Computer Aided Translation

Philipp Koehn

7 September 2012
Why Machine Translation?

**Assimilation** — reader initiates translation, wants to know content

- user is tolerant of inferior quality
- focus of majority of research (GALE program, etc.)

**Communication** — participants don’t speak same language, rely on translation

- users can ask questions, when something is unclear
- chat room translations, hand-held devices
- often combined with speech recognition, IWSLT campaign

**Dissemination** — publisher wants to make content available in other languages

- high demands for quality
- currently almost exclusively done by human translators
Why Machine Translation?

**Assimilation** — reader initiates translation, wants to know content

- user is tolerant of inferior quality
- focus of majority of research (GALE program, etc.)

**Communication** — participants don’t speak same language, rely on translation

- users can ask questions, when something is unclear
- chat room translations, hand-held devices
- often combined with speech recognition, IWSLT campaign

**Dissemination** — publisher wants to make content available in other languages

- high demands for quality
- currently almost exclusively done by human translators
Goal: Helping Human Translators

If you can’t beat them, join them.

- How can machine translation help human translators?
- First question: What do translators do?
Overview

• Human Translation

• Assistance to Human Translators

• User Study

• Assistance to Monolingual Translators

• Integration of Translation Memory and MT
Setup

• 10 students at the University of Edinburgh
  – half native French speakers
  – half native English speakers with advanced French

• Each student translated
  – news stories
  – French-English
  – about 40 sentences
  – easy task: familiar content, no specialized terminology

• Keystroke log
Keystroke Log

Input:  Au premier semestre, l’avionneur a livré 97 avions.
Output: The manufacturer has delivered 97 planes during the first half.

(37.5 sec, 3.4 sec/word)

black: keystroke, purple: deletion, grey: cursor move
height: length of sentence
Analysis

• We can observe
  – slow typing
  – fast typing
  – pauses

• Pauses
  – beginning pause: reading the input sentence
  – final pause: reviewing the translation
  – short pauses (2-6 seconds): hesitation
  – medium pauses (6-60 seconds): problem solving
  – big pauses (>60 seconds): serious problem
## Time Spent on Activities

<table>
<thead>
<tr>
<th>User</th>
<th>total</th>
<th>initial</th>
<th>final</th>
<th>short</th>
<th>medium</th>
<th>big</th>
<th>keystroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1a</td>
<td>3.3s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L1b</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
</tr>
<tr>
<td>L1c</td>
<td>3.9s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.7s</td>
<td>-</td>
<td>2.5s</td>
</tr>
<tr>
<td>L1d</td>
<td>2.8s</td>
<td>0.2s</td>
<td>0.0s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L1e</td>
<td>5.2s</td>
<td>0.3s</td>
<td>0.0s</td>
<td>0.3s</td>
<td>1.9s</td>
<td>0.5s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2a</td>
<td>5.7s</td>
<td>0.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>0.7s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2b</td>
<td>3.2s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2c</td>
<td>5.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>1.5s</td>
<td>0.3s</td>
<td>3.1s</td>
</tr>
<tr>
<td>L2d</td>
<td>3.4s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.6s</td>
<td>-</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2e</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
</tr>
</tbody>
</table>

L1 = native French, L2 = native English  
average time per input word
# Time Spent on Activities

<table>
<thead>
<tr>
<th>User</th>
<th>total</th>
<th>initial</th>
<th>final</th>
<th>short</th>
<th>medium</th>
<th>big</th>
<th>keystroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1a</td>
<td>3.3s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L1b</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
</tr>
<tr>
<td>L1c</td>
<td>3.9s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.7s</td>
<td>-</td>
<td>2.5s</td>
</tr>
<tr>
<td>L1d</td>
<td>2.8s</td>
<td>0.2s</td>
<td>0.0s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L1e</td>
<td>5.2s</td>
<td>0.3s</td>
<td>0.0s</td>
<td>0.3s</td>
<td>1.9s</td>
<td>0.5s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2a</td>
<td>5.7s</td>
<td>0.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>0.7s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2b</td>
<td>3.2s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2c</td>
<td>5.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>1.5s</td>
<td>0.3s</td>
<td>3.1s</td>
</tr>
<tr>
<td>L2d</td>
<td>3.4s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.6s</td>
<td>-</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2e</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
</tr>
</tbody>
</table>

L1 = native French, L2 = native English

average time per input word
**Time Spent on Activities**

<table>
<thead>
<tr>
<th>User</th>
<th>total</th>
<th>initial</th>
<th>final</th>
<th>short</th>
<th>medium</th>
<th>big</th>
<th>keystroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1a</td>
<td>3.3s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L1b</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
</tr>
<tr>
<td>L1c</td>
<td>3.9s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.7s</td>
<td>-</td>
<td>2.5s</td>
</tr>
<tr>
<td>L1d</td>
<td>2.8s</td>
<td>0.2s</td>
<td>0.0s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L1e</td>
<td>5.2s</td>
<td>0.3s</td>
<td>0.0s</td>
<td>0.3s</td>
<td>1.9s</td>
<td>0.5s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2a</td>
<td>5.7s</td>
<td>0.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>0.7s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2b</td>
<td>3.2s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2c</td>
<td>5.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>1.5s</td>
<td>0.3s</td>
<td>3.1s</td>
</tr>
<tr>
<td>L2d</td>
<td>3.4s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.6s</td>
<td>-</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2e</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
</tr>
</tbody>
</table>

L1 = native French, L2 = native English

average time per input word

Philipp Koehn  
Computer Aided Translation  
7 September 2012
### Time Spent on Activities

<table>
<thead>
<tr>
<th>User</th>
<th>total</th>
<th>initial</th>
<th>final</th>
<th>short</th>
<th>medium</th>
<th>big</th>
<th>keystroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1a</td>
<td>3.3s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L1b</td>
<td>7.7s</td>
<td><strong>1.3s</strong></td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
</tr>
<tr>
<td>L1c</td>
<td>3.9s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.7s</td>
<td>-</td>
<td>2.5s</td>
</tr>
<tr>
<td>L1d</td>
<td>2.8s</td>
<td>0.2s</td>
<td>0.0s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.8s</td>
</tr>
<tr>
<td>L1e</td>
<td>5.2s</td>
<td>0.3s</td>
<td>0.0s</td>
<td>0.3s</td>
<td>1.9s</td>
<td>0.5s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2a</td>
<td>5.7s</td>
<td>0.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>0.7s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2b</td>
<td>3.2s</td>
<td>0.1s</td>
<td>0.1s</td>
<td>0.2s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>2.2s</td>
</tr>
<tr>
<td>L2c</td>
<td>5.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>1.5s</td>
<td>0.3s</td>
<td>3.1s</td>
</tr>
<tr>
<td>L2d</td>
<td>3.4s</td>
<td><strong>0.7s</strong></td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.6s</td>
<td>-</td>
<td>1.8s</td>
</tr>
<tr>
<td>L2e</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
</tr>
</tbody>
</table>

