

ISTITUTO DI LINGUISTICA COMPUTAZIONALE "A. ZAMPOLLI"

2016



Rapporto Annuale



Consiglio Nazionale delle Ricerche

Sommario

1	Introduzione	5
2	L'ILC nel 2016: fatti e cifre	5
2.1	Personale.....	5
2.2	Organizzazione interna.....	7
2.2.1	Responsabili e Organi di Governo.....	7
2.2.2	Organizzazione interna	7
2.2.3	Organizzazione della ricerca	8
2.3	Finanziamenti	10
2.4	Progetti.....	11
2.4.1	Progetti europei	11
2.4.2	Progetti nazionali e regionali	12
2.5	Collaborazioni scientifiche	15
2.5.1	Accordi Bilaterali	15
2.5.2	Accordi e Convenzioni.....	15
2.5.3	Altre collaborazioni	18
2.6	Valutazione della ricerca dell'ILC.....	20
2.7	Premi	21
3	Attività di ricerca	22
3.1	Ricerca scientifica	22
3.2	Ricerca Istituzionale	24
3.3	Pubblicazioni	25
3.3.1	Contributi in rivista	25
3.3.2	Contributi in volume	26
3.3.3	Contributi in atti di convegno	26
3.3.4	Curatele.....	28
3.4	Comunicazioni a convegni senza pubblicazione degli atti.....	29
3.5	Internazionalizzazione.....	29
4	Attività di alta formazione	30
4.1	Corsi presso Università.....	30
4.2	Summer Schools.....	30
4.3	Supervisione di tesi di laurea e di dottorato	31
4.4	Tirocini.....	33
4.5	Formazione interna	34
5	Disseminazione scientifica	34
5.1	Workshop, conferenze, seminari	34
5.1.1	Workshop e conferenze organizzati e co-organizzati dall'ILC.....	34
5.1.2	Partecipazione a comitati scientifici di conferenze.....	36

	5.1.3	Comunicazioni e seminari su invito	36
	5.1.4	Seminari interni.....	38
6		Attività editoriali.....	39
7		Attività di terza missione	40
	7.1	Partecipazione a organismi tecnico-scientifici e normativi	40
	7.2	Partecipazione ad Associazioni e Comitati scientifici	41
	7.3	Partecipazione a comitati di valutazione	41
	7.4	Valorizzazione dei risultati e trasferimento tecnologico	42
	7.5	Attività di Public Engagement	42
	7.5.1	Rapporti con le Istituzioni	42
	7.5.2	Eventi pubblici.....	43
	7.5.3	Siti web e social media	44
	7.5.4	Trasmissioni radiofoniche	44
	7.5.5	Iniziative di interazione con scuole	44
	7.5.6	Biblioteca	45
8		English Summary	46
9		Appendice	49

Prefazione

A nome dell'Istituto è per me un onore introdurre il secondo Rapporto Annuale dell'Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) del Consiglio Nazionale delle Ricerche. Il rapporto 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 2016 su molteplici versanti, che spaziano dalla ricerca scientifica e l'alta formazione alla disseminazione e divulgazione scientifiche e la valorizzazione sul versante economico, sociale e culturale della ricerca dell'ILC. Nella panoramica offerta, le singole voci e i contributi individuali si ricompongono all'interno di un quadro complesso e dinamico dal quale emergono linee di ricerca strategiche e innegabili punti di forza.

Nel 2016, le attività di ricerca dell'Istituto 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.

La tabella che segue riassume in cifre le attività dell'anno:

I risultati 2016 della ricerca dell'ILC	
Articoli in rivista	19
Capitoli di libri / contributi in volumi	7
Contributi in atti di convegno	45
Curatele	9
Direzione Scientifica di riviste	3
Supervisione di tesi di LM e di dottorato	9
Insegnamento corsi universitari	2
Insegnamento di Moduli in Scuole estive	8
Progetti di ricerca	20
Infrastrutture di ricerca	2
Riconoscimenti e premi	8

Tra i risultati più significativi del 2016 si segnalano in particolare:

- il rafforzamento e l'estensione dei rapporti con la comunità scientifica nazionale, anche grazie al ruolo dell'ILC come rappresentante del consorzio italiano di CLARIN-ERIC, l'Infrastruttura di ricerca per le Scienze Umane e Sociali (SSH). Tramite il Coordinatore Nazionale per l'Italia è stata condotta una capillare opera di disseminazione presso la comunità dei potenziali utenti nell'ambito sia della Linguistica Computazionale sia delle SSH. Al rafforzamento e all'estensione dei rapporti con la comunità scientifica nazionale hanno contribuito anche la partecipazione di ricercatori dell'ILC alle attività dell'Associazione Italiana di Linguistica Computazionale (AIRC); la presentazione di domande per Bandi nazionali e regionali; la stipula di dichiarazioni di intenti, protocolli di intesa, accordi quadro e convenzioni di varia natura con università, enti di ricerca e istituzioni culturali, così come la collaborazione con altri istituti CNR per lo sviluppo di progetti interdisciplinari;*
- il consolidamento dell'ampia rete di relazioni internazionali sia nell'ambito della Linguistica Computazionale sia delle SSH, grazie alla posizione dell'ILC, oltre che come coordinatore nazionale di CLARIN-IT, anche come referente tecnologico nazionale nell'ambito dell'azione ELRC - European Language Resource Coordination. Al rafforzamento della rete di relazioni hanno contribuito anche: le visite di studiosi stranieri di chiara fama presso la sede dell'ILC; la stipula di accordi bilaterali di collaborazione scientifica con prestigiose istituzioni straniere; l'avvio o la prosecuzione di collaborazioni scientifiche internazionali extra-progettuali con università, enti di ricerca e istituzioni culturali internazionali (CNRS, Max Planck Institute, Université Paris-Sorbonne, European Science Foundation, Austrian Academy of Sciences, ecc.), così come con l'UNESCO e il Consiglio d'Europa;*

- *il proseguimento delle attività editoriali, di disseminazione e divulgazione scientifica condotte mediante: la direzione di tre riviste scientifiche del settore; l'organizzazione di importanti conferenze/workshop nazionali e internazionali; la partecipazione a eventi divulgativi rivolti al grande pubblico, tra cui è da segnalare la presentazione del primo volume della traduzione del Talmud Babilonese in italiano tenutasi a Roma presso l'Accademia dei Lincei il 5 aprile 2016; la collaborazione con scuole e università per avvicinare gli studenti ai temi della Linguistica Computazionale e delle Digital Humanities;*
- *nel rapporto con le Istituzioni governative, ricercatori dell'ILC sono stati interpellati in qualità di esperti a livello nazionale per audizioni presso la Camera dei Deputati e il Ministero del Lavoro e delle Politiche Sociali.*

In merito alla valutazione della ricerca dell'ILC, vale la pena menzionare l'esito positivo della rivalutazione condotta dal Prof. De Mauro incaricato di fornire un parere esterno a seguito delle anomalie di metodo e di merito segnalate dall'istituto per la valutazione CNR 2011-2014. Tale valutazione positiva trova conferma nei risultati relativi alla VQR 2011-2014: il 66% dei prodotti della ricerca dell'ILC hanno ricevuto un giudizio Eccellente/Elevato, con un voto medio superiore sia alla precedente VQR 2004-2010 per il DSU, sia – più significativamente - rispetto a quanto rilevato (internamente al CNR e globalmente) per l'Area disciplinare di afferenza.

Tra i fattori di rischio emersi con forza nel corso dell'anno, si segnala che la nomina dell'Istituto come National Representative e istituto esecutore della infrastruttura italiana CLARIN-IT, che ha fatto seguito all'adesione dell'Italia all'infrastruttura di ricerca europea CLARIN ERIC a fine 2015, rappresenta un elemento di complessità non indifferente, in considerazione del fatto che non vi è stato fino ad oggi nessun finanziamento dedicato, per cui l'Istituto si trova a fare fronte – da solo – alla costituzione del nodo nazionale, un'impresa estremamente onerosa in termini sia di attrezzature richieste sia di risorse umane.

Pisa, giugno 2017



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.

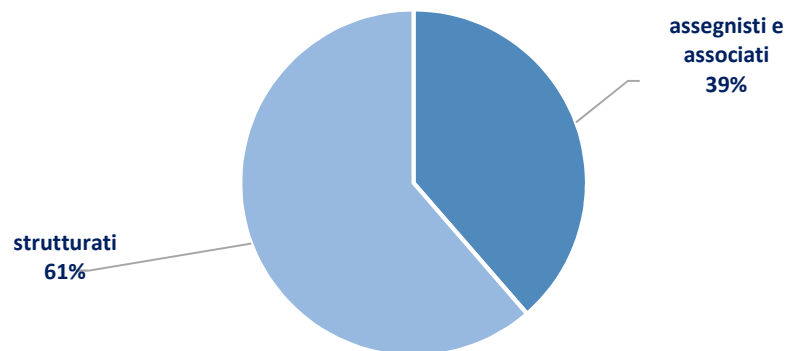
La ricerca all'ILC combina ricerca di base, con un investimento su temi di frontiera, e ricerca applicata, all'interno di un circolo virtuoso con ricadute significative sulla società e, in particolare, sul contesto socio-economico e culturale. Le attività sono condotte all'interno di una rete consolidata 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.

2 L'ILC nel 2016: fatti e cifre

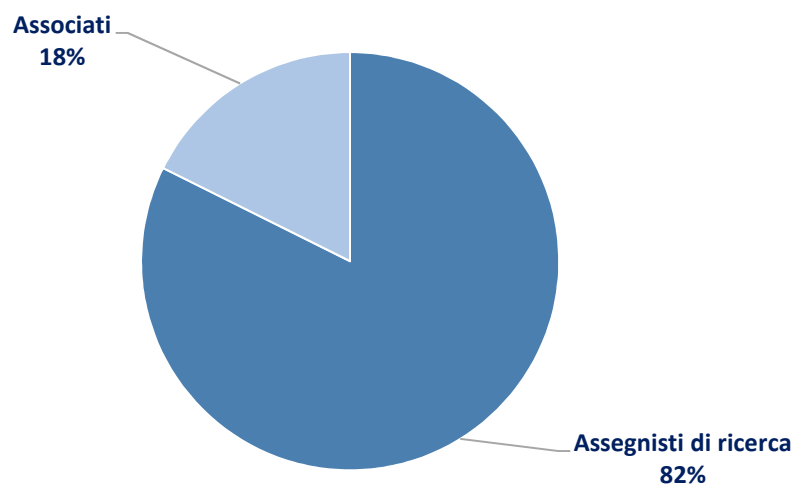
2.1 Personale

Il personale dell'ILC è composto da ricercatori, tecnologi e personale tecnico-amministrativo sia di supporto alla ricerca sia di area gestionale-amministrativa. Nel corso del 2016, il numero delle unità di personale ha subito delle variazioni: due unità sono andate in quiescenza (un collaboratore tecnico degli enti di ricerca e un tecnologo); un'unità con il profilo di collaboratore tecnico degli enti di ricerca (CTER) è risultata vincitrice del concorso riservato al personale con contratto a tempo determinato per posti a tempo indeterminato con lo stesso profilo (CTER); un'altra unità con profilo di ricercatore a tempo determinato ha rassegnato le dimissioni in seguito alla vincita di un concorso presso un'università straniera.

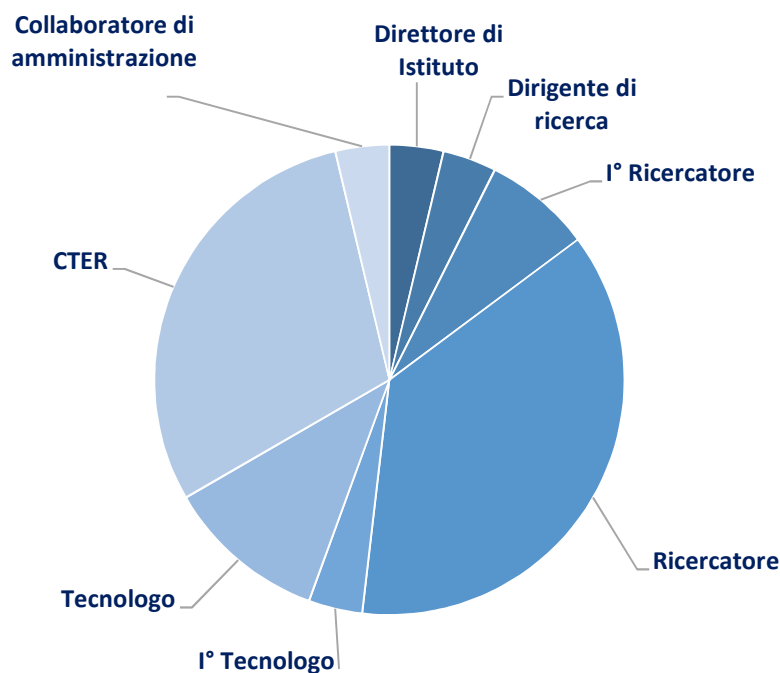
Oltre al Direttore dell'Istituto, lo staff è composto da 27 unità di personale strutturato (a tempo indeterminato e determinato) distribuite tra la sede di Pisa (25) e la sezione staccata di Genova (2), a cui si aggiungono 17 unità (15 a Pisa e 2 a Genova) tra assegnisti e personale associato: i dati fanno riferimento alla data del 31 dicembre 2016. Per maggiori dettagli v. appendice: *Personale ILC*.



RAPPORTO PERSONALE STRUTTURATO VS ALTRO PERSONALE



RIPARTIZIONE ALTRO PERSONALE



SUDDIVISIONE PERSONALE STRUTTURATO PER PROFILI E LIVELLI

2.2 Organizzazione interna

2.2.1 Responsabili e Organi di Governo

Direttore	Simonetta Montemagni
Consiglio di Istituto	Felice Dell'Orletta, Antonella Gadducci, Claudia Marzi, Monica Monachini, Simonetta Montemagni, Gabriella Pardelli, Vito Pirrelli
Responsabile amministrativo	Antonella Gadducci
Responsabile dei sistemi informativi	Alessandro Enea
Responsabile della sicurezza	Roberto Bartolini
Responsabile della biblioteca	Gabriella Pardelli
Referente per la formazione	Claudia Marzi

2.2.2 Organizzazione interna

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

Rappresentanti ILC all'interno dei Comitati dell'Area della Ricerca – CNR di Pisa:

Commissione Comunicazione	Paola Baroni, Claudia Soria
Commissione Relazioni Internazionali	Valeria Quochi
BRIGHT - La Notte dei Ricercatori in Toscana	Marcello Ferro
AREAPERTA - Comitato di Redazione	Felice Dell'Orletta <i>(fino a ottobre 2016)</i> , Emiliano Giovannetti <i>(da novembre 2016)</i>

2.2.3 Organizzazione della ricerca

Nel 2016 è proseguito il processo di razionalizzazione delle attività di ricerca e sviluppo, già avviato negli anni precedenti. In particolare, questo processo ha riguardato la riorganizzazione delle attività secondo quanto previsto dai nuovi Regolamenti CNR entrati in vigore l'anno precedente. La Direzione insieme agli organi di governo dell'istituto ha promosso una riflessione interna che ha coinvolto tutti i ricercatori e tecnologi, e in modo particolare i responsabili di Commessa, di Modulo così come di progetto e collaborazione scientifica, sui confini tematici e il contenuto delle varie linee di attività, nell'ottica di minimizzare le sovrapposizioni e di massimizzare le sinergie tra aree di ricerca. Il risultato di tale riflessione ha portato all'adozione di un modello organizzativo "a matrice", in cui aree di competenza e progetti rappresentano rispettivamente gli assi verticali e orizzontali della matrice: le competenze maturate nell'ambito delle singole aree si combinano in modo vario, innovativo e sinergico all'interno dei progetti di ricerca. All'interno di questo assetto organizzativo della ricerca dell'ILC, i Laboratori e i gruppi di ricerca rivestono un ruolo centrale e strategico come luoghi di sviluppo di competenze e come spazio per la ricerca "curiosity-driven", mentre l'asse orizzontale della matrice coincide con la riorganizzazione delle attività in progetti come richiesto dai regolamenti CNR, confluita nel Piano di Gestione Preliminare 2017-2019 (PdGP 2017-2019).

A fine 2016, in occasione della redazione del PdGP 2017-2019, è stato introdotto dal CNR un nuovo livello con la funzione di descrittore scientifico, denominato "attività di ricerca", a cui è stata assegnata la specifica funzione di aggregatore tematico dei progetti degli istituti. Tale riorganizzazione ha costituito un'importante occasione per consolidare la riorganizzazione in corso. La definizione delle attività di ricerca rispetto alle quali aggregare tematicamente i progetti dell'ILC è stata oggetto di attenta riflessione. Con una buona approssimazione, essa ha coinciso con le aree di attività dei gruppi di ricerca e dei laboratori con l'unica eccezione dell'area "Digital Humanities", frammentata tra diversi gruppi e per la quale è stata pianificata una razionalizzazione interna. Segue la lista delle quattro aree di attività di ricerca identificate:

- DH - Digital Humanities;
- LRI - Risorse, standard e infrastrutture di ricerca;
- MIND - Modelli (bio-)computazionali dell'uso linguistico;
- TAL - Trattamento automatico del linguaggio naturale ed estrazione di conoscenza.

