Introduction to Statistical Machine Translation

Fabienne Cap

UPPSALA UNIVERSITY SWEDEN
<table>
<thead>
<tr>
<th>Type</th>
<th>Date</th>
<th>Time</th>
<th>Place</th>
<th>Topic</th>
<th>Reading / Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>2016-03-30</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Introduction (SS)</td>
<td>Koehn 1; JM 25.1-2; Hutchins; CFMF</td>
</tr>
<tr>
<td>F</td>
<td>2016-03-30</td>
<td>14-16</td>
<td>6-K1031</td>
<td>MT evaluation (SS)</td>
<td>Koehn 8; JM 25.9</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-04</td>
<td>10-12</td>
<td>2-0076</td>
<td>MT in practice (Convertus) - guest lecture</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>2016-04-06</td>
<td>10-12</td>
<td>Chomsky</td>
<td>MT in practice (AS)</td>
<td>lab report 1</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-11</td>
<td>10-12</td>
<td>2-0076</td>
<td>Introduction to SMT (FC)</td>
<td>Koehn Ch 4, Ch 7, KK97</td>
</tr>
<tr>
<td>L</td>
<td>2016-04-13</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Word-based SMT (SS)</td>
<td>lab report 2</td>
</tr>
<tr>
<td>L</td>
<td>2016-04-18</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Word-based SMT (SS)</td>
<td>lab report 2</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-18</td>
<td>14-16</td>
<td>2-1077</td>
<td>Machine translation at Semantix, a translation provider - guest lecture</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>2016-04-20</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Parallel Corpora, Alignment (AS)</td>
<td>Koehn 2-4, JT 3-4, KK97, KK99</td>
</tr>
<tr>
<td>L</td>
<td>2016-04-25</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Parallel corpora &amp; alignment (AS)</td>
<td>lab report 3</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-27</td>
<td>10-12</td>
<td>2-0076</td>
<td>Phrase-based SMT (FC)</td>
<td>Koehn Ch 5</td>
</tr>
<tr>
<td>L</td>
<td>2016-05-02</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Phrase-based SMT (AS)</td>
<td>lab report 4</td>
</tr>
<tr>
<td>F</td>
<td>2016-05-04</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Decoding (CH)</td>
<td>Koehn Ch 6</td>
</tr>
<tr>
<td>L</td>
<td>2016-05-09</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Phrase-based SMT (AS)</td>
<td>lab report 4</td>
</tr>
<tr>
<td>F</td>
<td>2016-05-11</td>
<td>10-12</td>
<td>2-0076</td>
<td>Tree-based SMT &amp; MT for morphologically rich languages (SS, FC)</td>
<td>Koehn 10.2, 11</td>
</tr>
<tr>
<td>F</td>
<td>2016-05-16</td>
<td>10-12</td>
<td>2-0076</td>
<td>Document-wide decoding &amp; Neural MT (CH)</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>2016-05-18</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Document-wide decoding lab (AS)</td>
<td>oral lab report 5</td>
</tr>
<tr>
<td>S</td>
<td>2016-05-23</td>
<td>10-12</td>
<td>2-0076</td>
<td>Seminar - master student presentations</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>2016-05-25</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Seminar - master student presentations</td>
<td></td>
</tr>
</tbody>
</table>
## Where are we?

