Mauro Cettolo: Curriculum Vitae et Studiorum (updated at Apr 2016)

Full Name:	Mauro Cettolo
Date of Birth:	April, 22 1965
Place of Birth:	Udine
Nationality:	Italian
Marital Status:	married, two children
Legal residence:	via Bolghera, 29
	I-38122 Trento (TN), Italy
	tel. +39 0461 935694, cell. 346 4950466
Present employment:	permanent position as a senior researcher (PAT contract) at FBK
Work address:	Fondazione Bruno Kessler, FBK
	via Sommarive, 18
	I-38123 Povo (TN), Italy
	tel +39 0461 314551
E-mail addresses:	cettolo@fbk.eu, maurocettolo65@gmail.com
Web:	https://hlt-mt.fbk.eu/people/profile/cettolo

EDUCATION

Secondary Education:

When, Where:	1979-1984, Technical Institute "A. Malignani" of Udine, Italy
Education and Degrees Awarded:	technical school-leaving certificate, specialization in electronics
Final mark:	52/60

University:

When, Where:	1987-1991, University of Udine, Italy
Education and Degrees Awarded:	degree in computer science on June, 26 1991
Final mark:	110/110 with honours
Thesis title:	"Algoritmi per il controllo di sistemi periodici di distribuzione in presenza di incertezza sulla domanda" ("Algorithms for the control of periodic distribution systems in presence of uncertainty on demand")
Thesis Advisor:	prof. Franco Blanchini

SPECIALIZATION SCHOOLS

Jul 7 - Jul 18 1997 : NATO Advanced Study Institute school on computational models of speech pattern processing, in Jersey (UK)

Sep 5 - Sep 15 1995 : GII School for PhD, in Rovereto (TN), Italy, attending the following courses:

- Software Engineering I: Software Development (prof. C. Ghezzi)
- Software Engineering II: Software Testing (prof. A. Cimitile)
- Hardware and Software Supports for Distributed Multimedia Systems (prof. M. Maresca)
- Stochastic Models for Reliability and Performance Analysis (prof. V. Grassi)
- Advanced Methodologies for Digital Circuit Design (prof. P. Prinetto)

- Metropolitan Networks (prof. L. Lenzini)

Jun 27 - Jul 10 1993 : NATO Advanced Study Institute school on speech recognition and coding, in Granada (SPA)

TEACHING

academic years 2008/09, 09/10, 10/11, 11/12, 12/13 : some frontal lessons for the module *Multilingual Technology*, Master in Human Language Technology and Interfaces, University of Trento, Italy

academic year 2007/2008 : module *Multilingual Technology*, Master in Human Language Technology and Interfaces, University of Trento, Italy. Laboratory lessons only

academic year 2005/2006 : module *Cross Language Information Technologies* at Free University of Bozen-Bolzano, Italy. Laboratory lessons only

academic years 2002/2003, 2003/2004 : module *Laboratorio di Programmazione* (Programming Lab) at the Science Faculty of Verona University, Italy

academic year 2001/2002 : module *Linguaggi di Programmazione: Implementazione* (Programming Languages: Implementation) at the Science Faculty of Trento University, Italy

academic year 2000/2001 : module *Linguaggi di Programmazione: Implementazione* (Programming Languages: Implementation) at the Science Faculty of Trento University, Italy. Laboratory lessons only

academic years 1997/1998, 1998/1999 and 1999/2000 : module *Informatica Generale* (Fundamentals of Computer Science) at the Science Faculty of Verona University, Italy

RELEVANT PROFESSIONAL EXPERIENCES

Jan 2004 - today : FBK (formerly ITC-irst) as a senior researcher (level R2)

Dec 1996 - Dec 2003 : ITC-irst as a staff researcher (level R1)

Dec 1993 - Nov 1996 : ITC-irst as a junior researcher (level E)

Jan 1992 - Nov 1993 : ITC-irst, Centro per la Ricerca Scientifica e Tecnologica, with a grant at the Artificial Intelligence Division - Speech Recognition Group

OTHER WORKING EXPERIENCES

Nov 1988 - Dec 1989 : Wiener Hotel in S. Giovanni al Natisone (UD) as night porter

May 1986 - Dec 1986 : SIP-Italian Telecom as office worker in Udine

Jan 1985 - Apr 1986 : military service, with the rank of second lieutenant

JOURNAL, CONFERENCE, WORKSHOP COMMITTEES

Transactions of the ACL 2015- : currently, member of the standing review committee

IWSLT 2011,2012,2013,2014,2015,2016 : evaluation committee

EAMT 2012 : local organization chair

LaDaKT 2012 : program committee and local organization (www.trentorise.eu/LaDaKT12)

IWSLT 2007 : chair of local organization

AISV 2006 : program co-chair

OPENLAB 2006 : chair of shared tasks (tcstar.org/openlab2006)

In addition, I served as a reviewer for major international journals, conferences and workshops in the field of human language technology (in the last year: Transactions of the ACL, ACL, EMNLP, NAACL, WMT...).

