



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





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

SDL Translation Management System - VDEVOTETMS06 - / Inbox / Demo Environment / SDL Product Overview (Translation | EN-US > DE) - [SDL Product Overview.docx]

Back Refresh Help Info Print Log Out Save Revert Submit Reject QA Check Override

Preview Task Comments Return To Inbox TM

SDL Studio Online 2011

File Home Segment Format Tools Actions

Confirm Translation Cut Undo Copy Redo Paste Find and replace Concordance search Go to segment Previous segment Next segment

Document Edit Search

Display: All segments Containing:

When new content is written and submitted for translation SDL TMS automatically checks the content against previously translated content using the latest patented technology and advanced linguistic processing.

Segment ID	Source Text	Match %	Target Text
1	When new content is written and submitted for translation SDL TMS automatically checks the content against previously translated content using the latest patented technology and advanced linguistic processing.	100%	Wenn neue Inhalte automatisch überprüft werden, überprüft die Software die Inhalte mit der neuesten Sprachverarbeitungstechnologie und den bereits übersetzten Inhalten.
5900	Enables corporations to centralise all multilingual assets into a centralised repository.	100%	SDL TMS ermöglicht es Unternehmen, alle mehrsprachigen Inhalte in einer zentralisierten Datenbank zu speichern.
5901	When new content is written and submitted for translation SDL TMS automatically checks the content against previously translated content using the latest patented technology and advanced linguistic processing.	100%	Wenn neue Inhalte automatisch überprüft werden, überprüft die Software die Inhalte mit der neuesten Sprachverarbeitungstechnologie und den bereits übersetzten Inhalten.
5902	Any content matched is delivered back translated, whilst new content requiring translation is automatically delivered down into the translation supply chain for human translation.	100%	Bereits übersetzte Inhalte werden automatisch ausgegeben, neu zu übersetzende Inhalte werden in den normalen Übersetzungsprozess gegeben.
5903	For more information about SDL TMS please visit our translation management section.	100%	Weitere Informationen über SDL TMS finden Sie in der Rubrik „Translation Management“.
5904	SDL Knowledge-based Translation System (SDL KbTS™)		SDL Knowledge-based Translation System (SDL KbTS™)
5905	Provides high-quality translations, accelerated time-to-market and reduced total cost for the world's leading brands.	82%	SDL KbTS™ liefert führenden Unternehmen weltweit qualitativ hochwertige Übersetzungen, beschleunigt die Time-to-Market und ermöglicht eine Reduzierung der Gesamtkosten.
5906	The power of the solution lies in the combination of sophisticated machine translation technology with other translation automation	100%	Der Vorteil der Lösung liegt in der Kombination hochentwickelter maschineller Übersetzungstechnologie mit weiteren automatisierten

DemoSequence

Translation Results Concordance Search Comments Term Recognition

Search text: [] Search

TU(s): 11239 INS

Local intranet | Protected Mode: Off 100%



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%)



Machine Translation

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

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



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')



MT Technology Advances



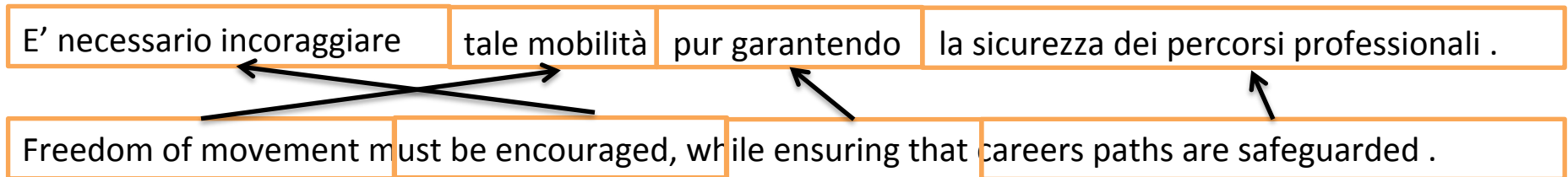
OVERVIEW

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



Statistical MT

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

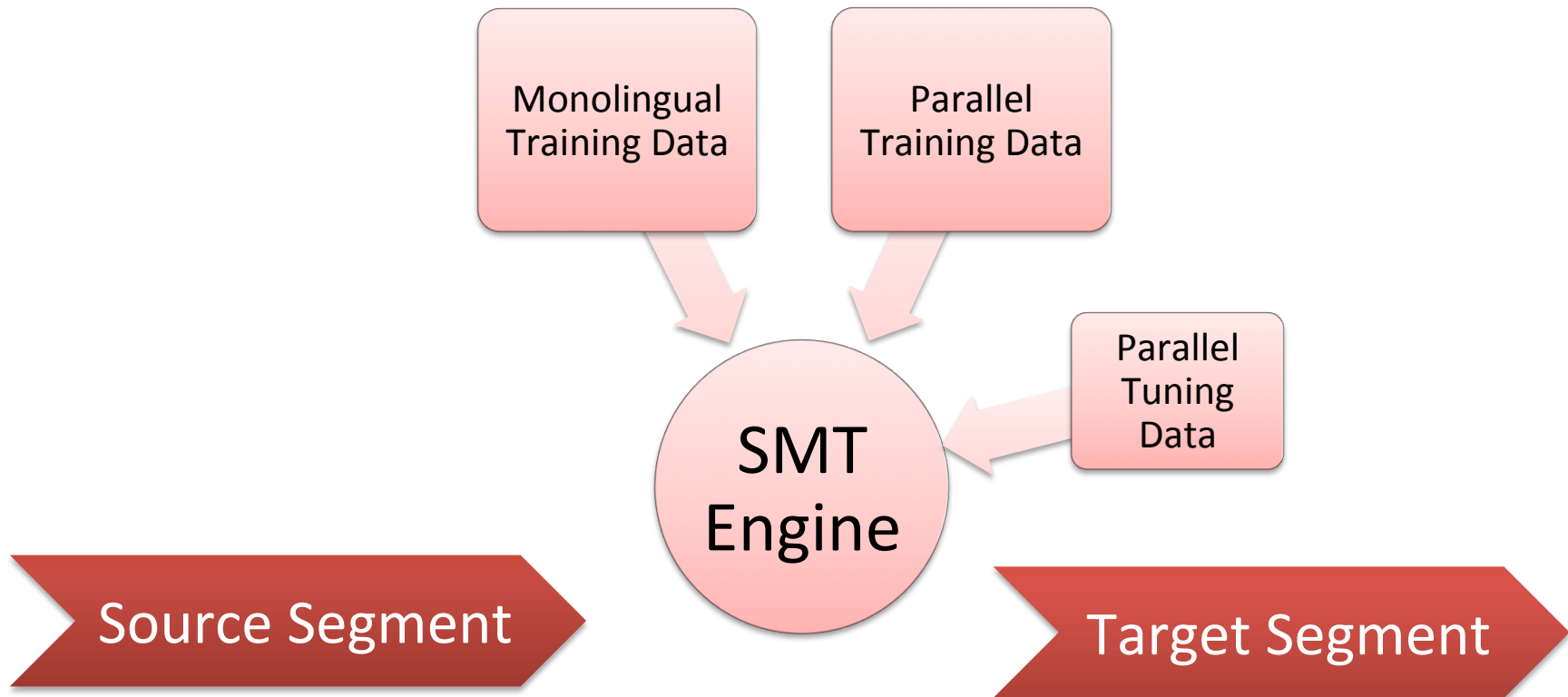


How phrase-based SMT works:

