



Curriculum vitae

Informazioni personali

Giulia Venturi

 –

 –  +39 050-3152847

 giulia.venturi@ilc.cnr.it

 [linkedin.com/pub/giulia-venturi/32/379/604](https://www.linkedin.com/pub/giulia-venturi/32/379/604)

Sesso Femminile | **Data di nascita** 31/01/1982 | **Nazionalità** Italiana

Dal 2007 il mio oggetto di studio e di ricerca è la definizione di metodologie di estrazione di informazione linguistica e di dominio da corpora testuali rappresentativi di diverse varietà d'uso della lingua impiegando e specializzando quando necessario strumenti di Trattamento Automatico della Lingua (TAL). I principali prodotti delle mie ricerche sono documentate nelle pubblicazioni elencate nell'ultima sezione di questo curriculum.

Estrazione di informazione linguistica: metodologia di analisi che consiste nella definizione di una serie di regole che, applicate all'output di una catena di analisi linguistica, permettono di estrarre caratteristiche informative rispetto a una ampia gamma di fenomeni linguistici caratterizzanti una determinata collezione di testi [51]. Tale metodologia di ricostruzione del profilo linguistico di una collezione di testi ha rappresentato il punto di partenza per lo sviluppo di una serie di applicazioni software:

- primo strumento di analisi automatica della leggibilità per la lingua italiana, basato su tecnologie TAL, e il primo a livello internazionale in grado di valutare la complessità di una singola frase
- sistema di riconoscimento automatico della lingua materna
- sistema di classificazione automatica del genere testuale
- sistema di classificazione di testi caratterizzati da un livello di riflessione sulla attività didattica
- sistema in grado di predire il livello di sviluppo delle competenze di scrittura di apprendenti L1

La metodologia mi ha permesso di condurre studi di interesse linguistico sui seguenti aspetti:

- aspetti caratteristici dell'interlingua
- quali sono i fenomeni linguistici che permettono di valutare il livello di complessità di un intero documento o di una singola frase
- quale livello di descrizione linguistica è più informativo per la classificazione del genere testuale
- i fenomeni linguistici che mutano nell'evoluzione delle abilità di scrittura di apprendenti L1
- elementi di complessità caratterizzanti i consensi informati
- elementi di complessità caratterizzanti i testi giuridici e amministrativi

All'interno di questo filone di ricerche lo studio della complessità linguistica costituisce un aspetto centrale come punto di partenza per la definizione di strategie e risorse testuali da impiegare in un processo di semplificazione automatica del testo. Più recentemente, questa attività di ricerca sta evolvendo verso lo studio di quali sono gli elementi linguistici correlati con la percezione umana della difficoltà di comprensione di un testo

Estrazione di informazione di dominio: questo filone di studi mi ha portata alla definizione di metodi di accesso al contenuto semantico di collezioni di documenti che tenessero conto delle specificità della varietà d'uso in esame. Mi sono in particolare dedicata alle seguenti attività:

- definizione di schemi di annotazione semantica che permettano di catturare il modo in cui l'informazione è organizzata nei testi, in particolare in testi di dominio biomedico e giuridico
- definizione di strategie di estrazione di informazione semantica da testi di dominio, in particolare da testi giuridici
- definizione di metodi di estrazione terminologica da testi di dominio, in particolare giuridici

Trasversali ai due principali filoni di ricerca sono i due seguenti ambiti:

A. specializzazione di strumenti di TAL alle specificità d'uso della lingua, in particolare al linguaggio biomedico e alla lingua del diritto

B. contributo alla definizione di schemi di annotazione linguistica, per quanto riguarda in particolare la definizione di metodologie di identificazione di errori di annotazione in treebank e l'indagine di fenomeni linguistici non chiaramente codificati in schemi di annotazione morfosintattica.

5 febbraio 2017 - 4 febbraio 2020

Progetto nazionale

Titolo: CHROME - Cultural Heritage Resources Orienting Multimodal Experiences

Ente finanziatore: Ministero dell'Istruzione, dell'Università e della Ricerca, PRIN 2015

Importo totale del finanziamento: 410.000,00 Euro

Importo finanziato per Unità Operativa: 47.432,68 Euro

Nominativo coordinatore del progetto: Francesco Cutugno (Università degli Studi di Napoli "Federico II")

Riferimenti: finanziamento approvato con Decreto Direttoriale 7 novembre 2016 n. 2634 Ammissione contributi, <http://www.chrome.unina.it/>

Periodo di attività: dal 05/02/2017 al 04/02/2020

Partecipanti al progetto: Università degli Studi di Napoli "Federico II", Università degli Studi Roma Tre, Università degli Studi di Salerno, Istituto di Scienze Applicate e Sistemi Intelligenti "Eduardo Caianiello" (CNR), Istituto di Linguistica Computazionale "A. Zampolli" (ILC) del CNR di Pisa.

Finalità del progetto: L'obiettivo principale del progetto è l'elaborazione di una metodologia per raccogliere, rappresentare e analizzare dati multimodali relativi ai beni culturali e presentarli attraverso agenti artificiali il cui comportamento è ispirato da un'analisi accurata di guide esperte, curatori di musei e tour operator.

Ruolo svolto: il ruolo da me svolto all'interno del progetto ha permesso la definizione di metodologie per l'estrazione di informazione di dominio da testi di interesse per il progetto quali guide e descrizioni di edifici di interesse storico-artistico. Mi sono inoltre occupata di definire una metodologia volta a individuare fenomeni linguistici (lessicali e sintattici) correlati con aspetti di chiarezza espositiva e comprensibilità che dovranno caratterizzare le guide turistiche realizzate nel corso del progetto.

La mia partecipazione al progetto è attestata dalla lettera di assunzione come ricercatrice a tempo determinato, protocollo n. 0000319 del 10/05/2018, dalla lettera di incarico personale dipendente in data 4/06/2018 e dalla certificazione rilasciata dal responsabile della collaborazione, dott. Felice Dell'Orletta, in data 3/09/2018, n. protocollo 0000563.

30 maggio 2016 - 29 maggio 2019

Collaborazione nazionale di ricerca

Titolo: protocollo di intesa tra Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa (INDIRE) e Istituto di Linguistica Computazionale "A. Zampolli" (ILC-CNR)

Ente finanziatore: Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa (INDIRE)

Importo totale del finanziamento: 40.000,00 Euro

Importo finanziato per Unità Operativa: 40.000,00 Euro

Nominativo coordinatore del progetto: responsabile scientifico per l'Istituto di Linguistica Computazionale (ILC-CNR) dott. Felice Dell'Orletta, per l'Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa (INDIRE) dott.ssa Maria Teresa Sagri

Riferimenti: finanziamento approvato come da PROTOCOLLO ILC 0000539 DEL 20/06/2016

Periodo di attività: dal 30/05/2016 al 29/05/2019

Partecipanti al progetto: Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa (INDIRE) e Istituto di Linguistica Computazionale "A. Zampolli" (ILC) del CNR di Pisa

Finalità del progetto: L'obiettivo principale della collaborazione consiste nello sviluppo di applicazioni di gestione della conoscenza (Knowledge Management) nel settore documentale educativo basate su tecnologie di Trattamento Automatico della Lingua.

Ruolo svolto: il ruolo da me svolto all'interno del progetto ha permesso specializzare strumenti di annotazione linguistica automatica del testo ed estrazione, organizzazione ed indicizzazione della conoscenza alle specificità dei documenti contenuti nelle banche dati INDIRE (Piani di Miglioramento, report dei progetti Alternanza scuola-lavoro, portfolio professionale del docente e delle attività raccolte dalla piattaforma Neoassunti).

La mia partecipazione al progetto è attestata dalla certificazione rilasciata dal responsabile della collaborazione, dott. Felice Dell'Orletta, in data 3/09/2018, n. protocollo 0000563.

Progetto regionale

Titolo: SMART NEWS - Social Sensing for breaking news

Ente finanziatore: Regione Toscana - Bando FAR FAS 2014

Importo totale del finanziamento: 654.861,37 Euro

Importo finanziato per Unità Operativa: 94.340,50 Euro

Nominativo coordinatore del progetto: Sauro Salvadori (Hyperborea S.r.l.)

