MTTT – Machine Translation Training Tool
A Tool to Teach MT, Evaluation and Post-editing

P. Bouillon (FTI), P. Estrella (FaMAF), R. Lafuente (FaMAF), M. Lemos (UBP), S. Girletti (FTI)

- **Collaboration**: FaMAF@UNC, UBP, Córdoba, Argentina & TIM/FTI@UniGe, Geneva, Switzerland
- **Goal**: integrate complete MT+PE workflow into one tool
- **Scope**: for pedagogical use
- **Advantages**: open-source tool, GUI, desktop and web versions, user guide provided
- **Current state**: on-going development
- **Further development**: access to intermediate results, tmx support, extensive testing

UNC FAMAF

UNIVERSITÉ DE GENÈVE
FACULTÉ DE TRADUCTION ET D'INTERPRÉTATION

Desktop version

Web version

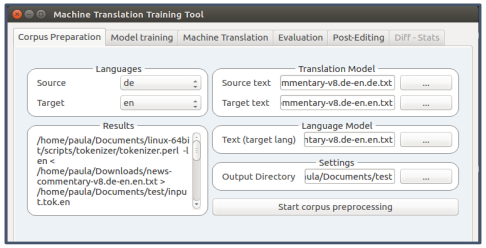
Corpus Preparation

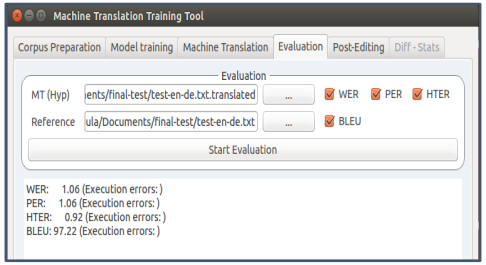
Training

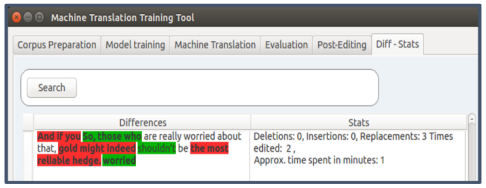
Machine Translation

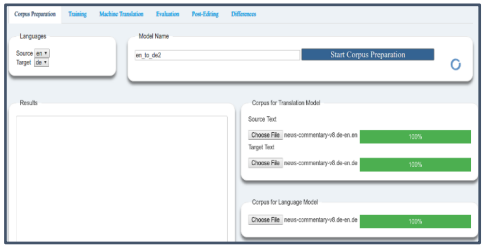
Evaluation

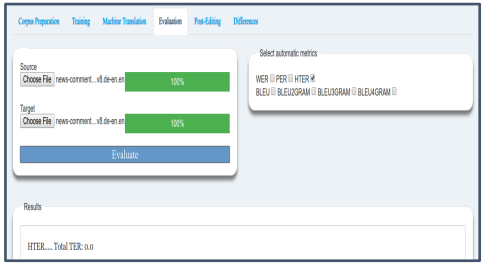
Post-Editing

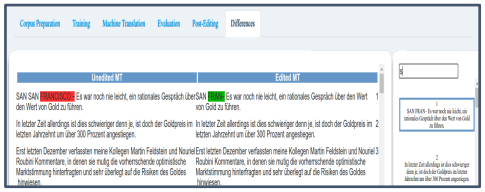












GitHub

roxana-lafuente/MTTT

PaulaEstrella/MTTT-PyQT

miguelemosreverte/MTTT_Web

UNC FAMAF

UNIVERSITÉ DE GENÈVE
FACULTÉ DE TRADUCTION ET D'INTERPRÉTATION

SCATE
Smart Computer-Aided Translation Environment

Year 3 / 4

http://www.ccl.kuleuven.be/scate

Vincent Vandeghinste, Tom Vanallemeersch, Liesbeth Augustinus, Frank Van Eynde, Joris Pelemans, Lyan Verwimp, Patrick Wambacq, Geert Heyman, Marie-Francine Moens, Iulianna van der Lek-Ciudin, Frieda Steurs, Ayla Rigouts Terryn, Els Lefever, Arda Tezcan, Lieve Macken, Sven Coppers, Jan Van den Bergh, Kris Luyten, Karin Coninx

Poster Topics

- Semantic-based alignment
- Syntactic concordancing
- Translation with probabilistic synchronous tree substitution grammar
- A neural network architecture for detecting grammatical errors in SMT
- Domain adaptation for ASR Post-Editing
- A deep classifier for bilingual lexicon induction
- Intelligible and in-context recommendations for translation environments

Your single-source partner for corporate product communication

20th Annual Conference of the European Association for Machine Translation (EAMT)

TM & MT – a happy couple …or how to calculate the potential benefit

Elsy Stolze, STAR Language Technology & Solutions

What added value does MT provide?

- For customers with...
  - an existing translation process
  - extensive Translation Memory and terminology
- We offer: 3-phase proof of concept
  - Initial analysis
  - Productive pilot phase
  - Productive analysis

FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

Massive, open & cross-disciplinary

FOUND CAT

Blended Learning
Flipped Classroom

Course material

LMS plugin

Quiz

Peer review

Weblate as CAT tool
+ our modifications

Suitable for most languages

Crowd sourced translation

Yandex
MyMemory

+ Machine translation

Scientific text

Prof. Dr. rer. nat. Alexander Ferrein
Gary Evans B.Sc.

foundcat.fh-aachen.de | cat.fh-aachen.de
www.fh-aachen.de

Part of the Fellowship -Programm 2016 | www.stifterverband.org/digital-lehrfellows

STIFTERVERBAND

Ministerium für Innovation,
Wissenschaft und Forschung
des Landes Nordrhein-Westfalen