- **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



Statistical MT

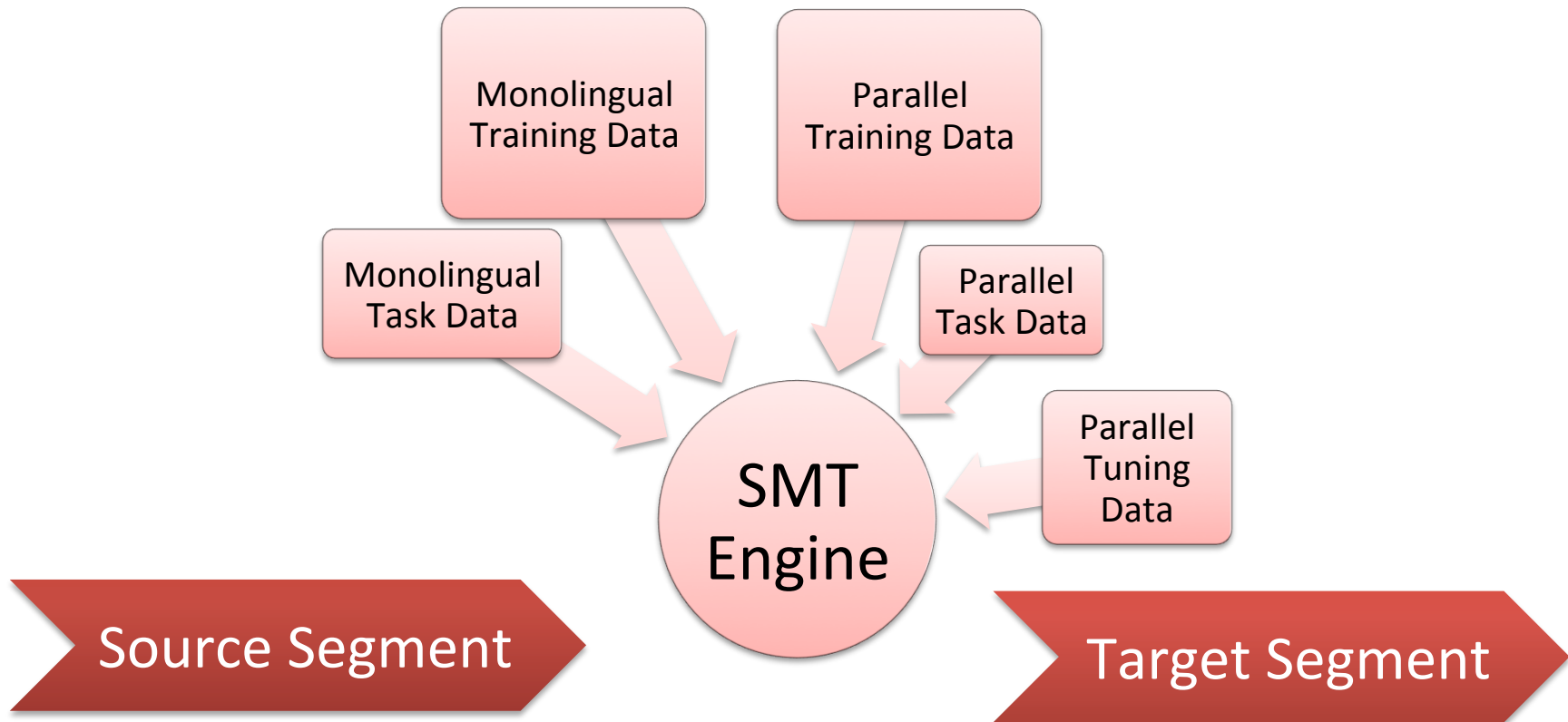


MT quality depends on

- *distance* of source and target languages,
- *amount* of training data,
- ... *and closeness* between training and task data



Domain Adaptation in SMT



Domain adaptation provides means to effectively integrate **small** amounts of task data in the training process!



MateCat Use Case

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

Translate - PROJ-359BF7E47E19CC9E903E2484864C5926 - 352

http://demo.matecat.com/translate/PROJ-359BF7E47E19CC9E903E2484864C5926/en-it/361-sldfjw322d

Reader Google

HLT 2011 MateCat Google Docs Wolfram|Alpha Reviews MTM FBK SSL VPN Intranet 2.0 | FBK

Translate - PROJ-359BF7E47E19...

matecat Jobs List > PROJ-359BF7E47E19CC9E903E2484864C5926 (352 - EN)

100%

197959 A CAT Tool for Your Business. Simple. Web-Based

Semplice. Web-based.

T→ DRAFT TRANSLATED

Suggestion from TM

Post-editing

Suggestion from SMT

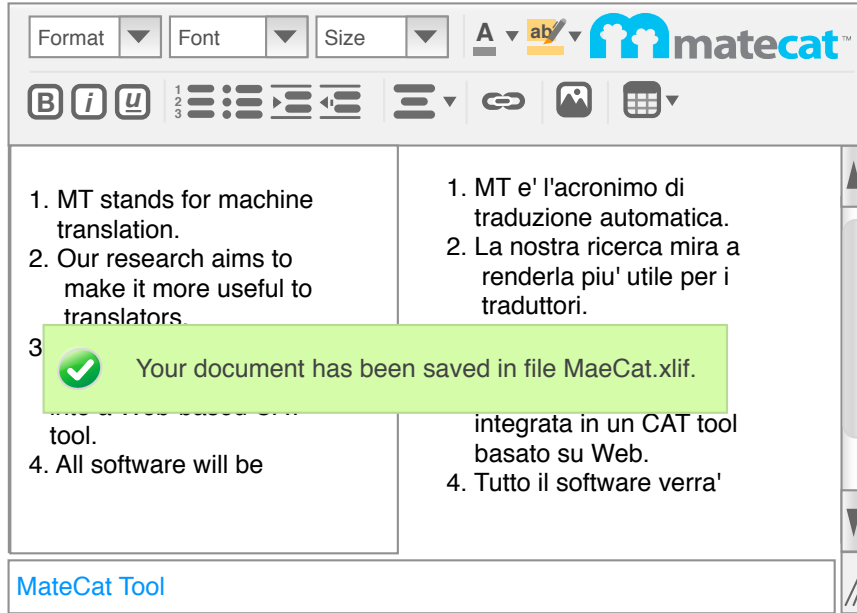
Translation matches

A CAT Tool for Your Business. Simple. Web-Based	Un CAT Tool per il tuo lavoro. Semplice. Web-based.	Source: MT 2012-09-13 MT
A CAT Tool for Your Business. Simple. Web-Based	Uno strumento CAT per il tuo business. Semplice. Web-Based	Source: rprosser 2006-09-19 31%
Tools for making your business competitive	Strumenti per rendere competitiva l'impresa	Source: rprosser 2006-09-19 31%

197960 MateCat integrates Statistical Machine Translation and Collaborative Translation Memories, within the Human Translation workflow. MateCat increases the productivity of professional translators and enhances their work experience with MT.