Riferimenti: finanziamento approvato con Decreto n.6649 del 10/12/2015 Regione Toscana, <http://www.smart-news.it/>

Periodo di attività: dal 14/03/2016 al 13/09/2018

Partecipanti al progetto: Hyperborea s.r.l., Web Application for the Future Internet (WAFI) Laboratory dell'Istituto di Informatica e Telematica (IIT) del CNR di Pisa, il gruppo Multimedia Information Retrieval (MIR) dell'Istituto di Scienze e Tecnologie dell'Informazione (ISTI) c, Istituto di Linguistica Computazionale "A. Zampolli" (ILC) del CNR di Pisa

Finalità del progetto: Il progetto mira a realizzare uno strumento per la gestione delle ultime notizie che aiuti i giornalisti nell'intero processo che consiste nell'individuazione delle ultime notizie, nella raccolta di informazioni in merito ad esse e nella scrittura di articoli.

Ruolo svolto: il ruolo da me svolto all'interno del progetto ha permesso di condurre un'analisi dettagliata delle peculiarità di un linguaggio settoriale quale quello usato nei social media che, in quanto rappresentativo di una scrittura *breve*, presenta caratteristiche diverse dalla prosa giornalistica su cui gli strumenti di annotazione linguistica automatica del testo e strumenti di estrazione e organizzazione della conoscenza sono solitamente sviluppati. I risultati di questo studio forniscono un ausilio indispensabile all'adattamento di tali strumenti allo scopo di migliorarne le prestazioni di analisi.

La mia partecipazione al progetto è attestata dall'atto di conferimento protocollo n. 0000827 del 31/10/2016 e dalla certificazione rilasciata dal responsabile della collaborazione, dott. Felice Dell'Orletta, in data 3/09/2018, n. protocollo 0000563.

Progetto regionale

Titolo: SCRIBE – Scritture brevi, semplificazione linguistica, inclusione sociale: modelli e applicazioni

Ente finanziatore: Ministero dell'Istruzione, dell'Università e della Ricerca, Progetto Prin (PRIN 2010FWM3B4 – Area 10)

Importo totale del finanziamento: 328.500,00 Euro

Importo finanziato per Unità Operativa: 55.482,00 Euro

Nominativo coordinatore del progetto: Trifone Pietro - Università degli Studi di Roma "Tor Vergata", Facoltà di Lettere e Filosofia, Dipartimento di Studi Filologici, Linguistici e Letterari

Riferimenti: inanziamento approvato con Decreto Direttoriale n. 719 del 23/10/2012

Periodo di attività: dal 01/02/2013 al 31/01/2016

Partecipanti al progetto: Università di Tor Vergata (coordinatore), Università "L'Orientale" di Napoli; Università Roma Tre; Università di Pisa; Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) del CNR di Pisa.

Finalità del progetto: L'obiettivo di SCRIBE è stato quello di proporre un modello esplicativo del fenomeno delle *Scritture Brevi* e di produrre un dizionario digitale per permettere la divulgazione delle forme e, con essa, favorire la diffusione della comprensione e della competenza relativa. Un secondo e altrettanto rilevante filone di ricerca nell'ambito del progetto è stato quindi dedicato alla semplificazione dei testi mediante scritture brevi, ampiamente intese, e al problema della loro leggibilità e comprensione nell'ottica del dialogo interno tra varietà di una stessa lingua (gerghi vs italiano standard, linguaggi giovanili vs italiano standard, tecnicismi e linguaggi specialistici vs italiano standard).

Ruolo svolto: Il ruolo da me svolto all'interno del progetto ha permesso di arrivare alla definizione di uno schema di annotazione di fenomeni linguistici indicatori di complessità del testo. Nell'ambito del progetto mi sono concentrata sulla definizione di un tagset di regole di semplificazione che coinvolgono i fenomeni individuati a diversi livelli di descrizione linguistica (lessicale, morfo-sintattica e sintattica). Tali regole includono ad esempio, la divisione del periodo in più periodi autonomi, regola che coinvolge fenomeni linguistici quali la coordinazione o la subordinazione; l'ordinamento di materiale linguistico, regola che coinvolge ad esempio casi di dislocazione e marcatezza sintattica; la cancellazione di materiale linguistico ridondante a livello informativo, regola che comporta ad esempio l'eliminazione di sequenze di modificatori, di frasi incidentali, ecc. Il tagset è stato applicato su due tipologie di testi rappresentativi di due diverse strategie di semplificazione: *i)* una collezione di storie per bambini con ridotte capacità di comprensione, semplificate da esperti in semplificazione del testo e *ii)* una raccolta di diverse tipologie di testi (manuali, romanzi, ecc.) utilizzati nella scuola come materiali educativi e semplificati da insegnanti per essere più comprensibili da parte di studenti l'italiano come L2.

La mia partecipazione al progetto è attestata dall'atto di conferimento dell'assegno, protocollo n. 0001035 del 21/11/2014 e dalla certificazione rilasciata dal direttore dell'Istituto di Linguistica Computazionale (ILC-CNR) dott.ssa Simonetta Montemagni in data 3/09/2018, n. protocollo 0000566.

1 aprile 2013 - 31 marzo 2014

Progetto premiale MIUR

Titolo: L'amministrazione della giustizia in Italia: il caso della neurogenetica e delle neuroscienze, un approccio interdisciplinare

Ente finanziatore: Ministero dell'Istruzione, dell'Università e della Ricerca

Importo totale del finanziamento: 900.000,00 Euro

Importo finanziato per Unità Operativa: 25.000,00 Euro

Nominativo coordinatore del progetto: Dott. Andrea Bozzi, Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) del CNR di Pisa

Riferimenti: finanziamento approvato con come da prot. n. 525, certificazione rilasciata dal referente del progetto in data 3 maggio 2013

Periodo di attività: dal 01/04/2013 al 31/03/2014

Partecipanti al progetto: Istituti CNR proponenti: Istituto di Studi Giuridici Internazionali; Istituto di Linguistica Computazionale "Antonio Zampolli"; Istituto di Ricerche sulla Popolazione e le Politiche Sociali; Istituto di Scienze e Tecnologie della Cognizione

Finalità del progetto: L'obiettivo del progetto era quello di condurre un'analisi multidisciplinare, rivolta alle dimensioni etiche, giuridiche e sociali dell'impiego della neurogenetica e delle neuroscienze nel sistema della giustizia in Italia, a partire dall'analisi di collezioni di sentenze recanti riferimenti neuroscientifici emesse dai tribunali italiani (sezioni civili e penali, in tutti i gradi di giudizio) e tenuto conto degli sviluppi delle conoscenze scientifiche e riferendosi al contesto europeo e internazionale relativamente alle applicazioni e alla loro regolazione.

Ruolo svolto: All'interno del progetto il mio ruolo è stato quello di definire una metodologia di analisi di corpora di sentenze emesse dai tribunali italiani (sezioni civili e penali, in tutti i gradi di giudizio), raccolte nell'ambito del progetto, basata sulla specializzazione di tecniche di estrazione di contenuti semantici rilevanti basate su strumenti di Trattamento Automatico del Linguaggio, di organizzazione strutturata della conoscenza estratta e di indicizzazione del testo finalizzata alla loro navigazione "intelligente".

La mia partecipazione al progetto è attestata dalla certificazione rilasciata dal direttore dell'Istituto di Linguistica Computazionale (ILC-CNR) dott.ssa Simonetta Montemagni in data 3/09/2018, n. protocollo 0000566.

1 novembre 2012 - 31 ottobre 2014

Progetto regionale

Titolo: JURNET - Legal Text Mining: costruzione di reti semantico-concettuali finalizzate a una navigazione intelligente di corpora di testi giuridici

Ente finanziatore: Regionale Toscana, POR CRO FSE 2007-2013 Asse IV – Capitale Umano

Importo totale del finanziamento: 30.000,00 Euro

Importo finanziato per Unità Operativa: 30.000,00 Euro

Nominativo coordinatore del progetto: Dott.ssa Simonetta Montemagni, Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) del CNR di Pisa

Riferimenti: Finanziamento approvato con Decreto N 2460 del 30 Maggio 2012

Periodo di attività: dal 01/11/2012 al 30/04/2014

Partecipanti al progetto: Istituto di Linguistica Computazionale "Antonio Zampolli" del CNR di Pisa; Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG-CNR); Scuola Superiore Sant'Anna di Pisa; European Center for Law, Science and New Technologies (ECLT), Università di Pavia

Finalità del progetto: L'obiettivo del progetto era quello di mettere a punto metodi e tecniche che, basati su strumenti di analisi automatica della lingua e di estrazione della conoscenza (Knowledge Extraction), consentissero di navigare in modo "intelligente" collezioni (corpora) di sentenze rendendo esplicita la rete che le lega, formata dai termini e dai concetti veicoli della conoscenza, nonché dai rinvii a sentenze precedenti e a testi legislativi..

