November 1, 2016

Computer Aided Translation: Advances and Challenges

Philipp Koehn
Computer Aided Translation
Advances and Challenges

Philipp Koehn

1 November 2016
Overview

• A practical introduction: the \textsc{casmacat} workbench

• Postediting

• Types of assistance

• Logging, eye tracking and user studies

• Implementation details of the \textsc{casmacat} workbench
part I

CASMACAT workbench
CASMACAT workbench
CASMACAT Project 2011-2014

• Cognitive studies of translators leading to insights into interface design
  \[→ \text{better understanding of translator needs}\]

• Workbench with novel types of assistance to human translators
  – interactive translation prediction
  – interactive editing and reviewing
  – adaptive translation models
  \[→ \text{better tools for translators}\]

• Demonstration of effectiveness in field tests with professional translators
  \[→ \text{increased translator productivity}\]
Core Modes

- Post-editing of MT
- Translation from scratch
- Review
Postediting Modes

- Post-editing of MT
  - CasMaCat PEMT
    - CasMaCat MT
    - External MT
    - TMs (MateCat)
  - External MT
- Translation from scratch
  - CasMaCat ITP
    - Floating prediction window
    - Text editing embedded
- Review
  - Machine learning techniques
    - Active
    - Online
Postediting Interface

- Source on left, translation on right
- Context above and below

Le Pakistan a donc été récompensé par l'assistance et les armes des États-Unis.

As a result, Pakistan was rewarded with American financial assistance and arms.

Pour mieux redistribuer ses cartes, Moucharraf a envoyé l'armée pakistanaise dans les zones ethniques qui longent l'Afghanistan, pour la première fois depuis l'indépendance du Pakistan.

In furtherance of his re-alignment, Musharraf sent the Pakistani army into the tribal areas bordering Afghanistan for the first time since Pakistan's independence.

Les opérations contre les forces des Talibans et d'Al-Qaed ont obtenu des résultats mitigés.
Confidence Measures

- Sentence-level confidence measures
  → estimate usefulness of machine translation output

- Word-level confidence measures
  → point posteditor to words that need to be changed
Incremental Updating

Machine Translation
Incremental Updating

Machine Translation

Postediting
Incremental Updating

Machine Translation

Postediting

Retraining
Pour la science, cela sert à vérifier la validité du Modèle standard (MS), et cela permet aux physiciens de scruter tout écart entre les observations et les prédictions du MS.

Ils sont d'ailleurs plusieurs à souhaiter ardemment qu'on en trouve, car la moindre différence pourrait ouvrir une porte sur une "nouvelle physique" et boucher certains trous du Modèle.
Pour mieux redistribuer ses cartes, Moucharraf a envoyé l'armée pakistanaise dans les zones ethniques qui longent l'Afghanistan, pour la première fois depuis l'indépendance du Pakistan.

In furtherance of his re-alignment, Musharraf sent the Pakistani army into the tribal areas bordering Afghanistan for the first time since Pakistan's independence.
Word Alignment

With interactive translation prediction

Shade off translated words, highlight next word to translate

Ils sont d'ailleurs plusieurs à souhaiter ardemment qu'on en trouve, car la moindre différence pourrait ouvrir une porte sur une "nouvelle physique" et boucher certains trous du Modèle.

There are elsewhere several who wish fervently that,, because
Translation Option Array

- **Visual aid**: non-intrusive provision of cues to the translator
- **Clickable**: click on target phrase → added to edit area
- **Automatic orientation**
  - most relevant is next word to be translated
  - automatic centering on next word
**Bilingual Concordancer**

- *abandonner* → *abandon*
- *give up*
- *to*
- *to abandon*
However, the European Central Bank (ECB) asked about it in a report on virtual currencies published in October.

Paraphrases for "However"
- on the other hand
- nevertheless
How do we Know it Works?

• Intrinsic Measures
  – word level confidence: user does not change words generated with certainty
  – interactive prediction: user accepts suggestions

• User Studies
  – professional translators faster with post-editing
  – ... but like interactive translation prediction better

• Cognitive studies with eye tracking
  – where is the translator looking at?
  – what causes the translator to be slow?
Logging and Eye Tracking

- Pre-loading MT suggestion
- Reading ST segment
- Reading TT segment
- Post-Editing activities (ins, del)
Home Edition

- Running **CASMACAT** on your desktop or laptop

- **Installation**
  - Installation software to run virtual machines (e.g., Virtualbox)
  - installation of Linux distribution (e.g., Ubuntu)
  - installation script sets up all the required software and dependencies
Administration through Web Browser

Administration

Translate

- Translate new document
- List documents

Engines

- Manage engines
- Upload engine
- Build new prototype

Settings

- Reset CAT and MT server
- CAT Settings
- Update Software

Deployed: fr-en-upload-1
Memory: 1.2 GB used, 6.6 GB free
Disk: 12.9 GB used, 10.2 GB free
Uptime: 22:24
Load: 0.01, 0.05, 0.08
Monday, 06 October 2014, 21:22:41
Training MT Engines

- Train MT engine on own or public data
## Managing MT Engines

### Manage Engines

#### English-French

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Size</th>
<th>Build date</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>NC+TED</td>
<td>2.3G</td>
<td>27 Mar 14</td>
<td>deploy delete download</td>
</tr>
</tbody>
</table>

Prototypes *(Inspect Details in Prototype Factory)*

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Status</th>
<th>Build date</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>NC+TED</td>
<td>done</td>
<td>Fri 20:34</td>
<td>delete</td>
</tr>
<tr>
<td>1</td>
<td>NC</td>
<td>done</td>
<td>Fri 20:34</td>
<td>create engine delete</td>
</tr>
</tbody>
</table>

#### English-Spanish

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Size</th>
<th>Build date</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>NC+TED</td>
<td>2.3G</td>
<td>27 Mar 14</td>
<td>deploy delete download</td>
</tr>
</tbody>
</table>

Prototypes *(Inspect Details in Prototype Factory)*

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Status</th>
<th>Build date</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>NC+TED+EP</td>
<td>stopped</td>
<td>Fri 20:34</td>
<td>resume delete</td>
</tr>
<tr>
<td>2</td>
<td>NC+TED</td>
<td>done</td>
<td>Fri 20:34</td>
<td>delete</td>
</tr>
<tr>
<td>1</td>
<td>NC</td>
<td>done</td>
<td>Fri 20:34</td>
<td>create engine delete</td>
</tr>
</tbody>
</table>
CAT Settings

Updated.

Interactive Translation Prediction ✔
Search and Replace ✔
Bilingual Concordancer ✔
Hide Contributions ✔
Floating Predictions ✔
Translation Options
Allow Change of Visualization Options ✔
Restrict ITP to Draft Stage
Show/Hide Visualization Preferences
displayMouseAlign
displayCaretAlign
displayShadeOffTranslatedSource ✔
displayConfidences
highlightValidated
highlightPrefix
highlightSuffix
highlightLastValidated
limitSuffixLength

update
part II

cat methods
post-editing
Productivity Improvements

(source: Autodesk)
MT Quality and Postediting Effort

- Postediting effort = number of words changed

- Evaluation metric at IWSLT 2014
  - TER = automatic metric, comparison against a reference translation
  - HTER = postediting metric, actual words changed

### English–German

<table>
<thead>
<tr>
<th>Ranking</th>
<th>HTER</th>
<th>TER</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU-BRIDGE</td>
<td>19.2</td>
<td>54.6</td>
</tr>
<tr>
<td>UEDIN</td>
<td>19.9</td>
<td>56.3</td>
</tr>
<tr>
<td>KIT</td>
<td>20.9</td>
<td>54.9</td>
</tr>
<tr>
<td>NTT-NAIST</td>
<td>21.3</td>
<td>54.7</td>
</tr>
<tr>
<td>KLE</td>
<td>28.8</td>
<td>59.7</td>
</tr>
</tbody>
</table>

