

# Curriculum Vitae Europass

## Informazioni personali

Cognome/i nome/i

Indirizzo/i

Telefono/i

Email

Nazionalità

Data di nascita

### Dominique Brunato

Via Francesco Pardi, 9, 56124, Pisa (PI)

+39 3402356841

dominique.brunato@yahoo.it

Italiana

29 Agosto 1982

## Aree di studio

### Linguistica teorica e computazionale, Acquisizione del linguaggio, Scienze Cognitive

## Sommario

Dominique Brunato è assegnista di ricerca dal 2013 presso l'Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) del CNR, dove collabora ai progetti e alle linee di ricerca del laboratorio *Italian Natural Language Processing Lab* (ItaliaNLP Lab, [www.italianlp.it](http://www.italianlp.it)). Ha conseguito la laurea specialistica in Linguistica nel 2010 e il Dottorato di Ricerca in Scienze Cognitive nel 2015, entrambi presso l'Università degli Studi di Siena. I suoi interessi di ricerca si collocano nell'ambito della linguistica teorica e computazionale e affrontano le seguenti tematiche:

- studio della **complessità linguistica** nel sistema e in rapporto ai processi cognitivi di comprensione e produzione del linguaggio;
- definizione di modelli basati su metodi e strumenti di Natural Language Processing (NLP) per il **trattamento e la misurazione della complessità linguistica** di un testo;
- **monitoraggio linguistico** di tipologie testuali e varietà d'uso della lingua (strutturali, funzionali, situazionali);
- studio dei **processi di acquisizione e sviluppo delle competenze linguistiche** nell'apprendimento della lingua madre a diversi livelli di scolarità.

La sua ricerca è caratterizzata da un forte legame con applicazioni di uso reale, in particolare nel settore delle tecnologie dell'educazione, quali la profilazione di materiali didattici rispetto alle competenze linguistiche dello studente, la creazione di percorsi didattici personalizzati anche non convenzionali o a distanza e lo sviluppo di metodologie basate su strumenti di NLP a supporto dell'autovalutazione dei processi di apprendimento.

## Esperienza professionale

Date

Funzione o posto occupato

Tipologia

Nome e indirizzo del datore di lavoro

Riferimento o numero di protocollo

15 febbraio 2018 - in corso

Titolare di assegno di ricerca

Assegno di ricerca di tipologia "Grant" per lo svolgimento di attività di ricerca inerenti l'ambito settoriale strategico "Smart, Secure and inclusive communities" sul progetto specifico dal titolo PERFORMA "Personalizzazione di pERCorsi FORMativi Avanzati" Supervisore interno: Dott. Felice Dell'Orletta

Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC), Consiglio Nazionale delle Ricerche (CNR), Via Moruzzi 1, Pisa, Italia

Progetto congiunto di alta formazione in regime di cofinanziamento con la Regione Toscana, POR FSE 2014-2020 Asse A Occupazione – Priorità di investimento A.2 – Obiettivo A.2.1 – Azione A.2.1.7 - Programma d'Intervento ARCO-CNR, CUP "B56J17001330004", Decreto n.14139 del 21/09/2017

Altre informazioni	Bando Regione Toscana n° ILC.ASS.006.2017.PI, Prot. ILC n. 0001016 - PROT. CNR n. 0084296 Atto di conferimento dell'incarico prot. ILC-CNR n. 0000119 del 13/02/2018 Importo totale finanziamento: €37.800,000 Importo finanziamento per Unità Operativa: ILC-CNR €27.000,00 Nominativo coordinatore del progetto: Istituto di Scienza e Tecnologie dell'Informazione, A. Faedo - CNR
Date	1 aprile 2017 - 9 febbraio 2018
Funzione o posto occupato	Titolare di assegno di ricerca
Tipologia	Assegno di ricerca Post Dottorale nell'ambito del progetto di ricerca "UBIquitous Massive Open Learning (UBImol)" per la tematica "Definizione di metodologie e tecniche di Trattamento Automatico della Lingua per lo studio dei processi di apprendimento linguistico e a supporto della valutazione e auto-valutazione in ambito didattico". Supervisore interno: Dott. Felice Dell'Orletta.
Nome e indirizzo del datore di lavoro	Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC), Consiglio Nazionale delle Ricerche (CNR), Via Moruzzi 1, Pisa, Italia
Riferimento o numero di protocollo	Progetto Regione Toscana, Bando POR FESR 2014-2020, Decreto n.3389 del 30/07/2014 Regione Toscana
Altre informazioni	Atto di conferimento dell'incarico, prot. ILC-CNR n. 0000276 del 31/03/2017 Importo totale finanziamento: €540.000,000 Importo finanziamento per Unità Operativa: Unità Operativa ILC-CNR €81.000,00 Nominativo coordinatore del progetto: Paolo Ongaro (M.E.T.A. S.r.l.)
Date	22 ottobre 2015 - 18 marzo 2017
Funzione o posto occupato	Titolare di assegno di ricerca
Tipologia	Assegno di ricerca professionalizzante nell'ambito del programma di ricerca SMART NEWS per la tematica "Definizione di metodologie e tecniche da utilizzarsi all'interno di piattaforme di e-learning destinate a scuole di diverso ordine e grado, con particolare attenzione i) all'estrazione e formalizzazione di conoscenza da testi e ii) all'analisi della leggibilità di un testo e a supporto della sua eventuale semplificazione linguistica." Supervisore interno: Dott.ssa Simonetta Montemagni.
Nome e indirizzo del datore di lavoro	Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC), Consiglio Nazionale delle Ricerche (CNR), Via Moruzzi 1, Pisa, Italia
Riferimento o numero di protocollo	Bando FAR FAS 2014 della Regione Toscana e azienda META srl, Decreto n.6649 del 10/12/2015 Regione Toscana
Altre informazioni	Atto di conferimento dell'incarico, prot. ILC-CNR, n°0001001 del 16/10/2015; I° rinnovo (dal 19/10/2016 al 18/01/2017), Protocollo ILC-CNR n°0000795 del 18/10/2016; II° rinnovo (dal 19/01/2017 al 18/03/2017): Protocollo ILC-CNR n°0000023 del 18/01/2017 Importo totale del finanziamento: €654.861,37 Importo finanziamento per Unità Operativa: Unità Operativa ILC-CNR €94.340,5 Nominativo coordinatore del progetto: Sauro Salvadori (Hyperborea S.r.l.)
Date	1 Ottobre 2013 - 30 Settembre 2015
Funzione o posto occupato	Titolare di assegno di ricerca
Tipologia	Assegno di ricerca professionalizzante nell'ambito del programma di ricerca "iSLe" (Intelligent Semantic Liquid eBook) per la tematica "Definizione e specializzazione di metodologie e tecniche per la valutazione della leggibilità di un testo e a supporto della sua eventuale semplificazione, in vista della loro integrazione come funzionalità avanzate della piattaforma software per l'editoria digitale scolastica del progetto iSLe". Supervisore interno: Dott.ssa Simonetta Montemagni.

