

ISTITUTO DI LINGUISTICA COMPUTAZIONALE “A. ZAMPOLLI”

2015



Rapporto Annuale



Consiglio Nazionale delle Ricerche

Sommario

1	Introduzione	4
2	L'ILC nel 2015: fatti e cifre	4
2.1	Personale.....	4
2.2	Organizzazione interna.....	6
2.2.1	Responsabili e Organi di Governo.....	6
2.2.2	Organizzazione gestionale-amministrativa	6
2.2.3	Organizzazione della ricerca	7
2.3	Finanziamenti	8
2.4	Progetti.....	9
2.4.1	Progetti europei	9
2.4.2	Progetti nazionali e regionali	11
2.5	Collaborazioni scientifiche	14
2.5.1	Accordi Bilaterali	14
2.5.2	Accordi e Convenzioni.....	14
2.5.3	Altre collaborazioni	16
2.6	Premi	17
3	Attività di ricerca	18
3.1	Ricerca scientifica	18
3.2	Ricerca Istituzionale	19
3.3	Workshop, conferenze, seminari	20
3.3.1	Workshop e conferenze organizzati e co-organizzati dall'ILC.....	20
3.3.2	Partecipazione a comitati scientifici di conferenze.....	20
3.3.3	Relazioni a convegni, conferenze, workshop e giornate di studio.....	21
3.3.4	Seminari interni.....	22
3.4	Pubblicazioni	26
3.4.1	Contributi in rivista	26
3.4.2	Contributi in volume	26
3.4.3	Contributi in atti di convegno	26
3.4.4	Curatele.....	29
3.4.5	Note Editoriali	29
3.4.6	Altri prodotti della ricerca	29
3.5	Internazionalizzazione.....	30
4	Attività di alta formazione	31
4.1	Corsi universitari	31
4.2	Summer School	31
4.3	Seminari invitati presso altre istituzionali nazionali e internazionali	32
4.4	Tesi di laurea	33

4.5	Tesi di dottorato	35
4.6	Tirocini	36
4.7	Visiting students	37
4.8	Convenzioni nell'ambito della formazione	37
4.9	Formazione interna	37
5	Attività editoriali	38
6	Attività di terza missione	41
6.1	Partecipazione a Organismi tecnico-scientifici e normativi	41
6.2	Partecipazione ad Associazioni e Comitati scientifici	41
6.3	Partecipazione a comitati di valutazione	42
6.4	Valorizzazione dei risultati e trasferimento tecnologico	42
6.5	Attività di Public Engagement	43
6.5.1	Eventi pubblici.....	43
6.5.2	Sito web	43
6.5.3	Trasmissioni radiofoniche	43
6.5.4	Iniziative di interazione con scuole e Università.....	43
7	APPENDICE: Personale ILC	45

Prefazione

Questo documento costituisce il primo rapporto annuale dell'Istituto di Linguistica Computazionale "Antonio Zampolli" del Consiglio Nazionale delle Ricerche. Con tutti i limiti di una prima uscita, si propone di raccontare un anno di idee, di progetti, di persone. In particolare, raccoglie e sintetizza - all'insegna della trasparenza e del desiderio di condivisione sia verso l'interno sia con il mondo esterno - le attività svolte nel corso del 2015 su molteplici versanti, che spaziano da quello della ricerca scientifica, dell'alta formazione, del trasferimento tecnologico e della divulgazione a quello gestionale-amministrativo. A nome dell'istituto è per me una grande soddisfazione introdurre questa panoramica delle attività svolte nel corso del 2015, che ricompone le singole voci e i contributi individuali all'interno di un quadro complesso e dinamico dal quale emergono innegabili punti di forza.

Nel 2015, le attività di ricerca dell'Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) sono proseguite con successo, portando al conseguimento di risultati apprezzabili, al consolidamento e all'estensione della visibilità nazionale e internazionale di linee di attività, all'estensione della rete di contatti scientifici e all'incremento delle iniziative di formazione specialistica e di coordinamento di attività scientifiche e istituzionali.

Tra i risultati più significativi dell'anno:

- *A seguito dell'adesione dell'Italia all'infrastruttura europea CLARIN-ERIC (Common Language Resources Infrastructure - European Research Infrastructure Consortium), l'ILC è stato nominato Istituto esecutore dell'infrastruttura italiana CLARIN-IT e un ricercatore è stato nominato dal MIUR come National Representative;*
- *Nel 2015, si è registrato un incremento della capacità di attrarre risorse esterne per la ricerca, sia su bandi competitivi (nazionali e internazionali) sia attraverso contratti e consulenze con committenza esterna (pubblica e industriale) e interna al CNR: dei 16 progetti attivi, 5 a fine anno risultavano terminati e 9 erano rappresentati da contratti attivati nel corso dell'anno;*
- *Le attività di alta formazione da parte di ricercatori ILC sono proseguite con successo. Tra queste: tre corsi di insegnamento accademico sono stati tenuti presso l'Università di Pisa e la Venice International University nell'ambito di corsi di laurea triennale e magistrale; quattro moduli didattici sono stati svolti a International Summer Schools; quattro tesi di dottorato sono giunte a compimento nel corso dell'anno; ricercatori dell'ILC hanno condotto la supervisione di 11 tesi di laurea triennale e di 6 tesi di laurea specialistica;*
- *Le attività editoriali, di disseminazione e divulgazione scientifica sono proseguite con successo con la direzione di tre riviste scientifiche del settore (di cui una avviata nel corso dell'anno), l'organizzazione di importanti conferenze internazionali e la partecipazione a numerosi eventi divulgativi rivolti al grande pubblico;*
- *Ricercatori dell'Istituto hanno contribuito, insieme a rappresentanti delle maggiori istituzioni italiane operanti nel settore della Linguistica Computazionale, alla fondazione dell'Associazione Italiana di Linguistica Computazionale (AILC).*

Pisa, novembre 2016



Simonetta Montemagni
Direttrice dell'ILC-CNR

1 Introduzione

L'Istituto di Linguistica Computazionale (ILC) è un centro di riferimento, a livello nazionale e internazionale, nel settore della Linguistica Computazionale. L'Istituto afferisce al Dipartimento Scienze Umane e Sociali, Patrimonio Culturale (DSU) del Consiglio Nazionale delle Ricerche (CNR) e svolge attività di ricerca nei settori scientifici strategici della disciplina, oltre ad attività editoriali, di formazione e di trasferimento tecnologico.

Fin dalle origini, la missione e le attività di ricerca dell'ILC si collocano programmaticamente all'interno dell'area umanistica, in un rapporto di costante interazione interdisciplinare con competenze di base eterogenee, che vanno dalle varie anime della Linguistica (formale, tipologica, cognitiva e applicata) all'informatica e alle infrastrutture digitali, dalla psicologia della cognizione allo studio dei sistemi complessi e alle neuroscienze. Ne sono testimonianza l'afferenza dell'Istituto al Dipartimento di Scienze Umane e Sociali, Patrimonio Culturale (DSU) e la sua collocazione nella mappa delle competenze disciplinari del CNR all'interno di un'area disciplinare di tipo umanistico (N), corrispondente al settore ERC SH4_6 e all'Area 10 della classificazione ANVUR.

La varietà delle linee di attività e dei progetti di ricerca rendono l'ILC una realtà unica nel panorama italiano e una delle poche a livello internazionale dove si affiancano: ricerche innovative nel settore delle Digital Humanities; attività volte alla definizione di standard e infrastrutture di ricerca distribuite; definizione di metodi e di tecniche avanzate per la ricerca e la gestione "intelligente" dell'informazione all'interno di basi documentali in linguaggio naturale disponibili sul Web o su Intranet locali; creazione di modelli computazionali dell'apprendimento linguistico in contesti ecologici di interazione comunicativa.

La ricerca all'ILC combina ricerca di base, con un investimento su temi di frontiera, e ricerca applicata, all'interno di un circolo virtuoso con ricadute significative sulla società e, in particolare, sul contesto socio-economico e culturale. Le attività sono condotte all'interno di una rete consolidata di collaborazioni a livello nazionale e internazionale con Istituti di ricerca, Università ed Enti Pubblici, così come con industrie e piccole e medie imprese, nell'ambito di numerosi progetti di ricerca.

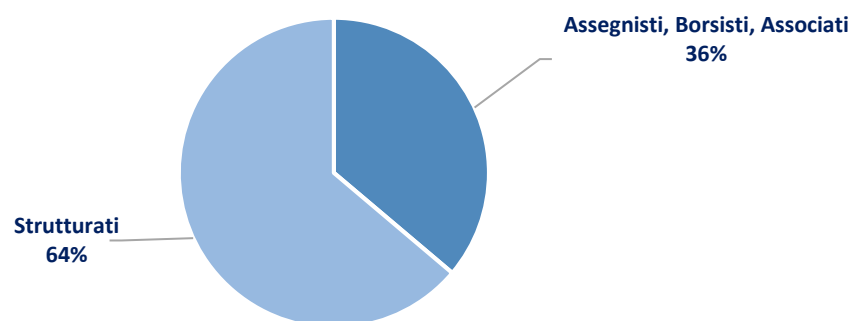
Nel 2015, le attività di ricerca e sviluppo dell'ILC interessano quattro aree di competenza principali:

- Trattamento automatico del testo e filologia computazionale;
- Trattamento automatico del linguaggio naturale ed estrazione di conoscenza;
- Risorse, standard e infrastrutture;
- Modelli computazionali dell'uso linguistico.

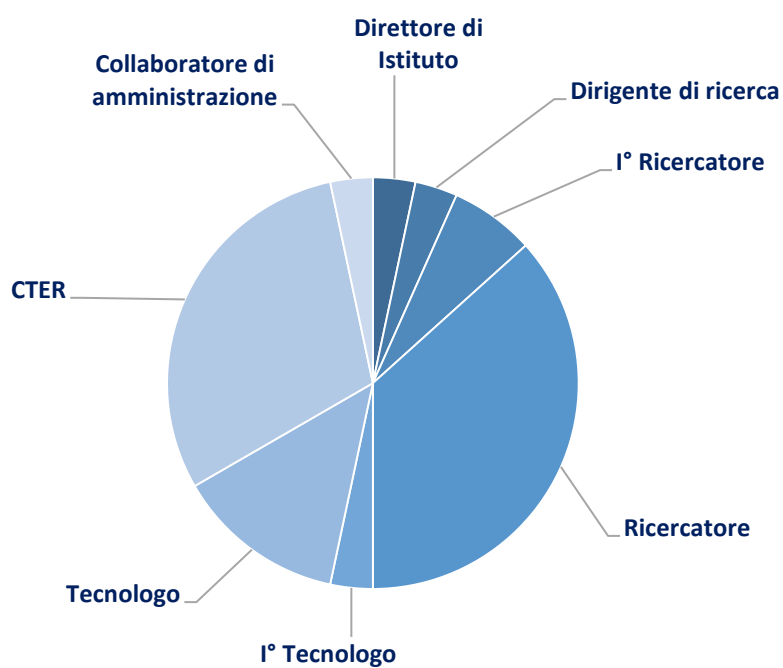
2 L'ILC nel 2015: fatti e cifre

2.1 Personale

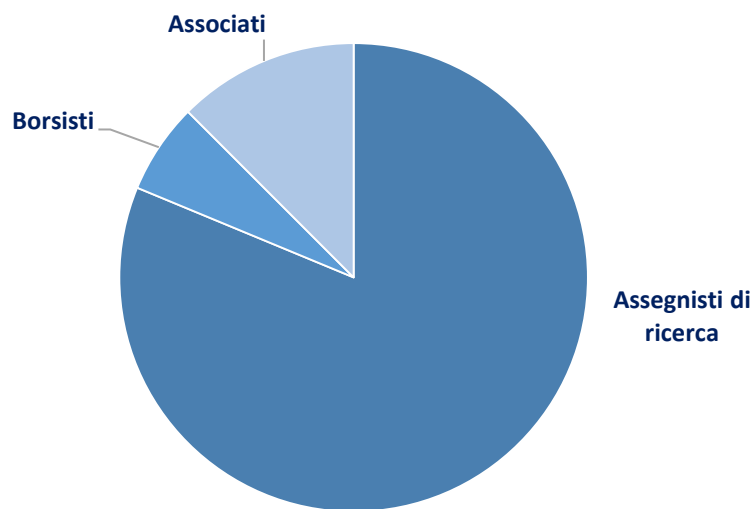
Il personale dell'ILC è composto da ricercatori, tecnologi e personale tecnico/amministrativo sia di supporto alla ricerca sia di area gestionale-amministrativa. Nel 2015, l'Istituto ha in totale 30 unità di personale strutturato (a tempo indeterminato e determinato) distribuite tra la sede di Pisa (28) e la sezione staccata di Genova (2), a cui se ne aggiungono 18 (16 a Pisa e 2 a Genova) tra assegnisti, contrattisti, borsisti e personale associato (cfr. Appendice 1).



RAPPORTO PERSONALE STRUTTURATO VS ALTRO PERSONALE



SUDDIVISIONE PERSONALE STRUTTURATO PER PROFILI E LIVELLI



RIPARTIZIONE ALTRO PERSONALE

2.2 Organizzazione interna

2.2.1 Responsabili e Organi di Governo

<i>Direttore</i>	Simonetta Montemagni
<i>Consiglio di Istituto</i>	Alessandro Enea, Claudia Marzi, Monica Monachini, Simonetta Montemagni, Vincenzo Parrinelli, Vito Pirrelli, Manuela Sassi
<i>Responsabile amministrativo</i>	Antonella Gadducci
<i>Responsabile dei sistemi informativi</i>	Alessandro Enea
<i>Responsabile della sicurezza</i>	Vincenzo Parrinelli

2.2.2 Organizzazione gestionale-amministrativa

Nel 2015 è proseguita con successo la riorganizzazione delle mansioni gestionali e amministrative avviata in precedenza. Tale riorganizzazione si è resa necessaria alla luce del sovraccarico di responsabilità in relazione a tutte le operazioni contabili e relativi adempimenti amministrativi, anche in forza dei cambiamenti amministrativo-burocratici prodotti dall'entrata in vigore delle ultime normative statali in materia di acquisti. Nell'ambito di tale riorganizzazione è stato istituito un *Ufficio supporto alla ricerca* per progetti finanziati in ambito nazionale e internazionale (USP), deputato a fornire supporto a ricercatori e tecnologi durante l'intero iter progettuale, dalla proposta alla rendicontazione.

Segue la lista dei comitati in essere nel 2015:

<i>Ufficio Supporto Progetti</i>	Paola Baroni, Antonella Gadducci, Eva Sassolini, Noemi Terreni
<i>Comitato di autovalutazione e premiale</i>	Monica Monachini, Simonetta Montemagni, Vito Pirrelli
<i>Commissione comunicazione</i>	Paola Baroni, Laura Cignoni, Alessandro Enea, Simonetta Montemagni, Claudia Soria
<i>Comitato scientifico per i seminari</i>	Federico Boschetti, Felice Dell'Orletta, Monica Monachini

Nel 2015, i rappresentanti ILC all'interno dei Comitati dell'Area della Ricerca – CNR di Pisa sono rimasti invariati:

Commissione Comunicazione

Paola Baroni, Claudia Soria

Commissione Relazioni Internazionali

Valeria Quochi

2.2.3 Organizzazione della ricerca

Nel 2015 le attività progettuali dell'Istituto sono strutturate per Commesse e Moduli di attività, come previsto dal Regolamento di organizzazione e funzionamento del CNR del 2005 (art. 44):

COMMESSA / MODULO	RESPONSABILE
Metodi e Strumenti Computazionali per l'Informatica Umanistica e la gestione della conoscenza - Metodi e strumenti computazionali per l'analisi del testo - Trattamento automatico del linguaggio ed estrazione di conoscenza - Metodi, risorse e strumenti per lo studio della storia della tradizione e per la critica del testo	Simonetta Montemagni Manuela Sassi Simonetta Montemagni Federico Boschetti
Modelli bio-computazionali del linguaggio e della cognizione - Architetture bio-computazionali del lessico mentale - Modelli teorici e computazionali di acquisizione lessicale in contesti mono- e multi-lingui	Vito Pirrelli Marcello Ferro Claudia Marzi
Risorse e tecnologie linguistiche: standardizzazione, sviluppo, infrastrutture, strategie - Risorse e tecnologie linguistiche: modelli, metodi di sviluppo, applicazioni, disegno di strategie internazionali - Infrastrutture per l'interoperabilità e l'integrazione di risorse e tecnologie linguistiche	Monica Monachini Monica Monachini Claudia Soria
Tecnologie della lingua per il Patrimonio Culturale Tecniche linguistico-semantiche per il supporto alla traduzione e alla consultazione di testi	Vito Pirrelli Emiliano Giovannetti

Moduli ILC all'interno di Commesse afferenti ad altri Istituti del CNR:

COMMESSA / MODULO	RESPONSABILE
Knowledge Discovery and Data Mining (Estrazione di conoscenza e data mining) Natural Language Processing and Text Mining	Chiara Renso (ISTI) Felice Dell'Orletta (ILC)
Web Applications for the Future Internet Social Language Resources for the Multilingual Web	Maria Claudia Buzzi (IIT) Monica Monachini (ILC)

COMMESSA / MODULO	RESPONSABILE
Commessa per la gestione delle attività del bando Progetto Interesse-Invecchiamento	Tullio Pozzan (BM)
▪ ILC	Lucia Marconi (ILC)

All'organizzazione ufficiale per Commesse e Moduli si sono affiancati nel tempo i **Laboratori**, che si configurano come strutture trasversali e dinamiche che svolgono la duplice funzione di interfaccia verso l'esterno delle linee di attività e di ricerca che vi afferiscono e, internamente all'istituto, di incubatori di nuove idee e soluzioni basate sull'integrazione e la condivisione di competenze, infrastrutture e tecnologie.

