

# Marco Turchi, PhD

## WORKING EXPERIENCES:

 FONDAZIONE BRUNO KESSLER October 2012 – onwards	Researcher. Fondazione Bruno Kessler, Trento, Italy  Topic: Statistical Machine Translation and Sentiment Analysis
 EUROPEAN COMMISSION October 2009 – October 2012	Researcher. Joint Research Centre - JRC - European Commission, Ispra, Italy  Topic: Statistical Machine Translation and Multilingual Text Analysis
 University of BRISTOL January 2007 – August 2009	Research Assistant. Department of Engineering Mathematics, University of Bristol, U.K.  Topic: Statistical Multilingual Analysis for Retrieval and Translation
 University of BRISTOL September 2006 – January 2007	Visiting Fellow. Department of Engineering Mathematics, University of Bristol, U.K.  Topic: Research activity on Machine Learning, Natural Language Processing and Web Mining. Advisor: Prof. Nello Cristianini.
 March 2006 – September 2006	Research Internship on Machine Learning and Natural Language Processing. Xerox Research Centre Europe, Grenoble France.  Topic: "Automatic Hyperlink Assistant for Wikipedia Writer" (Pump Priming: Semantic Information Extraction using Integrated Probabilistic Techniques - Pascal Project). Advisors: Dr. Eric Gaussier and Dr. Cyril Goutte
 June 2005– September 2005.	Research Internship on Machine Learning and Text Mining. Yahoo Research Lab, Pasadena California U.S.  Topic: "Feature Selection using unlabelled documents". Advisor: Dr. Dennis De Coste
 UNIVERSITY OF CALIFORNIA November 2004 – March 2006.	Visiting Fellow. Statistical Department, University of California, Davis, U.S.  Topic: Machine Learning, Natural Language Processing and Web Mining. Advisor: Prof. Nello Cristianini
December 1995 – current	Several experiences as: <ul style="list-style-type: none"> <li>• Lecturer and Network Administrator of Windows 2000</li> <li>• Programmer and System Administrator</li> <li>• Reporter</li> </ul>

## Education:

- March 2006.

Ph.D. in "Computer Engineering, adaptive information processing".

Dissertation on "Time Series Analysis of Textual Data". Supervisor: Prof. M. Maggini.

Department of Information Engineering, University of Siena, Italy

- October 2002.

MSc's degree in Information Engineering.

Dissertation on: "Detection of Web page components: an approach based on visual layout" Supervisors: Prof.

Ing. Marco Maggini, Ing. Michelangelo Diligenti

Department of Information Engineering, University of Siena, Italy.

## Skills:

Computing	<p>Excellent in Java, Python, matlab, C++, mySql and Linux. Basic Knowledge of Ansi c, html, Perl, scipy. Excellent user's experience of Java search engine: Apache Lucene, ElasticSearch, General Architecture for Text Engineering: Gate, Weka, Apache Cayenne, Statistical Machine Translation toolkit Moses, LingPipe and High Performance Computing (HPC) technology.</p>
Communication	<p>Several platform presentations to international conferences and workshops. Selected Conferences:</p> <ul style="list-style-type: none"><li>• 11<sup>th</sup> Biennial Conference of the Association for Machine Translation in the Americas (AMTA'14)</li><li>• 25<sup>th</sup> International Conference on Computational Linguistics (COLING'14)</li><li>• IEEE / WIC / ACM International Conferences on Web Intelligence (WI'09) and Intelligent Agent Technology (IAT'09)</li><li>• 3<sup>rd</sup> Workshop on Statistical Machine Translation, ACL'08</li><li>• IEEE/WIC/ACM International Joint Conference on Web Intelligence WI'04</li></ul>
Collaboration	<p>Several activities with the following researchers outside my group: Prof. Nello Cristianini - University of Bristol, Bristol, UK PhD. Cyril Goutte - National Research Council Canada, Gatineau, QC, Canada PhD. Lucia Specia - University of Sheffield, Sheffield, UK Prof. Marco Riani - Univirsity of Parma, Parma, Italy PhD. Domenico Perrotta - European Commission JRC, Ispra, Italy PhD. Elisa Ricci - Fondazione Bruno Kessler, Trento, Italy</p>
Organization	<p>Co-organizer of:</p> <ul style="list-style-type: none"><li>• 10<sup>th</sup>, 11<sup>th</sup> and 12<sup>th</sup> Workshop on Statistical Machine Translation, 2015, 2016 and 2017</li><li>• 9<sup>th</sup> Machine Translation Marathon, Trento, September 2014</li><li>• Intelligent Analysis and Processing of Web News Content workshop at WI-IAT - Milan 15 September 2009</li><li>• Statistical Multilingual Analysis for Retrieval and Translation associated workshop at EAMT - Barcelona 13 May 2009</li><li>• European Project SMART Meeting in Bristol, May, 2008.</li></ul> <p>Coordinator and head coach of basketball teams from September 1993.</p>

