MateCat: an Open Source CAT Tool for MT post-editing

Marcello Federico, Nicola Bertoldi
FBK Trento, Italy

Marco Trombetti, Alessandro Cattelan
Translated srl, Rome, Italy
Tutorial outline

- The research project (MF, 20’)
- MT technology advances (MF, 20’)
- The MateCat tool and how it works (AC, 30’)
- Use cases (MT 20’)
- Break (30’)
- Installing the tool (NB, 30’)
- Interactive session (All, 40’)
- Conclusions and future plans (MT+MF, 20’)

Tutorial outline

- The research project (MF, 20’)
- MT technology advances (MF, 20’)
- The MateCat tool and how it works (AC, 30’)
- Use cases (MT 20’)
- Break (30’)
- Installing the tool (NB, 30’)
- Interactive session (All, 40’)
- Conclusions and future plans (MT+MF, 20’)
The Research Project
OVERVIEW

- Motivation
- CAT scenario
- Project goals
- Roadmap
- Field tests
Motivation

- **Human translation (HT)**
  worldwide demand for translation services has accelerated, due to globalization and growth of the Information Society

- **Gap between MT and HT**
  MT has improved significantly but independently from HT
  MT research has not directly addressed how to improve HT
  Most professional translators barely use MT

- **The unavoidable adoption of MT**
  Post-editing experiments have shown great promise
  The challenge is how to smoothly integrate MT and HT!
Scenario

All our translators got a CAT tool!
Translation Project

I’m the project manager
Scenario

- **Computer assisted translation (CAT)**
  - dominant technology: **CAT tools**
  - supporting many file formats
  - spell checking, terminology, dictionaries, …
  - *translation memory (TM), machine translation* (MT).

- **CAT Tool: text editor for translators**
  - text is split into *segments*
  - translation suggestions of segments and/or words
Commercial CAT Tool
Translation Memory

- Incrementally **stores** translated segments
- **Retrieves** perfect or **fuzzy matches** of segments
- **Shared** among translators working on the same project
- **A TM models the style and terminology of the customer**
Translation Memory

When does it help?
- on repetitive texts, such as technical manuals
- when more translators work on the same project

How does it help?
- accelerates translation process
- ensures consistency across different translators

Limitations
- number of useful matches is generally small (5-10%)
Translation decomposed into a sequence of rule applications

**Statistical MT:**
- searches *optimal* sequence of translation rules
- translation rules learned from *parallel texts*
- *statistical model* defined over rules and fitted to data
- rule sequences generate *linear or hierarchical* structures
Machine Translation

When does it help?
- language pairs supported by large parallel data
- translation directions between close languages
- training data represent well task data

How does it help?
- provides good draft to post-edit
- avoids translating easy/repetitive fragments

Limitations
- translations may lack of global coherence
- may produce bad output that causes waste of time
# MateCat Project

<table>
<thead>
<tr>
<th>Project Acronym</th>
<th>MateCat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Title</td>
<td>Machine Translation Enhanced Computer Assisted Translation</td>
</tr>
<tr>
<td>Funding scheme</td>
<td>STREP FP7-ICT-2011-7 Grant # 287688</td>
</tr>
<tr>
<td>Duration</td>
<td>36 months, 1 Nov 2011 - 30 Oct 2014</td>
</tr>
<tr>
<td>Consortium</td>
<td>Fondazione Bruno Kessler - Italy Universite Le Mans - France The University of Edinburgh – United Kingdom Translated srl - Italy</td>
</tr>
<tr>
<td>Effort</td>
<td>349 person-month (= 9.7 full-time-equivalent/year)</td>
</tr>
<tr>
<td>Budget</td>
<td>3,368K €</td>
</tr>
<tr>
<td>Funding</td>
<td>2,650K €</td>
</tr>
</tbody>
</table>
MateCat Project

Strategic
- Seamless integration of machine and human translation
- Enhance productivity and user experience with CAT

Research
- New MT functionality
  - self-tuning, user-adaptive, informative

Technology
- Web-based CAT tool integrating new MT features
- Full open source solution (Moses, IRSTLM, …)
Why another CAT tool?

