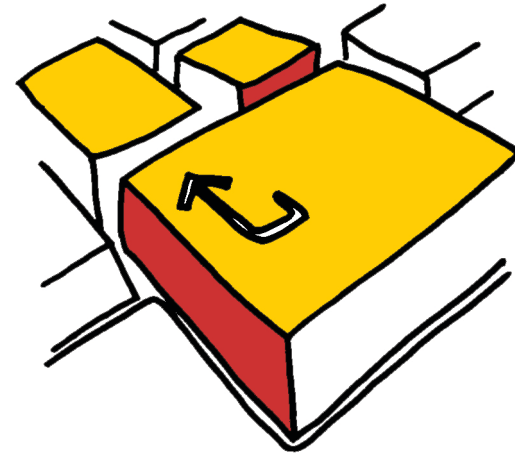


TAUS Moses Roundtable

Welcome and Aims

Rahzeb Choudhury
TAUS

*11-Sep-2013
Prague, Czech Republic*



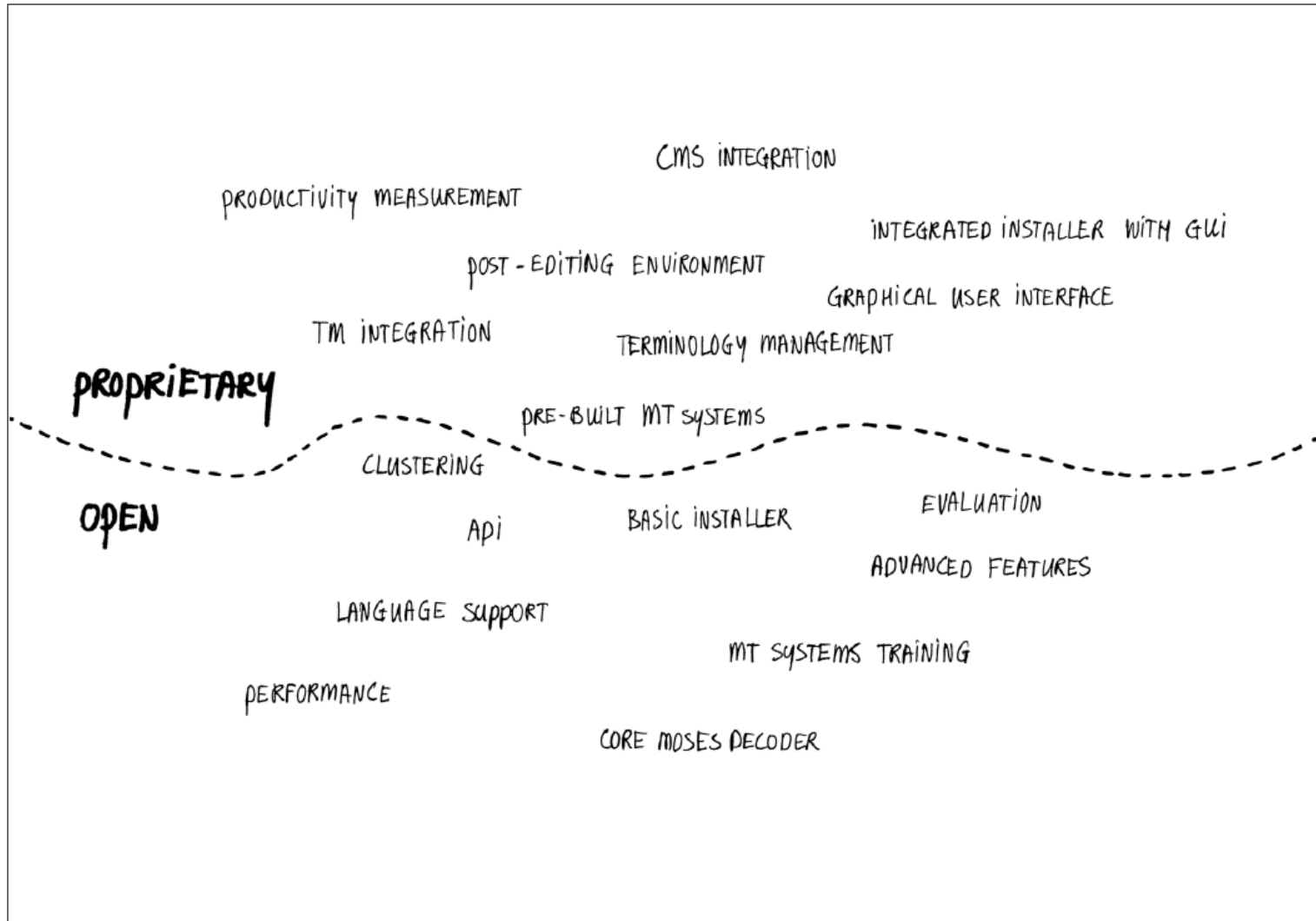
Moses Users – Finding Common Ground

Are there areas where Moses users (from industry) can cooperate? (beyond what is already done as part of MosesCore)

AREA	COOPERATION
Knowledge Sharing	Yes
Sharing Investment	?
Sharing Code	?



Open/Proprietary



Agenda

14:00/ Welcome

14:10/ Introductions

14:30/ Results Moses Survey

15:00/ Moses Roadmap

15:30 / Discussion on Areas for Cooperation

16:00 / BREAK

16:30/ Review/Prioritize Areas for Cooperation

17:15/ Wrap Up and Adjourn

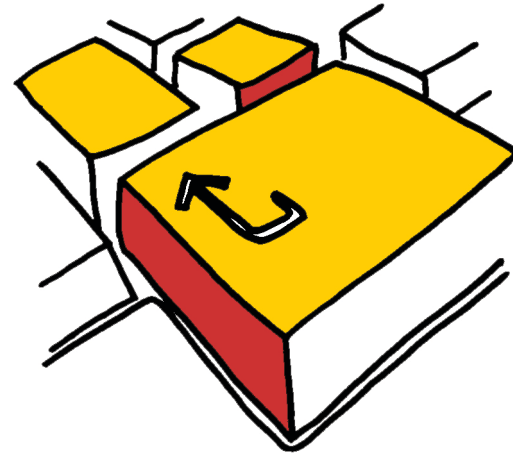


TAUS Moses Roundtable

Introductions

11-Sep-2013

Prague, Czech Republic



TAUS

MOSES  CORE

Introductions

- Konstantinos Chatzitheodorou, Alpha CRC
- Natalia Kljueva, Charles University
- Shadi Salen, Charles University
- Milan Condak, Condak.net s.r.o.
- Bonnie Dorr, DARPA
- Zdena Závůrková, IBM
- Anabela Barreiro, INESC-ID
- Adam Lopez Johns Hopkins University
- Christian Buck, Lantis
- Michal Kašpar, Lingea s.r.o.
- Jacek Skarbek, LocStar
- Tomas Fulatak, Moravia
- Niko Papula, Multilizer



Introductions

- Daniel Rosàs, Pactera
- Francis Tyers, Prompsit
- WonYoung Seo, Samsung Electronics
- SeungWook Lee, Samsung Electronics
- Falko Schaefer, SAP AG
- Alexander Semerenko, Seznam.cz, a.s.
- Jie Jiang, Capita T&I
- Martin Baumgärtner, STAR Langauge Technology & Solutions GmbH
- Ronald Horselenberg, TransIT BV
- Ulrich Germann, University of Edinburgh
- Varvara Logacheva, USFD
- Alex Yanishevsky, Welocalize
- Andrzej Zydrón, XTM-INTL



Introductions

Organisers

- Ondrej Bojar, Charles University
- Philipp Koehn, University of Edinburgh
- Barry Haddow, University of Edinburgh
- Hieu Hoang, University of Edinburgh

