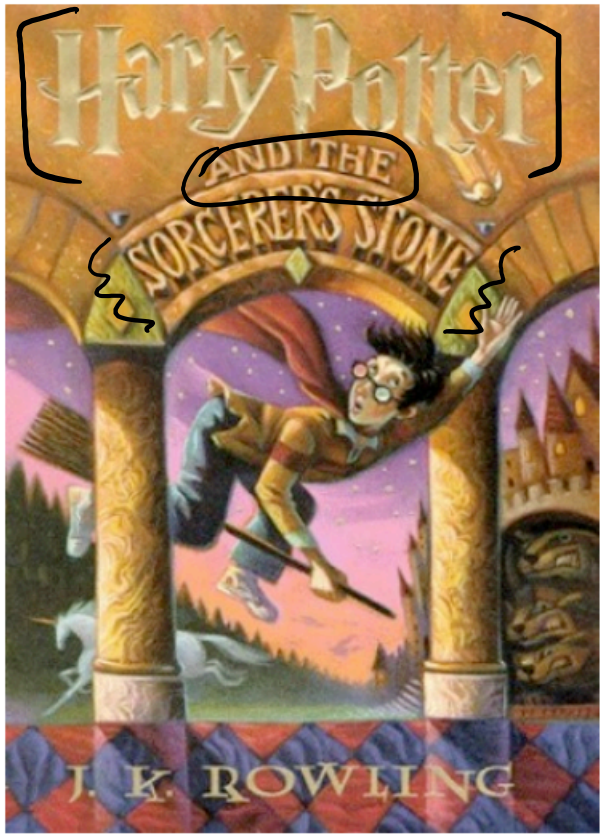
		CLASSIC SOUPS	Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄雷湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
雷花湯	64.	Egg Drop Soup	1.10	2.10
雲雷湯	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

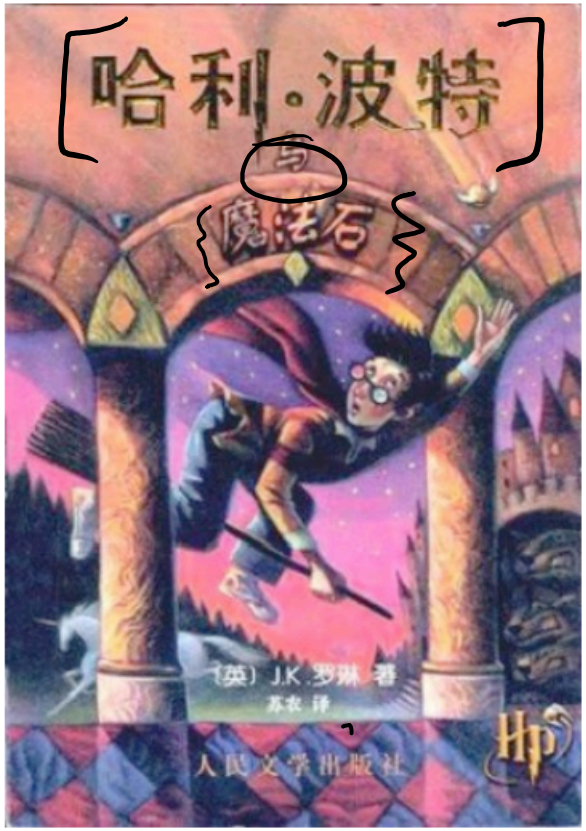
The Rosetta Stone

Warren Weaver (1949)

