Europass Curriculum Vitae

Personal information

Surname(s) / First name(s)	**Tonelli, Sara**
Address(es)	Fondazione Bruno Kessler, via Sommarive 18, I-38100 Trento
Telephone(s)	+39 0461 314 542 (office)
Email(s)	satonelli@fbk.eu
Nationality(-ies)	
Date of birth	
Employment status	Since August 2010, Researcher at Fondazione Bruno Kessler, Human Language Technology group, Trento, Italy.

Skills and competences

Mother tongue(s) **Italian**

Self-assessment
European level [*]

Understanding				Speaking				Writing	
Listening		Reading		Spoken interaction		Spoken production			
C2	Proficient user	C2	Proficient user	C2	Proficient user	C2	Proficient user	B2	Independent user
C1	Proficient user	C2	Proficient user	C2	Proficient user	C1	Proficient user	C2	Proficient user
B1	Independent user	B2	Independent user	A2	Basic user	A2	Basic user	A2	Basic user

German (first data row), **English** (second data row), **French** (third data row)

[*] *Common European Framework of Reference (CEF) level*

Research interests	In the area of Natural Language Processing, most notably in the fields of

- lexical and frame semantics, in particular the automatic creation of lexical resources both with rule-based and machine learning techniques

- corpus analysis with a focus on multilinguality and parallel texts

- multilingual key-concept extraction and evaluation

- discourse analysis, in particular the semantics of connectives

Programming languages	Very good knowledge of Perl; good knowledge of CGI and Prolog (SICStus prolog and SWI); basic knowledge of Python (natural language toolkit)

Education

Phd in Computational Linguistics	University of Venice, Department of Language Sciences, 22nd cycle (2006–2009) Title: *Semi-automatic Techniques for extending the FrameNet lexical database to new languages.* Advisors: Prof. Rodolfo Delmonte, Dott. Emanuele Pianta Degree obtained on 29th March 2010
Summer internship	Institute for Research in Cognitive Science, University of Pennsylvania, USA Jun - Aug 2009
Post-degree course (*Aufbaustudium*) in Computational Linguistics	University of Munich, Germany Oct 2000 - June 2002 Final evaluation: A
Master's degree (*laurea vecchio ordinamento*) in Foreign Languages	University of Bergamo, June 2000 Final evaluation: 110/110 *cum laude*
Socrates-Erasmus student exchange	Faculty of Social Sciences, University of Karlsruhe, Germany Sept 1997 - Jul 1998

Professional experiences

Research

Aug 2010 - Present
Human Language Technologies Group, Fondazione Bruno Kessler, Trento
Research Projects:

- *NewsReader* EU Project: From Jan 2013 to Dec 2015. Responsible of WP on benchmarking and evaluation of NLP tools for multilingual event extraction and processing.

- *TERENCE* EU Project (An adaptative learning system for reasoning about stories with poor comprehenders and their educators): From Oct 2010 to Jun 2013. Development of NLP tools for the automatic analysis of Italian children's stories. Development of a system for measuring text readability and classifying documents into language proficiency levels.

- *PESCaDO* EU Project (Personalized Environmental Service Configuration and Delivery Orchestration): From Jan 2010 to Dec 2012. Content extraction and processing from multilingual web-sites. Development of a multilingual concept extraction tool to support the creation of ontologies in the environmental domain.

Jun - Aug 2009
Institute for Research in Cognitive Science, University of Pennsylvania,
Philadelphia, USA
Adaptation of the Penn Discourse Treebank paradigm to the LUNA corpus of spoken Italian. Definition of the annotation manual and the annotation workflow for discourse relations in dialogs.

Jan - Mar 2008
University of Trento, Italy
http://www.ist-luna.eu/project_description.htm
Research activity for the LUNA European Project: pilot study on the annotation of Italian dialogues with frame information. Definition of the annotation manual and of the annotation workflow.

