

## Curriculum vitae di Cristina Bosco

### Studi ed attività di ricerca

- Da ottobre 2006, è ricercatrice afferente al dipartimento di Informatica presso la Facoltà di Scienze Matematiche Fisiche e Naturali dell'Università di Torino.
- Da dicembre 2004, è titolare di una **borsa post-dottorato Lagrange Fellow** della durata di due anni, finanziata dalla Fondazione ISI (Istituto per l'Interscambio Scientifico), per lo svolgimento, presso la Facoltà di Psicologia dell'Università di Torino e sotto la supervisione del prof. B. Bara e del prof. Lombardo, di un progetto di ricerca che si colloca, da un lato, nell'area del Natural Language Processing (NLP) e, dall'altro, nell'ambito della teoria della complessità, intitolato "*Human language as a complex system: study of variation phenomena with corpus-based techniques*".  
Descrizione della ricerca.  
La disponibilità di ampie collezioni di testi sintatticamente annotati (i.e. treebank) che consentono l'estrazione e la rappresentazione tramite strumenti statistici di dati linguistici, ha consentito lo sviluppo dei sistemi di NLP attualmente allo stato dell'arte, e.g. di parser (i.e. analizzatori sintattici) con prestazioni intorno al 90% di precisione ed accuratezza.  
Tuttavia, un'ampia varietà di dati sperimentali presenti in letteratura mostrano chiaramente la non riproducibilità di tali prestazioni su lingue diverse dall'inglese e su treebank contenenti dati con caratteristiche linguistiche ed annotazioni diverse da quelli contenuti nell'inglese Penn Treebank (<http://www.cis.upenn.edu/~treebank/home.html>), a tutt'oggi considerato lo standard per il testing di sistemi di NLP. Il progetto "*Human language as a complex system: study of variation phenomena with corpus-based techniques*" si propone di contribuire alla ricerca delle cause di tale non riproducibilità, sviluppando strumenti per il confronto di dati linguistici differenziati sia in termini e.g. di genere e stile, sia in termini di tipologia di annotazione, facendo riferimento alla risorsa treebank TUT di cui al punto seguente.
- Da dicembre 2002, è stata titolare di un **assegno di ricerca**, della durata di due anni, presso il Dipartimento di Informatica (Facoltà di Scienze Matematiche Fisiche e Naturali) dell'Università di Torino per il progetto "*Le risorse linguistiche per l'elaborazione del linguaggio naturale*".  
Descrizione della ricerca.  
Il progetto "*Le risorse linguistiche per l'elaborazione del linguaggio naturale*" ha contribuito all'ampliamento ed ingegnerizzazione della risorsa di tipo treebank per l'italiano denominata Turin University Treebank (TUT) (per ulteriori informazioni si veda <http://www.di.unito.it/~tutreeb/>), sul cui schema di annotazione e prime versioni prototipali si veda il punto seguente. In particolare, per far fronte al progressivo aumento delle dimensioni del treebank (che oggi include 2.000 frasi ovvero 59.000 token annotati), la ricerca affronta il problema della definizione di protocolli per il processo dell'annotazione e l'implementazione di programmi in linguaggio Perl per la verifica della correttezza dei dati annotati (e.g. assenza di alberi sintattici con cicli o nodi non connessi).  
Non solo il progetto contribuisce in modo cruciale alla ricerca nell'ambito del NLP per l'italiano supplendo alla mancanza di un treebank liberamente accessibile per la nostra lingua, ma, sviluppando un tipo di rappresentazione particolarmente ricca e calibrata sull'italiano, offre anche strumenti specifici per la rappresentazione dei fenomeni linguistici che ne sono tipici.
- Da aprile 1998, ha frequentato il **dottorato di ricerca in Informatica** (XIV ciclo) presso il

Dipartimento di Informatica dell'Università di Torino. Durante il dottorato ha lavorato nell'ambito del gruppo di NLP ed ha acquisito il titolo di dottore di ricerca nel 2004 discutendo la tesi di dottorato intitolata "*A grammatical relation system for treebank annotation*" (<http://www.di.unito.it/~bosco/phdthesis.zip>) (supervisore prof. V. Lombardo, revisori prof. J. Carroll (Department of Informatics - Sussex University, UK), e prof. M.T. Paziienza (Facoltà di Ingegneria - Università di Roma Tor Vergata).

Descrizione della tesi.