AREE DI ATTIVITÀ e DESCRIZIONE	
DH Digital Humanities	La linea di attività DH intende coniugare le acquisizioni e conoscenze delle scienze informatiche con gli approcci metodologici e i modelli teorici dell'analisi testuale e della filologia del testo, contribuendo così alla trasformazione delle modalità di conservazione, fruizione, studio e pubblicazione dei documenti letterari, archivistici e bibliotecari, e offrendo nuove prospettive di indagine e condivisione. Le soluzioni tecnologiche messe a punto in questo ambito si integrano in un sistema "multi-modulare" a componenti indipendenti ma interconnessi, da cui le diverse metodologie di accesso, gestione, studio e revisione del testo possono trarre vantaggio e opportunità di interazione/integrazione.
LRI Risorse, standard e infrastrutture di ricerca	La linea di attività LRI si propone di facilitare la ricerca nel settore dell'ingegneria delle lingue e di ottimizzare il ciclo di produzione delle risorse linguistiche. Questo chiama in causa l'adozione di standard, lo scambio di buone pratiche per l'interoperabilità, il riciclo e il riutilizzo dei risultati disponibili in termini di dati e strumenti. Le attività principali di quest'area riguardano la definizione di modelli per la creazione, rappresentazione, estensione e mantenimento di lessici computazionali, repertori terminologici e ontologici, corpora e tecnologie linguistiche, ma anche lo sviluppo di soluzioni tecnologiche per la creazione di un'infrastruttura di ricerca distribuita e cooperativa, volta a stabilire nuove funzionalità di accesso, interoperabilità e condivisione di risorse e strumenti linguistici.
MIND Modelli (bio-)computazionali dell'uso linguistico	La linea di attività MIND è dedicata allo studio dei fattori che governano i processi di comprensione, produzione, apprendimento e variazione di una lingua e le interazioni dinamiche tra di essi. In particolare, promuove lo sviluppo di modelli teorici dell'uso linguistico e la loro verifica empirica attraverso: l'uso di metodi probabilistici per lo studio di corpora, lessici e basi di dati; simulazioni computazionali; lo studio di evidenza linguistica di natura sperimentale, clinica e acquisizionale. Gli obiettivi di questa linea di attività sono perseguiti coniugando metodologie di rappresentazione formale e modellazione simbolica con i metodi, i dati e gli strumenti di indagine di settori disciplinari più orientati all'analisi dell'uso linguistico in contesti finalizzati e controllati, quali la psico- e neuro-linguistica, la sociolinguistica e la glottodidattica.
TAL Trattamento automatico del linguaggio naturale ed estrazione di conoscenza	La linea di attività TAL si occupa di sviluppare metodi e tecniche per consentire l'accesso automatico al contenuto di un testo, rendendo espliciti quei nuclei di conoscenza linguistica che rispondono a una vasta gamma di bisogni informativi dei parlanti, dall'accesso su base semantica al contenuto testuale, alla valutazione della struttura del testo come indicatore della sua accessibilità ed efficacia comunicativa. Le soluzioni tecnologiche sviluppate in questo ambito rispondono alle necessità di ricerca e gestione "intelligente" dell'informazione contenuta all'interno di grandi basi documentali in continua evoluzione, e sono propedeutiche allo sviluppo di un ampio ventaglio di applicativi commerciali. TAL si focalizza sullo studio di modelli probabilistici del linguaggio e sullo sviluppo di algoritmi di apprendimento automatico per l'annotazione del testo e l'estrazione di conoscenza.

A queste aree si è affiancata una quinta linea ART – "Attività di Ricerca Trasversali" riguardante il perfezionamento linguistico, statistico e informatico, nonché nel campo della gestione della ricerca e della conoscenza dei sistemi di ricerca europei e internazionali, della valorizzazione dei risultati della ricerca e protezione della proprietà intellettuale.

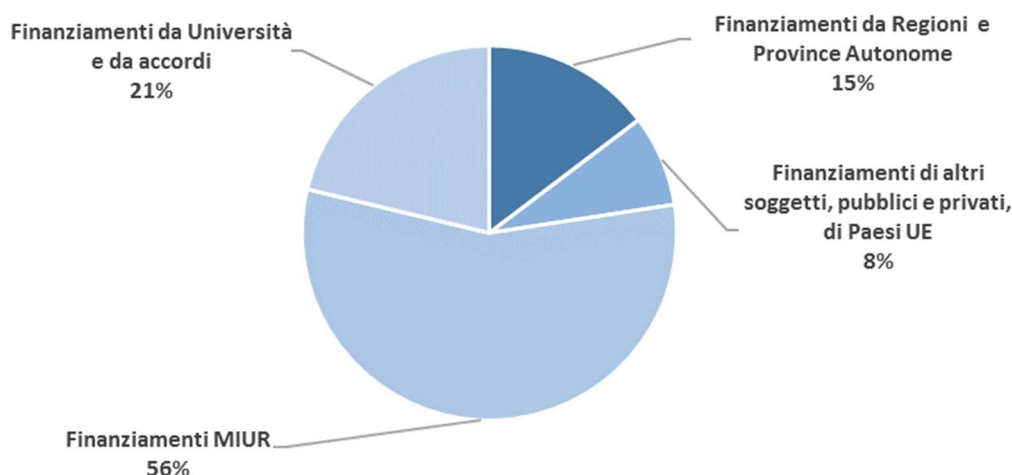
LABORATORI e GRUPPI di RICERCA	
ComPhys Lab	<i>Obiettivo:</i> Progettazione e sviluppo di modelli (bio-)computazionali del comportamento linguistico con l'obiettivo di comprendere e spiegare le relazioni tra competenza grammaticale, uso linguistico e correlati neuro- e psico-linguistici della comunicazione verbale e dei suoi disturbi. <i>Responsabile:</i> Vito Pirrelli <i>Sito:</i> www.comphyslab.it
CoPhiLab	<i>Obiettivo:</i> 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. <i>Responsabile:</i> Federico Boschetti <i>Sito:</i> http://cophilab.ilc.cnr.it:8080/CoPhiLabPortal

ItaliaNLP Lab	<i>Obiettivo:</i> 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. <i>Responsabile:</i> Felice Dell'Orletta <i>Sito:</i> www.italianlp.it
LaRI Group	<i>Obiettivo:</i> Il gruppo "Risorse e Infrastrutture Linguistiche" si propone di facilitare la ricerca nel settore dell'ingegneria delle lingue e di ottimizzare il ciclo di produzione delle risorse linguistiche. <i>Responsabile:</i> Monica Monachini <i>Sito:</i> http://lari.ilc.cnr.it
Literary Group Computing	<i>Obiettivo:</i> Il gruppo "Literary Computing" si occupa di modelli testuali, ontologie dei testi, traduzione assistita da computer, progettazione e sviluppo di applicazioni per il "Literary Computing". <i>Responsabile:</i> Emiliano Giovannetti <i>Sito:</i> http://licolab.ilc.cnr.it

2.3 Finanziamenti

Nel corso del 2016 si è registrato un significativo incremento dei progetti finanziati attivi dell'Istituto: in particolare, i progetti e sottoprogetti a fine anno risultavano essere 25, articolati in 14 progetti ILC e 11 progetti di altri istituti a cui ILC ha partecipato con un sottoprogetto. Tra questi, si annoverano un progetto terminato nell'anno e tre progetti strategici interni. A inizio 2016, ovvero quando è entrata in vigore la nuova organizzazione scientifica e gestionale dell'Ente, risultavano dal PdGP 2016-2018 9 progetti attivi, ripartiti in 6 progetti ILC e 3 sottoprogetti con coordinamento esterno, che nel corso dell'anno sono stati incrementati. Tutti i progetti sono stati acquisiti sia nell'ambito di bandi competitivi (nazionali e internazionali) sia tramite contratti e consulenze con committenza esterna (pubblica e industriale) e interna al CNR. Nel 2016 erano in vigore 29 accordi di collaborazione scientifica con istituzioni accademiche e culturali nazionali e internazionali così come con imprese (cfr. sezione "Collaborazioni scientifiche"): di questi, circa la metà prevedono un finanziamento. L'incremento del numero di progetti attivi purtroppo non ha trovato riscontro a livello dei finanziamenti incassati nel 2016: ciò trova giustificazione nel fatto che i progetti avviati nel corso del 2016 alla fine dell'anno non avevano ancora ricevuto alcun finanziamento.

RIPARTIZIONE DELLE ENTRATE - ANNO 2016



2.4 Progetti

2.4.1 Progetti europei

Progetti europei coordinati dall'ILC

DLDP - The Digital Language Diversity Project

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

Responsabile Scientifico del Progetto e Responsabile Scientifico Unità di Ricerca ILC: 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. Tra le attività condotte nel 2106 vi sono: elaborazione di una scala per la valutazione della vitalità linguistica digitale; progettazione e realizzazione di questionari per la raccolta di informazioni relative all'uso e usabilità delle lingue regionali e minoritarie su Internet e per mezzo di strumenti digitali; analisi dei dati raccolti attraverso questionari per ottenere informazioni relative all'uso e all'usabilità delle lingue regionali e minoritarie su Internet e per mezzo di strumenti digitali (1.293 questionari per 30 domande); progettazione della struttura e del contenuto di un corso on-line rivolto a parlanti di lingue regionali e minoritarie per promuovere la presenza e l'uso di queste lingue in ambito digitale; preparazione dei report di progetto; attività di networking e di disseminazione.

<http://www.dldp.eu>

Progetti europei ai quali l'ILC partecipa come partner

CADDY - Cognitive Autonomous Diving buddy

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

Responsabile Scientifico Unità di Ricerca 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. Le principali attività di ILC sono state rivolte alla creazione di un protocollo di comunicazione fra il robot e il subacqueo, alla creazione di un linguaggio per la loro interazione (composto da gesti, la loro trascrizione in forma scritta, sintassi e semantica) e allo sviluppo di un analizzatore sintattico in grado di associare sequenze di gesti sintatticamente corrette a un comando o lista di comandi da far eseguire al robot. L'attività è focalizzata alla realizzazione di un interprete di un linguaggio gestuale per l'interazione uomo-robot, in grado di associare sequenze di gesti base ad azioni singole e, tramite queste, di richiedere al robot lo svolgimento di missioni più complesse. Il lavoro ha visto lo sviluppo di un linguaggio dei gesti strutturato, indipendente dalla lingua, affine al linguaggio naturale e il più possibile conforme ai gesti convenzionali utilizzati dai subacquei.

<http://www.caddy-fp7.eu>

PARTHENOS - Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

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.

Nel corso del 2016 sono state redatte le linee guida per la creazione, il mantenimento e la preservazione a lungo termine dei contenuti digitali in fatto di Risorse Linguistiche e sono state definite le buone pratiche per la catalogazione e la ricerca in repository comuni, con politiche di accesso aperto (open data e open access). Si è

collaborato, inoltre, alla definizione di un framework semantico comune per le scienze umane, verificandone la compatibilità con le pratiche del proprio settore.

<http://www.parthenos-project.eu>

Iniziative europee in cui l'ILC è sottocontraente

ELRC - European Language Resource Coordination

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

Responsabile Scientifico Unità di Ricerca ILC: 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 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.

Nell'ambito del progetto, Simonetta Montemagni è il referente tecnologico nazionale per l'Italia. I referenti tecnologici nazionali dell'azione ELRC costituiscono il Language Resources Board, l'organo di governo di ELRC.

<http://www.lr-coordination.eu>

Altre iniziative internazionali in cui l'ILC è partner

Progetto DiTMAO - Dictionary of Old Occitan medico-botanical terminology

Progetto internazionale finanziato dalla DFG (Deutsche Forschungsgemeinschaft)

Responsabile Scientifico Unità di Ricerca ILC: Emiliano Giovannetti

Il progetto DiTMAO mira alla creazione di un sistema di lessicografia digitale contenente termini medico-botanici in occitano antico derivanti da testi medievali in latino e in caratteri ebraici. In particolare, l'attività dell'ILC è dedicata alla definizione di modelli lessico-ontologici e al supporto alla progettazione dell'editor LexΩ.

<https://www.uni-goettingen.de/en/ditmao/487498.html>

2.4.2 Progetti nazionali e regionali

Progetti finanziati a livello nazionale

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).

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

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.

<http://modelact.lablita.it>

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).

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

Progetto quinquennale finanziato dal MIUR.

Responsabile Scientifico Unità di Ricerca ILC: 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.

<http://www.talmud.it>

TOTUS MUNDUS - The World is our Home

Progetto nazionale finanziato dall'Università La Sapienza di Roma

Responsabile Scientifico Unità di Ricerca ILC: Emiliano Giovannetti

Il progetto è finalizzato alla creazione di una piattaforma digitale mediante la quale sarà possibile accedere all'edizione annotata da Pasquale D'Elia dell'Atlante mondiale prodotto nel 1602 a Pechino da Matteo Ricci e Li Zhizao. A tale scopo è stata attivata una sinergia tra l'Istituto di Informatica e Telematica del Consiglio Nazionale delle Ricerche - IIT-CNR (capofila del progetto) e l'Università La Sapienza. L'ILC, nell'ambito di un sottoprogetto, collabora in particolare per la definizione di modelli lessico-ontologici e per il supporto alla creazione di una risorsa lessico-ontologica bilingue cinese-italiano.

<http://www.totusmundus.it>

VOCI DELLA GRANDE GUERRA

Progetto di 18 mesi (2016-2018) finanziato dalla Presidenza Consiglio dei Ministri nell'ambito dell'avviso pubblico per la Commemorazione del Centenario della Grande Guerra.

Responsabile Scientifico Unità di Ricerca ILC: Simonetta Montemagni

Il progetto mira a preservare e diffondere le memorie della Prima Guerra Mondiale attraverso la costruzione e la pubblicazione online di un corpus digitale di testi (lettere, bollettini di guerra, giornali, diari, ecc.) rappresentativi delle diverse modalità di sentire e raccontare l'Italia in guerra da parte dei suoi protagonisti. Questo patrimonio culturale collettivo è valorizzato grazie all'utilizzo di tecniche avanzate di linguistica computazionale, web semantico e visualizzazione dell'informazione. L'archivio digitale, che comprende documenti storici appartenenti a un ampio spettro di registri e varietà linguistiche, permette di ricostruire la polifonia delle lingue dell'Italia in guerra: la voce

ufficiale della propaganda e la voce dei soldati, la voce dei giornali e la voce delle lettere, la voce delle élite degli intellettuali e la voce popolare, la voce del consenso e la voce del dissenso. Un'interfaccia online dotata di funzionalità di ricerca innovative consente di consultare i testi tramite chiavi di ricerca semantiche e, ove possibile, di visualizzare le immagini delle edizioni cartacee originarie. L'ILC, in particolare, si occupa dell'acquisizione digitale del testo mediante tecniche innovative di OCR, della specializzazione di componenti per l'annotazione linguistica del testo e per l'estrazione di informazione nella forma di eventi e georeferenziazione dei luoghi della guerra.
<http://www.vocidellagrandeguerra.it>

Progetti regionali

SMART NEWS - Social sensing for breaking news

Progetto biennale (2016-2018) finanziato dalla Regione Toscana nell'ambito del Bando FAR FAS 2014 – LINEA A.

Responsabile Scientifico Unità di Ricerca ILC: Felice Dell'Orletta

Il progetto mira a realizzare un tool per la gestione delle breaking news che aiuti i giornalisti nelle diverse fasi di questo processo: dall'individuazione della breaking news, alla raccolta di informazioni, alla scrittura. Il tool "ascolterà" i social media e sarà in grado di individuare automaticamente le breaking news. A tale scopo saranno messe a punto avanzate metodologie per il trattamento automatico della lingua in grado di: individuare ed estrarre dai documenti di social media entità rilevanti per la descrizione delle notizie; classificare i documenti in base alle informazioni estratte; identificare i testimoni di eventi; individuare il sentimento prodotto dai documenti; riepilogare automaticamente diversi documenti al fine di scrivere articoli di giornale.

<http://www.smart-news.it>

Progetti CNR

CLAVIUS - Clavius on the Web

Progetto nazionale.

Responsabile Scientifico Unità di Ricerca ILC: 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. Nel corso del 2016, in particolare, è stato finalizzato lo sviluppo della piattaforma per lo studio della corrispondenza di Cristoforo Clavio con astronomi come Galileo Galilei, Tycho Brahe, ecc. La piattaforma permette di trascrivere le lettere, di annotare semanticamente porzioni del testo collegando termini con risorse contestuali come il lessico sviluppato in seno al progetto e WikiData (www.wikidata.org). Nel corso del 2016 sono state svolte anche attività con studenti di alcuni licei romani che, durante le fasi di sviluppo della piattaforma, hanno dato pareri e suggerimenti sul software in divenire e testato in prima persona l'approccio di annotazione proposto, basato su un linguaggio con sintassi tipo Markdown.

