A Tookit for Visualizing the Coherence of Tree-based Reordering with Word Alignments

Authors:

Gideon Maillette de Buy Wenniger,
Maxim Khalilov, Khalil Sima'an

Statistical Language Processing and Learning Lab at the Institute for Logic Language and Computation (ILLC), University of Amsterdam, the Netherlands
Big Picture: Motivation

- Improving SMT:
  - Better models
  - Better learning methods and decoding

- Better models:
  - More sensible alignments
  - Explicit language-specific reordering models
  - Adding all sorts of extra information

- This work: support search better models

- Data visualization facilitates SMT research
Motivating Example

- **Tree Structures**: basis syntactic SMT

- **Un-cohesiveness** resulting from negation

Why has there been no Health and Safety Committee meeting since 1998?

```
WHADVP
  Why
  has
  there been
```

```
SQ
  VBN
```

```
VP
  NP
  ?
```

```
NP
  V.W.: since 1998
```

```
PP
  NP
```

```
DT NNP CC NNP NNP NN IN NP
  no Health and Safety Committee meeting since CD
```

Waarom heeft het Comité voor veiligheid en hygiëne sinds 1998 niet meer vergaderd?

- Is this big NP subtree appropriate?
  - For translation?
  - For reordering?

- Insight in **coherence** trees and alignments
Basic Alignment Visualization

- Alignment between source and target sentence
  - General: m-to-n mappings between words

Now, however, he is to go before the courts once more because the public prosecutor is appealing.

Hij schijnt echter weer voor de rechtbank te moeten verschijnen, omdat de officier van justitie in beroep gaat.
Alignment Span of Nodes

- Alignment mapping function

\[ A(n) \rightarrow \{1, \cdots, m\}^* \]

- Span of target positions covered by subtree rooted at node \( n \)

**Definition 2.1 (Alignment Span)**

\[ \text{AlignmentSpan}(n) := \left[ a_{n_{\text{min}}}, a_{n_{\text{max}}} \right] = \left[ \min_{x \in \text{LeafNodes}(n)} \left( \min_{a_x' \in A(x)} a_x' \right), \max_{y \in \text{LeafNodes}(n)} \left( \max_{a_y' \in A(y)} a_y' \right) \right] \]
Cohesiveness

- Alignment **Cohesive** nodes: source side syntactic phrase pair

- Un-cohesive / **Alignment Violation**: Two distinct subtrees align within same target range

---

**Definition 2.2** (Alignment Violation)

\( \text{violates}(n', n) := \text{terminal}(n') \land n' \notin \text{descendants}(n) \land \)

\( (\text{AlignmentSpan}(n) = [a_{n_{\text{min}}}, a_{n_{\text{max}}}] \land \)

\( (a_{n_{\text{min}}} \leq A(n') \leq a_{n_{\text{max}}}) \)
I would therefore once more ask you to ensure that we get a Dutch channel as well.

**Ask** and **once** violate the alignment span of the S subtree.
Reordering with ITG Constraints

- **Origin:**
  *Inversion Transduction Grammars* (Wu, 1997):
  Bilingual Parsing

- **Application for Reordering:**
  - Basic: child nodes binary tree may be inverted
  - General tree: permute child nodes arbitrarily
  - Restriction to constituency parse
Tree Constrained Reordering

- **Monotonization**: Reorder source to match target order

- Formal definition precedence used

  **Definition 2.3** (Alignment Span Precedence)

  \[
  \text{AlignmentSpan}(c1) = [a_{1\text{min}}, a_{1\text{max}}] < \\
  \text{AlignmentSpan}(c2) = [a_{2\text{min}}, a_{2\text{max}}] \\
  := (a_{1\text{min}} < a_{2\text{min}}) \land (a_{1\text{max}} < a_{2\text{min}})
  \]

- All positions covered by \( c1 \) strictly precede those covered by \( c2 \)

- **Un-cohesiveness** causes order problems
Tree-Constrained Reordering

Yes, Mrs. Schroedter, I shall be pleased to look into the facts of this case when I have received your letter.

Mevrouw Schroedter, nadat ik uw brief ontvangen heb, zal ik de feiten die op dit vraagstuk betrekking hebben, zeker bestuderen.

- Right subtree moved forward, violated subtree not touched
Yes, Mrs Schroeder, I shall be pleased to look into the facts of this case when I have received your letter.

Mevrouw Schroeder, nadat ik uw brief ontvangen heb, zal ik de feiten die op dit vraagstuk betrekking hebben, zeker bestuderen.

Yes, Mrs Schroeder, I shall when I have your letter received be pleased to into the facts of this case look.

Mevrouw Schroeder, nadat ik uw brief ontvangen heb, zal ik de feiten die op dit vraagstuk betrekking hebben, zeker bestuderen.
- Browsing through aligned sentences
- Insight into alignment mapping sub-trees
- Assess quality reordering tree-constrained ITG
- Get ideas for new tree-transduction operations
Related Work

- Visualization tools
  - *Cairo* (Smith and Jahr, 2000)
  - *Yawat* (Germann, 2008)
  - *Stockholm Tree Aligner (STA)* (Volk et al., 2007)
Cairo: IBM model visualization
Yawat: focus on clarity

- Visualization alignment matrix
- Support manual annotation
- Dynamic highlighting
- Visualization parallel treebanks
- Focus on hand-annotated trees
Conclusions

- Toolkit targeted especially SMT people
- Focus on automatically generated resources and syntactic SMT
- Offers new functionality
- Goal: support SMT research
Future Work

- Heuristics in subtree reordering
- Tree modification
- Alignment refinement
Questions?


