Computer Aided Translation
Advances and Challenges

Philipp Koehn

30 October 2015
Overview

• A practical introduction: the CASMACAT workbench

• Postediting

• Types of assistance

• Logging, eye tracking and user studies

• Implementation details of the CASMACAT workbench
part I

CASMACAT workbench
CASMACAT workbench
CASMACAT Project 2011-2014

• Cognitive studies of translators leading to insights into interface design
  → better understanding of translator needs

• Workbench with novel types of assistance to human translators
  – interactive translation prediction
  – interactive editing and reviewing
  – adaptive translation models
  → better tools for translators

• Demonstration of effectiveness in field tests with professional translators
  → increased translator productivity
Architecture

- 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
  - Machine learning techniques
    - Active
    - Online
- Review
  - GUI features
    - Search & Replace bar
    - Visualization options
    - Translation options tables
    - Biconcordancer
    - E-pen
  - Confidence measures
    - 3-colour code
    - Prefix / suffix highlight
  - ST-TT alignments
    - Cursor-based
    - Mouse-based
    - Coverage-based
  - TT highlights
    - Prefix
    - Suffix
    - Last word edited
Core Modes

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

- Post-editing of MT
  - CasMaCat PEMT
    - CasMaCat MT
    - External MT
    - TMs (MateCat)
  - CasMaCat ITP
    - Floating prediction window
    - Text editing embedded
- Translation from scratch
- Review
  - Machine learning techniques
    - Active
    - Online
• Source on left, translation on right
• Context above and below
Incremental Updating
Incremental Updating
Incremental Updating

Machine Translation

Postediting

Retraining
Word Alignment

• Caret alignment (green)
• Mouse alignment (yellow)
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
Forget it. It's too risky. I'm through doing that shit.

Olvidarlo. Es demasiado arriesgado. Estoy haciendo...

You always say that. The same thing every time.

"I'm through, never again, too dangerous."
Bilingual Concordancer

abandon

**give up**

**to**

**to abandon**
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
Paraphrasing

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 TT segment

Reading ST 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
• Train MT engine on own or public data
• MT engines can be
  – switched out
  – downloaded
  – uploaded
  – shared
• With own MT engine, all CASMACAT modes are available
part II

cat methods
post-editing
Productivity Improvements

(source: Autodesk)
MT Quality and Productivity

- What is the relationship between MT Quality and Postediting Speed
- One study (English–German, news translation, non-professionals)

<table>
<thead>
<tr>
<th>System</th>
<th>Speed</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sec./wrd.</td>
<td>wrds./hr.</td>
</tr>
<tr>
<td>ONLINE-B</td>
<td>5.46</td>
<td>659</td>
</tr>
<tr>
<td>UEDIN-SYNTAX</td>
<td>5.38</td>
<td>669</td>
</tr>
<tr>
<td>UEDIN-PHRASE</td>
<td>5.45</td>
<td>661</td>
</tr>
<tr>
<td>UU</td>
<td>6.35</td>
<td>567</td>
</tr>
</tbody>
</table>
Translator Variability

- Translator differ in
  - ability to translate
  - motivation to fix minor translation

- High variance in translation time
  (again: non-professionals)

<table>
<thead>
<tr>
<th>Post-editor</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sec./wrd.</td>
</tr>
<tr>
<td>1</td>
<td>3.03</td>
</tr>
<tr>
<td>2</td>
<td>4.78</td>
</tr>
<tr>
<td>3</td>
<td>9.79</td>
</tr>
<tr>
<td>4</td>
<td>5.05</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>English–German</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>

<table>
<thead>
<tr>
<th>English–French</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
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
Word Alignment

- Caret alignment (green)
- Mouse alignment (yellow)
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
Mouse Over 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
Pakistan 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

L’intervention israélienne dans la bande de Gaza et les bombardements américains en Irak pour lutter contre les djihadistes de l’État islamique en Irak et au Levant ont également ajouté de la nervosité sur les marchés.