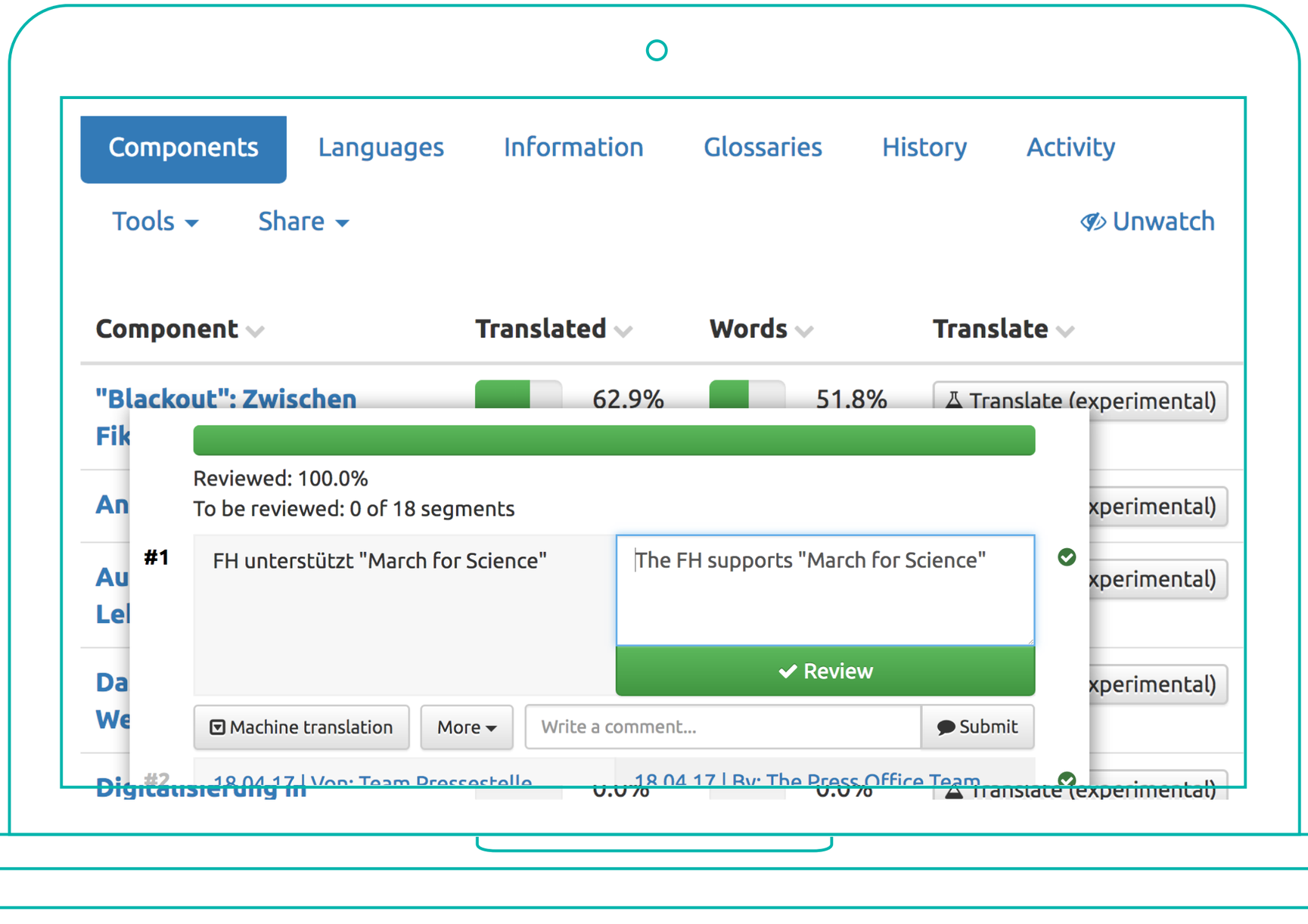
50 Students translated 10 scientific articles in 1 week homework

Peer reviewed.

Hi Guys,

This computer aided translation tool is a very good idea because it helps many people when they are searching for high-quality translations, which you can't find so much on the Internet.

Thanks to this project you can help other students and improve your English skill at the same time, so I did a couple of translations, I hope it will help somebody :)



Prof. Dr. rer. nat. Alexander Ferrein
Gary Evans B.Sc.

foundcat.fh-aachen.de | cat.fh-aachen.de
www.fh-aachen.de

Part of the Fellowship -Programm 2016 | www.stifterverband.org/digital-lehrfellows

STIFTERVERBAND

Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen

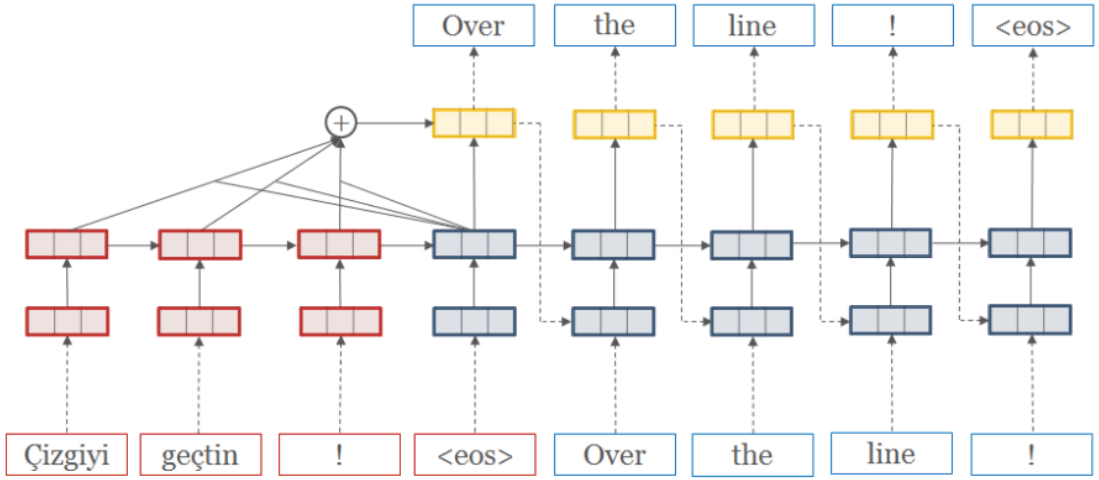
FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

An open-source neural machine translation system.

English Français 简体中文 한국어
日本語 Русский العربية

@harvardnlp
@systran

- OpenNMT is a industrial-strength, open-source (MIT) neural machine translation system utilizing the Torch/PyTorch toolkit.



- Features include:
  - Simple general-purpose interface, requires only src/tgt files.
  - Speed/memory optimizations for **multi-GPU** training.
  - A dependency-free **C++ translator** for model deployment.
  - Latest research features to improve translation performance.
  - Extensions to other sequence generation tasks:
    - Text **summarization** and
    - **Image-to-text** generation.
  - Active **community** with academic and industrial contributions.
  - Pretrained **models** available for several language pairs.

An open-source neural machine translation system.

