

The Computer and Natural Language (Ling 445/515)

Topic 5: Machine Translation

Markus Dickinson
Dept. of Linguistics, Indiana
Autumn 2008

- Computers and Language
- Topic 5: Machine Translation
- Introduction
- Examples for Translations
- Background: Dictionaries
- Transformer approaches
- Linguistic knowledge based systems
- Direct transfer systems
- Interlingua-based systems
- Machine learning based systems
- Alignment
- Statistical Modeling
- Phrase-Based Translation
- What makes MT hard?
- Evaluating MT systems
- References

1 / 61

What is Machine Translation?

Translation is the process of:

- ▶ moving texts from one (human) language (**source language**) to another (**target language**),
- ▶ in a way that preserves meaning.

Machine translation (MT) automates (part of) the process:

- ▶ Fully automatic translation
- ▶ Computer-aided (human) translation

- Computers and Language
- Topic 5: Machine Translation
- Introduction
- Examples for Translations
- Background: Dictionaries
- Transformer approaches
- Linguistic knowledge based systems
- Direct transfer systems
- Interlingua-based systems
- Machine learning based systems
- Alignment
- Statistical Modeling
- Phrase-Based Translation
- What makes MT hard?
- Evaluating MT systems
- References

2 / 61

What is MT good for?

- ▶ When you need the gist of something and there are no human translators around:
 - ▶ translating e-mails & webpages
 - ▶ obtaining information from sources in multiple languages (e.g., search engines)
- ▶ If you have a limited vocabulary and a small range of sentence types:
 - ▶ translating weather reports
 - ▶ translating technical manuals
 - ▶ translating terms in scientific meetings
 - ▶ determining if certain words or ideas appear in suspected terrorist documents → help pin down which documents need to be looked at closely
- ▶ If you want your human translators to focus on interesting/difficult sentences while avoiding lookup of unknown words and translation of mundane sentences.

- Computers and Language
- Topic 5: Machine Translation

- Introduction
- Examples for Translations
- Background: Dictionaries
- Transformer approaches
- Linguistic knowledge based systems
- Direct transfer systems
- Interlingua-based systems
- Machine learning based systems
- Alignment
- Statistical Modeling
- Phrase-Based Translation
- What makes MT hard?
- Evaluating MT systems
- References

3 / 61

Is MT needed?

- ▶ Translation is of immediate importance for multilingual countries (Canada, India, Switzerland, . . .), international institutions (United Nations, International Monetary Fund, World Trade Organization, . . .), multinational or exporting companies.
- ▶ The European Union used to have 11 official languages, since May 1, 2004 it has 20. All federal laws and other documents have to be translated into all languages.

- Computers and Language
- Topic 5: Machine Translation

- Introduction
- Examples for Translations
- Background: Dictionaries
- Transformer approaches
- Linguistic knowledge based systems
- Direct transfer systems
- Interlingua-based systems
- Machine learning based systems
- Alignment
- Statistical Modeling
- Phrase-Based Translation
- What makes MT hard?
- Evaluating MT systems
- References

4 / 61

Example translations

The simple case

- ▶ It will help to look at a few examples of real translation before talking about how a machine does it.
- ▶ Take the simple Spanish sentence and its English translation below:

(1) Yo hablo español.
I speak_{1st.sg} Spanish
'I speak Spanish.'
- ▶ Words in this example pretty much translate one-for-one
- ▶ But we have to make sure *hablo* matches with *Yo*, i.e., that the subject agrees with the form of the verb.

- Computers and Language
- Topic 5: Machine Translation

- Introduction
- Examples for Translations
- Background: Dictionaries
- Transformer approaches
- Linguistic knowledge based systems
- Direct transfer systems
- Interlingua-based systems
- Machine learning based systems
- Alignment
- Statistical Modeling
- Phrase-Based Translation
- What makes MT hard?
- Evaluating MT systems
- References

5 / 61

Example translations

A slightly more complex case

The order and number of words can differ:

- (2) a. Tú hablas español?
You speak_{2nd.sg} Spanish
'Do you speak Spanish?'
- b. Hablas español?
Speak_{2nd.sg} Spanish
'Do you speak Spanish?'

