Overview

Today

- some syntactic phenomena that make (machine) translation difficult
- some solutions from MT research

Word Order

- Languages differ in word order:
  - SVO: English, French, Mandarin, Russian, ...
  - SOV: Hindi, Latin, Japanese, Korean, ...
  - VSO, VOS, OSV, OSV exist, but less common
  - German is V2 in main clause, SOV in subordinate clause
  - Word order more flexible when function is morphologically marked

Example: German–English

der Mann, der den letzten Marathon gewonnen hat

the man who won the last marathon
Word Order

Translation units can be discontinuous

Example: German separable verb prefixes are clause-final

He proposes a trade
Er schlägt einen Handel vor
source side pre-reordering
- preprocess the source text to better match target language word order
- standard for some language pairs for phrase-based SMT
- various approaches based on syntactic analysis of source sentence:
  - hand-written rules [Nießen and Ney, 2000, Collins et al., 2005]
  - neural pre-reordering [Miceli Barone and Attardi, 2015]
- negative results for neural MT [Du and Way, 2017]

target side pre-reordering?
- in principle, we can reorder target side before training
- need second step to restore original target language word order
  → this is hard
- some research, but never became standard

Syntactic N-grams
- n-grams may not be meaningful if word order is flexible
- syntactic n-grams for evaluation: head-word-chain metric (HWCM) [Liu and Gildea, 2005]

Non-projective Structures
- classical example of context-sensitive structures in Swiss German (Zurich dialect)
dass mer em Hans s’Huus hälfed aastriiche
that we help Hans paint the house
**Non-projective Structures**

most syntax-based SMT systems are context-free
- either they can’t produce non-projective structures,
  or we use pseudo-projective arcs (dotted line)

- most syntax-based SMT systems are context-free
- either they can’t produce non-projective structures,
  or we use pseudo-projective arcs (dotted line)

**Subcategorization**

words occur with specific syntactic arguments
- complex mapping between syntactic arguments and meaning
- example
  - *remember* can have direct object or clausal object
  - *remind* can have direct object, and prep. or clausal object

  *he remembers his medical appointment.*
  *he remembers that he has a medical appointment.*
  *remind* can have direct object, and prep. or clausal object
  - direct object: recipient of information
  - prep. or clausal object: information
  - I remind him of his medical appointment.
  - I remind him that he has a medical appointment.

  ungrammatical (or semantically nonsensical):
  - “he remembers of his medical appointment.
  - “he reminds his medical appointment.

**Subcategorization**

words occur with specific syntactic arguments
- complex mapping between syntactic arguments and meaning
- example
  - *remember* can have direct object or clausal object
  - *remind* can have direct object, and prep. or clausal object

  *he remembers his medical appointment.*
  *he remembers that he has a medical appointment.*
  *remind* can have direct object, and prep. or clausal object
  - direct object: recipient of information
  - prep. or clausal object: information
  - I remind him of his medical appointment.
  - I remind him that he has a medical appointment.

  ungrammatical (or semantically nonsensical):
  - “he remembers of his medical appointment.
  - “he reminds his medical appointment.

**Subcategorization**

words occur with specific syntactic arguments
- complex mapping between syntactic arguments and meaning
- example
  - *remember* can have direct object or clausal object
  - *remind* can have direct object, and prep. or clausal object

  *he remembers his medical appointment.*
  *he remembers that he has a medical appointment.*
  *remind* can have direct object, and prep. or clausal object
  - direct object: recipient of information
  - prep. or clausal object: information
  - I remind him of his medical appointment.
  - I remind him that he has a medical appointment.

  ungrammatical (or semantically nonsensical):
  - “he remembers of his medical appointment.
  - “he reminds his medical appointment.

**Subcategorization**

words occur with specific syntactic arguments
- complex mapping between syntactic arguments and meaning
- example
  - *remember* can have direct object or clausal object
  - *remind* can have direct object, and prep. or clausal object

  *he remembers his medical appointment.*
  *he remembers that he has a medical appointment.*
  *remind* can have direct object, and prep. or clausal object
  - direct object: recipient of information
  - prep. or clausal object: information
  - I remind him of his medical appointment.
  - I remind him that he has a medical appointment.

  ungrammatical (or semantically nonsensical):
  - “he remembers of his medical appointment.
  - “he reminds his medical appointment.
**Syntax and Neural MT**

- Attentional encoder–decoder is less limited than previous approaches
  - Reordering can be learned by attention model
  - Consequence of subcategorization constraints: you should not translate syntactic arguments independently
  - Recurrent model can handle discontiguous and non-projective structures

**Linguistic Input Features**

- **disambiguate words by POS**
  - English | German
  - close<sub>verb</sub> | schließen
  - close<sub>adj</sub> | nah
  - close<sub>noun</sub> | Ende

**Experiments**

- **Features**
  - Lemmas
  - Morphological features
  - POS tags
  - Dependency tags
  - BPE tags

- **Data**
  - WMT16 training/test data
  - English↔German and English→Romanian
hypothesis:
• recurrent neural networks have recency bias
• instead, we want to induce syntactic bias

One theory of hierarchy

- Generate symbols sequentially using an RNN
- Add some “control symbols” to rewrite the history periodically
  - Periodically “compress” a sequence into a single “constituent”
  - Augment RNN with an operation to compress recent history into a single vector (→ “reduce”)
  - RNN predicts next symbol based on the history of compressed elements and non-compressed terminals (“shift” or “generate”)
  - RNN must also predict “control symbols” that decide how big constituents are
- We call such models recurrent neural network grammars.
Recurrent Neural Network Grammars

Trees as sequences

The hungry cat meows.