### English–French

<table>
<thead>
<tr>
<th>Ranking</th>
<th>HTER</th>
<th>TER</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU-BRIDGE</td>
<td>16.5</td>
<td>42.6</td>
</tr>
<tr>
<td>RWTH</td>
<td>16.6</td>
<td>41.8</td>
</tr>
<tr>
<td>KIT</td>
<td>17.6</td>
<td>42.3</td>
</tr>
<tr>
<td>UEDIN</td>
<td>17.2</td>
<td>43.3</td>
</tr>
<tr>
<td>MITLL-AFRL</td>
<td>18.7</td>
<td>43.5</td>
</tr>
<tr>
<td>FBK</td>
<td>22.3</td>
<td>44.3</td>
</tr>
<tr>
<td>MIRACL</td>
<td>32.9</td>
<td>52.2</td>
</tr>
</tbody>
</table>
Translator Variability

- Professional translators

<table>
<thead>
<tr>
<th>Posteditor</th>
<th>HTER</th>
<th>TER</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE 1</td>
<td>32.2</td>
<td>56.1</td>
</tr>
<tr>
<td>PE 2</td>
<td>19.7</td>
<td>56.3</td>
</tr>
<tr>
<td>PE 3</td>
<td>40.9</td>
<td>56.2</td>
</tr>
<tr>
<td>PE 4</td>
<td>27.6</td>
<td>55.9</td>
</tr>
<tr>
<td>PE 5</td>
<td>25.0</td>
<td>55.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Posteditor</th>
<th>HTER</th>
<th>TER</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE 1</td>
<td>35.0</td>
<td>42.6</td>
</tr>
<tr>
<td>PE 2</td>
<td>17.5</td>
<td>42.8</td>
</tr>
<tr>
<td>PE 3</td>
<td>23.7</td>
<td>43.0</td>
</tr>
<tr>
<td>PE 4</td>
<td>39.7</td>
<td>42.3</td>
</tr>
<tr>
<td>PE 5</td>
<td>19.7</td>
<td>42.9</td>
</tr>
</tbody>
</table>

- Also very high variability
MT Quality and Productivity

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU</th>
<th>Training Sentences</th>
<th>Training Words (English)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT1</td>
<td>30.37</td>
<td>14,700k</td>
<td>385m</td>
</tr>
<tr>
<td>MT2</td>
<td>30.08</td>
<td>7,350k</td>
<td>192m</td>
</tr>
<tr>
<td>MT3</td>
<td>29.60</td>
<td>3,675k</td>
<td>96m</td>
</tr>
<tr>
<td>MT4</td>
<td>29.16</td>
<td>1,837k</td>
<td>48m</td>
</tr>
<tr>
<td>MT5</td>
<td>28.61</td>
<td>918k</td>
<td>24m</td>
</tr>
<tr>
<td>MT6</td>
<td>27.89</td>
<td>459k</td>
<td>12m</td>
</tr>
<tr>
<td>MT7</td>
<td>26.93</td>
<td>230k</td>
<td>6.0m</td>
</tr>
<tr>
<td>MT8</td>
<td>26.14</td>
<td>115k</td>
<td>3.0m</td>
</tr>
<tr>
<td>MT9</td>
<td>24.85</td>
<td>57k</td>
<td>1.5m</td>
</tr>
</tbody>
</table>

- Paper at main conference [Sanchez-Torron and Koehn, AMTA2016]
- Same type of system (Spanish–English, phrase-based, Moses)
- Trained on varying amounts of data
## MT Quality and Productivity

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU</th>
<th>Training Sentences</th>
<th>Training Words (English)</th>
<th>Post-Editing Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT1</td>
<td>30.37</td>
<td>14,700k</td>
<td>385m</td>
<td>4.06 sec/word</td>
</tr>
<tr>
<td>MT2</td>
<td>30.08</td>
<td>7,350k</td>
<td>192m</td>
<td>4.38 sec/word</td>
</tr>
<tr>
<td>MT3</td>
<td>29.60</td>
<td>3,675k</td>
<td>96m</td>
<td>4.23 sec/word</td>
</tr>
<tr>
<td>MT4</td>
<td>29.16</td>
<td>1,837k</td>
<td>48m</td>
<td>4.54 sec/word</td>
</tr>
<tr>
<td>MT5</td>
<td>28.61</td>
<td>918k</td>
<td>24m</td>
<td>4.35 sec/word</td>
</tr>
<tr>
<td>MT6</td>
<td>27.89</td>
<td>459k</td>
<td>12m</td>
<td>4.36 sec/word</td>
</tr>
<tr>
<td>MT7</td>
<td>26.93</td>
<td>230k</td>
<td>6.0m</td>
<td>4.66 sec/word</td>
</tr>
<tr>
<td>MT8</td>
<td>26.14</td>
<td>115k</td>
<td>3.0m</td>
<td>4.94 sec/word</td>
</tr>
<tr>
<td>MT9</td>
<td>24.85</td>
<td>57k</td>
<td>1.5m</td>
<td>5.03 sec/word</td>
</tr>
</tbody>
</table>

- User study with professional translators
- Correlation between BLEU and post-editing speed?
MT Quality and Productivity

BLEU against PE speed and regression line with 95% confidence bounds
+1 BLEU $\leftrightarrow$ decrease in PE time of $\sim 0.16$ sec/word
MT Quality and PE Quality

better MT ↔ fewer post-editing errors
Translator Variability

<table>
<thead>
<tr>
<th></th>
<th>HTER</th>
<th>Edit Rate</th>
<th>PE speed (spw)</th>
<th>MQM Score</th>
<th>Fail</th>
<th>Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR1</td>
<td>44.79</td>
<td>2.29</td>
<td>4.57</td>
<td>98.65</td>
<td>10</td>
<td>124</td>
</tr>
<tr>
<td>TR2</td>
<td>42.76</td>
<td>3.33</td>
<td>4.14</td>
<td>97.13</td>
<td>23</td>
<td>102</td>
</tr>
<tr>
<td>TR3</td>
<td>34.18</td>
<td>2.05</td>
<td>3.25</td>
<td>96.50</td>
<td>26</td>
<td>106</td>
</tr>
<tr>
<td>TR4</td>
<td>49.90</td>
<td>3.52</td>
<td>2.98</td>
<td>98.10</td>
<td>17</td>
<td>120</td>
</tr>
<tr>
<td>TR5</td>
<td>54.28</td>
<td>4.72</td>
<td>4.68</td>
<td>97.45</td>
<td>17</td>
<td>119</td>
</tr>
<tr>
<td>TR6</td>
<td>37.14</td>
<td>2.78</td>
<td>2.86</td>
<td>97.43</td>
<td>24</td>
<td>113</td>
</tr>
<tr>
<td>TR7</td>
<td>39.18</td>
<td>2.23</td>
<td>6.36</td>
<td>97.92</td>
<td>18</td>
<td>112</td>
</tr>
<tr>
<td>TR8</td>
<td>50.77</td>
<td>7.63</td>
<td>6.29</td>
<td>97.20</td>
<td>19</td>
<td>117</td>
</tr>
<tr>
<td>TR9</td>
<td>39.21</td>
<td>2.81</td>
<td>5.45</td>
<td>96.48</td>
<td>22</td>
<td>113</td>
</tr>
</tbody>
</table>

- Higher variability between translators than between MT systems
Postediting and MT Metrics

- Goal of MT quality metrics not clear
  - understandability: do you get the meaning?
  - post-editing effort: how much effort to change?

