
Fabio Tamburini

nato a Bologna il 18/6/1968

Piazza Bonazzi, 5, 40133, Bologna

Tel: 051 560577

fabio.tamburini@unibo.it

<http://corpora.ficlit.unibo.it/People/Tamburini/>

Profilo Professore Associato Confermato (sett. disc. L-LIN/01)

Consegue la Laurea in Scienze dell'Informazione *cum laude* nel 1992 applicando modelli cognitivi basati su reti neurali al riconoscimento di pattern. Inizia la sua attività presso il Centro Interfacoltà di Linguistica Teorica e Applicata «L. Heilmann» (CILTA), Università di Bologna. Nel 2004 consegue il Dottorato di Ricerca in Ingegneria Elettronica, Informatica e Telecomunicazioni con un lavoro di analisi linguistica e computazionale in ambito fonetico/acustico e prosodico, in particolare riguardante il riconoscimento automatico della prominenza prosodica nella lingua parlata.

Attualmente è professore associato presso la Scuola di Lettere e Beni Culturali dell'Università di Bologna per la quale ha tenuto negli ultimi anni numerosi corsi, sia entro lauree triennali che magistrali, prevalentemente toccando argomenti di Linguistica Computazionale, Linguistica dei Corpora e Linguistica Applicata. I suoi principali interessi di ricerca vanno dal Trattamento Automatico delle Lingue alla creazione di Risorse Linguistiche.

Affiliazione Dipartimento di Filologia Classica e Italianistica – Alma Mater Studiorum, Università di Bologna, Via Zamboni, 32, 40126, Bologna

Formazione

- 1992:** Consegue la Laurea in *Scienze dell'Informazione*, in data 21/7/1992, presso la Facoltà di Scienze MM. FF. NN. dell'Università di Bologna, con la tesi "Studio e realizzazione di un sistema a reti neurali per il riconoscimento di immagini" e votazione 110/110 e lode. Relatori: Prof. Dante Bollini, dr. Renzo Davoli.
- 1997:** Partecipa a un corso sulla *Corpus Linguistics* presso il *Tuscan Word Centre* del prof. John Sinclair, nell'ambito del progetto *Bononia Legal Corpus* (BoLC).
- 1998:** Partecipa alla "7a Scuola estiva di calcolo parallelo" presso l'"*High Performance System Division*" del CINECA di Bologna.
- 2001:** Partecipa alla "*13th European Summer School in Logic, Language and Information*", Università di Helsinki.
- 2003:** Partecipa, nell'ambito del progetto FIRB2001, al ciclo di seminari *CTL@CILTA* tenuto dal prof. Michael Moortgat (per un totale di 55 ore). I seminari hanno affrontato problematiche relative all'applicazione della *Categorial Type logic* al linguaggio naturale, in particolare rispetto agli aspetti riguardanti l'analisi sintattica e semantica.
- 2004:** Consegue il titolo di Dottore di Ricerca in Ingegneria Elettronica e Informatica, in data 13/5/2004, presso la Facoltà di Ingegneria dell'Università degli Studi di Bologna con la tesi "Identificazione automatica della prominenza prosodica nella lingua parlata". Relatore: prof. Giovanni Emanuele Corazza.

Posizione Accademica

- 1992 - 97:** Contratto di collaborazione presso il Centro Interfacoltà di Linguistica Teorica e Applicata «L. Heilmann», Università di Bologna.
- 1997:** Presa di servizio, nel gennaio 1997, come *Collaboratore Tecnico*, 7° *qualifica funzionale* presso il Centro Interfacoltà di Linguistica Teorica e Applicata «L. Heilmann», Università di Bologna, come vincitore del relativo Concorso Pubblico.
- 2000:** Presa di servizio, nel maggio 2000, come *Funzionario di Elaborazione Dati*, 8° *qualifica funzionale* presso il Centro Interfacoltà di Linguistica Teorica e Applicata «L. Heilmann», Università degli Studi di Bologna, dopo il relativo Concorso Pubblico.
- 2001:** Presa di servizio come Ricercatore (SSD L-LIN/01, ex L09A), in data 3/4/2001, presso la Scuola Superiore di Lingue Moderne per Interpreti e Traduttori (per le esigenze del CILTA, Bologna), Università degli Studi di Bologna, dopo il relativo Concorso Pubblico.
- 2004:** Trasferimento alla Facoltà di Lettere e Filosofia dell'Università degli Studi di Bologna, in data 1/2/2004.
- 2004:** Conferma nel ruolo di Ricercatore Universitario in data 3/4/2004.

- 2011:** Presa di servizio come Professore Associato (SSD L-LIN/01), in data 1/3/2011, presso la Facoltà di Lettere e Filosofia, Università degli Studi di Bologna, dopo essere risultato vincitore della relativa valutazione comparativa (bando II tornata 2008).
- 2014:** Conferma nel ruolo di Professore Associato in data 2/3/2014.
- 2018:** Consegue l'Abilitazione Scientifica Nazionale alla I Fascia per il Sett. Conc. 10/G1 (dal 1/8/2018 al 1/8/2024).

Attività Didattica

A.A. 2000/01	Seminari di "Fondamenti di Linguistica Informatica" (12h) all'interno del corso di Linguistica Applicata della prof.ssa R. Rossini Favretti, Facoltà di Lettere e Filosofia dell'Università di Bologna.
A.A. 2001/02	Titolare, per affidamento d'incarico, del corso di "Informatica Applicata" (36h) presso il Corso di Laurea in Informatica, Facoltà di Scienze MM.FF.NN. dell'Università di Bologna. (Il programma ha riguardato argomenti di Linguistica Computazionale)
A.A. 2002/03 ... A.A. 2018/19	Titolare, per affidamento d'incarico, del corso di "Linguistica Computazionale" (6 CFU, 30h) presso il Corso di Laurea Specialistica in Informatica, Facoltà di Scienze MM.FF.NN. dell'Università di Bologna. Nell'A.A. 2009/10 il corso è stato rinominato "Elaborazione del Linguaggio Naturale" , mantenendo contenuti simili, e dall'A.A.2017/18 è mutuato da Text Retrieval, Analysis and Mining (LM) del Corso di Studio Magistrale in <i>Digital Humanities and Digital Knowledge</i> .
A.A. 2003/04 ... A.A. 2009/10	Titolare, per affidamento d'incarico, del corso di "Linguistica Informatica" (5 CFU, 30h) presso il Corso di Laurea in Lettere, curriculum moderno: linguistico, Facoltà di Lettere e Filosofia dell'Università di Bologna.
A.A. 2005/06 ... A.A. 2011/12	Titolare del corso di "Linguistica Computazionale" (5/6 CFU, 30h) presso il Corso di Laurea in Lettere, Facoltà di Lettere e Filosofia dell'Università di Bologna.
A.A. 2005/06 ... A.A. 2007/08	Titolare, per affidamento d'incarico, del corso di "Linguistica Informatica (LS)" (5 CFU, 30h) presso il Corso di Laurea Specialistica in Linguistica, Facoltà di Lettere e Filosofia e Facoltà di Lingue e Letterature Straniere dell'Università di Bologna.
A.A. 2005/06 ... A.A. 2007/08	Titolare, per affidamento d'incarico, del corso di "Linguistica Computazionale (LS)" (5 CFU, 30h) presso il Corso di Laurea Specialistica in Linguistica, Facoltà di Lettere e Filosofia e Facoltà di Lingue e Letterature Straniere dell'Università di Bologna.
A.A. 2008/09 ... A.A. 2018/19	Titolare del corso di "Linguistica Computazionale (LM)" (6 CFU, 30h) presso il Corso di Laurea Magistrale in Italianistica, Culture Letterarie Europee, Scienze Linguistiche, Scuola di Lettere e Beni Culturali dell'Università di Bologna. Il corso è mutuato anche dalla Scuola di Lingue e Letterature, Traduzione e Interpretazione del medesimo Ateneo.
Dic. 2009	Ha tenuto un seminario presso Il Dip. di Linguistica, Università di Pavia nell'ambito del Dottorato di ricerca in Linguistica su "Il contributo delle nuove tecnologie nell'analisi della lingua parlata".
A.A. 2010/11 ... A.A. 2011/12	Titolare del corso di "Linguistica Applicata" (6 CFU, 30h) presso il Corso di Laurea in Lettere, Facoltà di Lettere e Filosofia dell'Università di Bologna. Il corso è mutuato anche dalla Facoltà di Lingue del medesimo Ateneo.
Giu. 2011	Ha tenuto varie lezioni e laboratori all'Erasmus Intensive Program <i>"Natural Language Processing: Techniques, Applications and Challenges - NaLaProTAC"</i> tenutosi presso l'Università della Western Macedonia, http://nalaprotac.ict.e.uowm.gr . Il progetto è nato da un consorzio di Atenei europei: Università di Bologna, Università di Coimbra, Università di Cyprus, Università di Groningen, Università di Ioannina, Università di Primorska, Università di Toulouse 2 Le Mirail, Università di York.