Partendo da un'analisi sia della linguistica teorica sia della linguistica dei corpora sia dei treebank esistenti, la tesi affronta il problema dell'annotazione dei dati linguistici orientata alla modellazione probabilistica del linguaggio e finalizzata ai task del NLP. Poiché questa analisi dimostra la crucialità delle informazioni relative alla struttura predicativo-argomentale (che ruota principalmente attorno al verbo) nei sistemi di NLP allo stato dell'arte, la tesi sviluppa una metodologia, denominata Augmented Relational Structure (ARS), atta ad annotare treebank rappresentando specificamente questo tipo di informazione e tenendo conto delle varie componenti in essa presenti (i.e. morfo-sintattica, sintattica-funzionale e semantica). La tesi si conclude offrendo strumenti per la valutazione del costrutto teorico ARS. Da un lato, la costruzione di uno schema di annotazione e l'applicazione di tale schema ad un corpus di lingua italiana (il cui risultato è il TUT, di cui al punto precedente), offre gli strumenti per una valutazione qualitativa della metodologia ARS. Dall'altro, l'implementazione di una versione prototipale di convertitore dallo schema di annotazione applicato in TUT ad un altro schema ben noto in letteratura (i.e. quello del Penn Treebank), offre invece gli strumenti per una valutazione quantitativa di ARS.

- Da dicembre 1996 a dicembre 1997, ha frequentato il **master in Tecnologia e comunicazione multimediale** del COREP (Politecnico di Torino) conseguendo il relativo diploma.

- Da marzo 1996 a marzo 1998, è stata titolare di una **borsa di studio post-laurea** del Dipartimento di Discipline Filosofiche (Facoltà di Lettere e Filosofia) dell'Università di Torino, sotto la supervisione della prof. C. Bazzanella (allora docente di Filosofia del Linguaggio), per lo svolgimento di attività di ricerca e di supporto alla didattica, della durata di un anno e poi rinnovata l'anno successivo.

Descrizione della ricerca.

Si sono affrontati dal punto di vista teorico e pratico i temi relativi ai corpora di dati linguistici ed agli strumenti per la loro costruzione, annotazione ed analisi in ambito sia linguistico sia computazionale. In particolare, facendo riferimento al progetto CHILd Language Data Exchange System (CHILDES, <http://poppy.psy.cmu.edu/childes.html>), si sono affrontati i problemi della trascrizione del linguaggio parlato ed annotazione del linguaggio scritto o parlato assistite dal computer per la costruzione di banche di dati linguistici.

Inoltre, sono state affrontate tematiche relative alla teoria e pratica della scrittura ipertestuale, online in linguaggio html oppure offline con gli strumenti offerti da Toolbook.

- Nell'ottobre 1995, ha conseguito la **laurea in Filosofia con indirizzo in analisi delle lingue naturali e informatica** con una tesi in Filosofia del Linguaggio, intitolata "*Un lessico ed un interprete semantico per un programma di comprensione di storie*", relatore prof. D. Marconi (Facoltà di Lettere e Filosofia, Torino) e controrelatore prof. L. Lesmo.

Descrizione della tesi.

La tesi discute ed implementa in linguaggio LISP un progetto di linguistica computazionale. In particolare partendo da una disamina della letteratura relativa all'analisi automatica del linguaggio naturale in termini di morfologia, sintassi e semantica, la tesi sviluppa ed implementa un interprete semantico per la lingua italiana

che prende come input una collezione di questionari raccolti nell'ambito di una ricerca sociologica.

- Nel 1995 ha frequentato il **corso di formazione professionale per Tecnici per la produzione di materiali didattici con supporti informatici** presso il COSP, della durata di 600 ore.
- Ha conseguito il **diploma di maturità classica** presso il Liceo Vittorio Alfieri di Torino.

## Lingue conosciute

Inglese, francese.

