

Universitat Politècnica de València
Master in Artificial Intelligence, Pattern Recognition and Digital Imaging
2023-2024

MACHINE TRANSLATION

Francisco Casacuberta

Miguel Domingo

`fcn@prhlt.upv.es`

`midobal@prhlt.upv.es`

Topics

1. Introduction to machine translation
2. Statistical machine translation
3. Neural machine translation
4. Interactive machine translation
5. Advanced topics in neural machine translation
6. External knowledge in neural machine translation

Scheduling

Date	Session	Teacher	Topic
7/11/2023	1	F. Casacuberta	Introduction to machine translation
14/11/2023	2	F. Casacuberta	Statistical machine translation
21/11/2023	3	F. Casacuberta	Neural machine translation
28/11/2023			RAH
12/12/2023	4	F. Casacuberta	Neural machine translation
12/12/2023	5	M. Domingo	Practical exercise SMT
19/12/2023	6	M. Domingo	Practical exercise NMT
9/ 1/2024	7	F. Casacuberta	Interactive machine translation
16/ 1/2024	8	M. Domingo	Practical exercise SMT+NMT
23/ 1/2024	9	F. Casacuberta	Advanced topics in machine translation
30/ 1/2024	10	F. Casacuberta	External knowledge in NMT

Assessment

- The course must be passed through a test, courseworks and lab exercises:
 - Test (20%). February 9th, 2024.
 - Report of 3 courseworks (60%), deadline: February 1st, 2024
 - * SMT (3.5 out of 10 points -21%-)
 - * NMT (3.5 out of 10 points -21%-)
 - * NMT extension (3 out of 10 points -18%-)
 - Report of 2 labworks (20%), deadline: February 4th, 2024.
 - * SMT (10%).
 - * NMT (10%).
- “Actas”: February 21th, 2024.

Lab work

1. Statistical machine translation

1.1 Building a first SMT system.

1.2 Experiments with Moses.

<http://www.statmt.org/moses/>

2. Neural machine translation

2.1 Building a first NMT system.

2.2 Experiments with OpenNMT-py.

https://github.com/PRHLT/OpenNMT-py/tree/lab_sessions

Bibliography on MT

1. Philipp Koehn. *Statistical Machine Translation*. Cambridge University Press, 2010.
2. Philipp Koehn. *Neural Machine Translation*, Cambridge University Press. 2020.
3. S. Barrachina, O. Bender, F. Casacuberta, J. Civera, E. Cubel, S. Khadivi, A. Lagarda H. Ney, J. Tomás, and E. Vidal. *Statistical approaches to computer-assisted translation*. Computational Linguistics, 35(1):3-28, 2009.
4. A. Peris, M. Domingo, F. Casacuberta. *Interactive Neural Machine Translation*. Computer Speech and Language, 45:201-220. 2017.
5. Ye Jia, Ron J. Weiss, Fadi Biadisy, Wolfgang Macherey, Melvin Johnson, Zhifeng Chen, Yonghui Wu. Direct speech-to-speech translation with a sequence-to-sequence model. arXiv 2019.

Links

1. *Association for Machine Translation in the Americas (AMTA).*
URL: <https://amtaweb.org/>
2. *European Association for Machine Translation (EAMT).*
URL: <http://www.eamt.org/>
3. *Asia-Pacific Association for Machine Translation (AAMT).*
URL: <http://www.aamt.info/>
4. *TAUS Data Association.*
URL: <http://www.tausdata.org/>
5. *mgiza: A word alignment tool based on famous GIZA++*
URL: <https://github.com/moses-smt/mgiza>
6. *Thot: a toolkit to train phrase-based models for statistical machine translation.*
URL: <https://github.com/daormar/thot>
7. *Moses: a factored phrase-based beam-search decoder for machine translation.*
URL: <http://www.statmt.org/moses/>
8. *CasMaCat: the next generation translator's workbench.*
URL: <http://www.casmacat.eu/>
9. *NMT-Keras: Neural Machine Translation with Keras.*
URL: <https://github.com/lvapeab/nmt-keras/>
10. *MarianNMT: Fast Neural Machine Translation in C++.*
URL: <https://marian-nmt.github.io/>
11. *OpenNMT: An open source neural machine translation system.*
URL: <https://opennmt.net/>

Journals

- Machine Translation

<https://www.springer.com/journal/10590>

- Computational Linguistics

<http://www.mitpressjournals.org/loi/coli>

- Natural Language Engineering

<http://journals.cambridge.org/action/displayJournal?jid=NLE>

- Computer Speech & Language

<http://www.journals.elsevier.com/computer-speech-and-language/>

- IEEE/ACM Transactions on Audio, Speech, and Language Processing

<http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=10376>

- Transactions of the Association for Computational Linguistics

<http://www.transacl.org/>

Main conferences

- AMTA conferences
- EAMT conferences
- Machine Translation Summits (MTSummit).
- Annual Meeting of the Association for Computational Linguistics (ACL)
- Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)
- European Chapter of the ACL (EACL)
- International Conference on Computational Linguistics (COLING)
- Conference on Machine Translation (Old Workshop on Statistical Machine Translation -WMT-)
- International Workshop on Spoken Language Translation (IWSLT)
- The International Conference on Learning Representations (ICLR)

- ACL Anthology

<https://aclanthology.org/>

- Machine Translation Reading List

<https://github.com/THUNLP-MT/MT-Reading-List>

Some translation companies located in Spain

- **Pangeanic** <https://www.pangeanic.com/>
- **Webinterpret** <https://www.webinterpret.com/us/>
- **Ubiquus** (Celer Soluciones+Traducciones Políglota) <https://www.ubiquus.com/en-us/>
- **Prompsit** <http://www.prompsit.com/>
- **Gear translations** <https://www.geartranslations.com/en/>
- **Bigtranslation** <https://www.bigtranslation.com/>
- ...

Some demos & translators

- Neural Machine Translation

<https://demosmt.prhlt.upv.es/inmt/>

- CasMaCat: An Interactive Machine Translation Prototype

<https://demosmt.prhlt.upv.es/matecat-test/translate/demo-eutt2.xliff/es/1-fmttqpvc/demo@4002#1>

- SisHiTra

<https://demosmt.prhlt.upv.es/sishitra/>

- Moses

<http://demo.statmt.org/>

- GoogleTrans

http://www.google.com/language_tools

- Microsoft Bing

<https://www.bing.com/translator>

- DeepL

<https://www.deepl.com/en/translator>

- Translate.com

<https://www.translate.com/>