Syntax-based Translation

Part 1: Re-ordering for Phrase-based translation

March 25, 2014

Thanks to Chris Callison-Burch and Michael Collins for many of today’s slides.

mt-class.org/penn/

www.cs.columbia.edu/~cs4705/
Administrative

- Homework 3: due tomorrow, March 26 @ 6 PM, writeup Thursday in class (bring six copies)

- **April 1:** Project proposal (meet with one of us beforehand)

- Feedback
Course feedback

10 minutes

• What questions do you have? What do you want to know?

• How is the class going?
• Understand why syntax is important for reordering models
• Review non-syntactic reordering models for phrase-based machine translation
• Review the “Clause Restructuring” approach of Collins, Koehn, and Kucerova
• Understand why it is a good fit for phrase-based machine translation
• Discuss its limitations
• Foreign input is segmented in phrases
• Each phrase is translated into English
• Phrases are reordered
Some Reordering Already Captured

- Local reordering can be captured within phrases

Some Reordering Already Captured

natuerlich  hat  john  spass am  spiel

of course  john  has  fun with the  game

of course  john has  fun with the  game

of course  john has  fun with the  game

of course  john has  fun with the  game

natuerlich  hat john  spass am  spiel

of course  john has  fun with the  game
Phrase translation table

• Main knowledge source: table with phrase translations and their probabilities

• Example: phrase translations for natuerlich

| Source     | Translation      | Probability $\phi(e|f)$ |
|------------|------------------|------------------------|
| natuerlich | of course        | 0.5                    |
| natuerlich | naturally        | 0.3                    |
| natuerlich | of course ,      | 0.15                   |
| natuerlich | , of course ,    | 0.05                   |
Probabilistic Model

• Linear model:

\[(e^*, a^*) = \arg\max_{e,a} \sum_i \lambda_i h(f, a, e)\]

• Reordering score can be incorporated in the TM

\[h_{\text{reorder}} = \log d(a_i - b_i - 1)\]

–\(d\) is an exponential function of phrase move distance
Distance-based Reordering

Scoring function: $d(x) = \alpha|x|$ – exponential with distance
Values of $\alpha$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>
Distance-based reordering

• Small values of $\alpha$, severely discourage reordering
  – Limit reordering to monotonic or a narrow window
  – OK for languages with very similar word orders
  – Bad for languages with different word orders

• The distance-based penalty applies uniformly to all words and all word types
  – Doesn’t know that adjectives and nouns should swap when translating from French to English

• Puts most responsibility on the language model
How else could we model reordering?

- Why not assign a distinct reordering probability to each word/phrase in the phrase table?
  \[ p(\text{reorder} \mid f, e) \]
- This is known as **lexicalized reordering**
- How can we estimate that probability?
Lexicalized Reordering model

How much you should charge for your Facebook profile.
**Lexicalized Reordering model**

<table>
<thead>
<tr>
<th>Wieviel</th>
<th>sollte</th>
<th>man</th>
<th>aufgrund</th>
<th>seines</th>
<th>Profils</th>
<th>in</th>
<th>Facebook</th>
<th>verdienen</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much should you charge for your Facebook profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**m**: monotone (keep order)
Lexicalized Reordering model

<table>
<thead>
<tr>
<th>Wieviel</th>
<th>sollte</th>
<th>man</th>
<th>aufgrund</th>
<th>seines</th>
<th>Profils</th>
<th>in</th>
<th>Facebook</th>
<th>verdienen</th>
</tr>
</thead>
</table>

**m**: monotone (keep order)

**s**: swap order

How much should you charge for your Facebook profile?
Lexicalized Reordering model

m: monotone (keep order)
s: swap order
d: become discontinuous

How much you charge for your Facebook profile
**Lexicalized Reordering model**

<table>
<thead>
<tr>
<th>Wieviel</th>
<th>sollte</th>
<th>man</th>
<th>aufgrund</th>
<th>seines</th>
<th>Profils</th>
<th>in</th>
<th>Facebook</th>
<th>verdienen</th>
</tr>
</thead>
</table>

**m**: monotone (keep order)

**s**: swap order

**d**: become discontinuous

Reordering features are probability estimates of s, d, and m

---

How much should you charge for your Facebook profile?
Lexicalized Reordering table

- Identical phrase pairs `<f,e>` as in the phrase translation table
- Contains values for $p(\text{monotone}|e,f)$, $p(\text{swap}|e,f)$, $p(\text{discontinuous}|e,f)$

| Source        | Translation       | $p(m|e,f)$ | $p(s|e,f)$ | $p(d|e,f)$ |
|---------------|-------------------|-----------|-----------|-----------|
| naturlich     | of course         | 0.52      | 0.08      | 0.4       |
| naturlich     | naturally         | 0.42      | 0.1       | 0.48      |
| naturlich     | of course ,       | 0.5       | 0.001     | 0.499     |
| naturlich     | , of course       | 0.27      | 0.17      | 0.56      |
Discussion: Is this better?