L1 = native French, L2 = native English
average time per input word
Pauses Reconsidered

- Our classification of pauses is arbitrary (2-6sec, 6-60sec, >60sec)

- Extreme view: all you see is pauses
  - keystrokes take no observable time
  - all you see is pauses between action points

- Visualizing range of pauses:
  time \( t \) spent in pauses \( p \in P \) up to a certain length \( l \)

\[
sum(t) = \frac{1}{Z} \sum_{p \in P, l(p) \leq t} l(p)
\]
Results
Overview

- Human Translation
- **Assistance to Human Translators**
- User Study
- Assistance to Monolingual Translators
- Integration of Translation Memory and MT
Our Types of Assistance

- **Sentence completion**
  - tool suggests how to complete the translation
  - one phrase at a time

- **Translation options**
  - most likely translations for each word and phrase
  - ordered and color-highlighted by probability

- **Postediting machine translation**
  - start with machine translation output
  - user edits, tool shows changes
Technical Notes

- Online at http://www.caitra.org/

- User uploads source text, translates one sentence at a time

- Implementation
  - AJAX Web 2.0 using Ruby on Rails, mySQL
  - Back end: Moses machine translation system
Predicting Sentence Completion

- Tool makes a suggestion how to continue (in red)
- User can accept it (by pressing TAB), or type in her own translation
- Same idea as TransType, with minor modifications
  - show only short text chunks, not full sentence completion
  - show only one suggestion, not alternatives
How does it work?

• Uses search graph of SMT decoding

• Matches partial user translation against search graph, by optimizing
  1. minimal string edit distance between path in graph and user translation
  2. best full path probability, including best completion to end

• Technical notes
  – search graph is pre-computed and stored in database
  – matching is done server-side, typically takes less than 1 second
  – completion path is returned to client (web browser)
• For each word and phrases: suggested translations

• Ranked (and color-highlighted) by probability

• User may click on suggestion → appended to text box
Translation Options - How does it work?

- Uses phrase translation table of SMT system

- Translation score: future cost estimate
  - conditional probabilities $\phi(e|f), \phi(f|e)$
  - lexical probabilities $\text{lex}(e|f), \text{lex}(f|e)$
  - word count feature
  - language model estimate

- Ranking of shorter vs. longer phrases by including outside future cost estimate
Das erste schwarz-grüne Bündnis auf Landesebene rückt näher: Die Spitzen von CDU und Grünen in Hamburg halten ihre Differenzen für überwindbar.

Leaders of the Hamburger CDU and Greens open path to coalition negotiations. Then the CDU-leader Michael Freytag and Green party leader Anja Hajduk the division between the parties is bridgable.
Postediting Machine Translation

- Textbox is initially filled with machine translation
- User edits translation
- String edit distance to machine translation is shown (blue background)
Overview

- Human Translation
- Assistance to Human Translators
- User Study
- Assistance to Monolingual Translators
- Integration of Translation Memory and MT
Evaluation

• Recall setup
  – 10 students, half native French, half native English
  – each student translated French-English news stories
  – about 40 sentences for each condition of assistance

• Five different conditions
  – unassisted
  – prediction (sentence completion)
  – options
  – predictions and options
  – post-editing
Quality

- We want faster translators, but not worse

- Assessment of translation quality
  - show translations to bilingual judges, with source
  - judgment: fully correct? yes/no

  Indicate whether each user’s input represents a fully fluent and meaning-equivalent translation of the source. The source is shown with context, the actual sentence is bold.

- Average score: 50% correct — lower than expected
  - judges seemed to be too harsh
  - when given several translations, tendency to judge half as bad
## Example of Quality Judgments

<table>
<thead>
<tr>
<th>Src.</th>
<th>Sans se démonter, il s’est montré concis et précis.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT</td>
<td>Without dismantle, it has been concise and accurate.</td>
</tr>
<tr>
<td>1/3</td>
<td>Without fail, he has been concise and accurate. (Prediction+Options, L2a)</td>
</tr>
<tr>
<td>4/0</td>
<td>Without getting flustered, he showed himself to be concise and precise. (Unassisted, L2b)</td>
</tr>
<tr>
<td>4/0</td>
<td>Without falling apart, he has shown himself to be concise and accurate. (Postedit, L2c)</td>
</tr>
<tr>
<td>1/3</td>
<td>Unswayable, he has shown himself to be concise and to the point. (Options, L2d)</td>
</tr>
<tr>
<td>0/4</td>
<td>Without showing off, he showed himself to be concise and precise. (Prediction, L2e)</td>
</tr>
<tr>
<td>1/3</td>
<td>Without dismantling himself, he presented himself consistent and precise. (Prediction+Options, L1a)</td>
</tr>
<tr>
<td>2/2</td>
<td>He showed himself concise and precise. (Unassisted, L1b)</td>
</tr>
<tr>
<td>3/1</td>
<td>Nothing daunted, he has been concise and accurate. (Postedit, L1c)</td>
</tr>
<tr>
<td>3/1</td>
<td>Without losing face, he remained focused and specific. (Options, L1d)</td>
</tr>
<tr>
<td>3/1</td>
<td>Without becoming flustered, he showed himself concise and precise. (Prediction, L1e)</td>
</tr>
</tbody>
</table>
# Faster and Better

<table>
<thead>
<tr>
<th>Assistance</th>
<th>Speed</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>4.4s/word</td>
<td>47% correct</td>
</tr>
<tr>
<td>Postedit</td>
<td>2.7s (-1.7s)</td>
<td>55% (+8%)</td>
</tr>
<tr>
<td>Options</td>
<td>3.7s (-0.7s)</td>
<td>51% (+4%)</td>
</tr>
<tr>
<td>Prediction</td>
<td>3.2s (-1.2s)</td>
<td>54% (+7%)</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>3.3s (-1.1s)</td>
<td>53% (+6%)</td>
</tr>
</tbody>
</table>
### Faster and Better, Mostly