PROJECTS LED

Sep 2009 - Feb 2012 : EuroMatrixPlus (Bringing Machine Translation for European Languages to the User), funded by the European Union within the 7th Framework Programme

Oct 2009 - 2012 : MeteoTrentino, industrial contract for the development of a system for the automatic translation of weather forecast bulletins written in Italian into English and German

Dec 2009 - Jul 2011 : Aineva, industrial contract for the development of a system for the automatic translation of bulletins of conditions and avalanche forecasts written in Italian into English, French and German

Mar 2002 - Nov 2002 : ProtoRai, funded by the Italian broadcaster RAI for enhancing/updating the ITC-irst broadcast news transcription system system

Sep 2000 - May 2002 : Presto (Preservation Technology for European Broadcast Archives), funded by the European Union

SUPERVISION OF STUDENTS

2015 Summer internship on Machine Translation (MT) of Lorenzo De Mattei (MMT EU funded project)

2012 Summer internship on MT of Ondrei Platek (MateCat EU funded project)

2011 pre-PhD internship on MT of Prashant Mathur

2009 Summer internships on MT of German Sanchis-Trilles and Fahim Ahmed Salim

2008 Summer internships of Muhammad Arifur Rahman and Christian Hardmeier

2002-2003 Michele Vescovi (Trento University), thesis: "Algoritmi per la segmentazione audio basati sul criterio di informazione bayesiano: analisi, implementazione e sperimentazione" (BIC-based algorithms for audio segmentation: analysis, implementation and experimentation)

2002 Michele Vescovi (Trento University), stage: "Un algoritmo efficiente per la segmentazione audio basato sul criterio di informazione bayesiano: analisi, implementazione e sperimentazione" (An efficient algorithm for audio segmentation based on BIC: analysis, implementation and experimentation)

2001 Aditya (IIT - Kanpur, India), summer camp: "Click detection in Large Audio Files"

1998 Riccardo Delpari (Verona University), thesis: "Interfaccia Grafica e Monitoraggio di un Sistema per la Comunicazione Vocale Multilingua: Analisi, Progetto e Realizzazione" (Graphical Interface and Monitoring of a Multilingual Speech Communication System: Analysis, Design and Implementation)

1997 Christian Girardi (Trento University), thesis: "Sviluppo di un'Interfaccia Grafica per un Sistema di Comunicazione Vocale Multilingua" (Development of a Graphical Interface for a Multilingual Speech Communication System)

TECHNOLOGICAL TRANSFER

2015-2017 co-developer of MMT (www.modernmt.eu), an MT infrastructure that allows the automatic translation in a domain-sensitive way, without any need for prior adaptation nor definition of the domain

2012-2014 co-developer of MATECAT (www.matecat.com, en.wikipedia.org/wiki/MateCat), a web-based computer-assisted translation (CAT) tool integrating MT

2007-2008 co-developer of IRSTLM, an open source toolkit for handling large scale n-gram language models widely used by the research community

2002 : developer of the Audio Quality Monitor, released to both ACS (system integration company in Rome) and RAI Research Centre (Turin), within the PRESTO project. Moreover, ACS and ITC-irst signed a contract for the updates and commercial exploitation of the module

2002 : developer of the recognizer of initial sounds of TV/radio programs released to RAI Research Centre under the PROTORAI collaboration

1999, 2000, 2002 : developer of (three releases of) the module for audio preprocessing (segmentation, classification and clustering), provided to RAI Research Centre under the PROTORAI collaboration

1999 : responsible of technical aspects of the ITC-irst translation system developed for the `C-STAR` consortium; a worldwide demonstration of the prototype was showed to the media on July, 22 1999

1994 : developer of the compiler and optimizer of finite state networks for the representation of n-gram language models, released to GST company (Trento) within the technological transfer of A.Re.S., a dictation system for radiological reporting

1992 : developer of the compiler of regular grammars employed by SGS Informatica of Cavalese (TN) for implementing a speech based interface to Windows, showed at SMAU (Milan) in 1992

INVITED TALKS

- PER UNA STORIA DELLA TRADUZIONE AUTOMATICA, convegno "Tradotto/tradito", Liceo Classico "Beccaria", Milan, October, 23 2008

- AN INTRODUCTION TO HMMS, University of Siena, October, 29 2004

RESEARCH INTERESTS

Since I have joined FBK/ITC-irst in 1992, my scientific interests fall into the generic area of statistical pattern recognition, more precisely automatic speech recognition first and statistical machine translation now. Broadly, my activities can be divided in the following areas:

Search and Language Modeling

I considered the problem of n-gram language models estimation and their integration, together with the lexical knowledge, into the search space explored by decoding algorithms. I implemented a compiler for a grammar specification language, including a finite state network optimizer, that was used for implementing many speech recognition systems. Moreover, I contributed to the development of IRSTLM, an open source toolkit for handling large scale language models widely used by the research community.

