

Burr, Elisabeth (2003) "Informatica umanistica e Linguistica dell'italiano - Quo vadis?", in: Fiorimonte, Domenico (ed.): *Informatica Umanistica. Dalla ricerca all'insegnamento. Atti dei convegni Computers, Literature and Philology*, Università di Roma "La Sapienza" (1999) e Università di Alicante, Spagna (2000). Roma: Bulzoni 41-64.

INFORMATICA UMANISTICA E LINGUISTICA DELL'ITALIANO - QUO VADIS?

1. Introduzione

Benché l'importanza della rivoluzione di personal computer e Internet sia ormai sotto gli occhi di tutte e tutti, ancor oggi, nelle discipline umanistiche in generale e nella romanistica in particolare, le possibilità offerte dal computer o da Internet vengono sfruttate dalle e dai docenti per la ricerca ma non per l'insegnamento. Ciò risulta da una ricerca che ho svolto come membro del direttivo di *Textual scholarship and humanities computing*, una dei working groups che compongono la rete tematica *Advanced Computing in the Humanities* promossa dalla Comunità Europea nell'ambito del programma SOCRATES.¹ Il quadro che risulta da questo studio approfondito dei siti Internet delle facoltà di Lingue e Letterature in Germania ed altrove in Europa potrebbe, infatti, essere caratterizzato nel modo seguente:

- Internet viene usato soprattutto per diffondere informazioni su docenti, ore di ricevimento, piani di studio, programmi semestrali, contenuti dei corsi e simili;
- all'interno delle facoltà o dei dipartimenti aumenta il numero dei corsi che nel loro titolo fanno riferimento all'informatica umanistica ma che in verità si occupano esclusivamente del lato tecnico dell'uso del computer, ossia della cosiddetta *computer literacy*;
- tentativi di integrare una metodologia informatica o Internet negli insegnamenti umanistici tradizionali sono in genere molto rari.

Secondo me, le cause di questa situazione non sono tanto da cercare in un'eventuale carenza di tecnologia nelle varie facoltà, ma piuttosto nella mancanza di personale docente competente, ossia con esperienze di ricerche condotte con l'aiuto del computer o sulla base dell'informazione presente in Internet, come pure nella mancanza di supporti elettronici, e, quello che forse è più importante, in una determinata concezione dell'apprendimento.

L'apprendimento viene, infatti, inteso piuttosto come acquisizione di teorie e non come conseguimento di un intendimento basato sulla propria esperienza o, detto con le parole di Richard Saul Wurman (1999): "Our systems of learning are not based on systems of

¹ Informazioni su questo progetto si possono trovare su: <http://www.hd.uib.no/AcoHum/aco-hum.html>. I risultati ottenuti durante i tre anni della sua durata sono stati pubblicati in Smedt / Gardiner / Ore / Orlandi / Short / Souillot / Vaughan (1999).

understanding. Instead, it is [sic] based on a system of memorization." Ciò spiega, per esempio, il fatto che in corsi diretti a future e futuri insegnanti si teorizzi sulle eventuali possibilità educative e conseguenze dell'uso del computer o di Internet senza mai accendere un computer. L'oggetto trattato stesso è, dunque, del tutto assente.

2. Insegnamento della linguistica

Questa situazione si rispecchia naturalmente anche nell'insegnamento della linguistica delle varie lingue. Fa eccezione soltanto l'insegnamento della linguistica dell'inglese per i tanti supporti che sono stati creati per questa lingua ed il rispetto di cui gode, ormai anche all'interno dell'Anglistica tedesca, la linguistica dei corpora ed il filone empirico della linguistica dell'inglese in generale. L'insegnamento della linguistica delle lingue romanze in generale e della linguistica dell'italiano in particolare, invece, soffre fortemente delle mancanze alle quali ho accennato prima.

2.1 L'insegnamento tradizionale

Infatti, sebbene la linguistica come scienza umanistica si proponga di descrivere, capire e interpretare la lingua come oggetto culturale, prodotto cioè, come dice Tito Orlandi, "dalla mente umana, e, in parte, dalla volontà umana di comunicare con altri esseri umani, o comunque di agire su altri esseri umani o sul mondo del reale" (Orlandi 1997), questo oggetto non viene veramente studiato nei corsi tradizionali. Invece di guardare da vicino come funzionano il parlare naturale, scritto o orale, ossia l'unica manifestazione reale dell'oggetto culturale 'lingua', vengono discusse teorie o descrizioni della sua struttura o delle sue varietà ed i risultati di ricerche pubblicate. La lingua stessa è presente soltanto in veste di esempi isolati e persino molte volte inventati che vengono trattati nella letteratura considerata o riportati nelle grammatiche o nei dizionari. Anche quando viene richiesta l'applicazione di una teoria ai fatti della *parole* questi ultimi vengono prelevati manualmente da singoli testi. Metodi di ricerca empirica che, mettendo a profitto le possibilità offerte dalle nuove tecnologie, permetterebbero alle ed agli studenti di prendere sistematicamente coscienza delle regolarità inerenti al parlare e dei contenuti trasmessi non vengono insegnati e ciò malgrado il fatto che il piano di studio preveda, il più delle volte, una tesi di laurea basata su una ricerca empirica originale.

Al momento della tesi si scontrano dunque due realtà. Da un lato le laureande ed i laureandi di oggi non hanno soltanto frequentato università che garantiscono l'accesso alle nuove

tecnologie in generale e all'informazione digitale in particolare, ma la maggioranza di loro ha anche un computer a casa e lo usa, più o meno ogni giorno, sia per produrre testi in forma digitale (tesine, lettere, domande ecc.) sia per comunicare in forma digitale e per procurarsi informazioni da Internet; dall'altro, durante i loro studi le stesse persone hanno conosciuto soltanto un tipo di ricerca tradizionale, ossia l'analisi di manifestazioni della lingua archiviate su supporto cartaceo. Non c'è da stupirsi, dunque, se, come documenta l'esempio seguente, lo spreco di energie e di risorse è molto frequente:

Una studente ha scelto un tema nell'ambito dell'analisi del discorso. La ricerca empirica si doveva basare sui dialoghi di un determinato film. La studente ha trascritto meticolosamente tutti i dialoghi del film (chi ha intrapreso una trascrizione di questo tipo conosce bene il lavoro faticoso e noioso che ciò comporta). Per la trascrizione la studente ha usato ovviamente il computer. Non avendo, però, imparato una metodologia che le permettesse di creare un corpus codificato con questo materiale e di condurre un'analisi sistematica con l'aiuto del computer, la studente ha stampato le trascrizioni per poi svolgere l'analisi manualmente applicando le varie teorie studiate.

