Human Effort and Machine Learnability in Computer Aided Translation

Spence Green

Sida I. Wang    Jason Chuang*    Jeffrey Heer*
Sebastian Schuster    Christopher D. Manning

Stanford University    *University of Washington

EMNLP // 27 October 2014
For high accuracy...pure MT has to be given up in favor of mixed MT in which a human brain intervenes.
An Inauspicious Start...

The Georgetown University MT project tried to produce useful output in 1962, [but] they had to resort to postediting. The postedited translation took slightly longer to do and was more expensive than conventional human translation.

[ALPAC, 1966, p.19]
34+ Years of Interactive MT

Post-edit now better than unaided...but tedious for users
34+ Years of Interactive MT

Post-edit now better than unaided...but tedious for users

Clever interactive systems, disappointing user studies...

😊 Interactive Translation System (ITS) [Melby et al. 1980]
34+ Years of Interactive MT

Post-edit now better than unaided...but tedious for users

Clever interactive systems, disappointing user studies...

😊 Interactive Translation System (ITS)  [Melby et al. 1980]
😊 TransType  [Langlais and Lapalme 2002]
34+ Years of Interactive MT

Post-edit now better than unaided...but tedious for users

Clever interactive systems, disappointing user studies...

😊 Interactive Translation System (ITS) [Melby et al. 1980]
😊 TransType [Langlais and Lapalme 2002]
😊 TransType2 [Macklovitch 2006]
34+ Years of Interactive MT

Post-edit now better than unaided...but **tedious for users**

Clever interactive systems, disappointing user studies...

- Interactive Translation System (ITS) [Melby et al. 1980]
- TransType [Langlais and Lapalme 2002]
- TransType2 [Macklovitch 2006]
- Caitra [Koehn 2009]
<table>
<thead>
<tr>
<th>Quality</th>
<th>Slow</th>
<th>Fast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hi</td>
<td>Dissemination</td>
<td>Assimilation</td>
</tr>
</tbody>
</table>

Speed

- Slow
- Fast

Quality

- Lo
- Hi
This Paper

Compare post-edit vs. new interactive MT

[Green et al. 2014]
This Paper

Compare post-edit vs. new interactive MT [Green et al. 2014]

Two questions

1. Human productivity – time and quality?
2. Facilitates adaptive MT?
This Paper

Compare post-edit vs. new interactive MT [Green et al. 2014]

Two questions

1. Human productivity – time and quality?
2. Facilitates adaptive MT?

Result – Interactive works when MT model is good
Applied previous translation user study [Green et al. 2013b]
Insights | Interactive Mode

Applied previous translation user study [Green et al. 2013b]

Mixed-initiative principles [Horvitz 1999]
Insights | Interactive Mode

Applied previous translation user study  
[Green et al. 2013b]

Mixed-initiative principles  
[Horvitz 1999]

Significant Phrasal revision to support interactions
User Interface Walkthrough
Parliament Does Not Support Amendment Freeing Tymoshenko

Today, the Ukraine parliament dismissed, within the Code of Criminal Procedure amendment, the motion to revoke an article based on which the opposition leader, Yulia Tymoshenko, was sentenced.

Heute ist die Ukraine Parlament entlassen, innerhalb der Code der strafrechtlichen Verfahren Änderungsantrag, den Antrag auf Revoke einen Artikel basiert auf, die die Opposition leader, yulia Tymoshenko, war Sentenced.
Interactive Mode

Des enseignants se rendent régulièrement auprès des élèves de l’institut Jedličkův et leur proposent des activités qui les intéressent et les amusent.

Teachers regularly visit Jedličkův Institute students and offered them activities of interest to them and having fun.

Les étudiants eux-mêmes n’ont pas les moyens de se rendre à des cours, nous essayons de les aider de cette manière.

The students themselves cannot be required to attend courses, we are trying to help themselves cannot themselves could not themselves do not themselves cannot afford

Dans le cadre de l’institut Jedlička, nous transférons ce projet dans un prochain...
Interactions

Three classes:

Source comprehension

Target gisting

Target generation
Interactions

**Source comprehension**

Word-to-word translations

Source coverage – highlight translated words

**Target gisting**

**Target generation**
Source Comprehension

Horvitz #11 – maintaining working memory of recent interactions
Source Comprehension

Horvitz #6 – allowing efficient direct invocation and termination
Interactions

Source comprehension

Target gisting

Full best translation

Real-time updating – full translation generation

Target generation
À équiper le centre de formation Studeo qui est accessible aux personnes à mobilité réduite et dont nous travaillons à la réalisation dans le cadre de l'institut Jedlička, avec l'association Tap, et ça depuis six ans.

To equip studeo training centre which is accessible to people with reduced mobility and we work to achieve in the framework of the Institute jedlička, with tap, and been there for six years.