Segue la lista dei Laboratori attivi nel 2015:

ComPhys Lab

Progettazione e sviluppo di modelli (bio-)computazionali del comportamento linguistico con l'obiettivo di comprendere e spiegare le relazioni tra competenza grammaticale, uso linguistico e correlati neuro- e psico-linguistici della comunicazione verbale e dei suoi disturbi.

Sito: www.comphyslab.it

Responsabile: Vito Pirrelli

CoPhiLab

Formalizzazione delle entità e delle relazioni nel dominio della filologia collaborativa; creazione di risorse digitali; progettazione e sviluppo di componenti software, in particolare per le lingue classiche.

Sito: <http://cophilab.ilc.cnr.it:8080/CoPhiLabPortal>

Responsabile: Federico Boschetti

ItaliaNLP Lab

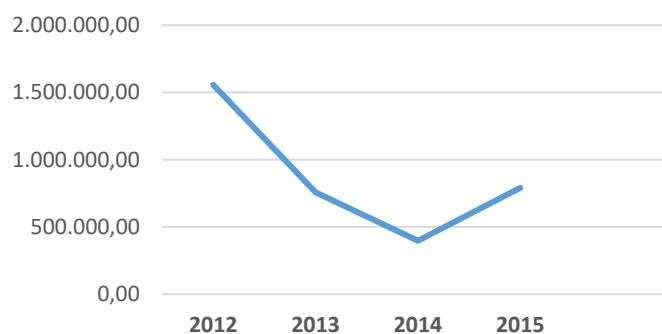
Progettazione e sviluppo di modelli, metodi, algoritmi e tecnologie per il Trattamento Automatico del Linguaggio e per l'estrazione di conoscenza, con particolare attenzione alla lingua italiana. Principali linee di attività: annotazione linguistica multi-livello di testi; estrazione di conoscenza da collezioni documentali; sviluppo di prototipi applicativi.

Sito: www.italianlp.it

Responsabile: Felice Dell'Orletta

2.3 Finanziamenti

Rispetto all'anno precedente, nel 2015 è quasi raddoppiata la capacità di attrarre risorse esterne per la ricerca, sia su bandi competitivi (nazionali e internazionali) sia attraverso contratti e consulenze con committenza esterna (pubblica e industriale) e interna al CNR.



TREND DEI FINANZIAMENTI ESTERNI

Come illustrato nel grafico, tale dato segna un'inversione dell'andamento discendente che ha caratterizzato gli ultimi anni dell'Istituto, un andamento purtroppo condiviso a livello nazionale per quanto riguarda sia il CNR e altri enti di ricerca sia l'università.

L'informazione relativa alla quantità di fondi attratti rappresenta tuttavia un dato parziale, che deve essere integrato con il numero di contratti esterni attivi, nel 2015 pari a 16 (cfr. www.ilc.cnr.it/it/content/progetti). Ai 5 progetti che a fine anno risultavano terminati, si sono affiancati 9 nuovi contratti attivati nel corso dell'anno: anche su questo versante, dunque, il saldo rimane positivo.

A questo dato, va inoltre affiancato il numero di proposte progettuali avanzate nel corso del 2015, all'interno di bandi estremamente competitivi a livello nazionale ed internazionale (per un totale di 24). A livello internazionale, 2 sono state le proposte sottomesse nell'ambito del Bando ERA-NET HERA JRP UP, 3 nel Bando ERASMUS+ e 10 nell'ambito di diverse Call di H2020, alle quali si aggiunge il coinvolgimento dell'ILC in una proposta ERC. Sul versante nazionale, l'ILC ha partecipato a 1 proposta in risposta a un Bando della Presidenza del Consiglio (Struttura di Missione per gli Anniversari di Interesse Nazionale), a 4 proposte nell'ambito di Bandi regionali (3 in PAR-FAS 2007-2013 e FAR FAS 2014, e 1 in POR FESR 2014-2020), a 2 Bandi di fondazioni private (Compagnia di San Paolo e Telecom Italia) e a un Bando di Ateneo (Università di Verona). Di queste 24 proposte, 3 sono state ammesse al finanziamento, mentre altre sono ancora in attesa dell'esito della selezione. A queste, si aggiungono le proposte preparate in risposta al Bando "Progetti di ricerca di Rilevante Interesse Nazionale" (PRIN 2015), che sono state inviate a inizio 2016: 9 di queste vedono coinvolta un'Unità di Ricerca ILC.

2.4 Progetti

2.4.1 Progetti europei

Progetti europei coordinati dall'ILC

DLDP - The Digital Language Diversity Project (<http://www.dldp.eu>)

Progetto triennale finanziato dalla Comunità Europea nell'ambito del programma Erasmus+.

Responsabile Scientifico: Claudia Soria

Il progetto Erasmus + (Attività KA2, Azione chiave 2 Partenariati Strategici) si propone di far avanzare la sostenibilità delle lingue regionali e minoritarie europee nel mondo digitale, incrementando nei loro parlanti la conoscenza e le abilità per creare e condividere contenuti sulle apparecchiature digitali usando le loro lingue minoritarie.

NETWORDS - The European Network on Word Structure (<http://www.networds-esf.eu>)

Programma quadriennale di Research Networking della European Science Foundation.

Responsabile Scientifico: Vito Pirrelli

NetWordS è un programma volto a incrementare le attuali conoscenze teoriche, tipologiche, psicolinguistiche, informatiche e neurofisiologiche relative alla struttura e all'elaborazione delle parole, con l'obiettivo di promuovere nuovi metodi di ricerca sul lessico mentale. Il programma, che coinvolge 16 Paesi europei e più di 40 centri di ricerca, ha una forte valenza interdisciplinare e persegue la formazione di giovani scienziati attraverso borse di studio e brevi soggiorni all'estero, incoraggiando l'integrazione sinergica di metodologie multi-disciplinari e diversi orientamenti teorici attraverso la condivisione di obiettivi trasversali, dati e infrastrutture di ricerca.

Progetti europei ai quali l'ILC partecipa come partner

CADDY - Cognitive Autonomous Diving buddy (<http://www.caddy-fp7.eu>)

Progetto triennale finanziato dalla Comunità Europea nell'ambito del Settimo Programma Quadro (Progetto FP7-ICT-2013-10: Cognitive Systems and Robotics). Coordinatore: Università di Zagabria-Facoltà di Elettronica.

Partner: Consiglio Nazionale delle Ricerche – ISSIA e ILC; Istituto Superior Tecnico; Jacobs University Bremen Gmbh; Universitaet Wien; University of Newcastle Upon Tyne; Divers Alert Network Europe Foundation

Referente Scientifico ILC: Lucia Marconi

Progetto europeo (STREP) di robotica marina finalizzato alla realizzazione di un sistema che contribuirà a rendere più sicure immersioni e attività subacquee. Il contributo ILC al progetto ha riguardato: la definizione di un protocollo di comunicazione fra il robot e il subacqueo; la creazione di un linguaggio per la loro interazione; lo sviluppo di un analizzatore sintattico in grado di associare sequenze di gesti sintatticamente corrette ad un comando o lista di comandi da far eseguire al robot.

PARTHENOS - Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies (<http://www.parthenos-project.eu>)

Progetto Horizon 2020 di durata quadriennale finanziato dalla Comunità europea.

Responsabile Scientifico Unità di Ricerca ILC: Monica Monachini

Il progetto mira a rafforzare la coesione della ricerca nell'ampio settore relativo a studi linguistici, scienze umane, patrimonio culturale, storia, archeologia e settori collegati, attraverso un cluster tematico di infrastrutture di ricerca europee, integrando iniziative, infrastrutture elettroniche e altre infrastrutture di punta, e costruendo ponti tra campi diversi, sebbene strettamente interconnessi. PARTHENOS raggiungerà questo obiettivo attraverso la definizione e il supporto di standard comuni, il coordinamento di attività congiunte, l'armonizzazione della definizione e dell'implementazione di policy e lo sviluppo di servizi comuni e di soluzioni condivise agli stessi problemi.

Iniziative europee in cui l'ILC è sottocontraente

ELRC - European Language Resource Coordination (<http://lr-coordination.eu>)

Azione finanziata nell'ambito del CEF SMART 2014/1074 Programme, contratto Ref. Ares(2014)2275366, dalla Commissione Europea

Referente tecnologico nazionale: Simonetta Montemagni

L'azione "European Language Resources Coordination" (ELRC) si colloca all'interno del programma 'Connecting Europe Facility' (CEF): finanziata dalla Commissione europea, si propone di migliorare ed estendere la copertura e la qualità del sistema di traduzione automatica MT@EC sviluppato dalla DG Translation in vista della sua integrazione nei servizi pubblici online di tutti i paesi europei. Tale obiettivo è perseguito attraverso l'individuazione e la raccolta di risorse linguistiche multilingui delle amministrazioni e istituzioni governative in tutti i 30 Paesi europei che partecipano al programma CEF. Nell'ambito di questa azione, l'ILC ha il compito di supportare il processo di individuazione e raccolta delle risorse linguistiche (lessici e corpora testuali, multi- e mono-lingui) in Italia.

2.4.2 Progetti nazionali e regionali

Progetti finanziati a livello nazionale

AAA offresi - Ascolto Accoglienza Azioni offresi

Progetto finanziato dal Ministero dell'Interno nell'ambito del Fondo Europeo per l'integrazione di cittadini di paesi terzi (FEI) 2007-2013 – Progetti a valenza territoriale 2013. Capofila: Comune di Genova - Ufficio Legalità e Diritti; Partner: ILC e COSPE (Cooperazione per lo Sviluppo dei Paesi Emergenti)

Responsabile Scientifico Unità di Ricerca ILC: Paola Cutugno

Progetto finalizzato all'integrazione sociale e linguistica di immigrati, con particolare attenzione al settore della Pubblica Amministrazione. Il contributo ILC al progetto è consistito nella realizzazione di strumenti documentali a supporto dei servizi del Comune di Genova per rendere accessibile anche all'utenza non madrelingua il linguaggio settoriale della Pubblica Amministrazione, caratterizzato - come tutti i linguaggi tecnici - da proprie convenzioni linguistiche. Sulla base della terminologia saliente emersa dai diversi ambiti di attività l'ILC ha realizzato un glossario strutturato con spiegazioni atte a facilitare la comprensione delle parole più rilevanti. Il glossario è stato tradotto e pubblicato nelle lingue più diffuse tra i cittadini immigrati presenti sul territorio genovese (albanese, arabo, cinese, spagnolo, ucraino). Si tratta di uno strumento utilizzabile dagli operatori dei servizi demografici, dai cittadini stranieri che si presentano agli sportelli per la fruizione dei servizi e dalle associazioni di immigrati.

MEMORATA POETIS - Memoria poetica e poesia della memoria. Ricorrenza lessicale e tematiche nella versificazione epigrafica e nel sistema letterario

Programma di Ricerca Scientifica di Rilevante Interesse Nazionale finanziato dal MIUR (PRIN-2010/2011-2010NK2ACB). Durata: 1 febbraio 2013 - 31 gennaio 2016.

Responsabile Scientifico Unità di Ricerca ILC: Federico Boschetti

Scopo del progetto è creare e mettere a disposizione della comunità scientifica un motore di ricerca plurilingue, semantico e tematico, oltre che lessicale, utile allo studio delle iscrizioni in versi, di qualunque provenienza ed epoca, genere e tipo. L'Unità di Ricerca ILC si è occupata di: condurre l'analisi linguistica dei materiali epigrafici e dei corpora letterari latini e greci su cui applicare il motore di ricerca intertestuale; mettere a punto metodologie di statistica esplorativa per lo studio dei testi; arricchire il corpus digitale con un numero limitato di epigrafi in lingua araba e una più estesa collezione di testi poetici coevi, nella medesima lingua, come gruppo di controllo.

MODELACT - Modelli semantici e pragmatici dell'ontologia azionale (<http://modelact.lablita.it>)

Progetto di durata quadriennale finanziato con il Fondo per gli Investimenti della Ricerca di Base - "Futuro in Ricerca 2012" del MIUR (Project Code RBFR12C6O8; 2013-2016)

Responsabile Scientifico Unità di Ricerca ILC: Irene Russo

Il progetto mira a far progredire le attuali conoscenze sulla cognizione umana, in particolare grazie all'individuazione delle qualità cognitivamente rilevanti per la segmentazione del continuum dello spazio azionale in unità discrete e codificate dal linguaggio naturale. L'obiettivo conoscitivo principale è di passare dall'identificazione dei concetti azionali prodotta dal precedente progetto regionale toscano IMAGACT (Ontologia per immagini dell'azione per l'acquisizione di L2 in ambiente multilingue) alla loro definizione, fornendo una modellazione che deve dimostrarsi valida, primariamente a fini del trattamento automatico del linguaggio e dell'interazione uomo-macchina. Il progetto propone un cambio di paradigma semantico, da deduttivo a induttivo, nella costruzione di ontologie. Il problema del tipo dei modelli di azione elaborati dal sistema cognitivo umano è affrontato integrando diversi punti di vista: acquisizionale, semantico e computazionale.

SCRIBE - Scritture Brevi, Semplificazione Linguistica, Inclusione Sociale: Modelli e Applicazioni

Progetto triennale (2013-2016) finanziato dal Ministero dell'Istruzione, dell'Università e della Ricerca (PRIN 2010FWM3B4 – Area 10). Partner del Progetto: Università di Tor Vergata, Università "L'Orientale" di Napoli, Università ROMA TRE, Università di Macerata, Università di Pisa, ILC-CNR.

Responsabile Scientifico Unità di Ricerca ILC: Simonetta Montemagni

Il progetto si propone di studiare il fenomeno della produzione relativa a messaggi sintetici e abbreviati dal punto di vista sia sincronico che diacronico, dalle espressioni contemporanee (brevi scritti usati per i messaggi elettronici, sms e chat) alle strategie di abbreviazione tipiche dell'italiano e dei sistemi dialettali grafici e linguistici. Il contributo ILC al progetto si articola lungo le seguenti direttrici: 1) specializzazione ed estensione di strumenti di trattamento automatico del linguaggio per l'annotazione linguistica automatica di "scritture brevi" e il monitoraggio dei fenomeni linguistici che le contraddistinguono; 2) definizione di una metodologia "avanzata" per la valutazione dell'accessibilità (leggibilità) di un testo e la sua eventuale semplificazione.

TALMUD - Traduzione Talmud Babilonese (<http://www.talmud.it>)

Progetto quinquennale finanziato dal MIUR. Partner coinvolti nel progetto: Presidenza del Consiglio dei Ministri; Ministero dell'Istruzione, dell'Università e della Ricerca; Unione delle Comunità Ebraiche Italiane; Collegio Rabbinico Italiano; Consiglio Nazionale delle Ricerche.

Responsabile Scientifico Unità di Ricerca ILC: Andrea Bozzi / Emiliano Giovannetti

Il progetto ha come obiettivo la traduzione in lingua italiana del Talmud Babilonese. La traduzione commentata, con testo originale a fronte in lingua ebraica e aramaica, è realizzata da un team di traduzione formato da circa 70 studiosi da tutte le parti del mondo tramite l'utilizzo di Traduco, una piattaforma Web collaborativa sviluppata dall'ILC che include strumenti editoriali avanzati e componenti per il trattamento del testo e della conoscenza basati su metodi e tecniche della linguistica computazionale.

Progetti regionali

ISLE - Intelligent Semantic Liquid eBook (<http://www.progettoisle.it>)

Progetto biennale finanziato dalla Regione Toscana nell'ambito del Bando POR CREO FESR 2007-2013 - Attività 1.5.a, in collaborazione con aziende IT (M.E.T.A. SRL, 01Servizi SRL, VIDITRUST SRL, SPACE SPA); Bando Unico R&S - Anno 2012 - Linea A.

Responsabile Scientifico Unità di Ricerca ILC: Simonetta Montemagni

Il progetto iSLe si propone di ripensare i processi di generazione, di fruizione e distribuzione dei contenuti per la didattica su piattaforma mobile, web e off-line. In particolare, iSLe si pone come obiettivo la realizzazione di un nuovo tipo di ebook per la didattica che possa adattarsi alle esigenze formative degli utenti e li conduca all'apprendimento attraverso percorsi cognitivi impliciti e espliciti. La piattaforma innovativa per l'editoria digitale scolastica sviluppata all'interno del progetto è provvista di funzionalità avanzate che permettono di creare una nuova generazione di liquid book nativi o seminativi, cioè generati anche a partire da contenuti tradizionali. Il contributo al progetto dell'unità ILC riguarda l'annotazione semantica dei contenuti multimediali e l'analisi dello stile di redazione dei contenuti sulla base delle esigenze formative.