Fund Raising	<p>I have been involved in the preparation of several research proposals.</p> <p>Fund Proposal:</p> <ul style="list-style-type: none"> <li>• QT21 (EU Project Horizon 2020)</li> <li>• “Adaptive Quality Estimation for SAP”. Industrial project with SAP, Germany.</li> <li>• European Association for Machine Translation (EAMT) summer internship project titled: “Open source software development for tidying up and extending translation memories”.</li> <li>• Innovative Project Competition: Automatic multilingual indexing of parliamentary documents Ralf Steinberger, Marco Turchi &amp; Erik van der Goot, IPSC-GlobeSec-OPTIMA</li> </ul>
Supervision	<p>Phd. Student:</p> <ul style="list-style-type: none"> <li>• 2017: Ruchit Agrawal</li> <li>• 2017: Amirohossein Tebbifakhr</li> <li>• 2014: Rajen Chatterjee</li> </ul> <p>Former Phd. Student:</p> <ul style="list-style-type: none"> <li>• Jose Guilherme Camargo de Souza (eBay)</li> </ul> <p>Internship Student:</p> <ul style="list-style-type: none"> <li>• 2016: Gebremedhen Gebremelak (University of Trento, Italy)</li> <li>• 2015: Masoud Jalili (University of Tehran, Iran)</li> <li>• 2014: Marion Weller (University of Stuttgart, Germany).</li> <li>• 2014: Hamed Zamani (University of Tehran, Iran)</li> <li>• 2014: Daniel Torregrosa Rivero (University of Alicante, Spain)</li> <li>• 2013: Antonis Anastasopoulos (National Technical University of Athens, Greece)</li> <li>• 2013: Mihael Arcan (Digital Enterprise Research Institute (DERI), National University of Ireland, Galway, Ireland)</li> </ul> <p>Bachelor Student Master Thesis:</p> <ul style="list-style-type: none"> <li>• 2013: Antonis Anastasopoulos (National Technical University of Athens, Greece)</li> </ul>
Other	<p>I have served as a PC Member for several international conferences and workshops and as reviewer for international journals.</p>

### Languages:

Italian: Mother Tongue	English: Proficient	French: Basic
------------------------	---------------------	---------------

### Selected Invited Talks:

Platform presentation at:

- Invited Speaker at Errare Workshop, Sinaia, Romania, September 2015
- Invited Speaker at TAUS Quality Evaluation Summit, Dublin, June 2014
- Invited Speaker at 9<sup>th</sup> MT Marathon, Trento, September 2014
- Invited Speaker at 6<sup>th</sup> MT Marathon, Trento, September 2011
- LIMSI/CNRS, Spoken Language Processing Group, Orsay July, 2011;
- Translation Forum, DG Translation, Luxembourg and Brussels June 2011;
- University of Zurich, Institute for Computational Linguistics, May 2010;
- University of Southampton, Information Signals, Images, Systems (ISIS) Research Group, June 2009;

- HPC Symposium, Bristol September 2008.

### **Research Activities:**

My research has been mostly focused on the development and application of machine learning techniques to NLP-related problems with particular emphasis to Machine Translation.

In recent years, I have analyzed different aspects of text/web mining, moving from classical classification tasks such as supervised and unsupervised learning including the exploration of different document representations to more computational linguistics tasks such as Statistical Machine Translation, multi-document summarization and sentiment analysis.

My current research is centered on the development of new machine translation technologies for professional translators in a Computer-Assisted Tool scenario. In particular, I'm working on automatic correcting the outputs of a machine translation system by leveraging human annotated data and post-edits and on neural machine Translation.

