Advances in Computer Aided Translation Beyond Post-Editing

Philipp Koehn

31 October 2015
Overview

• Post-editing

• Richer information
  – word alignment
  – confidence scores
  – translation option array
  – bilingual concordancer
  – paraphraser

• Interactive translation prediction

• Model adaptation

• Logging, eye tracking, and user studies

• CASMACAT Home Edition
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, Musharraf 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-Qaeda ont obtenu des résultats mitigés.

- Screenshot from CASMACAT post-editing mode (same as MATECAT)
- Source on left, translation on right / context above and below
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 (sec./wrd.)</th>
<th>Speed (wrds./hr.)</th>
<th>Metric (BLEU)</th>
<th>Metric (MANUAL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONLINE-B</td>
<td>5.46</td>
<td>659</td>
<td>20.7</td>
<td>0.637</td>
</tr>
<tr>
<td>UEDIN-SYNTAX</td>
<td>5.38</td>
<td>669</td>
<td>19.4</td>
<td>0.614</td>
</tr>
<tr>
<td>UEDIN-PHRASE</td>
<td>5.45</td>
<td>661</td>
<td>20.1</td>
<td>0.571</td>
</tr>
<tr>
<td>UU</td>
<td>6.35</td>
<td>567</td>
<td>16.1</td>
<td>0.361</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>
Overview

• Post-editing

• Richer information
  – word alignment
  – confidence scores
  – translation option array
  – bilingual concordancer
  – paraphraser

• Interactive translation prediction

• Model adaptation

• Logging, eye tracking, and user studies

• CASMACAT Home Edition
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

And on that the signs are mixed.

Y en que los indicios son desiguales.
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
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?
• 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>الزائر</td>
<td>the visitor</td>
</tr>
<tr>
<td>العلماء</td>
<td>the scholars</td>
</tr>
<tr>
<td>المكتبة</td>
<td>the library</td>
</tr>
</tbody>
</table>

### 2008:
- **defying once new president george w. bush**, which opposes the no date has been set for the
- **challenging again the new a defiant the first**, who opposes who opposes who opposes who opposes date.
- **in defiance of once again, challenging once again the president george bush, who opposed to setting any the date of the
- **in 2008, defying the again us president george w. bush**, who opposed to any the date of.

### 2008:
- **defiant once again president george bush**, who opposes who opposes who opposes who opposes date.
- **agains the new a defiant the first**, who opposes who opposes who opposes who opposes date.
- **in defiance of once again, challenging once again the president george bush, who opposed to setting any the date of the
- **in 2008, defying the again us president george w. bush**, who opposed to any the date of.

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

<table>
<thead>
<tr>
<th><strong>رسحب</strong></th>
<th><strong>القوات</strong></th>
<th><strong>المقاتلة</strong></th>
<th><strong>الاميركية</strong></th>
<th><strong>العراق</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>withdrawal of combat troops</td>
<td>us</td>
<td>iraq</td>
<td>the fighting forces the us from iraq</td>
<td></td>
</tr>
<tr>
<td>withdrawal of the fighting forces us from iraq</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>withdrawal of troops fighter</td>
<td>the us</td>
<td>of</td>
<td>from iraq</td>
<td></td>
</tr>
<tr>
<td>withdrawal of combat forces</td>
<td>the fighter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>withdrawal of forces the fighter</td>
<td></td>
<td></td>
<td>from iraq in</td>
<td></td>
</tr>
<tr>
<td>withdrawal of troops</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>withdrawal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the withdrawal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the withdrawal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Bilingual Concordancer

abandon

abandon

abandon

abandon

abandon

abandon

abandon

abandon

abandon

abandon

abandon

abandon

abandon

abandon
<table>
<thead>
<tr>
<th>92 traductions de take+ .. ride dans 106 occurrences</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Expression</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>take+ .. ride</td>
<td>3</td>
</tr>
</tbody>
</table>

| dindons de la farce | 4 |
| monté un bateau    | 3 |
| faire avoir        | 3 |
| se faire rouler    | 2 |
| fait bernier       | 2 |
| se faire jouer     | 2 |
| moqués de          | 2 |
| fait              | 2 |
| les a              | 2 |
| se sont fait avoir | 2 |
| le public pour attirer la | 1 |
| a fait une balade  | 1 |
| nous rouler dans ce projet nous tous | 1 |
| en train de monter un bateau à la population canadienne | 1 |
| tête des contribuables que se paie le | 1 |
| passer une petite vite | 1 |
| bourrer de l'autre côté de la chambre en | 1 |
| ont pris la voiture que pour faire une balade | 1 |

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 pâlissant 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.
machine translation

maschinelle Übersetzung
Maschinenübersetzung

translation machine
Übersetzungsmaschine

See also:
machine - Maschine
Gerät, Automat, Anlage, Apparat
bearbeiten, maschinel herstellen, spanen, zerspanen
Übersetzung, Translation, Übersetzen
Verschiebung, Sprachübersetzung

External sources (not reviewed)
The implementing provisions applicable to the machine translation system would have to be established by the Select Committee [...]