2005 - 2006
Laboratory of Computational Linguistics, University of Venice, Italy
http://project.cgm.unive.it
Automatic conversion of the phrase-based *Venice Italian Treebank* into a dependency treebank with grammatical relations, in cooperation with Toshiba Research Europe Ltd, UK

2003 - 2004
Laboratory of Computational Linguistics, University of Venice, Italy
Development of the VENEX corpus of anaphoric information in spoken and written Italian in cooperation with Prof. Massimo Poesio, University of Essex

2001 - 2002
Centrum fuer Informations-und Sprachverarbeitung, Munich University, Germany
http://www.cis.uni-muenchen.de
Development of an electronic lexicon of Italian verbs

Teaching	Mar 2012 - Oct 2012
	Faculty of Cognitive Science, University of Trento, Italy
	Adjunct Professor of *Computational Linguistics*
	Oct 2011 - Jan 2012 / Oct 2012 - Jan 2013
	Faculty of Computer Science, University of Bozen, Italy
	Teaching assistant of *Computational Linguistics*
	2003 - 2006
	Laboratory of Computational Linguistics, University of Venice, Italy
	Coordination and teaching activities for the distance learning course *Assisted and automatic translation (TATAE)* for professional translators
	Sept 2005 - Jul 2006
	Laboratory of Computational Linguistics, University of Venice, Italy
	Teaching assistant of *Introduction to Prolog*
	Sept 2001- Jul 2002
	Centrum fuer Informations-und Sprachverarbeitung, Munich University, Germany
	Teaching assistant of *Introduction to Perl Programming* (in German)
Other	2000 - 2002 (partial time)
	Sail Labs GmbH, Munich, Germany
	http://www.sail-technology.com
	Automatic translation software for *Langenscheidt*
	Linguistic Consultant

Other	Co-chair of the AI*IA workshop "Learning by Reading in the Real World" (LERREW 2011)
	Member of the steering committee of the task "Frame Labeling over Italian Texts", Third Evaluation Campaign of NLP tools for Italian (Evalita 2012)
	Reviewer for the following international conferences and workshops: ESWC 2013, ACL2013, EACL 2012, Coling 2012, NAACL 2012, EKAW2012, ebTEL 2012 (1st International Workshop on evidence-based Technology Enhanced Learning), LTEC 2012 (Workshop on Learning Technology for Education in Cloud), *SEM 2012, Coling 2010
	Reviewer for the following international journals: "Artificial Intelligence", "Discourse and Dialogue", "Constructions and Frames", "Language Resources and Evaluation"
	PC member of: AISV 2013 (9th National Conference of the Italian Association of Speech Science), Venice, Italy 1st Workshop on EVENTS, co-located with NAACL, Atlanta, USA
	Member of the local organizing committee of: ARQAS 2003 (International Symposium on Reference Resolution and its Application to Question Answering and Summarization) InSTIL/ICALL 2004 (International Symposium on NLP and Speech Technologies in Advanced Language Learning Systems) RTE-2 (Second Recognising Textual Entailment Challenge, 2006).
Invited seminars and talks	
May 2012	*Automatic mapping between FrameNet and WordNet*, Community Meeting on Aligning Lexical Resources (LexLink), Co-Located with LREC, Istanbul, Turkey
May 2012	*Wikipedia-based WSD for Multilingual Frame Annotation*, IBM Watson, Yorktown Heights, NY, USA
April 2012	*Annotation of Italian Dialogues following the Penn Discourse Treebank paradigm*, Penn Discourse Treebank Workshop, University of Pennsylvania, Philadelphia, USA
December 2010	*Semi-automatic techniques for extending the FrameNet lexical database to new languages*, Track "Distinguished Phd Theses", Workshop on Human Language Technologies for Italian, Italian Association for Artificial Intelligence, Brescia, Italy
January 2009	*A semi-automatic approach to the development of Italian FrameNet based on existing lexical resources*, Computational Linguistics Colloquium, Centrum für Informations-und Sprachverarbeitung der Ludwig-Maximilians-Universität, Munich, Germany

December 2008	*Towards Italian FrameNet: a proposal for the semi-automatic development of FrameNet-like resources*, Computational Linguistics Colloquium, University of Saarbrücken, Germany
May 2008	*Building Italian FrameNet through frame information transfer from English to Italian* Language and Communication Technologies Colloquia, Student Day, Free University of Bozen, Italy
February 2006	*Algorithms and Techniques for anaphora resolution and pronominal binding: a comparative evaluation*, Seminar series on Advanced Syntax, University of Venice, Italy