Large Vocabulary Speech Recognition

For several years I worked within the ITC-irst large vocabulary speech recognition team. Significant results have been achieved in the efficient tree-based organization of the search space for large vocabulary dictation tasks. The originality of the developed techniques has been certified by a US and European patent. In 1998, I have been involved in writing two chapters of the book on spoken language dialogues edited by prof. R. De Mori, published by Academic Press.

From an applicative viewpoint, I was involved in the implementation of a system for dictating radiological reports, that is a speaker-independent continuous-speech system with a 10K-word dictionary. That was done in a collaboration with the radiological department of Santa Chiara Hospital (Trento). The system was put on the market by the company GST of Trento.

Speech Understanding and Translation

During 1996-99, I dealt with the problems of speech understanding and translation, in particular of the transcription of sentences belonging to limited domains and of the detection of semantic boundaries within them. This activity was mainly accomplished within the international `C-STAR` consortium, whose applicative goal was to develop a distributed system to show the effectiveness of the automatic speech translation between a number of languages. During 1999, a number of demos took place involving laboratories from all around the world (in addition to us: USA, Germany, France, Korea and Japan), and respective languages. The final demo, kept on 22 July 1999, received relevant echo on national and international media.

Broadcast News Transcription

Until the end of 2002, I worked on the automatic transcription of audio from broadcast news programs, activity accomplished in collaboration with RAI and funded by the EC, through European projects. In particular, I developed techniques for the segmentation and classification of audio signals in terms of acoustic source (music, speech, gender of speaker) and of channel bandwidth (narrow/wideband), and on the speaker identification.

In the ambit of the project PRESTO (2000-2002), funded by the EU, I collaborated with RAI Research Centre and ACS (Rome) in the implementation of a system for the automatic audio quality control, to be employed during the digitization process of RAI archives.

Automatic Transcription of Lectures

Within the project FAME (2002-2004), funded by the EU, I did research on the transcription of scientific lectures given by, possibly non-native, English speakers.

Information Extraction

In the FAME project as well, I worked on information extraction problems, mainly text categorization and named entity recognition. In particular, I applied models estimated through maximum entropy techniques.

Statistical Machine Translation

Since 2004 I have focused my efforts in statistical machine translation. I actively contributed to the development of systems employed by our research team in many international evaluation campaigns (NIST 2004-2007; TC-STAR 2005-2007; IWSLT 2005-2015), covering various research aspects, like the design of new reordering models, the introduction of shallow syntax in the translation process, the system tuning, the exploitation of comparable corpora crawled from the Web, model adaptation, etc. My activity is funded by European projects, in the recent past by EUROMATRIXPLUS (2009-2011) and MATECAT (2012-2014), currently by MMT (2015-2017).

OTHER PROFESSIONAL SKILLS

For implementing algorithms and developing systems, I usually work on PCs equipped with Mac or Linux OS, writing code in *C/C++* and *Perl* languages, and using various Unix tools. I also know *Java*, having been the reference language during my teaching of *Programming Lab* module, and *Pascal*. I had short programming experiences with *ML*, *Lisp*, *Fortran* and *Clips*.

I run my experiments on Kore, a large high-performance computing system in FBK, consisting of more than one hundred compute nodes and two file servers, equipped with Sun Grid Engine for queueing and batch processing.

CITATION INDICES from GOOGLE SCHOLAR PROFILE (Apr 2016)

	all	since 2011
citations	1612	844
h-index	22	13
i10-index	41	18

PUBLICATION LIST

Journals

[1] M. Federico, L. Bentivogli, N. Bertoldi, M. Cettolo, M. Negri, and M. Turchi. On the evaluation of adaptive machine translation for human post-editing. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(2):388–399, 2016.

[2] M. Cettolo, N. Bertoldi, M. Federico, H. Schwenk, L. Barrault, and C. Servan. Translation project adaptation for mt-enhanced computer assisted translation. *Machine Translation*, 28(2):127–150, 2014.

[3] N. Bertoldi, P. Simianer, M. Cettolo, K. Wäschle, M. Federico, and S. Riezler. Online adaptation to post-edits for phrase-based statistical machine translation. *Machine Translation*, 28:309–339, 2014.

