Hybrid systems

- Human intelligence – not perfect.
- Artificial intelligence – still far from perfect.
- Combination of human + machine?
  - Need to combine the strengths of each, not the errors/weaknesses though.
  - So think about what humans do best and what machines do best.
  - Joint action.

Human-in-the-loop

- Non-simulation HILT.
- Previously, Wizard-of-Oz concept was used to mimic a smart artificial system.
- But what if we retain the human intelligence in the system rather than replace it with AI?
- Use technology to enhance or augment not entirely replace.

Homunculus

- Latin for “little man”.
- 17th Century – artificial dwarf made by alchemists.
- Illustrates the functioning of a system.
- Intelligent agents control embodied systems.

(Modular approach to cognition.)
Cyborg

Expectations shape interaction

• Expectations and beliefs about a system alter the way people interact with them.
  • Behaviour altered when told a computer system was less capable or more capable, even when performance was identical.

Example Edinburgh research:


Language translation

• As discussed previously, language communication is noisy (physical signal and information content).

• Translation adds another noisy channel to the process.
  • There can be multiple correct translations, i.e. equivalent choice.
  • Some differences down to style rather than grammar.

• Errors can multiply with multiple translations:
  • L1 ↔ L2 (e.g. English) ↔ L3.

• Speech translation involves a hidden layer of text translation.

Spoken dialects and accents add yet more noise

Burnistoun (BBC Scotland)
Faith in the tech

Don't Log Off - Series 4 - 2. Found in Translation
[Broadcast on BBC Radio 4, 10:30AM 26 Oct 2013 (Available until 1 Jan 2099)]

• The story of Bryan from the US and Anna from Russia who met online - using Google Translate.
• Bryan doesn't speak Russian and Anna doesn't speak English - they conduct their communication entirely via the online translation tool.
• Anna decided to move to the US with her two children. She sells her house in Russia and takes just three suitcases to set up home with Bryan. The couple’s understanding of each other’s languages remains minimal.
• She arrived in the US in July 2013 - and the couple had 90 days to get married or Anna would have to leave the country. The wedding date is set for 21st September - but then, suddenly, it’s called off... because Anna has concerns.

Justified?

• Is state-of-the-art Machine Translation good enough for every situation?
• How important is accuracy?
  • Conveying intimate information (love and marriage).
  • Legal information.
  • Talking to a doctor or reading medical instructions in an unfamiliar language.

Extreme examples but it is always worth having a human with knowledge of the target language check the Machine Translation output. (Mistakes are costly and embarrassing.)

HCI International Conference
Beijing 2007

Highlight of attention to attendee

Exit from the nearest door and DON'T crowd.
Please keep safe and caution when you start.
Please follow exit route and the volunteer ticket holder.
If you need any help please contact with staff at registration desk immediately.
Following the lion dance and enjoy.
Intuitively something wrong

Establishing context/discourse seems easy: it is a menu. How did you reach that conclusion?

Google can be controversial

Jorge Rivas ran into an offensive glitch when using Google Translate. In eight out of ten tries, the Spanish language word “indocumentado,” which translates to “undocumented,” was mistranslated by Google Translate as “illegal” when it appeared in a headline.

“As a journalist, when I use the term undocumented immigrant instead of illegal immigrant I’m doing so in order to remain more neutral and not use language charged with anti-immigrant sentiment. When you use the term illegal immigrant, it affects attitudes towards immigrants and people of colour.”

Current situation

• Human translators
  • relatively slow and expensive
  • limited cognitive resources
  • small set of proficient language models per individual
  • high quality – expert knowledge

• Machine translators
  • fast and cheap
  • easy access
  • mixed quality (OK for some jobs but not all – high “noise” level)

Solution? Post-editing

1. Automatic first pass of a translation.
2. Human translator then checks and cleans up the output (noise filter).
3. Correction concept also used on language learning websites (human updates/modifies human output).
   • Social network community.
• Multiple post-edited output: some overlap but many differences.
• How do people read and identify problems? Edits are not always just a matter of swapping one word for another.

Case Study: CASMACAT background

• In order to create an interactive system (Casmacat) we need to understand the progressive nature of post-editing and not just the final result. “Help along the way.”
• Human translation is an incremental, dynamic (time and space) and analogue process.
• Little known about the cognitive processes involved in detecting that a translation is wrong [error spotting].
  • Some literature on proofreading and on plausibility.
**Relevance Theory Model of Post-editing**

**NOTE:** Human translator is the hub or central component

**Methodology: eye-tracking**

1. Precise indication of where and when attention is focused.
2. Patterns of eye movements can reveal how problems are initially spotted, checked/verified and then resolved.
3. People do not read and parse the sentences normally and then generate a BLEU score (a common MT metric of success/accuracy).