Israeli intervention in the Gaza Strip and the American bombing in...
confidence measures
Levels

• Machine translation engine indicates where it is likely wrong

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

• What are we predicting?
  – how useful is the translation — on a scale of (say) 1–5
  – indication if post-editing is worthwhile
  – estimation of post-editing effort
  – pin-pointing errors
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
  - human judgement of usefulness for post-editing (2012, 2014)
  - post-editing time (2013, 2014)

- Also task for word-level quality estimation (2014, 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
word level confidence
Visualization

- Highlight words less likely to be correct
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
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>
<th>Baseline Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
<td>Precision</td>
</tr>
<tr>
<td>Deletion</td>
<td>7%</td>
<td>27%</td>
<td>11%</td>
</tr>
<tr>
<td>Insertion</td>
<td>-</td>
<td>-</td>
<td>5%</td>
</tr>
<tr>
<td>Any edit</td>
<td>-</td>
<td>-</td>
<td>20%</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
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 |
Input Sentence
Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator
He | has
Input Sentence

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

Professional Translator

He has | for months
**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

![Example sentence](image)

- Show rest of sentence
Spence Green’s Lilt System

• Show alternate translation predictions

• Show alternate translations predictions with probabilities
Search for best translation creates a graph of possible translations
Prediction from Search Graph

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
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.
Refinements

• Matching Last Word
  – more important to match last word in path
  – refinement of best path: search for last word

• Case-insensitive matching

• Approximate word matching
  – lower substitution cost for words that differ by a few letters
  – implemented at letter edit distance $\leq 10\%$

• Stemmed matching
  – allow for difference in word endings (last 3 letters)
  – assumed to be morphological variation
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
Some Results

• News translation produced by post-editing MT output

• Same MT system used for simulated interactive translation prediction

<table>
<thead>
<tr>
<th>#</th>
<th>Method</th>
<th>Word Acc.</th>
<th>Letter Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Baseline</td>
<td>56.0%</td>
<td>75.2%</td>
</tr>
<tr>
<td>2</td>
<td>1 + Matching last word</td>
<td>59.0%</td>
<td>80.6%</td>
</tr>
<tr>
<td>3</td>
<td>2 + Case insensitive matching</td>
<td>58.7%</td>
<td>80.4%</td>
</tr>
<tr>
<td>4</td>
<td>2 + Approximate word matching</td>
<td>60.5%</td>
<td>80.6%</td>
</tr>
<tr>
<td>5</td>
<td>2 + Stemmed matching</td>
<td>59.4%</td>
<td>80.5%</td>
</tr>
<tr>
<td>6</td>
<td>4 + ”Desperate” word completion</td>
<td>60.5%</td>
<td>84.5%</td>
</tr>
</tbody>
</table>

• Details see Koehn [ACL, 2014]
Open Challenges

• Better metric than string edit distance to account for moves

• Retranslation or search graph matching?

• Interactive translation prediction for syntax-based models
  – syntax-based models work better for German, Chinese
  – search lattice → search forest
  – some preliminary work...

• Are neural machine translation models better at this?

⇒ Lots of interesting work in this area to be done
bilingual concordancer
Bilingual Concordancer

**abandonner**

**abandon**

- Incerc des Etats-Unis à abandonner Musharraf -- et les con American reluctance to abandon Musharraf -- together
- juridique, il a décidé d'abandonner la constitutionnalité, d' abandon la constitutionality, removed
- simplement menacé d’abandonner ses accords commerci simply threatened to abandon or never to conclude t

**give up**

- serait donc contraint d’abandonner le droit de créer son pm would be required to give up the right to develop its
- n’ était pas disposé à abandonner ses fonctions militaire was not ready to give up his military post, but a

**to**

- t ne veulent donc pas abandonner leurs prérogatives dar policy and do not want to delegate this prerogat