Progetti di Ente

CLAVIUS - Clavius on the Web (<http://claviusontheweb.it>)

Progetto nazionale. Durata: 19 novembre 2012 - 11 agosto 2017

Responsabile Scientifico Unità di Ricerca ILC: Emiliano Giovannetti / Simone Marchi

L'obiettivo del progetto è la conservazione e la valorizzazione di una parte dei manoscritti conservati nell'Archivio Storico della Pontificia Università Gregoriana. Il progetto prende in esame alcuni manoscritti relativi a Christophorus Clavius (1538-1612), matematico e astronomo gesuita. Tali manoscritti sono stati digitalizzati, trascritti, tradotti e analizzati dai punti di vista linguistico, lessicale e semantico. La terminologia e le entità di dominio individuate nel testo sono state strutturate in un lessico e un'ontologia e sono state collegate a risorse già disponibili in rete secondo i principi dei Linked Data.

NINFA - iNtelligent Integrated Network For Aged people

Progetto di Interesse strategico CNR "Invecchiamento": Innovazioni tecnologiche e molecolari per un miglioramento della salute dell'anziano. Durata: 1 gennaio 2014 - 30 giugno 2018. Coordinatore: IEIIT - Istituto di Elettronica e Ingegneria dell'informazione e delle Telecomunicazioni.

Referente Scientifico Unità di Ricerca ILC: Lucia Marconi

Si tratta di un sotto-progetto del progetto CNR nazionale "Invecchiamento" che intende analizzare e proporre soluzioni ai problemi relativi all'applicazione delle Tecnologie dell'Informazione e della Comunicazione (TIC) per un invecchiamento attivo e al monitoraggio di danni cognitivi. Il progetto affronta l'accettabilità di nuove TIC, la valutazione della condizione di benessere degli utenti e la gestione di eventi critici con un impatto minimo sugli utenti durante la fornitura di servizi domiciliari, attraverso soluzioni differenti, calibrate principalmente su aspetti riguardanti gli utenti finali. Il contributo dell'ILC riguarda l'applicazione di tecnologie del linguaggio per il monitoraggio di danni cognitivi.

SM@RTINFRA-SSHCH - Infrastrutture integrate intelligenti per l'ecosistema dei dati delle scienze sociali, umane e del patrimonio culturale

Progetto nazionale di durata triennale finanziato con il "Fondo Ordinario per gli Enti di Ricerca - Quota finalizzata al Finanziamento Premiale di Specifici Programmi e Progetti" del MIUR.

Responsabile Scientifico Unità di Ricerca ILC: Monica Monachini

Il progetto premiale mira a creare una struttura di governance di coordinamento nazionale dei nodi italiani delle infrastrutture di ricerca (RI) europee di Social Sciences and Humanities, Cultural Heritage (SSHCH). Il risultato finale prevede il potenziamento delle RI nazionali e della partecipazione dell'Italia come membro agli ERIC già costituiti (CLARIN, ESS, SHARE) o in fase di costituzione (DARIAH). Le principali linee di attività sono: formazione per lo sviluppo delle competenze; ricerca nel settore delle tecnologie abilitanti fondamentali (Key Enabling Technologies - KETs) per l'avanzamento delle infrastrutture; networking, trasferimento tecnologico; diffusione dei risultati. All'interno del progetto, l'ILC ha predisposto azioni preparatorie in vista dell'adesione dell'Italia all'infrastruttura di ricerca CLARIN, avvenuta nell'autunno del 2015.

Progetti in cui l'ILC è sottocontraente

MAPS - Marine Planning and Service Platform

Progetto finanziato nell'ambito del bando "POR Liguria (2007-2013) – Asse 1 Innovazione e Competitività – Azione 1.2.2. DLTM – Ricerca industriale e sviluppo sperimentale a favore delle imprese del Distretto ligure per le Tecnologie Marine – Anno 2015 – pos. 19"

Referente ILC: Monica Monachini

Il progetto regionale MAPS ha come scopo la realizzazione di un motore di ricerca semantico per la documentazione di ambito oceanografico. Al progetto partecipano partner industriali liguri. L'ILC ha contribuito al progetto come sottocontraente, svolgendo per conto della società Delta Progetti 2000 Srl le seguenti attività: definizione delle specifiche dell'interfaccia di ricerca semantica; definizione dell'architettura del sistema; sviluppo e integrazione del motore semantico; supporto nella sperimentazione.

2.5 Collaborazioni scientifiche

2.5.1 Accordi Bilaterali

- **ACCORDO CNR/PAS (PAN) - POLONIA**

La protezione del patrimonio linguistico. Analisi comparativa delle attitudini in Italia e in Polonia verso la diversità linguistica

Responsabile ILC: Claudia Soria

2.5.2 Accordi e Convenzioni

- **Accademia della Crusca**

Recupero, conservazione e analisi di archivi testuali

Responsabili ILC: Simonetta Montemagni

- **Archivio Storico della Pontificia Università Gregoriana - APUG (Roma)**

Metodi e applicazioni web per la fruizione in rete di documenti, manoscritti e a stampa antichi sul tema della storia della scienza

Responsabile ILC: Andrea Bozzi

- **Associazione MIL Onlus Museo Internazionale Luzzati**

Definizione delle modalità di realizzazione dei moduli necessari per sviluppare l'architettura software eMoCHA (enhanced Museum of Cultural Heritage & Arts)

Responsabile ILC: Lucia Marconi

- **Associazione NeuroCare Onlus (Pisa)**

Collaborazione per lo sviluppo di sistemi per il monitoraggio in campo medico

Responsabile ILC: Vito Pirrelli

- **Austrian Academy of Sciences (Wien, Austria)**

Collaborazione in materia di "Corpus linguistics and Theoretical and computational modelling of Morphology"

Responsabile ILC: Vito Pirrelli

- **European Network of Ancient Greek and Latin Epigraphy - EAGLE (Grant Agreement n. 325122)**

Cooperation to WG1, WG2, WG3, WG4, WG5

Responsabile ILC: Federico Boschetti

- **European Science Foundation**

Piano di pubblicazione dei risultati del progetto NetWords (The European Network on Word Structure)

Responsabile ILC: Vito Pirrelli

- **Fondazione Bruno Kessler (FBK, Trento), Dip. di Studi Umanistici - Facoltà di Lettere e Filosofia - Università del Piemonte Orientale "Amedeo Avogadro" (Vercelli), Dip. di Informatica - Università degli Studi di Pisa, Dip. di Ingegneria dell'Impresa - Università degli Studi di Roma Tor Vergata, Dip. di Informatica - Università degli Studi di Torino**

Creazione di una rete di interesse comune nell'ambito delle tecnologie e delle risorse per il trattamento automatico del linguaggio

Responsabile ILC: Andrea Bozzi

- **Fondazione Bruno Kessler - FBK (Trento)**

Costruzione di un corpus per la lingua italiana - ITA-TimeBank - annotato manualmente con relazioni temporali

Responsabile ILC: Monica Monachini

- **Gruppo Editoriale l'Espresso S.p.A. (Roma), Istituto di Informatica e Telematica - Consiglio Nazionale delle Ricerche (IIT-CNR, Pisa) e Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" - Consiglio Nazionale delle Ricerche (ISTI-CNR, Pisa)**

Sperimentazione di nuove forme di giornalismo sfruttando le tecnologie e gli strumenti ICT oggetto di attività di ricerca e sviluppo negli istituti CNR coinvolti

Responsabile ILC: Felice Dell'Orletta

▪ **Istituto di Fisiologia Clinica (IFC-CNR)**

Collaborazione nei settori della adeguatezza ed efficacia pragmatica della comunicazione verbale e dei disturbi evolutivi della comunicazione verbale e non verbale

Responsabile ILC: Vito Pirrelli

▪ **M.E.T.A. Srl**

Sviluppo di sistemi per il supporto alla didattica con utilizzo di tecniche di annotazione linguistica ed estrazione di conoscenza e di strumenti e metodi per l'analisi della leggibilità di testi destinati a scuole di diverso ordine e grado

Responsabile ILC: Felice Dell'Orletta

▪ **Mount Allison University**, Dept. of Classics - (Sackville, NB, Canada)

Cooperation in the field of Digital Classics and Collaborative Philology

Responsabile ILC: Federico Boschetti

▪ **Museo Archeologico di Zagabria** (Zagabria, Croazia) e Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" - Consiglio Nazionale delle Ricerche (ISTI-CNR, Pisa)

Sviluppo congiunto di ricerche nei settori della linguistica computazionale e della computer graphics applicati ai beni culturali, archeologia e epigrafia digitali, informatica umanistica

Responsabile ILC: Federico Boschetti

▪ **Museo Galileo - Istituto e Museo di Storia della Scienza** (IMSS, Firenze)

Conservazione, analisi e valorizzazione del patrimonio testuale storico scientifico di IMSS

Responsabili ILC: Simonetta Montemagni, Eva Sassolini

▪ **Università Cattolica del Sacro Cuore** - Centro Interdisciplinare di Ricerche per la Computerizzazione dei Segni dell'Espressione (UCSC-CIRCSE, Milano)

Attività di ricerca nel settore del trattamento automatico delle lingue classiche

Responsabile ILC: Federico Boschetti

▪ **Università degli Studi di Napoli Federico II** – Centro Interdipartimentale L.U.P.T. (Laboratorio di Urbanistica e di Pianificazione del Territorio "Raffaele d'Ambrosio")

Azioni comuni di ricerca e formazione: monitoraggio e linee guida per la costruzione, uso e comprensione delle forme della comunicazione pubblica sincrona e asincrona in rete; corsi di formazione nell'ambito della pragmatica della comunicazione e dell'educazione digitale; approfondimenti scientifici e pianificazione di interventi formativi nell'ambito della comunicazione asincrona con riferimento alla Computer Mediated Communication e alla Keyboard-to-screen-communication; attività di collaborazione con centri deputati a sostegno delle specifiche disabilità linguistiche

Responsabile ILC: Simonetta Montemagni

▪ **Università degli Studi di Pavia**, Dip. di Studi Umanistici

Attività di collaborazione sulle seguenti tematiche di ricerca: progettazione e sviluppo di risorse linguistiche mono e multilingui (corpora annotati e non, lessici, ontologie), relative a lingue antiche e/o moderne; utilizzo di risorse linguistiche prodotte dalla collaborazione o già liberamente disponibili presso le Parti per fini di ricerca linguistica a vari livelli di analisi (morfologia, sintassi, semantica, lessico).

Responsabile ILC: Vito Pirrelli

▪ **Università di Pisa**, Dip. di Informatica e **Università degli Studi di Torino**, Dip. di Informatica

Costruzione di una Treebank dell'italiano con annotazione sintattica a dipendenze secondo lo schema "Stanford Dependencies" nell'ambito della collaborazione tra il Dipartimento di Informatica dell'Università degli Studi di Pisa e GOOGLE

Responsabile ILC: Simonetta Montemagni

- **Università di Pisa**, Dip. di Informatica e **Università degli Studi di Torino**, Dip. di Informatica
Estensione della Treebank dell'italiano con annotazione sintattica a dipendenze secondo lo schema "Universal Dependencies (UD), nell'ambito della collaborazione tra il Dipartimento di Informatica dell'Università degli Studi di Pisa e LINKEDIN.
Responsabile ILC: Simonetta Montemagni
- **Université Paris-Sorbonne**, LABEX OBVIL, Observatoire de la vie littéraire (Paris)
Attività di ricerca finalizzata allo studio e produzione di edizioni digitali di manoscritti di autori moderni e contemporanei
Responsabile ILC: Simonetta Montemagni
- **Université Sidi Mohammed Ben Abdellah** (Fès, Marocco)
Poursuivre et développer une coopération dans les domaines de l'enseignement et de la recherche en sciences humaines et sociales, en particulier sur le traitement numérique des langues et des manuscrits, collation, édition critique
Responsabile ILC: Vito Pirrelli
- **University of Patras**
Erasmus+ - Erasmus Charter for Higher Education (ECHE)
Responsabile ILC: Vito Pirrelli

2.5.3 Altre collaborazioni

Oltre alle collaborazioni che sono state oggetto di un accordo formale, nel corso del 2015 l'ILC ha ampliato la sua rete di contatti attraverso collaborazioni informali, sia con la rete scientifica CNR sia con la comunità scientifica nazionale e internazionale, riguardanti diversi ambiti disciplinari. Oltre a collaborare con altri gruppi di ricerca specializzati nel settore della linguistica computazionale, il personale dell'ILC collabora infatti con studiosi di altre discipline umanistiche e di altri settori che possono beneficiare delle tecnologie del linguaggio.

Nel corso dell'anno, l'ILC ha proseguito la collaborazione con la rete scientifica del CNR, in particolare con Istituti afferenti a tre Dipartimenti:

- **Scienze Umane e Sociali, Patrimonio Culturale**
 - Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG)
 - Istituto di Scienze e Tecnologie della Cognizione (ISTC)
- **Scienze Biomediche**
 - Istituto di Fisiologia Clinica (IFC)
- **Ingegneria, ICT e Tecnologie per l'Energia e i Trasporti**
 - Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI)
 - Istituto di Informatica e Telematica (IIT)
 - Istituto di elettronica e di ingegneria dell'informazione e delle telecomunicazioni (IEIT)
 - Istituto di studi sui sistemi intelligenti per l'automazione (ISSIA)

Alcune di queste collaborazioni sono state formalizzate con la stipula di accordi di collaborazione scientifica e convenzioni (cfr. supra).

Nel 2015 è proseguita l'attività di consolidamento ed estensione delle collaborazioni dell'ILC con la comunità scientifica italiana e internazionale, grazie anche alle attività di alta formazione in Università italiane ed estere e alla partecipazione a bandi di diversa natura (interni all'Ente, regionali, nazionali e di fondazioni private), da parte di singoli gruppi di ricerca o di più gruppi di ricerca. Tra le collaborazioni esterne sono da segnalare, in particolare:

- **Institut Géographique National**
Annotazione e referenziazione di Entità Nominate
Personale coinvolto: Francesca Frontini

- **Scuola Normale Superiore, Pisa – Gruppo di Ricerca coordinato da Amos Bertolacci**
Sviluppo di un componente software per l'editoria critica da integrare nella piattaforma di analisi testuale Omega
 Personale coinvolto: Emiliano Giovannetti, Simone Marchi
- **University of Alberta (Canada), Dept. of Psychology**
Definizione di un modello esplicativo "memory-based" dell'effetto di latenza nel compito di digitazione di parole composte inglesi in corrispondenza del confine di morfema, che delinea un effetto combinato di specializzazione predittiva e competizione inibitoria tra composti e loro costituenti memorizzati contestualmente.
 Personale coinvolto: Marcello Ferro, Claudia Marzi, Vito Pirrelli
- **University of Groningen (Paesi Bassi)**
Realizzazione di strumenti finalizzati all'estrazione mirata di dati da risorse dialettali online e predisposizione per analisi dialettometriche offerte dalla piattaforma GABMAP (www.gabmap.nl), come esempio di interoperabilità di risorse e strumenti all'interno di una infrastruttura di ricerca.
 Personale coinvolto: Sebastiana Cucurullo, Simonetta Montemagni, Eva Sassolini
- **University of Novi Sad (Serbia), Faculty of Philosophy, Dept. of Psychology**
Modelli psico-computazionali biologicamente ispirati di acquisizione ed elaborazione associative della conoscenza lessicale. Confronto tra modelli non-supervisionati di apprendimento correlativo con l'obiettivo di valutare possibili punti di contatto tra modelli diversi, nei termini delle equazioni differenziali che ne definiscono il comportamento
 Personale coinvolto: Marcello Ferro, Claudia Marzi, Vito Pirrelli
- **Università degli Studi di Roma "La Sapienza" - Dipartimento di Psicologia dei processi di Sviluppo e socializzazione - prof. Pietro Lucisano**
Definizione di una metodologia di monitoraggio delle competenze linguistico-comunicative degli studenti, e in modo particolare delle abilità di scrittura, in una prospettiva diacronica e articolata rispetto a diverse dimensioni (lessicale, morfo-sintattica e sintattica). Prime sperimentazioni, condotte con successo, hanno riguardato: lo studio dell'evoluzione delle abilità di scrittura nel passaggio dalla prima alla seconda classe della scuola secondaria di primo grado; la costruzione di un modello di analisi empirica in grado di consentire l'osservazione dei processi e dei prodotti della didattica della produzione scritta
 Personale coinvolto: Dominique Pierina Brunato, Felice Dell'Orletta, Simonetta Montemagni, Giulia Venturi

2.6 Premi

- **Distinguished Young Paper**
Conferenza CLIC-it 2015 (Conferenza Italiana di Linguistica Computazionale)
 Dominique Brunato e Felice Dell'Orletta, *ISACCO: a corpus for investigating spoken and written language development in Italian school-age children*
 Povo – Trento (Fondazione Bruno Kessler), 3-4 dicembre 2015
- **Best Poster Award**
GL17 - Seventeenth International Conference on Grey Literature. A New Wave of Textual and Non-Textual Grey Literature
 Sara Goggi, Monica Monachini, Francesca Frontini, Roberto Bartolini, Gabriella Pardelli, Maurizio De Mattei, Franco Bustaffa, Giuseppe Manzella (2015) *Marine Planning and Service Platform (MAPS): An Advanced Research Engine for Grey Literature in Marine Science*
 Amsterdam (Paesi Bassi), 1-2 dicembre 2015

3 Attività di ricerca

In conformità con quanto indicato nelle “Linee Guida per la gestione integrata del Ciclo della Performance degli Enti Pubblici di Ricerca” dell’ANVUR, le attività condotte nel corso del 2015 sono qui di seguito suddivise in *ricerca scientifica* e *ricerca istituzionale*.

