CMP462: Natural Language Processing

Lecture 13: Introduction to Machine Translation

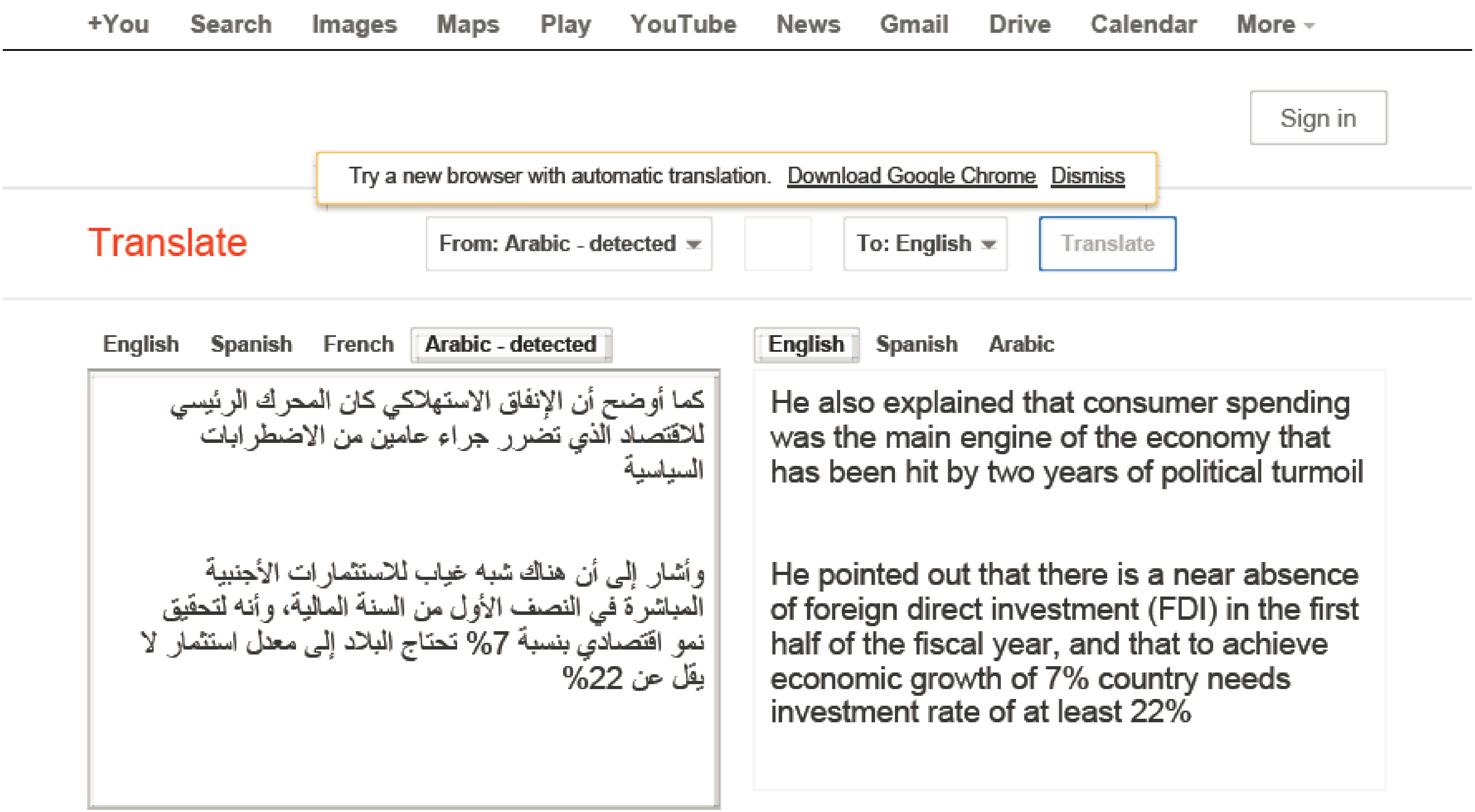
Mohamed Alaa El-Dien Aly
Computer Engineering Department
Cairo University
Spring 2013

Agenda

- Introduction
- Challenges of Machine Translation (MT)
- Classical Approaches
  - Direct MT
  - Transfer Based MT
  - Interlingua-Based MT
- Introduction to Statistical Machine Translation (SMT)

Acknowledgment:
Most slides adapted from Michael Collins NLP class on Coursera.

