

1 Curriculum

Dati personali

Nome:	Dell'Orletta Felice
Data di nascita:	30/05/1977
Residenza:	via Vecchia Fiorentina 196A, loc. Montione, 56023 Cascina (PI)
Telefono:	3347938570
E-mail:	felice.dellorletta@ilc.cnr.it

Attuale posizione professionale

Ricercatore presso l'Istituto di Linguistica Computazionale A. Zampolli (ILC) del Consiglio Nazionale delle Ricerche (CNR) di Pisa.

Curriculum scolastico

Dottorato: in Informatica, conseguito in data 15/12/2008 presso l'Università di Pisa, dipartimento di Informatica, scuola di dottorato in Scienze di base Galileo Galilei.

Tesi di dottorato dal titolo "*Improving the accuracy of Natural Language Dependency Parsing*". Relatore: Prof. Giuseppe Attardi.

Laurea specialistica: in Informatica, conseguita in data 26/04/2004 presso l'Università di Pisa.

Tesi di laurea dal titolo "*Compressione della struttura del grafo ad albero*". Relatore: Prof. Fabrizio Luccio.

Laurea: in Informatica (triennale), conseguita in data 12/12/2003 presso l'Università di Pisa.

Istruzione superiore: maturità scientifica conseguita presso il liceo scientifico "Marie Curie" di Giulianova (TE).

Scuole di Dottorato

- 2005, Marzo 7-18: "Bertinoro International Spring School" (BISS05)

Borse di Studio e Concorsi

- **16 Settembre 2011 - Oggi:** Ricercatore presso l'Istituto di Linguistica Computazionale A. Zampolli (ILC) del Consiglio Nazionale delle Ricerche (CNR) di Pisa
- **1 Agosto 2009 - 15 Settembre 2011:** Vincitore di contratto di prestazione d'opera triennale per giovane ricercatore presso l'Istituto di Linguistica Computazionale (ILC) del Consiglio Nazionale delle Ricerche (CNR) di Pisa, nell'ambito del progetto FIRB, approvato dal ministero dell'Università e della ricerca con decreto ministeriale n. 1277/RIC. del 17 novembre 2008, dal titolo: "PAISA: Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati". Contratto ottenuto per lo svolgimento della seguente attività: "*Annotazione multi-livello di un Corpus dell'Italiano Contemporaneo on-line mediante tecniche robuste di parsing incrementale*", sotto la responsabilità scientifica del Dott. Vito Pirrelli.
- **1 Giugno 2009 - 31 Luglio 2009:** Vincitore di Assegno di Ricerca presso l'istituto di Linguistica Computazionale (ILC) del Consiglio Nazionale delle Ricerche (CNR) di Pisa, nell'ambito del progetto di ricerca "Semantic Scouting", afferente alla Commessa IC.P02.004 "Linguaggio, Cognizione e Computazione", Modulo IC.P02.004.001 "Tecnologie linguistiche e gestione della conoscenza". attività di ricerca sul tema "*Sviluppo di strumenti adattativi per l'analisi a dipendenze dell'Italiano finalizzati all'estrazione di informazioni da testi di dominio e alla classificazione documentale*", sotto la responsabilità scientifica del Dott. Vito Pirrelli.
- **15 Gennaio 2008 - 31 Maggio 2009:** Vincitore di Borsa di Studio presso il Dipartimento di Informatica di Pisa per lo svolgimento di una ricerca dal titolo: "Tagging Semantico", provvedimento n.24/V del 22.11.2007. Tale ricerca è stata sviluppata all'interno del progetto: "SemaWiki: Analisi di testi per il Semantic Web e il Question Answering", sotto la responsabilità scientifica del Prof. Giuseppe Attardi.
<http://medialab.di.unipi.it/wiki/SemaWiki>.
- **Settembre 2005 - 1 Gennaio 2008:** Vincitore di Borsa di Studio presso l'Istituto di Linguistica Computazionale (ILC) del Consiglio Nazionale delle Ricerche (CNR) di Pisa. Borsa emessa per studi e ricerche nell'ambito della seguente tematica: "Metodi computazionali per l'analisi del linguaggio basati su modelli probabilistici e subsimbolici". (126.048.BS Prot. ILC n. 49/05), sotto la responsabilità scientifica del Dott. Vito Pirrelli.

Attività professionale

- **Settembre 2004 - Maggio 2005:** Contratto di Consulenza presso l'Istituto di Linguistica Computazionale (ILC) del Consiglio Nazionale delle Ricerche (CNR) di Pisa per lo sviluppo di modelli stocastici basati sul metodo della massimizzazione dell'Entropia per la disambiguazione e classificazione nel trattamento automatico del linguaggio naturale. Responsabilità scientifica del Dott. Vito Pirrelli.

Contratto di Consulenza presso il Dipartimento di Linguistica dell'Università degli studi di Pisa per lo studio e lo sviluppo di modelli stocastici per la disambiguazione del problema del Soggetto/Oggetto e lo sviluppo di sistemi in grado di estrarre informazione semantica da testi scritti in linguaggio naturale [5][6]. Responsabilità scientifica del Dott. Alessandro Lenci.

Attività didattica e seminari

Attività didattica

- Anno accademico 2012-2013. Professore a contratto nell'ambito del Corso di laurea triennale Informatica Umanistica (classe L-10) dell'Università di Pisa del corso di Linguistica Computazionale.
- Anno accademico 2011-2012. Professore a contratto nell'ambito del Corso di laurea triennale Informatica Umanistica (classe L-10) dell'Università di Pisa del corso di Linguistica Computazionale.
- Anno accademico 2010-2011. Professore a contratto nell'ambito del Corso di laurea triennale Informatica Umanistica (classe L-10) dell'Università di Pisa del corso di Linguistica Computazionale.
- Anno accademico 2010-2011. Cultore della materia del corso Tecnologie linguistiche per l'estrazione di Informazione della laurea magistrale Informatica Umanistica (classe LM-43) dell'Università di Pisa.

Seminari su invito

- 19/11/2009. Seminario dal titolo "Tecniche e Sistemi Automatici per il Reperimento di Eventi e Relazioni Temporalì", presso Dipartimento di Linguistica Teorica e Applicata dell'Università di Pavia. Convegno dal titolo "EVENTUALITÀ, TEMPORALITÀ, TESTUALITÀ: Aspetti Teorici e Computazionali", nell'ambito del corso di Dottorato di Ricerca in Linguistica.

http://lettere.unipv.it/uploads/convegni/1398926846_Programma.Pavia_sito.pdf

- 22/06/2009. Seminario dal titolo “Modelli e strumenti per l’annotazione linguistica stocastica del testo”, presso l’Istituto di Linguistica Computazionale (ILC) del Consiglio Nazionale delle Ricerche (CNR) di Pisa.

<http://www.ilc.cnr.it/viewpage.php/sez=manifestazioni/id=907/vers=ita>

- 14/06/2006. Seminario dal titolo “NLP ed estrazione di conoscenze”, al “Reading Group on Language and Intelligence”, organizzato dal prof. Giuseppe Attardi nell’ambito del corso di Dottorato in Informatica dell’Università di Pisa.

<http://medialab.di.unipi.it/web/Language+Intelligence/>

- 21/11/2005. Seminario dal titolo “Introduzione all’utilizzo di Hidden Markov Model per l’allineamento di sequenze biologiche”, Dipartimento di Informatica di Pisa.

