Lecture 21: Machine Translation

Julia Hockenmaier
juliahmr@illinois.edu
3324 Siebel Center
Machine Translation in 2018

Xi Jinping inspected in Shanghai

Xi Jinping stressed during his visit to Shanghai

Strengthening reform and opening up and starting to build confidence and determination

Accelerate the improvement of urban energy level and core competitiveness

Xinhua News Agency, Shanghai, November 7th, Xi Jinping, general secretary of the CPC Central Committee, president of the State Council and chairman of the Central Military Commission, stressed during his recent visit to Shanghai that he should adhere to the guidance of socialism with Chinese characteristics in the new era, resolutely implement the decision-making and deployment of the Party Central Committee, and strengthen reform and opening up, confidence and
Machine Translation in 2012

"Baby diaper, 384 yuan; cell phone foil, 10 yuan; newspaper, 1 yuan ..." Every day in the past decade, in the city of Beijing, there are 5,000 different income, constitute a different family carefully fill in a unified format Living Book - "Journal of Urban Households Living Journal."

This line of "needle and thread" really records their life trajectory, but also outlines the 'urban residents disposable income', "urban household survey income" and other statistics on people's livelihood.

For many years, grass-roots investigators who deal with these original ecological data can always personally know those people who live in the city make money and spend money ...

How do our daily lives become numbers, and how do numbers affect our lives?

Beijing has 5,000 households with book-keeping households. Their records will be the channels through which the government can understand various aspects of residents' incomes, living standards, prices and other information, and provide an important basis for formulating social development plans and making scientific decisions, including the minimum living standard and the minimum wage and many more. These families are selected by the statistical investigation unit following the principle of random sampling, and the bookkeeping households are rotated in their entirety in three years.

The data they fill in, after systematically arranging, aggregating and analyzing, each month will form a number of data reports such as "per capita disposable income of urban residents" and...
Why is MT difficult?
Some examples

John loves Mary.
*Jean aime Marie.*

John told Mary a story.
*Jean a raconté une histoire à Marie.*

John is a computer scientist.
*Jean est informaticien.*

John swam across the lake.
*Jean a traversé le lac à la nage.*
Correspondences

John loves Mary.
\[\text{Jean aime Marie.}\]

John told Mary a story.
\[\text{Jean [a raconté] une histoire [à Marie].}\]

John is a [computer scientist].
\[\text{Jean est informaticien.}\]

John [swam across] the lake.
\[\text{Jean [a traversé] le lac [à la nage].}\]
Correspondences

One-to-one:
  John = Jean, aime = loves, Mary=Marie

One-to-many/many-to-one:
  Mary = [à Marie]
  [a computer scientist] = informaticien

Many-to-many:
  [swam across___] = [a traversé ___ à la nage]

Reordering required:
  told Mary₁ [a story]₂ = a raconté [une histoire]₂ [à Marie]₁
Lexical divergences

The different senses of homonymous words generally have different translations:

English-German:  
(river) bank - Ufer
(financial) bank - Bank

The different senses of polysemous words may also have different translations:

I know that he bought the book:  Je sais qu’il a acheté le livre.
I know Peter:  Je connais Peter.
I know math:  Je m’y connais en maths.
Lexical divergences

Lexical specificity

German *Kürbis* = English *pumpkin* or *(winter) squash*
English *brother* = Chinese *gege* (older) or *didi* (younger)

Morphological divergences

English: *new book(s), new story/stories*
French: un *nouveau livre* (sg.m), une *nouvelle histoire* (sg.f), 
      des *nouveaux livres* (pl.m), des *nouvelles histoires* (pl.f)

- How much *inflection* does a language have? 
  (cf. Chinese vs. Finnish)
- How many *morphemes* does each word have? 
- How easily can the morphemes be *separated*?
Syntactic divergences

Word order: fixed or free?
If fixed, which one? [SVO (Sbj-Verb-Obj), SOV, VSO,... ]

Head-marking vs. dependent-marking

**Dependent-marking** (English)  the man’s house
**Head-marking** (Hungarian)  the man house-his

Pro-drop languages can omit pronouns:
Italian (with inflection):  l eat = mangio; he eats = mangia
Chinese (without inflection):  l/he eat: chīfàn
## Syntactic divergences: negation