MateCat integra la traduzione automatica statistica e le memorie di traduzione collaborativa, all'interno del flusso di lavoro di traduzione umana. MateCat aumenta la produttività dei traduttori professionisti e migliora la loro esperienza di lavoro con MT.

Progress: Eq. words: -- Translated: --% Approved: --% Words last hour: -- Expected completion: -- Hours Project Statistics



**Self-tuning
MT**

Translated documents

MT stands for machine translation. automatica. Our research aims to make it more useful to translators. Our MT technology will be seamlessly integrated into a Web-based CAT tool.	MT e' l'acronimo di traduzione automatica. La nostra ricerca mira a renderla piu' utile per i traduttori. La nostra tecnologia di traduzione automatica sara' perfettamente integrata in un CAT tool basato su Web.
---	---

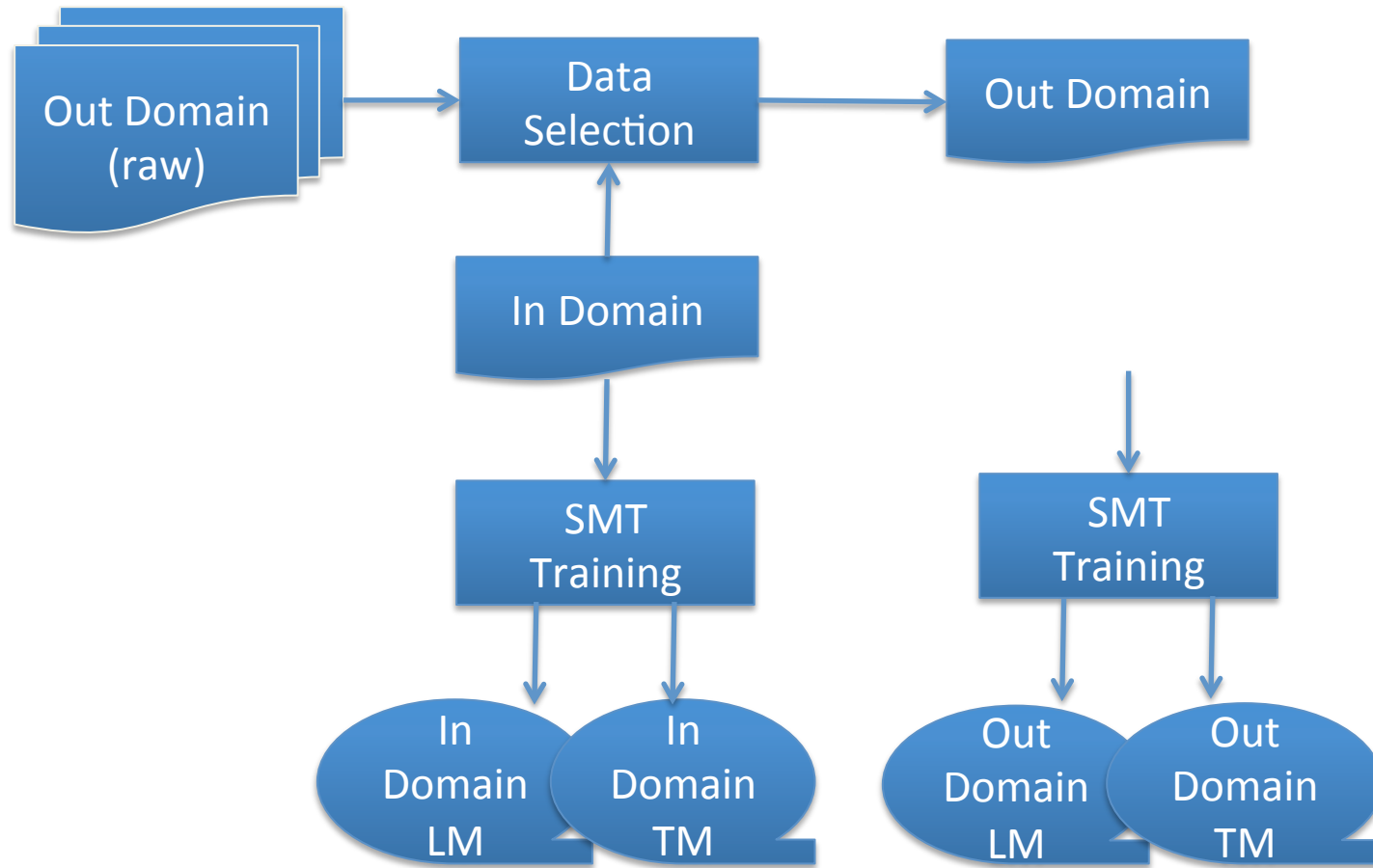
MT Server

Domain Adaptation

Models



Project adaptation



Different modalities for combining or merging models



Project Adaptation



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”,
MT Journal, 2014.



Project Adaptation: test protocol

Source	Score
Source: Anonymous 2012-09-05	100%
Source: MT 2012-09-13	51%
Source: rprosser 2006-09-10	31%

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

Source	Score
Source: Anonymous 2012-09-05	100%
Source: MT 2012-09-13	51%
Source: rprosser 2006-09-10	31%

Field-test session (FT)
Remaining 80% of doc
Test: domain-adapted MT vs.
project-adapted MT



