

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?

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

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 → 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 → 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*
 - Ø** *etw. beachten*

Overview of the morphology-aware translation process

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

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

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

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 ← 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!