

La presente raccolta di saggi (prevalentemente inediti, e tutti nati nell'ultimo quinquennio) presenta cinque prospettive sui diversi orizzonti della linguistica generale: dalle Americhe all'Oriente, dalla linguistica storica alla storia della linguistica, dal generativismo alla linguistica dei corpora.

Manuel Barbera è Ricercatore confermato alla Facoltà di Lingue dell'Università di Torino. Ha fatto parte di ricerche COFIN 1997, 1999 e 2001, PRIN 2002 e 2007, FIRB 2001 e Bando Regionale in Materia di Scienze umane e sociali 2008. Nel 1999 ha fondato l'associazione bmanuel.org, ed è membro dell'Associazione per la storia della lingua italiana dal 2004. Si è occupato di linguistica generale, linguistica storica, semantica e linguistica testuale, filosofia del linguaggio, storia della linguistica, linguistica dei corpora, lessicografia, critica testuale, metricologia, paleografia e creazione di font, romanistica, italianistica, uralistica, altaistica, amerindologia ed austronesianistica. È autore di varie monografie (*La gradazione baltofinnica*, 1993; *Introduzione storico-descrittiva alla lingua vota*, 1995 e 2012; *A Short Etymological Dictionary of the Votic Language*, 1994 e 2012; *Corpora e linguistica in rete*, 2007; *Schema e storia del "Corpus Taurinense". Linguistica dei corpora dell'italiano antico*, 2009) e di numerosi saggi su volume e rivista.

In copertina è un'immagine del *Habronattus cuspidatus*, un piccolo ma attivissimo saltacide che, come tutti i ragni saltatori, va famoso per i suoi otto occhi (quali i ragni in genere peraltro hanno) di notevole grandezza ed evidenza, grazie ai quali, tra l'altro, riesce invidiabilmente a guardarsi bene alle spalle (cfr. anche Michael F. Land - Dan-Eric Nilsson, *Animal Eyes*, Oxford - New York - etc., Oxford University Press, 2002 "Oxford animal biology series", soprattutto pp. 95-99).

Non sono purtroppo riuscito ad identificare l'autore dello scatto, e pertanto il mancato riconoscimento della sua paternità e dei suoi diritti è indipendente dalla mia volontà, ma sono dispostissimo a farlo qualora ne fossi posto in grado.

Manuel Barbera

***Molti occhi sono meglio di uno:  
saggi di linguistica generale 2008-12***



**Qu.A.S.A.R. s.r.l.**

Manuel Barbera

*Molti occhi sono meglio di uno:  
saggi di linguistica generale 2008-12*



**Qu.A.S.A.R. s.r.l.**

2013

ISBN-10: 88-87193-26-6  
ISBN-13: 978-88-87193-26-4

Il volume è distribuito con licenza Creative Commons Attribuzione -  
Condividi allo stesso modo 2.5 Italia



<http://creativecommons.org/licenses/by-sa/2.5/it/>

La versione e-book è scaricabile gratuitamente da  
<http://www.bmanuel.org/>

The logo for Bmanuel.org features the word 'Bmanuel' in a stylized, handwritten red font. Below it, the word 'ORG' is written in a smaller, bold, red, sans-serif font.

*I found myself agreeing with Goodman's insistence that the world does not have a "ready-made" or "built-in" description; many descriptions may "fit", depending on our interests and purposes. (This does not mean that anything we happen to like "fits". That more than one description may be right does not mean that every description is right, or that rightness is subjective).*

Hilary Putnam, *A Half Century of Philosophy. Viewed from Within*, in «Dedalus» CXXVI (1997)<sup>1</sup> 198.



## Presentazione.

I saggi qui opportunamente riuniti mostrano bene la caratteristica principale del loro Autore, una curiosità sorniona da gatto soriano che sembra dormire e invece guata il topo. Ampi sono gli orizzonti e lo sguardo, interessi antichi (quelli sul Voto, al quale Barbera ha dedicato un eccellente lavoro di dottorato) e mai sopiti (le lingue orientali e quelle amerinde) si affiancano a più recenti àmbiti di intervento (la linguistica dei corpora) in un insieme che fa andare il lettore sull'ottovolante. Ma non è una sensazione inquietante, perché Barbera affronta con garbo comunicativo i diversi problemi, pur sulla base di una documentazione ineccepibile. Il garbo si rivela al meglio nell'invito inesperto alla riflessione che l'Autore rivolge al lettore.

Così, per limitarmi ad un esempio, nel caso della discussione sul Prete Gianni, tema affascinante e monumentale. È certo che i semi della leggenda sono ben più antichi del XII secolo, ma non basta rifarsi alla chiesa siriana orientale nestoriana della Persia sassanide e al suo zelo missionario, poiché resta insolubile il problema degli *Atti di Tommaso*. D'accordo, è un testo apocrifo e certo non anteriore, nella redazione a noi giunta, al V secolo, tuttavia la filologia ci invita a credere che esso presuppone un testo del II secolo, sempre siriano (da Edessa? Nisibis?), altrettanto colorito e romanzato, dunque ben anteriore allo scisma nestoriano. Oltre tutto, le tradizioni indiane spesso prevedono un arrivo dell'Apostolo via mare, a Malankara nell'India meridionale, una delle sedi principali del commercio monsonico con l'Occidente e non c'è nulla di inverosimile in un arrivo di Ebrei cristianizzati con le navi romane. Si ricorderà infatti che da Quseir, sul Mar Rosso, partivano ogni settimana tre navi per l'India e inoltre che comunità israelitiche sono note anticamente nella medesima regione. Temi affascinanti, sui quali forse mai si potrà dire una parola definitiva (per ora cfr. Robert Eric Frykenberg, *History of Christianity in India*, Oxford University Press, 2008 "Oxford History of the Christian Church").

Lodevole e pacato è il contributo sui rapporti tra linguistica generativa e linguistica dei corpora, del quale va apprezzato la volontà di dialogo scientifico, anche se ci si può legittimamente interrogare sul successo che un'iniziativa come questa può ottenere, poiché la linguistica generativa è per sua natura catafratta alla critica e alla discussione, almeno secondo il mio parere, e tuttavia è opportuno stimolare risposte e favorire incontri: nella galassia che è oggi la linguistica ognuno pensa di essere al centro e rischia di parlare di sé tra sé e sé, dunque bene ha fatto Barbera a tener aperte le porte del dialogo.

Franco Crevatin



## Introduzione.

I cinque saggi che qui si presentano coprono cinque anni di attività scientifica (anche se a volte con radici anteriori) e sono sostanzialmente inediti, vale a dire: il primo è completamente inedito; del secondo è solo presente un PDF nell'Archivio della *Schweizerische Sprachwissenschaftliche Gesellschaft / Société Suisse de Linguistique* (SSG/SSL); del terzo è stata pubblicata unicamente una versione abbreviata; del quarto la versione finora a stampa è in realtà inservibile perché lordata da inaccettabili errori editoriali; e del quinto, infine, è edito giusto un adattamento francese, ma non l'originale italiano.

Tali contributi rappresentano complessivamente cinque diversi sguardi sulla linguistica generale, coprendone il vasto orizzonte con cinque occhi diversi. Geograficamente, vanno dalle Americhe (con la tassonomia delle lingue ameride) all'Asia (con le intricate vicende del Prete Gianni) con al centro la vecchia Europa (con l'anadeissi nelle lingue baltofiniche). Metodologicamente, invece, la scelta è incardinata sulla linguistica storica (con i lavori posti alle estremità della cernita, quello ameridio e quello asiatico, ma soprattutto con quello baltofinico posto al centro, che ne è il vero cuore pulsante); intorno a questi, in posizione mediana, ci sono sia la storia della linguistica (col confronto tra generativismo e linguistica del corpora) sia la più spiccata attualità sincronica (con i corpora di CMR).

La presentazione è, quindi, chiasticamente arrangiata come nello schema musicale cosiddetto "a ponte" del quarto e del quinto quartetto di Bartók, o della decima sinfonia di Mahler. In posizione esposta, di apertura, è però collocato il pezzo amerindio perché è forse il più "scomodo" ed è intorno ad esso che la raccolta è stata ideata.

Tutti i lavori sono presentati con le norme (specie bibliografiche) care all'autore, perlopiù disattese in base ai capricci redazionali degli editori: qualora versioni parziali o semiedite fossero già in esistenza, norme e bibliografia sono state ricorrette ed uniformate; il testo non è invece stato di solito davvero aggiornato, anche se non si è rinunciato ad ancora apportare poche correzioni ed aggiunte qui e là.

Torino, 4 dicembre 2012.





## Una introduzione ai NUNC: storia della creazione di un corpus\*.

*The forms of the Orange tree, the Cocoa nut, the Palms, the Mango, the Banana, will remain clear & separated, but the thousand beauties, which unite them all into one perfect scene, must perish: yet they will leave, like a tale heard in childhood, a picture full of indistinct, but most beautiful figures.*

Charles Darwin, *Diary of the Voyage of H.M.S. Beagle*, 1836 August 6th, Bahia.

0. EXCUSATIO NON PETITA. La ragione principale della mia presenza liminare, forse più che altro simbolica, rispetto alla serie di ricerche che si andrà qui a svolgere è che sono, ebbene sì, almeno questo, il padre dei NUNC: il primo iniziatore, cioè, di una ormai lunga catena di eventi, di cui non ultimi quelli qui in oggetto; il mittente, cioè, al quale, pur talora incolpevole, non mancherà a molti la tentazione di rispedire ogni addebito. E sia.

E poche chiacchiere introduttive a NUNC e newsgroup è tutto quello che vi ammanirò, senza molte altre pretese.

1. NASCITA DEI NUNC. Elaborai, infatti, il progetto dei NUNC (acronimo di *Newsgroup UseNet Corpora*) nell'estate 2002 in [bmanuel.org](http://bmanuel.org) (l'organizzazione che ho fondato: cfr. BARBERA 2007a), anche se indovinai le potenzialità dei newsgroup<sup>1</sup> ed incominciai i primi download sperimentali di testi già nell'inverno 2001.