Ruolo svolto: Il ruolo da me svolto all'interno del progetto ha permesso di raccogliere collezioni documentali rappresentative di diverse tipologie di sentenze e di domini diversi; dettagliata indagine di quegli aspetti di Trattamento Automatico della lingua del diritto che, relativi all'elaborazione di costrutti specifici usati nelle sentenze, possono essere responsabili di un calo nelle prestazioni delle tecniche di estrazione e organizzazione della conoscenza basati su strumenti di analisi linguistica automatica; di definire metodologie di estrazione e organizzazione della conoscenza volte a individuare all'interno di collezioni di sentenze elementi informativi rilevanti e rinvii giurisprudenziali e legislativi.

La mia partecipazione al progetto è attestata dall'atto di conferimento dell'assegno del 25/10/2012 con n. protocollo 0000955 e dalla certificazione rilasciata dal direttore dell'Istituto di Linguistica Computazionale (ILC-CNR) dott.ssa Simonetta Montemagni in data 3/09/2018, n. protocollo 0000566.

Progetto nazionale CNR

Titolo: Migrazioni - Determinanti cognitive, linguistiche, formative e tecnologiche dei processi di integrazione

Ente finanziatore: Dipartimento del Consiglio Nazionale delle Ricerche "Identità Culturale",

Importo totale del finanziamento: dato disponibile presso la sede centrale CNR

Importo finanziato per Unità Operativa: 30.000,00 Euro

Nominativo coordinatore del progetto: Prof. Tullio Gregory, Università di Roma "La Sapienza"

Riferimenti: IC.P10, http://www.cnr.it/progetti/Progetto.html?id_prog=7841

Periodo di attività: dal 10/2009 al 09/2012

Partecipanti al progetto: ISTC - Istituto di Scienze e Tecnologie della Cognizione del CNR; ITD - Istituto per le Tecnologie Didattiche del CNR (Genova); Istituto di Linguistica Computazionale "Antonio Zampolli" del CNR (Pisa)

Finalità del progetto: L'obiettivo del progetto ideato dal giurista Andrea Di Porto, Direttore del Dipartimento Identità Culturale del CNR negli anni 2006-08, è quello di costituire un'équipe di ricercatori e tecnici afferenti a 13 Istituti del CNR e all'Osservatorio sulla codificazione e sulla formazione del giurista in Cina, considerando i tradizionali settori di indagine degli Istituti CNR e le competenze scientifiche in essi presenti per monitorare le tematiche e i fenomeni migratori, fornendo spunti di indagine e riflessioni interdisciplinari.

Ruolo svolto: All'interno della Linea di ricerca, n.1.4 "Tecnologie avanzate e Integrazione linguistico-culturale nella scuola" del progetto il ruolo da me svolto ha permesso la definizione di una metodologia per il monitoraggio dell'evoluzione del processo di apprendimento linguistico di studenti attraverso la ricostruzione delle competenze linguistiche espresse in corpora di produzioni scritte in ambito scolastico ed estratte in modo automatico grazie all'ausilio di strumenti di Trattamento Automatico della Lingua.

La mia partecipazione al progetto è attestata dall'atto di conferimento assegno atto di conferimento protocollo n. 0000347 del 31/03/2010 e dalla certificazione rilasciata dal direttore dell'Istituto di Linguistica Computazionale (ILC-CNR) dott.ssa Simonetta Montemagni in data 3/09/2018, n. protocollo 0000566.

2008 - 2009

Collaborazione nazionale di ricerca

Titolo: accordo di collaborazione tra l'Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG-CNR) di Firenze e l'Istituto di Linguistica Computazionale "A. Zampolli" (ILC-CNR) di Pisa

Ente finanziatore: Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG-CNR) di Firenze

Importo totale del finanziamento: 10.000,00 Euro

Importo finanziato per Unità Operativa: 10.000,00 Euro

Nominativo coordinatore del progetto: responsabile scientifico per l'Istituto di Linguistica Computazionale (ILC-CNR) dott.ssa Simonetta Montemagni, per l'Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG-CNR) dott. Pierluigi Spinosa

Riferimenti: n. protocollo: 0000207 del 28/07/2008

Periodo di attività: dal 2008 al 2009

Partecipanti al progetto: Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG-CNR) e Istituto di Linguistica Computazionale "A. Zampolli" (ILC-CNR)

Finalità del progetto: progettazione e sviluppo di funzionalità avanzate di annotazione semantica del testo legislativo integrate nel NIR Editor nell'ambito del progetto nazionale "Norme in Rete (NIR)".

Ruolo svolto: Il ruolo da me svolto all'interno del progetto ha permesso di specializzare un modulo di Trattamento Automatico della Lingua (IDEAL), implementato come un compilatore di grammatiche a stati finiti definite su sequenze di gruppi sintattici (chunk) (es. gruppi nominali, verbali). La specializzazione è consistita nell'estensione della grammatica per l'individuazione delle seguenti tipologie di disposizioni legislative di modifica: abrogazione, sostituzione e integrazione a partire dalla loro realizzazione lessicale, morfo-sintattica e sintattica nel testo. L'obiettivo era quello di fornire un supporto all'attività di consolidamento editoriale del testo vigente.

La mia partecipazione al progetto è attestata dalla certificazione rilasciata dal direttore dell'Istituto di Linguistica Computazionale (ILC-CNR) dott.ssa Simonetta Montemagni in data 3/09/2018, n. protocollo 0000566.

Progetto europeo

Titolo: BOOTSTREP - Bootstrapping of Ontologies and Terminologies Strategic Research Project

Ente finanziatore: Commissione europea

Importo totale del finanziamento: 3.600.000,00 Euro

Importo finanziato per Unità Operativa: 440.000,00 Euro

Nominativo coordinatore del progetto: Udo Hahn, Friedrich Schiller, Universitaet Jena, Germania

Riferimenti: FP6-IST-2004 Grant Agreement No. 028099 - Data: 21/11/2005, https://cordis.europa.eu/project/rcn/79393_en.html

Periodo di attività: dal 01/04/2006 al 31/03/2009

Partecipanti al progetto: partecipanti accademici e industriali da 4 paesi europei (Germania, UK, Francia e Italia) e un paese asiatico (Singapore); partecipanti coinvolti nel Workpackage 4 'Bio-Events': Istituto di Linguistica Computazionale "Antonio Zampolli" del CNR di Pisa (coordinatore, Italia); School of Computer Science, The University of Manchester (UK); European Bioinformatics Institute, Cambridge (UK) e Istituto di Linguistica Computazionale "A. Zampolli" (ILC-CNR)

Finalità del progetto: Lo scopo del progetto è stato quello di riunire database di dati biologici già esistenti in aggiunta ad archivi terminologici ed implementare un sistema di analisi testuale che aumenti continuamente la loro copertura analizzando documenti e pubblicazioni medico biologiche. La progettata integrazione della conoscenza biologica in una struttura concettuale omogenea faciliterà l'accesso alla conoscenza altrimenti frammentata e ne aumenterà sostanzialmente l'utilizzo per scopi di Ricerca e Sviluppo, ad esempio nell'industria bio-tecnologica e farmaceutica europea.

Ruolo svolto: Il ruolo da me svolto all'interno del progetto ha permesso di arrivare alla definizione di uno schema di annotazione adeguato alla rappresentazione del contenuto proposizionale di eventi bio-medici presenti nel corpus di abstract di articoli scientifici in lingua inglese raccolti su MEDLINE. Tale schema ha permesso di rendere esplicita l'informazione relativa alla struttura predicato-argomento di ogni elemento linguistico (per lo più verbo/deverbale) considerato di interesse da esperti del dominio. Il mio ruolo è stato inoltre quello di definire uno schema di annotazione della modalità con cui viene espresso l'evento.

La mia partecipazione al progetto è attestata dalla lettera di conferimento borsa, protocollo n. 0000146 del 02/04/2007 e dalla certificazione rilasciata dal direttore dell'Istituto di Linguistica Computazionale (ILC-CNR) dott.ssa Simonetta Montemagni in data 3/09/2018, n. protocollo 0000566.

