Immediate Adaptation to User Corrections in Post-Editing SMT

Patrick Simianer, Sariya Karimova, Stefan Riezler
Heidelberg University, Germany

Oct 28, 2016
iMT 2016 : AMTA 2016 Workshop on Interacting with Machine Translation

TOC

Motivation

User-adaptation in computer-aided translation (CAT) is crucial

1. To overcome domain shifts between training data and translated materials
2. To prevent frustrations with translation technology, e.g. related to post-editing
3. To boost efficiency and (possibly) quality

WO 2007000372 A1

[title] Sheathed element glow plug

[abstract segment #1] A sheathed element glow plug (1) is to be placed inside a chamber (3) of an internal combustion engine.

[abstract segment #2] The sheathed element glow plug (1) comprises a heating body (2) that has a glow tube (6) connected to a housing (4).

⋮

WO 2007031371 A1

[title] Sheathed element glow plug

[abstract segment #1] A sheathed element glow plug (1) serves for arrangement in a chamber of an internal combustion engine.

[abstract segment #2] The sheathed element glow plug comprises a heating body (2) which has a glow tube (5) and a heating coil (8) which is arranged in the glow tube (5).

⋮

Motivation

- Translation memories naturally adapt to their users, this raises expectations
  - → But updating SMT-based CAT systems is not straight-forward
- Adaptation by re-training (overnight) is useful
  - → But it can't help during translation sessions, as it's a slow process
- Online adaptive SMT is well studied and there are even products[1] that implement it
  - → But most research is theoretical, user studies are scarce
  - → Adaptation is potentially inprecise due to automatic alignment methods

[1] lilt.com, SDL Trados

Proposed approach

We present an approach to online user adaptive post-editing with precise, immediate adaptation:

$\Rightarrow$ By leveraging user-generated alignments for phrase-table adaptation

$\Rightarrow$ We evaluate our approach to adaptation in a user study

Definition of online adaptation

For each example $t = 1, \ldots, |d|$

1. Receive input sentence x_t
2. Output translation $\hat{y}_t$ from current model
3. Receive user output y_t
4. Refine models on $(x, \hat{y}, y)_t$

Figure: Online learning procedure in computer-aided translation

Related work

W/o user study: Bertoldi et al. [2014][2], Ortiz-Martínez et al. [2010][3], Wuebker et al. [2015b][4]

W/ user study: Green et al. [2014][5], Denkowski [2015][6]

Automatic alignment model: Bertoldi et al. [2014], Denkowski [2015], Ortiz-Martínez et al. [2010]

Tuning only: Green et al. [2014], Wuebker et al. [2015b]

[2] Online adaptation to post-edits for phrase-based statistical machine translation
[3] Online Learning for Interactive Statistical Machine Translation
[4] Hierarchical Incremental Adaptation for Statistical Machine Translation
[5] Human Effort and Machine Learnability in Computer Aided Translation
[6] Machine Translation for Human Translators

Related work – Evaluation

Quality

- Measure BLEU/TER of post-edits wrt. given reference translations (not necessarily meaningful)

Simulated quality

- Measure BLEU/TER of unaltered MT outputs wrt. given reference translations (identical to standard MT evaluation)

Manual effort

1. Measure BLEU/TER of MT outputs wrt. post-edits [HTER]
2. Measure and normalize counts of clicks and keystrokes

Simulated manual effort

1. Measure TER/BLEU wrt. offline created post-edits
2. Use a model of user behavior to estimate number of clicks/keystrokes needed to produce reference translation from MT output

Related work – Evaluation

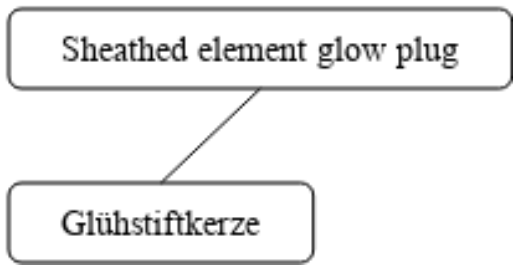
- Ortiz-Martínez et al. [2010]: Improved simulated quality and simulated manual effort compared to static systems
- Bertoldi et al. [2014]: Improved simulated quality compared to static systems
- Green et al. [2014]: Improved simulated manual effort compared to non-adapted system
- Wuebker et al. [2015b]: Improved simulated quality compared to baseline system
- Denkowski [2015]: Improved simulated quality and manual effort compared to static systems