English Français 简体中文 한국어
日本語 Русский العربية

@harvardnlp
@systran

- State-of-the-art results in WMT 2017
  (English-German news task)

- Ongoing research projects
  - Efficient data sampling strategies for NN training.
  - Better initialization for fast optimization convergence.
  - Probabilistic line search approach for SGD.
    (https://arxiv.org/abs/1703.10034)
  - Multi-encoders for neural machine translation
    (https://arxiv.org/abs/1601.00710)
  - Linear mapping (bridge) between encoder-decoder layers.
  - Modeling coverage for NMT. (https://arxiv.org/pdf/1601.04811)
  - Domain control for NMT. (https://arxiv.org/abs/1612.06140)
  - Hyper-specialization techniques for NMT

data2.unhcr.org/fr/situations/mediterranean

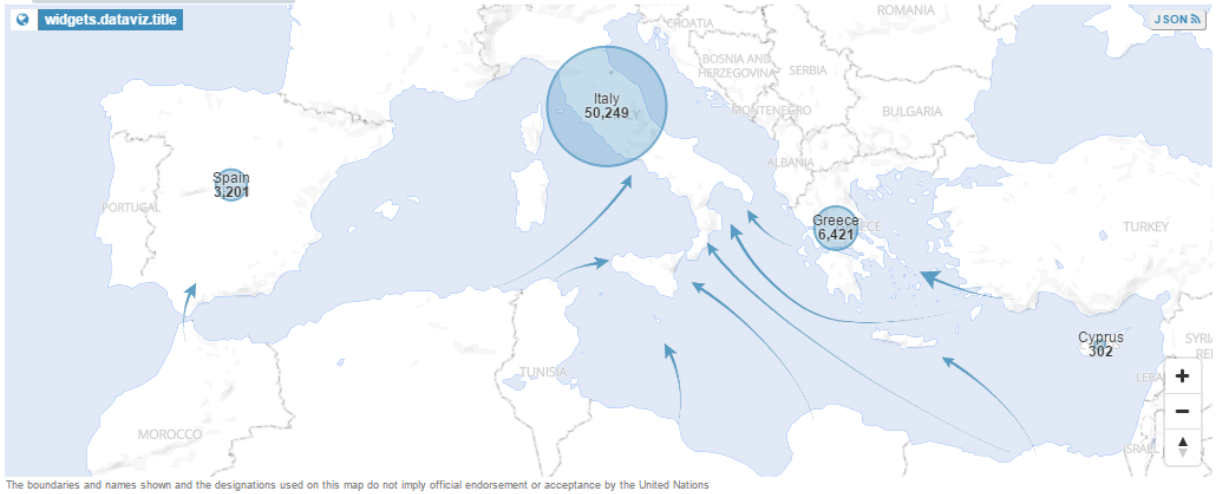
PORTAIL OPÉRATIONNEL CRISES DE RÉFUGIÉS

PAYS | URGENCES | PARTENAIRES | RECHERCHE | à propos | Contact | Maps | Boite a outils IM | API | Français

English
Français
عربي

Mediterranean Situation | Choix de Lieu

TAGS Population Group - / Various - Refugees/Migrants



Population

.CSV | JSON

Country of origin	Source	Data date	Population	
Nigeria		30 Apr 2017	12.2%	5,253
Bangladesh		30 Apr 2017	10.8%	4,645
Guinea		30 Apr 2017	9.7%	4,184
Côte d'Ivoire		30 Apr 2017	9.1%	3,920
Gambia		30 Apr 2017	6.6%	2,850
Syrian Arab Rep.		30 Apr 2017	6.5%	2,786
Senegal		30 Apr 2017	5.9%	2,555

Population

59,871
Last updated 24 May 2017

Population

JSON

362,753
Source - Government
Last updated 31 Dec 2016

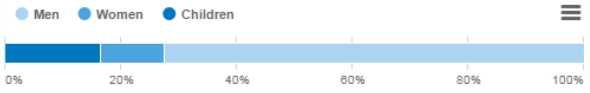
Population

JSON

1,364
Last updated 23 May 2017

Population

JSON



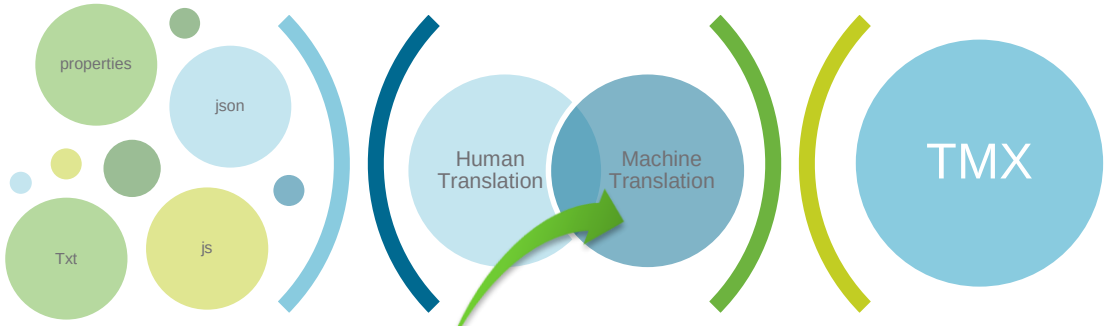
Featured Documents

Interagency Factsheet on refugee and migrant children and UASC in Europe 2016
Télécharger Détails | 1.59 MB |
Téléchargements: 9,675

Is there a case for the use of MT in the Third Social Sector?