- Example: dropping of the word “not”
  - understandability: big mistake
  - post-editing effort: quick add of just one word

- Not clear, what tradition manual metrics prefer (adequacy, fluency)

- Not clear, what BLEU score etc. prefer
word alignment
Pour mieux redistribuer ses cartes, Moucharraf a envoyé l'armée pakistanaise dans les zones ethniques qui longent l'Afghanistan, pour la première fois depuis l'indépendance du Pakistan.

In furtherance of his re-alignment, Musharraf sent the Pakistani army into the tribal areas bordering Afghanistan for the first time since Pakistan's independence.
Word Alignment from MT

- Machine translation output is constructed by phrase mappings
- Each phrase mapping has internal word alignment
  ⇒ This can be used to visualize word alignments
- But: word alignment points become invalid after user edits
Word Alignment from Alignment Tools

- During machine translation training, standard component is word alignment

- Standard tools
  - old workhorse: GIZA++
  - currently popular tool: fast-align

- These tools have been adapted to align new sentence pairs
Pour mieux redistribuer ses cartes, Moucharraf a envoyé l'armée pakistanaise dans les zones ethniques qui longent l'Afghanistan, pour la première fois depuis l'indépendance du Pakistan.

In furtherance of his re-alignment, Musharraf sent the Pakistani army into the tribal areas bordering Afghanistan for the first time since Pakistan's independence.

- Highlight the source word aligned to the word at the current mouse position
Pour mieux redistribuer ses cartes, Moucharraf a envoyé l'armée pakistanaise dans les zones ethniques qui longent l'Afghanistan, pour la première fois depuis l'indépendance du Pakistan.

In furtherance of his re-alignment, Musharraf sent the Pakistani army into the tribal areas bordering Afghanistan for the first time since Pakistan's independence.

- Highlight the source word aligned to the word at the current caret position
Shade Off Translated

- Use in interactive prediction mode
- Shade off words that are already translated
- Highlight words aligned to first predicted translation word
confidence measures
("quality estimation")
Levels

• Machine translation engine indicates where it is likely wrong

• Different Levels of granularity
  – document-level (SDL’s ”TrustScore”)
  – sentence-level
  – word-level
Sentence-Level Confidence

- Translators are used to "Fuzzy Match Score"
  - used in translation memory systems
  - roughly: ratio of words that are the same between input and TM source
  - if less than 70%, then not useful for post-editing

- We would like to have a similar score for machine translation

- Even better
  - estimation of post-editing time
  - estimation of from-scratch translation time
  - can also be used for pricing

- Very active research area
Quality Estimation Shared Task

• Shared task organized at WMT since 2012

• Given
  – source sentence
  – machine translation

• Predict
  – HTER score on post-edited sentences (2013–2016)
  – post-editing time (2013, 2014)

• Also task for word-level quality estimation (2014–2016) and document-level quality estimation (2015)
QuEst

• Open source tool for quality estimation

• Source sentence features
  – number of tokens
  – language model (LM) probability
  – 1–3-grams observed in training corpus
  – average number of translations per word

• Similar target sentence features

• Alignment features
  – difference in number of tokens and characters
  – ratio of numbers, punctuation, nouns, verbs, named entities
  – syntactic similarity (POS tags, constituents, dependency relationships)

• Scores and properties of the machine translation derivation

• Uses Python’s scikit-learn implementation of SVM regression
WMT 2016: Best System

- Yandex School of Data Analysis (Kozlova et al., 2016)

- QuEst approach with additional features
  - syntactically motivated features
  - language model and statistics on web-scale corpus
  - pseudo-references and back-translations
  - other miscellaneous features

- Performance
  - mean average HTER difference 13.53
  - ranking correlation 0.525
word level confidence
Visualization

- Highlight words less likely to be correct

And on that the signs are mixed.  

Y en que los indicios son desiguales.
Methods

• Simple methods quite effective
  – IBM Model 1 scores
  – posterior probability of the MT model

• Machine learning approach
  – similar features as for sentence-level quality estimation
Annotation

• Machine translation output

Quick brown fox jumps on the dog lazy.

• Post-editing

The quick brown fox jumps over the lazy dog.

• Annotation

Fast brown fox jumps on the dog lazy.

bad good good good bad good good good good

• Problems: dropped words? reordering?
Quality Requirements

- Evaluated in user study

- Feedback
  - could be useful feature
  - but accuracy not high enough

- To be truly useful, accuracy has to be very high

- Current methods cannot deliver this
WMT 2016: Best System

- Unbabel (Martins et al., 2016)

- Viewed as tagging task

- Features: black box and language model features

- Method: Combination of
  - feature-rich linear HMM model
  - deep neural networks
    (feed-forward, bi-directionally recurrent, convolutional)

- Performance
  - F-score for detecting good words: 88.45
  - F-score for detecting bad words: 55.99
automatic reviewing
Automatic Reviewing

- Can we identify errors in human translations?
  - missing / added information
  - inconsistent use of terminology

Input Sentence

Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Human Translation

Moreover, he planned for months to give a lecture in Miami.
Reviewing with E-Pen

• Intuition
  – reviewing more efficient with pen and paper
  – e-pen enables this work process in digital environment

• Work carried out
  – fronted modified for larger drawing area
  – backend support for hand-written text recognition (HTR)
  – development of methods for HTR

• Field trial carried out → corpus of reviewing edits
Analysis of Reviewer Edits

• 171 insertions — vast majority function words

• 152 deletions — about half substantial content

• 621 replacements — of which:
  – 75 changes to punctuation only
  – 28 change to lowercase / uppercase
  – 29 cases that are mostly deletions
  – 8 cases that are mostly insertions
  – 289 morphological/spelling changes (Levenshtein distance of less than 50%)
  – 190 other changes, about equal amounts function words and content words
Automatic Reviewing

• Focus on translation errors
  – not: basic spell checking
  – not: basic grammar checking

• Do not try the impossible
  – semantic errors
  – errors in function words

• What is left?
  – added content (insertions)
  – non-translated content (deletions)
  – inconsistency in terminology
Method

• Word alignment of human translation and source

• Detect unaligned words
  – insertion of content words: unaligned sequence of words in the draft translation
  – deletion of content words: unaligned sequence of words in the source sentence
  – inconsistent terminology: source word occurs multiple times, aligned to different word

• Only content words (minimum 4 characters)
Evaluation on Field Trial Data

- **Two evaluation metrics**
  - strict: predicted word X deleted / inserted
  - generous: predicted any deletion / insertion

<table>
<thead>
<tr>
<th>Edit type</th>
<th>Strict Scoring</th>
<th>Generous Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>Deletion</td>
<td>7%</td>
<td>27%</td>
</tr>
<tr>
<td>Insertion</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Any edit</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- **Good enough to be useful?**
Subjective Evaluation

- Evaluation on community translation platform data
- English–German
- Predict insertions and deletions
- Manually check if these are valid suggestions (i.e., precision only) by native German speaker
Results

- 4 cases of detection of valid errors (3 deletions, 1 insertion)
- 31 false alarms

<table>
<thead>
<tr>
<th>Count</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 cases</td>
<td>unaligned verb</td>
</tr>
<tr>
<td>6 cases</td>
<td>one-to-many alignment</td>
</tr>
<tr>
<td>2 cases</td>
<td>non-literal</td>
</tr>
<tr>
<td>6 cases</td>
<td>misalignment, often due to unknown word</td>
</tr>
<tr>
<td>1 case</td>
<td>valid verb ellipsis, repeated in sub clause</td>
</tr>
</tbody>
</table>

- Good enough to be useful?
interactive translation prediction
Interactive Translation Prediction

Input Sentence
Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator
Interactive Translation Prediction

Input Sentence
Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator
He
Interactive Translation Prediction

Input Sentence
Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator
He | has
Interactive Translation Prediction

Input Sentence
Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator
He has | for months
Interactive Translation Prediction

Input Sentence
Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator
He planned |
Input Sentence

Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator

He planned | for months
Visualization

• Show $n$ next words

Olvidarlo. Es demasiado

arriesgado. Estoy haciendo

• Show rest of sentence
Spence Green’s Lilt System

- Show alternate translation predictions

- Show alternate translations predictions with probabilities

To equip students to reduce mobility and Institute Jedlička, teachers regularly visit Jedličkov activity and make them and them ongoing make regular are regularly
Search for best translation creates a graph of possible translations
One path in the graph is the best (according to the model)
This path is suggested to the user
The user may enter a different translation for the first words

We have to find it in the graph
Prediction from Search Graph

We can predict the optimal completion (according to the model)
• Average response time based on length of the prefix and number of edits
• Main bottleneck is the string edit distance between prefix and path.
Word Completion

- Complete word once few letters are typed

- Example: predict college over university?