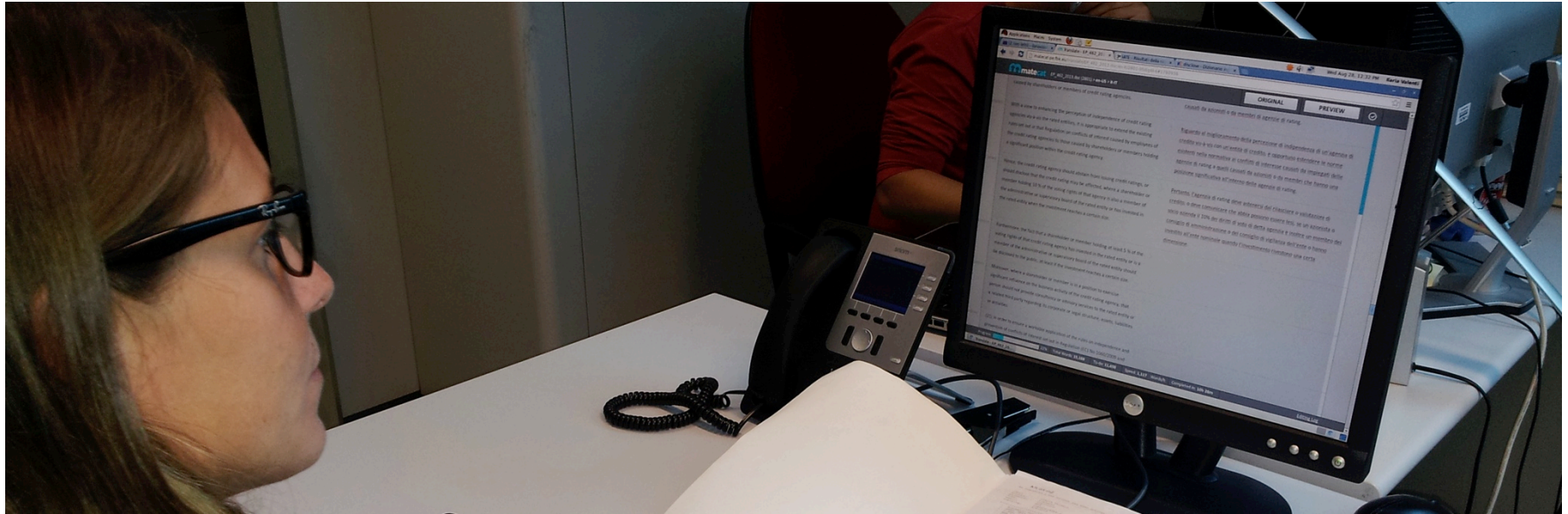
Project Adaptation: lab tests

Simulate post-edits with reference translations

pair	MT engine	IT domain			Legal domain		
		BLEU	TER	GTM	BLEU	TER	GTM
en→it	DA	55.3	29.2	77.8	31.0	53.1	61.8
	PA	57.5	26.3	78.6	35.0	49.1	64.6
en→fr	DA	41.3	38.3	69.5	33.9	52.2	63.0
	PA	41.4	37.9	69.9	36.4	49.1	65.1
en→es	DA	—	—	—	35.5	50.7	65.7
	PA	—	—	—	36.4	50.2	65.6
en→de	DA	—	—	—	19.3	65.0	52.6
	PA	—	—	—	20.1	64.7	52.8



Project Adaptation: field tests



Key performance indicators:

- TTE - Time to edit (words/hour)
- PEE - Post-editing effort (human TER)



Project Adaptation: field tests

matecat < Back to Translation Export All Data in CSV

Job 984 - Editing Log

Slowest 5.000 segments by time-to-edit

Summary

Words	Avg Secs per Word	% of MT	% of TM	Total Time-to-edit	Avg PE Effort %	% of words in too SLOW edits	% of words in too FAST edits
2641	5s	96%	4%	03h:38m:59s	33%	1%	0%

Editing Details

Secs/Word	Job ID	Segment ID	Words	Suggestion source	Match percentage	Time-to-edit	Post-editing effort
163.8	984	580414	26	Machine Translation	86%	10m:58s	20%

Segment You can move a volume to a new I/O group to balance the workload across the in the system without stopping host activity to the volumes.

Suggestion È possibile spostare un volume di I / E in un nuovo gruppo di bilanciare il carico di lavoro tra i nel sistema host senza dover arrestare l'attività per i volumi.

Translation È possibile spostare un volume in un nuovo gruppo I/O per bilanciare il carico di lavoro nel sistema senza arrestare l'attività dell'host sui volumi.

Diff View È possibile spostare un volume ~~di I / E~~ in un nuovo gruppo ~~di I/O~~ per bilanciare il carico di lavoro ~~tra i~~ nel sistema ~~host~~ senza ~~dover~~ arrestare l'attività ~~per i~~ dell'host sui volumi.

Secs/Word	Job ID	Segment ID	Words	Suggestion source	Match percentage	Time-to-edit	Post-editing effort
28.8	984	580432	27	Machine Translation	86%	12m:57s	31%

Data collection and logging for in-depth analysis

```
percentage;Time-to-edit;Post-editing effort;Segment;Suggestion;Tra  
";"20%";"You can move a volume to a new I/O group to balance the  
";"31%";"Mirrored, compressed, and thin-provisioned volumes as we  
";"65%";"It is common to observe high compression ratios in datab  
";"47%";"Establish paths to I/O groups on hosts After the system  
";"46%";"Select the preferred node in that I/O group that the hos  
";"45%";"The wizard only changes the I/O group for the volume and  
";"27%";"For volumes mapped to Fibre Channel hosts, the wizard al  
8 984;580424;"Machine Translation";50;"86%";"312327";"27%";"Like thin-provisioned volumes, compressed volumes have v  
9 984;580379;"Machine Translation";11;"86%";"252606";"66%";"Like thin-provisioned volumes, compressed volumes have v  
10 984;580411;"Machine Translation";37;"86%";"240454";"41%";"Planning for compression in pre-existing installations T  
11 984;580389;"Machine Translation";20;"86%";"235892";"15%";"You can also monitor information on compression usage to  
12 984;580449;"Machine Translation";22;"86%";"229242";"47%";"If a recommended service action is active, these events  
13 984;580419;"Machine Translation";28;"86%";"220055";"43%";"You can also use this wizard to move volumes to another  
14 984;580399;"Machine Translation";32;"86%";"217623";"30%";"After compression is applied to stored data, the require  
15 984;580461;"Machine Translation";29;"86%";"216299";"45%";"Email notifications The Call Home feature transmits oper  
16 984;580390;"Machine Translation";15;"86%";"198145";"44%";"To monitor system-wide compression savings and capacity,  
17 984;580465;"Machine Translation";12;"86%";"193919";"33%";"The system can send SNMP messages that notify personnel
```