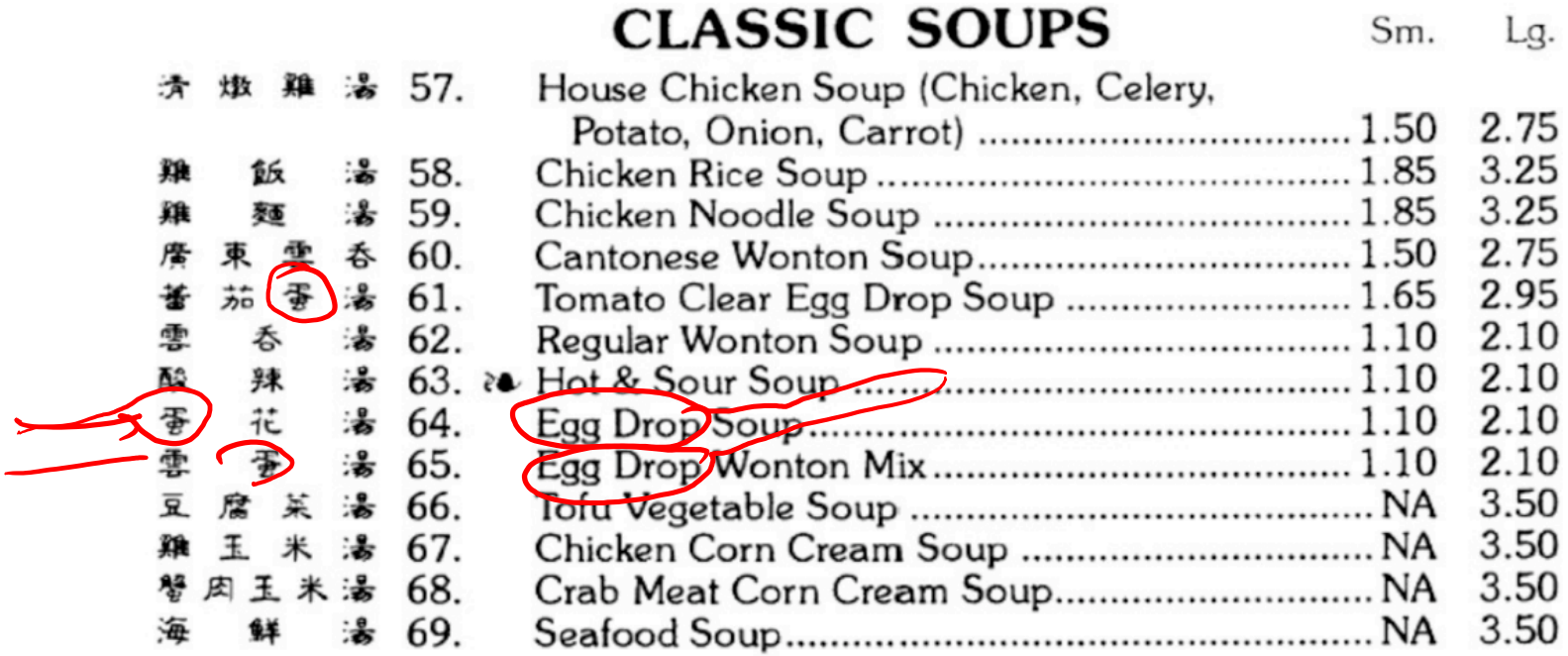
One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: '***This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.***'

