Curriculum vitae

Personal information

Giulia Venturi

📍 http://www.italianlp.it/people/giulia-venturi/

📱 +39 050-3152847

✉ giulia.venturi@ilc.cnr.it

in linkedin.com/pub/giulia-venturi/32/379/604

Gender Female | Date of birth 31 January 1982

RESEARCH INTERESTS

Extraction of linguistic and extra–linguistic knowledge from domain–specific corpora.
– NLP-based analysis of domain–specific linguistic varieties, in particular of the legal and medical language, with a specific view to the adaptation of NLP tools to the domain–specific peculiarities.
– NLP-based approaches to semantic processing of legal and biomedical texts.
– NLP-enabled readability assessment of texts across textual genres and domains.
– NLP-enabled "linguistic profiling" of texts.
– NLP-enabled analysis of the evolution of written language competence of L1 Italian learners.

WORK EXPERIENCE

1 July 2019 - Today
Permanent Researcher

Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) (Institute for Computational Linguistics) of the National Research Council (CNR), Pisa.

15 May 2018 - 30 June 2019
Temporary Researcher

Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) (Institute for Computational Linguistics) of the National Research Council (CNR), Pisa.

1 November 2016 - 10 May 2018
Research Grant Senior

Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) (Institute for Computational Linguistics) of the National Research Council (CNR) on the topic "Development of methodologies and advanced techniques for the identification, management and writing of breaking news to support journalists, with particular attention to the analysis of social media", in the framework of the regional project SMART NEWS "Social Sensing for breakingnews" - Regione Toscana - Bando FAR FAS 2014.
Supervision of Dr. Felice Dell'Orletta.

April-July 2016 and March-June 2017
Contract of Cooperation Autonomous Coordinated and Continuous

To support the didactic of Course of "Computational Linguistics II", master degree in "Linguistic Technologies (*Tecnologie del linguaggio*), University of Pisa.

24 November 2014 - 24 October 2016	**Research Grant** *Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC)* (Institute of Computational Linguistics) of the National Research Council (CNR) on the topic "Development of methodologies and advanced techniques for assessing the linguistic complexity of the texts in view of the definition of controlled languages and methods for text simplification, with particular attention to the analysis of specialized languages", in the framework of the national project "Short writings, linguistic simplification, social inclusion: models and applications (SCRIBE)". Supervision of Dr. Simonetta Montemagni.
1 November 2012 - 31 October 2014	**Research Grant** *Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC)* (Institute of Computational Linguistics) of the National Research Council (CNR) for the "Legal Text Mining: building semantic networks to support advanced queries in legal textual corpora (JURNET)" national Project. Supervision of Dr. Simonetta Montemagni.
1 July 2011 - 31 October 2012	**Research Grant** Scuola Superiore Sant'Anna of Pisa on the topic of "Law and Language: methods and techniques for linguistically formalizing the judicial dialogue amongst European courts". Supervision of Prof. Antonio Lazari.
1 April 2010 - 30 April 2011	**Research Grant** *Istituto di Ricerche sulle Attività Terziarie (IRAT)* of the National Research Council (CNR) for the "Migration" national Project on the topic of "Development of methodologies and techniques for monitoring the linguistic competences in Italian of foreign students in the primary and secondary schools through advanced technologies in the field of computational linguistics", as a result of collaboration between the *Istituto di Ricerche sulle Attività Terziarie* and the *Istituto di Linguistica Computazionale (ILC)* of the National Research Council (CNR). Supervision of Dr. Simonetta Montemagni.
1 September 2008 - 30 November 2008	**Collaboration contract** Technical report "A comparative analysis of the existing morphosyntactic annotation schemes (PoS–tagsets) for the Italian language". Department of Linguistics, University of Roma 3.
1 April 2007 - 31 March 2010	**Scholarship** *Istituto di Linguistica Computazionale (ILC)* of the National Research Council (CNR) on the topic of "Methods and techniques for the automatic extraction of lexical–semantic information from corpora of legal and biomedical texts, with a view to the acquisition of domain–specific events". Supervision of Dr. Simonetta Montemagni.

TEACHING EXPERIENCES	■
2013–today	**Academic Course** Within the course of Dr. Simonetta Montemagni "Computational Linguistics II" Master degree in "Linguistic Technologies (*Tecnologie del linguaggio*)", University of Pisa.
2013–today	**Academic Course** Title: 'Natural Language Processing and Legal Knowledge Extraction' Advanced module of the *International Summer School: Managing lexical resources in the semantic web (LEX)* Faculty of Law, University of Bologna (Ravenna).
12 June 2017	**Academic Course** Title: 'Valutazione della complessità linguistica a sostegno dell'attività didattica' Summer school "Strumenti digitali per umanisti", organized by AIUCD (Associazione per l'Informatica Umanistica e la Cultura Digitale)

. 4 May 2017	**Supervision of thesis**
	Master degree in "Linguistic Technologies (*Tecnologie del linguaggio*)", University of Pisa.
	Candidate: Chiara Alzetta
	Title: "Studio linguistico-computazionale per l'analisi dei tipi linguistici. Similarità e differenze nel confronto fra Universal Dependencies Treebanks"

16 febbraio 2017	**Supervision of thesis**
	Master degree in "Linguistic Technologies (*Tecnologie del linguaggio*)", University of Pisa.
	Candidate: Sabrina Rinnone
	Title: "La leggibilità dei consensi informati: un approccio linguistico–computazionale"

23 June 2016	**Doctoral Committe**
	Member of the evaluation committee of Itziar Gonzalez Dios's Ph.D. thesis "Readability Assessment and Automatic Text Simplification. The Analysis of Basque Complex Structures", Prot. n. 201600004647 del 6 giugno 2016.
	University of the Basque Country.