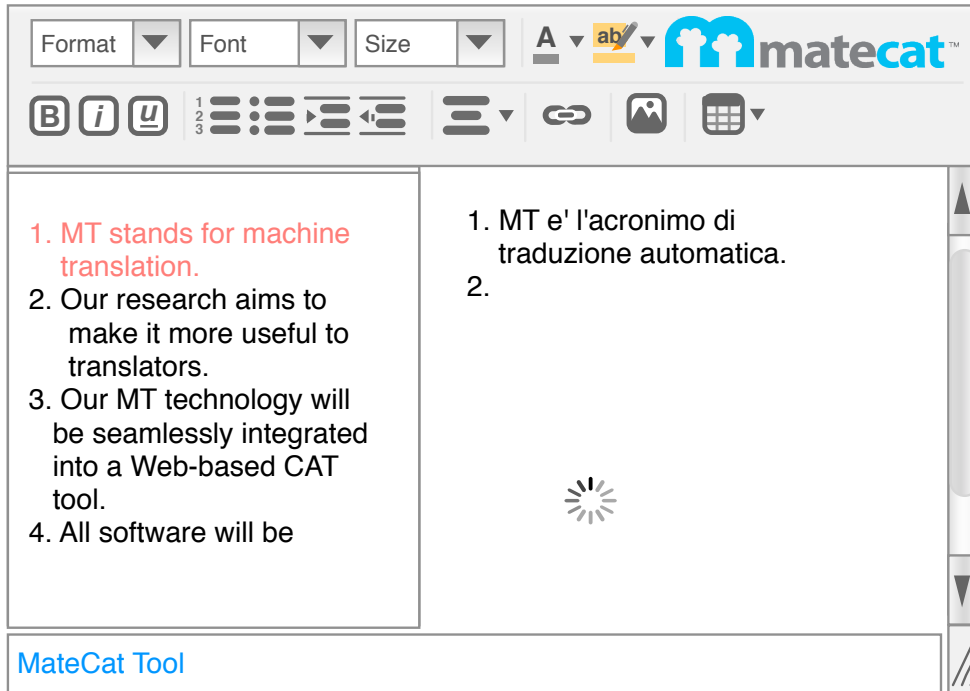

Project Adaptation: field tests

domain	user	TTE (sec/word)				PEE			
		warm-up	field-test	p-value	Δ	warm-up	field-test	p-value	Δ
IT	t1	4.70	3.36	0.001	28.51%	34.27	30.99	0.060	9.57%
	t2	2.26	2.47	0.220	-9.29%	38.50	39.52	0.330	-2.65%
	t3	3.17	3.11	0.450	1.89%	32.53	30.17	0.133	7.25%
	t4	4.77	3.64	0.006	23.69%	32.22	28.44	0.040	11.73%
Legal	t1	5.20	5.63	0.222	-8.27%	26.47	24.57	0.212	7.18%
	t2	5.42	3.92	0.002	27.68%	29.11	26.25	0.140	9.82%
	t3	5.86	4.32	0.000	26.28%	35.65	34.11	0.247	4.32%
	t4	6.60	3.73	0.000	43.48%	22.72	18.07	0.011	20.47%

English to Italian direction - 8 professional translators

Real working conditions - MateCat tool

Post-edits: 97-98% from MT, 2-3% from TM suggestions



User-adaptive MT

User Feedback

SRC	MT stands for machine translation.
MT	MT sta per traduzione automatica.
USR	MT e' l'acronimo di traduzione automatica.

MT Server

On-Line Learning

Models



Online model adaptation



Generative Cache Model

*Extract and store phrases and n-grams
in cache models: features with decaying scores*

Discriminative re-ranking

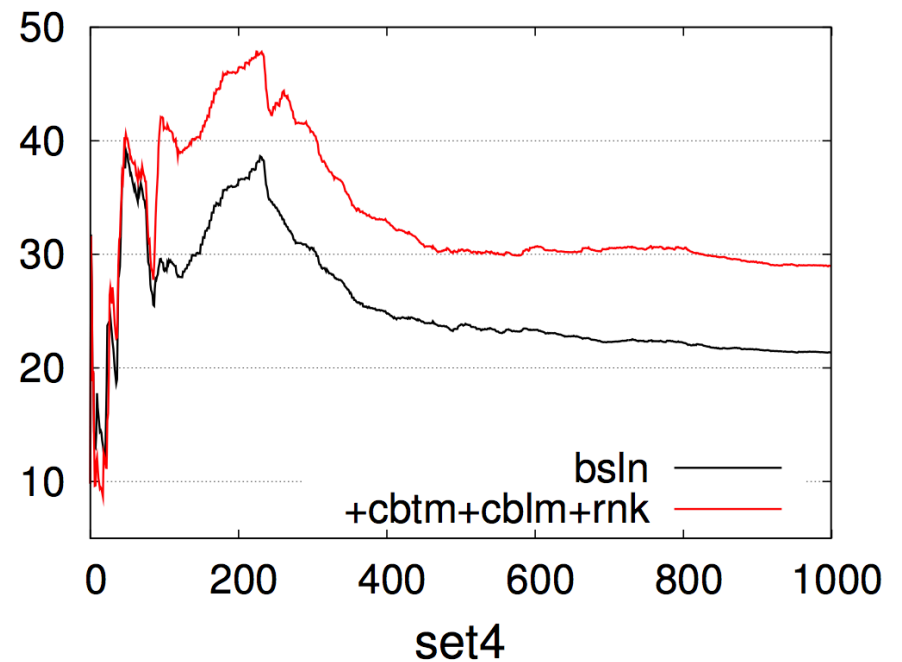
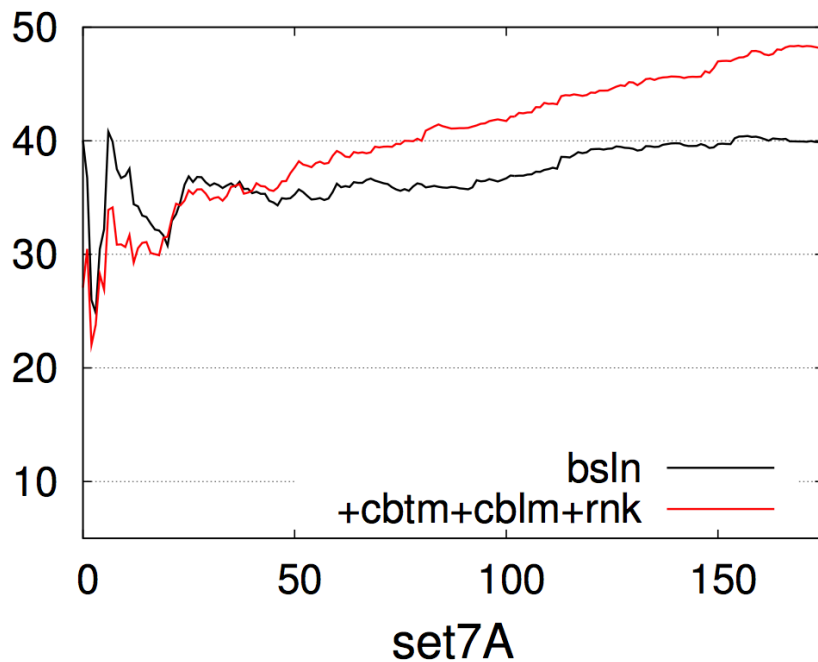
*Learn phrases and n-grams found in the post-edits and
use them to re-rank MT n-best outputs*

N. Bertoldi, P. Simianer, M. Cettolo, K. Wäschle, M. Federico,
S.Riezler “Online Adaptation to Post-Edits for Phrase-Based
Statistical Machine Translation”, *MT Journal*, 2014.