During my career, I have actively contributed with my research to several Eu projects, i.e. SMART, EU-Bridge, CoSyne, Matecat, QT21 and Cracker.

**References:** available on request.

### **Selected Publications:**

- S. Jalalvand, M. Negri, D. Falavigna, M. Matassoni, and M. Turchi "Automatic quality estimation for ASR system combination", *Computer Speech & Language*, 47, pp.214-239. 2018
- M. Arcan, M. Turchi, S. Tonelli, P. Buitelaar "Leveraging bilingual terminology to improve machine translation in a CAT environment", *Natural Language Engineering*, 1-26. 2017
- R. Chatterjee, G. Gebremelak, M. Negri, M. Turchi "Online Automatic Post-editing for MT in a Multi-Domain Translation Environment", In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers (Vol. 1, pp. 525-535)*, 2017
- D. Falavigna, M. Matassoni, S. Jalalvand, M. Negri, M. Turchi "DNN adaptation by automatic quality estimation of ASR hypotheses", *Computer Speech & Language*, Elsevier, 2016
- M. Sabet Jalili, M. Negri, Matteo, M. Turchi, and E. Barbu (2016) "An Unsupervised Method for Automatic Translation Memory Cleaning" *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, Germany
- L. Bentivogli, N. Bertoldi, M. Cettolo, M. Federico, M. Negri and M. Turchi (2016) "On the Evaluation of Adaptive Machine Translation for Human Post-Editing", *IEEE Transaction Audio Speech and Language Processing*.
- L. Gatti, M. Guerini and M. Turchi (2015) "SentiWords: Deriving a High Precision and High Coverage Lexicon

for Sentiment Analysis.", IEEE Transactions on Affective Computing Journal.

- S. Jalalvand, M. Negri, D. Falavigna and M. Turchi (2015) "Driving Rover with segment-based ASR quality estimation.", In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL), Beijing, China.
- J. G. C. de Souza, M. Negri, E. Ricci and M. Turchi (2015) "Online multitask learning for machine translation quality estimation.", In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL), Beijing, China.
- M. Arcan, M. Turchi and P. Buitelaar (2015) "Knowledge portability with semantic expansion of ontology labels.", In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL), Beijing, China.
- J. G. C. de Souza, H. Zamani, M. Negri, M. Turchi and D. Falavigna (2015) "Multitask Learning for Adaptive Quality Estimation of Automatically Transcribed Utterances.", In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015), Denver, Colorado, USA.
- M. Turchi, A. Anastasopoulos, J.G. Camargo de Souza and M. Negri (2014) "Adaptive Quality Estimation for Machine Translation.", In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), Baltimore, USA.
- M. Guerini, L. Gatti and M. Turchi (2013) "Sentiment Analysis: How to Derive Prior Polarities from SentiWordNet.", In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Seattle, USA.
- A. Balahur and M. Turchi (2013) "Comparative Experiments Using Supervised Learning and Machine Translation for Multilingual Sentiment Analysis.", Computer Speech and Language.
- R. Steinberger, M. Ebrahim, and M. Turchi (2012) "JRC Eurovoc Indexer JEX - A freely available multi-label categorisation tool.", In Proceedings of the Language Resources and Evaluation Conference (LREC), Istanbul, Turkey.
- M. Turchi, T. DeBie, N. Cristianini (2012) "An Intelligent Web Agent that Autonomously Learns how to Translate.", Web Intelligence and Agent Systems: An International Journal (WIAS).
- I. Flaounas, M. Turchi, O. Ali, N. Fyson, T. De Bie, N. Mosdell, J. Lewis, N. Cristianini (2011) "The Structure of the EU Mediasphere.", PLoS ONE Journal.
- L. Specia, D. Raj, M. Turchi (2010) "Machine Translation Evaluation versus Quality Estimation.", Machine Translation Journal.
- L. Specia, M. Turchi, N. Cancedda, M. Dymetman, N. Cristianini (2009) "Estimating the Sentence-Level Quality of Machine Translation Systems.", In Proceedings of EAMT 2009, Barcelona, Spain.
- M. Turchi, T. De Bie, N. Cristianini (2008) "Learning Performance of a Machine Translation System: a Statistical and Computational Analysis.", In Proceedings of the ACL 2008 Third Workshop on Statistical Machine Translation, Columbus, USA.