Syntax-Based Alignment: Supervised or Unsupervised?

Hao Zhang and Daniel Gildea

Computer Science Department
University of Rochester
Tree-based Alignment

IP

⇒

IP

⇒

IP

NP

wo

dui

ni

jiang

⇒

NP

VP

⇒

NP

VP

⇒

NP

VP

I

tell

you

Zhang and Gildea
Two Examples of Syntax-Based Alignment Models

- Stochastic Inversion Transduction Grammars (Wu 1997)
  - Synchronous parsing
  - Binary bracketing grammar
  - Either straight or inverted

- Tree-to-String transformation (Yamada and Knight 2001, 2002)
  - Fixed parse on one side
  - Penn Treebank grammar
  - All possible re-orderings
Training of ITG

\[ N_s \]

\[ N_t \]

you

tell

I

wo  dui  ni  jiang

Z'

Y'

I

Z

[wo dui ni jiang]

Zhang and Gildea
One Parse for One Alignment
The Unambiguous Inversion Bracketing Grammar

- 4 Nonterminals: $S$, $A$, $B$, $C$
- $S$, the dedicated root symbol
- $C$, the dedicated preterminal
- $A$, for straight rules
- $B$, for inverted rules
- $A$ and $B$ alternate on the right
An Example of ITG Parse Tree
Training of Tree-to-String Model

\[
\begin{align*}
N & \quad \downarrow \quad l \\
& \quad \downarrow \quad k \\
& \quad \downarrow \quad 1 \\
A_1 & \quad A_2 & \quad A_3 & \quad A_4
\end{align*}
\]
Of the six possible re-orderings of the three terminals, the two are not allowed.
Allow Subtrees to be “Cloned”

Constituents of sentence can move to arbitrary locations, at a cost in probability.
## Categorisation of Alignment Parameters

|                      | ITG                                      | Tree-to-String                                      |
|----------------------|------------------------------------------|-----------------------------------------------------|
| lexical translation  | $P(C' \rightarrow e_i/f_j)$             | $P_t(f|e)$                                           |
| insertion            | $P(C' \rightarrow \epsilon/f_j)$       | $P_{\text{ins}}(\text{left}, \text{right}, \text{none}|\epsilon)$, $P_t(f|\text{NULL})$ |
| cloning              |                                          | $P_{\text{ins}}(\text{clone}|\epsilon)$, $P_{\text{clone}}(\epsilon_i|\text{clone} = 1)$ |
| deletion             | $P(C' \rightarrow e_i/\epsilon)$       | $P_t(\text{NULL}|e)$                                |
| re-ordering          | $P(A \rightarrow [AB])$                 | $P_{\text{order}}(\rho|\epsilon \Rightarrow \text{children}(\epsilon))$ |
|                      | $P(B \rightarrow \langle BA\rangle)$    |                                                     |
|                      | ...                                      |                                                     |
## Experiments

|                | Training Data | Evaluation Data | Cross-Validation Data |
|----------------|---------------|-----------------|-----------------------|
| English-Chinese| 18,773        | 48              | 49                    |
| English-French | 20,000        | 447             | 37                    |

\[
AER = 1 - \frac{|A \cap G_P| + |A \cap G_S|}{|A| + |G_S|}
\]

\[
G = G_S = G_P \text{ for English-Chinese gold standard alignments}
\]
ITG Training Curve

- Perplexity vs. Iterations
- AER
### Result: English-Chinese

| Model                                      | Precision | Recall | Alignment Error Rate |
|--------------------------------------------|-----------|--------|----------------------|
| IBM Model 1                                | .56       | .42    | .52                  |
| IBM Model 4                                | .67       | .43    | .47                  |
| Inversion Transduction Grammar             | .68       | .52    | .40                  |
| Tree-to-String w/ Clone                    | .65       | .43    | .48                  |
| Tree-to-String w/o Clone                   | .63       | .41    | .50                  |
### Result: Chinese-English

| Method                                      | Precision | Recall | Alignment Error Rate |
|---------------------------------------------|-----------|--------|----------------------|
| IBM Model 4                                 | .56       | .59    | .42                  |
| Inversion Transduction Grammar              | .68       | .52    | .40                  |
| Tree-to-String, automatic parses            | .61       | .48    | .46                  |
| Tree-to-String, gold parses                 | .61       | .52    | .44                  |
### Result: English-French

| Method                          | Precision | Recall | Alignment Error Rate |
|---------------------------------|-----------|--------|----------------------|
| IBM Model 1                     | .63       | .71    | .34                  |
| IBM Model 4                     | .83       | .83    | .17                  |
| Inversion Transduction Grammar  | .82       | .87    | .16                  |
| Tree-to-String w/ Clone        | .84       | .85    | .15                  |

Zhang and Gildea
Punctuation Raising

We see them.

Nous les voyons.
**Aditional Results: Tree2String Without Punctuation Raising**

|                           | Precision | Recall | Alignment Error Rate |
|---------------------------|-----------|--------|----------------------|
| Tree-to-String w/ Clone  | .84       | .85    | .15                  |
| Tree-to-String w/ Clone w/o PR | .71       | .75    | .27                  |
**Additional Results: Using Ambiguous ITG**

|                                | Precision | Recall | Alignment Error Rate |
|--------------------------------|-----------|--------|----------------------|
| unambiguous grammar            | .82       | .87    | .16                  |
| ambiguous, single constituent grammar | .80       | .87    | .18                  |
Summary

- Trees can help alignment
- Loosening constraints necessary

Future Work

- Synchronous parsing using realistic grammars
- Looking at large pieces of tree
Thanks

Rebecca Hwa
Mary Swift
Big-Picture Seminar @ URCS