**to abandon**

- es tout en refusant d’abandonner son arsenal nucléaire drawal while refusing to abandon its nuclear weapons a
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>German</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windkraft (noun, feminine) (also: Windenergie)</td>
<td>wind power (noun)</td>
</tr>
<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>
</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>
</tr>
<tr>
<td>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>
</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>
</tr>
<tr>
<td>Nur wenige befürchten hingegen, dass dies auch bei erneuerbaren Energieträgern wie Biomasse oder Windkraft der Fall sein wird.</td>
<td></td>
</tr>
</tbody>
</table>

- Context shows when each translation is used
- Indication of source supports trust in translations
92 traductions de take+ .. ride dans 106 occurrences

dindons de la farce
monté a bateau
faire avoir
se fasse rouler
faire berner
se faire jouer
moquer de
faire
les a
se sont fait avoir
le public pour attirer la
a fait une balleade
nous rouler dans ce projet
nous tous
en train de monter un
bateau a la population
canadienne
tete des contribuables que
se paie le
passer une petite vite
bourrer de l'autre cote de
la chambre en
ont pris la voiture que pour
faire une balade

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 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 pourtant de 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 Transspotting

- 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

- 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

\[
\text{the first time} \\
\text{das erste mal} \\
tm:-0.56, lm:-2.81 \\
d:-0.74. all:-4.11 \\
-9.3 + \\
-4.11 = \\
-13.41
\]
Improving Rankings

- Removal of duplicates and near duplicates

<table>
<thead>
<tr>
<th>bad</th>
<th>good</th>
</tr>
</thead>
<tbody>
<tr>
<td>erupted</td>
<td>climbing</td>
</tr>
<tr>
<td>ausbrach</td>
<td>Klettern</td>
</tr>
<tr>
<td>ausbrach,</td>
<td>Bergsteigen</td>
</tr>
<tr>
<td>platzte</td>
<td>Aufstieg</td>
</tr>
<tr>
<td>Ausbruch</td>
<td>abhalten,</td>
</tr>
<tr>
<td>ausgebrochen</td>
<td>Erklimmen</td>
</tr>
<tr>
<td>ausgebrochen ist</td>
<td>beim Besteigen</td>
</tr>
</tbody>
</table>