- Achim Ruopp, TAUS
- Rahzeb Choudhury, TAUS



TAUS Moses Roundtable

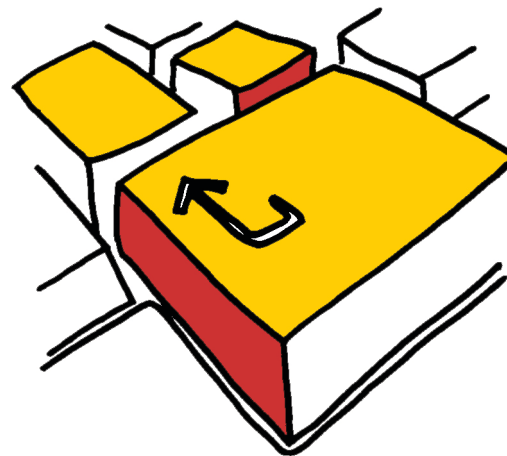
MT @ Alpha CRC

CHATZITHEODOROU Konstantinos

Alpha CRC

11-Sep-2013

Prague, Czech Republic



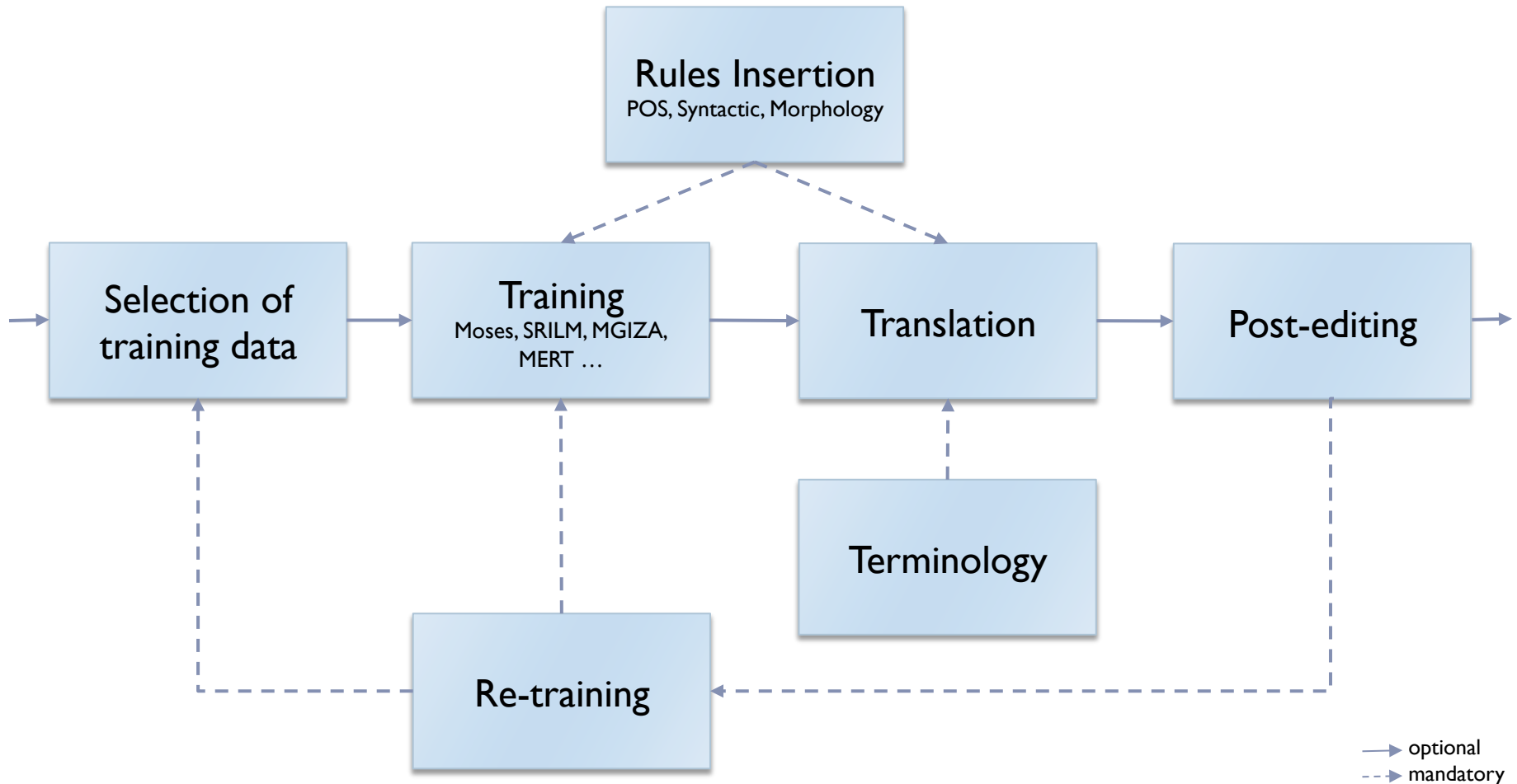
 **TAUS**

MOSES  CORE

MT @ Alpha CRC

- Working with MT since 2006
- Hybrid phrase-based MT system
- Post-editing cost evaluation
 - Have developed Reverse Analysis, a methodology to evaluate the post editing effort on the basis of how much MT output was edited

Alpha MT flow



TAUS Moses Roundtable

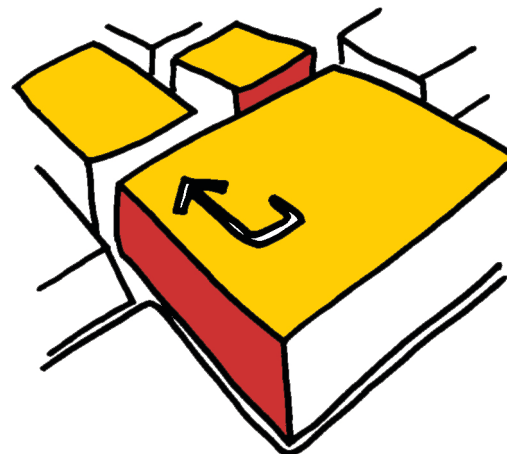
<My MTs and my CATs>

<Milan Čondák>

<Condak.net s.r.o. Petřvald>

11-Sep-2013

Prague, Czech Republic



< My MTs and my CATs >

<PC Translator>

- My first MT was Czech program PC Translator. This SW run in Windows, one language is foreign language and second language is Czech. PC Translator has the bidirectional indexes and can translate in both directions.
- There was 3 main modules: a Dictionary, an Editor and a Dictionary Manager.
- PC Translator worked in two modes: translating of an entire file or translating of text in Editor.
- By text translating was visible a terminology of opened sentence.

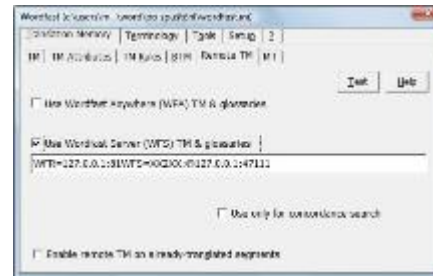


Wordfast Classic in MS Word

- Wordfast Classic (WFC) have been offering integration MT which works in MS Word
- So I asked a developer of PC Translator to create API for MS Word. He created three APIs: for MS Word, for MS Outlook and MS IE. Later he added APIs for more email clients and web browsers.
- WFC begun to use new feature, a Companion. In new window is displayed terminology of opened segment which is found in Wordfast glossary.