- User types the letter $u \rightarrow$ change prediction

- "Desperate" word completion: find any word that matches
Redecoding

• Translate the sentence again, enforce matching the prefix

• Recent work on this: Wuebker et al. [ACL 2016]

Models and Inference for Prefix-Constrained Machine Translation

Joern Wuebker, Spence Green, John DeNero, Saša Hasan
Lilt, Inc.

first_name@lilt.com

Minh-Thang Luong
Stanford University

lmthang@stanford.edu
Prefix-Matching Decoding

• Prefix-matching phase
  – only allow translation options that match prefix
  – prune based on target words matched

• Ensure that prefix can be created by system
  – add synthetic translation options from word aligned prefix
    (but with low probability)
  – no reordering limit

• After prefix is match, regular beam search

• Fast enough?
  ⇒ Wuebker et al. [ACL 2016] report 51-89ms per sentence
Tuning

- Optimize to produce better predictions
- Focus on next few words, not full sentence
- Tuning metric
  - prefix BLEU (ignoring prefix to measure score)
  - word prediction accuracy
  - length of correctly predicted suffix sequence
- Generate diverse n-best list to ensure learnability
- Wuebker et al. [ACL 2016] report significant gains
Neural Interactive Translation Prediction

• Recent success of neural machine translation (see WMT 2016)

• For instance, attention model
Neural MT: Sequential Prediction

• The model produces words in sequence

\[ p(\text{output}_t|\{\text{output}_1, \ldots, \text{output}_{t-1}\}, \text{input}) = g(\hat{\text{output}}_{t-1}, \text{context}_t, \text{hidden}_t) \]

• Translation prediction: feed in user prefix
**Example**

**Input:** Das Unternehmen sagte, dass es in diesem Monat mit Bewerbungsgesprächen beginnen wird und die Mitarbeiterzahl von Oktober bis Dezember steigt.

<table>
<thead>
<tr>
<th>Correct</th>
<th>Prediction</th>
<th>Prediction probability distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>the</td>
<td>the (99.2%)</td>
</tr>
<tr>
<td>✓</td>
<td>company</td>
<td>company (90.9%), firm (7.6%)</td>
</tr>
<tr>
<td>✓</td>
<td>said</td>
<td>said (98.9%)</td>
</tr>
<tr>
<td>✓</td>
<td>it</td>
<td>it (42.6%), this (14.0%), that (13.1%), job (2.0%), the (1.7%), ...</td>
</tr>
<tr>
<td>✓</td>
<td>will</td>
<td>will (77.5%), is (4.5%), started (2.5%), ’s (2.0%), starts (1.8%), ...</td>
</tr>
<tr>
<td>✓</td>
<td>start</td>
<td>start (49.6%), begin (46.7%)</td>
</tr>
<tr>
<td></td>
<td>inter@@</td>
<td>job (16.1%), application (6.1%), en@@ (5.2%), out (4.8%), ...</td>
</tr>
<tr>
<td>✗</td>
<td>viewing</td>
<td>state (32.4%), related (5.8%), viewing (3.4%), min@@ (2.0%), ...</td>
</tr>
<tr>
<td>✗</td>
<td>applicants</td>
<td>talks (61.6%), interviews (6.4%), discussions (6.2%), ...</td>
</tr>
<tr>
<td>✓</td>
<td>this</td>
<td>this (88.1%), so (1.9%), later (1.8%), that (1.1%)</td>
</tr>
<tr>
<td>✓</td>
<td>month</td>
<td>month (99.4%)</td>
</tr>
<tr>
<td>✗</td>
<td>,</td>
<td>and (90.8%), , (7.7%)</td>
</tr>
<tr>
<td>✗</td>
<td>with</td>
<td>and (42.6%), increasing (24.5%), rising (6.3%), with (5.1%), ...</td>
</tr>
<tr>
<td>✓</td>
<td>staff</td>
<td>staff (22.8%), the (19.5%), employees (6.3%), employee (5.0%), ...</td>
</tr>
<tr>
<td>✗</td>
<td>levels</td>
<td>numbers (69.0%), levels (3.3%), increasing (3.2%), ...</td>
</tr>
<tr>
<td>✗</td>
<td>rising</td>
<td>increasing (40.1%), rising (35.3%), climbing (4.4%), rise (3.4%), ...</td>
</tr>
<tr>
<td>✓</td>
<td>from</td>
<td>from (97.4%)</td>
</tr>
<tr>
<td>✓</td>
<td>October</td>
<td>October (81.3%), Oc@@ (12.8%), oc@@ (2.9%), Oct (1.2%)</td>
</tr>
<tr>
<td>✗</td>
<td>through</td>
<td>to (73.2%), through (15.6%), until (8.7%)</td>
</tr>
<tr>
<td>✓</td>
<td>December</td>
<td>December (85.6%), Dec (8.0%), to (5.1%)</td>
</tr>
<tr>
<td>✓</td>
<td>.</td>
<td>. (97.5%)</td>
</tr>
</tbody>
</table>
• Better prediction accuracy, even when systems have same BLEU score (state-of-the-art German-English systems, compared to search graph matching)

<table>
<thead>
<tr>
<th>System</th>
<th>Configuration</th>
<th>BLEU</th>
<th>Word Prediction Accuracy</th>
<th>Letter Prediction Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural</td>
<td>no beam search</td>
<td>34.5</td>
<td>61.6%</td>
<td>86.8%</td>
</tr>
<tr>
<td></td>
<td>beam size 12</td>
<td>36.2</td>
<td>63.6%</td>
<td>87.4%</td>
</tr>
<tr>
<td>Phrase-based</td>
<td>-</td>
<td>34.5</td>
<td>43.3%</td>
<td>72.8%</td>
</tr>
</tbody>
</table>

• Better recovery from failure

• Fast enough with GPU
  – translation speed with CPU: 100 ms/word
  – translation speed with GPU: 7ms/word
bilingual concordancer
Bilingual Concordancer

abandon

abandoner

give up

to

t to abandon

es tout en refusant d’ abandonner son arsenal nucléaire

Musharraf -- together

Musharraf -- et les cos, Musharraf -- together

American reluctance to abandon constitutionality, remove

simply threatened to abandon or never to conclude the

conclusion.

...
How does it Work?