Parallel Corpus: Examples



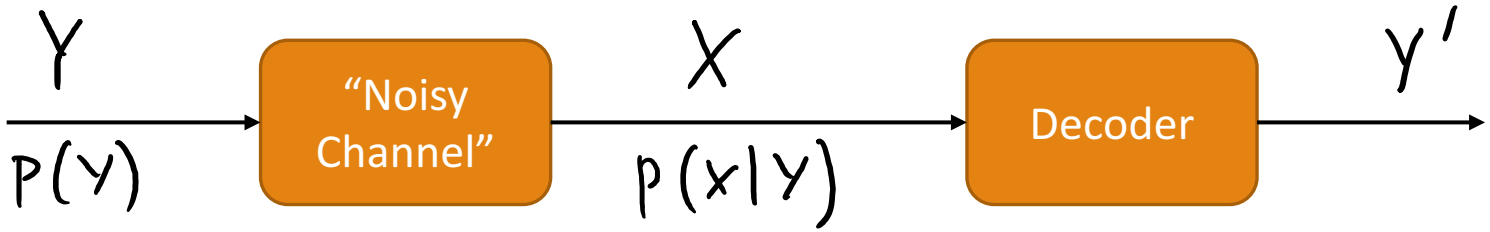


Parallel Corpus: Examples



		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

Noisy Channel Model



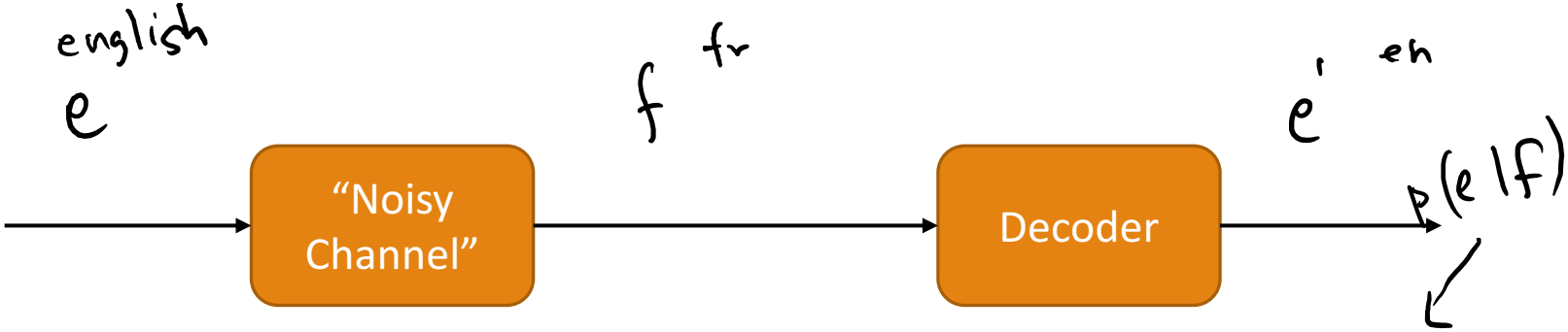
$$y' = \operatorname*{argmax}_y P(y|x)$$

$$= \operatorname*{argmax}_y \frac{P(x|y)\,P(y)}{P(x)}$$

$$= \operatorname*{argmax}_y P(x|y)\,P(y)$$

Noisy Channel Model

fr → en



$$e' = \arg\max_e p(f|e)\, p(e)$$

Translation Model

Language Model

Example: Noisy Channel

Que hambre tengo yo

$\rightarrow$

What hunger have $p(s|e) = 0.000014$

Hungry I am so $p(s|e) = 0.000001$

I am so hungry $p(s|e) = 0.0000015$

Have i that hunger $p(s|e) = 0.000020$

...

Example: Noisy Channel

Que hambre tengo yo

$\rightarrow$

What hunger have	$p(s\|e)p(e) = 0.000014 \times 0.000001$
Hungry I am so	$p(s\|e)p(e) = 0.000001 \times 0.0000014$
I am so hungry	$p(s\|e)p(e) = 0.0000015 \times 0.0001$
Have i that hunger	$p(s\|e)p(e) = 0.000020 \times 0.00000098$

...