L'insegnamento tradizionale della linguistica non ha, però, soltanto implicazioni negative per gli stessi studi, ma comporta anche gravi conseguenze nel momento in cui laureate e laureati cercano lavoro. Fatto sta che studenti che concludono i loro studi senza aver avuto la possibilità di riflettere sul come sfruttare al meglio le possibilità offerte dalle nuove tecnologie per un determinato tipo di lavoro, già oggi come oggi non sono in grado, il più delle volte, di trovare un lavoro che permetta loro di usare le conoscenze specifiche degli oggetti culturali acquisite in modo proficuo e soddisfacente, e ciò nonostante che l'industria delle tecnologie dell'informazione cerchi disperatamente persone in possesso delle stesse conoscenze. Quest'apparente contraddizione si spiega abbastanza facilmente se teniamo conto del fatto che studenti che sanno sì usare un *word processor*, scrivere messaggi elettronici e navigare su Internet, ma non hanno conseguito una certa dimestichezza con l'applicazione di metodi informatici alla soluzione di un problema o con il trattamento automatico di contenuti, il più delle volte non vedono affatto in questo tipo di lavoro possibile prospettiva, oppure, se la vedono, non sono in grado, quando si presentano ad un posto di lavoro, di dimostrarsi fiduciose e fiduciosi della loro capacità di poter applicare le nuove tecnologie alle loro conoscenze.

Qui vorrei, invece, dimostrare in che modo si può insegnare la linguistica partendo dalle realizzazioni naturali di una lingua, sfruttando le possibilità che ci offrono il computer e l'avvento e la diffusione di Internet, e fornendo, allo stesso tempo, le qualifiche chiave richieste dalla società dell'informazione.

2.2 Un po' di storia personale

Il progetto che verrà presentato è nato dalle mie ricerche condotte con l'aiuto del computer e si basa soprattutto sui corpora elettronici da me creati proprio al fine di condurre queste ricerche. La creazione stessa di questi corpora è il risultato di un'esperienza che ho fatto alla fine dei miei studi, al momento della tesi. Avendo conosciuto tante teorie e descrizioni anche contraddittorie sull'uso dei tempi verbali volevo vedere io stessa come il *passé simple* ed il *passé composé* da una parte e l'*indefinido* e il *perfecto compuesto* dall'altra venivano usati in vari giornali francesi e spagnoli e nelle loro suddivisioni interne (cf. Burr 1979). Ciò mi ha portata ad un lavoro massacrante. Ho contato manualmente tutte le occorrenze dei due tempi verbali nei giornali prescelti e quando mi sono resa conto che per poter applicare metodi di statistica avevo bisogno di conoscere anche le occorrenze delle altre forme verbali - nessuno mi aveva insegnato una metodologia da applicare ad una tale ricerca - ho dovuto ricominciare e contare anche le rimanenti forme. Se non fosse stato per la tesi non avrei continuato. Ciò che, però, più tardi mi è parso più grave era che il corpus analizzato non poteva essere usato anche per altre ricerche e che la ricerca stessa non poteva essere ripetuta e controllata da altre persone.

Quando nell'80 ho cominciato a riflettere su un eventuale dottorato di ricerca, i computer erano in agguato ed il loro uso mi è subito apparso come l'unico rimedio se volevo continuare con le ricerche sulla lingua naturale scritta o parlata; però costavano molto e avevo grandi timori che l'unica via verso un loro uso passasse attraverso l'apprendimento del programmare. Intanto mi sono messa alla ricerca di corpora già esistenti di giornali italiani che mi avrebbero permesso di raggiungere lo scopo che mi ero prefissa, ossia la delimitazione della variazione linguistica propria di un determinato registro di stili.² Siccome non ho trovato un corpus già esistente mi sono finalmente decisa a comprare un computer e a creare io stessa un corpus dei giornali italiani. Anche questo lavoro è stato massacrante. La grande differenza era, però, che una volta conclusa la ricerca (cf. Burr 1993b, 1998a u. 1998b), ci sarebbe stato un corpus (cf. Burr 1993a, 1994, 1996) che poteva essere usato anche per altre ricerche e, appunto, per l'insegnamento (cf. per esempio Burr 1999).

2.3 I primi passi verso corsi di informatica umanistica

Il progetto di usare questo corpus di giornali italiani nell'insegnamento è cominciato già nel 1990 e ha attraversato, nel frattempo, varie fasi e vari cambiamenti di prospettiva determinati

sia dalle esperienze man mano acquisite sia dai mezzi tecnologici e metodologici a mia disposizione.³

2.3.1. La prima fase

Lo scopo dei corsi della prima fase era di indirizzare le e gli studenti verso l'analisi dell'uso naturale della lingua insegnando loro il lato tecnico e metodologico di una ricerca sistematica con l'aiuto del computer. In un secondo tempo i metodi imparati dovevano essere applicati ad un piccolo progetto di ricerca.

Nei corsi di questa fase sono stati usati diversi programmi di analisi testuale, scelti per ragioni diverse. Nel primo corso tenuto nel 1990 è stato usato *Micro-OCP* perché era il programma che io stessa conoscevo a fondo avendolo usato per la mia ricerca (cf. Burr 1993b). *Micro-OCP* risultava, però, troppo complesso per un corso della durata di pochi mesi e frequentato da più di 25 studenti, che prima del corso non avevano mai acceso un computer. I problemi incontrati, il lavoro faticosissimo ma necessario per risolverli in modo da poter portare il corso ad una felice conclusione e le resistenze alle quali andavo incontro quando cercavo di far acquistare un computer per il mio ufficio all'università sono fra le cause che hanno portato ad una lunga interruzione del progetto in questione.

