

Andrea Amelio Ravelli

Curriculum vitae et studiorum

ESPERIENZE DI RICERCA	
03/2023 - in corso	Ruolo: Assegno di ricerca post-dottorale (durata: 12 mesi) Istituzione: Alma Mater Studiorum Università di Bologna Tematica: “Word Specificity and Text Readability”, all’interno del Progetto ABSTRACTION (ERC-2021-STG- 101039777) Area CUN: 10 Link progetto: https://site.unibo.it/abstraction Responsabile scientifico: Prof.ssa Marianna Marcella Bolognesi Studio dell’influenza di specificità e concretezza, intese come caratteristiche semantiche intrinseche di concetti e parole, sulla leggibilità dei testi, in prospettiva multilingue (IT, EN). Costruzione di corpora e dataset per il confronto tra domini linguistici e generi testuali lungo le assi di riferimento della specificità e della concretezza (e della loro reciproca influenza), creazione di modelli computazionali e validazione con modelli cognitivi.
06/2021- 02/2023	Ruolo: Assegno di ricerca post-dottorale (durata: 12 mesi + rinnovo 12 mesi) Istituzione: Istituto di Linguistica Computazionale “A. Zampolli” (ILC-CNR) Pisa Tematica: “New strategies for promoting attendance in colorectal cancer screening programmes of Tuscany (BESTcc)” e “Trattamento automatico della lingua ed estrazione di informazione” Link progetto: https://www.cnr.it/it/progetti-di-ricerca/progetto/41648 Responsabile scientifico: Dott.ssa Giulia Venturi Adattamento al dominio specialistico di ambito medico di sistemi di valutazione della leggibilità dei testi. Analisi automatica della leggibilità di lettere di invito allo screening per il cancro colon-rettale. Raccolta giudizi umani sulla leggibilità di lettere di invito allo screening, variate in base a norme di riscrittura specifiche, e validazione dei nuovi sistemi di analisi adattati al dominio medico. Prosecuzione delle ricerche precedenti sulla previsione multimodale del potenziale di coinvolgimento nella comunicazione verbale.

06/2020-06/2021	Ruolo: Assegno di ricerca post-dottorale (durata: 12 mesi) Istituzione: Istituto di Linguistica Computazionale “A. Zampolli” (ILC-CNR) Pisa Tematica: “Definizione di metodologie di Trattamento Automatico del Linguaggio per la rappresentazione multimodale della conoscenza” Responsabile scientifico: Dott.ssa Giulia Venturi Costruzione di un corpus e un dataset multimodale (testo, audio, video), a partire da registrazioni di visite guidate in siti di interesse culturale, annotato manualmente con informazione affettiva (attenzione ed engagement dei partecipanti) finalizzato alla modellazione, l’analisi e la previsione del coinvolgimento stimolato dalla guida turistica. Estrazione di caratteristiche esplicite specifiche per modalità e loro combinazione al fine di addestrare un classificatore automatico di espressioni linguistiche (considerate nell’insieme delle informazioni linguistiche, visuali e acustiche che compongono l’atto linguistico in contesto comunicativo faccia a faccia) in base alla portata del livello di coinvolgimento che queste possono generare nell’ascoltatore.
-----------------	---

TITOLI DI STUDIO	
10/2016-10/2019	Dottorato di Ricerca in: Filologia, Letteratura Italiana, Linguistica Curriculum: Linguistica (Internazionale) Settore Scientifico Disciplinare: L-LIN/01 Area CUN: 10 Conseguito presso: UNIVERSITÀ DEGLI STUDI DI FIRENZE Borsa di studio interdipartimentale (Dipartimento di Lettere e Filosofia, Dipartimento di Ingegneria dell’Informazione) Data di discussione: 30/04/2020 Giudizio: “<i>Excellent cum laude</i>” – menzione di “<i>Doctor Europaeus</i>” Titolo della Tesi: “<i>Annotation of linguistically derived action concepts in computer vision datasets</i>” Supervisor: Massimo Moneglia, Lorenzo Seidenari La Tesi riporta una serie di esperimenti, condotti nel corso del Dottorato di Ricerca, focalizzati sullo sfruttamento dell’ontologia multilingue e multimodale dei verbi d’azione IMAGACT in task di Trattamento Automatico del Linguaggio Naturale. Tra questi, particolare attenzione è stata riservata alla descrizione di un esperimento di annotazione automatica, attraverso tecniche di semantic similarity, dei concetti azionali proposti dalla risorsa all’interno del dataset LSMDC (Large Scale Movie Description Challenge), col duplice scopo di popolare l’ontologia IMAGACT con nuove istanze multimodali (video e caption) e di arricchire semanticamente le occorrenze di verbi d’azione nel dataset LSMDC con i concetti azionali corrispondenti.