Components of an MT system

Language Model

$p(e)$

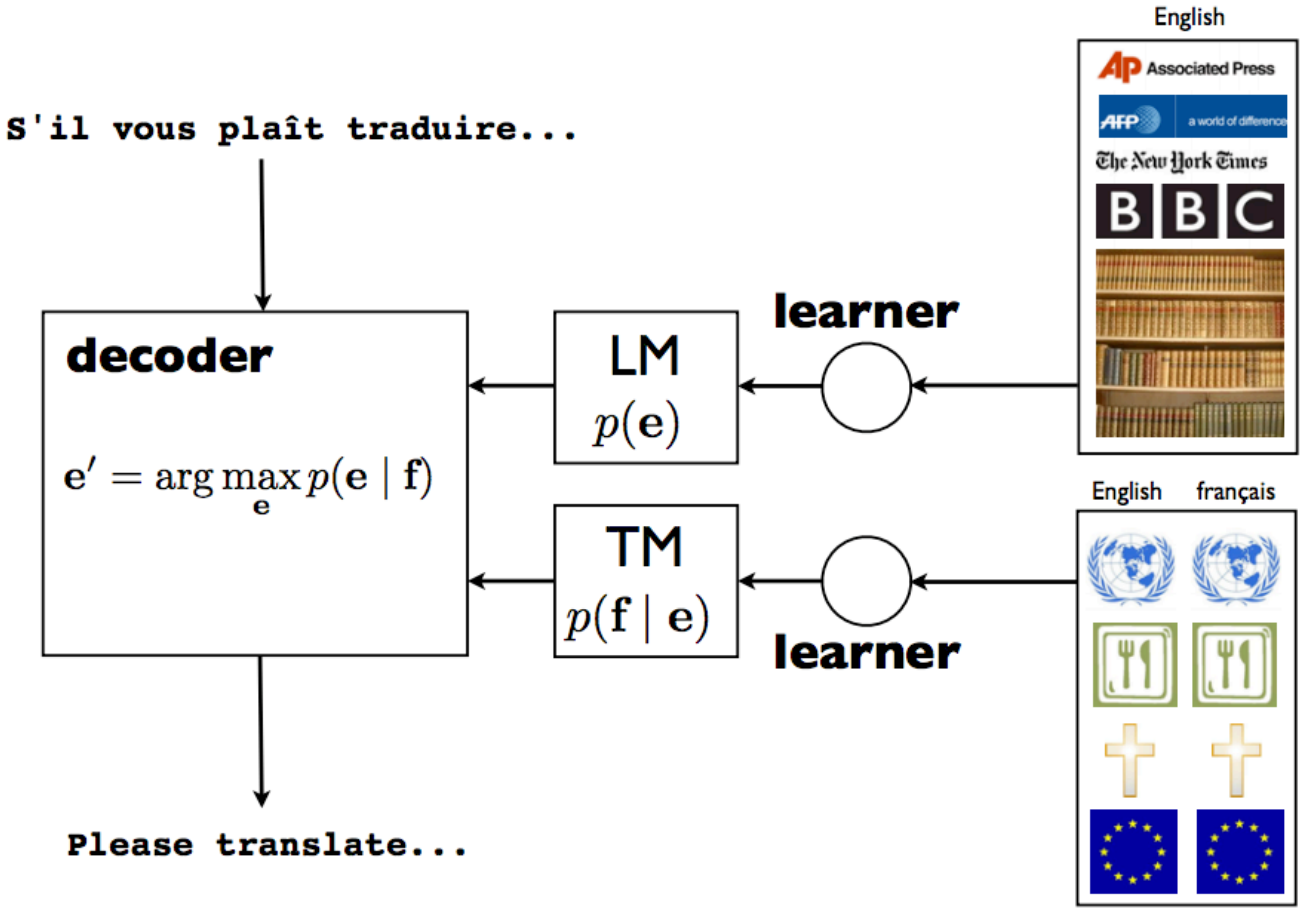
Translation Model

$p(f \mid e)$

Decoding Algo

$\underset{e}{\operatorname{argmax}}\ p(f \mid e)\, p(e)$

Components of an MT system



Evaluating MT

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

- Yorrick Wilks

Human Evaluation

$f \rightarrow e$

Fluency

Adequacy

"Does e look like english?"

"Does it mean same as french!"