Wordfast Classic in MS Word



```
WFS1=000010;https://www.opolcapiz.com/language/translate/
WFS2=000010;https://www.opolcapiz.com/language/translate/
WFS3=000010;https://www.opolcapiz.com/language/translate/
WFS4=000010;https://www.opolcapiz.com/language/translate/
WFS5=000010;https://www.opolcapiz.com/language/translate/
WFS6=000010;https://www.opolcapiz.com/language/translate/
WFS7=000010;https://www.opolcapiz.com/language/translate/
WFS8=000010;https://www.opolcapiz.com/language/translate/
WFS9=000010;https://www.opolcapiz.com/language/translate/
WFS10=000010;https://www.opolcapiz.com/language/translate/
WFS11=000010;https://www.opolcapiz.com/language/translate/
WFS12=000010;https://www.opolcapiz.com/language/translate/
WFS13=000010;https://www.opolcapiz.com/language/translate/
WFS14=000010;https://www.opolcapiz.com/language/translate/
WFS15=000010;https://www.opolcapiz.com/language/translate/
WFS16=000010;https://www.opolcapiz.com/language/translate/
WFS17=000010;https://www.opolcapiz.com/language/translate/
WFS18=000010;https://www.opolcapiz.com/language/translate/
WFS19=000010;https://www.opolcapiz.com/language/translate/
WFS20=000010;https://www.opolcapiz.com/language/translate/
WFS21=000010;https://www.opolcapiz.com/language/translate/
WFS22=000010;https://www.opolcapiz.com/language/translate/
WFS23=000010;https://www.opolcapiz.com/language/translate/
WFS24=000010;https://www.opolcapiz.com/language/translate/
WFS25=000010;https://www.opolcapiz.com/language/translate/
WFS26=000010;https://www.opolcapiz.com/language/translate/
WFS27=000010;https://www.opolcapiz.com/language/translate/
WFS28=000010;https://www.opolcapiz.com/language/translate/
WFS29=000010;https://www.opolcapiz.com/language/translate/
WFS30=000010;https://www.opolcapiz.com/language/translate/
WFS31=000010;https://www.opolcapiz.com/language/translate/
WFS32=000010;https://www.opolcapiz.com/language/translate/
WFS33=000010;https://www.opolcapiz.com/language/translate/
WFS34=000010;https://www.opolcapiz.com/language/translate/
WFS35=000010;https://www.opolcapiz.com/language/translate/
WFS36=000010;https://www.opolcapiz.com/language/translate/
WFS37=000010;https://www.opolcapiz.com/language/translate/
WFS38=000010;https://www.opolcapiz.com/language/translate/
WFS39=000010;https://www.opolcapiz.com/language/translate/
WFS40=000010;https://www.opolcapiz.com/language/translate/
WFS41=000010;https://www.opolcapiz.com/language/translate/
WFS42=000010;https://www.opolcapiz.com/language/translate/
WFS43=000010;https://www.opolcapiz.com/language/translate/
WFS44=000010;https://www.opolcapiz.com/language/translate/
WFS45=000010;https://www.opolcapiz.com/language/translate/
WFS46=000010;https://www.opolcapiz.com/language/translate/
WFS47=000010;https://www.opolcapiz.com/language/translate/
WFS48=000010;https://www.opolcapiz.com/language/translate/
WFS49=000010;https://www.opolcapiz.com/language/translate/
WFS50=000010;https://www.opolcapiz.com/language/translate/
WFS51=000010;https://www.opolcapiz.com/language/translate/
WFS52=000010;https://www.opolcapiz.com/language/translate/
WFS53=000010;https://www.opolcapiz.com/language/translate/
WFS54=000010;https://www.opolcapiz.com/language/translate/
WFS55=000010;https://www.opolcapiz.com/language/translate/
WFS56=000010;https://www.opolcapiz.com/language/translate/
WFS57=000010;https://www.opolcapiz.com/language/translate/
WFS58=000010;https://www.opolcapiz.com/language/translate/
WFS59=000010;https://www.opolcapiz.com/language/translate/
WFS60=000010;https://www.opolcapiz.com/language/translate/
WFS61=000010;https://www.opolcapiz.com/language/translate/
WFS62=000010;https://www.opolcapiz.com/language/translate/
WFS63=000010;https://www.opolcapiz.com/language/translate/
WFS64=000010;https://www.opolcapiz.com/language/translate/
WFS65=000010;https://www.opolcapiz.com/language/translate/
WFS66=000010;https://www.opolcapiz.com/language/translate/
WFS67=000010;https://www.opolcapiz.com/language/translate/
WFS68=000010;https://www.opolcapiz.com/language/translate/
WFS69=000010;https://www.opolcapiz.com/language/translate/
WFS70=000010;https://www.opolcapiz.com/language/translate/
WFS71=000010;https://www.opolcapiz.com/language/translate/
WFS72=000010;https://www.opolcapiz.com/language/translate/
WFS73=000010;https://www.opolcapiz.com/language/translate/
WFS74=000010;https://www.opolcapiz.com/language/translate/
WFS75=000010;https://www.opolcapiz.com/language/translate/
WFS76=000010;https://www.opolcapiz.com/language/translate/
WFS77=000010;https://www.opolcapiz.com/language/translate/
WFS78=000010;https://www.opolcapiz.com/language/translate/
WFS79=000010;https://www.opolcapiz.com/language/translate/
WFS80=000010;https://www.opolcapiz.com/language/translate/
WFS81=000010;https://www.opolcapiz.com/language/translate/
WFS82=000010;https://www.opolcapiz.com/language/translate/
WFS83=000010;https://www.opolcapiz.com/language/translate/
WFS84=000010;https://www.opolcapiz.com/language/translate/
WFS85=000010;https://www.opolcapiz.com/language/translate/
WFS86=000010;https://www.opolcapiz.com/language/translate/
WFS87=000010;https://www.opolcapiz.com/language/translate/
WFS88=000010;https://www.opolcapiz.com/language/translate/
WFS89=000010;https://www.opolcapiz.com/language/translate/
WFS90=000010;https://www.opolcapiz.com/language/translate/
WFS91=000010;https://www.opolcapiz.com/language/translate/
WFS92=000010;https://www.opolcapiz.com/language/translate/
WFS93=000010;https://www.opolcapiz.com/language/translate/
WFS94=000010;https://www.opolcapiz.com/language/translate/
WFS95=000010;https://www.opolcapiz.com/language/translate/
WFS96=000010;https://www.opolcapiz.com/language/translate/
WFS97=000010;https://www.opolcapiz.com/language/translate/
WFS98=000010;https://www.opolcapiz.com/language/translate/
WFS99=000010;https://www.opolcapiz.com/language/translate/
WFS100=000010;https://www.opolcapiz.com/language/translate/
```



Web translation services in my MT and CATs

- PC Translator can show offers from Google and Bing:
- http://www.condak.net/machine_t/cs/comprendo/cs/07.html
- MetaTaxis for Word 2007 + Web MT Servers:
- http://www.condak.net/cat_other/virtaal/20130821/cs/02.html
- Free Translation via Internet works without registration.