<http://claviusontheweb.it>

NINFA - iNtelligent Integrated Network For Aged people

Progetto di Interesse strategico CNR "Invecchiamento": Innovazioni tecnologiche e molecolari per un miglioramento della salute dell'anziano.

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.

Progetto ItaliaNLP – WAFI

Responsabile Scientifico Unità di Ricerca ILC: Felice Dell’Orletta

Progetto di collaborazione con il laboratorio WAFI dell’Istituto di Informatica e Telematica del CNR di Pisa (IIT-CNR) che mira a potenziare il sistema di analisi dei testi derivanti da social media nell’ambito della Cyber Intelligence. Nello specifico, l’ILC svolge le seguenti attività: consulenza su approcci e metodologie di analisi del testo; integrazione dello strumento T2K nella piattaforma di Cyber-Intelligence WAFI; sviluppo di prototipi per finalità specifiche, ad esempio per la rilevazione di hate-speech nei contenuti postati dagli utenti; realizzazione congiunta di pubblicazioni scientifiche.

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.

2.5 Collaborazioni scientifiche

2.5.1 Accordi Bilaterali

▪ Polish Academy of Sciences - POLONIA

La protezione del patrimonio linguistico. Analisi comparativa delle attitudini in Italia e in Polonia verso la diversità linguistica (triennio 2014-2016)

Confronto dello stato attuale di vitalità etnolinguistica della lingua casciuba e di quella piemontese, attraverso la progettazione di un questionario di vitalità etnolinguistica, la sua diffusione in Casciubia e Piemonte, e l’analisi ed il confronto delle risposte ottenute (120 questionari)

Responsabili: Claudia Soria (ILC), Nicole Dolowy-Rybinska

▪ Sidi Mohamed Ben Abdellah University - MAROCCO

L’Oceano Universale: un sistema semi-automatico per la classificazione morfologica, sintattica e semantica dei verbi trilitteri arabi (accordo CNR/CNRST anni 2016-2017)

Acquisizione di informazione lessicale araba, morfologia, trattamento automatico della lingua araba

Responsabili: Vito Pirrelli (ILC), Mohammed El Mohajir

2.5.2 Accordi e Convenzioni

▪ Accademia della Crusca

Recupero, conservazione e analisi di archivi testuali

Responsabile ILC: Simonetta Montemagni

▪ Associazione NeuroCare Onlus

Collaborazione per lo sviluppo di sistemi per il monitoraggio in ambito neuro-linguistico e neuro-cognitivo

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
- **Comunità Sant'Egidio - Ufficio Scolastico Regionale per la Liguria - Liceo Cassini di Genova - Liceo Andrea D'Oria di Genova - Liceo Scientifico Enrico Fermi di Genova - Istituto Professionale "Nino Bergese" di Genova**
Progetto Storie di una diversa giovinezza
Responsabili ILC: Paola Cutugno, Lucia Marconi
- **Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)**
Raccolta di risorse linguistiche italiane che sono rilevanti per lo European Language Resource Coordination (ELRC)
Responsabile ILC: Simonetta Montemagni
- **European Language Resources Association - ELRA**
Gestione, personalizzazione, aggiornamento e documentazione di risorse linguistiche e attività di divulgazione scientifica
Responsabili ILC: Nicoletta Calzolari, Sara Goggi, Monica Monachini, Claudia Soria
- **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**, Dip. di Studi Umanistici - Facoltà di Lettere e Filosofia - **Università del Piemonte Orientale "Amedeo Avogadro"**, 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
- **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
- **Istituto di Informatica e Telematica - Consiglio Nazionale delle Ricerche (IIT-CNR)**
Attività di Social Media Intelligence, Cyber-Intelligence
Responsabile ILC: Felice Dell'Orletta
- **Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" - Consiglio Nazionale delle Ricerche (ISTI-CNR)**
Meta Opac Pisano (MOP): aggiornamento, mantenimento e miglioramento dell'accesso ai cataloghi delle biblioteche partecipanti
Responsabile ILC: Monica Monachini
- **Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa (INDIRE, Firenze) - Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG-CNR)**
Integrazione delle competenze scientifiche e delle risorse intellettuali e tecnico-strumentali esistenti presso INDIRE, ILC e ITTIG sul tema dello sviluppo e implementazione di tecnologie della conoscenza per l'innovazione nel settore educativo e formativo.
Responsabili ILC: Simonetta Montemagni
- **Istituto Nazionale di Documentazione, Innovazione e Ricerca Educativa (INDIRE, Firenze)**
Studio e sviluppo prototipale di applicazioni di gestione della conoscenza (Knowledge Management-KM) nel settore documentale educativo basate su tecnologie di trattamento automatico del linguaggio (Natural Language Processing-NLP).
Responsabile ILC: Felice Dell'Orletta

- **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
- **Museo Archeologico di Zagabria e Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" - Consiglio Nazionale delle Ricerche (ISTI-CNR)**
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)**
Conservazione, analisi e valorizzazione del patrimonio testuale storico scientifico di IMSS
Responsabili ILC: Simonetta Montemagni, Eva Sassolini
- **Sociedad Cubana de Derecho e Informática (SCDI)**
Studio e realizzazione di moduli software di analisi testuale per gestire e interrogare corpora testuali di dominio
Responsabili ILC: Eva Sassolini
- **Universidad Nacional de Educación a Distancia (UNED)**
Scambio di competenze scientifiche nell'ambito della gestione di materiale testuale di valore storico-culturale. La collaborazione è finalizzata alla conversione in formato XML-TEI dei testi e alla loro interrogazione con un sistema di analisi testuale per testi lemmatizzati.
Responsabile ILC: Eva Sassolini
- **Università Cattolica del Sacro Cuore - Centro Interdisciplinare di Ricerche per la Computerizzazione dei Segni dell'Espressione (UCSC-CIRCSE)**
Attività di ricerca nel settore del trattamento automatico delle lingue classiche
Responsabile ILC: Federico Boschetti
- **Università degli Studi di Firenze, Dipartimento di Lettere e Filosofia (DILEF)**
Azione congiunta basata sull'integrazione delle competenze linguistiche e lessicografiche e linguistico-computazionali
Responsabile ILC: Eva Sassolini
- **Università degli Studi di Genova - Dipartimento di Scienze per l'Architettura DSA**
Studio e realizzazione di moduli software per accedere, gestire ed estrarre informazioni dai dati relativi alla ricerca: Censimento e schedatura di complessi di architettura moderna e contemporanea in Liguria
Responsabili ILC: Paola Cutugno, Lucia Marconi
- **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) e definizione dello standard di annotazione e strumenti correlati per quanto riguarda la lingua italiana, nell'ambito della collaborazione tra il Dipartimento di Informatica dell'Università degli Studi di Pisa e LINKEDIN.
Responsabile ILC: Simonetta Montemagni
- **Université Paris-Sorbonne - Observatoire de la vie littéraire (OBVIL)**
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)
Sviluppo di una collaborazione nei settori dell'istruzione e della ricerca in Scienze Umane e Sociali, in particolare per quanto riguarda il trattamento automatico della lingua e l'elaborazione digitale dei manoscritti, della collazione e dell'edizione critica.
Responsabile ILC: Vito Pirrelli
- **University of Göttingen**
DiTMAO "An XML-based Information System for Old Occitan Medical Terminology": creazione di un sistema di lessicografica digitale contenente termini medico-botanici in occitano antico derivanti da testi medievali in latino e in caratteri ebraici.
Responsabile ILC: Emiliano Giovannetti

2.5.3 Altre collaborazioni

Oltre alle collaborazioni che sono state oggetto di un accordo formale (cfr. *supra*), nel corso del 2016 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.

Per quanto riguarda, nello specifico, la comunità scientifica italiana, nel 2016 l'attività di consolidamento ed estensione delle collaborazioni dell'ILC è stata condotta mediante:

- la partecipazione dell'ILC a numerose proposte progettuali sottomesse in risposta a bandi regionali, nazionali e di fondazioni private, da parte sia di singoli gruppi di ricerca sia di più gruppi congiuntamente: quest'ultima opzione è stata fortemente incoraggiata per potenziare sinergie interne a partire dalla definizione di proposte progettuali. Tra queste, va menzionata la collaborazione con altri istituti CNR per lo sviluppo di progetti interdisciplinari:

DIPARTIMENTO DI SCIENZE UMANE E SOCIALI, PATRIMONIO CULTURALE (DSU)

- Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG)
- Istituto di Scienze e Tecnologie della Cognizione (ISTC)

DIPARTIMENTO DI SCIENZE BIOMEDICHE (DSB)

- Istituto di Fisiologia Clinica (IFC)

DIPARTIMENTO INGEGNERIA, ICT E TECNOLOGIE PER L'ENERGIA E I TRASPORTI (DIITET)

- Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI)
- Istituto di Informatica e Telematica (IIT) di Pisa - Laboratorio WAFI

- Istituto di elettronica e di ingegneria dell'informazione e delle telecomunicazioni (IEIT)
- Istituto di studi sui sistemi intelligenti per l'automazione (ISSIA)
- la promozione di un'azione che vede coinvolte l'Accademia della Crusca, l'Opera di Santa Maria del Fiore di Firenze e altre importanti istituzioni culturali nazionali finalizzata a garantire un futuro a importanti archivi testuali costruiti nel tempo in collaborazione con l'ILC, ora minacciati dall'inevitabile processo di obsolescenza tecnologica. Tali risorse, spesso arricchite con un variegato e prezioso apparato di annotazioni da parte di studiosi, rappresentano un patrimonio culturale di inestimabile valore da salvaguardare e valorizzare;
- la partecipazione attiva alla vita di associazioni scientifiche nazionali operanti nei settori della Linguistica Computazionale e dell'Informatica Umanistica, in particolare: l'*Associazione Italiana di Linguistica Computazionale* (AILC, <http://www.ai-lc.it/it/>), all'interno della quale Simonetta Montemagni svolge il ruolo di vice-presidente e un ricercatore dell'istituto, Felice Dell'Orletta, è membro del Direttivo; l'*Associazione per l'informatica umanistica e la cultura digitale* (AIUCD, <http://www.aiucd.it/associazione/>), all'interno della quale Federico Boschetti, ricercatore dell'Istituto, è membro del Direttivo. Tali azioni sono promosse nella convinzione che l'ILC debba rappresentare un partner attivo e trainante nell'ambito di iniziative associative. Entrambe le associazioni, infatti, stanno svolgendo un ruolo strategico nella creazione di comunità scientifiche nazionali dei settori della Linguistica Computazionale e delle Digital Humanities, con particolare attenzione alla promozione delle attività scientifiche e formative all'interno dei due settori, al consolidamento dei legami con altre iniziative, nazionali, europee e internazionali che operano in questi ambiti e alla loro promozione nell'ambito della politica nazionale, in particolare per quanto riguarda l'università e la ricerca scientifica;
- il ruolo di rappresentante del consorzio italiano di CLARIN: tramite il Coordinatore Nazionale per l'Italia dell'Infrastruttura di ricerca per le Scienze Umane e Sociali CLARIN-ERIC, ruolo assegnato a Monica Monachini, ricercatrice dell'ILC, l'ILC ha condotto una capillare opera di disseminazione presso la comunità dei potenziali utenti nell'ambito sia della Linguistica Computazionale sia delle Scienze Umane e Sociali. Per quanto il consorzio CLARIN-IT sia tuttora in corso di formazione, sono stati già stipulati i primi accordi formali e/o firmate le manifestazioni di interesse da parte di partner nazionali.

Altre collaborazioni informali, nazionali e internazionali, riguardano:

- **CNRS Toulouse** (Francia)
Modellizzazione dell'accesso lessicale in chiave bilingue: evidenza psico-linguistica e simulazioni computazionali con codifica fonologica per tratti, per evidenziare e monitorare in chiave incrementale gli aspetti legati al riconoscimento e ricodifica dell'input della L2
Referente: Claudia Marzi
- **Max Planck Institute for the History of Science** (Berlino, Germania)
Collaborazione per l'integrazione delle tecnologie e dei software implementati dall'ILC nella loro piattaforma di studio dei testi
Referente: Simone Marchi
- **Russian Committee for UNESCO Information for All Programme (IFAP)/ Interregional Library Cooperation Centre**
Scambio di informazioni e know-how per la promozione del multilinguismo sul web
Referente: Claudia Soria
- **Università di Pisa, Laboratorio di Antropologia del Mondo Antico (LAMA)**
Collaborazione con il LAMA tramite seminari, implementazione di software (Euphoria 2.0) per l'annotazione di termini rilevanti per lo studio dei riti in tragedia, stesura di un articolo in rivista, partecipazione a convegni (presso l'Università di Göttingen, e presso AIUCD 2016 a Venezia), tutoraggio per gli stage curriculari
Referente: Federico Boschetti
- **Università di Pisa, Dip. di Filologia, letteratura e linguistica**
Commerce Numérique – Digitalizzazione della rivista letteraria "Commerce"
Referente: Federico Boschetti

▪ **University of Alberta - Dept. of Psychology** (Canada)

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

Referente: Vito Pirrelli

Collaboratori ILC: Claudia Marzi (ILC) e Marcello Ferro (ILC)

▪ **University of Patras, Laboratory of Modern Greek Dialects** (Grecia)

Studio della morfologia del greco moderno e dei suoi dialetti, con particolare riferimento all'analisi dei paradigmi verbali e alla loro modellizzazione teorica e computazionale.

Referente: Vito Pirrelli

▪ **University of Strathclyde** (Glasgow, UK)

Progetto volto a ridurre i tempi di ricerca all'interno di spazi metrici sui quali viene costruito un indice di ricerca.

Referente: Franco Alberto Cardillo

2.6 Valutazione della ricerca dell'ILC

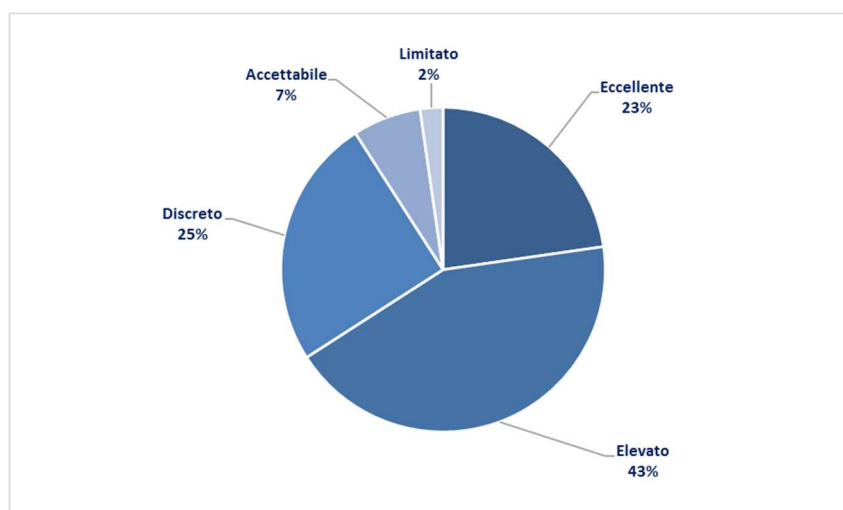
A fine 2015, al termine della valutazione CNR 2011-2014, veniva rilasciato il Rapporto di valutazione dell'ILC. L'esito conseguito dall'Istituto è risultato essere "Low quality", sintetizzato nel punteggio di 57/100. Tale esito contrasta nettamente con i risultati conseguiti in precedenza dall'Istituto nell'ambito della valutazione CNR 2004-2007, e anche con la VQR 2004-2010 il cui esito globale di istituto è stato ricostruito a partire dai risultati conseguiti dai singoli ricercatori. A nostro avviso, hanno contribuito a tale risultato negativo importanti anomalie di metodo e di merito riguardanti i criteri che hanno guidato la valutazione, in particolare il fatto di aver applicato indicatori bibliometrici di area informatica per la valutazione di ricerca condotta in area umanistica, e di non aver tenuto in dovuto conto la missione primariamente umanistica dell'Istituto come definita all'interno dello Statuto e dei piani di programmazione scientifica. Tali anomalie sono state segnalate dalla Direzione e dal Consiglio di Istituto ai vertici dell'Ente, che nel corso del 2016 hanno deciso di procedere con verifiche e integrazioni in merito. A tal fine, è stato interpellato il Prof. Tullio De Mauro che ha accettato l'incarico di fornire un parere circa l'esito della valutazione dell'Istituto. Si riporta di seguito il passaggio finale del parere fornito, che - come si può notare - è in netto contrasto con l'esito "ufficiale" conseguito:

Non pare dubbio che il Rapporto si sia indotto, dati i criteri extraumanistici che ha scelto di assumere, a una valutazione non adeguata dell'attività dell'ILC. Il CNR e i suoi organi bene faranno a correggere drasticamente il giudizio di Low Quality in un giudizio che, a mio avviso, sarebbe tuttavia ancora riduttivo e non adeguato se solo si limitasse a giudicare media la qualità dell'ILC che, come risulta dalla documentazione fornita, nel contesto della comunità scientifica nazionale e internazionale è piuttosto riconosciuta high.