<table>
<thead>
<tr>
<th>Type</th>
<th>Date</th>
<th>Time</th>
<th>Place</th>
<th>Topic</th>
<th>Reading / Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>2016-03-30</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Introduction (SS)</td>
<td>Koehn 1; JM 25.1-2; Hutchins; CFMF</td>
</tr>
<tr>
<td>F</td>
<td>2016-03-30</td>
<td>14-16</td>
<td>6-K1031</td>
<td>MT evaluation (SS)</td>
<td>Koehn 8; JM 25.9</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-04</td>
<td>10-12</td>
<td>2-0076</td>
<td>MT in practice (Convertus) - guest lecture</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>2016-04-06</td>
<td>10-12</td>
<td></td>
<td>MT in practice (AS)</td>
<td>lab report 1</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-11</td>
<td>10-12</td>
<td>2-0076</td>
<td>Introduction to SMT (FC)</td>
<td>Koehn Ch 4, Ch 7, KK97</td>
</tr>
<tr>
<td>L</td>
<td>2016-04-13</td>
<td>10-12</td>
<td></td>
<td>Word-based SMT (SS)</td>
<td>lab report 2</td>
</tr>
<tr>
<td>L</td>
<td>2016-04-18</td>
<td>10-12</td>
<td></td>
<td>Word-based SMT (SS)</td>
<td>lab report 2</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-18</td>
<td>14-16</td>
<td>2-1077</td>
<td>Machine translation at Semantix, a translation provider - guest lecture</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>2016-04-20</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Parallel Corpora, Alignment (AS)</td>
<td>Koehn 2-4, JT 3-4, KK97, KK99</td>
</tr>
<tr>
<td>L</td>
<td>2016-04-25</td>
<td>10-12</td>
<td></td>
<td>Parallel corpora &amp; alignment (AS)</td>
<td>lab report 3</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-27</td>
<td>10-12</td>
<td>2-0076</td>
<td>Phrase-based SMT (FC)</td>
<td>Koehn Ch 5</td>
</tr>
<tr>
<td>L</td>
<td>2016-05-02</td>
<td>10-12</td>
<td></td>
<td>Phrase-based SMT (AS)</td>
<td>lab report 4</td>
</tr>
<tr>
<td>F</td>
<td>2016-05-04</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Decoding (CH)</td>
<td>Koehn Ch 6</td>
</tr>
<tr>
<td>L</td>
<td>2016-05-09</td>
<td>10-12</td>
<td></td>
<td>Phrase-based SMT (AS)</td>
<td>lab report 4</td>
</tr>
<tr>
<td>F</td>
<td>2016-05-11</td>
<td>10-12</td>
<td>2-0076</td>
<td>Tree-based SMT &amp; MT for morphologically rich languages (SS, FC)</td>
<td>Koehn 10.2, 11</td>
</tr>
<tr>
<td>F</td>
<td>2016-05-16</td>
<td>10-12</td>
<td>2-0076</td>
<td>Document-wide decoding &amp; Neural MT (CH)</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>2016-05-18</td>
<td>10-12</td>
<td></td>
<td>Document-wide decoding lab (AS)</td>
<td>oral lab report 5</td>
</tr>
<tr>
<td>S</td>
<td>2016-05-23</td>
<td>10-12</td>
<td></td>
<td>Seminar - master student presentations</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>2016-05-25</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Seminar - master student presentations</td>
<td></td>
</tr>
</tbody>
</table>
Goals for Today

High-level introduction to Statistical Machine Translation

Word-based Translation Models

Noisy Channel Model

Language Models
The Dark Side (e.g. Siths) [STAR WARS POSTER] The Light Side (e.g. Jedis)
The language of the Sith

Tegu mus kelias antai kash.

Translation? Anyone?

**Problem:**
→ Human translators may not be available
→ Human translators are expensive

**Possible solution:**
We found a collection of translated texts!
15min - 20min

May the force be with you!
Lessons Learned

What do we learn from this exercise?

- 1-to-1 translations easier to identify than 1-to-n, n-to-1 or n-to-m
- unseen words cannot be translated
- ambiguity: some words have more than one correct translation → the context determines which one
- sometimes words need to be re-orderend
Goals for Today

High-level introduction to Statistical Machine Translation

Word-based Translation Models

Noisy Channel Model

Language Models
Today, word-based translation models are outdated, but they introduce some important concepts which are still relevant for state-of-the-art SMT models:

- generative modelling
- noisy-channel model
- IBM Models 1-5
- expectation maximisation algorithm

→ more details in Aaron’s lecture on word alignment!
Generative Model: source language words are generated by target language words

```
1 2 3 4
tave dury kash itinwit
```

```
| | /
the door is very small
```

Introduce an **alignment function** \( a: i \rightarrow j \)

\( a: \{1 \rightarrow 2, 2 \rightarrow 2, 3 \rightarrow 3, 4 \rightarrow 4, 5 \rightarrow 4\} \)
Generative Modelling

**Generative Model:** source language words are generated by target language words

```
1 2 3 4
```

tave dury kash itinwit

```
1 2 3 4
```

the door is very small

Introduce an **alignment function** \( a: i \rightarrow j \)

\[ a: \{1 \rightarrow 2, 2 \rightarrow 2, 3 \rightarrow 3, 4 \rightarrow 4, 5 \rightarrow 4\} \]

**Translation:** Decode what kind of English word sequence has generated the Sith word sequence
Special Cases in Word Alignment

Dropping words:

```
1 2 3 4
tave dury kash itinwit
   
door is very small
```
a:{1→2, 2→3, 3→4, 4→4}
Special Cases in Word Alignment

Inserting words: Introduce a special NULL word

NULL tave dury kash nwit

the door is just small

\{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3, 4 \rightarrow 0, 5 \rightarrow 4\}
Lexical Translations

- tave → the
- dury → door
- kash → is, in
- nwit → smal

In case of multiple translation options:
- use the most common one in that context
Count translation statistics:

How often is **dury** translated into...