Vi erano allora solo due precedenti: (1) il tedesco ELWIS (cfr. HINRICHS *et alii* 1995 e FELDWEIG - KIBIGER - THIELEN 1995), storicamente assai importante non fosse perché è stato coinvolto nella creazione del vecchio e glorioso Stuttgart-Tübingen Tagset (STTS) per il POS-tagging del tedesco, tutt'ora largamente usato; (2) il *CMU Text Learning Group Data Archive* di Tom Mitchell del 1993, comunemente noto come *20 Newsgroups*, che non è però un vero precedente poiché, a norma della definizione che correntemente uso di *corpus* (e che riporto qui sotto), non è realmente un corpus, ma bensì una collezione di testi predisposta per *machine learning*.

---

\* Una prima versione di questa breve introduzione fu dapprima presentata alle *Giornate di studio: Scritto e parlato, formale e informale. La comunicazione mediata dalla rete. Torino, 29-30 ottobre 2010*; una seconda versione fu poi presentata ad un incontro presso l'Istituto di italianistica dell'Universität Basel il 20 maggio 2011; un adattamento francese, infine, dal titolo *Une introduction au NUNC: histoire de la création d'un corpus* è stato pubblicato in *Variétés syntaxiques dans la variété des textes online en italien: aspects micro- et macrostructuraux*, édité par Angela Ferrari e Letizia Lala, Nancy, Université de Nancy II - Editions universitaires de Lorraine, 2011 = "Verbum" XXXIII (2011)<sup>1-2</sup> 9-36. Questa versione italiana è inedita.

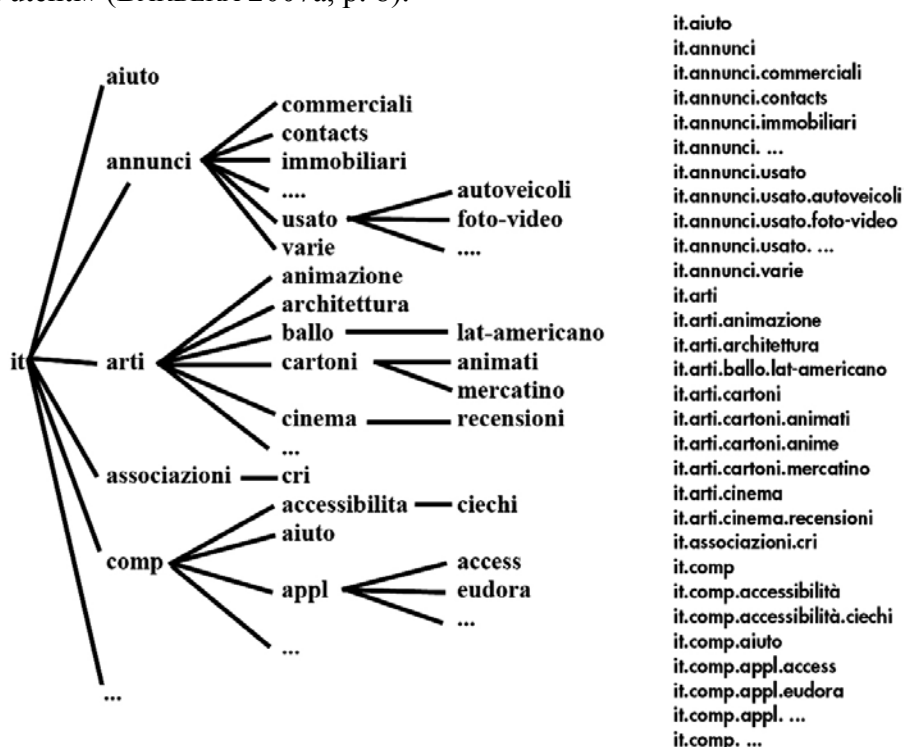
<sup>1</sup> Su forestierismi e prestiti adattati e non in italiano ho largamente argomentato in BARBERA 2009, pp. 7-13 (§ 1.4 *La resa dei forestierismi in italiano*; cfr. anche BARBERA 2003), e non giova qui ripetermi, se non per dire che mi conformerò in tutto alle decisioni allora prese.

Raccolta di testi (scritti, orali o multimediali) o parti di essi in numero finito in formato elettronico trattati in modo uniforme (ossia tokenizzati<sup>2</sup> ed addizionati di markup adeguato) così da essere gestibili ed interrogabili informaticamente; se (come spesso) le finalità sono linguistiche (descrizione di lingue naturali o loro varietà), i testi sono perlopiù scelti in modo da essere autentici e rappresentativi.

Tav. 1. La definizione di corpus secondo BARBERA - CORINO - ONESTI 2007b, p. 70 e BARBERA 2009, p. 126.

Entrambi, però, usavano, per i loro diversi scopi, i newsgroup meramente come comodo serbatoio di testi, non *in qua tale*. Cosa che invece era precisamente il mio intento.

2. I NEWSGROUP PER I NUNC. In breve «un newsgroup è un forum telematico a libero accesso, gratuito, disponibile su Internet, che si manifesta nella forma di testi scritti, ed il cui funzionamento è assai semplice: ogni utente scrive un messaggio, il post, e lo invia ad una specie di “bacheca elettronica” mantenuta presso una rete di server (i newserver che costituiscono UseNet), dai quali gli altri utenti del gruppo possono scaricarlo, leggerlo e rispondervi, costruendo anche articolate catene (thread) di botte e risposte. La facilità d’uso garantisce la grande diffusione dello strumento tra le categorie più diverse di utenti e giustifica la grande quantità di traffico esistente su UseNet. Queste “bacheche elettroniche” che sono i newsgroup sono poi articolate in una tassonomia precisa, ossia in un sistema di cornici argomentative che si chiamano “gerarchie”, a base geografico-nazionale e/o tematica; anche queste gerarchie, peraltro, nascono dal basso in base alla iniziativa degli utenti» (BARBERA 2007a, p. 8).



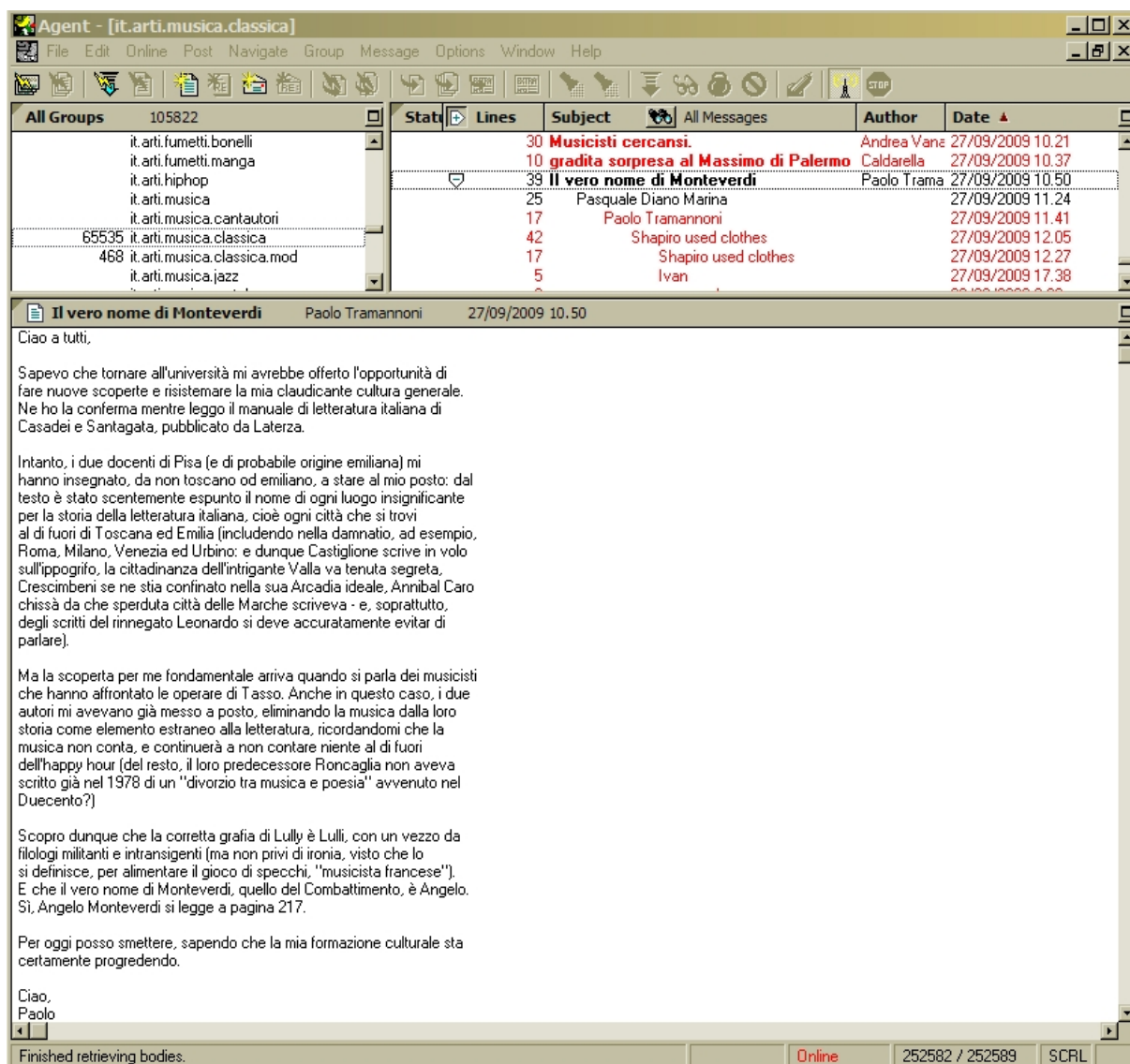
Tav. 2. Porzione iniziale della gerarchia it.\*: (a) struttura arborescente e (b) tabulato lineare dei newsgroup.