Introduction



Challenges: Lexical Ambiguity

Example1:

book the flight reservar
read the book libro

Example2:
the box was in the pen
the pen was on the table

Example3:
kill a man matar
kill a process acabar

Challenges: Differing Word Orders

English word order is *subject–verb–object*
Japanese word order is *subject–object–verb*

English: IBM bought Lotus
Japanese: *IBM Lotus bought*

English: Sources said that IBM bought Lotus yesterday
Japanese: *Sources yesterday IBM Lotus bought that said*

Challenges: Syntactic Structure Not Preserved Across Translation

(Example from Dorr et al. 1999)

The bottle floated into the cave

⇓

La botella entro a la cuerva flotando
(the bottle entered the cave floating)

Challenges: Syntactic Ambiguity

(Example from Dorr et al. 1999)

John hit the dog with the stick

$\Downarrow$

John golpeo el perro con el palo/que tenia el palo

(hit with the stick OR the dog with the stick)

Challenges: Pronoun Resolution

(Example from Dorr et al. 1999)

The computer outputs the data; it is fast.

⇓

La computadora imprime los datos; es rapida

The computer outputs the data; it is stored in ascii.

⇓

La computadora imprime los datos; estan almacendos en ascii

Direct Machine Translation

- Translation is word-by-word
- Very little analysis of the source text (e.g.,no syntactic or semantic analysis)
- Relies on a large bilingual dictionary. For each word in the source language, the dictionary specifies a set of rules for translating that word
- After the words are translated, simple reordering rules are applied (e.g. move adjectives after nouns when translating from English to French)

Example of a set of Direct Translation Rules

(From Jurafsky and Martin, edition 2, chapter 25. Originally from a system from Panov 1960)

Rules for translating much or many into Russian:

if preceding word is *how* **return** *skol'ko*

elseif preceding word is *as* **return** *stol'kozhe*

elseif word is *much*

  if preceding word is *very* **return** nil

  elseif following word is a noun **return** *mnogo*

else(wordismany)

  if preceding word is a preposition and following word is noun **return** *mnogii*

  else return *mnogo*

Problems with Direct Machine Translation

- Lack of any analysis of the source language causes several problems, for example:
  - Difficult or impossible to capture long-range reorderings

    English: Sources said that IBM bought Lotus yesterday

    Japanese: Sources yesterday IBM Lotus bought that said

- Words are translated without disambiguation of their syntactic role e.g., *that* can be a complementizer or determiner, and will often be translated differently for these two cases:

  They said *that* ...

  They like *that* ice-cream

Transfer Based Approaches

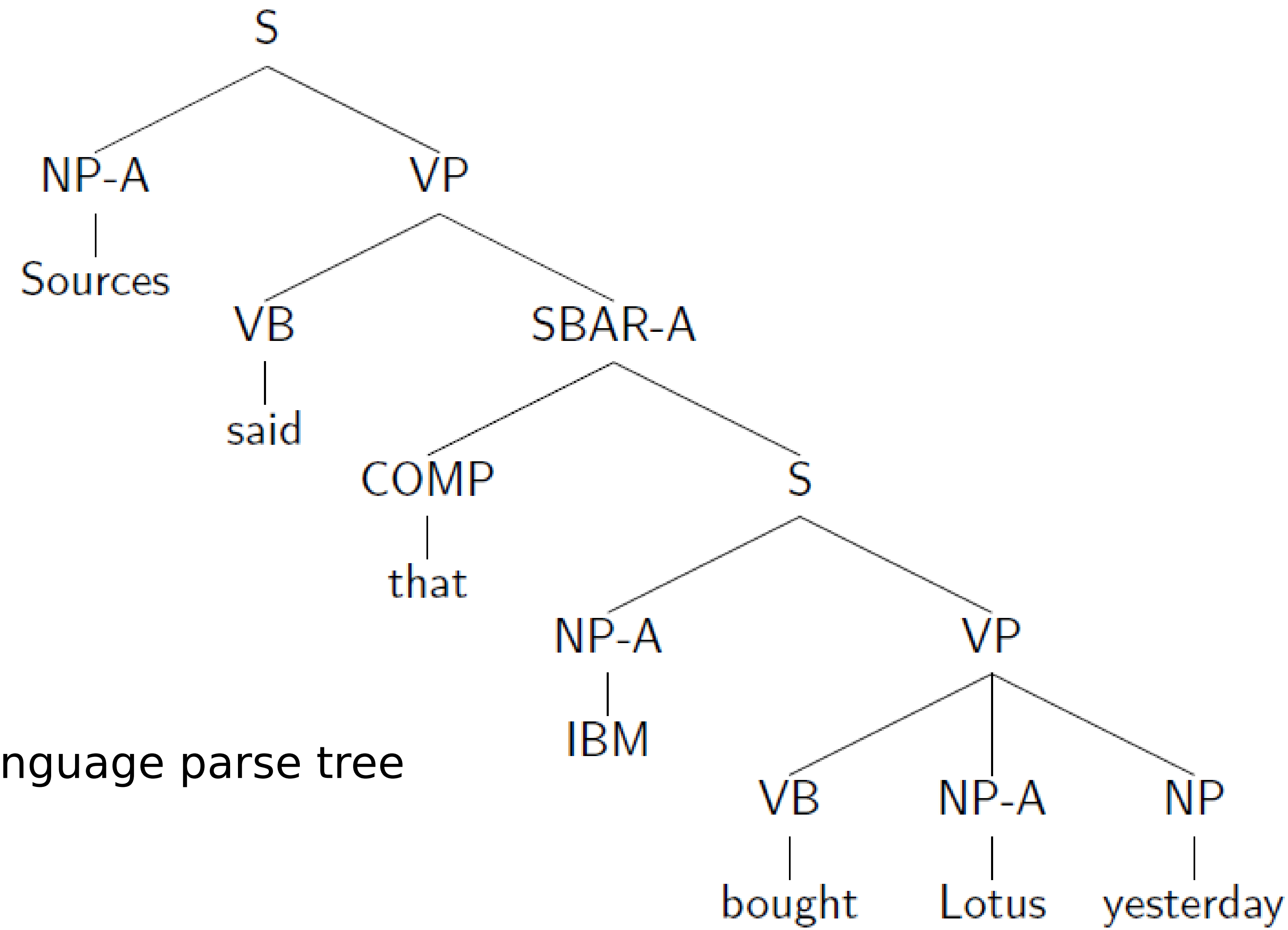
Three phases in translation:

- Analysis: Analyze the source language sentence; for example, build a syntactic analysis of the source language sentence.
- Transfer: Convert the source-language parse tree to a target-language parse tree.
- Generation: Convert the target-language parse tree to an output sentence.

Transfer Based Approaches

- The "parse trees" involved can vary from shallow analyses to much deeper analyses (even semantic representations).
- The transfer rules might look quite similar to the rules for direct translation systems. But they can now operate on syntactic structures.
- It's easier with these approaches to handle long-distance reorderings
- The *Systran* systems are a classic example of this approach

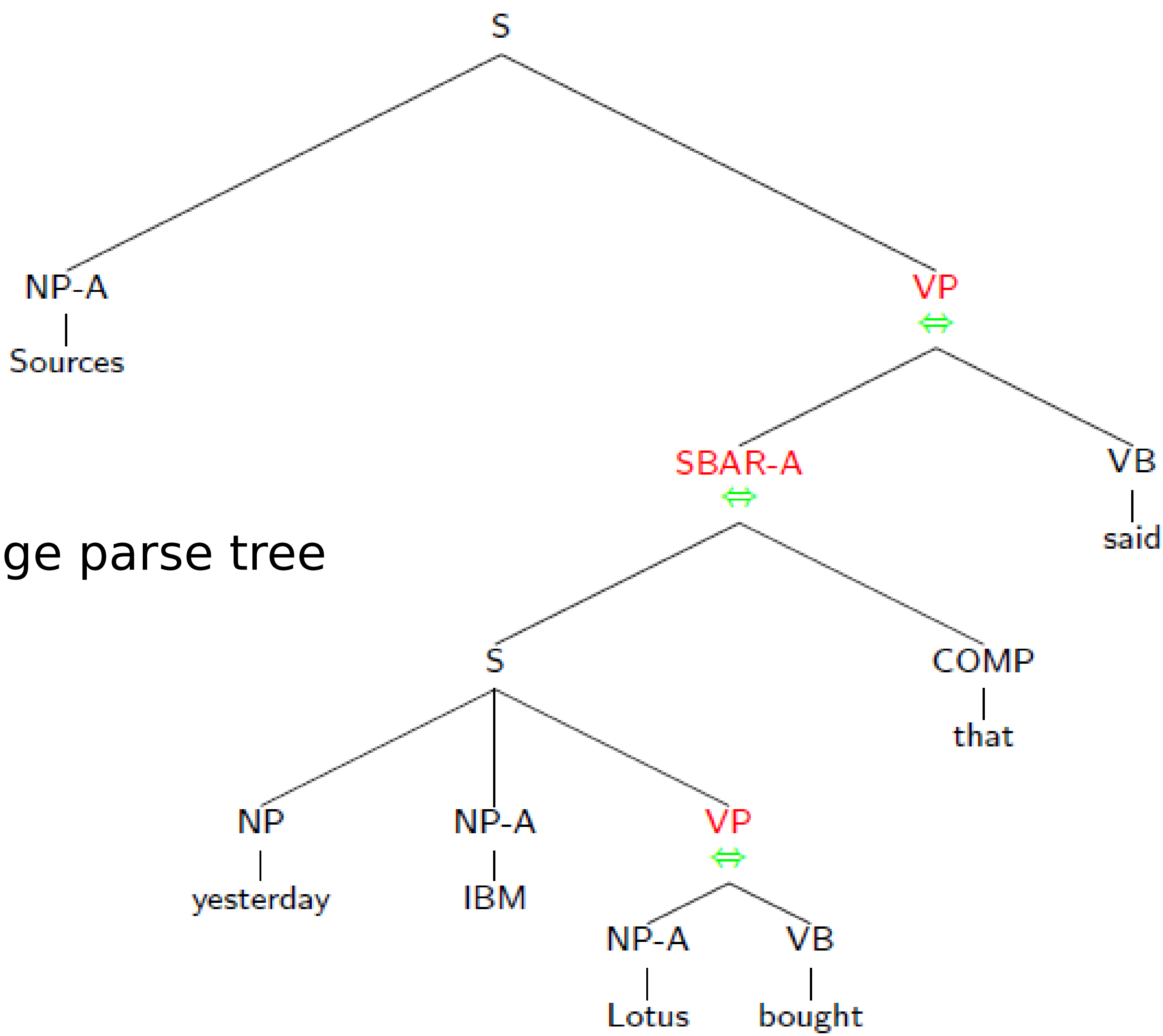
Example



Source language parse tree

English: Sources said that IBM bought Lotus yesterday

Example



Target language parse tree

→ Japanese: Sources yesterday IBM Lotus bought that said

Interlingua-Based Translation

Two phases in translation:

- Analysis: Analyze the source language sentence into a (language-independent) representation of its meaning.
- Generation: Convert the meaning representation into an output sentence.

Interlingua-Based Translation

One Advantage: If we want to build a translation system that translates between n languages, we need to develop n analysis and generation systems. With a transfer based system, we'd need to develop $O(n^2)$ sets of translation rules.

Disadvantage: What would a language-independent representation look like?

Interlingua-Based Translation

- How to represent different concepts in an interlingua?
- Different languages break down concepts in quite different ways:
  - German has two words for *wall*: one for an internal wall, one for a wall that is outside
  - Japanese has two words for *brother*: one for an elder brother, one for a younger brother
  - Spanish has two words for *leg*: pierna for a human's leg, pata for an animal's leg, or the leg of a table

Introduction to Statistical MT

- Parallel corpora are available in several language pairs
- Basic idea: use a parallel corpus as a training set of translation examples