Nome e indirizzo del datore di lavoro  
Riferimento o numero di protocollo  
Altre informazioni

Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC), Consiglio Nazionale delle Ricerche (CNR), Via Moruzzi 1, Pisa, Italia  
Bando della Regione Toscana POR CRReO FESR 2007-2013, linea di intervento 1.5.a e 1.6, Decreto n. 5874 del 10/12/2012, <http://www.progettoisle.it/il-progetto-isle/>  
Atto di conferimento dell'incarico, prot. ILC-CNR n°0000949 del 30/09/2013; I° rinnovo (dal 31/03/2014 al 31/03/2015), Protocollo ILC-CNR n°0000843 del 29/09/2014; II° rinnovo (01/04/2015-30/09/2015), Protocollo ILC-CNR n°0000405 del 31/03/2015  
Importo totale finanziamento: €1.038.917,42  
Importo finanziamento per Unità Operativa: ILC-CNR €120.610,00  
Nominativo coordinatore del progetto: Paolo Ongaro (M.E.T.A. S.r.l.)

Date  
Funzione o posto occupato

Dicembre 2010 - Dicembre 2012  
Linguista Computazionale junior, Contratto di prestazione di lavoro autonomo occasionale

Principali mansioni e responsabilità

– sviluppo di algoritmi basati su regole linguistiche implementati in linguaggi proprietari per la categorizzazione automatica di testi in linguaggio naturale, l'estrazione di informazioni, il riconoscimento di entità singole e di relazioni tra entità, tramite l'uso di software basati sulla tecnologia semantica;  
– analisi di termini, espressioni e costruzioni tipiche dell'espressione di un commento, giudizio e opinioni ("sentiment");  
– arricchimento della rete semantica con i termini relativi al "sentiment";  
– costruzione e annotazione di corpora con informazione di natura lessicale, morfosintattica e sintattica.

Nome e indirizzo del datore di lavoro  
Tipo o settore d'attività

Expert System S.p.A., Via Virgilio 56/q, 41123, Modena, Italia  
Information & Communication Technology, NLP, Tecnologie semantiche

## Capacità e competenze professionali

Madrelingua/e

Autovalutazione  
Livello europeo(\*)

Inglese  
Spagnolo

### Italiano

Comprensione		Parlato		Scritto
Ascolto	Lettura	Interazione	Produzione orale	
C1	C1	C1	C1	C1
A2	A2	A2	A2	A2

(\*) Quadro comune europeo di riferimento per le lingue

Capacità e competenze informatiche

Conoscenza avanzata dei sistemi operativi MS Windows™ (XP, 2000, Vista, Windows 7) e Linux/Ubuntu;  
Software e applicativi: MS Office/ LibreOffice;  
Editor di testo: LateX, Emacs, Word, Notepad++;  
Linguaggi di programmazione: Python (conoscenza base), UNIX shell scripting;  
Buona conoscenza della sintassi delle espressioni regolari.

## Istruzione e formazione

Date  
Certificato o diploma ottenuto

2010 – 2015  
Dottorato di Ricerca in Scienze Cognitive (indirizzo Linguistica), 26° ciclo, conseguito il 21/07/2015

Titolo della tesi

*A study on linguistic complexity from a computational linguistics perspective. A corpus-based investigation of Italian bureaucratic texts.*  
Relatore: Prof.ssa Adriana Belletti

Nome e tipo d'istituto di istruzione o formazione	Facoltà di Lettere e Filosofia, Dipartimento di Scienze Sociali, Politiche e Cognitive, Università degli studi di Siena
Livello nella classificazione nazionale o internazionale	Post Doctoral Degree
Date	2007 - 2010
Certificato o diploma ottenuto	Laurea specialistica in Linguistica, conseguita il 13/04/2010
Titolo della tesi	<i>Effetti del tratto di genere nella comprensione di frasi relative soggetto e oggetto in italiano: uno studio sperimentale in età prescolare.</i>
	Relatore: Prof.ssa Adriana Belletti; Correlatore: Prof. Luigi Rizzi
	Votazione: 110/110 e Lode
Principali materie/competenze apprese	Linguistica teorica, Linguistica applicata, Psicolinguistica, Acquisizione del linguaggio.
Nome e tipo d'istituto di istruzione o formazione	Facoltà di Lettere e Filosofia, Dipartimento di Scienze Sociali, Politiche e Cognitive, Università degli studi di Siena
Livello nella classificazione nazionale o internazionale	Master Degree
Date	2001-2005
Certificato o diploma ottenuto	Laurea di primo livello in Scienze della Comunicazione
Titolo della tesi	<i>Sovrana: Una rivista femminile negli anni del fascismo.</i>
	Votazione: 110/110
Nome e tipo d'istituto di istruzione o formazione	Interfacoltà Lettere e Filosofia – Scienze Politiche, Università degli Studi di Padova
Livello nella classificazione nazionale o internazionale	Bachelor Degree
Date	1996-2001
Certificato o diploma ottenuto	Diploma socio-psico pedagogico sperimentazione Progetto Brocca
Nome e tipo d'istituto di istruzione o formazione	Istituto Magistrale Statale "Don Giuseppe Fogazzaro", Vicenza