- Computers and Language
- Topic 5: Machine Translation

- Introduction
- Examples for Translations
- Background: Dictionaries
- Transformer approaches
- Linguistic knowledge based systems
- Direct transfer systems
- Interlingua-based systems
- Machine learning based systems
- Alignment
- Statistical Modeling
- Phrase-Based Translation
- What makes MT hard?
- Evaluating MT systems
- References

6 / 61

What goes into a translation

Some things to note about these examples and thus what we might need to know to translate:

- ▶ Words have to be translated. → dictionaries
- ▶ Words are grouped into meaningful units (cf. our discussion of syntax for grammar checkers).
- ▶ Word order can differ from language to language.
- ▶ The forms of words within a sentence are systematic, e.g., verbs have to be conjugated, etc.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

7 / 61

Different approaches to MT

We'll look at some basic approaches to MT:

- ▶ Transformer systems
- ▶ Systems based on linguistic knowledge
 - ▶ Direct transfer systems
 - ▶ Interlinguas
- ▶ Machine learning approaches, i.e., statistical machine translation (SMT)
 - ▶ SMT is by far the most popular form of MT right now

Most of these use dictionaries in one form or another, so we will start by looking at dictionaries.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

8 / 61

Dictionaries

An MT **dictionary** differs from a "paper" dictionary:

- ▶ must be computer-usable (electronic form, indexed)
- ▶ needs to be able to handle various word inflections
- ▶ contains (syntactic and semantic) restrictions that a word places on other words
 - ▶ e.g., subcategorization information: *give* needs a giver, a person given to, and an object that is given
 - ▶ e.g., selectional restrictions: if *X eats*, *X* must be animate
- ▶ may also contain frequency information
- ▶ can be hierarchically organized, e.g.:
 - ▶ all nouns have person, number, and gender
 - ▶ verbs (unless irregular) conjugate in the past tense by adding *ed*.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

9 / 61

Transformer approaches

- ▶ **Transformer** architectures transform example sentences from one language into another.
- ▶ They consist of
 - ▶ a grammar for the source/input language
 - ▶ a source-to-target language dictionary
 - ▶ source-to-target language rules
- ▶ Note that there is no grammar for the target language, only mappings from the source language.
 - ▶ By their very nature, transformer systems are non-**reversible** because they lack a target language grammar.
 - ▶ These systems do not require much knowledge of the target language, so they are usually very **robust** = they will return a result for nearly any input sentence.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

10 / 61

An example for the transformer approach

We'll walk through a German-to-English example.

- (3) a. Drehen Sie den Knopf eine Position zurück.
b. Turn the button back one position.
 1. Using grammar, assign parts-of-speech & basic structure:
 - (4) Drehen Sie [den Knopf] [eine Position] zurück.
verb pro. art. noun art. noun prep.
 2. Using dictionary, find the target language words
 - (5) Drehen Sie [den Knopf] [eine Position] zurück.
turn you the button one position back
 3. Using source-to-target rules, "translate":
 - ▶ 'turn' and 'back' form one unit.
 - ▶ 'Drehen . . . zurück' is a command, so remove 'you'
- ⇒ End result: *Turn back the button one position.*

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

11 / 61

Linguistic knowledge-based systems

- ▶ Linguistic knowledge-based systems include knowledge of both the source and the target languages.
- ▶ We will look at direct transfer systems and then the more specific instance of interlinguas.
 - ▶ Direct transfer systems
 - ▶ Interlinguas

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

12 / 61

Direct transfer systems

A direct transfer systems consists of:

- ▶ A source language grammar
- ▶ A target language grammar
- ▶ Rules relating source language underlying representation (UR) to target language UR
 - ▶ A direct transfer system has a **transfer component** which relates a source language representation with a target language representation.
 - ▶ This can also be called a **comparative grammar**.

We'll walk through the following French to English example:

(6) Londres plaît à Sam.
 London is pleasing to Sam
 'Sam likes London.'

Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems

Direct transfer systems

Interlingua-based systems

Machine learning based systems

Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

13 / 61

Steps in a transfer system

1. source language grammar analyzes the input and puts it into an **underlying representation** (UR).
Londres plaît à Sam → *Londres plaire Sam* (source UR)
2. The transfer component relates this source language UR (French UR) to a target language UR (English UR).