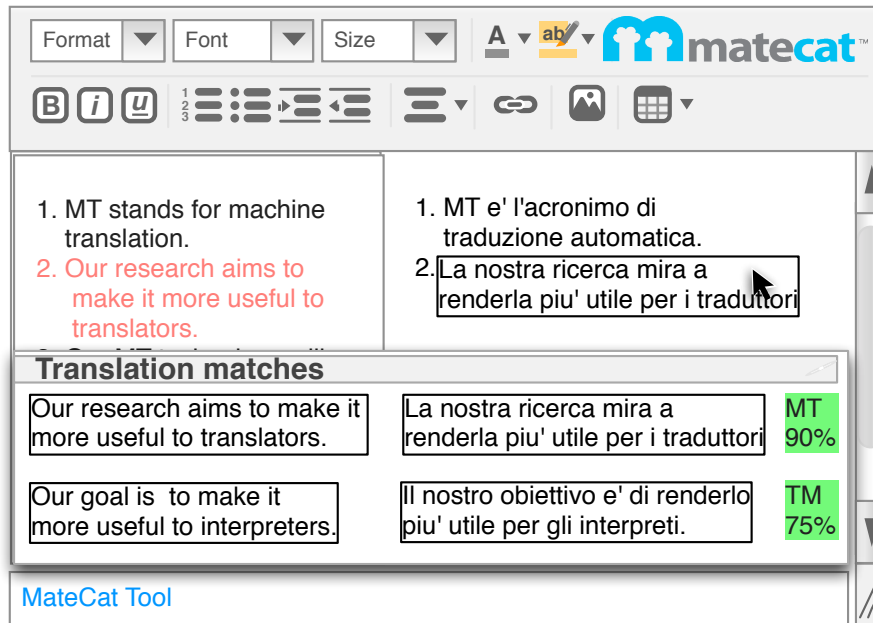
Online model adaptation

BLEU of baseline vs. user-adaptive system on increasing portions of two documents of English-Italian IT domain





Informative MT



Source

SRC Our research aims to make it more useful to translators.

Informative MT

MT and TM suggestions

Filtering and ranking

MT Server

MT decoder

QE engine

TM Server



Informative MT

Methods and system
adaptation (SMT) systems.

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).



TRANSLATED

System architectures to adapt

L'objectif de cet atelier de travail est de développer des méthodes et des architectures de système d'adapter les systèmes de traduction automatique s

Traduction automatique

Collecteurs/tuyauteries

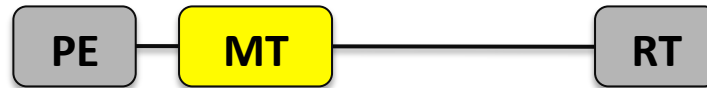
Source: MT	2013-08-28	MT
Source: Anonymous	2012-08-26	18%
Source: TRANSLATED	0000-00-00	15%



What is a poor and what is a good MT suggestions?



Good translation: $sim(TGT, PE) \approx 1$



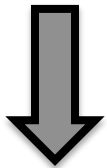
Wrong translation: $sim(TGT, PE) \approx sim(TGT, RT)$

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



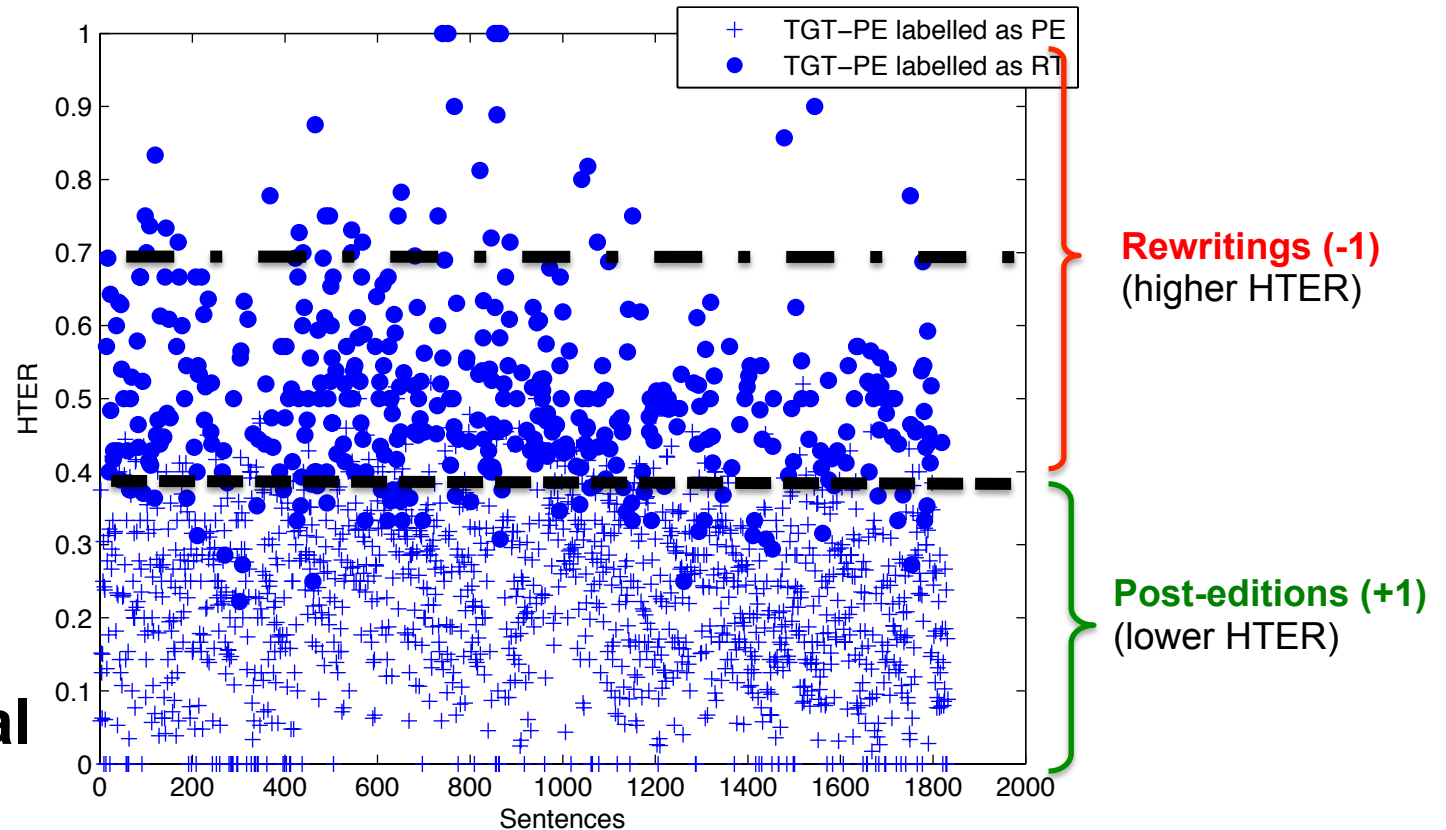
Automatic data labeling

WMT-12
subjective
threshold
0.7



~0.4

best empirical
threshold



M.Turchi, M. Negri, M. Federico, "Data-driven Annotation of Binary MT Quality Estimation Corpora Based on Human Post-editions", MT Journal 2014.



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!

	Document		Post-editor		SMT System		MAE
	Training	Test	Training	Test	Training	Test	
Real	Doc A	Doc B	Alice	Bob	Sys1	Sys2	0.2240
...	Doc A	Doc B	Alice	Bob	<i>Sys1</i>	<i>Sys1</i>	0.1542
...	Doc A	Doc B	<i>Alice</i>	<i>Alice</i>	<i>Sys1</i>	<i>Sys1</i>	0.1396
Ideal	<i>Doc A</i>	<i>Doc A</i>	<i>Alice</i>	<i>Alice</i>	<i>Sys1</i>	<i>Sys1</i>	0.1074



We need to adaptive QE!

No, we need to multi-task learning!