Existing tools
- Deploy generic and static MT engines
- Difficult to integrate/evaluate new MT functionality

MateCat Tool
- Enterprise level CAT tool for real use
- *Interoperability* across MT and TM engines
- Supporting new MT functionalities
- Supporting document formats, tags
- Automatic collection of usage statistics
Tutorial outline

- The research project (MF, 20’)
- **MT technology advances** (MF, 20’)
- The MateCat tool and how it works (AC, 30’)
- Use cases (MT 20’)
- Break (30’)
- Installing the tool (NB, 30’)
- Interactive session (All, 40’)
- Conclusions and future plans (MT+MF, 20’)

**Ø**
MT Technology Advances
OVERVIEW

- Self-tuning MT
- User-adaptive MT
- Informative MT
Statistical MT

Search for the optimal (=best scoring) translation using phrase-pairs

E’ necessario incoraggiare tale mobilità pur garantendo la sicurezza dei percorsi professionali.

Freedom of movement must be encouraged, while ensuring that careers paths are safeguarded.

How phrase-based SMT works:

- **Search steps**: select source segment, translate, attach to target
- **Decoder**: explores the search space and scores translations hypotheses
- **Scores**: linear combination of feature functions
- **Features**: phrase-pairs, target n-grams, relative phrase-movement
- Features and linear-combination weights are machine learned
MT quality depends on

- *distance* of source and target languages,
- *amount* of training data,
- ... *and closeness* between training and task data
Domain adaptation provides means to effectively integrate small amounts of task data in the training process!
Domain Adaptation
   ... before translation project starts (baseline system)

Self-tuning MT [Project adaptation]
   .... incrementally during the lifetime of a translation project

User-adaptive MT [Online adaptation]
   ... instantly after each sentence is post-edited.
Application scenario

- Post-editing
- Suggestion from TM
- Suggestion from SMT
1. MT stands for machine translation.
2. Our research aims to make it more useful to translators.
3. Your document has been saved in file MaeCat.xlif.
4. All software will be integrated into a Web-based CAT tool.

1. MT e' l'acronimo di traduzione automatica.
2. La nostra ricerca mira a renderla più utile per i traduttori.
3. La tua ricerca viene salvata nel file MaeCat.xlif.
4. Tutto il software verrà integrato in un CAT tool basato su Web.
Different modalities for combining or merging models
Data selection
- Cross-entropy difference (Axelrod et. al, 2012)		
  After project has started (source + post-edits)

TM Adaptation
- Fill-up & back-off (Bisazza et al., 2011, Niehues and Waibel, 2011)

LM Adaptation
- Linear interpolation

M. Cettolo, N. Bertoldi, M. Federico, C. Servan, H. Schwenk,
“Translation Project Adaptation for MT-Enhanced CAT”,
Project Adaptation: test protocol

Warm-up session (WU)
First 20% of doc
Post-edit domain adapted MT

Field-test session (FT)
Remaining 80% of doc
**Test:** domain-adapted MT vs. project-adapted MT
Simulate post-edits with reference translations

<table>
<thead>
<tr>
<th>pair</th>
<th>MT engine</th>
<th>IT domain</th>
<th>Legal domain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BLEU</td>
<td>TER</td>
</tr>
<tr>
<td>en→it</td>
<td>DA</td>
<td>55.3</td>
<td>29.2</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>57.5</td>
<td>26.3</td>
</tr>
<tr>
<td>en→fr</td>
<td>DA</td>
<td>41.3</td>
<td>38.3</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>41.4</td>
<td>37.9</td>
</tr>
<tr>
<td>en→es</td>
<td>DA</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>en→de</td>
<td>DA</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Key performance indicators:

- TTE - Time to edit (words/hour)
- PEE - Post-editing effort (human TER)
## Project Adaptation: field tests

Data collection and logging for in-depth analysis

<table>
<thead>
<tr>
<th>Job 984 - Editing Log</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slowest 5,000 segments by time-to-edit</td>
</tr>
</tbody>
</table>

