Syntax-Based Machine Translation

Luke Zettlemoyer

Slides from Philipp Koehn, Matt Post
Levels of Transfer

$P(\text{i will do it tomorrow} | \text{Yo lo haré mañana}) = 0.8$

| English (E)       | $P(E | \text{lo haré})$ |
|------------------|------------------------|
| will do it       | 0.8                    |
| will do so       | 0.2                    |

| English (E)       | $P(E | \text{mañana})$ |
|------------------|------------------------|
| tomorrow         | 0.7                    |
| morning          | 0.3                    |
Goals of Translating with Syntax

- **Reordering driven by syntactic**
  - E.g., move German verb to final position
- **Better explanation for function words**
  - E.g., prepositions and determiners
- **Allow long distance dependencies**
  - Translation of verb may depend on subject or object, which can have high string distance
- **Will allow for the use of syntactic language models**
Syntactic Language Models

- Allows for long distance dependencies

- Left translation would be preferred!

Philipp Koehn SMT Tutorial 4 April 2006

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

Philipp Koehn SMT Tutorial 4 April 2006
String to Tree Translation

- Create English syntax trees during translation [Yamada and Knight, 2001]
  - very early attempt to learn syntactic translation models
  - use state-of-the-art parsers for training
  - allows us to model translation as a parsing problem, reusing algorithms, etc.
Yamada and Knight [2001]

- $p(f|e)$ is a generative process from an English tree to a foreign string

Kare ha ongaku wo kiku no ga daisuki desu
Learned Model

- 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
- Pick Japanese words
- Translate into tree stumps
Yamada and Knight: Decoding

- A Parsing Problem
  - Can use CKY Algorithm, with rules that encode reordering, inserted works

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

- Finished when all foreign words covered
Yamada and Knight: Decoding

- A Parsing Problem
  - Can use CKY Algorithm, with rules that encode reordering, inserted works

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

- Finished when all foreign words covered
Yamada and Knight: Training

- Want $P(f|e)$, where $e$ is an English parse tree
  - Parse the English side of bi-text
  - Use parser output as gold standard
- Many different derivations from $e$ to $f$ (for a fixed pair)
  - Use EM training approach
  - Same idea as IBM Models (but a bit more complex)
Is The Model Realistic?

- Do English trees align well onto foreign string?
- Crossings between French-English [Fox, 2002]
  - ~1-5 per sentence (depending on how you count)
- Can be reduced by
  - Flattening tree, as done by Yamada and Knight
  - Mixing in phrase level translations
  - Special casing many constructions
What about tree-to-tree?

- Consider the following trees:

Mary did not slap the green witch

- We might merge them as follows:

Mary did not slap * the green witch

Maria * no daba una bofetada a la bruja verde

Maria * no daba una bofetada a la bruja verde
Inversion Transduction Grammars (ITGs)

- Simultaneously generates two trees (English and Foreign) [Wu, 1997]

- Rules, binary and unary
  - $X \rightarrow X_1X_2 \parallel X_1X_2$
  - $X \rightarrow X_1X_2 \parallel X_2X_1$
  - $X \rightarrow e\parallel f$
  - $X \rightarrow e\parallel^*$
  - $X \rightarrow ^*\parallel f$

- Builds a common binary tree
  - Limits the possible reorderings
  - Challenging to model complete phrases
  - But, can do decoding as parsing, just like before!
Hierarchical Phrase Model [Chiang, 2005]

- Hybrid of ITGs and phrase based translation
- Word rules
  - \( X \rightarrow \text{maison} \parallel \text{house} \)
- Phrasal Rules
  - \( X \rightarrow \text{daba una bofetada} \parallel \text{slap} \)
- Mixed Terminal / Non-terminal Rules
  - \( X \rightarrow X \text{ bleue} \parallel \text{blue }X \)
  - \( X \rightarrow \text{ne }X \text{ pas} \parallel \text{not }X \)
  - \( X \rightarrow X_1 X_2 \parallel X_2 \text{ of }X_1 \)
- Technical Rules
  - \( S \rightarrow S \ X \parallel \ S \ X \)
  - \( S \rightarrow X \parallel X \)
Hierarchical Rule Extraction

- Include all word and phrase alignments
  - $X \rightarrow$ verde $||$ green
  - $X \rightarrow$ bruja verde $||$ green witch
  - ...

- Consider every possible rule, with variable for subphrases
  - $X \rightarrow X$ verde $||$ green $X$
  - $X \rightarrow$ bruja $X$ $||$ X witch
  - $X \rightarrow$ a la $X$ $||$ the $X$
  - $X \rightarrow$ daba una botefada $||$ slap $X$
  - ...

---

Learning a Phrase Translation Table

- Task: learn the model from a parallel corpus
- Three stages:
  - word alignment: using IBM models or other method
  - extraction of phrase pairs
  - scoring phrase pairs

Miles Osborne Machine Translation 13 February 2012
The Rest of The Details

- See paper [Chiang, 2005]
  - Model is done much like phrase-based systems
  - Too many rules → Need to prune

- Efficient parsing algorithms for decoding

- How well does it work?
  - Chinese-English: 26.8 → 28.8 BLEU
  - Competitive with phrase-based systems on most other language pairs, but lags behind when the language pair has modest reordering
  - There has been significant work on better ways of extracting translation rules, and estimating parameters
Tree to Tree Translation [Chiang, 2010]

- Very brief sketch, see paper for details!
Tree to Tree Translation

- Key idea: Learn synchronous tree substitution grammar
Tree to Tree Translation

- To make it work: Allow many different tree structures (when syntax doesn’t align directly)

Figure 4: Fuzzy tree-to-tree extraction effectively restructures the Chinese tree from Figure 2 in two ways but does not commit to either one.
Tree to Tree Translation

- And, the paper has tons of other details…
- But, let’s see the results!