## Altre attività formative

### Corsi di formazione con attestato di frequenza

02-05/05/2017: Corso su "Statistical Methods and quantitative analyses of lexical data" organizzato dall'Istituto di Linguistica Computazionale "A. Zampolli" e tenuto dal Dr. Martijn Wieling. I principali argomenti trattati sono stati: Introduzione all'analisi statistica con R; Test Statistici (t-test, ANOVA); Regressione lineare e logistica; Regressione Mixed Effect; Regressione non lineare;

25-27/02/2013: corso base di 21 ore sull'uso del software TalTaC2 (Trattamento Automatico Lessicale e Testuale del Contenuto per l'Analisi di un Corpus), organizzato dall'Università La Sapienza di Roma;

gennaio 2009 – giugno 2009: Corso LIS (Lingua italiana dei segni) di primo livello, organizzato dalla sede ENS (ente nazionale sordomuti) di Siena.

## **Partecipazione a scuole e seminari di ricerca**

24/04/2015: Workshop “Measuring linguistic complexity: A multidisciplinary perspective”, organizzato dalla *Linguistic Research Unit of the Institute of Language and Communication*, Université catholique de Louvain, Louvain-la-Neuve;

16-20/09/2013: IQLA-GIAT 1st Summer School in Quantitative Analysis of Textual Data, organizzata dal Dipartimento di Statistica dell'Università di Padova. Il programma formativo ha trattato le diverse fasi dell'analisi quantitativa di dati testuali: dalla raccolta di corpora (anche da web), alla teoria linguistica alla base delle analisi quantitative, all'utilizzo dei software di analisi.

16-20/07/2013: UCREL Summer School in Corpus Linguistics, organizzata dal *Department of Linguistics and English Language della Lancaster University*, UK. Il programma formativo ha trattato aspetti di ricerca e applicazione della linguistica dei corpora, presentati nell'ambito di lezioni teoriche e sessioni di laboratorio dedicate alla sperimentazione di strumenti informatici per la creazione, l'interrogazione e l'analisi di corpora.

16-20/07/2012: First Summer School on “Ontological Analysis” (Povo, TN), organizzata dalla International Association for Ontology and its applications (IAOA);

23-27/01/2012: Scuola Invernale TRIPLE [SIT] 2012 – Classi di parole: modelli, tipologia, computazione, presso Università Roma Tre – Dipartimento di Linguistica;

## **Attività scientifica e di ricerca**

### **Iniziative e incarichi scientifici**

Da ottobre 2013 ad oggi: è membro del laboratorio di ricerca *Italian Natural Language Processing Lab* (ItaliaNLP Lab) dell'Istituto di Linguistica Computazionale “Antonio Zampolli” (ILC-CNR) di Pisa.

Da maggio 2017 ad oggi: è cultore della materia per l'insegnamento di Linguistica Computazionale 1 del CDL di Informatica Umanistica presso l'Università di Pisa (Rif: Delibera n. 66 del 17 maggio 2017 del Consiglio del dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa).

### **Incarichi di docenza in corsi universitari, master di alta formazione e corsi professionali**

a.a. 2017/2018: ha svolto attività di docenza nell'ambito del Master “La lingua del diritto: comprensione, elaborazione e applicazioni professionali”, organizzato dall'Università degli Studi di Pavia, Dipartimento di Giurisprudenza; Materia di insegnamento: 1) Le tecnologie del linguaggio per la valutazione della qualità degli atti normativi, all'interno del modulo “Linee guida per la formulazione degli atti normativi”; 2) Prove di redazione di atti normativi (laboratorio) (Nr. ore: 4 complessive)

Riferimenti: In possesso di lettera di affidamento di docenza rilasciata dall'Università degli Studi di Pavia in data 27/02/2018, certificata da prot. n° 1560.

a.a. 2016/2017: ha svolto attività di supporto alla didattica nell'ambito dell'insegnamento di Linguistica Computazionale II del CDL in Tecnologie del Linguaggio tenuto dalla dott.ssa Simonetta Montemagni e dalla dott.ssa Giulia Venturi nel CDL Magistrale in Informatica Umanistica, Università di Pisa; Materia di insegnamento: 1) La semplificazione automatica e semi-automatica del testo: metodi, risorse e prospettive applicative; 2) Annotazione di operazioni di semplificazione all'interno di corpora allineati (Nr. 4 ore complessive).

Riferimenti: In possesso di lettera di affidamento dell'incarico rilasciata dal docente titolare del corso, dott.ssa Simonetta Montemagni, in data 29/06/2016.

23-26 gennaio 2017: ha svolto attività di docenza nell'ambito delle giornate formative "La qualità degli atti normativi e amministrativi" rivolte a dipendenti di enti locali finalizzato all'uso di tecnologie del linguaggio per la valutazione della complessità linguistica di atti normativi e amministrative e organizzate dall'Agenzia Formativa Unione dei Comuni della Versilia, Querceta Seravezza (LU).

Riferimenti: contratto per prestazione professionale occasionale conferito dall'agenzia formativa dell'Unione dei Comuni della Versilia in data 19/01/2016.

26 ottobre 2016: ha svolto attività di supporto alla formazione degli studenti nell'ambito dell'insegnamento di "Tecniche di redazione dei regolamenti e degli atti amministrativi" del CDL in Scienze Politiche dell'Università di Pisa; Materia di insegnamento: Le tecnologie linguistico-computazionali per la leggibilità della comunicazione pubblica.

Riferimenti: In possesso di lettera di affidamento dell'incarico sottoscritta dal Direttore del Dipartimento, Prof. Paolo Nello, in data 18/10/2016, prot. 0003309/2016.