- Virtaal Plugins - models for TM and MT:
- http://www.condak.net/cat_other/virtaal/20130821/cs/03.html

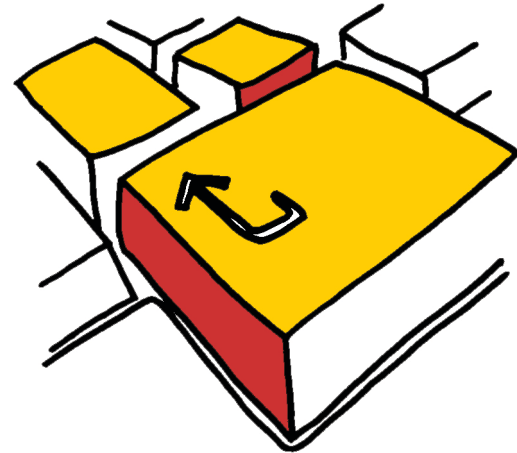


TAUS Moses Roundtable

MT in localization company

Jacek Skarbek
LocStar

11-Sep-2013
Prague, Czech Republic



Our experience with MT

- We are a software localization provider which has been using CAT tools for 17 years – we have large TMs
- Some of our clients provide us work with MTranslated content (both for no matches and as alternative to TM fuzzies) using their own MT solutions – our job is to work on MT like with fuzzy matches (no typical post-editing)
- For about 2 years we have been used MT for one of our main customer. We buy MT content from third party that use their own solution based on Moses and TMs provided by customer.
- We have large TMs collected and we test our Moses based internal solution to use it in production environment



Moses related problems/areas to improve

- Tag handling/inline markup – only partially resolved by M4Loc solutions
- Lack of API to better/easier integrate Moses into production workflow
- We need better terminology/software items handling. I.e something like <zone>, but phrase in zone treated separately from rest of sentence at the level of TM and as a part of sentence at the level of LM
- Inflections in Slavic languages



Other problems

- Weird approach to MT rate – customers tend to decrease translation rate in the same percent as measured (or estimated) acceleration of translation work itself, while translation rates covers also project management and all other linguistic and technical tasks that are not accelerated by MT
- We are not allowed to use MT for some customers – it is restricted by work agreement . Although we treat MTranslated content as fuzzy matches, they afraid that it would impact on final quality of translation



TAUS Moses Roundtable

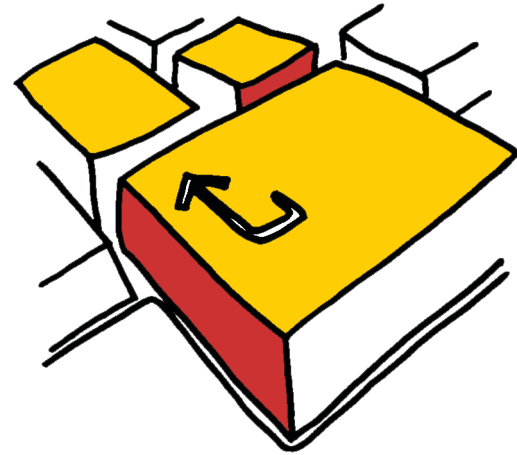
Samsung Electronics
Cooperation and SMT

Seung-Wook Lee, Wonyoung Seo
Samsung Electronics Corporation

11-Sep-2013

Prague, Czech Republic

 **TAUS**



MOSES  CORE

Samsung Electronics Cooperation and Machine Translation

- Our team provide translation services for various internal groupware applications (e.g., instance messenger) for the department
- One of the main concerns of ours is to expand language pairs
 - There are very little of bilingual corpus available for the most of languages, such as Asian languages
 - Is indirect translation the solution? how do we deal with the error propagation?
 - Working groups and developer meetings for those languages may necessary



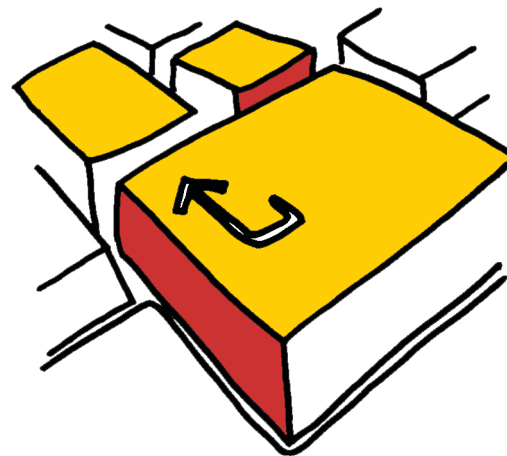
TAUS Moses Roundtable

Statistical Machine
Translation at SAP

Dr. Falko Schaefer
SAP Language Services

11-Sep-2013

Prague, Czech Republic



SMT Project at SLS

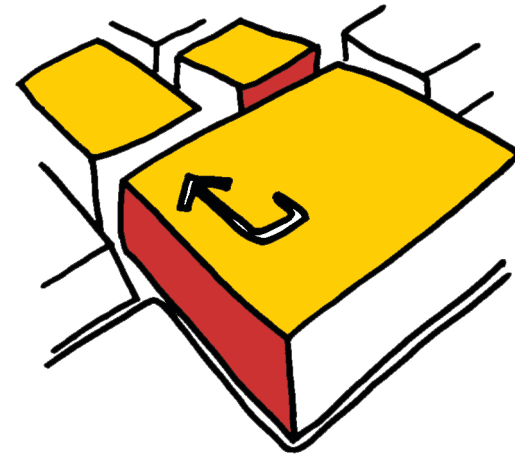
- SAP Language Services (SLS) has successfully worked with rule-based MT for over 20 years
- However, the growing demand for a new breed of MT meant that SLS began to embark on Moses-based SMT in early 2013
- The SLS MT project aims to establish a new MT service to reduce translation throughput time and cost
- To that end SLS works with an external partner to support implementation and knowledge transfer



TAUS Moses Roundtable

Alex Yanishevsky
Welocalize

11-Sep-2013
Prague, Czech Republic



Company Intro

- One of top 10 LSPs (language service providers)
- Department dedicated to MT and Language Tools (evaluation of MT, productivity workbench, corpus preparation, vendor selection, vendor training and certification)
- MT agnostic
- MT integrated into TMS/GMS



Areas of Interest

- Productization of Moses
 - lower barrier of entry
 - interoperability
- Integration into TMS/GMS
- Tag handling
- Predictive modeling