Gen. 2012	Ha tenuto un ciclo di lezioni e laboratori (in coll. con Malvina Nissim) all'interno della Scuola Invernale TRIPLE (SIT2012) "Classi di parole: modelli, tipologia, computazione", tenutasi presso l'Università di Roma3,
A.A. 2012/13	Titolare del corso di " Linguistica Applicata " (12 CFU, 60h) presso il Corso di Laurea in ... Lettere, Scuola di Lettere e Beni Culturali dell'Università di Bologna. Il corso è mutuato
A.A. 2018/19	anche dalla Facoltà di Lingue del medesimo Ateneo.
A.A. 2012/13	Titolare del corso di " Linguistica Informatica (LM) " (6 CFU, 30h) presso il Corso di ... Laurea Magistrale in Italianistica, Culture Letterarie Europee, Scienze Linguistiche,
A.A. 2018/19	Scuola di Lettere e Beni Culturali dell'Università di Bologna.
Mag. 2017	Ha tenuto il seminario " <i>Morphological Analysis, PoS-Tagging and Lemmatisation</i> ", nell'ambito delle "Lectures on Computational Linguistics" organizzate dal Corso di Dottorato in Scienze linguistiche delle Università di Pavia/Bergamo, in collaborazione con la Fondazione Bruno Kessler.
A.A. 2018/19	Titolare del corso di " Text Retrieval, Analysis and Mining (LM) " (6 CFU, 30h) presso il Corso di Studio Magistrale in <i>Digital Humanities and Digital Knowledge</i> , Scuola di Lettere e Beni Culturali dell'Università di Bologna.

Tutti i corsi istituzionali sono stati integrati dalla progettazione e realizzazione di siti web e CD-ROM specifici, contenenti i materiali didattici e i principali strumenti tecnologici del settore, atti a supportare lo sviluppo di risorse linguistiche etichettate in laboratorio e a sperimentare le più avanzate tecnologie nel campo del trattamento automatico delle lingue.

Altri incarichi accademici

- 2017/ - Presidente della Commissione Tirocini per il CdS in Digital Humanities and Digital Knowledge (LM43) - Università di Bologna.
- 2015/ - Responsabile del Curriculum in Scienze Linguistiche della Laurea Magistrale in Italianistica, Culture Letterarie Europee, Scienze Linguistiche (LM14) e membro della commissione per la revisione del Corso di Studio.
- 2015/ - Presidente della Commissione Tirocini per il CdS in Lettere (L10) - Università di Bologna.
- 2013/ - Membro del comitato scientifico della biblioteca del FICLIT - Università di Bologna.
- 2015/2018 - Membro del Direttivo dell'Associazione Italiana di Linguistica Computazionale.
- 2015/2018 - Membro della Giunta del Dipartimento di Filologia Classica e Italianistica, FICLIT, Università di Bologna.
- 2008/2012 - Membro della Giunta del Dipartimento di Studi Linguistici e Orientali, Università di Bologna.
- 2008/2012 - Coordinatore della sezione di Linguistica del Dipartimento di Studi Linguistici e Orientali, Università di Bologna.
- 2007/2012 - Membro delle commissioni per la revisione dei Corsi di Studio (DM270/2004) della Laurea in Lettere e del Curriculum Magistrale in Scienze Linguistiche.
- 2003/2006 - Membro del comitato scientifico della biblioteca del CILTA - Università di Bologna.
- 1998/2006 - Coordinamento scientifico della *Sezione di Linguistica computazionale, Informatica umanistica e Comunicazione* (SLIC) costituita presso il CILTA - Università di Bologna.

Attività professionali e di consulenza

- 2019 - Lezione BBS/Ferrari
- 2008/2011 - Consulente presso il CINECA nell'ambito del progetto di ricerca PAPHYRUS - *Cultural and historical digital libraries dynamically mined from news archives* finanziato nell'ambito del 7° programma quadro dell'Unione Europea (<http://www.ict-papyrus.eu>). Il progetto si pone come obiettivo la creazione di biblioteche digitali attraverso l'analisi automatica di *news* di tipo testuale, parlate e audiovisive. La complessità delle operazioni di estrazione dell'informazione da tali media e le elaborazioni necessarie ad una loro corretta gestione, richiedono l'applicazione delle più innovative tecnologie sviluppate nell'ambito della linguistica computazionale.
- 2004/2005 - Collaborazione con l'azienda EXPERTSYSTEM per lo scambio e lo sviluppo di risorse linguistiche.

Attività di ricerca

Nella sua attività di ricerca ha esplorato un'area di studio fortemente interdisciplinare quale la linguistica computazionale, che coinvolge ambiti disciplinari diversi, tra cui la linguistica, l'informatica, la matematica e la statistica, l'ingegneria del suono, ecc., concentrandosi prevalentemente sugli aspetti modellistico-tecnologici della disciplina.

Ha coordinato, nell'ambito di vari progetti, gli aspetti di innovazione progettuale legati all'applicazione delle tecnologie informatiche ai vari ambiti della ricerca linguistica, in particolare alla linguistica dei *corpora*, al trattamento automatico delle lingue, sia scritte che parlate, e alla didattica delle lingue in modalità *e-learning* o *blended learning*.

Principali linee di ricerca

ASPETTI PROSODICI DELLA LINGUA PARLATA

Identificazione automatica della prominenzia frasale nella lingua parlata, anche in una prospettiva interlinguistica e tipologica. Individuazione della struttura informativa degli enunciati sia tramite metodologie fonetico/acustiche sul parlato sia utilizzando metodologie proprie dell'analisi testuale computazionale sulla trascrizione.

ANNOTAZIONE MORFOLOGICA DI GRANDI CORPORA E STRUMENTI AUTOMATICI PER L'ANALISI MORFOLOGICA

Recentemente ha iniziato un nuovo progetto volto all'annotazione morfologica del corpus CORIS/CODIS ovvero all'inserimento automatico di opportune annotazioni in grado di consentire ricerche mirate su singoli morfemi e/o su combinazioni di essi. A tal fine ha operato profonde modifiche al lemmario di un analizzatore/generatore morfologico, introducendo, tra le altre, le informazioni relative alla struttura morfologica dei lemmi. Questo progetto è stato inserito e finanziato nell'ambito dei bandi Galileo per la mobilità di studiosi tra Italia e Francia.

ANALISI DEL LINGUAGGIO PATOLOGICO

Negli ultimi anni si è occupato dell'identificazione precoce della malattia di Alzheimer mediante l'analisi computazionale della produzione spontanea dei pazienti. Il progetto si occupa dell'identificazione di opportuni indicatori linguistici in grado di marcare in maniera affidabile l'inizio della malattia e della loro misurazione con strumenti automatici.

Nello stesso ambito, si occupa dello sviluppo di sistemi di riconoscimento vocale per il parlato disartrico.

TECNICHE DI MACHINE LEARNING E DEEP NEURAL NETWORK

Dopo le recenti rivoluzioni tecnologiche nell'ambito del machine learning si occupa dello sviluppo di metodologie e algoritmi per il Deep Learning applicati all'analisi delle lingue naturali. Ha anche esplorato approcci decisamente innovativi che si basano sul calcolo della probabilità quantistica.

ANALISI SINTATTICA DELLA LINGUA CON STRUMENTI AUTOMATICI

Creazione, anche mediante procedure automatiche di learning, di una grammatica categoriale (*Categorical Type Logic Grammar*) per l'italiano; studio e realizzazione di strumenti automatici per l'analisi sintattica della lingua (*parsing*) nell'ambito dei modelli categoriali. Sviluppo di nuovi modelli formali per la gestione automatizzata dei fenomeni linguistici (*Unification Categorical Grammar*)

MODELLI SEMANTICI DISTRIBUZIONALI

Ha affrontato gli ambiti connessi con la costruzione di modelli geometrici per la gestione computazionale del significato – *Word Space* o *Semantic Vector Space* o *Word Embedding* – sia a livello lessicale che di frase esplorando i possibili strumenti matematici più idonei a rappresentare modelli semantici compositivi (prodotto tensore e algebre quantistiche).

STRUMENTI PER L'ANNOTAZIONE DI GRANDI CORPORA

Induzione automatica di un insieme di categorie lessicali per la lingua italiana, con le quali poter annotare i corpora rispetto alle parti del discorso; studio e implementazione di annotatori automatici per parti del discorso.

LINGUISTICA DEI CORPORA

Progettazione, costruzione e annotazione di grandi corpora.

Attività di coordinamento in progetti internazionali/nazionali di ricerca

2016: Ha partecipato al coordinamento del task relativo alla PoS-tagging di tweet in EVALITA 2016 (<http://evalita.fbk.eu/2016/>).

2011: Ha coordinato il task relativo alla Lemmatizzazione in EVALITA 2011 (<http://evalita.fbk.eu/2011/>).

2007: Ha coordinato il task relativo al PoS-tagging in EVALITA 2007 (<http://evalita.fbk.eu/2007/>), un'iniziativa per la valutazione di sistemi automatici per l'analisi della lingua italiana. Un numero rilevante di studiosi ha partecipato alla valutazione, sperimentando i vari sistemi sui dati forniti dagli organizzatori.

2005: Coordinamento linguistico-informatico del progetto AIR (Altair-In-Rete) finalizzato alla sperimentazione di ambienti e-learning per la didattica delle lingue a distanza in ambito universitario.

