Machine Translation with Source-Predicted Target Morphology

Joachim Daiber

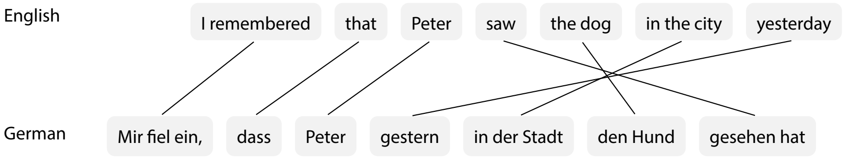
Institute for Logic, Language and Computation

University of Amsterdam

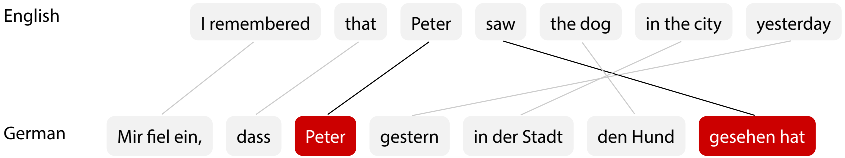
Translation into morphologically rich languages

English	I remembered	that	Peter	saw	the dog	in the city	yesterday
German	Mir fiel ein,	dass	Peter	gestern	in der Stadt	den Hund	gesehen hat

Translation into morphologically rich languages



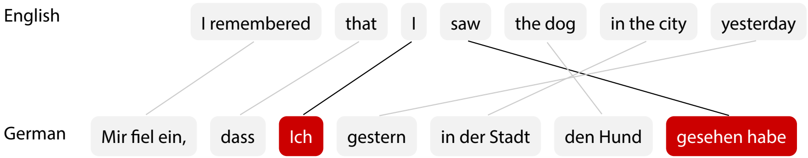
Translation into morphologically rich languages



Challenges:

- Morphological agreement over long distances

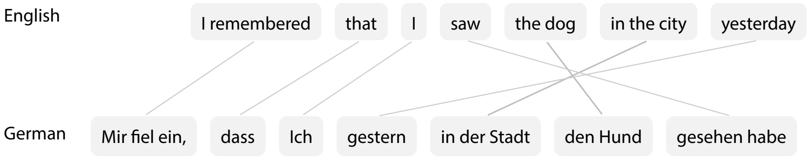
Translation into morphologically rich languages



Challenges:

- Morphological agreement over long distances

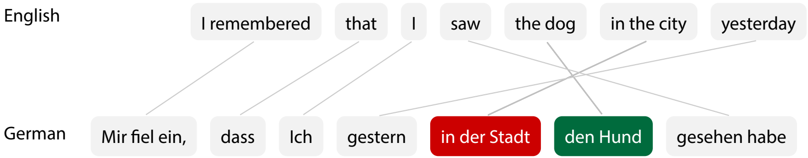
Translation into morphologically rich languages



Challenges:

- Morphological agreement over long distances

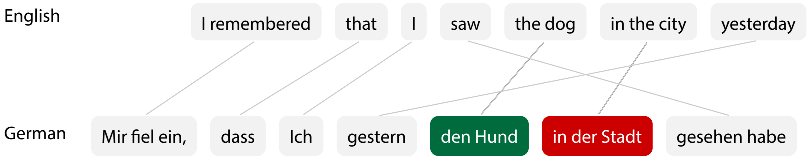
Translation into morphologically rich languages



Challenges:

- Morphological agreement over long distances
- Relatively freer word order

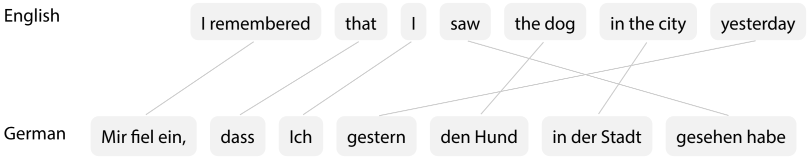
Translation into morphologically rich languages



Challenges:

- Morphological agreement over long distances
- Relatively freer word order

Translation into morphologically rich languages



Challenges:

- Morphological agreement over long distances
- Relatively freer word order
- Data sparsity

Translation into morphologically rich languages

- Established methods often do not work well
- One example: Source-side reordering

Source-predicted target morphology?