- 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>the</td>
<td>the</td>
</tr>
<tr>
<td>us house of representatives</td>
<td>the house of representatives</td>
</tr>
<tr>
<td>adopted</td>
<td>adopted</td>
</tr>
<tr>
<td>thursday</td>
<td>thu</td>
</tr>
<tr>
<td>legally</td>
<td>the legally</td>
</tr>
<tr>
<td>calls for the withdrawal of combat troops</td>
<td>demands withdrawal of forces</td>
</tr>
<tr>
<td>us</td>
<td>us</td>
</tr>
<tr>
<td>iraq</td>
<td>iraq</td>
</tr>
<tr>
<td>in</td>
<td>in</td>
</tr>
<tr>
<td>no later than</td>
<td>not later than</td>
</tr>
<tr>
<td>the</td>
<td>of</td>
</tr>
<tr>
<td>first</td>
<td>first</td>
</tr>
<tr>
<td>april</td>
<td>april</td>
</tr>
<tr>
<td>it was</td>
<td>he was</td>
</tr>
<tr>
<td>the</td>
<td>us house</td>
</tr>
<tr>
<td>was adopted</td>
<td>adopted</td>
</tr>
<tr>
<td>thursday, the</td>
<td>a</td>
</tr>
<tr>
<td>the law</td>
<td>a law</td>
</tr>
<tr>
<td>demands withdrawal of troops</td>
<td>calls for withdrawal of forces</td>
</tr>
<tr>
<td>the</td>
<td>of</td>
</tr>
<tr>
<td>fighter</td>
<td>from</td>
</tr>
<tr>
<td>first of</td>
<td></td>
</tr>
<tr>
<td>april</td>
<td></td>
</tr>
<tr>
<td>earlier,</td>
<td></td>
</tr>
<tr>
<td>he</td>
<td>us</td>
</tr>
<tr>
<td>was</td>
<td>adopted</td>
</tr>
<tr>
<td>by</td>
<td>the</td>
</tr>
<tr>
<td>on</td>
<td>thursday’s</td>
</tr>
<tr>
<td>a</td>
<td>legally</td>
</tr>
<tr>
<td>calls for withdrawal</td>
<td>demands withdrawal of</td>
</tr>
<tr>
<td>of</td>
<td>of</td>
</tr>
<tr>
<td>the</td>
<td>from</td>
</tr>
<tr>
<td>us</td>
<td>iraq</td>
</tr>
<tr>
<td>in</td>
<td>in</td>
</tr>
<tr>
<td>the</td>
<td>not</td>
</tr>
<tr>
<td>first of</td>
<td></td>
</tr>
<tr>
<td>april</td>
<td></td>
</tr>
<tr>
<td>it was the</td>
<td></td>
</tr>
<tr>
<td>he</td>
<td></td>
</tr>
<tr>
<td>was</td>
<td></td>
</tr>
<tr>
<td>adopted</td>
<td>adopted</td>
</tr>
<tr>
<td>thursday the</td>
<td>the</td>
</tr>
<tr>
<td>legally,</td>
<td>legal</td>
</tr>
<tr>
<td>demands withdrawal of</td>
<td>calls for withdrawal</td>
</tr>
<tr>
<td>of</td>
<td>of</td>
</tr>
<tr>
<td>the</td>
<td>from</td>
</tr>
<tr>
<td>us</td>
<td>iraq</td>
</tr>
<tr>
<td>in</td>
<td>in</td>
</tr>
<tr>
<td>the</td>
<td>not</td>
</tr>
<tr>
<td>first of</td>
<td></td>
</tr>
<tr>
<td>april</td>
<td></td>
</tr>
<tr>
<td>earlier, the</td>
<td></td>
</tr>
<tr>
<td>he</td>
<td></td>
</tr>
<tr>
<td>was</td>
<td></td>
</tr>
<tr>
<td>adopted, the</td>
<td></td>
</tr>
<tr>
<td>thursday, the</td>
<td></td>
</tr>
<tr>
<td>legally,</td>
<td></td>
</tr>
<tr>
<td>demands withdrawal of</td>
<td></td>
</tr>
<tr>
<td>of</td>
<td></td>
</tr>
<tr>
<td>the</td>
<td></td>
</tr>
<tr>
<td>american</td>
<td>by the</td>
</tr>
<tr>
<td>the</td>
<td>first</td>
</tr>
<tr>
<td>of</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>بالسحب</th>
<th>القوات المقاتلة الأمريكية من العراق</th>
<th>withdrawal of</th>
<th>combat troops</th>
<th>us</th>
<th>iraq</th>
</tr>
</thead>
<tbody>
<tr>
<td>رواد</td>
<td>the fighting forces the us from iraq</td>
<td>the fighting forces</td>
<td>us</td>
<td>iraq</td>
<td></td>
</tr>
<tr>
<td>بالسحب</td>
<td>withdrawal of troops</td>
<td>fighter</td>
<td>the us</td>
<td></td>
<td></td>
</tr>
<tr>
<td>بالسحب</td>
<td>withdrawal of combat forces</td>
<td>the fighter</td>
<td>of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>بالسحب</td>
<td>withdrawal of forces the fighter</td>
<td>of</td>
<td>from</td>
<td></td>
<td></td>
</tr>
<tr>
<td>بالسحب</td>
<td>withdrawal of troops</td>
<td>the american</td>
<td>from iraq</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
No big difference — once significantly better
Monolingual Translation Triage

- Study on Russian–English (Schwartz, 2014)