PROFESSIONAL FORMATION ACTIVITIES	
13 October, 10 and 27 November 2017	Training activity aimed at teaching hospital employees the use of language technologies for assessing the linguistic complexity of legal texts, in the framework of the course "Digital Culture: natural language and legal language for websites" organized by the University Hospital of Pisa.
23 and 26 January 2017	Training activity aimed at teaching public employees the use of language technologies for assessing the linguistic complexity of legal and administrative texts, in the framework of the course "The quality of legal and administrative texts" organized by the Unione dei Comuni della Versilia, Querceta.
28 October, 17 November and 1 December 2017	Training activity aimed at teaching hospital employees the use of language technologies for assessing the linguistic complexity of legal texts, in the framework of the course "The user, the usability and perceived quality of the health WEB contents: theory and operational tools" organized by the University Hospital of Pisa.

– 12 September 2019, Title: 'Why is this sentence complex? Cherry-pick the optimal set of features!', "Interactive Workshop on Measuring Language Complexity", 12-12 September, Freiburg, Germany.

– 14 February 2019, Title: 'Natural Language Processing Methods and Instruments for the Extraction of Information from Domain Corpora', Conference "DR2 Distant Reading and Data-Driven Research in the History of Philosophy", 13-14 february, University of Torino.

– 6 March 2019, Title: 'Natural Language Processing methods to support legal communication', Workshop "Legal communication in a multilingual and multicultural context", 5-7 March at Bergische Universität Wuppertal, Germany

– 19 March 2016, Title: 'Italian legal corpora seen through the lenses of linguists and jurists', International Conference "The fabric of law. Discovering patterns through Legal Corpus Linguistics", 18-19 March Heidelberger Akademie Der Wissenschaften.

– 14 May 2015, Title: 'Construction of corpora of legal texts: what, how, what for', Faculty of Law, University 'Pablo de Olavide', Seville (Spain).

– 13 May 2015, Title: 'Natural Language Processing and Linguistic Profiling of Texts: methods and applications', Departamento de Filología Francesa e Inglesa de la Universidad de Cádiz (Spain).

– 30 April 2014, Title: 'Natural Language Processing and Knowledge Extraction', Departamento de Filología Francesa e Inglesa de la Universidad de Cádiz (Spain).

– 11 December 2013, Title: 'Oltre il contenuto tecnologie linguistico–computazionali per l'analisi della struttura linguistica del testo', "Seminario di Cultura Digitale", Department of Computer Science, University of Pisa.

– 5-7 December 2012, Title: 'Computational Linguistics as bridge between the language of law and legal knowledge', within the course of Dr. Antonio Lazari "Introduction to European Private Comparative Law", Scuola Superiore Sant'Anna of Pisa.

– 11 and 17 May 2012, Title: 'Natural Language Processing and Law: methods, techniques and applications', Scuola Superiore Sant'Anna of Pisa.

– 18 May 2012, Title: 'Natural Language Processing and Law: methods, techniques and applications', within the seminar *Linguistic technologies and Law. Towards semantic queries in legal texts*, Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG) (Institute of Legal Information Theory and Techniques) of the National Research Council (CNR), Firenze.

– 2 December 2010, Title: 'Language and law: a computational linguistics perspective', within the seminar *Legal Information and linguistic technologies*, Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG) (Institute of Legal Information Theory and Techniques) of the National Research Council (CNR), Firenze.

– 30 March 2007, Title: 'Methods and techniques for the automatic extraction of lexical–semantic information from corpora of legal texts belonging to the environmental domain. From the construction of a domain–specific TermBank to the acquisition of classes of legal and environmental events', Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG) (Institute of Legal Information Theory and Techniques) of the National Research Council (CNR), Firenze.

– Coordinator of the Student Scholarship and Volunteer Program, during the 57th Annual Meeting of the Association for Computational Linguistics (ACL), 28 July-2 August, 2019.

– Member of the Program Committee of the *Second Workshop on Automated Detection, Extraction and Analysis of Semantic Information in Legal Texts (ASAIL)*, hold in conjunction with the "XVI International Conference on AI and Law (ICAIL 2017)", 16 June 2017, London, UK.

– Member of the Program Committee of the *International Conference on Dependency Linguistics (Depling)*, 18-20 settembre 2017, Pisa.

– Member of the Program Committee of the *Conference on Empirical Methods in Natural Language Processing (EMNLP)* for the years 2016 e 2017.

– Member of the Program Committee of the *International Conference on Computational Linguistics (COLING)* for the year 2016.

– Co–organizer of the workshop on *Computational Linguistics for Linguistic Complexity (CL4LC)*, held in the framework of the *26th International Conference on Computational Linguistics* (COLING 2016), 11 December 2016.

– Cultore della materia for the scientific disciplinary sector L/LIN01.

– Member of the *ItaliaNLP Laboratory* (www.italianlp.it) at the Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR).

– Member of the *Laboratory of Forensic Linguistics (Laboratorio di Linguistica Giudiziaria, LALIGI)* of the Univeristy of Firenze.

– Member of the Organizing Committee of the workshop *Computational Linguistics for Linguistic Complexity*, hosted by the *26th International Conference on Computational Linguistics* (COLING 2016), 11 December 2016.

– Member of the Organizing Committee of the *Workshop on Innovative Use of NLP for Building Educational Applications* since 2014.