1 Motivation

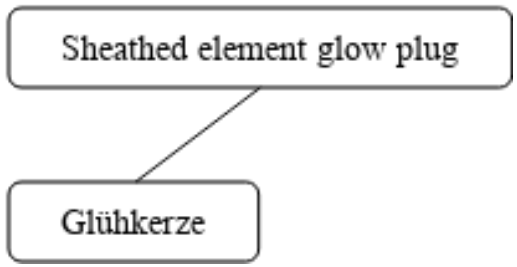
2 Proposed approach

3 User study

Example – MT output #1

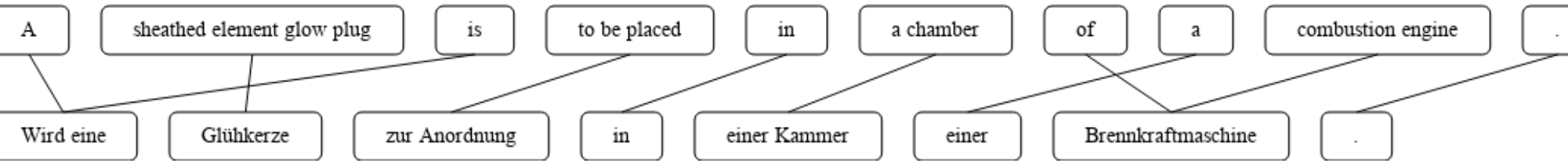


Example – User correction #1

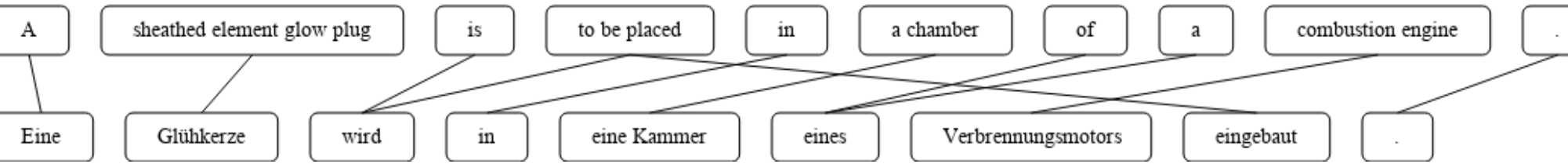


- sheathed element glow plug $\rightarrow$ Glühkerze

Example – MT output #2



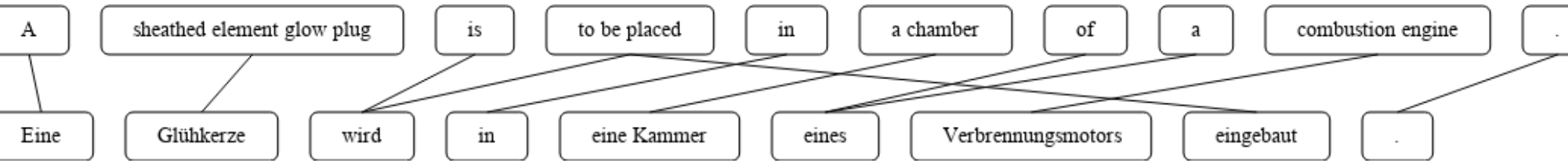
Example – User correction #2



Immediately learned translation rules:

- $a_0 \rightarrow$ eine
- is to be placed$_{2,3}$ $X_1 \rightarrow$ wird X_1 eingebaut
- a chamber$_5 \rightarrow$ eine Kammer
- of $a_{6,7} \rightarrow$ eines
- combustion engine$_8 \rightarrow$ Verbrennungsmotors

Example – User correction #2

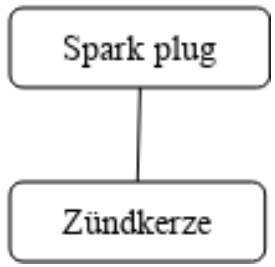


Derived translation rules:

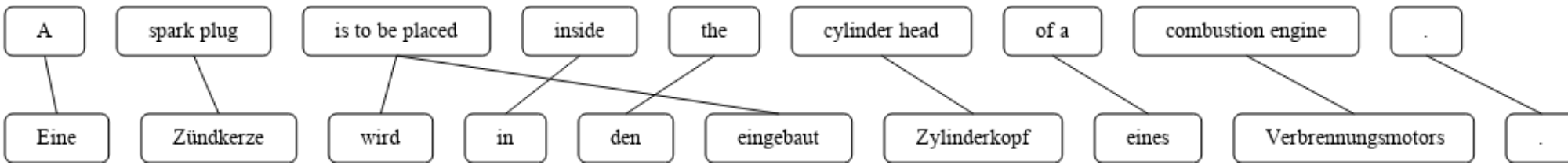
- in a chamber $\rightarrow$ in eine Kammer
- of a combustion engine $\rightarrow$ eines Verbrennungsmotors
- in a chamber of a combustion engine $\Rightarrow$ in eine Kammer eines Verbrennungsmotors
- in a chamber of X_1 combustion engine $\rightarrow$ in eine Kammer X_1 Verbrennungsmotors

