Lorenzo Gregori

Curriculum Vitae

✉ lorenzo.gregori@unifi.it

Education

2011–2015 **PhD in Computer Engineering**, *University of Florence*, Italy.
Area of study: Natural Language Processing

2009–2010 **Master degree in Digital Humanities**, *University of Florence*, Italy.
Area of study: Computational Linguistics

2003–2008 **Bachelor degree in Computer Science**, *University of Florence*, Italy.
Area of study: Algorithms and data structures

Job Experience

2019–Present **Lab Technician**, University of Florence, Italy.
Technician at Digital Humanities Lab, Humanities Department

2017–2019 **Research Fellow**, University of Florence, Italy.
Within the IMAGACT-MED Project (PAR/FAS Regione Toscana), 2017-2019.

2017 **Research Fellow**, Cognitive System Research Institute, Athens, Greece.
3 months scientific mission at CSRI (Athens) within the The European Network on Integrating Vision and Language (iV&L Net) ICT COST Action IC1307.

2011–2017 **Research Fellow**, University of Florence, Italy.
- MODELACT Project (Futuro in Ricerca 2012), 2013-2016
  - Ontology linking;
  - Semantic analysis of action verbs;
  - IMAGACT Database rebuilding and migration.
- IMAGACT Project (PAR/FAS Regione Toscana), 2011-2013
  - Linguistic annotation for the development of IMAGACT Ontology of Action;
  - Collaboration in RIDIRE Corpus creation;
  - Spoken Language Database development.

2010–2011 **Administrative Assistance**, University of Florence, Italy.
400 hours of administrative job at Humanities Department of the University of Florence.

2009–2010 **Teacher of Computer Science**, Centro Studi Mugello snc, Borgo S. Lorenzo, Florence, Italy.
1 year of computer science teaching in an Industrial Technical Institute classroom.

2007–2009 **Software analyst and tester**, Fondiaria-SAI, Florence, Italy.
Employed in IT area at Fondiaria-SAI Insurance Company.

2007–2009 **Teacher of Computer Science**, Municipal Library of Vaglia, Florence, Italy.
3 courses of Basic Computing, 30 hours each.

2006–2007 **Librarian**, Municipal Library of Vaglia, Florence, Italy.
Employed for 1 year within the National civil service programme.

Language Skills

Italian Native
English Fluently

Computer Skills

Platforms Windows, Linux
Languages Java, C, PHP, Linux shell, CQL, SQL, XPath, XQuery, Cypher, HTML, XML, RDF, OWL, CSS, TEI
Databases MySQL, DB2, eXist, Neo4J

Computational Linguistics Skills

Skills developed in MA degree program, PhD course, within IMAGACT and MODELACT Projects, covering the following research areas:
○ Ontology mapping and linking;
○ Linguistic annotation;
○ Corpus querying and management;
○ Verb semantics;
○ Text processing;
○ Word sense disambiguation;
○ Information retrieval;
○ Knowledge representation.

Talks in Conferences

2018 GSCP 2018: La comunicazione parlata 2018. December 12-14, 2018, Naples (Italy). Corsi, G., Gagliardi, G., Gregori, G. *DILLo: Database Italiano del Lessico per Logopedisti*.

2018 Korp Workshop 2018. October 16, 2018, Stockholm (Sweden). Cimmino, D., Gregori, L. *Using Korp for the simplification of Italian professional language (CITPRO Project)*.

2018 XII Congresso FLI: Il Logopedista alla sfida delle conoscenze e delle nuove competenze. October 4-6, 2018, Palermo (Italy). Corsi, G., Gagliardi, G., Gregori, L., Lippi, A., Shabana, S., Panunzi, A. *La Logopedia incontra Linguistica e Informatica. Ricerca interdisciplinare per lo sviluppo di nuovi strumenti per la pratica logopedica*.

2018 Workshop on Shortcomings in Vision and Language (SiVL) @ European Conference on Computer Vision (ECCV 2018). September 8, 2018, Munich (Germany). Ravelli, A.A., Gregori, L., Seidenari, L., *Semantic Action Discrimination in Movie Description Dataset*.