<table>
<thead>
<tr>
<th>User</th>
<th>Unassisted</th>
<th>Postedit</th>
<th>Options</th>
<th>Prediction</th>
<th>Prediction+Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1a</td>
<td>3.3sec/word</td>
<td>1.2s -2.2s</td>
<td>2.3s -1.0s</td>
<td>1.1s -2.2s</td>
<td>2.4s -0.9s</td>
</tr>
<tr>
<td></td>
<td>23% correct</td>
<td>39% +16%</td>
<td>45% +22%</td>
<td>30% +7%</td>
<td>44% +21%</td>
</tr>
<tr>
<td>L1b</td>
<td>7.7sec/word</td>
<td>4.5s -3.2s</td>
<td>4.5s -3.3s</td>
<td>2.7s -5.1s</td>
<td>4.8s -3.0s</td>
</tr>
<tr>
<td></td>
<td>35% correct</td>
<td>48% +13%</td>
<td>55% +20%</td>
<td>61% +26%</td>
<td>41% +6%</td>
</tr>
<tr>
<td>L1c</td>
<td>3.9sec/word</td>
<td>1.9s -2.0s</td>
<td>3.8s -0.1s</td>
<td>3.1s -0.8s</td>
<td>2.5s -1.4s</td>
</tr>
<tr>
<td></td>
<td>50% correct</td>
<td>61% +11%</td>
<td>54% +4%</td>
<td>64% +14%</td>
<td>61% +11%</td>
</tr>
<tr>
<td>L1d</td>
<td>2.8sec/word</td>
<td>2.0s -0.7s</td>
<td>2.9s (+0.1s)</td>
<td>2.4s (-0.4s)</td>
<td>1.8s -1.0s</td>
</tr>
<tr>
<td></td>
<td>38% correct</td>
<td>46% +8%</td>
<td>59% (+21%)</td>
<td>37% (-1%)</td>
<td>45% +7%</td>
</tr>
<tr>
<td>L1e</td>
<td>5.2sec/word</td>
<td>3.9s -1.3s</td>
<td>4.9s (-0.2s)</td>
<td>3.5s -1.7s</td>
<td>4.6s (-0.5s)</td>
</tr>
<tr>
<td></td>
<td>58% correct</td>
<td>64% +6%</td>
<td>56% (-2%)</td>
<td>62% +4%</td>
<td>56% (-2%)</td>
</tr>
<tr>
<td>L2a</td>
<td>5.7sec/word</td>
<td>1.8s -3.9s</td>
<td>2.5s -3.2s</td>
<td>2.7s -3.0s</td>
<td>2.8s -2.9s</td>
</tr>
<tr>
<td></td>
<td>16% correct</td>
<td>50% +34%</td>
<td>34% +18%</td>
<td>40% +24%</td>
<td>50% +34%</td>
</tr>
<tr>
<td>L2b</td>
<td>3.2sec/word</td>
<td>2.8s (-0.4s)</td>
<td>3.5s +0.3s</td>
<td>6.0s +2.8s</td>
<td>4.6s +1.4s</td>
</tr>
<tr>
<td></td>
<td>64% correct</td>
<td>56% (-8%)</td>
<td>60% -4%</td>
<td>61% -3%</td>
<td>57% -7%</td>
</tr>
<tr>
<td>L2c</td>
<td>5.8sec/word</td>
<td>2.9s -3.0s</td>
<td>4.6s (-1.2s)</td>
<td>4.1s -1.7s</td>
<td>2.7s -3.1s</td>
</tr>
<tr>
<td></td>
<td>52% correct</td>
<td>53% +1%</td>
<td>37% (-15%)</td>
<td>59% +7%</td>
<td>53% +1%</td>
</tr>
<tr>
<td>L2d</td>
<td>3.4sec/word</td>
<td>3.1s (-0.3s)</td>
<td>4.3s (+0.9s)</td>
<td>3.8s (+0.4s)</td>
<td>3.7s (+0.3s)</td>
</tr>
<tr>
<td></td>
<td>49% correct</td>
<td>49% (-0%)</td>
<td>51% (+2%)</td>
<td>53% (-4%)</td>
<td>58% (+9%)</td>
</tr>
<tr>
<td>L2e</td>
<td>2.8sec/word</td>
<td>2.6s -0.2s</td>
<td>3.5s +0.7s</td>
<td>2.8s (-0.0s)</td>
<td>3.0s +0.2s</td>
</tr>
<tr>
<td></td>
<td>68% correct</td>
<td>79% +11%</td>
<td>59% -9%</td>
<td>64% (-4%)</td>
<td>66% -2%</td>
</tr>
<tr>
<td>avg.</td>
<td>4.4sec/word</td>
<td>2.7s -1.7s</td>
<td>3.7s -0.7s</td>
<td>3.2s -1.2s</td>
<td>3.3s -1.1s</td>
</tr>
<tr>
<td></td>
<td>47% correct</td>
<td>55% +8%</td>
<td>51% +4%</td>
<td>54% +7%</td>
<td>53% +6%</td>
</tr>
</tbody>
</table>
• Unassisted
  – more than 5 seconds per input word
  – very bad (35%, 16%)

• With assistance
  – much faster and better
  – reaching roughly average performance
Slow Users 2: Only Faster

- Unassisted
  - more than 5 seconds per input word
  - average quality

- With assistance
  - faster and but not better
Fast Users

- Unassisted
  - fast: 3-4 seconds per input word
  - L1a is very bad (23%), L1c is average (50%)

- With assistance
  - faster and better
  - L1a closer to average (30-45%), L1c becomes very good (54-61%)
• Use the assistance sparingly or not at all, and see generally no gains

• The two best translators are in this group

• Postediting
  – mixed on quality (2 better, 1 worse, 1 same), but all faster
  – best translator (L2e, 68%) becomes much better (record 79%)
Further Analysis

- How does the assistance change translator behaviour?
- How do translators utilize assistance?
- How is the translation produced?
Analysis: Segment into periods of activity: typing, tabbing, clicking, pauses

one second before and after a keystroke is part of typing interval
# Activities: Native French User L1b

<table>
<thead>
<tr>
<th>User: L1b</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>4.5s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.0s</td>
<td>0.4s</td>
<td>1.1s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>4.5s</td>
<td>0.6s</td>
<td>0.1s</td>
<td>0.4s</td>
<td>0.9s</td>
<td>0.7s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.7s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.6s</td>
<td>-</td>
<td>0.4s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>4.8s</td>
<td>0.6s</td>
<td>0.4s</td>
<td>0.4s</td>
<td>1.3s</td>
<td>0.5s</td>
<td>0.9s</td>
<td>0.5s</td>
<td>0.2s</td>
</tr>
</tbody>
</table>
### Activities: Native French User L1b

