

Elisabeth Burr

INSEGNARE LINGUISTICA CON UN CORPUS DI GIORNALI

1. Introduzione

Benché l'importanza della rivoluzione di personal computer e Internet¹ sia ormai sotto gli occhi di tutte e tutti, ancor oggi, nelle discipline umanistiche in generale e nella romanistica in particolare, tentativi di integrare una metodologia informatica o Internet negli insegnamenti umanistici tradizionali sono in genere molto rari. Ciò risulta da una ricerca che ho svolto all'interno della rete tematica *Advanced Computing in the Humanities* promossa dalla Comunità Europea.²

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 con esperienze di ricerche condotte con l'aiuto del computer o sulla base dell'informazione presente in Internet, nella mancanza di supporti elettronici e in una determinata concezione dell'apprendimento. L'apprendimento viene, infatti, inteso piuttosto come acquisizione di teorie e non come conseguimento di un sapere basato sulla propria esperienza. Per esprimerlo con le parole di Richard Saul Wurman (1999): "Our systems of learning are not based on systems of understanding. Instead, it is [sic] based on a system of memorization."

Questa situazione si rispecchia naturalmente anche nell'insegnamento della linguistica dell'italiano. 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 (1997), "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", questo oggetto non viene veramente studiato nei corsi tradizionali. Invece di guardare da vicino come funzioni il

¹ Si tratta, secondo me, di una rivoluzione paragonabile a quella dell'invenzione dell'alfabeto nell'antichità e poi della stampa Gutenberghiana nel rinascimento. Questa rivoluzione creerà a lungo andare, come ci dicono le teorie dei media ed anche Allen Renear (1999), nuove forme di organizzazione sociale e politica, nuove pratiche sociali e scientifiche, nuove forme di cultura e d'arte e comporterà, come hanno fatto le rivoluzioni precedenti, che diventino dominanti altre forme del pensare e dell'esperienza.

² 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).

parlare naturale, scritto o orale, ossia l'unica manifestazione reale dell'oggetto '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 nella veste degli esempi isolati e persino molte volte inventati che vengono trattati nella letteratura o sono riportati nelle grammatiche o nei dizionari. Se 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.

In questa sede 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.

2. Insegnamento a base di un corpus di giornali

Il tipo di insegnamento che vorrei presentare è nato dalle mie ricerche condotte con l'aiuto del computer e si basa soprattutto su un corpus elettronico da me creato proprio al fine di condurre queste ricerche. Iniziato nel 1990, il progetto di usare questo corpus nell'insegnamento 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.

La fase attuale è cominciata nel 1998 con l'avvento di *TACTWeb*, una software che ha reso possibile la consultazione del corpus in rete (vedi sotto 2.2). Questo fatto ha facilitato l'integrazione dell'analisi di corpora nei miei seminari di linguistica sincronica dell'italiano e con ciò l'introduzione di un altro tipo d'insegnamento. I seminari, infatti, non si prefiggono più soltanto di portare ad una conoscenza delle teorie, del sistema e della norma della lingua e delle sue varietà ma vogliono invece mettere le e gli studenti in grado di porre domande ad un corpus che rispecchia una varietà dell'italiano, di svolgere un'analisi empirica sistematica, controllabile e ripetibile sulla "superficie linguistica dei testi" (Sabatini 1999: 145), di dedurre dai dati rilevati le regolarità dell'uso e di discutere, poi, in modo critico e fondato, in base ai

risultati e delle cognizioni da loro personalmente acquisite, sul funzionamento della lingua e sulle varie teorie proposte al riguardo.