M. Turchi, A. Anastasopoulos, J.G. Camargo de Souza and M. Negri “**Adaptive Quality Estimation for Machine Translation**”, Proc. ACL 2014.

J. G. Camargo de Souza, M. Turchi and M. Negri “**Predicting Machine Translation Quality Estimation Across Domains**”, Proc. Coling 2014.



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')



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

How to install the tool

Nicola Bertoldi - FBK

MateCAT is a STREP project funded by the EC (grant 287688) under the 7th Framework Programme (FP7-ICT-2011-7).



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')

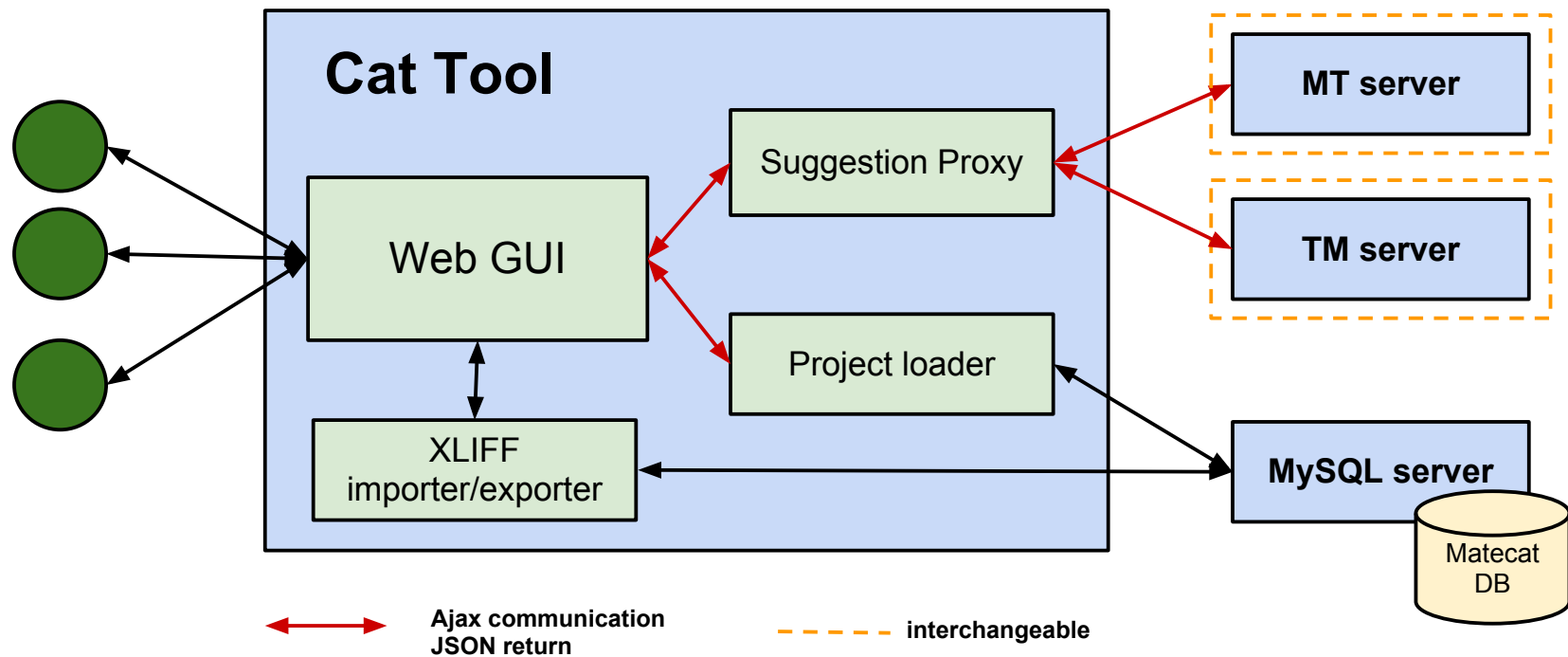


Outline

- 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





MySQL DB server

- MySQL settings

URI	<code>mysql.server.url</code>
username	<code>matecat</code>
password	<code>matecat01</code>



TM server

- MyMemory-compliant REST APIs:

<http://mymemory.translated.net/doc/spec.php>

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

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



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

Matecat Tool installation



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
```

```
INSERT INTO `engines` VALUES
(1,'MyMemory (All Pairs)', 'TM', 'MyMemory',
'http://memory.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);
```

Id

Type

URI



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;
```

mysql shell

```
+-----+
| Database |
+-----+
| information_schema |
| matecat |
| test |
+-----+
```



CAT tool installation

Test

mysql shell

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

```
+-----+----- ... -----+-----+  
| id | name          | type | description          |  
base_url          | ... | penalty|  
+-----+----- ... -----+-----+  
| 1 | MyMemory     | TM   | MyMemory           ... |  
http://mymemory.translated.net/api | ... | 0 |  
| 2 | MT server    | MT   | En-It MT server ... |  
http://my.mtserver.url:8080      | ... | 14 |  
+-----+----- ... -----+-----+
```



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
```

URL
of CAT tool

```
@@@path@@@ /MATECAT/cattool
```




CAT tool configuration

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