Progetto europeo**Titolo:** DALOS - Drafting Legislation with Ontology-based Support**Ente finanziatore:** Unione Europea, eParticipation project**Importo totale del finanziamento:** 387.568,00 Euro**Importo finanziato per Unità Operativa:** 41.600,00 Euro**Nominativo coordinatore del progetto:** Dott.ssa Daniela Tiscornia, Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG-CNR) di Firenze**Riferimenti:** project n. 2006/01/024, certificazione rilasciata dal coordinatore del progetto in data 14 settembre 2009, <http://ep-momentum.epu.ntua.gr/eParticipationProjects/DALOS/tabid/61/Default.aspx>**Periodo di attività:** dal 01/04/2007 al 30/06/2008**Partecipanti al progetto:** Istituto di Linguistica Computazionale "Antonio Zampolli" del CNR di Pisa; Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG-CNR) (coordinatore, Italia); Department of Computer Science University of Sheffield - USFD (UK); Autonomous University of Barcelona - IDT-UAB (Spagna); Leiden University (Paesi Bassi); European University Institute (Unione Europea); The Leibniz Center for Law University of Amsterdam UvA (Paesi Bassi); Camera dei Deputati Italian Parliament (Italia)**Finalità del progetto:** L'obiettivo del progetto era quello di fornire al legislatore europeo una serie di risorse lessicali e ontologiche che costituissero un supporto all'uso adeguato di terminologia giuridica in fase di drafting legislativo multilingue. L'uso di metodi di estrazione automatica della conoscenza e ontology learning aveva l'esplicito intento di facilitare il compito del legislatore europeo che deve far fronte alla diversa realizzazione linguistica di uno stesso concetto. Le metodologie sviluppate nell'ambito del progetto sono state applicate ad un caso di studio specifico, quello della legislazione europea in materia di protezione del consumatore per quattro diverse lingue (italiano, inglese, spagnolo e olandese).**Ruolo svolto:** All'interno del progetto il mio ruolo è stato quello di specializzare gli strumenti di estrazione automatica di conoscenza da testi di dominio e ontology learning basati su strumenti di Trattamento Automatico del Linguaggio alle necessità del dominio giuridico. In particolare, mi sono occupata della specializzazione della mini-grammatica implementata negli strumenti per l'identificazione e l'acquisizione di unità mono e polirematiche, e di varianti terminologiche, impiegate nella costruzione di un lessico multilingue in materia di diritto di protezione del consumatore.

La mia partecipazione al progetto è attestata dalla lettera di conferimento borsa, protocollo n. 0000146 del 02/04/2007 e dalla certificazione rilasciata dal direttore dell'Istituto di Linguistica Computazionale (ILC-CNR) dott.ssa Simonetta Montemagni in data 3/09/2018, n. protocollo 0000566.

**INCARICHI CONFERITI PER
ATTIVITÀ DI FORMAZIONE
PROFESSIONALE****Struttura:** Azienda Ospedaliera Universitaria Pisana, Pisa**Periodo di attività:** 13 ottobre 2017 - 27 novembre 2017**Durata complessiva:** 6 ore complessive**Descrizione:** incarico di formazione rivolta a dipendenti di azienda ospedaliero-universitaria finalizzato all'uso di tecnologie del linguaggio per la valutazione della complessità linguistica di testi giuridici, ricevuto nell'ambito del corso di formazione *Cultura digitale: linguaggio naturale e linguaggio giuridico per i siti web*.**Riferimenti:** lettera di conferimento incarico di docenza conferita dall'Azienda Ospedaliera Universitaria Pisana, n. protocollo 45598, il 24/08/2017**Struttura:** agenzia formativa dell'Unione dei Comuni della Versilia, Querceta (Lucca)**Periodo di attività:** 3 gennaio 2017 - 26 gennaio 2017**Durata complessiva:** 4 ore complessive**Descrizione:** incarico di formazione rivolta a dipendenti di enti locali finalizzato all'uso di tecnologie del linguaggio per la valutazione della complessità linguistica di atti normativi e amministrativi, ricevuto nell'ambito del corso di formazione *La qualità degli atti normativi e amministrativi*.**Riferimenti:** contratto per prestazione professionale occasionale conferito dall'agenzia formativa dell'Unione dei Comuni della Versilia in data 19/01/2016

Struttura: Azienda Ospedaliera Universitaria Pisana, Pisa

Periodo di attività: 28 ottobre 2016 - 1 dicembre 2016

Durata complessiva: 6 ore complessive

Descrizione: incarico di formazione rivolta a dipendenti di azienda ospedaliero-universitaria finalizzato all'uso di tecnologie del linguaggio per la valutazione della complessità linguistica della comunicazione sanitaria, ricevuto nell'ambito del corso di formazione *L'utente, l'usabilità e la qualità percepita dei contenuti WEB della sanità: teoria e strumenti operativi*.

Riferimenti: lettera di conferimento incarico di docenza conferita via email dal dott. Paolo Garzella, Ufficio Comunicazione, P.O. tec. Comunicazione Pubblica Utilità, U.O. Relazioni con il Pubblico dell'Azienda Ospedaliera Universitaria Pisana

Struttura: Università di Pisa, Dipartimento di Filologia, Letteratura, Linguistica

Tipologia di corso: insegnamento universitario

Periodo di attività: Dall'anno accademico 2018-2019 - oggi

Durata complessiva: 12 ore complessive

Materia di insegnamento: Linguistica Computazionale II, Corso di Laurea Magistrale in Informatica Umanistica dell'Università di Pisa.

Struttura: Università di Pavia, Facoltà di Giurisprudenza

Tipologia di corso: master universitario di II livello

Periodo di attività: Dal 2017 - oggi

Durata complessiva: 9 ore complessive

Materia di insegnamento: strumenti di Trattamento Automatico della Lingua per l'estrazione di informazione di dominio da collezioni di documenti giuridici e per la misura della complessità della lingua del diritto, nell'ambito del master universitario *La lingua del diritto: comprensione, elaborazione e applicazioni professionali* dell'Università di Pavia con il Senato della Repubblica e in collaborazione con il CNR.

Riferimenti: n. protocollo della lettera di affidamento docenza 16560, titolo III, classe 5, fascicolo 13-1/2017, anche per l'anno 2018

Struttura: Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD)

Tipologia di corso: scuola estiva

Periodo di attività: 12 giugno 2017

Durata complessiva: 2 ore complessive

Materia di insegnamento: introduzione all'impiego di strumenti di Trattamento Automatico della Lingua per la valutazione della complessità linguistica a sostegno dell'attività didattica, nell'ambito della Scuola Estiva *Strumenti digitali per umanisti* organizzata dall'AIUCD in collaborazione con il Laboratorio di Cultura Digitale (LABCD) del Corso di Laurea in Informatica Umanistica, Università di Pisa.

Riferimenti: <http://dhtools.labcd.unipi.it/programma/>

Struttura: Università di Pisa, Dipartimento di Scienze Politiche

Tipologia di corso: insegnamento universitario

Periodo di attività: 26 ottobre 2016

Durata complessiva: 2 ore complessive

Materia di insegnamento: introduzione al ruolo delle tecnologie linguistico-computazionali per la leggibilità della comunicazione istituzionale, nell'ambito dell'insegnamento di *Tecnica di Redazione dei Regolamenti e degli Atti Amministrativi* (Cod. 080NN), Laurea Magistrale in Scienze Delle Pubbliche Amministrazioni.

Riferimenti: n. protocollo della lettera di incarico 0003311/2016, in data 18/10/2016

Struttura: Università "Pablo de Olavide" di Siviglia (Spagna), Facoltà di Giurisprudenza

Tipologia di corso: insegnamento universitario

Periodo di attività: 14 maggio 2015

Durata complessiva: 2 ore complessive

Materia di insegnamento: strategie e applicazioni di costruzione di corpora di documenti giuridici.