– Member of the Organizing Committee of the *International Conference on Language Resources and Evaluation (LREC)* since 2014.

– Member of the Organizing Committee of the *Italian Conference on Computational Linguistics (CLiC-it)* since 2014.

– Member of the Program Committee of the *Workshop on Innovative Use of NLP for Building Educational Applications (BEA)* since 2014.

– Member of the Program Committee of the Special Track on *Rules and Human Language Technology*, hosted by the *8th International Web Rule Symposium (RuleML 2014)*, Prague, Czech Republic, August 18–22, 2014.

– Member of the Organizing Commitee of the *5th Workshop on Semantic Processing of Legal Texts (SPLeT–2014)*, in conjunction with the *9th Language Resources and Evaluation Conference (LREC 2014)*, 27 May 2014, Reykjavik, Iceland.

– Member of the Program Committee of the Special Track on *Translating between Human Language and Formal Rules: Business, Law, and Government*, hosted by the *7th International Web Rule Symposium (RuleML 2013)*, University of Washington, Seattle, Washington, U.S.A., July 11-13th, 2013.

– Member of the Organizing Commitee of the *First Shared Task on Dependency Parsing of Legal Texts*, held during the *4th Workshop on Semantic Processing of Legal Texts (SPLeT–2012)*.

– Member of the Program Committee of the *4th Workshop on Semantic Processing of Legal Texts (SPLeT–2012)*, held in conjunction with the *8th Language Resources and Evaluation Conference (LREC 2012)*, 27 May 2012, Istanbul.

– Member of the Organizing Commitee of the shared task *Domain Adaptation for Dependency Parsing*, held during *Evalita 2011 (Evaluation of NLP and Speech Tools for Italian)*.

– Member of the Organizing Commitee of the shared task *Frame Labeling over Italian Texts Task*, held during *Evalita 2011 (Evaluation of NLP and Speech Tools for Italian)*.

EDUCATION AND TRAINING

2008-2011

Ph.D. degree in Linguistics, Applied Linguistics, Linguistic Engineering

Thesis Title: 'Lingua e Diritto: una prospettiva linguistico–computazionale (Language and law: a computational linguistics perspective)'

University of Torino (Italy)
Supervisor: Prof. Mario Squartini
Members of examination committee: Prof. Mario Squartini, Dr. Simonetta Montemagni, Prof. Mirko Tavoni, Prof. Antonio Lazari.

Master degree in Italian Language and Literature

2004-2006

Thesis Title: 'L'ambiente, le norme, il computer. Uno studio linguistico–computazionale per la creazione di ontologie giuridiche in materia ambientale (The environment, the norms, the computer. A computational–linguistics study for the creation of legal ontologies in the environmental domain)'

University of Pisa (Italy)
Mark: 110/110 cum laude
Supervisors: Prof. Mirko Tavoni, Dr. Simonetta Montemagni, Prof. Alessandro Lenci.

Bachelor degree in Literature

2001-2004

Thesis Title: 'Ognunu parla comu sa'. Analisi sociolinguistica della sceneggiatura di *Così ridevano* di Gianni Amelio' ('Ognunu parla comu sa'. Sociolinguistic analysis of the screeplay of the movie *Così ridevano* by Gianni Amelio)

University of Pisa (Italy)
Mark: 110/110 cum laude
Supervisors: Prof. Mirko Tavoni and Prof. Lorenzo Cuccu.

PERSONAL SKILLS

Mother tongue(s)   Italian

Other language(s)

Understanding		Speaking		Writing
Listening	Reading	Spoken interaction	Spoken production	
English				
B2	B2	B2	B2	B2
French				
A3	A3	A3	A3	A3

Levels: A1/A2: Basic user - B1/B2: Independent user - C1/C2: Proficient user
Common European Framework of Reference (CEF) level

PUBLICATIONS

Chapters in books

Agnoloni T. e Venturi G. (2018), *Semantic Processing of Legal Texts*. In Visconti J. (ed.), *Handbook of Communication in the Legal Sphere*, Walter de Gruyter, pp. 109-137, ISBN 978-1-948087-84-1.

Adorni G., Dell'Orletta F., Koceva F., Torre I. e Venturi G. (2018), *Extracting Dependency Relations from Digital Learning Content*. In Serra G. and Tasso G. (eds.), *Digital Libraries and Multimedia Archives. 14th Italian Research Conference on Digital Libraries (IRCDL 2018)*, pp. 114-119, ISBN 978-3-319-73164-3.

Brunato D., Dell'Orletta F., Montemagni S. e Venturi G. (2018), *Monitoraggio di scritture brevi: aspetti metodologici e primi risultati*. A. Manco e A. Mancini (eds.), *Scritture Brevi: segni, testi e contesti. Dalle iscrizioni antiche ai tweet*, Collana di studi Quaderni di AION-Linguistica, Università di Studi di Napoli "L'Orientale", Napoli, pp. 149-176, ISBN 978-88-6719-140-6.

Brunato D. e Venturi G. (2016), *Le tecnologie linguistico-computazionali per la leggibilità della comunicazione istituzionale*. In: Saulle Panizza (ed.), *Profili attuali di qualità degli atti normativi e amministrativi*, Pisa University Press, pp. 119–157, ISBN 978-886741-6271.

Dell'Orletta F., Montemagni S. e Venturi G. (2016), *Esplorazioni computazionali nello spazio dell'interlingua: verso una nuova metodologia di indagine*. In R. Bombi e V. Orioles (eds.) *Atti del XLVIII Congresso Internazionale di Studi della Società di Linguistica Italiana* (SLI 2014), 25–27 settembre Udine, Bulzoni Editore, pp. 143-161, ISBN: 978-88-6897-029-1.