[4] M. Cettolo, M. Vescovi, and R. Rizzi. Evaluation of BIC-based Algorithms for Audio Segmentation. *Computer Speech and Language*, 19:147–170, 2005.

[5] N. Bertoldi, F. Brugnara, M. Cettolo, M. Federico, and D. Giuliani. Cross-Task Portability of a Broadcast News Speech Recognition System. *Speech Communication*, 38(3-4):335–347, 2002.

[6] M. Cettolo, A. Corazza, and R. De Mori. Language Portability of a Speech Understanding System. *Computer Speech and Language*, 12:1–21, 1998.

[7] M. Federico, M. Cettolo, F. Brugnara, and G. Antoniol. Language Modeling for Efficient Beam-Search. *Computer Speech and Language*, 9(4):353–379, 1995.

Patents

[8] G. Antoniol, F. Brugnara, M. Cettolo, and M. Federico. A System for Recognizing Continuous Speech. US Patent n. US05765133, August 1998.

[9] G. Antoniol, F. Brugnara, M. Cettolo, and M. Federico. A System for Recognizing Continuous Speech. European Patent n. 96104031.8-2215, May 1995.

In Books

[10] M. Cettolo, R. Gretter, and R. De Mori. *Knowledge Integration*, pages 231–256. Academic Press, San Diego, US-CA, 1998. Chapter 8.

[11] M. Cettolo, R. Gretter, and R. De Mori. *Search and Generation of Word Hypotheses*, pages 257–309. Academic Press, San Diego, US-CA, 1998. Chapter 9.

[12] M. Cettolo and A. Corazza. *History Integration into Semantic Classification*, volume 169 of *NATO ASI Series F*, pages 356–361. Springer Verlag, 1998.

[13] G. Antoniol, F. Brugnara, M. Cettolo, and M. Federico. *Language Models Comparison in a Robot Telecontrol Application*, pages 244–247. NATO ASI Series F. Springer Verlag, Berlin, Germany, 1995.

Supervised Thesis

[14] M. Vescovi. Algoritmi per la segmentazione audio basati sul criterio di informazione bayesiano: analisi, implementazione e sperimentazione (BIC-based algorithms for audio segmentation: analysis, implementation and experimentation). Thesis, March 2003. Supervisors: R. Rizzi and M. Cettolo. Trento University. In Italian.

[15] R. Delpari. Interfaccia Grafica e Monitoraggio di un Sistema per la Comunicazione Vocale Multilingua: Analisi, Progetto e Realizzazione (Graphical Interface and Monitoring of a Multilingual Speech Communication System: Analysis, Design and Implementation). Thesis, March 1999. Supervisor: M. Cettolo. Verona University. In Italian.

International Conferences and Workshops

[16] L. Bentivogli., M. Cettolo, M. Amin Farajian, and M. Federico. WAGS: A Beautiful English-Italian Benchmark Supporting Word Alignment Evaluation on Rare Words. In *Proc. of LREC*, Portorož, Slovenia, 2016.

[17] C. Hardmeier, P. Nakov, S. Stymne, J. Tiedemann, Y. Versley, and M. Cettolo. Pronoun-Focused MT and Cross-Lingual Pronoun Prediction: Findings of the 2015 DiscoMT Shared Task on Pronoun Translation. In *Proc. of DiscoMT*, Lisbon, Portugal, 2015.

[18] M. Cettolo, L. Bentivogli, R. Cattoni, J. Niehues, S. Stüker, and M. Federico. Report on the 12th IWSLT Evaluation Campaign, IWSLT 2015. In *Proc. of IWSLT*, 2015.

[19] M. Cettolo, N. Bertoldi, and M. Federico. The Repetition Rate of Text as a Predictor of the Effectiveness of Machine Translation Adaptation. In *Proc. of AMTA*, pages 166–179, Vancouver, Canada, 2014.

[20] M. Federico, N. Bertoldi, M. Cettolo, M. Negri, M. Turchi, M. Trombetti, A. Cattelan, A. Farina, D. Lupinetti, A. Martines, A. Massidda, H. Schwenk, L. Barrault, F. Blain, P. Koehn, C. Buck, and U. Germann. The MateCat Tool. In *Proc. COLING: System Demonstrations*, pages 129–132, Dublin, Ireland, 2014.

[21] P. Mathur, M. Cettolo, M. Federico, and J.G.C. de Souza. Online Multi User Adaptive SMT. In *Proc. of AMTA*, Vancouver, Canada, 2014.

[22] P. Mathur and M. Cettolo. Optimized MT Online Learning in Computer Assisted Translation. In *Proc. of the AMTA workshop on Interactive and Adaptive Machine Translation*, Vancouver, Canada, 2014.