Non è qui, infatti, la sede per ritracciare la storia di Usenet, dalla istituzione delle cosiddette *Big Eight*<sup>3</sup>, gerarchie tematiche in lingua inglese, alla nascita delle varie gerarchie “nazionali”: è storia ormai sotto gli occhi di tutti, raccontata e riassunta mille volte sul web. Quello da cui il mio progetto iniziale era partito era invece proprio l’esistenza di queste gerarchie geonazionali, quasi

<sup>2</sup> «E – come dicevo in BARBERA 2011/08, n. 20 – l’attualità ed irrinunciabilità della coppia concettuale (e quindi, in seconda istanza, anche terminologica) token-type è confermata anche da bibliografia recente come WETZEL 2009».

<sup>3</sup> E cioè: comp., news., sci., humanities., soc., talk., misc. più l’anarchica alt.: cfr. ad es. CORINO 2007, pp. 225-227.

sempre “in lingua” ed articolate tematicamente solo al loro interno, un’idea di cui potete avere dalla tavola 2 qui sopra, che riproduce l’inizio della gerarchia ufficiale italiana, *it*.



Tav. 3. Una schermata del newsreader Agent<sup>4</sup>: in alto a destra le gerarchie (con il numero dei post), in alto a sinistra i thread dei post (col il numero di righe per post), ed in basso il testo del post.

In bmanuel.org ho scaricato<sup>5</sup> materiali<sup>6</sup> (dal 2002 ad oggi con varia completezza<sup>7</sup>) per croato (*hr.*), danese (*dk.*), estone (*ee.*), finnico (*sf.*, ecc.<sup>8</sup>), francese (*fr.*), inglese (britannico, *uk.* e *free.uk.*, ed australiano, *au.*), italiano (*it.*, *free.it.*, *it-alt.*, ecc.<sup>9</sup>), norvegese (*no.*), olandese (*nl.*, ecc.), polacco (*pl.* e *free.pl.*), portoghese (continentale, *pt.*, e brasiliano, *br.*), slovacco (*sk.*), spagnolo (di Spagna, *es.* e cileno, *chile.*), tedesco (*de.*, ecc.) ed ungherese

<sup>4</sup> Il post riprodotto (dal newsgroup *it.arti.musica.classica*), conforme all’interesse di VALERE per i “registri alti”, è, appunto, di questo tipo. Ma non troppo ...

<sup>5</sup> Col programma (newsreader) Agent della Forté: cfr. la Tav. 3.

<sup>6</sup> Naturalmente dei soli newsgroup testuali. Molta parte del traffico su Usenet avviene infatti nei newsgroup cosiddetti binari, in cui invece che testo possono venire postati allegati di varia natura (prevalentemente immagini o filmati pornografici, ma non solo), rigorosamente esclusi dai newsgroup testuali.

<sup>7</sup> Utilizzando fino ad otto newserver contemporaneamente: Giganews, Newsreader, Supernews, Newshosting, Active-news, Newsfeeds, Individual.de e X-privat.

<sup>8</sup> Menziono solo le gerarchie principali.

<sup>9</sup> Per le ragioni di questa molteplicità di gerarchie, cfr. *infra* § 6.1.

(hu.). Lo studio delle grandi gerarchie tematiche sovranazionali, linguisticamente, potrebbe essere interessante a vagliare l'esistenza o meno di quell'inglese "globale" lanciato da Crystal: ma, non sentendomi probabilmente nel cuore abbastanza anglista, è il guanto di una sfida che non mi sono ancora sentito di raccogliere<sup>10</sup>.

I NUNC costituiscono così una innovativa suite multilingue (in cui l'italiano è solo il cuore) di corpora costituiti con i testi dei newsgroup.

3. DURA LEX. Il problema dello status legale dei corpora, e quindi quello della loro libera accessibilità, ha sempre rappresentato una mia personale ossessione, che ho cercato di trasmettere a tutto il gruppo di ricerca che ruota intorno a bmanuel.org ed a quello (in larga parte coincidente) dell'Università di Torino che ruota intorno a Carla Marengo e me: le conclusioni di ciò (sia per quanto riguarda l'acquisizione dei testi e le licenze dei corpora, con riferimento all'esperienza di GNU nel software e del copyleft in genere) sono state compendiate in BARBERA - CORINO - ONESTI 2007a: cfr., in particolare, le riflessioni riassunte in ALLORA - BARBERA 2007; la definizione legale della situazione in ZANNI 2007; e le "soluzioni" contrattuali presentate in CIURCINA - RICOLFI 2007.

Ora, «UseNet per definizione e tradizione è il regno del pubblico dominio, quindi [la disponibilità legale del materiale] sembrerebbe una ovvia assunzione; in realtà, se lo si dovesse sostenere legalmente, le cose potrebbero non essere così pacifiche (talvolta si è ricorso ad un cosiddetto "diritto implicito"), ma dato che il comune sentire sostiene comunque la nostra *bonam fidem*, e che non vi sono ad ogni buon conto interessi rilevanti lesi, è certo assai improbabile che contestazioni significative possano essere sollevate. In effetti sono anni che Google mantiene commercialmente archivi di newsgroup senza che ciò sia avvenuto» (BARBERA 2007a, p. 8 in nota).

3.1 EFFETTO COPYLEFT. Ciò assunto, tutti i NUNC sono stati licenziati da bmanuel.org (che ne è il soggetto proprietario) con licenza Creative Commons Share Alike (cfr. CIURCINA - RICOLFI 2007), il che ne garantisce la libera diffusione a chiunque (ad esempio a corpora.unito.it, che ne è il distributore principale) purché i risultati derivati dal lavoro su questi corpora siano rilasciati con la medesima licenza (copyleft): siano cioè resi pubblici alle stesse condizioni.

I NUNC sono così diventati la base testuale per molti progetti, almeno due dei quali hanno visto me o studiosi della nostra scuola tra i proponenti e partecipanti: *in primis* il FIRB RBAU014XCF "L'italiano nella varietà dei testi. L'incidenza della variazione diacronica, testuale e diafasica nell'annotazione e interrogazione di corpora generali e settoriali", coordinatore nazionale Carla Marengo (i cui risultati sono stati affidati alla distribuzione di corpora.unito.it); ed attualmente anche il progetto VALERE (*Varietà Alte di Lingue Europee in REte*), Regione Piemonte Bando Scienze umane e sociali 2008.

3.2 EFFETTO TESTUALE. Un grave limite, spesso lamentato in molti corpora "più o meno" disponibili<sup>11</sup> ma il cui status legale è più incerto dei nostri, è la limitata fruibilità del contesto, cosa che li ha resi perlopiù inutilizzabili per ricerche di tipo testuale. Con la conseguenza che linguistica dei corpora e linguistica testuale si sono sostanzialmente ignorati fino a tempi recenti.

Il problema mi ha colpito particolarmente, non solo perché l'antica amicizia con la sempre compianta Maria-Elisabeth Conte ha impresso la testualità nel DNA mio come di tutti quelli che la conobbero<sup>12</sup>, ma soprattutto perché, teoreticamente, mi sono sempre più convinto che una corretta definizione dei fondamenti della linguistica dei corpora (così come del nodo storiografico, fondamentale, dei rapporti tra strutturalismo e generativismo) non può prescindere dalla corretta comprensione ed individuazione del livello *testuale* del linguaggio (cfr. un primo affondo in

---

<sup>10</sup> Per ora sono stati effettuati scarichi sperimentali ma continuativi solo per alcune aree tematiche di mio più diretto interesse: musica classica, enogastronomia, filosofia, fonts.

<sup>11</sup> Per l'italiano si pensi ad esempio al CORIS.

<sup>12</sup> Valori, anzi, che Carla Marengo ed io abbiamo cercato di trasmettere alle più giovani componenti del nostro gruppo, che più non ebbero la fortuna di conoscerla in persona.

BARBERA 2011 § 1.2.2: ma intendo svolgere in futuro l'argomento in una apposita sede monografica).

La precoce diagnosi del problema pratico legale-testuale di cui sopra, ad ogni buon conto, ha consentito di porvi prestamente rimedio (cfr. BARBERA 2009 § 2.2.3 p. 22), tanto che della recente inversione di tendenza credo di avere una diretta responsabilità, essendo tutti i corpora da me prodotti od ispirati, NUNC in testa, legalmente sicuri, e quindi fruibili senza limitazioni alcune, tantopiù di contesto, che è sempre allargabile fin all'intero testo. A prova di ciò si vedano le ormai molte applicazioni testuali prodotte dalla nostra scuola e, soprattutto, da quella dei testualisti di Basilea, come ad esempio quelle raccolte in BARBERA - CORINO - ONESTI 2007a.

4 LE ATTRATTIVE DEI NEWSGROUP. Molte ragioni, come accennavo (cfr. § 1), mi facevano desiderare di studiare i newsgroup in quanto tali. Alcune sono ovvie (la grande abbondanza testuale; l'organizzazione in gerarchie nazionali che è garanzia di uniformità diacorica; la verosimile disponibilità legale del materiale, ecc.), sicché basta avervi cursoriamente accennato. Altre forse meno, sicché vale la pena di spendervi qualche parola.

4.1 FOLK TAXONOMY ED ENCICLOPEDIA POPOLARE. Una delle caratteristiche più interessanti dei newsgroup è che nascono sempre dal basso in base alla iniziativa degli utenti stessi: la decisione di quali tematiche debbano ricevere una propria bacheca, e di come le bacheche si organizzino all'interno di un dato dominio (nazionale o linguistico) non è decisa dall'alto da una qualche autorità (ministeri, accademie, "specialisti" o lobbies di varia natura).

L'effetto è che una gerarchia di geonazionale di newsgroup si presenta così come una sorta di "enciclopedia popolare" di una data cultura, un vero ritratto spontaneo della società che l'ha prodotta. Ciò la rende, linguisticamente, di speciale interesse lessicografico, tanto per lo studio dei neologismi, quanto per quello dei lessici specialistici, aspetti che furono sfruttati soprattutto per il FIRB "L'italiano nella varietà dei testi".

