Predicting Prepositions for SMT

Marion Weller, Alexander Fraser and Sabine Schulte im Walde

wellermn@ims.uni-stuttgart.de

May 11, 2015
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.* → *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.* → *an etw. glauben*

- **Content-bearing prepositions**: largely determined by source-side preposition
  
  *to sit under/on the table* → *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 $\rightarrow$ PP
  - empty preposition $\rightarrow$ 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 → PP
  - empty preposition → NP
- Arguments are then inflected accordingly
- Realization of prepositions is independent of structural mismatches

*to pay attention to sth.* → *auf etw. achten*
  *∅ etw. beachten*
Overview of the morphology-aware translation process

| input          | lemmatized SMT output | prep       | morph. feat.     | inflected          | gloss    |
|----------------|-----------------------|------------|------------------|--------------------|----------|
| Ø              | → PREP                | Ø-Acc      | –                | welche             | which    |
| what           | welch<PWAT>           | Acc        | Acc.Fem.Sg.Wk    | Rolle              | role     |
| role           | Rolle <+NN><Fem><Sg>  | Ø-Nom      | –                |                   |          |
| Ø              | → PREP                | Ø-Nom      | Nom              |                   |          |
| the            | die <+ART><Def>       | Nom        | Nom.Masc.Sg.St   |                   |          |
| giant          | riesig<ADJ>           | Nom        | Nom.Masc.Sg.Wk   |                   |          |
| planet         | Planet <+NN><Masc><Sg>| Nom        | Nom.Masc.Sg.Wk   |                   |          |
| has            | gespielt<VVPP>        | –          | –                |                   |          |
| played         | hat <+VAFIN>          | –          | –                |                   |          |
| in             | → PREP                | bei-Dat    | –                | beid              | for      |
| the            | die <+ART><Def>       | Dat        | Dat.Fem.Sg.St    |                   |          |
| development     | Entwicklung <+NN><Fem><Sg> | Dat     | Dat.Fem.Sg.Wk    |                   |          |
| of             | → PREP                | Ø-Gen      | –                |                   |          |
| the            | die <+ART><Def>       | Gen        | Gen.Neut.Sg.St   |                   |          |
| solar system    | Sonnensystem <+NN><Neut><Sg> | Gen    | Gen.Neut.Sg.Wk   |                   |          |

The table illustrates the process of morphology-aware translation, showing how lemmatized input is transformed and inflected based on the morphological features.
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
  - 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
  - annotation of governing verb/noun
  - annotation of subcategorization status
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)
Outcome and Discussion

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

- **Experiments with enriched place-holder representation**
  - annotation of grammatical case \(\iff\) 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?**
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)

- **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!