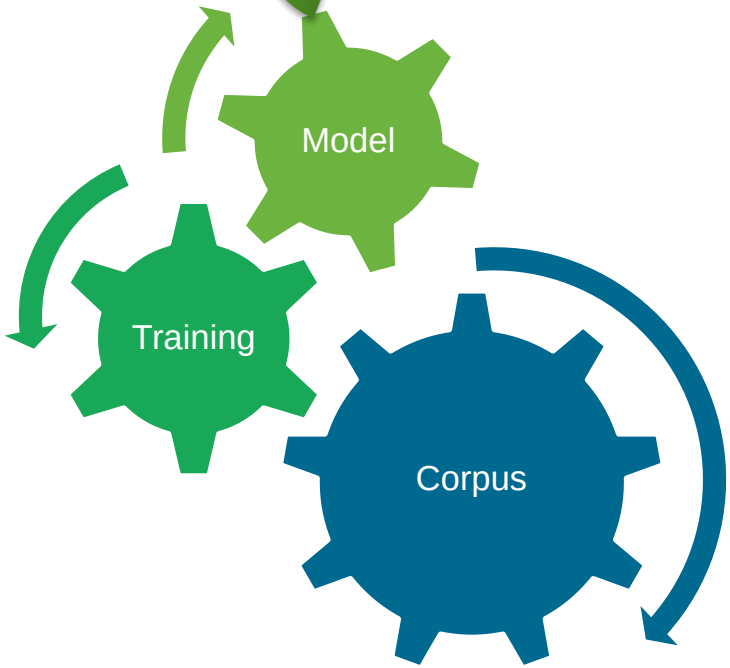
@celiaricoperez
Universidad Europea

L10n



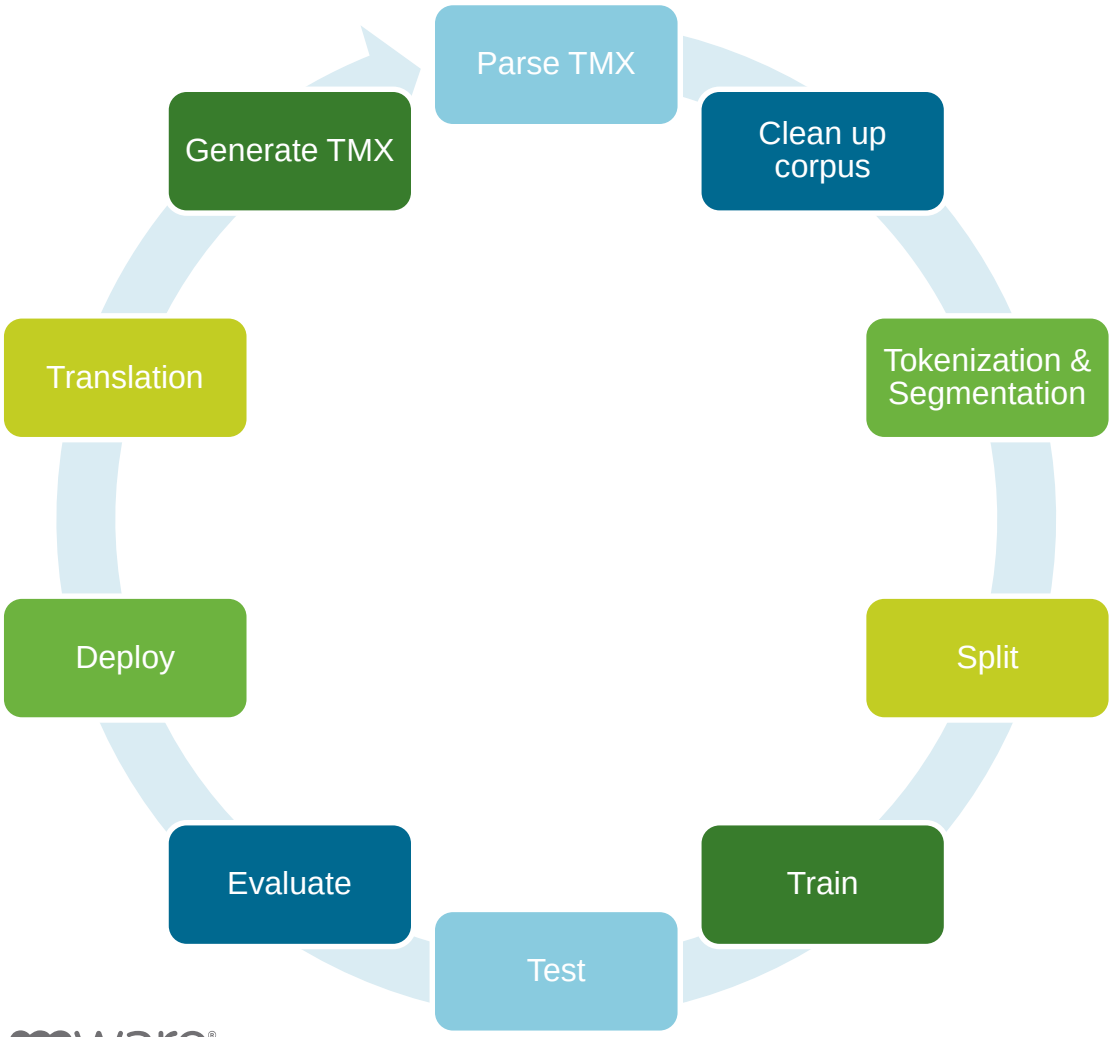
Parsing Translation TM

MT



vmware

OpenNMT Toolset (OMTS)



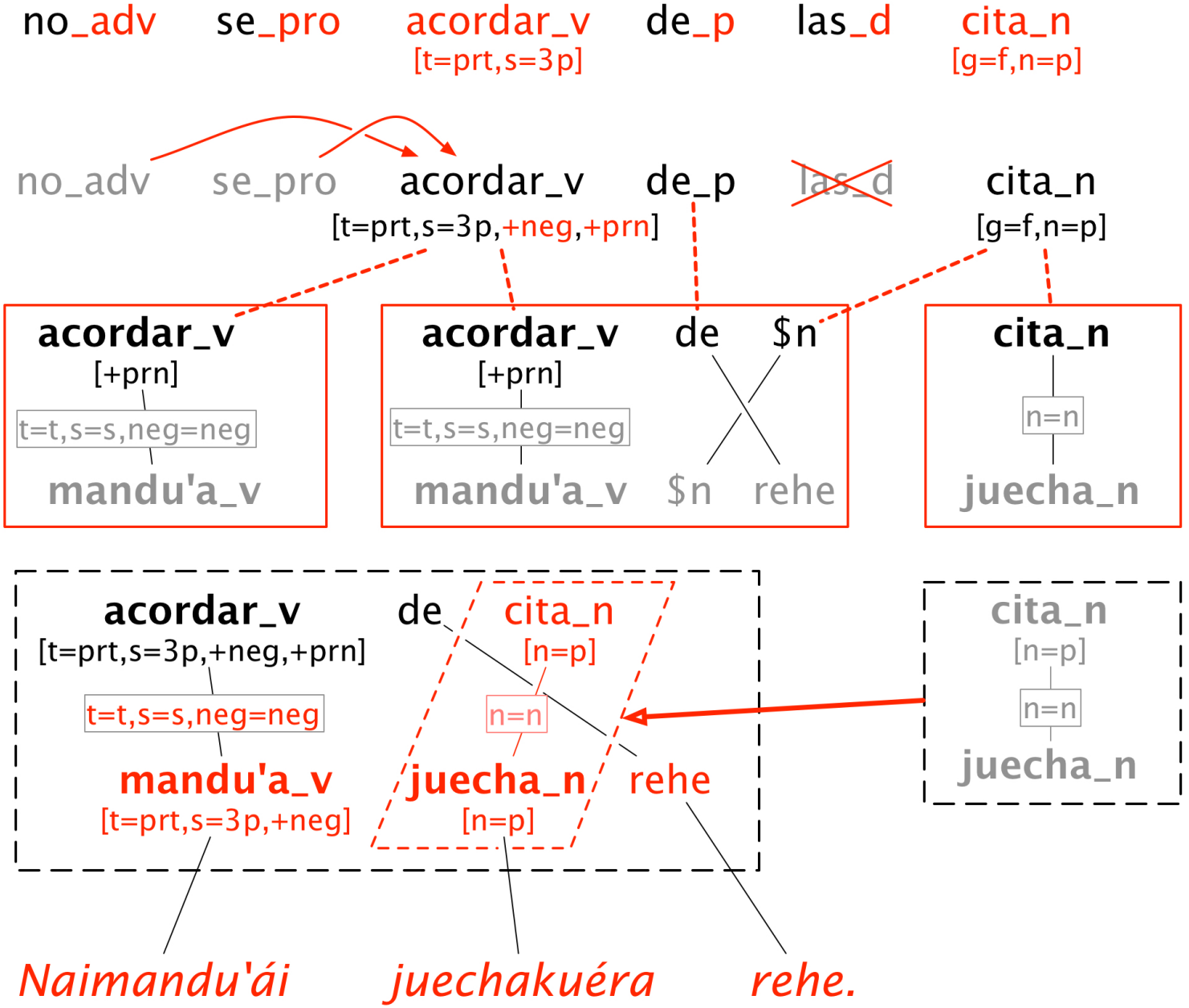
- Streamline the process of creating workable NMT models
- Help users choose the best model by evaluating OpenNMT output
- Integrate neural machine translation into enterprise localization process
- Enable users to try the latest machine translation technology with least effort

vmware

Minimal Dependency Translation:
a Framework for CAT for Under-Resourced Languages

- For under-resourced languages (URLs), lack of written material
- Translation: a partial solution
- But insufficient resources for CAT for URLs
- MDT: phrase-based RBMT for URLs
- Assumed resources
  - SL: tokeniser, POS tagger, morphological analyser
  - TL: morphological generator
- Project goals
  - Open-source tools for implementing MDT
  - Implementations for Spanish–Guarani, English–Amharic

No se acordaron de las citas.