TAUS Moses Roundtable

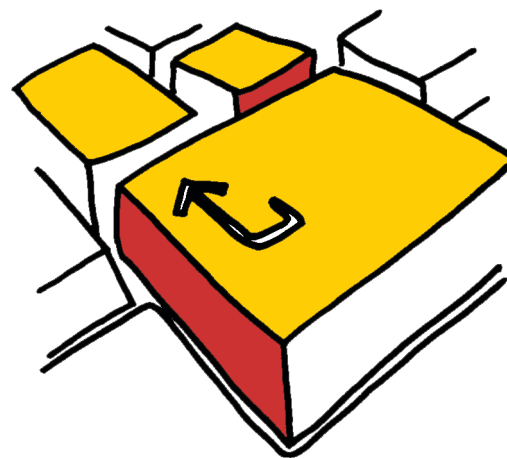
Results: Moses Survey

Achim Ruopp

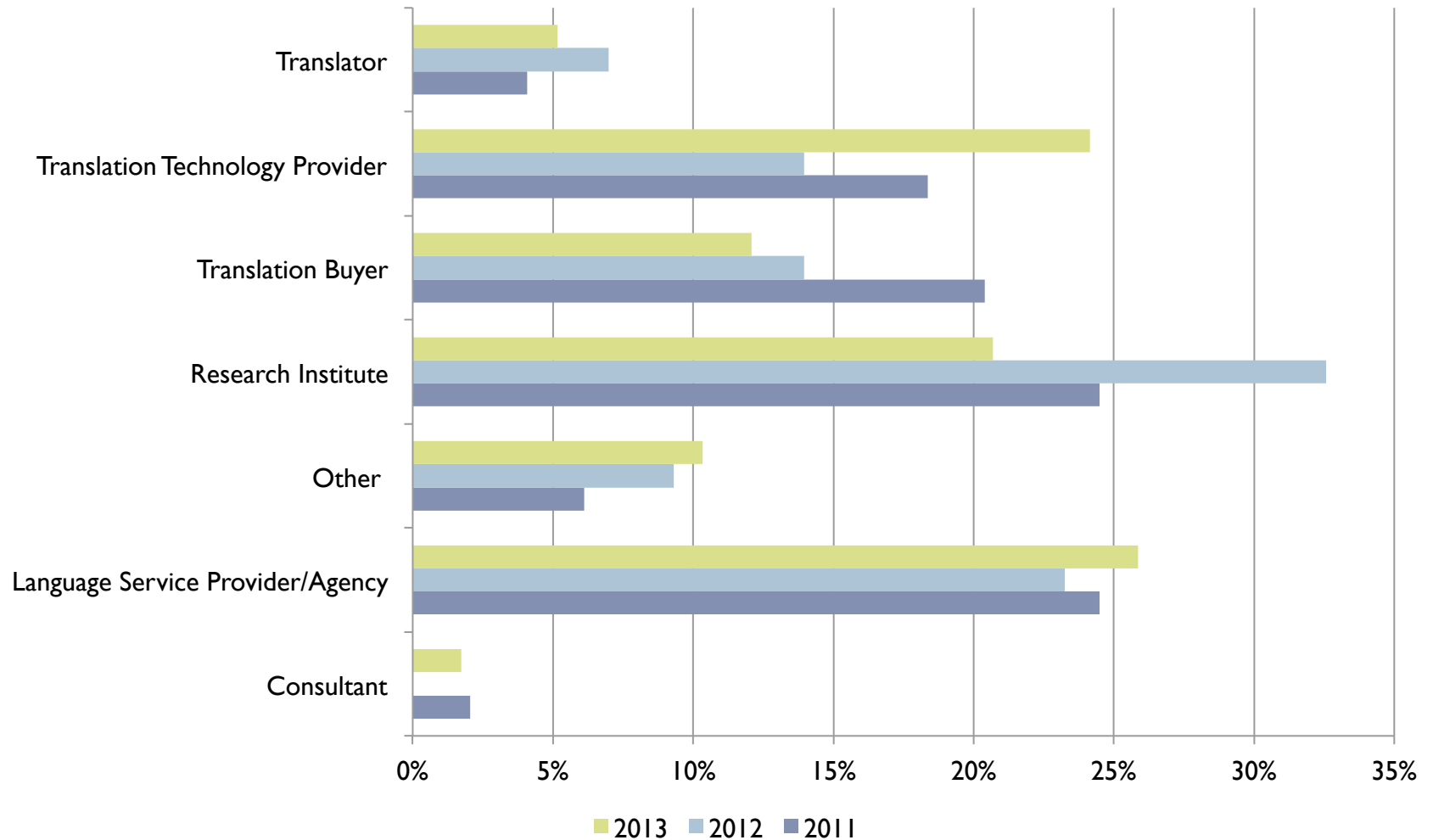
TAUS

11-Sep-2013

Prague, Czech Republic



Demographic Composition



Ranking of Requested Moses Improvements

2013 Rank	2012 Rank	2011 Rank	
1	1	3	Training and translation speed
2	4	1	Integrating Moses into existing workflow/system (e.g. TM integration)
3	3	2	Installing and using Moses
4			Terminology Management
5	2	4	Evaluation results (e.g. evaluating productivity)
6	5	5	Language-specific issues
7			Advanced features (e.g. tree-based translation)
8	7	7	Customer support
	6	6	Easier to get the right human resources



#1 Requested Moses Improvement

Training and Translation Speed

- Users are aware that SMT requires a considerable amount of computing resources
- Request driven by management and user demands
 - Fast-turn-around/online translation
 - Frequent re-training of systems with new/updated data
- Recommendation
 - Integrate recent training speed improvements into the training tool chain
 - Document recommendations how to best use the training speed improvements
 - Further optimize performance for multi-threaded decoding



#2 Requested Moses Improvement

Integrating Moses into Existing Workflows/Systems

- Integration into growing number of diverse systems
 - TMS/CaT/TenT
 - Content Management Systems
 - Automated Speech Recognition
 - Dialog Systems
 - ...
- Recommendation
 - Comprehensive, stable and well documented APIs to the decoder and data produced by it
 - RESTful HTTP API (Google/Bing compatible?)
 - Finish Okapi/M4Loc file format support



#3 Requested Moses Improvement

Installing and Using Moses

- Still a range of installation experiences from “No problem” to “very complex to understand and to implement”
- Should Moses team provide installable packages?
- Windows support?
- UI?
- Recommendation
 - Occasional stable releases of Moses as installable packages across different platforms – consistency is key
 - Take on the maintenance and release of required components abandoned by their original developers



#4 Requested Moses Improvement

Terminology Management

- Terminology injection from terminological resources
 - Term bases
 - Named entity recognizers
- Recommendation:
 - Better documentation of the XML input feature
 - Ensuring that the XML input feature minimally impacts translation quality of the overall sentence
 - Handling input with named entities marked up by named entity recognizers
 - Ensure XML input can be handled in the complete tool chain (e.g. tokenizer)



#5 Requested Moses Improvement

Evaluation

- Expansion of the metrics that can be used to tune Moses MT systems
 - Specifically for MT+post-editing scenario
- Evaluation and productivity testing systems
 - Can be external
- Recommendation
 - Integrate tuning metrics into Moses that allow optimizing systems for the MT+post-editing usage scenario
 - Ensure interoperability with external evaluation/productivity testing systems, e.g. TAUS DQF, QT Launchpad



#6 Requested Moses Improvement

Language-Specific Issues

- Moses is focused on a relatively small set of European languages
- Survey participants would like to see tools for more languages included
- Full Unicode support
- Recommendation
 - Test and improve Unicode support in the language-independent core
 - Recommend and document use of additional language tools
 - Encourage users to report Unicode issues and provide language-specific data



#7 Requested Moses Improvement

Advanced Features for Moses Commercialization

- Received a broad cross-section of requests
- Researchers develop cutting-edge technologies that could benefit industry
- Too often conversations still happen in distinct academic and industry silos
- Recommendation
 - Start a conversation explaining how newly developed methods and technologies can help the industry to address critical MT issues, e.g.
 - Tree-based/syntax-based models
 - Morphologically rich languages



#8 Requested Moses Improvement

Customer Support

- Moses support mailing list considered excellent
- Few requests for professional support or faster support response times
- Recommendation
 - Continue excellent support on mailing list
 - Improve documentation for some industry-relevant features to allow easier adoption



Moses Open Source Project

Strengths

- Additions/updates to “core” Moses: decoder, training, LM
 - Latest methods
 - Benefiting all users
- Documentation and support
- MosesCore funded releases and tutorial

Weaknesses

- Few contributions by long-time industry Moses users
 - Adobe Moses Tools
 - DoMY CE
- Complexity of installation/use for entry-level users



Moses Future

Not mutually exclusive!

Academic Project

- Sharing platform for research progress
- Unstable code base
- Complex use
- Integration by few sophisticated technology providers
- Similar OSS project
 - HTK speech recognition toolkit

Broad Adoption

- Ease of installation
- Ease of use for diverse scenarios
- Pre-trained engines
- Similar OSS projects
 - PostgreSQL
 - NLTK (Natural Language Toolkit)
 - CMU Sphinx



PostgreSQL

Object-relational database management system

- Started in 1986 by Michael Stonebraker at UC Berkeley
- Evolved from research project into universal RDBMS
- Used by Apple, BASF, Skype, Redhat, governments, universities ...
- Broad contributor base
 - Often industry funded
- PostgreSQL license (similar to MIT license)
- Commercial support through EnterpriseDB
- Consulting/training available