#### **Altre attività legata alla collaborazione in ambito accademico**

Correlatore per le seguenti tesi di laurea:

a.a. 2016-2017: *Definizione di un metodo e creazione di un corpus per la valutazione di un sistema di semplificazione automatica del testo*, Tesi di Laurea Triennale in Informatica Umanistica, Università di Pisa. Laureanda: Giulia Cantoni. Relatore: Felice Dell'Orletta.

a.a. 2016-2017: *Lessico e Sintassi: studio delle variazioni tra generi e complessità*, Tesi di Laurea Triennale in Informatica Umanistica, Università di Pisa. Laureando: Pietro dell'Oglio. Relatore: Felice Dell'Orletta.

a.a. 2017-2018: *Creazione di un corpus e primi esperimenti per l'identificazione delle parti informative di un testo a supporto di un sistema di generazione di domande di comprensione*, Tesi di Laurea Magistrale in Informatica Umanistica, Università di Pisa. Laureanda: Martina Valeriani. Relatore: Felice Dell'Orletta.

#### **Premi e Riconoscimenti scientifici**

4 dicembre 2015: *Distinguished Young Paper mention* per l'articolo "Isacco: a corpus for investigating spoken and written language development in Italian school-age children", pubblicato negli atti della seconda Conferenza Italiana di Linguistica Computazionale, 3-4 dicembre 2015, Fondazione Bruno Kessler, Trento.

2010-2013: Vincitrice di una borsa di studio triennale per la frequenza del Dottorato di Ricerca in Scienze Cognitive, 26° ciclo, erogata dall'Università degli Studi di Siena.

luglio 2009: Vincitrice di una borsa di studio per la frequenza di un corso intensivo di lingua inglese della durata di un mese presso il King's College (UK), erogata dall'Università degli Studi di Siena.

#### **Organizzazione di conferenze internazionali/nazionali, partecipazione a comitati scientifici di conferenze internazionali/nazionali**

11/12/2016: membro del Comitato organizzativo del Workshop "CL4LC - Computational Linguistics for Linguistic Complexity", tenutosi nell'ambito della "26<sup>a</sup> International Conference on Computational Linguistics (COLING 2016)", Osaka, Giappone, (<https://sites.google.com/site/cl4lc2016/home>)

15-16/10/2012: membro del Comitato organizzativo del "9° Workshop of Syntax and Semantics", tenutosi presso il CISCL (Centro Interdipartimentale di Studi Cognitivi sul Linguaggio), Università degli Studi di Siena.

31/10/2018-4/11/2018: membro del Comitato scientifico in qualità di reviewer della conferenza internazionale "Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)", Bruxelles, Belgio.

10-12/12/2018: membro del Comitato scientifico della "Fifth edition of the Italian Computational Linguistics Conference" (CLiC-it) per l'area "Morphology and Syntax", Torino.

05/06/2018: membro del Comitato scientifico della terza edizione del “Workshop on Innovative Use of NLP for Building Educational Applications (BEA13)”, nell’ambito della 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics. (Riferimenti: <http://www.cs.rochester.edu/~tetreaul/naacl-bea13.html#committee>)

08/08/2018: membro del Comitato scientifico della seconda edizione del “Workshop on Innovative Use of NLP for Building Educational Applications (BEA12)”, nell’ambito della 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Danimarca. (Riferimenti: <http://www.cs.rochester.edu/u/tetreaul/emnlp-bea12.html#committee>)

07-11/09/2017: membro del Comitato scientifico della “2017 Conference on Empirical Methods in Natural Language Processing”, Copenhagen, Danimarca. (Riferimenti: <http://www.aclweb.org/anthology/D/D17/D17-1.pdf>)

11-13/12/2017: membro del Comitato scientifico della “Fourth Italian Conference on Computational Linguistics (CLiC-it 2017)”, Roma. (Riferimenti: <http://ceur-ws.org/Vol-2006/part1.pdf>)

05-06/12/2016: membro del Comitato scientifico della “Third Italian Conference on Computational Linguistics (CLiC-it 2017)”, Napoli (Riferimenti: <http://ceur-ws.org/Vol-1749/preface.pdf>)

## Attività di referaggio di articoli, libri e progetti

da gennaio 2018: reviewer per la rivista Language Resources and Evaluation (LREV) (rivista ANVUR di fascia A per l’area 10)

2017: reviewer per la rivista *Journal of the Association for Information Science and Technology (JASIST)*

2016: reviewer per lo special issue in Digital Humanities and Computational Linguistics, pubblicato dalla rivista Italian Journal on Computational Linguistics, (2)2.

## Temi di ricerca

I temi di ricerca affrontati nel corso della formazione accademica e nei successivi incarichi di collaborazione scientifica sono elencati di seguito:

### **Acquisizione della prima lingua e teoria sintattica**

A partire dalla tesi di laurea magistrale si è occupata di aspetti relativi all’acquisizione della prima lingua, con particolare interesse per i fenomeni connessi all’acquisizione di strutture sintattiche complesse nella prima infanzia. In particolare, ha approfondito lo studio delle teorie grammaticali di ispirazione minimalista che riconducono la difficoltà posta da frasi complesse quali le frasi relative oggetto al principio di località sintattica della Minimalità Relativizzata (Friedmann et al., 2009). In questo quadro teorico di riferimento, ha condotto uno studio sperimentale per l’italiano volto a indagare il ruolo del tratto di genere nella comprensione di frasi relative oggetto da parte di bambini in età prescolare. (Pubblicazioni di riferimento: 3 e 4 della sezione dedicata).

### **Studio della complessità linguistica e definizione di metriche per la sua misurazione automatica**

Come parte delle ricerche per la tesi di dottorato, ha approfondito gli studi sulla complessità linguistica, in particolare sintattica, in rapporto ai processi cognitivi di comprensione e produzione del linguaggio, al fine di definire modelli robusti computazionalmente e dotati di plausibilità cognitiva, per la misurazione automatica della complessità linguistica di un testo. All’interno del laboratorio *ItaliaNLP Lab* dell’Istituto di Linguistica Computazionale, dal 2013 collabora alle ricerche sulla valutazione automatica della leggibilità dei testi e sulla semplificazione automatica e semi-automatica, con particolare riferimento alla lingua italiana. In questo contesto, si è occupata della costruzione di risorse linguisticamente annotate adatte all’addestramento di algoritmi di apprendimento automatico per la valutazione della complessità linguistica e la semplificazione dei testi. (Pubblicazioni di riferimento: 7, 13, 15, 17, 18).