TraMOOC project

Aims:

- a translation platform offering reliable translation services for 11 languages
- target languages include weakly supported and/or morphologically rich languages
- various types of educational MOOC texts – presentations, assignments, video lecture subtitles, forum blog texts

TraMOOC project

Recent results:

- translation systems based on PBMT and NMT
- comparative analysis of two approaches:
  - automatic evaluation
  - human rankings of adequacy and fluency
  - error annotation
  - technical and temporal post-editing effort
- NMT: better ranking, better fluency, lower error count
  however, several aspects are not clear

SkyCode MT – a translation system using deep syntactic and semantic analysis

Luchezar Jackov, SkyCode Ltd.

- **Universal dictionary based on Princeton WordNet**
  The morphological analysis and lexical synthesis are based on lexicalizations bound to PWN synsets

- **Language-independant semantic knowledge base**
  The knowledge base is built around dependency relations on PWN synsets

- **Analysis is split from synthesis and analysis and synthesis rules are shared among similar languages**

- **Compact data representation**
  The total space needed for 42 translation directions is approx. 150 MB

SkyCode MT – a translation system using deep syntactic and semantic analysis

Luchezar Jackov, SkyCode Ltd.

- **Portable implementation**

The system is implemented in C++ which makes it portable across various platforms

- **Currently available for English, German, French, Italian, Spanish, Bulgarian and Turkish**

- **Easy adding of new languages allowing translation from and to all of the languages in the system**

- **The rule-based nature of the system makes it easy to implement under-resourced languages**

ModernMT: A New Open-Source Translation Platform for the Translation Industry

Translated.net

FONDAZIONE BRUNO KESSLER

THE UNIVERSITY of EDINBURGH

TAUS

Real time adaptation to content

No more training or set-up time

Designed to scale for data and users

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 645487.

www.modernmt.eu

ModernMT: A New Open-Source Translation Platform for the Translation Industry

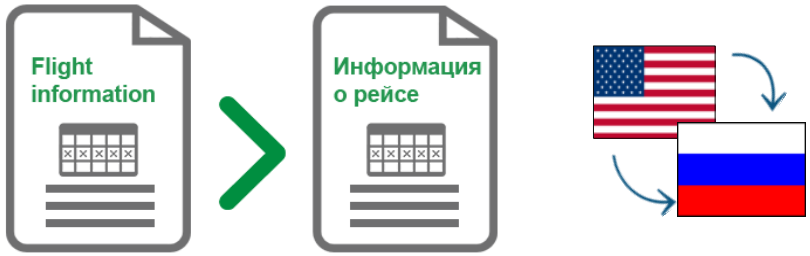
- Currently: classical phrase-based approach
- New data can be added at any time
- Phrase table entries are created on the fly ...
  ... by phrase extraction from a selection of parallel text most similar to the input
- Interpolation of several language models chosen from a large pool of domain / task-specific LMs

- On Github: http://github.com/modernmt

amadeus

GLOBAL DISTRIBUTION SYSTEM

Challenges



Provide custom machine translation of Fare Quote Notes for 5,000+ travel agencies located in Russia, Armenia and Belarus

Business requirements

- Understandable ("good enough") translation of Fare Quote Notes (FQN) from English into Russian with focus on frequent terms and abbreviations
- High performance and reliability even in case of the large number of requests
- Integration in the Amadeus system's interface

PROMT Cloud API

Cloud solution with programming interface (API), ready to process large number of translation requests

Customization

Pre-Processing	Additional Dictionaries	Translation Memory
Special algorithms that take into account FQN format and structure	**~1,200 terms** specific for FQN **~20,000 names** of airline companies and airport codes	professional human translation of the most frequent sentences

Implementation results

- Integration in the Amadeus system:
  - Translation of an entire FQN
  - Translation of current page of an FQN
- PROMT translation takes into account linguistic features of FQN
- Additional functions:
  - Translation of special terms on hover
  - End-users feedback

Usage statistics

4 million words **per week**

6k requests **per week**

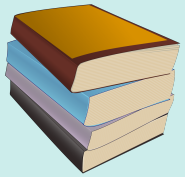
~700 words in every request

Automatic Building of Machine Translation
(2013-2016)

Selected final results

- **Goal** rapid and cost-effective MT for new languages
- **Work** academia ↔ industry secondments
- **Case study** Croatian as a new official EU language
- **Results** publicly available tools and datasets

Data acquisition



Machine Translation

Google Summer of Code

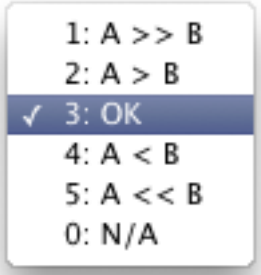
Evaluation

Dissemination

Appraise on Azure

- What is this about?
  - A cloud-based, multi-purpose evaluation framework
  - Used for WMT evaluation campaign – but supports arbitrary annotation tasks!

- What has changed?
  - Latest frameworks: Python 3, Django 1.11
  - Now runs on Azure
    - either inside Ubuntu VM, or
    - natively on Windows Server



- Why should I care?
  - Appraise on Azure allows focus on evaluation, not install/build an eval system

Azure for Research

Cloud computing awards

Big data, real HPC, machine learning, deep learning (GPUs), IoT, and more….