**Cognitive experiments**

- Investigate the cognitive processes involved in checking for lexical, syntactic and semantic violations in translated text.
- Establish clear baselines.
- Contrast monolingual, multilingual readers and professional translators.
- Establish whether there are “levels of processing difficulty” between classes of errors.
- Test visual alignment cues and interactive aspects of the User Interface.

**Error classifications**

1) **TE:** Transposition (Easy).
   Hypothesised to be the easiest and to provide a baseline measure.
   
   *Picasso said that good artists ocpy [copy], great artists steal.*
   [Picasso sagte, dass gute Künstler kopieren, großartige Künstler klauen.]

2) **TD:** Transposition (Difficult).
   Two internal letters switched to produce an incorrect but legitimate word.
   
   *I have decided to write all my deepest thoughts in a dairy [diary] again.*
Error classifications

3) WO: Word Order. Transposition at the word level rather than letter level.

   Mostly affected [affected were] the vegetable, corn and chickpea crops.
   [Betroffen waren vor allem der Gemüse-, Mais- und Kichererbsenanbau.]

4) MT: Mistranslation of Tense or agreement. Violation in verb tense or a mismatch in gender or number agreement.

   Many of our friend [friends] are surfers and I have a great friend who lives in Tamarindo.
   [Viele unserer Freund sind Surfer und ich habe einen großartigen Freund, der in Tamarindo lebt.]
   The cuts were [would] ultimately hit the combat troops.
   [Die Kürzungen wurden letztendlich die Kampftruppen treffen.]

Error classifications

5) ML: Mistranslated Lexical item. Semantically connected but contextually odd or inappropriate.

   Judge Torkjel Nesheim cancelled [interrupted] Breivik during his monologue.
   [Richter Torkjel Nesheim unterbrach Breivik während diesem Monolog.]

Basic experimental design

• Stimuli used drawn from the German-to-English Machine Translation Marathon 2012 (MTM12) competition dataset (therefore based on real translations).
• Participants had to read each sentence and decide whether there was an error (yes/no decision).
• If yes, they had to click on the first word of where the problem began.

Contrasts

• Monolinguals (native English speakers) versus bilinguals (only translated text shown).
• Bilinguals versus professional translators (source and translation displayed).
  • Expert user analysis.
• Presentation of visual alignment between source text and translation.
Monolingual comparison example
Verb agreement incorrect (Jobs is a name here).

No error version (total reading times on each word in ms)

<table>
<thead>
<tr>
<th>Steve</th>
<th>Jobs</th>
<th>was</th>
<th>certainly</th>
<th>a</th>
<th>designer,</th>
<th>a</th>
<th>dreamer</th>
<th>and</th>
<th>self-proclaimed</th>
<th>Hippie.</th>
</tr>
</thead>
<tbody>
<tr>
<td>231</td>
<td>453</td>
<td>487</td>
<td>1254</td>
<td>151</td>
<td>624</td>
<td>0</td>
<td>1097</td>
<td>192</td>
<td>2624</td>
<td>688</td>
</tr>
</tbody>
</table>

Steve Jobs was certainly a designer, a dreamer and self-proclaimed Hippie.

Error version (total reading times on each word in ms)

<table>
<thead>
<tr>
<th>Steve</th>
<th>Jobs</th>
<th>were</th>
<th>certainly</th>
<th>a</th>
<th>designer,</th>
<th>a</th>
<th>dreamer</th>
<th>and</th>
<th>self-proclaimed</th>
<th>Hippie.</th>
</tr>
</thead>
<tbody>
<tr>
<td>434</td>
<td>590</td>
<td>1902</td>
<td>538</td>
<td>0</td>
<td>403</td>
<td>0</td>
<td>421</td>
<td>0</td>
<td>810</td>
<td>302</td>
</tr>
</tbody>
</table>

Steve Jobs were certainly a designer, a dreamer and self-proclaimed Hippie.

Source & target example (no error)

Verb agreement incorrect (Jobs is a name here).