<table>
<thead>
<tr>
<th>User: L1b</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>4.5s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.0s</td>
<td>0.4s</td>
<td>1.1s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>4.5s</td>
<td>0.6s</td>
<td>0.1s</td>
<td>0.4s</td>
<td>0.9s</td>
<td>0.7s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.7s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.6s</td>
<td>-</td>
<td>0.4s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>4.8s</td>
<td>0.6s</td>
<td>0.4s</td>
<td>0.4s</td>
<td>1.3s</td>
<td>0.5s</td>
<td>0.9s</td>
<td>0.5s</td>
<td>0.2s</td>
</tr>
</tbody>
</table>

Slightly less time spent on typing
## Activities: Native French User L1b

<table>
<thead>
<tr>
<th>User: L1b</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>4.5s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.0s</td>
<td>0.4s</td>
<td>1.1s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>4.5s</td>
<td>0.6s</td>
<td>0.1s</td>
<td>0.4s</td>
<td>0.9s</td>
<td>0.7s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.7s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.6s</td>
<td>-</td>
<td>0.4s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>4.8s</td>
<td>0.6s</td>
<td>0.4s</td>
<td>0.4s</td>
<td>1.3s</td>
<td>0.5s</td>
<td>0.9s</td>
<td>0.5s</td>
<td>0.2s</td>
</tr>
</tbody>
</table>

- Less pausing
- Slightly less time spent on typing
### Activities: Native French User L1b

<table>
<thead>
<tr>
<th>User: L1b</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>7.7s</td>
<td>1.3s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>1.8s</td>
<td>1.9s</td>
<td>2.3s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>4.5s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>0.1s</td>
<td>1.0s</td>
<td>0.4s</td>
<td>1.1s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>4.5s</td>
<td>0.6s</td>
<td>0.1s</td>
<td>0.4s</td>
<td>0.9s</td>
<td>0.7s</td>
<td>1.5s</td>
<td>0.4s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.7s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.7s</td>
<td>0.1s</td>
<td>0.6s</td>
<td>-</td>
<td>0.4s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>4.8s</td>
<td>0.6s</td>
<td>0.4s</td>
<td>0.4s</td>
<td>1.3s</td>
<td>0.5s</td>
<td>0.9s</td>
<td>0.5s</td>
<td>0.2s</td>
</tr>
</tbody>
</table>

- Less pausing
- Especially less time in big pauses
- Slightly less time spent on typing
# Activities: Native English User L2e

<table>
<thead>
<tr>
<th>User: L2e</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>2.6s</td>
<td>0.4s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>0.7s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>3.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.4s</td>
<td>0.6s</td>
<td>0.2s</td>
<td>1.7s</td>
<td>0.1s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.8s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>-</td>
<td>1.4s</td>
<td>-</td>
<td>0.3s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>3.0s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>-</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
# Activities: Native English User L2e

<table>
<thead>
<tr>
<th>User: L2e</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>2.6s</td>
<td>0.4s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>0.7s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>3.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.4s</td>
<td>0.6s</td>
<td>0.2s</td>
<td>1.7s</td>
<td>0.1s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.8s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>-</td>
<td>1.4s</td>
<td>-</td>
<td>0.3s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>3.0s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>-</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Little time spent on assistance
Does not use both assistances, little overall change

Little time spent on assistance
Activities: Native English User L2e

<table>
<thead>
<tr>
<th>User: L2e</th>
<th>total</th>
<th>init-p</th>
<th>end-p</th>
<th>short-p</th>
<th>mid-p</th>
<th>big-p</th>
<th>key</th>
<th>click</th>
<th>tab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unassisted</td>
<td>2.8s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.2s</td>
<td>0.3s</td>
<td>0.1s</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Postedit</td>
<td>2.6s</td>
<td>0.4s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>1.0s</td>
<td>0.1s</td>
<td>0.7s</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Options</td>
<td>3.5s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.4s</td>
<td>0.6s</td>
<td>0.2s</td>
<td>1.7s</td>
<td>0.1s</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>2.8s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>0.3s</td>
<td>-</td>
<td>1.4s</td>
<td>-</td>
<td>0.3s</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>3.0s</td>
<td>0.1s</td>
<td>0.3s</td>
<td>0.2s</td>
<td>0.5s</td>
<td>-</td>
<td>1.9s</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Does not use both assistances, little overall change

Postediting: less typing (-1.2s) more medium pauses (+0.7s)

Little time spent on assistance
# Origin of Characters: Native French L1b

<table>
<thead>
<tr>
<th>User: L1b</th>
<th>key</th>
<th>click</th>
<th>tab</th>
<th>mt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postedit</td>
<td>18%</td>
<td>-</td>
<td>-</td>
<td>81%</td>
</tr>
<tr>
<td>Options</td>
<td>59%</td>
<td>40%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>14%</td>
<td>-</td>
<td>85%</td>
<td>-</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>21%</td>
<td>44%</td>
<td>33%</td>
<td>-</td>
</tr>
</tbody>
</table>
Origin of Characters: Native French L1b

<table>
<thead>
<tr>
<th>User: L1b</th>
<th>key</th>
<th>click</th>
<th>tab</th>
<th>mt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postedit</td>
<td>18%</td>
<td>-</td>
<td>-</td>
<td>81%</td>
</tr>
<tr>
<td>Options</td>
<td>59%</td>
<td>40%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>14%</td>
<td>-</td>
<td>85%</td>
<td>-</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>21%</td>
<td>44%</td>
<td>33%</td>
<td>-</td>
</tr>
</tbody>
</table>

Translation comes to large degree from assistance
## Origin of Characters: Native English L2e

<table>
<thead>
<tr>
<th>User: L2e</th>
<th>key</th>
<th>click</th>
<th>tab</th>
<th>mt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postedit</td>
<td>20%</td>
<td>-</td>
<td>-</td>
<td>79%</td>
</tr>
<tr>
<td>Options</td>
<td>77%</td>
<td>22%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>61%</td>
<td>-</td>
<td>38%</td>
<td>-</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>100%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>User: L2e</td>
<td>key</td>
<td>click</td>
<td>tab</td>
<td>mt</td>
</tr>
<tr>
<td>-----------</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>Postedit</td>
<td>20%</td>
<td>-</td>
<td>-</td>
<td>79%</td>
</tr>
<tr>
<td>Options</td>
<td>77%</td>
<td>22%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prediction</td>
<td>61%</td>
<td>-</td>
<td>38%</td>
<td>-</td>
</tr>
<tr>
<td>Prediction+Options</td>
<td>100%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Although hardly any time spent on assistance, fair amount of characters produced by it.
Pauses: French-Native User L1b