2003-2008: Ha coordinato un gruppo di giovani studiosi all'interno di un progetto FIRB2001 volto a) all'individuazione di un appropriato insieme di categorie lessicali per la lingua italiana, b) all'analisi sintattica della lingua con strumenti automatici e c) allo sviluppo di un corpus diacronico per l'italiano dall'Unità d'Italia alla fine della seconda guerra mondiale (in collaborazione con studiosi dell'Accademia della Crusca).

2000: Coordinamento linguistico-informatico del progetto FIRB finalizzato alla sperimentazione di ambienti e-learning nei cicli scolastici primario e secondario.

1998-2006: Coordinamento scientifico della Sezione di Linguistica computazionale, Informatica umanistica e Comunicazione (SLIC) costituita presso il CILTA.

1998: Coordinamento linguistico-informatico del progetto MISSILE – ideato e sviluppato presso il CILTA – volto all'elaborazione di materiali multimediali per l'apprendimento della lingua inglese rivolti ai militari di leva e ai volontari in ferma breve.

1997: Coordinamento linguistico-informatico dei progetti DIAPASON, finalizzato alla costruzione di percorsi didattici multimediali per gli studenti delle Facoltà scientifiche, e BoLC, volto alla costituzione ed all'analisi di corpora giuridici paralleli e confrontabili.

Partecipazione a progetti internazionali/nazionali di ricerca

- FIRB 2001 (RRF), FIRB 2009 (Modena)
- PRIN 2015 (Unibo Paolucci)
- PRIN 2017 (Unibo Grandi)
- ERC Inscribe (S.Ferrara), Scientific Advisor.

Collaborazioni con studiosi e Gruppi di ricerca

- John Sinclair, *Tuscan Word Centre*.
- Michael Moortgat, OTS – Università di Utrecht.
- Edoardo Lombardi Vallauri, Dipartimento di Linguistica, Università degli Studi Roma Tre.
- Francesco Sabatini, Domenico Proietti, Marco Biffi, Accademia della Crusca, Firenze.
- Rodolfo Delmonte, Dipartimento di Scienze del Linguaggio, Università "Cà Foscari" di Venezia.
- Petra Wagner, Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld, Germany.
- Matteo Capelletti, LIX - École Polytechnique, France.
- Dipartimento di "Information and Knowledge Management Services" del CINECA, Bologna.
- Johan Bos, Università "La Sapienza" di Roma.
- Fabio Montermini, CLLE-ERSS, CNRS Tolosa.
- Fabio Vitali, Università di Bologna.
- Cristina Bosco, Alessandro Mazzei, Viviana Patti, Università di Torino.

Attività di revisore/valutatore

Collane

- *Linguistica e Linguistiche*, Pàtron Editore

Riviste e volumi:

- *Italian Journal of Computational Linguistics*,
- *IEEE Transaction on Audio, Speech & Language Processing*,
- *International Journal of Artificial Intelligence*,
- *Computers & Education*,
- *Journal of Speech Sciences*,
- *Artificial Intelligence*.
- *PLOS-ONE*.

- *Cognitive Science*.
- *Studi e Saggi Linguistici*.
- *Rhapsodie: a Prosodic and Syntactic Treebank for Spoken French*, Benjamin's Studies in Corpus Linguistics.

Comitato scientifico dei seguenti convegni:

- *International Conference on Language Resources and Evaluation*:
 - o *LREC2008*, Marrakech; *LREC2010*, Malta; *LREC2012*, Istanbul; *LREC2014*, Reykjavik; *LREC2016*, Portorož; *LREC2018*, Miyazaki, Giappone.
- *Annual Conference of the International Speech Communication Association (ISCA)*:
 - o *InterSpeech2012*, Portland (OR); *InterSpeech2013*, Lyon; *InterSpeech2014*, Singapore; *InterSpeech2015*, Dresda; *InterSpeech2016*, San Francisco, CA; *InterSpeech2017*, Stockholm; *InterSpeech2018*, Hyderabad, India; *InterSpeech2019*,
- *IEEE International Conference on Acoustics, Speech and Signal Processing*:
 - o *ICASSP2018*, Calgary, Canada
- *International Joint Conference on Artificial Intelligence*:
 - o *IJCAI 2019*, Macao, China;
- *Italian Conference on Computational Linguistics*:
 - o *CLIC-it2014*, Pisa; *CLIC-it2015*, Trento (Area co-chair: *Morphology, Syntax and Parsing*); *CLIC-it2016*, Napoli; *CLIC-it2017*, Roma (Area Chair: *Spoken Language Processing and Automatic Speech Understanding*); *CLIC-it2018*, Torino (Program co-chair).
- *Prosodic Prominence: Perceptual and Automatic Identification (Speech Prosody 2010 workshop)*.
- *Speech Prosody 2018 Conference*, Poznań.
- *XXXVIII Convegno annuale della Società Italiana di Glottologia*, Bologna, 2013.
- *Convegno Nazionale dell'Associazione Italiana di Scienze della Voce*:
 - o *AISV2012*, Roma; *AISV2013*, Venezia; *AISV2015*, Bologna; *AISV2016*, Salerno; *AISV2017*, Pisa.
- *LangTech 2008*, Roma.
- Conferenza internazionale *pS-prominenceS*, Viterbo, 2013.

Premi

- Membro della giuria del premio Ferrero 2014 dell'Associazione Italiana di Scienze della Voce (con Salza P.L., Sigona, F.).
- Membro della giuria del premio Zampolli associato al Convegno TAL 2014, Torino (con Delmonte R., Aliprandi C.).

Progetti

- Nel 2004, è stato valutatore esterno per il progetto "TONES-2" presentato nell'ambito del bando "ESRC National Centre for E-social Science", coordinato dal prof. Kurshid Ahmad presso l'Università del Surrey, Guilford, UK.
- Nel 2012 è stato revisore esterno del progetto "Automatic annotation of prosody in speech corpora" presentato da Piet Mertens per essere ammesso a finanziamento presso la Katholieke Universiteit Leuven, Belgio.

Pubblicazioni principali

1. Basili R., Brambilla S., Croce D., Tamburini F. (2017). Developing a Large Scale FrameNet for Italian: the IFrameNet Experience In *Proc. of the Fourth Italian Conference on Computational Linguistics (CLiC-it 2017)*, Roma, 11-13 December 2017, 59-64, ISBN 978-88-99982-76-8, ISSN 1613-0073, scopus 2-s2.0-85037348145.
2. Tamburini F. (2017). Semgrex-Plus: a Tool for Automatic Dependency-Graph Rewriting. In *Proc. of the Fourth International Conference on Dependency Linguistics - Depling 2017*, Pisa, 18-20 September 2017, 248-254, ISBN 978-91-7685-467-9, ISSN 1650-3686.
3. Sanguinetti M., Bosco C., Mazzei A., Lavelli A., Tamburini F. (2017). Annotating Italian Social Media Texts in Universal Dependencies In *Proc. of the Fourth International Conference on Dependency Linguistics - Depling 2017*, Pisa, 18-20 September 2017, 229-239, ISBN 978-91-7685-467-9, ISSN 1650-3686.