[23] M. Freitag, J. Wuebker, S. Peitz, M. Huck, A. Birch, N. Durrani, P. Koehn, M. Mediani, I. Slawik, J. Niehues, E. Cho, A. Waibel, N. Bertoldi, M. Cettolo, and M. Federico. Combined spoken language translation. In *Proc. of IWSLT*, Lake Tahoe, US-CA, 2014.

[24] M. Cettolo, J. Niehues, S. Stüker, L. Bentivogli, and M. Federico. Report on the 11th IWSLT Evaluation Campaign, IWSLT 2014. In *Proc. of IWSLT*, 2014.

[25] M. Cettolo, C. Girardi, and M. Federico. WIT3: il Corpus dei Sottotitoli Multilingue degli Interventi alle Conferenze TED. In *Proceedings of AISV*, Venice, Italy, 2013.

[26] N. Bertoldi, M. Cettolo, and M. Federico. Cache-based Online Adaptation for Machine Translation Enhanced Computer Assisted Translation. In *Proc. of MT Summit*, pages 35–42, Nice, France, 2013.

[27] M. Cettolo, N. Bertoldi, and M. Federico. Project adaptation for mt-enhanced computer assisted translation. In *Proc. of MT Summit*, Nice, France, 2013.

[28] P. Mathur, M. Cettolo, and M. Federico. Online Learning Approaches in Computer Assisted Translation. In *Proc. of WMT*, pages 301–308, Sofia, Bulgaria, 2013.

[29] M. Cettolo, C. Servan, N. Bertoldi, M. Federico, L. Barrault, and H. Schwenk. Issues in incremental adaptation of statistical mt from human post-edits. In *Proc. of the MT Summit XIV Workshop on Post-editing Technology and Practice (WPTP-2)*, Nice, France, 2013.

[30] M. Freitag, S. Peitz, J. Wuebker, H. Ney, N. Durrani, M. Huck, P. Koehn, T.-L. Ha, J. Niehues, M. Mediani, T. Herrmann, A. Waibel, N. Bertoldi, M. Cettolo, and M. Federico. EU-BRIDGE MT: Text translation of talks in the EU-BRIDGE project. In *Proc. of IWSLT*, pages 128–135, Heidelberg, Germany, 2013.

[31] M. Cettolo, J. Niehues, S. Stüker, L. Bentivogli, and M. Federico. Report on the 10th IWSLT Evaluation Campaign. In *Proc. of IWSLT*, 2013.

[32] M. Federico, M. Cettolo, L. Bentivogli, M. Paul, and S. Stüker. Overview of the IWSLT 2012 evaluation campaign. In *Proc. of IWSLT*, 2012.

[33] M. Cettolo, C. Girardi, and M. Federico. WIT³: Web Inventory of Transcribed and Translated Talks. In *Proc. of EAMT*, pages 261–268, Trento, Italy, 2012.

[34] M. Federico, S. Stüker, L. Bentivogli, M. Paul, M. Cettolo, T. Herrmann, J. Niehues, and G. Moretti. The IWSLT 2011 Evaluation Campaign on Automatic Talk Translation. In *Proc. of LREC*, Istanbul, Turkey, 2012.

[35] N. Bertoldi, M. Cettolo, M. Federico, and C. Buck. Evaluating the Learning Curve of Domain Adaptive Statistical Machine Translation Systems. In *Proc. of WMT*, pages 433–441, Montréal, Canada, 2012.

[36] M. Cettolo, N. Bertoldi, and M. Federico. Bootstrapping Arabic-Italian SMT through Comparable Texts and Pivot Translation. In *Proc. of EAMT*, Leuven, Belgio, 2011.

[37] M. Cettolo, N. Bertoldi, and M. Federico. Methods for Smoothing the Optimizer Instability in SMT. In *Proc. of MT Summit*, Xiamen, China, 2011.

[38] N. Bertoldi, M. Cettolo, and M. Federico. Statistical machine translation of texts with misspelled words. In *Proc. of NAACL*, pages 412–419, Los Angeles, US-CA, 2010.

[39] A. Bisazza, I. Klasinas, M. Cettolo, and M. Federico. FBK @ IWSLT 2010. In *Proc. of IWSLT*, Paris, France, 2010.

[40] M. Cettolo, M. Federico, and N. Bertoldi. Mining Parallel Fragments from Comparable Texts. In *Proc. of IWSLT*, Paris, France, 2010.

[41] G. Sanchis-Trilles and M. Cettolo. Online Language Model adaptation via N-gram Mixtures for Statistical Machine Translation. In *Proc. of EAMT*, Saint-Raphaël, France, 2010.

[42] N. Bertoldi, A. Bisazza, M. Cettolo, M. Federico, and G. Sanchis-Trilles. FBK @ IWSLT 2009. In *Proc. of IWSLT*, Tokyo, Japan, 2009.