Deadline: 15 June 2017

(and every two months)

- Email:
  - azurerfp@microsoft.com

Microsoft Azure for Research Group

@azure4research

HimL: Health in my Language

Type:	Horizon 2020 Innovation Action
Duration:	February 2015 to January 2018
Website:	`www.himl.eu`

THE UNIVERSITY of EDINBURGH

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Charles University, Prague

LINGEA

NHS 24

Cochrane

Health in my Language HimL

HimL: Aims

We will make public health information available to consumers in their own language by:

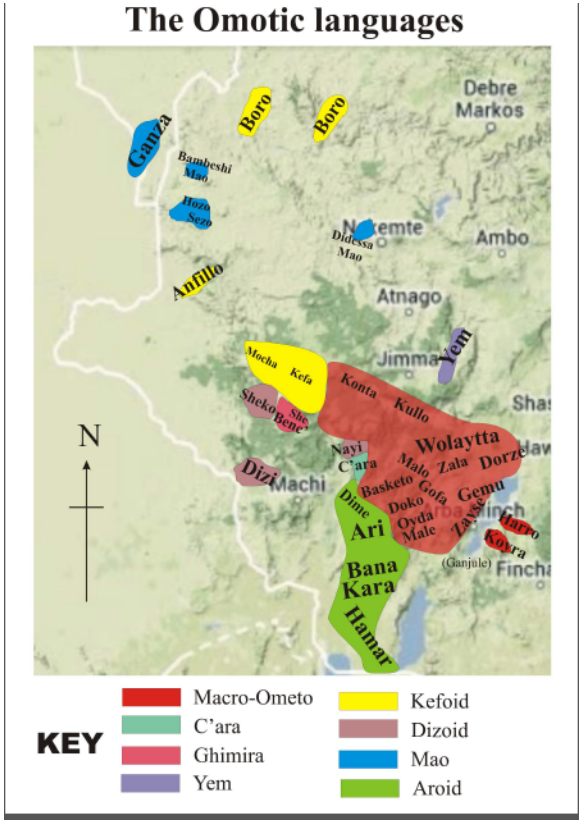
- Deploying MT which is:
  - Adapted to the domain
  - Semantics- and discourse-aware – preserves meaning
  - Can translate into languages with rich morphology found in central and eastern Europe
- Integrating the MT into the content management workflow of the two users (Cochrane and NHS 24)
- Carefully managing user expectations
- Comprehensively evaluating user satisfaction and impact of the new functionality

Health in my Language HimL

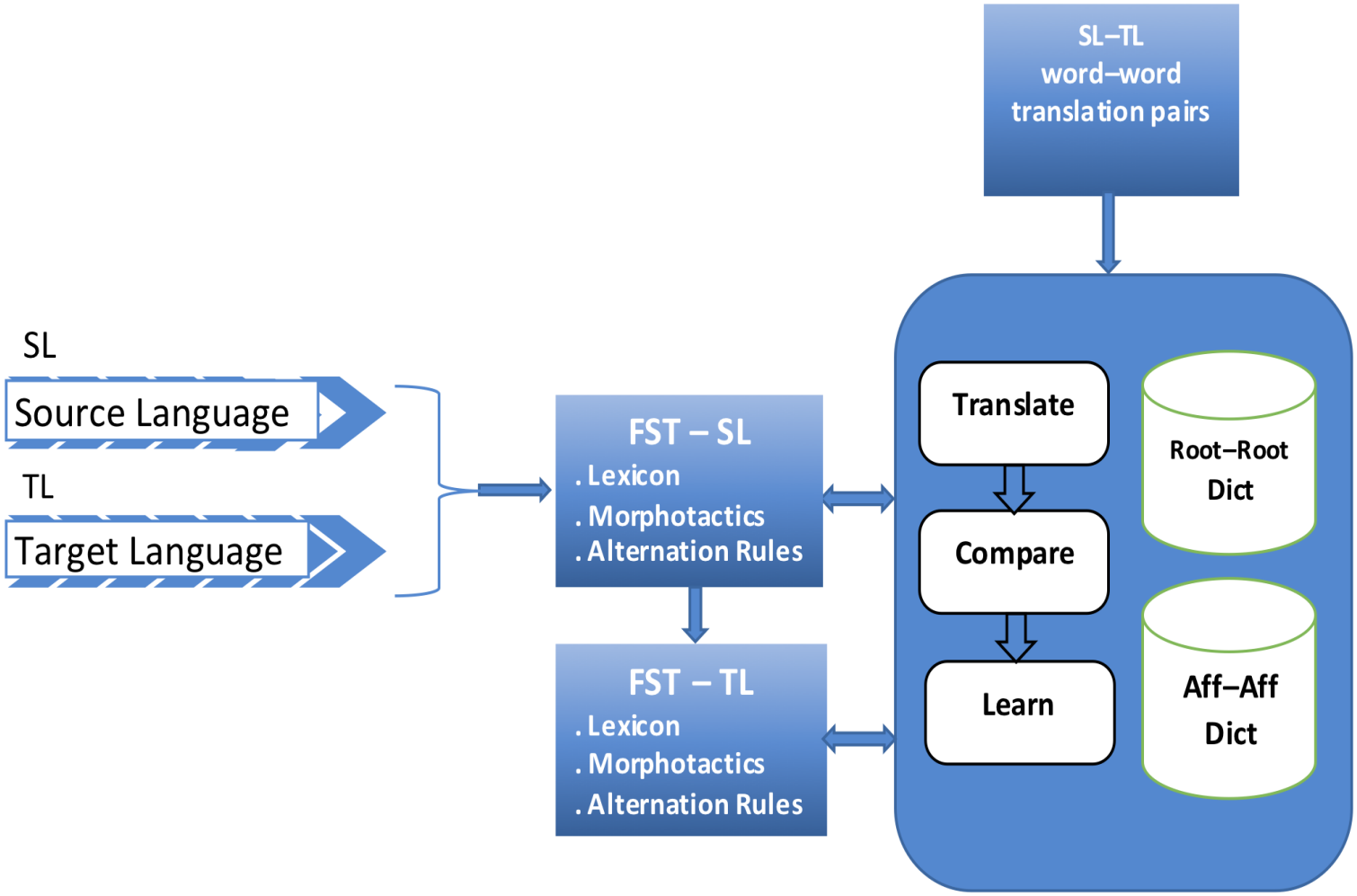
A translation-based approach to the learning of the morphology of an under-resourced language

Tewodros Abebe, Addis Ababa University
Michael Gasser, Indiana University

- Most languages have no computational morphological resources and insufficient data for creating them
- Most under-resourced languages are closely related to better-resourced languages
- Objective: learn the morphology of an under-resourced (target) language using translation from a related well-resourced (source) language
- Wolaytta (source) and Gofa (target)
  - Omotic family of southwestern Ethiopia
  - Wolaytta: best studied, Gofa: very few resources



Architecture and Processing



Your single-source partner for corporate product communication

20th Annual Conference of the European Association for Machine Translation (EAMT)