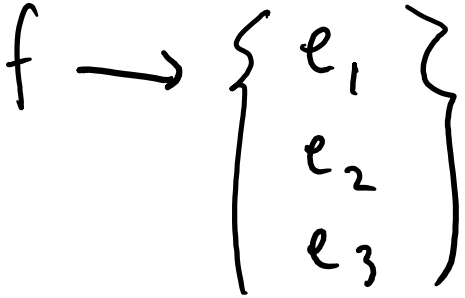
A: furious nAgA on wednesday , the tribal minimum pur of ten schools also was burnt

B: furious nAgA on wednesday the tribal pur mini ten schools of them was also burnt

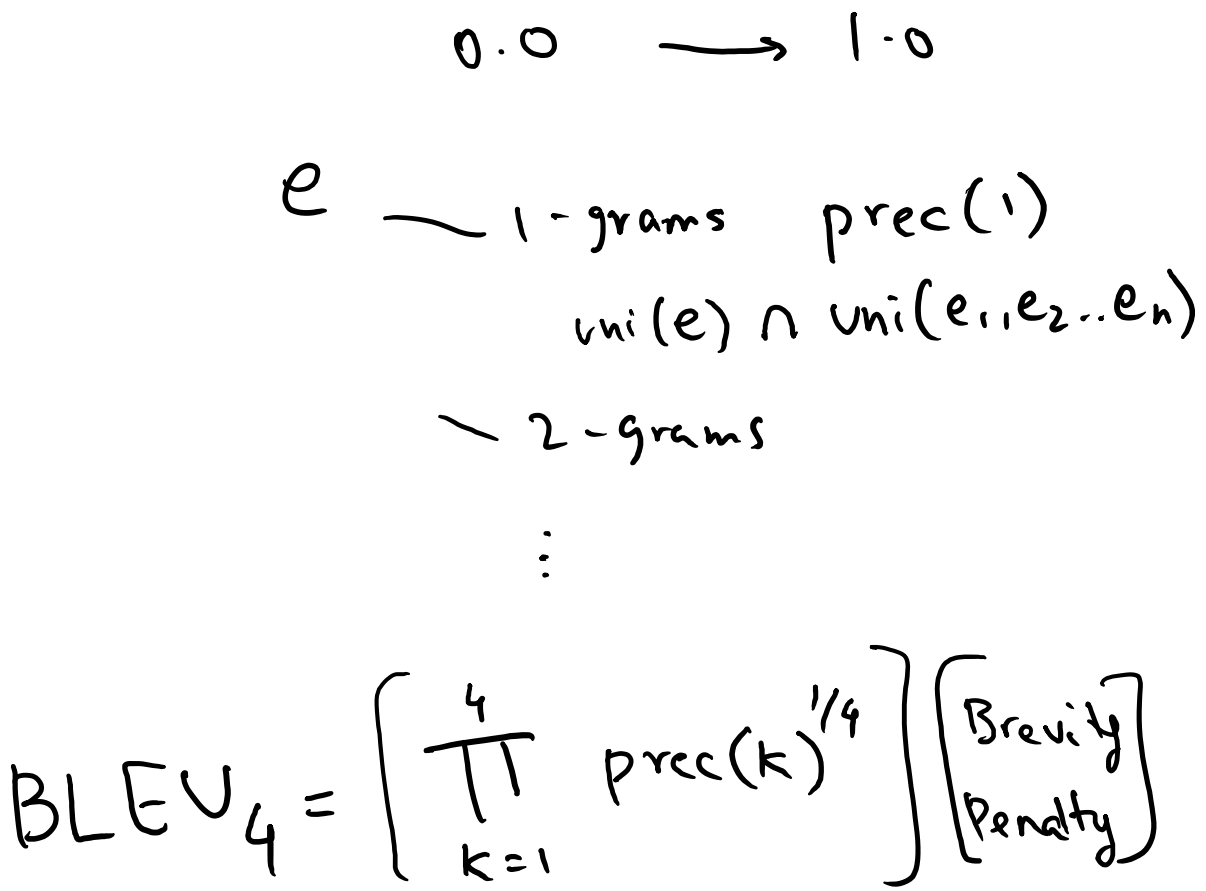
Automated Evaluation

Fluency

Adequacy



BLEU Score


0.0 → 1.0
e
1-grams prec(1)
uni(e) ∩ uni(e1,e2..en)
2-grams
BLEU4 = [∏ k=1 to 4 prec(k)^1/4] [Brevity Penalty]