Dell'Orletta F., Venturi G. (2016), *ULISSE: una strategia di adattamento al dominio per l'annotazione sintattica automatica*. In E. M. Ponti e M. Baudassi (eds.) "Compter parler soigner: tra linguistica e intelligenza artificiale", atti del convegno 15-17 dicembre 2014, Pavia University Press, pp. 55-79, ISBN: 978-88-6952-038-9.

Sagri M.T., Tiscornia D., Montemagni S., Venturi G. (2015), *Neuroscienze e genetica comportamentale in un corpus di sentenze italiane alla luce dei risultati di elaborazioni linguistico-computazionali*. In: E. Fabiani, S. Faro, N. Lettieri (eds.), *Diritto, Neuroscienze, Scienze della Cognizione*, Napoli, Edizioni Scientifiche Italiane, pp. 339-366, ISBN: 978-88-495-2932-6.

Venturi G. (2013), *Semantic annotation of Italian legal texts: A FrameNet-based approach*. In: Mirjam Fried; Nikiforidou Kiki (eds.), *Advances in Frame Semantics*, Benjamins Current Topics, John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 51-84, ISBN: 978-90-272-0277-2. Lo stesso contenuto è stato pubblicato anche nell'articolo su rivista in *Constructions and Frames*.

Dell'Orletta F., Marchi S., Montemagni S., Venturi G., Agnoloni T. e Francesconi E. (2013), *Domain Adaptation for Dependency Parsing at Evalita 2011*. In Magnini B., Cutugno F., Falcone M., Pianta E. (eds.), *Evaluation of Natural Language and Speech Tool for Italian*, LNCS-LNAI, Vol. 7689, Springer-Verlag Berlin Heidelberg, pp. 58-69, ISBN 978-3-642-35827-2.

Basili R., Lenci A., De Cao D. e Venturi G. (2013), *EvalIta 2011: the Frame Labeling over Italian Texts Task*. In Magnini B., Cutugno F., Falcone M., Pianta E. (eds.), *Evaluation of Natural Language and Speech Tool for Italian*, LNCS-LNAI, Vol. 7689, Springer-Verlag Berlin Heidelberg, pp. 195-204, ISBN 978-3-642-35827-2.

Bonin F., Dell'Orletta F., Venturi G. and Montemagni S. (2012), *Lessico settoriale e lessico comune nell'estrazione di terminologia specialistica da corpora di dominio*. In S. Ferreri (eds.), *Atti del XLIV Congresso Internazionale di Studi della Società di Linguistica Italiana* (SLI 2010), 27-29 settembre Viterbo, pp. 207-220, ISBN: 978-88-7870-655-2.

Dell'Orletta F., Montemagni S., Vecchi E. M. e Venturi G. (2011), *Tecnologie linguistico-computazionali per il monitoraggio della competenza linguistica italiana degli alunni stranieri nella scuola primaria e secondaria*. In Giovanni Carlo Bruno, Immacolata Caruso, Manuela Sanna, Immacolata Vellecco (eds.), *Percorsi migranti: uomini, diritto, lavoro, linguaggi*, McGraw-Hill, pag. 319-336, ISBN 978-88-386-7296-5.

Venturi G. (2010), *Legal Language and Legal Knowledge Management Applications*. In: E. Francesconi, S. Montemagni, W. Peters and D. Tiscornia (eds.), *Semantic Processing of Legal Texts. Where the Language of Law Meets the Law of Language*, Lecture Notes in Computer Science, vol. 6036, Berlin Heidelberg, Springer-Verlag, p. 3-26, ISBN: 3-642-12836-X.

Agnoloni T., Fernández Barrera M., Sagri M.T., Tiscornia D. and Venturi G. (2010), *When a FrameNet-style knowledge description meets an ontological characterization of Fundamental Legal Concepts*. In P. Casanovas, U. Pagallo, G. Ajani, G. Sartor (eds.), *AI Approaches to the Complexity of Legal Systems* (AICOL 2009), LNCS, Springer, pp. 93-112, ISBN 978-3-642-16523-0.

Agnoloni T., Bacci L., Francesconi E., Peters W., Montemagni S., Venturi G. (2009), *A two-level knowledge approach to support multilingual legislative drafting*. In: Joost Breuker, Pompeu Casanovas, Michel C.A. Klein, Enrico Francesconi (eds.), *Law, Ontologies and the Semantic Web - Channelling the Legal Information Flood*, Frontiers in Artificial Intelligence and Applications, Amsterdam, IOS Press, p. 177-198, ISBN: 978-1-58603-942-4.

Venturi G., Montemagni S., Marchi S., Sasaki Y., Thompson P., McNaught J. and Ananiadou S. (2009), *Bootstrapping a Verb Lexicon for Biomedical Information Extraction*. In Alexander Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing*, LNCS 5449, Springer-Verlag Berlin Heidelberg, pp. 137-148, ISBN 3-642-00381-8.

Lenci A., Montemagni S., Pirrelli V., Venturi G. (2009), *Ontology learning from Italian legal texts*. In Joost Breuker, Pompeu Casanovas, Michel C.A. Klein, Enrico Francesconi (eds.), *Law, Ontologies and the Semantic Web - Channelling the Legal Information Flood*, Frontiers in Artificial Intelligence and Applications, Springer, Volume 188, pp. 75-94, ISBN 978-1-58603-942-4.