<http://compass2.di.unipi.it/prenotazioneEventi/seminari.aspx?id=349>

Incarichi scientifici

Membro del Comitato Scientifico, presieduto dalla Prof.ssa Savina Raynaud, del “Centro interdisciplinare di ricerche per la Computerizzazione dei segni dell’espressione (CIRCSE)” dell’Università Cattolica del Sacro Cuore. Quadriennio 2009/2010-2012/2013.

Iniziative scientifiche

- Co-organizzatore della campagna di valutazione dei sistemi di analisi sintattica a dipendenze di testi giuridici scritti in italiano ed in inglese (“First Shared Task on Dependency Parsing of Legal Texts”), alla Conferenza “Semantic Processing of Legal Texts” (SPLeT 2012), Istanbul, 27 Maggio 2012.
- Co-organizzatore della campagna di valutazione di metodi di adattamento a nuovi domini dei sistemi di analisi sintattica a dipendenze dell’italiano (“Domain Adaptation for Dependency Parsing track”), alla Conferenza “Evalita’11 - Evaluation of NLP and Speech Tools for Italian”, Roma, 24-25 Gennaio 2012.

- Co-organizzatore della campagna di valutazione dei sistemi di analisi sintattica a dipendenze dell'italiano (“Dependency parsing track”), alla Conferenza “Evalita’09 - Evaluation of NLP and Speech Tools for Italian”, Reggio Emilia, 12 Dicembre 2009.
<http://evalita.fbk.eu/parsing.html>.

Riconoscimenti

- “Best Verifiability, Reproducibility, and Working Description award” assegnato dalla Commissione di Controllo del Software (“Software Reviewing Committee”) della conferenza “14th International Conference on Intelligent Text Processing and Computational Linguistics” (CICLing-2013) per il software descritto nell’articolo [36]. I criteri presi in considerazione sono stati: la chiarezza, la semplicità, la completezza e la qualità generale del codice che accompagna l’articolo.
- Prima posizione nella competizione sul **Part-of-Speech Tagging** della campagna di valutazione di strumenti per il trattamento automatico dell’Italiano “Evaluation of NLP and Speech Tools for Italian 2009 (EVALITA 2009)”. Articolo: [21].
<http://evalita.fbk.eu/pos.html>
- Prima posizione nella competizione sul **Dependency Parsing** della campagna di valutazione di strumenti per il trattamento automatico dell’Italiano “Evaluation of NLP and Speech Tools for Italian 2009 (EVALITA 2009)”. Articolo: [20].
<http://evalita.fbk.eu/parsing.html>
- Terza posizione nella competizione congiunta su **Dependency Parsing** e **Semantic Role Labeling** della conferenza “Conference on Computational Natural Language Learning (CoNLL) 2008”. Articolo: [14].
<http://www.yr-bcn.es/conll2008/>

Competenze

Esami sostenuti durante il corso di Dottorato in Informatica

Felice Dell’Orletta durante il corso di dottorato in Informatica alla scuola di dottorato in Scienze di base “Galileo Galilei”, ha seguito i seguenti esami superando

le relative verifiche conclusive:

- “Machine Learning oriented Natural Language Processing”, Prof. Kiril Ribarov, Charles University, Praga
- “Probabilità, Statistica e Processi Stocastici”, Prof. Francesco Flandoli
- “Logica per l’informatica”, Prof. Andrea Masini
- “An introduction to dynamical systems (a comparative study of discrete and continuous dynamical systems and of their stability)”, Prof. Frederico de Oliveira-Pinto
- “Computational Intelligence”, Prof. Beatrice Lazzerini
- “Data Mining and Soft Computing”, Prof. Francisco Herrera

Lingue straniere

- Buona conoscenza dell’inglese scritto e parlato, soprattutto tecnico/scientifico. Superamento del corso “Intermediate English” (attuale livello B2), presso il Centro Linguistico Interdipartimentale dell’università di Pisa nell’anno 2006-2007.

Linguaggi di programmazione conosciuti

- Ottima conoscenza dei linguaggi c/c++, java, python, perl. Conoscenza acquisita soprattutto attraverso la realizzazione di strumenti avanzati di elaborazione del linguaggio naturale.

Curriculum dell’attività scientifica e di ricerca

A partire dal 2004, il dott. Felice Dell’Orletta ha svolto attività di ricerca nel settore del trattamento automatico del linguaggio, in particolare dell’annotazione linguistica automatica, a partire dalla segmentazione del testo in frasi e successivamente in parole (tokens) fino ai livelli di annotazione linguistica, morfo-sintattica, sintattica a dipendenze e semantica. Nel corso di questo periodo, ha progettato e sviluppato nuove funzioni, metodi e algoritmi per migliorare le prestazioni in termini di accuratezza ed efficienza dei vari sistemi ed algoritmi di annotazione del testo.

Nel contesto generale appena delineato, l’attività scientifica di Felice Dell’Orletta spazia tra diversi settori del trattamento automatico del linguaggio naturale. A

parte la progettazione e lo sviluppo di sistemi basati su metodi di apprendimento automatico supervisionato per l'annotazione linguistica del testo di cui sopra, ha messo a punto innovative metodologie di estrazione della conoscenza a partire da corpora annotati (ad esempio, terminologia di dominio, neologismi) che sono state implementate in prototipi applicativi. Ha inoltre progettato e sviluppato modelli computazionali dell'apprendimento del linguaggio. Tra le attuali linee di ricerca si annoverano lo studio di nuovi metodi automatici per l'auto-apprendimento (Self-Training) e adattamento al dominio (Domain Adaptation) di strumenti di annotazione linguistica basati su diversi metodi di apprendimento automatico (Machine Learning).

Segue un elenco delle principali linee di ricerca di Felice Dell'Orletta, corredato da una breve descrizione delle attività svolte, con riferimento ai risultati conseguiti, in particolare le pubblicazioni e gli algoritmi sviluppati:

- Progettazione e sviluppo di algoritmi per l'annotazione linguistica automatica, basati su metodi di Machine Learning:
 - annotazione morfo-sintattica, sintattica a costituenti non ricorsivi e sintattica a dipendenze
 - annotazione semantica (riconoscimento di entità nominali, annotazione di ruoli semantici)
- Studio e sviluppo di modelli computazionali dell'apprendimento del linguaggio
- Ontology Learning
- Studio e sviluppo di metodi per l'apprendimento automatico (Self-Training) e adattamento al dominio (Domain Adaptation) di strumenti di annotazione linguistica.
- Sviluppo di modelli computazionali basati su tecniche statistico-distribuzionali (Word Space Models)
- Sviluppo di modelli computazionali per il reperimento di Eventi e Relazioni Temporali

Progettazione e sviluppo di algoritmi per l'annotazione linguistica automatica

Si occupa dello studio, sviluppo ed estensione di strumenti per l'annotazione robusta di corpora testuali dell'Italiano. Annotazione a livello morfo-sintattico, sintattico a costituenti non ricorsivi, sintattico a dipendenze e semantico (“semantic role labelling” e “named entity recognition”).

