

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 314 574 (office)

Email(s)

satonelli@fbk.eu

Nationality(-ies)

Italian

Date of birth

October 5, 1976

Employment status

Since January 2010, Junior Researcher, Human Language Technologies group, Fondazione Bruno Kessler (FBK-Irst), Trento, Italy (*Pescado* European project). Ongoing collaboration with Prof. Rodolfo Delmonte, Laboratory of Computational Linguistics, Ca' Foscari, Venezia (*StilVen* project funded by the Venetan Regional Government).

Skills and competences

Mother tongue(s)

Italian

*Self-assessment
European level ^(*)*

German

English

French

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	B2	Independent user	C1	Proficient user	C2	Proficient user
B1	Independent user	B2	Independent user	A2	Basic user	A2	Basic user	A2	Basic user

^(*) 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 keyphrase 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)

Other

Experience in the editing of scientific documents with L^AT_EX;

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 (*scuola di specializzazione biennale*) 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

Invited seminars and talks

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

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.

Sara Tonelli, Emanuele Pianta, Rodolfo Delmonte, Michele Brunelli. *Ven-Pro: a Morphological Analyzer for Venetan*. 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

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

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

2009

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

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 3rd 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.

Professional experiences

Research

Jan 2010 - Present

Human Language Technologies Group, Fondazione Bruno Kessler, Trento
Research activity as a junior researcher for the European project *PESCaDO* (Personalized Environmental Service Configuration and Delivery Orchestration, <http://www.pescado-project.eu>) aimed at the development of resources and tools for the analysis of multilingual texts in the environmental domain.

March 2009 - Present

Laboratory of Computational Linguistics, University Ca' Foscari, Venice
Research consultant for the *StilVen* project, funded by the Venetan Regional Government, for the development of natural language processing tools for Venetan, in particular of an English - Venetan machine translation system.

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 project: pilot study on the annotation of Italian dialogues with frame information. Definition of the annotation manual and of the annotation workflow.

2005 - 2006

Laboratorio di Linguistica Computazionale, 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

Laboratorio di Linguistica Computazionale, 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	<p>2003- 2006 <i>Laboratorio di Linguistica Computazionale, University of Venice, Italy</i> Coordination and teaching activities for the distance learning course <i>Assisted and automatic translation (TATAE)</i> for professional translators</p> <p>Sept 2005- Jul 2006 <i>Laboratorio di Linguistica Computazionale, University of Venice, Italy</i> Assistantship of the course <i>Fundamentals of Prolog programming</i></p> <p>Sept 2001- Jul 2002 <i>Centrum fuer Informations-und Sprachverarbeitung, Munich University, Germany</i> Assistantship of the course <i>Introduction to Perl Programming</i></p>
Tester and linguistic consultant	<p>2000-2002 (partial time) <i>Sail Labs GmbH, Munich, Germany</i> http://www.sail-technology.com Automatic translation software for <i>Langenscheidt</i> Quality Assurance department</p>
Other	<p>I have been part 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) and RTE-2 (Second Recognising Textual Entailment Challenge, 2006).</p> <p>Reviewer of the “23rd International Conference on Computational Linguistics” (Coling-2010). Reviewer of “Constructions and Frames”, John Benjamins Publishing Company</p> <p><i>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.</i></p> <p>7 settembre 2010</p>