### **Studio della variazione linguistica**

Ha condotto studi di monitoraggio della lingua italiana basati su analisi generate da strumenti di Trattamento Automatico del Linguaggio per la caratterizzazione dei tratti linguistici della lingua nelle sue diverse varietà d'uso (strutturali, funzionali, situazionali), comprese varietà di lingua "non canoniche" (Pubblicazioni di riferimento: 6, 8, 11, 12). Come parte della tesi di dottorato, ha approfondito la letteratura sui linguaggi specialistici, in particolare quello burocratico-amministrativo, con l'obiettivo di distinguere i tratti di complessità linguistica connaturati al testo burocratico dalle istanze tipiche del cosiddetto "burocratese", come requisito preliminare per la specializzazione di un indice di leggibilità automatico sulle caratteristiche lessicali, morfo-sintattiche e sintattiche di questa tipologia di testi. (Pubblicazioni di riferimento: 2, 5, 16).

### **Studio dei processi di apprendimento ed evoluzione delle competenze linguistiche**

Si è occupata dello studio e della modellazione computazionale dei processi di apprendimento ed evoluzione delle competenze linguistiche a partire dall'analisi di produzioni linguistiche di studenti (*learner corpora*) a diversi livelli di scolarità. In questo contesto, ha curato la creazione di un primo corpus sperimentale per l'italiano, che raccoglie riassunti in forma scritta e orale di bambini della scuola primaria. Il corpus raccolto ha permesso di condurre un'indagine volta alla valutazione dell'effetto della variazione diamesica sull'evoluzione delle competenze linguistiche degli studenti attraverso la misurazione di indicatori di complessità linguistica estratti automaticamente dal testo. (Pubblicazioni di riferimento: 1, 14)

### **Partecipazione a progetti di ricerca**

I temi di ricerca documentati nella sezione precedente sono stati affrontati anche nell'ambito di progetti di ricerca a cui ha preso parte come titolare di assegno di ricerca o collaboratore presso l'Istituto di Linguistica Computazionale del CNR. (La partecipazione ai progetti di seguito elencati è attestata dalla lettera di certificazione rilasciata dai responsabili scientifici dei progetti coinvolti per l'Unità ILC-CNR, Dott. Felice Dell'Orletta (lettera prot.ILC-CNR n° 0000579 del 5/09/2018) e Dott.ssa Simonetta Montemagni (lettera con protocollo ILC-CNR n° 0000618 del 04/09/2018):

1. **PERFORMA - Personalizzazione di percorsi FORMativi Avanzati** (Progetto Congiunto di Alta Formazione della Regione Toscana, Bando POR FSE 2014-2020 Asse A Occupazione – Priorità di investimento A.2 – Obiettivo A.2.1 – Azione A.2.1.7 - Programma d'Intervento ARCO-CNR).

**Finalità del progetto:** Il progetto, attualmente in corso, si propone di definire nuove metodologie basate sull'uso di tecnologie del linguaggio per la creazione di percorsi di formazione a distanza personalizzabili a diversi livelli. Il risultato sarà lo sviluppo di un sistema di profilazione dei materiali didattici fruibili all'interno di piattaforme di e-learning, deputato a valutarne l'adeguatezza rispetto alle competenze linguistiche dei corsisti e rispetto al dispositivo di lettura (tablet, pc, smartphone) dal quale il corso può essere fruito.

**Ruolo svolto:** Nell'ambito del progetto PERFORMA l'attività svolta si è concentrata sui seguenti aspetti: ricognizione dello stato dell'arte rispetto all'uso delle tecnologie del linguaggio in ambito educativo e didattico, con particolare interesse per la sperimentazione di tali tecnologie all'interno di piattaforme di e-learning di ultima generazione, tra cui i cosiddetti MOOC (Massive Open Online Courses); acquisizione e annotazione linguistica di un corpus sperimentale rappresentativo del genere testuale di riferimento; prime analisi qualitative a quantitative volte a identificare: le caratteristiche linguistiche che contraddistinguono il genere in esame, spaziando dal livello lessicale a quello morfosintattico e sintattico; la loro distribuzione all'interno delle varie tipologie di testi in cui si può articolare il genere (es. testi narrativi, testi espositivi, testi di verifica); gli adattamenti da apportare agli strumenti di annotazione linguistica automatica per consentire l'analisi adeguata delle strutture linguistiche tipiche del dominio. Pubblicazioni collegate: 9, 8.

2. **UBImol - Ubiquitous Massive Open Learning** (Bando POR FESR 2014-2020 della Regione Toscana)

**Finalità del progetto:** Il progetto UBImol si colloca nell'ambito della tematica "Definizione di metodologie e tecniche di Trattamento Automatico della Lingua per lo studio dei processi di apprendimento linguistico e a supporto della valutazione e auto-valutazione in ambito didattico". Si propone di sviluppare una piattaforma di E-learning arricchita con tecnologie innovative in grado di offrire corsi di lingua personalizzati rispetto al livello delle competenze linguistiche specifiche per ciascun profilo di discente.

**Ruolo svolto:** Nell'ambito del progetto UBImol la sua attività si è focalizzata sulla definizione e sperimentazione di una metodologia volta a specializzare sistemi per la valutazione automatica della complessità linguistica di un testo rispetto al dispositivo di fruizione. Una seconda linea di attività ha permesso di definire i requisiti di una metodologia di "self-assessment" personalizzato dell'apprendimento, in grado di valutare l'evoluzione delle competenze linguistiche degli studenti, dall'inizio alla fine del corso, attraverso il monitoraggio di una serie di parametri linguistici articolati su più livelli (lessicale, morfosintattico e sintattico), estratti automaticamente dalle produzioni linguistiche di ciascun corsista tramite strumenti per il Trattamento Automatico della Lingua. Ha curato la redazione dei *deliverables* tecnici elencati ai punti 19, 20, 21.