Agnoloni T., Bacci L., Francesconi E., Spinosa P., Tiscornia D., Montemagni S., Venturi G. (2008), *Building an ontological support for multilingual legislative drafting*. In Arno R. Lodder and Laurens Mommers (eds.), *Legal Knowledge and Information Systems - JURIX 2007: The Twentieth Annual Conference on Legal Knowledge and Information Systems*, Leiden, The Netherlands, 12–15 December 2007, Frontiers in Artificial Intelligence and Applications, vol. 165, IOS Press, pp. 9-18, ISBN 978-1-58603-810-6.

Curatorships

Brunato D., Dell'Orletta F., Venturi G., François T., Blache P. (2016), *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC 2016)*, workshop della 26th International Conference on Computational Linguistics (COLING), Osaka, Japan, 11 December 2016, ISBN: 978-4-87974-709-9.

Journal papers

Dell'Orletta F., Greco S., Montemagni S., Morini E., Rossi F., Sagri M.T., Venturi G. (2019), *Le parole del miglioramento. Come le scuole descrivono il cambiamento*. In *Psicologia dell'Educazione*, n. 1/2019, pp. 47-68, ISSN: 1971-3711.

Venturi G., Dell'Orletta F., Montemagni S., Flore E., Bellandi T. (2017), *La qualità dei consensi informati. Un'analisi linguistico-computazionale della leggibilità dei testi*. In *Salute e Territorio*, anno XXXVIII, Pacini Editore, pp. 35-39, ISSN 0392-4505.

Barbagli A., Lucisano P., Dell'Orletta F., Montemagni S., Venturi G. (2015), *Il ruolo delle tecnologie del linguaggio nel monitoraggio dell'evoluzione delle abilità di scrittura: primi risultati*. In R. Basili e S. Montemagni (eds.), *Italian Journal of Computational Linguistics*, 1(1), Accademia University press, pp. 99-117, ISSN: 2499-4553, ISBN: 978-88-99200-63-3.

Brunato D., Venturi G. (2014), *Le tecnologie linguistico-computazionali nella misura della leggibilità di testi giuridici*. In: Tiscornia D., Romano F., Sagri M.T. (eds.), *Diritto, Linguaggio e tecnologie dell'informazione*, fascicolo monografico di *Informatica e Diritto*, numero 2014/1, pp. 111-142, ISSN: 0390-0975.

Dell'Orletta F., Montemagni S., Venturi G. (2014), *Assessing document and sentence readability in less resourced languages and across textual genres*. In: T. François, D. Bernhard (eds.), *International Journal of Applied Linguistics (ITL)*, John Benjamins Publishing Company, Amsterdam/Philadelphia, vol. 165(2), pp. 163-193, ISSN: 0019-0829.

Dell'Orletta F., Venturi G., Montemagni S. (2013), *Linguistically-driven Selection of Correct Arcs for Dependency Parsing*. In: *Computacón y Sistemas*, vol. 17, p. 125-136, ISSN: 1405-5546. **Best Verifiability, Reproducibility, and Working Description award**.

Venturi G. (2011), *Semantic annotation of Italian legal texts: a FrameNet-based approach*. In: K. Ohara, K. Nikiforodou (eds.), *Constructions and Frames*, John Benjamins Publishing Company, Amsterdam/Philadelphia, vol. 3(1), pp. 46-79, ISSN: 1876-1933. Lo stesso contenuto è stato pubblicato anche nell'articolo contenuto nel libro *Advances in Frame Semantics*.

P. Thompson, S. Montemagni, J. McNaught, N. Calzolari, R. del Gratta, V. Lee, S. Marchi, M. Monachini, P. Pezik, V. Quochi, D. Rebholz-Schuhmann, C.J. Rupp, Y. Sasaki, G. Venturi and S. Ananiadou (2011), *The BioLexicon: a large-scale terminological resource for biomedical text mining*. In *Journal BMC Bionformatics*, pp. 12-397, ISSN: 1471-2105.

Bonin F., Dell'Orletta F., Venturi G., Montemagni S. (2010), *Singling out Legal Knowledge from World Knowledge*. In: E. Francesconi (ed.), *Informatica e Diritto*, vol 19(1-2), pp. 217-229, ISSN: 0390-0975.

Dell'Orletta F., Lenci A., Marchi S., Montemagni S., Pirrelli V., Venturi G. (2008), *Dal testo alla conoscenza e ritorno: estrazione terminologica e annotazione semantica di basi documentali di dominio*. In *AIDA Informazioni*, Atti del Convegno Nazionale Ass.I.Term "I–TerAnDo", Università della Calabria, 5–7 giugno 2008, Roma: AIDA, n. 1-2/2008, ISSN 1121-0095, pp. 185-206.

Referred conference papers

Cocciu E., Brunato D., Venturi G., Dell'Orletta F. (2018), *Gender and Genre Linguistic profiling: a case study on female and male journalistic and diary prose*. In: *Proceedings of the 5th Italian Conference on Computational Linguistics (CLiC-it)*, Torino, 10-12 dicembre.

De Felice I., Dell'Orletta F., Venturi G., Lenci A., Montemagni S. (2018), *Italian in the Trenches: Linguistic Annotation and Analysis of Texts of the Great War*. In: *Proceedings of the 5th Italian Conference on Computational Linguistics (CLiC-it)*, Torino, 10-12 dicembre.