09/2012-02/2016	Laurea Magistrale in: Filologia Moderna Curriculum: Linguistica Italiana e Informatica Umanistica Classe di Laurea: LM-14 Conseguita presso: UNIVERSITÀ DEGLI STUDI DI FIRENZE Data di discussione: 25/02/2016 Votazione: 110/110 e lode Titolo della Tesi: <i>“La disambiguazione del riferimento dei verbi azionali attraverso la risorsa multilingue IMAGACT: prospettive di sfruttamento e integrazione dei dati”</i> Supervisor: Alessandro Panunzi, Massimo Moneglia La Tesi ha come obiettivo la revisione integrale del corpus parallelo italiano-inglese delle standardizzazioni, e relative traduzioni, delle occorrenze estratte dai corpora di riferimento per la costruzione della risorsa IMAGACT. Partendo da un'introduzione alla risorsa e un'iniziale esposizione della metodologia elaborata per la revisione, sono descritte nel dettaglio le tipologie di segnalazioni registrate ed esposte riflessioni sui dati e proposte di intervento. Concludono la Tesi alcuni esperimenti di sfruttamento della risorsa in task di traduzione automatica e disambiguazione.
09/2007-07/2012	Laurea Triennale in: Lettere Moderne Curriculum: Linguistico Classe di Laurea: L-10 Conseguita presso: UNIVERSITÀ DEGLI STUDI DI FIRENZE Data di discussione: 05/07/2012 Votazione: 108/110 Titolo della Tesi: <i>“La lingua in cattedra: riflessioni sociolinguistiche sul parlato accademico”</i> Supervisor: Marco Biffi, Angela Frati La Tesi presenta la raccolta di un corpus di italiano parlato specifico del dominio del parlato accademico, con l'obiettivo di analizzare la variazione diafasica nel contesto delle lezioni accademiche. Il corpus si compone di trascrizioni di registrazioni effettuate presso l'Ateneo fiorentino, e sono state considerate 2 lezioni di docenti differenti (una integrale di ~2 ore, più un campione di ~20 min. per una seconda lezione) per i seguenti Dipartimenti: Agraria, Architettura, Economia, Lettere e Filosofia, Psicologia. È stato inoltre condotto un esperimento di riflessione autovalutativa sui soggetti del campionamento attraverso la somministrazione di una breve intervista, finalizzata a sondare il grado di consapevolezza della variazione diafasica della lingua.
09/2002-07/2007	Diploma: Maturità scientifica Conseguito presso: Liceo scientifico “Pier Paolo Pasolini” – Potenza Votazione: 75/100