<table>
<thead>
<tr>
<th>Translation of <strong>dury</strong></th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>door</td>
<td>8,000</td>
</tr>
<tr>
<td>portal</td>
<td>1,600</td>
</tr>
<tr>
<td>entrance</td>
<td>200</td>
</tr>
<tr>
<td>doorway</td>
<td>150</td>
</tr>
<tr>
<td>gate</td>
<td>50</td>
</tr>
</tbody>
</table>
Estimate translation probabilities:

- Maximum Likelihood Estimation (MLE)

\[
t(\text{english}|\text{sith}) = \frac{\text{count}(\text{english},\text{sith})}{\text{count}(\text{sith})}
\]

- for \( \text{sith} = \text{dury} \):

\[
t(\text{english}|\text{sith}) = \begin{cases} 
0.8 & \text{if english = door} \\
0.16 & \text{if english = portal} \\
0.02 & \text{if english = entrance} \\
0.015 & \text{if english = doorway} \\
0.005 & \text{if english = gate}
\end{cases}
\]
Goals for Today

High-level introduction to Statistical Machine Translation

Word-based Translation Models

Noisy Channel Model

Language Models
What is a noisy channel?

https://www.youtube.com/watch?v=OMUsVcYhERY
Noisy Channel Model

- origin in acoustics and information theory
- idea: foreign language sentence is a message distorted through a noisy channel
- decode distorted message and restore original message
- use two models:
  - source model $p(\text{Source})$ (= language model)
  - channel model $p(\text{Received}|\text{Source})$ (= translation model)

Caution Confusing Terminology!!!
# Noisy Channel Model vs. SMT

<table>
<thead>
<tr>
<th>Noisy Channel Model</th>
<th>SMT</th>
<th>our example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source signal</td>
<td>(desired) SMT output</td>
<td>English text</td>
</tr>
<tr>
<td>(noisy) Channel</td>
<td>Translation model</td>
<td></td>
</tr>
<tr>
<td>Receiver (distorted message)</td>
<td>SMT input source language text</td>
<td>Sith text</td>
</tr>
</tbody>
</table>
Use Bayes’ rule to decompose $P(\text{english}|\text{sith})$ into

- Translation Model $P(\text{sith}|\text{english})$
- Language Model $P(\text{english})$

$$\arg\max_{e} P(e|s) = \arg\max_{e} \frac{P(s|e) * P(e)}{P(s)}$$

$$= \arg\max_{e} P(s|e) * P(e)$$
Modelling Statistical Machine Translation

Sith – English Parallel Corpus

Statistical Analysis

Sith Input → Translation Model → Broken English

English Corpus

Statistical Analysis

Language Model → Fluent English
Introduction to Statistical Machine Translation

Let's climb in there
Let's climb in there
Let's climb there in
Tegu mus kelias antai kash

Sith – English
Parallel Corpus

Statistical Analysis

Sith Input

Translation Model

Broken English

Statistical Analysis

English Corpus

Language Model

Fluent English

Fabienne Cap

Parallel Corpus

Sith – English
Broken
Corpus

Statistical Analysis

Input

Sith
English
Broken
Corpus

Fabric Language
Model English

Fluent

Let’s in there climb
Let’s climb in there
Let’s climb there in
There in let’s climb

Let’s climb in there
Modelling Statistical Machine Translation

Sith – English Parallel Corpus

Statistical Analysis

Sith Input → Translation Model → Broken English

P(sith | english)

Decoding Algorithm

P(english)

Let’s in there climb
Let’s climb in there
Let’s climb there in
There in let’s climb

Tegu mus kelias antai kash

English Corpus

Statistical Analysis

Broken English → Language Model → Fluent English

P(english)