A seguito di ciò, la Direzione si è attivata affinché l'esito della rivalutazione potesse essere acquisito formalmente agli Atti da parte del Consiglio Scientifico dell'Ente.

Nel corso del 2016 la Direzione, coadiuvata dal Responsabile della Biblioteca Gabriella Pardelli, ha coordinato la selezione e la sottomissione delle pubblicazioni del personale dell'ILC per la campagna di valutazione nazionale VQR 2011-2014 alla quale hanno contribuito 18 tra ricercatori e tecnologi, con un totale di 45 prodotti della ricerca: di questi, 42 prodotti sono stati sottomessi per il settore disciplinare L-LIN/01 di area umanistica e 3 per quello di area informatica INF/01. I risultati conseguiti dall'Istituto nell'ambito della VQR 2011-2014, rilasciati a inizio 2017, sono nettamente in contrasto con l'esito conseguito nella valutazione CNR 2011-2014 e sono in linea con quanto emerge dal parere del Prof. De Mauro: come si può osservare nel grafico che segue, il 66% dei prodotti della ricerca (45) dell'ILC hanno ricevuto un giudizio Eccellente (punteggio 1) o Elevato (0,7), con un voto medio di 0,64. Si noti che il periodo delle due valutazioni, CNR e VQR, coincide.

RIPARTIZIONE DEI PRODOTTI ILC PER PUNTEGGIO CONSEGUITO



Tale risultato è nettamente superiore rispetto al voto medio conseguito nella VQR 2004-2010 dal DSU, che era 0,29 per l'Area 10 (è noto che tale risultato è stato condizionato dalla mancata sottomissione dei prodotti richiesti da parte di molti ricercatori). Focalizzandosi sull'Area 10, per la quale sono stati sottomessi 42 dei 45 prodotti dell'ILC, si rileva un voto medio di 0,59 (leggermente inferiore a quello globale) con gli indici R e X (corrispondenti, rispettivamente, al rapporto tra voto medio della sottoistituzione nell'Area e voto medio di Area, e al rapporto tra la frazione di prodotti Eccellenti/Elevati della sottoistituzione nell'Area e la frazione di prodotti Eccellenti/Elevati dell'Area) superiori a 1 (rispettivamente $R=1,09$; $X=1,11$). Ovvero, il voto medio e la frazione di prodotti Eccellenti/Elevati dell'ILC sono maggiori rispetto a quanto rilevato globalmente per l'Area. Si fa inoltre notare che il risultato conseguito è anche superiore alla media CNR per l'Area. Per quanto riguarda l'indicatore finale di qualità della ricerca dell'istituzione, ovvero IRFD, l'ILC è risultato avere un peso quali-quantitativo superiore alla quota di prodotti attesi: si colloca infatti tra i 6 istituti CNR con $IRFD > n/N$, e in particolare risulta essere quello che registra lo scarto maggiore (0,53 vs 0,35). Il Rapporto VQR 2011-2014 completo per quanto riguarda il CNR è disponibile nel sito dell'ANVUR, all'indirizzo <http://www.anvur.org/rapporto-2016/files/Enti/99.CNR.pdf>.

2.7 Premi

BEST PAPER

- **Piccini S.** (2016). *CLAVIUS: verso la modellazione di una risorsa termino-ontologica diacronica del dominio matematico-astronomico del XVII secolo*. Atti del XXVI Convegno internazionale Ass.I.Term "Terminologia e organizzazione della conoscenza nella conservazione della memoria digitale" 14-16 aprile 2016, Università della Calabria, Rende (CS), AIDA informazioni, N. 1-2- Anno 34 – gennaio-giugno 2016, ISBN 978-88-548-9402-0, ISSN 1121-0095.
- Weingart A., **Giovannetti E.** (2016). *A Lexicon for Old Occitan medico-botanical terminology in lemon*. In Proceedings of the Second International Workshop on Semantic Web for Scientific Heritage (SW4SH 2016). Draelants I., Faron Zucker C., Monnin A., Zucker A. (Eds). Heraklion (Greece), 30 maggio 2016, <http://ceur-ws.org/Vol-1595/paper3.pdf>
- Connor R., Vadicamo L., **Cardillo F.A.**, Rabitti F., *Supermetric Search with the Four-Point Property*, Proceedings della conferenza SISAP 2016, 24-26 ottobre 2016, Tokyo, Giappone.
- **Pardelli G., Goggi S., Monachini M., Bartolini R., Russo I.** (2016). *A Geographical Visualization of GL Community: a Snapshot*, in Dominic Farace, Jerry Frantzen (eds.), Eighteenth International Conference on Grey Literature: Leveraging Diversity in Grey Literature, Abstract Book, New York, Best Poster Conference. Amsterdam, TextRelease, p.67, ISBN: 978-90-77484-29-6

SECOND BEST PAPER

- Valsecchi F., Abrate M., Bacciu C., **Piccini S.**, Marchetti A. (2016) *Text Encoder and Annotator: an all-in-one editor for transcribing and annotating manuscripts with RDF*. In Draelants I., Faron Zucker C., Monnin A., Zucker A. (eds), *Proceedings of the Second International Workshop on Semantic Web for Scientific Heritage (SW4SH 2016)*, co-located with 13th Extended Semantic Web Conference (ESWC 2016), May 30th, 2016, Heraklion (Greece), CEUR Workshop Proceedings, vol. 1595, ISSN 1613-0073.

DISTINGUISHED YOUNG PAPER

- Bompalas S., **Ferro M.**, **Marzi C.**, **Cardillo F.A.**, **Pirrelli V.** (2016) *Reassessing Inflectional Regularity in Modern Greek Conjugation*. CLIC-it & EVALITA 2016 Proceedings. CEUR proceedings volume 1749

ALTRI PREMI

- **EVALITA 2016** - Campagna di valutazione per le tecnologie del linguaggio per la lingua italiana scritta e parlata
Cimino A., Dell'Orletta F.
1^a posizione - PoSTWITA task (POS tagging for Italian Social Media Texts)
2^a e 3^a posizione - SENTIPOLC task (SENTiment POLarity Classification)
Napoli, 7 dicembre 2016
- **NVIDIA Hardware Grant** - Dedicated to empowering and collaborating with professors and researchers at universities worldwide to inspire cutting-edge technological innovation
Dell'Orletta F.

3 Attività di ricerca

In conformità con quanto indicato nelle “Linee Guida per la gestione integrata del Ciclo della Performance delle università statali e degli enti pubblici di ricerca italiani” dell’ANVUR (<http://www.anvur.org/attachments/article/833/Linee%20Guida%20EPR.pdf>), le attività condotte nel corso del 2016 sono qui di seguito suddivise in *ricerca scientifica* e *ricerca istituzionale*.

3.1 Ricerca scientifica

Nel 2016 le attività di ricerca scientifica hanno permesso di conseguire risultati apprezzabili, consolidando ed estendendo la visibilità nazionale e internazionale delle varie linee di attività. La rete di contatti e collaborazioni di ricerca è stata estesa e si sono moltiplicate le iniziative di formazione specialistica e di coordinamento delle attività scientifiche.

Segue una sintesi dei principali risultati ottenuti nelle diverse aree di ricerca nel corso del 2016.

DH - Digital Humanities

Nel 2016 le attività svolte presso il Laboratorio CoPhiLab hanno riguardato, in particolare, l’approfondimento teorico relativo alla definizione di filologia collaborativa e cooperativa e lo sviluppo di applicazioni sia per l’annotazione collaborativa di testi letterari greci e latini sia per la validazione della Wordnet basata sul lessico omerico. I risultati più significativi sono stati lo sviluppo di Euporia 2.0, un’applicazione per l’annotazione di tragedie greche per lo studio dei rituali sulla scena, e le nuove iniziative di digitalizzazione di testi moderni, all’interno dei progetti “Voci della Grande Guerra” e “Commerce Numérique” (rivista letteraria francese Commerce 1924-1932).

Presso il Gruppo *Literary Computing* sono stati ulteriormente sviluppati i modelli per la traduzione e lo studio di testi adottati nell’ambito dei progetti “Traduzione del Talmud Babilonese” e “Clavius on the Web”. Per quanto riguarda le tecnologie di Computer-Assisted Translation, il sistema Traduco, di supporto per il lavoro di traduzione, revisione e redazione del testo, è stato esteso con funzionalità avanzate e innovative di supporto alla traduzione basate metodi e tecniche della semantica distribuzionale. Nel 2016 sono inoltre proseguite le attività finalizzate alla costruzione di

risorse lessico-semantiche: è stato rilasciato un primo prototipo per la creazione di lessici e ontologie (LexO) ed è proseguito lo sviluppo del motore morfologico della lingua araba.

È proseguita l'estensione dell'Archivio Storico Testuale dell'ILC, con il recupero di importanti archivi testuali per diverse lingue. Nel corso dell'anno sono state anche condotte sperimentazioni per la valorizzazione di archivi testuali già recuperati, mediante la loro interrogazione attraverso una piattaforma web e la visualizzazione dei contenuti con tecniche di Visual Analytics.

LRI - Risorse, standard e infrastrutture di ricerca

Tra le numerose attività di ricerca svolte nel corso del 2016 vi sono:

- la prosecuzione dello sviluppo di modelli e formati di rappresentazione per la creazione di risorse linguistiche (risultati importanti includono anche la creazione di lessici computazionali diacronici e ontologie);
- la raccolta e conversione di corpora testuali e risorse terminologiche multilingui per la specializzazione dei servizi di traduzione automatica offerti dalla Commissione Europea;
- la prosecuzione degli studi in ambito Grey Literature con la definizione di terminologie specifiche e l'applicazione di tecniche di visualizzazione per il monitoraggio della comunità del settore;
- lo sviluppo delle attività di ricerca sulla Diversità Linguistica Digitale, attraverso la costituzione nell'ambito dell'European Language Resources Association (ELRA) di un gruppo di lavoro internazionale sulle lingue "less-resourced", coordinato da Claudia Soria, ricercatrice dell'ILC.

MIND - Modelli (bio-)computazionali dell'uso linguistico

Il 2016 ha visto un consolidamento delle attività e dei risultati afferenti all'area. Per quanto concerne lo sviluppo di una piattaforma neuro-computazionale per la simulazione di processi di elaborazione seriale, le attività di ricerca si sono articolate come segue:

- validazione e valutazione delle architetture correnti su più lingue (italiano, tedesco, arabo, greco moderno);
- analisi comparativa dei risultati ottenuti su più lingue (italiano, tedesco, arabo, greco moderno);
- sperimentazione del porting delle attuali architetture su sistemi hardware distribuiti per l'elaborazione parallela;
- disegno di architetture a più livelli;
- disegno di piattaforme interattive di supporto alla glottodidattica acquisizionale.

Relativamente agli approcci interdisciplinari a modelli teorici e computazionali di acquisizione lessicale in contesti mono- e multi-lingui, le principali attività di ricerca hanno riguardato:

- la definizione di un modello di memorizzazione e accesso lessicale di forme morfologicamente complesse in contesti monolingui su diverse lingue (tedesco, italiano, inglese, arabo e greco moderno);
- simulazioni incrementali in chiave bilingue che hanno portato ad approfondire il comportamento competitivo tra l'aspettativa per suoni della prima lingua (L1) e la percezione dei fonemi esclusivi della L2.

TAL - Trattamento automatico del linguaggio naturale ed estrazione di conoscenza

Le attività del laboratorio ItaliaNLP Lab si sono focalizzate principalmente sullo studio di modelli probabilistici del linguaggio e sullo sviluppo di algoritmi di apprendimento automatico per il trattamento automatico del linguaggio naturale. I risultati più significativi includono:

- la progettazione e l'adattamento di algoritmi per l'annotazione linguistica automatica, con particolare attenzione al trattamento del linguaggio dei Social Media;
- lo sviluppo di metodi semi-automatici per la creazione di risorse parallele monolingua di frasi semplici e complesse;
- la creazione di sistemi per l'analisi dei Social Media finalizzati alla Sentiment Analysis e al riconoscimento dei testimoni di un determinato evento e per l'analisi automatica di documentazione tecnica (estrazione di informazione da brevetti; analisi dei requisiti scritti in linguaggio naturale; studio diacronico di documentazione tecnica per l'identificazione di tecnologie emergenti);
- l'implementazione di sistemi per la semplificazione automatica del testo attraverso metodi di generazione basati sulle reti neurali.

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)

A seguito dell'adesione dell'Italia a CLARIN-ERIC a fine 2015, l'ILC è stato nominato istituto esecutore e coordinatore nazionale dell'infrastruttura CLARIN-IT e una ricercatrice dell'Istituto, Monica Monachini, ha ottenuto la nomina ministeriale di Coordinatore Nazionale per l'Italia dell'infrastruttura di ricerca. Nel 2016, le attività relative hanno riguardato sia aspetti tecnologici, ovvero la costruzione e il potenziamento di un nodo italiano all'interno della federazione di centri CLARIN-ERIC, sia aspetti strategico-politici, riguardanti l'avvio della creazione del consorzio nazionale sul duplice versante delle comunità della linguistica computazionale e delle Scienze Umane e Sociali.

È stato attivato il primo centro italiano, ILC4CLARIN, con l'installazione del repository DSpace (<http://dspace-clarin-it.ilc.cnr.it>) e il suo collegamento al Registro dei Centri di CLARIN (<https://centres.clarin.eu>). Il centro ILC4CLARIN offre a tutto il consorzio servizi di catalogazione degli strumenti e delle risorse linguistiche italiane esistenti, secondo i metadati e i protocolli stabiliti da CLARIN-ERIC. Il catalogo è collegato al Virtual Language Observatory di CLARIN-ERIC, così da rendere le risorse italiane reperibili nella piattaforma di tutto il network europeo. Si è iniziato altresì a inserire in catalogo le risorse disponibili presso l'ILC. Il centro ha guadagnato la certificazione di tipo C secondo le procedure di CLARIN (repository) ed è in procinto di ottenere il livello superiore B (erogazione di servizi e "data seal of approval"). L'attività ha comportato lo studio approfondito dei Metadati CMDI e l'analisi dei protocolli per l'integrazione dei servizi linguistici negli ambienti e nelle piattaforme per la concatenazione di moduli di analisi linguistica erogati da CLARIN. Sul fronte delle attività di networking, il coordinatore nazionale ha condotto capillare opera di disseminazione di CLARIN presso la comunità dei potenziali utenti attraverso i vari canali di distribuzione: interventi alle Conferenze del settore, relazioni ad invito, scuole estive presso Università e Istituzioni, interviste per riviste specialistiche. Il consorzio è tuttora in fieri, ma sono stati stipulati i primi accordi formali e/o firmate le manifestazioni di interesse da parte dei partner.

DARIAH (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) che mira a fornire servizi avanzati basati sull'uso delle tecnologie dell'informazione e della comunicazione (ICT) per la ricerca nel campo delle Arti e Scienze Umane. L'ILC, in quanto centro esecutore per il CNR del nodo italiano dell'infrastruttura CLARIN, dedicata alle risorse e le tecnologie linguistiche per le Scienze Umane e Sociali e in quanto partner del progetto infrastrutturale Parthenos, promuove la comunicazione, la collaborazione e le sinergie all'interno di DARIAH-IT e CLARIN-IT, nonché la diffusione e la condivisione di esperienze e strumenti nel campo delle Digital Humanities per la trasmissione e l'utilizzo del patrimonio culturale digitale. L'Istituto, in particolare, fornisce contenuti digitali con particolare attenzione all'interoperabilità, alla conservazione e alla sicurezza dei dati stessi. Inoltre, mette a disposizione sotto forma di servizi web le funzionalità di software legacy prodotto presso l'Istituto per l'accesso e la consultazione on-line di contenuti digitali di interesse per la comunità; fornisce strumenti per l'analisi delle lingue classiche e per varietà storiche della lingua, l'estrazione di informazione da testi e il collegamento a thesauri e ontologie. Offre inoltre moduli e catene di annotazione e acquisizione automatica di conoscenza linguistica, nonché competenze nel settore degli standard di rappresentazione e dei formati di annotazione dei dati linguistici.

Ricercatori dell'ILC (Federico Boschetti, Monica Monachini e Simonetta Montemagni) contribuiscono alla governance di DARIAH-IT mediante la partecipazione a diversi Working Groups.