- Allow monolingual translators to assess their translation
  - confident $\rightarrow$ accept the translation
  - verify $\rightarrow$ proofread by bilingual
  - partially unsure $\rightarrow$ part of translation handled by bilingual
  - completely unsure $\rightarrow$ 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
Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

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 $e'_i, e''_i, \ldots$
  ⇒ 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
  \[ \rightarrow \] create bad translations by back-translation

- **Process**
  - Train machine translation system for both directions
  - Translate test set target \[ \rightarrow \] source \[ \rightarrow \] 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
Backoff

In Domain Model
Out-of Domain Model
Look up phrase
If found, return
If not found
If found, return
Look up phrase
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
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
Suffixes

government of the people, by the people, for the people
of the people, by the people, for the people
the people, by the people, for the people
people, by the people, for the people
, by the people, for the people
by the people, for the people
the people, for the people
people, for the people
, for the people
for the people
the people
people
<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>, 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

<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>, 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
### Querying the Suffix Array

<p>| | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>9</td>
<td>6</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>4</td>
<td>8</td>
<td>11</td>
<td>3</td>
<td>7</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

, by the people, for the people
, for the people
by the people, for the people
for the people
government of the people, by the people, for the people
of the people, by the people, for the people
people
people, by the people, for the people
people, for the people
the people
the people, by the people, for the people
the people, for the people

**Query:** *people*
## Querying the Suffix Array

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

**Query:** 
*people*

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

<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>, 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

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

#### Query: **people**

**Binary search:** middle of remaining space

<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>, 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>
# Querying the Suffix Array

<table>
<thead>
<tr>
<th></th>
<th>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>2</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>3</td>
<td>the people</td>
</tr>
<tr>
<td>7</td>
<td>the people, by the people, for the people</td>
</tr>
<tr>
<td>3</td>
<td>the people, for the people</td>
</tr>
</tbody>
</table>

**Query:** people

**Binary search:** match
Querying the Suffix Array

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)
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>used by</td>
<td>used by</td>
</tr>
<tr>
<td>human translator</td>
<td>target language information seeker</td>
<td></td>
</tr>
<tr>
<td>restricted domain</td>
<td>open domain translation</td>
<td></td>
</tr>
<tr>
<td>(e.g. product manual)</td>
<td>(e.g. news)</td>
<td></td>
</tr>
<tr>
<td>very repetitive content</td>
<td>huge diversity (esp. web)</td>
<td></td>
</tr>
<tr>
<td>corpus size:</td>
<td>corpus size:</td>
<td></td>
</tr>
<tr>
<td>1 million words</td>
<td>100-1000 million words</td>
<td></td>
</tr>
<tr>
<td>commercial developers</td>
<td>academic/commercial research</td>
<td></td>
</tr>
<tr>
<td>(e.g., SDL Trados)</td>
<td>(e.g., Google)</td>
<td></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 deuxiéme 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 translation="À l’article "> 21
<xml translation=", 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.

- More compact formalism for the purposes of this presentation:

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

- XML frames

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

The second paragraph of Article 21 is deleted.

- Very large hierarchical rule

\[
( \text{The second paragraph of Article } x \text{ is deleted.} \\
; \text{À l’article } x , \text{le texte du deuxième alinéa est supprimé.} )
\]
Result: Acquis
logging and eye tracking
Logging functions

