Human Effort and Machine Learnability in Computer Aided Translation

Spence Green

Sida I. Wang   Jason Chuang*   Jeffrey Heer*

Sebastian Schuster   Christopher D. Manning
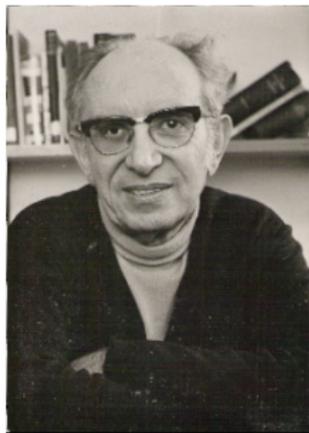
Stanford University   *University of Washington

THE PRESENT STATE OF RESEARCH ON MECHANICAL TRANSLATION

YEHOSHUA BAR-HILLEL*

[Bar-Hillel 1951]

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]

Speed

	Slow	Fast
Lo		Assimilation
Hi	Dissemination →	

Quality

5

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

Insights | Interactive Mode

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

Post-edit Mode

Parliament Does Not Support Amendment Freeing Tymoshenko

Parlament unterstützt keine Änderungsantrag tymoshenko befreien

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 hel

| themselves cannot |
| themselves could not |
| themselves do not |
| themselves cannot afford |

Dans le cadre de l'...........................'institut Jedlička, nous transférerons ce projet dans un no...........................

UI presented at UIST 2014

Interactions

Three classes:

Source comprehension

Target gisting

Target generation

Interactions

Source comprehension
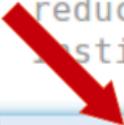
    Word-to-word translations

    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

Target Gisting

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 hel

themselves cannot
themselves could not
themselves do not
themselves cannot afford

Dans le cadre de l'institut Jedlička, nous transférerons ce projet dans un no

Target Gisting

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 hel
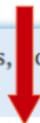
themselves cannot
themselves could not
themselves do not
themselves cannot afford

Dans le cadre de l'... 'institut Jedlička, nous transférerons ce projet dans un no...

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

Target Generation | Target Reordering

t that would lead to freeing the imprisoned former Prime Minister was revoked reading of the proposal for mitigation of sentences for economic offences.

gsantrag das würde dazu führen, dass die Befreiung der
n W Premierminister war Revoked während zweite Lesung der
ür Wirtschaftliche Offences. 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 hel

themselves cannot
themselves could not
themselves do not
themselves cannot afford

Dans le cadre de l'... ...'institut Jedlička, nous transférerons ce projet dans un no...

Horvitz #6 – allowing efficient direct invocation and termination

21

User Study Results

Experimental Design

Task translate French-English or English-German

Source Text ≈3,000 tokens of News/Medical/Software

Experimental Design

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

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)
Expert Subjects	16 per language pair

Measure human productivity – quality vs. time

Question #1 | Quality

Expected Wins

[Federmann 2010]