4. Basile I., Tamburini F. (2017). Towards Quantum Language Models. In *Proc. of Conference on Empirical Methods in Natural Language Processing - EMNLP2017*, Copenhagen, 9-11 September 2017, 1841-1850, ISBN 978-1-945626-83-8.
5. Tamburini F. (2017). Quali risorse usa (o crea) il linguista? In Masini F., Grandi N. (eds.), *Tutto ciò che hai sempre voluto sapere sul linguaggio e sulle lingue*, Caissa: Cesena, 137-140, ISBN 978-88-6729-044-4.
6. Malavasi M., Turri E., Motolese M.R., Marxer R., Farwer J., Christensen H., Desideri L., Tamburini F., Green P. (2017). An Innovative Speech-Based Interface to Control AAL and IoT Solutions to Help People with Speech and Motor Disability. In Cavallo F. et al. (eds.) *Ambient Assisted Living, Lecture Notes in Electrical Engineering*, Springer International Publishing, 269-278, ISSN 1876-1100, ISBN 978-3-319-54282-9, scopus: 2-s2.0-85018685142, DOI 10.1007/978-3-319-54283-6_20.
7. D'Errico M., Grandi N., Paternes Meloni S., Tamburini F. (2016). Induzione di categorie grammaticali e lessicali. In F. Dedè (ed.) *Categorie grammaticali e classi di parole. Statuto e riflessi metalinguistici*, Roma: Il Calamo, 115-137, ISBN 9788898640188.
8. Tamburini F. (2016). (Better than) State-of-the-Art PoS-tagging for Italian Texts. In *Proc. 3rd Italian Conference on Computational Linguistics - CLiC-IT 2016*, Napoli, 5-6 December 2016, 280-284, ISBN 978-88-99982-08-9, ISSN 1613-0073.
9. Alfieri L., Tamburini F. (2016). (Almost) Automatic Conversion of the Venice Italian Treebank into the Merged Italian Dependency Treebank Format. In *Proc. 3rd Italian Conference on Computational Linguistics - CLiC-IT 2016*, Napoli, 5-6 December 2016, 19-23, ISBN 978-88-99982-08-9, ISSN 1613-0073.
10. Bosco C., Tamburini F., Bolioli A., Mazzei A. (2016). Overview of the EVALITA 2016 Part Of Speech on TWitter for ITALian Task. In *Proc. Fifth International Workshop EVALITA 2016*, Napoli, 7 December 2016, 78-84, ISBN 978-88-99982-08-9, ISSN 1613-0073.
11. Tamburini F. (2016). A BiLSTM-CRF PoS-tagger for Italian tweets using morphological information. In *Proc. Fifth International Workshop EVALITA 2016*, Napoli, 7 December 2016, 120-123, ISBN 978-88-99982-08-9, ISSN 1613-0073.
12. Green P., Marxer R., Cunningham S., Christensen H., Rudzicz F., Yancheva M., Coy A., Malavasi M., Desideri L., Tamburini F. (2016). CloudCAST - Remote Speech Technology for Speech Professionals. In *Proc. Interspeech 2016*, San Francisco, 8-12 September 2016, 1608-1612, ISSN 1990-9772, DOI: 10.21437/Interspeech.2016-148
13. Tamburini F. (2016). Specialising Paragraph Vectors for Text Polarity Detection. In *Proc. 10th International Conference on Language Resources and Evaluation - LREC 2016*, Portorož, 23-28 May 2016, ISBN 978-2-9517408-9-1.
14. Beltrami D., Calzà L., Gagliardi G., Ghidoni E., Marcello N., Rossini-Favretti R., Tamburini F. (2016). Automatic Identification of Mild Cognitive Impairment through the Analysis of Italian Spontaneous Speech Productions. In *Proc. 10th International Conference on Language Resources and Evaluation - LREC 2016*, Portorož, 23-28 May 2016, ISBN 978-2-9517408-9-1.
15. Vayra M., Avesani C., Tamburini F. (2015). *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio* (a cura di.), Milano:AISV, ISBN: 978-88-97657-11-8.
16. Beltrami D., Calzà L., Gagliardi G., Ghidoni E., Marcello N., Rossini Favretti R., Tamburini F. (2015). Strumenti di screening linguistico per l'identificazione precoce della fragilità cognitiva. In Vayra M., Avesani C., Tamburini F. (a cura di.), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio*, Milano:AISV, 329-344, ISBN: 978-88-97657-11-8, DOI: 10.17469/O2101AISV000020.
17. Beltrami D., Calzà L., Gagliardi G., Ghidoni E., Marcello N., Rossini-Favretti R., Tamburini F. (2015). Marche linguistiche come indicatori del decadimento cognitivo, *Nea Science*, 2: 29-35, ISSN 2282-6009.
18. Tamburini F. (2015). Automatic Detection of Prosodic Prominence by Means of Acoustic Analyses, *Lingue e Linguaggio*, XIV.1, 131-148, ISSN 1720-9331, doi:10.1418/80758, scopus: 2-s2.0-84942616813.
19. Calzà L., Beltrami D., Gagliardi G., Ghidoni E., Marcello N., Rossini-Favretti R., Tamburini F. (2015). Should we screen for cognitive decline and dementia? *Maturitas*, 82(1): 28-35, ISSN 0378-5122, doi:10.1016/j.maturitas.2015.05.013, scopus: 2-s2.0-84939566809.
20. Basile V., Peroni S., Tamburini F., Vitali F. (2015). Topical tags vs non-topical tags: Towards a bipartite classification? *Journal of Information Science*, 41(4): 486-505, ISSN 0165-5515, doi:10.1177/0165551515585283, scopus: 2-s2.0-84936989835.

21. Tamburini F. (2014). Are Quantum Classifiers Promising? In Proceedings of the *First Italian Conference on Computational Linguistics CLiC-it 2014*, Pisa, 360-364, ISBN 978-886741-472-7.
22. Di Gennaro P., Rossi A., Tamburini F. (2014). The FICLIT+CS@UniBO System at the EVALITA 2014 Sentiment Polarity Classification Task. In Proceedings of the *Fourth International Workshop EVALITA 2014*, Pisa, 93-97, ISBN 978-886741-472-7.
23. Tamburini F. (2014). Emotion Recognition with a Kernel Quantum Classifier. In Proceedings of the *Fourth International Workshop EVALITA 2014*, Pisa, 120-123, ISBN 978-886741-472-7.
24. Tamburini F., Bertini, C., Bertinetto, P.M. (2014). Prosodic prominence detection in Italian continuous speech using probabilistic graphical models. In Proc. of the *7th International Conference on Speech Prosody*, Science Foundation Ireland, 285-289, ISSN: 2333-2042, scopus: 2-s2.0-84904687334.
25. Grandi N., Nissim M., Tamburini F., Vayra M. (2014). La nozione di classico in linguistica. Atti del XXXVIII Convegno della Società Italiana di Glottologia (a cura di). In *Biblioteca della Società Italiana di Glottologia*, pp. 1-208, ISBN 9788898640034.
26. Tamburini F. (2013). Una valutazione oggettiva dei metodi più diffusi per l'estrazione automatica della frequenza fondamentale. In *Atti dell'IX Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV2013)*, Bulzoni:Roma, 427-434, ISBN: 978-88-7870-901-0.
27. Tamburini F. (2013). The Lemmatisation Task at the EVALITA 2011 Evaluation Campaign. In B. Magnini, F. Cutugno, M. Falcone and E. Pianta (eds.), *Evaluation of Natural Language and Speech Tools for Italian*, LNAI, Springer, 7689, 230-238, ISBN: 978-3-642-35827-2.
28. Tamburini F. (2013). The AnIta-Lemmatiser: a tool for accurate lemmatisation of Italian texts. In B. Magnini, F. Cutugno, M. Falcone and E. Pianta (eds.), *Evaluation of Natural Language and Speech Tools for Italian*, LNAI, Springer, 7689, 266-273, ISBN: 978-3-642-35827-2.
29. Lombardi Vallauri E., Tamburini F. (2012). Topic and Focus marking in an Italian corpus: some results of algorithmic measurements and structural interpretation. In H. Mello, M. Pettorino, T. Raso (eds), *Proc. of the VII GSCP International Conference: Speech and Corpora*, Belo Horizonte, Brazil, Firenze University Press, 191-196, ISBN: 978-88-665-5351-9.
30. Gagliardi G., Lombardi Vallauri E., Tamburini F. (2012). La prominente in italiano: demarcazione più che culminazione. In *Atti dell'VIII Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV2012)*, Bulzoni:Roma, 255-270, ISBN: 978-88-7870-774-0.
31. Wagner P., Tamburini F., Windmann A. (2012). Objective, Subjective and Linguistic Roads to Perceptual Prominence. How are they compared and why? In *Proc. 13th Annual Conference of the International Speech Communication Association (InterSpeech2012)*, Portland (OR), SS09.02, ISSN: 1990-9770, scopus: 2-s2.0-84878549887.
32. Tamburini F., Melandri M. (2012). AnIta: a powerful morphological analyser for Italian. In *Proc. 8th International Conference on Language Resources and Evaluation - LREC 2012*, Istanbul, 941-947, ISBN: 978-2-9517408-7-7.
33. Gagliardi G., Lombardi Vallauri E., Tamburini F. (2012). A topologic view of Topic and Focus marking in Italian. In *Proc. 8th International Conference on Language Resources and Evaluation - LREC 2012*, Istanbul, 948-955, ISBN: 978-2-9517408-7-7.
34. Tamburini F. (2012). The EVALITA 2011 Lemmatisation Task. Working Notes of EVALITA 2011, ISSN: 2240-5186.
35. Tamburini F. (2012). The AnIta-Lemmatiser. Working Notes of EVALITA 2011, ISSN: 2240-5186.
36. Grandi N., Montermini F., Tamburini F. (2011). "Annotating large corpora for studying Italian derivational morphology". *Lingue e Linguaggio*, X.2, 227-244, ISSN: 1720-9331.
37. Grandi N., Montermini F., Tamburini F. (2011). Prefazione al numero speciale "Morphology meets computational linguistics" di *Lingue e Linguaggio*, X.2, 181-182, ISSN: 1720-9331.
38. Windmann A., Jaul I., Tamburini F., Wagner P. (2011). "Prominence-Based Prosody Prediction for Unit Selection Speech Synthesis". In *Proc. Interspeech 2011*, Florence, Italy, August 2011, 325-328, ISSN(CD): 1990-9772.