2.1. Il corpus

Il corpus computerizzato sul quale si basa questo tipo di seminari è un corpus di giornali italiani costituito nella sua forma originale dalle seguenti componenti (nella tabella vengono elencati anche le singole edizioni ed il numero complessivo delle parole della rispettiva componente):

Corpus italiano «Unificazione tedesca 1989»³

| <u>Testata</u> | <u>Edizioni</u> | <u>Numero parole</u> |
|---------------------|----------------------|----------------------|
| 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 |

Componente italiana del corpus romanzo «Elezioni Europee 1994»

| <u>Testata</u> | <u>Edizioni</u> | <u>Numero parole</u> |
|---------------------|----------------------|----------------------|
| Corriere della Sera | 13., 14., 15.06.1994 | 303.641 |

Fanno parte del corpus soltanto edizioni complete dei vari giornali; ogni sub-corpus rispecchia quindi il prodotto ‘giornale’ così 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 un corpus annotato. 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 stesso 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, delle diverse presentazioni al loro interno quali testo corrente e citazione di fonti scritte o parlate:

³ Il corpus può essere acquisito dall'*Oxford Text Archive* dietro la firma di un contratto (cf. Burr 1993); 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 joint venture? 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) cioè dal "parlato stilizzato" (Dardano 1987: 64).

2.2. TACTWeb

Come detto sopra (cf. 2), l'avvento di *TACTWeb* sta all'origine di questo tipo d'insegnamento. *TACTWeb* è la versione per la rete di *Tact 2.1 (Text-Analysis Computing Tools)*, ancora oggi considerato uno dei migliori programmi di analisi testuale.⁴

Per poter analizzare un corpus con *TACTWeb* questo deve prima essere trasformato in una banca dati con l'aiuto dell'apposito modulo di *Tact 2.1*. Questa banca dati viene poi messa su

⁴ Per una descrizione del programma ed una introduzione alle sue funzioni vedasi Lancashire (1996).

un server e resa accessibile tramite una pagina web.⁵ Attualmente con *TACTWeb* possono essere analizzati i seguenti componenti del corpus indicato (vedi 2.1):

| Testata | Edizione | Numero parole |
|---------------------|------------|---------------|
| Corriere della Sera | 15.06.1994 | 98.224 |
| Corriere della Sera | 21.10.1989 | 82.787 |
| Il Mattino | 21.10.1989 | 80.091 |

Il screen shot seguente dimostra la pagina Web dalla quale si accede ai corpora italiani:



fig. 1: Corpora dell'italiano accessibili tramite Internet per un'analisi con l'aiuto di *TACT Web*

Nella fig. 2 si vede la matrice che viene proposta in un unico momento per tutte le banche dati italiane messe in rete. La componente da analizzare si sceglie in una piccola finestra *drop down*. Le istruzioni per l'uso di *TactWeb* e le informazioni sulla sintassi da rispettare si possono consultare attivando l'aiuto (Hilfe zur Query Syntax):

⁵ Cf. <http://www.uni-duisburg.de/FB3/ROMANISTIK/PERSONAL/Burr/humcomp/analysis.htm>.