Results | Quality

Question #2 | Time
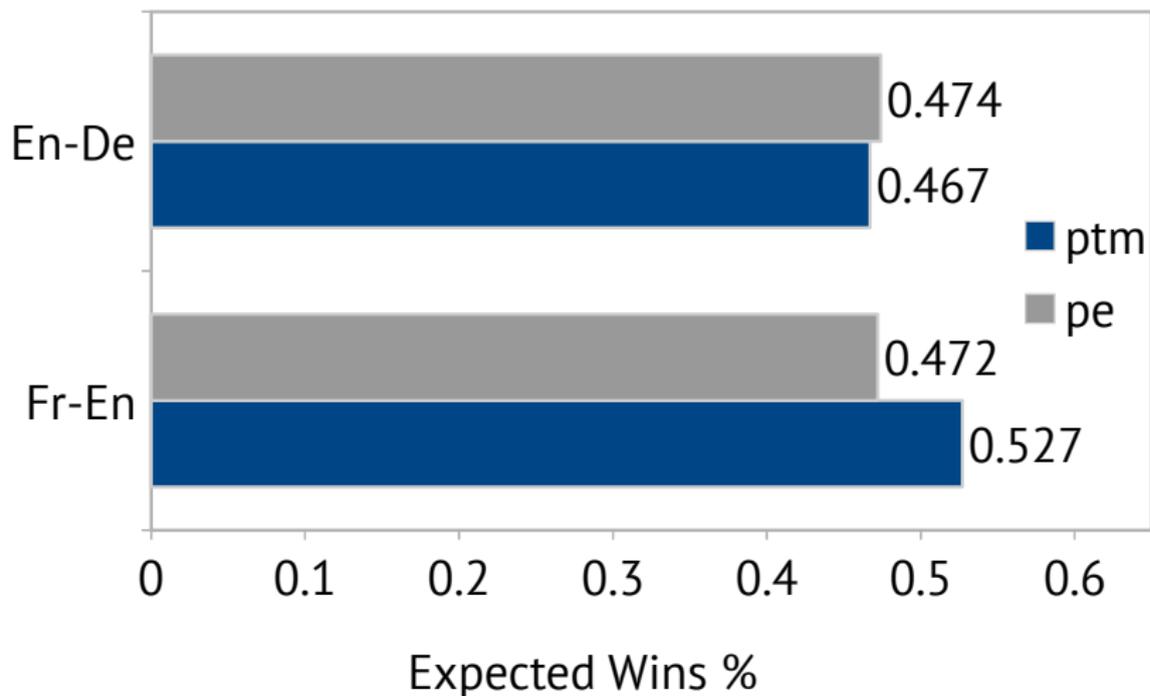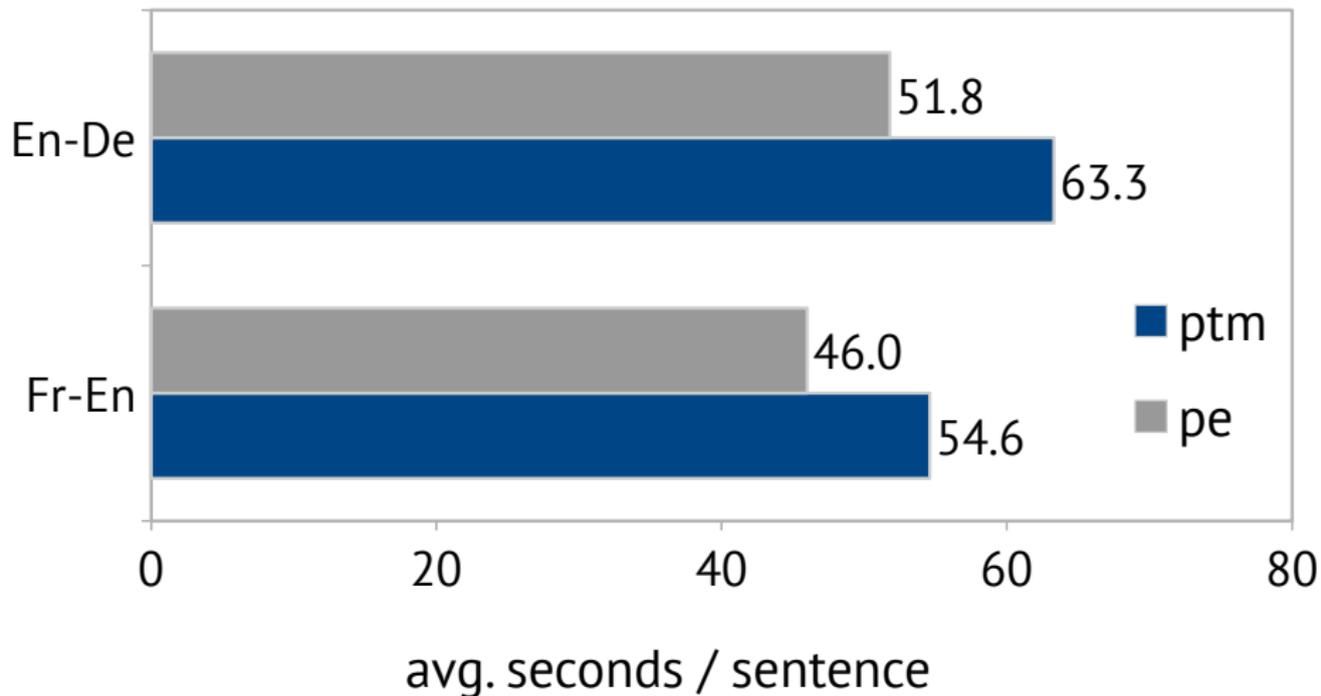
avg. seconds / sentence