$\vdots$

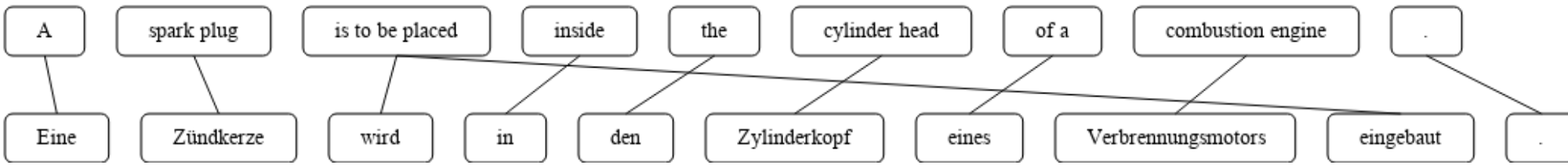
Example – MT output #3



Example – MT output #4



Example – User correction #4



Approach – Weight updates

- Pairwise ranking updates to weigh many sparse features, e.g. rule ids
- Per coordinate learning rates used to prevent too harsh changes
- Default learning rate for id features of newly extracted rules is the overall median
- Leave-one-out: Derived translation rules are only added to subsequent grammars to prevent overfitting

Approach – Summary

1. (User correction received)
2. Extract immediate corrections from post-edit and alignment and add to current grammar
3. Re-translate input with new grammar to generate k-best list
4. Pairwise ranking update using k-best
5. Add N-grams of post-edit to adaptive language model (following Denkowski et al. [2014])
6. Derive all possible rules from user correction

⋮

1. Motivation
2. Proposed approach
3. User study

User study – Setup

Subjects

19 students, 13 prospective translators, 6 CS students, 4 different mother tongues

Data

Titles and abstracts of patent documents, filtered by length, clustered by similarity

Environment

Controlled environment in a computer pool, 90 minute sessions

Machine translation

Hierarchical phrase-based system built from title/abstract training data, good baseline translation results

Task

Post-edit about 500 words from English into German, each task is shared by two subjects

User study – Results

response variable	estimated Δ	
HBLEU_{+1}	$+6.8 \pm 2.0$ [%]	$\mathrm{p} < 0.001$
HTER	-5.3 ± 1.9 [%]	$\mathrm{p} < 0.01$
normalized time	-118 ms	—

Table: Estimated differences in the response variables contrasting non-adaptive to adaptive systems. MT metrics calculated by comparing original MT outputs to user corrections.

Summary

- Novel graphical interface with (phrase-) alignments for a new form of interactive post-editing
- Alignment can be used for immediate and bulk adaptation of the translation model
- User study shows significant reductions in manual effort and slight speed improvement

Our code open source: `https://github.com/pks/lfpe`

Questions?

Thank you!

References I

Nicola Bertoldi, Patrick Simianer, Mauro Cettolo, Katharina Wäschle, Marcello Federico, and Stefan Riezler. Online adaptation to post-edits for phrase-based statistical machine translation. *Machine Translation*, 28, 2014.

M. Denkowski. *Machine Translation for Human Translators*. PhD thesis, Carnegie Mellon University, 2015.

Michael Denkowski, Alon Lavie, Isabel Lacruz, and Chris Dyer. Real time adaptive machine translation for post-editing with cdec and transcenter. In *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation*, 2014.

Spence Green, Sida Wang, Jason Chuang, Jeffrey Heer, Sebastian Schuster, and Christopher D. Manning. Human effort and machine learnability in computer aided translation. In *Empirical Methods in Natural Language Processing*, 2014.

References II

Benjamin Marie and Aurélien Max. Touch-based pre-post-editing of machine translation output. In *EMNLP*, 2015.

Daniel Ortiz-Martínez, Ismael García-Varea, and Francisco Casacuberta. Online learning for interactive statistical machine translation. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 2-4, 2010, Los Angeles, California, USA*, 2010.

Katharina Wäschle and Stefan Riezler. Analyzing Parallelism and Domain Similarities in the MAREC Patent Corpus. *Multidisciplinary Information Retrieval*, 2012.

Joern Wuebker, Spence Green, and John DeNero. Hierarchical incremental adaptation for statistical machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015a.

References III

Joern Wuebker, Spence Green, and John DeNero. Hierarchical incremental adaptation for statistical machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015b.