Infatti, il progetto è stato ripreso soltanto nel 1995. Questa volta, invece di basare il corso su un unico programma di analisi testuale, sono stati usati tre programmi diversi, ossia *Logiciel St. Chef* (francese), *DBT 1.0* (italiano) e *TACT 2.1* (inglese), e ciò per due motivi. Da un lato si cercava di risolvere così almeno in parte l'anacronismo, divenuto evidente già nel primo corso, che studenti delle lingue romanze dovevano adoperare programmi con interfaccia inglese. Dall'altro si sperava che sfruttando lo specifico di ogni singolo programma e portando in questo modo le e gli studenti da un elenco di parole e di frequenze alle concordanze e poi ad un'analisi variata ed interattiva, si poteva far capire non soltanto lo stretto legame esistente fra il tipo di ristrutturazione del materiale linguistico e le conoscenze alle quali dava accesso ma anche l'interdipendenza fra il programma e la codifica del corpus.

Volendo fare una valutazione critica di questa fase si può concludere che i corsi erano, in generale, rivolti piuttosto verso l'aspetto tecnico e troppo poco verso quello concettuale. Soffrivano, inoltre, della loro posizione precaria all'interno del programma dei corsi di laurea.

² Per una definizione di „registro di stili“ vedi Coseriu (1981/1988: 25).

³ Cf. <http://www.uni-duisburg.de/FB3/ROMANISTIK/PERSONAL/Burr/humcomp/home.htm>.

I corsi avevano, infatti, lo status di esercitazioni supplementari⁴ che richiedevano soltanto una frequenza regolare ma non un esame finale. Era molto difficile, dunque, convincere le ed i studenti a portare avanti un progetto di ricerca che le avrebbe permesso di fare esperienze con l'applicazione dei metodi insegnati e di prendere coscienza di che cosa significa un corpus, di che cos'è un'analisi computerizzata e perché una tale analisi dovrebbe essere interessante.

2.3.2. *La seconda fase*

La seconda fase del progetto si può caratterizzare come il tentativo d'inserire l'informatica umanistica fra le materie d'esame del corso di laurea, tentativo secondo Aimée Morrison (1999) vitale se si vuole promuovere l'insegnamento dell'informatica umanistica. Si è cercato, dunque, d'agganciare il corso di metodologia informatica a seminari tradizionali di linguistica delle lingue romanze offerti, nello stesso semestre, da altre o altri docenti. Ossia, una programmazione comune dei temi da trattare nei seminari di linguistica avrebbe dovuto garantire che i metodi di analisi computerizzata insegnati nel corso di metodologia informatica potevano essere sfruttati direttamente per ricerche empiriche da intraprendere nell'ambito dei corsi di linguistica delle lingue romanze. Per entrambi i corsi era previsto un esame riconosciuto nel corso di laurea. L'esame doveva consistere in un'unica tesina da esaminare sotto due angoli diversi, ossia, mentre nell'ambito del corso di informatica umanistica il voto sarebbe stato attribuito esaminando l'impostazione della ricerca empirica, l'applicazione dei metodi informatici e l'esposizione della ricerca, nell'ambito del corso di linguistica si sarebbe esaminato lo svolgimento del tema stesso e la discussione dei risultati. L'unico programma usato in questa fase fu *TACT 2.1*.

Anche se il tentativo d'istituzionalizzazione di un corso d'informatica umanistica e il suo riconoscimento a fini di esame è fallito dovuto o alla mancata collaborazione fra docenti al momento della programmazione dei temi da trattare nei due corsi o alla loro mancata disponibilità di confrontarsi con le possibilità offerte da tecnologie non ancora di uso comune, ed anche se sussistevano i problemi dovuti all'interfaccia inglese di *TACT*, l'andamento dei corsi di questo tipo merita, in generale, un giudizio positivo. In fatti, i corsi a base di un programma interattivo come *TACT 2.1* portavano un buon numero di studenti a sviluppare un interesse per la ricerca computerizzata e di intuire le sue potenzialità. Anche se allora ciò sembrava poco visto i pochi progetti empirici conclusi a causa della posizione sempre periferica dei corsi, nel frattempo questa visione pessimista dei risultati raggiunti si è

⁴ Allora, l'unica via a disposizione quando si voleva introdurre un nuovo tipo di insegnamento nel mio

dimostrata infondata. Alcune studenti hanno, in fatti, saputo trarre profitto delle loro conoscenze allora acquisite e condurre una ricerca sistematica e approfondita per la tesi di laurea.

2.3.3. *La fase attuale*

La terza fase, cominciata nel 1998, rappresenta una vera e propria svolta per quanto riguarda sia la posizione dell'informatica umanistica nei programmi regolari dei corsi di laurea sia l'insegnamento. La svolta è stata resa possibile dall'avvento di *TACTWeb*, una software che collega *TACT* con Internet.⁵ Questa software ha, infatti, permesso di integrare l'analisi di corpora nei miei seminari tradizionali di linguistica e, siccome questi erano materia di esame, il problema del riconoscimento accademico, che per tanti anni aveva ostacolato lo sfruttamento delle nuove tecnologie e metodologie per l'insegnamento, da un momento all'altro non c'era più. Da allora, e sempre tramite i miei seminari di linguistica, l'informatica umanistica fa parte dei programmi regolari dei corsi di laurea in italiano e nelle altre lingue romanze. In più, questo tipo di seminario suscita ogni semestre più interesse e viene sempre più richiesto. Ed anche il problema dell'interfaccia non si pone più, permettendo Internet la traduzione di un'interfaccia inglese in qualsiasi lingua.

A parte questi aspetti più tecnici e formali l'integrazione di metodi informatici ha anche comportato una trasformazione dello scopo dei miei seminari. Infatti, questi non si prefiggono più soltanto di portare ad una conoscenza della teoria linguistica, del sistema e della norma della lingua e delle sue varietà oppure di insegnare una determinata metodologia da applicare in una ricerca sull'uso della lingua, ma vogliono invece mettere le e gli studenti in grado di porre domande ad un corpus, di svolgere un'analisi empirica sistematica, controllabile e ripetibile sulla "superficie linguistica dei testi" (Sabatini 1999: 145), di dedurre le regolarità dell'uso e di discutere, poi, in modo critico e fondato, sulla base delle cognizioni da loro personalmente acquisite sul funzionamento della lingua e dei risultati raggiunti, delle varie teorie o descrizioni e della loro fondatezza.