Dei molti interessi sociologici ed antropologici che ovviamente suscita un simile fenomeno, dobbiamo qui in genere sorvolare. Ma almeno di uno non possiamo completamente tacere, visti i costanti intrecci che l'antropologia ha sempre tessuto con la linguistica (si pensi anche solo a Sapir ed alla sua tradizione): questa "enciclopedia popolare" assomiglia molto nella sua organizzazione ad una *folk taxonomy*; non a caso comune alle *folk taxonomies* ed ai newsgroup è l'organizzazione gerarchica. Il tema è da tempo ben noto agli antropologi (risale, nientemeno, a DURKHEIM 1912), e si è specializzato soprattutto nella etnobotanica e nella etnobiologia in genere (tra la molta bibliografia cfr. almeno BERLIN *et alii* 1973, BROWN 1986 e HEALEY 1993). Scenari più generali («se ne sono infatti già avute interessanti e più generali applicazioni cognitive alla "antropologia della scienza" *tout court* (cfr. Atran 2001)», come notava CORINO 2007, p. 242 nota 23) sono però ben ipotizzabili, e dai NUNC, credo, potrebbe venire un importante contributo.

4.2 UMGANGSSPRACHE. La natura testuale e diamesica dei post dei newsgroup è cosa che ha richiesto parecchie riflessioni. Il quadro generale per le CMR (Comunicazioni Mediate dalla Rete) è stato approntato da Adriano Allora (cfr. ALLORA 2005 e 2009), ed in base a questo, e talora al modello complementare di Koch-Oesterreicher, una prima sistemazione del problema è stata data da CORINO 2007, che ne sottolineava anche la amplissima forbice di registri, tra quelli alti, saggistici (oggetto privilegiati del progetto VALERE) e quelli informali al limite della chat.

Re: lol pügÄ@iöf 08/04/2007 2.36

"Calsifer" <ss@ss> ha scritto nel messaggio  
[news:46177460\\$0\\$36439\\$4fafbaef@reader5.news.tin.it..](mailto:news:46177460$0$36439$4fafbaef@reader5.news.tin.it..)  
 >  
 > "pügÄ@iöf" <alfredo2@gazeta.pl> ha scritto nel messaggio  
 > [news:ev6pql\\$4tp\\$1@inews.gazeta.pl..](mailto:news:ev6pql$4tp$1@inews.gazeta.pl..)  
 >>  
 >> "Calsifer" <ss@ss> ha scritto nel messaggio  
 >> [news:46150111\\$0\\$4797\\$4fafbaef@reader4.news.tin.it..](mailto:news:46150111$0$4797$4fafbaef@reader4.news.tin.it..)  
 >>>  
 >>> "pügÄ@iöf" <alfredo2@gazeta.pl> ha scritto nel messaggio  
 >>> [news:eutvpb\\$hr3\\$1@inews.gazeta.pl..](mailto:news:eutvpb$hr3$1@inews.gazeta.pl..)  
 >>>>!  
 >>>>  
 >>>>  
 >>>> Che lolli?  
 >>>>  
 >>>>  
 >>>>  
 >> lollo xchè con tutto questo casino mi è impossibile concentrarmi....  
 >>  
 >> ^ \_ ^  
 >>  
 > Ma ciao!!  
 > Che casino?  
 >  
 >> ciao fatina,  
 >> felicissima Pasqual!!!  
 >>  
 >  
 > Grazie!! Anche a te felicissima Pasqual! Smack!!  
 > ^ \_ ^  
 >  
 >> (e la prox volta che fai sesso pensamì!)  
 >>  
 >> ^ \_ ^  
 >>  
 >>  
 >  
 > O\_O  
 >  
 >  
 O\_O  
 non fai sesso e non mi vuoi pensare?  
 non fai sesso ma mi vuoi pensare?  
 fai sesso e non mi vuoi pensare?  
 :-)

Tav. 4. Un post "informale", del tipo escluso da VALERE, dal newsgroup *free.it.4amicialbar*<sup>13</sup>.

Su questa questione generale preferirei non ritornare qui, sostando semmai su una caratteristica di questa (queste?) varietà. In BARBERA 2007a, p. 8, parlavo per i newsgroup di «una *Umgangssprache* assolutamente contemporanea e reale molto variata per registri e temi», precisando in nota «La nozione è vetusta, legata soprattutto alle problematiche sorte intorno al cosiddetto “latino volgare” tra i grandi *patres* della romanistica; già lo Spitzer, inoltre, in diversa ma confrontabile ottica, la aveva applicata all’italiano (“*italienische Umgangssprache*”: cfr. SPITZER 1922/2007); e, comunque, è stata riproposta anche recentemente (cfr. KIESLER 2006). L’analogia sembra abbastanza buona, in quanto si tratta, molto in soldoni, di una lingua comune, usuale e media, non tematicamente o sociologicamente delimitabile, più vicina al parlato ma di fatto scritta,

<sup>13</sup> Si noti, peraltro, come l’esempio sia strutturalmente assai interessante: ad una prima proposta il cui testo contiene solo un «!», seguono due battute dal contenuto irrilevante, una domanda disattesa, degli auguri di Pasqua, ed un paio di innocui scherzi sessuali.

e per la quale, in realtà la dicotomia scritto-parlato non è veramente pertinente». Poi, con, appunto, la traduzione italiana del 2007 della *Italienische Umgangssprache*<sup>14</sup> esplose la spitzermania (non saprei come altrimenti chiamarla), e la, peraltro corretta, impostazione di quella nota andava ampliata e ribadita in sede monografica, contro le molte “appropriazioni indebite” che di Spitzer si stavano facendo: nessuna migliore occasione per fare ciò si poteva presentare della commemorazione nencioniana, del grande maestro padre di ogni “parlato-scritto”.

In BARBERA - MARELLO 2008, infatti, ribadii la centralità della tradizione dei latinisti e dei romanisti, culminata con Hofmann, rispetto a quella ballyana, al solito geniale quanto *estrampa*, dello Spitzer, disegnando una prima mappa del ricco humus otto-novecentesco su cui Hofmann e Spitzer si innestano, e prendendo le distanze dalla scelta (interpretativa ancor più che traduzionale) di rendere *Umgangssprache* con *lingua della conversazione* per presentare Spitzer come un precursore di Grice (idealmente sovrapponendo *Stilkritik* e *Speech Acts*, riducendo il “dialogico” al *conversational*, la *Höflichkeit* al *Cooperative Principle*, ecc., con conseguenze a mio parere nefaste).

4.3 THREAD E QUOTING. La modalità medesima con cui gli utenti interagiscono tra di loro con i vari post crea delle strutture testuali e macrotestuali particolarmente interessanti. I vari post si organizzano, infatti, in catene (*thread*) in cui uno risponde all’altro, a volte allontanandosi anche in ramificazioni laterali spesso assai lontane dal capocatena. La disponibilità alla ricerca testuale dei NUNC ha reso possibile studiare alcuni di questi meccanismi.

---

<sup>14</sup> Per un cui apprezzamento cfr. GRAFFI 2008.