French UR English UR
 X plaire Y ↔ Eng(Y) like Eng(X)
 (where Eng(X) means the English translation of X)

Londres plaire Sam (source UR) → *Sam like London* (target UR)

3. target language grammar translates the target language UR into an actual target language sentence.
Sam like London → *Sam likes London*

Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems

Direct transfer systems

Interlingua-based systems

Machine learning based systems

Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

14 / 61

Things to note about transfer systems

- ▶ The transfer mechanism is in theory reversible; e.g., the *plaire* rule works in both directions
 - ▶ Not clear if this is desirable: e.g., Dutch *aanvangen* should be translated into English as *begin*, but *begin* should be translated as *beginnen*.
- ▶ Because we have a separate target language grammar, we are able to ensure that the rules of English apply; *like* → *likes*.
- ▶ Word order is handled differently than with transformers: the URs are essentially unordered.
- ▶ The underlying representation can be of various levels of abstraction – words, syntactic trees, meaning representations, etc.; we will talk about this with the **translation triangle**.

Levels of abstraction

- ▶ There are differing levels of abstraction at which transfer can take place. So far we have looked at URs that represent only word information.
- ▶ We can do a full syntactic analysis, which helps us to know how the words in a sentence relate.
- ▶ Or we can do only a partial syntactic analysis, such as representing the dependencies between words.

Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems

Direct transfer systems

Interlingua-based systems

Machine learning based systems

Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

15 / 61

Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems

Direct transfer systems

Interlingua-based systems

Machine learning based systems

Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

16 / 61

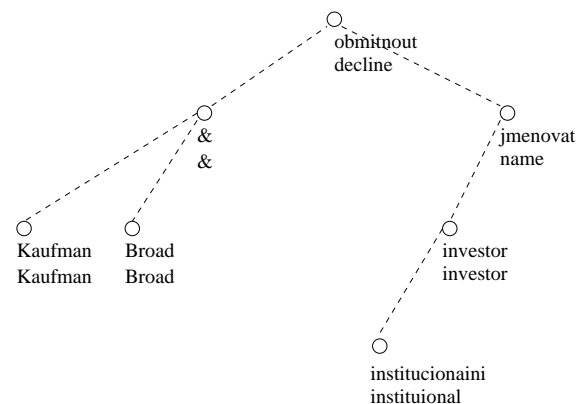
Czech-English example

(7) Kaufman & Broad odmítla institucionální investory
 Kaufman & Broad declined institutional investors
 to name/identify
 'Kaufman & Broad refused to name the institutional investors.'

Example taken from Čmejrek, Cuřín, and Havelka (2003).

- ▶ They find the base forms of words (e.g., *obmidout* 'to decline' instead of *odmítla* 'declined')
- ▶ They find which words depend on which other words and represent this in a tree (e.g., the noun *investory* depends on the verb *jmenovat*)
- ▶ This dependency tree is then converted to English (comparative grammar) and re-ordered as appropriate.

Dependency tree for Czech-English example



Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems

Direct transfer systems

Interlingua-based systems

Machine learning based systems

Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

17 / 61

Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems

Direct transfer systems

Interlingua-based systems

Machine learning based systems

Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

18 / 61

Interlinguas

- ▶ Ideally, we could use an **interlingua** = a language-independent representation of meaning.
- ▶ **Benefit:** To add new languages to your MT system, you merely have to provide mapping rules between your language and the interlingua, and then you can translate into any other language in your system.
- ▶ What your interlingua looks like depends on your goals; an example for *I shot the sheriff.* is shown on the following slide.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems

Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

19 / 61

Interlingual problems

- ▶ What exactly should be represented in the interlingua?
 - ▶ e.g., English *corner* = Spanish *rincón* = 'inside corner' or *esquina* = 'outside corner'
- ▶ A fine-grained interlingua can require extra (unnecessary) work:
 - ▶ e.g., Japanese distinguishes *older brother* from *younger brother*, so we have to disambiguate English *brother* to put it into the interlingua.
 - ▶ Then, if we translate into French, we have to ignore the disambiguation and simply translate it as *frère*, which simply means 'brother'.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems

Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

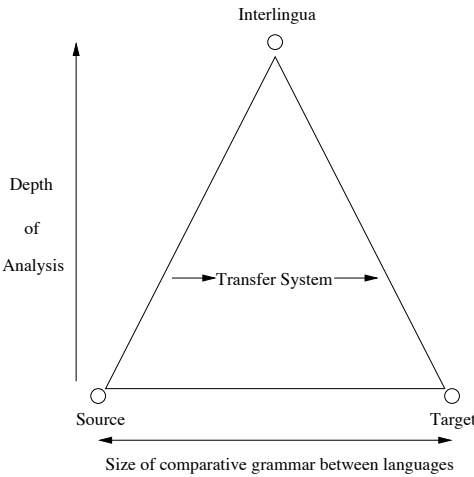
What makes MT hard?