BLEU Score: Example

e ‘ extension of isi in uttar pradesh ’

e_1 ‘ isi ’s expansion in uttar pradesh ’

e_2 ‘ the spread of isi in uttar pradesh ’

e_3 ‘ isi spreading in uttar pradesh ’

e_4 the spread of isi in uttar pradesh

$$\text{prec}(1) = \frac{1+1+1+1+1+1+1}{1+1+1+1+1+1+1+1} = \frac{7}{8} = 0.875$$

BLEU Score: Example

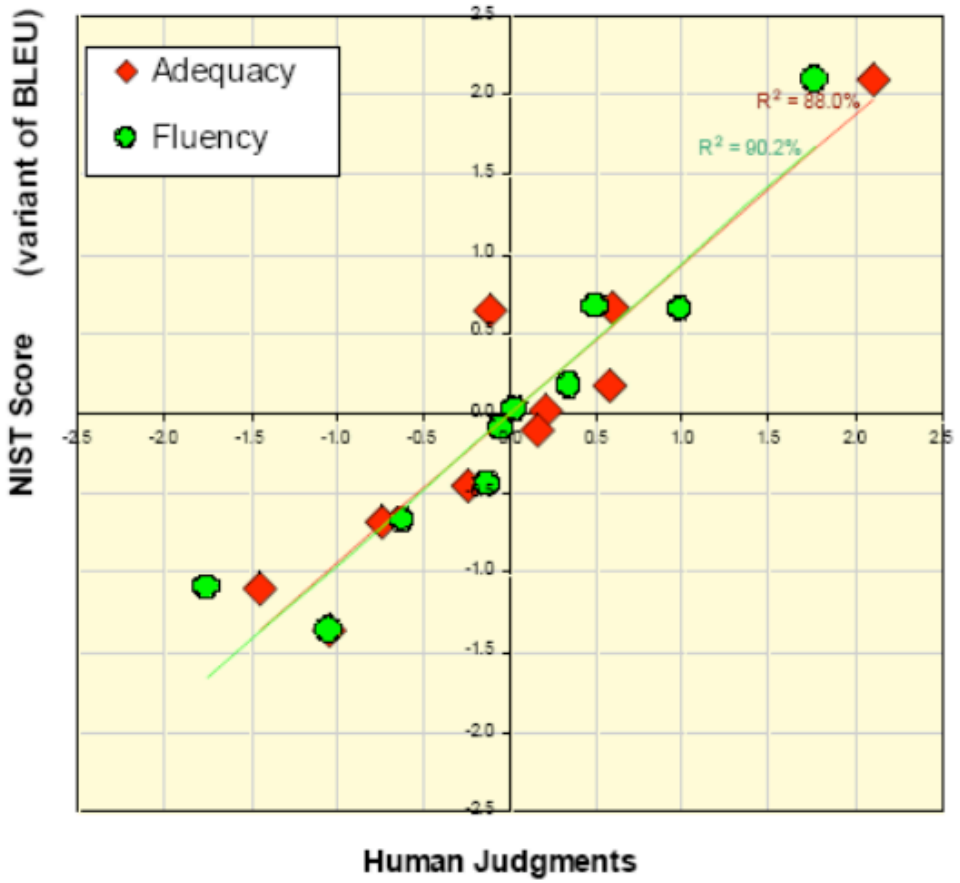
c

‘ extension of isi in uttar pradesh ’