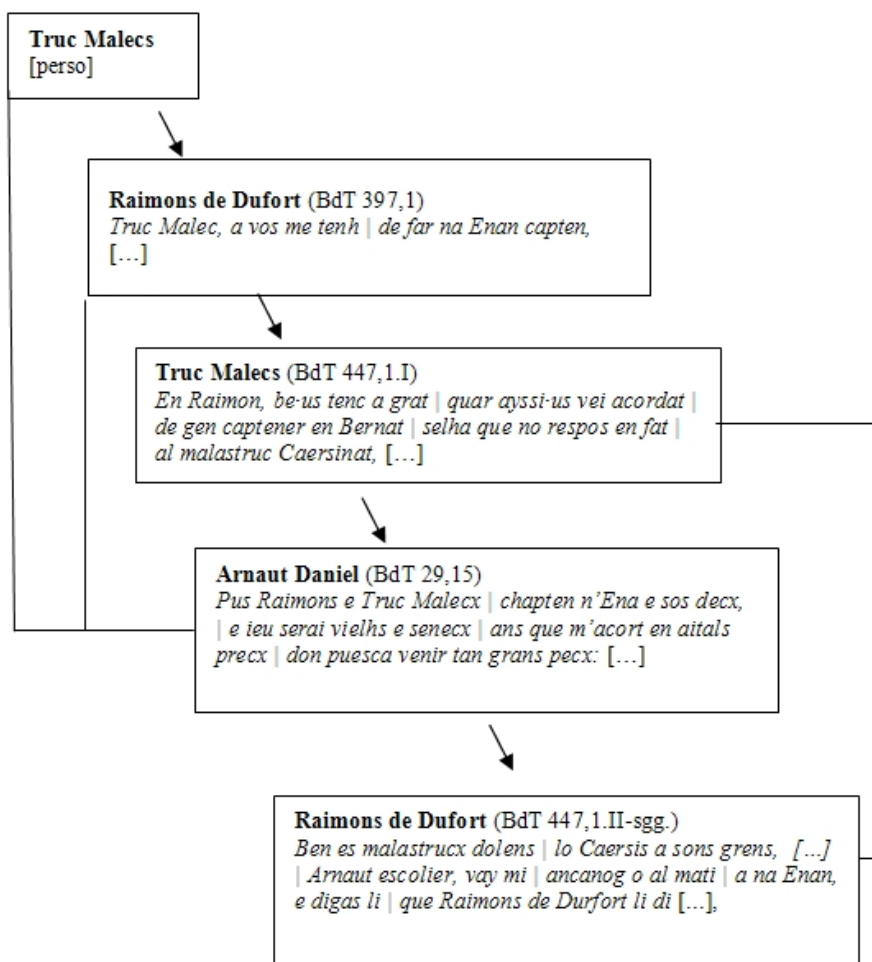


62	Su quali parametri si agisce per "interpretare" ?	Gianluca Billo	14/07/2003 12.12
36	Erewhon		14/07/2003 12.24
47	Gianluca Billo		14/07/2003 13.11
18	Sandro Bisotti		14/07/2003 15.19
29	Giorgia		15/07/2003 12.52
122	Zazie dans le métro		14/07/2003 13.12
122	Gianluca Billo		14/07/2003 13.47
92	Andrea T		14/07/2003 19.03
19	Zazie dans le métro		14/07/2003 21.22
74	Gianluca Billo		15/07/2003 9.46
229	Zazie dans le métro		14/07/2003 21.22
141	Gianluca Billo		15/07/2003 9.46
158	Zazie dans le métro		15/07/2003 13.23
22	Neo		14/07/2003 14.40
32	Sandro Bisotti		14/07/2003 16.33
35	Zazie dans le métro		14/07/2003 21.22
12	Danny Rose		14/07/2003 18.32
33	Andrea Trave		14/07/2003 21.44
26	Zazie dans le métro		14/07/2003 21.52
30	Danny Rose		14/07/2003 21.59
22	Zazie dans le métro		15/07/2003 13.23
82	Sandro Bisotti		15/07/2003 0.26
143	Zazie dans le métro		15/07/2003 13.23
138	Sandro Bisotti		15/07/2003 18.16
47	Massimiliano Vono		15/07/2003 23.25
34	Zazie dans le métro		16/07/2003 12.27
199	Zazie dans le métro		16/07/2003 19.51
24	Moosbrugger		16/07/2003 22.27
29	Zazie dans le métro		16/07/2003 22.57
33	Moosbrugger		16/07/2003 23.42
42	Zazie dans le métro		17/07/2003 13.10
61	Zazie dans le métro		16/07/2003 22.57
31	Zazie dans le métro		17/07/2003 20.10
85	Zazie dans le métro		18/07/2003 12.36
65	Danny Rose		18/07/2003 14.27
67	Zazie dans le métro		18/07/2003 17.31
31	"pp7dd" <p p 7 d d NObScPAMe@N:		17/07/2003 2.04
20	Erewhon		17/07/2003 15.43
24	Erewhon		17/07/2003 15.46
14	Zazie dans le métro		17/07/2003 20.10
28	Sandro Bisotti		17/07/2003 17.40
24	Moosbrugger		17/07/2003 21.16
68	Zazie dans le métro		17/07/2003 21.48
51	Moosbrugger		17/07/2003 22.08
27	Sandro Bisotti		18/07/2003 1.59
15	Zazie dans le métro		18/07/2003 12.36
22	Jashugan		16/07/2003 23.04
34	leonardo polato		14/07/2003 18.57
65	Danny Rose		14/07/2003 21.59
122	Sandro Bisotti		15/07/2003 1.49
53	Danny Rose		15/07/2003 14.30
27	Gianluca Billo		15/07/2003 15.37
33	Zazie dans le métro		15/07/2003 13.23
11	Danny Rose		15/07/2003 14.08
14	Zazie dans le métro		15/07/2003 21.07
45	Danny Rose		15/07/2003 14.06
32	Gianluca Billo		15/07/2003 15.37

Tav. 5. Un thread del newsgroup it.arti.musica classica come visualizzato dal newsreader Agent.

Anche se non intendo indugiarvi oltre<sup>15</sup>, potendo rimandandare al molto già pubblicato (cfr. almeno MARELLO 2007 e COSTANTINO - MARELLO - ONESTI 2009), vorrei solo ricordare come questo tipo di organizzazione responsoriale di testi e macrotesti, *mutatis mutandis*, non è una assoluta novità del mondo della CMR, trovandosene esempi già nella tradizione romanza medioevale: ad Elisa Corino avevo infatti suggerito per il suo importante studio del 2007 di citare il caso della famosa (e famigerata) tenzone del sirventese osceno di Arnaut Daniel:

<sup>15</sup> Per la selezione di post "alti" effettuata pensando a VALERE l'esuberanza del fenomeno è peraltro più contenuta che non per quelle "medi" e "bassi".



TAV. 6. La tenzone tra Truc Malecs, Raimons de Durfort ed Arnaut Daniel su Na Ena. Schema di CORINO 2007, p. 235.  
Testi da CONTINI 1936, pp. 228-30 ed EUSEBI 1984, p. 4.

Tra i fenomeni più interessanti che questa organizzazione responsoriale scatena è quello del **quoting**, ossia la tecnica con cui si riporta (e si cancella) il testo cui si risponde. Un esempio (restando nel registro alto selezionato dal progetto VALERE, cui alcuni dei contributi qui presentati si riconnettono) potrebbe essere il seguente, ricordo di una non memorabile serata della Scala, in cui il noto musicologo Arrigo Quattrocchi risponde ad un denigratore di Donizetti:

Newsgroups: it.arti.musica.classica.mod

Subject: La Fille du =?ISO-8859-1?Q?r=E9giment?= alla Scala

From: rodolfo.canaletti@tin.it (Rudy)

Date: Mon, 26 Feb 2007 17:45:27 +0100

Serata non certo memorabile quella di ieri sera alla Scala. La Fille du | régiment, secondo quanto ne posso capire, è un'opera che non offre | grandi emozioni se non quella di contenere alcune arie che richiedono un | grande virtuosismo vocale. Già posso dire che questo non rientra nei | miei gusti, in genere molto lontani dal puro esibizionismo. | Ma, evidentemente la direzione artistica della scala ha ritenuto, | giustamente credo, necessario inserire in una stagione operistica, oltre | a opere d'avanguardia, a opere di grande tradizione, anche opere di | caratura più leggera e capaci di soddisfare innanzi tutto gli amanti del | bel canto. | E la Fille du Régiment ha tutti i numeri per farlo. | Come struttura operistica è un ibrido: comprende parti recitate, | recitativi veri e propri, arie, pezzi d'insieme, etc. Quindi già | diventa difficile stabilire la sua appartenenza alla Opéra-Comique (come | detto dal sottotitolo) o all'Opera buffa italiana. In più, riesce | difficile stabilire se di opera buffa vera e propria si tratta o di | opera semi-seria, come la presenza di alcune arie struggenti ("Il faut | partir" alla fine del primo atto, e "Par le rang et par l'opulence" e | "Pour me rapprocher de Marie" nel secondo) farebbero pensare. La regia | di Crivelli (con scene e costumi di Zeffirelli), mi è sembrato che abbia | privilegiato la parte buffa. |

Tuttavia c'è da registrare un successo solo parziale di questa | operazione-proposta. Il successo (e la giustificazione) di tale | proposta, proprio per la natura prettamente belcantistica dell'opera, | dovrebbe essere condizionato da un'offerta di qualità di realizzazione | straordinariamente alta: grandissimi interpreti, nuova e

rutilante messa | in scena, direttore d'orchestra capace di valorizzare le voci e nel | contempo di trasmettere emozione. | Tutto ciò non c'è stato. | L'offerta originale si basava soprattutto su due interpreti che sulla | carta dovrebbero essere il meglio che oggi offre il mercato in quanto a | canto di coloratura: Juan Diego Florez e la Natalie Dessay. Ma | evidentemente l'ingaggio di queste due "stelle" deve aver finito i | quattrini, perché si è dovuto ricorrere a una messa in scena che ha | ormai almeno 30 anni sul groppone, e francamente li dimostra tutti. | A questo primo impiccio, se ne è aggiunto subito un secondo (sembra | quasi di assistere a una crisi del governo Prodi!): la Dessay ha | (giustamente) ritenuto sconveniente cantare in una messa in scena così | vetusta, e ha ruscato la partecipazione. | Quindi alla fine, per sostenere la qualità dell'offerta, è rimasto il | solito JDF che in questi ultimi anni non so quante volte abbia | interpretato (bis o non bis) la parte di Tonio. Qualche maligno dice che | in famiglia ormai lo chiamino con questo nome :-)). |

E puntualmente, ieri sera ha fatto capolino la noia. | La messa in scena si avvale di una scenografia stilizzata che | ricostruisce un villaggio tirolese nel primo atto, e un salone del | castello nel secondo. I costumi sono divise militari dell'esercito | napoleonico per gli uomini, e caricature di "suntuosi" vestiti e | parrucche del primo Ottocento per le donne e per i civili. | Il movimenti di recitazione sono spesso spinti fino alla farsa, come i | comportamenti sguaiati di Marie, vivandiera nel primo atto, le corse | agilmente ginniche di Tonio, i comportamenti da sturmtruppen dei | militari, le prove di ballo e di canto del secondo atto, la storpiata | caricaturale pronuncia dei nomi degli ospiti tedeschi, i versacci qua e | là, etc. | La musica e la direzione orchestrale non offrono, almeno a me, alcuno | spunto per un commento. | L'interpretazione del canto ha mostrato una Rancatore che ha fatto il | suo dovere senza una particolare eccellenza; e un Florez che ha fatto | sì, la sua grande figura (ci sarebbe mancato altro che non la facesse!), | ma francamente non mi ha entusiasmato. La celebre aria "Mes amis" | musicalmente nulla di eccezionale, ma resa famosa dai sette (o più?) do | di petto, ha avuto più di tre minuti di applausi e roboanti richieste di | bis (fortunatamente non concesso), ma si è dimostrata solo un esibizione | muscolare. | Florez secondo me non ha una bella voce: il timbro è molto chiaro, direi | adolescenziale (pochi armonici?), il volume piuttosto piccolo. In questo | modo, certo, ne guadagna l'agilità, che indubbiamente è moltissima, e | l'estensione verso l'alto. Ma ne perde la caratura del personaggio. Per | Tonio, personaggio di fatto inesistente, questo può andar bene. Ma ho | sentito dire che intende interpretare personaggi veri come il duca di | Mantova. Sarà, ma la cosa mi lascia molto scettico. |

Per finire, spero di avere un maggior coinvolgimento il 10 marzo con | l'offerta scaligera di una Salome diretta da Harding, con la regia di | Luc Bondy. |

Saluti a tutti |

Rudy! ----- | rodolfo.canaletti@tin.it | <http://www.dicoseunpo.it>

Newsgroups: it.arti.musica.classica.mod

Subject: Re: La Fille du régiment alla Scala

From: "Arrigo Quattrocchi" <ar.quattrocchi@NOSPAMmclink.it>

Date: Sat, 3 Mar 2007 01:54:36 +0100

"Rudy" <rodolfo.canaletti@tin.it> ha scritto nel messaggio |  
news:1hu5lah.md91ea35magwN%rodolfo.canaletti@tin.it... |

*Serata non certo memorabile quella di ieri sera alla Scala.* |

Nulla posso dire sulla serata, ma rimango perplesso di fronte a numerose | affermazioni. |