Publications

2013	Gianni Barlacchi and Sara Tonelli. *ERNESTA: A sentence simplification tool for children's stories in Italian*. To appear in Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CicLing 2013).

| 2012 | Sara Tonelli, Claudio Giuliano and Kateryna Tymoshenko. *Wikipedia-based WSD for Multilingual Frame Annotation*. Accepted for publication in Artificial Intelligence Journal. |

Marco Rospocher, Sara Tonelli, Luciano Serafini and Emanuele Pianta. *Corpus-based Terminological Evaluation of Ontologies*. Accepted for publication in Applied Ontology, special issue on Ontologies and Terminologies, IOS Press.

Sara Tonelli and Elena Cabrio. *Hunting for Entailing Pairs in the Penn Discourse Treebank*. In Proceedings of Coling, Mumbay, India.

Sara Tonelli, Ke Tran Manh and Emanuele Pianta. *Making Readability Indices Readable*. In Proceedings of the NAACL Workshop on Predicting and Improving Readability for Target Reader Population (PITR2012), Montreal, Canada.

Sara Tonelli, Elena Cabrio and Emanuele Pianta. *Key-concept extraction from French articles with KX*. In Proceedings of the French Text Mining Evaluation Workshop (DEFT 2012), Grenoble, French.

Sara Tonelli, Volha Bryl, Claudio Giuliano and Luciano Serafini. *Investigating the Semantics of Frame Elements.* In Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW-2012), Galway, Ireland

Marco Rospocher, Sara Tonelli, Luciano Serafini and Emanuele Pianta. *Key-concept extraction for ontology engineering.* (demo paper) In Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW-2012), Galway, Ireland

Volha Bryl, Sara Tonelli, Claudio Giuliano and Luciano Serafini. *A Novel FrameNet-based resource for the Semantic Web.* In Proceedings of the 27th ACM Symposium on Applied Computing (SAC-2012), Riva del Garda (TN), Italy

Anastasia Moumtzidou, Stefanos Vrochidis, Sara Tonelli, Ioannis Kompatsiaris and Emanuele Pianta. *Discovery of Environmental Nodes in the Web.* In Proceedings of the 5th Information Retrieval Facility Conference, Vienna, Austria

Sucheta Ghosh, Richard Johansson, Giuseppe Riccardi and Sara Tonelli. *Improving the Recall of a Discourse Parser by Constraint-based Postprocessing.* In Proceedings of LREC, Istanbul, Turkey

2011   Sara Tonelli, Marco Rospocher, Emanuele Pianta and Luciano Serafini. *Boosting collaborative ontology building with key-concept extraction.* In Proceedings of the IEEE Fifth International Conference on Semantic Computing (ICSC-2011), Stanford, CA

Sucheta Ghosh, Sara Tonelli, Giuseppe Riccardi and Richard Johansson. *End-to-end Discourse Parser Evaluation.* In Proceedings of the IEEE Fifth International Conference on Semantic Computing (ICSC-2011), Stanford, CA

Sara Tonelli, Rodolfo Delmonte. *Desperately Seeking Implicit Arguments in Text.* In Proceedings of the ACL Workshop on Relational Models of Semantics, Portland, Oregon

Sara Tonelli, Emanuele Pianta. *Matching documents and summaries using key-concepts.* In Proceedings of the French Text Mining Evaluation Workshop (DEFT 2011), Montpellier, France

Sucheta Ghosh, Richard Johansson, Giuseppe Riccardi and Sara Tonelli. *Shallow Discourse Parsing with Conditional Random Fields.* In Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP-2011), Chiang Mai, Thailand

Suhel Jaber, Sara Tonelli, Rodolfo Delmonte. *Venetan to English machine translation: issues and possible solutions.* In Proceedings of the 8th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2011), Copenhagen, Denmark

2010

Sara Tonelli. *Semi-automatic Techniques for Extending the FrameNet Lexical Database to New Languages*. PhD thesis, Dept. of Language Sciences, Università Ca' Foscari, Venezia.