Riferimenti: lettera di incarico del titolare del corso Prof. Antonio Lazari, Departamento de Derecho Privado, Università "Pablo de Olavide", Siviglia, in data 16/03/2015

Struttura: Università di Cadice (Spagna), Departamento de Filología Francesa e Inglesa
Tipologia di corso: insegnamento universitario
Periodo di attività: 13 maggio 2015
Durata complessiva: 2 ore complessive
Materia di insegnamento: introduzione alle strategie e applicazioni dell'uso di strumenti di Trattamento Automatico della Lingua per la ricostruzione del profilo linguistico di un testo.
Riferimenti: lettera di incarico della titolare del corso Angeles Zarco Tejada, Prof. Titular Universidad, Dpto. Filología Francesa e Inglesa, Universidad de Cátrategie e applicazioni dell'uso di strumenti di Trattamento Automatico del Linguaggio per la ricostruzione del profilo linguistico di un testodiz, in data 29/01/2015

Struttura: Università di Cadice (Spagna), Departamento de Filología Francesa e Inglesa
Tipologia di corso: insegnamento universitario
Periodo di attività: 30 aprile 2014
Durata complessiva: 2 ore complessive
Materia di insegnamento: introduzione all'uso di strumenti di Trattamento Automatico della Lingua per l'estrazione e organizzazione di conoscenza da corpora documentali.
Riferimenti: lettera di incarico della titolare del corso Angeles Zarco Tejada, Prof. Titular Universidad, Dpto. Filología Francesa e Inglesa, Universidad de Cátrategie e applicazioni dell'uso di strumenti di Trattamento Automatico del Linguaggio per la ricostruzione del profilo linguistico di un testodiz, in data 13/01/2014

Struttura: Università di Pisa, Dipartimento di Filologia, Letteratura, Linguistica
Tipologia di corso: supporto alla didattica universitaria
Periodo di attività: Anno accademico 2013-2014 - Anno accademico 2017-2018
Durata complessiva: 10 ore all'anno
Materia di insegnamento: Linguistica Computazionale II, Corso di Laurea Magistrale in Informatica Umanistica dell'Università di Pisa.
Riferimenti: come certificato dalla titolare del corso dott.ssa Simonetta Montemagni in data 29/03/2018, disponibile al link: <https://drive.google.com/open?id=1qnnGfJzfXzMf6Pw5qXZQwtNy2pX3KJXm>

Struttura: Università di Bologna, Facoltà di Giurisprudenza, sede di Ravenna
Tipologia di corso: scuola estiva
Periodo di attività: Dal 2013 - oggi
Durata complessiva: 1 ora all'anno
Materia di insegnamento: strumenti di Trattamento Automatico della Lingua per l'estrazione di informazione di dominio da collezioni di documenti giuridici, nell'ambito del modulo avanzato dell'*International Summer School: Managing lexical resources in the semantic web (LEX)*.
Riferimenti: <http://summerschoollex.cirfid.unibo.it/>

Struttura: Scuola Superiore Sant'Anna di Pisa, Classe accademica di Scienze Sociali, Settore di Scienza Giuridiche, Istituto di Diritto, Politica, Sviluppo
Tipologia di corso: insegnamento universitario
Periodo di attività: 17 maggio 2012
Durata complessiva: 2 ore complessive
Materia di insegnamento: introduzione ai metodi, strumenti e applicazioni per il Trattamento Automatico della Lingua del diritto, nell'ambito del corso *Introduzione al diritto privato comparato ed europeo* del Prof. Antonio Lazari.
Riferimenti: <https://drive.google.com/open?id=1mLsxFcrAScnKG1fluZ1VT20JF34y6hz>

Struttura: Scuola Superiore Sant'Anna di Pisa, Classe accademica di Scienze Sociali, Settore di Scienza Giuridiche, Istituto di Diritto, Politica, Sviluppo
Tipologia di corso: insegnamento universitario
Periodo di attività: 6-7 dicembre 2012
Durata complessiva: 4 ore complessive
Materia di insegnamento: introduzione alla linguistica computazionale come ponte tra la lingua del diritto e la conoscenza giuridica, nell'ambito del corso *Introduzione al diritto privato comparato ed europeo* del Prof. Antonio Lazari.
Riferimenti: <https://www.santannapisa.it/it/event/la-linguistica-computazionale-come-ponte-tra-la-lingua-del-diritto-e-conoscenza-giuridica-ii> Conferimento incarico in data 23/11/2012, n. protocollo 0010734 (P)

Ruolo svolto: membro della commissione valutatrice della tesi di dottorato di Itziar Gonzalez Dios dal titolo "Readability Assessment and Automatic Text Simplification. The Analysis of Basque Complex Structures", presso Università dei Paesi Baschi

Periodo di attività: 23 giugno 2016

Riferimenti: Prot. n. 201600004647 del 6 giugno 2016.

SUPERVISIONE DI TESI DI
LAUREA

Titolo tesi: "La leggibilità dei consensi informati: un approccio linguistico-computazionale";
candidata: Sabrina Rinnone

Struttura: Università di Pisa, Dipartimento di Filologia, Letteratura, Linguistica, Corso di laurea magistrale in Informatica Umanistica

Periodo di attività: luglio 2016 - febbraio 2017

Riferimenti: <https://etd.adm.unipi.it/t/etd-01202017-181656/>

Titolo tesi: "Studio linguistico-computazionale per l'analisi dei tipi linguistici. Similarità e differenze nel confronto fra Universal Dependencies Treebanks"; candidata: Chiara Alzetta

Struttura: Università di Pisa, Dipartimento di Filologia, Letteratura, Linguistica, Corso di laurea magistrale in Informatica Umanistica

Periodo di attività: gennaio-maggio 2017

Riferimenti: <https://etd.adm.unipi.it/t/etd-04122017-120113/>

ORGANIZZAZIONE DI INIZIATIVE
SCIENTIFICHE

Titolo dell'evento: coordinamento dello 'Student Scholarship and Volunteer Program' nell'ambito della *57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*

Luogo e data di svolgimento: Firenze, 28 luglio-2 agosto 2019

Riferimenti: <http://www.acl2019.org/EN/student-scholarship-applications-volunteers.xhtml>

Titolo dell'evento: Workshop "Computational Linguistics for Linguistic Complexity (CL4LC)", organizzato nell'ambito della *26th International Conference on Computational Linguistics (COLING 2016)*

Luogo e data di svolgimento: Osaka (Giappone), 11 dicembre 2016

Riferimenti: <https://sites.google.com/site/cl4lc2016/home>

Titolo dell'evento: Workshop "Semantic Processing of Legal Texts (SPLeT-2014)", organizzato nell'ambito della *9th Language Resources and Evaluation Conference (LREC 2014)*

Luogo e data di svolgimento: Reykjavik (Islanda), 27 maggio 2014

Riferimenti: <https://sites.google.com/site/splet2014workshop/contact>

Titolo dell'evento: "First Shared Task on Dependency Parsing of Legal Texts", organizzato nell'ambito della *4th Workshop on Semantic Processing of Legal Texts (SPLeT-2012)*

Luogo e data di svolgimento: Istanbul (Turchia), 27 maggio 2012

Riferimenti: <https://sites.google.com/site/splet2012workshop/home>

Titolo dell'evento: "Domain Adaptation for Dependency Parsing Shared Task", organizzato nell'ambito della *Evalita 2011 (Evaluation of NLP and Speech Tools for Italian)*

Luogo e data di svolgimento: Roma, 24 gennaio 2012

Riferimenti: http://www.evalita.it/2011/tasks/dependency_parsing

Titolo dell'evento: "Frame Labeling over Italian Texts Task", organizzato nell'ambito della *Evalita 2011 (Evaluation of NLP and Speech Tools for Italian)*

Luogo e data di svolgimento: Roma, 24 gennaio 2012

Riferimenti: <http://sag.art.uniroma2.it/flait/>

Titolo della rivista: "Numéro spécial Traitement automatique de la langue juridique", vol. 58, n. 2, della rivista *Traitement Automatique des Langues*, Adeline Nazarenko et Adam Wyner (eds.), edita dall' *Association pour le Traitement Automatique des Langues (ATALA)*

Data di pubblicazione: 2017

Riferimenti: <http://www.atala.org/sites/default/files/1-%20TAL-58-2-legal%20NLP-introduction.pdf>

Titolo dell'evento: 3rd IEEE Conference on "Machine Learning and Natural Language Processing: Models, Systems, Data and Applications"

Luogo e data di svolgimento: Marrakech (Morocco), 21-24 ottobre 2018

Riferimenti: <http://www.ieee.ma/cist18/the-conference/committees/program-committee>

Titolo dell'evento: "International Conference on Dependency Linguistics (Depling)"