Ha studiato, progettato e sviluppato numerosi algoritmi (implementati all'interno di piattaforme funzionanti) basati su metodi di classificazione stocastica per il trattamento e l'analisi del linguaggio naturale, nei suoi diversi generi e domini testuali. Si rimanda alla sezione *Sviluppo di moduli software* per un breve elenco e una descrizione dei vari moduli software. Alcuni di questi tool definiscono oggi lo stato dell'arte per l'Italiano (vedi sezione *Riconoscimenti*).

Ha progettato e sviluppato numerose architetture per sistemi di analisi dei testi. Queste architetture incorporano tecniche basate sull'analisi linguistica per l'interpretazione, l'estrazione di conoscenze, e la formalizzazione dei testi. Vengono utilizzate per analizzare testi scritti in linguaggio naturale e crearne una rappresentazione strutturata adatta a migliorare le prestazioni in termini di accuratezza dei sistemi che successivamente la sfrutteranno, ad esempio sistemi per il Question Answering, per l'Ontology Learning, etc. Nella sezione *Partecipazione a progetti* è possibile trovare una breve rassegna di queste architetture e dei vari sistemi che le utilizzano in contesti applicativi reali.

In maniera più specifica durante il periodo del dottorato Felice Dell'Orletta si è occupato del Parsing Sintattico multilingue a dipendenze del testo scritto in linguaggio naturale [1]. Nella tesi ha introdotto un nuovo metodo di parsing ed un nuovo metodo di combinazione di alberi sintattici, che definiscono oggi lo stato dell'arte per l'Italiano [20] e per diverse altre lingue [16], sia dal punto di vista dell'accuratezza del sistema, che dell'efficienza computazionale (si rimanda alla sezione *Breve descrizione della tesi di dottorato*). Gli algoritmi proposti nella tesi sono stati considerati i migliori metodi di parsing e combinazione di parsing, dal punto di vista dell'efficienza e dell'accuratezza, da una valutazione riportata nell'articolo "Ensemble Models for Dependency Parsing: Cheap and Good?", scritto da Mihai Surdeanu and Christopher D. Manning (Proceedings of the North American Chapter of the Association for Computational Linguistics Conference (NAACL-2010), 2010).

Per quanto spesso l'attenzione nello studio di nuovi algoritmi o sistemi di analisi si sia focalizzata sull'analisi linguistica dell'italiano quasi tutti gli strumenti sviluppati sono stati testati con successo anche su altre lingue ottenendo prestazioni in linea (dove non oltre) con lo stato dell'arte. Alcuni esempi di lingue analizzate sono: Inglese, Arabo, Bulgaro, Cinese, Ceco, Danese, Olandese, Tedesco, Giapponese, Portoghese, Sloveno, Spagnolo, Svedese, Turco, Latino, etc. Alcuni articoli che illustrano i risultati di analisi condotte su queste lingue sono [1][9][13][14][16][23][31].

Modelli computazionali dell'apprendimento del linguaggio

Si occupa dello sviluppo di modelli computazionali per simulare le dinamiche di apprendimento del linguaggio, con particolare riferimento all'acquisizione delle

relazioni grammaticali e delle costruzioni sintattiche. Sperimenta vari metodi di apprendimento automatico e di modellizzazione stocastica. I risultati di questa linea di attività sono documentati negli articoli [5][6][7].

Ontology Learning

Si occupa di estrazione automatica di terminologia e dell'acquisizione di ontologie e classi semantiche da corpora testuali. In particolare si occupa dell'estrazione e acquisizione di unità terminologiche mono- e polirematiche di Dominio da testi specifici. I risultati di questa linea di attività sono documentati in [4][11][12].

In questo settore ha messo a punto un innovativo metodo di estrazione terminologica su base contrastiva che permette di discriminare tra terminologia “metalingua” e del “linguaggio oggetto” (es. nel caso di testi giuridici tra termini giuridici e termini relativi al dominio legislativo), oppure di identificare neologismi o parole in via di recessione. Tale metodo è stato sviluppato nel sistema: **TermExtractor** [24].

Auto Apprendimento (Self-Training) e adattamento al dominio (Domain Adaptation)

Si occupa di adattamento di strumenti di analisi e annotazione linguistica a specifici domini di applicazione: referti medici e corpora bio-medici in generale (progetto “ConnectToLife”), corpora Giuridici, testi di Letteratura, testi estratti dal web, testi scientifici (progetto “Semantic Scouting”), etc.

Ha sviluppato e sta sviluppando metodi di individuazione automatica degli errori (Automatic Error Detection) allo scopo di semplificare e velocizzare il lavoro di annotazione manuale di grandi corpora testuali (progetto FIRB “PAISA”), e allo scopo di estendere automaticamente corpora di addestramento per gli strumenti di analisi del testo basati su metodi di Machine Learning (progetto “SemaWiki” e “PAISA”). Questo lavoro è particolarmente importante per quelle lingue, come ad esempio l'Italiano, dove i corpora a disposizione non sono sufficientemente grandi per permettere agli strumenti di ottenere prestazioni adeguate in termini di accuratezza di analisi. In breve, l'idea sviluppata è quella di utilizzare una grande quantità di testo annotato automaticamente e identificarne all'interno le analisi corrette, inserendo poi tali analisi identificate nel corpus di addestramento degli strumenti. Tale processo può essere iterato fino a quando si ottengono miglioramenti in termini di accuratezza da parte degli strumenti di analisi.

Sviluppo di modelli computazionali basati su tecniche statistico-distribuzionali

Si occupa di modelli computazionali basati su tecniche statistico-distribuzionali (Word Space Models) sia nell'ottica di individuare similarità semantica tra "elementi" all'interno di un testo, che tra diversi testi. Si vedano in proposito:[8][10]. Nello specifico in questo momento sta studiando nuove rappresentazioni del testo (o degli elementi che lo compongono) e nuove funzioni per il calcolo della similarità e per il raggruppamento o la divisione delle entità testuali in "cluster semantici".

Sviluppo di modelli computazionali per il reperimento di Eventi e Relazioni Temporali

Si occupa dello sviluppo di un sistema di analisi delle espressioni temporali nei testi scritti in lingua italiana. Ha sviluppato un riconoscitore e normalizzatore di espressioni temporali sulle specifiche TimeML.[17] Inoltre ha sviluppato e continua sviluppare un classificatore statistico di relazioni tra le espressioni temporali. [15]

Altre linee di attività in corso

Sviluppo di modelli computazionali per il monitoraggio dell'uso linguistico tra varietà diamesiche, diafasiche e diacroniche a partire da testi annotati linguisticamente.

Breve descrizione della tesi di dottorato, titolo: Improving the accuracy of Natural Language Dependency Parsing

Lo scopo della tesi di dottorato di Felice Dell'Orletta è stato quello di migliorare l'accuratezza dei parser statistici multilingue a dipendenze, sia in termini di accuratezza che di prestazioni rispetto ai metodi ed ai sistemi di parsing presenti nello stato dell'arte. Nello specifico, nella tesi Dell'Orletta è partito utilizzando un parser a dipendenze Shift/Reduce con costi computazionali lineari e accuratezze vicino allo stato dell'arte (DeSR <http://sites.google.com/site/desrparser/>) e ha sviluppato nuovi metodi per aumentarne l'accuratezza evitando di incrementarne i costi computazionali.

