Data Intensive Linguistics — Lecture 18
Machine translation (V): Syntax-Based Models

Philipp Koehn

13 March 2006
Syntax-based SMT

• Why Syntax?

• Yamada and Knight: *translating into trees*

• Wu: *tree-based transfer*

• Chiang: *hierarchical transfer*

• Collins, Kucerova, and Koehn: *clause structure*

• Other approaches
The Challenge of Syntax

- The classical machine translation *pyramid*
Advantages of Syntax-Based Translation

- **Reordering** for syntactic reasons
  - e.g., move German object to end of sentence

- Better explanation for *function words*
  - e.g., prepositions, determiners

- Conditioning to *syntactically related words*
  - translation of verb may depend on subject or object

- Use of *syntactic language models*
  - ensuring grammatical output
Syntactic Language Model

- *Good syntax tree* → good English

- Allows for *long distance constraints*

- Left translation preferred by syntactic LM
String to Tree Translation

- Use of English *syntax trees* [Yamada and Knight, 2001]
  - exploit *rich resources* on the English side
  - obtained with statistical parser [Collins, 1997]
  - *flattened tree* to allow more reorderings
  - works well with syntactic language model
Yamada and Knight [2001]

Kare ha ongaku wo kiku no ga daisuki desu

[from Yamada and Knight, 2001]
## Reordering Table

| Original Order | Reordering       | p(reorder|original) |
|----------------|------------------|-------------|
| PRP VB1 VB2    | PRP VB1 VB2      | 0.074       |
| **PRP VB1 VB2**| **PRP VB2 VB1**  | **0.723**   |
| PRP VB1 VB2    | VB1 PRP VB2      | 0.061       |
| PRP VB1 VB2    | VB1 VB2 PRP      | 0.037       |
| PRP VB1 VB2    | VB2 PRP VB1      | 0.083       |
| PRP VB1 VB2    | VB2 VB1 PRP      | 0.021       |
| VB TO          | VB TO            | 0.107       |
| VB TO          | TO VB            | 0.893       |
| TO NN          | TO NN            | 0.251       |
| TO NN          | NN TO            | 0.749       |
Decoding as Parsing

- Chart Parsing

[Diagram of PRP: he, followed by Japanese words: kare ha ongaku wo kiku no ga daisuki desu]

- Pick Japanese words

- Translate into tree stumps
Decoding as Parsing

- Chart Parsing

- Pick Japanese words

- Translate into tree stumps
Decoding as Parsing

- Adding some *more entries*...
Decoding as Parsing

• Combine entries
Decoding as Parsing

\[ \text{PRP} \quad \text{he} \quad \text{PP} \quad \text{to} \quad \text{PP} \quad \text{to} \quad \text{VB} \quad \text{listening} \]

\[ \text{kare} \quad \text{ha} \quad \text{ongaku} \quad \text{wo} \quad \text{kiku} \quad \text{no} \quad \text{ga} \quad \text{daisuki desu} \]
Decoding as Parsing

He adores listening to music.
Decoding as Parsing

- **Finished** when all foreign words covered

Philipp Koehn  
DIL Lecture 18  
13 March 2006
Yamada and Knight: Training

- **Parsing** of the English side
  - using Collins statistical parser

- **EM training**
  - translation model is used to map training sentence pairs
  - EM training finds low-perplexity model
  → *unity of training and decoding* as in IBM models
Is the Model Realistic?

• Do English trees *match* foreign strings?

• Crossings between French-English [Fox, 2002]
  – 0.29-6.27 per sentence, depending on how it is measured

• Can be reduced by
  – *flattening tree*, as done by [Yamada and Knight, 2001]
  – detecting *phrasal* translation
  – *special treatment* for small number of constructions

• Most coherence between *dependency structures*
Inversion Transduction Grammars

- Generation of *both* English and foreign trees [Wu, 1997]

- Rules (binary and unary)
  
  - $A \to A_1A_2 \parallel A_1A_2$
  - $A \to A_1A_2 \parallel A_2A_1$
  - $A \to e \parallel f$
  - $A \to e \parallel *$
  - $A \to * \parallel f$

⇒ *Common binary tree* required

  - limits the complexity of reorderings
Syntax Trees

Mary did not slap the green witch

- English binary tree
Syntax Trees

- Spanish binary tree

Maria no daba una bofetada a la bruja verde
• *Combined tree* with reordering of Spanish
Inversion Transduction Grammars

• Decoding by parsing (as before)

• Variations
  – may use *real syntax* on either side or both
  – may use *multi-word* units at leaf nodes
Chiang: Hierarchical Phrase Model

- **Chiang** [ACL, 2005] (best paper award!)
  - context free bi-grammar
  - *one non-terminal* symbol
  - right hand side of rule may include non-terminals and terminals

- *Competitive* with phrase-based models in 2005 DARPA/NIST evaluation
Types of Rules

- **Word** translation
  - $X \rightarrow \textit{maison} \parallel \textit{house}$

- **Phrasal** translation
  - $X \rightarrow \textit{daba una bofetada} \mid \textit{slap}$