39. Rossini Favretti R., Tamburini F., Zaninello A. (2011). "Exploiting corpus evidence for automatic sense induction". In *Proc. III Congreso Internacional de Linguística de Corpus (CILC2011)*, Valencia, Spain, April 2011, 751-760, ISBN: 978-84-694-6225-6.
40. Grandi N., Nissim M., Tamburini F. (2011). "Noun-clad adjectives. On the adjectival status of non-head constituents of Italian attributive compounds". *Lingue e Linguaggio*, X.1, 161-176, ISSN: 1720-9331.
41. Windmann A., Wagner P., Tamburini F., Arnold D., Oertel C. (2010). "Automatic Prominence Annotation for German Speech Syntesis Corpus: Towards Prominence-Based Prosody Generator for Unit Selection Syntesis". In *Proc. 7th ISCA Speech Syntesis Workshop*, September 2010, Kyoto, Japan.
42. Di Matteo N., Peroni S., Tamburini F., Vitali F. (2010). "Of Mice and Terms: Clustering Algorithms on Ambiguous Terms in Folksonomies". In *Proc of 25th ACM Symposium On Applied Computing (SAC 2010)*, Sierre, Switzerland, March 2010, 844-848.
43. Tamburini F. (2009). "Prominenza frasale e tipologia prosodica: un approccio acustico". In *Proc. Linguistica e modelli tecnologici di ricerca, XL congresso internazionale di studi, Società di Linguistica Italiana*, Vercelli, 437-455.
44. Di Matteo N., Peroni S., Tamburini F., Vitali F. (2009). "A parametric architecture for tags clustering in folksonomic search engines". In *Proc of 9th International Conference on Intelligent Systems Design and Applications (ISDA 2009)*, Pisa, Novembre 2009, 279-282.
45. Tamburini F. (2009). "PoS-tagging Italian texts with CORISTagger". In *Proc of EVALITA 2009. AI*IA Workshop on Evaluation of NLP and Speech Tools for Italian*, Reggio Emilia, December 2009.
46. Capelletti M., Tamburini F. (2009). "Polymorphic Categorical Grammars: expressivity and computational properties". In *Proc of ESSLLI Workshop on Parsing with Categorical Grammars*, Bordeaux, France.
47. Rossini Favretti R., Tamburini F. (2009). "Exploring register features through statistical analysis and interpretation of corpus evidence". In *Abstracts of 31st Annual Meeting of the Linguistics Association of Germany - Dgfs "Corpus, Colligation, Register Variation"*, Osnabrück, marzo 2009.
48. Capelletti M., Tamburini F. (2009). "Parsing with Polymorphic Categorical Grammars". *Advances in Computational Linguistics. Research in Computing Science*, 41, 87-98
49. Tamburini F. (2008). "La Linguistica Computazionale: un crogiolo di esperienze multidisciplinari". *Griselda Online*. Luglio 2008.
50. Tamburini F., Seidenari C., Bolognesi A., Bernardi R. (2008). "Italian Lexical-Classes Definition Using Automatic Methods". In Rossini Favretti R. (ed.), *Frames, Corpora and Knowledge Representation*, Bologna: Bononia University Press, 95-120.
51. Bernardi R., Bolognesi A., Seidenari C., Tamburini F. (2008). "Learning an Italian Categorical Grammar". In Rossini Favretti R. (ed.), *Frames, Corpora and Knowledge Representation*, Bologna: Bononia University Press, 185-200.
52. Magnini B., Cappelli A., Tamburini F., Bosco C., Mazzei A., Lombardo V., Bertagna F., Calzolari N., Toral A., Bartalesi Lenzi V., Sprugnoli R., Speranza M. (2008). "Evaluation of Natural Language Tools for Italian: EVALITA 2007". In *Proc. 6th International Conference on Language Resources and Evaluation - LREC 2008*, Marrakech.
53. Tamburini F. (2007). "EVALITA 2007: the Part-of-Speech Tagging Task". *Intelligenza Artificiale*, IV(2), 4-7.
54. Tamburini F. (2007). "CORISTagger: a high-performance PoS tagger for Italian". *Intelligenza Artificiale*, IV(2), 14-15.
55. Tamburini F., Wagner P. (2007). "On Automatic Prominence Detection for German". In *Proc. InterSpeech 2007*, Antwerp, 1809-1812.
56. Rossini Favretti R., Tamburini F., Martelli E. (2007). "Words from Bononia Legal Corpus". in W. Teubert (ed.), *Text Corpora and Multilingual Lexicography*, John Benjamins Publishing Company, 11-30.
57. Onelli C., Proietti D., Seidenari C., Tamburini F. (2006). "The DiaCORIS project: a diachronic corpus of written Italian". In *Proc. 5th International Conference on Language Resources and Evaluation - LREC 2006*, Genova, 1212-1215.
58. Bernardi R., Bolognesi A., Seidenari C., Tamburini F. (2006). "POS tagset design for Italian". In *Proc. 5th International Conference on Language Resources and Evaluation - LREC 2006*, Genova, 1396-1401.

59. Tamburini F. (2006). "Reliable Prominence Identification in English Spontaneous Speech". In *Proc. Speech Prosody 2006*, Dresden, PS1-9-19.
60. Bernardi R., Bolognesi A., Seidenari C., Tamburini F. (2005). "Automatic induction of a POS tagset for Italian". In *Proc. Australasian Language Technology Workshop 2005*, Sydney, 176-183.
61. Tamburini F. (2005). "Identificazione automatica della prominente frasale nel linguaggio parlato". In *Atti del 1° Convegno Nazionale AISV2004 - Associazione Italiana di Scienze della Voce*, Padova, 725-754.
62. Tamburini F. (2005). "Automatic Prominence Identification and Prosodic Typology". In *Proc. InterSpeech 2005*, Lisbon, 1813-1816.
63. Tamburini F. (2005). *Fenomeni prosodici e prominente: un approccio acustico*. Bologna: BUP.
64. Tamburini F. (2005). "Elaborazione automatica del linguaggio parlato", In Rossini Favretti R. (cur.), *E-Learning: comunicazione mediata e presenziale*. Bologna: BUP, 39-54.
65. Tamburini F., Caini C. (2005). "An automatic system for detecting prosodic prominence in American English continuous speech", *International Journal of Speech Technology*, 8, 33-44.
66. Bernardi R., Bolognesi A., Tamburini F., Moortgat M. (2004). "Categorial Type Logic meets Dependency Grammar to annotate an Italian corpus", In *Proc. Recent Advances in Dependency Grammars Workshop - COLING 2004*, Ginevra, 57-64.
67. Tamburini F., Caini C. (2004). "Automatic Annotation of Speech Corpora for Prosodic Prominence", In *Proc. Compiling and Processing Spoken Language Corpora workshop - LREC-CPSLC*, Lisbona, 53-58.
68. Tamburini F. (2004). "Building Distributed Language Resources by Grid Computing", In *Proc. 4th International Conference on Language Resources and Evaluation - LREC 2004*, Lisbona, 1217-1220.
69. Tamburini F., Paci S. (2004). "Dynamic generation of web-based multimedia language learning materials through on-line templates", In Satchell R., Chenik, N. (eds.), *University Language Centres: Forging the Learning Environments of the future. Papers from the 7th CercleS International Conference*, Parigi, 357-363.
70. Tamburini F. (2003). "Automatic Prosodic Prominence Detection in Speech using Acoustic Features: an Unsupervised System", In *Proc. 8th European Conference on Speech Communication and Technology - Eurospeech 2003*, Ginevra, 129-132.
71. Tamburini F. (2003). "Prosodic prominence detection in speech", In *Proc. Seventh International Symposium on Signal Processing and its Applications - ISSPA03*, Parigi, 385-388.
72. Tamburini F., Paci S. (2002). "Web-based language learning: authoring and assessment technologies", In *Proc. International Conference on Information and Communication Technologies in Education - ICTE2002*, Badajoz, Spagna, 872-877.
73. Tamburini F. (2002). "Automatic detection of prosodic prominence in continuous speech", In *Proc. Third International Conference on Language Resources and Evaluation - LREC2002*, Las Palmas, Spagna, 301-306.
74. Tamburini F. (2002). "A dynamic model for reference corpora structure definition", In *Proc. Third International Conference on Language Resources and Evaluation - LREC2002*, Las Palmas, Spagna, 1847-1850.
75. Tamburini F., De Santis C., Zamuner E. (2002). "Identifying phrasal connectives in Italian using quantitative methods", In S.Nuccorini (ed.), *Phrases and Phraseology - Data and Descriptions*, Peter Lang, 45-64.
76. Rossini Favretti R., Tamburini F., De Santis C. (2002). "CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model", In Wilson A., Rayson P. and McEnery T. (eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, Lincom-Europa, Munich, 27-38.
77. Rossini Favretti R., Tamburini F., Martelli E (2001). "Words from Bononia Legal Corpus", *International Journal of Corpus Linguistics*, Vol. 6, 13-34.
78. Tamburini F. (2000). "Annotazione grammaticale e lemmatizzazione di corpora in italiano", In Rossini Favretti R. (cur.), *Linguistica e informatica: multimedialita', corpora e percorsi di apprendimento*, Bulzoni, Roma, 57-73.
79. Paci S., Tamburini F. (1999). "A multimedia application for self-access learning: the MISSILE software package", In Poppi F. (a cura di), *Percorsi assistiti nell'auto-apprendimento*, CLUEB, Bologna, 141-149.

80. Tamburini F. (1999). "A multimedia framework for second language teaching in self-access environments", *Computers & Education*, 32(2), 137-149.