2.4 I corpora

I corpora computerizzati sui quali si basa maggiormente l'insegnamento sono costituiti, al momento attuale, dai seguenti giornali (nella tabella vengono elencati anche le singole edizioni ed il numero complessivo delle parole della rispettiva componente):

dipartimento era di offrire un corso di esercitazioni supplementari.

Corpus italiano - Unificazione tedesca 1989⁶

Testata	Edizioni	Numero parole
Corriere della Sera	19., 20., 21.10.1989	258.287
Il Mattino	20., 21.10.1989	171.501
La Repubblica	20., 21.10.1989	174.958
La Stampa	20., 21.10.1989	119.771
Totale		724.517

Corpus romanzo - Elezioni Europee 1994

Testata	Edizioni	Numero parole
Corriere della Sera	13., 14., 15.06.1994	303.641
Le Monde	12./13., 14., 15.06.1994	236.236
La Vanguardia	13., 14., 15.06.1994	261.133
Totale		801.010

Fanno parte del corpus soltanto edizioni complete dei vari giornali, cosicché ogni sub-corpus rispecchia il prodotto ‘giornale’ come è stato venduto alle lettrici ed ai lettori. L'unica cosa che distingue i giornali del corpus dalla loro versione stampata è la pubblicità, che per vari motivi, soprattutto tecnici, non è stata integrata nel corpus stesso.

Si tratta di corpora annotati. Ossia, è stato sviluppato un sistema di *markup* con cui il corpus è stato arricchito. Questo sistema è composto, come si vede di seguito, da una serie di variabili e dai loro rispettivi valori. Siccome vuole rispecchiare la complessità della struttura dei giornali, il sistema deve essere molto complesso. Infatti, deve tenere conto non soltanto delle varie testate ma anche delle loro parti tematiche, ossia dei vari settori, poi dei tipi di testo che compongono un giornale e, non da ultimo, della distinzione che si adopera nei vari testi fra testo corrente e testo scritto o parlato citato:

⁵ Per quanto riguarda in particolare *TACTWeb* vedi 2.5.

⁶ Il corpus può essere acquisito dall'*Oxford Text Archive* dietro la firma di un contratto (cf. Burr 1993a); una parte è stata pubblicata anche su CD-ROM (cf. Burr 1994).

unità	variabile	esempio
testata	<Z>	<Z Corriere della Sera>
edizione	<E>	<E 130694>
settore	<S>	<S Politica>
fonte del testo	<A>	
firmato		<A firmato>
anonimo		<A non firmato>
nome dell'autrice/autore	<N>	<N D' Errico Enzo >
pagina	<C>	<C 01>
lingua	<L>	<L Inglese>
tipo di testo	<T>	
occhiello		<T Occhiello>
titolo		<T Titolo>
sottotitolo		<T Sottotitolo>
sommario		<T Sommario>
catenaccio		<T Catenaccio>
civetta		<T Civetta>
articolo		<T Articolo>
spalla		<T Spalla>
fondo		<T Fondo>
corsivo		<T Corsivo>
notizia		<T Notizia>
messaggio breve		<T Flash>
notizia breve		<T Breve>
programma (TV, cinema)		<T Programma>
contenuto di film		<T Film>
intervista		<T Intervista>
rubrica		<T Rubrica>
critica		<T Critica>
lettera di lettrici/lettori		<T Lettera>
elenco		<T Elenco>
il tempo		<T Tempo>
titolo (libro, film, canzone ecc.)		<T Nome>
didascalia		<T Foto>
presentazione del parlare	<P>	
testo corrente		<P Prosa>
citazione di una fonte scritta		<P Citazione>
citazione di una fonte orale		<P Discorso>
domanda in un'intervista		<P Domanda>
risposta in un'intervista		<P Risposta>

Nell'esempio seguente viene riportato un estratto da un testo annotato con il sistema sopra evidenziato. Le codifiche appaiono in neretto:

<Z Corriere della Sera>
 <E 140694>
 <L Italiano>
 <C 19>
 <S Cronache Italiane>
 <A firmato>
 <N Buccini Goffredo>
 <T Occhiello>
 <P Prosa>
 <r nonid>
 DI PIETRO A PARIGI
 <T Titolo>
 A caccia de+i segreti di Gardini
 <T Sottotitolo>
 Interrogato Vernes socio de+l contadino
 <T Articolo>
 PARIGI - <L Francese> <P Discorso> "Monsieur Di Pietro? Non, je ne vous dis rien (<L Italiano> Non vi dico nulla)". <P Prosa> Gentile ma inflessibile, Jean Zamponi. Il capo del+l' ufficio istruzione parigino mette al+la porta i giornalisti. Il palazzo e' blindato perche' arriva Tonino Di Pietro, il giudice di <L Francese> <T Nome> "Main Propres" <T Articolo> <L Italiano> per un interrogatorio in cantiere da 8 mesi. Era l' ottobre '93 quando Di Pietro avvio' la rogatoria per sentire Jean Marc Vernes, banchiere amico de+l premier neogollista Eduard Balladur, socio fedele di Gardini e suo probabile compagno di scalate. In ballo c' e' una domanda: Vernes acquisto' titoli Enimont per conto di Gardini, con i soldi del+la Montedison? Si presto', cio+e', a un' operazione di <L Francese> "portage" <L Italiano> che doveva contribuire al+l' assalto de+l suo amico di Ravenna sul+le azioni del+la <L Inglese> joint venture? <L Italiano> E una questione centrale, a un mese e mezzo da+l processo Cusani e a poche settimane dal+l' inizio de+l processone Enimont.

Una codifica di questo tipo è essenziale se si intendono condurre studi sulla variazione linguistica presente all'interno dei giornali o all'interno del corpus in generale, sulla lingua usata in un determinato settore o in un determinato tipo di testo, sul 'parlato' che ci propongono i giornali o su ciò che differenzia *la vera oralità* dalla *finta* (cf. Dardano 2000) ossia da quella che Dardano chiama anche "parlato stilizzato" (Dardano 1987: 64).