*La Fille du | régiment, secondo quanto ne posso capire, è un'opera che non offre | grandi emozioni se non quella di contenere alcune arie che richiedono un | grande virtuosismo vocale. Già posso dire che questo non rientra nei | miei gusti, in genere molto lontani dal puro esibizionismo.* |

Nulla da ridire sui gusti, ma non capisco 1) perché il virtuosismo vocale | dovrebbe essere puro esibizionismo e non avere, invece, una funzione | espressiva e un ruolo di drammaturgia, 2) quale tipo di emozioni ci si | aspetti da un'opera di genere comico, certo non quelle di Tristano e Isotta. |

*Come struttura operistica è un ibrido: comprende parti recitate, | recitativi veri e propri, arie, pezzi d'insieme, etc.*

*Quindi già | diventa difficile stabilire la sua appartenenza alla Opéra-Comique (come | detto dal sottotitolo) o all'Opera buffa italiana. In più, riesce | difficile stabilire se di opera buffa vera e propria si tratta o di | opera semi-seria, come la presenza di alcune arie struggenti ("Il faut | partir" alla fine del primo atto, e "Par le rang et par l'opulence" e | "Pour me rapprocher de Marie" nel secondo) farebbero pensare. La regia | di Crivelli (con scene e costumi di Zeffirelli), mi è sembrato che abbia | privilegiato la parte buffa.* |

La fille du régiment è a tutti gli effetti una opéra-comique, tanto che per | farla circolare sui palcoscenici italiani Donizetti dovette fare numerosi | adattamenti, e la versione italiana non ha mai convinto fino in fondo | proprio perché sembra troppo francese. La versione francese ha infatti i | dialoghi recitati, e la presenza dell'elemento patetico non è anomala nel | genere francese. La presenza di brani patetici in un'opera comica era | peraltro abituale anche in Italia, e proprio Donizetti aveva accentuato | questa contaminazione, basta pensare all'Elisir d'amore. | Il patetico, d'altronde, è un fortissimo elemento di contrasto rispetto al | comico, e quest'ultimo punta sull'eleganza e la leggerezza, non sulla farsa. | Anche per questo penso che La fille du régiment sia un capolavoro di | drammaturgia, comica s'intende, e non sia affatto un'opera di mero | esibizionismo. Forse bisognerebbe cercare di capire i meccanismi del comico, | e di | apprezzarli per quello, e non è poco, che possono comunicare al pubblico. |

Una postilla sullo spettacolo di Crivelli-Zeffirelli. Credo sia del 1959, e | dunque è credo il più antico spettacolo in circolazione sui palcoscenici | italiani e forse stranieri. Lo vidi in teatro una quindicina di anni fa e | già allora mi sembrò da rottamare senza indugio. Oltretutto nacque per la | versione italiana dell'opera, che ha un carattere diverso. E' sconcertante | che la Scala abbia deciso di riesumarlo ancora una volta, invece di passare | a qualcosa di più moderno e, soprattutto, meno farsesco. Penso che il | vecchiume di questo allestimento possa inficiare la migliore delle | esecuzioni musicali. Bene ha fatto la Dessay a lasciare questa produzione al | suo destino. |

Tav. 7. Il quoting tra due post di `it.arti.musica.classica`  
(indico con la barra | l'andata a capo e col corsivo il testo quotato).

Essendo una maestra di ellissi, è ben comprensibile che quest'ordine di fenomeni sia stato (e continui ad essere) particolarmente studiato da Carla Marellò, ai cui lavori non posso che rinviare (cfr. soprattutto MARELLÒ 2007).

5. PROBLEMI. A fianco di molte caratteristiche interessanti, non mancano certo i problemi nel maneggiare questo tipo di materiali; problemi che tre anni fa così riassumevo: «(a) peculiarità linguistiche mediate dal mezzo (gergo informatico, abbreviazioni, emoticon, ecc.); (b) frequenti “sporature” del testo dovute alla trasmissione (passaggio da charset diversi, ecc.) od alla battitura; (c) presenza di spam, post OT (“out of topic”) e crossposting; (d) l'abbondanza di testo ripetuto, a volte (quando effetto del quoting) testualmente rilevante e quindi “buono”, ma comunque sempre per statistiche lessicali dannoso» (BARBERA 2007a, p. 8).

Nel corso del tempo (da CASAVECCHIA 2005 a BARBERA - COLOMBO 2010, giusto per indicare le due tappe più salienti) a questi problemi abbiamo trovato rimedi via via migliori, ma una soluzione realmente efficace resta per alcuni aspetti ancora di là da venire: le versioni che abbiamo messo finora online riescono comunque a funzionare ed a contenere i danni entro l'accettabile, od almeno così spero.

Non voglio qui entrare nei dettagli della struttura e delle funzioni dei filtri di pulizia allestiti, né in quello della creazione dell'interfaccia (adeguate informazioni sono comunque ricavabili dalla bibliografia), se non per segnalare un paio di caratteristiche (o piuttosto limiti) delle versioni attualmente disponibili online particolarmente rilevanti a chi utilizzi questi corpora a fini testuali. Il primo è che per consentire un uso tradizionale (cioè statistico-lessicografico) dei corpora, andava eliminato il più possibile il testo ripetuto: e così nei corpora standard i testi originari sono stati fortemente potati riducendoli ad un solo messaggio per thread (ma alcune versioni “intatte” dei più piccoli sono state messe online con la dicitura *uncut*); il secondo è che nelle versioni “vecchie” ancora online non sono marcati i confini di frase, problema comunque che abbiamo recentemente risolto (cfr. BARBERA - COLOMBO - ONESTI 2008 *i.s.*), ma non ancora esportato retroattivamente. Questi problemi, inoltre, va ricordato, riguardano solo le versioni online; in locale è tutto diverso.

6. SPECIFICHE. I corpora di newsgroup italiani della suite NUNC attualmente disponibili online (home: <http://www.bmanuel.org/projects/ng-HOME.html>; distribuiti tutti da [corpora.unito.it](http://corpora.unito.it)), fra corpora generalisti e corpora specialistici di cucina, motori e fotografia digitale, ammontano finora a 280.587.779 token (cfr. BARBERA 2007, p. 6, Tav. 2), e coprono il traffico postato nella prima metà del corrente decennio.

A questi se ne stanno affiancando altri prodotti per il progetto VALERE che estenderanno la loro copertura anche alla seconda metà del decennio. In locale, comunque, tale copertura è già totale.

6.1 LA USENET ITALIANA. La Usenet italiana cui i NUNC-IT attingono ha delle caratteristiche proprie che imprimono uno speciale valore ai corpora italiani, rispetto al panorama internazionale, e su cui meriterebbe farne qualche cenno.

La cosa che più colpisce un esploratore che traguardi per la prima volta il panorama della Usenet italiana, è la straordinaria abbondanza e varietà dei suoi panorami. Alla gerarchia geonazionale storica, *it.*, che aveva delle procedure per la creazione di newsgroup nuovi troppo rigorose, si affiancò ben presto (sul modello degli *alt.* internazionali) una gerarchia *free.it.*, dopo alcuni anni di vita tumultuosa ufficialmente chiusa (che significa che non vengono più creati newsgroup nuovi, ma quelli esistenti sono in molti casi attivi ancora oggi) perché troppo anarchica e sostituita

da una più tranquilla gerarchia *it-alt*. A fianco a queste fu presto istituita la gerarchia *italia*, che comprende solo newsgroup di interesse municipale. Ma ancora non basta: le principali ditte di servizi di comunicazione hanno spesso creato delle gerarchie autonome, perlopiù a scarsa vitalità, dalle storiche *agora.*, *vol.*, *tin.* e *tiscali.* alla più recente *fastweb.*; e su questo modello lo stesso hanno fatto alcune organizzazioni private, come la gerarchia *scout*. Solo la Germania, in Europa, si avvicina ad una simile situazione, pur restandone largamente al di sotto.

La caratteristica, sia pure estrinseca, che ne deriva ai NUNC-IT è quella di una ancora maggiore rappresentatività della *Umgangssprache* nazionale, e quella di una maggiore duttilità alla creazione di corpora specifici per usi particolari, che magari non sarebbe stato possibile creare per altre lingue europee.

6.2. CORPORA CHIUSI. Con tutto il parlare che recentemente si fa di *Web Corpora*, gioverà ribadire che i NUNC sono corpora nel senso proprio della definizione data in Tav. 1: «sono quindi chiusi, fissati, rielaborati e variamente annotati, non fluttuanti come la rete su cui si fanno ricerche estemporanee, ad esempio con Google, o come i cosiddetti *Web Corpora*, spesso presentati come l'ultima frontiera della linguistica dei corpora (cfr. BARBERA - CORINO - ONESTI 2007b, § 1.5 *I corpora futuribili: Web as a corpus?*, pp. 44-45, con bibliografia)» (BARBERA - MARELLO 2008).

6.3 IL GESTORE CQP. Tutti i corpora NUNC, che siano quelli “vecchi” creati per il FIRB “La varietà dei testi” che quelli nuovi, creati per il progetto VALERE, sono gestiti col Corpus Workbench (CWB), il software open source creato dell'IMS Stuttgart, basato sul Corpus Query Processor (CQP) e sul POS-tagger TreeTagger, anche se interfaccia differenziate sono allo studio.

Su CQP e CWB la letteratura non manca di certo e posso rimandare a CHRIST - SCHULZE 1996, HEID 2007, BARBERA 2009 ¶21 pp. 985-1021, oltre naturalmente alla homepage del software ed alla manualistica che vi si può scaricare; per il TreeTagger cfr. SCHMID 1994 e la manualistica sul sito.

7 ATTUALITÀ DEI NUNC. Un dubbio inquietante, a quel che sento, sta serpeggiando e prendendo vieppiù piede: che Usenet sia ormai un fenomeno sorpassato, sostituito negli ultimi anni dai blog, anche per i settori più tecnici.

Mi sento di poter smentire abbastanza categoricamente questa leggenda metropolitana: Usenet gode di ottima salute, il volume del postato non è affatto minore a quello degli anni passati (fatto salvo il picco di metà decennio)<sup>16</sup>, i newsgroup più tecnici non languono affatto, e complessivamente gli italiani si confermano come la popolazione d'Europa che più usa i newsgroup.