Luogo e data di svolgimento: Pisa, 18-20 settembre 2017

Riferimenti: <http://depling.org/depling2017/committee.html>

Titolo dell'evento: Second Workshop on "Automated Detection, Extraction and Analysis of Semantic Information in Legal Texts (ASAIL)", nell'ambito dell' *XVI International Conference on AI and Law (ICAIL 2017)*

Luogo e data di svolgimento: Londra, 16 giugno 2017

Riferimenti: <http://www.iaail.org/?q=article/asail-2017-2nd-workshop-automated-semantic-analysis-information-legal-texts>

Titolo dell'evento: "Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)"

Luogo e data di svolgimento: Copenhagen (Danimarca), 7-11 settembre 2017

Riferimenti: lettera di invito da parte degli organizzatori della conferenza, Sebastian Riedel and Rebecca Hwa, in data 24/02/2017

Titolo dell'evento: "Italian Conference on Computational Linguistics (CLiC-it)"

Luogo e data di svolgimento: Dal 2016 - oggi

Titolo dell'evento: "Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)"

Luogo e data di svolgimento: Austin (Texas, USA), 1-5 novembre 2016

Riferimenti: lettera di invito da parte degli organizzatori della conferenza Xavier Carreras and Kevin Duh, in data 7/03/2016

Titolo dell'evento: Workshop on "Innovative Use of NLP for Building Educational Applications (BEA)",

Luogo e data di svolgimento: Dal 2014 - oggi

Titolo dell'evento: "International Language Resources and Evaluation Conference (LREC)"

Luogo e data di svolgimento: Dal 2014 - oggi

Titolo dell'evento: "Special Track Human Language Technology", organizzato nell'ambito dell' *8th International Web Rule Symposium (RuleML 2014)*

Luogo e data di svolgimento: Praga (Repubblica Ceca), 18-22 agosto 2014

Riferimenti: <http://2014.ruleml.org/rules-and-human-language-technology>

Titolo dell'evento: "Translating between Human Language and Formal Rules: Business, Law, and Government", organizzato nell'ambito del *7th International Web Rule Symposium (RuleML 2013)*

Luogo e data di svolgimento: Seattle (Washington, USA), 11-13 luglio 2013

Riferimenti: http://wiki.ruleml.org/index.php/Human_Language_Technology

Titolo dell'evento: 4th Workshop on "Semantic Processing of Legal Texts (SPLeT-2012)", nell'ambito della *International Language Resources and Evaluation Conference (LREC 2012)*

Luogo e data di svolgimento: Istanbul (Turchia), 27 maggio 2012

Riferimenti: <https://sites.google.com/site/splet2012workshop/program-committee-1>

Nome del laboratorio: "Italian Natural Language Processing Laboratory (ItaliaNLP Lab)" dell'Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR) di Pisa

Data: dal 2013

Riferimenti: <http://www.italianlp.it/people/giulia-venturi/>

Nome del laboratorio: "Laboratorio di Linguistica Giudiziaria (LALIGI)" del Dipartimento di Linguistica dell'Università di Firenze

Data: dal 2013

Riferimenti: <http://www.patriziabellucci.it/laligi/partecipanti.htm>

PREMI E RICONOSCIMENTI

Distinguished Young Paper mention per l'articolo *Tecnologie del linguaggio e monitoraggio dell'evoluzione delle abilità di scrittura nella scuola secondaria di primo grado*, pubblicato negli atti della *First Italian Conference on Computational Linguistics*, 9-10 dicembre 2014, Università di Pisa.

Culture della materia per il corso di "Linguistica Computazionale II", settore disciplinare L/LIN01, delibera n. 39 del Consiglio di Dipartimento, Università di Pisa, seduta del 26 marzo 2014.

Best Verifiability, Reproducibility, and Working Description award per il software descritto nell'articolo *Linguistically-driven Selection of Correct Arcs for Dependency Parsing*, assegnato dal Software Reviewing Committee della *14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013)*, 24-30 marzo 2013, University of the Aegean, Samos, Greece.

PARTECIPAZIONE SU INVITO A EVENTI SCIENTIFICI

Titolo dell'intervento: "Why is this sentence complex? Cherry-pick the optimal set of features!"

Evento: *Interactive Workshop on Measuring Language Complexity*

Luogo e data di svolgimento: Friburgo (Germania), 12 settembre 2019

Riferimenti: http://www.christianbentz.de/MLC2019_index.html

Titolo dell'intervento: "Natural Language Processing methods to support legal communication"

Evento: workshop *Legal communication in a multilingual and multicultural context*

Luogo e data di svolgimento: Bergische Universität Wuppertal (Germania), 6 marzo 2019

Riferimenti: <https://www.sprachforschung.uni-wuppertal.de/de/workshop.html>

Titolo dell'intervento: "Natural Language Processing Methods and Instruments for the Extraction of Information from Domain Corpora"

Evento: conferenza *DR2 Distant Reading and Data-Driven Research in the History of Philosophy*

Luogo e data di svolgimento: Università di Torino, 14 febbraio 2019

Riferimenti: <https://dr2blog.hcommons.org/category/dr2/>

Titolo dell'intervento: "Italian legal corpora seen through the lenses of linguists and jurists"

Evento: conferenza *The fabric of law. Discovering patterns through Legal Corpus Linguistics*, organizzato dalla Heidelberger Akademie Der Wissenschaften

Luogo e data di svolgimento: Heidelberger (Germania), 19 marzo 2016

Riferimenti: <https://www.cal2.eu/index.php/conference-programme-fabric-of-law-and-language>

Titolo dell'intervento: "Il ruolo delle tecnologie linguistico-computazionali nella misurazione della complessità linguistica di testi giuridici: principi metodologici e prime applicazioni"

Evento: seminario nazionale *Giornata di studio e formazione sulla qualità degli atti normativi e amministrativi*, organizzata dal Dipartimento di Scienze Politiche dell'Università di Pisa, in collaborazione con l'Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG) del CNR di Firenze

Luogo e data di svolgimento: Pisa, 22 ottobre 2016

Riferimenti: <https://www.unipi.it/index.php/unipieventi/event/2130-la-qualita-degli-atti-normativi-e-amministrativi>

Titolo dell'intervento: "Improving Medical Communication with Language Technologies"

Evento: conferenza internazionale *Communication QuickStart* organizzata dal Association Against Melanoma

Luogo e data di svolgimento: Pisa, 11 giugno 2016

Riferimenti: <http://www.melanomapatientnetwork.eu/communication-quickstart-2016.html>

Titolo dell'intervento: "Tecnologie del linguaggio per la valutazione della comunicazione sanitaria"

Evento: giornata di studio *Sanità. La comunicazione, le relazioni e la cultura digitale*, organizzato dall'Azienda Ospedaliera Universitaria Pisana e patrocinato dall'Associazione italiana della comunicazione pubblica e istituzionale

Luogo e data di svolgimento: Pisa, 20 maggio 2016

Riferimenti: <https://www.ienonline.org/ien/convegni-e-percorsi-sulla-comunicazione-sanitaria/>

Titolo dell'intervento: "Burocratese e comunicazione pubblica: un'analisi linguistico-computazionale della leggibilità"

Evento: seminario nazionale *Reperire, capire e comunicare l'informazione giuridica*, organizzato dall'Istituto di teoria e tecniche dell'informazione giuridica (ITTIG) del CNR, dall'Associazione per la qualità degli atti amministrativi (AQuAA) e dall'Ordine dei giornalisti della Toscana

Luogo e data di svolgimento: Firenze, 14 novembre 2014 <http://www.ittig.cnr.it/reperire-capire-e-comunicare-linformazione-giuridica/>

Riferimenti: <http://www.ittig.cnr.it/reperire-capire-e-comunicare-linformazione-giuridica/>

Titolo dell'intervento: "Oltre il contenuto: tecnologie linguistico-computazionali per l'analisi della struttura linguistica del testo. Cosa, come perché"

Evento: seminario nazionale dal titolo *Seminari di cultura digitale*, organizzato dal Dipartimento di Informatica, Università di Pisa

Luogo e data di svolgimento: Pisa, 11 dicembre 2013

Riferimenti: <http://www.labcd.unipi.it/seminari/brunato-dellorletta-venturi-oltre-il-contenuto-tecnologie-linguistico-computazionali-per-lanalisi-della-struttura-linguistica-del-testo/>

Titolo dell'intervento: "Trattamento Automatico del Linguaggio e Diritto: metodi, strumenti e applicazioni"

Evento: seminario nazionale *Tecnologie del linguaggio e Diritto. Verso un accesso semantico a testi giuridici*, organizzato dall'Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG-CNR), Firenze.

