Statistical Machine Translation

Introduction and Overview

Jakub Waszczuk
Heinrich-Heine-Universität Düsseldorf

Winter Semester 2018/19

Today

1. Introduction to *Machine Translation*
   a. Approaches
   b. Applications
   c. Examples
   d. Existing Systems
   e. ...
2. Overview of the Semester
   a. Covered Topics
   b. Format of Course
   c. Passing the Course
   d. ...

Machine Translation

dt. *Maschinelle Übersetzung*

Abbreviation: MT (or dt. MÜ)

Basic Problem:
given a sentence in an *input langage* (also called *source language*), translate it automatically (e.g. using a computer program) into a *target language*

Particularity:
There are many „right“ translations!

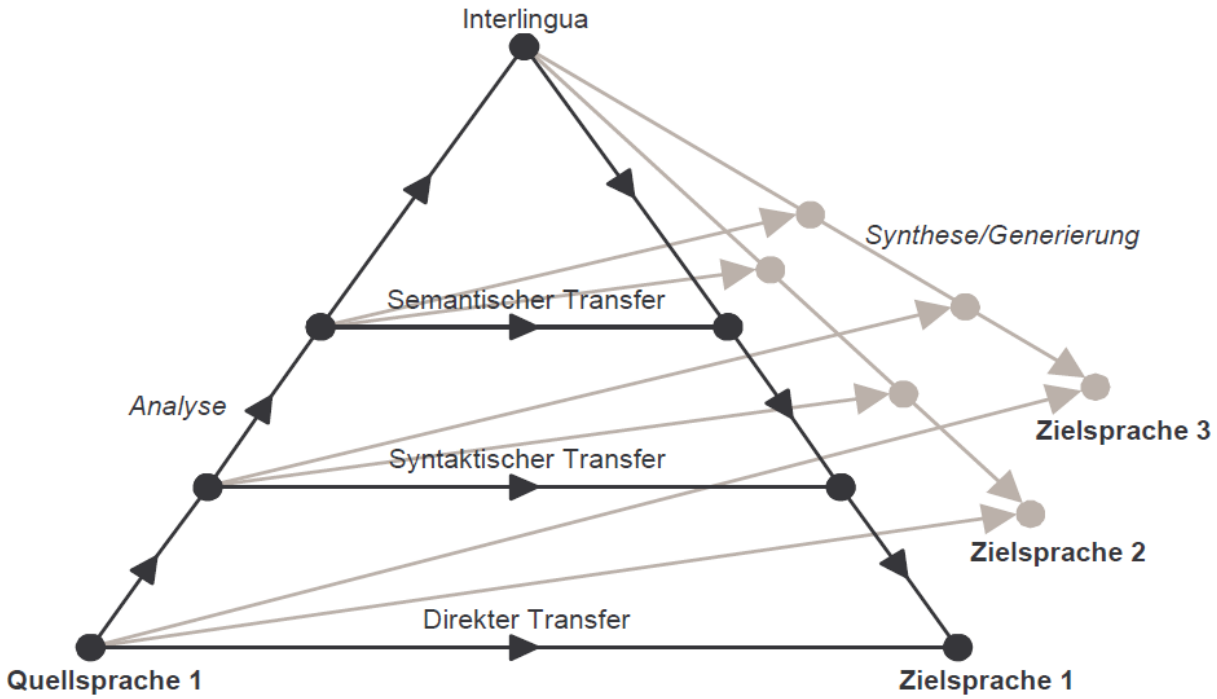
Predominant Paradigms

1. **Rule-based Approaches**
   - Language is a restricted, rule-based system
   - Automatic language processing can be handled with rules
   - Rules are formulated on the basis of human intuition
2. **Statistical Approaches**
   - Language is the total of its instances
   - Generalizations about languages are possible based on large collections of texts (*corpora*), which are considered as representative samples
3. **Hybrid Approaches**

Following U. Reinke (2005)

Potential (rule-based) MT Architectures

Vauquois-Dreieck/Pyramid



From U. Reinke (2005)

⇒ One analysis and one synthesis module per language and one transfer module per language pair

Also: transfer and synthesis together in one module

Rule-based Approaches

Historically, before the statistical approaches

Motivation:

- Good translations require linguistic knowledge to (a) understand the source sentence and (b) generate a well-formed target sentence.
- (Rule-based) analysis/synthesis modules for some levels (e.g., morphology, syntax) already exist.