MT in real-world practice: Challenges & solutions at Swiss Federal Railways

Nadira Hofmann, STAR Language Technology & Solutions

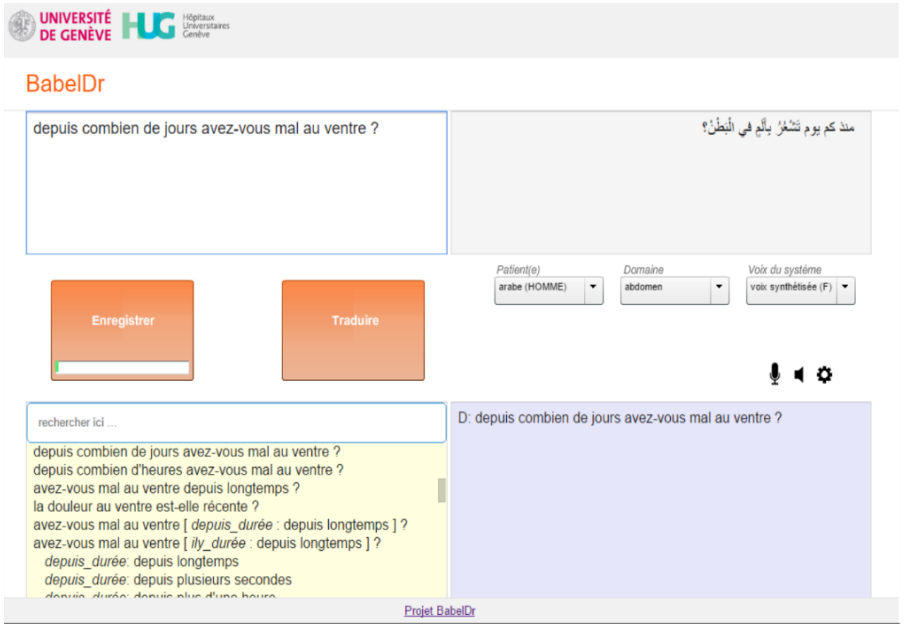
Introducing MT at the Swiss Federal Railways...

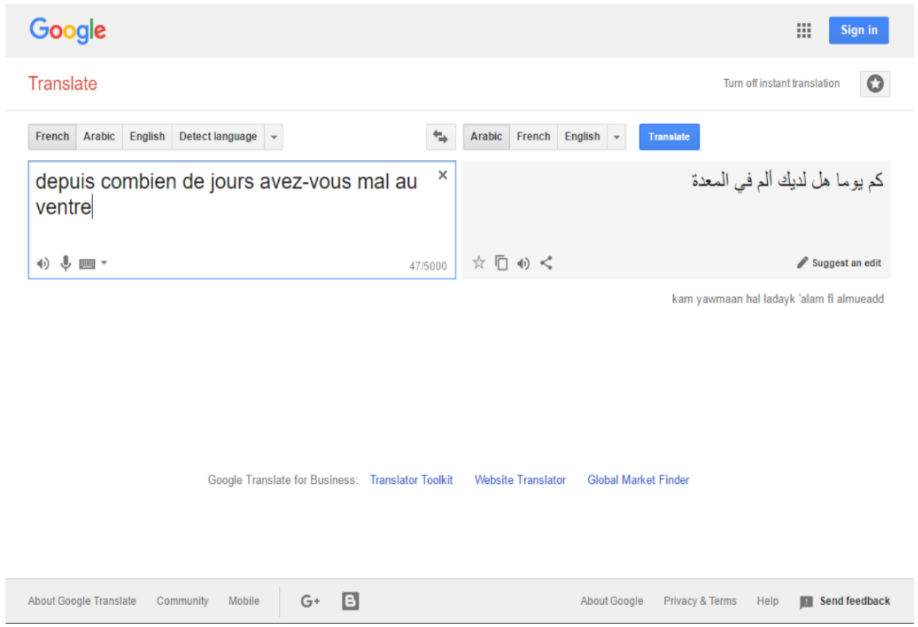
- Preconditions & motivation
- Approaches & requirements
- Deployment
  - Workflow system
  - SBB Translate

Babeldr vs Google Translate: A User Study at Geneva University Hospitals

P. Bouillon, J. Gerlach, H. Spechbach, N. Tsourakis and S. Halimi

- Aim: compare two translation tools for diagnosis in emergency settings





Results

- 18 diagnoses
  - 9 French doctors
  - 2 standardized Arabic-speaking patients
- GT is not precise enough for the anamnesis, even if some doctors could reach a correct conclusion
- BabelDr
  - Was not perceived as less robust and limited
  - Allowed more correct diagnoses

Dissecting Human Pre-Editing toward Better Use of Off-the-Shelf Machine Translation Systems

Rei Miyata (Nagoya University) Atsushi Fujita (NICT)

Objectives

1) Investigate the **capability of the pre-editing strategy**
   - A human-in-the-loop protocol to collect pre-edit instances
   - Japanese-to-English translation tasks on 4 datasets
2) Provide an **overview of possible edit operations**
   - A typology of edit operations

Source text (ST)

出生届を居住区の市区町村の役所に提出

Alternative lexical choice

Phrase reordering

お住まいの市区町村の役所に出生届を

ご提出ください。

Disuse of clause-ending noun
Use of honorific expression

MT output

Birth registrations submitted to the public office of the municipality where you live.

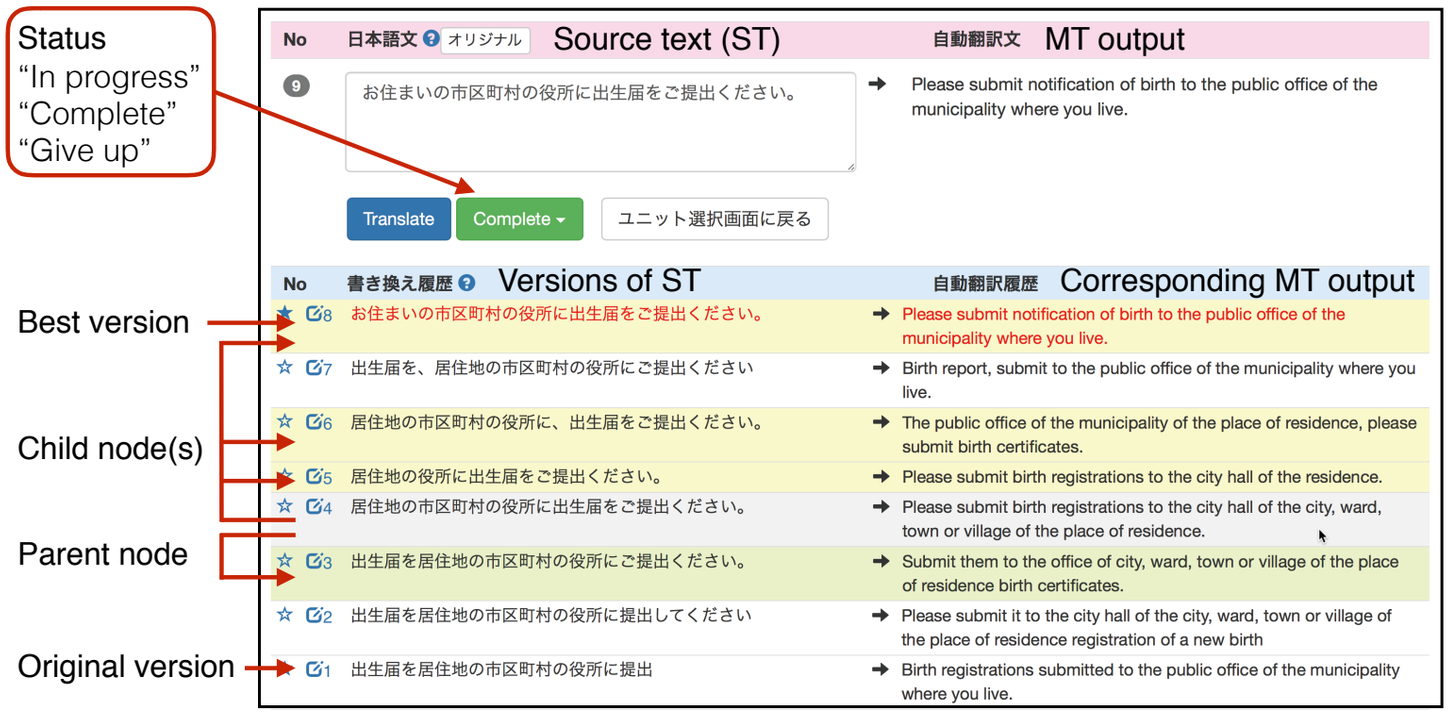
?