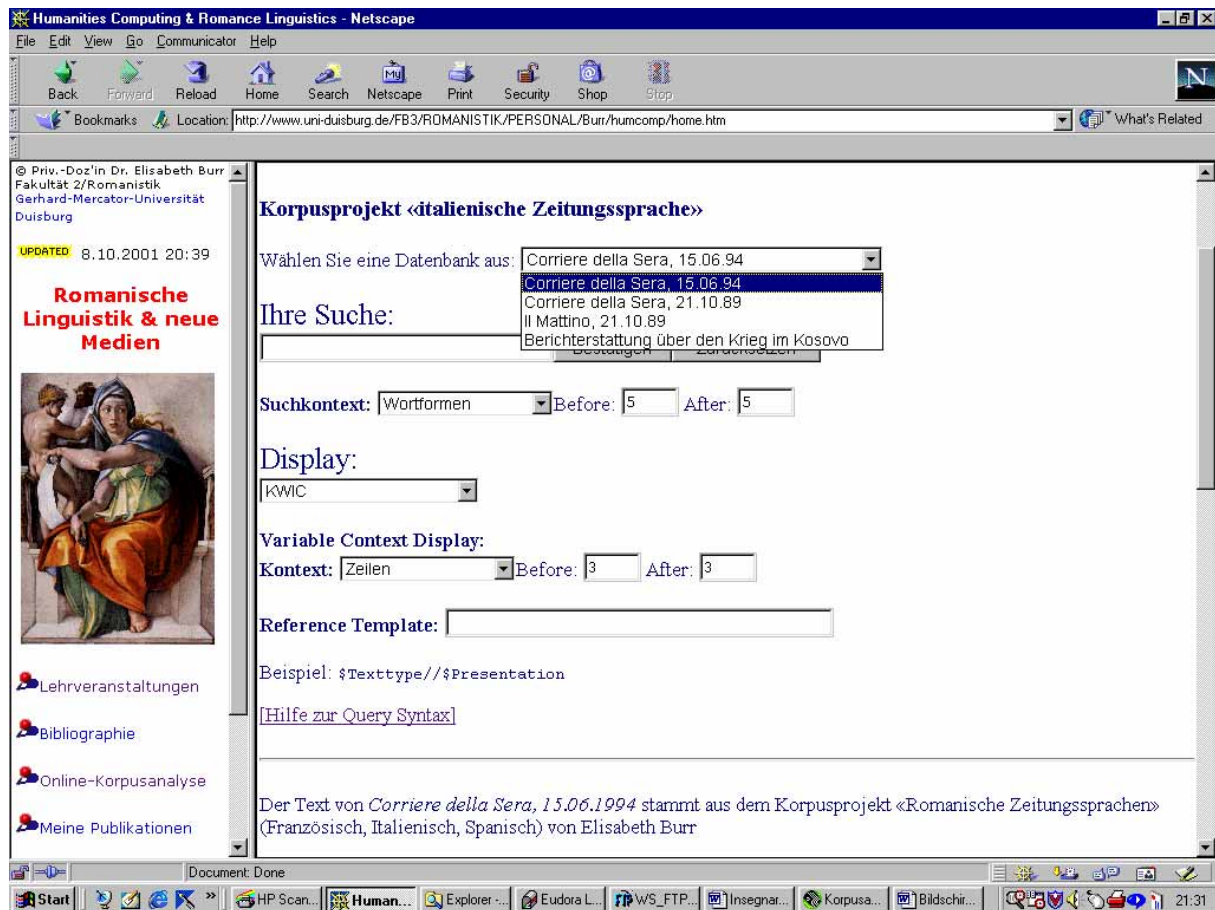


fig. 2: Matrice unica per l'accesso ai corpora italiani

Come *Tact 2.1*, *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 un'apposita maschera:

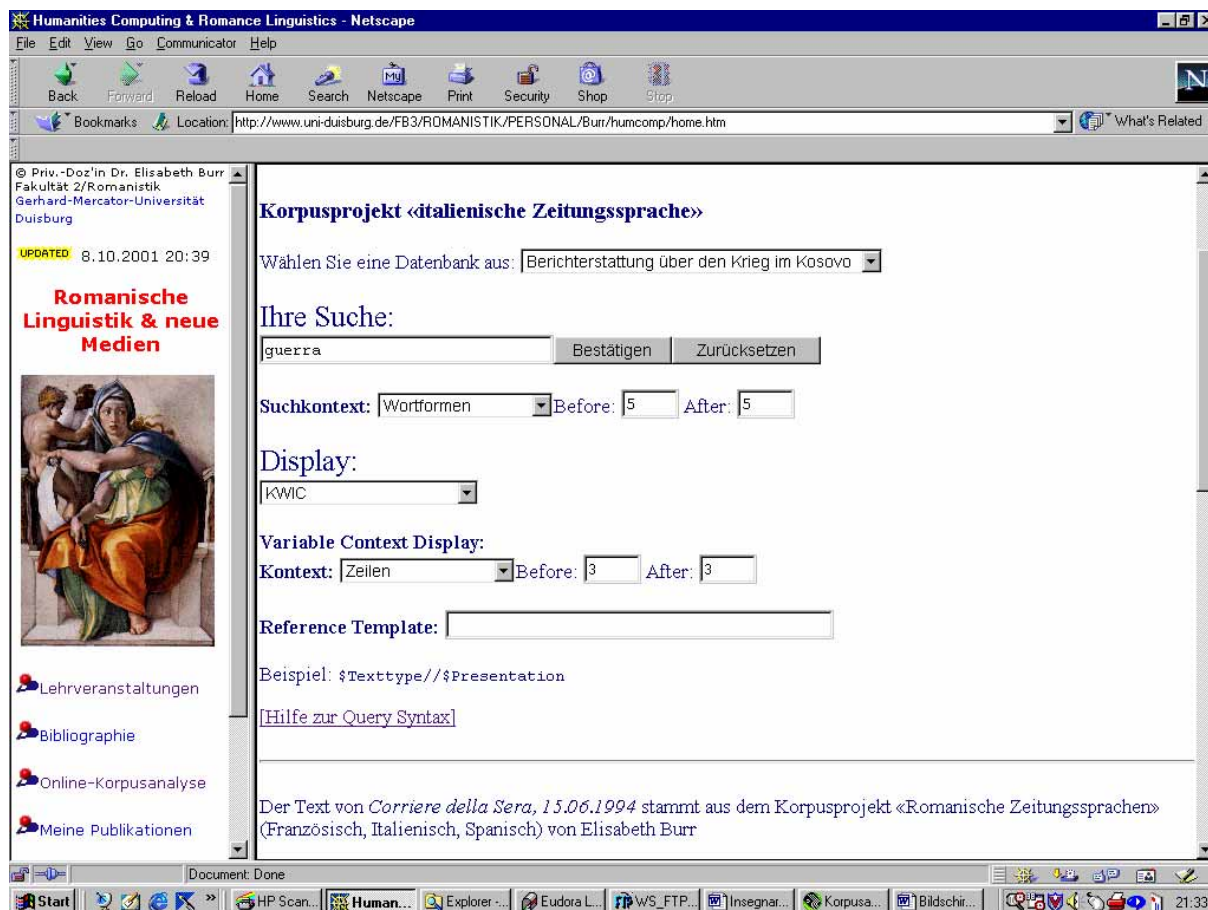


fig. 3: Maschera di ricerca

2.3. 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 dimestichezza con la ricerca, avrei potuto portare qui un esiguo numero di ricerche condotte nell'ambito dei vari seminari sulla variazione, il lessico, le categorie tempo, aspetto e modo e ultimamente la grammatica del verbo. Il progetto che presenterò è stato scelto non soltanto per gli interessanti risultati raggiunti riguardo un tema di nuovo tristemente molto attuale, ma perché dimostra anche che i corpora già in rete non devono, per forza, rimanere le uniche banche dati da esplorare nell'ambito di un determinato seminario, ma possono persino fungere da modelli per la creazione di nuove banche dati. Il progetto è stato elaborato durante un seminario sulle varietà della lingua dei giornali italiani offerto a studenti del primo biennio del corso di laurea in italiano.

A causa della guerra in Kosovo allora in atto, uno dei temi da trattare in questo seminario 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 trattare questo tema unicamente sulla base di 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 analizzare i rapporti del momento sulla guerra in atto. Uno studente ha dunque scaricato articoli da giornali italiani in linea e li ha arricchiti con un *markup* elaborato sulla base del sistema sopra spiegato (cf. 2.1). Il piccolo corpus di 3.513 parole è stato poi trasformato in una banca dati e messo in rete (cf. Rizzo 1999). Questo corpus è poi stato analizzato da una studente.