Nella parte iniziale della tesi sono stati presentati gli esperimenti ed i risultati ottenuti durante la partecipazione (insieme al professore Giuseppe Attardi, alla

professoressa Maria Simi, Atanas Chanev e Massimiliano Ciaramita) al *multilingual dependency shared task* della conferenza *Computational Natural Language Learning* (CoNLL) del 2007 [9]. Nel 2007 la CoNLL ha proposto un task sul parsing a dipendenze multilingue. Ogni parser doveva essere in grado di analizzare 10 lingue diverse, per le quali l'organizzazione forniva un corpus di addestramento ed un test non annotato su cui valutare i diversi sistemi. Il gruppo di Felice Dell'Orletta è risultato al settimo posto su 23 partecipanti. In questa parte della tesi oltre a descrivere i risultati ottenuti nella competizione, è stata effettuata una accurata analisi degli errori dei migliori sistemi presentati alla conferenza, per rivelare gli aspetti critici dei vari sistemi di parsing applicati alle differenti lingue.

Questo studio ha permesso di definire un nuovo metodo di parsing e un nuovo metodo di combinazione di alberi sintattici e quindi di migliorare le accuratèzze del parser di partenza raggiungendo prestazioni che sono lo stato dell'arte, senza incrementarne i costi computazionali lineari [16]. Il nuovo metodo di parsing, chiamato **Reverse Revision Parsing**, sfrutta un parser Shift Reduce che analizza il testo da sinistra verso destra seguito da un secondo Shift Reduce parser che analizza al contrario la frase utilizzando come informazioni (features) addizionali le analisi fatte dal primo parser. Il nuovo algoritmo di combinazione di alberi a dipendenze, chiamato **Linear Parser Combination**, sfrutta la caratteristica di dover combinare alberi (quindi di avere in input strutture ad albero) per diminuire il costo computazionale quadratico del noto algoritmo di calcolo dell'albero con peso massimo all'interno di un grafo pesato (*Maximum Spanning Tree algorithm*), utilizzato in tutti i sistemi di combinazione di alberi a dipendenze.

Tali metodi sono stati utilizzati per la partecipazione allo shared task della CoNLL 2008 [14], in collaborazione con Giuseppe Attardi, Massimiliano Ciaramita e Mihai Surdeanu. Il task nel 2008 riguardava sia il problema dell'analisi a dipendenze dell'Inglese, che la determinazione del marcamento dei ruoli semantici (Semantic Role Labeling) basati su una rappresentazione a dipendenze per le relazioni semantiche. I risultati ottenuti utilizzando sia il nuovo metodo di parsing che di combinazione di alberi sviluppato nella tesi, ci hanno permesso di piazzarci al terzo posto e di sviluppare l'unico sistema a costo computazionale lineare tra quelli piú accurati.

Parte del lavoro di tesi è stato anche dedicato a una serie di esperimenti finalizzati a determinare un insieme di features per massimizzare le prestazioni di un parser a dipendenze Shift Reduce. È stato, ad esempio, riportato il miglioramento in termini di accuratezza di un parser quando vengono utilizzate informazioni estratte dal testo annotato a "chunk" (siano essi rivisti manualmente oppure annotati in modo automatico). Ed è stato quindi calcolato l'impatto, in termini di accuratezza, del *chunking* nel migliorare le prestazioni di un parser a dipendenze. [13]

È stata anche analizzata la prestazione dei diversi modelli di parsing quando applicati a lingue che sono caratterizzate da un più alto grado di libertà di posizione delle parole all'interno della frase e da una ricca morfologia. Sono stati individuati i casi più critici da analizzare per un parser e attraverso una dettagliata e quantitativa analisi distribuzionale dei dati linguistici è stato messo in evidenza il contributo relativo di un insieme distribuito di fattori grammaticali e semantici nel parsing. Inoltre è stata introdotta l'*animatezza* nel parsing, una feature semantica di solito non presente nei vari corpora a disposizione annotati a dipendenze, e ne è stato discusso e provato l'utilizzo per la disambiguazione di particolari classi complesse come il riconoscimento e la disambiguazione del *soggetto* e l'*oggetto*. [6]

Sviluppo di moduli software

Felice Dell'Orletta ha sviluppato numerosi componenti software per il trattamento automatico del linguaggio naturale. Tali componenti sono stati sviluppati in C, C++, Python e Java e nella maggior parte dei casi sono basati su metodi di apprendimento automatico (Machine Learning). Segue un elenco di alcuni dei vari componenti sviluppati nel tempo, tale elenco non include le numerose personalizzazioni sviluppate:

- **Language Identifier.** Componente basato su metodi di Machine Learning (metodo della massimizzazione dell'Entropia) capace di individuare la lingua nella quale è scritto il testo da analizzare. [25]
- **Sentence Splitter.** Componente basato su metodi di Machine Learning (metodo della massimizzazione dell'Entropia) capace di dividere il testo analizzato in frasi, classificando statisticamente i vari segni di punteggiatura forte. [28]
- **Tokenizer.** Componente basata su un sistema a regole per la segmentazione del testo in frasi. [26][31]
- **Part-of-Speech tagger.** Componente basato su metodi di Machine Learning (massimizzazione dell'entropia, macchine a vettori di supporto) capace di annotare a livello morfo-sintattico le parole (o token) all'interno di una frase.[3][21][28] (stato dell'arte per l'Italiano [21])
- **Chunker.** Componente basata su metodi di Machine Learning (metodo della massimizzazione dell'entropia) capace di annotare costituenti sintattici non ricorsivi (o chunk) all'interno di una frase. [13][29]

- **Dependency Parser.** Componente basato su metodi di Machine Learning (averaged perceptron, massimizzazione dell'entropia, macchine a vettori di supporto) capace di annotare sintatticamente una frase.[9][13][14][16][20] (stato dell'arte per l'Italiano [20]) (In questo caso il mio ruolo, più che lo sviluppo del software nel suo insieme, è stato quello di ricercare e sviluppare nuovi algoritmi per migliorare le prestazioni e l'efficienza del software di partenza). Studio e sviluppo di questi algoritmi sono stati parte del lavoro della tesi di dottorato [1].
- **TermExtractor.** Estrattore di terminologia di dominio, basato su algoritmi statistici non supervisionati. [24]
- **Text-2-Knowledge (T2K).** Co-autore della piattaforma linguistico-computazionale "T2K". Piattaforma sviluppata per l'estrazione di conoscenza da testi, basata su algoritmi statistici supervisionati e non, capace di indicizzare automaticamente il testo utilizzando i termini di dominio estratti, e di creare relazioni semantiche tra le entità estratte dal testo. [11]
- **Indexer.** Componente che, data una lista di termini ed una collezione di dati, calcola la forma prototipica dei termini, calcola la specificità di ogni termine rispetto ai documenti della collezione e indicizza la collezione utilizzando i termini passati.
- **Named Entity Tagger.** Componente basata su metodi di Machine Learning (massimizzazione dell'entropia, macchine a vettori di supporto) che riconosce e classifica le entità nominali. [19][30]
- **TETI.** Componente basata su un complesso sistema a regole per riconoscere e normalizzare le espressioni temporali in base alle specifiche TimeML. [17]
- **TRS-Learner.** Componente basata su metodi di Machine Learning (metodo della massimizzazione dell'entropia) capace di classificare le relazioni tra le espressioni temporali. [15]
- **Felice-Classifiers.** Librerie e moduli per l'implementazione di vari algoritmi di Machine Learning sotto forma di funzioni software. Attualmente utilizzato presso l'ILC-CNR di Pisa.