- 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>

| 1/3  | Without fail, he has been concise and accurate. (Prediction+Options, L2a) |
| 4/0  | Without getting flustered, he showed himself to be concise and precise. (Unassisted, L2b) |
| 4/0  | Without falling apart, he has shown himself to be concise and accurate. (Postedit, L2c) |
| 1/3  | Unswayable, he has shown himself to be concise and to the point. (Options, L2d) |
| 0/4  | Without showing off, he showed himself to be concise and precise. (Prediction, L2e) |
| 1/3  | Without dismantling himself, he presented himself consistent and precise. (Prediction+Options, L1a) |
| 2/2  | He showed himself concise and precise. (Unassisted, L1b) |
| 3/1  | Nothing daunted, he has been concise and accurate. (Postedit, L1c) |
| 3/1  | Without losing face, he remained focused and specific. (Options, L1d) |
| 3/1  | Without becoming flustered, he showed himself concise and precise. (Prediction, L1e) |
## 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>3.3sec/word</td>
<td>1.2s</td>
<td>2.3s</td>
<td>1.1s</td>
<td>2.4s</td>
</tr>
<tr>
<td></td>
<td>23% correct</td>
<td>-2.2s</td>
<td>-1.0s</td>
<td>-2.2s</td>
<td>-0.9s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>39%</td>
<td>45%</td>
<td>30%</td>
<td>44%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+16%</td>
<td>+22%</td>
<td>+7%</td>
<td>+21%</td>
</tr>
<tr>
<td>L1b</td>
<td>7.7sec/word</td>
<td>4.5s</td>
<td>4.5s</td>
<td>2.7s</td>
<td>4.8s</td>
</tr>
<tr>
<td></td>
<td>35% correct</td>
<td>-3.2s</td>
<td>-3.3s</td>
<td>-5.1s</td>
<td>-3.0s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48%</td>
<td>55%</td>
<td>61%</td>
<td>41%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+13%</td>
<td>+20%</td>
<td>+26%</td>
<td>+6%</td>
</tr>
<tr>
<td>L1c</td>
<td>3.9sec/word</td>
<td>1.9s</td>
<td>3.8s</td>
<td>3.1s</td>
<td>2.5s</td>
</tr>
<tr>
<td></td>
<td>50% correct</td>
<td>-2.0s</td>
<td>-0.1s</td>
<td>-0.8s</td>
<td>-1.4s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>61%</td>
<td>54%</td>
<td>64%</td>
<td>61%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+11%</td>
<td>+4%</td>
<td>+14%</td>
<td>+11%</td>
</tr>
<tr>
<td>L1d</td>
<td>2.8sec/word</td>
<td>2.0s</td>
<td>2.9s</td>
<td>2.4s</td>
<td>1.8s</td>
</tr>
<tr>
<td></td>
<td>38% correct</td>
<td>-0.7s</td>
<td>(+0.1s)</td>
<td>(-0.4s)</td>
<td>(-1.0s)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>46%</td>
<td>59%</td>
<td>37%</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+8%</td>
<td>(+21%)</td>
<td>(-1%)</td>
<td>+7%</td>
</tr>
<tr>
<td>L1e</td>
<td>5.2sec/word</td>
<td>3.9s</td>
<td>4.9s</td>
<td>3.5s</td>
<td>4.6s</td>
</tr>
<tr>
<td></td>
<td>58% correct</td>
<td>-1.3s</td>
<td>(-0.2s)</td>
<td>-1.7s</td>
<td>(-0.5s)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>64%</td>
<td>56%</td>
<td>62%</td>
<td>56%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+6%</td>
<td>(-2%)</td>
<td>+4%</td>
<td>(-2%)</td>
</tr>
<tr>
<td>L2a</td>
<td>5.7sec/word</td>
<td>1.8s</td>
<td>2.5s</td>
<td>2.7s</td>
<td>2.8s</td>
</tr>
<tr>
<td></td>
<td>16% correct</td>
<td>-3.9s</td>
<td>-3.2s</td>
<td>-3.0s</td>
<td>-2.9s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50%</td>
<td>34%</td>
<td>40%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+34%</td>
<td>+18%</td>
<td>+24%</td>
<td>+34%</td>
</tr>
<tr>
<td>L2b</td>
<td>3.2sec/word</td>
<td>2.8s</td>
<td>3.5s</td>
<td>6.0s</td>
<td>4.6s</td>
</tr>
<tr>
<td></td>
<td>64% correct</td>
<td>(-0.4s)</td>
<td>+0.3s</td>
<td>+2.8s</td>
<td>+1.4s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>56%</td>
<td>60%</td>
<td>61%</td>
<td>57%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-8%)</td>
<td>-4%</td>
<td>-3%</td>
<td>-7%</td>
</tr>
<tr>
<td>L2c</td>
<td>5.8sec/word</td>
<td>2.9s</td>
<td>4.6s</td>
<td>4.1s</td>
<td>2.7s</td>
</tr>
<tr>
<td></td>
<td>52% correct</td>
<td>-3.0s</td>
<td>(-1.2s)</td>
<td>-1.7s</td>
<td>-3.1s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53%</td>
<td>37%</td>
<td>59%</td>
<td>53%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+1%</td>
<td>(-15%)</td>
<td>+7%</td>
<td>+1%</td>
</tr>
<tr>
<td>L2d</td>
<td>3.4sec/word</td>
<td>3.1s</td>
<td>4.3s</td>
<td>3.8s</td>
<td>3.7s</td>
</tr>
<tr>
<td></td>
<td>49% correct</td>
<td>(-0.3s)</td>
<td>(+0.9s)</td>
<td>(+0.4s)</td>
<td>(+0.3s)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>49%</td>
<td>51%</td>
<td>53%</td>
<td>58%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+0%)</td>
<td>(+2%)</td>
<td>(+4%)</td>
<td>(+9%)</td>
</tr>
<tr>
<td>L2e</td>
<td>2.8sec/word</td>
<td>2.6s</td>
<td>3.5s</td>
<td>2.8s</td>
<td>3.0s</td>
</tr>
<tr>
<td></td>
<td>68% correct</td>
<td>-0.2s</td>
<td>+0.7s</td>
<td>(-0.0s)</td>
<td>+0.2s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>79%</td>
<td>59%</td>
<td>64%</td>
<td>66%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+11%</td>
<td>-9%</td>
<td>(-4%)</td>
<td>-2%</td>
</tr>
<tr>
<td>avg.</td>
<td>4.4sec/word</td>
<td>2.7s</td>
<td>3.7s</td>
<td>3.2s</td>
<td>3.3s</td>
</tr>
<tr>
<td></td>
<td>47% correct</td>
<td>-1.7s</td>
<td>-0.7s</td>
<td>-1.2s</td>
<td>-1.1s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55%</td>
<td>51%</td>
<td>54%</td>
<td>53%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+8%</td>
<td>+4%</td>
<td>+7%</td>
<td>+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>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.
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 $l$