3.3 Pubblicazioni

3.3.1 Contributi in rivista

Basili R., Montemagni S., [Nota Editoriale](#), IJCoL Italian Journal of Computational Linguistics, Vol. 2 n. 1, 2016

Brunato D., Dell'Orletta F., [ISACCO: a corpus for investigating spoken and written language development in Italian school-age children](#), 2016.

Connor R., Cardillo F. A., Vadicamo L. e Rabitti F., [Hilbert exclusion: improved metric search through finite isometric embeddings](#), 2016.

Frontini F., Brando C., Riguet M., Jacquot C., Jolivet V., [Annotation of Toponyms in TEI Digital Literary Editions and Linking to the Web of Data](#), 2016.

Frontini F., Carmen B., Ganascia J. G., [REDEN: Named Entity Linking in Digital Literary Editions Using Linked Data Sets](#), 2016.

Giovannetti E., Albanesi D., Bellandi A., Benotto G., [Traduco: A collaborative web-based CAT environment for the interpretation and translation of texts](#), 2016.

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](#), 2016.

Ide N. e Calzolari N., [Editors' introduction to the special issue: papers from LREC 2014](#), 2016.

Khan, F., Arrigoni, S., Boschetti, F., Frontini, F. (2016) *Restructuring a Taxonomy of Literary Themes and Motifs for More Efficient Querying*. In *Materialities of Literature*, Vol 4. No. 2, 2016.

Lenci A., Labanca N., Marazzini C., Montemagni S., [Voci della Grande Guerra: An Annotated Corpus of Italian Texts on World War I](#), 2016.

Marzi C., Ferro M., Cardillo F. A., Pirrelli V., [Effects of frequency and regularity in an integrative model of word storage and processing](#), 2016.

Marzi C., Ferro M., Nahli O., [Arabic word processing and morphology induction through adaptive memory self-organisation strategies](#), 2016.

Marzi C., Pirrelli V., [Word knowledge and word usage: a foreword](#), in *Italian Journal of Linguistics (IJL)*, Volume 28 (1), Pacini Editore, 2016.

Marzi C., Pirrelli V., [Word knowledge and word usage: a foreword](#), in *Lingue e Linguaggio*, Volume XV (1), Il Mulino, 2016.

Monachini M., Frontini F., [CLARIN, l'infrastruttura europea delle risorse linguistiche per le scienze umane e sociali e il suo network italiano CLARIN-IT](#), IJCoL Italian Journal of Computational Linguistics, Vol 2., n. 2, 2016

Rehm G., Uszkoreit H., Ananiadou S., Bel N., Bieleveciene A., Borin L., Branco A., Budin G., Calzolari N., Daelemans W., Garabik R., Grobelnik M., Garcia-Mateo C., Van Genabith J., Hajic J., Hernaez I., Judge J., Koeva S., Krek S., Krstev C., Linden K., Magnini B., Mariani J., McNaught J., Melero M., Monachini M., Moreno A., Odijk J., Ogrodniczuk M., Pezik P., Piperidis S., Przepiorkowski A., Rognvaldsson E., Rosner M., Pedersen B. S., Skadina I., De Smedt K., Tadic M., Thompson P., Tufis D., Varadi T., Vasiljevs A., Vider K. e Zabarskaite J., [The strategic impact of META-NET on the regional, national and international level](#), 2016.

Rhem G., Uszkoreit H., Calzolari N., Monachini M., [The strategic impact of META-NET on the regional, national and international level](#), 2016.

Soria C. "Language Technologies for Digital Language Diversity." *Scriptus Manet: humanitāro un mākslas zinātņu žurnāls*, Nr. 2. Liepāja: LiePA, pp. 47–60, 2016

Weingart A., Giovannetti E., [Extending the Lemon Model for a Dictionary of Old Occitan Medico-Botanical Terminology](#), 2016.

3.3.2 Contributi in volume

Brunato D., Venturi G., [Le tecnologie linguistico-computazionali per la leggibilità della comunicazione istituzionale](#), 2016.

Del Gratta R., Boschetti F., Del Grosso A., Khan F., Monachini M., [Cooperative philology on the way to web services: The case of the cophiwordnet platform](#), 2016.

Frontini F., Del Gratta R., Monachini M., [GeoDomainWordNet: Linking the Geonames Ontology to WordNet](#), 2016.

Montemagni S., Wieling M., [Tracking linguistic features underlying lexical variation patterns: A case study on Tuscan dialects](#), 2016.

Quochi V., [Development and representation of Italian light-fare constructions](#), 2016.

Soria C. [Speakers' attitudes on Piedmontese: a case for vitality re-assessment](#). In Day, D., Rewi, P. and Higgins, R. (Eds.), *The Journeys of Besieged Languages*. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 135-156, 2016

Wieling M., Simonetta M., [Infrequent forms: Noise or not?](#), 2016.

3.3.3 Contributi in atti di convegno

Arrigoni S., Khan F., Monachini M., Boschetti F., [Misurare Memorata Poetis: prime statistiche](#), 2016.

Barbagli A., Lucisano P., Dell'Orletta F., Montemagni S., Venturi G., [CltA: an L1 Italian Learners Corpus to Study the Development of Writing Competence](#), 2016.

Bartolini R., Pardelli G., Goggi S., Giannini S., Biagioni S., [A terminological "journey" in the Grey Literature domain](#), 2016.

Bellandi A., Benotto G., Di Segni G., Giovannetti E., [Investigating the Application and Evaluation of Distributional Semantics in the Translation of Humanistic Texts: a Case Study](#), 2016.

Benotto G., Giovannetti E., Marchi S., [Investigating the Application of Distributional Semantics to Stylometry](#), 2016.

Benotto G., Giovannetti E., Nahli O., [An application of distributional semantics for the analysis of the Holy Quran](#), 2016.

Berti M., Gregory C., Tariq Y., Yuri B., Federico B., Del Gratta R., [Ancient Greek WordNet meets the Dynamic Lexicon: the example of the fragments of the Greek Historians](#), 2016.

Bompolas S., Ferro M., Marzi C., Cardillo F. A., Pirrelli V., [Reassessing Inflectional Regularity in Modern Greek Conjugation](#), 2016.

Brando C., Abadie N., Frontini F., [Linked Data Quality for Domain-Specific Named-Entity Linking](#), 2016.

Brunato D., Cimino A., Dell'Orletta F., Venturi G., [PaCCSS-IT: A Parallel Corpus of Complex-Simple Sentences for Automatic Text Simplification](#), 2016.

Calzolari N., [Introduction of the Conference Chair and ELRA President Nicoletta Calzolari](#), 2016.

Calzolari N., [Preface: General Chair](#), 2016.

Cardillo R. C. F. A., [Quantifying the specificity of near-duplicate image classification functions](#), 2016.

Connor R., Vadicamo L., Cardillo F., Rabitti F., [Supermetric search with the four-point property](#), 2016.

Corazza A., Montemagni S., Semeraro G., *Preface*, CEUR-WS.org, Vol. 1749 [Proceedings CLiC-it 2016 and EVALITA 2016](#), 2016.

- Del Gratta R., Frontini F., Monachini M., Pardelli G., Russo I., Bartolini R., Khan F., Soria C., Calzolari N., [LREC as a Graph: People and Resources in a Network](#), 2016.
- Del Mario G. A., Federico B., Marchi S., Giovannetti E., [Vantaggi dell'Astrazione attraverso l'Approccio Orientato agli Oggetti per il Digital Scholarly Editing](#), 2016.
- Dell'Orletta F., Venturi G., [ULISSE: una strategia di adattamento al dominio per l'annotazione sintattica automatica](#), 2016.
- Dell'Orletta F., Venturi S. E. G. M., [Esplorazioni computazionali nello spazio dell'interlingua: verso una nuova metodologia di indagine](#), 2016.
- Dolowy-Rybinska N., Soria C., [Surveying the ethnolinguistic vitality of two regional collateral languages: the case of Kashubian and Piedmontese](#), 2016.
- Ferro M., Cardillo F. A., Pirrelli V., Gagné C. L., Spalding T. L., [Written word production and lexical self-organisation: evidence from English \(pseudo\)compounds](#), 2016.
- Frontini F., Carmen B., Ganascia J. G., [REDEN ONLINE: Disambiguation, Linking and Visualisation of References in TEI Digital Editions](#), 2016.
- Giannini S., Biagioni S., Goggi S., Pardelli G., [Grey Literature citations in the age of Digital Repositories and Open Access](#), 2016.
- 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](#), 2016.
- Khan, F., Bellandi, A., Benotto, G., Frontini, F., Giovannetti, E., Reboul, M. (2016). Leveraging A Narrative Ontology to Query a Literary Text. In 7th International Workshop on Computational Models of Narrative (CMN'16).
- Khan A. F., Bellandi A., Monachini M., [Tools and Instruments for Building and Querying Diachronic Computational Lexica](#), 2016.
- Khan F., Díaz-Vera J., Monachini M., [The Representation of an Old English Emotion Lexicon as Linked Open Data](#), 2016.
- Khan F., Javier E. D., Monachini M., [Representing Polysemy and Diachronic Lexico-Semantic Data on the Semantic Web](#), 2016.
- Manzella G. M. R., Bartolini R., Bustaffa F., D'Angelo P., De Mattei M., Frontini F., Maltese M., Medone D., Monachini M., Novellino A., Spada A., [Marine Planning and Service Platform: Specific Ontology Based semantic Search Engine Serving Data Management and Sustainable Development](#), 2016.
- Nahli O., Boschetti F., Arrigoni S., Tessarolo L., [Il corpus di testi arabi in Memorata Poetis](#), 2016.
- Nahli O., Frontini F., Monachini M., Khan F., Zarghili A., Khalfi M., [Al Qamus al Muhit, a Medieval Arabic Lexicon in LMF](#), 2016.
- Pardelli G., Goggi S., Giannini S., Biagioni S., [Two decades of terminology: European framework programmes titles](#), 2016.
- Pardelli G., Goggi S., Monachini M., Bartolini R., Russo I., [A Geographical Visualization of GL Community: a Snapshot](#), 2016.
- Piccini S., Giovannetti E., Bellandi A., Ruimy N., [Le lexique électronique de la terminologie de Ferdinand de Saussure: une première](#), 2016.
- Pieri G., Brunato D., Dell'Orletta F., [Studio sull'ordinamento dei costituenti nel confronto tra generi e complessità](#), 2016.
- Popescu V., Liu L., Del Gratta R., Choukri K., Calzolari N., [New Developments in the LRE Map](#), 2016.

- Russo I., Pisano S., Soria C. 2016. ``[Sardinian on Facebook: Analysing Diatopic Varieties through Translated Lexical Lists](#)``. In Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016, Napoli, Italy, 5-6/12/2016, pp. 263-267
- Soria C., Russo I., Quochi V., Hicks D., Gurrutxaga A., Sarhimaa A., Tuomisto M., [Fostering digital representation of EU regional and minority languages: the Digital Language Diversity Project](#), 2016.
- Stanzione A., Re G., Mugelli G., Taddei A., Boschetti F., Del Gratta R., [Homeric Greek WordNet: costruire una risorsa lessico-semantica fra ricerca e didattica](#), 2016.
- Tusa E., Dell'Orletta F., Montemagni S., Venturi G., [Dieci sfumature di marcatezza sintattica: Verso una nozione computazionale di complessità](#), 2016.
- Recchia V., Ferro M., Maglie R., Dodaro A., [Readability of current patient information leaflets for informed consent in UK radiotherapy centers](#), 2016.
- Rizzetto M., Trevisiol A., Falcone D., Pilon N., Tomè P., Boschetti F., Springmann U., [Nuove frontiere delle Digital Humanities in classe: esperienze dal campo](#), 2016.
- Valsecchi F., Abrate M., Bacciu C., Marchetti A., Piccini S., [Text Encoder and Annotator: an all-in-one editor for transcribing and annotating manuscripts with RDF](#), 2016.
- Weingart A., Giovannetti E., [A lexicon for Old Occitan medico-botanical terminology in lemon](#), 2016.
- Wieling M., Sassolini E., Cucurullo S., Montemagni S., [ALT Explored: Integrating an Online Dialectometric Tool and an Online Dialect Atlas](#), 2016.

3.3.4 Curatele

- Boschetti F. (a cura di), [Book of Abstracts](#), AIUCD 2016
- Branco A., Calzolari N. e Choukri K. (a cura di), [4REAL Workshop: Workshop on Research Results Reproducibility and Resources Citation in Science and Technology of Language](#), 2016.
- Brunato D., Dell'Orletta F., Venturi G., François T., Blache P. (a cura di), [Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity \(CL4LC 2016\)](#), 2016.
- Calzolari N., Choukri K., Declerck T., Goggi S., Grobelnik M., Maegaard B., Mariani J., Mazo H., Moreno A., Odijk J., Stelios P. (a cura di), [Proceedings of the Tenth International Conference on Language Resources and Evaluation \(LREC 2016\)](#), 2016.
- Khan F., Vintar P., Araúz P. L., Faber P., Frontini F., Parvizi A., Gri-Simeunovi L., Unger C. (a cura di), [Language and Ontology \(LangOnto2\) & Terminology and Knowledge Structures \(TermiKS\)](#), 2016.
- Marzi C. e Pirrelli V. (a cura di), [Word knowledge and word usage](#), Special Issue "NetWordS Final Conference Word knowledge and word usage: representations and processes in the mental lexicon", in *Italian Journal of Linguistics (IJL)*, Volume 28 (1), Pacini Editore, 2016
- Marzi C., Pirrelli V. (a cura di) [Word knowledge and word usage](#), Special Issue "NetWordS Final Conference Word knowledge and word usage: representations and processes in the mental lexicon", in *Lingue e Linguaggio*, Volume XV (1), Il Mulino, 2016.
- Montemagni S. (a cura di), CEUR-WS.org, Vol. 1749 [Proceedings CLiC-it 2016 and EVALITA 2016](#), 2016.
- Soria C., Pretorius L., Declerck T., Mariani J., Scannell K., Wandl-Vogt E. (a cura di), [CCURL 2016 Collaboration and Computing for Under-Resourced Languages: Towards an Alliance for Digital Language Diversity](#), 2016.

3.4 Comunicazioni a convegni senza pubblicazione degli atti

- **(Digital) Libraries: Building Participative Communities**
Federico Boschetti – *Risorse digitali per uno sviluppo sostenibile delle comunità*
Catania (Italia), 18 marzo 2016
- **XIV Congresso della Società internazionale di Linguistica e Filologia italiana (SILFI)**
Dominique Pierina Brunato, Felice Dell’Orletta, Simonetta Montemagni, Giulia Venturi - *Tecnologie del linguaggio per l’analisi di corpora di produzioni di apprendenti l’italiano come L1 e L2: metodologie e prospettive*
Madrid (Spagna), Università Complutense di Madrid (UCM), 4-6 aprile 2016
- **Workshop Contested Languages in the Old World 2 (CLOW2)**
Claudia Soria - *Surveying the ethnolinguistic vitality of two regional collateral languages: the case of Kashubian and Piedmontese*
Torino, Università degli Studi di Torino, 5-6 maggio 2016
- **COMET2016 - The 14th International Communication, Medicine and Ethics Conference**
Poster - *Readability of current Patient Information Leaflets for Informed Consent in UK radiotherapy centers*
Marcello Ferro et al.
Aalborg (Danimarca), 4-6 luglio 2016
- **CLARIN Annual Conference - CLARIN Bazaar 2016**
Federico Boschetti - *Web Services for Latin Morphological Analysis*
Aix-en-Provence (Francia), 26-28 ottobre 2016
- **CLARIN workshop "DSpace digital repository"**
Riccardo Del Gratta - *Repository DSPACE-LINDAT*
Lubiana (Slovenia), 8-10 novembre 2016
- **CLARIN workshop on Multilingual corpus annotation tools: development and integration**
Riccardo Del Gratta - *Italian data & tools*
Lubiana (Slovenia), 10-11 novembre 2016
- **GL18 - Eighteenth International Conference on Grey Literature. Leveraging Diversity in Grey Literature**
Sara Goggi:
 - relazione orale: Bartolini R.; Pardelli G.; Goggi S.; Giannini S.; Biagioni S. (2016.) *A terminological "journey" in the Grey Literature domain*
 - presentazione poster: Gabriella Pardelli, Sara Goggi, Monica Monachini, Roberto Bartolini, Irene Russo (2016). *A Geographical Visualization of GL Community: a Snapshot*New York (USA), New York Academy of Medicine, 28-29 novembre 2016
- **CLARIN-PLUS workshop on Oral History (OH)**
Riccardo Del Gratta - *Infrastructure*
Utrecht (Paesi Bassi), 6-7 dicembre 2016

3.5 Internazionalizzazione

Nel 2016 l'Istituto ha proseguito nell'azione di promozione dell'internazionalizzazione della ricerca scientifica e tecnologica nel settore della Linguistica Computazionale. Questo obiettivo è stato perseguito: partecipando a programmi di ricerca e a organismi a livello internazionale; fornendo competenze scientifiche su richiesta di autorità governative; garantendo la collaborazione con enti e 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.