Evaluating MT systems

References

20 / 61

The translation triangle



Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems

Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

21 / 61

Machine learning

- ▶ Instead of trying to tell the MT system how we're going to translate, we might try a **machine learning** approach
 - ▶ the computer will learn how to translate based on example translations.
- ▶ For this, we need
 - ▶ examples of translations as **training data**, and
 - ▶ a way of learning from that data.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

22 / 61

Using frequency (statistical methods)

- ▶ We can look at how often a source language word is translated as a target language word, i.e., the **frequency** of a given translation, and choose the most frequent translation.
 - ▶ But how can we tell what a word is being translated as? There are two different cases:
 - ▶ We are told what each word is translated as: **text alignment**
 - ▶ We are not told what each word is translated as: use a **bag of words**
- We can also attempt to learn alignments, as a part of the process, as we will see.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

23 / 61

Sentence alignment

- ▶ **sentence alignment** = determine which source language sentences align with which target language ones (what we assumed in the bag of words example).
- ▶ Intuitively easy, but can be difficult in practice since different languages have different punctuation conventions.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

24 / 61

Word alignment

- ▶ **word alignment** = determine which source language words align with which target language ones
 - ▶ Much harder than sentence alignment to do automatically.
 - ▶ But if it has already been done for us, it gives us good information about what a word's translation equivalent is.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

25 / 61

Different word alignments

- ▶ One word can map to one word or to multiple words. Likewise, sometimes it is best for multiple words to align with multiple words.
- ▶ English-Russian examples:
 - ▶ one-to-one: *khorosho* = *well*
 - ▶ one-to-many: *kniga* = *the book*
 - ▶ many-to-one: *to take a walk* = *gulyat'*
 - ▶ many-to-many: *at least* = *khotya by* ('although if/would')

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

26 / 61

Calculating probabilities

- ▶ With word alignments, it is relatively easy to calculate probabilities.
- ▶ e.g., What is the probability that *run* translates as *correr* in Spanish?
 1. Count up how many times *run* appears in the English part of your bi-text. e.g., 500 times
 2. Out of all those times, count up how many times it was translated as (i.e., aligns with) *correr*. e.g., 275 (out of 500) times.
 3. Divide to get a probability: $275/500 = 0.55$, or 55%
- ▶ Word alignment gives us some frequency numbers, which we can use to align new cases, using other information, too (e.g., contextual information)

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

27 / 61

Word alignment difficulties

- ▶ Sometimes it is not clear that word alignment is possible.

(8) Ivan aspirant.
Ivan graduate student
'Ivan is a graduate student.'
- ▶ What does *is* align with?
- ▶ In cases like this, a word can be mapped to a "null" element in the other language.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

28 / 61

The "bag of words" method

- ▶ What if we're not given word alignments?
- ▶ How can we tell which English words are translated as which German words if we are only given an English text and a corresponding German text?
 - ▶ We can treat each sentence as a **bag of words** = unordered collection of words.
 - ▶ If word A appears in a sentence, then we will record all of the words in the corresponding sentence in the other language as appearing with it.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

29 / 61

Example for bag of words method

- ▶ English *He speaks Russian well.*
- ▶ Russian *On khorosho govorit po-russki.*

Eng	Rus	Eng	Rus
He	On	speaks	On
He	khorosho	speaks	khorosho
He	govorit
He	po-russki	well	po-russki

The idea is that, over thousands, or even millions, of sentences, *He* will tend to appear more often with *On*, *speaks* will appear with *govorit*, and so on.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

30 / 61

Example for bag of words method

Calculating probabilities: sentence 1

So, for *He* in *He speaks Russian well/On khorosho govorit po-russki*, we do the following:

1. Count up the number of Russian words: 4.
2. Assign each word equal probability of translation: $1/4 = 0/25$, or 25%.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

Example for bag of words method

Calculating probabilities: sentence 2

If we also have *He is nice./On simpatich'nyi.*, then for *He*, we do the following:

1. Count up the number of possible translation words: 4 from the first sentence, 2 from the second = 6 total.
 - Note that we are NOT counting the number of English words: we count the number of *possible translations*
2. Count up the number of times *On* is the translation = 2 times out of 6 = $1/3 = 0.33$, or 33%.

