

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/>

Bmanuel
ORG

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)¹ 198.

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.

Per una soluzione teorica e storica dei rapporti tra grammatica generativa e linguistica dei corpora*.

*Se il mondo ha la struttura del linguaggio
e il linguaggio la forma della mente
la mente con i suoi pieni e i suoi vuoti
è niente o quasi e non ci rassicura.*

*Così parlò Papirio. Era già scuro
e pioveva. Mettiamoci al sicuro
disse e affrettò il passo senza accorgersi
che il suo era il linguaggio del delirio.*

Eugenio Montale, *Diario del '71, La forma del mondo.*

*Della filosofia che ne fai? Tutti vaniloqui, chiacchiere, per citare poi
nomi insoliti Kotarbinski, Tarski, Wittgenstein, marche, sembrano, di
pianoforti.*

Antonio Pizzuto, *Ravenna, 1962, [capo 10].*

Vent'anni fa Charles Fillmore iniziava un suo articolo con le seguenti parole:

[...] These two [armchair linguist and corpus linguist] don't speak to each other very often, but when they do the corpus linguist says to the armchair linguist, "Why should I think that what you tell me is true?", and the armchair linguist says to the corpus linguist, "Why should I think that what you tell me is interesting?" [...] I have two major observations to make. The first is that I don't think there can be any corpora, however large, that contain information about all of the areas of English lexicon and grammar that I want to explore; all that I have seen are inadequate. The second observation is that every corpus that I've had a chance to examine, however small, has taught me facts that I couldn't imagine finding out about in any other way. My conclusion is that the two kinds of linguists need each other. Or better, that the two kinds of linguists, wherever possible, should exist in the same body. (FILLMORE 1992, p. 35).

Sembrerebbero conclusioni talmente ragionevoli e di buon senso che ci si sarebbe aspettato che qualsiasi persona, non dico linguista, dovrebbe essersi trovato d'accordo.

Ed invece così non fu. Come mai?

A questa domanda¹ (ed ad altre consimili) cercherà di rispondere il presente articolo offrendo alcune interpretazioni e proposizioni.

* Mi corre l'obbligo di ringraziare, per ragioni diverse, ma tutte valide, nel passato, nel presente e (spero a lungo) nel futuro, Alfredo Rizza, Angela Ferrari, Eva Cappellini, Federica Venier, Giorgio Graffi e Lorenzo Renzi. Naturalmente calanchi, dirupi e precipizi sono di mia libera elezione e responsabilità, e, se ci casco, sono solo fatti (e cocci) miei.

Il contributo era stato presentato a Lugano il 14 settembre 2012 alle *7es Journées suisses de Linguistique. L'empirie en linguistique: variété et complexité des approches. Lugano, Università della Svizzera italiana, 13-14 settembre 2012*, organizzate dalla Société Suisse de Linguistique | "Schweizerische Sprachwissenschaftliche Gesellschaft, e ne è stato depositato un PDF nell'*Archive* della Société Suisse de Linguistique (SSG/SSL): <http://www.sagw.ch/fr/ssg/taetigkeiten/7e-Giornate-svizzere-della-Linguistica.html>.

¹ Una via differente, ma non priva di analogie, è stata quella perseguita nell'Accademia ungherese delle scienze (MTA: *Magyar Tudományos Akadémia*), dove però, invece di porsi storiograficamente il problema, si è puntato, partendo dalla considerazione dei dati, empirici vs. intuitivi (a séguito della questione sulla (a)grammaticalità che aveva fatto esplodere Sampson 1987/2001: cfr. KERTÉSZ - RAKOSI 2008), alla formulazione di un nuovo modello di ricerca linguistica, il

Che letteralmente Chomsky abbia detto che da un corpus² non si può imparare nulla, come spesso si riporta, è certo possibilissimo, ma almeno non l'avrebbe letteralmente scritto (od almeno io non ne ho trovato il luogo); però ha detto e scritto qualcosa che vi va molto vicino, come lo spesso citato passo seguente:

Any natural corpus will be skewed. Some sentences won't occur because they are obvious, others because they are false, still others because they are impolite. The corpus, if natural, will be so wildly skewed that the description would be no more than a mere list. (CHOMSKY 1962/58, p. 159).

Ed indubitabilmente, come scrisse LEECH 1991, p. 8, «Chomsky had, effectively, put to flight the corpus linguistics of the earlier generation». E le cose non sembrano essere molto cambiate se ancora nel 2004 ha ripetuto che «Corpus linguistics doesn't mean anything» (ANDOR 2004, p. 97).