The diagram shows the average translation time per word (y-axis) over the length of pauses (x-axis) for different conditions:
- Unassisted
- Postedit
- Options
- Prediction
- Prediction+Options

The graph indicates that the average translation time increases as the length of pauses increases, with the Prediction+Options condition having the highest average time overall.
Pauses: English-Native User L2e

The diagram shows the average translations time (sec/word) against the length of pauses (sec) for different conditions: Unassisted, Postedit, Options, Prediction, and Prediction+Options. The graph illustrates how the time increases with the length of pauses for each condition.
Learning Curve

users become better over time with assistance

![Graph showing the Learning Curve with different lines representing Unassisted, Postedit, Options, Prediction, and Prediction+Options. The x-axis represents the sentence number, and the y-axis represents the average time per word (sec/word).]
User Feedback

• Q: In which of the five conditions did you think you were most accurate?
  – predictions+options: 5 users
  – options: 2 users
  – prediction: 1 user
  – postediting: 1 user

• Q: Rank the different types of assistance on a scale from 1 to 5, where 1 indicates not at all and 5 indicates very helpful.
  – predictions+options: 4.6
  – prediction: 3.9
  – options: 3.7
  – postediting: 2.9
User Feedback

• Q: In which of the five conditions did you think you were most accurate?
  – predictions+options: 5 users
  – options: 2 users
  – prediction: 1 user
  – postediting: 1 user

• Q: Rank the different types of assistance on a scale from 1 to 5, where 1 indicates not at all and 5 indicates very helpful.
  – prediction+options: 4.6
  – prediction: 3.9
  – options: 3.7
  – postediting: 2.9

• Note: does not match empirical results
Summary

• Assistance made translators faster
  – average speed improvement from 4.4s/word to 2.7-3.7s/word
  – reduction of big pauses
  – reduction of typing effort in post-editing

• Assistance made translators better
  – average judgment increased from 47% to 51-55% with help
  – even good translators get better with postediting

• Some good translators ignored the assistance

• Fastest and (barely) best with postediting, but did not like it
Outlook: More analysis

• What do translators think about when they are pausing?

• What are the hard problems?
  – unknown words
  – words without direct translation
  – syntactic re-arrangement

• What do translators change in post-editing?

⇒ We will investigate this in a new EU project
Related Work: Tools used by Translators

- Translators often use standard text editors and additional tools
- Bilingual dictionary
- Spell checker, grammar checker
- Monolingual concordancer
- Terminology database
- Web search to establish and verify meaning of terms
## Bilingual Concordancer

<table>
<thead>
<tr>
<th>Examples</th>
<th>Windkraft (noun, feminine) (also: Windenergie)</th>
<th>wind power (noun)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zum Vergleich: Windkraft schafft fast sieben Mal mehr.</td>
<td>By way of comparison, wind power generates almost seven times as much.</td>
<td></td>
</tr>
<tr>
<td>Windkraft ist eine etablierte, wettbewerbsfähige Technologie mit hoher Zuverlässigkeit</td>
<td>Wind power is an established, competitive technology with high reliability</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.powergeneration.siemens.de/abo...nservices/">German: www.powergeneration.siemens.de/abo...nservices/</a></td>
<td><a href="http://www.powergeneration.siemens.com/abo...nservices/">English: www.powergeneration.siemens.com/abo...nservices/</a></td>
<td></td>
</tr>
<tr>
<td>Windkraft (noun, feminine) (also: Windenergie)</td>
<td>wind energy (noun)</td>
<td></td>
</tr>
<tr>
<td>Je mehr aber klimapolitische Sonntagsreden von der Politik auch in Taten umgesetzt werden, desto höher steigt dieser Preis und desto wettbewerbsfähiger werden saubere Energien wie die Windkraft.</td>
<td>But as the focus of the climate change issue shifts increasingly from policy to action, this price will increase and cleaner energy sources like wind will become more competitive.</td>
<td></td>
</tr>
<tr>
<td>[German: emagazine.credit-suisse.com/app/art...4382 (=DE](<a href="http://emagazine.credit-suisse.com/app/art...4382">http://emagazine.credit-suisse.com/app/art...4382</a> (=DE)</td>
<td>[English: emagazine.credit-suisse.com/app/art...4382 (=en](<a href="http://emagazine.credit-suisse.com/app/art...4382">http://emagazine.credit-suisse.com/app/art...4382</a> (=en)</td>
<td></td>
</tr>
<tr>
<td>Nur wenige befürchten hingegen, dass dies aber bei erneuerbaren Energieträgern wie Biomasse oder Windkraft der Fall sein wird.</td>
<td>However, only a few fear that this will also be the case with renewable energy sources such as biomass or wind energy.</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.eu2006.gv.at/de/News/Press_Rele...1proell.html">German: www.eu2006.gv.at/de/News/Press_Rele...1proell.html</a></td>
<td><a href="http://www.eu2006.gv.at/en/News/Press_Rele...1proell.html">English: www.eu2006.gv.at/en/News/Press_Rele...1proell.html</a></td>
<td></td>
</tr>
</tbody>
</table>

Show translations in context ([www.linguee.com](http://www.linguee.com))
Overview

- Human Translation
- Assistance to Human Translators
- User Study
- **Assistance to Monolingual Translators**
- Integration of Translation Memory and MT
Enabling Monolingual Translators

- Monolingual translator
  - wants to understand a foreign document
  - has no knowledge of foreign language
  - uses a machine translation system

- Questions
  - Is current MT output sufficient for understanding?
  - What else could be provided by a MT system?
Good Enough

- An MT system might produce this:
  The girl entered into room.

- We know what is meant:
  The girl entered the room.

- We understood.
Process

• MT system translates foreign story

• Person A edits it
  – goal: fluent translation
    that represents the meaning (as it was understood)
  – without access to reference translation

• Person B checks if edited sentences are correct
  – with access to reference translation
Example

- MT system translates foreign sentence
  The girl goes the room.

- Person A edits it
  The girl goes into the room.

- Reference
  The girl enters the room.

- Person B checks edited sentence: CORRECT
Real Example

• MT system output:

The study also found that one of the genes in the improvement in people with prostate cancer risk, it also reduces the risk of suffering from diabetes.