Disadvantages/Problems:

- Requires a lot of manual work → high development costs
- Quickly becomes complex → high maintenance costs
- Difficult to account for the flexibility of languages → coverage
- A lot of manual effort for each new language pair

Nevertheless: successfully[a] used in commercial products (e.g. Systran)

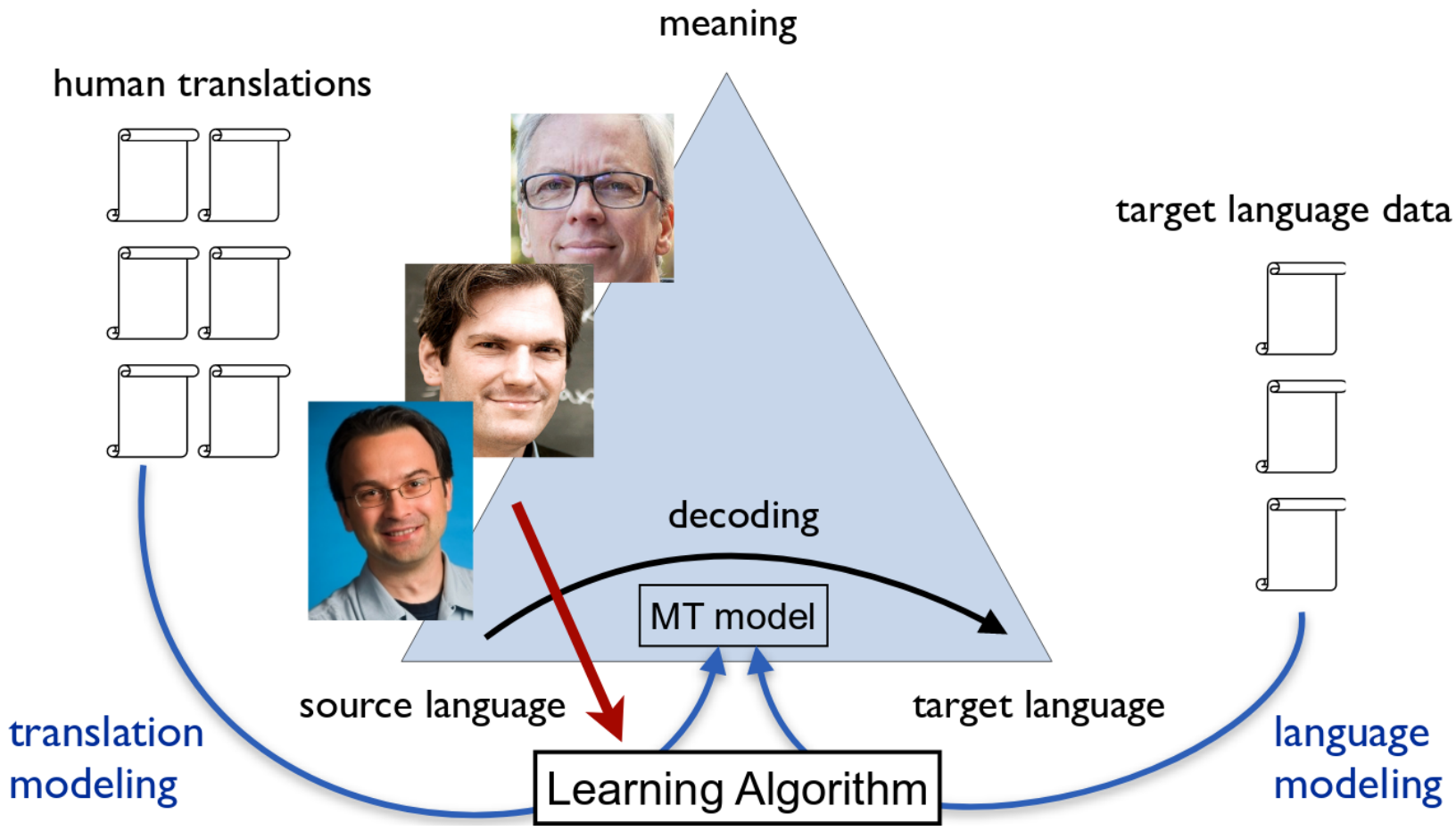
[a] Well, at least in the past

Statistical Approaches (1)