La ricerca parte da undici parole scelte dalla studente stessa perché giudicate parole chiave quando si parla di una guerra. Nella tabella seguente queste parole sono elencate assieme alle frequenze con le quali appaiono nel corpus:

| <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 ciò che si chiama comunemente una concordanza in formato *KWIC*:

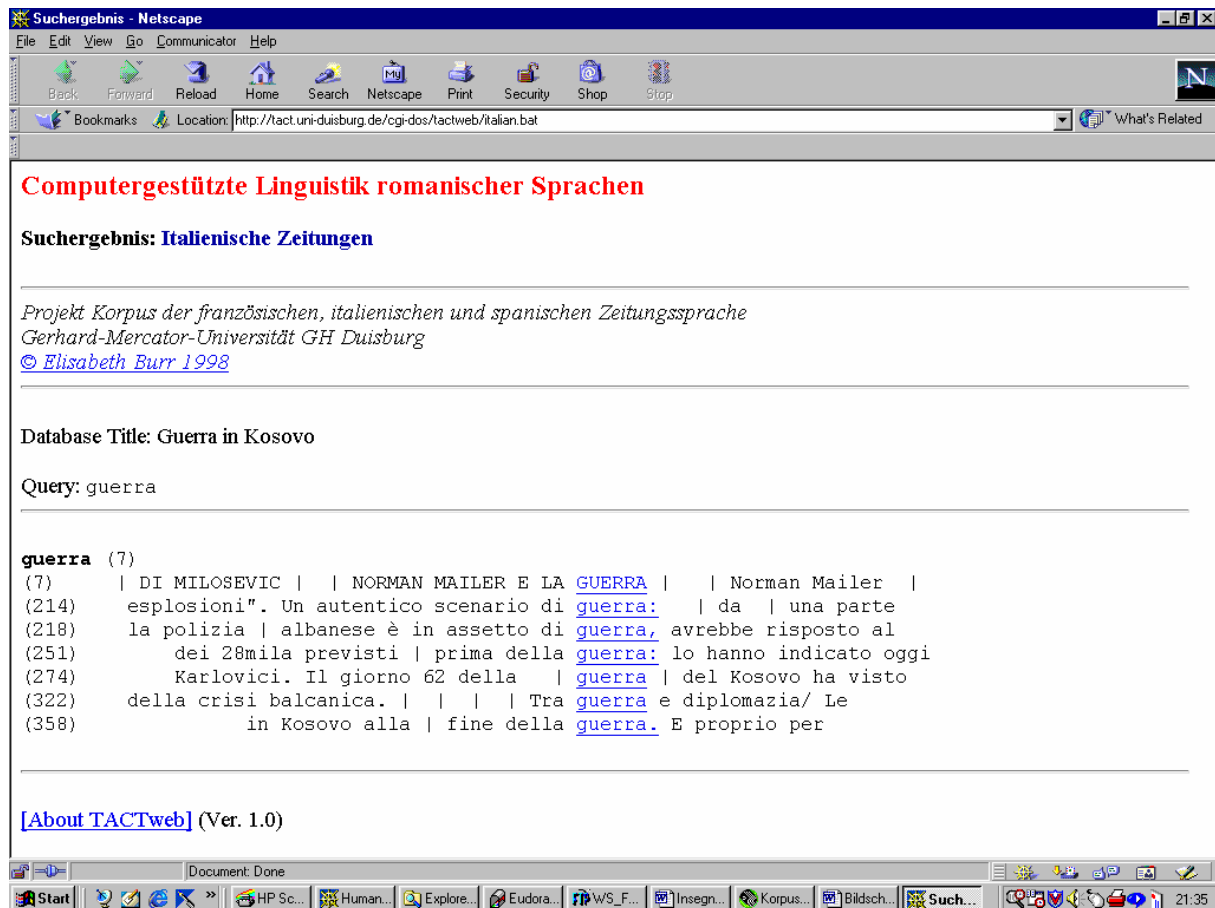


fig. 4: Concordanza in formato KWIC

Come si vede nella fig. 5, *TACTWeb* permette, inoltre, di accedere al testo intero da ogni singola occorrenza