Le condizioni per cui la lingua dei newsgroup possa fornire quella lingua dell'uso medio, quel parlato-scritto, quella *Umgangssprache* nazionale che, almeno per l'italiano, sulla scorta di Sabatini e Nencioni, teorizzavamo in BARBERA - MARELLO 2008, e che sembrava una irraggiungibile pantera, sulla cui odorosa usta tanti grandi linguisti e storici della lingua si erano mossi senza pure mai giungerne alla cattura, sembrano continuare a sussistere e sorridere al futuro della ricerca, invogliandoci ad estendere la copertura cronologica dei NUNC al presente e quella diatopica all'Europa tutta, ed oltre, come si sta effettivamente cercando di fare.

---

<sup>16</sup> Infatti se verso la metà del primo decennio del secolo (con minori variazioni a seconda dei paesi) si è verificata la massima espansione del medium, negli anni successivi vi è stata una generale stabilizzazione su buoni valori medi, con consolidamento od addirittura recupero nel volume del postato, specie nelle gerarchie italiane. Se la ripresa è più marcatamente visibile nella sommatoria complessiva delle gerarchie nazionali (essendo fortemente aumentati il numero di newsgroup attivi), anche in singole gerarchie la curva di consolidamento-recupero è abbastanza chiara; si considerino ad esempio i newsgroup campione *it.hobby.cucina* (2004: 44.3 MB; 2005: 67.2 MB; 2006: 49.6 MB; 2007: 57.6 MB; 2008: 54.6 MB; 2009: 51.7 MB; 2010: 52.5 MB) e *de.rec.fotografie* (2004: 92.1 MB; 2005: 84.1 MB; 2006: 65.5 MB; 2007: 44.5 MB; 2008: 45.6 MB; 2009: 56.4 MB).

BIBLIOGRAFIA.

ALLORA

- 2005 Adriano Allora, *A Tentative Typology of Net Mediated Communication*, comunicazione presentata alla *Corpus Linguistics 2005 Conference, Birmingham July 14-17 2005*, disponibile online alla pagina <http://www.corpus.bham.ac.uk/PCLC/>
- 2009 Adriano Allora, *Variazione diamesica generale nelle Comunicazioni Mediate dalla Rete*, in «Rassegna Italiana di Linguistica Applicata» III (2009) 147-170.

ALLORA - BARBERA

- 2007 Adriano Allora - Manuel Barbera, *Il problema legale dei corpora. Prime approssimazioni*, in BARBERA - CORINO - ONESTI 2007a, pp. 109-118.

ATRAN

- 2001 Scott Atran, *Folk Biology and the Anthropology of Science. Cognitive Universals and Cultural Particulars*, preprint to «Behavioral and Brain Sciences», deposited on 30th April 2001, online at <http://www.bbsonline.org/Preprints/OldArchive/bbs.atran.html>.

BALLY

- 1937/09 Charles Bally, *Traité de stylistique française*, 2. édition, Carl Winter - C. Klincksieck, Heidelberg - Paris, 1937<sub>2</sub> [1909<sub>1</sub>; 1951<sub>3</sub>, Genève, Georg & cie].

BARBERA

- 2003 Manuel Barbera, *Review to Manfred Görlach, A Dictionary of European Anglicisms. A Usage Dictionary of Anglicisms in Sixteen European Languages*, in «International Journal of Lexicography» XVI (2003)<sup>2</sup> 208-216.
- 2004 Manuel Barbera, *Il progetto FIRB. Stato dei lavori*, documento interno inedito, Ver. 7 aggiornata al febbraio 2004.
- 2007a Manuel Barbera, *Per la storia di un gruppo di ricerca. Tra bmanuel.org e corpora.unito.it*, in BARBERA - CORINO - ONESTI 2007a, pp. 3-20.
- 2007b Manuel Barbera, *I NUNC-ES: strumenti nuovi per la linguistica dei corpora in spagnolo*, in «Cuadernos de filología italiana» XIV (2007) 11-32.
- 2009 Manuel Barbera, *Schema e storia del "Corpus Taurinense". Linguistica dei corpora dell'italiano antico*, Alessandria, Edizioni dell'Orso, 2009.
- 2011/08 Manuel Barbera, *"Partes Orationis", "Parts of Speech", "Tagset" e dintorni. Un prospetto storico-linguistico*, in BORGHI - RIZZA 2011, tomo I, pp. 113-145. Rielaborazione di una lezione inedita, *Parti del discorso ed annotazione di corpora elettronici*, tenuta a Basilea il 9 maggio 2008 presso l'Istituto di Italianistica dell'Universität Basel.
- 2011 Manuel Barbera, *Intorno a "Schema e storia del Corpus Taurinense"*, comunicazione al *III Incontro di filologia digitale, Verona, 3-5 marzo 2010*, ora in COTTICELLI KURRAS 2011, pp. 27-48.

BARBERA - COLOMBO

- 2010 Manuel Barbera - Simona Colombo, *Pre-Processing Normalization Procedures for Newsgroup Corpora*, in MONEGLIA - PANUNZI 2010, pp.175-191.

BARBERA - COLOMBO - ONESTI

- 2008 *i.s.* Manuel Barbera - Simona Colombo - Cristina Onesti, *Einige Vorschläge zur Identifizierung der Satzgrenzen in Newsgroup-Diskursen: computer- und textlinguistische Probleme*, relazione alla *Tagung Deutsche Sprachwissenschaft in Italien, Rom, 15.2.2008*; in corso di pubblicazione.

BARBERA - CORINO - ONESTI

- 2007a Manuel Barbera - Elisa Corino - Cristina Onesti (a cura di), *Corpora e linguistica in Rete*, Perugia, Guerra Edizioni, 2007.
- 2007b Manuel Barbera - Elisa Corino - Cristina Onesti, *Cosa è un corpus? Per una definizione*

*più rigorosa di corpus, token, markup*, in BARBERA - CORINO - ONESTI 2007a, pp. 25-88.

BARBERA - MARELLO

2003/12 Manuel Barbera - Carla Marellò, *Corpo a corpo con l'inglese della corpus linguistics, anzi, della linguistica dei corpora*, in NESI - DE MARTINO 2012, pp. 357-370.

2008 Manuel Barbera - Carla Marellò, *Tra scritto-parlato, Umgangssprache e comunicazione in rete: i corpora NUNC*, in «Studi di Grammatica Italiana» XXVII (2008, *recte* 2011) = *Per Giovanni Nencioni. Convegno Internazionale di Studi*. Pisa - Firenze, 4-5 Maggio 2009, a cura di Anna Antonini e Stefania Stefanelli, Firenze, Le Lettere, 2011, pp. 157-185.

BERLIN *et alii*

1973 Brent Berlin - Dennis E. Breedlove - Peter H. Raven, *General Principles of Classification and Nomenclature in Folk Biology*, in «American Anthropologist» VII (1973) 214-242.

BORGHI - RIZZA

2011 *Anatolistica Indoeuropeistica e Oltre – nelle Memorie dei Seminari offerti da Onofrio Carruba (Anni 1997-2002), al Medesimo presentate*, a cura di Guido Borghi ed Alfredo Rizza, Milano, Qu.A.S.A.R., 2011 “Antiqui Aevi grammaticae artis studiorum consensus. Series maior” 1.

BROWN

1986 Cecil H. Brown, *The growth of ethnobiological nomenclature*, in «Current Anthropology» XXVII (1986)<sup>1</sup> 1-19.

CASAVECCHIA

2005 Sara Casavecchia, *Progettazione ed implementazione di corpora di lingua inglese basati sui newsgroups*, Tesi di laurea, Facoltà di lingue e letterature straniere, Università di Torino, a.a. 2004-2005, relatore Emanuele Barbera.

CHRIST - SCHULZE

1996 Oliver Christ - Bruno Maximilian Schulze, *CWB. Corpus Work Bench, Ein flexibles und modulares Anfragesystem für Textcorpora*, in FELDWEG - HINRICHS 1996; disponibile online alla pagina <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/Papers/christ+schulze:tuebingen.94.ps.gz>.

CIGNETTI

2007 Luca Cignetti, *Alcune forme di polifonia testuale nei notiziari accademici di Athenaeum. Aspetti funzionali ed argomentativi*, in BARBERA - CORINO - ONESTI 2007a, pp. 199-207.

CIURCINA - RICOLFI

2007 Marco Ciurcina - Marco Ricolfi, *Le Creative Commons Public Licences per i corpora. Una suite di modelli per la linguistica dei corpora*, in BARBERA - CORINO - ONESTI 2007a, pp. 127-132.

CONTE

1999/88 Maria-Elisabeth Conte, *Condizioni di coerenza*, Alessandria, Edizioni dell'Orso, 1999. Nuova edizione, con l'aggiunta di due saggi a cura di Bice Mortara Garavelli, di Maria-Elisabeth Conte, *Condizioni di coerenza. Ricerche di linguistica testuale*, Firenze, La Nuova Italia Editrice, 1988 “Pubblicazioni della Facoltà di Lettere e filosofia dell'Università di Pavia” 46.

CONTINI

1936 Gianfranco Contini, *Per la conoscenza di un sirventese di Arnaut Daniel*, in «Studi Medievali» ns. IX (1936), pp. 223-231.

CORINO

2007 Elisa Corino, *NUNC (Newsgroup UseNet Corpora). Questioni metodologiche ed aspetti della testualità*, in BARBERA - CORINO - ONESTI 2007a, pp. 225-252.

COSTANTINO - MARELLO - ONESTI

2009 Mauro Costantino - Carla Marellò - Cristina Onesti, *La cucina discussa in rete. Analisi di gruppi di discussione italiani relativi alla cucina*, in FROSINI - ROBUSTELLI 2009, pp. 717-727.

COTTICELLI KURRAS

2011 *Linguistica e filologia digitale: aspetti e progetti*, a cura di Paola Cotticelli Kurras, Alessandria, Edizioni dell'Orso, 2011.

