

Machine Translation Course

**Cristina España-Bonet,
Koel Dutta Chowdhury,
Yasser Hamidullah**

DFKI & UdS

Summer Semester 2023

11th April 2023

Neural Machine Translation reaches historic milestone: human parity for Chinese to English translations

Rate this article ★★★★★

Microsoft Translator March 14, 2018

Share 12

7

in 0

0



Outline

1 Basic Questions

2 Evaluation

3 Resources

4 Stay up-to-date

Basic Questions

Who?

Cristina España i Bonet
Koel Dutta Chowdhury
Yasser Hamidullah



Translation-Oriented Language Technologies,
in Department of Language Science and Technology



Multilingual Technologies MLT Lab,
in Language Technology Lab

Basic Questions

Who?

Prof. Dr. Josef van Genabith Chair

Lehrstuhl Translationsorientierte Sprachtechnologie

Work related to

- Machine Translation
- Information Retrieval
- and Multilingual Technologies in general

Basic Questions

Who?

The team

**Josef van Genabith, Cristina España-Bonet, Yasser
Hamidullah, Koel Dutta Chowdhury, Daria Pylypenko,
Uliana Sentsova, Sepideh Mollanorozy,
+ masters, +HiWis!**

We usually have HiWi positions (**not now** :-())
and/or thesis master topics
—examples later related to MT if we have time.

Basic Questions

What and for who?

- Introduction to machine translation going from statistical to neural machine translation systems and paying special attention to the current state of the art.
- A strong background on mathematics, statistics and machine learning is not required but a basic knowledge is expected to follow the course.

Basic Questions

Main Blocks

- 1 Introduction, Basics & Evaluation
- 2 Statistical Machine Translation
- 3 Neural Machine Translation
- 4 State of the Art (selected topics!)

Basic Questions

How?

- Classes with slides
 - Introduction to MT, SMT basics...
 - Some slides public, some protected with pwd (not sharable because of non-open content)
- Classes with slides+laptop??
 - Embeddings
- 2 Labs (leading to a mandatory assignment)
 - SMT, NMT
- Bonus: a challenge!

Evaluation

Get your credits with...

2 assignments (30%) and a written exam (70%)

- The MT challenge is a bonus for just a person (+10%)
- You need to submit the labs to do the exam, you need to pass the exam to get the bonus

Evaluation

Assignments

- 1 Report on SMT Lab (15%)
- 2 Report on NMT Lab (15%)
- 3 Written exam (70%)

Evaluation

Assignments

1 Report on SMT Lab (15%)

~ week 8

2 Report on NMT Lab (15%)

~ week 13

3 Written exam (70%)

week 15: 17/07/2023

There is no textbook for the course, but you can find general information in the following references. More specific texts or slides will be provided for each session.

- Statistical Machine Translation book by Philipp Koehn
- From SMT to NMT, three posts in a blog to understand the basics of NMT
- Chapter on Neural Machine Translation by Philipp Koehn

Resources

Classic Book Chapters

- Manning & Schütze, *Foundations of Statistical Natural Language Processing*, 1999, Chapter 13: Statistical Alignment and Machine Translation
- Jurafsky & Martin, *Speech and Language Processing*, Second Edition, 2009, Chapter 25: Machine Translation
- *Handbook of Natural Language Processing*, Second Edition, 2010, (eds.) Nitin Indurkha and Fred J. Damerau, Chapter 17: Statistical Machine Translation (Abraham Ittycheria)
- *Handbook of Computational Linguistics and Natural Language Processing*, 2010, (eds.) Alexander Clarke, Chris Fox and Shalom Lappin. Chapter 19: Machine Translation (Any Way)

Resources

Conferences & Journals

State of the art in **conferences** such as:

- ACL, EMNLP, COLING, CoNLL, ...
- WMT, MT-SUMMIT, AAMT, EAMT, ...

See last advances also in **journals**:

- Computational Linguistics, Machine Translation, ...

Not many journals in the field!

- **WMT**: <http://www.statmt.org/wmt23>
- **IWSLT**: <https://iwslt.org/2023/>
- **WMT-SL**: <https://www.wmt-slt.com/>
(coorganised by DFKI)

You can find data, baseline systems, know what others do, what's best, differences among language pairs, etc.