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Negated</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English</strong></td>
<td><em>I drank coffee.</em></td>
<td><em>I didn’t drink (any) coffee.</em></td>
<td><em>do-support, any</em></td>
</tr>
<tr>
<td><strong>French</strong></td>
<td><em>J’ai bu du café</em></td>
<td><em>Je n’ai pas bu de café.</em></td>
<td><em>ne..pas du → de</em></td>
</tr>
<tr>
<td><strong>German</strong></td>
<td><em>Ich habe Kaffee getrunken</em></td>
<td><em>Ich habe <strong>keinen</strong> Kaffee getrunken</em></td>
<td><em>keinen Kaffee = ‘no coffee’</em></td>
</tr>
</tbody>
</table>
Semantic differences

Aspect:
- English has a **progressive aspect**: ‘Peter swims’ vs. ‘Peter *is swimming*’
- German can only express this with an **adverb**: ‘Peter schwimmt’ vs. ‘Peter schwimmt gerade’ (‘swims currently’)

**Motion events** have two properties:
- **manner** of motion (swimming)
- **direction** of motion (across the lake)

Languages express either the manner with a verb and the direction with a ‘satellite’ or vice versa (L. Talmy):
- English (satellite-framed): *He [swam]*\textsubscript{MANNER} *[across]*\textsubscript{DIR} *the lake*
- French (verb-framed): *Il a [traversé]*\textsubscript{DIR} *le lac [à la nage]*\textsubscript{MANNER}
An exercise
Knight’s Centauri and Arctuan

1a. ok-voon ororok sprok.
1b. at-voon bichat dat.

2a. ok-drubel ok-voon anok plok sprok.
2b. at-drubel at-voon pippat rrat dat.

3a. erok sprok izok hihok ghirok.
3b. totat dat arrat vat hilat.

4a. ok-voon anok drok brok jok.
4b. at-voon krat pippat sat lat.

5a. wiwok farok izok stok.
5b. totat jjat quat cat.

6a. lalok sprok izok jok stok.
6b. wat dat krat quat cat.

7a. lalok farok ororok lalok sprok izok enemok.
7b. wat jjat bichat wat dat vat eneat.

8a. lalok brok anok plok nok.
8b. iat lat pippat rrat nnat.

9a. wiwok nok izok kantok ok-yurp.
9b. totat nnatquat oloat at-yurp.

10a. lalok mok nok yorok ghirok clok.
10b. wat nnat gat mat bat hilat.

11a. lalok nok crrrok hihok yorok zanzanok.
11b. wat nnat arrat mat zanzanat.

12a. lalok rarok nok izok hihok mok.
12b. wat nnat forat arrat vat gat.
The original corpus

1a. Garcia and associates.
1b. Garcia y asociados.

2a. Carlos Garcia has three associates.
2b. Carlos Garcia tiene tres asociados.

3a. his associates are not strong.
3b. sus asociados no son fuertes.

4a. Garcia has a company also.
4b. Garcia tambien tiene una empresa.

5a. its clients are angry.
5b. sus clientes están enfadados.

6a. the associates are also angry.
6b. los asociados tambien están enfadados.

7a. the clients and the associates are enemies.
7b. los clientes y los asociados son enemigos.

8a. the company has three groups.
8b. la empresa tiene tres grupos.

9a. its groups are in Europe.
9b. sus grupos están en Europa.

10a. the modern groups sell strong pharmaceuticals.
10b. los grupos modernos venden medicinas fuertes.

11a. the groups do not sell zanzanine.
11b. los grupos no venden zanzanina.

12a. the small groups are not modern.
12b. los grupos pequeños no son modernos.
1a. Garcia and associates.
1b. Garcia y asociados.
2a. Carlos Garcia has three associates.
2b. Carlos Garcia tiene tres asociados.
3a. his associates are not strong.
3b. sus asociados no son fuertes.
4a. Garcia has a company also.
4b. Garcia tambien tiene una empresa.
5a. its clients are angry.
5b. sus clientes estan enfadados.
6a. the associates are also angry.
6b. los asociados tambien estan enfadados.
7a. the clients and the associates are enemies.
7b. los clientes y los asociados son enemigos.
8a. the company has three groups.
8b. la empresa tiene tres grupos.
9a. its groups are in Europe.
9b. sus grupos estan en Europa.
10a. the modern groups sell strong pharmaceuticals
10b. los grupos modernos venden medicinas fuertes
11a. the groups do not sell zanzanine.
11b. los grupos no venden zanzanina.
12a. the small groups are not modern.
12b. los grupos pequenos no son modernos.
Machine translation approaches
The Rosetta Stone

Three different translations of the same text:
- Hieroglyphic Egyptian (used by priests)
- Demotic Egyptian (used for daily purposes)
- Classical Greek (used by the administration)