CRYSTAL

1997 David Crystal, *English as a Global Language*, Cambridge, Cambridge University Press, 1997 [2003<sup>r</sup>].

2001 David Crystal, *Language and the Internet*, Cambridge University Press, 2001 [2006<sub>2</sub>].

DURKHEIM

1912/2003 Émile Durkheim, *Les formes élémentaires de la vie religieuse: le système totémique en Australie*, Paris, F. Alkan, 1912. [riedizione moderna: Paris, PUF, 2003 "Quadrige"].

EUSEBI

1984 Arnaut Daniel, *Il sirventese e le canzoni*, a cura di Mario Eusebi, Milano, All'insegna del pesce d'oro, 1984.

FELDWEG - HINRICHS

1996 *Lexikon und Text: wiederverwendbare Methoden und Ressourcen zur linguistischen Erschließung des Deutschen*, herausgegeben von Helmut Feldweg und Erhard W. Hinrichs, Tübingen, Max Niemeyer Verlag, 1996 "Lexicographica. Series maior" 73.

FELDWEG - KIBIGER - THIELEN

1995 Helmut Feldweg - Ralf Kibiger - Christine Thielen, *Zum Sprachgebrauch in deutschen Newsgruppen*, in «Osnabrücker Beiträge zur Sprachtheorie» L (1995) 143-154, disponibile anche online <http://www.sfs.uni-tuebingen.de/Elwis/news.ps>.

FERRARI - MANDELLI

2007 Angela Ferrari - Magda Mandelli, *Note sull'impiego dei connettivi nei notiziari accademici del corpus Athenaeum. Aspetti quantitativi e qualitativi*, in BARBERA - CORINO - ONESTI 2007a, pp. 183-198.

FIORI

2004 Silvia Fiori, *An Analysis of Linguistics Newsgroups through their Paratexts*, in «Rassegna italiana di linguistica applicata» XXXVI (2004)<sup>2-3</sup> 67-81.

FROSINI - ROBUSTELLI

2009 *Storia della lingua e storia della cucina. Parola e cibo: due linguaggi per la storia della società italiana. Atti del convegno ASLI 2007, Modena 20-22 settembre*, a cura di Cecilia Robustelli e Giovanna Frosini, Modena, Cesati, 2009.

GHENO

2004 Vera Gheno, *Prime osservazioni sulla grammatica dei gruppi di discussione telematici di lingua italiana*, in «Studi di Grammatica Italiana» XXII (2004) 267-308.

2005 Vera Gheno, *Mini-compendio sulla lingua dei newsgroup*, in «Italiano Accessibile», online alla pagina <http://www.italianoaccessibile.it/detail.asp?idn=2871>.

GRAFFI

2008 Giorgio Graffi, *Recensione a Leo Spitzer, Lingua italiana del dialogo*, a cura di Claudia Caffi e Cesare Segre, trad. di Livia Tonelli, Milano, il Saggiatore, 2007, in "Incontri linguistici" XXXI (2008) 196-201.

GRICE

1989 [Herbert] Paul Grice, *Studies in the Way of Words*, Cambridge (MA), Harvard University Press, 1989.



GÜNTHER - LUDWIG

1994 *Schrift und Schriftlichkeit | Writing and Its Use*, herausgegeben von | edited by Hartmut Günther und | and Otto Ludwig, Berlin, de Gruyter, 1994.

HEALEY

1993 Christopher Healey, *Folk Taxonomy and Mythology of Birds of Paradise in the New Guinea Highlands*, in «Ethnology» XXXII (1993) 19-35.

HEID

2007 Ulrich Heid, *Il Corpus WorkBench come strumento per la linguistica dei corpora. Principi ed applicazioni*, in BARBERA - CORINO - ONESTI 2007a, pp. 89-108.

HINRICHS *et alii*

1995 *Abschlußbericht [zu ELWIS Projekte]*, Projektleiter Prof Dr Erhard W. Hinrichs, Mitarbeiter Helmut Feldweg, Marie Boyle-Hinrichs und Ralf Hauser, PS file online <http://www.sfs.uni-tuebingen.de/Elwis/abschlussbericht.ps>.

HOFMANN

1926 Johann Baptist Hofmann, *Lateinische Umgangssprache*, Heidelberg, Carl Winter, 1929 "Indogermanische Bibliothek" I.1.17.

1929 Johann Baptist Hofmann, *Der Begriff Umgangssprache*, in «Indogermanische Forschungen» XLVII (1929) 209-213.

HOLTUS - RADTKE

1985 *Gesprochenes Italienisch in Geschichte und Gegenwart*, herausgegeben von Günter Holtus und Edgar Radtke, Tübingen, Gunter Narr Verlag, 1985 "Tübinger Beiträge zur Linguistik" Bd. 252

KIESLER

2006 Reinhard Kiesler, *Einführung in die Problematik des Vulgärlateins*, Tübingen, Niemeyer, 2006.

KOCH - OESTERREICHER

1994 Peter Koch - Wulf Oesterreicher, *Funktionale Aspekte der Schriftkultur*, in GÜNTHER - LUDWIG 1994, pp. 587-604.

KORZEN - LUNDQUIST

2007 *Comparing Anaphors between Sentences, Texts and Languages. Proceedings of the international symposium held at the Copenhagen Business School, September 1<sup>st</sup>-3<sup>rd</sup> 2005*, edited by Iørn Korzen and Lita Lundquist, Frederiksberg, Samfundslitteratur Press, 2007 "Copenhagen Studies in Language" 34.

MARELLO

2007 Carla Marellò, *Does Newsgroups "Quoting" Kill or Enhance Other Types of Anaphors?*, in KORZEN - LUNDQUIST 2007, pp. 145-157.

MITCHELL

1997 Tom M[ichael] Mitchell, *Machine Learning*, New York, McGraw-Hill, 1997 "McGraw-Hill Series in Computer Science".

MONEGLIA - PANUNZI

2010 *Bootstrapping. Information from Corpora in a Cross-Linguistic Perspective*, edited by Massimo Moneglia and Alessandro Panunzi, Firenze, Firenze University Press, 2010.

NENCIONI

1983/1976 Giovanni Nencioni, *Parlato-parlato, parlato-scritto, parlato-recitato*, in «Strumenti critici» XXIX (1976) 1-56; poi in NENCIONI 1983, pp. 126-179.

1983 Giovanni Nencioni, *Di scritto e di parlato. Discorsi linguistici*, Bologna, Zanichelli, 1983 "La parola letteraria" 6.

NESI - DE MARTINO

2012 *Lingua italiana e scienze. Atti del convegno internazionale, Firenze, Villa Medicea di Castello, 6-8 febbraio 2003*, a cura di Annalisa Nesi e Domenico De Martino, Firenze, Accademia della Crusca, 2012.

PILLET - CARSTENS

1933 Alfred Pillet - Henry Carstens, *Bibliographie der Troubadours*, Halle a. S., Max Niemeyer, 1933 "Schriften der Königsberger Gelehrten Gesellschaft. Sonderreihe" 3.

RICOLFI

2007 Marco Ricolfi, *Il terribile diritto. La proprietà intellettuale: un incentivo od un ostacolo all'innovazione ed alla creatività?*, in BARBERA - CORINO - ONESTI 2007a, pp. xj-xiiij.

SABATINI

1985 Francesco Sabatini, *L' "italiano dell'uso medio": una realtà tra le varietà linguistiche italiane*, in HOLTUS - RADTKE 1985, pp. 154-184.

SAPIR

1916/94 E[dward] Sapir, *Time Pespective in Aboriginal American Culture, a Study in Method*, Ottawa, Government Printing Bureau, 1916 "Canada Department of mines. Geological survey" memoirs 90, anthropological series 30. Poi in *The Collected Works of Edward Sapir*, Volume IV. *Ethnology*, edited by Regna Darrell and Judith Irvine, Berlin - New York, Mouton de Gruyter, 1994, pp. 31-120.

SCHMID

1994 Helmut Schmid, *Probabilistic Part-of-Speech Tagging Using Decision Trees*, paper presented at the *International Conference on New Methods in Language Processing, Manchester (UK), 1994*; versione revisionata PS/PDF online sul sito dell'IMS Stuttgart: <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html>.

SPITZER

1922 Leo Spitzer, *Italienische Umgangssprache*, Bonn - Leipzig, Kurt Schroeder Verlag, 1922 "Veröffentlichungen des romanischen Auslandsinstituts der rheinischen Friedrich Wilhelms-Universität Bonn" 1.

2007/1922 Leo Spitzer, *Lingua italiana del dialogo*, a cura di Claudia Caffi e Cesare Segre, traduzione [di SPITZER 1922] di Livia Tonelli, Milano, il Saggiatore, 2007.

WETZEL

2009 Linda Wetzel, *Types and Tokens: on Abstract Objects*, Cambridge (Massachusetts) - London (England), the MIT Press, 2009.

ZANNI

2007 Samantha Zanni, *Corpora elettronici e copyright. Lo stato legale della questione*, in BARBERA - CORINO - ONESTI 2007a, pp. 119-126.

CORPORA E SITI DI RIFERIMENTO.

20 Newsgroups <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/data/news20.html>  
<http://kdd.ics.uci.edu/databases/20newsgroups/20newsgroups.data.html>

bmanuel.org <http://www.bmanuel.org>

corpora.unito.it <http://www.corpora.unito.it/>

CWB <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/>

ELWIS <http://www.sfs.uni-tuebingen.de/Elwis/>

GNU <http://www.gnu.org>

Google Groups <http://groups.google.it/>

## Indice.

<i>Presentazione di Franco Crevatin</i> .....	5
Introduzione .....	7
1. <i>Tassonomia, filogenesi ed altro: la classificazione linguistica del Nordamerica</i> .....	9
2. <i>Per una soluzione teorica e storica dei rapporti tra grammatica generativa e linguistica dei corpora</i> .....	27
3. <i>Anafora e deissi in diacronia: il caso del voto</i> .....	47
4. <i>Una introduzione ai NUNC: storia della creazione di un corpus</i> .....	97
5. <i>Il Prete Gianni ed i kitan neri: una nota</i> .....	115





Finito di stampare  
nel marzo 2013  
da Status S.r.l. - Genova