2017 Workshop: Linking Data, Ontologies and Distributional Models for the Representation of Lexical Meaning. June 9, 2017, Florence (Italy). L. GREGORI, A. PANUNZI, A.A. RAVELLI, *Linking IMAGACT Ontology of Actions to Semantic Resources and Knowledge Basis.*

2016 MODELACT Conference. Action, Language and Cognition. June 6-7, 2016, Rome (Italy). L. GREGORI, *A new IMAGACT data model to address the action verb translation challenge.*

2016 The Luxembourg BabelNet Workshop. March 2-3, 2016, Luxembourg (Luxembourg). L. GREGORI, A.A. RAVELLI, A. PANUNZI, *Enriching BabelNet verbal entities with videos.*

2015 IX LABLITA and IV LEEL International Workshop. Units of Reference for Spontaneous Speech Analysis and their correlations across languages. August 3-7, 2015, Belo Horizionte (Brazil). A. PANUNZI, B. ROCHA, L. GREGORI, *Comparing annotations for the prosodic segmentation of spontaneous speech. Focus on the reference units.*

2015 4th MODELACT Meeting. June 11, 2015, Florence (Italy). L. GREGORI, Y. PAN, A. PANUNZI, *Modellizzazione dello spazio azionale e confronti cross-linguistici.*

2015 LABLITA Workshop in honor of I. Korzen. The Annotation of Nordic Languages in IMAGACT4ALL. April 15-17, 2015, Florence (Italy). Y. PAN, L. GREGORI, *What we learned implementing Chinese in the IMAGACT4ALL Heavy interface.*

2014 MODELACT Midterm Meeting. October 29, 2014, Florence (Italy). A. PANUNZI, L. GREGORI, *LABLITA Report.*

2011 VII LABLITA International Workshop in Corpus Linguistics. New annotation paradigms for spoken romance languages databases. A. PANUNZI, L. GREGORI, *IPIC: an XML database for the study of Information Patterning.*

Publications

2020 Gagliardi, G., Gregori, L., Ravelli, A.A. (2020). An NLP Pipeline as Assisted Transcription Tool for Speech Therapists. In D. Kokkinakis, K. Lundholm Fors, C. Themistocleous, M. Antonsson, M. Eckerström (eds), *Proceedings of the LREC 2020 Workshop on:Resources and Processing of Linguistic, Para-linguisticand Extra-linguistic Data from People with Various Forms ofCognitive/Psychiatric/Developmental Impairments (RaPID-3)*. ELRA, Paris. pp. 124-130.

2020 Gregori, L., Panunzi, A., Rocha, B. (2020). SLAC Database. `http://dx.doi.org/10.1075/scl.94.slac`

2020 Panunzi, A., Gregori, L., Rocha, B. (2020). Comparing annotations for the prosodic segmentation of spontaneous speech. In S. Izre'el, H. Mello, A. Panunzi, T. Raso (eds), *In Search of Basic Units of Spoken Language*. John Benjamins Publishing Company, Amsterdam.

2019 Ravelli, A.A., Gregori, L., Varvara, R. (2019). Comparing Ref-Vectors and word embeddings in a verb semantic similarity task. In Mehwish Alam, Valerio Basile, Felice Dell'Orletta, Malvina Nissim & Nicole Novielli (eds), *Proceedings of the 3rd Workshop on NaturalLanguage for Artificial Intelligence co-locatedwith AIxIA 2019*.

2019 Gregori, L., Varvara, R., Ravelli, A.A. (2019). Action Type induction from multilingual lexical features. *Procesamiento Del Lenguaje Natural, 63*, 85-92.

2018 Cresti, E., Moneglia, M., Gregori, L., Panunzi, A. (2018). *The Language into Act Theory: A Pragmatic Approach to Speech in Real-Life*. In H. Koiso & P. Paggio (eds), Proceedings of LB-ILR2018 and MMC2018 Joint Workshop: Language and Body in Real Life Multimodal Corpora 2018: Multimodal Data in the Online World. ELRA, Paris.

2018 Moneglia, M., Panunzi, A., Gregori, L. (2018). *"Taking Events" in Hindi. A Case Study from the Annotation of Indian Languages in IMAGACT*. In G. Nath Jha et al. (eds), WILDRE4 - 4th Workshop on Indian Language Data: Resources and Evaluation. ELRA, Paris.

2018 Panunzi, A., Moneglia, M., Gregori, L. (2018). *Action Identification and Local Equivalence of Action Verbs: The Annotation Framework of the IMAGACT Ontology*. In J. Pustejovsky & I. van der Sluis (eds), Proceedings of the LREC 2018 Workshop "AREA - Annotation, Recognition and Evaluation of Actions". ELRA, Paris.

2018 Panunzi, A., Gregori, L., Ravelli, A.A. (2018). *One Event, Many Representations. Mapping Action Concepts through Visual Features.* In N. Calzolari et al (eds), Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). ELRA, Paris.