Every other word has the probability $1/6 = 0.17$, or 17%, so *On* is clearly the best translation for *He*.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

Probabilities used in IBM models

Probabilistic models are generally more sophisticated, treating the problem as the source language generating the target and taking into account probabilities such as:

- $n(\#|word)$ = probability of the number of words in the target language that the source word generates
- $pnull$ = probability of a null word appearing
- $t(word|sword)$ = probability of a target word, given the source word [i.e., what we've seen so far]
- $d(tposition|sposition)$ = probability of a target word appearing in position *tposition*, given the source position *sposition*

But we need alignments to estimate these parameters.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

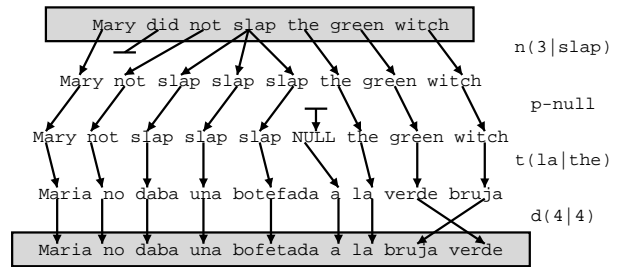
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

A Generative Story (IBM Models)



Source: Introduction to Statistical Machine Translation, Chris Callison-Burch and Philipp Koehn, <http://www.iccs.inf.ed.ac.uk/~pkoeHN/publications/esslli-slides-day3.pdf>

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

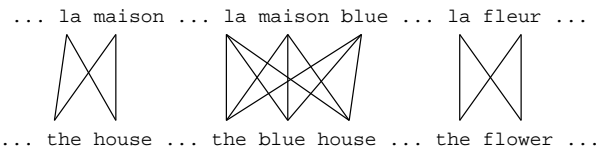
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

Beyond Bags of Words



A chicken-and-egg problem

- If we had the word alignments, we could estimate the parameters of our generative story.
- If we had the parameters, we could estimate the alignments.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

Expectation Maximization Algorithm

The Expectation Maximization (EM) algorithm works forwards and backwards to estimate the probabilities:

EM in a nutshell

1. initialize model parameters (e.g. uniform)
2. (re-)assign probabilities to the missing data
3. (re-)estimate model parameters from completed data (**weighted counts**)
4. iterate, i.e., repeat steps 2&3 until you hit some stopping point

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems

Alignment

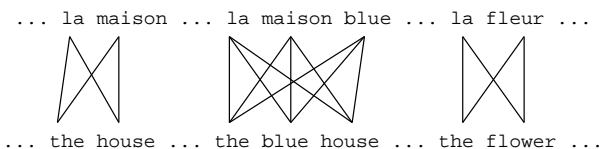
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

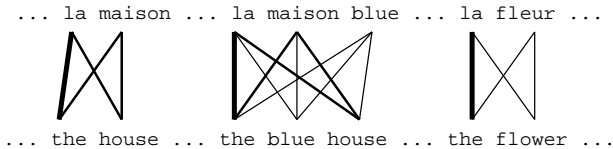
Initial Step



► All connections equally likely.

Computers and Language
 Topic 5: Machine Translation
 Introduction
 Examples for Translations
 Background: Dictionaries
 Transformer approaches
 Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems
 Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation
 What makes MT hard?
 Evaluating MT systems
 References

After 1st Iteration



► Connections between e.g. *la* and *the* are more likely.

Computers and Language
 Topic 5: Machine Translation
 Introduction
 Examples for Translations
 Background: Dictionaries
 Transformer approaches
 Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems
 Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation
 What makes MT hard?
 Evaluating MT systems
 References

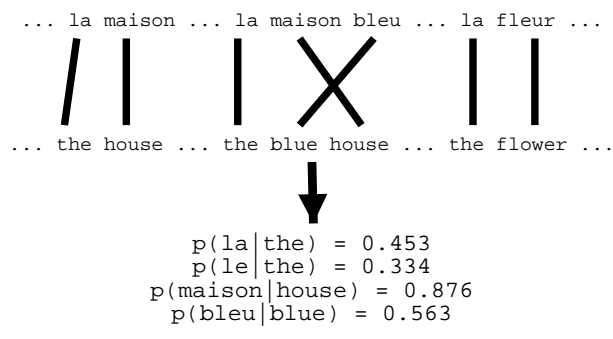
After Another Iteration



► Connections between e.g. *fleur* and *flower* are more likely (pigeon hole principle).

