Predicting Prepositions for SMT

Marion Weller, Alexander Fraser and Sabine Schulte im Walde

wellermn@ims.uni-stuttgarg.de

May 11, 2015
Introduction

Translating prepositions is difficult in SMT

- Convey the source-side meaning
- Meet target-side requirements
Introduction

Translating prepositions is difficult in SMT
- Convey the source-side meaning
- Meet target-side requirements

How are prepositions generated in translation?
Introduction

Translating prepositions is difficult in SMT
- Convey the source-side meaning
- Meet target-side requirements

How are prepositions generated in translation?
- **Functional prepositions**: determined by target-side requirements
  
  \[
  to \ believe \ in \ sth. \quad \rightarrow \quad \text{an etw. glauben}
  \]
Introduction

Translating prepositions is difficult in SMT

- Convey the source-side meaning
- Meet target-side requirements

How are prepositions generated in translation?

- **Functional prepositions**: determined by target-side requirements
  
  *to believe in sth.* \(\rightarrow\) *an etw. glauben*

- **Content-bearing prepositions**: largely determined by source-side preposition
  
  *to sit under/on the table* \(\rightarrow\) *unter/auf dem Tisch sitzen*
Translating prepositions is difficult in SMT

- Convey the source-side meaning
- Meet target-side requirements

How are prepositions generated in translation?

- **Functional prepositions**: determined by target-side requirements
  
  *to believe in sth.* → *an etw. glauben*

- **Content-bearing prepositions**: largely determined by source-side preposition
  
  *to sit under/on the table* → *unter/auf dem Tisch sitzen*

- **“In-between”**: source- and target-side play a role
  
  *go to the cinema/to the beach* → *ins Kino/an den Strand gehen*
Generating prepositions on the target-side

**Objective:** Model all subcategorized elements (PP/subject/object)
Generating prepositions on the target-side

**Objective:** Model all subcategorized elements (PP/subject/object)

**Abstract representation** in a morphology-aware SMT system

- Prepositions are substituted with place-holders
- All subcategorized elements are available in an abstract form
Generating prepositions on the target-side

**Objective:** Model all subcategorized elements (PP/subject/object)

**Abstract representation** in a morphology-aware SMT system
- Prepositions are substituted with place-holders
- All subcategorized elements are available in an abstract form

**Generation of prepositions** as post-processing step based on a combination of source- and target-side information
Generating prepositions on the target-side

**Objective:** Model all subcategorized elements (PP/subject/object)

**Abstract representation** in a morphology-aware SMT system
- Prepositions are substituted with place-holders
- All subcategorized elements are available in an abstract form

**Generation of prepositions** as post-processing step based on a combination of source- and target-side information
- Arguments are assigned their respective function
  - overt preposition → PP
  - empty preposition → NP
- Arguments are then inflected accordingly
Generating prepositions on the target-side

**Objective:** Model all subcategorized elements (PP/subject/object)

**Abstract representation** in a morphology-aware SMT system
- Prepositions are substituted with place-holders
- All subcategorized elements are available in an abstract form

**Generation of prepositions** as post-processing step based on a combination of source- and target-side information
- Arguments are assigned their respective function
  - overt preposition $\rightarrow$ PP
  - empty preposition $\rightarrow$ NP
- Arguments are then inflected accordingly
- Realization of prepositions is independent of structural mismatches

\[ \text{to pay attention to sth. } \rightarrow \text{auf etw. achten} \]
\[ \emptyset \text{ etw. beachten} \]
Overview of the morphology-aware translation process