• What does this mean?

• Monolingual translator:

The research also found that one of the genes increased people’s risk of prostate cancer, but at the same time lowered people’s risk of diabetes.

• Document context helps
Experiment

- Language pairs
  - Arabic–English
  - Chinese–English

- Machine translation systems
  - Edinburgh’s 2009 GALE systems
  - Moses system with all available parallel data

- Stories taken from NIST 2008 test sets
<table>
<thead>
<tr>
<th>Story</th>
<th>Headline</th>
<th>Sent.</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: chi</td>
<td>White House Pushes for Nuclear Inspectors to Be Sent as Soon as Possible to Monitor North Korea’s Closure of Its Nuclear Reactors</td>
<td>6</td>
<td>207</td>
</tr>
<tr>
<td>2: chi</td>
<td>Torrential Rains Hit Western India, 43 People Dead</td>
<td>10</td>
<td>204</td>
</tr>
<tr>
<td>3: chi</td>
<td>Research Shows a Link between Arrhythmia and Two Forms of Genetic Variation</td>
<td>7</td>
<td>247</td>
</tr>
<tr>
<td>4: chi</td>
<td>Veteran US Goalkeeper Keller May Retire after America’s Cup</td>
<td>10</td>
<td>367</td>
</tr>
<tr>
<td>5: ara</td>
<td>Britain: Arrests in Several Cities and Explosion of Suspicious Car</td>
<td>7</td>
<td>224</td>
</tr>
<tr>
<td>6: ara</td>
<td>Ban Ki-Moon Withdraws His Report on the Sahara after Controversy Surrounding Its Content</td>
<td>8</td>
<td>310</td>
</tr>
<tr>
<td>7: ara</td>
<td>Pakistani Opposition Leaders Call on Musharraf to Resign.</td>
<td>11</td>
<td>312</td>
</tr>
<tr>
<td>8: ara</td>
<td>Al-Maliki: Iraqi Forces Are Capable of Taking Over the Security Dossier Any Time They Want</td>
<td>8</td>
<td>255</td>
</tr>
</tbody>
</table>
Experiment

- Monolingual translators
  - 10 students/staff at the University of Edinburgh
  - none knew Arabic or Chinese
  - have access to full stories at a time, may correct prior sentences

- Bilingual translators
  - 3 of the 4 reference translations in NIST test set

- Remaining reference translation as truth
Results: Arabic

percentage of sentences judged as correct
Results: Arabic

compared to bilingual translators
Results: Arabic

best monolinguals as good as worst bilingual
Results: Arabic and Chinese

mostly worse performance for Chinese
performance differs widely per story
Results per Story

one story: monolinguals as good as bilinguals

Chinese Politics
Chinese Weather
Chinese Science
Chinese Sports
Arabic Terror
Arabic Diplomacy
Arabic Politics

Bilingual
Mono Post-Edit
Offering more assistance

- Progress in computer aided translation

- Interactive machine translation (TransType)
  - show prediction of sentence completion
  - recompute when user types own translation

- Alternative translations [Koehn and Haddow, 2009]
  - display translation options from translation model
  - ranked by translation score
### Translation Options

<table>
<thead>
<tr>
<th>the</th>
<th>the us house of representatives</th>
<th>adopted</th>
<th>thursday</th>
<th>legally</th>
<th>calls for the withdrawal of</th>
<th>combat troops</th>
<th>us</th>
<th>iraq</th>
<th>no later than the first from april</th>
</tr>
</thead>
<tbody>
<tr>
<td>the us house of representatives</td>
<td>the</td>
<td>thursday</td>
<td>law</td>
<td>the</td>
<td>the</td>
<td>fighting forces</td>
<td>the</td>
<td>us</td>
<td>from</td>
</tr>
<tr>
<td>it was</td>
<td>us house of representatives</td>
<td>was adopted</td>
<td>thursday</td>
<td>the</td>
<td>the</td>
<td>law</td>
<td>demands withdrawal of</td>
<td>troops</td>
<td>fighter</td>
</tr>
<tr>
<td>he was</td>
<td>the us house</td>
<td>adopted by</td>
<td>thursday’s</td>
<td>a</td>
<td>a</td>
<td>law</td>
<td>calls for withdrawal of</td>
<td>combat forces</td>
<td>forces</td>
</tr>
<tr>
<td>earlier, the</td>
<td>us house</td>
<td>adopted the</td>
<td>thursday, the</td>
<td>legally</td>
<td>calls for withdrawal</td>
<td>forces</td>
<td>the</td>
<td>us</td>
<td>from</td>
</tr>
<tr>
<td>was</td>
<td>us</td>
<td>adopted a</td>
<td>on</td>
<td>the</td>
<td>legally</td>
<td>demands withdrawal of</td>
<td>troops</td>
<td>from</td>
<td>iraq</td>
</tr>
<tr>
<td>it was</td>
<td>the</td>
<td>adopted</td>
<td>the</td>
<td>the</td>
<td>legal</td>
<td>calls for withdrawal</td>
<td>of</td>
<td>the</td>
<td>from</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2008</th>
<th>defying</th>
<th>once</th>
<th>new president george w. bush, which opposes</th>
<th>no date</th>
<th>has been set for the</th>
<th>president</th>
<th>george bush</th>
<th>who opposes</th>
<th>no date</th>
<th>has been set for</th>
</tr>
</thead>
<tbody>
<tr>
<td>the 2008</td>
<td>defiant</td>
<td>once again</td>
<td>the new</td>
<td>president</td>
<td>george bush</td>
<td>who</td>
<td>opposes</td>
<td>no date</td>
<td>has been set for</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>challenging</td>
<td>again</td>
<td>the new</td>
<td>president</td>
<td>george bush</td>
<td>who</td>
<td>opposes</td>
<td>no date</td>
<td>has been set for</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 2008 | defying the | again | us president george w. bush | opposed to | setting | any | the | date | of | any |
|------|-------------|------|-----------------------------|-----------|----------|--------|------|------|------|------|------|
| the | defying | once again the | us president | george w. bush | opposed to | setting | any | the | date | of | any |
| | | | | | | | | | | | | |