Computers and Language
 Topic 5: Machine Translation
 Introduction
 Examples for Translations
 Background: Dictionaries
 Transformer approaches
 Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems
 Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation
 What makes MT hard?
 Evaluating MT systems
 References

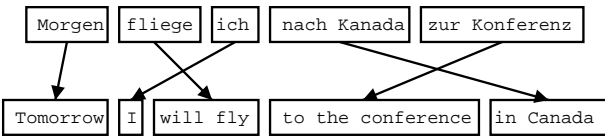
Convergence



Computers and Language
 Topic 5: Machine Translation
 Introduction
 Examples for Translations
 Background: Dictionaries
 Transformer approaches
 Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems
 Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation
 What makes MT hard?
 Evaluating MT systems
 References

Phrase-Based Translation Overview

But this word-based translation doesn't account for many-to-many mappings between languages

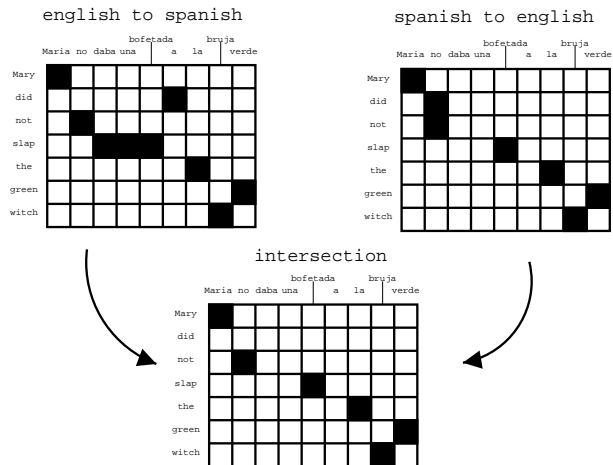


► Foreign "phrases" are translated into English.
 ► Phrases may be reordered.

Current models allow for many-to-one mappings → we can use those to induce many-to-many mappings

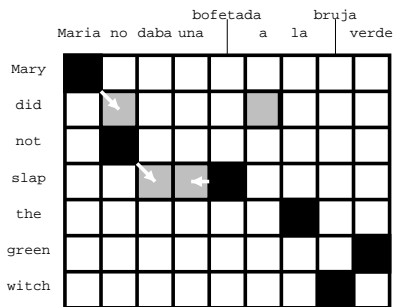
Computers and Language
 Topic 5: Machine Translation
 Introduction
 Examples for Translations
 Background: Dictionaries
 Transformer approaches
 Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems
 Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation
 What makes MT hard?
 Evaluating MT systems
 References

Intersecting Alignments



Computers and Language
 Topic 5: Machine Translation
 Introduction
 Examples for Translations
 Background: Dictionaries
 Transformer approaches
 Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems
 Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation
 What makes MT hard?
 Evaluating MT systems
 References

Growing Alignments



- Heuristically add alignments along the diagonal (Och & Ney, *Computational Linguistics*, 2003)

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

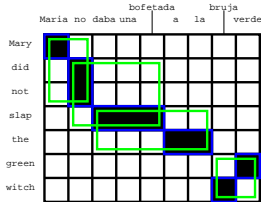
What makes MT hard?

Evaluating MT systems

References

43 / 61

Induced Phrases



(Maria, Mary), (no, did not), (slap, daba una bofetada), (a la, the), (bruja, witch), (verde, green), (Maria no, Mary did not), (no daba una bofetada, did not slap), (daba una bofetada a la, slap the), (bruja verde, green witch)

We can now use these phrase pairs as the units of our probability model.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

44 / 61

Advantages of Phrase-Based Translation

- Many-to-many translation can handle non-compositional phrases.
- Use of local context.
- The more data, the longer the phrases that can be learned.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

45 / 61

What makes MT hard?

