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

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

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 VB1 VB2    | PRP VB1 VB2| 0.974         |
| 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         |
| TO VB          | TO VB      | 0.893         |
| TO NN          | TO NN      | 0.251         |
| NN TO          | NN TO      | 0.749         |
Decoding as Parsing

- Chart Parsing

```
 PRP
 he

PP
 to

VB
 listening

VB
 adores
```

- Pick Japanese words
- Translate into tree stumps

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

- Finished when all foreign words covered
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) – $A \rightarrow A_1A_2 \parallel A_1A_2$ – $A \rightarrow A_1A_2 \parallel A_2A_1$ – $A \rightarrow e \parallel f$ – $A \rightarrow e \parallel \ast$ – $A \rightarrow \ast \parallel f$

$\Rightarrow$ Common binary tree required – limits the complexity of reorderings

Syntax Trees

- English binary tree
- Spanish binary tree

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 – $X \rightarrow \text{maison} \parallel \text{house}$
- Phrasal translation – $X \rightarrow \text{daba una bofetada} \parallel \text{slap}$
- Mixed non-terminal / terminal – $X \rightarrow X \text{ bleue} \parallel \text{blue X}$
  – $X \rightarrow \text{ne X pas} \parallel \text{not X}$
  – $X \rightarrow X_1X_2 \parallel X_2 \text{ of X}_1$
- Technical rules – $S \rightarrow S \parallel X$ – $S \rightarrow X \parallel X$
Learning Hierarchical Rules

Maria no daba una botefada a la bruja verde

Details of Chiang's Model

- Too many rules
  
  → filtering of rules necessary

- Efficient parse decoding possible
  
  – hypothesis stack for each span of foreign words
  
  – 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: SV0, Arabic: VSO, German: fairly free order;
  
  a lot details differ: position of adverbs, sub clauses, etc.)

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

- Details see [Collins, Kucerova and Koehn, ACL 2005]

Clause Level Reordering

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

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

System BLEU

- baseline system: 25.2%
  
  - with manual rules: 26.8%
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
- 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