### Summary

<table>
<thead>
<tr>
<th>Words</th>
<th>Avg Secs per Word</th>
<th>% of MT</th>
<th>% of TM</th>
<th>Total Time-to-edit</th>
<th>Avg P.Effort %</th>
<th>% of words in too SLOW cells</th>
<th>% of words in too FAST cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>2541</td>
<td>3</td>
<td>56%</td>
<td>4%</td>
<td>03:18:25</td>
<td>33%</td>
<td>1%</td>
<td>0%</td>
</tr>
</tbody>
</table>

### Editing Details

<table>
<thead>
<tr>
<th>Job ID</th>
<th>Segment ID</th>
<th>Words</th>
<th>Suggestion source</th>
<th>Match percentage</th>
<th>Time-to-edit</th>
<th>Post-editing effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>984</td>
<td>358032</td>
<td>26</td>
<td>Machine Translation</td>
<td>56%</td>
<td>10m:58s</td>
<td>20%</td>
</tr>
</tbody>
</table>

- **Segments**: You can move a volume to a new I/O group to balance the workload across the I/O group without stopping host activity to the volumes.
- **Suggestions**: You can monitor information on compression usage to:
  - “Machine Translation; 15: 86%; 198145; 44%: To monitor system-wide compression savings and capacity, 984; 580390; "Machine Translation": 15: 86%; 198145; 44%: To monitor system-wide compression savings and capacity, 984; 580465; "Machine Translation": 12: 86%; 193919; 33%: The system can send SNMP messages that notify personnel.”

- **Data**: Mirrored, compressed, and thin-provisioned volumes all enable the wizard used in Fibre Channel hosts can be mapped.
- **Time**: “Planning for compression in pre-existing installations 18455: "Machine Translation": 37: 86%; 204045; "Email notifications: The Call Home feature transmits operation.”

---

*Note: The table and text are placeholders and are not actual data or information.*
Project Adaptation: field tests

<table>
<thead>
<tr>
<th>domain</th>
<th>user</th>
<th>TTE (sec/word)</th>
<th>PEE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>warm-up</td>
<td>field-test</td>
<td>p-value</td>
</tr>
<tr>
<td>IT</td>
<td>t1</td>
<td>4.70</td>
<td>3.36</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>t2</td>
<td>2.26</td>
<td>2.47</td>
<td>0.220</td>
</tr>
<tr>
<td></td>
<td>t3</td>
<td>3.17</td>
<td>3.11</td>
<td>0.450</td>
</tr>
<tr>
<td></td>
<td>t4</td>
<td>4.77</td>
<td>3.64</td>
<td>0.006</td>
</tr>
<tr>
<td>Legal</td>
<td>t1</td>
<td>5.20</td>
<td>5.63</td>
<td>0.222</td>
</tr>
<tr>
<td></td>
<td>t2</td>
<td>5.42</td>
<td>3.92</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>t3</td>
<td>5.86</td>
<td>4.32</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>t4</td>
<td>6.60</td>
<td>3.73</td>
<td>0.000</td>
</tr>
</tbody>
</table>

English to Italian direction - 8 professional translators

Real working conditions - MateCat tool

Post-edits: 97-98% from MT, 2-3% from TM suggestions
1. MT stands for machine translation.
2. Our research aims to make it more useful to translators.
3. Our MT technology will be seamlessly integrated into a Web-based CAT tool.
4. All software will be

MateCat Tool

User Feedback

<table>
<thead>
<tr>
<th>SRC</th>
<th>MT stands for machine translation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT</td>
<td>MT sta per traduzione automatica.</td>
</tr>
<tr>
<td>USR</td>
<td>MT 'e l'acronimo di traduzione automatica.</td>
</tr>
</tbody>
</table>
Online model adaptation

Generative Cache Model
Extract and store phrases and n-grams in cache models: features with decaying scores

Discriminative re-ranking
Learn phrases and n-grams found in the post-edits and use them to re-rank MT n-best outputs

Online model adaptation

BLEU of baseline vs. user-adaptive system on increasing portions of two documents of English-Italian IT domain
1. MT stands for machine translation.
2. Our research aims to make it more useful to translators.
3. Our MT technology will be seamlessly integrated into a Web-based CAT tool.