We've seen how MT systems can work, but MT is a very difficult task because languages are vastly different. They differ:

- Lexically: In the words they use
- Syntactically: In the constructions they allow
- Semantically: In the way meanings work
- Pragmatically: In what readers take from a sentence.

In addition, there is a good deal of real-world knowledge that goes into a translation.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

46 / 61

Lexical ambiguity

Words can be **lexically ambiguous** = have multiple meanings.

- *bank* can be a financial institution or a place along a river.
- *can* can be a cylindrical object, as well as the act of putting something into that cylinder (e.g., *John cans tuna.*), as well as being a word like *must*, *might*, or *should*.

⇒ We have to know which meaning before we translate.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

47 / 61

Semantic relationships

Often we find (rough) **synonyms** between two languages:

- English *book* = Russian *kniga*
- English *music* = Spanish *música*

But words don't always line up exactly between languages.

- English **hypernyms** = words that are more general in English than in their counterparts in other languages
 - English *know* is rendered by the French *savoir* ('to know a fact') and *connaître* ('to know a thing')
 - English *library* is German *Bücherei* if it is open to the public, but *Bibliothek* if it is intended for scholarly work.
- English **hyponyms** = words that are more specific in English than in their foreign language counterparts.
 - The German word *berg* can mean either *hill* or *mountain* in English.
 - The Russian word *ruka* can mean either *hand* or *arm*.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

48 / 61

Semantic overlap

And then there's just fuzziness, as in the following English and French correspondences (Jurafsky & Martin 2000, Figure 21.2)

- ▶ *leg* = *etape* (journey), *jambe* (human), *pied* (chair), *patte* (animal)
- ▶ *foot* = *pied* (human), *patte* (bird)
- ▶ *paw* = *patte* (animal)

Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems

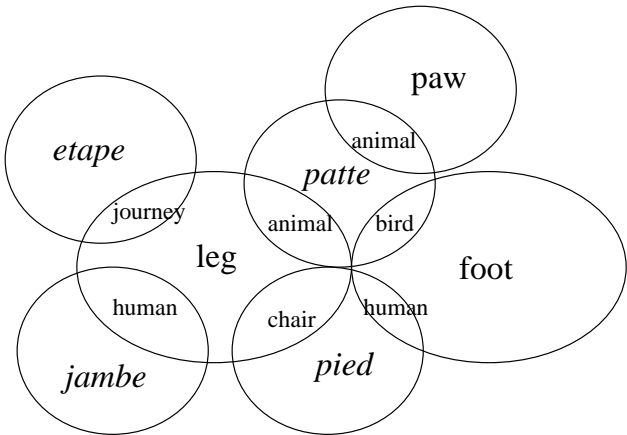
Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

Venn diagram of semantic overlap



Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems

Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

Semantic non-compositionality

Some verbs carry little meaning, so-called **light verbs**

- ▶ French *faire une promenade* is literally 'make a walk,' but it has the meaning of the English *take a walk*
- ▶ Dutch *een poging doen* 'do an attempt' means the same as the English *make an attempt*

And we often face **idioms** = expressions whose meaning is not made up of the meanings of the individual words.

- ▶ e.g., English *kick the bucket*
 - ▶ approximately equivalent to the French *casser sa pipe* ('break his/her pipe')
 - ▶ but we might want to translate it as *mourir* ('die')
 - ▶ and we want to treat it differently than *kick the table*

Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems

Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

Idiosyncratic differences

Some words do not exist in a language and have to be translated with a more complex phrase: **lexical gap** or **lexical hole**.

- ▶ French *gratiner* means something like 'to cook with a cheese coating'
- ▶ Hebrew *stam* means something like 'I'm just kidding' or 'Nothing special.'

There are also idiosyncratic **collocations** among languages, e.g.:

- ▶ English *heavy smoker*
- ▶ French *grand fumeur* ('large smoker')
- ▶ German *starker Raucher* ('strong smoker')

Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems

Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

Structure and word order differences

- ▶ Word order (and syntactic structure) differs across languages.

- ▶ E.g., in English, we have what is called a subject-verb-object (SVO) order, as in (9).

(9) John punched Bill.
 SUBJECT VERB OBJECT

- ▶ In contrast, Japanese is SOV. Arabic is VSO. Dyrbal (Australian aboriginal language) has free word order.
- ▶ MT systems have to account for these differences.

Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems

Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

More on word order differences

- ▶ Sometimes things are conceptualized differently in different languages, e.g.:

- (10) a. His name is Jerome.
 b. Er heißt Jerome. (German)
 He goes-by-name-of Jerome
 c. Il s' appelle Jerome. (French)
 He himself call Jerome.

- ▶ Words don't really align here.

Computers and Language
 Topic 5: Machine Translation

Introduction
 Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
 Direct transfer systems
 Interlingua-based systems

Machine learning based systems
 Alignment
 Statistical Modeling
 Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

How syntactic grouping and meaning relate (Syntax/Semantics)

Even within a language, there are syntactic complications. We can have **structural ambiguities** = sentences where there are multiple ways of interpreting it.

(11) John saw the boy with the binoculars.

with the binoculars can refer to either *the boy* or to how John saw the boy.

- ▶ This difference in structure corresponds to a difference in what we think the sentence means
- ▶ Do we translate only one interpretation? Or do we try to preserve the ambiguity in the target language?

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

55 / 61

How language is used (Pragmatics)

Translation becomes even more difficult when we try to translate something in context.

- ▶ *Thank you* is usually translated as *merci* in French, but it is translated as *s'il vous plaît* 'please' when responding to an offer.
- ▶ *Can you drive a stick-shift?* could be a request for you to drive my manual transmission automobile, or it could simply be a request for information about your driving abilities.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

56 / 61

Real-world knowledge

- ▶ Sometimes we have to use **real-world knowledge** to figure out what a sentence means.

(12) Put the paper in the printer. Then switch it on.

- ▶ We know what *it* refers to only because we know that printers, not paper, can be switched on.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

57 / 61

Evaluating MT systems

- ▶ We've seen some translation systems and we know that translation is hard.
- ▶ The question now is: How do we evaluate MT systems, in particular for use in large corporations as likely users?
 - ▶ How much change in the current setup will the MT system force?
 - ▶ How will it fit in with word processors and other software?
 - ▶ Will the company selling the MT system be around in the next few years for support and updates?
 - ▶ How fast is the MT system?
 - ▶ How good is the MT system (quality)?

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

58 / 61

Evaluating quality

Two main components in evaluating quality:

- ▶ **Intelligibility** = how understandable the output is
- ▶ **Accuracy** = how faithful the output is to the input
 - ▶ A common (though problematic) evaluation metric is the BLEU metric, based on *n*-gram comparisons

And some methods we can use to gauge these properties:

- ▶ **Error analysis** = how many errors we have to sort through and how they affect intelligibility & accuracy
- ▶ **Test suite** = a set of sentences that our system should be able to handle

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

59 / 61

Intelligibility

Intelligibility Scale (from Arnold et al., 1994)

1. The sentence is perfectly clear and intelligible. It is grammatical and reads like ordinary text.
2. The sentence is generally clear and intelligible. Despite some inaccuracies or infelicities of the sentence, one can understand (almost) immediately what it means.
3. The general idea of the sentence is intelligible only after considerable study. The sentence contains grammatical errors and/or poor word choices.
4. The sentence is unintelligible. Studying the meaning of the sentence is hopeless; even allowing for context, one feels that guessing would be too unreliable.

Computers and Language
Topic 5: Machine Translation

Introduction
Examples for Translations

Background: Dictionaries

Transformer approaches

Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems

Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation

What makes MT hard?

Evaluating MT systems

References

60 / 61

Further reading

Some of the examples are adapted from the following books:

- ▶ Doug J. Arnold, Lorna Balkan, Siety Meijer, R. Lee Humphreys and Louisa Sadler (1994). *Machine Translation: an Introductory Guide*. Blackwells-NCC, London. 1994. Available from <http://www.essex.ac.uk/linguistics/clmt/MTbook/>
- ▶ Jurafsky, Daniel, and James H. Martin (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Prentice-Hall. More info at <http://www.cs.colorado.edu/~martin/slp.html>.

Computers and Language
Topic 5: Machine Translation
Introduction
Examples for Translations
Background: Dictionaries
Transformer approaches
Linguistic knowledge based systems
Direct transfer systems
Interlingua-based systems
Machine learning based systems
Alignment
Statistical Modeling
Phrase-Based Translation
What makes MT hard?
Evaluating MT systems
References