3.1 Ricerca scientifica

Nel 2015 le attività di ricerca scientifica hanno permesso di conseguire risultati apprezzabili, consolidando ed estendendo la visibilità nazionale e internazionale delle linee di attività. La rete di contatti scientifici e collaborazioni di ricerca è stata estesa e sono state moltiplicate le iniziative di formazione specialistica e di coordinamento delle attività scientifiche. Qui di seguito i risultati ottenuti nelle diverse aree di ricerca.

Metodi e strumenti computazionali per l'informatica umanistica e la gestione della conoscenza

Per quanto concerne le attività di recupero, conservazione e valorizzazione del patrimonio culturale dell'ILC è da segnalare l'ampliamento del corpus bilingue italiano-latino dei Digesta di Giustiniano, attuato nell'ambito della collaborazione con l'Università di Roma Tor Vergata. Sono state sviluppate soluzioni tecnologiche avanzate su temi centrali del trattamento automatico del linguaggio e per l'estrazione di conoscenza. Tali tecnologie svolgono un ruolo chiave nell'ambito del Knowledge Management, della didattica delle lingue e della ricerca umanistica e, grazie al grado di maturità raggiunto, sono state oggetto di interesse da parte di istituzioni pubbliche e di industrie. Inoltre, sono stati notevolmente potenziati gli strumenti e le risorse per lo studio delle lingue classiche, in particolare latino, greco e arabo. Da segnalare il porting in Java di LEMLAT per la lemmatizzazione e l'analisi morfologica di testi in lingua latina e l'ulteriore sviluppo del motore morfologico dell'arabo.

Modelli bio-computazionali del linguaggio e della cognizione

Tra i risultati scientifici più rilevanti vi sono l'estensione e l'ottimizzazione dell'architettura TSOM (Temporal Self-organising Map) e gli esperimenti comparativi sui sistemi flessionali di tedesco, arabo e italiano, volti allo studio del comportamento funzionale e dell'organizzazione morfologica di mappe temporali auto-organizzanti addestrate con lessici delle diverse lingue al fine di valutare diversi compiti correlati direttamente e indirettamente all'acquisizione e all'elaborazione lessicali. Per quanto concerne gli aspetti di coordinamento scientifico e gestionale, è da segnalare che Vito Pirrelli ha diretto la rete scientifica internazionale legata al progetto ESF RNP NetWordS, che ha visto coinvolte più di 40 istituzioni di 16 Paesi europei. La ricerca condotta da tale rete interdisciplinare è stata incentrata sul lessico mentale, sulla struttura della parola e sull'elaborazione delle parole nelle lingue europee.

Risorse e tecnologie linguistiche: standardizzazione, sviluppo, infrastrutture, strategie

È stato sviluppato un nuovo paradigma basato su infrastrutture di ricerca che consentono agli studiosi l'accesso distribuito, la condivisione e l'utilizzo di dati linguistici in combinazione con metodi e strumenti per il trattamento automatico del linguaggio. I risultati di spicco si individuano nella definizione di modelli e metodologie per estrarre, annotare e rappresentare nuovi livelli di conoscenza linguistica ed extra-linguistica, nonché nella valorizzazione e preservazione del patrimonio linguistico. Oltre allo sviluppo delle risorse e tecnologie linguistiche e alla definizione di strategie e raccomandazioni volte a alla loro promozione, sono state intensificate le iniziative di standardizzazione a livello nazionale, mediante la partecipazione attiva come esperti ai gruppi del Comitato Nazionale di Unificazione (UNI) e ai gruppi internazionali ISO e W3C. Sul versante delle risorse, le attività hanno riguardato primariamente la comunità delle Scienze Umane, con particolare attenzione alle lingue classiche. Grazie a collaborazioni con l'industria, sono state sviluppate risorse e terminologie per la gestione e il recupero intelligente di informazione da documentazione del settore oceanografico. Per quanto riguarda l'apprendimento delle lingue e la diversità linguistica, sono da segnalare il progetto Erasmus+ sulla Diversità Linguistica Digitale e la costituzione di un gruppo di lavoro (Special Interest Group) sulle lingue “less-resourced” nell'ambito dell'Associazione ELRA (European Language Resources Association)

Modelli, strutture e metodi computazionali per lessici specialistici antichi

Sono stati ulteriormente sviluppati i modelli per la traduzione e lo studio di testi adottati nell'ambito dei progetti “Talmud (Traduzione del Talmud Babilonese)” e “Clavius on the Web” (vedi infra). I sistemi realizzati possono essere facilmente adattati a testi di altra natura e a lingue diverse da quelle per le quali sono stati inizialmente concepiti.

Per quanto concerne le ricerche condotte nell'ambito delle tecnologie di Computer-Assisted Translation (CAT), è da segnalare il sistema Web collaborativo Traduco, sviluppato nell'ambito del progetto Talmud. Si tratta di una soluzione all'avanguardia che fornisce supporto a traduzioni collaborative e permette di facilitare, migliorare e velocizzare la traduzione di opere complesse e articolate. Oltre ai componenti standard tipici degli strumenti CAT tradizionali, include alcune specifiche funzionalità atte a semplificare il processo di traduzione di testi che pongono specifiche problematiche strutturali, stilistiche, linguistiche ed ermeneutiche. Il sistema, in costante sviluppo, è accessibile via web ed è in grado di fornire al traduttore, in modo automatico, suggerimenti di possibili traduzioni prodotte sulla base del repertorio di traduzioni già effettuate, sia dal traduttore stesso sia da altri traduttori.

Traduco è adatto per studi avanzati in ambito filologico o paleografico, in quanto consente di: incorporare le immagini oltre alle risorse testuali; inserire note, commenti, riferimenti bibliografici, annotazioni e citazioni; marcare il testo su base semantica; creare automaticamente rubriche tematiche; effettuare ricerche complesse all'interno del testo stringato grazie a tecniche avanzate per il trattamento automatico della lingua e l'estrazione della conoscenza; visualizzare preimpaginati ad hoc ed esportare le traduzioni per la produzione immediata di edizioni cartacee; monitorare in tempo reale lo svolgimento del lavoro tramite un pannello di supervisione.

3.2 Ricerca Istituzionale

L'Istituto di Linguistica Computazionale (ILC) partecipa attivamente a due importanti infrastrutture di ricerca di livello europeo, con ruoli diversi.

CLARIN (Common Language Resources and Technology Infrastructure)

L'ILC ha partecipato alla fase preparatoria di CLARIN nel periodo 2008-2011. Il 1° ottobre 2015 l'Italia è diventata il XVI Membro a Pieno Titolo di CLARIN-ERIC, infrastruttura di ricerca dedicata alle Scienze Umane e Sociali, gestita a livello europeo dall'ente di coordinamento, European Research Infrastructure Consortium.

Nel 2015 è stata avviata la fase di costituzione del centro italiano di CLARIN, coordinato dall'ILC. Monica Monachini, ricercatrice dell'ILC, ha ottenuto la nomina ministeriale di Coordinatore Nazionale per l'Italia dell'Infrastruttura di ricerca per le Scienze Umane e Sociali con il compito di formare il consorzio italiano, CLARIN-IT. Tale consorzio, ancora in formazione, includerà molti dei principali Atenei e Istituti di ricerca che si occupano di studi linguistici e letterari, filologia e informatica, così come gruppi principalmente coinvolti nel campo delle scienze umane digitali e nei settori della documentazione, della digitalizzazione e delle tecnologie linguistiche per le scienze umane.

DARIAH-IT (Digital Research Infrastructure for the Arts and Humanities)

L'ILC è partner di DARIAH-IT, il nodo Italiano di DARIAH-EU (Digital Research Infrastructure for the Arts and Humanities), di cui condivide la missione principale. L'infrastruttura eroga servizi avanzati basati sull'uso delle tecnologie dell'informazione e della comunicazione (TIC) per la ricerca nel campo delle Arti e Scienze Umane. L'ILC, in particolare, offre contenuti digitali e competenze nelle aree degli standard di rappresentazione di dati linguistici, con particolare attenzione all'interoperabilità, alla conservazione e alla sicurezza dei dati. DARIAH-IT fornisce strumenti di ricerca digitale all'avanguardia e promuove il loro uso da parte di un'ampia e multidisciplinare comunità di ricercatori. Oltre a combinare sviluppo tecnologico con istruzione e formazione, questa infrastruttura promuove collaborazione e ricerca congiunta, comunicazione e disseminazione, condivisione di esperienze e strumenti nei settori delle scienze umane digitali e del patrimonio culturale digitale.

3.3 Workshop, conferenze, seminari

3.3.1 Workshop e conferenze organizzati e co-organizzati dall'ILC

L'Istituto è da sempre coinvolto nell'organizzazione di alcuni dei principali eventi scientifici del settore della Linguistica Computazionale. Tra gli eventi del 2015 sono da segnalare, in particolare:

- **Word Knowledge and Word Usage - Representations and Processes in the Mental Lexicon**
Conferenza finale di **NetWordS** (www.networds-esf.eu)
Pisa, Scuola Normale Superiore, 30 marzo – 1 aprile 2015
- **LRL 2015, Language Technologies in support of Less-Resourced Languages**
4° workshop biennale sulle lingue sottorappresentate (www.ilc.cnr.it/lrl2015/)
Poznań - Polonia (Adam Mickiewicz University), 28 novembre 2015

3.3.2 Partecipazione a comitati scientifici di conferenze

Numerosi ricercatori dell'ILC sono membri dei comitati scientifici di conferenze nazionali e internazionali nei settori della Linguistica Computazionale e delle Digital Humanities:

Federico Boschetti

- membro dei comitati scientifici per la valutazione dei contributi inoltrati alla conferenza *Corpus-based Research in the Humanities (CRH)*
- membro del comitato scientifico per l'organizzazione del convegno annuale 2016 dell'Associazione per l'informatica umanistica e la cultura digitale (AIUCD) a Venezia

Nicoletta Calzolari

- nomina come conference chair della conferenza internazionale *LREC 2016 (International Conference on Language Resources and Evaluation)*
- nomina come general chair della conferenza internazionale *COLING 2016* (Osaka)

Paola Cutugno

- membro del Comitato organizzatore e scientifico del *XIV Simposio Internacional de Comunicacion Social*, organizzato dal Centro de Lingüística Aplicada (CLA)

Riccardo Del Gratta

- membro del comitato scientifico per la valutazione dei contributi inoltrati alle conferenze *WLSI2015* e *LREC 2016*

Felice Dell'Orletta

- membro del comitato scientifico e revisore della *Conferenza Italiana di Linguistica Computazionale (CLIC-IT 2015)*
- membro del comitato scientifico e revisore della *Conference on Empirical Methods in Natural Language Processing (EMNLP) 2015*

Emiliano Giovannetti

- membro del program committee di *Language Resources and Evaluation Conference (LREC 2016)*

Lucia Marconi

- membro del Comitato organizzatore e scientifico del *XIV Simposio Internacional de Comunicacion Social*, organizzato dal Centro de Lingüística Aplicada (CLA)

Monica Monachini

- membro del Comitato di Programma *Language Resources and Evaluation Conference (LREC)*
- membro del Comitato di Programma della *Conferenza Italiana di Linguistica Computazionale (CLIC-IT 2015)*
- nomina come Workshop Chair della Conferenza Internazionale *COLING 2016*

Ouafae Nahli

- membro del comitato scientifico dell'*International Colloquium in Information Science and Technology (IEEE CIST'16)*, October 24-26, 2016, in Tangier-Assilah (Marocco)
- membro del Communication and Publicity Committee e Scientific Committee dell'*International Conference on Arabic Language Processing (ICALP)*

Vito Pirrelli

- membro del comitato scientifico della conferenza *SALC V - Scandinavian Association for Language and Cognition 2015*
- membro del comitato scientifico della conferenza internazionale *Décembrettes 9*
- membro del comitato scientifico per il *17th International Morphology Meeting (IMM 17)*
- programme chair della *Final NetWords Conference* (www.networks-esf.eu)

Claudia Soria

- membro del Comitato Scientifico del *Workshop TaLaRe 2015*, Caen, Francia
- membro del Comitato Scientifico della conferenza *Language & Technology Conference (LTC2015)*
- co-chair del workshop *Language Technologies in Support of Less-Resourced Languages (LRL 2015)*, Poznan, Polonia
- co-chair del *Collaboration and Computing for Under-Resourced Languages Workshop (CCURL II)* "Towards an Alliance for Digital Language Diversity", Portoroz (Slovenia), 23 maggio 2016
- area chair per il settore "Risorse Linguistiche" della *Conferenza Italiana di Linguistica Computazionale (CLIC-IT 2015)*
- membro del Program Committee di *Traitement Automatique des Langues Régionales de France et d'Europe (TALARE2015)*, Caen, Francia
- membro del Comitato Scientifico della conferenza *Language Resources and Evaluation Conference (LREC2016)*

3.3.3 Relazioni a convegni, conferenze, workshop e giornate di studio

Comunicazioni a convegni su invito

- **MAPLEX 2015 - Multiple Approaches for Multilingual frame Semantics wordnet generative**
Nicoletta Calzolari - *Language Resources Challenges at the Crossroads of Data, Language and Infrastructure* (relazione su invito)
Yamagata (Giappone), 9 -10 febbraio 2015
- **Quantitative Morphology Meeting**
Vito Pirrelli - *For a biologically-motivated model of the morphological lexicon*
Belgrado (Serbia), 10-11 luglio, 2015
- Giornata di studi **TAL e Beni Culturali**, volta a promuovere la cooperazione e la sinergia tra gli aspetti umanistici e gli aspetti tecnologici nel Trattamento Automatico della Lingua.
Francesca Frontini - *TAL per E-Books, libri elettronici ed enhanced books*
Gardone Riviera, Fondazione "Il Vittoriale degli Italiani", 14 settembre 2015
- **Third International Conference for the History and Philosophy of Computing (HAPOC 3)**
Alessandro Lenci, Simonetta Montemagni - *Words for Machines, Machines for Words. A brief history of computational linguistics and digital humanities in Pisa*
Pisa, 8-11 ottobre, 2015
- Giornata di studio **Il ritorno dello studio del greco nell'Europa del XV secolo: antiche e nuove prospettive per una ripresa degli studi classici**
Federico Boschetti - *Spunti di filologia cooperativa*
Venezia, Auditorium Santa Margherita, 13 novembre 2015
- **Commemorazione dell'Anniversario della morte del Libertador Simón Bolívar**

Manuela Sassi - *La lessicografia costituzionale bolivariana oggi*
Roma, Campidoglio, 17-18 dicembre 2015

Comunicazioni a convegni senza pubblicazione degli atti

- Workshop **WLSI 2015**, The Second International Workshop on Worldwide Language Service Infrastructure
Federico Boschetti - *Collaborative Philology on the way to Web Services: the case of CoPhiWordnet*
Kyoto (Giappone), Kyoto University, 22-23 gennaio 2015
- **MAPLEX2015 Multiple Approaches to Lexicon Conference**
Frontini F., Quochi V., Monachini M. - *Generative Lexicon and polysemy: inducing logical alternations*, 2015
Yamagata (Giappone), 9-10 febbraio 2015
- Conferenza internazionale **ELD 2015 - Digital Literary Studies**
Francesca Frontini, Carmen Brando, Clémence Jacquot, Marine Riguët and Vincent Jolivet - *Linked Open Data for the Annotation of Toponyms in Digital Editions*
Coimbra (Portogallo), Università di Coimbra, 14-15 maggio 2015
- **OCEANS 2015 MTS/IEEE**
Chiarella D., Bibuli M., Bruzzone G., Caccia M., Ranieri A., Zereik E., Marconi L., Cutugno P. - *Gesture-based Language for Diver-Robot Underwater Interaction*
Genova, 18-21 maggio 2015
- **Digital Humanities 2015**
Albanesi A., Bellandi A., Benotto G., Giovannetti E. - *Translation, Annotation and Knowledge Modelling of the Babylonian Talmud: the Traduco System*, 2015
Sydney (Australia), 29 giugno - 3 luglio 2015
- **Workshop Digital Humanities 2016**
Francesca Frontini e Carmen Brando (IGN) - *A place for Places: Current Trends and Challenges in the Development and Use of Geo-Historical Gazetteers*
Cracovia (Polonia), 11-16 luglio 2016
- Conferenza **DHANT - Digital Humanities and Antiquity**
Federico Boschetti - *Few notes on the domain and subdomains of Collaborative and Cooperative Philology*
Grenoble (Francia), 2-4 settembre 2015
- **JLC2015 - 8es Journées Internationales de Linguistique de Corpus**
Carmen Brando, Francesca Frontini, Alaa Abi Haidar and Jean-Gabriel Ganascia - *Reconnaissance d'entités nommées: adaptation au domaine de la littérature française du XIXe siècle*
Orléans (Francia), 2-4 settembre 2015
- Convegno **DiXiT 2015 - Technology, Software, Standards for the Digital Scholarly Edition**
Riccardo Del Gratta - *The role of digital scholarly editors in the design of components for cooperative philology*
L'Aia (Paesi Bassi), 14 - 18 settembre 2015
- Workshop **Digital Arabic & Digital Persian Research (Leipzig Workshop Week)**
Federico Boschetti - *Enhancing the Arabic OCR workflow*
Ouafae Nahli - *Improved Written Arabic Word Parsing through Orthographic, Syntactic and Semantic constraints*
Lipsia (Germania), Università di Lipsia, 17 dicembre 2015

3.3.4 Seminari interni

L'ILC promuove attività di studio e divulgazione della ricerca scientifica nei settori d'interesse attraverso una serie di incontri di natura prevalentemente interdisciplinare. Oltre ai seminari su temi specifici e alla discussione di articoli

scientifici, sono organizzate lezioni tenute da esperti esterni e brevi presentazioni interne, prevalentemente legate a progetti nazionali e internazionali in corso presso l'Istituto.