Sara Tonelli, Giuseppe Riccardi, Rashmi Prasad, Aravind Joshi. *Annotation of Discourse Relations for Conversational Spoken Dialogs*. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta.

Emanuele Pianta, Sara Tonelli. *KX: A flexible system for Keyphrase eXtraction*. In Proceedings of SemEval 2010, Task 4: Keyword extraction from Scientific Articles, Uppsala, Sweden

Sara Tonelli, Rodolfo Delmonte. *VENSES++: Adapting a deep semantic processing system to the identification of null instantiations*. In Proceedings of SemEval 2010, Task 10: Linking events and their participants in discourse, Uppsala, Sweden

Alexei Ivanov, Giuseppe Riccardi, Sucheta Ghosh, Sara Tonelli and Evgeny Stepanov. *Acoustic Correlates of Meaning Structure in Conversational Speech*. In Proceedings of Interspeech, Makuhari, Japan.

Sara Tonelli, Emanuele Pianta, Rodolfo Delmonte, Michele Brunelli. *VenPro: a Morphological Analyzer for Venetan*. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta.

Sara Tonelli, Giuseppe Riccardi. *Guidelines for Annotating the LUNA corpus with Frame Information*. Technical Report # DISI-10-017, Department of Information Engineering and Computer Science, University of Trento

2009	Sara Tonelli, Claudio Giuliano. *Wikipedia as frame information repository*. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2009), Singapore

Sara Tonelli, Daniele Pighin. *New features for FrameNet - WordNet mapping*. In Proceedings of the Thirteenth Conference on Computational Language Learning (CoNLL), Boulder, Colorado

Sara Tonelli, Daniele Pighin, Claudio Giuliano, Emanuele Pianta. *Semi-automatic Development of FrameNet for Italian*. In Proceedings of the FrameNet Workshop and Masterclass, Co-located with the Seventh International Workshop on Treebanks and Linguistic Theories, Milano, Italy

Sara Tonelli, Emanuele Pianta. *A novel approach to mapping FrameNet lexical units and WordNet synsets*. In Proceedings of the Eighth International Conference on Computational Semantics (IWCS 2009), Tilburg, Netherlands

Sara Tonelli, Emanuele Pianta. *Three issues in cross-language frame information transfer*. In Proceedings of the Seventh International Conference on Recent Advances in Natural Language Processing (RANLP 2009), Borovets, Bulgaria

Sara Tonelli, Emanuele Pianta. *Analisi della Qualità della Traduzione basata su frame*. XLIII International Conference of the Italian Linguistic Society (SLI 09), Verona, Italy

Sara Tonelli, Rodolfo Delmonte. *Knowledge-poor and knowledge rich approach in anaphora resolution algorithms: a comparison*. In 'Linguistica e Modelli Tecnologici di Ricerca - XL Congresso Internazionale di Studi della Società di Linguistica Italiana', Ed. Bulzoni, Rome, Italy, ISBN 978-88-7870-469-5.

Rodolfo Delmonte, Antonella Bristot, Sara Tonelli, Emanuele Pianta. *English / Veneto Resource Poor Machine Translation with STILVEN*. In International Review BULAG - Special issue on Data Sense Mining, Machine Translation and Controlled Languages, Ed. Presses universitaires de Franche-Comté, Besançon, France, ISBN 978-2-84867-261-8

Rodolfo Delmonte, Sara Tonelli, Rocco Tripodi. *Semantic Processing for Text Entailment with VENSES*. In Proceedings of the Fifth PASCAL Recognizing Textual Entailment Challenge, Gaithersburg, Maryland, USA

Marco Dinarelli, Silvia Quarteroni, Sara Tonelli, Alessandro Moschitti, Giuseppe Riccardi *Annotating Spoken Dialogs: from Speech Segments to Dialog Acts and Frame Semantics* In Proceedings of the EACL Workshop on Semantic Representation of Spoken Language (SRSL), Athens, Greece