2.5 TACTWeb

Come detto sopra (cf. 2.3.3), *TACTWeb* è alla base della svolta che ha portato l'informatica umanistica a diventare parte integrante dei miei seminari tradizionali di linguistica dell'italiano e delle altre lingue romanze. L'integrazione stessa è stata possibile perché con *TACTWeb* le potenzialità di Internet potevano essere sfruttate per l'insegnamento, permettendo questo programma l'analisi di corpora attraverso la rete. Una parte dei corpora sopra elencati (cf. 2.4) è stata, dunque, trasformata in base di dati con l'aiuto dell'apposito

modulo di *TACT* 2.1 e resa accessibile da una pagina web.⁷ Attualmente con *TACTWeb* possono essere analizzati i corpora seguenti:

<u>Testata</u>	<u>Edizione</u>	<u>Numero parole</u>
Le Monde	15.06.1994	75.514
Corriere della Sera	15.06.1994	98.224
La Vanguardia	15.06.1994	80.414
Corriere della Sera	21.10.1989	82.787
Il Mattino	21.10.1989	80.091

Come *TACT*, il programma usato nei corsi della seconda fase (cf. 2.3.2), *TACTWeb* è un programma interattivo e permette di formulare ricerche molto complesse. La sintassi da rispettare quando si formulano queste ricerche è la stessa di *TACT* 2.1. Le ricerche vengono inserite in una maschera d'interrogazione che viene proposta per ogni base di dati.

Poter accedere a queste basi di dati attraverso la rete comporta un enorme vantaggio per l'insegnamento. Prima di tutto, l'esito di un'interrogazione non è più dipendente dal fatto che i singoli computer siano stati configurati nel modo richiesto da *TACT* ma unicamente dalle realizzazioni linguistiche contenute nella base di dati e dalle ricerche. Ciò significa che l'attenzione delle e degli studenti non viene più presa da problemi tecnici collegati alla macchina come tale e alla sua manipolazione ma si possono concentrare pienamente sui vari aspetti del loro tema, sull'impostazione della loro ricerca, sulla metodologia da applicare, sui metodi adatti ad un determinato tipo di lavoro e sull'interpretazione dei dati ricavati. In più, siccome la base di dati è facilmente accessibile ad ogni ora del giorno e da ogni computer collegato con la rete, le e gli studenti hanno anche più libertà nel gestire il loro lavoro, di tornare e rilevare altri dati per punti rimasti incerti e di seguire strade non ancora esplorate. Lo sfruttamento delle nuove tecnologie permette, dunque, di creare corsi dove le e gli studenti attraverso le loro proprie esperienze cominciano a capire che la linguistica è veramente una scienza rivolta dapprima alla cognizione dei fatti linguistici naturali e dei loro legami con la cultura e la società e poi alla sistemazione delle conoscenze acquisite con l'aiuto di teorie interpretative.

2.6 Un esempio

Per dimostrare le potenzialità inerenti a corsi di questo tipo e l'interesse investigativo che questi possono suscitare persino presso studenti che non hanno ancora acquisito una grande

⁷ Cf. <http://www.uni-duisburg.de/FB3/ROMANISTIK/PERSONAL/Burr/humcomp/analysis.htm>.

dimestichezza con la ricerca, vorrei presentare qui uno dei progetti elaborati durante un seminario sulle varietà della lingua dei giornali italiani che è stato offerto a studenti del primo biennio del corso di laurea in italiano. A parte gli interessanti risultati raggiunti, il progetto in questione è stato scelto in modo particolare perché dimostra anche che i corpora già in rete non devono, per forza, rimanere le uniche basi di dati da esplorare nell'ambito di un determinato corso, ma possono persino funzionare come modelli per la creazione di nuove basi di dati.

A causa della guerra in Kosovo allora in atto, uno dei temi da trattare in questo corso erano i rapporti sulla guerra. Siccome i corpora da me creati negli anni ottanta e novanta non contenevano, ovviamente, materiale adatto, si pensava all'inizio di trattarlo unicamente sulla base delle ricerche già pubblicate. Una volta, però, che le e gli studenti sperimentando con l'analisi dei corpora in rete avevano capito le possibilità offerte dai metodi informatici per la ricerca, è sorta l'idea di creare un apposito corpus per poter controllare se la lingua usata nei rapporti sulla guerra in Kosovo corrispondesse a quella usata, secondo le ricerche studiate, in occasione di altre guerre. Alla fine, uno studente si è lasciato convincere a creare un piccolo corpus scaricando articoli sulla guerra in Kosovo da giornali italiani in linea e arricchendoli con un *markup* elaborato sulla base del sistema sopra delineato (cf. 2.4). Il piccolo corpus di 3.513 parole è stato poi trasformato in una base di dati e messo in rete perché potesse essere analizzato anche da altre persone (cf. Rizzo 1999).

Riferisco qui di seguito alcuni dei risultati più importanti di una delle analisi condotte, nell'ambito dello stesso seminario, sulla base di questo corpus (cf. Gironi 1999), perché mi sembrano documentare che le varie possibilità offerte da programmi come *TACTWeb* non fruttano meramente dati quantitativi, ma che una ricerca sistematica sulla superficie dei testi può aprire anche una strada per accedere al contenuto e all'ideologia trasportata dai messaggi stessi.

La ricerca parte da undici parole scelte dalla studente perché giudicate parole chiave quando si parla di una guerra. Nella tabella seguente queste parole sono elencate assieme alle loro frequenze:

<u>parola chiave</u>	<u>frequenza</u>
guerra	7
pace	4
bomba (e derivati)	15
morti	2
forza/e	17
uomini	16
donna/e	1
truppe	13
vittoria	1
sconfitta	1
civili	2

Una delle tante possibilità offerta da *TACTWeb* per un'analisi approfondita dell'uso di queste parole è di farsi rilevare le occorrenze di ogn'una di loro in modo tale che appaiano al centro del loro contesto, ossia di richiedere che siano rappresentate in formato *KWIC*. Grazie al *markup* con il quale il corpus è stato arricchito è possibile di farsi mostrare, allo stesso tempo, anche in che giornale e/o tipo di testo le singole occorrenze vengono usate. *TACTWeb* permette, inoltre, di accedere al testo intero da ogni singola occorrenza oppure di stabilire la quantità di contesto (in linee) da rilevare assieme alle occorrenze di una parola chiave.