Partecipazione a progetti

Le attività di ricerca delineate nella precedente sezione sono state spesso condotte all'interno di progetti nazionali che hanno visto coinvolti partecipanti accademici ed industriali.

- **1 Agosto 2009 - Oggi:** progetto FIRB, titolo: “PAISA: Piattaforma per l’Apprendimento dell’Italiano Su corpora Annotati”, approvato dal ministero dell’Università e della ricerca con decreto ministeriale n. 1277/RIC. del 17 novembre 2008, presso l’istituto di Linguistica Computazionale (ILC) del Consiglio Nazionale delle Ricerche (CNR) di Pisa. Compito di Dell’Orletta è lo svolgimento della seguente attività: “*annotazione multi-livello di un Corpus dell’Italiano Contemporaneo on-line mediante tecniche robuste di parsing incrementale*”.

Nel corso di questo primo anno di progetto, l’attività di Felice Dell’Orletta è consistita e consiste in:

- Sviluppo e adattamento al dominio web di tool di annotazione linguistica capaci di analizzare enormi quantità di dati testuali scaricati dalla rete in maniera robusta, accurata ed efficiente. Scopo dello sviluppo di tali strumenti è la creazione di un corpus campione dell’uso dell’Italiano contemporaneo di circa 100 milioni di parole. Gli strumenti già sviluppati in questa prima fase di progetto sono capaci di analizzare il testo e annotarlo ai seguenti livelli di analisi linguistica: segmentazione in frasi (**Sentence Splitter**), segmentazione in token (**Tokenizer**), etichettatura morfo-sintattica (**POS-Tagger**), annotazione di costituenti sintattici non ricorsivi (o chunk) (**Chunker**), dipendenze sintattiche (**Dependency Parser**). Questi strumenti oggi sono parte dei tool di analisi linguistica dell’Istituto di Linguistica Computazionale (ILC) del CNR di Pisa.
- Studio e sviluppo di metodi di auto-apprendimento (self-training) e di adattamento automatico al dominio (domain adaptation) degli strumenti di annotazione linguistica. Attraverso lo studio delle diverse distribuzioni statistiche morfo-grammaticali e sintattiche prodotte dagli strumenti di analisi, viene associata ad ogni scelta di annotazione delle percentuali di *Plausibilità*. Tali percentuali permettono di determinare eventuali errori (o isole di errori) nell’analisi dei testi o anche porzioni di testo annotate in maniera particolarmente accurata. Questo permette di sviluppare dei brocessi di bootstrapping automatico per incrementare i corpora di apprendimento degli strumenti o per adattare i corpora a nuovi domini, e di conseguenza per adattare gli strumenti stessi. Inoltre tale processo permette anche di determinare quali frasi, tra tutte quelle contenute in una collezione di testi, massimizzerebbero le funzioni di apprendimento di uno strumento statistico (chiaramente se introdotte nel training-set del tool) per permettere di annotare manualmente solo quelle, evitando la correzione manuale di enormi quantità di frasi che non introdurrebbero nessuna conoscenza in più al sistema.

- **1 Agosto 2009 - Oggi:** progetto “ConnectToLife”, in collaborazione con NoemaLife Spa. Il progetto mira alla progettazione e alla implementazione di un sistema sperimentale per la classificazione automatica di testi di tipo medico. Nel corso di questo progetto l’attività di Felice Dell’Orletta è consistita in [32]:
 - Adattamento degli strumenti di analisi del testo al dominio medico.
 - Sviluppo di metodi di clustering basati su *Shingling* e distanza di *Levenshtein*.
 - Studio dell’insieme delle caratteristiche (*feature model*) da estrarre dai referti medici per massimizzare l’accuratezza degli algoritmi basati sul Machine Learning nella classificazione di nuovi testi medici.
- **1 Giugno 2009 - 31 Luglio 2009:** progetto di ricerca dal titolo: “Semantic Scouting”, afferente alla Commessa ILC “Linguaggio, Cognizione e Computazione”, Modulo “Tecnologie linguistiche e gestione della conoscenza”. Compito di Dell’Orletta è stato lo svolgimento della seguente attività: “Sviluppo di strumenti adattativi per l’analisi a dipendenze dell’Italiano finalizzati all’estrazione di informazioni da testi di dominio e alla classificazione documentale”.

Nel corso di questo progetto l’attività di Felice Dell’Orletta è consistita in:

- Adattamento degli strumenti di analisi del testo al dominio scientifico dei testi analizzati
- Sviluppo di un estrattore terminologico (**TermExtractor**) particolarmente adatto per l’estrazione di terminologia da testi di dominio specifico. Tale strumento utilizza dei filtri linguistici configurabili per selezionare candidati termini (un termine rilevante, ad esempio, potrebbe essere una sequenza di aggettivi e sostantivi contigui all’interno di una frase). Successivamente, dei filtri statistici basati sulla combinazione di più funzioni statistiche (ad esempio C-Value e NC-Value) ordinano i candidati termini per forza associativa delle parole che li compongono. Una soglia decide quali candidati possono essere davvero considerati termini complessi. Infine, è stato sviluppato un nuovo metodo di *contrasto* che confronta la distribuzione di un termine complesso all’interno di un dominio e lo contrasta con la distribuzione dello stesso in un dominio generico. Da tale contrasto viene determinato un ordinamento dei termini sulla base della loro specificità nel dominio. Tale strumento è il frutto di studi sviluppati in [24]. TermExtractor è oggi utilizzato all’interno dell’Istituto di Linguistica Computazionale (ILC) del CNR di Pisa ed è tutt’ora in una fase di sviluppo e studio di nuove funzioni statistiche e di nuovi algoritmi di contrasto.

- **15 Gennaio 2008 - 31 Maggio 2009:** Progetto di ricerca dal titolo: “SemaWiki: Analisi di testi per il Semantic Web e il Question Answering”, presso il Dipartimento di Informatica di Pisa. <http://medialab.di.unipi.it/wiki/SemaWiki>.

Nel corso di questo progetto l'attività di Felice Dell'Orletta è consistita in:

- Partecipazione attiva alla creazione di nuovi algoritmi e allo sviluppo degli stessi per la realizzazione di un sistema per l'analisi e l'annotazione linguistica della Wikipedia italiana e la creazione di una piattaforma accessibile via Web per le interrogazioni semantiche sui dati presenti nella Wikipedia. Tale sistema è stato sviluppato all'interno del gruppo di ricerca Analisi di testi per il Semantic Web e il Question Answering, presso il Dipartimento di Informatica dell'Università di Pisa (<http://medialab.di.unipi.it/wiki/Projects>).
- Sviluppo di una pipeline di strumenti avanzati di elaborazione del linguaggio naturale, per lo più basati su algoritmi statistici e di apprendimento automatico. Tale pipeline, chiamata Tanl (Text Analytics and Natural Language <http://medialab.di.unipi.it/wiki/Tanl>), è una suite di moduli software composta da: Segmentatore di frasi (**Sentence Splitter**), segmentatore di token (**Tokenizer**), analizzatore morfo-sintattico (**Part-of-Speech tagger**), parser a dipendenze (**Dependency Parser**), riconoscitore e classificatore di entità nominali (**Named Entity Tagger**).
- Sviluppo di nuove tecniche automatiche di bootstrapping per l'incremento di corpora di apprendimento dei parser a dipendenze multilingue.