‘ isi ’s expansion in uttar pradesh ’
‘ the spread of isi in uttar pradesh ’
‘ isi spreading in uttar pradesh ’
the spread of isi in uttar pradesh

$prec(2) = \frac{5}{7}$ $prec(3) = \frac{4}{6}$ $prec(4) = \frac{3}{5}$

BLEU's not bad…



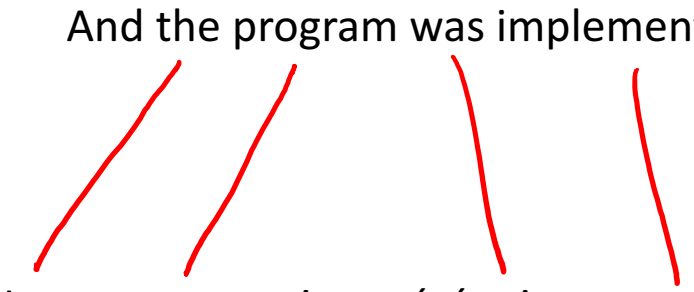
Outline

Machine Translation

Introduction to Statistical MT

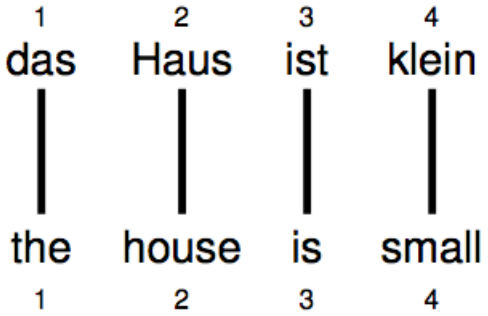
IBM Translation Models

Statistical Translation Model



$p(\text{fle}) = p(\text{la} \mid \text{the}) \; p(\text{programme} \mid \text{program}) \; \ldots$

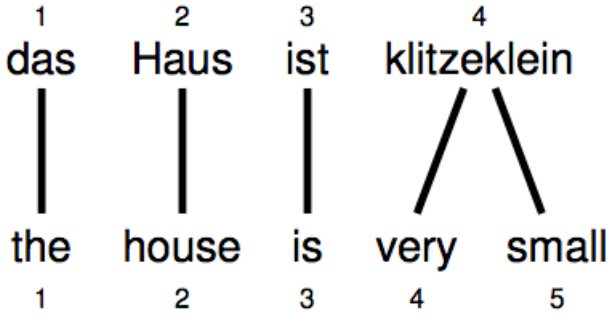
Word Alignment: Direct



Alignment Function

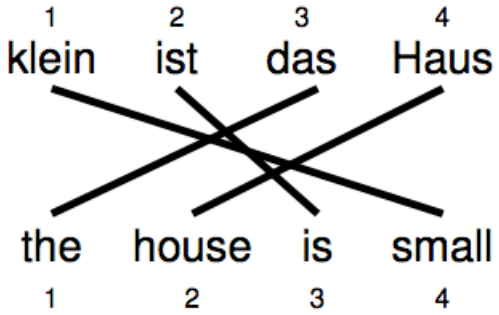
$a = \{1 \to 1, 2 \to 2, 3 \to 3, 4 \to 4\}$

Word Alignment: 1-to-Many



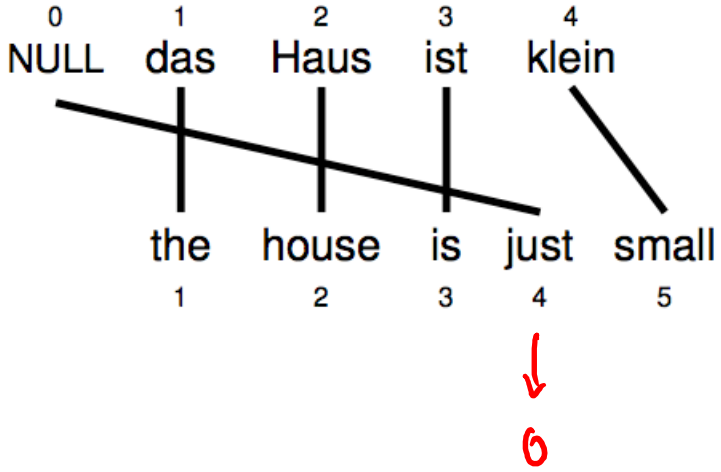
$a = \{1 \to 1, 2 \to 2, 3 \to 3, 4 \to 4, 5 \to 4\}$