Sfruttando questi stessi metodi di rilevamento ed organizzazione dell'informazione nella sua ricerca sull'uso delle parole scelte ed analizzando le regolarità che dimostrano i loro contesti la studente è arrivata ai seguenti risultati:⁸

- L'argomento è la guerra, ma di guerra non si parla mai direttamente con le parole che la designano, quasi come se si tentasse di mascherare la verità, di renderla più accettabile:

guerra (7)

La Stampa/Titolo | | NORMAN MAILER E LA **GUERRA** | | Norman Mailer |
 La Repubblica/Articolo autentico **scenario di guerra:** | da | una parte
 La Repubblica/Articolo è in **assetto di guerra**, avrebbe risposto al
 IlSole24Ore/Articolo previsti | **prima della guerra:** lo hanno indicato oggi
 IlSole24Ore/Articolo **Il giorno 62 della | guerra** | del Kosovo ha visto
 Il Messaggero/Occhiello | | | Tra **guerra** e diplomazia/ Le
 Il Messaggero/Articolo **alla | fine della guerra.** E proprio per

- di guerra si parla in termini di numeri di uomini che la Nato spedisce in Kosovo, raramente in numeri di morti, sebbene la morte sia la prima delle conseguenze della guerra,

uomini (16)

La Repubblica di circa cinquantamila | **uomini** da mandare in Kosovo.
 La Repubblica avanguardia di quei 28.000 **uomini** | e donne che
 La Stampa BRUXELLES | Fino a cinquantamila **uomini** dislocati in Macedonia,
 La Stampa Wesley Clark. Cinquantamila **uomini** che, | dice il |
 La Stampa scopo". Ma anche cinquantamila | **uomini** che non escludono la
 La Stampa l'Alleanza prevedeva che 28 mila **uomini** | bastassero per questo

⁸ Riferisco di seguito alcune delle conclusioni alle quali arriva Francesca Gironi (1999) nella sua tesina.

La Stampa quasi il | doppio | degli **uomini**: alla Nato si parla di
 La Stampa ha già oggi circa 14 mila **uomini** che | fanno parte della
 La Stampa dichiara le sue disponibilità di **uomini** e mezzi rispetto | alle
 La Stampa l'apporto più alto, con circa 9000 **uomini**, | mentre | gli Usa
 La Repubblica e | attaccato | con duecento **uomini** i villaggi albanesi di
 IlSole24Ore forza di interposizione a 45mila **uomini** | | Rallentati i raid
 IlSole24Ore jugoslavi | | | 45MILA **UOMINI** | PER LA FORZA | DI
 IlSole24Ore in dispiegamento di 45mila **uomini**, invece dei 28mila
 Il Messaggero alla Nato l'aumento a | 50.000 **uomini** della forza di pace
 Il Messaggero al minimo le | perdite | di **uomini** e mezzi Nato, Short ha

- di guerra si parla soprattutto in termini dei bombardamenti della Nato e raramente in termini di vittime civili:

bombarda (1)

La Repubblica | | | | Milosevic **bombarda** villaggi nel Kosovo.

bombardamenti (7)

La Repubblica | | | | Ancora una notte di **bombardamenti**. In giornata la
 La Stampa piedi una regione | martoriata dai **bombardamenti** e ridare strade
 La Stampa Nato insiste sul successo dei | **bombardamenti**. A | Bruxelles
 La Stampa le ragioni per sospendere i **bombardamenti** - dice - | ma
 La Stampa la sua strategia o sospenda i | **bombardamenti**". E l'idea di
 IlSole24Ore militari dopo 60 giorni di **bombardamenti** alleati e la
 Il Messaggero è vicina. Ancora due mesi di **bombardamenti** a tappeto |

bombardare (1)

Il Messaggero | Quanto alla decisione di **bombardare** ad alta quota per

bombardato (3)

La Repubblica per tutta la giornata - hanno **bombardato** e | attaccato |
 IlSole24Ore | | Gli aerei della Nato hanno **bombardato** una delle residenze
 IlSole24Ore velivoli | Nato | avevano già **bombardato** un'altra residenza

bombe (3)

La Repubblica morti | | MORINI - Piovono le **bombe** della Nato, continuano
 Il Messaggero Nato | | Ancora due mesi di **bombe** | e Milosevic
 Il Messaggero le unità colpite della **bombe** dell'Alleanza.

- *pace* viene usato soltanto in relazione all'esercito della Nato:

pace (4)

La Stampa "Kfor plus", la **forza di pace** | potenziata chiesta dal
 La Stampa utilizzati per | una **missione di pace** e per nessuna altro
 IlSole24Ore futura **forza internazionale di pace** per il | Kosovo, | la
 Il Messaggero | 50.000 uomini della **forza di pace** destinata ad entrare in

- *forza/forze* non viene mai usato per indicare gli eserciti di Milosevic:

forza (14)

La Repubblica/Articolo il momento - di una **forza** da impiegare per |
 La Repubblica/Articolo inviare da subito la **forza** ai | confini del Kosovo,
 La Repubblica/Articolo | partecipazione alla **forza** anche di militari di
 La Repubblica/Articolo dell'eventuale | **forza** di interposizione in
 La Stampa/Articolo "Kfor plus", la **forza** di pace | potenziata
 La Stampa/Articolo si riconosce che la **forza** avrà un | compito ben
 La Stampa/Articolo | finora. | Sarà una **forza** - spiega il portavoce
 IlSole24Ore/Titolo 1999) | | | | | La **forza** di interposizione a
 IlSole24Ore/Catenaccio UOMINI | PER LA **FORZA** | DI INTERPOSIZIONE |
 IlSole24Ore/Articolo | I nuovi piani per la **forza** militare di
 IlSole24Ore/Articolo di profughi. La **forza** di interposizione
 IlSole24Ore/Articolo esatta della **forza** e i contributi dei vari
 IlSole24Ore/Articolo della futura **forza** internazionale di pace
 Il Messaggero/Articolo | 50.000 uomini della **forza** di pace destinata ad

forze (3)