```
$> gedit config.inc.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



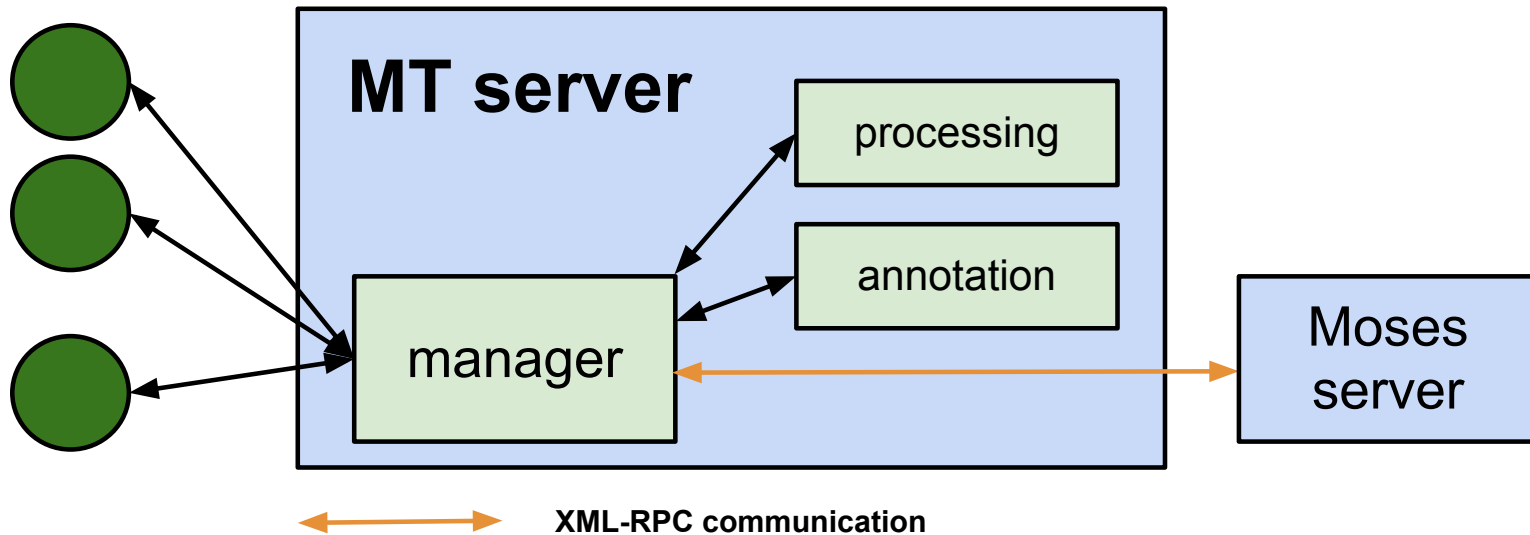
MT server

Matecat Tool installation



MT server

non-adaptive

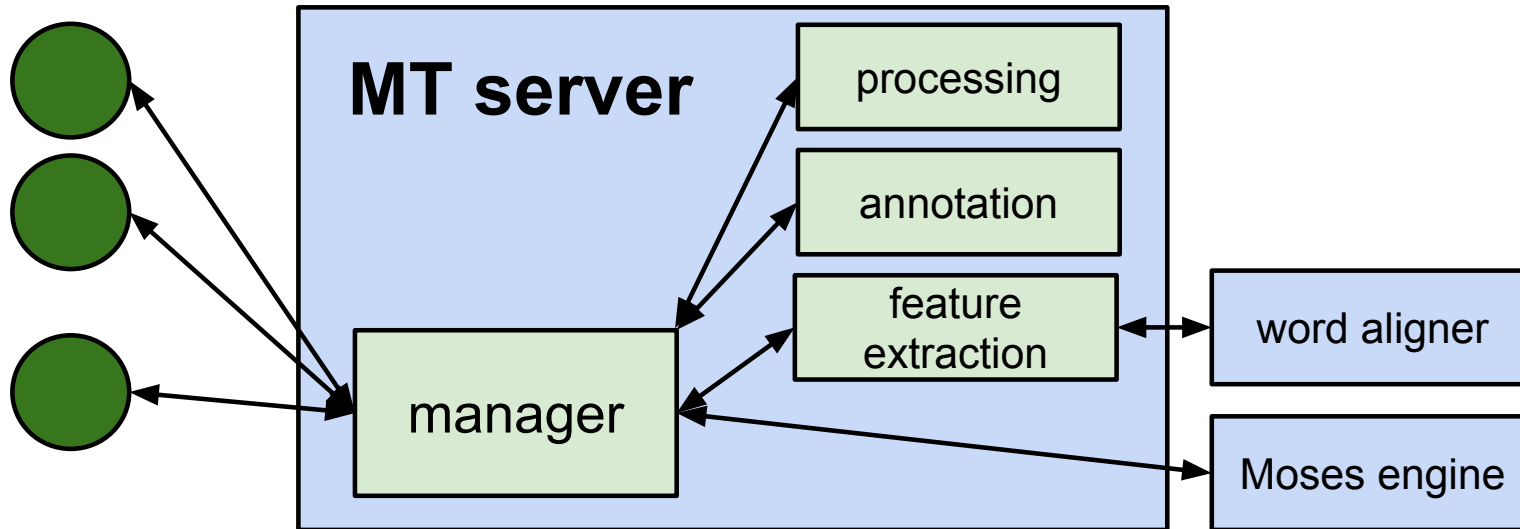


- asynchronous translation requests
- text processing and translation annotation
- Moses server can run remotely



MT server

adaptive



- Moses engine runs locally
- Word aligner runs locally:
 - pivot-based, onlineMgiza++



Non-adaptive MT server



Moses configuration

non-adaptive

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
```



Moses configuration

non-adaptive

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

non-adaptive

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

non-adaptive

Download:

```
$> cd /MATECAT
```

```
$> wget
```

```
www.matecat.com/download/mtserver.zip
```

```
$> tar xzf mtserver.zip
```

Main directory
of MT server



MT server installation

non-adaptive

Configure Moses server

```
$> cd /MATECAT/mtserver  
$> cp template_server.config server.config
```



MT server installation

non-adaptive

```
$> edit server.config
```

```
MOSES_ROOT=/MOSES
```

Full path

```
MOSES_MODELS=/MATECAT/models
```

Moses server
remote URL

```
MOSES_URL=my.mosesserver.url
```

```
MOSES_PORT=7777
```

Port

```
MOSES_LOG=/MATECAT/mtserver/log
```

```
XMLRPC_ROOT=/XMLRPC
```

If not static-linked

```
$> source server.config
```



MT server installation

non-adaptive

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

non-adaptive

Test Moses server

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

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

From another shell

Source text
according to models

```
$> source server.config
```

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

```
Parlamento europeo
```

Expected output



MT server installation

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
```

Full path

accepts queries from

Port

```
$> source server.config
```




MT server installation

non-adaptive

Start MT server

```
$> python_server/start-mtserver.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

non-adaptive

Source text
according to models

Test MT server

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

Expected output
in JSON format

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



Adaptive MT server

Matecat Tool installation



Moses configuration

adaptive

Configure Moses

```
$> 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

adaptive

Test Moses

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

European Parliament

Enter a source text
according to models

Parlamento europeo

Expected output



Moses configuration

adaptive

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

adaptive

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

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

Install models

```
$> unzip sample_updater.zip  
$> cd updater  
$> cp template_updater.ini updater.ini
```



Feature extraction configuration

adaptive

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

adaptive

```
$> gedit SRC2TRG_gizacfg.online  
$> gedit TRG2SRC_gizacfg.online
```

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



MT server installation

adaptive

Download:

```
$> cd /MATECAT
```

```
$> wget
```

```
www.matecat.com/download/mtserver-adaptive.zip
```

```
$> tar xzf mtserver-adaptive.zip
```

Main directory
of MT server



MT server installation

adaptive

Configure MT server

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



MT server installation

adaptive

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/mtserver-adaptive/log
```

```
$> source server-adaptive.config
```



MT server installation

adaptive

Configure feature extraction

```
$> edit server-adaptive.config
```

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

```
$> source server-adaptive.config
```



MT server installation

adaptive

Configure MT server

```
$> edit server-adaptive.config
```

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

```
$> source server-adaptive.config
```



MT server installation

adaptive

Start MT server

```
$> 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



MT server installation

adaptive

Source text
according to models

Test MT server (translate)

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

Expected output
in JSON format

```
{ "data":  
  { "translations":  
    [ { "segmentID": "0000",  
        "translatedText": "parlamento europeo.",  
        "systemName": "system_adaptive",  
        "phraseAlignment": [[[0-1], [0, 1]], ...
```




MT server installation

adaptive

Test MT server (update)

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

Source text
and post-edit

```
{"data":
{"code": "0",
"systemName": "system_adaptive",
"string": "OK", ...}}
```

Expected output
in JSON format



Open in Chrome

<http://my.matecat.tool>

Enjoy!