Idea: translation via machine learning

1. Design a language-independent model that describes how to translate from one language to another and assign a probability/cost to each possible translation.
2. A model has a large number of parameters. These are learned from bilingual data (parallel corpora). → „Training“
3. Once the model is trained, input sentences can be translated automatically. A search algorithm is needed to determine the target sentences with the highest probability (or, equivalently, with the lowest cost).

Data-Driven Machine Translation



After Jörg Tiedemann (2nd PARSEME Training School, La Rochelle, 2016)

Statistical Approaches (2)

- Model/architecture only needs to be designed once
- Translation system for a new language pair can be trained quickly; Adaptation to new domains is also possible
- Even if you re-design your architecture, you can quickly train new models (provided that you have your training data)
- Results, of course, dependend on the language pair and the amount of training data, but comparable to (better than?) those of rule-based systems, see e.g. Google Translate
- More and more parallel corpora available
- Subject of active research

How good is statistical machine translation? (1)

French input

Nous savons trés bien que les Traités actuels ne suffisent pas et qu'il sera nécessaire à l'avenir de développer une structure plus efficace et différente pour l'Union, une structure plus constitutionnelle qui indique clairement quelles sont les compétences des États membres et quelles sont les compétences de l'Union.

Statistical machine translation

We know very well that the current treaties are not enough and that in the future it will be necessary to develop a different and more effective structure for the union, a constitutional structure which clearly indicates what are the responsibilities of the member states and what are the competences of the union.

Human translation

We know all too well that the present Treaties are inadequate and that the Union will need a better and different structure in future, a more constitutional structure which clearly distinguishes the powers of the Member States and those of the Union.

From Koehn (2010)

System: fr-en, University of Edinburgh (WMT 2005 shared task)

How good is statistical machine translation? (2)

Chinese input

伦敦每日快报指出,两台记载黛安娜王妃一九九七年巴黎死亡车祸调查资料的手提电脑,被从前大都会警察总长的办公室里偷走.

Statistical machine translation

The London Daily Express pointed out that the death of Princess Diana in 1997 Paris car accident investigation information portable computers, the former city police chief in the offices of stolen.

Human translation

London's Daily Express noted that two laptops with inquiry data on the 1997 Paris car accident that caused the death of Princess Diana were stolen from the office of a former metropolitan police commissioner.

From Koehn (2010)

System: cn-en, University of Edinburgh (NIST 2006 campaign)

How good is statistical machine translation? (3)

Chinese input [a]

伦敦“每日快报”指出，两名笔记本电脑上有一张导致戴安娜王妃死亡的1997年巴黎车祸调查数据被从一名前都市警察局长办公室偷走

Google Translate [b]

London's „Daily Express“ pointed out that a 1997 laptop accident survey data on two laptops that caused Princess Diana's death was stolen from a former urban police chief's office.

Human translation

London's Daily Express noted that two laptops with inquiry data on the 1997 Paris car accident that caused the death of Princess Diana were stolen from the office of a former metropolitan police commissioner.

[a] Obtained via reversed translation from the reference translation

[b] As of 9 October 2018

Is machine translation even useful?!