ULTERIORI ESPERIENZE FORMATIVE	
29-31/05/2023	Partecipazione a “Lectures on Computational Linguistics” Ente organizzatore: Associazione Italiana di Linguistica Computazionale, Università di Pisa, Istituto di Linguistica Computazionale ILC-CNR (Pisa) Luogo: Università di Pisa
08-10/06/2022	Partecipazione a “Lectures on Computational Linguistics” Ente organizzatore: Associazione Italiana di Linguistica Computazionale, Università di Napoli “L’Orientale”, Università di Napoli “Federico II” Luogo: Università di Napoli “L’Orientale”
05-09/07/2021	Partecipazione a “Computer Vision Crash Course (CVCC)” Ente organizzatore: Università di Genova/MaLGa Luogo: Università di Genova (Online)
16-18/06/2021	Partecipazione a “Lectures on Computational Linguistics” Ente organizzatore: Associazione Italiana di Linguistica Computazionale, Università di Roma Tor Vergata Luogo: Università di Roma Tor Vergata (Online)
08-10/05/2019	Partecipazione a “Lectures on Computational Linguistics” Ente organizzatore: Associazione Italiana di Linguistica Computazionale, Università di Pavia, Fondazione Bruno Kessler Luogo: Università di Pavia
09/2018-12/2018	Periodo di studio all’estero (<i>visiting PhD</i>) presso IXA Taldea, Unità di Ricerca dell’Università Euskal Herriko Unibertsitatea di Donostia-San Sebastián (Spagna) Durante il periodo trascorso come ospite presso l’IXA Taldea (ora HiTZ), sotto la supervisione del dott. Oier Lopez de Lacalle e del prof. Eneko Agirre, mi sono occupato dello sviluppo degli esperimenti di annotazione dei concetti azionali di IMAGACT nel dataset LSMDC, confluiti nella stesura della Tesi di Dottorato. Durante la mia permanenza ho rafforzato e approfondito le mie conoscenze nel campo del Trattamento Automatico del Linguaggio Naturale, con particolare attenzione a tecniche di similarità semantica del testo e all’uso di modelli di rappresentazione vettoriale di parole e frasi.
04-06/09/2017	Partecipazione a “3rd Summer School on Integrating Vision and Language: Vision and Language Integration in Cognitive Robotics” Ente organizzatore: The European Network on Integrating Vision and Language (iV&L Net – ICT COST Action IC1307) Luogo: Cognitive Systems Research Institute (CSRI) - Atene (Grecia)
04-05/05/2017	Partecipazione a “Lectures on Computational Linguistics” Ente organizzatore: Associazione Italiana di Linguistica Computazionale, Università di Pavia, Fondazione Bruno Kessler Luogo: Università di Pavia

13-17/02/2017	Partecipazione a “Scuola Invernale TRIPLE (SIT 2017) - Espansioni del lessico. Nuove parole, nuovi formati, nuovi significati” Ente organizzatore / Luogo: Università degli Studi “Roma 3”
12/2014-03/2015	Tirocinio Universitario presso Dr. Wolf s.r.l. (Firenze) Attività: Revisione dati paralleli (italiano e inglese) della risorsa IMAGACT, sviluppo di un tagset specifico per la catalogazione dei problemi rilevati, annotazione errori di trascrizione e/o allineamento, revisione generale dei dati della risorsa

COMPETENZE LINGUISTICHE
Italiano (Madrelingua) Inglese (Comprensione: C1; Comunicazione orale: B2/C1; Scrittura: C1)

COMPETENZE INFORMATICHE
Sistemi operativi: Linux, macOS, Windows Programmi di produttività: Suite MS Office, MS Teams, LibreOffice/OpenOffice, PyCharm IDE, Visual Studio Code Linguaggi di programmazione/scripting: Python (ottimo), R (buono), Java (base), e relative librerie Linguaggi di markup: xml, html, TEI, LaTeX Database: MySQL, SQLite, neo4j/Cypher Corpora: SketchEngine, NoSketchEngine, CQL Software di annotazione e analisi linguistica e multimodale: ANVIL, ELAN, Praat, AntConc, WinPitch

ATTIVITÀ DIDATTICA – Supervisione tesi	
09/06/2022	Titolo: Studio e classificazione multimodale del coinvolgimento tramite l'implementazione di modelli di apprendimento automatico Candidato: Daniele Borghesi Relatore: Felice Dell'Orletta Correlatori: Andrea Amelio Ravelli, Alessandro Lenci Corso di Laurea: Informatica Umanistica (Triennale) Università: Università di Pisa Anno Accademico: 2020-2021

03/02/2022	Titolo: Predittori di attenzione e di categorie tematiche: uno studio sulla comunicazione in un corpus multimodale di visite turistiche guidate Candidata: Ludovica Binetti Relatore: Felice Dell’Orletta Correlatore: Andrea Amelio Ravelli Controrelatore: Mirko Tavosanis Corso di Laurea: Informatica Umanistica (Magistrale) Università: Università di Pisa Anno Accademico: 2020-2021
30/09/2021	Titolo: Multimodalità del coinvolgimento: creazione di un dataset ed esperimenti sull’engagement tra guida e visitatori in siti di interesse culturale Candidato: Federico Boggia Relatore: Felice Dell’Orletta Correlatori: Alessandro Lenci, Andrea Amelio Ravelli Corso di Laurea: Informatica Umanistica (Triennale) Università: Università di Pisa Anno Accademico: 2020-2021
04/02/2021	Titolo: Costruzione di un corpus multimodale per lo studio delle strategie di coinvolgimento di guide turistiche in siti culturali Candidato: Luca Poggianti Relatore: Felice Dell’Orletta Correlatori: Alessandro Lenci, Andrea Amelio Ravelli Corso di Laurea: Informatica Umanistica (Triennale) Università: Università di Pisa Anno Accademico: 2019-2020