Decoding Algorithm

Fabienne Cap
Introduction to Statistical Machine Translation
Modelling Statistical Machine Translation

P(sith | english) * P(english)

argmax P(sith | english) * P(english)

Let's climb in there

Tegu mus kelias antai kash

Statistical Analysis

Parallell Corpus

Sith – English

Broken Language

English

Input

Translation Model

Fluent English

Statistical Analysis

Parallel Corpus

English

Broken Corpus

Language Model

Statistical Analysis

Input

Sith

English

Fluent

English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

Broken Corpus

Fluent English

Statistical Analysis

Input

Sith

English

B
The role of the Translation and the Language Model

**Translation Model:** prefers adequate translations

- \( P(\text{Tegu mus kelias antai kash} \mid \text{Let’s climb in there}) > \)
- \( P(\text{Tegu mus kellias antai kash} \mid \text{Let’s climb in here}) > \)
- \( P(\text{Tegu mus kelias antai kash} \mid \text{Let’s clamber in there}) \)

**Language Model:** prefers grammatical/fluent sequences

- \( P(\text{Let’s climb in there}) > P(\text{Let’s there climb in}) \)
Goals for Today

High-level introduction to Statistical Machine Translation

Word-based Translation Models

Noisy Channel Model

Language Models
Prefer one string over another (ensure fluency)
- "small step": 5,880,000 hits on Google
- "little step": 1,780,000 hits on Google

**Language Model:**

estimate how likely a string is in a given language:

\[
\begin{align*}
P_{LM}(\text{the door is small}) & > P_{LM}(\text{small the is door}) \\
P_{LM}(\text{let’s climb in there}) & > P_{LM}(\text{let’s climb is there})
\end{align*}
\]
N-Gram Language Models

Markov chain

\[ p(w_1, w_2, \ldots, w_n) = p(w_1)p(w_2|w_1)\ldots p(w_n|w_1, w_2, \ldots, w_{n-1}) \]

Markov assumption

\[ p(w_n|w_1, w_2, \ldots, w_{n-1}) = p(w_n|w_{n-m}, \ldots, w_{n-2}, w_{n-1}) \]

Maximum likelihood estimation (e.g. 3-gram)

\[ p(w_3|w_1, w_2) = \frac{\text{count}(w_1, w_2, w_3)}{\sum_w \text{count}(w_1, w_2, w)} \]
Add **special markers** at the start and the end of the sentence! → certain tokens often appear at the start or at the end

**Smoothing**
- big problem: unseen n-grams → \( p(e) = 0 \)
- smoothing: reserve probability mass for unseen events

**Interpolation and backoff**
→ combine higher-order and lower-order models
N-Gram Language Models - Example

\[ p(\text{he, has, there, not}) = 0.037 \times 0.00025 \times 1.2 \times e^{-5} \times 0.00061 \times 0.15 \times 0 = 0 \]

\[ p(\text{he, has, there, not, been}) = 0.037 \times 0.029 \times 0.385 \times 0.020 \times 0.297 \times 0.210 = 5.15 \times e^{-7} \]
Goals for Today

High-level introduction to Statistical Machine Translation

Word-based Translation Models

Noisy Channel Model

Language Models

IBM Models for Word Alignment
A short overview on IBM Models for Word Alignment
IBM Model 1 only uses lexical translation

\[
p(e, a | f) = \frac{\epsilon}{4^3} \times t(\text{the} | \text{das}) \times t(\text{house} | \text{Haus}) \times t(\text{is} | \text{ist}) \times t(\text{small} | \text{klein})
\]

\[
= \frac{\epsilon}{4^3} \times 0.7 \times 0.8 \times 0.8 \times 0.4
\]

\[
= 0.0028 \epsilon
\]
Adding a model of alignment

natürlich ist das haus klein

of course is the house small

of course the house is small

lexical translation step

alignment step
IBM Model 3

Adding a model of fertility

1. ich
2. gehe
3. ja
4. nicht
5. zum
6. haus

fertility step

1. ich
2. gehe
3. NULL
4. nicht
5. zum
6. zum
7. haus

NULL insertion step

1. ich
2. NULL
guge
3. nicht
4. zum
5. zum
6. haus

lexical translation step

1. I
do
goge
not
to
the
house

1. I
do
not
go
to
the
house
distortion step
IBM Model 4

Motivation

- Absolute position for distortion feels wrong
- Words do not move independently
- Some words tend to move and some not

→ Introduce a relative distortion model

→ Introduce a dependence on word classes
IBM Models 1-4 are deficient

- some impossible translations have positive probabilities
- multiple output words may be placed in the same place
- probability mass is wasted!