Chiriatti G., Della Gala V., Dell'Orletta F., Montemagni S., Pettenati M.C., Sagri M.T., Venturi G. (2018), *A NLP-based Analysis of Reflective Writings by Italian Teachers*. In: *Proceedings of the 5th Italian Conference on Computational Linguistics (CLiC-it)*, Torino, 10-12 dicembre.

Cimino A., Dell'Orletta F., Brunato D., Venturi G. (2018), *Sentences and Documents in Native Language Identification*. In: *Proceedings of the 5th Italian Conference on Computational Linguistics (CLiC-it)*, Torino, 10-12 dicembre.

Brunato D., De Mattei L., Dell'Orletta F., Iavarone B., Venturi G. (2018), *Is this sentence difficult? Do you agree?*. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)*, October 31-November 4, Bruxelles, pp. 1-12, ISBN: 978-1-948087-84-1.

Alzetta C., Dell'Orletta F., Montemagni S., Venturi G. (2018), *Assessing the Impact of Iterative Error Detection and Correction. A Case Study on the Italian Universal Dependency Treebank*. In: *Proceedings of the Universal Dependencies Workshop 2018 (UDW 2018)*, workshop della "Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)", November 1, Bruxelles, pp. 1-8, ISBN: 978-1-948087-84-1.

Alzetta C., Dell'Orletta F., Montemagni S., Venturi G. (2018), *Acquiring Typological Evidence from Universal Dependency Treebanks*. In: *11th Edition of International Conference on Language Resources and Evaluation (LREC 2018)*, 7-12 May, Miyazaki, Japan, pp. 4540-4549, ISBN: 979-10-95546-00-9.

Alzetta C., Dell'Orletta F., Montemagni S., Venturi G. (2018), *Dangerous Relations in Dependency Treebanks*. In: *16th International Workshop on Treebanks and Linguistic Theories*, 23-24 January, Praga, Repubblica Ceca, pp. 201-210, ISBN: 978-80-88132-04-2.

Cimino A., Wieling M., Dell'Orletta F., Montemagni S. e Venturi G. (2017), *Identifying Predictive Features for Textual Genre Classification: the Key Role of Syntactic Features*. In *Proceedings of the 4th Italian Conference on Computational Linguistics (CLiC-it)*, Napoli, 11-13 dicembre, pp. 107-112, ISSN: 1613-0073.

Tusa E., Dell'Orletta F., Montemagni S., Venturi G. (2016), *Dieci sfumature di marcatezza sintattica: verso una nozione computazionale di complessità*. In *Proceedings of the 3nd Italian Conference on Computational Linguistics (CLiC-it)*, 5-6 Dicembre, Napoli, pp. 297-302, ISSN: 1613-0073.

Brunato D., Cimino A., Dell'Orletta F. e Venturi G. (2016), *PaCCSS–IT: A Parallel Corpus of Complex-Simple Sentences for Automatic Text Simplification*. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)*, November 2-6, Austin, Texas, USA, pp. 351-361, ISBN: 978-1-945626-25-8.

Barbagli A., Lucisano P., Dell'Orletta F., Montemagni S., Venturi G. (2016), *CItA: an L1 Italian Learners Corpus to Study the Development of Writing Competence*. In *Proceedings of 10th Edition of International Conference on Language Resources and Evaluation (LREC 2016)*, 23-28 May, Portorož, Slovenia, pp. 88–95, ISBN: 978-2-9517408-9-1.

Richter S., Cimino A., Dell'Orletta F., Venturi G. (2015), *Tracking the Evolution of Written Language Competence: an NLP–based Approach*. In *Proceedings of the 2nd Italian Conference on Computational Linguistics (CLiC-it)*, 3-4 December, Trento, Italy, pp. 236–240, ISBN: 978-88-99200-62-6.

Barbagli A., Lucisano P., Dell'Orletta F., Montemagni S., Venturi G. (2015), *CItA: un Corpus di Produzioni Scritte di Apprendenti l'Italiano L1 Annotato con Errori*. In *Proceedings of the 2nd Italian Conference on Computational Linguistics (CLiC-it)*, 3-4 December, Trento, Italy, pp. 31–35, ISBN: 978-88-99200-62-6.

Venturi G., Bellandi T., Dell'Orletta F., Montemagni S. (2015), *NLP–Based Readability Assessment of Health-Related Texts: a Case Study on Italian Informed Consent Forms*. In *Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis (Louhi 2015)*, workshop della "Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)", 17 September, Lisbon, Portugal, 131–141, ISBN: 978-1-941643-32-7.

Brunato D., Dell'Orletta F., Venturi G., Montemagni S. (2015), *Design and Annotation of the First Italian Corpus for Text Simplification*. In *Proceedings of the 9th Linguistic Annotation Workshop (LAW'15)*, workshop della "Conference of North American Chapter of the Association for Computational Linguistics (NAACL 2015)", 5 June, Denver, Colorado, USA, pp. 31-34, ISBN: 978-1-941643-47-1.

Brunato D., Dell'Orletta F., Venturi G., Montemagni S. (2014), *Defining an annotation scheme with a view to automatic text simplification*. In *Proceedings of the First Italian Conference on Computational Linguistics (CLiC-it)*, 9-10 December, Pisa, Italy, pp. 87-92, ISBN:978-8-86741-472-7.

Barbagli A., Lucisano P., Dell'Orletta F., Montemagni S., Venturi G. (2014), *Tecnologie del linguaggio e monitoraggio dell'evoluzione delle abilità di scrittura nella scuola secondaria di primo grado*. In *Proceedings of the First Italian Conference on Computational Linguistics (CLiC-it)*, 9-10 December, Pisa, Italy, pp. 23-27, ISBN:978-8-86741-472-7. **Distinguished Young Paper mention**.