• Have word-aligned parallel corpus

• Efficient data structure to quickly look up queried phrases
  (suffix arrays, we’ll come back to them later)

• Translation spotting
  – look up queried phrase
  – use word alignment to identify target phrase
  – some edge cases (unaligned words at beginning/end)
Verification of Terminology

- Translation of German *Windkraft*

<table>
<thead>
<tr>
<th>Example</th>
<th><strong>Windkraft</strong> (noun, feminine) (also: Windenergie)</th>
<th><strong>wind power</strong> (noun)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zum Vergleich: <em>Windkraft</em> schafft fast sieben Mal mehr.</td>
<td>By way of comparison, <em>wind power</em> generates almost seven times as much.</td>
<td></td>
</tr>
<tr>
<td><em>Windkraft</em> ist eine etablierte, wettbewerbsfähige Technologie mit hoher Zuverlässigkeit.</td>
<td><em>Wind power</em> is an established, competitive technology with high reliability</td>
<td></td>
</tr>
<tr>
<td>German: <a href="http://www.powergeneration.siemens.de/about...ns-services/">www.powergeneration.siemens.de/about...ns-services/</a></td>
<td>English: <a href="http://www.powergeneration.siemens.de/about...ns-services/">www.powergeneration.siemens.de/about...ns-services/</a></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example</th>
<th><strong>Windkraft</strong> (noun, feminine) (also: Windenergie)</th>
<th><strong>wind energy</strong> (noun)</th>
</tr>
</thead>
<tbody>
<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 <em>Windkraft</em>.</td>
<td>But as the focus of the climate change issue shifts increasingly from policy to action, this price will increase and cleaner <em>energy</em> sources like <em>wind</em> will become more competitive.</td>
<td></td>
</tr>
<tr>
<td>German: <a href="http://emagazine.credit-suisse.com/app/art...4382">emagazine.credit-suisse.com/app/art...4382</a> (=DE)</td>
<td>English: <a href="http://emagazine.credit-suisse.com/app/art...4382">emagazine.credit-suisse.com/app/art...4382</a> (=en)</td>
<td></td>
</tr>
<tr>
<td>Nur wenige befürchten hingegen, dass dies auch bei erneuerbaren Energieträgern wie Biomasse oder <em>Windkraft</em> der Fall sein wird.</td>
<td>However, only a few fear that this will also be the case with renewable <em>energy</em> sources such as biomass or <em>wind energy</em>.</td>
<td></td>
</tr>
</tbody>
</table>

- Context shows when each translation is used
- Indication of source supports trust in translations
<table>
<thead>
<tr>
<th>expression</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>dinndons de la farce</td>
<td>4</td>
</tr>
<tr>
<td>monté un bateau</td>
<td>3</td>
</tr>
<tr>
<td>faire avoir</td>
<td>3</td>
</tr>
<tr>
<td>se fasse rouler</td>
<td>2</td>
</tr>
<tr>
<td>faire berner</td>
<td>2</td>
</tr>
<tr>
<td>se fai jouer</td>
<td>2</td>
</tr>
<tr>
<td>moqués de</td>
<td>2</td>
</tr>
<tr>
<td>fait</td>
<td>2</td>
</tr>
<tr>
<td>les a</td>
<td>2</td>
</tr>
<tr>
<td>se sont fait avoir</td>
<td>2</td>
</tr>
<tr>
<td>le public pour attirer la</td>
<td>1</td>
</tr>
<tr>
<td>a fait une balade</td>
<td>1</td>
</tr>
<tr>
<td>nous rouler dans ce projet</td>
<td>1</td>
</tr>
<tr>
<td>nous tous</td>
<td>1</td>
</tr>
<tr>
<td>en train de monter un bateau à la population canadienne</td>
<td>1</td>
</tr>
<tr>
<td>tête des contribuables que se paie le</td>
<td>1</td>
</tr>
<tr>
<td>passer une petite vite</td>
<td>1</td>
</tr>
<tr>
<td>bourrer de l'autre côté de la chambre en</td>
<td>1</td>
</tr>
<tr>
<td>ont pris la voiture que pour faire une balade</td>
<td>1</td>
</tr>
</tbody>
</table>

92 traductions de *take+ .. ride* dans 106 occurrences

**dindons de la farce**

Emissions continue to rise and taxpayers are being taken along for the ride.

They are left with nothing. Now they are here illegally with no documentation. Canadians are being taken for a ride.

This would affect close to 400,000 Canadians, 80,000 of them Quebecers, who have been the ones taken for a ride.

I think that this is a prime example of a tainted system in which people who cannot afford to invest in sectors eligible for tax credits are urged to do so through all kinds of scams and end up being taken for a ride.

Les émissions continuent d’augmenter et c’est le contribuable qui est le dindon de la farce.

Ces personnes se trouvent ici illégalement, elles n’ont aucun document et nous, les Canadiens, sommes les dindons de la farce.

Il s’agit d’une mesure qui toucherait près de 400 000 Canadiens, dont 80 000 Québécois, qui ont été les dindons de la farce.

Je pense que c’est un exemple parfait d’un système vicié, où des gens qui n’ont pas les moyens d’investir dans des domaines où on peut obtenir des crédits d’impôt se voient, par toutes sortes de subterfuges, invités à le faire et, en bout de ligne, ils se trouvent à être les dindons de la farce.
TransSearch: Improved Transpotting

• Used to solve difficult translation problems
  – 7.2 million queries submitted to the system over a 6-year period
  – 87% contain at least two words
  – mainly search for idiomatic expressions such as *in keeping with*

• Improved translation spotting [Bourdaillet et al., MT Journal 2011]

• Filtering with classifier (45 features, trained on annotated data)
  – relative word count
  – word alignment scores
  – ratio of function words

• Merging of translations that only differ in function words, morphology

• Pseudo-relevance feedback
translation options
Translation Option Array

<table>
<thead>
<tr>
<th>Translation Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kletterei sind schwer verletzt, und zehn Menschen werden vermisst, nachdem Mount Ontake (御嶽山, Ontake-san), ein beliebter Kletterplatz im zentralen Japan, ausbrach, zum ersten Mal in fünf Jahren.</td>
</tr>
<tr>
<td>Philipp Koehn Computer Aided Translation 1 November 2016</td>
</tr>
</tbody>
</table>

- Visual aid: non-intrusive provision of cues to the translator
- Trigger passive vocabulary
Visualization

• Show up to 6 options per word or phrase

• Rank best option on top

• Use color highlighting to show likelihood
  (grey = less likely to be useful)

• Clickable: click on target phrase → added to edit area

• Automatic orientation
  – most relevant is next word to be translated
  – automatic centering on next word
How to Rank

• Basic idea: best options on top

• Problem: how to rank word translation vs. phrase translations?

• Method: utilize future cost estimates

• Translation score
  – sum of translation model costs
  – language model estimate
  – outside future cost estimate

```
the first time
das erste mal
tm:-0.56,lm:-2.81
  d:-0.74. all:-4.11

-9.3 +
-4.11 =
-13.41
```
Improving Rankings

- Removal of duplicates and near duplicates

- Ranking by likelihood to be used in the translation
  → can this be learned from user feedback?
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?
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
Example: Arabic

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>.example</td>
<td><strong>defying</strong> once new president george w. bush, who opposes the no date has been set for the in defiance of once again, a a defiant the first in defiance of once again, once again the us president george w. bush's, which opposes the date of the in 2008, defying the again us president george w. bush, who opposed to setting any the date of the in 2008, deflecting the time, defying once again, the</td>
</tr>
</tbody>
</table>

**up to 10 translations for each word / phrase**
Example: Arabic

withdrawal of troops

US from Iraq

the fighting forces

US from Iraq

the US
No big difference — once significantly better
Monolingual Translation Triage

- Study on Russian–English (Schwartz, 2014)

- Allow monolingual translators to assess their translation
  - confident → accept the translation
  - verify → proofread by bilingual
  - partially unsure → part of translation handled by bilingual
  - completely unsure → handled by bilingual

- Monolingual translator highly effective in triage
Monolingual Translation: 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
paraphrasing
Paraphrasing

Input Sentence
Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator
He planned for months to **give a lecture** in Miami in October.