Qualitative Results | Time

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…*

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…*
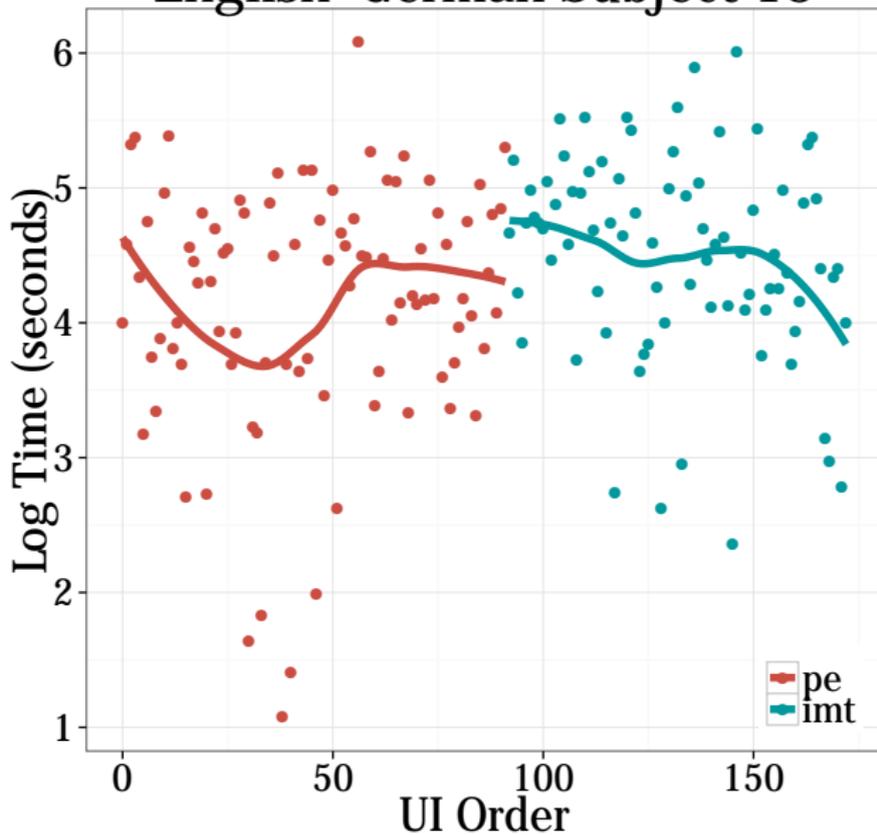
I am used to this [post-edit]*, this is how Trados works.*

English–German Subject 18

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   [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 Adaptation Learning Setup

BLEU for baseline tuning   [Cer et al. 2010]

HTER for adaptation

    Interpretable measure of human effort

Online algorithm

    SGD-based   [Green et al. 2013a]

    Loss – **cross-entropy** for switching metrics   (see paper)

Adapting the Representation

Baseline tuning

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

Adapting the Representation

Baseline tuning

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

Adaptation

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

Adapting the Representation

Baseline tuning

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

Adaptation

$h :=$ 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

$h :=$ 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

$$\hat{e} = \arg\max_{e} w^{\top}\phi(e, f)$$

Prefix Decoding

$$\hat{e} = \arg\max_{e} w^\top \phi(e, f)$$

Define operator $\mathrm{pref}(e, h)$

$$\hat{e} = \arg\max_{e \text{ s.t. } \mathrm{pref}(e,h)} w^\top \phi(e, f)$$

Prefix Decoding

$$\hat{e} = \arg\max_{e} w^{\top}\phi(e, f)$$

Define operator $\operatorname{pref}(e, h)$

$$\hat{e} = \arg\max_{e \text{ s.t. } \operatorname{pref}(e,h)} w^{\top}\phi(e, f)$$

By-product – alignment for feature extraction

OOVs – Dynamic phrase table augmentation

Test Results | French-English

		BLEU↑	HTER↓	
baseline	bleu	39.3	28.3	
adapted	hter	*40.3*	**26.4**	−1.9

Test Results | English-German

		BLEU↑	HTER↓	
baseline	bleu	23.1	44.1	
adapted	hter	21.7	**42.4**	—1.7

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

depending	on	the	file
abhängig	von	der	datei

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

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