Risultati software e di risorse annotate prodotte in questo progetto possono essere trovate in <http://medialab.di.unipi.it/wiki/Tanl>. Risultati in termini di pubblicazioni scientifiche sono: [13][14][16][19][20]. Tali pubblicazioni riguardano soprattutto lo sviluppo di nuovi metodi di analisi a livello di parsing a dipendenze, riconoscimento di entità nominali e pos-tagging. Riportano inoltre lo studio e la selezione di nuove informazioni estratte dai testi da utilizzare nelle fasi di addestramento e analisi di algoritmi basati sull'apprendimento automatico, per migliorarne le accuratezza. Particolare importanza in questi studi oltre all'accuratezza degli strumenti è stata data anche alla loro efficienza computazionale.

- **Gennaio 2007 - Gennaio 2008:** progetto “Smartext”. Progetto finanziato dal partner industriale TotalWire s.r.l. Il progetto mirava alla progettazione e sviluppo di un sistema sperimentale per l'analisi linguistica ed estrazione dell'informazione dalle trascrizioni di intercettazioni telefoniche. Nel corso di questo progetto l'attività di Felice Dell'Orletta è consistita in:

- Partecipazione attiva alla progettazione e sviluppo di nuovi algoritmi per la realizzazione di un sistema di analisi e annotazione linguistica dei testi scritti in lingua italiana, presso il Dipartimento di Linguistica Computazionale (ILC) del Consiglio Nazionale delle Ricerche (CNR) di Pisa. Tale sistema ha richiesto la realizzazione di questi componenti software: **Language Identifier**, componente capace di individuare la lingua nella quale è scritto il testo da analizzare; **Sentence Splitter**, componente che dato un testo lo divide in frasi; **Tokenizzatore**, componente capace di analizzare una frase e dividerla in parole o meglio token; **Part-of-speech tagger**, componente capace di analizzare una frase e assegnare ad ogni parola la propria analisi grammaticale; **Chunker**, componente che prende in input una frase analizzata a livello grammaticale ed esegue una prima analisi funzionale, dividendo la frase in insiemi di token contigui denominati "chunk"; **Named Entity Tagger**, componente che prende in input una frase divisa in chunk e identifica e classifica all'interno della frase le entità nominali (nomi propri di persona, nomi propri di luoghi, organizzazioni, prodotti) e varie altre entità considerate interessanti per il particolare dominio del progetto (date, codici fiscali, numeri di telefono, etc.)
- Adattamento degli strumenti appena elencati al dominio delle trascrizioni di intercettazioni telefoniche.

Risultati dal punto di vista software per questo progetto possono essere trovati nei rapporti tecnici e documenti di progettazione: [25][26][27][28][29][30][31]. I risultati in termini di pubblicazioni scientifiche hanno riguardato soprattutto lo studio di alcune classi di ambiguità linguistica particolarmente complesse nella comprensione di quelle lingue che permettono una maggiore libertà nell'ordine delle parole all'interno della frase, ed in particolare modo dell'Italiano. Alcune pubblicazioni sono: [2][3][6][7].

Publicazioni

Tesi di Dottorato

1. "Improving the accuracy of Natural Language Dependency Parsing"
Autore: Felice Dell'Orletta
PhD Thesis. Dipartimento di Informatica, Università di Pisa, 2008.

Articoli in Libro

2. "Corpus-based modelling of grammar variation"

Autori: Dell'Orletta Felice, Lenci Alessandro, Montemagni Simonetta, Pirrelli Vito

Pubblicata in: Andrea Sansò (ed.), *Language Resources and Linguistic Theory*, 38-55. [Materiali Linguistici 59.] Milano: Franco Angeli, 2007.

3. “Tecnologie linguistico-computazionali per il monitoraggio della competenza linguistica italiana degli alunni stranieri nella scuola primaria e secondaria.”

Autori: Dell'Orletta Felice, Simonetta Montemagni, Vecchi Eva Maria, Giulia Venturi

Pubblicata in: *Percorsi migranti: uomini, diritto, lavoro, linguaggi*. p. 319-336, McGraw-Hill, ISBN: 978-88-386-7296-5, 2011.

4. “Domain Adaptation for Dependency Parsing at Evalita 2011”

Autori: Dell'Orletta Felice, Marchi Simone, Montemagni Simonetta, Venturi Giulia, Agnoloni Tommaso e Francesconi Enrico.

Pubblicata in: *Evaluation of Natural Language and Speech Tool for Italian*, LNCS–LNAI, Vol. 7689, Springer–Verlag Berlin Heidelberg, pp. 58–69, (2013).

Articoli in Rivista

5. “Maximum Entropy for Italian Pos Tagging”

Autori: Dell'Orletta Felice, Federico Maria, Lenci Alessandro, Montemagni Simonetta, Pirrelli Vito

Pubblicata in: *Evalita parsing task: an analysis of the first parsing system contest for Italian*. In *Intelligenza Artificiale*, anno IV, num 2 giugno 2007.

6. “Dal testo alla conoscenza e ritorno: estrazione terminologica e annotazione semantica di basi documentali di dominio.”

Autori: Felice Dell'Orletta, Alessandro Lenci, Simone Marchi, Simonetta Montemagni, Vito Pirrelli, Giulia Venturi

Pubblicata in: *Atti del Convegno Nazionale Ass.I.Term 'I-TerAnDo'*, Università della Calabria, 5-7 giugno 2008, Roma, AIDA Informazioni, Anno 26, n. 1-2/2008, 197-218.

Articoli in Conferenza

7. “Climbing the path to grammar: a maximum entropy model of subject/object learning.”

Autori: Dell'Orletta Felice, Lenci Alessandro, Montemagni Simonetta, Pirrelli Vito

Pubblicata in: Proceedings of the ACL-2005 Workshop "Psychocomputational Models of Human Language Acquisition (PsychoCompLA-2005)", University of Michigan, Ann Arbor (USA), 29-30 June 2005.

8. "Probing the space of grammatical variation: induction of cross-lingual grammatical constraints from treebanks."

Autori: Dell'Orletta Felice, Lenci Alessandro, Montemagni Simonetta, Pirrelli Vito

Pubblicata in: Atti del Coling/ACL 2006 Workshop "Frontiers in Linguistically Annotated Corpora", Sydney Convention and Exhibition Centre, Sydney, Australia, 22 July 2006, 21-28.

9. "Searching treebanks for functional constraints: cross-lingual experiments in grammatical relation assignment."

Autori: Dell'Orletta Felice, Lenci Alessandro, Montemagni Simonetta, Pirrelli Vito

Pubblicata in: LREC 2006: 5th International Conference on Language Resources and Evaluation. Genoa, Italy, 24-25-26 May 2006. Proceedings, Paris, The European Language Resources Association (ELRA). CD-ROM, 433-438.

10. "Computational linguistics meets philosophy: a Latent Semantic Analysis of Giordano Bruno's texts."