Altre Pubblicazioni

- Tamburini F., Paci S. (2002). "On-line templates for the production of web-based multimedia language learning material", In Proc. *International Conference on New Educational Environments - ICNEE2002*, Lugano.
- Rossini Favretti R., Gianninoni R., Paci S., Tamburini F. (2000). "Interfacce utente per la didattica delle lingue straniere: analisi di un'applicazione multimediale monitorata", In Andronico A., Casadei G., Sacerdoti G. (cur.), *Didattica 2000. Informatica per la didattica*, 2, 215-221.
- Rossini Favretti R., Bromwich W., Greenwood A., Poppi F., Tamburini F., Thursfield A. (1999). *Laboratorio DIAPASON for science*, Bologna: CLUEB.
- Davoli R., Tamburini F., Gaioni R. (1999). "A translation, rotation and scale invariant transform for grey scale images: a parallel implementation", In Proc. *Fifth European SGI / Cray MPP Workshop*, Bologna.
- Davoli R., Tamburini F., Frascari A. (1996). "Niche Genetic Algorithms: a massively parallel approach to Schedulability Checking in Real-Time Systems", *Science and Supercomputing at C.I.N.E.C.A.: 1995-Report*, Erbacci G., Voli M. (eds.), C.I.N.E.C.A., Bologna, 208-216.
- Davoli R., Tamburini F., Giachini L.A. (1996). "Scheduling Data Flow Programs in Hard Real-Time Environments", *Formal Techniques in Real-Time and Fault Tolerant Systems*, Lecture Notes in Computer Science, Heidelberg: Springer, 1135, 263-278.
- Davoli R., Tamburini F., Giachini L.A., Fiumana F. (1995). "Schedulability Checking in Real-Time Systems using Neural Networks", *Journal of Artificial Neural Networks*, 2, 421-430.
- Francis R., Tamburini F. (1994). "Identifying salient prosodic parameters in connected speech", In Rossini Favretti R. et al. (eds.), *The Self Access Facility in a Language Centre*, C.I.L.T.A., Bologna, 31-40.
- Rossini Favretti R., Silver M., Gasser R., Tamburini F. (eds.) (1994). *The Self Access Facility in a Language Centre*, C.I.L.T.A., Bologna.
- Tamburini F., Davoli R. (1994). "An algorithmic method to build good training sets for neural-network classifiers", *UBLCS Technical Report Series*, 94-18.
- Davoli R., Tamburini F. (1993). "DATA algorithm: a numerical method to extract shape information from gray scale images", In Pözlleitner W. and Wenger E. (eds.), *Image Analysis and Synthesis*, Vienna, 219-230.

Convegni, conferenze e seminari

Relazioni/Seminari su invito:

- "Morphological Analysis, PoS-Tagging and Lemmatisation", nell'ambito delle "Lectures on Computational Linguistics" organizzate dal Corso di Dottorato in Scienze linguistiche delle Università di Pavia/Bergamo, in collaborazione con la Fondazione Bruno Kessler, maggio 2017.
- "Automatic Identification of Prosodic Prominence by means of Acoustic Analyses", Conferenza internazionale *pS-prominenceS*, Viterbo, dicembre 2013.
- "Alla ricerca di un modello per l'analisi della prominente frasale nella lingua parlata", *Scuola Normale Superiore*, Pisa, marzo 2013.
- "Funzione demarcativa della prominente prosodica", seminario tenuto all'interno del ciclo *Linguaggio a Roma Tre*, Dip. di Linguistica, Università degli Studi Roma Tre, gennaio 2012 (in coll. con Lombardi Vallauri E.).
- "Automatic Prominence Identification and Prosodic Typology". Seminario tenuto presso la *Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld*, Germany, ottobre 2011.

- “Metodologie per la costruzione di grandi corpora: controllo o anarchia?”, Presentato alla *Giornata di Studi su Corpora di italiano L1 e L2: risorse linguistiche e strumenti di analisi*, Dip. di Lingue e Letterature Straniere, Università di Verona, maggio 2011.
- “Using Computational Methods to Analyse some Linguistic Phenomena”, Seminario tenuto al *CLLE-ERSS del CNRS* di Tolosa, febbraio 2010.
- “Il contributo delle nuove tecnologie nell’analisi della lingua parlata”, Seminario tenuto presso il Dip. di Linguistica, Università di Pavia nell’ambito del Dottorato di ricerca in Linguistica, dicembre 2009.
- “Computational Methods for different linguistic phenomena”, Seminario tenuto al *Laboratoire MoDyCo, Université Paris X*, dicembre 2008.
- “The grid experiments and Language Resources at CILTA”, Seminario tenuto presso il *Department of Computer Science, University of Surrey*, Guilford, settembre 2004.
- “CORIS/CODIS: a corpus of written Italian”, Seminario tenuto presso il *Tuscan Word Centre* del prof. John Sinclair, Pescia, giugno 2002.
- “Tecnologie per la distribuzione di risorse multimediali sul Web”, presentato al convegno *Scuola 5.0*, Bologna, maggio 2001.
- “Getting and handling corpus source documents”, Seminario tenuto presso il *Tuscan Word Centre* del prof. John Sinclair, Pescia, giugno 1999.

Relazioni:

- “Semgrex-Plus: a Tool for Automatic Dependency-Graph Rewriting”. *Fourth International Conference on Dependency Linguistics - Depling 2017*, Pisa, 18-20 Settembre 2017.
- “Towards Quantum Language Models”. *Conference on Empirical Methods in Natural Language Processing - EMNLP2017*, Copenhagen, 9-11 Settembre 2017 (in coll. con Basile I.).
- “(Better than) State-of-the-Art PoS-tagging for Italian Texts”. In *Proc. 3rd Italian Conference on Computational Linguistics - CLiC-IT 2016*, Napoli, 5-6 dicembre 2016.
- “(Almost) Automatic Conversion of the Venice Italian Treebank into the Merged Italian Dependency Treebank Format”. *3rd Italian Conference on Computational Linguistics - CLiC-IT 2016*, Napoli, 5-6 December 2016 (in coll. con Alfieri L).
- “Overview of the EVALITA 2016 Part Of Speech on TWitter for ITALian Task”. *Fifth International Workshop EVALITA 2016*, Napoli, 7 dicembre 2016 (in coll. con Bosco C., Bolioli A., Mazzei A.).
- “A BiLSTM-CRF PoS-tagger for Italian tweets using morphological information”. *Fifth International Workshop EVALITA 2016*, Napoli, 7 dicembre 2016.
- “Specialising Paragraph Vectors for Text Polarity DetectionE”. *10th International Conference on Language Resources and Evaluation - LREC 2016*, Portorož, 23-28 May 2016.
- “Automatic Identification of Mild Cognitive Impairment through the Analysis of Italian Spontaneous Speech Productions”. *10th International Conference on Language Resources and Evaluation - LREC 2016*, Portorož, 23-28 May 2016 (in coll. con Beltrami D., Calzà L., Gagliardi G., Ghidoni E., Marcello N., Rossini-Favretti R.).
- “Are Quantum Classifiers Promising?” *First Italian Conference on Computational Linguistics CLiC-it 2014*, Pisa, dicembre 2014.
- “Emotion Recognition with a Kernel Quantum Classifier”. *Fourth International Workshop EVALITA 2014*, Pisa, dicembre 2014.
- “Prosodic prominence detection in Italian connected speech”, presentato alla Conferenza internazionale *pS-prominenceS*, Viterbo, 13 dicembre 2013 (in coll. con Bertinetto P.M., Bertini C.).
- “Una valutazione oggettiva dei metodi più diffusi per l’estrazione automatica della frequenza fondamentale”, Presentato all’*IX Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce*, Venezia, gennaio 2013.