Tra le iniziative organizzate nel corso del 2015 si segnalano:

Data: 18 febbraio 2015

Titolo: [Sguardi sull'odissea dell'Omero italiano](#)

Relatore: Yuri Bizzoni

Descrizione: presentazione di uno studio che, adottando metodi di Semantica Distribuzionale, ha lo scopo di misurare come le traduzioni di Omero in lingua italiana fra il Diciottesimo e il Ventunesimo secolo tendano a fare scelte lessicali sempre più aderenti ai pattern distribuzionali dell'originale greco.

Data: 24 febbraio 2015

Titolo: [Aghi in un pagliaio e come trovarli? Il caso dell'accesso lessicale](#)

Relatore: Michael Zock, Università di Aix-Marseille, C.N.R.S. – Laboratorio LIF

Descrizione: In che modo un dizionario in formato digitale può essere trasformato e utilizzato come strumento di supporto automatico alla ricerca lessicale? I dizionari elettronici possono offrire un grande aiuto in tal senso. Conoscere una parola, cioè averla registrata in un database (sia esso rappresentato dalla mente umana o da un dizionario), non implica necessariamente la capacità di avere accesso a quella parola in tempo reale. Il computer può offrire un valido strumento per stimolare e accelerare questo processo attraverso l'uso di indici di associazione, percorsi pre-strutturati di navigazione del dizionario, contesti d'uso reale delle parole.

Data: 27 febbraio 2015

Titolo: [Echi poetici e visivi fra intertestualità e iconotestualità](#)

Relatore: Aرسالane Zarghili

Descrizione: illustrazione di come metodi computazionali di analisi testuale possano essere estesi e applicati agli ambiti d'interesse di chi si occupa del riconoscimento di immagini di alto valore storico e culturale. L'obiettivo è passare dalla ricerca del "detto" (tramite l'annotazione esplicita di temi e motivi testuali o visivi) alla ricerca del "non detto" (tramite il riconoscimento di strutture intrinsecamente presenti nell'oggetto di studio). Riflessione sulle peculiarità dell'epigrafia digitale, perché è proprio l'epigrafia il luogo specifico dove testo, oggetto visivo e contesto richiedono per definizione un approccio olistico e un trattamento integrato dei diversi livelli di informazione.

Data: 23 aprile 2015

Titolo: [Apprendimento discriminativo e morfologia implicita](#)

Relatore: Petar Milin, Università di Novi Sad, Facoltà di Filosofia, Dip. di Psicologia

Descrizione: presentazione di un modello computazionale di acquisizione della competenza morfologica costruito sulla base delle equazioni di Rescorla-Wagner. Il modello impara ad attivare più fortemente quei contrasti lessicali che sono effettivamente codificati nel segnale ortografico, senza decomporre il segnale stesso in una sequenza di morfemi. Il modello si dimostra capace di fare un uso ottimale del ricco potenziale di segnali sublessicali che servono a discriminare i vari contrasti morfologici codificati nell'input e utilizzati durante il riconoscimento delle parole.

Data: 22 maggio 2015

Titolo: [Valutazione della leggibilità e semplificazione dei testi in lingua basca presso Ixa Group](#)

Relatore: Itziar Gonzalez-Dios

Descrizione: illustrazione delle attività di Ixa Group, un gruppo di ricerca dell'Università dei Paesi Baschi che ha sviluppato risorse, tool e applicazioni per il basco e presentazione di alcune informazioni tipologiche e sociolinguistiche di questa lingua. In particolare, descrizione di ErreXail (un sistema per la valutazione della leggibilità che combina informazioni linguistiche e statistiche), EuTS (un sistema di semplificazione delle strutture sintattiche basato sulla conoscenza) e del corpus di testi semplificati in fase di costruzione.

Data: 11 giugno 2015

Titolo: [Analisi testuale e linguistica di lirici arcaici e adespoti giambici ed elegiaci: ipotesi di ricerca di applicazione della Filologia Computazionale al greco antico](#)

Relatore: Anika Nicolosi

Descrizione: nell'ambito di un'ipotesi di ricerca basata sull'applicazione della filologia computazionale al greco antico, discussione dei possibili sviluppi di un'analisi testuale e linguistica compiuta su una scelta di autori greci, partendo dalle rispettive edizioni moderne di riferimento, con particolare attenzione ai lirici arcaici e agli adespoti giambici ed elegiaci.

Data: 18 giugno 2015

Titolo: [L'elaborazione dei composti e dei sintagmi nominali in inglese. Evidenza basata su tentativi automatici di costruzione semantica.](#)

Relatori: Christina L. Gagné e Thomas L. Spalding, Università di Alberta, Dip. di Psicologia

Descrizione: panoramica sulle problematiche legate alla composizionalità lessicale e illustrazione di un quadro teorico volto a spiegare il ruolo dell'informazione relazionale nella combinazione concettuale. Illustrazione delle indagini più recenti sul ruolo della competizione relazionale nei composti morfologici semanticamente non trasparenti e descrizione del ruolo e della natura della competizione relazionale durante l'elaborazione sia di nuovi sintagmi modificatore-sostantivo sia di composti già conosciuti.

Data: 25 giugno 2015

Titolo: [Arricchimento semantico di edizioni digitali in TEI, riconoscimento e disambiguazione di menzioni di autori in testi di critica letteraria](#)

Relatore: Francesca Frontini

Descrizione: illustrazione dei lavori condotti per adattare ed estendere algoritmi di riconoscimento, classificazione e linking di entità nominate nel "Corpus Critique", un insieme di testi di critica letteraria francese che il laboratorio Labex OBVIL di Parigi sta pubblicando in edizione digitale. Lo scopo è produrre risorse testuali annotate per facilitare la ricerca nell'ambito della storia della critica letteraria e della storia delle idee in generale.

Data: 9 luglio 2015

Titolo: [Studio della complessità linguistica e prospettive applicative](#)

Relatore: Dominique Brunato

Descrizione: illustrazione dei presupposti teorici e della metodologia seguita per la costruzione di una risorsa per lo studio della semplificazione automatica e semi-automatica dei testi in lingua italiana. Si tratta di un primo passo per la creazione di un sistema di semplificazione semi-automatica flessibile, ovvero specializzato per lettori e domini diversi.

Data: 23 luglio 2015

Titolo: [Traduco: un ambiente web collaborativo per la traduzione assistita di testi antichi](#)

Relatore: Andrea Bellandi

Descrizione: illustrazione delle principali caratteristiche che differenziano il sistema Traduco, progettato e realizzato nell'ambito del Progetto Traduzione Talmud Babilonese, rispetto ad altri ambienti di traduzione assistita (CAT).

Data: 28 luglio 2015

Titolo: [Metodi e tecniche per il monitoraggio dell'evoluzione delle competenze linguistiche di apprendenti l'italiano come L1 nella scuola secondaria di primo grado](#)

Relatore: Stefan Richter

Descrizione: presentazione di metodi e tecniche per monitorare l'evoluzione delle abilità di scrittura degli studenti italiani che possono essere usati con successo in studi di pedagogia sperimentale. Illustrazione della metodologia messa a punto basandosi su algoritmi di apprendimento automatico applicata su un corpus di produzioni scritte raccolte in scuole secondarie di primo grado di Roma; descrizione del corpus di produzioni e delle caratteristiche linguistiche che possono essere automaticamente estratte e dei criteri di selezione.

Data: 30 ottobre 2015

Titolo: [Storylines: Modelli Computazionali per Big Data](#)

Relatore: Tommaso Caselli

Descrizione: illustrazione dei principi e delle problematiche per la creazione di storylines da grandi quantità di dati testuali secondo il paradigma dei Big Data. Con storylines si fa riferimento a un indice strutturato di descrizioni di eventi, relativi partecipanti e opinioni ordinato cronologicamente e rispetto a una certa tematica.

3.4 Pubblicazioni

3.4.1 Contributi in rivista

Barbagli A., Lucisano P., Dell'Orletta F., Montemagni S., Venturi G., *Il ruolo delle tecnologie del linguaggio nel monitoraggio dell'evoluzione delle abilità di scrittura: primi risultati*, in "Italian Journal of Computational Linguistics", Vol. 1, n. 1, 2015, pagg. 99-117

Del Gratta R., Frontini F., Khan F. e Monachini M., *Converting the PAROLE SIMPLE CLIPS Lexicon into RDF with lemon*, in "Semantic Web", n. 6, 2015, pagg. 387-392

Goggi S., Pardelli G., Giannini S., Biagioni S. e Battisti M., *La littérature grise des projets de recherche européens*, in "I2D - Information, données & documents. Pratiques & recherches", n. 52, 2015, pag. 34

Giannini S., Biagioni S., Goggi S. e Pardelli G., *Mapping Italian grey communities: what is there beyond the Academy?*, in "The Grey journal", n. 11, 2015, pagg. 17-28

Goggi S., Monachini M., Frontini F., Bartolini R., Pardelli G., De Mattei M., Bustaffa F., Manzella G., *Marine Planning and Service Platform (MAPS) An Advanced Research Engine for Grey Literature in Marine Science*, in "The Grey journal", n. 11, 2015, pagg. 171-178

Marzi C., Pirrelli V., *A Neuro-Computational Approach to Understanding the Mental Lexicon*, in "Journal of cognitive science (Seoul. Online)", n. 16, 2015, pagg. 493-535

3.4.2 Contributi in volume

Brando C., Frontini F. e Ganascia J., *Disambiguation of Named Entities in Cultural Heritage Texts Using Linked Data Sets*, in "New Trends in Databases and Information Systems – Communications in computer and information science", Springer International Publishing, n. 539, 2015, pagg. 505-514

Marchi S., *Greek Into Arabic, a Research Infrastructure Based on Computational Modules to Annotate and Query Historical and Philosophical Digital Texts. Part II: System Components and Features*, in "Digital Texts, Translations, Lexicons in the Web", Firenze, Leo Olschki editore, 2015, pagg. 43-60

Morgavi G., Nerino R., Marconi L., Cutugno P., Ferraris C., Cinini A. e Morando M., *An Integrated Approach to the Well-Being of the Elderly People at Home*, in "Ambient Assisted Living - Italian Forum 2014 - Series Biosystems & Biorobotics", Springer International Publishing, Vol. 11, 2015, pagg. 265-274

Pirrelli V., Ferro M. e Marzi C., *Computational complexity of abstractive morphology*, in "Understanding and Measuring Morphological Complexity", Oxford, Oxford University Press, 2015, pagg. 141-166

Simi M., Montemagni S. e Bosco C., *Harmonizing and merging Italian treebanks: Towards a merged Italian dependency treebank and beyond*, in "Harmonization and Development of Resources and Tools for Italian Natural Language Processing within the PARLI Project - Studies in Computational Intelligence 589", Springer International Publishing, 2015, pagg. 3-23

Soria C., *Assessing the effect of official recognition on the vitality of minority and regional languages: a case study from Italy*, in "Policy and Planning for Endangered Languages", Cambridge, Cambridge University Press, 2015, pagg. 123-137

3.4.3 Contributi in atti di convegno

Albanesi D., Bellandi A., Benotto G., Di Segni G., Giovannetti E., *When Translation Requires Interpretation: Collaborative Computer-Assisted Translation of Ancient Texts*, in "Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities" (LaTeCH 2015, Beijing, 30 luglio 2015), 2015, pagg. 84-88

- Barbagli A., Lucisano P., Dell'Orletta F., Montemagni S., Venturi G., *ClitA: un Corpus di Produzioni Scritte di Apprendenti l'Italiano L1 Annotato con Errori*, Torino, Accademia University Press, 2015, pagg. 31-35 (2nd Italian Conference on Computational Linguistics (CLiC-it), Trento, 3-4 dicembre 2015)
- Bellandi A., Bellusci A., Giovannetti E., *Computer Assisted Translation of Ancient Texts: the Babylonian Talmud Case Study*, in "NLPCS 2014: 11th International Workshop on Natural Language Processing and Cognitive Science", 2015 (Venezia, 27-29 ottobre 2014), Berlin/Munich, De Gruyter Saur, 2015
- Boschetti F., Del Gratta R., Del Grosso A., Monachini M., Diakoff H. e Nahli O., *Collaborative Philology on the way to Web Services: the case of CoPhiWordnet*, in "Proceedings of The Second International Workshop on Worldwide Language Service Infrastructure, WLSI 2015", 2015 (Kyoto, 22-23 gennaio 2015)
- Boukhaled M., Frontini F., Ganascia J., *A Peculiarity-based Exploration of Syntactical Patterns: a Computational Study of Stylistics*, in "CEUR workshop proceedings", Vol. 1410, 2015 (Workshop on Interactions between Data Mining and Natural Language Processing 2015 co-located with European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases -ECML PKDD 2015, Porto, Portogallo, 7 settembre 2015)
- Boukhaled M., Frontini F., Ganascia J., *Une mesure d'intérêt à base de surreprésentation pour l'extraction des motifs syntaxiques stylistiques*, in "Actes de La 22e Conférence Sur Le Traitement Automatique Des Langues Naturelles", 2015, pagg. 391-396 (22e Conférence Sur Le Traitement Automatique Des Langues Naturelles - TALN 2015, Caen, Francia, 22-25 giugno 2015)
- Brando C., Frontini F., Ganascia J., *Linked data for toponym linking in French literary texts*, in "GIR '15 Proceedings of the 9th Workshop on Geographic Information Retrieval", 2015 (Parigi, 26-27 novembre 2015)
- Brunato D., Dell'Orletta F., *ISACCO: a corpus for investigating spoken and written language development in Italian school-age children*, Torino, Accademia University Press, 2015, pagg. 31-35 (2nd Italian Conference on Computational Linguistics (CLiC-it), Trento, 3-4 dicembre 2015)
- Brunato D., Dell'Orletta F., Venturi G., Montemagni S., *Design and Annotation of the First Italian Corpus for Text Simplification*, 2015, in "Proceedings of LAW IX - The 9th Linguistic Annotation Workshop", 2015 (Denver, Colorado, 5 giugno 2015), pagg. 31-34
- Cignoni L., Fornaciari G., Fornaciari A., *Many hands make light work: collaborative CLIL activities for University courses in Medieval funerary archaeology*, 2015, pagg. 2271-2279 (9th International Technology, Education and Development Conference Madrid, 2-4 marzo 2015)
- Cutugno P., Lucentini R., Marconi L., Chiarella D., *Relaciones sin violencia: lenguaje, estereotipos y sexismo benévolo*, in "Comunicación Social: retos y perspectivas", Vol. I, 2015, pagg. 200-204 (XIV Simposio Internacional de Comunicación Social: retos y perspectivas, Santiago de Cuba, 19 - 23 gennaio 2015)
- De Felice I., *GraDes: a corpus of grasp descriptions*, 2015 (AISV 2015, Bologna, 28-30 gennaio 2014),
- Del Gratta R., Frontini F., Monachini M., Pardelli G., Russo I., Bartolini R., Goggi S., Khan F., Quochi V., Soria C., Calzolari N., *Visualising Italian Language Resources: a Snapshot*, in "Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it 2015", 2015, pagg. 100-104 (II Italian Conference on Computational Linguistics CLiC-it 2015, Trento, 3-4 dicembre 2015)
- Ferro M., Marzi C., Pirrelli V., *Lexical parsability and morphological structure*, in "MMM9 - Morphology and Semantics", 2015, pagg. 22-37 (Ninth Mediterranean Morphology Meeting, Dubrovnik, Croatia, 15-18 settembre 2013)
- Frontini F., *Analyse et extraction des motifs syntaxiques dans la prose de Robert Challe et de ses apocryphes*, 2015 (Robert Challe: approches numériques des questions d'auctorialité, Parigi, 28 marzo 2015)
- Frontini F., *Mining for characterising patterns in literature using correspondence analysis: an experiment on French novels*, 2015 (Göttingen Dialog in Digital Humanities, Gottinga, Germania, 14 luglio 2015)
- Frontini F., *Trattamento automatico del linguaggio per le Digital Humanities. Riconoscimento e disambiguazione di menzioni di autori in testi di critica letteraria*, 2015 (Seminario di Cultura Digitale, Pisa, 14 novembre 2015)