Instrumental in our understanding of ancient Egyptian

This is an instance of parallel text:
The Greek inscription allowed scholars to decipher the hieroglyphs
MT History

WW II: Code-breaking efforts at Bletchley Park, England (Alan Turing)
1948: Shannon/Weaver: Information theory
1949: Weaver’s memorandum defines the task
1954: IBM/Georgetown demo: 60 sentences Russian-English
1960: Bar-Hillel: MT to difficult
1966: ALPAC report: human translation is far cheaper and better: kills MT for a long time
1980s/90s: Transfer and interlingua-based approaches
1990: IBM’s CANDIDE system (first modern statistical MT system)
2000s: Huge interest and progress in wide-coverage statistical MT: phrase-based MT, syntax-based MT, open-source tools
Now: Neural machine translation
The Vauquois triangle

Analysis

Interlingua

Semantic transfer

Generation

Semantics

Syntactic transfer

Transfer

Syntax

Direct transfer

Words

Source

Target

CS447 Natural Language Processing
Direct translation

Maria non dió una bofetada a la bruja verde.

1. **Morphological analysis of source string**
   Maria non\textsubscript{Neg} dar\textsubscript{3sgF-Past} una bofetada a la bruja verde
   (usually, a complete morphological analysis)

2. **Lexical transfer** *(using a translation dictionary):*
   Mary not slap\textsubscript{3sgF-Past} to the witch green.

3. **Local reordering:**
   Mary not slap\textsubscript{3sgF-Past} the green witch.

4. **Morphology:**
   Mary did not slap the green witch.
Limits of direct translation: Phrasal reordering

Adverb placement in German:
The green witch is at home this week.
Diese Woche ist die grüne Hexe zuhause.

Japanese SOV order:
He adores listening to music
Kare ha ongaku wo kiku no ga daisuki desu

PPs in Chinese:
Jackie Cheng went to Hong Kong
Cheng Long dao Xianggang qu
Syntactic transfer

Requires a syntactic parse of the source language, followed by reordering of the tree

Local reordering:

Nonlocal reordering:
Semantic transfer

Done at the level of **predicate-argument structure**
(some people call this syntactic transfer too…):

\[
\begin{align*}
[ & \text{SUBJ} \ [ \text{PRED} \ Hans]_2] \\
[ & \text{PRED} \ kochen(\uparrow \text{SUBJ})] \\
[ & \text{ADJN} \ \{[ \text{PRED} \ \text{gerne}]_3] \\
\end{align*}
\]

\textit{Hans kocht gerne}

or at the level of **semantic representations** (e.g. DRSs):

\[
\begin{align*}
l_T : \begin{array}{c}
x_2 \\
\text{Hans}(x_2) \\
\end{array} & \quad \rightarrow & \quad l_T : \begin{array}{c}
x_2 \\
\text{Hans}(x_2) \\
\end{array} \\
l_3 : \begin{array}{c}
x_3 \\
\text{gerne}(x_3) \\
\end{array} & \quad \rightarrow & \quad l_3 : \begin{array}{c}
x_3 \\
\text{like}(x_2, x_3) \\
\end{array} \\
l_1 : \begin{array}{c}
x_1 \\
\text{kochen}(x_1) \\
\end{array} & \quad \rightarrow & \quad l_1 : \begin{array}{c}
x_1 \\
\text{cook}(x_2) \\
\end{array}
\end{align*}
\]

\textit{Hans kocht gerne} \quad \textit{Hans likes cooking}

\textit{Dorna et al. 1998}
Interlingua approaches

Based on the assumption that there is one **common meaning representation** (e.g. predicate logic) that abstracts away from any difference in surface realization.

- Semantic transfer: each language produces its own meaning representation
- Was thought useful for multilingual translation

![Diagram showing the process of translation through an interlingua]

*Leavitt et al. 1994*
Statistical Machine Translation
Statistical Machine Translation

We want the best (most likely) [English] translation for the [Chinese] input:

\[ \text{argmax}_{\text{English}} P(\text{English} \mid \text{Chinese}) \]

We can either model this probability directly, or we can apply Bayes Rule. Using Bayes Rule leads to the “noisy channel” model.