3. **SMART NEWS - Social sensing for breakingnews** (Bando FAR FAS 2014 della Regione Toscana)

**Finalità del progetto:** Il progetto mira a realizzare uno strumento basato sul paradigma del social sensing per la gestione delle ultime notizie, che aiuti i giornalisti nell'intero processo di individuazione delle breaking news, di raccolta di informazioni in merito ad esse e di scrittura di articoli. Tale strumento prevede l'integrazione di tecnologie del linguaggio per l'analisi della leggibilità dei testi e l'estrazione di informazione.

**Ruolo svolto:** Nell'ambito del progetto è stata condotta una approfondita analisi delle peculiarità linguistiche di varietà di lingua non canonica (in particolare testi tratti dai social media) anche ai fini dell'adattamento di strumenti di Trattamento Automatico del Linguaggio deputati all'annotazione linguistica multilivello e all'estrazione e organizzazione della conoscenza dal testo.

4. **iSLe (Intelligent Semantic Liquid eBook)** (POR CRoO FESR 2007-2013) (periodo 2013-2015):

**Finalità del progetto:** Obiettivo del progetto è stato quello di studiare e sperimentare un nuovo tipo di liquid ebook per l'editoria scolastica e professionale multiplatforma. La piattaforma iSLe ha sfruttato le potenzialità della annotazione semantica dei contenuti digitali e la capacità di ragionamento di un motore inferenziale per implementare in applicazioni multimodali metodi di apprendimento espliciti e impliciti. Nell'ambito del progetto sono state inoltre definite nuove metafore di fruizione e costruzione di percorsi didattici basate sulla social interaction dei discenti.

**Ruolo svolto:** Come titolare dell'assegnato di ricerca legato al progetto "iSLE" per la tematica definita dal bando n°126.048/ASS.009.2013, ha svolto attività di ricerca legate alla definizione e specializzazione di metodologie e tecniche per la valutazione automatica della leggibilità di un testo e a supporto della sua eventuale semplificazione. Le attività condotte nell'ambito della tematica del progetto hanno contemplato la raccolta di un corpus sperimentale di testi rappresentativo del genere di interesse ("materiali didattici") e lo studio qualitativo e quantitativo delle sue peculiarità formali, lessicali e sintattiche, anche al fine di valutare gli adattamenti richiesti agli strumenti di analisi linguistica automatica per il trattamento corretto di questi testi. Ha inoltre collaborato alla definizione di una metodologia sperimentale a supporto della semplificazione automatica del testo definendo un nuovo schema di annotazione per la lingua italiana. Ha curato la redazione dei *deliverables* tecnici elencati ai punti 22, 23, 24. Pubblicazioni scientifiche collegate alle tematiche: 15, 17)

5. **SCRIBE – Scritture Brevi, Semplificazione Linguistica, Inclusione Sociale: Modelli e Applicazioni** (PRIN-2010/2011-2010 FWM3B4), (periodo 2015-2016):

**Finalità del progetto:** L'obiettivo del progetto è stato lo studio interdisciplinare del fenomeno relativo alla produzione di messaggi sintetici e abbreviati dal punto di vista sia sincronico che diacronico, delle espressioni contemporanee (ovvero, brevi scritti usati per i messaggi elettronici, sms e chat) e delle strategie di abbreviazione tipiche dell'italiano e dei sistemi dialettali grafici e linguistici.

**Ruolo svolto:** Ha condotto analisi di monitoraggio linguistico volte a caratterizzare il fenomeno delle "Scritture Brevi" all'interno del quadro di variazione linguistica dell'italiano; queste analisi hanno considerato sia aspetti teorici, quali la ricostruzione del profilo linguistico di testi riconducibili alla tipologia di interesse (es. testi pubblicati all'interno di piattaforme di micro-blogging), sia aspetti più applicativi legati alla valutazione della complessità linguistica e all'individuazione degli aspetti di maggior criticità che questa tipologia di testi pone all'accuratezza degli strumenti di TAL. Le riflessioni scaturite da queste indagini sono documentate nella pubblicazione al punto 6.

## Comunicazioni nell'ambito di conferenze e seminari

29/07/2017: "Manual and automatic approaches for building parallel corpora", seminario a invito nell'ambito della Scuola di Dottorato in Information Technology and Electrical Engineering (ITEE), Università degli Studi di Napoli "Federico II".

29/09/2017: "Metodi automatici di costruzione di un corpus parallelo per lo studio della semplificazione del testo", presentazione al Workshop "SIMPLICITAS. Semplificazione Linguistica della Comunicazione Istituzionale per facilitare l'accessibilità ai contenuti informativi", "LI Congresso internazionale di Studi della Società di Linguistica italiana (SLI)", Università degli studi di Napoli "L'Orientale".

04/04/2016: "Tecnologie del linguaggio per l'analisi di corpora di produzioni di apprendenti l'italiano come L1 e L2: metodologie e prospettive", presentazione nell'ambito del "XIV Congresso della Società internazionale di Linguistica e Filologia italiana (SILFI)", Facoltà di Filologia dell'Università Complutense di Madrid (UCM).

11/12/2015: "Tecnologie del linguaggio e scienza medica", poster presentato nell'ambito del II° Incontro di Studi "Tra medici e linguisti. Lingua e patologia: le frontiere interdisciplinari del linguaggio", Università degli Studi di Napoli Federico II°.

25/09/2015: "The development of grammatical complexity in oral and written child language: some empirical observations from a broad methodological perspective", presentazione al Workshop "Hard sentences: how to measure clausal complexity", "XLIX Congresso internazionale di Studi della Società di Linguistica Italiana (SLI)", Università di Malta.

09/07/2015: "Studio della complessità linguistica e prospettive applicative", presentazione nell'ambito del ciclo *ILC-Spring Seminars*, Istituto di Linguistica Computazionale (ILC-CNR), Pisa.

29/05/2015: "Costruzione di un corpus parallelo di testi in versione originale e semplificata per la lingua italiana: primi risultati, prospettive applicative e di ricerca", presentazione orale nell'ambito del Convegno interannuale PRIN SCRIBE–Scritture Brevi, Università di Macerata.