Nel corso del 2016 presso l'Istituto sono stati ospitati visiting scholar provenienti da diverse Università europee per attività di insegnamento e/o ricerca:

- **Christina Gagné Department of Psychology, University of Alberta (Canada)**
Comprensione e produzione delle parole composte
Referente ILC: Vito Pirrelli
24 gennaio 2016 - 3 febbraio 2016
- **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 2015 - gennaio 2016
- **Marianne Reboul - Labex Obvil dell'Università Sorbona IV**
Periodo di formazione e collaborazione alla ricerca presso il gruppo di Literary Computing dell'ILC. Le ricerche condotte hanno permesso di pubblicare, in collaborazione con il gruppo LARI, un articolo sulla rappresentazione di lessici e ontologie diacroniche nella narratologia computazionale
Referente ILC: Emiliano Giovannetti
22 febbraio - 4 marzo 2016

Nel 2016 una ricercatrice dell'ILC è stata ospitata come *visiting scholar* per attività di insegnamento e ricerca presso una università straniera:

- **Francesca Frontini - University of Copenhagen, Center for Sprogteknologi (Danimarca)**
CLARIN Short term mobility
12-16 giugno 2016

4 Attività di alta formazione

4.1 Corsi presso Università

- **Università di Pisa - Informatica Umanistica (classe L-10)**
Corso di Laurea Triennale: insegnamento di Linguistica Computazionale (12 CFU)
Docente: Felice Dell'Orletta (professore a contratto)
a.a. 2016/2017
- **Università di Pisa - Informatica Umanistica (classe L-10)**
Corso di Laurea Magistrale: insegnamento di Linguistica Computazionale II (6 CFU)
Docenti: Simonetta Montemagni (professore a contratto), Giulia Venturi
a.a. 2015/2016

4.2 Summer Schools

- **Université d'été en humanités numériques - Labex OBVIL**
Francesca Frontini – *EREMOS: une outil pour l'exploration des motifs syntaxiques*
Parigi (Francia), Université Sorbonne Paris IV, 4-6 luglio 2016

▪ **CODEFEST 2016**

Claudia Soria - *The Digital Language Diversity Project*
Donostia, 6 luglio 2016

▪ **Trieste Encounters on Cognitive Science (TEX2016), Scuola Internazionale Superiore di Studi Avanzati (SISSA)**
Vito Pirrelli

- *From memory processes to lexical self-organisation: a biologically-motivated integrative view of the morphological lexicon*
- *Word co-activation and competition and the perception of morphological structure*
Trieste, 7-14 luglio 2016

▪ **International Summer School LEX 2016 - MANAGING LEXICAL RESOURCES IN THE SEMANTIC WEB**

Giulia Venturi - *Natural Language processing and legal knowledge extraction*
Ravenna, Università di Bologna - Sede di Ravenna, 8 settembre 2016

▪ **Summer course "New trends in quantitative and computational linguistics"**

Monica Monachini - *Language Resources and Infrastructures for Digital Humanities*
Irene Russo - *Standards and Tools Methods and Model for Automatic Annotation and Extraction of Sentiment*
Fahad Khan - *Electronic Corpora, Digital Editions of Texts and Representation of Diachronic information in Lexicons*
Ciudad Real (Spagna), Universidad de Castilla-La Mancha (UCLM), 28 - 29 settembre 2016

4.3 Supervisione di tesi di laurea e di dottorato

Laurea magistrale

Pietro Bertocchini

Università di Bologna, Dip. di Filologia Classica e Italianistica

Titolo: Ricerche sul Clitofonte

Relatore: Camillo Neri; *Correlatori:* Federico Condello, Federico Boschetti (ILC)

Lorenzo De Mattei

Università di Pisa, Informatica Umanistica

Titolo: Complesso per l'uomo e per la macchina? Uno studio comparativo dei fenomeni di complessità sintattica nei giudizi umani e nell'analisi linguistica automatica

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

Alberto Stefanini

Università degli Studi di Parma, Dip. di Discipline Umanistiche (tesi in corso)

Titolo: Indagine e sperimentazione sulle pratiche d'uso di risorse e strumenti digitali nell'ambito della filologia greca

Controrelatore: Monica Monachini

Ricercatori dell'ILC si sono occupati anche della supervisione di numerose tesi di laurea triennale, la cui lista è riportata in *Appendice II - Supervisione di tesi di laurea triennale*.

Dottorato

▪ *discusse nel 2016*

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

Abstract 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. È 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.

Supervisore ILC: Lucia Marconi

▪ *in corso nel 2016*

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.

Supervisore ILC: Felice Dell'Orletta

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).

Supervisore ILC: Vito Pirrelli

Nell'ambito di accordi di collaborazione tra ILC e Università straniere, ricercatori dell'Istituto svolgono attività di supervisione di studenti di dottorato:

Mustafa Khalfi

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

Argomento: acquisizione del lessico medioevale arabo Al-qamuws al-muhyt in Lemon

Relatore: Arsalan Zarghili

Correlatore: Ouafae Nahli

Younes Lasri

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

Argomento: Optical Character Recognition (OCR) applicato a testi arabi a stampa

Relatore: Arsalan Zarghili

Correlatore: Federico Boschetti

4.4 Tirocini

Nel corso del 2016 sono stati attivati numerosi tirocini incentrati su un'ampia gamma di argomenti e rivolti a studenti iscritti a corsi di laurea di università italiane ed estere:

Tutor: Federico Boschetti

- Università di Pisa, Lettere antiche
Tema: Validazione di *synset di Homeric Greec WordNet*
Tirocinante: Giulia Re
- Università di Pisa, Lettere antiche
Tema: *Validazione di synset di Homeric Greec WordNet*
Tirocinante: Antonio Stanzone

Tutor: Felice Dell'Orletta

- Università di Pisa, Informatica Umanistica
Tema: *Definizione di metodi per rintracciare fenomeni linguistici di complessità del testo in vista della sua semplificazione*
Tirocinante: Sergio Castiglia
- Università di Pisa, Informatica Umanistica
Tema: *Definizione di modelli per la valutazione dell'evoluzione delle competenze sintattiche orali e scritte in bambini italiani della scuola primaria*
Tirocinante: Cesare Cei
- Università di Pisa, Informatica Umanistica
Tema: *Definizione di metodi per lo studio della complessità linguistica per l'uomo e per la macchina*
Tirocinante: Lorenzo De Mattei
- Università di Pisa, Informatica Umanistica
Tema: *Studio di fenomeni linguistici di complessità del testo in vista della sua semplificazione*
Tirocinante: Giulia Pieri
- Università di Pisa, Informatica Umanistica
Tema: *Specializzazione di un metodo di annotazione di entità nominate di dominio alle specificità della lingua dei social media*
Tirocinante: Chiara Scarpitta

Tutor: Monica Monachini

- Università degli Studi di Parma - Convenzione Unica di Tirocinio Formativo
Tema: *Applicazione di metodi e strumenti della filologia computazionale allo studio del greco antico*
Tirocinante: Alberto Stefanini

Tutor: Simonetta Montemagni

- Università di Pisa, Informatica Umanistica
Tema: *Revisione, estensione e armonizzazione della UD Treebank per la lingua italiana*
Tirocinante: Chiara Alzetta
- Università di Pisa, Informatica Umanistica
Tema: *Definizione di modelli computazionali per lo studio dell'evoluzione dell'apprendimento delle capacità di scrittura degli studenti nella scuola secondaria di primo grado*
Tirocinante: Alessio Miaschi
- Università di Pisa, Informatica Umanistica

Tema: *Specializzazione di una metodologia di analisi automatica della leggibilità del testo alle specificità della lingua dei consensi informati*

Tirocinante: Sabrina Rinnone

Tutor: Vito Pirrelli

- University of Patras

Erasmus+ - Erasmus Charter for Higher Education (ECHE)

Tema: *Simulazione computazionale di processi acquisizionali del sistema verbale del greco moderno*

Tirocinante: Stavros Bompolas

Tutor: Claudia Soria

- Università di Pisa, Informatica Umanistica

Tema: *Sviluppo di un'applicazione web di serious gaming per la raccolta di dati sulla percezione della somiglianza linguistica*

Tirocinante: Francesca Danti

4.5 Formazione interna

Nel corso del 2016 è stata condotta un'attenta valutazione dei bisogni formativi del personale ILC dedicato alle attività di ricerca che ha portato alla pianificazione delle attività di formazione interna. In particolare, la ricognizione dei fabbisogni formativi ha identificato come approfondimento e aggiornamento prioritario l'analisi quantitativa del dato linguistico. Il corso programmato e organizzato con diversi moduli e un impegno temporale di 4 giorni, con lezioni frontali sia la mattina sia il pomeriggio, è stato poi erogato nel mese di maggio 2017. Il principale obiettivo individuato è l'acquisizione e l'approfondimento di competenze di analisi quantitative e statistiche indispensabili alla descrizione e alla modellazione della complessità delle informazioni relative ai dati linguistici.

5 Disseminazione scientifica

5.1 Workshop, conferenze, seminari

5.1.1 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 2016 sono da segnalare, in particolare:

- **ELRC Italia - Verso un accesso ai servizi pubblici senza barriere linguistiche**
Workshop dell'European Language Resource Coordination - Edizione italiana
<http://lr-coordination.eu/it/italy>
Roma, Spazio Europa - Rappresentanza in Italia della Commissione Europea, 15 marzo 2016
- **LREC 2016 - Language Resources and Evaluation Conference**
10ª Conferenza sulle Risorse Linguistiche e sulla Valutazione
<http://lrec-conf.org/lrec2016/lrec2016.htm>
Conference Chair: Nicoletta Calzolari
Conference Editorial Committee Member: Sara Goggi
Portorož (Slovenia), 23-28 maggio 2016
- **LangOnto2 + TermiKS - Language and Ontology & Terminology and Knowledge Structures ()**
2° Workshop Congiunto su Lingua e Ontologia & Terminologia e Struttura della Conoscenza (LREC 2016 Workshop)
<http://langandonto.github.io/LangOnto2-TermiKS/>
Workshop Co-Chairs: Fahad Khan, Francesca Frontini
Portorož (Slovenia), Grand Hotel Bernardin Conference Center, 23 maggio 2016

- **CCURL 2016 - Collaboration and Computing for Under-Resourced Languages Workshop**
2° Workshop su Collaborazione ed Uso del Computer per Lingue con Risorse Insufficienti “Verso un’Alleanza per la Diversità Linguistica Digitale” (LREC 2016 Workshop)
<http://www.ilc.cnr.it/ccurl2016>
Workshop Co-Chair: Claudia Soria
Portorož (Slovenia), Grand Hotel Bernardin Conference Center, 23 maggio 2016
- **Workshop di introduzione alle Digital Humanities**
Organizzato dal Laboratorio di Cultura Digitale (LabCD) dell’Università di Pisa, in collaborazione con l’ILC-CNR e con l’Associazione per l’informatica Umanistica e la Cultura Digitale (AIUCD)
<http://www.labcd.unipi.it/evento/seminarioworkshop-sulle-digital-humanities>
Pisa, Università di Pisa, 31 maggio – 1 giugno 2016
- **MODELACT Conference on “Action, Language and Cognition”**
<http://modelact.lablita.it/main.pages/conference.html>
Roma, Consiglio Nazionale delle Ricerche, 6-7 giugno 2016
- **A Place for Places**
Workshop sulle tendenze e sulle sfide attuali nello sviluppo e nell'uso delle basi di dati storico-geografiche (DH 2016 Pre-conference Workshop)
<http://aplace4places.github.io>
Workshop Co-Chair: Francesca Frontini
Cracovia (Polonia), Jagiellonian University and Pedagogical University, 11 luglio 2016
- **Digital Literary Stylistics**
Workshop sulla Stilistica Letteraria Digitale (DH 2016 Pre-conference Workshop)
<http://digitalliterarystylistic.github.io>
Workshop Co-Chair: Francesca Frontini
Cracovia (Polonia), Jagiellonian University and Pedagogical University, 12 luglio 2016
- **AIUCD 2016**
Co-organizzazione del V convegno annuale dell’Associazione per l’Informatica Umanistica e le Culture Digitali (AIUCD)
<http://www.aiucd2016.unive.it>
Venezia, Università Ca’ Foscari, 7-9 settembre 2016
- **ANLP’16 - IEEE CIST’16**
Second Invited Session on “Arabic Natural Language Processing: Models, Systems, Data and Applications”
Workshop internazionale dedicato al trattamento computazionale della lingua e della letteratura araba
Workshop Co-Chair: Vito Pirrelli
Tangeri (Marocco), 24 - 26 ottobre 2016
- **CLiC-it 2016**
3ª Conferenza Italiana di Linguistica Computazionale
<http://clic-it2016.dieti.unina.it>
Conference Co-Chair: Simonetta Montemagni
Napoli, Università Federico II, Biblioteca dell’Accademia Pontaniana, 5-6 dicembre 2016
- **COLING 2016 - 26th International Conference on Computational Linguistics**
26ª Conferenza internazionale sulla linguistica computazionale
<http://coling2016.anlp.jp>
Conference Chair: Nicoletta Calzolari
Workshop Chair: Monica Monachini
Osaka (Giappone), Osaka International Convention Center, 11-16 dicembre 2016
- **CL4LC 2016 - Computational Linguistics for Linguistic Complexity**

Workshop su Linguistica Computazionale e Complessità Linguistica
(COLING 2016 Workshop)
<https://sites.google.com/site/cl4lc2016/>
Workshop Co-Chairs: Dominique Brunato, Felice Dell'Orletta, Giulia Venturi
Osaka (Giappone), Osaka International Convention Center, 11 dicembre 2016

5.1.2 Partecipazione a comitati scientifici di conferenze

Numerosi ricercatori dell'ILC sono membri dei comitati scientifici di conferenze e workshop nazionali e internazionali in settori strategici della Linguistica Computazionale e delle Digital Humanities, tra i quali:

- DH 2016 - Conferenza internazionale Digital Humanities
- DH 2017 - Conferenza internazionale Digital Humanities
- AIUCD 2016 - V Convegno annuale dell'Associazione per l'Informatica Umanistica e le Culture Digitali
- AIUCD 2017 - VI Convegno annuale dell'Associazione per l'Informatica Umanistica e le Culture Digitali
- AI*IA2016 - XV Conferenza internazionale dell'Associazione Italiana per l'Intelligenza Artificiale
- LREC 2016 - X Conferenza sulle Risorse Linguistiche e sulla Valutazione
- COLING 2016 - XXVI Conferenza internazionale sulla linguistica computazionale
- CLIC-IT 2016 - III Conferenza Italiana di Linguistica Computazionale
- CLARIN 2016 - Conferenza Annuale
- AnaMORPHOSys 2016 - Conferenza "Analyzing Morphological Systems"
- IMM'17 - International Morphology Meeting
- IEEE CIST'16 - International Colloquium in Information Science and Technology
- LAPI 2016 - I Simposio internazionale "Endangered languages and language varieties in the Iberian Peninsula: Documentation and revitalization"
- GDDH - Goettingen Dialog for Digital Humanities
- CogaLex'16 - Workshop "Cognitive Aspects of the Lexicon" nell'ambito di COLING 2016
- LO2TKS 2016 - Workshop "Language and Ontology & Terminology and Knowledge Structures"

5.1.3 Comunicazioni e seminari su invito

- **Fondazione Bruno Kessler di Trento (FBK-DH)**
Emiliano Giovannetti e Angelo Mario Del Grosso - *Presentazione al gruppo di Digital Humanities di FBK delle attività di ricerca e sviluppo condotte dal gruppo di Literary Computing dell'ILC*
Povo (TN), 13 - 14 gennaio 2016
- **Biblioteca V. Bellini**
Federico Boschetti - *(Digital) Libraries building partecipative communities*
Evento organizzato nell'ambito della Settimana della Cultura digitale - DiCultHer da: Comune di Catania, Ambasciata USA in Italia, Consolato USA a Napoli, Associazione per l'Informatica umanistica e la Cultura Digitale (AIUCD), Istituto di Scienze e Tecnologie della Cognizione (ISTC-CNR), in collaborazione con Associazione italiana biblioteche (AIB)
Catania, Biblioteca V. Bellini, 18 marzo 2016
- **Università di Pisa**
Seminario organizzato in collaborazione con il Laboratorio di Antropologia del Mondo Antico – LAMA (referente: Prof. Andrea Taddei)
Federico Boschetti - *Filologia collaborativa e cooperativa: osservazioni teoriche e pratiche*
Pisa, 31 marzo 2016
- **Université Sorbonne Paris IV**
Seminario "*Les méthodes de la critique d'attribution du XVI au XXI siècle*"
Francesca Frontini - *Stato di avanzamento delle ricerche su Robert Challe*
Parigi, 2 aprile 2016