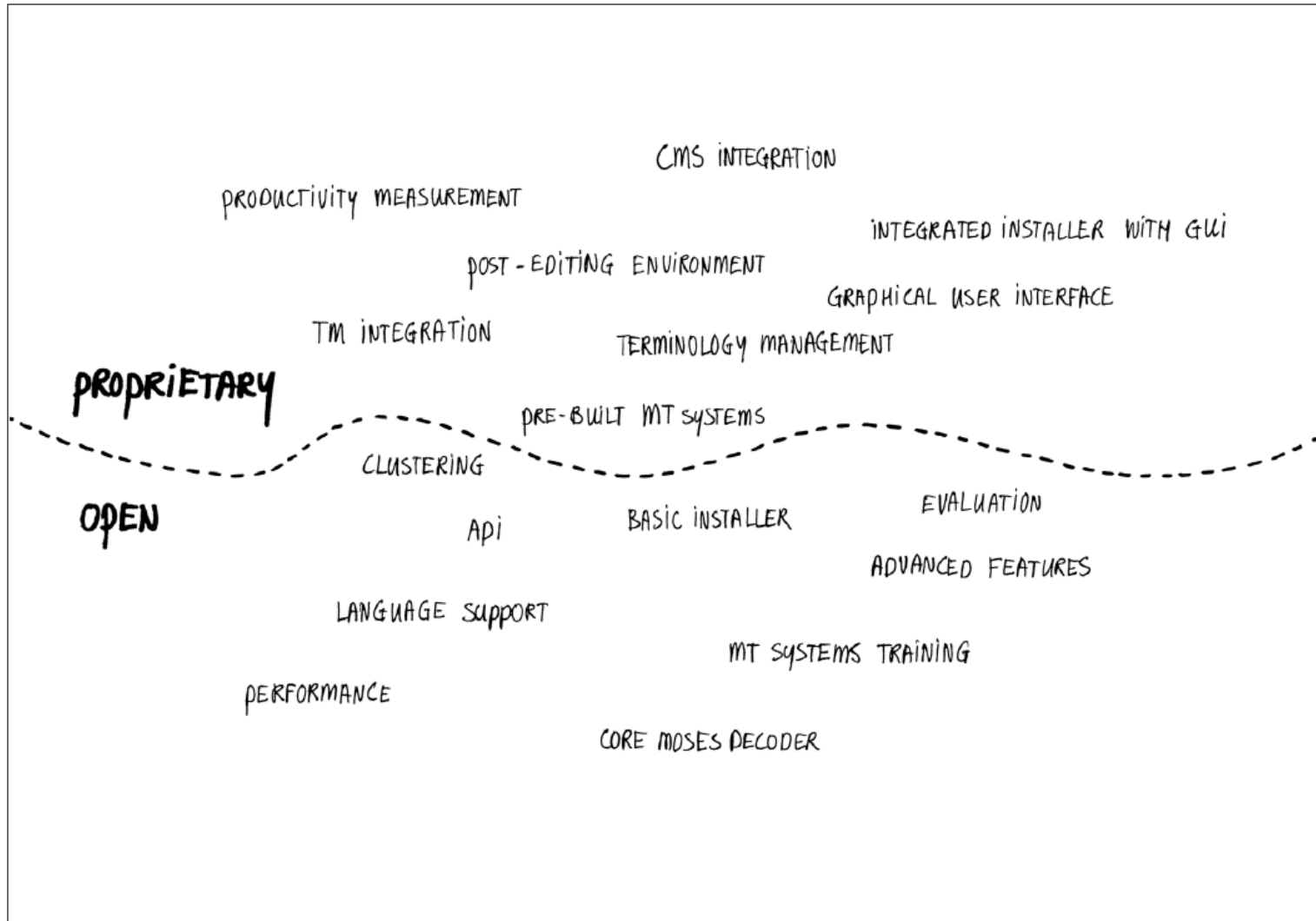


Discussion To Follow

- Discuss industry needs
- Identify areas of industry cooperation
- Discuss line between open source project and proprietary add-ons



Open/Proprietary



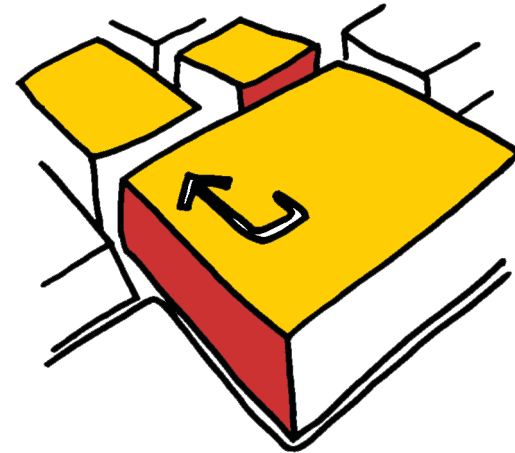
TAUS Moses Roundtable

Moses Roadmap

Philipp Koehn

11-Sep-2013

Prague, Czech Republic



TAUS

MOSES  CORE

Moses Roadmap

Philipp Koehn

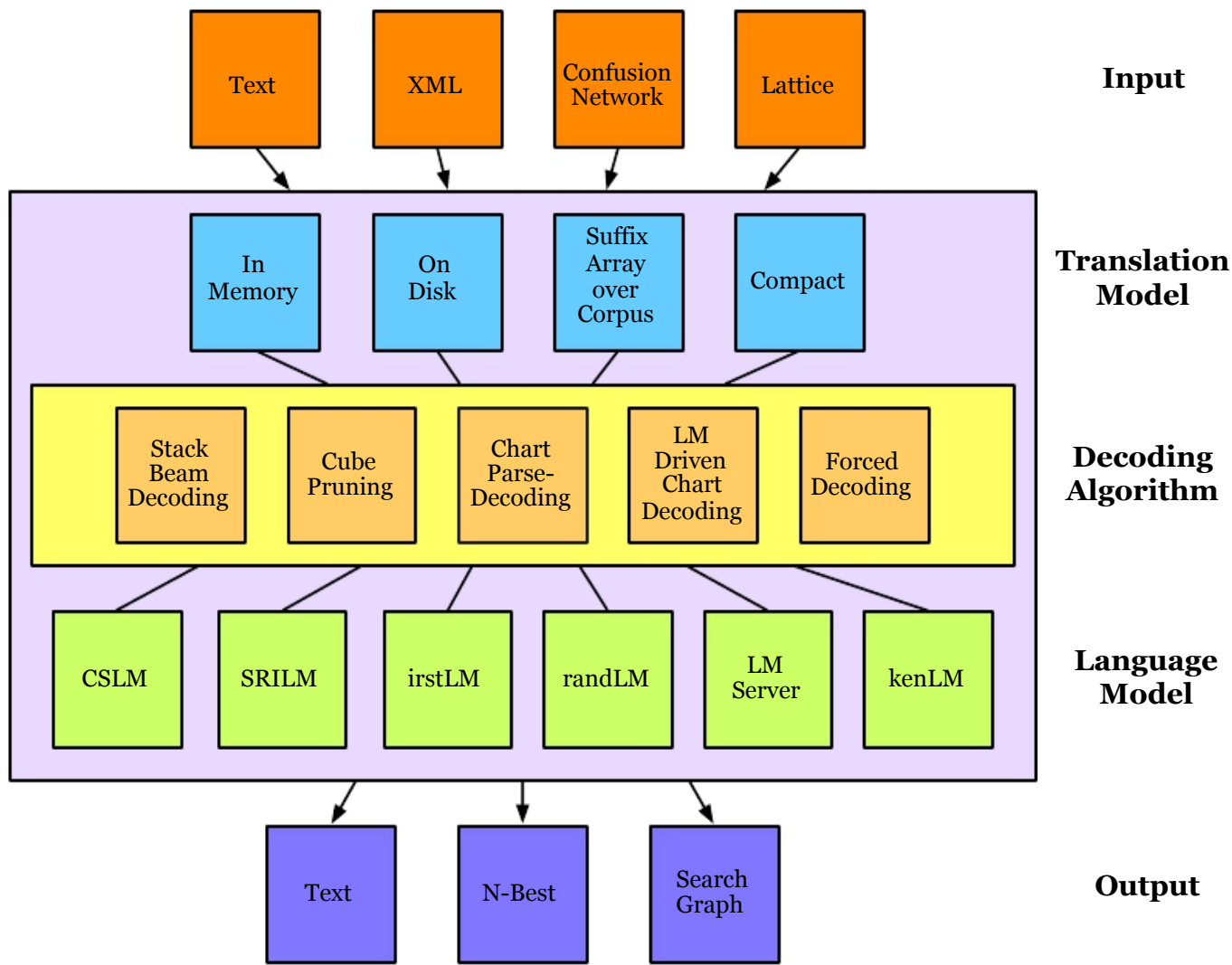
11 September 2013



Development in Moses

- Moses is mainly developed in academia
- Academic research progress is somewhat un-predictable
- Biases
 1. quality
 2. scalability
 3. usability
- Sometimes research use cases do not match industry use cases (e.g., translation of news vs. technical documentation)

