Machine Translation with Source-Predicted Target Morphology

Joachim Daiber and Khalil Sima'an
ILLC, University of Amsterdam

OVERVIEW

- Novel pipeline for translation into morphologically rich languages
- Source enriched with target morphology
- Challenges:
  - Predicting target morphology
  - Learning salient attributes
  - Integration into MT systems

MOTIVATION

- Knowing relevant morphological target properties helps translation
- Possible improvements in both lexical selection and reordering

MORPHOLOGICAL ATTRIBUTES

Word type	Manual selection	Autom. selection
noun	gender† number case	gender number case
adj	gender† number‡ case‡ declension	gender number case synpos degree
verb	number‡* person‡* tense* mode*	-

† Transferred with lemma. ‡ Propagated from noun. * Dropped in later work.

MT INTEGRATION

- Strategies:
  - Training on Viterbi predictions
  - Training on gold projections
- Integration as sparse features, such as

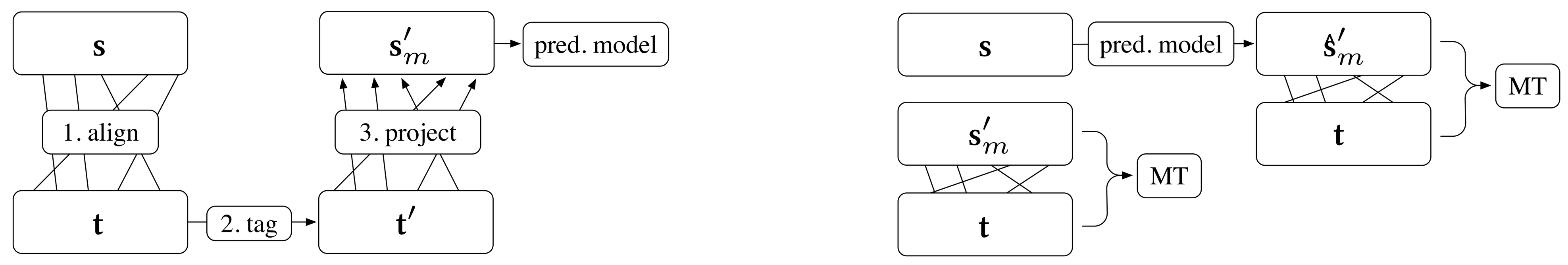
`gender=fem+number=sing+case=dat X` $\rightarrow$ `-er X`

REFERENCES

[1] Thomas Müller, Helmut Schmid, and Hinrich Schütze. Efficient higher-order CRFs for morphological tagging. In *Proceedings of EMNLP 2013*, pages 322–332, Seattle, USA, 2013.

[2] Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of ACL 2006*, pages 433–440, Sydney, Australia, 2006.

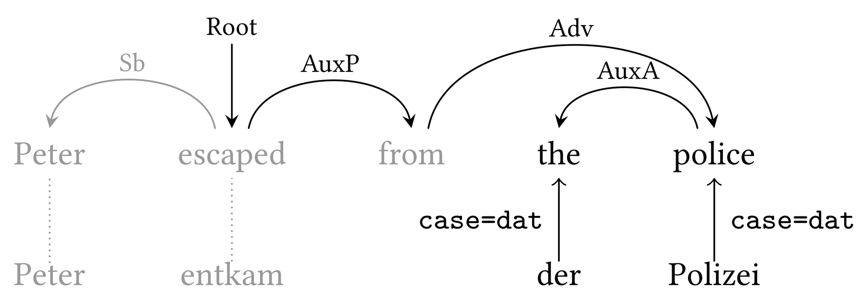
SYSTEM TRAINING AND TRANSLATION



(a) Morphology projection and pred. model training.

(b) Machine translation system training.

MODELING TARGET-SIDE MORPHOLOGY



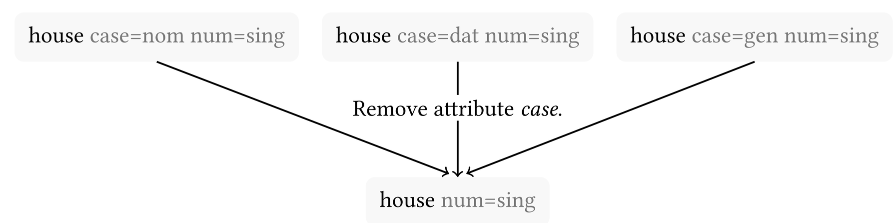
- Source-side dependency chains:
  - Word order might differ significantly
  - Source predicate-argument structure more informative for predicting target morph.
- $P(\mathbf{s}'_m \mid \tau, \mathbf{s})$: Source-side dependency chain model to predict morph. enriched source $\mathbf{s}'_m$.
  - Estimation: Coarse-to-fine CRF [1]
  - Decoding: root $\rightarrow$ leaves
  - Features as in morph. tagging and additionally: dependency labels, number of children, source POS and child tokens.
- Best performance:
  - 5^{th} order CRF
  - Trained on 50k-100k dependency chains
  - Dep. chains from non-isomorphic trees

LEARNING SALIENT MORPHOLOGICAL ATTRIBUTES

- Consider only attributes helpful for language pair (less sparsity, better predictions)
- Salient attribute: attribute that enables better lexical selection
- Determine best attributes:
  - Simple IBM model 1 translation model:

$$P(\mathbf{t} \mid \mathbf{s}) = \sum_{\mathbf{s}'_m} P(\mathbf{s}'_m \mid \mathbf{s}) P(\mathbf{t} \mid \mathbf{s}'_m)$$

  - Find best attributes by merging tags [2]
  - Merging tag occurrences $\rightarrow$ removing morph. attribute



EVALUATION

		Translation		Word order	Lexical choice
Morphological attributes	Training decor.	METEOR	BLEU	Kendall's τ	BLEU-1
No morphology	-	35.74	15.12	45.26	49.86
Manual selection	Predicted	35.85	15.19	45.43	50.01
	Projected	34.63[A]	14.00[A]	44.07	48.75
Autom. selection	Predicted	35.99[AC]	15.23[B]	45.88	50.27
	Projected	35.98[AC]	15.22[C]	45.89	50.27

[A] Statistically significant against baseline at $p < 0.05$ [B] Statistically significant against baseline at $p < 0.06$ [C] Statistically significant against Manual selection at $p < 0.05$

Phrase-based MT setup for English-to-German.

ACKNOWLEDGEMENTS

EXPERT

The first author is supported by the EXPERT (EXPloiting Empirical appRoaches to Translation) Initial Training Network (ITN) of the European Union's Seventh Framework Programme.