EMNLP 2023 EIGHTH CONFERENCE ON MACHINE TRANSLATION (WMT23)

December 6-7, 2023
Singapore

Home

[\[HOME\]](#)

TRANSLATION TASKS: [\[GENERAL MT \(NEWS\)\]](#) [\[TERMINOLOGY\]](#) [\[INDIC MT\]](#) [\[SIGN LANGUAGE\]](#)

EVALUATION TASKS: [\[QUALITY ESTIMATION\]](#)

OTHER TASKS: [\[WORD-LEVEL AUTOCOMPLETION\]](#)

This conference builds on a series of annual workshops and conferences on statistical machine translation, going back to 2006:

- the [NAACL-2006 Workshop on Statistical Machine Translation](#),
- the [ACL-2007 Workshop on Statistical Machine Translation](#),
- the [ACL-2008 Workshop on Statistical Machine Translation](#),
- the [EACL-2009 Workshop on Statistical Machine Translation](#),
- the [ACL-2010 Workshop on Statistical Machine Translation](#),
- the [EMNLP-2011 Workshop on Statistical Machine Translation](#),
- the [NAACL-2012 Workshop on Statistical Machine Translation](#),
- the [ACL-2013 Workshop on Statistical Machine Translation](#),
- the [ACL-2014 Workshop on Statistical Machine Translation](#),
- the [EMNLP-2015 Workshop on Statistical Machine Translation](#),
- the [First Conference on Machine Translation \(at ACL-2016\)](#),
- the [Second Conference on Machine Translation \(at EMNLP-2017\)](#),
- the [Third Conference on Machine Translation \(at EMNLP-2018\)](#),
- the [Fourth Conference on Machine Translation \(at ACL-2019\)](#),
- the [Fifth Conference on Machine Translation \(at EMNLP-2020\)](#),
- the [Sixth Conference on Machine Translation \(at EMNLP-2021\)](#),
- the [Seventh Conference on Machine Translation \(at EMNLP-2022\)](#).

is from @iwslt

IWSLT
@iwslt · Apr 5

Shared Tasks
ion period has started!
em submission due
ril 18, 2023
e info: iwslt.org/2023/

participants who
registered yet please
r for the Evaluation
ign via the registration



6

IWSLT Retweeted

EMCLT
@E... · Mar 15

py to promote the
Task on low-resource
translation at [IWSLT](#)

20 Years of IWSLT

13-14 July, 2023 - Toronto

The International Conference on Spoken Language Translation (IWSLT) is an annual scientific conference, associated with an open evaluation campaign on spoken language translation, where both scientific papers and system descriptions are presented.

The 20th edition of IWSLT will be co-located with [ACL 2023](#) in Toronto, Canada on 13-14 July, 2023, and run as a hybrid event.

Shared Tasks

The IWSLT 2023 Evaluation Campaign will host shared tasks featuring the following focus areas:

- **Speech translation of talks:**
 - [Offline track](#) (Marco Turchi, Zoom, Germany, Matteo Negri, FBK, Italy)
 - [Simultaneous track](#) (Xutai Ma, Juan Pino, Meta; Katsuhito Sudoh, NAIST, Japan)
 - [Subtitling track](#) (Mauro Cettolo, FBK, Italy; Evgeny Matusov, AppTek, Germany)
 - [Multilingual track: ACL 60-60 Initiative](#) (Elizabeth Salesky, JHU; Jan Niehues, KIT)
- **Speech-to-speech translation of multi-source data:**
 - [Speech-to-speech track](#) (Qianqian Dong, Tom Ko, Mingxuan Wang, Bytedance, China)

Important Links

13-14 July, 2023 - Toronto

Shared Tasks

Registration

Paper Submission

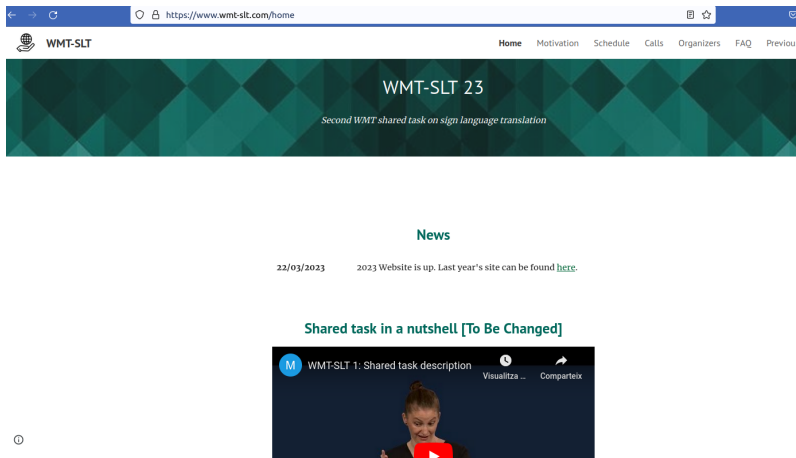
Important Dates

Timeline

Organizers

Program Committee

Sponsorship Opportunities



The screenshot shows the homepage of the WMT-SLT 23 website. The browser's address bar displays the URL <https://www.wmt-slt.com/home>. The website's header includes the WMT-SLT logo and a navigation menu with links to Home, Motivation, Schedule, Calls, Organizers, FAQ, and Previous. The main banner features a green geometric pattern and the text "WMT-SLT 23" and "Second WMT shared task on sign language translation". Below the banner, a "News" section contains a post dated 22/03/2023 stating that the 2023 website is up and that the previous year's site can be found [here](#). A section titled "Shared task in a nutshell [To Be Changed]" contains a video player. The video player shows a thumbnail of a woman speaking, with a red play button in the center. Above the video, the title "WMT-SLT 1: Shared task description" is visible, along with icons for "Visualiza ..." and "Comparteix".