As with sequence labeling, Bayes Rule simplifies the modeling task, so this was the first approach for statistical MT.
The noisy channel model

Translating from Chinese to English:

\[
\text{argmax}_{\text{Eng}} P(\text{Eng}|\text{Chin}) = \text{argmax}_{\text{Eng}} P(\text{Chin}|\text{Eng}) \times P(\text{Eng})
\]

Translation Model

Language Model

\( P(O|I) \)

\( P(\text{Eng}) \)

Decoder

\( \hat{I} = \text{argmax}_{I} P(O|I)P(I) \)
The noisy channel model

This is really just an application of Bayes’ rule:

\[
\hat{E} = \arg \max_E P(E|F)
\]

\[
= \arg \max_E \frac{P(F|E) \times P(E)}{P(F)}
\]

\[
= \arg \max_E \left( P(F|E) \right) \times \left( \frac{P(E)}{P(F)} \right)
\]

\[
\hat{E} \propto P(F|E) \times P(E)
\]

Translation Model Language Model

The translation model \(P(F \mid E)\) is intended to capture the faithfulness of the translation. It needs to be trained on a parallel corpus.

The language model \(P(E)\) is intended to capture the fluency of the translation. It can be trained on a (very large) monolingual corpus.
Statistical MT

Parallel corpora

Monolingual corpora

Translation Model

Language Model

Input

Decoding algorithm

Translation

Chair: Members, good morning.

Chair: 各位議員，早晨。

MOTION: PRESIDENT (in Cantonese): Good morning, Honourable Members. We will now start the meeting. First of all, the motion on the "Appointment of the Chief Justice of the Court of Final Appeal of the Hong Kong Special Administrative Region". Secretary for Justice.

President: Good morning, Honourable Members.
n-gram language models for MT

With training on data from the web and clever parallel processing (MapReduce/Bloom filters), $n$ can be quite large:
- Google (2007) uses 5-grams to 7-grams,
- This results in huge models, but the effect on translation quality levels off quickly:

Size of models

![Graph showing the size of models vs. number of n-grams.](image)

Effect on translation quality

![Graph showing BLEU scores vs. training data size.](image)

Figure 3: Number of $n$-grams (sum of unigrams to 5-grams) for varying amounts of training data.

Figure 5: BLEU scores for varying amounts of data using Kneser-Ney (KN) and Stupid Backoff (SB).
Translation probability $P(fp_i \mid ep_i)$

Phrase translation probabilities can be obtained from a **phrase table:**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP</td>
<td>FP</td>
<td></td>
</tr>
<tr>
<td>green witch</td>
<td>grüne Hexe</td>
<td>…</td>
</tr>
<tr>
<td>at home</td>
<td>zuhause</td>
<td>10534</td>
</tr>
<tr>
<td>at home</td>
<td>daheim</td>
<td>9890</td>
</tr>
<tr>
<td>is</td>
<td>ist</td>
<td>598012</td>
</tr>
<tr>
<td>this week</td>
<td>diese Woche</td>
<td>…</td>
</tr>
</tbody>
</table>

This requires **phrase alignment** on a parallel corpus.
Creating parallel corpora

A **parallel corpus** consists of the same text in two (or more) languages.

Examples: Parliamentary debates: Canadian Hansards; Hong Kong Hansards, Europarl; Movie subtitles (OpenSubtitles)

In order to train translation models, we need to **align the sentences** (Church & Gale ’93)

<table>
<thead>
<tr>
<th>English</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>According to our survey, 1988 sales of mineral water and soft drinks were much higher than in 1987, reflecting the growing popularity of these products. Cola drink manufacturers in particular achieved above-average growth rates. The higher turnover was largely due to an increase in the sales volume. Employment and investment levels also climbed. Following a two-year transitional period, the new Foodstuffs Ordinance for Mineral Water came into effect on April 1, 1988. Specifically, it contains more stringent requirements regarding quality consistency and purity guarantees.</td>
<td>Quant aux eaux minérales et aux limonades, elles rencontrent toujours plus d’adeptes. En effet, notre sondage fait ressortir des ventes nettement supérieures à celles de 1987, pour les boissons à base de cola notamment. La progression des chiffres d’affaires résulte en grande partie de l’accroissement du volume des ventes. L’emploi et les investissements ont également augmenté. La nouvelle ordonnance fédérale sur les denrées alimentaires concernant entre autres les eaux minérales, entrée en vigueur le 1er avril 1988 après une période transitoire de deux ans, exige surtout une plus grande constance dans la qualité et une garantie de la pureté.</td>
</tr>
</tbody>
</table>
Today’s key concepts

Why is machine translation hard?
   Linguistic divergences: morphology, syntax, semantics

Different approaches to machine translation:
   Vauquois triangle
   Statistical MT (more on this next time)