<table>
<thead>
<tr>
<th>task</th>
<th>extraction</th>
<th>dist. lim.</th>
<th>rules</th>
<th>features</th>
<th>BLEU dev</th>
<th>BLEU test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Eng</td>
<td>string-to-string</td>
<td>10</td>
<td>440M</td>
<td>1k</td>
<td>32.7</td>
<td>23.4</td>
</tr>
<tr>
<td></td>
<td>string-to-string</td>
<td>20</td>
<td>440M</td>
<td>1k</td>
<td>33.3</td>
<td>23.7</td>
</tr>
<tr>
<td></td>
<td>tree-to-tree exact</td>
<td>20</td>
<td>50M</td>
<td>5k</td>
<td>32.8</td>
<td>23.9</td>
</tr>
<tr>
<td></td>
<td>tree-to-tree fuzzy</td>
<td>20</td>
<td>440M</td>
<td>160k</td>
<td>33.9</td>
<td>24.3</td>
</tr>
<tr>
<td></td>
<td>+ nesting</td>
<td>20</td>
<td>180M</td>
<td>79k</td>
<td>33.9</td>
<td>24.3</td>
</tr>
<tr>
<td>Ara-Eng</td>
<td>string-to-string</td>
<td>10</td>
<td>790M</td>
<td>1k</td>
<td>48.7</td>
<td>48.9</td>
</tr>
<tr>
<td></td>
<td>tree-to-tree exact</td>
<td>10</td>
<td>38M</td>
<td>5k</td>
<td>46.6</td>
<td>47.5</td>
</tr>
<tr>
<td></td>
<td>tree-to-tree fuzzy</td>
<td>10</td>
<td>790M</td>
<td>130k</td>
<td>49.4</td>
<td>49.7</td>
</tr>
<tr>
<td></td>
<td>+ nesting</td>
<td>10</td>
<td>190M</td>
<td>66k</td>
<td>49.2</td>
<td>49.8</td>
</tr>
</tbody>
</table>

Table 3: On both the Chinese-English and Arabic-English translation tasks, fuzzy tree-to-tree extraction outperforms exact tree-to-tree extraction and string-to-string extraction. Brackets indicate statistically insignificant differences ($p \geq 0.05$).
Clause Level Restructuring

- **Approach:**
  - Still use phrase-based system
  - First, parse the input sentence and reorder it
  - Then, pass it to the phrase-based translator

- **Why?**
  - Most long distance re-ordering is at the clause level
  - E.g., English: SVO, Arabic: VSO, German: relatively free order
  - Most other phenomena can be captured by the large phrase tables!
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.
Pre-ordering Model

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.)

Step 2: Apply the phrase-based machine translation pipeline to the reordered input.
I will pass on the Report to you.
Clause Restructuring

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
Rule 2: Verbs follow complementizers
In a subordinated clause move the head of the clause
to follow the complementizer
Clause Restructuring

Rule 2: Verbs follow complementizers
In a subordinated clause mote the head of the clause to follow the complementizer

S-MO

KOUS-CP

damit
so-that

VINF-HD

do
en

PPER-SB

g
i
en

VP-OC

VVINF-HD
uebernehmen
adopt

...
Rule 3: Move subject
The **subject** is moved to directly precede the **head** of the clause

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

- **S-MO**
  - **VINF-HD**
  - **koennen**
  - **PPer-SB**
  - **Sie**
  - **VP-OC**
  - **can**
  - **you**
  - **VVINF-HD**
  - **uebernehmen**
  - **adopt**
  - ...
Clause Restructuring

Rule 3: Move subject
The subject is moved to directly precede the head of the clause

KOUS-CP

damit
so-that

PPER-SB
Sie
you

VINF-HD
koennen
can

VP-OC

VVINF-HD
uebernehmen
adopt

...
Rule 4: Particles
In verb particle constructions, the particle is moved to precede the finite verb.
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.

Wir
we could not it
konnten nicht einreichen
es not submit

VP-OC

OOER-OA

VVINF-HD

PTK-NEG

S
Rule 5: Infinitives

Infinitives are moved to directly follow the finite verb within a clause.

```
PPER-SB  VVINF-HD  VVINF-HD  OOER-OA  PTK-NEG  VP-OC
Wir      konnten  einreichen es  nicht  es  nicht
we       could    submit     it    it    it
```
Rule 6: Negation

Negative particle is moved to directly follow the finite verb.

Wir konnten einreichen es nicht...
Clause Restructuring

Rule 6: Negation

Negative particle is moved to directly follow the finite verb

Wir konnten nicht einreichen es it
we could not submit it
"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
I will pass_on to_you the report, so_that you can adopt it perhaps.

Now that seems less like the ravings of a madman.
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
Bleu score increase

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>Annotator</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
Examples

| Reference | 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.
<table>
<thead>
<tr>
<th>Reference</th>
<th>The current difficulties should encourage us to redouble our efforts to promote coorperation in the Euro-Mediterranean framework.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The current problems should spur us, our efforts to promote coorperation within the framework of the e-prozesses to be intensified.</td>
</tr>
<tr>
<td></td>
<td>The current problems should spur us to intensify our efforts to promote cooperation within the framework of the e-prozesses.</td>
</tr>
</tbody>
</table>
Examples

Reference

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.
<table>
<thead>
<tr>
<th>Reference</th>
<th>We have voted against the report by Mrs. Lalumiere for reasons that include the following:</th>
</tr>
</thead>
</table>

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