- **Università di Pisa**
Seminario nell'ambito del CdS Magistrale di Linguistica e traduzione per il corso di "Teorie e pratiche traduttive" del Prof. Alessandro Lenci
Federico Boschetti – *I testi classici greci e latini con testi a fronte nell'era digitale*
Pisa, 14 aprile 2016
- **Università di Pisa**
Seminario nell'ambito del CdS Magistrale di Linguistica e traduzione per il corso di "Teorie e pratiche traduttive" del Prof. Alessandro Lenci
Emiliano Giovannetti – *La traduzione del Talmud Babilonese in Italiano: soluzioni moderne per sfide antiche*
Pisa, 17 maggio 2016
- **PICGL4 - 4th Patras International Conference of Graduate Students in Linguistics**
Vito Pirrelli – *Co-activation and competition effects in lexical storage and processing* (key-note speech)
Patrasso, 20 maggio 2016
- **IDEM DAY 2016**
Monica Monachini – *CLARIN-IT: Services for the Italian Community of Humanities and Social Sciences*
Giornata di incontro dedicata a tutta la comunità IDEM, l'infrastruttura di Autorizzazione e Autenticazione della rete GARR, e alle diverse comunità di settori quali astrofisica, beni culturali, bio-medicina, scienze umane e sociali, ecc. per presentare le attività degli utenti, i progetti di ricerca e le storie di successo grazie alla collaborazione con IDEM
Roma, Università degli Studi di Roma Tre, 8 giugno 2016
- **Università di Pisa**
Seminario organizzato in collaborazione con il Laboratorio di Cultura Digitale dell'Università di Pisa (referente: Prof.ssa Enrica Salvatori) e il patrocinio dell'Associazione di Informatica Umanistica e Cultura Digitale (AIUCD)
Federico Boschetti – *L'acquisizione del testo da immagini digitali tramite Optical Character Recognition (OCR) e la relativa correzione su piattaforme collaborative*
Pisa, 12 e 27 giugno 2016
- **5° Convegno Annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD)**
Convegno annuale dell'associazione AIUCD, nata per promuovere e diffondere la riflessione metodologica e teorica, la collaborazione scientifica e lo sviluppo di pratiche, risorse e strumenti condivisi nel campo dell'informatica umanistica e nell'uso delle applicazioni digitali in tutte le aree delle Scienze Umane e Sociali
Monica Monachini – *CLARIN-IT, l'Infrastruttura di Ricerca per le Scienze Umane e Sociali*
Venezia, 7-9 settembre 2016
- **3ª Festa della Lingua Sarda Ufficiale**
Claudia Soria – *Su Sardu: una limba digitale?*
Bonarcadu (OR), 24-25 settembre 2016
- **ELEN General Assembly**
Claudia Soria – *The Digital Language Diversity Project*
Barcellona (Spagna), 5 ottobre 2016
- **Università di Parma**
Incontro "Le risorse informatiche applicate alle discipline umanistiche", patrocinato dalla Delegazione di Parma dell'Associazione Italiana di Cultura Classica (A.I.C.C.)
Federico Boschetti – *Risorse informatiche e studi classici: un esempio didattico*
Parma, 10 ottobre 2016
- **Università di Parma**
Lezione a invito, all'interno dei corsi di Storia della Lingua Greca e di Didattica della lingua e letteratura greca
Monica Monachini – *Infrastrutture di ricerca nel settore umanistico*
Parma, 10 ottobre 2016

- **Convegno “Typology, acquisition, grammaticalization studies”**
Claudia Marzi - *Morpho-phonotactic typicality and second language acquisition and processing*
Pavia, 14-15 ottobre 2016
- **IEEE CIST’16 - International Colloquium in Information Science and Technology**
Vito Pirrelli - *Machine learning and neuro-linguistic issues in language processing* (key-note speech)
Tangeri, 25 ottobre 2016
- **Università Milano-Bicocca**
Emiliano Giovannetti - *Relazione nell’ambito del ciclo di lezioni “Scopriamo il Talmud”*
Milano, 26 ottobre 2016
- **International Workshop on Arabic Linguistics and Digital Humanities - Università Roma Tre**
Vito Pirrelli - *Implications of Discriminative Learning for Arabic Morphology* (invited talk)
Roma, 17 novembre 2016

5.1.4 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, 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 2016 sono da segnalare i seguenti seminari:

Data: 28 gennaio 2016

Titolo: [Investigare il ruolo della morfologia e della trasparenza semantica nella produzione scritta delle parole composte e pseudo-composte inglesi](#)

Relatore: Christina Gagné, Department of Psychology, University of Alberta (Canada)

Descrizione: panoramica generale della comprensione e della produzione delle parole composte; discussione sull'importanza della trasparenza semantica per le teorie psicolinguistiche delle parole morfologicamente complesse; presentazione dell'evidenza empirica risultante da un lavoro di digitazione utilizzato per misurare la produzione scritta di composti, pseudo-composti e parole monomorfemiche lettera per lettera.

Data: 4 marzo 2016

Titolo: [Il Progetto Odysseus: uno strumento per lo studio delle traduzioni](#)

Relatore: Marianne Reboul, Labex OBVIL, Sorbonne Paris-IV (Francia)

Descrizione: presentazione dell’Odysseus Tool, uno strumento rapido e accurato per lo studio dell’evoluzione delle traduzioni. Tale strumento permette di analizzare testi monolingui e plurilingui e di accedere contemporaneamente a molte traduzioni di uno stesso testo sorgente, per vedere i possibili pattern evolutivi nelle pratiche di traduzione.

Data: 12 luglio 2016

Titolo: [La validazione di Ancient Greek WordNet: un punto di incontro fra ricerca e didattica](#)

Relatore: Antonio Stanzone, tirocinante presso il laboratorio CoPhiLab

Descrizione: illustrazione del lavoro di validazione della Homeric Greek WordNet (HGWN), costruita a partire dalla Ancient Greek WordNet (AGWN) prendendo in considerazione il solo lessico omerico. Tale lavoro rientra in un progetto pilota che ha l’obiettivo di creare WordNets dei maggiori autori classici, dando alle WordNets risultanti e collegate fra di loro la prospettiva diacronica.

6 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)**

Lanciata nel 2015 come iniziativa editoriale dell'Associazione Italiana di Linguistica Computazionale (AILC), la Rivista Italiana di Linguistica Computazionale 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. 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; diverse applicazioni del trattamento automatico del linguaggio (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***

*Segreteria di redazione: **Sara Goggi***

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

- **LANGUAGE RESOURCES AND EVALUATION (LRE)**

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 50 (2016)

ISSN Stampa: 1574-020X

ISSN Online: 1574-0218

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

*Segreteria di redazione: **Sara Goggi***

Casa editrice: 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 XV (2016)

ISSN: 1720-9331

Direzione Scientifica: Vito Pirrelli, Sergio Scalise

Segreteria di redazione: Claudia Marzi

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

Numerosi ricercatori dell'ILC sono membri di comitati scientifici e redazionali di riviste nazionali e internazionali, così come di collane editoriali, tra i quali:

- Rivista di filologia classica *Lexis-Poetica, retorica e comunicazione nella tradizione classica*, Adolf Hakkert Editore
- *Italian Journal of Computational Linguistics (IJCOL)*, Accademia University Press
- *Language Resources and Evaluation (LRE)*, Springer
- *AI Communications* - journal on Artificial Intelligence, IOS Press
- *ACM Journal on Computing and Cultural Heritage (JOCCH)*, ACM
- *ACM Transactions on Internet Technology (TOIT)*, ACM
- *Computational Linguistics*, MIT Press Journals
- Rivista *Umanistica Digitale*, ABIS - AlmaDL
- Rivista *Frontiers in Digital Humanities - Section Digital Literary Studies*, Frontiers
- *The Mental Lexicon*, John Benjamins
- *Italian Journal of Linguistics*, Pacini Editore
- *Frontiers Human Neuroscience*, Frontiers
- *Journal of King Saudi University – Computer and Information Sciences*, Elsevier
- Collana editoriale *Language Variation*, Language Science Press
- Collana di *Cultura Digitale*, Pacini Editore

7 Attività di terza missione

7.1 Partecipazione a organismi tecnico-scientifici e normativi

- **Nicoletta Calzolari**
Chair del Comitato **ISO/TC37/SC4 on Language Resource Management**
Membro dell'Advisory Group **ISO/TC 37/AG**
- **Francesca Frontini e Fahad Khan**
Membri del **Comitato W3C-Ontolex** per la creazione di un set di raccomandazioni per la rappresentazione di lessici e ontologie per il web semantico
- **Monica Monachini**
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/CT 014 Terminologia della Commissione Tecnica Documentazione e Informazione Automatica.
 - delegato italiano UNI con diritto di voto in rappresentanza dell'Italia all'interno del Comitato ISOTC37/SC4 Language Resource Management

ISO TC37/SC4 – Organizzazione Internazionale di Standardizzazione

- esperto nel gruppo di lavoro ISOTC37/SC4/WG4 LMF standard per risorse lessicali
- esperto all'interno del gruppo di lavoro ISOTC37/SC4 MetaData e ISOCAT

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

7.2 Partecipazione ad Associazioni e Comitati scientifici

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

Federico Boschetti

- Membro del Comitato Direttivo dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD)

Nicoletta Calzolari

- Presidente Onorario 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)

Felice Dell'Orletta

- Socio-fondatore e membro del Direttivo dell'Associazione Italiana di Linguistica Computazionale (AILC)

Monica Monachini

- CLARIN (Common Language Resources and Technology Infrastructure), Coordinatore Nazionale di CLARIN-IT

Simonetta Montemagni

- Socio-fondatore e vice-presidente dell'Associazione Italiana di Linguistica Computazionale (AILC)
- Membro del Language Resources Board (LRB) del Tender European Language Resources Coordination (ELRC, www.lr-coordination.eu) e punto di riferimento nazionale Tecnologico (Technology National Anchor Point- NAP) per l'azione ELRC
- Membro del gruppo di lavoro italiano nell'ambito dell'iniziativa internazionale Universal Dependencies (UD)
- Membro del Board dell'Associazione ELRA (European Language Resources Association)

Claudia Soria

- Vice-presidente dell'European Language Equality Network (ELEN)
- Membership Secretary 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
- Convenor del "Less-Resourced Languages Special Interest Group" creato all'interno dell'associazione ELRA

7.3 Partecipazione a comitati di valutazione

Alcuni ricercatori dell'ILC sono membri di comitati nazionali e internazionali per la valutazione e selezione di dottorati, assegnisti di ricerca, professori e progetti di ricerca, nonché per la Valutazione della Qualità della Ricerca.

Nicoletta Calzolari

Membro del Consiglio di Dottorato di Linguistica dell'Università di Pisa

Simonetta Montemagni

Revisore per la VQR 2011-14, Area 10 (Scienze dell'Antichità, Filologico-Letterarie e Storico-artistiche)

Vito Pirrelli

Revisore per la VQR 2011-14, Area 10 (Scienze dell'Antichità, Filologico-Letterarie e Storico-artistiche)

7.4 Valorizzazione dei risultati e trasferimento tecnologico

Nel corso del 2016 sono proseguite le attività volte alla valorizzazione dei risultati della ricerca dell'ILC a beneficio della società e delle imprese. Esse includono contatti e collaborazioni sia con imprese italiane e internazionali (piccole e medie imprese così come grandi realtà industriali multinazionali) sia con enti pubblici locali, finalizzate al trasferimento tecnologico sia delle competenze acquisite in diversi settori della linguistica computazionale sia di risorse linguistiche e tecnologie linguistico-computazionali. In tal modo è stato possibile diffondere le conoscenze e le competenze acquisite in diversi ambiti della linguistica computazionale e trasferire le risorse linguistiche e le tecnologie acquisite nel settore. Grazie alla collaborazione con specialisti di altre discipline, il personale ILC ha infatti sviluppato metodi e tecnologie innovative che possono trovare applicazione in numerosi ambiti, ad esempio: Pubblica Amministrazione; Istruzione e Formazione; Sanità; Patrimonio Culturale e Turismo; Imprese; Terzo Settore.

Tra queste vale la pena menzionare per il 2016:

- l'accordo di collaborazione con LINKEDIN, insieme all'Università di Pisa e l'Università di Torino per l'estensione della Treebank dell'italiano con annotazione sintattica a dipendenze secondo lo schema "Universal Dependencies" (UD);
- la prosecuzione delle attività finalizzate all'integrazione di funzionalità avanzate basate su tecnologie linguistico-computazionali sviluppate all'ILC all'interno della piattaforma per l'editoria scolastica digitale di META srl attualmente in uso presso Mondadori Education;
- l'integrazione dei sistemi di analisi linguistica e di estrazione di informazione con la piattaforma software sviluppata dalla società Hyperborea per l'analisi dei Social Media allo scopo di rilevare le breaking news (attività svolta nell'ambito del progetto SMART NEWS);
- l'accordo di collaborazione scientifica con INDIRE per lo sviluppo di analisi del dominio educativo e formativo mediante metodi e tecniche di trattamento automatico del linguaggio.

L'ILC ha continuato a contribuire in modo significativo alla conservazione e valorizzazione del patrimonio culturale "invisibile" italiano, in particolare, linguistico, filologico, storico e letterario. Nel corso del 2016, è proseguito il recupero e la conversione in formato standard XML/TEI di importanti archivi testuali con formati di rappresentazione obsoleti.

7.5 Attività di Public Engagement

Nel corso del 2016 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 stakeholder.

7.5.1 Rapporti con le Istituzioni

Nel 2016, ricercatori dell'ILC sono stati interpellati in qualità di esperti di Linguistica Computazionale a livello nazionale per le seguenti audizioni:

- Felice Dell'Orletta, Giulia Venturi
Ministero del Lavoro e delle Politiche Sociali
Direzione Generale per l'inclusione e le politiche sociali

Comitato Tecnico Scientifico dell'Osservatorio nazionale sulla condizione delle persone con disabilità

Tema: valutazione della leggibilità dei quiz per le patenti in vista della loro semplificazione

27 gennaio 2016

- Simonetta Montemagni

Camera dei deputati

Commissioni riunite IX (Trasporti, poste e telecomunicazioni) e X (Attività produttive, commercio e turismo)

Tema: il ruolo delle Tecnologie del Linguaggio nella formazione di un Mercato Unico Digitale

23 marzo 2016

Nel 2016 è inoltre proseguita la collaborazione con il **Centro Gestione Rischio Clinico e Sicurezza dei Pazienti della Regione Toscana** finalizzata all'utilizzo di tecnologie del linguaggio sviluppate all'ILC per il monitoraggio dell'efficacia comunicativa di documenti di ambito biomedico.

7.5.2 Eventi pubblici

- 5 aprile 2016 - Roma, Accademia Nazionale dei Lincei

La prima edizione italiana del Talmud Babilonese

Cerimonia di consegna al Presidente della Repubblica Italiana Sergio Mattarella della prima copia del primo volume del "Trattato di Rosh haShanà", edito da La Giuntina. L'opera è la prima traduzione del Talmud Babilonese in italiano, effettuata da un team di traduttori, revisori editoriali e informatici grazie al sistema "Traduco", strumento all'avanguardia nell'ambito delle traduzioni collaborative. "Traduco" è stato sviluppato dall'ILC nell'ambito del progetto "Traduzione Talmud Babilonese", siglato da un protocollo d'intesa tra la Presidenza del Consiglio dei Ministri, il Ministero dell'Istruzione, dell'Università e della Ricerca, il Consiglio Nazionale delle Ricerche e l'Unione delle Comunità Ebraiche Italiane, e coordinato dal Rabbino Capo di Roma, Riccardo Di Segni.

Nel corso della cerimonia, oltre a Massimo Inguscio, Presidente del Consiglio Nazionale delle Ricerche, sono intervenuti: Alberto Quadrio Curzio, Presidente dei Lincei; Clelia Piperno, Direttore del progetto Traduzione Talmud Babilonese; Alberto Melloni, Segretario della Fondazione per le scienze religiose Giovanni XXIII; Gianni Letta, Presidente del Comitato d'onore del progetto; Renzo Gattegna, Presidente dell'Unione delle Comunità Ebraiche Italiane (UCEI); Adin Even Israel Steinsaltz, Fondatore del Israel Institute for Talmudic Publications; Riccardo Di Segni, Presidente del progetto Traduzione Talmud Babilonese; Stefania Giannini, Ministro dell'Istruzione, dell'Università e della Ricerca.

- 30 settembre 2016 - Pisa, Area della Ricerca di Pisa - CNR

BRIGHT 2016 - Notte Europea dei Ricercatori: in occasione dell'evento l'Istituto ha organizzato i seguenti seminari:

PERCORSO: PATRIMONIO CULTURALE

Da Omero ai fratelli Grimm: grandi e piccoli eroi... a parole!

Esplorare il vocabolario di Omero e confrontare le edizioni delle fiabe dei Grimm.

Comprendere le infrastrutture di ricerca

Comprendere le infrastrutture per la Linguistica Computazionale: un modo per esplorarne metodi e strumenti.