User requests alternative translations for parts of sentence.
Visualization in CASMACAT

• User marks part of translation
• Clicks on paraphrasing button
• Alternative translations appear
Paraphrasing Research

- Somewhat popular research area

- Popular method: extract from parallel data
  - goal: find paraphrases for phrase $e$
  - look up likely translations $f_1, f_2, \ldots$ for $e$
  - for each $f_i$, look up likely translations $e_i', e_i'', \ldots$
  $\Rightarrow$ these are the paraphrases

- Refinement: collect over several foreign languages, intersect

- Paraphrase database for several languages:
  http://paraphrase.org/
Paraphrasing in Context

• Our problem: paraphrasing in context
  – driven by source
  – considers sentence context
  – ranking and diversity important
  – real time performance

• Approach
  – target span is mapped to source span
  – search graph is consulted for alternative translations for source span
  – additional translations generated by combining translation options
  ⇒ initial list of translations
  – various components to distill $n$-best paraphrases
Components

• Filtering: remove some translations
  – with extraneous punctuation
  – too similar to others
  – additional function words

• Scoring: score translations
  – translation model scores
  – language model score in context
  – compare alternate translations against best path

• Sorting: rank list
  – cluster translations by similarity
  – picks best translation from each cluster
Automatic Evaluation

• Motivation
  – alternative translations should fix translation errors
  → create bad translations by back-translation

• Process
  – Train machine translation system for both directions
  – Translate test set target → source → target*
  – Spot differences between target and target*
  – Use span in target* as “marked by user”, span in target as correct
Example

• Translate

Unlike in Canada, the American states are responsible for the organisation of federal elections.

• Into

в отличие от канады, американские штаты ответственны за организацию федеральных выборов в соединенных штатах.

• Back into English

Unlike in Canada, US states are responsible for the organization of federal elections.
Manual Evaluation

- Web based interactive evaluation tool

- Same setup as automatic evaluation
  - shows target span
  - 5 selectable paraphrases
  - user accepts one $\rightarrow$ correct

- Four users (U1–U4)

- Number of instances where one translation is correct

<table>
<thead>
<tr>
<th>Method</th>
<th>U1</th>
<th>U2</th>
<th>U3</th>
<th>U4</th>
<th>average score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>6/50</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>17</td>
<td>12</td>
<td>10</td>
<td>13/50</td>
</tr>
<tr>
<td>10</td>
<td>24</td>
<td>20</td>
<td>26</td>
<td>29</td>
<td>26/50</td>
</tr>
</tbody>
</table>
adaptation
Adaptation

- Machine translation works best if optimized for domain
- Typically, large amounts of out-of-domain data available
  - European Parliament, United Nations
  - unspecified data crawled from the web
- Little in-domain data (maybe 1% of total)
  - information technology data
  - more specific: IBM’s user manuals
  - even more specific: IBM’s user manual for same product line from last year
  - and even more specific: sentence pairs from current project
- Various domain adaptation techniques researched and used
Combining Data

- Too biased towards out of domain data
- May flag translation options with indicator feature functions
Interpolate Models

- \( p_c(e|f) = \lambda_{\text{in}} p_{\text{in}}(e|f) + \lambda_{\text{out}} p_{\text{out}}(e|f) \)
- Quite successful for language modelling
Multiple Models

- Multiple models → multiple feature functions

Use both

Out-of Domain Model

In Domain Model
Backoff

Diagram:
- Look up phrase
  - If found, return
  - If not found
    - In Domain Model
    - If found, return
    - Out-of Domain Model
      - If found, return
Fill-Up

- Use translation options from in-domain table
- Fill up with additional options from out-of-domain table
Sentence Selection

- Select out-of-domain sentence pairs that are similar to in-domain data
- Score similarity with language model, other means
Project Adaptation

- Method developed by the Matecat project
- Update model during translation project
  - After each day
    - collected translated sentences
    - add to model
    - optimize
- Main benefit after the first day
Adaptable Translation Model

- Store in memory
  - parallel corpus
  - word alignment

- Adding new sentence pair
  - word alignment of sentence pair
  - add sentence pair
  - update index (suffix array)

- Retrieve phrase translations on demand
On-Demand Word Alignment

- Needed: word alignment method that scores a sentence pairs

- Online EM algorithm
  - keep sufficient statistics of corpus in memory
  - run EM iteration on single sentence pair
  - update statistics
  - return word alignment

- For efficiency reason, a static model may be sufficient

- Implementations in both mGIZA and fast-align
<table>
<thead>
<tr>
<th></th>
<th>government of the people, by the people, for the people</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>of the people, by the people, for the people</td>
</tr>
<tr>
<td>3</td>
<td>the people, by the people, for the people</td>
</tr>
<tr>
<td>4</td>
<td>people, by the people, for the people</td>
</tr>
<tr>
<td>5</td>
<td>, by the people, for the people</td>
</tr>
<tr>
<td>6</td>
<td>by the people, for the people</td>
</tr>
<tr>
<td>7</td>
<td>the people, for the people</td>
</tr>
<tr>
<td>8</td>
<td>people, for the people</td>
</tr>
<tr>
<td>9</td>
<td>, for the people</td>
</tr>
<tr>
<td>10</td>
<td>for the people</td>
</tr>
<tr>
<td>11</td>
<td>the people</td>
</tr>
<tr>
<td>12</td>
<td>people</td>
</tr>
</tbody>
</table>
Sorted Suffixes

5    , by the people, for the people
9    , for the people
6    by the people, for the people
10   for the people
1    government of the people, by the people, for the people
2    of the people, by the people, for the people
12   people
4    people, by the people, for the people
8    people, for the people
11   the people
3    the people, by the people, for the people
7    the people, for the people
## Suffix Array

<table>
<thead>
<tr>
<th></th>
<th>the people, for the people</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>, for the people</td>
</tr>
<tr>
<td>9</td>
<td>, for the people</td>
</tr>
<tr>
<td>6</td>
<td>by the people, for the people</td>
</tr>
<tr>
<td>10</td>
<td>for the people</td>
</tr>
<tr>
<td>1</td>
<td>government of the people, by the people, for the people</td>
</tr>
<tr>
<td>2</td>
<td>of the people, by the people, for the people</td>
</tr>
<tr>
<td>12</td>
<td>people</td>
</tr>
<tr>
<td>4</td>
<td>people, by the people, for the people</td>
</tr>
<tr>
<td>8</td>
<td>people, for the people</td>
</tr>
<tr>
<td>11</td>
<td>the people</td>
</tr>
<tr>
<td>3</td>
<td>the people, by the people, for the people</td>
</tr>
<tr>
<td>7</td>
<td>the people, for the people</td>
</tr>
</tbody>
</table>

**suffix array:** sorted index of corpus positions
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>by the people, for the people</td>
</tr>
<tr>
<td>9</td>
<td>by the people, for the people</td>
</tr>
<tr>
<td>6</td>
<td>for the people</td>
</tr>
<tr>
<td>10</td>
<td>government of the people, by the people, for the people</td>
</tr>
<tr>
<td>1</td>
<td>people</td>
</tr>
<tr>
<td>2</td>
<td>people, by the people, for the people</td>
</tr>
<tr>
<td>12</td>
<td>people, for the people</td>
</tr>
<tr>
<td>4</td>
<td>the people</td>
</tr>
<tr>
<td>8</td>
<td>the people, by the people, for the people</td>
</tr>
<tr>
<td>11</td>
<td>the people, for the people</td>
</tr>
<tr>
<td>3</td>
<td>the people, by the people, for the people</td>
</tr>
<tr>
<td>7</td>
<td>the people, for the people</td>
</tr>
</tbody>
</table>

Query: **people**
Querying the Suffix Array

<table>
<thead>
<tr>
<th>2</th>
<th>of the people, by the people, for the people</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>by the people, for the people</td>
</tr>
<tr>
<td>9</td>
<td>, for the people</td>
</tr>
<tr>
<td>6</td>
<td>by the people, for the people</td>
</tr>
<tr>
<td>10</td>
<td>for the people</td>
</tr>
<tr>
<td>1</td>
<td>government of the people, by the people, for the people</td>
</tr>
<tr>
<td>12</td>
<td>people</td>
</tr>
<tr>
<td>4</td>
<td>people, by the people, for the people</td>
</tr>
<tr>
<td>8</td>
<td>people, for the people</td>
</tr>
<tr>
<td>11</td>
<td>the people</td>
</tr>
<tr>
<td>3</td>
<td>the people, by the people, for the people</td>
</tr>
<tr>
<td>7</td>
<td>the people, for the people</td>
</tr>
</tbody>
</table>