The better the quality of machine translation, the more useful the systems

But: even **low quality machine translations** can be quite useful

$\rightarrow$ depending on the application

Applications of machine translation

Use of (automatic) translation

1. ***Assimilation***
   Translation of a foreign-language text to understand the content
   → coverage, robustness
2. ***Dissemination***
   Translation of a text to publish it in a foreign language
   → quality
3. ***Communication***
   Translation of emails, chat room discussions, even conversations (→ speech recognition)
   → speed, context-sensitivity

Risks associated with dissemination of „translated“ texts

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

From A. Eisele (2010)

Applications (1)

Fully-automatic, high-quality machine translation

- The translation problem is only partially of a linguistic nature!
  → World and contextual knowledge is required.
- So far only possible for *limited domains*, e.g., weather forecasts, summaries of sport events, flight information systems
- Solution: *controlled language*

Gisting - understand the essence of a foreign-language text

- Translation quality doesn't have to be perfect
- Target: internet users, intelligence services, ...
- Meanwhile, also used for dissemination (e.g. for translating Microsoft Support articles)

Applications (2)

Links with speech technology

Translation of telephone conversations, audio transmissions, etc.

- In speech recognition and SMT similar ideas and models are used → direct composition possible
- Real-time language translation is, hence, possible

Post-editing

Generic term: *human-aided machine translation*

- Goal: high-quality text suitable for publication
- First MT, then human correction
- Can save translation costs if the cost of post-processing is lower than for the complete translation

Applications (3)

Tools for translators

interactive environment for human translators → higher productivity

- *translation memory*
- MT system calculates its own confidence in the translation. When too low → human translation.
- ...

Existing resources and systems

In order to build an SMT system for a particular language pair, you need:

1. a generic SMT toolkit, e.g. Moses
2. a parallel corpus for the language pair
3. a monolingual corpus for the target language

Free Software

- **GIZA++**: word-based IBM models
- **SRILM**, **IRST**: language model software
- **Moses**: phrase-based SMT
- **CDec**, **travatar**: syntax-based SMT
- **OpenNMT**, **Marian**: neural machine translation (NMT)
- **BLEU**, **METEOR**: evaluating the quality of an MT system
- ...

Bilingual Data

Freely available

- **Europarl**: debates of the European Parliament, 21 languages, up to 2 million sentences per language
- **OPUS**: a huge collection of freely available parallel Corpora, $>$ 90 languages, $>$ 40 billion tokens

For a payment

- Many parallel corpora available, e.g. for English-Arabic, English-Chinese
- Most are accessible through the LDC (Lingustic Data Consortium, University of Pennsylvania)

Data Preprocessing (1)

Text vs. sentences

- Research mainly focused on translation of individual sentences
- Often important context is lost.

  The window is open. **It** is blue. → La fenêtre est ouverte. **Elle** est bleue.

  He is trying. → Er versucht es. / Er bemüht sich. / Er ist ein anstrengender Mensch.

Data Preprocessing (2)

Preprocessing

- **Sentence segmentation**: segmentation of text into sentences
- **Tokenization**: segmentation of sentence into words

  Hans-Joachim kauft in New York Fish'n'Chips für $2.50. →

  Hans-Joachim | kauft | in | New York | Fish'n'Chips | für | $ | 2.50 | .

  Problem: definition of a „word“ (what about *kindergarten* or *Neuwortgenerierung*?)

Data Preprocessing (3)

Optional preprocessing steps

- Normalization of upper/lower case
  - lowercase in general
    → less data sparcity, but some information loss
  - *true-casing* – „true“ upper- and lowercase
- Identification of numbers (`zwei`, `2.0`), dates, proper names, etc.

After translation: steps in the inverse order

Programs for these steps are mostly freely available.

Additional Reading

Philipp Koehn, *Statistical Machine Translation*, Cambridge University Press, reprint. with corr., 2011

→ some copies in the ULB

Also may be helpful:

- C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*, MIT Press, 1999
- Some Online-Tutorials, e.g. from Kevin Knight (`https://kevincrawfordknight.github.io/papers/wkbk-rw.pdf`) or Michael Collins (`http://www.cs.columbia.edu/~mcollins/notes-spring2013.html`)

Overview of the Semester

1. Probability theory
2. Language models
3. Word-based translation models
4. Phrase-based translation models
5. Decoding
6. Evaluation
7. Advanced topics: tree-based models, NMT, ...