Modular Design



Progress in Models

1990

2000

2010

word-based models

phrase-based models

formal grammar-based models

linguistic grammar-based models

semantics

Progress in Methods

1990

2000

2010

probabilistic models

parameter tuning

**large-scale
discriminative training**

Quality



- Better machine learning methods
- Linguistically motivated models
- More data



Quality

Some examples from UEDIN systems in WMT 2013

- Better machine learning methods
operation sequence model
- Linguistically motivated models
syntax-based machine translation model
- More data
training a language model on 130 billion words

Operation Sequence Model

5-gram sequence model over operations
(minimal phrase translations, reordering)

$p(o_1) p(o_2|o_1) p(o_3|o_1,o_2) \dots p(o_{10}|o_6,o_7,o_8,o_9)$

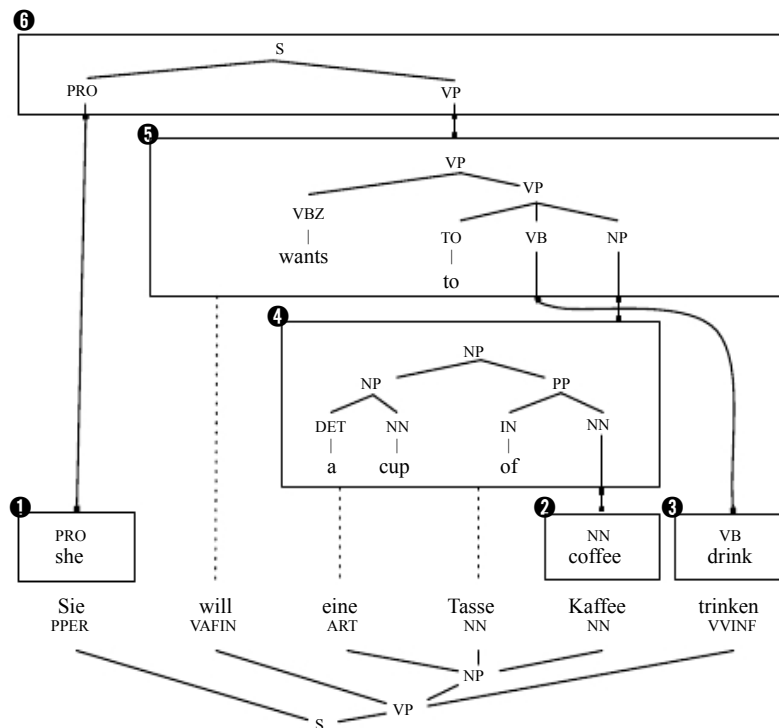
Feature function in Moses [Durrani et al., ACL 2013]

o ₁	Generate(Ich,I)	Ich↓ I
o ₂	GenerateTargetOnly(do)	Ich↓ Ido
o ₃	InsertGap	Ichnicht↓
o ₄	Generate(nicht,not)	Idonot
o ₅	JumpBack(1)	Ichgehe↓nicht
o ₆	Generate(gehe,go)	Idonotgo
o ₇	GenerateSourceOnly(ja)	Ichgeheja↓nicht Idonotgo
o ₈	JumpForward	Ichgehejanicht↓ Idonotgo
o ₉	Generate(zum,tothe)	...gehejanichtzum↓ ...notgotothe
o ₁₀	Generate(haus,house)	...janichtzumhaus↓ ...gotothehouse

	2012	2013
de-en	+0.26	+0.29
fr-en	+0.19	+0.37
es-en	+0.49	+0.90
cs-en	+0.33	+0.09
ru-en	+0.46	+0.33
en-de	+0.07	+0.20
en-fr	+0.60	+0.36
en-es	+0.57	+0.44
en-cs	+0.35	+0.27
en-ru	+0.30	+0.40

Syntax-Based Machine Translation

[Nadejde et. al, WMT2013]



German–English

English–German

manual score	system
0.608	UEDIN-SYNTAX
0.586	UEDIN-PHRASE

manual score	system
0.614	UEDIN-SYNTAX
0.571	UEDIN-PHRASE

Huge, I Say Huge!, Language Model

- Unpruned 5-gram language model trained on 130 billion words
- Training straightforward [Heafield et al., ACL 2013]
- Decoding requires 1TB RAM machine
- Best performance at WMT2013 (manual judgment)

Spanish–English		French–English		Czech–English	
score	system	score	system	score	system
0.624	UEDIN-HEAFIELD	0.638	UEDIN-HEAFIELD	0.607	UEDIN-HEAFIELD
0.595	”ONLINE-B”	0.604	UEDIN	0.582	”ONLINE-B”
0.570	UEDIN	0.591	”ONLINE-B”	0.562	UEDIN

Usability



- Main uses of Moses
 - productivity tool for professional translators
 - gisting for information discovery

- Research driven by real-world use
 - incremental training
 - handling of tags
 - terminology management
 - quality estimation



- Integration of statistical MT and collaborative translation memories
- Novel technology
 - Self-tuning machine translation
 - User adaptive machine translation
 - Informative machine translation
- Open source workbench
- Extensive testing by translation agency



- Cognitive studies of translator behaviour based on key logging and eye tracking
- Novel types of assistance to human translators
 - interactive translation prediction
 - interactive editing
 - adaptive translation models
- Open source workbench
- Field tests by translation agency and online volunteer translation platforms

AC+OCEPT

- Use of machine translation for community content
- Novel technology
 - Pre-editing of content
 - Monolingual and bilingual post-editing
 - Development of feedback loops
- Use in
 - commercial product forum relating to Symantec network security products
 - content in community of volunteer translators Traducteurs sans Frontières

The Future: Better Models

- Syntax-based and semantic statistical models
 - improvements to basic tools of natural language processing
 - requires annotated data resources, annotation standards
 - new models, training methods, inference algorithms
- Exploitation of data — machine learning
 - different types: parallel, comparable, monolingual, interactive
 - scaling up of existing machine learning methods
 - adaptation to user needs
- Integration with other technologies
 - human translation and localization workflows
 - speech recognition
 - dialog systems
 - information retrieval
 - data mining
 - communication systems

The Future: Better Usability

- Installation
 - MOSESCORE installer, pre-built binaries
 - pre-installed virtual machines for Amazon EC et al.
- Resources
 - ongoing efforts to make data publicly available
 - memory and time efficient training and decoding
- Integration into workflows
 - addressing requirements of professional translators
 - industry-led projects on handling tags, untranslated terms, terminology
 - MOSESCORE "arrows" workflow management
 - various server process implementation, e.g., based on Google API

Thank You



questions?