WMT-SLT

Home Motivation Schedule Calls Organizers FAQ Previous

WMT-SLT 23

Second WMT shared task on sign language translation

News

22/03/2023 2023 Website is up. Last year's site can be found [here](#).

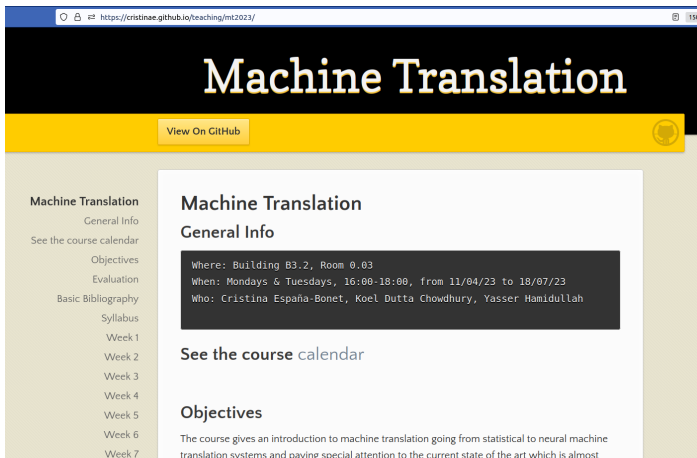
Shared task in a nutshell [To Be Changed]

M WMT-SLT 1: Shared task description Visualiza ... Comparteix

0

Stay up-to-date

Webpage



The screenshot shows a web browser window with the address bar displaying `https://cristinae.github.io/teaching/mt2023/`. The page has a black header with the title "Machine Translation" in a large, yellow-outlined font. Below the header is a yellow bar containing a "View On GitHub" button and a GitHub logo. The main content area has a light beige background. On the left, there is a sidebar with a list of links: "Machine Translation", "General Info", "See the course calendar", "Objectives", "Evaluation", "Basic Bibliography", "Syllabus", "Week 1", "Week 2", "Week 3", "Week 4", "Week 5", "Week 6", and "Week 7". The main content area features a section titled "Machine Translation" with a sub-section "General Info". Below this, a dark grey box contains the following text: "Where: Building B3.2, Room 0.03", "When: Mondays & Tuesdays, 16:00-18:00, from 11/04/23 to 18/07/23", and "Who: Cristina España-Bonet, Koel Dutta Chowdhury, Yasser Hamidullah". Below the dark grey box, there is a link "See the course calendar" and a section titled "Objectives" with a paragraph: "The course gives an introduction to machine translation going from statistical to neural machine translation systems and paying special attention to the current state of the art which is almost".

Machine Translation

General Info

See the course calendar

Objectives

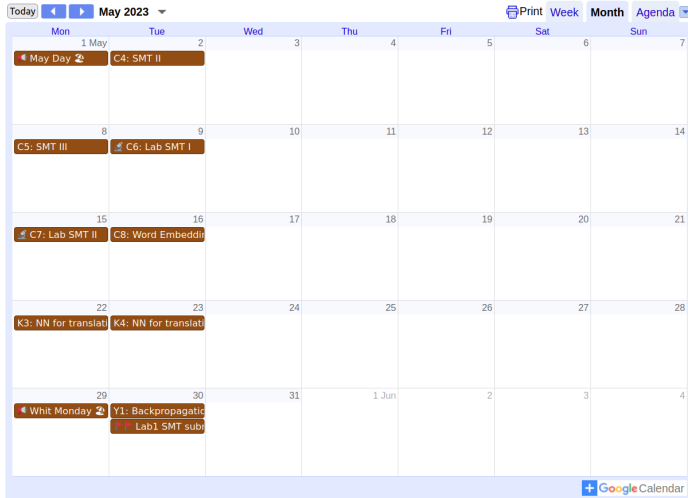
The course gives an introduction to machine translation going from statistical to neural machine translation systems and paying special attention to the current state of the art which is almost

<https://cristinae.github.io/teaching/mt2023/>


Stay up-to-date

Calendar

MT



Stay up-to-date

A close-up photograph of a typewriter keyboard. The focus is on a sheet of white paper with the word "Questions?" typed in a dark, serif font. The typewriter's metal frame and keys are visible in the background and foreground, creating a sense of depth. The lighting is soft, highlighting the texture of the paper and the metallic surfaces.

Questions?

Machine Translation Course

**Cristina España-Bonet,
Koel Dutta Chowdhury,
Yasser Hamidullah**

DFKI & UdS

Summer Semester 2023

11th April 2023