[43] G. Sanchis-Trilles, M. Cettolo, and N. Bertoldi. Online Language Model Adaptation for Spoken Dialog Translation. In *Proc. of IWSLT*, Tokyo, Japan, 2009.

[44] M. Federico, N. Bertoldi, and M. Cettolo. IRSTLM: an Open Source Toolkit for Handling Large Scale Language Models. In *Proc. of Interspeech*, pages 1618–1621, Melbourne, Australia, 2008.

[45] M. Cettolo, M. Federico, D. Pighin, and N. Bertoldi. Shallow-Syntax Phrase-Based Translation: Joint versus Factored String-to-Chunk Models. In *Proc. of AMTA*, Waikiki, US-HI, 2008.

[46] N. Bertoldi, M. Cettolo, R. Cattoni, and M. Federico. FBK @ IWSLT 2007. In *Proc. of IWSLT*, pages 76–83, Trento, Italy, 2007.

[47] M. Cettolo and M. Federico, editors. *Proc. of IWSLT 2007*. Fondazione B. Kessler, 2007.

[48] M. Federico and M. Cettolo. Efficient Handling of N-gram Language Models for Statistical Machine Translation. In *Proc. of WMT*, pages 88–95, Prague, Czech Republic, 2007.

[49] D. Gupta, M. Cettolo, and M. Federico. POS-based Reordering Models for Statistical Machine Translation. In *Proc. of MT Summit*, pages 207–213, Copenhagen, Denmark, 2007.

[50] B. Chen, M. Federico, and M. Cettolo. Better N-best Translations through Generative n-gram Language Models. In *Proc. of MT Summit*, pages 105–110, Copenhagen, Denmark, 2007.

[51] G. Falavigna, N. Bertoldi, F. Brugnara, R. Cattoni, M. Cettolo, B. Chen, M. Federico, D. Giuliani, R. Gretter, D. Gupta, and D. Seppi. The IRST English-Spanish Translation System for European Parliament Speeches. In *Proc. of Interspeech*, pages 2833–2836, Antwerp, Belgium, 2007.

[52] N. Bertoldi, R. Cattoni, M. Cettolo, B. Chen, and M. Federico. ITC-irst at the 2006 TC-STAR SLT Evaluation Campaign. In *Proc. of the TC-STAR Workshop*, pages 18–24, Barcelona, Spain, 2006.

[53] R. Cattoni, N. Bertoldi, M. Cettolo, B. Chen, and M. Federico. A Web-based Demonstrator of a Multi-lingual Phrase-based Translation System. In *Proc. of EACL (vol. Poster and Demontrations)*, pages 91–94, Trento, Italy, 2006.

[54] R. Cattoni, N. Bertoldi, M. Cettolo, B. Chen, and M. Federico. A Web-based Interface to a Multi-lingual Phrase-based Translation System. In *Proc. of ECAI*, pages 91–94, Riva del Garda, Italy, 2006.

[55] M. Cettolo and M. Federico. Text Segmentation Criteria for Statistical Machine Translation. In *Proc. of FinTAL*, pages 664–673, Turku, Finland, 2006.

[56] B. Chen, R. Cattoni, N. Bertoldi, M. Cettolo, and M. Federico. The ITC-irst SMT system for IWSLT 2006. In *Proc. of IWSLT*, pages 53–58, Kyoto, Japan, 2006.

[57] B. Chen, M. Cettolo, and M. Federico. Reordering rules for phrase-based statistical machine translation. In *Proc. of IWSLT*, pages 182–189, Kyoto, Japan, 2006.

[58] V. Sandrini, M. Federico, and M. Cettolo. Maximun Entropy Tagging with Binary and Real-Valued Features. In *Proc. of EACL - Workshop on Learning Structured Information in Natural Language Applications*, Trento, Italy, 2006.

[59] M. Cettolo, M. Federico, N. Bertoldi, R. Cattoni, and B. Chen. A Look Inside the ITC-irst SMT System. In *Proc. of MT Summit*, pages 451–457, Phuket, Thailand, 2005.

[60] B. Chen, R. Cattoni, N. Bertoldi, M. Cettolo, and M. Federico. The ITC-irst SMT system for IWSLT-2005. In *Proc. of IWSLT*, pages 98–104, Pittsburgh, US-PA, 2005.

[61] V. H. Quan, M. Cettolo, and M. Federico. Integrated n-best re-ranking for spoken language translation. In *Proc. of Interspeech*, pages 3181–3184, Lisbon, Portugal, 2005.

[62] N. Bertoldi, R. Cattoni, M. Cettolo, and M. Federico. The ITC-irst statistical machine translation system for IWSLT-2004. In *Proc. of IWSLT*, pages 51–58, Kyoto, Japan, 2004.