<table>
<thead>
<tr>
<th>input</th>
<th>lemmatized SMT output</th>
<th>prep</th>
<th>morph. feat.</th>
<th>inflected</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>PREP</td>
<td>Ø-Prep</td>
<td>Ø-Nom</td>
<td>Ø-Gen</td>
<td>Ø-Gram</td>
</tr>
<tr>
<td>what</td>
<td>welch&lt;PWAT&gt;</td>
<td>Acc</td>
<td>Acc</td>
<td>Acc.Fem.Sg.Wk</td>
<td>welche</td>
</tr>
<tr>
<td>role</td>
<td>Rolle&lt;+NN&gt;&lt;Fem&gt;&lt;Sg&gt;</td>
<td>Ø-Nom</td>
<td>Ø-Nom</td>
<td>Nom.Masc.Sg.St</td>
<td>Rolle</td>
</tr>
<tr>
<td>Ø</td>
<td>PREP</td>
<td>Ø-Nom</td>
<td>Nom</td>
<td>Nom.Masc.Sg.Wk</td>
<td>der</td>
</tr>
<tr>
<td>the</td>
<td>die&lt;+ART&gt;&lt;Def&gt;</td>
<td>-</td>
<td>-</td>
<td>Nom.Masc.Sg.Wk</td>
<td>riesige</td>
</tr>
<tr>
<td>giant</td>
<td>riesig&lt;ADJ&gt;</td>
<td>bei-Dat</td>
<td>-</td>
<td>Nom.Masc.Sg.Wk</td>
<td>Planet</td>
</tr>
<tr>
<td>planet</td>
<td>Planet&lt;+NN&gt;&lt;Masc&gt;&lt;Sg&gt;</td>
<td>bei-Dat</td>
<td>-</td>
<td>Nom.Masc.Sg.Wk</td>
<td>gespielt</td>
</tr>
<tr>
<td>has</td>
<td>gespielt&lt;VVPP&gt;</td>
<td>bei-Dat</td>
<td>-</td>
<td>Nom.Masc.Sg.Wk</td>
<td>hat</td>
</tr>
<tr>
<td>played</td>
<td>hat&lt;VAFIN&gt;</td>
<td>bei-Dat</td>
<td>-</td>
<td>Nom.Masc.Sg.Wk</td>
<td>bei</td>
</tr>
<tr>
<td>in</td>
<td>PREP</td>
<td>bei-Dat</td>
<td>-</td>
<td>Nom.Masc.Sg.Wk</td>
<td>der</td>
</tr>
<tr>
<td>the</td>
<td>die&lt;+ART&gt;&lt;Def&gt;</td>
<td>bei-Dat</td>
<td>-</td>
<td>Nom.Masc.Sg.Wk</td>
<td>Entwicklung</td>
</tr>
<tr>
<td>development</td>
<td>Entwicklung&lt;+NN&gt;&lt;Fem&gt;&lt;Sg&gt;</td>
<td>bei-Dat</td>
<td>-</td>
<td>Nom.Masc.Sg.Wk</td>
<td>der</td>
</tr>
<tr>
<td>of</td>
<td>PREP</td>
<td>bei-Dat</td>
<td>-</td>
<td>Nom.Masc.Sg.Wk</td>
<td>Entwicklung</td>
</tr>
<tr>
<td>the</td>
<td>die&lt;+ART&gt;&lt;Def&gt;</td>
<td>bei-Dat</td>
<td>-</td>
<td>Nom.Masc.Sg.Wk</td>
<td>des</td>
</tr>
<tr>
<td>solar system</td>
<td>Sonnensystem&lt;+NN&gt;&lt;Neut&gt;&lt;Sg&gt;</td>
<td>bei-Dat</td>
<td>-</td>
<td>Nom.Masc.Sg.Wk</td>
<td>Sonnensystems</td>
</tr>
</tbody>
</table>
Outcome and Discussion

- Basic place-holder representation  
  → negative impact on translation quality
Outcome and Discussion

- **Basic place-holder representation**
  - negative impact on translation quality

- **Experiments with enriched place-holder representation**
  - annotation of grammatical case
  - annotation of governing verb/noun
  - annotation of subcategorization status

- **No improvement over the morphology-aware baseline** (BLEU and evaluation of preposition accuracy)

- **How to improve the current system?**
  - Abstract representation
    - case: light semantic annotation
    - more semantically motivated information to obtain a more meaningful representation of prepositions
Outcome and Discussion

- **Basic place-holder representation**
  → negative impact on translation quality

- **Experiments with enriched place-holder representation**
  - annotation of grammatical case ➔ best system
  - annotation of governing verb/noun
  - annotation of subcategorization status

How to improve the current system?

- **Abstract representation**
  - case: light semantic annotation
  - more semantically motivated information to obtain a more meaningful representation of prepositions
Outcome and Discussion

- **Basic place-holder representation**
  - negative impact on translation quality

- **Experiments with enriched place-holder representation**
  - annotation of grammatical case ⇐ best system
  - annotation of governing verb/noun
  - annotation of subcategorization status

- **No improvement over the morphology-aware baseline**
  (BLEU and evaluation of preposition accuracy)
Outcome and Discussion

• Basic place-holder representation
  → negative impact on translation quality

• Experiments with enriched place-holder representation
  – annotation of grammatical case
  – annotation of governing verb/noun
  – annotation of subcategorization status

  ⇩ best system

• No improvement over the morphology-aware baseline
  (BLEU and evaluation of preposition accuracy)

• How to improve the current system?
Outcome and Discussion

- **Basic place-holder representation**
  → negative impact on translation quality

- **Experiments with enriched place-holder representation**
  - annotation of grammatical case \( \Leftarrow \) best system
  - annotation of governing verb/noun
  - annotation of subcategorization status

- **No improvement over the morphology-aware baseline**
  (BLEU and evaluation of preposition accuracy)

- **How to improve the current system?**
  **Abstract representation**
  - case: light semantic annotation
  - more semantically motivated information to obtain
    a more meaningful representation of prepositions
Thank you!