- Classic example: IBM work on French-English translation, using the Canadian Hansards. (1.7 million sentences of 30 words or less in length).
- Idea goes back to Warren Weaver (1949): suggested applying statistical and cryptanalytic techniques to translation.

Introduction to Statistical MT

... 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."

(Warren Weaver, 1949, in a letter to Norbert Wiener)

The Noisy Channel Model

- Goal: translation system from French to English
- Have a model $p(e \mid f)$ which estimates conditional probability of any English sentence e given the French sentence f. Use the training corpus to set the parameters.
- A Noisy Channel Model has two components:
  - $p(e)$ the language model
  - $p(f \mid e)$ the translation model
- Which gives us:

$$p(e|f) = \frac{p(e,f)}{p(f)} = \frac{p(e)\,p(f|e)}{\sum_e p(e)\,p(f|e)}$$

and

$$\text{argmax}_e\, p(e|f) = \text{argmax}_e\, p(e)\,p(f|e)$$

The Noisy Channel Model

- The language model $p(e)$ could be a trigram model, estimated from any data (parallel corpus not needed to estimate the parameters)
- The translation model $p(f \mid e)$ is trained from a parallel corpus of French/English pairs.
- Note:
  - The translation model is backwards!
  - The language model can make up for deficiencies of the translation model.
  - Later we'll talk about how to build $p(f \mid e)$
  - Decoding, i.e. finding $\text{argmax}_e \, p(e)\, p(f \mid e)$ is also a challenging problem.

Example from Koehn and Knight tutorial

Translation from Spanish to English, candidate translations based on $p(Spanish \mid English)$ alone:

Que hambre tengo yo

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 from Koehn and Knight tutorial

With $p(Spanish \mid English) \times p(English)$:

Que hambre tengo yo

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	$p(s\|e)p(e) = 0.000020 \times 0.00000098$

Recap

- Introduction
- Challenges of Machine Translation (MT)
- Classical Approaches
  - Direct MT
  - Transfer Based MT
  - Interlingua-Based MT
- Introduction to Statistical Machine Translation (SMT)