- “Objective, Subjective and Linguistic Roads to Perceptual Prominence. How are they compared and why?” *13th Annual Conference of the International Speech Communication Association (InterSpeech2012)*, Portland (OR), settembre 2012 (in coll. con Wagner P., Windmann A.).
- “AnIta: a powerful morphological analyser for Italian”. *8th International Conference on Language Resources and Evaluation - LREC 2012*, Istanbul, maggio 2012 (in coll. con Melandri, M).
- “A topologic view of Topic and Focus marking in Italian”. *8th International Conference on Language Resources and Evaluation - LREC 2012*, Istanbul, maggio 2012 (in coll. con Gagliardi, G. e Lombardi Vallauri E.)
- “Topic and Focus marking in an Italian corpus: some results of algorithmic measurement and structural interpretation”, Presentato al *VII GSCP International Conference: Speech and Corpora*, Belo Horizonte, Brasile, febbraio 2012 (in coll. con Lombardi Vallauri E.).
- “Topic e Focus in Italiano: demarcazione più che culminazione?”, Presentato all’*VIII Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce*, Roma, gennaio 2012 (in coll. con Gagliardi G. e Lombardi Vallauri E.).
- EVALITA 2011: the Lemmatisation Task”, *Presentato al Final Workshop of EVALITA 2011*, Roma, gennaio 2012.
- “Risorse linguistiche e strumenti computazionali per l’analisi morfologica dell’italiano”, Presentato al workshop *Morphology meets computational linguistics*, Bologna, ottobre 2010.
- “Noun-Clad Adjectives. On the adjectival status of non head constituents of Italian attributive compounds”, Presentato al *2nd TRIPLE International Conference, Word Classes. Nature, typology, computational representations*, Roma, marzo 2010 (in coll. con Grandi N. e Nissim M.).
- “Parsing with Polymorphic Categorical Grammars”, Presentato al *10th International Conference on Intelligent Text Processing and Computational Linguistics - CICLing 2009*, Città del Messico, marzo 2009 (in coll. con Capelletti M).
- “Italian Lexical-Classes Definition Using Automatic Methods”, Presentato a *Frames, Corpora and Knowledge Representation*, Bologna, giugno 2006 (in coll. con Seidenari C.).
- “EVALITA 2007: the Part-of-Speech Tagging Task”, *Presentato al Final Workshop of EVALITA 2007*, Roma, settembre 2007.
- “On Automatic Prominence Detection for German”, Presentato a *InterSpeech 2007*, Antwerp, agosto 2007 (in coll. con Wagner P.).
- “Prominenza frasale e tipologia prosodica: un approccio acustico”, Presentato al *XL congresso internazionale di studi, Società di Linguistica Italiana*, Vercelli, settembre 2006.
- “The DiaCORIS project: a diachronic corpus of written Italian”, Presentato al *5th International Conference on Language Resources and Evaluation - LREC 2006*, Genova, maggio 2006 (in coll. con Onelli C., Proietti D., Seidenari C.).
- “POS tagset design for Italian”, Presentato al *5th International Conference on Language Resources and Evaluation - LREC 2006*, Genova, maggio 2006 (in coll. con Bernardi R., Bolognesi A., Seidenari C.).
- “Reliable Prominence Identification in English Spontaneous Speech”, Presentato a *Speech Prosody 2006*, Dresden, maggio 2006.
- “Automatic Prominence Identification and Prosodic Typology”, Presentato a *InterSpeech 2005*, Lisbona, settembre 2005.
- “Identificazione automatica della prominenzia frasale nella lingua parlata”, Presentato al *1° Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce*, Padova, dicembre 2004.
- “POS tagset semi-automatic definition using the distributional information of CTL-supertags”, Seminario tenuto presso il CILTA nell’ambito del progetto FIRB2001, settembre 2004.
- “Categorial Type Logic meets Dependency Grammar to annotate an Italian corpus”, Presentato al *Recent Advances in Dependency Grammars Workshop - COLING 2004*, Ginevra, agosto 2004 (in coll. con Bernardi R., Bolognesi A., Moortgat M.).

- “Automatic Annotation of Speech Corpora for Prosodic Prominence”, Presentato al *Compiling and Processing Spoken Language Corpora workshop - LREC-CPSLC*, Lisbona, maggio 2004 (in coll. con Caini C.).
- “Building Distributed Language Resources by Grid Computing”, Presentato al *4th International Conference on Language Resources and Evaluation - LREC 2004*, Lisbona, maggio 2004.
- “Automatic Prosodic Prominence Detection in Speech using Acoustic Features: an Unsupervised System”, Presentato al *8th European Conference on Speech Communication and Technology - Eurospeech 2003*, Ginevra, settembre 2003.
- “Prosodic prominence detection in speech”, Presentato al *7th International Symposium on Signal Processing and its Applications - ISSPA03*, Parigi, luglio 2003.
- “Web-based language learning: authoring and assessment technologies”, Presentato all’*International Conference on Information and Communication Technologies in Education - ICTE2002*, Badajoz, Spagna, novembre 2002 (in coll. con Paci S.).
- “Dynamic generation of web-based multimedia language learning materials through on-line templates”, Presentato al *7th CercleS International Conference*, Parigi, settembre 2002 (in coll. con Paci S.).
- “Quantitative Analysis of Word Distributional Behaviour. Italian Adverbs of Manner: a case study”, Presentato al *7th TELRI Seminar "Information in Corpora"*, Dubrovnik, settembre 2002.
- “Corpus resources at CILTA: CORIS/CODIS a corpus of written Italian”, Seminario presso il *Tuscan Word Centre* del prof. John Sinclair nell’ambito di un corso sulla costruzione di *corpora*, giugno 2002.
- “Automatic detection of prosodic prominence in continuous speech”, Presentato al *3rd International Conference on Language Resources and Evaluation - LREC2002*, Las Palmas, Spagna, maggio 2002.
- “A dynamic model for reference corpora structure definition”, Presentato al *3rd International Conference on Language Resources and Evaluation - LREC2002*, Las Palmas, Spagna, maggio 2002.
- “On-line templates for the production of web-based multimedia language learning material”, Presentato all’*International Conference on New Educational Environments - ICNEE2002*, Lugano, maggio 2002 (in coll. con Paci S.).
- “Identifying phrasal connectives in Italian using quantitative methods”, Presentato al *4th International Symposium on Phraseology*, Roma, Aprile 2001 (in coll. con De Santis C., Zamuner E.).
- “CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model”, Presentato a *Corpus Linguistics 2001*, Lancaster UK, marzo 2001 (in coll. con Rossini Favretti R., De Santis C.).
- “Annotazione grammaticale e lemmatizzazione di corpora in italiano”, Presentato al *III Seminario AICLU - Linguistica e informatica: multimedialita', corpora e percorsi di apprendimento*, Bologna, Aprile 2000.
- “Applicazioni innovative dell’informatica nell’insegnamento delle lingue al CILTA”, Seminario tenuto presso il Dip. di Scienze dell’Informazione, febbraio 1999.

Organizzazione:

- Organizzazione (in coll. con Elena Cabrio e Alessandro Mazzei) della *Fifth Italian Conference on Computational Linguistics - CLiC-it 2018*, Torino, 10-12 dicembre 2018.
- Organizzazione (in coll. con Mario Vayra e Cinzia Avesani) del XI Convegno Nazionale dell’Associazione Italiana di Scienze della Voce, Bologna, 28-30 gennaio 2015.
- Organizzazione (in coll. con Nicola Grandi) del XXXVIII Convegno annuale della Società Italiana di Glottologia, La nozione di classico in linguistica, Bologna, 24-26 ottobre 2013.
- Organizzazione (in coll. con Petra Wagner) della *Special Session on Prosodic Prominence: Annotation, Prediction, Applications* al *13th Annual Conference of the International Speech Communication Association - InterSpeech2012*, Portland (OR), settembre 2012.
- Organizzazione (in coll. con Nicola Grandi e Fabio Montermini) del Workshop *Morphology meets computational linguistics*, Bologna, ottobre 2010.
- Organizzazione e presentazione dei seminari per “USCKI-Incognito Study trip” con seminari tenuti da R. Rossini Favretti, J. Bos, M. Nissim e F.Tamburini presso il Dip. SLO, aprile 2008.

- Organizzazione e presentazione del seminario "Parsing with Non-Associative Lambek Grammars" tenuto dal dott. M. Capelletti, presso il Dip. SLO, aprile 2008.
- Organizzazione e presentazione della giornata di studio "CORIS-DiaCORIS: nuovi strumenti e metodi per l'analisi e la diffusione della lingua italiana" con seminari tenuti dai prof. F. Sabatini, R. Rossini Favretti, D. Proietti, M. Biffi, A. Dardi, presso il Dip. SLO, maggio 2007.
- Organizzazione e presentazione del seminario "Dal Question-Answering alle applicazioni per l'Apprendimento delle lingue attraverso la creazione di Risorse Linguistiche" tenuto dal prof. R. Delmonte presso il CILTA, aprile 2005.
- Organizzazione e presentazione del seminario "Comunicazione e Multimodalità" tenuto dal prof. G. Ferrari presso il CILTA, marzo 2005.
- Organizzazione, in collaborazione col prof. G. Sandri del Dip. di Filosofia, Università di Bologna, del seminario "Gödel, Turing, and Intelligent Machinery" tenuto dal prof. W. Sieg presso il CILTA, novembre 2004.
- Organizzazione del seminario "Reversing time's arrow in grammar" tenuto dal prof. M. Moortgat presso il CILTA, settembre 2004.
- Organizzazione scientifica e presentazione della Giornata di Studio su "Mathematical and logical tools for natural language", con seminari tenuti dai prof. M. Moortgat, W. Buszkowski, R. Moot, presso il CILTA, marzo 2004.
- Organizzazione scientifica e presentazione del corso "CTL@CILTA" tenuto dal prof. M. Moortgat e dalla dott.ssa R. Bernardi presso il CILTA (per un totale di 55 ore di seminari), settembre 2003.
- Collaborazione con la prof.ssa R. Rossini Favretti nell'organizzazione scientifica del *III Seminario AICLU - Linguistica e informatica: multimedialità, corpora e percorsi di apprendimento*, con interventi, tra gli altri, dei prof. J. Sinclair, E. Tognini Bonelli, G. Ferrari, Bologna, Aprile 2000.
- Organizzazione di una Giornata di Studio sulle Grammatiche Categoriali, con seminari tenuti, tra altri, dai prof. J. Lambek, M. Moortgat, C. Casadio, presso il CILTA, marzo 2000.

Prodotti e risorse linguistiche sviluppate

PROMINENCE:

"*Prominence*" è un programma che raccoglie i risultati ottenuti dall'autore in numerosi anni di ricerca sull'identificazione automatica delle sillabe prominenti nel parlato continuo.

E' stato progettato per essere inserito all'interno di una nota e molto diffusa applicazione *open-source* - *WaveSurfer* - per l'analisi fonetico-acustica degli enunciati prodotta al Royal Institute of Technology - KTH - di Stoccolma, (<http://www.speech.kth.se/wavesurfer/>).