Informative MT

MT and TM suggestions

Filtering and ranking

MT Server

MT decoder

QE engine

TM Server

Our research aims to make it more useful to translators.
Informative MT

L'objectif de cet atelier de travail est de développer des méthodes et des architectures de système d'adapter les systèmes de traduction automatique statistique (TAS).

<table>
<thead>
<tr>
<th>Source: MT</th>
<th>2013-08-28</th>
<th>MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traduction automatique</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collecteurs/tuyauteries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source: Anonymous</td>
<td>2012-08-26</td>
<td>18%</td>
</tr>
<tr>
<td>Source: TRANSLATED</td>
<td>0000-00-00</td>
<td>15%</td>
</tr>
</tbody>
</table>
What is a poor and what is a good MT suggestions?

**Good translation:** \( \text{sim}(TGT,PE) \approx 1 \)

**Wrong translation:** \( \text{sim}(TGT,PE) \approx \text{sim}(TGT,RT) \)

Then we can label training data[PE,MT,RF] with a classifier into good and poor MT examples!
Automatic data labeling

WMT-12 subjective threshold
0.7

~0.4

best empirical threshold

Rewritings (-1) (higher HTER)

Post-editions (+1) (lower HTER)

Learning Algorithms

- Learning algorithms for:
  - regression: predict real values [HTER]
  - classification: predict labels [GOOD-BAD]
  - ranking: predict the rank of a set of items

- Not clear evidence that one works better than the others

- Conventional methods work in batch mode: training -> test

- Main problems to face:
  - achieving a “readable” model
  - over-fitting small samples of data with large feature sets
**QE in practice**

Large difference in performance between the ideal and real conditions!

<table>
<thead>
<tr>
<th>Document</th>
<th>Post-editor</th>
<th>SMT System</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>Test</td>
<td>Training</td>
<td>Test</td>
</tr>
<tr>
<td>Real</td>
<td>Doc A</td>
<td>Alice</td>
<td>Bob</td>
</tr>
<tr>
<td>...</td>
<td>Doc A</td>
<td>Alice</td>
<td>Bob</td>
</tr>
<tr>
<td>...</td>
<td>Doc A</td>
<td>Alice</td>
<td>Alice</td>
</tr>
<tr>
<td>Ideal</td>
<td>Doc A</td>
<td>Alice</td>
<td>Alice</td>
</tr>
</tbody>
</table>
We need to adaptive QE!

No, we need to multi-task learning!


Summary

Integration of HT and MT introduces **new:**

- **operating conditions** for MT
  - incremental adaptation on batches of translations
  - online adaptation from user feedback
- **functional requirements** for MT
  - 5-7 seconds latency to pre-fetch next translation
  - **real-time** online adaptation and quality estimation
- **evaluation issues** for MT
  - simulated translation sessions (with references)
  - field tests comparing different translation conditions
Conclusions

- Seamless integration of human and machine translation is probably the most relevant and promising challenge for the field of machine translation.
- Adaptive systems lower the operating cost of MT and improve utility and usability of technology.
- MateCat has delivered project and on-line adaptive MT.
- MT software available under the Moses distribution!
- We are ready for an online demonstration now ....
Tutorial outline

- The research project (MF, 20’)
- MT technology advances (MF, 20’)
- The MateCat tool and how it works (AC, 30’)
- Use cases (MT 20’)
- Break (30’)
- Installing the tool (NB, 30’)
- Interactive session (All, 40’)
- Conclusions and future plans (MT+MF, 20’)

Tutorial outline
MateCat: an Open Source CAT Tool for MT Post-Editing

How to install the tool

Nicola Bertoldi - FBK
Tutorial outline

- The research project (MF, 20’)
- MT technology advances (MF, 20’)
- The MateCat tool and how it works (AC, 30’)
- Use cases (MT 20’)
- Break (30’)
- Installing the tool (NB, 30’)
- Interactive session (All, 40’)
- Conclusions and future plans (MT+MF, 20’)