2018 Pan Y., Moneglia M., Panunzi A., Gregori L. (2018). *IMAGACT4ALL. Una ontologia per immagini dell'azione per l'apprendimento del lessico verbale di base delle lingue seconde*. In Proceedings of the XVII AITLA Conference, February 23-25, 2017, University of Naples "L'Orientale".

2018 Ravelli A.A., Gregori L., Panunzi A. (2018). *Collecting Information for Action Understanding: The Enrichment of the IMAGACT Ontology of Action*. In Proceedings of the Joint Ontology Workshops 2017 Episode 3: The Tyrolean Autumn of Ontology.

2017 Dovetto F.M., Gregori L., Panunzi A. (2017). *Sull'annotazione di un corpus orale mistilingue non standard (patologico schizofrenico)*. In A. De Meo, F.M Dovetto (eds), La comunicazione parlata / Spoken communication. Aracne Editrice, 345–361.

2017 Ravelli A.A., Gregori L., Feltracco A. (2017). *Evaluating a Rule Based Strategy to Map IMAGACT and T-PAS*. In Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-it 2017), Rome, Italy, December 11-13, 2017.

2017 Gregori L., Panunzi A. (2017). *Measuring the Italian-English lexical gap for action verbs and its impact on translation*. In Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications. Valencia (Spain), 102–109.

2016 Gregori L., Panunzi A., Ravelli A.A. (2016). *Linking IMAGACT ontology to BabelNet through action videos*. In A. Corazza, S. Montemagni, G. Semeraro (eds), Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016. 5-6 December 2016, Napoli. Accademia University Press, 162–167.

2016 Gregori L. (2016). *La forma dell'ontologia dell'azione IMAGACT: dal modello gerarchico al modello insiemistico in un DB a grafo*. PhD diss., University of Florence.

2015 Gregori L., Ravelli A.A., Panunzi A. (2015). *Linking dei contenuti multimediali tra ontologie multilingui: i verbi di azione tra IMAGACT e BabelNet*. In C. Bosco, S. Tonelli, F.M. Zanzotto (eds), Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it 2015. Accademia University Press, 150–154.

2014 Panunzi A. Cresti E., Gregori L. (2014). *RIDIRE. Corpus and tools for the acquisition of Italian L2*. In A. Abel, C. Vettori, N. Ralli (eds), Proceedings of the XVI EURALEX International Congress: The user in focus. EURAC research, 447–462.

2014 Panunzi A., De Felice I., Gregori L., Jacoviello S., Monachini M., Moneglia M., Quochi V. (2014). *Translating action verbs using a dictionary of images: the IMAGACT ontology*. In A. Abel, C. Vettori, N. Ralli (eds), Proceedings of the XVI EURALEX International Congress: The user in focus. EURAC research, 1163–1170.

2012 Gregori L., Panunzi A. (2012). *DB-IPIC: AN XML Database for Informational Patterning Analysis*. In H. Mello, M. Pettorino, T. Raso (eds), Proceedings of the VIIth GSCP International Conference: Speech and Corpora. Firenze University Press, 121–127.

2012 Moneglia M., Gagliardi G., Gregori L., Panunzi A., Paladini S., Williams A. (2012). *La variazione dei verbi generali nei corpora di parlato spontaneo. L'ontologia IMAGACT*. In H. Mello, M. Pettorino, T. Raso (eds), Proceedings of the VIIth GSCP International Conference: Speech and Corpora. Firenze University Press, 406–411.

2012 Panunzi A., Fabbri M., Moneglia M., Gregori L., Paladini S. (2012). *RIDIRE-CPI: an Open Source Crawling and Processing Infrastructure for Web Corpora Building*. In N. Calzolari et al (eds), Proceedings of the Eigth International Conference on Language Resources and Evaluation (LREC 2012). ELRA, Paris. 2274–2279.

2012 Panunzi A., Gregori L. (2012). *DB-IPIC. An XML Database for the Representation of Information Structure in Spoken Language*. In H. Mello, A. Panunzi, T. Raso (eds), Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation. Firenze University Press, Firenze. 133–150.

Research Projects

2017–2019 **IMAGACT-MED**, *FAS Salute 2014 - Regione Toscana*,
https://www.micc.unifi.it/projects/imagact-med/.
Role: Research Fellow

2013–2017 **MODELACT**, *Futuro in Ricerca 2012 programme (Project Code RBFR12C6O8)*,
http://modelact.lablita.it/.
Role: Research Fellow

2011–2014 **IMAGACT**, *PAR/FAS Regione Toscana, Linea di Azione 1.1.a.3*,
http://imagact.lablita.it/.
Role: Research Fellow