Luogo e data di svolgimento: Firenze, 18 maggio 2018

Riferimenti: <https://www.cnr.it/it/evento/12959/tecnologie-del-linguaggio-e-diritto-verso-un-accesso-semantico-a-testi-giuridici>

Titolo dell'intervento: "Lingua e diritto: una prospettiva linguistico-computazionale"

Evento: seminario nazionale *Informazione giuridica e tecnologie linguistiche digitali*, Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG-CNR), Firenze

Luogo e data di svolgimento: Firenze, 2 dicembre 2010

Riferimenti: <http://www.ittig.cnr.it/EditoriaServizi/EventiConvegni/seminario2dicembre2010.pdf>

ATTIVITÀ DI REFERAGGIO PER RIVISTE

Titolo rivista: "Rivista Umanistica Digitale"

Periodo di attività: 6 giugno 2018

Riferimenti: lettera del journal editor Marilena Daquino in data 6/06/2018

Titolo rivista: "Languages"

Periodo di attività: dal 7 marzo 2018 al 18 giugno 2018

Riferimenti: lettera del journal editor Michele Cardani in data 7 marzo 2018

Titolo rivista: "International Journal of Artificial Intelligence in Education"

Periodo di attività: 19 luglio 2016 al 29 luglio 2016

Riferimenti: lettera del journal editor Becky Passonneau in data 19 luglio 2016

Titolo rivista: "Language Resources and Evaluation"

Periodo di attività: dal 28 settembre 2015, in corso

Riferimenti: lettera del journal editor disponibile a questo link: https://drive.google.com/open?id=1ETiyfO0oZ_jdc4OQBBPLrdQ_y3rLdhWn

Titolo rivista: "Artificial Intelligence and Law"

Periodo di attività: dal 24 agosto 2012 al 22 agosto 2013

Riferimenti: lettera del journal editor Kevin Ashley in data 24 agosto 2012

IDONEITÀ CONCORSI

Risultata idonea nell'ambito del concorso CNR per **ricercatore III livello** a tempo indeterminato, di cui al bando n. 368.14 – Area strategica: la mente umana e la sua complessità, n. 2 posti, Prot. AMMCNT n. 85544 del 22/12/2016.

Risultata idonea nell'ambito del concorso CNR per **collaboratore tecnico VI livello** a tempo indeterminato presso l'Istituto di Teorie e Tecniche dell'Informazione Giuridica (ITTIG) di Firenze, di cui al bando n. 367.109, n. 1 posto, Prot. AMMCNT n. 0051136 del 27/07/2017.

ESPERIENZA PROFESSIONALE

15 maggio 2019 - oggi

Ricercatore a tempo indeterminato

Struttura: Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR) di Pisa

15 maggio 2018 - 14 maggio 2019

Ricercatore a tempo determinato

Struttura: Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR) di Pisa

Tematica: Strumenti e tecniche avanzate di NLP per la definizione di nuove strategie di monitoraggio del settore educativo

Riferimenti: provvedimento n. 38/2018 protocollo n. 0000315 del 8/05/2018, del direttore f.f. dell'Istituto di Linguistica Computazionale "A. Zampolli", lettera di assunzione protocollo n. 0000319 del 10/05/2018, contratto n. protocollo 0000322 del 14/05/2018, n. protocollo AMMCE 0033767 del 14/05/2018.

17 marzo 2017 - 16 giugno 2017

Contratto di Supporto alla Didattica

Struttura: Università di Pisa, Dipartimento di Filologia, Letteratura, Linguistica

Tematica: supporto alla didattica dell'insegnamento di "Linguistica computazionale II", nell'ambito del corso di laurea magistrale in Informatica Umanistica

Riferimenti: num. prot. 0000695/2017 (17/03/2017).

1 novembre 2016 - 10 maggio 2018

Assegno di ricerca Senior

Struttura: Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR) di Pisa

Tematica: Definizione di metodologie e tecniche di trattamento automatico del linguaggio per l'identificazione, la gestione e la scrittura di breaking news a supporto dei giornalisti, con particolare attenzione all'analisi dei contenuti veicolati dai social media, nell'ambito del progetto di ricerca SMART NEWS "Social Sensing for breakingnews" - Regione Toscana - Bando FAR FAS 2014.

Riferimenti: atto di conferimento dell'assegno per il periodo dal 1/11/2016 al 30/10/2017, protocollo n. 0000827 del 31/10/2016; proroga dal 1/11/2017 al 31/03/2018, n. protocollo 0000895 del 31/10/2017; proroga dal 1/04/2018 al 31/08/2018, n. protocollo 0000244 del 30/03/2018 dimissioni in data 9/05/2018, n. protocollo 0000317 del 09/05/2018

- 11 aprile 2016 - 10 luglio 2016 **Contratto di Supporto alla Didattica**
Struttura: Università di Pisa, Dipartimento di Filologia, Letteratura, Linguistica
Tematica: supporto alla didattica dell'insegnamento di "Linguistica computazionale II", nell'ambito del corso di laurea magistrale in Informatica Umanistica
Riferimenti: num. prot. 0001102/2016 (07/04/2016).
- 24 novembre 2014 - 23 ottobre 2016 **Assegno di ricerca**
Struttura: Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR) di Pisa
Tematica: Sviluppo di metodologie e tecniche avanzate per la valutazione della complessità linguistica di testi in vista della definizione di linguaggi controllati e di metodi di semplificazione testuale, con particolare attenzione all'analisi di linguaggi specialistici, nell'ambito del progetto di ricerca PRIN "Scritture brevi, semplificazione linguistica, inclusione sociale: modelli e applicazioni (SCRIBE)".
Riferimenti: atto di conferimento dell'assegno per il periodo dal 24/11/2014 al 23/11/2015, protocollo n. 0001035 del 21/11/2014 rinnovo dal 24/11/2015 al 23/05/2016, n. protocollo 0001097 del 20/11/2015; proroga dal 24/05/2016 al 23/08/2016, n. protocollo 0000453 del 23/05/2016; proroga dal 24/08/2016 al 23/10/2016, n. protocollo 0000663 del 24/08/2016
- 1 novembre 2012 - 31 ottobre 2014 **Assegno di ricerca**
Struttura: Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR) di Pisa
Tematica: Legal Text Mining: costruzione di reti semantico-concettuali finalizzate a una navigazione intelligente di corpora di testi giuridici, nell'ambito del progetto JURNET cofinanziato dalla Regione Toscana nell'ambito del POR CRO FSE 2007-2013 Asse IV – Capitale Umano
Riferimenti: atto di conferimento dell'assegno per il periodo dal 1/11/2012 al 31/10/2014, protocollo n. 0000955 del 25/10/2012
- 1 luglio 2011 - 31 ottobre 2012 **Assegno di ricerca**
Struttura: Scuola Superiore Sant'Anna di Pisa
Tematica: Law and Language: strumenti di formalizzazione linguistica del dialogo giudiziale in ambito comunitario, settore disciplinare Ius/02, Diritto Privato Comparato
Riferimenti: approvazione della selezione per titoli dell'assegno per il periodo dal 1/07/2011 al 30/06/2012, protocollo n. 0006836 del 27/06/2011; proroga dal 1/07/2012 al 31/12/2012, conferita in data 21/06/2012
- 1 aprile 2010 - 30 aprile 2011 **Assegno di ricerca**
Struttura: Istituto di Ricerche sulle Attività Terziarie (IRAT-CNR) di Napoli
Tematica: Sviluppo di metodologie e tecniche per il monitoraggio della competenza linguistica italiana di alunni stranieri della scuola primaria e secondaria mediante tecnologie avanzate nel campo della linguistica computazionale, nell'ambito del progetto nazionale CNR "Migrazioni", all'interno della linea di ricerca "Tecnologie avanzate e Integrazione linguistico-culturale nella scuola"
Riferimenti: atto di conferimento assegno per il periodo dal 1/04/2010 al 30/09/2010, protocollo n. 0000347 del 31/03/2010; proroga dal 1/10/2010 al 28/02/2011, n. protocollo 0000948 del 30/09/2010; proroga dal 1/03/2011 al 30/04/2011, n. protocollo 0000263 del 8/03/2011
- 1 settembre 2008 - 30 novembre 2008 **Contratto di Collaborazione**
Struttura: Dipartimento di Linguistica dell'Università di Roma 3
Tematica: Stesura di un rapporto tecnico contenente un'analisi comparativa degli schemi di annotazione morfosintattica e sintattica (tagging) per la lingua italiana
Riferimenti: atto di conferimento Prot. n. 153 REP. N. 1/2008.
- 1 aprile 2007 - 31 marzo 2010 **Borsa di Studio**
Struttura: Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR) di Pisa
Tematica: Metodi e tecniche per l'estrazione automatica di informazione semantico-lessicale da collezioni documentali di ambito giuridico e biomedico, con particolare enfasi sull'acquisizione di informazione relativa agli eventi di dominio
Riferimenti: lettera conferimento borsa, protocollo n. 0000146 del 02/04/2007