14/10/2014: "Burocratese e comunicazione pubblica: un'analisi linguistico-computazionale della leggibilità", Giornata di Studio "Reperire, capire e comunicare l'informazione giuridica" promossa dall'Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG-CNR), dall'Associazione per la qualità degli atti amministrativi (AQuAA) e dall'Ordine dei giornalisti della Toscana, Firenze, sede ITTIG-CNR.

11/12/2013: "Oltre il contenuto: tecnologie linguistico-computazionali per l'analisi della struttura linguistica del testo", nell'ambito del Seminario di Cultura Digitale, Dipartimento di Informatica, Università degli studi di Pisa.

26/11/2013: "Le tecnologie linguistico-computazionali nella valutazione della leggibilità dei testi: una indagine sul linguaggio amministrativo", CISCL, Università degli Studi di Siena.

21/11/2013: "Le tecnologie linguistico-computazionali nella misura della leggibilità di atti amministrativi: un caso di studio", Giornata di studio "I testi della Pubblica Amministrazione. Indici di leggibilità e valutazione della qualità", Palazzo Vecchio, Sala dei Dugento, Firenze.

## Produzione Scientifica

### Tesi di dottorato

Titolo: A study on linguistic complexity from a computational linguistics perspective. A corpus-based investigation of Italian bureaucratic texts.

Soggetto della tesi: A partire da una ricostruzione teorica del concetto di complessità linguistica definito dai modelli linguistici di ispirazione cognitiva, la tesi ha approfondito il punto di vista linguistico-computazionale al trattamento e alla misurazione della complessità linguistica, con particolare interesse per le applicazioni basate su strumenti di Trattamento Automatico del Linguaggio naturale (TAL) per la valutazione della leggibilità e semplificazione dei testi. In questo quadro di riferimento sono stati affrontati due casi di studio: un primo caso di studio, basato su un'analisi qualitativa e quantitativa di un corpus di testi italiani della pubblica amministrazione, si è proposto di distinguere i tratti di complessità linguistica "connaturati" al testo burocratico dalle istanze tipiche del cosiddetto burocratese, con la finalità di specializzare un indice automatico per il calcolo della leggibilità sulle caratteristiche (lessicali, morfo-sintattiche e sintattiche) di questa tipologia di testi; un secondo caso di studio ha invece approfondito la letteratura più recente sulla semplificazione automatica del testo contribuendo alla creazione della prima risorsa italiana specifica per questo compito. Per lo sviluppo della risorsa è stato definito un nuovo schema di annotazione contenente un insieme di regole volte all'identificazione e alla classificazione dei principali fenomeni di semplificazione linguistica che possono ricorrere all'interno di testi semplificati manualmente e che si prestano ad essere riprodotti in maniera affidabile con strumenti e metodologie di TAL.

### Articoli in rivista:

1. Dominique Brunato e Felice Dell'Orletta. (2016). Isacco: a corpus for investigating spoken and written language development in Italian school-age children. In *Italian Journal of Computational Linguistics (IJCoL), Emerging Topics at the Second Italian Conference on Computational Linguistics*, vol. 2, n. 1, Giugno 2016, ISSN: 2499-4553.
2. Dominique Brunato e Giulia Venturi. (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*, n. 2014/1, pp. 111–142, ISSN: 0390-0975.
3. Adriana Belletti, Naama Friedmann, Dominique Brunato e Luigi Rizzi. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian, *Lingua*, Volume 122, Numero 10, Agosto 2012, pp. 1053–1069, ISSN: 0024-3841.
4. Dominique Brunato, (2011). Acquisizione del linguaggio e complessità sintattica. Il caso delle frasi relative nei bambini, (2011). *RIFL - Rivista Italiana di Filosofia del Linguaggio*, n. 4, pp. 4–20, Open Access Journal: <http://www.rifl.unical.it/index.php/rifl/article/view/98>, ISSN 2036-6728.

### Capitoli di libri:

5. Dominique Brunato e Giulia Venturi. (2016). Le tecnologie linguistico-computazionali per la leggibilità della comunicazione istituzionale. In: *Saulle Panizza (eds.), Profili attuali di qualità degli atti normativi e amministrativi*, Pisa University Press, pp. 119–157, ISBN 978-886741-6271.
6. Dominique Brunato, Felice Dell'Orletta, Giulia Venturi e Simonetta Montemagni. (2016). Monitoraggio linguistico di Scritture Brevi: aspetti metodologici e primi risultati, In: (a cura di): *ALBERTO MANCO, AZZURRA MANCINI, Scritture brevi: segni, testi e contesti. Dalle iscrizioni antiche ai tweet. QUADERNI DI AION, Università degli studi di Napoli L'Orientale*, pp. 149-176, ISBN: 1825-2796.

### Articoli in Atti di Convegno:

7. Dominique Brunato, Lorenzo De Mattei, Felice Dell'Orletta, Benedetta Iavarone, Giulia Venturi (2018). Is this Sentence Difficult? Do you Agree?. In *Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)*, Association for Computational Linguistics, October 31–November 4, Brussels, Belgium, ISBN: 978-1-948087-84-1.
8. Pietro dell'Oglio, Dominique Brunato, Felice Dell'Orletta (2018). Lexicon and Syntax: Complexity across Genres and Language Varieties. In *Proceedings of 5th Italian Conference on Computational Linguistics (CLiC-it)*, 10-12 December, 2018, Turin, Italy.
9. Dominique Brunato, Martina Valeriani, Felice Dell'Orletta (2018). DARC-IT: a DATaset for Reading Comprehension in ITALian. In *Proceedings of 5th Italian Conference on Computational Linguistics (CLiC-it)*, 10-12 December, 2018, Turin, Italy.
10. Eleonora Cocciu, Dominique Brunato, Giulia Venturi, Felice Dell'Orletta. (2018). Gender and Genre Linguistic profiling: a case study on female and male journalistic and diary prose. In *Proceedings of 5th Italian Conference on Computational Linguistics (CLiC-it)*, 10-12 December, 2018, Turin, Italy.