La Repubblica/Articolo "conferenza sulle | **forze**" che dovrà stendere un
 La Stampa/Articolo Usa che comanda le **forze** dell'Alleanza in Europa
 La Stampa/Articolo di generazione delle **forze**", cioè l'operazione | in

- la presenza delle donne nelle truppe della Nato viene a malapena presa in considerazione come se fosse una svista il fatto che anche le donne vivono la guerra

donne (1)

La Repubblica

di quei 28.000 uomini | e **donne** che dovrebbero

Con l'aiuto della sua breve ricerca questa studente non ha, dunque, soltanto fatto l'esperienza di un rilevamento sistematico di dati linguistici da un corpus di manifestazioni naturali della lingua italiana e delimitato così l'uso di certe parole; questo rilevamento sistematico di dati dalla superficie dei testi assieme con il tipo di organizzazione dei dati che permette un programma come *TACTWeb* l'ha anche portata a prendere coscienza del fatto che l'uso della lingua non è neutrale. Invece, come dimostrano i risultati, nei rapporti sulla guerra in Kosovo si nascondono giudizi e prese di posizione che non sono evidenti quando si guardano le occorrenze di una determinata parola individualmente, ma si dischiudono soltanto quando si riescono a vedere tutte le occorrenze insieme. Solo attraverso una ricerca computerizzata, cioè, diventa evidente che la sistematicità con la quale appaiono questi giudizi e prese di posizione è l'espressione di una certa visione dei fatti; comune, fra l'altro, come si vede dalle concordanze sopra riportate, ai vari giornali.

3. Conclusione

Penso di aver dimostrato che un insegnamento che sfrutta le possibilità offerte dalle nuove tecnologie integrando l'informatica umanistica nei seminari tradizionali di linguistica dell'italiano e delle altre lingue romanze mette le e gli studenti in grado di acquisire una coscienza sistematica dei dati linguistici naturali e del fatto che essi trasportano contenuti legati alla cultura e alla società e che queste conoscenze le/li aiutano a capire lo stesso oggetto culturale 'lingua' procurando anche un'eccellente base sulla quale occuparsi in modo critico delle teorie e dei risultati di ricerche esistenti.

Questa è, però, soltanto una delle ragioni per le quali mi sembra importante integrare l'informatica umanistica nei corsi tradizionali. Un'altra ragione è che anche l'insegnamento della linguistica delle varie lingue deve tenere in considerazione che stiamo vivendo una rivoluzione dell'informazione causata dallo sviluppo dei media digitali in generale e di Internet in particolare. Non per niente questa rivoluzione ricorda la rivoluzione in atto nel rinascimento a causa prima dell'arrivo della carta e poi dell'invenzione della stampa e la sua rapida diffusione. Anche se la rivoluzione attuale, così come quella di allora, non porterà certo ad una sostituzione totale dei media tradizionali con quelli nuovi, Internet si assumerà,

nondimeno, così come ha fatto il libro nel passato, molte delle funzioni svolte fin'ora dai media tradizionali, integrando, come già avviene, i vari metodi di trasportare l'informazione, e quindi i contenuti. Metterà a disposizione, inoltre, e lo sta già facendo, tutt'un altro tipo di accesso all'informazione e alla comunicazione.

Ciò significa che le e gli studenti oggi non devono più soltanto imparare come sfruttare al meglio ed in modo responsabile, critico e cosciente, l'informazione trasportata dai media tradizionali come il libro, la stampa, il film ecc., e come comunicarla, per esempio, su supporto cartaceo o attraverso le onde sonore. Invece, devono imparare anche come trattare l'informazione trasportata dai media digitali e come comunicarla su supporto digitale. Siccome la linguistica è la scienza della lingua, “veicolo privilegiato per la produzione, la gestione, la trasmissione, l'accesso all'informazione” (Zampolli 2000: 25) spetta per primo alle ed ai docenti che se ne occupano rendere i supporti informatici e l'informazione ivi trasportata così come i metodi per il suo trattamento parte integrante del loro insegnamento e contribuire in questo modo anche alla formazione delle qualifiche chiave di cui le ed i discenti avranno bisogno per poter intraprendere non soltanto le attività tradizionalmente aperte a studenti delle discipline umanistiche ma anche qualcuna delle tante attività professionali che la rivoluzione tecnologica ha già creato nel campo dell'informazione e della comunicazione e nelle industrie della lingua e della cultura.

È chiaro, però, che anche se l'integrazione dell'informatica umanistica nei seminari tradizionali di linguistica dell'italiano e delle altre lingue romanze ed il suo riconoscimento a fini di esame, visto le tante resistenze che dovevano essere vinte, è un grande passo in avanti, non basterà, al lungo andare, offrire, per esempio, ogni due semestri un tale corso. Già ora con 12-15 studenti l'affluenza a questi corsi è troppo forte per poter dare ad ogni studente l'attenzione di cui avrebbe bisogno soprattutto quando fa l'esperienza di una ricerca computerizzata per la prima volta. In più, c'è anche il pericolo che un corso di questo tipo o diventi troppo complesso a causa dei tanti concetti nuovi che deve introdurre o non vada oltre un trattamento superficiale sia della linguistica sia dell'informatica umanistica. Mi sembra, dunque, che Aimée Morrison (1999) abbia ragione quando vede l'integrazione di metodi informatici in seminari tradizionali delle materie umanistiche come una soluzione soltanto intermedia, volta sia ad incrementare la visibilità e la credibilità delle persone che se ne occupano all'interno dell'università sia ad attirare l'interesse delle e dei studenti, creando così, man mano, una domanda per corsi di laurea in informatica umanistica.