S( NP( The hungry cat ) VP( meows ) )

Tree traversals correspond to stack operations
Recurrent Neural Network Grammars

Trees as sequences

The hungry cat meows.

S( NP( The hungry cat ) VP( meows ) . )

Tree traversals correspond to stack operations

R. Sennrich
MT – 2018 – 13
24/26
Recurrent Neural Network Grammars

Trees as sequences

S( NP( The hungry cat ) VP( meows ) )

Tree traversals correspond to stack operations

R. Sennrich MT – 2018 – 13 24/26
## Trees as sequences

The hungry cat meows.

S( NP( The hungry cat ) VP( meows ) . )

Tree traversals correspond to stack operations

<table>
<thead>
<tr>
<th>Terminals</th>
<th>Stack</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>NT(S)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Terminals</th>
<th>Stack</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>NT(S)</td>
<td>(S)</td>
<td>NT(NP)</td>
</tr>
</tbody>
</table>

Recurrent Neural Network Grammars
<table>
<thead>
<tr>
<th>Terminals</th>
<th>Stack</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S)</td>
<td>NT(S)</td>
<td>NT(NP)</td>
</tr>
<tr>
<td>(S, NP)</td>
<td>GEN(The)</td>
<td>GEN(hungry)</td>
</tr>
</tbody>
</table>

The table and diagram illustrate the process of Recurrent Neural Network Grammars, showing the transitions between non-terminals (NT) and terminals as actions are applied to the stack.
<table>
<thead>
<tr>
<th>Terminals</th>
<th>Stack</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>The</td>
<td>(S)</td>
<td>NT(S)</td>
</tr>
<tr>
<td>The</td>
<td>(S NP)</td>
<td>NT(NP)</td>
</tr>
<tr>
<td>The</td>
<td>(S NP The)</td>
<td>GEN(The)</td>
</tr>
<tr>
<td>The</td>
<td>(S NP The hungry)</td>
<td>GEN(hungry)</td>
</tr>
<tr>
<td>The</td>
<td>(S NP The hungry cat)</td>
<td>GEN(cat)</td>
</tr>
<tr>
<td>REDUCE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Compress “The hungry cat” into a single composite symbol.
### Syntactic Composition

Need representation for: \((NP \ The \ hungry \ cat)\)
Syntactic Composition

Need representation for: (NP *The hungry cat*)

What head type?

Syntactic Composition

Need representation for: (NP *The hungry cat*)

What head type?

Syntactic Composition

Need representation for: (NP *The hungry cat*)

What head type?
Syntactic Composition

Need representation for: (NP *The hungry cat*)
Recurrent Neural Network Grammars

**Syntactic Composition**

Need representation for: (NP *The hungry cat*)

![Syntactic Composition Diagram](image)

Recurrent Neural Network Grammars

**Recursion**

Need representation for: (NP *The hungry cat*)
(NP *The (ADJP very hungry) cat*)

![Recursion Diagram](image)
Recurrent Neural Network Grammars

RNNGs in MT

Eriguchi et al., Feb 2017

- Basic idea: learn decoder-encoder MT model and RNNG on parallel data with parsed target side, sharing target word embedding parameters. (multi-task learning). To translate, just use MT model.

### Table 2: BLEU and RIBES scores by the baseline and proposed models on the test set. We use the bootstrap resampling method from Koehn (2004) to compute the statistical significance. We use ↑ to mark those significant cases with p < 0.05.

<table>
<thead>
<tr>
<th></th>
<th>De-En</th>
<th>Re-En</th>
<th>Cs-En</th>
<th>Jp-En</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BLEU</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NMT</td>
<td>16.61</td>
<td>12.03</td>
<td>11.22</td>
<td>17.88</td>
</tr>
<tr>
<td>NMT+RG</td>
<td>16.41</td>
<td>12.46</td>
<td>12.06</td>
<td>18.84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Jp-En (Dev)</th>
<th>BLEU</th>
<th>NMT+RG</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o Buffer</td>
<td>18.02</td>
<td>NMT+RG</td>
<td></td>
</tr>
<tr>
<td>w/o Action</td>
<td>17.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/o Stack</td>
<td>17.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NMT</td>
<td>17.75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: Effect of each component in RNNG.

Bibliography I


Bibliography II