11. Dominique Brunato, Felice Dell'Orletta (2017). On the order of words in Italian: a study on genre vs complexity. In *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, vol. 1749, p. 25-31, Linköping University Electronic Press, ISBN: 978-91-7685-467-9.
12. Giulia Pieri, Dominique Brunato, Felice Dell'Orletta. Studio sull'ordinamento delle strutture grammaticali nel confronto tra generi e complessità. In *Proceedings of the Third Italian Conference on Computational Linguistics (CLiC-IT 2016)*, Napoli, 5–6 dicembre 2016, ISSN: 1613-0073.
13. Dominique Brunato, Andrea Cimino, Felice Dell'Orletta e Giulia Venturi. (2016). PaCCSS–IT: A Parallel Corpus of Complex–Simple Sentences for Automatic Text Simplification, In: *Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, November 1–5, 2016, Austin, Texas, USA, ISBN: 978-1-937284-96-1.
14. Dominique Brunato e Felice Dell'Orletta. (2015). Isacco: a corpus for investigating spoken and written language development in Italian school-age children. In *Proceedings of the Second Italian Conference on Computational Linguistics (CLiC-IT 2015)*, Trento, 3-4 dicembre 2015, pp. 61–65, ISBN: 978-88-99200-62-6.
15. Dominique Brunato, Felice Dell'Orletta, Giulia Venturi e Simonetta Montemagni. (2015). Design and Annotation of the First Italian Corpus for Text Simplification. In: *Proceedings of the NAACL 9th Linguistic Annotation Workshop (LAW'15)*, 5 June, Denver, Colorado, USA, pp. 31–34, ISBN: 978-1-941643-47-1.
16. Brunato Dominique. (2014). Complessità necessaria o stereotipi del “burocratese”? Un'indagine sulla leggibilità del linguaggio amministrativo da una prospettiva linguistico-computazionale, in *La lingua variabile nei testi letterari, artistici e funzionali contemporanei: analisi, interpretazione, traduzione: atti del XII Congresso SILFI: riassunti e testi in CD*, Palermo: Centro di studi filologici e linguistici siciliani, Dipartimento di scienze umanistiche, Università degli studi di Palermo, 2014, ISBN: 978-88-96312-56-8.
17. Dominique Brunato, Felice Dell'Orletta, Giulia Venturi e Simonetta Montemagni. (2014). Defining an annotation scheme with a view to automatic text simplification, in *Proceedings of the First Italian Conference on Computational Linguistics (CLiC-it 2014)*, Pisa, 9-11 dicembre 2014, pp. 87–92, ISBN:978-8-86741-472-7.  
*Prefazione in Atti di Congresso:*
18. Dominique Brunato, Felice Dell'Orletta, Giulia Venturi, Thomas François, Philippe Blache (2016). Preface of Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC 2016). In: *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC 2016)*, co-located with the 26th International Conference on Computational Linguistics (COLING 2016), 11-16 December 2016, Osaka (Japan).

*Rapporti tecnici:*

19. UBImol-D3.2: Deliverable di progetto regionale dal titolo “Auto-valutazione automatica assistita del contenuto”, 01/10/2018. Descrizione: Il deliverable descrive approfonditamente le attività di sviluppo di una metodologia basata su tecnologie per l'estrazione di informazione dal testo, a supporto dell'auto-valutazione automatica assistita del contenuto delle risposte fornite da studenti attraverso la piattaforma di apprendimento; queste attività sono state realizzate come parte dell'obiettivo operativo 3, attività 2, nell'ambito del progetto regionale “UBIMOL – Ubiquitous Massive Open Learning”.
20. UBImol-D3.3: Deliverable di progetto regionale dal titolo “Auto-valutazione automatica assistita della forma linguistica”, 01/10/2018. Descrizione: Il deliverable descrive approfonditamente le attività di sviluppo di una metodologia basata su tecniche di monitoraggio linguistico delle produzioni scritte degli studenti, finalizzata a fornire un supporto all'auto-valutazione automatica assistita delle competenze linguistiche; queste attività sono state realizzate come parte dell'obiettivo operativo 3, attività 3, nell'ambito del progetto regionale “UBIMOL – Ubiquitous Massive Open Learning”.
21. UBImol-D3.1: Deliverable di progetto regionale dal titolo “Analisi della leggibilità e supporto per l'adattamento dei testi al dispositivo di lettura”, 8/01/2018. Descrizione: Il deliverable descrive approfonditamente le attività di adattamento di uno strumento di valutazione automatica della leggibilità rispetto al dispositivo da cui viene fruito un testo, realizzate come parte dell'obiettivo operativo 3, attività 1, nell'ambito del progetto regionale “UBIMOL – Ubiquitous Massive Open Learning”.
22. ISLE-D3.2: Deliverable per il progetto “iSLe” (Intelligent Semantic Liquid eBook) dal titolo “Ontology learning”, 31/01/2014; Descrizione: Il deliverable descrive le attività di sviluppo e specializzazione dei componenti Trattamento Automatico del Linguaggio per l'estrazione di informazione da testi del dominio di interesse progettuale, realizzate come parte dell'obiettivo operativo 3, attività 2, nell'ambito del progetto regionale “iSLe” (Intelligent Semantic Liquid eBook).
23. ISLE-D3.3: Deliverable per il progetto dal titolo “Risorse linguistiche”, 31/01/2014. Descrizione: Il deliverable descrive approfonditamente le attività di sviluppo e specializzazione della catena di analisi linguistica automatica dei testi, realizzate come parte dell'obiettivo operativo 3, attività 3, nell'ambito del progetto regionale “iSLe” (Intelligent Semantic Liquid eBook).
24. ISLE-D3.4: Deliverable per il progetto “iSLe” (Intelligent Semantic Liquid eBook) dal titolo “Metodologie e tecniche per la valutazione della leggibilità”, 31/07/2014. Descrizione: Il deliverable descrive approfonditamente le attività di adattamento di uno strumento di valutazione automatica della leggibilità dei testi, realizzate come parte dell'obiettivo operativo 3, attività 4, nell'ambito del progetto regionale “iSLe” (Intelligent Semantic Liquid eBook).

Pisa, 05 dicembre 2018

In fede,  
Dominique Brunato