IBM Model 5

- fix deficiency by keeping track of vacancies
- details: see text book
Models with increasing complexity

Higher models include more information

<table>
<thead>
<tr>
<th>IBM Model 1</th>
<th>lexical translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM Model 2</td>
<td>adds absolute alignment model</td>
</tr>
<tr>
<td>IBM Model 3</td>
<td>adds fertility model</td>
</tr>
<tr>
<td>IBM Model 4</td>
<td>relative alignment model</td>
</tr>
<tr>
<td>IBM Model 5</td>
<td>fixes deficiency</td>
</tr>
</tbody>
</table>
Take Home Messages from Today??
<table>
<thead>
<tr>
<th>Type</th>
<th>Date</th>
<th>Time</th>
<th>Place</th>
<th>Topic</th>
<th>Reading / Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>2016-03-30</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Introduction (SS)</td>
<td>Koehn 1; JM 25.1-2; Hutchins; CFMF</td>
</tr>
<tr>
<td>F</td>
<td>2016-03-30</td>
<td>14-16</td>
<td>6-K1031</td>
<td>MT evaluation (SS)</td>
<td>Koehn 8; JM 25.9</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-04</td>
<td>10-12</td>
<td>2-0076</td>
<td>MT in practice (Convertus) - guest lecture</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>2016-04-06</td>
<td>10-12</td>
<td>Chomsky</td>
<td>MT in practice (AS)</td>
<td>lab report 1</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-11</td>
<td>10-12</td>
<td>2-0076</td>
<td>Introduction to SMT (FC)</td>
<td>Koehn Ch 4, Ch 7, KK97</td>
</tr>
<tr>
<td>L</td>
<td>2016-04-13</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Word-based SMT (SS)</td>
<td>lab report 2</td>
</tr>
<tr>
<td>L</td>
<td>2016-04-18</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Word-based SMT (SS)</td>
<td>lab report 2</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-18</td>
<td>14-16</td>
<td>2-1077</td>
<td>Machine translation at Semantix, a translation provider - guest lecture</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>2016-04-20</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Parallel Corpora, Alignment (AS)</td>
<td>Koehn 2-4, JT 3-4, KK97, KK99</td>
</tr>
<tr>
<td>L</td>
<td>2016-04-25</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Parallel corpora &amp; alignment (AS)</td>
<td>lab report 3</td>
</tr>
<tr>
<td>F</td>
<td>2016-04-27</td>
<td>10-12</td>
<td>2-0076</td>
<td>Phrase-based SMT (FC)</td>
<td>Koehn Ch 5</td>
</tr>
<tr>
<td>L</td>
<td>2016-05-02</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Phrase-based SMT (AS)</td>
<td>lab report 4</td>
</tr>
<tr>
<td>F</td>
<td>2016-05-04</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Decoding (CH)</td>
<td>Koehn Ch 6</td>
</tr>
<tr>
<td>L</td>
<td>2016-05-09</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Phrase-based SMT (AS)</td>
<td>lab report 4</td>
</tr>
<tr>
<td>F</td>
<td>2016-05-11</td>
<td>10-12</td>
<td>2-0076</td>
<td>Tree-based SMT &amp; MT for morphologically rich languages (SS, FC)</td>
<td>Koehn 10.2, 11</td>
</tr>
<tr>
<td>F</td>
<td>2016-05-16</td>
<td>10-12</td>
<td>2-0076</td>
<td>Document-wide decoding &amp; Neural MT (CH)</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>2016-05-18</td>
<td>10-12</td>
<td>Chomsky</td>
<td>Document-wide decoding lab (AS)</td>
<td>oral lab report 5</td>
</tr>
<tr>
<td>S</td>
<td>2016-05-23</td>
<td>10-12</td>
<td>2-0076</td>
<td>Seminar - master student presentations</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>2016-05-25</td>
<td>10-12</td>
<td>6-K1031</td>
<td>Seminar - master student presentations</td>
<td></td>
</tr>
</tbody>
</table>
https://www.youtube.com/watch?v=GMi4MtyDg40