Outline

• Tool architecture
  – CAT tool
  – MySQL server
  – TM and MT server

• CAT tool
  – Installation and basic configuration

• MT server
  – Installation and basic configuration

• Advanced configuration
Tool architecture

Cat Tool

Web GUI

Suggestion Proxy

Project loader

XLIFF importer/exporter

MT server

TM server

MySQL server

Matecat DB

Ajax communication
JSON return

interchangeable
MySQL DB server

- MySQL settings

<table>
<thead>
<tr>
<th>URI</th>
<th>mysql.server.url</th>
</tr>
</thead>
<tbody>
<tr>
<td>username</td>
<td>matecat</td>
</tr>
<tr>
<td>password</td>
<td>matecat01</td>
</tr>
</tbody>
</table>
TM server

• MyMemory-compliant REST APIs:
  - http://api.mymemory.translated.net/get
  - http://api.mymemory.translated.net/set

Matecat Tool installation
TM server

http://api.mymemory.translated.net/get

Mandatory attributes:

- **q**: text to translate
- **langpair**: source and target languages (en|it, ISO 639-1)
- **user**: name for identification
- **key**: password for identification
TM server

http://api.mymemory.translated.net/set

Mandatory attributes:

seg  sentence to add in the source language
tra  sentence to add in the target language
langpair  source and target languages  (en|it, ISO 639-1)
user  name for identification
key  password for identification
MT server

- Google-compliant REST API (version 2)
  
  http://my.mtserver.url:8080/translate

- Additional API, mimicking MyMemory “set”
  
  http://my.mtserver.url:8080/update
MT server

http://my.mtserver.url:8080/translate

Mandatory attributes:

q text to translate
source source language (ISO 639-1)
target target language (ISO 639-1)
key password for identification
MT server

http://my.mtserver.url:8080/update

Mandatory attributes:

- **segment**: sentence to add in the source language
- **translation**: sentence to add in the target language
- **source**: source language (ISO 639-1)
- **target**: target language (ISO 639-1)
- **key**: password for identification
CAT tool
CAT tool installation

• Guidelines and requirements:
  http://docs.matecat.com/installation-guide

• Installation steps:
  1. Install git and clone the repository
  2. **Initialize the database**
  3. Create the virtual host
  4. Install the virtual host
  5. **Create and customize CAT tool configuration**
  6. Configure memory-cached location
CAT tool installation

Initialize the database

```
$> cd /MATECAT/cattool/lib/model

$> gedit matecat.sql
```

```sql
INSERT INTO `engines` VALUES
(1,'MyMemory (All Pairs)', 'TM', 'MyMemory', 'http://api.mymemory.translated.net', 'get', 'set', 'delete', NULL, '1', 0);

INSERT INTO `engines` VALUES
(2,'MT server','MT','En-It MT server for Legal ', 'http://my.mtserver.url:8080', 'translate', 'update', NULL, NULL, '2', 14);
```
CAT tool installation

Create the Matecat DB

```
$> mysql -u root -p root_pw < matecat.sql
```

Handle carefully!
Use for the first installation only!

- The CAT tool is linked to one specific DB
- One TM and several MT server can be set
- Everything can be modified at any time (via mysql)
CAT tool installation

Test

$> mysql -u matecat -p matecat01

mysql> show databases;

+---------------------+
<table>
<thead>
<tr>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>information_schema</td>
</tr>
<tr>
<td>matecat</td>
</tr>
<tr>
<td>test</td>
</tr>
</tbody>
</table>
+---------------------+
CAT tool installation

Test

```
mysql> use matecat;
mysql> select * from engines;
```

```
+----+---------+---+---------------------------------+-----+
| id | name    | type | description                      |     |
|    |         |      |                                 |     |
| ...|         |      |                                 |     |
+----+---------+---+---------------------------------+-----+
|  1 | MyMemory | TM   | MyMemory                          |     |
|    |          |      | http://mymemory.translated.net/api |     |
|    |          |      |                                  |     |
|  2 | MT server | MT   | En-It MT server                   |     |
|    |          |      | http://my.mtserver.url:8080       |     |
```