• Do you think that this is a more sensible reordering model than the distance-based one?
• How could you determine if it is better or not?
• What do you think that it still lacks?
Empirically, yes!

Koehn et al, IWSLT 2005
“The Germans have another kind of parenthesis, which they make by splitting a verb in two and putting half of it at the beginning of an exciting chapter and the OTHER HALF at the end of it. Can any one conceive of anything more confusing than that? These things are called ‘separable verbs.’ The wider the two portions of one of them are spread apart, the better the author of the crime is pleased with his performance.”

Mark Twain
Ich werde Ihnen den Report aushändigen.
I will to_you the report pass_on.
Ich werde Ihnen den Report aushändigen.
I will to you the report pass on.

Ich werde Ihnen die entsprechenden Anmerkungen aushändigen.
I will to you the corresponding comments pass on.
Ich werde Ihnen den Report aushändigen.

I will to_you the report pass_on.

Ich werde Ihnen die entsprechenden Anmerkungen aushändigen.

I will to_you the corresponding comments pass_on.

Ich werde Ihnen die entsprechenden Anmerkungen am Dienstag aushändigen.

I will to_you the corresponding comments on Tuesday pass_on.
German free word order

The finite verb always appears in 2nd position, but *Any constituent* (not just the subject) can appear in the 1st position

I  will  to_you  the report  pass_on

To_you  will  I  the report  pass_on

The report  will  I  to_you  pass_on
Main clause

Ich werde Ihnen den Report aushaendigen,  
I will to_you the report pass_on  ,
Main clause

Ich werde Ihnen den Report aushaendigen,  
I will to_you the report pass_on,  

Subordinate clause

damit Sie den eventuell uebernehmen koennen .  
so_that you it perhaps adopt can.
Phrase-based models have an overly simplistic way of handling different word orders.

We can describe the linguistic differences between different languages.

Collins defines a set of 6 simple, linguistically motivated rules, and demonstrates that they result in significant translation improvements.
Step 1: Reorder the source language

Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen koennen.
Step 1: Reorder the source language

Ich werde Ihnen den Report **aussenden** , damit Sie den eventuell **übernehmen** können .

Ich **werde aussenden** Ihnen den Report , damit Sie **können übernehmen** den eventuell .
Step 1: Reorder the source language

Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen koennen.

(I will pass_on to_you the report, so_that you can adopt it perhaps.)
Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen koennen.

(I will pass_on to_you the report, so_that you can adopt it perhaps.)

Step 2: Apply the phrase-based machine translation pipeline to the reordered input.
I will pass on the Report to you.
Rule 1: **Verbs are initial in VPs**
Within a VP, move the head to the initial position
Clause Restructuring

Rule 1: **Verbs are initial in VPs**
Within a VP, move the head to the initial position

- **VVINF-HD** "uebernehmen" adopt
- **PDS-OA** "den" that
- **ADJD-MO** "eventuell" perhaps
- **VINF-HD** "koennen" can
Rule 2: **Verbs follow complementizers**
In a subordinated clause move the head of the clause to follow the complementizer.

- **KOUS-CP**
  - *damit*
  - so-that

- **PPER-SB**
  - *Sie*
  - you

- **VP-OC**
  - *uebernehmen*
  - adopt

- **VINF-HD**
  - *koennen*
  - can

Diagram:

```
S-MO
  └── KOUS-CP
      └── *damit*
          └── so-that

  ┐     └── PPER-SB
        └── *Sie*
            └── you

  ┐     └── VP-OC
        └── *uebernehmen*
            └── adopt

  └── VINF-HD
      └── *koennen*
          └── can
```

...
Rule 2: **Verbs follow complementizers**  
In a subordinated clause move the head of the clause to follow the complementizer.
Rule 3: **Move subject**