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

Users become better over time with assistance

![Graph showing learning effects with different types of assistance]
CASMACAT longitudinal study

Productivity projection as reflected in Kdur taking into account six weeks

(Kdur = user activity excluding pauses > 5 seconds)
Eye Tracking

- 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
Eye Tracking

- Problem: Accuracy and precision of gaze samples

![Diagram showing eye tracker results with annotations for good precision, poor accuracy, good accuracy, poor precision, and target looked at.]
Gaze-to-Word Mapping

• Recorded gaze locations and fixations

Right eye gaze samples

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. To make matters worse, escalating prices are racing ahead of inflation, healthcare professionals, who have suffered from the government's below-inflation salary increases. In addition to fuel and food, elec...

Left eye gaze samples

• Gaze-to-word mapping
Logging and Eye Tracking

focus on target word (green) or source word (blue) at position $x$
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
- 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>Style 1</th>
<th>Style 2</th>
<th>Style 3</th>
<th>Style 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>target / source-fix</td>
<td>source-target</td>
<td>target only</td>
<td>wider context</td>
</tr>
<tr>
<td>P</td>
<td>PI</td>
<td>PIA</td>
<td>P</td>
</tr>
<tr>
<td>P02</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>P03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P04</td>
<td>●</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>P05</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>P07</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>P08</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>P09</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

GUI

web server

CAT server

MT server

Javascript

HTTP

web socket

PHP

Python

HTTP

Python

Python
Web Server

- 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
Controller

• 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?