Word Alignment: Reordering

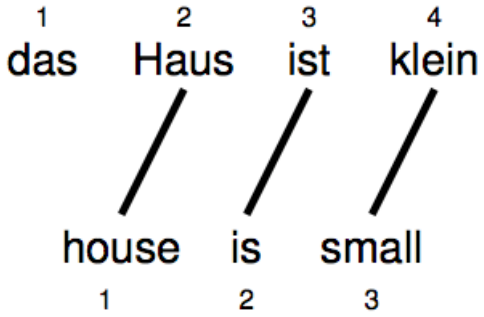


$a = \{1 \to 3, 2 \to 4, 3 \to 2, 4 \to 1\}$

Word Alignment: Inserting



Word Alignment: Dropping

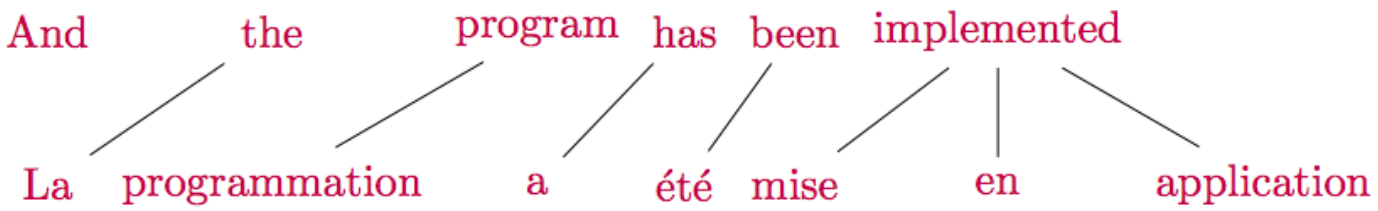


Translating with Alignments

$$p(f|e) = \sum_a p(f, a|e)$$

$$= \pi_i \sum_{a_i} q(a_i|i, l, m)\, t(f_i|e_{a_i})$$

Example: Translation Prob



$$
\begin{aligned}
P(\boldsymbol{f}, \boldsymbol{a}|\boldsymbol{e}) = & q(2|1,6,7) \times t(\text{La}|\text{the}) \\
& \times q(3|2,6,7) \times t(\text{Programmation}|\text{program}) \\
& \times q(4|3,6,7) \times t(\text{a}|\text{has}) \\
& \times q(5|4,6,7) \times t(\text{été}|\text{been}) \\
& \times q(6|5,6,7) \times t(\text{mise}|\text{implemented}) \\
& \times q(6|6,6,7) \times t(\text{en}|\text{implemented}) \\
& \times q(6|7,6,7) \times t(\text{application}|\text{implemented})
\end{aligned}
$$

IBM Models

Model 1

Model 2

Model 3/4/5

$$q(j \mid i, l, m) = \frac{1}{l}$$

$$q(j \mid i, l, m) = \frac{c(j, i, l, m)}{c(i, l, m)}$$

HMMs $q(j-1) \sim q(j)$

add new words ...

Word Alignment Algorithm

$$a^* = \underset{a}{\operatorname{argmax}}\ p(a \mid f, e)$$

q is simple

$$a_i = \underset{j}{\operatorname{argmax}}\ q(j \mid i, l, m)\, t(f_i \mid e_j)$$