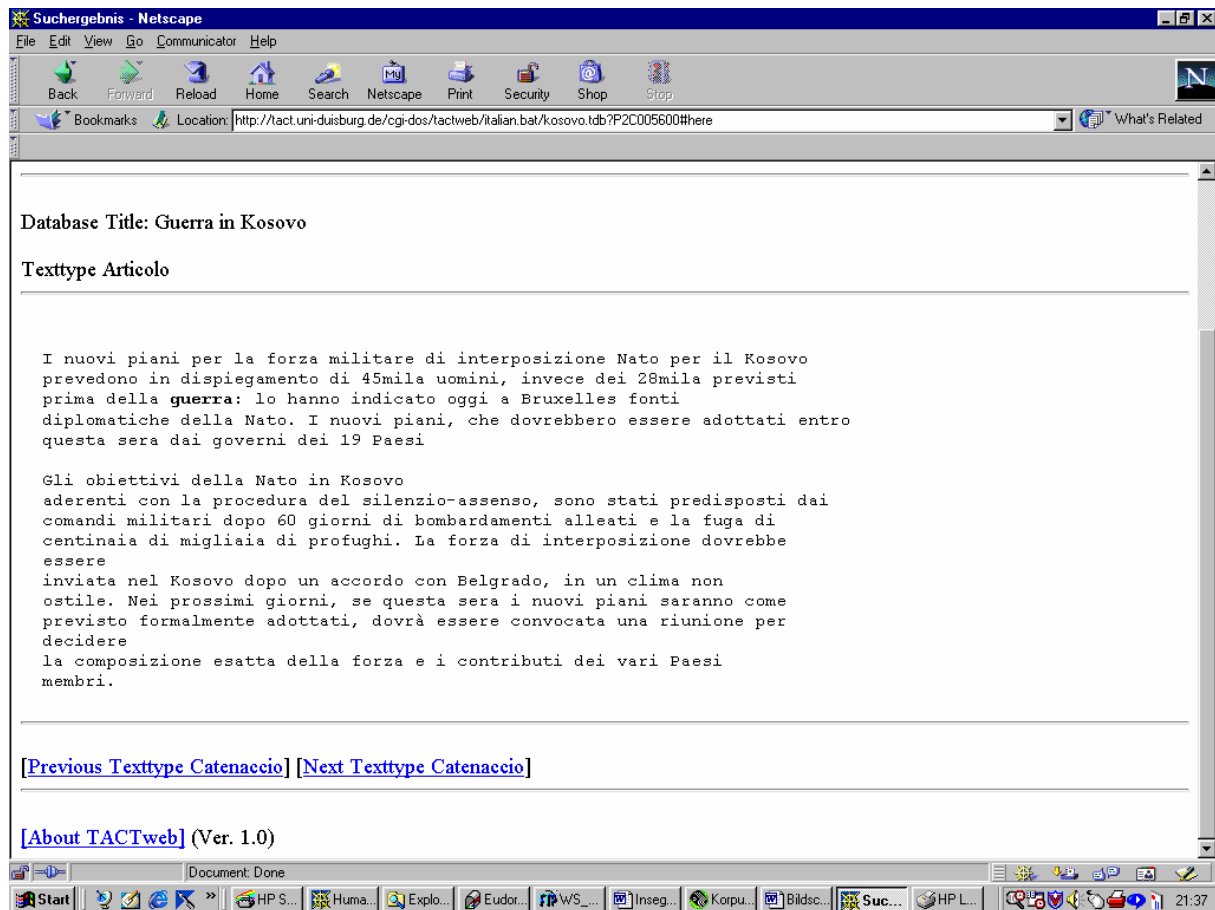


fig. 5: Accesso al testo intero dalla quarta occorrenza di *guerra*

oppure di stabilire direttamente la quantità (in linee) di contesto da rilevare assieme alle occorrenze di una parola chiave:

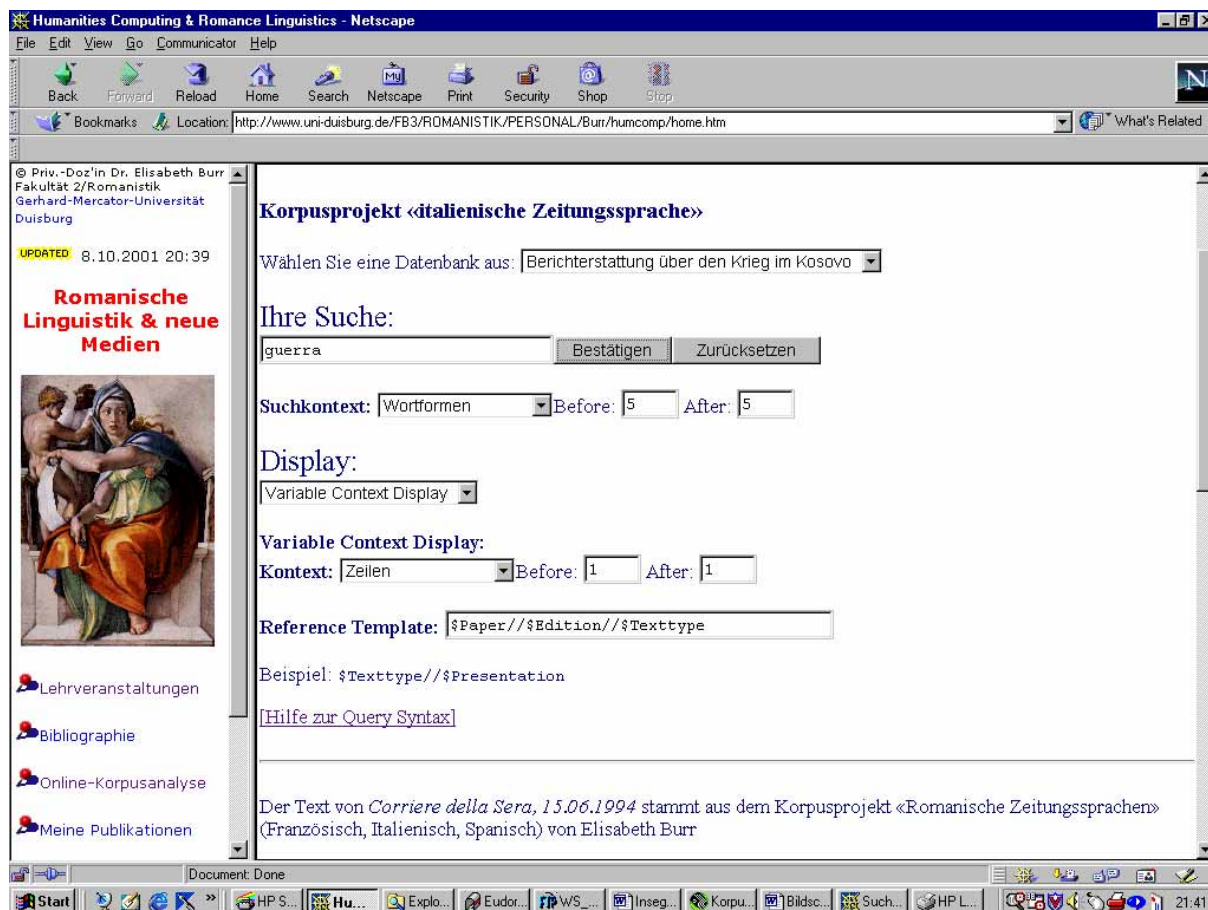


fig. 6: Quantità di contesto (in linee) da rilevare sotto “Kontext”