The subject is moved to directly precede the head of the clause

```
KOUS-CP
  |       |
  |       |
damit   so-that

VINF-HD
  |       |
  |       |
koennen can

PPER-SB
  |       |
  |       |
Sie     you

VP-OC
  |       |
  |       |
VVINF-HD
  |       |
uebernehmen    adopt
```
Rule 3: **Move subject**  
The subject is moved to directly precede the head of the clause.
Rule 4: **Particles**
In verb particle constructions, the particle is moved to precede the finite verb

\[
S \\
\downarrow \\
	ext{PPER-SB: } Wir \\
\text{VVINF-HD: } fordem \\
\text{NP-OA: } das Praesidium \\
\text{PTKVZ-SVP: auf *PARTICLE*}
\]
Rule 4: **Particles**

In verb particle constructions, the particle is moved to precede the finite verb.
Rule 5: **Infinitives**

Infinitives are moved to directly follow the finite verb within a clause.
**Rule 5: Infinitives**

Infinitives are moved to directly follow the finite verb within a clause.
Rule 6: **Negation**

Negative particle is moved to directly follow the finite verb

\[
\begin{align*}
S & \\
\text{PPER-SB} & \quad \text{VVINF-HD} \quad \text{VVINF-HD} \quad \text{OOER-OA} \quad \text{PTK-NEG} \quad \text{VP-OC} \\
\text{Wir} & \quad \text{konnten} \quad \text{eintragen} \quad \text{es} \quad \text{nicht} \quad \text{es} \\
\text{we} & \quad \text{could} \quad \text{submit} \quad \text{it} \quad \text{not} \\
\end{align*}
\]
Rule 6: **Negation**

Negative particle is moved to directly follow the finite verb
I will to you the report pass on, so that you it perhaps adopt can.
Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen koennen.

I will pass_on to_you the report, so_that you can adopt it perhaps.
Ich werde Ihnen den Report aushändigen, damit Sie den eventuell übernehmen können.

Now that seems less like the ravings of a madman.

I will pass_on to_you the report, so_that you can adopt it perhaps.
Experiments

- Parallel training data: Europarl corpus (751k sentence pairs, 15M German words, 16M English)
- Parsed German training sentences
- Reordered the German training sentences with their 6 clause reordering rules
- Trained a phrase-based model
- Parsed and reordered the German test sentences
- Translated them
- Compared against the standard phrase-based model without parsing/reordering
Significant improvement at $p<0.01$ using the sign test
Human Translation Judgments

• 100 sentences (10-20 words in length)
• Two annotators
• Judged two different versions
  – Baseline system’s translation
  – Reordering system’s translation
• Judgments: Worse, better or equal
• Sentences were chosen at random, systems’ translations were presented in random order
## Human Translation Judgments

<table>
<thead>
<tr>
<th></th>
<th>+</th>
<th>=</th>
<th>–</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotator 1</td>
<td>40%</td>
<td>40%</td>
<td>20%</td>
</tr>
<tr>
<td>Annotator 2</td>
<td>44%</td>
<td>37%</td>
<td>19%</td>
</tr>
</tbody>
</table>

+ = reordered translation better
– = baseline better
= = equal
I think it is wrong in principle to have such measures in the European Union.

I believe that it is wrong in principle to take such measures in the European Union.

I believe that it is wrong in principle, such measure in the European Union to take.
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reference</strong></td>
<td>I think it is wrong in principle to have such measures in the European Union</td>
</tr>
<tr>
<td><strong>Reordered</strong></td>
<td>I believe that it is wrong in principle to take such measures in the European Union</td>
</tr>
<tr>
<td><strong>Baseline</strong></td>
<td>I believe that it is wrong in principle, such measure in the European Union to take.</td>
</tr>
</tbody>
</table>
The current difficulties should encourage us to redouble our efforts to promote cooperation in the Euro-Mediterranean framework.

The current problems should spur us, our efforts to promote cooperation within the framework of the e-prozesses to be intensified.

The current problems should spur us to intensify our efforts to promote cooperation within the framework of the e-prozesses.
### Reference
The current difficulties should encourage us to redouble our efforts to promote cooperation in the Euro-Mediterranean framework.

### Baseline
The current problems should spur us, our efforts to promote cooperation within the framework of the e-prozesses to be intensified.

### Reordered
The current problems should spur us to intensify our efforts to promote cooperation within the framework of the e-prozesses.
To go on subsidizing tobacco cultivation at the same time is a downright contradiction.

At the same time, continue to subsidize tobacco growing, it is quite schizophrenic.