Hypothesis:

- Predicate-argument structure (PAS) of source and target are similar
- Linguistic information necessary for determining morph. target inflection resides in source sentence

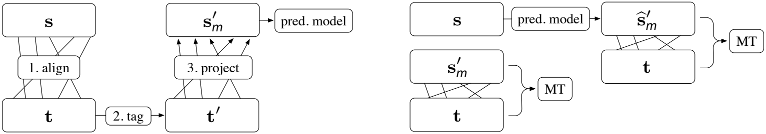
We explore:

- Target morphology as source-side prediction task
- Enriching source sentence with useful target properties

Three questions

1. Does knowing morphological target properties help?
2. Can we predict target morphology on the source PAS?
3. Which properties should we predict?

Does knowledge of morph. target properties help?



(a) Morphology projection. (b) MT system training.

Does knowledge of morph. target properties help?

		Translation	
Decoration	Tags	MTR	BLEU
None (baseline)	-	35.74	15.12
Proj. manual set	77	+2.43	+1.39
Proj. automatic set	225	+2.42	+1.20
Proj. full set	846	+2.72	+1.39

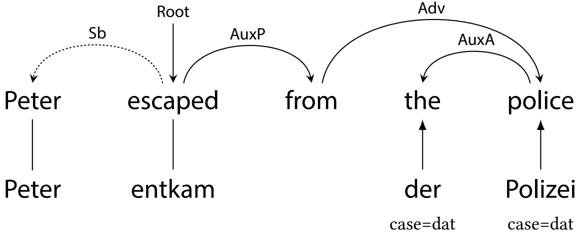
Table: Translation with various subsets of projected morphology (all $p < 0.01$).

Does knowledge of morph. target properties help?

		Word order	Lexical choice
Decoration	Tags	Kendall's τ	BLEU-1
None (baseline)	-	45.26	49.86
Proj. manual set	77	+4.20	+3.87
Proj. automatic set	225	+4.18	+3.39
Proj. full set	846	+4.57	+3.62

Table: Translation with various subsets of projected morphology (all $p < 0.01$).

Predicting target morphology on source trees



Source dependency chains

Prediction model:

- Conditional random field morphological tagger
- Instead of left-to-right: move down the dependency tree

Advantages of using source dependency chains:

- Access to syntactic information
- Soft enforcement of morphological agreement
- Combating data sparsity due to incomplete alignments

Which properties should we predict?

Problem: Many possible morphological target attributes:

- 846 combinations for German
- Might be redundant for the language pair
- Might be hard or even impossible to predict

Idea: Only include attributes if they lead to *better lexical selection*

Learning salient attributes

Procedure:

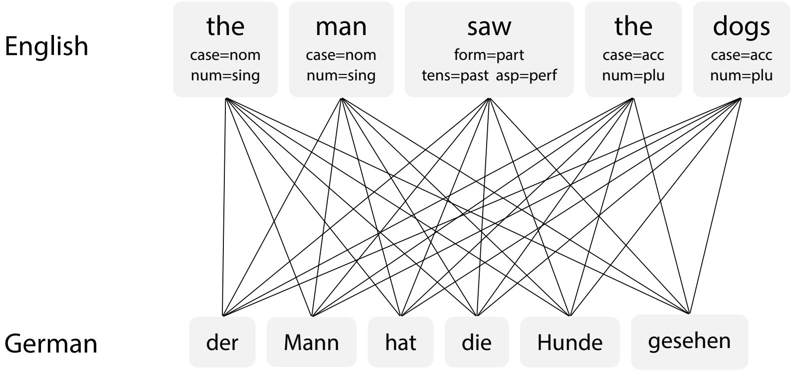
1. Decorate the source sentence with *all* attributes
2. Calc. likelihood of heldout set with word-based system (IBM model 1)
3. As long as the likelihood increases:
   - Find worst attribute by merging tags + recal. likelihood
   - Remove attribute, re-align
   - Repeat