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.⁶

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:⁷

- L’argomento è la guerra, ma di guerra non si parla quasi 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

```

⁶ Vedasi le concordanze riportate più giù.

⁷ Riferisco di seguito alcune delle conclusioni alle quali arriva Francesca Gironi (1999) nella sua tesina.

- 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
 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 di 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

- il fatto che anche le donne vivono la guerra viene a malapena presa in considerazione:

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 insieme al tipo di organizzazione dei dati permesso da 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 singolarmente, ma si rivelano soltanto quando si riescono a vedere insieme tutte le occorrenze. 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

Come si è visto, un insegnamento che sfrutta le possibilità offerte dalle nuove tecnologie integrando l'analisi di corpora computerizzati nei seminari di linguistica dell'italiano 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. Questa conoscenza le/li aiuta a capire lo stesso oggetto culturale 'lingua' procurando, perciò, un'eccellente base sulla quale occuparsi in modo critico delle teorie e dei risultati di ricerche esistenti.

La realizzazione di questo tipo d'insegnamento dipende, però, per prima cosa dalla disponibilità di corpora del italiano come viene scritto e parlato in situazioni naturali e poi dalla possibilità di accedere a questi corpora in rete. Infatti, solo così l'esito di una consultazione non sarà più dipendente dal fatto che i singoli computer siano stati configurati nel modo richiesto da un determinato programma, ma unicamente dalle realizzazioni

linguistiche contenute nella banca dati e dal tipo della ricerca adottato. Ciò significa che l'attenzione delle e degli studenti non verrà attirata da problemi tecnici collegati alla macchina come tale e alla sua manipolazione; essi quindi si potranno 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 una banca dati disponibile in rete è facilmente accessibile ad ogni ora del giorno e da ogni computer collegato con la rete, le e gli studenti avranno anche tutta la libertà nel gestire il loro lavoro, di tornare su punti rimasti incerti rilevandone altri dati e seguendo strade non ancora esplorate.

La realizzazione di un siffatto corso dipende però anche dalla presenza di un personale docente competente, che sappia sfruttare i nuovi media per condurre ricerche computerizzate sulla base di corpora, che sappia svegliare la curiosità delle e degli studenti per l'uso naturale della lingua e per ricerche di questo tipo come pure proporre soluzioni per i problemi incontrati. Solo così lo sfruttamento delle nuove tecnologie permetterà di creare corsi dove le e gli studenti attraverso le loro proprie esperienze cominceranno a capire che la linguistica è veramente una scienza rivolta prima alla cognizione dei fatti linguistici naturali e dei loro legami con la cultura e la società e poi alla sistemazione delle conoscenze acquisite.

Bibliografia

- Burr, Elisabeth (1993): *Corpus of Italian newspapers*. Oxford: Oxford Text Archive.
- Burr, Elisabeth (1994): "Italian Newspaper Corpus (ita03)", in: Association for Computational Linguistics: *European Corpus Initiative Multilingual Corpus 1 (ECI/MCI)* CD-ROM: \data\eci1\
- 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 / Lauta, Gianluca (2000): "Parlato e finzione di parlato nella stampa di oggi", *Tradizione e Innovazione*, VI Convegno Internazionale SILFI, Duisburg 28.06.-02.07.2000 (in stampa).
- Lancashire, Ian (1996): *Using TACT with Electronic Texts*. A Guide to Text-Analysis Computing Tools, Version 2.1 for MS-DOS and PC-DOS. New York: The Modern Language Association of America.
- Orlandi, Tito (1997): "I nuovi orizzonti della filologia. Ecdotica, critica testuale, editoria scientifica e mezzi informatici elettronici", *Convegno dell'Accademia Nazionale dei Lincei*, <rmcisadu.let.uniroma1.it/>.
- Renear, Allen (1999): "A Conversation With Allen Renear", in: Fish, Tatjana (ed.): *Outlook: Information and the Humanities in the Next Millennium*, FoxNews: September <www.foxnews.com/science/millennium/september/interview_renear1.sml>.

Sabatini, Francesco (1999): "'Rigidità-esplicitzza' vs 'elasticità-implicitzza': 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. Copenhagen: Museum Tusulanum Press 141-172.

TACT 2.1, sviluppato da John Bradley, Lidio Presutti, Michael Stairs e Centre for Computing in the Humanities, University of Toronto.

TactWeb, sviluppato da Geoffrey Rockwell e John Bradley.

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 <www.foxnews.com/science/millennium/september/interview_wurman.sml>.