## Attività didattica

- A partire dell'a.a. 2006/2007 è titolare del corso di Informatica applicata per Scienze Biologiche presso la Facoltà di Scienze MFN dell'Università di Torino.
- Nell'a.a. 2004/2005, come titolare di contratto per l'insegnamento presso la Facoltà di Lingue e Letterature Straniere dell'Università di Torino, ha svolto la parte del **corso di Informatica applicata alla comunicazione multimediale** relativa al trattamento automatico del linguaggio naturale ed alla linguistica dei corpora, equivalente a 30 ore di attività in aula per i corsi di laurea biennali specialistici.
- Nell'a.a. 2004/2005, come titolare di contratto per l'insegnamento presso la Facoltà di Lingue e Letterature Straniere dell'Università di Torino, ha svolto **laboratori di Informatica**, equivalenti a 40 ore di attività in aula per i corsi di laurea triennali.
- Nell'estate 2003, come titolare di contratto di prestazione d'opera con la Facoltà di Scienze Matematiche Fisiche e Naturali dell'Università di Torino, ha collaborato con il Prof. Werbrouck e la Dott. Griffin alla **redazione di testi per l'ampliamento del Scientific English Test (SET)** per gli studenti di Informatica, nell'ambito del progetto per l'Innovazione Didattica, area Insegnamento delle Lingue.
- Nel gennaio 2002, ha svolto presso la Facoltà di Lingue e Letterature Straniere dell'Università di Torino, il **seminario Rappresentazione del linguaggio naturale e sviluppo di risorse linguistiche** (10 ore in aula + 20 laboratorio) rivolto agli studenti del corso di Didattica delle Lingue Moderne.
- Nell'a.a. 2000/2001, come borsista art. 33, ha svolto due **corsi di Alfabetizzazione Informatica** presso la Facoltà di Lettere e Filosofia dell'Università di Torino, ed un **laboratorio di HTML** presso la Facoltà di Scienze della Comunicazione (sede di Ivrea), equivalenti a 100 ore di attività in aula.
- A febbraio 2001, ha svolto il **corso Testo ed Ipertesto** della durata di 10 ore nell'ambito del master per Autori di Opere Multimediali presso il Virtual Reality & MultiMedia Park di Torino.

- Negli ultimi anni ha tenuto, dietro richiesta dei docenti, **lezioni sulle risorse linguistiche e il trattamento automatico del linguaggio naturale** sia, nell'ambito del corso di Intelligenza Artificiale e Interazione Uomo Macchina (prof. L. Lesmo), per gli studenti del corso di laurea in Informatica, sia, nell'ambito del corso di Informatica Generale (dott. D. Cavagnino), per gli studenti della Facoltà di lingue e Letterature Straniere.
- Durante i due anni della borsa di studio post-laurea, ha svolto attività seminariali per i laureandi del Dipartimento di Discipline Filosofiche, in particolare un **seminario su progettazione e ambienti di sviluppo ipertestuale** ed uno sulla **trascrizione ed analisi informatizzata del linguaggio naturale**.

## Altre esperienze e collaborazioni

### · All'estero

- Da ottobre a dicembre 2001, è stata **ospite, come Honorary Research Fellow, del Department of Phonetics and Linguistics della University College London (London, UK)** sotto la supervisione del prof. R. Hudson, per approfondire lo studio dei formalismi sintattici a dipendenze.
- Da luglio a novembre 1997, come **stagista** (stage relativo al Master di cui sopra) **presso il Centre d'Etudes franco-italiennes (CEFI-CNRS) dell'Université de Savoie (Chambéry, Francia)**, ha partecipato alla realizzazione del sito web del Centro, all'adattamento al formato online di testi sulla cultura franco-italiana, ed ha svolto attività di ricerca sull'evoluzione della comunicazione interattiva in Internet e soprattutto sugli utilizzi di questa a fini didattici.

### · In Italia

- Nel 2005 è stata nominata **cultrice della materia presso la Facoltà di Lingue e Letterature Straniere dell'Università di Torino**.
- **Collabora** dal 1996 con la prof. C. Bazzanella (ora docente di Linguistica Generale presso il Dipartimento di Filosofia, Torino) **in attività di ricerca sui temi attinenti ai corpora di dati linguistici**, l'annotazione del contesto anche in dati multimediali, e lo studio di fenomeni specifici dell'italiano tramite dati estratti da corpora di dati linguistici.
- Da gennaio 1998 ad aprile 1999, è stata dipendente di una società informatica di Torino dove ha sviluppato in Smalltalk un'interfaccia per l'interrogazione di un database distribuito.
- Da gennaio 1997 a ottobre 2003, ha collaborato alla realizzazione della rivista online Servizio Web Italiano per la Filosofia (SWIF) del Laboratorio di Epistemologia Informatica dell'Università di Bari, occupandosi in particolare del *Quaderno di Filosofia del Linguaggio*.
- Da giugno a dicembre 1995, è stata stagista (stage relativo al corso di formazione professionale di cui sopra) presso una cooperativa di Torino attiva nel settore informatico, collaborando alla progettazione e sviluppo del dimostrativo di un

ipertesto per un museo d'arte.

Torino, 20 marzo 2006