Step 1: Decorate the source sentence with *all* attributes

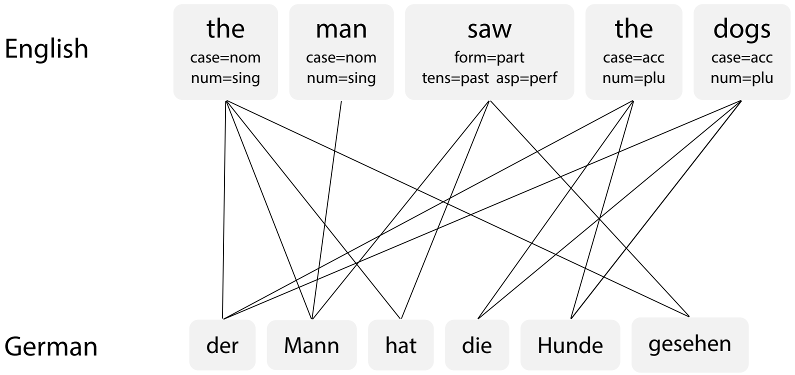
	the	man	saw	the	dogs
English	case=nom num=sing	case=nom num=sing	form=part tens=past asp=perf	case=acc num=plu	case=acc num=plu

German: der Mann hat die Hunde gesehen

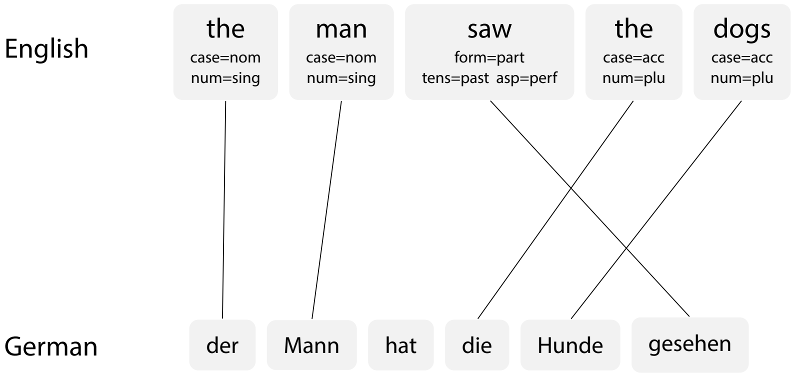
Step 2: Calc. heldout likelihhood with word-based MT



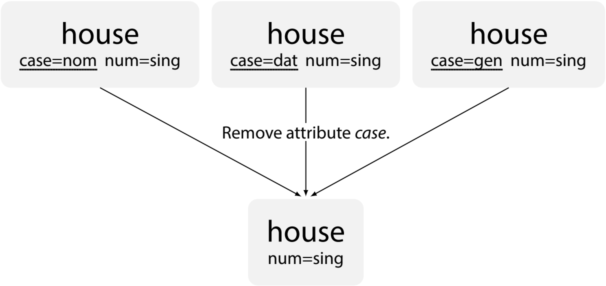
Step 2: Calc. heldout likelihhood with word-based MT



Step 2: Calc. heldout likelihhood with word-based MT



Step 3: Remove attributes by merging tags



Resulting morph. attributes (English–German)

Part of speech	Manual selection	Automatic selection
noun	gender† number case	gender number case
adj	gender† number‡ case‡ declension	gender number case synpos degree
verb	number‡* person‡* tense* mode*	-

Resulting morph. attributes (English–German)

	Manual selection	Automatic selection	All
Training time, 50k	36m	45m	77m
Training time, 100k	58m	82m	2h51m
Training time, 200k	1h54m	3h5m	6h44m
Tags	77	225	846
Best F_1	72.67	74.67	62.18

Integrating the predictions into the MT system

- Use dependency chain model to make predictions for test sentence
- Add sparse features to words and phrase:
  - Source morphology $\rightarrow$ target string suffixes and prefixes
  - Example: pos=det+gender=fem+number=sing+case=dat X $\rightarrow$ -er X

Results

	Translation		Word order	Lexical choice
Morph. attributes	MTR	BLEU	Kendall's τ	BLEU-1
No morphology	35.74	15.12	45.26	49.86
Manual selection	+0.74	+0.25	+2.10	+1.47
Autom. selection	+0.72	+0.27	+1.98	+1.35

Table: Translation with predicted test decorations (all $p < 0.05$).

Conclusion

- Novel approach: target morphology projection
- Realized as:
    1. Dependency chain model for predicting arbitrary target morphology
    2. Learning procedure to determine salient morphological attributes
    3. Strategies for integration into MT systems
- Current research direction: Interaction with word order.

Thank You!

Any questions?