Des enseignants se rendent régulièrement auprès des élèves de l'institut Jedličkův et leur proposent des activités qui les intéressent et les amusent.

Teachers regularly visit Jedličkův Institute students and offered them activities of interest to them and having fun.

Les étudiants eux-mêmes n’ont pas les moyens de se rendre à des cours, nous essayons de les aider de cette manière.

The students themselves cannot be required to attend courses, we are trying to help them themselves cannot
themselves could not
themselves do not
themselves cannot afford

Dans le cadre de l'
projet dans un nor
l'institut Jedlička, nous transférerons ce

themselfs cannot afford
Target Gisting

Teachers regularly visit Jedličkův Institute students and offered them activities of interest to them and having fun.

The students themselves cannot be required to attend courses, we are trying to help them.

Horvitz #9 – providing mechanisms for efficient agent-user collaboration to refine results
Interactions

Source comprehension

Target gisting

Target generation

  Real-time autocomplete dropdown

  Target reordering

  Insert complete translation
Target Generation | Autocomplete

**Horvitz #5** – employing dialog to resolve key uncertainties
That would lead to freeing the imprisoned former Prime Minister was revoked reading of the proposal for mitigation of sentences for economic offences.

Premierminister war Revoked während zweite Lesung der für Wirtschaftliche Offences.
Target Generation | Complete Translation

À équiper le centre de formation Studeo qui est accessible aux personnes à mobilité réduite et dont nous travaillons à la réalisation dans le cadre de l’institut Jedlička, avec l’association Tap, et ça depuis six ans.

To equip studeo training centre which is accessible to people with reduced mobility and we work to achieve in the framework of the Institute jedlička, with tap, and been there for six years.

Des enseignants se rendent régulièrement auprès des élèves de l’institut Jedličkův et leur proposent des activités qui les intéressent et les amusent.

Teachers regularly visit Jedličkův Institute students and offered them activities of interest to them and having fun.

Les étudiants eux-mêmes n’ont pas les moyens de se rendre à des cours, nous essayons de les aider de cette manière.

The students themselves cannot be required to attend courses, we are trying to help.

Horvitz #6 – allowing efficient direct invocation and termination
User Study Results
## Experimental Design

<table>
<thead>
<tr>
<th>Task</th>
<th>translate French-English or English-German</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Text</td>
<td>≈3,000 tokens of News/Medical/Software</td>
</tr>
</tbody>
</table>
Experimental Design

Task       translate French-English or English-German
Source Text ≈3,000 tokens of News/Medical/Software
Conditions post-edit (pe) and interactive (imt)
Experimental Design

<table>
<thead>
<tr>
<th>Task</th>
<th>translate French-English or English-German</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Text</td>
<td>≈3,000 tokens of News/Medical/Software</td>
</tr>
<tr>
<td>Conditions</td>
<td>post-edit (pe) and interactive (imt)</td>
</tr>
<tr>
<td>Expert Subjects</td>
<td>16 per language pair</td>
</tr>
</tbody>
</table>

**Measure human productivity** – quality vs. time
Depending on the file, the document into which you want to place the file may need to be opened first. **Scripts listed vary depending on the Creative Suite 4 components you’ve installed.**

— Source

**Expected Wins**

Je nach Datei muss das Dokument, in das diese eingefügt werden soll, erst geöffnet werden. **Welche Skripte aufgelistet werden, hängt davon ab, welche Creative Suite 4-Komponenten installiert sind.**

— Reference

[Best ← Rank 1  Rank 2  Rank 3  Rank 4  Rank 5 → Worst]

Aufgelistete Stichwörter hängen von den von Ihnen für Creative Suite 4 installierten Komponenten ab.

— Translation 1

[Best ← Rank 1  Rank 2  Rank 3  Rank 4  Rank 5 → Worst]

Die aufgeführten Skripte variieren je nach den Creative-Suite-4-Komponenten, die Sie installiert haben.

— Translation 2

[Federmann 2010]
Results | Quality

<table>
<thead>
<tr>
<th>Language Pair</th>
<th>ptm</th>
<th>pe</th>
<th>Expected Wins %</th>
</tr>
</thead>
<tbody>
<tr>
<td>En-De</td>
<td>0.467</td>
<td>0.474</td>
<td></td>
</tr>
<tr>
<td>Fr-En</td>
<td>0.527</td>
<td>0.472</td>
<td></td>
</tr>
</tbody>
</table>
Question #2 | Time

Fr-En  |
| 0  | 20  | 40  | 60  | 80  |
| 54.6| 63.3|

En-De  |
| 0  | 20  | 40  | 60  | 80  |
| 51.8|

Fr-En  |
| 0  | 20  | 40  | 60  | 80  |
| 46.0|

avg. seconds / sentence

- ptm
- pe
Post-edit mode was easier at first, but the interactive mode was better once I got used to it.
**Qualitative Results | Time**

*Post-edit mode was easier at first, but the interactive mode was better once I got used to it.*

*If I had time to use the interactive tool and grow accustomed to its way of functioning, it would be quite useful...*
Post-edit mode was easier at first, but the interactive mode was better once I got used to it.