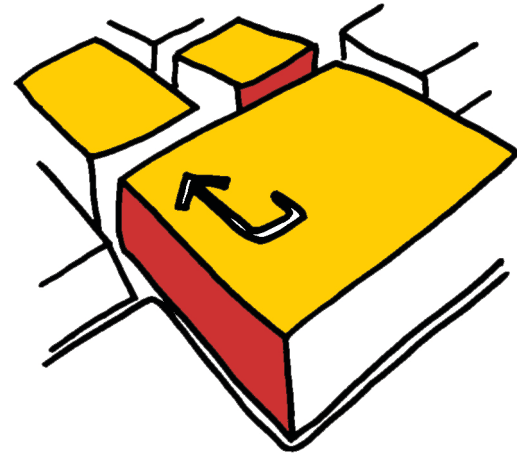
TAUS Moses Roundtable

Review and Discussion of
Sharing Options in the
Industry

Rahzeb Choudhury, Achim Ruopp
TAUS

11-Sep-2013

Prague, Czech Republic



Sharing Knowledge

- TAUS Machine Translation Showcases
 - Co-located with Localization World Conferences
 - Familiarizing the industry with Moses/SMT
 - Users share experiences
 - Panel discussions
- TAUS Machine Translation and Moses Tutorial
 - Online tutorial teaching theory and practice
 - 300+ registered users
 - Developed in collaboration with UEdin
- This TAUS Moses Roundtable

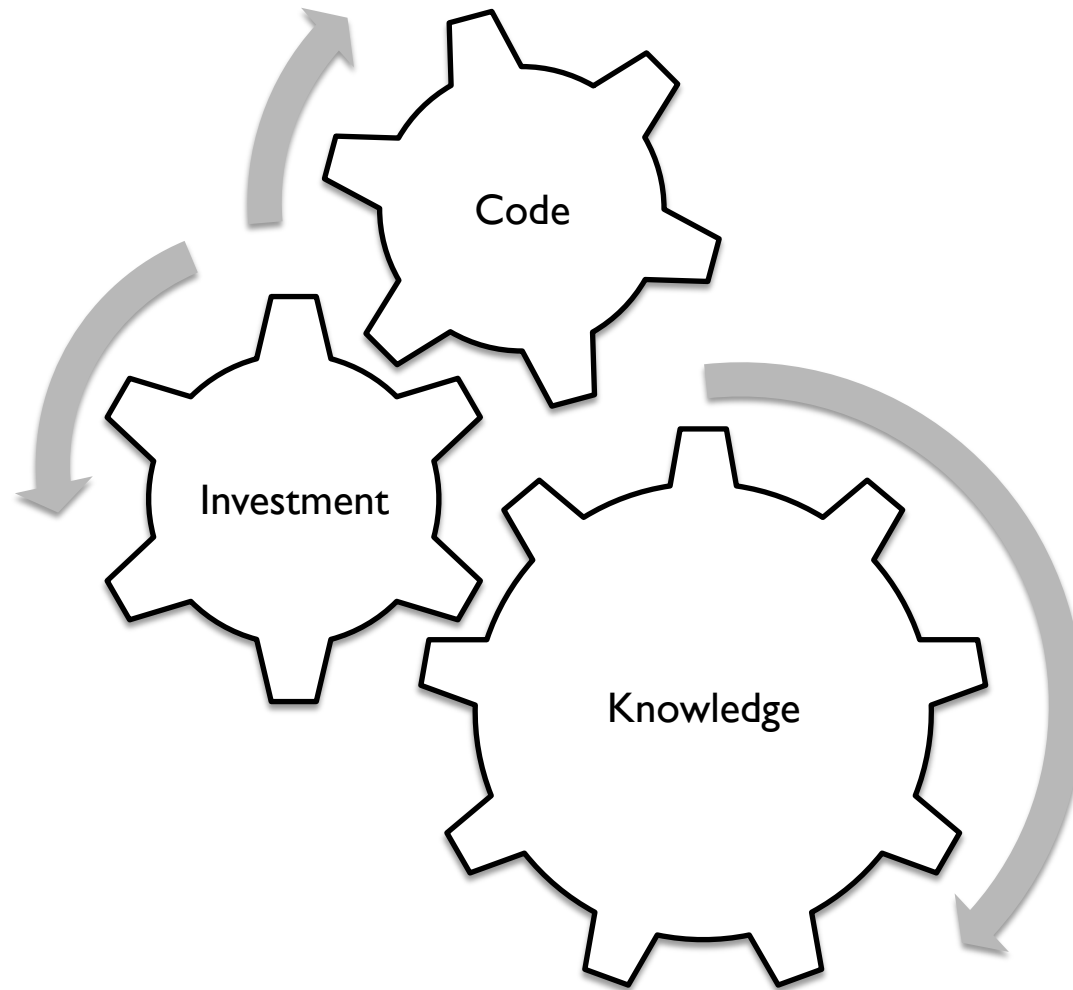


Sharing Code

- DoMY CE
 - Prepare training corpora
 - Train & tune SMT models
 - Manage SMT resources
 - Translate documents
- M4Loc – Moses for Localization
 - Integration with popular open source Okapi localization framework
 - Adobe Moses Tools
- In Moses /contrib folder
 - Moses for Mere Mortals
 - Several web APIs
- Language-specific non-breaking prefix files



Industry Sharing



Discussion: Ideas for Sharing

- What are common use scenarios?
 - (among participants)
- MT as a productivity enhancer
 - Beginner, Pilot, Implementation, Production, Ongoing rollout
- MT to gist – no participants involved in the scenario
- How do we make them easier to achieve?



Beginners -Installing and Using Moses

- Still a range of installation experiences from “No problem” to “very complex to understand and to implement”
- Should Moses team provide installable packages?
- Windows support?
- UI?
- Recommendation
 - Occasional stable releases of Moses as installable packages across different platforms – consistency is key
 - Take on the maintenance and release of required components abandoned by their original developers



Beginners - Installing and Using Moses

- Conclusions during meeting:
 - The resources available (Moses site, support list, MT and Moses Tutorial) are sufficient
 - The v1 release is very welcome and look forward to future releases
 - Try to ensure these resources are more easily discoverable, ensure documentation stays up to date, and easy to use



Implementation

Integrating Moses into Existing Workflows/Systems

- Integration into growing number of diverse systems
 - TMS/CaT/TenT
 - Content Management Systems
 - Automated Speech Recognition
 - Dialog Systems
 - ...
- Recommendation
 - Comprehensive, stable and well documented APIs to the decoder and data produced by it
 - RESTful HTTP API (Google/Bing compatible?)
 - Finish Okapi/M4Loc file format support



Implementation

Integrating Moses into Existing Workflows/Systems

- Conclusions during meeting:
 - Main areas of cooperation (APIs and formatting) covered by current activity
 - TAUS to help with next steps for Moses4Loc (Formatting) to help ensure there is thorough testing



Production

Training and Translation Speed

- Users are aware that SMT requires a considerable amount of computing resources
- Request driven by management and user demands
 - Fast-turn-around/online translation
 - Frequent re-training of systems with new/updated data
- Recommendation
 - Integrate recent training speed improvements into the training tool chain
 - Document recommendations how to best use the training speed improvements
 - Further optimize performance for multi-threaded decoding



Production

Training and Translation Speed

- Conclusions during meeting:
 - Participants did not have any specific ideas beyond what the MosesCore consortium members are already doing



Other Issues/Ideas Raised

- Lack of Data
 - The TAUS Data repository was shown as a potential source of training data
- Interoperability
 - Going forward it would be good to be able to share translation and language models
 - Participants briefly discussed the complexity of the challenge
- Shared engines
 - It was suggested that baseline language/industry/domain engines be made available
 - Making the engines built as part of the TAUS Developing Talent project available may be a good start. TAUS will look into this.



TAUS Moses Roundtable

Thank you!

Achim Ruopp (achim@taus.net)

Rahzeb Choudhury (rahzeb@taus.net)