Query: people

Binary search: start in the middle
### Querying the Suffix Array

<table>
<thead>
<tr>
<th>Index</th>
<th>Suffix</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>the people</td>
<td>For the people</td>
</tr>
<tr>
<td>3</td>
<td>people</td>
<td>The people</td>
</tr>
<tr>
<td>4</td>
<td>people , by the people</td>
<td>Government of the people</td>
</tr>
<tr>
<td>5</td>
<td>people , by the people</td>
<td>For the people</td>
</tr>
<tr>
<td>6</td>
<td>by the people</td>
<td>For the people</td>
</tr>
<tr>
<td>7</td>
<td>by the people</td>
<td>By the people</td>
</tr>
<tr>
<td>8</td>
<td>people , for the people</td>
<td>The people</td>
</tr>
<tr>
<td>9</td>
<td>, for the people</td>
<td>By the people</td>
</tr>
<tr>
<td>10</td>
<td>, for the people</td>
<td>For the people</td>
</tr>
</tbody>
</table>

**Query:** **people**

**Binary search:** discard upper half
Querying the Suffix Array

Query: people

Binary search: middle of remaining space
Querying the Suffix Array

Query: **people**

Binary search: match
Querying the Suffix Array

<table>
<thead>
<tr>
<th>Index</th>
<th>Suffix</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>, by the people, for the people</td>
</tr>
<tr>
<td>9</td>
<td>, for the people</td>
</tr>
<tr>
<td>6</td>
<td>by the people, for the people</td>
</tr>
<tr>
<td>10</td>
<td>for the people</td>
</tr>
<tr>
<td>1</td>
<td>government of the people, by the people, for the people</td>
</tr>
<tr>
<td>2</td>
<td>of the people, by the people, for the people</td>
</tr>
<tr>
<td>12</td>
<td>people</td>
</tr>
<tr>
<td>4</td>
<td>people, by the people, for the people</td>
</tr>
<tr>
<td>8</td>
<td>people, for the people</td>
</tr>
<tr>
<td>11</td>
<td>the people</td>
</tr>
<tr>
<td>3</td>
<td>the people, by the people, for the people</td>
</tr>
<tr>
<td>7</td>
<td>the people, for the people</td>
</tr>
</tbody>
</table>

Query: **people**

Finding matching range with additional binary searches for start and end
Bias Towards User Translation

- Cache-based models

- Language model
  → give bonus to n-grams in previous user translation

- Translation model
  → give bonus to translation options in previous user translation

- Decaying score for bonus (less recent, less relevant)

- More details: Bertoldi et al. [MT Summit 2013]
integration of translation memories
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
<table>
<thead>
<tr>
<th><strong>TM</strong></th>
<th><strong>vs.</strong></th>
<th><strong>SMT</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>used by</td>
<td></td>
<td>used by</td>
</tr>
<tr>
<td>human translator</td>
<td></td>
<td>target language information seeker</td>
</tr>
<tr>
<td>restricted domain</td>
<td></td>
<td>open domain translation</td>
</tr>
<tr>
<td>(e.g. product manual)</td>
<td></td>
<td>(e.g. news)</td>
</tr>
<tr>
<td>very repetitive content</td>
<td></td>
<td>huge diversity (esp. web)</td>
</tr>
<tr>
<td>corpus size:</td>
<td></td>
<td>corpus size:</td>
</tr>
<tr>
<td>1 million words</td>
<td></td>
<td>100-1000 million words</td>
</tr>
<tr>
<td>commercial developers</td>
<td></td>
<td>academic/commercial research</td>
</tr>
<tr>
<td>(e.g., SDL Trados)</td>
<td></td>
<td>(e.g., Google)</td>
</tr>
</tbody>
</table>
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.
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.

= 

À l’article 5, le texte du deuxième 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
texttv-2016-01-11_20161027-224944.p1a.xml
<xml translation="À l'article 21"/>
<xml translation="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.

• More compact formalism for the purposes of this presentation:

< À l’article 21 >, le texte du deuxième alinéa est supprimé. >
Two Solutions

• XML frames

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

for input

The second paragraph of Article 21 is deleted .

• Very large hierarchical rule

( The second paragraph of Article x is deleted .
; À l’ article x , le texte du deuxième alinéa est supprimé . )
Result: Acquis

The diagram shows the BLEU scores for different fuzzy match ranges and various translation methods (TM, SMT, XML, VLR). The range from 70-79% shows moderate BLEU scores, while the range from 90-99% shows significantly higher scores. The 100% range indicates perfect matches.
logging and eye tracking
• Different types of events are saved in the logging.
  – configuration and statistics
  – start and stop session
  – segment opened and closed
  – text, key strokes, and mouse events
  – scroll and resize
  – search and replace
  – suggestions loaded and suggestion chosen
  – interactive translation prediction
  – gaze and fixation from eye tracker
Logging functions

• In every event we save:
  – Type
  – In which element was produced
  – Time

• Special attributes are kept for some types of events
  – Diff of a text change
  – Current cursor position
  – Character looked at
  – Clicked UI element
  – Selected text

⇒ Full replay of user session is possible
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
## 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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score</th>
<th>Translation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/3</td>
<td>Without fail, he has been concise and accurate.</td>
<td><em>(Prediction+Options, L2a)</em></td>
</tr>
<tr>
<td>4/0</td>
<td>Without getting flustered, he showed himself to be concise and precise.</td>
<td><em>(Unassisted, L2b)</em></td>
</tr>
<tr>
<td>4/0</td>
<td>Without falling apart, he has shown himself to be concise and accurate.</td>
<td><em>(Postedit, L2c)</em></td>
</tr>
<tr>
<td>1/3</td>
<td>Unswayable, he has shown himself to be concise and to the point.</td>
<td><em>(Options, L2d)</em></td>
</tr>
<tr>
<td>0/4</td>
<td>Without showing off, he showed himself to be concise and precise.</td>
<td><em>(Prediction, L2e)</em></td>
</tr>
<tr>
<td>1/3</td>
<td>Without dismantling himself, he presented himself consistent and precise.</td>
<td><em>(Prediction+Options, L1a)</em></td>
</tr>
<tr>
<td>2/2</td>
<td>He showed himself concise and precise.</td>
<td><em>(Unassisted, L1b)</em></td>
</tr>
<tr>
<td>3/1</td>
<td>Nothing daunted, he has been concise and accurate.</td>
<td><em>(Postedit, L1c)</em></td>
</tr>
<tr>
<td>3/1</td>
<td>Without losing face, he remained focused and specific.</td>
<td><em>(Options, L1d)</em></td>
</tr>
<tr>
<td>3/1</td>
<td>Without becoming flustered, he showed himself concise and precise.</td>
<td><em>(Prediction, L1e)</em></td>
</tr>
</tbody>
</table>
Main Measure: Productivity

<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>3.3 sec/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>7.7 sec/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>3.9 sec/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>2.8 sec/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>5.2 sec/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>5.7 sec/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>3.2 sec/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>5.8 sec/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>3.4 sec/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>2.8 sec/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>4.4 sec/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 Novice Translators

L1 = native French, L2 = native English, average time per input word only typing
Unassisted Novice Translators

L1 = native French, L2 = native English, average time per input word typing, initial and final pauses
Unassisted Novice Translators

L1 = native French, L2 = native English, average time per input word typing, initial and final pauses, short, medium, and long pauses.