Dell'Orletta F., Wieling M., Cimino A., Venturi G., Montemagni S. (2014), *Assessing the readability of sentences: which corpora and features?*. In: Proceedings of the *9th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2014)*, workshop della "Annual Conference of the Association for Computational Linguistics (ACL 2014)", 26 giugno, Baltimore, Maryland, pp. 163-173, ISBN: 978-1-941643-03-7.

Dell'Orletta F., Venturi G., Cimino A., Montemagni S. (2014), *T2K: a System for Automatically Extracting and Organizing Knowledge from Texts*. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14*, pp. 2062-2070, Reykjavik, Iceland, 26-31 May, ISBN: 978-2-9517408-8-4.

Boschetti F., Cimino A., Dell'Orletta F., Lebani G.E., Passaro L., Picchi P., Venturi G., Montemagni S., Lenci A. (2014),*Computational Analysis of Historical Documents: An Application to Italian War Bulletins in World War I and War II*. In:*Proceedings of the Language resources and technologies for processing and linking historical documents and archives Workshop - Deploying Linked Open Data in Cultural Heritage*, workshop della "International Conference on Language Resources and Evaluation (LREC 2014)," pp. 70-75, Reykjavik, Iceland, 26 May ISBN: 978-2-9517408-8-4.

Venturi G. (2013), *Investigating legal language peculiarities across different types of Italian legal texts: an NLP-based approach*. In *Bridging the Gap(s) between Language and the Law. Proceedings of the 3rd European Conference of the International Association of Forensic Linguistics*, Porto, Portugal, 15–18 October 2012, pp. 138–156, ISBN: 978-989-8648-14-3.

Dell'Orletta F., Montemagni S., Venturi G. (2013), *Linguistic Profiling of Texts Across Textual Genre and Readability Level. An exploratory Study on Italian Fictional Prose*. In: *Proceedings of the Recent Advances in Natural Language Processing International Conference (RANLP 2013)*, pp. 189-197, Hissar, Bulgaria, 7-13 September, ISSN: 1313-8502.

Cimino A., Dell'Orletta F., Venturi G., Montemagni S. (2013), *Linguistic Profiling based on General-purpose Features and Native Language Identification*. In: *Proceedings of 8th Workshop on "Innovative Use of NLP for Building Educational Applications"*, workshop della "Conference of North American Chapter of the Association for Computational Linguistics (NAACL 2013)", pp. 207-215, Atlanta (Georgia), 13 June, ISBN: 978-1-937284-47-3.

Dell'Orletta F., Venturi G., Montemagni S. (2013), *Unsupervised Linguistically-Driven Reliable Dependency Parses Detection and Self-Training for Adaptation to the Biomedical Domain*. In: *Proceedings of the 12th Workshop on "Biomedical Natural Language Processing" (BioNLP-2013)*, workshop della "Conference of North American Chapter of the Association for Computational Linguistics (ACL 2013)", pp. 45-53, Sofia, Bulgaria, 8-9 August, ISBN: 978-1-937284-55-8.

Venturi G. (2012), *Design and Development of TEMIS: a Syntactically and Semantically Annotated Corpus of Italian Legislative Texts*. In: *Proceedings of the 4th Workshop on "Semantic Processing of Legal Texts"*, workshop della "International Conference on Language Resources and Evaluation (LREC 2012)", pp. 1-12, Istanbul, Turchia, 27 May, ISBN: 978-2-9517408-7-7.

Dell'Orletta F., Marchi S., Montemagni S., Plank B. e Venturi G. (2012), *The SPLeT–2012 Shared Task on Dependency Parsing of Legal Texts*. In *Proceedings of the 4th Workshop on "Semantic Processing of Legal Texts"*, workshop della "International Conference on Language Resources and Evaluation (LREC 2012)", pp. 42-51, Istanbul, Turkey, 27 May, ISBN: 978-2-9517408-7-7.

Lenci A., Montemagni S., Venturi G. e Cutrullà M.R. (2012), *Enriching the ISST–TANL Corpus with Semantic Frames*. In *Proceedings of the 8th International Conference on Language Resources and Evaluation* (LREC 2012), Istanbul, Turkey, 21–27 May, pp. 3719-3726, ISBN: 978-2-9517408-7-7.

Basili R., Lenci A., De Cao D. e Venturi G. (2012), *Evalita 2011: the Frame Labeling over Italian Texts Task*. In *Working Notes of EVALITA 2011*, 24-25 January 2012, Rome, Italy, ISSN 2240-5186.

Dell'Orletta F., Marchi S., Montemagni S., Venturi G., Agnoloni T. e Francesconi E. (2012), *Domain Adaptation for Dependency Parsing at Evalita 2011*. In *Working Notes of EVALITA 2011*, 24-25 January 2012, Rome, Italy, ISSN 2240-5186.

Dell'Orletta F., Montemagni S. and Venturi G. (2011), *READ–IT: Assessing Readability of Italian Texts with a View to Text Simplification*. In *Proceedings of Speech and Language Processing for Assistive Technologies (SLPAT 2011)*, workshop della "Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)", Edinburgh, July 30, pp. 73-83, ISBN: 978-1-937284-14-5.

Dell'Orletta F., Venturi G. and Montemagni S. (2011), *ULISSE: an Unsupervised Algorithm for Detecting Reliable Dependency Parses*. In *Proceedings of the 15th Conference on Computational Natural Language Learning* (CoNLL 2011), Portland, Oregon, USA, June 23–24, pp. 115-124, ISBN: 978-1-932432-92-3.