Matecat Tool installation
CAT tool installation

Set Apache2 Web Server

$> cd /MATECAT/cattool/INSTALL
$> cp matecat-vhost-sample matecat-vhost

$> gedit matecat-vhost

ServerName my.matecat.tool
ServerAdmin admin@matecat.tool

@@@path@@@ /MATECAT/cattool
$> cd /MATECAT/cattool/inc
$> cp config.inc.sample.php config.inc.php

$> gedit config.inc.php

```php
self::$DB_SERVER = "mysql.server.url"
self::$DB_DATABASE = "matecat"
self::$DB_USER = "matecat"
self::$DB_PASS = "matecat01"
```
CAT tool

Open in Chrome

http://my.matecat.tool

Without MT support
MT server

Matecat Tool installation
• asynchronous translation requests
• text processing and translation annotation
• Moses server can run remotely
• Moses engine runs locally
• Word aligner runs locally:
  – pivot-based, onlineMgiza++
Non-adaptive MT server
Download example models

```
$> cd /MATECAT
$> wget
    www.matecat.com/download/sample_models.zip
```

Install example models

```
$> unzip sample_models.zip
$> cd models
$> cp template_moses.ini moses.ini
```
Configure Moses

$> gedit moses.ini

Change all occurrences of @@PATH@@ to the current location of the models

Change parameters according to your preferences
Moses configuration

Test Moses

$> ${MOSES_ROOT}/bin/moses -f models/moses.ini

European Parliament

Enter a source text according to models

Parlamento europeo

Expected output
MT server installation

Operating systems
• Linux (Ubuntu, RedHat)
• Mac OSx (10.6 or higher)

Third–party software:
• Moses, featuring XMLRPC, Boost
• Python 2.7 (or higher)
• Perl 5.10 (or higher)
• Bash 3.2 (or higher)
• word-aligner (onlineMGIZA++), for adaptive version only
MT server installation

Download:

```bash
$> cd /MATECAT
$> wget
    www.matecat.com/download/mtserver.zip
$> tar xzf mtserver.zip
```

non-adaptive

Main directory of MT server
MT server installation

Configure Moses server

`$> cd /MATECAT/mtserver`

`$> cp template_server.config server.config`
MT server installation

```bash
$> edit server.config

MOSES_ROOT=/MOSES
MOSES_MODELS=/MATECAT/models
MOSES_URL=my.mosesserver.url
MOSES_PORT=7777
MOSES_LOG=/MATECAT/mtserver/log

XMLRPC_ROOT=/XMLRPC
```

If not static-linked

```bash
$> source server.config
```
MT server installation

Start Moses server

`$> python_server/start-mosesserver.sh`

Defined parameters (per moses.ini or switch)

...  ...

Listening on port 7777

Waiting for input from client

Expected output
MT server installation

Test Moses server

```bash
$> gedit python_server/testing/moses_client.py
```

```python
text = u"European Parliament"
```

From another shell

```bash
$> source server.config
$> python_server/testing/moses_client.py
```

Parlamento europeo

Source text according to models

Expected output

non-adaptive
Configure MT server

$> edit server.config

MTSERVER_ROOT=/MATECAT/mtserver
MTSERVER_URL=0.0.0.0
MTSERVER_PORT=8080
MTSERVER_SRCLNG=en
MTSERVER_TGTLNG=it

accepts queries from
Full path
Port
non-adaptive
MT server installation

Start MT server

$> python_server/start-mtsserver.sh

loading external source processors ...
loading external target processors ...
...
[date:time] ENGINE Serving on my.mtserver.url:8080
...