At the same time, to continue to subsidize tobacco growing is schizophrenic.
## Examples

<table>
<thead>
<tr>
<th>Reference</th>
<th>Baseline</th>
<th>Reordered</th>
</tr>
</thead>
<tbody>
<tr>
<td>To go on subsidizing tobacco cultivation at the same time is a downright contradiction.</td>
<td>At the same time, continue to subsidize tobacco growing, it is quite schizophrenic.</td>
<td>At the same time, to continue to subsidize tobacco growing is schizophrenic.</td>
</tr>
</tbody>
</table>
We have voted against the report by Mrs. Lalumiere for reasons that include the following:

We have voted, amongst other things, for the following reasons against the report by Mrs. Lalumiere:

We have, among other things, for the following reasons against the report by Mrs. Lalumiere voted:
We have voted against the report by Mrs. Lalumiere for reasons that include the following:

We have voted, amongst other things, for the following reasons against the report by Mrs. Lalumiere:

We have, among other things, for the following reasons against the report by Mrs. Lalumiere voted:
Discussion: Clause Restructuring

• Are you convinced that German-English translation has improved?
• Do you think that this is a good fit for phrase-based machine translation?
• What limitations does this method have?

(Discuss with your neighbor.)
Limitations

• Requires a parser for the source language
  – We have parsers for only a small number of languages
  – Penalizes “low resource languages”
  – Fine for translating from English into other languages

• Involves hand crafted rules

• Removes the nice language-independent qualities of statistical machine translation
Learning the Rules Automatically

- Great term project idea!
- “Improving a statistical MT system with automatically learned rewrite patterns”
  by Fei Xia and Michael McCord (COLING 2004)
Syntactic LMs

• Our goal is reorder the translated phrases so that they are grammatical English
• Isn’t the language model probability supposed to do that already?
• Instead of an n-gram model, could we augment the LM with syntactic information?
Problem: bottom up parsing requires whole sentence
We need the LM to be able to score partial translations
One possibility: Incremental parsing

```
S
- S/NP
  - S/PP
    - S/VP
      - VP/NN
        - NN
          - VP/NP
            - DT
              - president
            - NN
              - board
      - VP/NN
        - DT
          - the
    - NP/NP
      - NN
        - VP/NP
          - DT
            - meets
    - NP/NP
      - NN
        - VP/NP
          - DT
            - president
          - NN
            - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - the
          - NN
            - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - meets
      - NP/NP
        - NN
          - VP/NP
            - DT
              - president
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - the
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - meets
      - NP/NP
        - NN
          - VP/NP
            - DT
              - president
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - the
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - meets
      - NP/NP
        - NN
          - VP/NP
            - DT
              - president
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - the
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - meets
      - NP/NP
        - NN
          - VP/NP
            - DT
              - president
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - the
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - meets
      - NP/NP
        - NN
          - VP/NP
            - DT
              - president
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - the
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - meets
      - NP/NP
        - NN
          - VP/NP
            - DT
              - president
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - the
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - meets
      - NP/NP
        - NN
          - VP/NP
            - DT
              - president
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - the
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - meets
      - NP/NP
        - NN
          - VP/NP
            - DT
              - president
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - the
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - meets
      - NP/NP
        - NN
          - VP/NP
            - DT
              - president
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - the
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - meets
      - NP/NP
        - NN
          - VP/NP
            - DT
              - president
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - the
            - NN
              - board
      - NP/NP
        - NN
          - VP/NP
            - DT
              - meets
```
• We will talk more about syntax based LMs, and why they do a poor job selecting grammatical output later in the term.

• Thursday, we’ll move away from phrase-based MT and talk synchronous grammar models
• Questions about this material?
• Questions about the Project Proposals (due April 1, one week from today)?
• Reminder: set up meeting with Prof. Lopez or Post
“Some German words are so long that they have a perspective. Freundschaftsbezeigungen. Dilettantenaufdringlichkeiten. Stadtverordnetenversammlungen. These things are not words, they are alphabetical processions. And they are not rare; one can open a German newspaper at any time and see them marching majestically across the page—and if he has any imagination he can see the banners and hear the music, too.”

Mark Twain
“A dog is *der Hund*; now you put that dog in the genitive case, and is he the same dog he was before? No, sir; he is *des Hundes*; put him in the dative case and what is he? Why, he is *dem Hund*. Now you snatch him into the accusative case and how is it with him? Why, he is *den Hunden*. But suppose he happens to be twins and you have to pluralize him- what then? Why, they'll swat that twin dog around through the 4 cases until he'll think he's an entire international dog show. I don't like dogs, but I wouldn't treat a dog like that.”
“The Germans have an inhuman way of cutting up their verbs. Now a verb has a hard time enough of it in this world when it's all together. It's downright inhuman to split it up. But that's just what those Germans do. They take part of a verb and put it down here, like a stake, and they take the other part of it and put it away over yonder like another stake, and between these two limits they just shovel in German.”

Mark Twain