Die Durchführungsbestimmungen für das System der maschinellen Übersetzung müssten vom engeren Ausschuss des EPO-Verwaltungsrats [...]
Verification of Terminology

- Translation of German *Windkraft*

- Context shows when each translation is used
- Indication of source supports trust in translations
Paraphrasing

- User marks part of translation
- Clicks on paraphrasing button
- Alternative translations appear

However, the European Central Bank (ECB) asked about it in a report on virtual currencies published in October.
Overview

- Post-editing

- Richer information
  - word alignment
  - confidence scores
  - translation option array
  - bilingual concordancer
  - paraphraser

- Interactive translation prediction

- Model adaptation

- Logging, eye tracking, and user studies

- CASMACAT Home Edition
Interactive Translation Prediction

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."
Shade Off Translated

- Word alignment visualization for interactive translation prediction
- 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.
Visualization

• Show $n$ next words

![Example of visualization]

• Show rest of sentence
Spence Green’s Lilt System

- Show alternate translation predictions

![Example translation predictions]

- Show alternate translations predictions with probabilities

![Another example of translation predictions]
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
We can predict the optimal completion (according to the model)
Overview

• Post-editing

• Richer information
  – word alignment
  – confidence scores
  – translation option array
  – bilingual concordancer
  – paraphraser

• Interactive translation prediction

• **Model adaptation** Interactive translation prediction

• Logging, eye tracking, and user studies

• CASMACAT Home Edition
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

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

- Multiple models $\rightarrow$ multiple feature functions

Use both
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
Incremental Updating
Incremental Updating

Machine Translation

Postediting

Retraining
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
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)
Overview

• Post-editing

• Richer information
  – word alignment
  – confidence scores
  – translation option array
  – bilingual concordancer
  – paraphraser

• Interactive translation prediction

• Model adaptation

• **Logging, eye tracking, and user studies**

• CASMACAT Home Edition
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?
Input:  *Au premier semestre, l’avionneur a livré 97 avions.*
Output:  *The manufacturer has delivered 97 planes during the first half.*
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
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>Activity</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**
In this section, we report the origin of characters when using L1b.

### User: L1b

<table>
<thead>
<tr>
<th></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
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
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
• 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
• Local backtracking
  – immediate repetition
  – local alternation
  – local orientation

• Long-distance backtracking
  – long-distance alternation
  – text final backtracking
  – in-text long distance backtracking
Overview

• Post-editing

• Richer information
  – word alignment
  – confidence scores
  – translation option array
  – bilingual concordancer
  – paraphraser

• Interactive translation prediction

• Model adaptation

• Logging, eye tracking, and user studies

• CASMACAT Home Edition
- European research project 2011-2014

- All describe methods implemented in CASMACAT workbench
  - builds on MATECAT open source implementation
  - typical web application: LAMP (Linux, Apache, MySQL, PHP)
  - uses model, view, controller breakdown

- Workbench freely available at http://www.casmacat.eu/
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

### Build New Prototype

<table>
<thead>
<tr>
<th>Name</th>
<th>Segments</th>
<th>Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Central Bank</td>
<td>102,980</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>European Medicines Agency</td>
<td>372,824</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>EU Bookshop</td>
<td>3,618,897</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>European Constitution</td>
<td>6,667</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>European Parliament</td>
<td>1,260,689</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>KDE4</td>
<td>126,141</td>
<td>OPUS uploaded</td>
</tr>
<tr>
<td>KDE4 (el-en_GB)</td>
<td>125,537</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>Open Subtitles</td>
<td>220,445</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>Open Subtitles 2011</td>
<td>10,693,456</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>Open Subtitles 2012</td>
<td>12,984,773</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>Open Subtitles 2013</td>
<td>14,626,890</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>South-East European Times</td>
<td>165,532</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>South-East European Times v2</td>
<td>224,808</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>SPC</td>
<td>7,035</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>Tatoeba</td>
<td>2,469</td>
<td>OPUS upload</td>
</tr>
<tr>
<td>DGT-Translation Memory</td>
<td>3,016,402</td>
<td>JRC upload</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Use</th>
<th>ID</th>
<th>Name</th>
<th>Segments</th>
<th>Uploaded</th>
</tr>
</thead>
<tbody>
<tr>
<td>KDE4</td>
<td>all</td>
<td>1</td>
<td>KDE4</td>
<td>126141</td>
<td>21:39:27</td>
</tr>
</tbody>
</table>

- Previous setting: none
- Tuning set: KDE4
- Evaluation set: KDE4
- Name of the prototype: build
Thank You

questions?