Waiting for input from client

Expected output
MT server installation

Test MT server

$> curl --data q='European Parliament.'
   --data source='en' --data target='it'
   --data key='DUMMY'
   http://my.matecat.tool:8080/translate

{ "data":
   "translations":
   [ { "sourceText": "European Parliament.",
       "translatedText": "Parlamento europeo.",
       "tokenization": { "src": [[0,7],[9,18],[19,19]],
                      "tgt": [[0,9],[11,17],[18,18]]} } ]

Expected output in JSON format

non-adaptive
Source text according to models
Adaptive MT server
Moses configuration

Configure Moses

```bash
$> cd models
$> cp template_moses-adaptive.ini moses-adaptive.ini
$> gedit moses-adaptive.ini
```

Change all occurrences of `@@@PATH@@@`
to the current location of the models

Change parameters according to your preferences
Moses configuration

Test Moses

$> ${MOSES_ROOT}/bin/moses -f models/moses.ini

European Parliament

Enter a source text according to models

Parlamento europeo

Expected output

Matecat Tool installation
Moses configuration

Test Moses

```
$> ${MOSES_ROOT}/bin/moses -f models/moses-adaptive.ini
```

```
<dlt type="cbtm" id="MYCBTM0"
cbtm="Parliament|||parlamento">
<dlt type="cblm" id="MYCBLM0"
cblm="||parlamento">
European Parliament
```

```
parlamento europeo
```
Feature extraction configuration

Download models for feature extraction (based on MGIZA++)

```bash
$> cd /MATECAT
$> wget www.matecat.com/download/sample_updater.zip
```

Install models

```bash
$> unzip sample_updater.zip
$> cd updater
$> cp template_updater.ini updater.ini
```
Feature extraction configuration

Configure feature extraction

$> gedit updater.ini

mgiza_path = /MGIZA
extractor_path = /MOSES/bin/extract

Change all occurrences of
@@@SRC2TRG@@@, @@@TRG2SRC@@@, @@@PATH@@@
according to the languages and the models

Change parameters
according to your preferences
Feature extraction configuration

Change all occurrences of `@@@PATH@@@` according to the models

```bash
$> gedit SRC2TRG_gizacfg.online
$> gedit TRG2SRC_gizacfg.online
```
MT server installation

Download:

```
$> cd /MATECAT
$> wget www.matecat.com/download/mtserver-adaptive.zip
$> tar xzf mtserver-adaptive.zip
```
Configure MT server

```bash
$> cd /MATECAT/mtserver-adaptive
$> cp template_server-adaptive.config server-adaptive.config
```
MT server installation

Configure Moses

$> edit server-adaptive.config

MOSES_ROOT=/MOSES
MOSES_MODELS=/MATECAT/models
MOSES_THREADS=2
MOSES_CONFIG=/MATECAT/models/moses-adaptive.ini
MOSES_LOG=/MATECAT/mts_server-adaptive/log

$> source server-adaptive.config
Configure feature extraction

$> edit server-adaptive.config

```
UPDATER_MODELS=/MATECAT/updater
UPDATER_CONFIG=/MATECAT/updater/updater.ini
```

$> source server-adaptive.config
Configure MT server

$> edit server-adaptive.config

$> source server-adaptive.config

MTSERVER_ROOT=/MATECAT/mtserver
MTSERVER_URL=0.0.0.0
MTSERVER_PORT=8080
MTSERVER_SRCLNG=en
MTSERVER_TGTLNG=it
MT server installation

Start MT server

```bash
$> SERVER/start-mtserver-adaptive.sh
```

...  

```
[date:time] ENGINE Serving on 0.0.0.0:8080
[date:time] ENGINE Bus STARTED
...  
```

**Waiting for input from client**

**Expected output**
Test MT server (translate)

```bash
$> curl --data q='European Parliament.' --data source='en' --data target='it' --data key='DUMMY'
http://my.matecat.tool:8080/translate
```

```json
{
  "data": {
    "translations": [{
      "segmentID": "0000",
      "translatedText": "parlamento europeo.",
      "systemName": "system_adaptive",
      "phraseAlignment": [[0-1], [0, 1]], ...
    }
  }
}
```
Test MT server (update)

```bash
$> curl
    --data segment='European Parliament.'
    --data translation='Parlamento Europeo.'
    --data source='en' -data target='it'
    --data key='DUMMY'
http://my.matecat.tool:8080/update
```

Expected output in JSON format

```json
{"data":
{"code": "0",
 "systemName": "system_adaptive",
 "string": "OK", ...}
}
```
Open in Chrome

http://my.matecat.tool

Enjoy!