2008	Sara Tonelli, Emanuele Pianta. *Frame information transfer from English to Italian.* In Proceedings of LREC 2008, Marrakech, Morocco
	Sara Tonelli, Rodolfo Delmonte, Antonella Bristot. *Enriching the Venice Italian Treebank with Dependency and Grammatical relations.* Proceedings of LREC 2008, Marrakech, Morocco
	Bonaventura Coppola, Alessandro Moschitti, Sara Tonelli, Giuseppe Riccardi *Automatic FrameNet-based annotation of Conversational Speech.* In Proceedings of the Second IEEE Spoken Language Technology (SLT) Workshop, Goa, India
	Arianna Bisazza, Marco Dinarelli, Silvia Quarteroni, Sara Tonelli, Alessandro Moschitti, Giuseppe Riccardi *Semantic Annotations for Conversational Speech: from Speech transcriptions to Predicate Argument Structures.* In Proceedings of the Second IEEE Spoken Language Technology (SLT) Workshop, Goa, India
2007	Octavian Popescu, Sara Tonelli, Emanuele Pianta. *Preposition Disambiguation based on Chain Clarifying Relationships Contexts.* In Proceedings of SemEval-2007 Workshop, co-located with ACL 2007, Prague
	Rodolfo Delmonte, Antonella Bristot, Sara Tonelli, Marco Aldo Piccolino Boniforti. *Overlaps in AVIP/IPAR, the Italian Treebank of Spontaneous Speech.* In Proceedings of the 1st Workshop on Semantic Representation of Spoken Language (SRSL), Salamanca, Spain
	Rodolfo Delmonte, Antonella Bristot, Sara Tonelli. *VIT - Venice Italian Treebank: Syntactic and Quantitative Features.* In Proceedings of the Sixth International Workshop on Treebanks and Linguistic Theories, Norway
	Rodolfo Delmonte, Antonella Bristot, Marco Aldo Piccolino Boniforti and Sara Tonelli. *Entailment and Anaphora Resolution in RTE3.* In Proceedings of the 3nd PASCAL RTE challenge, co-located with ACL 2007, Prague
2006	Rodolfo Delmonte, Antonella Bristot, Marco Aldo Piccolino Boniforti and Sara Tonelli. *Another evaluation of Anaphora Resolution Algorithms and a Comparison with GETARUNS' knowledge rich approach.* In Proceedings of the 4th International Workshop on Robust Methods in Analysis of Natural Language Data (Romand 06), Trento
	Rodolfo Delmonte, Antonella Bristot, Marco Aldo Piccolino Boniforti and Sara Tonelli. *Coping with Semantic Uncertainty with VENSES.* In Proceedings of the 2nd PASCAL RTE challenge, Venezia
	Massimo Poesio, Rodolfo Delmonte, Antonella Bristot, Luminita Chiran and Sara Tonelli. *The VENEX corpus of anaphora and deixis in spoken and written Italian.* Essex NLE Technical Note 2003-1.

2005	Rodolfo Delmonte, Sara Tonelli, Marco Piccolino Boniforti , Antonella Bristot, Emanuele Pianta. *VENSES - A Linguistically-Based System for Semantic Evaluation*. In Proceedings of the 1st PASCAL RTE challenge, Southampton.
	Rodolfo Delmonte, Antonella Bristot, Marco Piccolino Boniforti, Sara Tonelli. *Le sovrapposizioni: sintassi, semantica e prosodia*. In Proceedings of the 1st AISV Conference - MISURA DEI PARAMETRI, Padova
	Rodolfo Delmonte, Antonella Bristot, Luminita Chiran, Ciprian Bacalu, Sara Tonelli. *Verso un'Analisi Sintattico-Semantica delle Sovrapposizioni*, in F. Albano Leoni, R. Giordano (a cura di) Italiano Parlato. Analisi di un Dialogo, Liguori Ed.
2004	Rodolfo Delmonte, Antonella Bristot, Luminita Chiran, Ciprian Bacalu, Sara Tonelli. *Parsing the Oral Corpus AVIP/API*. In Atti del Convegno Internazionale 'Il Parlato Italiano', Napoli.

Le informazioni contenute nel presente documento vengono rese ai sensi e per gli effetti degli artt. 46 e 47 del DPR 455-200 e corrispondono a verità.

18 febbraio 2013