Prominence è in grado di identificare le sillabe prominenti contenute all'interno degli enunciati utilizzando unicamente le informazioni ricavate dalla componente acustica dell'enunciato (*waveform* o oscillogramma), raggiungendo prestazioni confrontabili con quelle ottenute da esperti annotatori umani.

Il programma può essere configurato ed adattato a lingue differenti modificando il contributo prodotto dai differenti parametri acustici considerati nel modello di prominenza utilizzato (energia, enfasi spettrale, durata e accenti di *pitch*). E' stato utilizzato con successo per l'identificazione della prominenza in enunciati inglesi, italiani e tedeschi e i risultati ottenuti utilizzando questo programma sono già stati presentati in un gran numero di convegni e riviste internazionali.

L'interesse di studiosi a livello internazionale nei confronti di questa applicazione e piuttosto alto e varie collaborazioni sono già state avviate.

CORIS/CODIS:

Un corpus di italiano scritto - *CORIS/CODIS* - è stato realizzato presso l'Università di Bologna ed è disponibile online da alcuni anni.

Il progetto, elaborato e coordinato da R. Rossini Favretti e F. Tamburini, ha la finalità di costruire un corpus generale di riferimento dell'italiano scritto, ben dimensionato e rappresentativo, accessibile e di facile fruizione.

Il CORIS contiene ad oggi 130 milioni di parole e viene aggiornato tramite un corpus di monitoraggio inglobato con cadenza triennale (è in preparazione il monitor corpus 2011/13). E' costituito da una raccolta di testi, autentici e ricorrenti nell'uso, in formato elettronico, selezionati come rappresentativi dell'italiano contemporaneo.

Il progetto, al momento attuale, è il risultato di una ricerca svolta nell'ambito dell'Università di Bologna. Ciò è stato reso possibile dallo sviluppo tecnologico, dal possibile riferimento ad esperienze precedenti e dal lungo periodo di preparazione che ha preceduto la fase progettuale e costitutiva.

Prevede come destinatari e fruitori un ampio numero di persone che si collocano dagli studiosi dell'italiano come L1 o L2, agli studenti italiani e stranieri impegnati in un'analisi linguistica basata su dati autentici e, in una prospettiva più ampia, a tutti coloro che sono interessati ad un'analisi intralinguistica e/o interlinguistica.

CODIS rappresenta una nuova articolazione di CORIS e permette una rimodulazione dinamica dei componenti testuali costitutivi del corpus in modo da poterlo adattare ad ogni necessità di comparazione interlinguistica.

Il corpus e' liberamente accessibile in rete per scopi di ricerca, e, ad oggi, centinaia di studiosi, di numerosi paesi, utilizzano CORIS/CODIS nelle loro ricerche come fonte primaria di dati empirici sulla lingua italiana.

ANITA:

AnIta è un potente analizzatore morfologico per l'italiano, implementato nell'ambito della *two-level morphology*. E' fornito di un lessico contenente più di 110.000 lemmi; le dimensioni estremamente rilevanti del lessico gli consentono di fornire informazioni morfologiche circa la flessione e la derivazione per gran parte delle forme incontrate in testi autentici.

AnIta implementa una nuova metodologia per rappresentare i processi morfologici di derivazione, i "Grafici di Derivazione" presentati in (Grandi, Montermini, Tamburini, 2011).

Basandosi sull'analizzatore morfologico AnIta è stato sviluppato un lemmatizzatore in grado di associare, in congiunzione con CORISTAGGER, il lemma di riferimento per ogni forma in contesto. Il lemmatizzatore è risultato tra i migliori strumenti di questo tipo (2° posto) nella competizione EVALITA 2011 per la valutazione degli strumenti computazionali sulla lingua italiana.

CORISTAGGER:

CorisTagger è un programma per l'annotazione automatica di testi in lingua italiana rispetto alle categorie lessicali. E' in grado di associare ad ogni termine contenuto nel testo la sua parte del discorso risolvendo le ambiguità con un livello di precisione estremamente elevato, anche grazie all'incorporazione di un analizzatore morfologico altamente sofisticato basato su un lemmario composto da circa 110.000 lemmi.

L'annotatore ha ottenuto risultati eccellenti nella recente campagna di valutazione EVALITA2007 classificandosi tra i migliori sistemi per la lingua italiana.

DiaCORIS:

L'obiettivo del progetto DiaCORIS, finanziato nell'ambito dell'iniziativa ministeriale FIRB2001, ha riguardato la costruzione di un corpus diacronico rappresentativo dell'italiano scritto più ampio rispetto a quelli esistenti, in grado di coprire un esteso arco temporale (dal 1861 al 1945) e concepito per essere fruibile anche in congiunzione al corpus sincronico CORIS/CODIS.

La costruzione del *corpus diacronico di italiano scritto* ha richiesto: a) una fase di indagine storica per l'identificazione e la scelta dei materiali da inserire nel corpus; b) una fase di progettazione del corpus vero e proprio nel rispetto della rappresentatività delle varie tipologie testuali; c) una fase di raccolta, digitalizzazione e predisposizione dei documenti in formato elettronico.

Sulla base del quadro d'articolazione del corpus (1861-1898 "Dopo l'unità", 1899-1922 "L'età liberale", 1923-1945 "Il fascismo", ciascuno articolato nelle sezioni Stampa, Narrativa, Saggistica, Prosa giuridica, Miscellanea) l'intero corpus è stato completamente digitalizzato per un totale di 15 milioni di parole e sono stati inseriti i metadati necessari a una corretta caratterizzazione e individuazione dei documenti (autore, titolo, editore, genere, ecc.) in formato XML; il corpus è già da tempo accessibile in rete tramite l'utilizzazione di una complessa architettura di information retrieval.

Recentemente si è avviato un nuovo progetto, sempre in collaborazione con l'Accademia della Crusca, per completare DiaCORIS inserendo due nuovi subcorpora che andranno a coprire l'arco temporale 1946-2001 (più precisamente 1946-1967 e 1968-2001).

BoLC:

Il Bononia Legal Corpus - BoLC - è un progetto di ricerca interdisciplinare finalizzato alla costituzione ed all'analisi di corpora giuridici paralleli e confrontabili, coordinato dalla prof.ssa Rema Rossini Favretti. Il progetto di ricerca ha avuto inizio nel 1997 ed è stato sviluppato presso l'Università di Bologna con la consulenza scientifica di John Sinclair.

Al momento attuale il corpus si articola in due sottocorpora, uno italiano e uno inglese. La rappresentatività del corpus, nella sua globalità, trova giustificazione in due momenti principali. Il subcorpus costituito dai dati del linguaggio giuridico italiano ed il subcorpus costituito dai dati del linguaggio giuridico inglese sono posti come rappresentativi di due diversi sistemi giuridici e delle differenze esistenti fra il sistema di civil law ed il sistema di common law. Inoltre, è fatto riferimento al linguaggio giuridico sviluppato nell'ambito dell'Unione Europea, visto come base comune, per la compresenza di varie lingue e di vari sistemi che in questo si realizza. E' quindi prevista la comparazione di testi giuridici in lingue diverse, le cui forme sono espressione, da un lato, di un ordinamento giuridico comune, quale quello comunitario, e, dall'altro, di ordinamenti e culture giuridiche diverse, quali quelle sviluppate nei singoli stati. Si intende, in tal modo, tenere in conto sia la progressiva formazione di un diritto uniforme a livello europeo sia la pluralità di sistemi normativi nazionali esistenti nell'ambito dell'Unione Europea.

Secondo tale duplice prospettiva si è costituito, per un primo livello di indagine orientativa, un corpus pilota, formato da corpora paralleli in lingua italiana ed in lingua inglese. A questa prima fase ha fatto seguito la costituzione di corpora comparabili in cui hanno trovato rappresentazione documenti propri dell'ambito legislativo, giudiziario e amministrativo, per analizzare le forme linguistiche prodotte nei due sistemi in situazioni comunicative comparabili.

Semgrex-Plus:

Creare un *treebank*, annotando ogni frase con la sua struttura sintattica, è certamente un compito complesso, pesante e non privo di errori. Per queste ragioni i *treebank* spesso richiedono complesse operazioni di manutenzione e revisione per correggere errori o per adattare l'annotazione per usi diversi o per diverse interpretazioni dei fenomeni sintattici. In grossi progetti, come quello delle *Universal Dependencies*, una modifica alle *guideline* di annotazione per tener conto dell'aggiunta di nuove lingue o cambi di prospettiva teorici richiedono spesso la revisione dell'intero corpus per adattare le strutture sintattiche coinvolte.

Semgrex-Plus consente proprio l'automatizzazione di queste procedure: grazie alla definizione di un linguaggio formale per la specificazione delle regole di riscrittura delle strutture sintattiche a dipendenza e allo sviluppo di opportuni algoritmi per l'applicazione di tali regole e la verifica delle loro interazioni, consente la riscrittura automatica di intere porzioni di *treebank*, riducendo al minimo il lavoro richiesto al linguista. Il software è liberamente scaricabile all'indirizzo <https://github.com/ftamburin/Semgrex-Plus.git>

Bologna, li 15/6/2019

Fabio Tamburini