Most time difference on intermediate pauses.
## 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
### Origin of Characters: Native French L1b

<table>
<thead>
<tr>
<th></th>
<th>key</th>
<th>click</th>
<th>tab</th>
<th>mt</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Postedit</strong></td>
<td>18%</td>
<td>-</td>
<td>-</td>
<td>81%</td>
</tr>
<tr>
<td><strong>Options</strong></td>
<td>59%</td>
<td>40%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Prediction</strong></td>
<td>14%</td>
<td>-</td>
<td>85%</td>
<td>-</td>
</tr>
<tr>
<td><strong>Prediction+Options</strong></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
• 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 $l$

$$\text{sum}(t) = \frac{1}{Z} \sum_{p \in P, l(p) \leq t} l(p)$$
Results

![Graph showing average translation time vs. length of pauses (sec)]
Learning Effects

Users become better over time with assistance

![Graph showing learning effects](graph.png)
Learning Effects: Professional Translators

CASMACAT longitudinal study
Productivity projection as reflected in Kdur taking into account six weeks
(Kdur = user activity excluding pauses > 5 seconds)
Eye Trackers extensively used in cognitive studies of, e.g., reading behavior

- Overcomes weakness of key logger: what happens during pauses
- Fixation: where is the focus of the gaze
- Pupil dilation: indicates degree of concentration
• Problem: Accuracy and precision of gaze samples

![Eye tracker result diagram]

- ○ = target looked at
- × = eye tracker result

Good precision, poor accuracy
Good accuracy, poor precision
Gaze-to-Word Mapping

• Recorded gaze locations and fixations

Families hit with increase in cost of living
British families have to cough up an extra £31,300 a year as food in supermarkets have climbed at an alarming rate over the past year, still, making it hard for the Bank of England to cut interest rates control. To make matters worse, escalating prices are racing against healthcare professionals, who have suffered from the government below-inflation salary increases. In addition to fuel and food, elec

• Gaze-to-word mapping

Families hit with increase in cost of living
British families have to cough up an extra £31,300 a year as food in supermarkets have climbed at an alarming rate over the past year, still, making it hard for the Bank of England to cut interest rates control. To make matters worse, escalating prices are racing against healthcare professionals, who have suffered from the government below-inflation salary increases. In addition to fuel and food, elec
Logging and Eye Tracking

focus on target word (green) or source word (blue) at position $x$
Cognitive Studies: User Styles

- User style 1: Verifies translation just based on the target text, reads source text to fix it
Cognitive Studies: User Styles

- User style 2: Reads source text first, then target text
Cognitive Studies: User Styles

- User style 3: Makes corrections based on target text only
• User style 4: As style 1, but also considers previous segment for corrections
## Users and User Styles

<table>
<thead>
<tr>
<th></th>
<th>Style 1</th>
<th></th>
<th>Style 2</th>
<th></th>
<th>Style 3</th>
<th></th>
<th>Style 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>target / source-fix</td>
<td></td>
<td>source-target</td>
<td></td>
<td>target only</td>
<td></td>
<td>wider context</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>P</td>
<td>PI</td>
<td>PIA</td>
<td>P</td>
<td>PI</td>
<td>PI</td>
<td>PIA</td>
<td>P</td>
</tr>
<tr>
<td>P02</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>P03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P04</td>
<td>●</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>●</td>
<td>*</td>
<td>●</td>
</tr>
<tr>
<td>P05</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>P07</td>
<td>●</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>P08</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>P09</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

- Individual users employ different user styles
- But: consistently across different types of assistance
  (P = post-editing, PI = interactive post-editing, PIA = interactive post-editing with additional annotations)
Backtracking

- **Local backtracking**
  - **Immediate repetition**: the user immediately returns to the same segment (e.g. AAAA)
  - **Local alternation**: user switches between adjacent segments, often singly (e.g. ABAB) but also for longer stretches (e.g. ABC-ABC).
  - **Local orientation**: very brief reading of a number of segments, then returning to each one and editing them (e.g. ABCDE-ABCDE).

- **Long-distance backtracking**
  - **Long-distance alternation**: user switches between the current segment and different previous segments (e.g. JCJDJFJG)
  - **Text final backtracking**: user backtracks to specific segments after having edited all the segments at least once
  - **In-text long distance backtracking**: instances of long distance backtracking as the user proceeds in order through the text.
part III

CASMACAT workbench implementation
Components

- Javascript
- PHP
- Python
- GUI
- web server
- CAT server
- MT server
- web socket
- HTTP
- Python
- Python
- Python
• Builds on Matecat open source implementation

• Typical web application: LAMP (Linux, Apache, MySQL, PHP)

• Uses model, view, controller breakdown
Model

- Relevant data is stored in MySQL database `matecat_sandbox`

- Major database tables
  - Projects are stored in `projects`
  - They have a corresponding entry in `jobs`
  - Raw files (XLIFF) are stored in `files`
  - Segments are stored in `segments`
  - Translations of segments are stored in `segment_translations`
  - Log events are stored in `*_event`
  - etc.

- The major change from Matecat is the logging
Typical request: get information about a segment:
POST http://192.168.56.2:8000/?action=getSegments&time=1446185242727

Script index.php selects corresponding action in lib/controller
e.g., getSegmentsController.php

Response is HTML or JSON

The main action is really in the Javascript GUI public/js
- core functionality from Matecat public/js/cat.js
- CASMACAT extensions public/js/casmacat
• To a large degree middleware
• Calls external services such as
  – MT server
  – word aligner
  – interactive translation prediction
• Caches information about a sentence translation
• Google-style API to MT Server

• Python wrapper for Moses
  – basic translation request
  – includes pre and post processing pipeline
  – other functions: word alignment, incremental updating, etc.

• Uses mosesserver XMLRPC server
server.py

- Requires mosesserver to run as a service
  
  `mosesserver -config $MODELDIR/moses.ini --server-port 9010`

- Script server.py requires a lot of parameters
  - preprocessing tools (tokenizer, truecaser, etc.)
  - IP address and port
  - URL of the mosesserver API
  - etc.

- Request to the script
  
  `http://127.0.0.1:9000//translate?q=Un+test&key=0&source=xx&target=xx`

- Response
  
  ```json
  {"data": {"translations": [{"translatedText": "A test", "translatedTextRaw": "a test", "annotatedSource": "un test", "tokenization": {"src": [[0, 1], [3, 6]], "tgt": [[0, 0], [2, 5]]}}]}
  ```
Home Edition

- Moses is installed in /opt/moses

- CASMACAT is installed in /opt/casmacat
  - web server / GUI in /opt/casmacat/web-server
  - MT server (server.py) in /opt/casmacat/mt-server
  - CAT server in /opt/casmacat/cat-server
  - installation scripts in /opt/casmacat/install
  - log files in /opt/casmacat/logs

- Home Edition
  - admin web server in /opt/casmacat/admin
  - corpus data in /opt/casmacat/data
  - prototype training in /opt/casmacat/experiment
  - engines stored in /opt/casmacat-engines
Home Edition MT Engine

- Demo engine in /opt/casmacat/engines/fr-en-upload-1

- Files
  - biconcor.1
  - biconcor.1.align
  - biconcor.1.src-vcb
  - biconcor.1.tgt
  - biconcor.1.tgt-vcb
  - corpus-1.binlm.1
  - fast-align.1
  - fast-align.1.log
  - fast-align.1.parameters
  - fast-align-inverse.1
  - fast-align-inverse.1.log
  - fast-align-inverse.1.parameters
  - info
  - moses.tuned.ini.1
  - phrase-table-mmsapt.1
  - reordering-table.1.wbe-msd-bidirectional-fe.minlexr
  - RUN
  - truecase-model.1.en
  - truecase-model.1.fr

- The script RUN starts the engine
Thank You

questions?