[63] M. Cettolo and M. Federico. Minimum Error Training of Log-Linear Translation Models. In *Proc. of IWSLT*, pages 103–106, Kyoto, Japan, 2004.

[64] F. Brugnara M. Cettolo and M. Federico. Advances in the Automatic Transcription of Lectures. In *Proc. of ICASSP*, pages 769–772, Montreal, Canada, 2004.

[65] M. Federico E. Leeuwis and M. Cettolo. Language Modeling and Transcription of the TED Corpus Lectures. In *Proc. of ICASSP*, pages 232–235, Hong Kong, 2003.

[66] N. Bertoldi, F. Brugnara, M. Cettolo, M. Federico, D. Giuliani, E. Leeuwis, and V. Sandrini. The ITC-irst News on Demand Platform. In *Proc. of ECIR*, pages 520–527, Pisa, Italy, 2003.

[67] M. Cettolo and M. Vescovi. Efficient Audio Segmentation Algorithms based on the BIC. In *Proc. of ICASSP*, pages 537–540, Hong Kong, 2003.

[68] M. Vescovi, M. Cettolo, and R. Rizzi. A DP Algorithm for Speaker Change Detection. In *Proc. of Eurospeech*, pages 2997–3000, Geneva, Switzerland, 2003.

[69] F. Brugnara, M. Cettolo, M. Federico, and D. Giuliani. Issues in automatic transcription of historical audio data. In *Proc. of ICSLP*, Denver, US-CO, 2002.

[70] M. Cettolo. Porting an audio partitioner across domains. In *Proc. of ICASSP*, pages I:301–304, Orlando, US-FL, 2002.

[71] N. Bertoldi, F. Brugnara, M. Cettolo, M. Federico, and D. Giuliani. From broadcast news to spontaneous dialogue transcription: Portability issues. In *Proc. of ICASSP*, Salt Lake City, US-UT, 2001.

[72] M. Cettolo. Speaker tracking in a broadcast news corpus. In *Proc. of the 2001: A Speaker Odyssey - The Speaker Recognition Workshop*, Chania, Greece, 2001.

[73] F. Brugnara, M. Cettolo, M. Federico, and D. Giuliani. Advances in automatic transcription of broadcast news. In *Proc. of ICSLP*, pages II:660–663, Beijing, China, 2000.

[74] F. Brugnara, M. Cettolo, M. Federico, and D. Giuliani. A system for the segmentation and transcription of Italian radio news. In *Proc. of RIAO Content-Based Multimedia Information Access*, Paris, France, 2000.

[75] F. Brugnara, M. Cettolo, M. Federico, and D. Giuliani. A Baseline for the Transcription of Italian Broadcast News. In *Proc. of ICASSP*, Istanbul, Turkey, 2000.

[76] M. Cettolo. Segmentation and Classification of an Italian Broadcast News Corpus. In *Proc. of RIAO Content-Based Multimedia Information Access*, Paris, France, 2000.

[77] M. Cettolo and M. Federico. Model Selection Criteria for Acoustic Segmentation. In *Proc. of the ISCA ASR2000 Workshop*, Paris, France, 2000.

[78] M. Cettolo, A. Corazza, G. Lazzari, F. Pianesi, E. Pianta, and L.M. Tovena. A Speech-to-Speech Translation based Interface for Tourism. In *Proc. of ENTER*, Innsbruck, Austria, 1999.

[79] J. Haas, V. Warnke, H. Niemann, M. Cettolo, A. Corazza, D. Falavigna, and G. Lazzari. Semantic Boundaries in Multiple Languages. In *Proc. of Eurospeech*, pages 535–538, Budapest, Hungary, 1999.

[80] A. Corazza, M. Cettolo, G. Lazzari, E. Pianta, F. Pianesi, and L.M. Tovena. The ITC-irst Speech Translation System. In *Workshop AI*IA - Elaborazione del Linguaggio e Riconoscimento del Parlato*, pages 30–32, Trento, Italy, 1999.

[81] M. Cettolo and D. Falavigna. Automatic Detection of Semantic Boundaries based on Acoustic and Lexical Knowledge. In *Proc. of ICSLP*, pages 261–264, Sidney, Australia, 1998.

[82] M. Cettolo and D. Falavigna. Automatic Recognition of Spontaneous Speech Dialogues. In *Proc. of ICSLP*, pages 1551–1554, Sidney, Australia, 1998.

[83] B. Angelini, M. Cettolo, A. Corazza, D. Falavigna, and G. Lazzari. Multilingual Person to Person Communication at IRST. In *Proc. of ICASSP*, Munich, Germany, 1997.

