Statistical Machine Translation
Lecture 5
Syntax-Based Models
Philipp Koehn
pkoehn@inf.ed.ac.uk
School of Informatics
University of Edinburgh

Outline
- Reminder: Modeling and Decoding
- Why Syntax?
- Yamada and Knight: translating into trees
- Wu: tree-based transfer
- Chiang: hierarchical transfer
- Koehn: clause structure
- Other approaches

Phrase-Based Translation Model

Decoding

Search Space for Decoding Too Big

Word-Based Translation Model

- Translation process is broken up into small step:
  word translation, reordering, duplication, insertion
- Decoding can be done similarly to phrase-based decoding
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

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]

Reordering Table

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

- Chart Parsing

```
he
```
```
kare ha ongaku wo kiku no ga daisuki desu
```

- Pick Japanese words
- Translate into tree stumps

Adding some more entries...

Combine entries
Decoding as Parsing

- Finished when all foreign words covered

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)
  - $A \rightarrow A_1A_2 \parallel A_1A_2$
  - $A \rightarrow A_1A_2 \parallel A_2A_1$
  - $A \rightarrow e \parallel f$
  - $A \rightarrow e \parallel s$
  - $A \rightarrow e \parallel f$
  - Common binary tree required
  - limits the complexity of reorderings

Syntax Trees

- English binary tree

Syntax Trees (2)

- Spanish binary tree
Syntax-Based Statistical Machine Translation

Syntax Trees (3)

- Combined tree with reordering of Spanish

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

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
- Reordering constraints of ITG used in phrase-based systems

Types of Rules

- Word translation
  - \( X \rightarrow \text{maison} \parallel \text{house} \)
- Phrasal translation
  - \( X \rightarrow \text{daba una bofetada} \parallel \text{slap} \)
- Mixed non-terminal / terminal
  - \( X \rightarrow X \text{bleue} \parallel \text{blue} X \)
  - \( X \rightarrow \text{ne X pas} \parallel \text{not X} \)
  - \( X \rightarrow X1 X2 \parallel X2 \text{of X1} \)
- Technical rules
  - \( S \rightarrow S X \parallel S X \)
  - \( S \rightarrow X \parallel X \)

Learning Hierarchical Rules

- \( X \rightarrow X \text{verde} \parallel \text{green} X \)
- \( X \rightarrow a \text{la} X \parallel \text{the} X \)
Details

- 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

Syntax-Based Statistical Machine Translation

Clause Level Restructuring

- Why clause structure?
  - languages differ vastly in their clause structure
    - (English: SVO, Arabic: VSO, German: fairly free order; a lot of 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
- ACL 2005 paper [Collins, Koehn, Kucerova]

Syntax-Based Statistical Machine Translation

Clause Structure

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

Syntax-Based Statistical Machine Translation

Reordering When Translating

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

Syntax-Based Statistical Machine Translation

Clause Level Reordering

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

Syntax-Based Statistical Machine Translation

Syntax-Aided Phrase-Based MT [Koehn]

- Approach:
  - stick with phrase-based system
  - special treatment for special syntactic problems
- Noun Phrase Translation
- Clause Level Restructuring
Systematic Reordering German → English

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

⇒ 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

Syntax: Does it help?

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

- Well, maybe...
  - work on reordering German
  - 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 understand syntactic theories