- Frontini F., *What makes them different: the extraction of distinctive linguistic patterns for the protagonists of Molière's plays*, 2015 (Cycle des séminaires ILES LIMSI 2015, Parigi, 3 febbraio 2015)
- Frontini F., Bénard E., *The Syntax of Stage. Studying Linguistic Patterns in Molière*, 2015 (Göttinger philologisches Forum, Gottinga, Germania, 3 dicembre 2015)
- Frontini F., Brando C., Ganascia J., *Semantic Web based Named Entity Linking for Digital Humanities and Heritage Texts*, in "CEUR workshop proceedings", Vol-1364, 2015, pagg. 77-88 (Semantic Web for Scientific Heritage Workshop - SW4SH 2015, Portorose, Slovenia, 1 giugno 2015)
- Frontini F., Brando C., Ganascia J., *Domain-adapted named-entity linker using Linked Data*, in "CEUR workshop proceedings", Vol. 1386, 2015 (Workshop on NLP Applications: Completing the Puzzle co-located with the 20th International Conference on Applications of Natural Language to Information Systems - NLDB 2015, Passau, Germania, 17-19 giugno 2015)
- Frontini F., Boukhaled M. A., Ganascia J., *Linguistic Pattern Extraction and Analysis for Classic French Plays*, in "Atti della Journée ConSciLa (Confrontations en Sciences du Langage) Grammaire des genres et des styles: quelles approches privilégier?", 2015 (Parigi, Francia, 16 gennaio 2015)
- Frontini F., Boukhaled M. A., Ganascia J. G., *Moliere's Raisonneurs: a quantitative study of distinctive linguistic patterns*, in "Corpus Linguistics 2015 - Abstract Book 2015", 2015, pagg. 114-117 (Corpus Linguistics 2015, Lancaster, Regno Unito, 21-24 luglio 2015)
- Giannini S., Biagioni S., Goggi S., Pardelli G., *Grey Literature citations in the age of Digital Repositories and Open Access*, 2015, in "GL17 Program Book, GL-conference series", Vol. 17, 2015, pagg. 109-110 (XVII International Conference on Grey Literature. A New Wave of Textual and Non-Textual Grey Literature, Amsterdam, Paesi Bassi, 1-2 dicembre 2015)
- Giannini S., Biagioni S., Goggi S., Pardelli G., *Mapping Italian grey communities: what is there beyond the Academy?*, in "Grey Literature Lobby: Engines and Requesters for Change, The GL-conference series. Conference proceedings", Vol. 16, 2015, pagg. 17-29 (GL16 - XVI International Conference on Grey Literature Grey Literature Lobby: Engines and Requesters for Change, Washington DC, USA, 8-9 dicembre 2014)
- Goggi S., Monachini M., Frontini F., Bartolini R., Pardelli G., De Mattei M., Bustaffa F., Manzella G., *Marine Planning and Service Platform (MAPS): An Advanced Research Engine for Grey Literature in Marine Science*, in "Grey Literature Lobby: Engines and Requesters for Change, The GL-conference series. Conference proceedings", Vol. 16, 2015, pagg. 108-114 (XVI International Conference on Grey Literature Grey Literature Lobby: Engines and Requesters for Change, Washington DC, USA, 8-9 dicembre 2014)
- Goggi S., Pardelli G., Bartolini R., Frontini F., Monachini M., Manzella G., De Mattei M., Bustaffa F., *A semantic engine for grey literature retrieval in the oceanography domain*, in "GL17 Program Book", Vol. 17, 2015, pagg. 76-77 (XVII International Conference on Grey Literature. A New Wave of Textual and Non-Textual Grey Literature, Amsterdam, 1-2 dicembre 2015)
- Goggi S., Pardelli G., Sassi M., Giannini S., Biagioni S., *A terminological survey on the titles of the Seventh Framework Programme (FP7)*, 2015, pagg. 223-227 (XIV International Symposium on Comunicación Social: retos y perspectivas, Santiago de Cuba, 19-23 gennaio 2015)
- Khan F., Frontini F., *Using Ontologies to Model Polysemy in Lexical Resources*, in "Proceedings of the Workshop on Language and Ontologies", 2015 (Londra, Regno Unito, 14 aprile 2015)
- Marconi L., Cutugno P., Lucentini R., Chiarella D., Morgavi G., Morando M., *La tecnología como sostén de la organización de datos lingüísticos concernientes a las plantas medicinales*, in "Comunicación Social: retos y perspectivas", Vol 2, 2015, pagg. 605-609 (XIV Simposio Internacional de Comunicación Social: retos y perspectivas, Santiago de Cuba, 19 - 23 gennaio 2015)
- Marzi C., Ferro M. e Pirrelli V., *Lexical emergentism and the "frequency-by-regularity" interaction*, in "Word Knowledge and Word Usage", in "CEUR Workshop Proceedings", Vol. 1347, 2015, pagg. 37-41 (NetWordS Final Conference on

Word Knowledge and Word Usage: Representations and Processes in the Mental Lexicon, Pisa, 30-31 marzo, 1 aprile 2015)

Montemagni S., Wieling M., Nerbonne J., *The contribution of dialectometry to the study of the dialects of Italy. A case study on Tuscan*, abstract con peer review (Italian Dialect Meeting 2015 & CIDSM X, Leida, Paesi Bassi, 23 giugno 2015)

Nahli O., Marchi S., *Improvements on Written Arabic Word Parsing through Orthographic, Syntactic and Semantic constraints*, 2015, pagg. 210-214 (II Italian Conference on Computational Linguistics CLiC-it 2015, Trento, 3-4 dicembre)

Pirrelli V., Nahli O., Boschetti F., Del Gratta R. e Marzi C., *Computational Linguistics and Language Physiology: Insights from Arabic NLP and Cooperative Editing*, in "Third AIUCD Annual Conference - Humanities and Their Methods in the Digital Ecosystem, ICPS: ACM International Conference Proceeding Series", 2015, pagg. 1-8 (Third AIUCD Annual Conference - Humanities and Their Methods in the Digital Ecosystem, Bologna, 18-19 settembre 2015)

Richter S., Cimino A., Dell'Orletta F., Venturi G., *Tracking the Evolution of Written Language Competence: an NLP-based Approach*, 2015, pagg. 236-240 (2nd Italian Conference on Computational Linguistics - CLiC-it, Trento, 3-4 dicembre 2015)

Russo I., Caselli T. e Monachini M., *Extracting and Visualising Biographical Events from Wikipedia*, in "BD2015 Biographical Data in a Digital World 2015", Torino, Accademia University Press, Vol. 1399, 2015, pagg. 111-115 (BD2015 Biographical Data in a Digital World 2015, Amsterdam, 9 aprile 2015)

Soria C., *Towards an Alliance for Digital Language Diversity: Vision, Goals, and Challenges*, in "Linguistic and Cultural Diversity in Cyberspace - Proceedings of the 3rd International Conference 2015", pagg. 111-125 (3rd International Conference on Linguistic and Cultural Diversity in Cyberspace, Yakutsk, Russian Federation, 30 giugno-3 luglio 2014)

Soria C., Russo I., *The Digital Language Diversity Project*, in "Proceedings of the 7th Language & Technology Conference", 2015, pagg. 329-332 (Poznan, Polonia, 27-29 novembre 2015)

Venturi G., Bellandi T., Dell'Orletta F. e Montemagni S., *NLP-Based Readability Assessment of Health-Related Texts: a Case Study on Italian Informed Consent Forms*, 2015, pagg. 131-141 (Sixth International Workshop on Health Text Mining and Information Analysis (Louhi), Lisbona, 17 settembre 2015)

Venturi G., Rinnone S., Montemagni S., Sassi M., Terranova G., Flore E. e Bellandi T., *Language technologies for automatic readability assessment of health-related Information: a preliminary investigation into the informed consent forms used in a regional health service*, 2015 (ISCOM 2015 Conference: "The Golden Bridge: Communication and Patient Safety", Montecatini Terme, 15-16 giugno 2015)

3.4.4 Curatele

Pirrelli V., Marzi C. e Ferro M., *Proceedings of the NetWordS Final Conference on Word Knowledge and Word Usage: Representations and Processes in the Mental Lexicon*, Vol. 1347, CEUR Workshop Proceedings, 2015, pagg. 1-189

Italian Journal of Computational Linguistics - Special Issue "Emerging Topics at the First Italian Conference on Computational Linguistics", a cura di Roberto Basili, Alessandro Lenci, Bernardo Magnini, Simonetta Montemagni Associazione Italiana di Linguistica Computazionale, Vol. 1, n.1, December 2015

3.4.5 Note Editoriali

Roberto Basili, Alessandro Lenci, Bernardo Magnini, Simonetta Montemagni
Italian Journal of Computational Linguistics (IJCoL), Volume 1, Number 1, December 2015, pp. 7-15

3.4.6 Altri prodotti della ricerca

Baroni P., 2015-1-IT02-KA204-015090 DLDP: *Interactive Web Site*, 2015.

3.5 Internazionalizzazione

L'Istituto promuove l'internazionalizzazione della ricerca scientifica e tecnologica nel settore della Linguistica Computazionale: partecipa a programmi di ricerca e a organismi a livello internazionale; fornisce competenze scientifiche su richiesta di autorità governative; garantisce la collaborazione con enti ed istituzioni di altri Paesi nel campo scientifico-tecnologico e nella definizione della normativa tecnica.

L'ILC è il referente tecnologico nazionale (Technology NAP) dell'azione ELRC - European Language Resource Coordination e ospita infrastrutture di ricerca al servizio della comunità per la condivisione dei risultati delle attività di ricerca, al fine di promuovere l'utilizzo degli standard e delle buone pratiche, nonché la diffusione e il riutilizzo delle risorse.

Oltre ad essere esecutore dell'infrastruttura italiana CLARIN-IT, nonché National Representative di CLARIN-ERIC, l'Istituto, in qualità di partner principale del network europeo Meta-Net, ha contribuito allo sviluppo di uno dei core nodes (ora ospitato presso l'Istituto) della federazione MetaShare per la condivisione di risorse linguistiche e alla definizione del set di metadati per la loro descrizione, mettendo anche a disposizione un primo nucleo di risorse sulla piattaforma stessa.

Da segnalare, inoltre, che l'ILC ha coordinato il progetto NetWords - European Network on Word Structure (ESF Research Net-working Programme), con lo scopo di indagare approcci interdisciplinari mirati alla comprensione della struttura della parola nelle lingue europee e di sviluppare una rete scientifica per lo studio interdisciplinare dell'impatto della struttura morfologica sull'organizzazione del lessico mentale. Riunendo esperti di diversi settori di ricerca e di differenti ambiti teorici, il progetto ha inteso far progredire la conoscenza attuale dell'evidenza teorica, psicolinguistica, computazionale e neurofisiologica nell'ambito della struttura e dell'elaborazione delle parole, al fine di promuovere nuovi metodi di ricerca e di valutazione per le architetture della grammatica e la fisiologia del linguaggio. Attualmente la rete collega più di 40 laboratori e istituti di ricerca distribuiti in 16 paesi europei e ha contatti con Canada e Israele.

Nel corso del 2015 presso l'Istituto sono stati ospitati *visiting scholars* provenienti da diverse Università europee per attività di insegnamento e/o ricerca:

- **Anja Weingart - University of Göttingen**
Modellazione e codifica di risorse lessico-ontologiche diacroniche; elaborazione dello schema dell'applicativo Pinakes nella sua istanza per il lessico DiTMAO di occitano antico
Referente ILC: Emilio Giovannetti
gennaio 2015
- **Michael Zock - Aix-Marseille University**
Modelli computazionali del lessico mentale e tecnologie di supporto alla comunicazione verbale
Referente ILC: Vito Pirrelli
22-28 febbraio 2015
- **Petar Milin - University of Novi Sad**
Studio comparativo di modelli di apprendimento correlativo del lessico mentale
Referente ILC: Vito Pirrelli
19-25 aprile 2015
- **Christina L. Gagné e Thomas L. Spalding - University of Alberta**
Modelli funzionali di rappresentazione ed elaborazione di parole composte
Referente ILC: Vito Pirrelli
15-24 giugno 2015
- **Javier Enrique Díaz-Vera - University of Castilla-La Mancha**
Annotazione di un corpus di linguaggio emozionale e rappresentazione dello shift diacronico in lessici computazionali, con particolare riferimento all'inglese antico
Referente ILC: Monica Monachini
ottobre-dicembre 2015

- **Martijn Wieling - University of Groningen, Computer Science**
Estensione della risorsa dialettale online ALT-Web con funzionalità di analisi dialettometrica dei dati dialettali
Referente ILC: Simonetta Montemagni
ottobre 2015

4 Attività di alta formazione

4.1 Corsi universitari

- **Università di Pisa - Informatica Umanistica (classe L-10)**
Corso di Laurea Triennale: insegnamento di Linguistica Computazionale (12 CFU)
Docente: Felice Dell'Orletta
a.a. 2015/2016, I semestre
- **Università di Pisa - Informatica Umanistica (classe L-10)**
Corso di Laurea Magistrale: insegnamento di Linguistica Computazionale II (6 CFU)
Docenti: Simonetta Montemagni, Giulia Venturi
a.a. 2014/2015, II semestre
- **Venice International University**
Corso *Digital Tools for Humanities* nell'ambito del Globalization Program
Docente: Federico Boschetti
a.a. 2015/2016, I semestre
- **Nicoletta Calzolari**
membro del Consiglio di Dottorato di Linguistica dell'Università di Pisa

4.2 Summer School

- **International Summer School DARIAH**
Francesca Frontini - *Indexing Names in Digital Editions*
Como, Villa Vigoni, giugno 2015
- **First Edition of the Maghrebi Natural Language Processing (NLP) Summer School**
Eva Sassolini - *Texts and corpora: acquisition, representation, markup and tagging, types of corpora*
Eva Sassolini - NLP basic techniques and applications. Morpho-syntactic and semantic tagging; Information retrieval and extraction applications
Ouafae Nahli - *Arabic NLP: Special Challenges*
Ouafae Nahli - *Special Applications of Arabic NLP in the Humanities: "Greek into Arabic" project; "Memorata Poetis" project*
Ouafae Nahli *Invited Talk: "Le traitement lexical dans une perspective neuro-computationnelle."*
Ifrane (Marocco), Al Akhawayn University, 28 giugno – 10 luglio 2015
- **International Summer School of Scientific Communication**
Vito Pirrelli - *Interdisciplinarity in language sciences: an integrative perspective*
Spalato (Croazia), 30 giugno - 3 luglio, 2015
- **International Summer School LEX 2015 - MANAGING LEXICAL RESOURCES IN THE SEMANTIC WEB**
Giulia Venturi - *Natural Language processing and legal knowledge extraction*
Ravenna, Università di Bologna - Sede di Ravenna, 10 settembre 2015

4.3 Seminari invitati presso altre istituzionali nazionali e internazionali

- **Accademia della Crusca, Villa di Castello**
Giornata di studio: Presente e futuro di un corpus dell'italiano
Vito Pirrelli - Frequenza, struttura e processi morfologici nel lessico mentale: una prospettiva neo-conneccionista
Firenze, 14 dicembre 2015
- **Università di Pisa**
Seminario di cultura digitale: Il problema dell'elaborazione lessicale in una prospettiva neuro-computazionale
Corso seminariale della Laurea Magistrale in Informatica Umanistica
Relatore: Claudia Marzi
25 marzo 2015
- **Università di Cadice, Dip. di Filologia francese e inglese** (Cadice, Spagna)
Seminario: Natural Language Processing and Linguistic Profiling of Texts: methods and applications
Relatore: Giulia Venturi
13 maggio 2015
- **Università Pablo de Olavide, Facoltà di Giurisprudenza** (Siviglia, Spagna)
Seminario: Construction of corpora of legal texts: what, how, what for
Relatore: Giulia Venturi
14 maggio 2015
- **Paris Sorbonne, Labex Observatoire de la vie littéraire – OBVIL** (Parigi, Francia)
Seminario: Digital Humanities Research at the Institute of Computational Linguistics “A. Zampolli”: DH@ILC-CNR
Relatore: Simonetta Montemagni
20 maggio 2015
- **Università di Pisa, Dip. di Scienze Politiche**
Giornata di studio e formazione “La qualità degli atti normativi e amministrativi”
Le tecnologie linguistico-computazionali per la leggibilità della comunicazione istituzionale
Relatori: Dominique Brunato e Giulia Venturi
22 ottobre 2015
- **Università di Pisa – Dip. di Informatica**
Seminario: Computational Linguistics and Language Physiology: insights from word processing
Relatori: Claudia Marzi, Ouafae Nahli, Vito Pirrelli
25 marzo 2015
- **Università di Pisa**
Seminario di cultura digitale: Trattamento automatico del linguaggio per le Digital Humanities. Riconoscimento e disambiguazione di menzioni di autori in testi di critica letteraria
Centro interdipartimentale di formazione e ricerca “Laboratorio di Cultura Digitale”
Relatore: Francesca Frontini
2015
- **Scuola Normale Superiore, Laboratorio di Linguistica “G. Nencioni”**
Seminario: Metodi computazionali per lo studio della variazione dialettale. Risultati e prospettive di sviluppo
Relatori: Simonetta Montemagni
11 dicembre 2015