[84] M. Cettolo and A. Corazza. Automatic Detection of Semantic Boundaries. In *Proc of Eurospeech*, pages 919–922, Rhodes, Greece, 1997.

[85] B. Angelini, M. Cettolo, A. Corazza, D. Falavigna, G. Lazzari, A. Lavelli, F. Pianesi, and O. Stock. Preliminary Results of the C-STAR Project at IRST. In *Proc. of the C-STAR Workshop*, Kyoto, Japan, 1996.

[86] M. Cettolo, A. Corazza, and R. De Mori. A Mixed Approach to Speech Understanding. In *Proc. of ICSLP*, pages 841–844, Philadelphia, US-PA, 1996.

[87] G. Antoniol, F. Brugnara, M. Cettolo, and M. Federico. Language Model Representations for Beam-Search Decoding. In *Proc of ICASSP*, pages 588–591, Detroit, US-MI, 1995.

[88] F. Brugnara and M. Cettolo. Improvements in Tree-Based Language Model Representation. In *Proc of Eurospeech*, pages 1797–1800, Madrid, Spain, 1995.

[89] M. Cettolo, A. Corazza, and R. De Mori. Automatic Learning of Sentence Dependencies in Spoken Dialogues. In *Proc. of the ESCA Workshop on Spoken Dialogue Systems*, pages 77–80, Vigso, Denmark, 1995.

[90] B. Angelini, G. Antoniol, F. Brugnara, M. Cettolo, M. Federico, R. Fiutem, and G. Lazzari. Radiological Reporting by Speech Recognition: the A.Re.S. System. In *Proc of ICSLP*, pages III:1267–1270, Yokohama, Japan, 1994.

[91] G. Antoniol, F. Brugnara, M. Cettolo, and M. Federico. Language Model Estimations and Representations for Real-Time Continuous Speech Recognition. In *Proc. of ICSLP*, pages II:859–862, Yokohama, Japan, 1994.

[92] G. Antoniol, M. Cettolo, and M. Federico. Techniques for Robust Recognition in Restricted Domains. In *Proc. of Eurospeech*, pages III:2219–2221, Berlin, Germany, 1993.

[93] G. Antoniol, M. Cettolo, and M. Federico. Robust and Reliable Speech Understanding in Restricted Domains. In *Proc. of the IEEE ASR Workshop*, pages 103–104, Snowbird, US-UT, 1993.

[94] G. Antoniol, R. Cattoni, M. Cettolo, and M. Federico. Robust Speech Understanding for Robot Tele-control. In *Proc. of ICAR*, pages 205–209, Tokyo, Japan, 1993.

[95] G. Antoniol, F. Brugnara, M. Cettolo, R. Fiutem, R. Flor, and G. Lazzari. Comparison of Automatic Speech Recognition Systems based on New High Speed RISC Workstations. In *Proc. of ICSPAT*, pages 1473–1477, Santa Clara, US-CA, 1993.

[96] G. Antoniol, F. Brugnara, M. Cettolo, R. Fiutem, R. Flor, and G. Lazzari. ASR System Comparison on New High Performance Workstation. In *Proc. of AVIOS*, pages 17–24, San Jose, US-CA, 1993.

[97] G. Antoniol, G. Carli, M. Cettolo, R. Fiutem, R. Flor, and G. Lazzari. Finite State Network, Isolated Word, Real Time Automatic Speech Recognizer based on DSP32C. In *Proc. of ICSPAT*, pages II:1000–1004, Boston, US-MA, 1992.

[98] G. Antoniol, G. Carli, M. Cettolo, and R. Fiutem. Tools for Development, Test and Analysis of ASRs. In *Proc. of ICSPAT*, pages II:1005–1010, Boston, US-MA, 1992.

[99] F. Blanchini, M. Cettolo, and P. Righini. Disturbance Rejection Problem for Discrete-Time Linear Systems: a Comparison between Linear and Nonlinear Control. In *Proc. of the International Symposium on Implicit and Robust Systems*, pages 28–32, Warsaw, Poland, 1991.

Relevant Technical Reports

[100] M. Cettolo. You talk, I translate! Technical Aspects of the ITC-irst Speech Translation System. Technical Report 0003-25, ITC-irst, Centro per la Ricerca Scientifica e Tecnologia, Trento, Italy, March 2000.

[101] M. Cettolo and A. Corazza. Versione Italiana di ATIS: Apprendere dai Dati la Comprensione del Linguaggio Parlato. Technical Report 9511-10, ITC-irst, Centro per la Ricerca Scientifica e Tecnologia, Trento, Italy, November 1995.

[102] G. Antoniol and M. Cettolo. Algorithms for the FSN Compiling Problem. Technical Report 9210-01, ITC-irst, Centro per la Ricerca Scientifica e Tecnologia, Trento, Italy, October 1992.