**up to 10 translations for each word / phrase**
## Translation Options

<table>
<thead>
<tr>
<th>بالسحب</th>
<th>القوات</th>
<th>المقاتلة</th>
<th>الاميركية</th>
<th>العراق</th>
</tr>
</thead>
<tbody>
<tr>
<td>withdrawal of</td>
<td>combat troops</td>
<td>the US</td>
<td>from Iraq</td>
<td></td>
</tr>
<tr>
<td>the fighting forces</td>
<td>the US</td>
<td>from Iraq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>withdrawal of troops</td>
<td>fighter</td>
<td>the US</td>
<td>of Iraq</td>
<td></td>
</tr>
<tr>
<td>withdrawal of forces</td>
<td>combat forces</td>
<td>the fighter</td>
<td>from Iraq</td>
<td></td>
</tr>
<tr>
<td>withdrawal of troops</td>
<td>from Iraq</td>
<td>of the</td>
<td>from Iraq</td>
<td></td>
</tr>
<tr>
<td>withdrawal</td>
<td></td>
<td></td>
<td>the American</td>
<td></td>
</tr>
</tbody>
</table>
no big difference — once significantly better
Error Analysis
(a) Critical Judges

- Reference
  Torrential Rains Hit Western India, 43 People Dead

- Bilingual translator
  Heavy Rains Plague Western India Leaving 43 Dead
Error Analysis

(b) Mistakes by the professional translators

• Reference

   Over just two days on the 29th and 30th, rainfall in Mumbai reached 243 mm.

• Bilingual translator

   The rainfall in Mumbai had reached 243 cm over the two days of the 29th and 30th alone.
Error Analysis

(b2) Domain knowledge vs. language skills

- Bilingual translator
  
  With Munchen-Gladbach falling to the German Bundesliga 2, ...

- Monolingual translator

  The Mönchengladbach team fell into the second German league, ...
Error Analysis

(c) Bad English by monolingual translators

- Monolingual translator

  The western region of India heavy rain killed 43 people.
Error Analysis

(d) Mistranslated / untranslated name

- Reference

  Johndroe said that the two leaders ...

- Machine translation

  Strong zhuo, pointing out that the two presidents ...

- Monolingual translator

  Qiang Zhuo pointed out that the two presidents ...
Error Analysis
(e) Wrong relationship between entities

- Machine translation
  The colombian team for the match, and it is very likely that the united
  states and kai in the americas cup final performance.

- Monolingual translator 6
  The Colombian team and the United States are very likely to end up
  in the Americas Cup as the final performance.

- Monolingual translator 8
  The next match against Colombia is likely to be the United States’
  and Keller’s final performance in the current Copa America.
Error Analysis

(f) Badly muddled machine translation

• Reference

In the current America’s cup, he has, just as before, been given an important job to do by head coach Bradley, but he clearly cannot win the match singlehanded. The US team, made up of ”young guards,”...

• Machine translation

He is still being head coach bradley appointed to important, it’s even a fist ”, four young guards at the beginning of the ”, the united states is...
Conclusions

- **Main findings**
  - monolingual translators may be as good as bilinguals
  - widely different performance by translator / story
  - named entity translation critically important

- **Various human factors important**
  - domain knowledge
  - language skills
  - effort
Overview

- Human Translation
- Assistance to Human Translators
- User Study
- Assistance to Monolingual Translators
- Integration of Translation Memory and MT
Progress in Translation Automation

- **Translation Memory (TM)**
  - translators store past translation in database
  - when translating new text, consult database for similar segments
  - fuzzy match score defines similarity
    - widely used by translation agencies

- **Statistical Machine Translation (SMT)**
  - collect large quantities of translated text
  - extract automatically probabilistic translation rules
  - when translating new text, find most probable translation given rules
    - wide use of free web-based services
    - not yet used by many translation agencies
**TM vs. SMT**

- **TM**
  - Used by human translator
  - Restricted domain (e.g., product manual)
  - Very repetitive content
  - Corpus size: 1 million words
  - Commercial developers (e.g., SDL Trados)

- **SMT**
  - Used by target language information seeker
  - Open domain translation (e.g., news)
  - Huge diversity (esp. web)
  - Corpus size: 100-1000 million words
  - Academic/commercial research (e.g., Google)
Our Goal

Better TM

using SMT methods
Main Idea

- Input

  The second paragraph of Article 21 is deleted.

- Fuzzy match in translation memory

  The second paragraph of Article 5 is deleted.

⇒ Part of the translation from TM fuzzy match

Part of the translation with SMT

  The second paragraph of Article 21 is deleted.
Related Work

- Work inspired by EBMT
  
  [Smith and Clark, 2009]
  [Zhechev and van Genabith, 2010]
  - uses syntactic information in alignment
  - lower performance than reported here

- Encode fuzzy match as rule with gaps
  
  [Biçici and Dymetman, 2008]
  - similar to our second method
  - impressive improvements, but weak baseline SMT
Two Solutions

• XML frames

• Very large hierarchical rules
Example

• Input sentence:

The second paragraph of Article 21 is deleted.
Example

• Input sentence:

   The second paragraph of Article 21 is deleted .

• Fuzzy match in translation memory:

   The second paragraph of Article 5 is deleted .
   
   =

   À l’ article 5 , le texte du deuxième alinéa est supprimé .
Example

- Input sentence:

  The second paragraph of Article 21 is deleted.

- Fuzzy match in translation memory:

  The second paragraph of Article 5 is deleted.

  =

  À l’article 5, le texte du deuxième alinéa est supprimé.

- Detect mismatch (string edit distance)
Example

• Input sentence:

The second paragraph of Article 21 is deleted.

• Fuzzy match in translation memory:

The second paragraph of Article 5 is deleted.

= 
À l’article 5, le texte du deuxième alinéa est supprimé.

• Detect mismatch (string edit distance)

• Align mismatch (using word alignment from GIZA++)
Example

• Input sentence:
  The second paragraph of Article 21 is deleted.

• Fuzzy match in translation memory:
  The second paragraph of Article 5 is deleted.
  $=\quad$ À l’article 5, le texte du deuxiéme alinéa est supprimé.

Output word(s) taken from the target TM
Example

- Input sentence:

  The second paragraph of Article 21 is deleted.

- Fuzzy match in translation memory:

  The second paragraph of Article 5 is deleted.

  =

  À l’article 5, le texte du deuxième alinéa est supprimé.

Output word(s) taken from the target TM

Input word(s) that still need to be translated by SMT
Example

• Input sentence:

   The second paragraph of Article 21 is deleted.

• Fuzzy match in translation memory:

   The second paragraph of Article 5 is deleted.
   =

   À l’article 5, le texte du deuxième alinéa est supprimé.

• XML frame (input to Moses)

   <xml translation="À l’article "> 21
   <xml translation=", le texte du deuxième alinéa est supprimé.">
Input sentence:

The second paragraph of Article 21 is deleted.