4.4 Tesi di laurea

▪ Supervisione di tesi di laurea triennale

Cesare Cei

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Oralità e scrittura nella scuola primaria. Costruzione e analisi linguistica di un corpus di produzioni scritte e orali

Relatore: Felice Dell'Orletta

Jonathan Cuomo

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Amore, occhi, vita e altri universi. Un viaggio nella canzone d'autore. Analisi linguistico-computazionale del canzoniere di De André, De Gregori, Battiato

Relatore: Felice Dell'Orletta

Francesca Danti

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Diversamente Italiano – Sviluppo di un'applicazione Web di Serious Gaming

Correlatore: Claudia Soria

Relatore: Andrea Marchetti (IIT-CNR)

Niccolò Degl'Innocenti

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Comunicazione e analisi linguistica: il caso di Facebook Messenger

Correlatore: Felice Dell'Orletta

Relatore: Mirko Tavoanis

Michele Mallia

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Correttore ortografico applicato a testi storici

Correlatore: Felice Dell'Orletta

Relatore: Alessandro Lenci

Simona Ortolani

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Voci d'Italia

Correlatore: Claudia Soria

Relatore: Maurizio Tesconi (IIT-CNR)

Giulia Pieri

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Fenomeni di complessità sintattica: uno studio linguistico-computazionale sull'ordinamento delle strutture grammaticali all'interno di diverse varietà linguistiche

Relatore: Felice Dell'Orletta

Federica Semplici

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Information Extraction applicata a testi storici

Correlatore: Felice Dell'Orletta

Relatore: Alessandro Lenci

Irene Sucameli

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Analisi computazionale delle relazioni semantiche: uno studio della lingua italiana

Correlatore: Felice Dell'Orletta

Relatore: Alessandro Lenci

Valeriya Slovikovskaya

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Domain Adaptation for Part-of-Speech Tagging. Quality scoring of automatically tagged tweets

Relatore: Felice Dell'Orletta

Giulia Zamponi

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: La titolazione dei quotidiani online: analisi linguistica

Correlatore: Felice Dell'Orletta

Relatore: Mirko Tavosanis

- **Supervisione di tesi di laurea specialistica**

Yuri Bizzoni

Università di Pisa, Informatica Umanistica

Titolo: The Italian Homer: the evolution of translation patterns between the XVIII and the XXI century

Abstract: La tesi descrive metodologie informatiche per allineare testi bilingui (greco - Italiano) e molteplici traduzioni dello stesso testo, al fine di osservare trends traduttivi nel corso dei secoli.

Relatore: Alessandro Lenci

Correlatori: Federico Boschetti, Andrea Taddei

Lorenzo Marinelli

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Studio della complessità e della semplificazione linguistica a partire da un'analisi computazionale di un corpus parallelo di testi italiani

Relatore: Felice Dell'Orletta

Stefan Richter

Università di Lipsia - Informatica

Titolo: Methods and techniques for tracking the evolution of language competence in L1 and L2 language learners at school

Relatore: Felice Dell'Orletta

Mariagiovanna Scarale

Università di Pisa - Informatica Umanistica, Dip. di Filologia Letteratura Linguistica

Titolo: Il web 2.0 e la blogosfera italiana. Metodi statistici applicati ai corpora dell'italiano dei blog giornalistici e letterari

Controrelatore: Felice Dell'Orletta

Relatore: Mirko Tavosanis

- **Supervisione di tesi di laurea specialistica (in corso)**

Alberto Stefanini

Università degli Studi di Parma

Titolo: Indagine sulle pratiche d'uso di risorse e strumenti digitali nell'ambito degli studi di filologia classica

Relatore: Anika Nicolosi

Correlatore: Monica Monachini

- **Supervisione di tesi di dottorato**

Angelo Mario Del Grosso

Università di Pisa - Dottorato di Ricerca in Ingegneria dell'Informazione

Titolo: Designing a Library of Components for Textual Scholarship

Tutor: Andrea Bozzi, Federico Boschetti, Emiliano Giovannetti, Francesco Marcelloni

Younes Lasri

Sidi Mohamed Ben Abdellah University of Fez, Faculty of Sciences and Technology, PhD in Computer Science

Titolo: Optical Character Recognition (OCR) applied to Arabic printed texts

Tutor: Arsalan Zarghili

Co-tutor: Federico Boschetti

Status: in corso

4.5 Tesi di dottorato

▪ Tesi di dottorato discusse nel 2015

Giulia Benotto

Università di Pisa - Dottorato di Ricerca in Filologia, letteratura e linguistica

Titolo: Distributional Models for Semantic Relations: A Study on Hyponymy and Antonymy

Abstract: Studio ed esplorazione delle potenzialità e dei limiti dell'approccio distribuzionale rispetto alla semantica lessicale. Analisi delle misure distribuzionali sviluppate per il riconoscimento delle relazioni semantiche paradigmatiche allo stato dell'arte per valutare il grado di successo riportato da tali misure. I dati raccolti sono utilizzati per sviluppare nuove misure distribuzionali atte a classificare la relazione di iponimia/iperonimia e la relazione di antonimia.

Dominique Brunato

Università degli studi di Siena - Dottorato di Ricerca in Scienze Cognitive (indirizzo Linguistica)

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

Abstract: Monitoraggio linguistico di varietà di lingua speciale (linguaggio amministrativo) basato su analisi generate da strumenti di Trattamento Automatico del Linguaggio; definizione di metodologie computazionali per la valutazione della complessità linguistica e della semplificazione dei testi.

Irene De Felice

Università di Pisa - Dottorato in Discipline Umanistiche - Programma di ricerca in linguistica generale, storica, applicata, computazionale e delle lingue moderne

Titolo: Language and Affordances

Abstract: Scopo della ricerca è esaminare, all'interno dello specifico dominio delle azioni di presa, se il modo in cui i soggetti descrivono linguisticamente la presa di un dato oggetto sia influenzato dagli stessi fattori che risultano coinvolti nella modulazione dell'attività cerebrale in relazione alla presentazione visiva di oggetti manipolabili. Lo studio evidenzia come il concetto di affordance, ampiamente studiato in psicologia e neuropsicologia, possa essere efficacemente applicato in linguistica, e come un punto di vista prettamente linguistico su questo tema possa gettare nuova luce sui risultati emersi in altri campi di ricerca.

Angelo Mario Del Grosso

Università di Pisa - Dottorato di Ricerca in Ingegneria dell'Informazione

Titolo: Designing a Library of Components for Textual Scholarship

Abstract: Il lavoro affronta temi legati all'applicazione di nuove tecnologie, di metodologie informatiche e di progettazione software volti allo sviluppo di strumenti innovativi per le Digital Humanities (DH). In particolare, questo contributo definisce alcuni specifici requisiti relativi al dominio del Literary Computing e al settore del Digital Textual Scholarship.

▪ Tesi di dottorato in corso nel 2015

Andrea Cimino

Università di Pisa - Dottorato di Ricerca in Ingegneria dell'Informazione

Titolo: Strumenti e metodologie basati su natural language processing per l'analisi automatica di documenti tecnici

Tema: Creazione di una pipeline di analisi linguistica per l'analisi automatica di brevetti focalizzata sull'estrazione di entità di tipo utenti, vantaggi e svantaggi che può essere integrata in applicazioni di alto livello. L'estrazione di informazioni dai brevetti permette di creare una serie di applicazioni come, ad esempio, strumenti per identificare trends tecnologici.

Mustafa Khalfi

Università Sidi Mohamed Ben Abdellah (Marocco), Facoltà di Scienze e Tecnologie, Dip. di Informatica

Titolo: Acquisition du lexique medioevale arabe Al-qamuws al-muhiyt en Lemon

Correlatore: Ouafae Nahli

Roberta Lucentini

Università degli Studi di Genova - Dottorato di Ricerca in Arti, Spettacolo e Tecnologie Multimediali

Titolo: La catalogazione del patrimonio storico artistico quale strumento di tutela dei beni culturali. Due casi studio relativi alla grafica per la produzione tessile e il teatro: la Manifattura Italiana Tessuti Artistici (M.I.T.A.) e il Museo Luzzati

Tema: Lo studio ha indagato la storia della tutela e della catalogazione dei beni culturali in Italia a partire dal 1800 ai giorni nostri. Ha definito lo stato dell'arte della catalogazione in Italia a partire dagli anni Sessanta del '900 definendo gli standard e i metodi di catalogazione presenti sia a livello nazionale che internazionale. E' stata affrontata la catalogazione informatizzata di due archivi museali di arte contemporanea: uno dedicato ad una manifattura tessile di particolare rilievo nazionale, la M.I.T.A.; il secondo dedicato ad un importante artista e scenografo del Novecento, Emanuele Luzzati. Per la seconda attività catalogografica, è stato strutturato un sistema informatico di catalogazione ispirato agli standard ICCD, proponendo e sperimentando campi integrativi relativi alle attività teatrali.

Ouafae Nahli

Università: Università degli Studi di Roma "La Sapienza" - Dottorato in Lingua Araba

Titolo: Verso un'ontologia della cultura araba-islamica

Tema: Sviluppo di una rete di conoscenze per la cultura islamica araba sulla base di un processo di estrazione automatica dei dati da testi classici fondamentali. La rete di conoscenze sarà costituita da una ricca ontologia formale legata ad ontologie "general-purpose" esistenti (SUMO).

4.6 Tirocini

▪ **Saint Petersburg University - Venice International University**

Tema: Categorizzazione degli errori nei synsets di AncientGreekWordNet relativi ai colori (150 ore)

Tirocinante: Lidia Yatluk

Tutor: Federico Boschetti

▪ **Università di Pisa, Lingue e letterature straniere**

Tema: Creazione di una risorsa digitale collegata ad AncientGreekWordNet per lo studio di neologismi neoclassici in Italiano e Inglese (150 ore)

Tirocinante: Eleonora Cecchini

Tutor: Federico Boschetti

▪ **Università di Pisa, Lingue e letterature straniere**

Tema: Implementazione di semplici procedure informatiche per lo studio delle riforme ortografiche della lingua tedesca (150 ore)

Tirocinante: Nicola Nunes

Tutor: Federico Boschetti

▪ **Università di Pisa, Informatica Umanistica, Dip. di Filologia Letteratura Linguistica**

Tema: Costruzione di un corpus annotato di produzioni orali e scritte di bambini italiani frequentanti la scuola primaria

Tirocinante: Cesare Cei

Tutor: Felice Dell'Orletta

- **Università di Pisa, Informatica Umanistica, Dip. di Filologia Letteratura Linguistica**
Tema: *Analisi qualitativa e specializzazione di una risorsa per la semplificazione automatica del testo*
Tirocinante: Giulia Pieri
Tutor: Felice Dell'Orletta
- **Università di Pisa, Dip. di Filologia, Letteratura e Linguistica**
Tema: *Organizzazione della struttura dei dizionari di ambito romanzo*
Tirocinante: Erminio Maraia
Tutor: Emiliano Giovannetti
- **Università di Pisa, Informatica Umanistica, Dip. di Filologia Letteratura Linguistica**
Tema: *Progettazione e implementazione di una base di dati per la visualizzazione di metadati descrittivi relativi alle lingue regionali e minoritarie parlate sul territorio italiano*
Tirocinante: Simona Ortolani
Tutor: Claudia Soria
- **Università di Pisa, Informatica Umanistica, Dip. di Filologia Letteratura Linguistica**
Tema: *Sviluppo di un'applicazione web di Serious Gaming per la raccolta dati sulla percezione della somiglianza linguistica*
Tirocinante: Francesca Danti
Tutor: Claudia Soria

4.7 Visiting students

- *Commessa "Tecnologie della lingua per il Patrimonio Culturale" – Modulo "Tecniche linguistico-semantiche per il supporto alla traduzione e alla consultazione di testi"*
 - Marianne Reboul - **Paris- Sorbonne University - Paris IV**, Labex OBVIL (Francia)
gennaio-marzo 2015
- *ItaliaNLP Lab*
 - Stefan Richter - **University of Leipzig** (Germania)
marzo-agosto
 - Itziar Gonzalez-Dios - **University of the Basque Country**, Linguistics, IXA Group (Spagna)
aprile-giugno 2015

4.8 Convenzioni nell'ambito della formazione

- Accordi preliminari per la **Convenzione per attività alternanza scuola lavoro**
Supervisione di stagisti e studenti di liceo in stage di formazione alternanza scuola-lavoro presso l'ILC per attività nel settore delle Digital Humanities.
Liceo Classico Scientifico XXV Aprile di Pontedera (PI)
Referente ILC: M. Monachini
- **Convenzione Unica di Tirocinio Formativo - Università di Parma**
Applicazione di metodi e strumenti della filologia computazionale allo studio del greco antico
Referente ILC: M. Monachini

4.9 Formazione interna

Per quanto riguarda le attività di formazione del personale, l'Istituto ha promosso un corso di inglese scientifico (con incontri a cadenza settimanale), tenuto dalla ricercatrice dell'ILC Laura Cignoni, madrelingua inglese.

5 Attività editoriali

Le attività editoriali dell'Istituto vanno dalla direzione scientifica di riviste alla partecipazione ai comitati scientifici e attività redazionali.

Direzione scientifica di riviste:

- **ITALIAN JOURNAL OF COMPUTATIONAL LINGUISTICS (IJCoL)**

Nata nel 2015 come iniziativa editoriale dell'Associazione Italiana di Linguistica Computazionale (AIRC), IJCoL si propone come forum aggiornato di discussione attorno alla Linguistica Computazionale, con l'obiettivo di alimentare sinergie tra studi legati ad aree diverse del trattamento automatico del linguaggio. IJCoL si propone come continuazione ideale della rivista *Linguistica Computazionale*, fondata nel 1981 da Antonio Zampolli e non più pubblicata dal 2006, che ha rappresentato il primo e fino ad oggi l'unico forum autorevole per le diverse anime della linguistica computazionale in Italia. La rivista copre temi che ruotano attorno a linguaggio e computazione, affrontati da prospettive diverse, ad esempio: trattamento e apprendimento automatico del linguaggio; modelli computazionali del linguaggio, della cognizione e della variazione linguistica; acquisizione di conoscenza da testi; costruzione di risorse linguistiche; sviluppo di infrastrutture per l'interoperabilità e l'integrazione di risorse e tecnologie linguistiche; ricadute applicative di risorse e tecnologie linguistiche (Information Extraction, Question Answering, sommarizzazione automatica e traduzione automatica, ecc.).

Tipo: Rivista peer-reviewed, open access

Periodicità: semestrale

ISSN: 2499-4553

*Direzione scientifica: Roberto Basili, **Simonetta Montemagni***

Editore: Accademia University Press (www.aaccademia.it) per l'Associazione Italiana di Linguistica Computazionale (www.ai-lc.it)

- **LANGUAGE RESOURCES AND EVALUATION**

La prima rivista internazionale dedicata all'acquisizione, alla creazione, all'annotazione e all'uso delle risorse linguistiche, insieme a metodi per la valutazione delle risorse, delle tecnologie e delle applicazioni. Le risorse linguistiche includono dati linguistici e descrizioni in forma leggibile dal computer utilizzate per assistere e incrementare le applicazioni di elaborazione linguistica, quali corpora e lessici dello scritto o del parlato, risorse multimodali, grammatiche, basi di dati e dizionari terminologici o a dominio specifico, ontologie, basi di dati multimediali ecc., come pure strumenti software di base per la loro acquisizione, la loro preparazione, la loro annotazione, la loro gestione, il loro adattamento e il loro utilizzo. La valutazione delle tecnologie linguistiche consiste nel valutare lo stato dell'arte di una determinata tecnologia confrontando diversi approcci a un dato problema, nel valutare la disponibilità di risorse e tecnologie per una data applicazione e nel valutare l'utilizzabilità del sistema e la soddisfazione degli utenti.