- **Mixed** non-terminal / terminal
  - $X \rightarrow X \textit{bleue} \parallel \textit{blue X}$
  - $X \rightarrow \textit{ne X pas} \parallel \textit{not X}$
  - $X \rightarrow X1 X2 \parallel X2 \text{ of } X1$

- **Technical rules**
  - $S \rightarrow S X \parallel S X$
  - $S \rightarrow X \parallel X$
Learning Hierarchical Rules

X → X verde || green X
Learning Hierarchical Rules

<table>
<thead>
<tr>
<th>Mary</th>
<th>did</th>
<th>not</th>
<th>slap</th>
<th>the</th>
<th>green</th>
<th>witch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maria no daba una</td>
<td>a la bruja verde</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ X \rightarrow a \text{ la } X \parallel \text{ the } X \]
Details of Chiang’s Model

• Too many rules
  → filtering of rules necessary

• Efficient parse decoding possible
  – hypothesis stack for each span of foreign words
  – only one non-terminal → hypotheses comparable
  – length limit for spans that do not start at beginning
Clause Level Restructuring [Collins et al.]

• Why clause structure?
  – languages differ vastly in their clause structure
    (English: SVO, Arabic: VSO, German: fairly free order;
     a lot details differ: position of adverbs, sub clauses, etc.)
  – large-scale restructuring is a problem for phrase models

• Restructuring
  – reordering of constituents (main focus)
  – add/drop/change of function words

• Details see [Collins, Kucerova and Koehn, ACL 2005]
Clause Structure

- Syntax tree from German parser
  - statistical parser by Amit Dubay, trained on TIGER treebank
Reordering When Translating

- Reordering when translating into English
  - tree is flattened
  - clause level constituents line up
Clause Level Reordering

<table>
<thead>
<tr>
<th>S</th>
<th>Clause: Ich werde Ihnen die entsprechenden Anmerkungen ausgeben.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAFIN-HD</td>
<td>I will you the corresponding comments pass on.</td>
</tr>
<tr>
<td>PPER-DA</td>
<td>so that you that perhaps in the vote include can.</td>
</tr>
<tr>
<td>NP-OA</td>
<td>damit Sie das eventuell APRD-MO bei ART-DA der NN-NK Abstimmung</td>
</tr>
<tr>
<td>ADJ-NK</td>
<td>entsprechenden</td>
</tr>
<tr>
<td>NN-NK</td>
<td>Anmerkungen</td>
</tr>
<tr>
<td>VVFIN</td>
<td>aushaendigen</td>
</tr>
<tr>
<td>$.</td>
<td>.</td>
</tr>
</tbody>
</table>

- Clause level reordering is a **well defined task**
  - label German constituents with their *English order*
  - done this for 300 sentences, two annotators, high agreement
Systematic Reordering German $\rightarrow$ English

- Many types of reorderings are **systematic**
  - *move verb group together*
  - *subject - verb - object*
  - *move negation in front of verb*

$\Rightarrow$ *Write rules by hand*
- apply rules to test and training data
- train standard *phrase-based* SMT system

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline system</td>
<td>25.2%</td>
</tr>
<tr>
<td>with manual rules</td>
<td>26.8%</td>
</tr>
</tbody>
</table>
Improved Translations

- we must also this criticism should be taken seriously.
  → we must also take this criticism seriously.

- i am with him that it is necessary, the institutional balance by means of a political revaluation of both the commission and the council to maintain.
  → i agree with him in this, that it is necessary to maintain the institutional balance by means of a political revaluation of both the commission and the council.

- thirdly, we believe that the principle of differentiation of negotiations note.
  → thirdly, we maintain the principle of differentiation of negotiations.

- perhaps it would be a constructive dialog between the government and opposition parties, social representative a positive impetus in the right direction.
  → perhaps a constructive dialog between government and opposition parties and social representative could give a positive impetus in the right direction.
Other Syntax-Based Approaches

- ISI: extending work of Yamada/Knight
  - more complex rules
  - performance approaching phrase-based

- Prague: Translation via dependency structures
  - parallel Czech–English dependency treebank
  - tecto-grammatical translation model [EACL 2003]

- U.Alberta/Microsoft: treelet translation
  - translating from English into foreign languages
  - using dependency parser in English
  - project dependency tree into foreign language for training
  - map parts of the dependency tree (“treelets”) into foreign languages
Other Syntax-Based Approaches

- **Reranking** phrase-based SMT output with syntactic features
  - create n-best list with phrase-based system
  - POS tag and parse candidate translations
  - rerank with syntactic features
  - see [Koehn, 2003] and JHU Workshop [Och et al., 2003]

- JHU Summer workshop 2005
  - **Genpar**: tool for syntax-based SMT
Syntax: Does it help?

• Not yet
  – best systems still phrase-based, treat words as tokens

• Well, maybe...
  – work on reordering German
  – ISI: better for short sentences Chinese–English
  – automatically trained tree transfer systems promising

• Why not yet?
  – if real syntax, we need good parsers — are they good enough?
  – syntactic annotations add a level of complexity
    → difficult to handle, slow to train and decode
  – few researchers good at statistical modeling and syntactic theories