ATTIVITÀ DIDATTICA – Tutoraggio tirocinanti	
05/2022-in corso	Tirocinante: Eleonora Cocciu (Corso di Laurea Magistrale in Informatica Umanistica del Dipartimento di Filologia, Letteratura e Linguistica dell'Università di Pisa) Progetto: Studio della coerenza nel discorso attraverso reti neurali e verifica competenze specifiche in modelli di lingua neurali Sede tirocinio: Istituto di Linguistica Computazionale “A. Zampolli” (ILC-CNR) Descrizione: La tirocinante si è occupata dello studio della coerenza all'interno di discorsi in lingua italiana da differenti domini (sedute parlamentari, trascrizioni di seminari) attraverso metodi computazionali. La tirocinante ha eseguito numerosi esperimenti volti a modellare la coerenza all'interno del discorso e verificare fino a che punto un modello di lingua neurale di recente generazione (Transformer, es. BERT) è in grado di riconoscere alterazioni a livello di coerenza locale (frase per frase) o globale (coerenza interna del documento intero).
09/2021-12/2021	Tirocinante: Daniele Borghesi (Corso di Laurea Triennale in Informatica Umanistica del Dipartimento di Filologia, Letteratura e Linguistica dell'Università di Pisa) Progetto: Studio e classificazione multimodale del coinvolgimento tramite l'implementazione di modelli di apprendimento automatico. Sede tirocinio: Istituto di Linguistica Computazionale “A. Zampolli” (ILC-CNR) Descrizione: Il tirocinante ha studiato e implementato diversi metodi di selezione multimodale di caratteristiche salienti (tra linguistiche e acustiche) all'interno di un dataset di visite guidate in siti di interesse culturale. Su questo dataset, composto dalle frasi pronunciate da una guida turistica e arricchito di caratteristiche linguistiche e acustiche, ha compiuto in seguito una serie di esperimenti per ottimizzare le prestazioni di sistemi di classificazione automatica, e valutare l'incremento nelle prestazioni in base al filtraggio di caratteristiche applicato.

03/2021-07/2021	Tirocinanti: Federico Boggia , Ludovica Binetti (Corso di Laurea Triennale e Magistrale in Informatica Umanistica del Dipartimento di Filologia, Letteratura e Linguistica dell’Università di Pisa – tirocinio congiunto) Progetto: Costruzione di un dataset multimodale per lo studio del coinvolgimento nel corso di visite guidate in siti di interesse culturale. Sede tirocinio: Istituto di Linguistica Computazionale “A. Zampolli” (ILC-CNR) Descrizione: I tirocinanti si sono occupati della validazione, revisione e segmentazione in unità frasali delle trascrizioni del parlato di una guida turistica nel contesto di visite guidate in siti di interesse culturale. Il tirocinio ha portato alla segmentazione dei file audio allineati alle trascrizioni al fine di ottenere, per ogni singola frase, una rappresentazione testuale e una acustica sotto forma di features esplicite e interpretabili specifiche per modalità, per poi sperimentare le prestazioni di un classificatore a seconda dell’uso di una singola modalità o della fusione di entrambi i livelli di rappresentazione.
06/2020-09/2020	Tirocinanti: Luca Poggianti, Mario Gomis (Corso di Laurea Triennale in Informatica Umanistica del Dipartimento di Filologia, Letteratura e Linguistica dell’Università di Pisa) Progetto: Annotazione del livello di attenzione ed engagement in corpora multimodali. Sede tirocinio: Istituto di Linguistica Computazionale “A. Zampolli” (ILC-CNR) Descrizione: I tirocinanti hanno annotato manualmente le variazioni nel livello di attenzione nei partecipanti a visite guidate sui video di una porzione del dataset del progetto “Cultural Heritage Resources Orienting Multimodal Experiences” (CHROME). In seguito, hanno eseguito un monitoraggio linguistico sui testi allineati ai video, correlando l’andamento dell’attenzione del pubblico alle caratteristiche linguistiche della guida turistica.