I contenuti Internet per creare servizi al cittadino

Realizzazione di rassegne stampa e corpora testuali da internet e applicazioni web per navigare efficacemente i contenuti.

PERCORSO: SALUTE E BENESSERE | NUOVE TECNOLOGIE

Leggere per capire

Lettura, leggibilità, comprensione, dislessia: le tecnologie per un potenziamento personalizzato.

PERCORSO: NUOVE TECNOLOGIE

Traduco... Il Talmud in italiano

Il software "Traduco" utilizzato da 70 utenti nella traduzione del Talmud Babilonese in Italiano.

Am piasc! M'agradat! Indovina la lingua...

Un gioco per scoprire la diversità linguistica dell'Italia, attraverso lo specchio dei social network.

Due lingue in una. Interazione bilingue in una rete neurale artificiale

Come reti neurali artificiali memorizzano e acquisiscono parole di lingue diverse: pronuncia e interferenza.

Di cosa parli e come parli?

Tecnologie linguistico-computazionali per accedere al contenuto e alla forma linguistica del testo.

Per i contenuti degli interventi cfr. <http://www.ilc.cnr.it/it/content/ilc-bright-2016>

7.5.3 Siti web e social media

Nel 2016 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. Per divulgare le attività svolte, sono stati costantemente aggiornati anche i siti specifici dei diversi laboratori attivi presso l'Istituto. È stato inoltre realizzato un nuovo sito web (<http://licolab.ilc.cnr.it>) per presentare le attività svolte dal gruppo di ricerca "Literary Computing".

Le informazioni relative ai progetti a cui ha preso parte l'ILC sono state rese disponibili sui relativi siti, oltre che in alcune pagine specificatamente create sui social media, ad esempio la pagina Facebook e l'account Twitter "Diversità Linguistica" e la pagina Facebook e l'account Twitter "DLDPProject". Nel caso del progetto DLDP è stata anche prevista la progettazione, preparazione e diffusione di una newsletter a scopo divulgativo, con cadenza mensile.

7.5.4 Trasmissioni radiofoniche

- **Radio Aula 40 – Punto Radio**

<http://radioaula40.cnr.it>

Puntata n. 16 del 5 maggio 2016 - *I disturbi e le difficoltà di apprendimento nell'età scolare*

7.5.5 Iniziative di interazione con scuole

Nel corso del 2016 l'Istituto ha collaborato con Scuole Secondarie di secondo grado per avvicinare gli studenti alla Linguistica Computazionale e ai temi delle Digital Humanities e della filologia collaborativa e cooperativa:

- **Liceo Classico "Rosario Livatino"** di San Marco dei Cavoti (BN)
Progetto "Cultura digitale": presentazione dello strumento sviluppato presso l'ILC per lo studio e l'annotazione di testi in lingua originale con traduzione a fronte
Responsabile ILC: Emiliano Giovannetti
- **Istituto di Istruzione Superiore "Marco Polo" – Liceo Classico, Musicale e Artistico** di Venezia
Liceo Classico e Linguistico "Eugenio Montale" di San Donà di Piave (VE)
Liceo Classico "XXV aprile" di Portogruaro (VE)
Collaborazione nell'ambito del "Greek Studies in 15th Century Europe" (Marie Curie individual research project 2015-17; <http://greek15century.mml.ox.ac.uk>)
Ricercatore ILC: Federico Boschetti (<http://greek15century.mml.ox.ac.uk/people>)
- **Liceo Scientifico "XXV aprile"** di Pontedera (PI)
Attività di tutoraggio per l'Alternanza Scuola Lavoro (studenti: Niccolò Daddi, Mattia Mangini, Tommaso Mansani, Marta Baldini, Davide Lemmi, Elisa Casini, Francesca Andreoli, Elisa Banti - 3 Liceo Scientifico)
Responsabile ILC: Monica Monachini

Nell'ambito del progetto "Clavius on the Web" è stato organizzato il **Workshop Clavius@School**. Si tratta di una serie di incontri a cui hanno partecipato alcune classi di tre licei di Roma (liceo classico e linguistico "Aristofane"; liceo classico "Ennio Quirino Visconti"; liceo ginnasio statale "Virgilio"), coordinati dall'archivio storico APUG e dal CNR (Istituti ILC e IIT). Gli studenti degli ultimi anni dei licei coinvolti hanno potuto discutere e addentrarsi nei problemi

legati allo sviluppo di strumenti software per le scienze umane e coglierne i limiti, migliorando la consapevolezza del mezzo digitale non solo per le sue potenzialità, ma anche per i suoi rischi. Questa discussione ha dato vita a un approfondito dibattito sull'uso responsabile della tecnologia. Nel 2016 sono stati organizzati tre incontri presso la Pontificia Università Gregoriana (PUG):

- *Consapevolezza dei mezzi digitali e strumento di annotazione* (10-11 febbraio 2016)
- *Le ontologie: da Aristotele al web semantico* (21-22 marzo 2016)
- *Dalla annotazione alla ricerca* (27-28 aprile 2016)

7.5.6 Biblioteca

L'ILC ospita una biblioteca ad accesso libero, previo appuntamento. I cataloghi, costantemente aggiornati, sono liberamente consultabili on line in BIBLOS (http://www.biblos.cnr.it/05_ILC.html) e in MOP (<http://mop.isti.cnr.it>).

La Biblioteca è costituita da circa 6.400 volumi, pubblicati a partire dal secondo dopoguerra a oggi. Il patrimonio librario include circa 3.000 monografie, più di 2.000 volumi di letteratura grigia (rapporti di ricerca nazionali e internazionali, atti delle conferenze del settore della linguistica computazionale organizzate dalle principali associazioni internazionali) e una raccolta di 24 titoli tra i principali periodici italiani e stranieri del settore.

Al patrimonio librario della Biblioteca ILC si aggiunge il "Fondo Antonio Zampolli", lasciato in eredità dal fondatore dell'Istituto. Si tratta di una raccolta di inestimabile valore storico, costituita da 1.400 testi specialistici, alcuni dei quali rappresentano esemplari unici in Italia e tra i pochi nel mondo che documentano la nascita della Linguistica Computazionale.

8 English Summary

The main **AREAS OF RESEARCH** at the Institute of Computational Linguistics “Antonio Zampolli” (ILC) of the Italian National Research Council (CNR) are:

- **Natural Language Processing and Knowledge Extraction**
- **Digital Humanities**
- **Resources, Standards and Research Infrastructures**
- **(Bio-)Computational Models of language use**

ILC research is carried out in the framework of research **GROUPS** and **LABORATORIES**:

- **ComPhys Lab** - Design and development of (bio-) computational models of language behaviour with the aim to investigate the juncture of grammatical competence, language usage and neurophysiological and psycho-cognitive correlates of verbal communication and language and communication disorders (<http://www.comphyslab.it>)
- **CoPhiLab** - Formal description of the entities and relations involved in the domain of collaborative philology; creation of digital resources; design and implementation of software components, in particular for classical languages (<http://cophilab.ilc.cnr.it:8080/CoPhiLabPortal>)
- **ItaliaNLP Lab** - Design and development of models, methods, algorithms and technologies for Natural Language Processing, with particular emphasis on the Italian language. Main lines of activity: multi-level linguistic annotation of texts; domain-specific information extraction; development of application prototypes (<http://www.italianlp.it>)
- **LaRI Group** (LARI -Language Resources and Infrastructures, <http://lari.ilc.cnr.it>), that aims to foster research in the field of language engineering and to optimize the production of language resources
- **Literary Computing Group** (<http://licolab.ilc.cnr.it>), that aims to apply Computational Linguistics and Knowledge Engineering to Humanistic Texts and have the following main research topics: textual models; text ontologies; computer-assisted translation; object-oriented design and development of applications for literary computing.

At the end of 2016, ILC **STAFF** included researchers, technologists and technical staff with different backgrounds distributed as follows: **27** units of permanent and temporary staff, and **17** units represented by research fellows and associated personnel.

Research activities carried out during 2016 have led to significant results, consolidating and extending the national and international visibility of the Institute. The network of scientific contacts and research collaborations has been extended as testified by numerous scientific collaboration agreements with Universities and other Institutions signed during the year.

ILC actively participates in two important European **RESEARCH INFRASTRUCTURES**, with different roles:

- **CLARIN-IT** (www.clarin-it.it), the Italian node of *CLARIN - Common Language Resources and Technology Infrastructure* (www.clarin.eu). As CLARIN ERIC member and Leading NC partner, ILC creates and provides access to digital language data collections and digital tools and expertise for researchers. In 2016 ILC has set up the 1st CLARIN-IT National Centre: ILC4CLARIN (ilc4clarin.ilc.cnr.it), that will assist scholars in documenting resources using harmonized metadata descriptions, depositing resources and sharing them with clear licensing policies and secure federated access to protected resources.
- **DARIAH-IT** (<http://it.dariah.eu>), the Italian node of DARIAH-EU (Digital Research Infrastructure for the Arts and Humanities) *DARIAH-EU - Digital Research Infrastructure for the Arts and Humanities* (www.dariah.eu). ILC is partner of this pan-european infrastructure for arts and humanities scholars working with computational methods that supports digital research as well as the teaching of digital research methods.

EXTERNALLY FUNDED RESEARCH PROJECTS

During 2016, ILC was involved in the following externally funded projects:

- **4 European projects/programmes:**

ILC Role	Project
Coordinator	• DLDP - The Digital Language Diversity Project, www.dldp.eu
Partner	• CADDY - Cognitive Autonomous Diving budDY, www.caddy-fp7.eu • PARTHENOS - Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies, www.parthenos-project.eu
Subcontractor	• ELRC-European Language Resource Coordination, http://lr-coordination.eu. ILC is Technology National Anchor Point for Italy and is member of the Language Resources Board of the ELRC.

- **7 national projects:**; TALMUD; TOTUS MUNDUS; VOCI DELLA GRANDE GUERRA; 1 FIRB Project - Future in Research: MODELACT and 3 research projects of national interest (PRIN): MEMORATA POETIS; SCRIBE; CHROME
- **1 international project** as a partner: DiTMAO - Dictionary of Old Occitan medico-botanical terminology
- **2 bilateral agreement** for scientific and technological cooperation with the Polish Academy of Sciences and the Sidi Mohamed Ben Abdellah University
- **2 regional projects:** SMARTNEWS; UBIMOL
- **4 CNR projects:** CLAVIUS - Clavius on the Web; NINFA - iNtelligent Integrated Network For Aged people; SM@RTINFRA-SSHCH; ItalianNLP – WAFI Project on Cyber Intelligence

ILC researchers have won 8 **AWARDS**: 4 Best Paper; 1 Second Best Paper; 1 Distinguished Young Paper; 2 other awards (EVALITA 2016 and NVIDIA Hardware Grant).

More specifically, in 2016 ILC **DISSEMINATION AND OUTREACH ACTIVITIES** included:

- **workshops, conferences, seminars.** ILC researchers have worked for the organization and co-organization of workshops, conferences, seminars such as the *ELRC Workshop in Italy*, and have participated in scientific committees of Italian and international conferences, such as the *Fifth Annual Conference of the Associazione per l'informatica umanistica e la cultura digitale (AIUCD)* and the *Italian Conference on Computational Linguistics-CLIC-IT 2016*. They organized also internal seminars and have been invited for the presentation of papers, talks and communications in many national and international conferences.
- **high-education activities**, such as: courses on Computational Linguistics and Digital Humanities at Pisa University; international summer schools in Italy, France and Spain; external seminars at Pisa University; tutoring and co-tutoring of Bachelor's, Master's and PhD thesis in Italy and abroad; traineeships programs and hosting of visiting students and scholars.
- **editorial activities:** scientific direction of three journals: *Italian Journal of Computational Linguistics – IJCOL*; *Language Resources and Evaluation* and *Lingue e Linguaggio*; different editorial activities and responsibilities in scientific committees of other journals, e.g. *The Mental Lexicon*.
- **participation in technical-scientific and regulatory bodies**, such as UNI; ISO TC37/SC4; ISO/TC 37/AG; CLARIN), **in evaluation committees and in associations and scientific committees** like AILC, ELRC, ELRA and UNESCO.
- **exploitation of research results and technology transfer.** ILC, for example, worked for the implementation of an Italian Treebank with syntactically annotated dependencies within the collaboration between the University of Pisa, Google and LinkedIn.
- **public engagement activities:** dissemination of scientific results through press and public lectures like the Bright 2016 European Night of Researchers; constant updating of the institutional website; development of websites and web services open to the community; common activities with schools and universities.
- **archival and library services:** ILC Library contains about 6,400 volumes and the “Antonio Zampolli Collection”, donated by the founder of the Institute, that is made up of 1,400 specialist texts, some of which are unique examples in Italy and among the few ones in the world representing the origins of Computational Linguistics.

The table below provides a quantitative summary of ILC Research output in 2016:

ILC Research output 2016 – Overview	
Journal article	19
Book chapter/contributions in books	7
Conference proceeding	45
Edited book/proceedings	9
Scientific Direction of journals	3
Master's degree and doctoral thesis supervision	9
Teaching in university courses	2
Teaching modules in summer schools	8
Research projects	20
Research Infrastructures	2
Honors and awards	8

9 Appendice

Appendice I - Personale ILC

SEDE DI PISA	Profilo	N.
		(aggiornato al 31/12/16)
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 2
		Marzi Claudia
		Sassi Manuela (<i>in quiescenza dal 1/11/16</i>)
	Collaboratore tecnico enti di ricerca	Cucurullo Sebastiana (dal 15/11/16) 6
		Gadducci Antonella
		Parrinelli Vanessa
		Parrinelli Vincenzo (<i>in quiescenza dal 1/4/16</i>)
		Picchi Paolo
		Sassolini Eva
		Terreni Noemi
	Collaboratore di amministrazione	Pieri Antonella 1
Personale a tempo determinato	Ricercatore	Del Gratta Riccardo 1
		Frontini Francesca (<i>fino al 31/08/16</i>)
	Tecnologo	Baroni Paola 1
	Collaboratore tecnico enti di ricerca	Albanesi Davide 1
		Cucurullo Sebastiana (<i>fino al 15/11/16</i>)
		(personale in servizio il 31/12/16)
		Totale Personale a tempo indeterminato 21
		Totale Personale a tempo determinato 3
		Totale Contratto Lavoro Diritto Privato 1

SEDE DI GENOVA	Profilo	N. (aggiornato al 31/12/16)
Personale a tempo indeterminato	I° Ricercatore	Marconi Lucia 1
	Collaboratore tecnico enti di ricerca	Cutugno Paola 1
(personale in servizio il 31/12/16) Totale Personale a tempo indeterminato		2

ASSEGNISTI DI RICERCA		N. (aggiornato al 31/12/16)
Sede di Pisa	Bellandi Andrea	12
	Benotto Giulia (fino al 15/11/16)	
	Brunato Dominique Pierina	
	Cardillo Franco Alberto	
	Carlino Michela	
	Cimino Andrea	
	Del Grosso Angelo Mario	
	Mancini Lorenzo	
	Nahli Ouafae	
	Pecchioli Alessandra	
	Piccini Silvia	
	Russo Irene	
	Venturi Giulia	
Sede di Genova	Cinini Alessandra	2
	Lucentini Roberta	
Totale Assegnisti di ricerca (dati aggiornati al 31/12/16)		14

ASSOCIATI	Nome	N. (aggiornato al 31/12/16)
Sede di Pisa	Calzolari Zamorani Nicoletta Parrinelli Vincenzo <i>(dal 1/04/16)</i> Sassi Manuela <i>(dal 1/11/16)</i>	3

Collaboratori	Nome	
Sede di Pisa	Fabio Affè – collaborazione professionale (dal 1/10/16) Khan Anas Fahad - assegno di ricerca co-finanziato con l'Università Ca' Foscari di Venezia (fino al 30/11/2016)	

Appendice II - Supervisione di tesi di laurea triennale

Francesca Danti

Università di Pisa, Informatica Umanistica

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

Relatore: Andrea Marchetti

Correlatore: Claudia Soria

Nicola Nunes

Università di Pisa - Dip. di Filologia, letteratura e linguistica

Titolo: Sguardi sulla tradizione a stampa della fiaba dei fratelli Grimm “Märchen von einem, der auszog, das Fürchten zu lernen” con metodi computazionali

Relatore: Marianne Hepp

Correlatore: Federico Boschetti

Giulia Pieri

Università di Pisa, Informatica Umanistica

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

Relatore: Felice Dell'Orletta

Erica Tusa

Università di Pisa, Informatica Umanistica

Titolo: Verso una nozione computazionale di complessità sintattica: un'analisi linguistico-computazionale del passaggio da prototipico a marcato

Relatore: Felice Dell'Orletta

Redazione a cura di Michela Carlino

Giugno 2017

© 2017 Istituto di Linguistica Computazionale «A. Zampolli» | Tutti i diritti riservati



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