2008-2011 **Scuola di Dottorato in Studi Umanistici – Dottorato in Studi Euro-Asiatici: Indologia, Linguistica, Onomastica (XXIII ciclo), indirizzo di Linguistica, Linguistica Applicata e Ingegneria Linguistica**

Titolo della tesi: "Lingua e Diritto: una prospettiva linguistico-computazionale"

Ente: Università di Torino

Data: 7 ottobre 2011

Relatore: Prof. Mario Squartini

Membri della commissione giudicatrice: Prof. Mario Squartini, Dr. Simonetta Montemagni, Prof. Mirko Tavoni, Prof. Antonio Lazari.

2004-2006 **Laurea specialistica in Lingua e Letteratura Italiana**

Titolo della tesi: "L'ambiente, le norme, il computer. Uno studio linguistico-computazionale per la creazione di ontologie giuridiche in materia ambientale"

Ente: Università di Pisa

Classe di laurea: Classe delle lauree specialistiche in Lingua e Cultura Italiana (40/S)

Data: 5 dicembre 2006

Votazione: 110/110 e lode

Relatori: Prof. Mirko Tavoni, Dr. Simonetta Montemagni, Prof. Alessandro Lenci.

2001-2004 **Laurea triennale in Lettere**

Titolo della tesi: "Ognunu parla comu sa. Analisi sociolinguistica della sceneggiatura di *Così ridevano* di Gianni Amelio"

Ente: Università di Pisa

Classe di laurea: Classe delle lauree in Lettere (5)

Data: 23 ottobre 2004

Votazione: 110/110 e lode

Relatori: Prof. Mirko Tavoni e Prof. Lorenzo Cuccu.

luglio 2001 **Maturità classica**

Liceo Classico *C. Cavour* di Torino.

Votazione: 92/100.

COMPETENZE PERSONALI

Lingua madre Italiano

Altre lingue

Inglese

Francese

	Comprensione		Parlato		Scritto
	Ascolto	Lettura	Interazione	Produzione orale	
Inglese	B2	B2	B2	B2	B2
Francese	A3	A3	A3	A3	A3

Livelli: A1/A2: Livello base - B1/B2: Livello intermedio - C1/C2: Livello avanzato
Quadro Comune Europeo di Riferimento delle Lingue

Competenze informatiche

Conoscenza dei sistemi operativi Windows e Linux OS, del pacchetto Office e Open Office.
Conoscenza dei più diffusi browser web: Internet Explorer, Mozilla/Firefox, Google Chrome.
Conoscenza degli editor di testo Notepad++, Word, Emacs.
Conoscenza della sintassi delle espressioni regolari.
Conoscenza di linguaggi di marcatura: XML.
Conoscenza di LaTeX, linguaggio di markup usato per la preparazione di testi basato sul programma di composizione tipografica TeX.

Lingua e diritto: una prospettiva linguistico-computazionale.

Sommario: La tesi intende indagare i modi e le forme del rapporto tra lingua e diritto, tra analisi linguistica di testi giuridici e accesso al loro contenuto informativo, adottando una prospettiva innovativa rispetto a quella storicamente abbracciata da linguisti, giuristi e ricercatori in Intelligenza Artificiale e Diritto: la prospettiva linguistico-computazionale. L'obiettivo era quello di suggerire una risposta alle seguenti domande di ricerca:

- considerati i ben noti caratteri di complessità della lingua del diritto, in che misura gli strumenti di annotazione linguistica automatica del testo sviluppati per analizzare la lingua comune sono accurati nel trattamento automatico della lingua del diritto?

Risultati: costruzione di un corpus di testi legislativi in lingua italiana linguisticamente annotato fino al livello sintattico a dipendenze grazie alla specializzazione di criteri di annotazione seguiti per l'analisi di testi giornalistici per poter trattare le peculiarità del linguaggio giuridico. Ciò ha reso possibile *i)* quantificare l'accuratezza dell'annotazione linguistica di strumenti statistici e *ii)* individuare le principali strutture sintattiche dei testi giuridici che ne rendono complessa l'analisi da parte di un parser sintattico diminuendone l'accuratezza. Il corpus è stato messo liberamente a disposizione per la valutazione di strumenti di Trattamento Automatico del Linguaggio su testi giuridici e per poter essere utilizzato come training corpus in approcci di adattamento supervisionato al dominio.

- nonostante le complessità della lingua del diritto individuate e le relative difficoltà di analisi, gli strumenti di annotazione linguistica automatica sono in grado di ricostruire un articolato profilo linguistico di testi giuridici?

Risultati: definizione di una metodologia di ricostruzione del profilo linguistico di testi giuridici. Essa consiste nel monitoraggio comparativo tra un'ampia gamma di caratteristiche lessicali, morfo-sintattiche e sintattiche rintracciate in modo automatico in corpora di testi rappresentativi della lingua del diritto e della lingua ordinaria linguisticamente annotati in modo automatico. I risultati ottenuti hanno permesso di suggerire una possibile risposta operativa alla dibattuta e aperta questione circa i non lievi problemi di delimitazione tra lingua del diritto e linguaggio comune.

- in un'ottica di analisi incrementale del testo, in che modo i vari livelli di annotazione linguistica automatica del testo costituiscono il punto di partenza per accedere alla semantica di un testo giuridico?

Risultati: definizione di una metodologia di accesso incrementale al contenuto informativo di testi giuridici articolata in due passaggi: *i)* estrazione automatica della terminologia rilevante per il dominio dei testi in input e *ii)* annotazione semantica (descritta al punto seguente). La metodologia di estrazione automatica della terminologia di dominio ha permesso di discriminare le diverse tipologie di lessico presenti in un testo giuridico: quello relativo ai fatti oggetto del diritto (espressione della realtà 'extra-giuridica'), quello relativo ai concetti giuridici (espressione della realtà 'giuridica') e il lessico comune. La questione si inserisce nell'aperto dibattito nella comunità di ricerca in materia di Intelligenza Artificiale e Diritto impegnata nella costruzione di ontologie di dominio nelle quali le due tipologie di lessico non siano indiscriminatamente mischiate.

- qual è il modello di organizzazione e rappresentazione del significato più adatto a rendere esplicito il modo in cui i principali elementi informativi sono linguisticamente realizzati e organizzati in un testo giuridico?

Risultati: costruzione di un corpus di testi legislativi italiani annotato con informazione relativa a scenari deontici di *obbligo*, *permesso* e *divieto* sulla base di un'innovativa modalità di annotazione semantica. Il corpus è stato annotato seguendo e specializzando i principi di rappresentazione semantico-lessicale basata sui presupposti teorici della Frame Semantics Theory di Charles Fillmore e sul modello del progetto FrameNet. Tale metodologia si è dimostrata particolarmente espressiva per la rappresentazione del significato e della conoscenza in ambito giuridico rendendo esplicite le relazioni tra contenuto semantico-informativo di un testo giuridico e la sua corrispondente realizzazione linguistica. L'importanza della costruzione di tale corpus, unico nel suo genere per la lingua italiana, è tanto più evidente considerandone i vantaggi in un'ottica di futura automatizzazione del processo di annotazione semantica.

- 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.

Curatele

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.

Articoli in riviste

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: *Computación 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. Nikiforidou (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. Rebolz-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 Bioinformatics*, 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.

Articoli in atti di convegno

Coccu 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 3rd 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), *CLtA: 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), *CLtA: 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 1st 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.
- 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.

Deliverables e rapporti tecnici

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.

Pisa, 05 dicembre 2018