Bibliografia

- Bradley, John / Lancashire, Ian / Presutti, Lidio / Stairs, Michael (1996): *TACT 2.1*. Toronto: University of Toronto, Centre for Computing in the Humanities.
- Bradley, John / Rockwell, Geoffrey (1997): *TACTWeb 1.0*. 30.08.2000 <<http://tactweb.humanities.mcmaster.ca/tactweb/doc/tact.htm>>.
- Burr, Elisabeth (1979): *Präteritum und Perfekt im Französischen und Spanischen*. Experimentelle-statistische Untersuchung der Zeitungssprache. (Tesi di Laurea non pubblicata).
- Burr, Elisabeth (1993a): *Corpus of Italian newspapers*. Oxford: Oxford Text Archive.
- Burr, Elisabeth (1993b): *Verb und Varietät*. Ein Beitrag zur Bestimmung der sprachlichen Variation am Beispiel der italienischen Zeitungssprache (= Romanische Texte und Studien 5). Hildesheim: Olms.
- Burr, Elisabeth (1994): "Italian Newspaper Corpus (ita03)", in: Association for Computational Linguistics: *European Corpus Initiative Multilingual Corpus 1* (ECI / MCI) CD-ROM: \data\eci1\
- Burr, Elisabeth (1996): "A Computer Corpus of Italian Newspaper Language", in: *Research in Humanities Computing 4*. Selected Papers from the ALLC / ACH Conference, Christ Church, Oxford, April 1992. Oxford: Clarendon Press 216-239.
- Burr, Elisabeth (1998a): "«Italiano dell'uso medio» e «lingua dei giornali»", in: Navarro Salazar, Teresa (ed.): *Italica Matritensia*. Atti del IV Convegno SILFI, Società Internazionale di Linguistica e Filologia Italiana (Madrid, 27-29 giugno 1996). Firenze: Franco Cesati 115-133.
- Burr, Elisabeth (1998b): "«Questione della lingua», «lingua media» und Zeitungssprache", in: *Zeitschrift für Romanische Philologie* 114, 2: 269-289.
- Burr, Elisabeth (1999): *Teaching Romance Linguistics with On-line French, Italian and Spanish Corpora* (= LAUD 1998: 283, Series B: Applied and Interdisciplinary Papers).
- Burr, Elisabeth (1990-2000): *Romanische Linguistik & neue Medien*. 02.09.2000 <http://www.uni-duisburg.de/FB3/ROMANISTIK/PERSONAL/Burr/humcomp/home.htm>.
- Coseriu, Eugenio (1981/1988): "Die Begriffe "Dialekt", "Niveau" und "Sprachstil" und der eigentliche Sinn der Dialektologie", in: Albrecht, Jörn / Lüdtke, Jens / Thun, Harald (eds.): *Energeia und Ergon. Sprachliche Variation - Sprachgeschichte - Sprachtypologie*. Studia in honorem Eugenio Coseriu I: Albrecht, Jörn (ed.): *Schriften von Eugenio Coseriu (1965-1987)* (= TBL 300). Tübingen: Narr 15-43.
- Dardano, Maurizio (1987): "Il linguaggio dei giornali", in: Jacobelli, Jader (ed.): *Dove va la lingua italiana?* Roma/Bari: Laterza 58-65.
- Dardano, Maurizio / Frenguelli, Gianluca / Laut, Gianluca (2000): "Parlato e finzione di parlato nella stampa di oggi", in: Comitato scientifico della SILFI (ed.): *«Tradizione & Innovazione». La linguistica e filologia italiana alle soglie di un nuovo millennio*. Reader VI Convegno Internazionale della Società Internazionale di Linguistica e Filologia Italiana, Gerhard-Mercator-Universität Duisburg, 28.06-02.07.2000. Duisburg: Universität Duisburg 94-95.
- Gironi, Francesca (1999) „Die Sprache der Kriegsberichterstattung“, tesina, semestre estivo, Gerhard-Mercator-Universität Duisburg.
- Morrison, Aimée (1999): "Teaching Humanities Computing... A Graduate Student Perspective". 12.07.2000 <<http://www.humanities.ualberta.ca/amorrison/otherwork/cosh-coch99.htm>>.

- Orlandi, Tito (1997): "Informatica, formalizzazione e discipline umanistiche". 05.07.2000 <<http://rmcisadu.let.uniroma1.it/~orlandi/formaliz.html>>.
- Rizzo, Angelo (1999): Corpus „Berichterstattung über den Krieg im Kosovo“, semestre estivo, Gerhard-Mercator-Universität Duisburg. 31.08.2000 <<http://www.uni-duisburg.de/FB3/ROMANISTIK/PERSONAL/Burr/humcomp/corpusit.htm>>.
- Sabatini, Francesco (1999): "‘Rigidità-esplicitezza’ vs ‘elasticità-implicitezza’: possibili parametri massimi per una tipologia dei testi", in: Skytte, Gunver / Sabatini, Francesco (eds.): *Linguistica Testuale Comparativa*. In memoriam Maria-Elisabeth Conte. Atti del Convegno interannuale della Società di Linguistica Italiana, Copenaghen 5-7- febbraio 1998. Copenaghen: Museum Tusculanum Press 141-172.
- Smedt, Koenraad de / Gardiner, Hazel / Ore, Espen / Orlandi, Tito / Short, Harold / Souillot, Jacques / Vaughan, William (eds.) (1999): *Computing in Humanities Education. A European Perspective*. SOCRATES/ERASMUS thematic network project on Advanced Computing in the Humanities. Bergen: University of Bergen.
- Wurman, Richard Saul (1999): "A Conversation With Richard Saul Wurman", in: Fish, Tatjana (ed.): *Outlook: Information and the Humanities in the Next Millennium*, FoxNews: September. 05.07.2000 <http://www.foxnews.com/science/millennium/september/interview_wurman.sml>.
- Zampolli, Antonio (2000): "Risorse linguistiche per il trattamento dell'italiano scritto" in: Comitato scientifico della SILFI (ed.): «*Tradizione & Innovazione*». *La linguistica e filologia italiana alle soglie di un nuovo millennio*. Reader VI Convegno Internazionale della Società Internazionale di Linguistica e Filologia Italiana, Gerhard-Mercator-Universität Duisburg, 28.06-02.07.2000. Duisburg: Universität Duisburg 25-27.