Please submit notification of birth to the public office of the municipality where you live.

Protocol for Collecting Pre-Edit Instances

Human editors incrementally edit source texts (STs) relying on their introspection, so that improved MT quality is achieved. **(Miyata et al. 2015)**

New features: (1) Record every minimal edit, (2) Allow reversion of past edits



4 domains (400 sentences)
Ja-En SMT
Ja editor

↓

12,287 pre-edit instances

✓ More than 85% STs achieved satisfactory MT quality

✓ English-translatable Japanese STs are also Chinese- and Korean-translatable

Typology of Edit Operations

979 pre-edit instances

↓

53 types of edit operation

7 major categories

Structure, Content word, Functional word, Terminology, Orthography, Information, Edit that causes/resolves error in ST

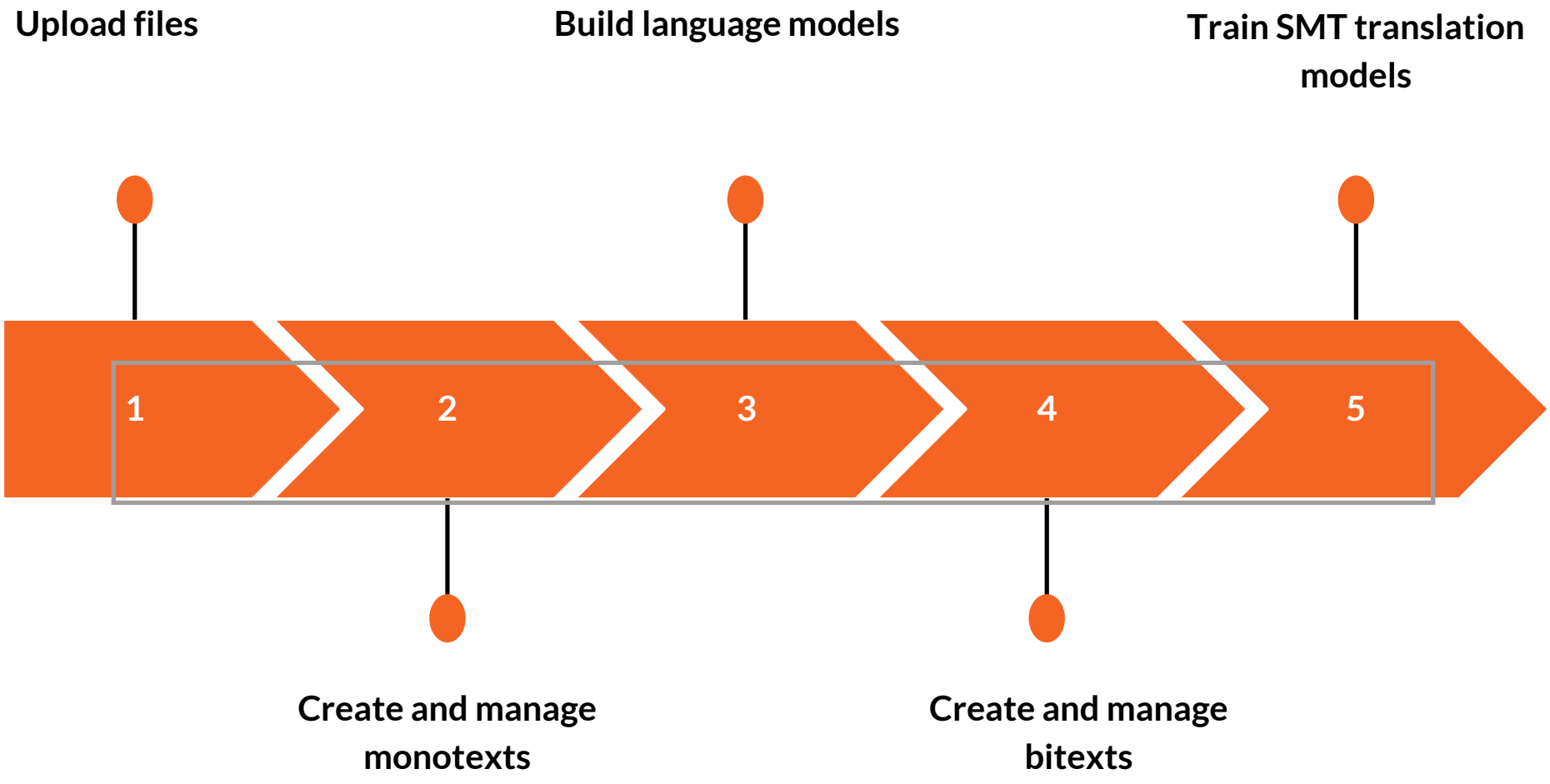
Mtradumàtica

Adrià Martín-Mor (UAB)
Sergio Ortiz-Rojas (Prompsit)
Gökhan Doğru (UAB)

Statistical Machine Translation Customisation for Translators

- Free (GPL) Moses-based platform
- User-friendly web interface
- Installable in private servers
- Understand how the components of the system work together
- m.tradumatica.net

Prepare your SMT system in 5 simple steps



MMT: New Open Source MT for the Translation Industry

Fast initial training (10x faster than Moses)
Seamless assimilation of new user data (while decoding)
Real-time learning (feature computation)
On-the-fly adaptation (from context)
Elastic architecture (from single to multi-node)
Ready-to-install package (docker or binaries)