SEMINARI SU INVITO	
07/07/2022	“Caratterizzazione multimodale della comunicazione: presentazione di un nuovo dataset e primi esperimenti”, nella serie “ILC Seminars” Luogo: Istituto di Linguistica Computazionale “A. Zampolli” ILC-CNR (Pisa)
21/06/2019	“Tra visione e linguaggio: classificazione semantica di azioni in un dataset per Computer Vision” Luogo: Università degli Studi di Firenze

19/09/2018	“Semantic action discrimination at the intersection between Vision and Language”, nella serie “IXA NLP Group Seminars” Luogo: Euskal Herriko Unibertsitatea (Spagna)
29/03/2018	“Risorse linguistiche e semantica computazionale”, parte del Workshop “Lingua e Computer – Esperienze della Ricerca in Elaborazione del Linguaggio Naturale” Luogo: Università degli Studi della Basilicata
02/03/2016	“Sfruttamento e integrazione dei dati dell’ontologia IMAGACT per la disambiguazione e per la traduzione automatica”, nella serie “Seminari LABLITA di Linguistica” Luogo: Università degli Studi di Firenze
20/03/2015	“Corpus IMAGACT e Machine Translation: creazione di un dataset italiano-inglese”, nella serie “Seminari LABLITA di Linguistica” Luogo: Università degli Studi di Firenze

RESPONSABILITÀ SCIENTIFICHE

01/2017-04/2020	Ruolo: Assistente editoriale di Rivista Scientifica Rivista: “CHIMERA: Romance Corpora and Linguistic Studies” ISSN: 2386-2629 https://revistas.uam.es/chimera
------------------------	---

ASSOCIAZIONI SCIENTIFICHE

Associazione Italiana di Linguistica Computazionale (AILC) - membro Associazione Italiana per l’Intelligenza Artificiale (AIxIA) – membro
--

ATTIVITÀ DI REFERAGGIO PER CONFERENZE INTERNAZIONALI

Annual Meeting of the Association for Computational Linguistics (ACL 2023) Global WordNet Conference (GWC 2022) Italian Conference of Computational Linguistics (CLiC-it 2020, 2021) Annual Meeting of the Association for Computational Linguistics (ACL 2021) Natural Language for Artificial Intelligence Workshop (NL4AI 2021, 2022) Multimodal WordNet Workshop (MMWN 2020) International Conference on Language Resources and Evaluation (LREC 2020, 2022) Contextual Representation of Events and Objects in Language Workshop (CREOL 2019)

ATTIVITÀ DI REFERAGGIO PER RIVISTE SCIENTIFICHE

LRE: Language, Resources and Evaluation (ISSN: 1574-0218)

CHIMERA: Romance Corpora and Linguistic Studies (ISSN: 2386-2629)

OSNM: Online Social Networks and Media (ISSN: 24686964)

ORGANIZZAZIONE DI EVENTI

07-08/09/2023	“Assessing Discourse Coherence in Italian Texts – DiScoTex” (Task@EVALITA2023) Ruolo: Co-organizzatore Luogo: Parma Sito web: https://sites.google.com/view/discotex/
30/11-03/12/2020	“Concreteness in Context – CONcreTEXT” (Task@EVALITA2020) Ruolo: Co-organizzatore Luogo: Bologna Sito web: https://lablita.github.io/CONcreTEXT/
23-25/09/2019	Workshop “Contextual Representation of Events and Objects in Language – CREOL@JOWO2019” Ruolo: Co-organizzatore, Session Chair Luogo: Medical University of Graz (Austria) Sito web: https://creol2019.di.unito.it
25-26/06/2019	Workshop “Language technologies for research in Digital Humanities: an opportunity for linguists – Oier Lopez de Lacalle (Euskal Herriko Unibertsitatea – IXA Taldea)”, nella serie “Seminari LABLITA di Linguistica” Ruolo: Organizzatore Luogo: Università degli Studi di Firenze
09/06/2017	Workshop “Linking Data, Ontologies and Distributional Models for the Representation of Lexical Meaning” Ruolo: Co-organizzatore Luogo: Università degli Studi di Firenze