Autori: Bassi Simonetta, Dell'Orletta Felice, Esposito D., Lenci Alessandro

Pubblicata in: LREC 2006: 5th International Conference on Language Resources and Evaluation. Genoa, Italy, 22-23 May 2006. Workshops and Tutorials. Paris, European Language Resources Association (ELRA). CD-ROM.

11. "Multilingual Dependency Parsing and Domain Adaptation using DeSR."

Pubblicata in: Proceedings of Conference on Natural Language Learning - Shared Task Session of EMNLP-CoNLL 2007, pp 1112-1118, Prague, 2007.

Autori: Attardi Giuseppe, Simi Maria, Dell'orletta Felice, Chaney Atanas, Ciaramita Massimiliano

12. "Combinazioni di parole e spazi semantici: un'analisi computazionale dei testi di Giordano Bruno"

Autori: Lenci Alessandro, Dell'Orletta Felice

- Publicata in: Atti dei Seminari di Signum. Scuola Normale di Pisa, Pisa, 2007.
13. "Text-2-Knowledge: una piattaforma linguistico-computazionale per l'estrazione di conoscenza da testi"
Autori: Dell'Orletta Felice, Lenci Alessandro, Marchi Simone, Montemagni Simonetta, Pirrelli Vito
Publicata in: In Atti del XL Congresso Internazionale di Studi della Società di Linguistica Italiana, (SLI 2006), Vercelli, 20-22 settembre 2006. Giacomo Ferrari (ed.), edizioni Bulzoni, numero pagine 16, Vercelli, Roma, 2007.
Giacomo Ferrari (ed.), edizioni Bulzoni, numero pagine 16, Roma, 2007.
 14. "Acquiring Legal Ontologies from Domain-specific Texts."
Autori: Felice Dell'Orletta, Alessandro Lenci, Simone Marchi, Simonetta Montemagni, Vito Pirrelli, Giulia Venturi
Publicata in: Proceedings of LangTech 2008, Rome, 28-29 February 2008. CD-ROM.
 15. "Chunking and Dependency Parsing"
Autori: Giuseppe Attardi, Felice Dell'Orletta
Publicata in: Proceedings of LREC 2008 Workshop on Partial Parsing, Marrakech, 2008.
 16. "DeSRL: A Linear-Time Semantic Role Labeling System"
Autori: Massimiliano Ciaramita, Giuseppe Attardi, Felice Dell'Orletta, Mihai Surdeanu
Publicata in: Proceedings of the Twelfth Conference on Natural Language Learning, CoNLL-2008 - Shared Task Session, Manchester, 2008.
 17. "Temporal Relations with Signals: the case of Italian Temporal Prepositions"
Autore: Tommaso Caselli, Felice Dell'Orletta, Irina Prodanof
Publicata in: Proceedings of the 16th International Symposium on Temporal Representation and Reasoning (TIME-2009), C. Lutz and J.-F. Raskin (eds), Conference Proceedings, IEEE: 125-132.
 18. "Reverse Revision and Linear Tree Combination for Dependency Parsing"
Autori: Giuseppe Attardi, Felice Dell'Orletta

Publicata in: North American Chapter of the Association for Computational Linguistics - Human Language Technologies short paper (NAACL HLT) 2009 conference.

19. "TETI: a TimeML compliant TimEx tagger for Italian"
Autori: Tommaso Caselli, Felice Dell'Orletta, Irina Prodanof
Publicata in: Proceedings of International Multiconference on Computer Science and Information Technology (IMCSIT 2009), Mragowo, Poland, 12-14 Oct. 2009 Page(s):185-192
20. "Evalita'09 Parsing Task: comparing dependency parsers and treebanks"
Autori: Cristina Bosco, Simonetta Montemagni, Alessandro Mazzei, Vincenzo Lombardo, Felice Dell'Orletta, Alessandro Lenci
Publicata in: Proceedings of Evalita'09 - Evaluation of NLP and Speech Tools for Italian, Reggio Emilia, December 2009.
21. "The Tanl Named Entity Recognizer at Evalita 2009"
Autori: Giuseppe Attardi, Felice Dell'Orletta, Stefano Dei Rossi, Eva Maria Vecchi
Publicata in: Proceedings of Evalita'09 - Evaluation of NLP and Speech Tools for Italian, Reggio Emilia, December 2009.
22. "Accurate Dependency Parsing with a Stacked Multilayer Perceptron"
Autori: Giuseppe Attardi, Felice Dell'Orletta, Maria Simi, and Joseph Turian
Publicata in: Proceedings of Evalita'09 - Evaluation of NLP and Speech Tools for Italian, Reggio Emilia, December 2009.
23. "Ensemble system for Part-of-Speech tagging"
Autori: Felice Dell'Orletta
Publicata in: Proceedings of Evalita'09 - Evaluation of NLP and Speech Tools for Italian, Reggio Emilia, December 2009.
24. "Comparing the influence of different treebank annotations on dependency parsing performance."
Autori: Cristina Bosco, Simonetta Montemagni, Alessandro Mazzei, Vincenzo Lombardo, Felice Dell'Orletta, Alessandro Lenci, Giuseppe Attardi, Maria Simi, Alberto Lavelli, Johan Hall, Jens Nilsson, Joakim Nivre
Publicata in: LREC 2010: 7th International Conference on Language Resources and Evaluation. 17-23 Maggio, Malta.

25. "Improvements in Parsing the Index Thomisticus Treebank Revision, Combination and a Feature Model for Medieval Latin"
Autori: Marco Passarotti, Felice Dell'Orletta
Pubblicata in: LREC 2010: 7th International Conference on Language Resources and Evaluation. 17-23 Maggio, Malta.
26. "A Contrastive Approach to Multi-word Extraction from Domain specific Corpora"
Autori: Francesca Bonin, Felice Dell'Orletta, Simonetta Montemagni, Giulia Venturi
Pubblicata in: LREC 2010: 7th International Conference on Language Resources and Evaluation. 17-23 Maggio, Malta.
27. "Singling out Legal Knowledge from World Knowledge. An NLP-based approach"
Autori: Francesca Bonin, Felice Dell'Orletta, Simonetta Montemagni, Giulia Venturi
Pubblicata in: Proceedings of IV Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT'10). Fiesole, 7 luglio 2010.
28. "Contrastive Filtering of Domain-Specific Multi-Word Terms from Different Types of Corpora"
Autori: Francesca Bonin, Felice Dell'Orletta, Simonetta Montemagni, Giulia Venturi
Pubblicata in: Proceedings of Multiword Expressions: from Theory to Applications (MWE 2010), the 23rd International Conference on Computational Linguistics (COLING2010). Beijing, China, 28 Agosto, 2010.
29. "ULISSE: an Unsupervised Algorithm for Detecting Reliable Dependency Parses"
Autori: Felice Dell'Orletta, Giulia Venturi, Simonetta Montemagni
Pubblicata in: Proceedings of Fifteenth Conference on Computational Natural Language Learning, CoNLL-2011. Portland, Oregon, USA, 23-24 Giugno, 2011.
30. "READ-IT: Assessing Readability of Italian Texts with a View to Text Simplification"
Autori: Felice Dell'Orletta, Giulia Venturi, Simonetta Montemagni
Pubblicata in: Proceedings of EMNLP 2011, Second Workshop on Speech and Language Processing for Assistive Technologies (SLPAT). Edinburgh, Scotland, UK, 27-31 Luglio, 2011