Source & target example (no error)

Error example with visual alignment

Heatmap (aggregate of participants)
Heatmap (aggregate of participants)

Detection results
Detection rates for each error type (maximum 12) and false positives (max 120)

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Monolinguals</th>
<th>Multilinguals</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE</td>
<td>11.85</td>
<td>11.50</td>
</tr>
<tr>
<td></td>
<td>98.75</td>
<td>95.83</td>
</tr>
<tr>
<td>TD</td>
<td>8.90</td>
<td>9.40</td>
</tr>
<tr>
<td></td>
<td>74.17</td>
<td>78.33</td>
</tr>
<tr>
<td>WO</td>
<td>10.65</td>
<td>10.85</td>
</tr>
<tr>
<td></td>
<td>88.75</td>
<td>90.42</td>
</tr>
<tr>
<td>MT</td>
<td>10.50</td>
<td>9.95</td>
</tr>
<tr>
<td></td>
<td>87.50</td>
<td>82.92</td>
</tr>
<tr>
<td>ML</td>
<td>8.25</td>
<td>7.65</td>
</tr>
<tr>
<td></td>
<td>68.75</td>
<td>63.75</td>
</tr>
<tr>
<td>False Positives</td>
<td>16.85</td>
<td>20.85</td>
</tr>
<tr>
<td></td>
<td>14.04</td>
<td>17.38</td>
</tr>
</tbody>
</table>

Note the level of false positive decisions – adding more noise to the data!

Some key findings

- Bilinguals suffer a time penalty if they read the translated sentence first; translators do not.
- Bilinguals spend longer on their initial reading of the source text compared to translation; opposite true for translators.
- Bilinguals flipped between the sentence pairs less often than translators.
- Professional translators make only limited use of alignment visualizations (no effect on error rate or speed).
- But non-experts benefit from visualization (target reading time and production speed).
- We can predict expertise automatically and could use this to adjust UI to maximize productivity.

Field tests

- Testing also conducted in a real-world setting.
- Professional translation company in Madrid, Spain.
- Multiple users used system over several weeks.
- Working prototype of final system.
UI design consequences

<table>
<thead>
<tr>
<th>Cognitive findings</th>
<th>Implication for UI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignment links helpful for non-experts.</td>
<td>Visualisation highly configurable.</td>
</tr>
<tr>
<td>Syntactic and lexical priming by Machine Translation suggestion.</td>
<td>Interaction Translation Prediction (ITP) configurable.</td>
</tr>
<tr>
<td>ITP requires user training.</td>
<td>Implement training mode.</td>
</tr>
<tr>
<td>Confidence measures in MT output only useful when reliable.</td>
<td>Binary indicator.</td>
</tr>
<tr>
<td>Monolingual error detection as accurate as bilingual.</td>
<td></td>
</tr>
</tbody>
</table>

Interactive Translation Prediction (a.k.a. intelligent auto-completion)

- ITP takes place every time a keystroke is detected by the system.
- The system produces a prediction for the rest of the sentence according to the text that the user has already entered.
- This prediction is placed at the right of the text cursor.

Confidence measures

Lisboa y Madrid desee embarcarse en un camino diferente del adoptado por Grecia e Irlanda.

The goal of confidence measures highlighting is twofold.
1. It helps the user to spot wrong translations (words with very low confidence are rendered in red).
2. It can also inform the user about the dubious translations (rendered in orange).

Prediction length control

Lisboa y Madrid quieren emprender un camino diferente del adoptado por Grecia e Irlanda.

Providing the user with a new prediction whenever a key is pressed has been proved to be cognitively demanding. Therefore:
1. The workbench displays the prediction up to the first wrong word according to the confidence measures.
2. The rest of the words in the translation are displayed in grey.
Word alignment information

- Alignment of source and target words is an important part of the translation process.
- To display their correspondence, aligned words are highlighted every time the user places the mouse or the text cursor on a word.

Austin, Texas: 1st November 2016

Computer Aided Translation: Advances and Challenges

- Moving beyond post-editing machine translation, a number of recent research efforts have advanced computer aided translation methods that allow for more interactivity, richer information such as confidence scores, and the completed feedback loop of instant adaptation of machine translation models to user translations. This tutorial will explain the main techniques for several aspects of computer aided translation:
  - confidence measures
  - interactive machine translation (interactive translation prediction)
  - bilingual concordancers
  - translation option display
  - online adaptation
  - eye tracking, logging, and cognitive user models
- For each of these, the state of the art and open challenges are presented.
- The tutorial will also look under the hood of the open source CASMACAT toolkit that is based on MATECAT, and available as a “Home Edition” to be installed on a desktop machine. The target audience of this tutorial are researchers interested in computer aided machine translation and practitioners who want to use or deploy advanced CAT technology.