Tipo: Rivista

Periodicità: Trimestrale

Copertura: Volume 1 (1967) - Volume 49 (2015)

ISSN Stampa: 1574-020X

ISSN Online: 1574-0218

*Direzione scientifica: Nancy Ide, **Nicoletta Calzolari***

Editore: Springer, Netherlands (www.springer.com)

- **LINGUE E LINGUAGGIO**

È un luogo di discussione di respiro internazionale sulla linguistica generale e teorica, con una particolare attenzione alle aree di interazione con altre discipline, dalla psicologia alle neuroscienze, dall'informatica alle scienze cognitive. Pubblica saggi in inglese e in italiano operando nel rispetto dei più alti standard qualitativi. La rivista si occupa centralmente di teoria del linguaggio, dei vari livelli dell'analisi linguistica (fonologia, morfologia, sintassi, semantica, pragmatica) e della loro interazione, ma anche di linguistica computazionale, acquisizionale, variazionale e diacronica e di storia del pensiero linguistico. *Lingue e linguaggio* pubblica contributi di natura sia teorica che descrittiva ed è aperta ai diversi orientamenti teorici offerti dalla linguistica contemporanea.

Tipo: Rivista

Periodicità: Semestrale

Copertura: Volume I (2002) - Volume XIII (2014)

ISSN: 1720-9331

Direzione scientifica: Vito Pirrelli, Sergio Scalise

Editore: Società Editrice Il Mulino, Bologna (www.mulino.it)

Partecipazione a comitati scientifici di riviste e collane e attività redazionali:

Federico Boschetti

- Membro della redazione della rivista *Lexis-Poetica, retorica e comunicazione nel mondo antico*

Nicoletta Calzolari

- Co-editor in chief, *Language Resources and Evaluation (LRE)*, Springer
- Member of the Advisory Board, *Italian Journal of Computational Linguistics (IJCOL)*

Felice Dell'Orletta

- Member of the Editorial Board, *Italian Journal of Computational Linguistics (IJCOL)*
- Reviewer per le seguenti riviste:
 - *ACM Journal on Computing and Cultural Heritage (JOCCH)*
 - *ACM Transactions on Internet Technology (TOIT)*
 - *AI Communications* - journal on Artificial Intelligence, IOS Press
 - *Computational Linguistics*, MIT Press Journals
 - *Language Resources and Evaluation (LRE)*, Springer

Sara Goggi

- Assistant Editor, *Language Resources and Evaluation (LRE)*, Springer
- Member of the Editorial Office, *Italian Journal of Computational Linguistics (IJCOL)*

Claudia Marzi

- Segreteria di redazione, *Lingue e Linguaggio*, Il Mulino, Bologna

Simonetta Montemagni

- Member of the Editorial Board of the Book Series "Language Variation", Language Science Press
- Membro del Comitato Scientifico della Collana di Cultura Digitale – Pacini Editore, Pisa
- Associate Editor all'interno dell'Editorial Board della rivista internazionale *Frontiers in Digital Humanities* - Section Digital Literary Studies

Vito Pirrelli

- Member of the Scientific Board, *The Mental Lexicon* (John Benjamins)
- Member of the Scientific Board, *Italian Journal of Linguistics*
- Member of the Scientific Board, *Italian Journal of Computational Linguistics (IJCOL)*

Claudia Soria

- Co-chair di uno Special Issue della Rivista “Language Resources and Evaluation” dal titolo “Under-resourced Languages, Collaborative Approaches and Linked Open Data: Resources, Methods and Applications” Consulente linguistico per la collana “Città in gioco” di Franco Cosimo Panini Editore

Nel 2015 è stato siglato un contratto con la casa editrice De Gruyter per la pubblicazione di una raccolta dei più significativi contributi inter-disciplinari del progetto NetWordS - The European Network on Word Structure – (Vito Pirrelli) e sono state avviate negoziazioni per la pubblicazione di una selezione dei migliori contributi della Final Conference del progetto NetWords sulle riviste internazionali *Lingue e linguaggio* (Claudia Marzi) e *Italian Journal of Linguistics* (Vito Pirrelli).

6 Attività di terza missione

6.1 Partecipazione a Organismi tecnico-scientifici e normativi

▪ **Monica Monachini**

Riveste i seguenti incarichi in Comitati per la definizione di standard e norme tecniche:

UNI – Ente Nazionale Italiano di Unificazione

- rappresentante designata dal CNR nel Gruppo di Lavoro UNI/CT014/GL04 "Terminologia" della Commissione Tecnica UNI/CT014 "Documentazione e Informazione"
- delegato italiano UNI con diritto di voto in rappresentanza dell'Italia all'interno del Comitato ISO/TC37/SC4 "Language Resource Management"

ISO TC37/SC4 – Organizzazione Internazionale di Standardizzazione

- esperto nel gruppo di lavoro ISO/TC37/SC4/WG4 "Lexical resources" di LMF, standard per risorse lessicali
- esperto nel gruppo di lavoro ISO/TC37/SC4 "Lexical resources" di MetaData e ISOCAT

CLARIN (Common Language Resources and Technology Infrastructure)

- membro del Comitato di Standardizzazione Internazionale

Meta-Net (www.meta-net.eu)

- membro dell'Executive Board di META-NET, una rete di eccellenza composta da 60 centri di ricerca appartenenti a 34 Paesi che mira a costruire le basi tecnologiche per una società europea dell'informazione multilingue

▪ **Nicoletta Calzolari**

Chair del Comitato ISO/TC37/SC4 "Language Resource Management"

Membro dell'Advisory Group ISO/TC37/AG

6.2 Partecipazione ad Associazioni e Comitati scientifici

Ricercatori dell'ILC rivestono ruoli di rilievo in associazioni e comitati scientifici di livello internazionale:

Nicoletta Calzolari

- Presidente dell'Associazione ELRA (European Language Resources Association)
- Presidente dell'Associazione Internazionale PAROLE
- Vice Presidente dell'Associazione META-TRUST (the legal person of the network of excellence META-NET and the Multilingual Europe Technology Alliance META)
- membro del Board della UNDL Foundation (Universal Networking Digital Language Foundation)
- membro del Language Resources Board (LRB) del Tender European Language Resources Coordination (ELRC)
- membro permanente dell'ICCL (International Committee of Computational Linguistics)
- membro dell'Advisory Board di LIDER (Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe)

Simonetta Montemagni

- riferimento nazionale nell'ambito dell'azione European Language Resources Coordination (ELRC, www.lrc-coordination.eu) all'interno del programma 'Connecting Europe Facility'
- membro del gruppo di lavoro italiano nell'ambito dell'iniziativa internazionale Universal Dependencies (UD)

Claudia Soria

- membro dello Steering Committee dell'European Language Equality Network (ELEN)
- membro della Foundation for Endangered Languages-FEL
- membro del Comitato Esecutivo di MAAYA - The World Network for Linguistic Diversity
- collaboratore esperto per la revisione dell'UNESCO Language Vitality Index

Ricercatori dell'ILC rivestono ruoli di rilievo anche in associazioni e comitati scientifici di livello nazionale:

Andrea Bozzi

- Presidente del Comitato Scientifico del "Progetto Traduzione Talmud Babilonese"

Felice Dell'Orletta

- Membro del Comitato Scientifico presieduto da Savina Raynaud del **Centro interdisciplinare di ricerche per la computerizzazione dei segni dell'espressione (CIRCSE)** dell'Università Cattolica del Sacro Cuore

Ricercatori dell'ILC, inoltre, sono stati tra i soci fondatori dell'**Associazione Italiana di Linguistica Computazionale (AILC)**, rivestendo ruoli di responsabilità (Simonetta Montemagni, vice-presidente; Felice Dell'Orletta, membro del Direttivo). L'associazione svolge un ruolo strategico nella creazione di una comunità scientifica nazionale del settore della linguistica computazionale. Particolare attenzione è volta alla promozione delle attività scientifiche e formative nel settore, al consolidamento dei legami con altre iniziative che operano nel campo della linguistica computazionale (a livello nazionale, europeo e internazionale) e alla promozione di questa disciplina nell'ambito della politica nazionale, in particolare per quanto riguarda il settore dell'università e della ricerca scientifica. Tra gli strumenti di cui l'Associazione si è dotata per il perseguimento dei propri obiettivi c'è la conferenza annuale CLIC-it (Italian Conference on Computational Linguistics), la campagna di valutazione per le tecnologie del linguaggio per la lingua italiana scritta e parlata (EVALITA) e la rivista *Italian Journal of Computational Linguistics* (IJCoL).

6.3 Partecipazione a comitati di valutazione

Simonetta Montemagni

- revisore di proposte di progetto per il Research Council KU Leuven (University of Leuven, Belgium)
- revisore esterno per la tesi di dottorato di Matthew Shardlow, School of Computer Science, National Centre for Text Mining (NaCTeM), Manchester Institute of Biotechnology, University of Manchester (UK)

Vito Pirrelli

- commissione internazionale per la valutazione di dirigente di ricerca CNRS (Toulouse)

6.4 Valorizzazione dei risultati e trasferimento tecnologico

Nel 2015 l'Istituto ha consolidato contatti e collaborazioni, sia con imprese italiane e internazionali (piccole e medie imprese e grandi industrie, anche multinazionali) sia con enti pubblici locali, volti alla valorizzazione dei risultati della ricerca. In tal modo è stato possibile diffondere le conoscenze e le competenze acquisite in diversi ambiti della linguistica computazionale, con particolare attenzione a risorse e tecnologie linguistiche sviluppate dall'Istituto.

Da segnalare, in particolare:

- la partecipazione a progetti sponsorizzati da **Google** e **Linkedin** per la costruzione di una *Treebank dell'italiano con annotazione sintattica a dipendenze secondo lo schema "Universal Dependencies"*;
- l'integrazione di funzionalità avanzate basate su tecnologie linguistico-computazionali sviluppate all'ILC all'interno della piattaforma per l'editoria scolastica digitale della società **META**, attualmente in uso presso **Mondadori Education**;
- l'utilizzo di tecnologie del linguaggio sviluppate all'ILC per il monitoraggio dell'efficacia comunicativa di documenti di ambito biomedico, in collaborazione con il **Centro Gestione Rischio Clinico e Sicurezza dei Pazienti** della **Regione Toscana**;
- la sponsorizzazione da parte di due **case farmaceutiche** (Lundbeck Italia Spa e Otsuka Pharmaceutical Italy Srl) di un portale con una rassegna stampa, aggiornata quotidianamente, sul tema della salute mentale <http://dbtvm1.ilc.cnr.it/salutem>

L'ILC continua a contribuire in modo significativo alla conservazione e valorizzazione del patrimonio culturale "invisibile" italiano, in particolare, linguistico, filologico, storico e letterario. Nel corso del 2015, in particolare, i formati obsoleti di numerosi archivi testuali sono stati recuperati e convertiti in formato standard XML/TEI.

6.5 Attività di Public Engagement

Nel corso del 2015 sono state realizzate diverse attività di divulgazione scientifica con valore educativo, culturale e di sviluppo della società. Scopo principale di tali iniziative, rivolte non solo agli esperti del settore, ma anche alle scuole e al grande pubblico, sono la disseminazione dei risultati delle ricerche e un maggior coinvolgimento di tutti gli stakeholders.

6.5.1 Eventi pubblici

In occasione di **BRIGHT 2015 - Notte Europea dei Ricercatori** (25 settembre 2015, CNR - Area della Ricerca di Pisa) l'Istituto ha organizzato interventi sui seguenti temi:

- *Come scrivi? Strumenti automatici per esplorare lo stile di un testo*
- *Di cosa parla? Strumenti automatici per esplorare il contenuto di un testo*
- *E io... traduco! Il computer al servizio della traduzione di testi antichi*
- *La fabbrica delle parole antiche. Latino e greco nei tempi della rete*
- *Una parola tira l'altra: lessico e reti neurali*
- *Dati collegati: reti di parole, reti di concetti*
- *Dal verbo all'azione: come un computer può comprendere le istruzioni*
- *Nel mare non ci sono solo pesci ma anche parole e numeri*
- *Reti di Parole. Reti di Concetti*

Per i contenuti degli interventi cfr. <http://www.ilc.cnr.it/it/content/interventi-ilc-bright-2015>

6.5.2 Sito web

Nel 2015 è proseguito lo sviluppo del sito web dell'Istituto. Il sito, disponibile in italiano e in inglese, è stato costantemente aggiornato con la segnalazione delle attività realizzate e arricchito con le informazioni relative ai progetti, alle collaborazioni, alla produzione scientifica, agli eventi e alle ultime notizie. È stato inoltre avviato lo studio di un nuovo layout "responsive" in vista di una futura ristrutturazione del sito.

6.5.3 Trasmissioni radiofoniche

- **Radio Aula 40 – Punto Radio**
<http://radioaula40.cnr.it>
Puntata del 29 gennaio 2015 - *Computer che insegnano, computer che imparano*
- **Radio Aula 40 – Punto Radio**
<http://radioaula40.cnr.it>
Puntata del 5 maggio 2015 - *I disturbi e le difficoltà di apprendimento nell'età scolare*
- **Radio Linguaggio – Parlairealmicrofono.it**
<https://soundcloud.com/radiolinguaggio/il-discorso-di-insediamento-di-sergio-mattarella-con-giulia-venturi-del-cnr-di-pisa>
Intervista a Giulia Venturi: *Il discorso di insediamento di Sergio Mattarella*

6.5.4 Iniziative di interazione con scuole e Università

Nel corso del 2015 l'Istituto ha collaborato con Scuole Secondarie di secondo grado per avvicinare gli studenti ai temi delle Digital Humanities e della filologia collaborativa. Da segnalare, in particolare:

- **Liceo Classico Marco Polo** (Venezia)
Annotazione sintattica (treebanking) di testi greci e latini
Docente: Federico Boschetti
Corso di aggiornamento per insegnanti, in collaborazione con: Paola Tomè, Oxford University, Chiara Palladino, Open Philology Project (Leipzig University), Mariantonietta Rizzetto, Liceo Classico "Marco Polo" di Venezia
- **Liceo Classico "Rosario Livatino"** di San Marco dei Cavoti (BN)
Cultura digitale: analisi e traduzione dell'opera di Ippocrate "Sulle arie, sulle acque e sui luoghi"
Responsabile ILC: Emiliano Giovannetti
- **Liceo classico e linguistico statale Aristofane, Liceo ginnasio statale Virgilio, Liceo classico statale Ennio Quirino Visconti** (Roma)
Tecnologia e Sapere: una battaglia dell'intelligenza
Progetto Clavius@school (nell'ambito del progetto Clavius on the Web, in collaborazione con IIT-CNR e APUG di Roma)
Referente ILC: Emiliano Giovannetti
- **Venice International University (VIU)**
Attivazione di tirocini
Responsabile ILC: F. Boschetti

7 APPENDICE: Personale ILC

SEDE DI PISA	Profilo		N.
Contratto Lavoro Diritto Privato	Direttore di Istituto	Montemagni Simonetta	1
Personale a tempo indeterminato	Dirigente di ricerca	Pirrelli Vito	1
	I° Ricercatore	Monachini Monica	1
	Ricercatore	Bartolini Roberto	9
		Boschetti Federico	
		Dell'Orletta Felice	
		Ferro Marcello	
		Giovannetti Emiliano	
		Marchi Simone	
		Pardelli Gabriella	
		Quochi Valeria	
		Soria Claudia	
	I° Tecnologo	Enea Alessandro	1
	Tecnologo	Goggi Sara	3
		Marzi Claudia	
		Sassi Manuela	
	Collaboratore tecnico enti di ricerca	Gadducci Antonella	6
		Parrinelli Vanessa	
		Parrinelli Vincenzo	
		Picchi Paolo	
		Sassolini Eva	
		Terreni Noemi	
Collaboratore di amministrazione	Pieri Antonella	1	
Personale a tempo determinato	Ricercatore	Del Gratta Riccardo	2
		Frontini Francesca	
	Tecnologo	Baroni Paola	1
	Collaboratore tecnico enti di ricerca	Albanesi Davide	2
		Cucurullo Sebastiana	
Totale Personale a tempo indeterminato			22
Totale Personale a tempo determinato			5
Totale Contratto Lavoro Diritto Privato			1

SEDE DI GENOVA	Profilo		N.
Personale a tempo indeterminato	I° Ricercatore	Marconi Lucia	1
	Collaboratore tecnico enti di ricerca	Cutugno Paola	1
Totale Personale a tempo indeterminato			2

ASSEGNISTI DI RICERCA		N.
Sede di Pisa	Bellandi Andrea	12
	Benotto Giulia	
	Brunato Dominique Pierina	
	Cardillo Franco Alberto	
	Cimino Andrea	
	Del Grosso Angelo Mario	
	Khan Anas Fahad	
	Mancini Lorenzo	
	Nahli Ouafae	
	Piccini Silvia	
	Russo Irene	
	Venturi Giulia	
Sede di Genova	Cinini Alessandra	2
	Lucentini Roberta	
Totale Assegnisti di ricerca		14

BORSISTI	Nome	N.
Sede di Pisa	De Felice Irene	1

ASSOCIATI	Nome	N.
Sede di Pisa	Bozzi Andrea	3
	Calzolari Zamorani Nicoletta	
	Cignoni Laura (dal 1 agosto 2015)	

Redazione a cura di Michela Carlino

Novembre 2016

© 2016 Istituto di Linguistica Computazionale «A. Zampolli» | Tutti i diritti riservati



<http://www.ilc.cnr.it>