If I had time to use the interactive tool and grow accustomed to its way of functioning, it would be quite useful...

I am used to this [post-edit], this is how Trados works.
Model Adaptation
A Mixed-Initiative System Should Learn

Horvitz #11  – Maintain working memory of interactions

Horvitz #12  – Continuing to learn by observing
Online Adaptation Learning Setup

**BLEU** for baseline tuning

[Ger et al. 2010]
Online Adaptation Learning Setup

**BLEU** for baseline tuning

**HTER** for adaptation

Interpretable measure of human effort

[Cer et al. 2010]
Online Adaptation Learning Setup

**BLEU** for baseline tuning

[Cer et al. 2010]

**HTER** for adaptation

Interpretable measure of human effort

Online algorithm

SGD-based

Loss – *cross-entropy* for switching metrics

[Green et al. 2013a] (see paper)
Adapting the Representation

Baseline tuning

\[ \phi(f, \hat{e}) \]
Adapting the Representation

Baseline tuning

\[ \phi(f, \hat{e}) \]

Adaptation

\[ h := \text{human translation related to } \hat{e} \]
Adapting the Representation

Baseline tuning

\[ \phi(f, \hat{e}) \]

Adaptation

\[ h := \text{human translation related to } \hat{e} \]

\[ \phi(f, \hat{e}) \leftarrow \phi(f, \hat{e}) \cap \phi(f, h) \]
Adapting the Representation

Baseline tuning

$$\phi(f, \hat{e})$$

**Adaptation**

$$h := \text{human translation related to } \hat{e}$$

$$\phi(f, \hat{e}) \leftarrow \phi(f, \hat{e}) \cap \phi(f, h)$$

**Templates** – alignment cliques, OOV context

Where do we get alignments?
Prefix Decoding

\[ \hat{e} = \arg \max_e w^T \phi(e, f) \]
Prefix Decoding

\[ \hat{e} = \arg\max_e w^T \phi(e, f) \]

Define operator \( \text{pref}(e, h) \)

\[ \hat{e} = \arg\max_e w^T \phi(e, f) \text{ s.t. } \text{pref}(e, h) \]
Prefix Decoding

\[ \hat{e} = \arg \max_e w^T \phi(e, f) \]

Define operator \( \text{pref}(e, h) \)

\[ \hat{e} = \arg \max_{e \text{ s.t. } \text{pref}(e, h)} w^T \phi(e, f) \]

By-product – alignment for feature extraction
<table>
<thead>
<tr>
<th>tarceva</th>
<th>was</th>
<th>thus</th>
<th>âble</th>
<th>to</th>
<th>halt</th>
<th>the</th>
<th>growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>tarceva</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>parvient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ainsi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>à</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stopper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>la</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>croissance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OOVs – Dynamic phrase table augmentation
## Test Results | French-English

<table>
<thead>
<tr>
<th></th>
<th>BLEU↑</th>
<th>HTER↓</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>baseline</strong></td>
<td>bleu</td>
<td>39.3</td>
</tr>
<tr>
<td><strong>adapted</strong></td>
<td>hter</td>
<td>40.3</td>
</tr>
</tbody>
</table>
## Test Results | English-German

<table>
<thead>
<tr>
<th></th>
<th>BLEU↑</th>
<th>HTER↓</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>baseline</strong></td>
<td>bleu</td>
<td>23.1</td>
</tr>
<tr>
<td><strong>adapted</strong></td>
<td>hter</td>
<td>21.7</td>
</tr>
</tbody>
</table>
Analysis | Fine-grained Corrections

depending on the file
abhängig von der datei
Analysis | Fine-grained Corrections

depending on the file
abhängig von der datei

Human corrections

(6) je nach datei
(1) das dokument
(1) abhängig von der datei
Analysis | Fine-grained Corrections

<table>
<thead>
<tr>
<th>depending on the file</th>
</tr>
</thead>
<tbody>
<tr>
<td>abhängig von der datei</td>
</tr>
</tbody>
</table>

Human corrections

(6) je nach datei
(1) das dokument
(1) abhängig von der datei

Adapted model

je nach dokument
Conclusion

**Quality** – IMT better when MT is good (Fr-En)
Conclusion

**Quality** – IMT better when MT is good (Fr-En)

**Time** – Post-edit is faster (for minimal UI training)
Conclusion

**Quality** – IMT better when MT is good (Fr-En)

**Time** – Post-edit is faster (for minimal UI training)

**Adaptation** – IMT better for re-tuning to HTER