31. “Lessico settoriale e lessico comune nell'estrazione di terminologia specialistica da corpora di dominio”
Autori: Francesca Bonin, Felice Dell'Orletta, Giulia Venturi, Simonetta Montemagni
Pubblicata in: Proceedings of XLIV Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI 2010), Viterbo, 27-29 settembre, 2010. (2012)
32. “Tecnologie linguistico-computazionali per la valutazione delle competenze linguistiche in ambito scolastico”
Autori: Felice Dell'Orletta, Simonetta Montemagni
Pubblicata in: Proceedings of XLIV Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI 2010), Viterbo, 27-29 settembre, 2010. (2012)
33. “Domain Adaptation for Dependency Parsing at Evalita 2011”
Autori: Felice Dell'Orletta, Simone Marchi, Simonetta Montemagni, Giulia Venturi, Tommaso Agnoloni, and Enrico Francesconi
Pubblicata in: Proceedings of Evalita'11 - Evaluation of NLP and Speech Tools for Italian. Roma, Italia, 24th-25th January, 2012
34. “Genre-oriented Readability Assessment: a Case Study”
Autori: Felice Dell'Orletta, Simonetta Montemagni, Giulia Venturi
Pubblicata in: Proceedings of the Workshop on Speech and Language Processing Tools in Education (SLP-TED), pp. 91–98, 2012
35. “Domain Adaptation for Dependency Parsing at Evalita 2011”
Autori: Felice Dell'Orletta, Simone Marchi, Simonetta Montemagni, Giulia Venturi, Tommaso Agnoloni, Enrico Francesconi
Pubblicata in: Magnini B., Cutugno F., Falcone M., Pianta E. (eds.), Evaluation of Natural Language and Speech Tool for Italian, LNCS–LNAI, Vol. 7689, Springer–Verlag Berlin Heidelberg, pp. 58–69. (2013)
36. “Linguistically-driven Selection of Correct Arcs for Dependency Parsing”
Autori: Felice Dell'Orletta, Simonetta Montemagni, Giulia Venturi
Pubblicata in: Computación y Sistemas, ISSN 1405-5546, vol. 17, No. 2, pp. 125-136. (2013)
37. “Automatic extraction of Function-Behaviour-State information from patents”
Autori: Gualtiero Fantoni, Riccardo Aprea, Felice Dell'Orletta, Maurizio Monge

- Publicata in: C.H. Chen and W. O'Brien (eds.) Journal of Advanced Engineering Informatics. (2013)
38. "Linguistic Profiling based on General-purpose Features and Native Language Identification"
- Autori: Andrea Cimino, Felice Dell'Orletta, Simonetta Montemagni, Giulia Venturi
- Publicata in: Proceedings of Eighth Workshop on Innovative Use of NLP for Building Educational Applications, Atlanta, Georgia, 13 June 2013, pp. 207-215.
39. "Unsupervised Linguistically-Driven Reliable Dependency Parses Detection and Self-Training for Adaptation to the Biomedical Domain"
- Autori: Felice Dell'Orletta, Simonetta Montemagni, Giulia Venturi
- Publicata in: Proceedings of the 12th workshop on Biomedical Natural Language Processing (BioNLP-2013), Sofia, Bulgaria, 8-9 August 2013, pp. 45-53.
40. "Mining Commonalities and Variabilities from Natural Language Documents"
- Autori: Alessio Ferrari, Giorgio Oronzo Spagnolo, Felice Dell'Orletta
- Publicata in: Proceedings of the 17th International Software Product Line Conference (SPLC-2013), Tokyo, Japan, 26-30 August 2013.
41. "Linguistic Profiling of Texts Across Textual Genre and Readability Level. An Exploratory Study on Italian Fictional Prose"
- Autori: Felice Dell'Orletta, Simonetta Montemagni, Giulia Venturi
- Publicata in: Proceedings of the Recent Advances in Natural Language Processing Conference (RANLP-2013), Hissar, Bulgaria, 7-13 September 2013.

Rapporti Tecnici

42. "Language Recognition Tool, Specifiche di Implementazione."
- Autori: Felice Dell'Orletta, Maria Federico, Emiliano Giovannetti, Alessandro Lenci, Simone Marchi, Andrea Trabucco, Vito Pirrelli
- Publicata in: Documento di progettazione ILC_TW_2007_01- Progetto Smarttext, Rapporto tecnico.

43. “Segmentazione di un Testo Italiano in Token.”
Autori: Felice Dell’Orletta, Maria Federico, Emiliano Giovannetti, Alessandro Lenci, Simone Marchi, Andrea Trabucco, Vito Pirrelli
Pubblicata in: Documento di progettazione ILC_TW_2007_02-Progetto Smarttext, Rapporto tecnico.
44. “Analisi Morfosintattica per l’Italiano.”
Autori: Felice Dell’Orletta, Maria Federico, Emiliano Giovannetti, Alessandro Lenci, Simone Marchi, Andrea Trabucco, Vito Pirrelli
Pubblicata in: Documento di progettazione ILC_TW_2007_03-Progetto Smarttext, Rapporto tecnico.
45. “Part-of-Speech Tagging per l’Italiano.”
Autori: Felice Dell’Orletta, Maria Federico, Emiliano Giovannetti, Alessandro Lenci, Simone Marchi, Andrea Trabucco, Vito Pirrelli
Pubblicata in: Documento di progettazione ILC_TW_2007_04-Progetto Smarttext, Rapporto tecnico.
46. “Specifiche di Chunking per l’Italiano.”
Autori: Felice Dell’Orletta, Maria Federico, Emiliano Giovannetti, Alessandro Lenci, Simone Marchi, Andrea Trabucco, Vito Pirrelli
Pubblicata in: Documento di progettazione ILC_TW_2007_05-Progetto Smarttext, Rapporto tecnico.
47. “Specifiche di Named Entity Recognition per l’Italiano.”
Autori: Felice Dell’Orletta, Maria Federico, Emiliano Giovannetti, Alessandro Lenci, Simone Marchi, Andrea Trabucco, Vito Pirrelli
Pubblicata in: Documento di progettazione ILC_TW_2007_06-Progetto Smarttext, Rapporto tecnico.
48. “Segmentazione di un Testo Inglese in Token.”
Autori: Felice Dell’Orletta, Maria Federico, Emiliano Giovannetti, Alessandro Lenci, Simone Marchi, Andrea Trabucco, Vito Pirrelli
Pubblicata in: Documento di progettazione ILC_TW_2007_07-Progetto Smarttext, Rapporto tecnico.
49. “Connect-to-Life - Rapporto tecnico intermedio. Progetto Connect-to-Life.”
Autori: Felice Dell’Orletta, Emiliano Giovannetti, Alessandro Lenci, Simone Marchi, Simonetta Montemagni, Vito Pirrelli

Il sottoscritto Felice Dell'Orletta, nato a Atri il 30/05/1977, residente in via Vecchia Fiorentina 196a loc. Montione 56023 Cascina, consapevole della responsabilità penale prevista, dall'art. 76 del D.P.R. 445/2000, per le ipotesi di falsità in atti e dichiarazioni mendaci ivi indicate, dichiara che quanto riportato nel presente curriculum vitae et studiorum corrisponde a verità.

27 agosto 2013

Felice Dell'Orletta