Fuzzy match in translation memory:

The second paragraph of Article 5 is deleted.

More compact formalism for the purposes of this presentation:

< À l’article 5, le texte du deuxième alinéa est supprimé. >
Steps

• Fuzzy matching
  – based on string edit distance on words

  \[
  FMS = 1 - \frac{\text{edit-distance}(\text{source, tm-source})}{\max(|\text{source}|, |\text{tm-source}|)}
  \]

  – string edit distance on letters as tie breaker
  – details see [Koehn and Senellart, AMTA 2010]

• Word alignment of TM source and target
  standard method

• Construction of XML frame
  – linking mismatch( input, TM source ) to TM target
  can be tricky
Construction of XML Frame

• Basic principles
  – start with fully specified XML frame
  – all mismatched source words have to be translated by SMT
  – all TM target words aligned to mismatched TM source words are removed
  – if the alignment to the TM target words fails
    → go to the previous TM source word and follow its alignment

• See paper for algorithm
À l'article 5, le texte du deuxième alinéa est supprimé.

The second paragraph of Article 21 is deleted.

The second paragraph of Article 5 is deleted.

À l'article 5, le texte du deuxième alinéa est supprimé.

<À l'article> 21 <, le texte du deuxième alinéa est supprimé.>
Special Case: Insertion

Source
String Edit
TM Source
Word Alignment
TM Target
XML Frame

<les>   big  <poissons>

<les>    big    <poissons>
Special Case: Deletion

```
les   gros   poissons
the   fish
<les>  <poissons>
```

```
les   gros   poissons
the   big   fish
<les>  <poissons>
```
Special Case: Unaligned Mismatch

the green fish

the big fish

les poissons

<les> green <poissons>
Special Case: Disconnected Alignments

- Jim will eat.
- Jim does not eat.
- Jim ne mange pas.

Diagram:

- Joe will eat.
- Joe does not eat.
- Joe ne mange pas.
- <Joe> will <mange>
Experiments

- Baseline 1: Unmodified TM matches
- Baseline 2: SMT system trained on TM data
- Our XML frame method
### Corpora: Size

#### Acquis

<table>
<thead>
<tr>
<th></th>
<th>Corpus</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>segments</td>
<td>1,165,867</td>
<td>4,107</td>
</tr>
<tr>
<td>English words</td>
<td>24,069,452</td>
<td>129,261</td>
</tr>
<tr>
<td>French words</td>
<td>25,533,259</td>
<td>135,224</td>
</tr>
</tbody>
</table>

#### Product

<table>
<thead>
<tr>
<th></th>
<th>Corpus</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>segments</td>
<td>83,461</td>
<td>2,000</td>
</tr>
<tr>
<td>English words</td>
<td>1,038,762</td>
<td>24,643</td>
</tr>
<tr>
<td>French words</td>
<td>1,110,284</td>
<td>26,248</td>
</tr>
</tbody>
</table>
# Corpora: Fuzzy Matches

## Acquis

<table>
<thead>
<tr>
<th>Sentences</th>
<th>Words</th>
<th>W/S</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>1395</td>
<td>14,559</td>
</tr>
<tr>
<td>90-99%</td>
<td>433</td>
<td>12,775</td>
</tr>
<tr>
<td>80-89%</td>
<td>154</td>
<td>5,347</td>
</tr>
<tr>
<td>70-79%</td>
<td>250</td>
<td>6,767</td>
</tr>
</tbody>
</table>

## Product

<table>
<thead>
<tr>
<th>Sentences</th>
<th>Words</th>
<th>W/S</th>
</tr>
</thead>
<tbody>
<tr>
<td>95-99%</td>
<td>230</td>
<td>3,006</td>
</tr>
<tr>
<td>90-94%</td>
<td>225</td>
<td>2,968</td>
</tr>
<tr>
<td>85-89%</td>
<td>177</td>
<td>2,000</td>
</tr>
<tr>
<td>80-84%</td>
<td>185</td>
<td>1,950</td>
</tr>
<tr>
<td>75-79%</td>
<td>152</td>
<td>1,350</td>
</tr>
<tr>
<td>70-74%</td>
<td>98</td>
<td>987</td>
</tr>
</tbody>
</table>
Results: Acquis
Results: Product

[Bar chart showing BLEU scores for different fuzzy match ranges (70-74%, 75-79%, 80-84%, 85-89%, 90-94%, 95-99%) with TM, SMT, and XML methods.]
Recap

• TM provides fuzzy matches

• SMT translates mismatched words

• TM match encoded in XML frame

... but is that not just a very large translation rule?
Background: Hierarchical Phrase Rules

• Given: sentence pair with monotone 1-to-1 alignment

  the big fish = les gros poissons

• Phrase translation rules

  ( the ; les )
  ( the big ; les gros )
  ( the big fish ; les gros poissons )
  ( big ; gros )
  ( big fish ; gros poissons )
  ( fish ; poissons )

• Hierarchical phrase-based rule are constructed by subtraction

  – large rule: ( the big fish ; les gros poissons )
  – small rule: ( big ; gros ) (contained in large rule)

  → hierarchical rule: ( the x fish ; les x poissons )
XML Frame as Very Large Rule

- XML frame

\[<\text{À l’ article} \ 21 <, \text{ le texte du deuxième alinéa est supprimé .} >\]

for input

The second paragraph of Article 21 is deleted .

- Very large rule

( The second paragraph of Article x is deleted . \\
; À l’ article x , le texte du deuxième alinéa est supprimé . )
Very Large Rules in SMT

- Rule size limited in SMT
  - maximum number of words, e.g. 5
  - maximum number of non-terminals ($x$), e.g. 2

- ... but only due to storage limitations for large rule rule tables

- Rules may be generated on the fly [Lopez, 2007]
Advantage over XML Method

• Choices
  1. multiple fuzzy matches in TM with same score
  2. same TM source with multiple translations
  3. multiple SMT translations

• Decisions in XML frame method
  1. randomly chosen
  2. most frequent
  3. highest model score (tried others, see paper)

• Decisions for very large rules
  1. all
  2. all
  3. integrated scoring of VLR rules and basic translation rules (tunable)
Result: Acquis
Future Work: User Studies

- Significant increases in BLEU

- To do: validation in user studies

- Additional benefit: possible to highlight mismatch in translation
  - input
    The second paragraph of Article 21 is deleted.
  - suggested translation
    À l’article 21, le texte du deuxiéme alinéa est supprimé.
Thank You

questions?