Bonin F., Dell'Orletta F., Venturi G. and Montemagni S. (2010), *A Contrastive Approach to Multi–word Extraction from Domain–specific Corpora*. In *Proceedings of the 7th International Conference on Language Resources and Evaluation* (LREC 2010), La Valletta, Malta, 19-21 May, pp. 3222-3229, ISBN: 2-9517408-6-7.

Bonin F., Dell'Orletta F., Venturi G. and Montemagni S. (2010), *Contrastive filtering of domain–specific multi–word terms from different types of corpora*. In *Proceedings of Multiword Expressions: from Theory to Applications (MWE 2010)*, workshop della "International Conference on Computational Linguistics (COLING 2010)", Beijing, China, August 28, pp. 76-79, ISBN: 978-7-900268-00-6.

Venturi G., Lenci A., Montemagni S., Vecchi E.M., Sagri M.T., Tiscornia D., Agnoloni T. (2009), *Towards a FrameNet Resource for the Legal Domain*. In *Proceedings of the III Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT'09)*, workshop della "International Conference on Artificial Intelligence and Law (ICAIL 2009)", Barcellona, 8 June, pp. 1-10, ISSN: 1613-0073.

Spinosa P., Giardiello G., Cherubini M., Marchi S., Venturi G., Montemagni S. (2009), *NLP–based Metadata Extraction for Legal Text Consolidation*. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law* (ICAIL 2009), Barcellona, 8-12 June, pp. 40-49, ISBN: 978-1-60558-597-0.

Venturi G. (2008), *Parsing Legal texts. A contrastive study with a View to Knowledge Management Applications*. In *Proceedings of the Ist Workshop Semantic Processing of Legal Texts*, workshop della "International Language Resources and Evaluation Conference (LREC 2008)", Marrakech, Morocco, May 26-1, pp, 1-10, ISBN: 2-9517408-4-0.

Thompson P., Venturi G., McNaught J., Montemagni S. and Ananiadou S. (2008), *Categorising Modality in Biomedical texts*. In *Proceedings of the Workshop Building and Evaluating Resources for Biomedical Text Mining*, workshop della "International Language Resources and Evaluation Conference (LREC 2008)", Marrakech, Morocco, May 26-1, pp. 27-34, ISBN: 2-9517408-4-0.

Thompson P., Cotter P., Ananiadou S., McNaught J., Montemagni S., Trabucco A. and Venturi G. (2008), *Building a bio–event annotated corpus for the acquisition of semantic frames from biomedical corpora*. In *Proceedings of the 6th International Conference on Language Resources and Evaluation* (LREC), Marrakech, Morocco, May 26-1 June, pp. 2159-2166, ISBN: 2-9517408-4-0.

Dell'Orletta F., Lenci A., Montemagni S., Marchi S., Pirrelli V., Venturi G. (2008), *Acquiring Legal Ontologies from Domain-specific Texts*. In *Proceeding of LangTech 2008*, Rome, 28–29 February.

Lenci A., Montemagni S., Pirrelli V., Venturi G. (2007), *NLP-based ontology learning from legal texts. A case study*. In *Proceedings of the II Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT'07)*, workshop della International Conference on Artificial Intelligence and Law (ICAIL 2007)", 4th June, Stanford University, Stanford, CA USA, pp. 113-130, ISBN: 978-1-59593-680-6.

Deliverables and technical reports

Carlino M., Cimino A., Dell'Orletta F., Venturi G. (2018), *Text Mining*, deliverable dell'obiettivo operativo 3 "Social Media Analysis/Mining" nell'ambito del progetto regionale SMARTNEWS (Social sensing for breaking news, PAR FAS 2007-2013 - LINEA A), Documento D3.2, marzo 2018.

Venturi G. (2009), *Rassegna comparativa degli schemi di annotazione morfosintattica per la lingua italiana*, Rapporto Tecnico TRIPLE - RTT/1, disponibile sul sito TRIPLE alla pagina http://host.uniroma3.it/laboratori/triple/Risorse_Resources.html

Montemagni S., Venturi G., Thompson P., Ananiadou S., McNaught J., Kim J., Rebholz D., Pezik P. (2008), *Augmented version of the bio-lexicon extended with bio-event information and term-to-term weighted links*, BOOTStrep project (Bootstrapping Of Ontologies and Terminologies STrategic REsearch Project, FP6 - 028099), Deliverable 3.2, settembre 2008.

Montemagni S., Trabucco A., Venturi G., Thompson P., Cotter P., Ananiadou S., McNaught J., Kim J., Rebholz D., Pezik P. (2007), *Event annotation of domain corpora*, BOOTStrep project (Bootstrapping Of Ontologies and Terminologies STrategic REsearch Project, FP6 - 028099), Deliverable 4.1, novembre 2007.

Montemagni S., Trabucco A., Venturi G. (2007), *Bio–Event Linguistic Annotation Tool*. User Manual. BOOTStrep project (Bootstrapping Of Ontologies and Terminologies STrategic REsearch Project, FP6 - 028099), WP4, settembre 2007.

Marchi S., Montemagni S., Venturi G., Bartolini R., Bertagna F., Ruffolo P., Peters W., Tiscornia D. (2007), *Report on Ontology learning tool and testing*, DALOS project (Drafting Legislation with Ontology-based Support Project), Deliverable D3.3, dicembre 2007.