Sembrerebbe, quindi, non esserci molto spazio per discussioni, ma in realtà non solo la posizione di Chomsky è molto più articolata di quanto queste estrapolazioni facciano supporre, ma la chomskyana non è neppure la sola posizione esistente nel generativismo, neanche nella “mainstream generative grammar”.

Ora basta constatare che non è inverosimile che si sia, specie in ambienti anglofoni, assistito ad una vera costruzione della linguistica dei corpora come una sorta di antigerativismo radicale; posizione che ha trovato la sua storica codifica nel classico manuale di Tony Mc Enery ed Andrew Wilson del 1996 (cfr. pp. 5-20; cito dalla seconda edizione del 2001): non vi mancano molte osservazioni assennate e perfettamente condivisibili, ma il generale inquadramento della vicenda in una opposizione razionalismo vs. empirismo da manuale di filosofia del liceo è teoreticamente insoddisfacente; e l'enfasi posta (posizione liminare e spazio accordabile) conduce a ridurre la linguistica dei corpora ad una mera risposta al generativismo, “finalmente sgombra di teorie”. Come se ciò fosse mai possibile o comunque espedito ...

Di altro spessore sono state le teorizzazioni di Geoffrey Sampson (soprattutto tra 1997 e 2001; ma, come vedremo, anche prima), in cui il richiamo alla tradizione empirista discende rettamente dalla lettura di Hume data da Popper (Popper 1975/72, capitolo 1 *Conoscenza congetturale: la mia soluzione al problema dell'induzione*), e come tale è metodologicamente affatto irreprensibile. Nonostante la caricatura che Chomsky insiste nel darne (cfr. ANDOR 2004, p. 97), epistemologicamente il ricorso di Sampson al metodo della *prova ed errore* ed al fallibilismo popperiano è altrettanto giustificabile di quello al modello hempeliano, strettamente nomologico-deduttivo, cui implicitamente si richiama Chomsky; ed i due modelli non sono affatto incompatibili, come Popper con la *Logica della scoperta scientifica* ben dimostra e testimonia. E come le argomentazioni samsoniane contro Steven Pinker, abilissimo divulgatore, ma “gridato” ed estremista, sono assolutamente convincenti, pure contro Chomsky *lui-même* (di ben diversa statura e spessore) o contro ogni possibile forma di generativismo non sempre colgono l'obiettivo.

Le ragioni vere di questa contrapposizione sono complesse, legate a più fattori, ma (almeno in parte) credo neutralizzabili nei fatti proprio seguendo le linee di Fillmore da cui siamo partiti. In buona parte questa neutralizzazione ha già, mi sembra, avuto luogo, ma questo non dispensa dal porsene il problema, storiograficamente e teoricamente (che è in realtà più una dittologia sinonimica che una reale duplicità di atteggiamenti).

cosiddetto *p-model*, i cui risultati, assai interessanti e non necessariamente in contraddizione con quanto qui argomentato, sono esposti nel recente KERTÉSZ - RÁKOSI 2012.

² Per il trattamento dei forestierismi è alla proposta globale di BARBERA 2009 che ci atteniamo nel presente contributo, con tutte le conseguenze pratiche che ciò implica: «Tondo (invariabile) o corsivo (con plurale in -s)? Prestito non adattato (ma comunque accettato, fosse anche *faute de mieux*) o fastidioso termine straniero se non da puristicamente evitare almeno da porre nella quarantena del corsivo?» (BARBERA - MARELLO 2009, n. 3).

Un primo fattore è legato ad un fraintendimento, banale, ma assai diffuso tra i linguisti non generativi, che spesso si fermano al *per graecum non legitur*: che l'oggetto studiato dalle "due linguistiche" sia il medesimo (perché entrambe si chiamerebbero *linguistica*, sennò?). Invece un conto è essere interessati alle lingue come organismi storici, sociali e culturali concreti (chiamiamole col consueto termine saussuriano di *langues*), come l'italiano, il thai od il navajo; ed un conto è essere invece interessati alle facoltà umane del linguaggio (si intendano *I-language*, come nella tradizione generativa recente, cfr. CHOMSKY 2000, o *grammaire générale* come nelle tradizioni settecentesche, cui non a caso Chomsky si ricollega esplicitamente, cfr. CHOMSKY 1966). Fraintendimento³ che certo non va riportato alla porta dei chomskyani, ché anzi, anche recentemente, Chomsky ha avuto su ciò parole chiarissime:

So people want to study performance or discourse or social interactions, or national languages or whatever — I don't see how there could be any issue about it — study anything that's worth studying. However, the study of I-language is a very specific topic. It's the study of a language as a part of human biology, trying to find out what it is that each individual has, that enables that individual to participate in larger social interactions, or to perform discourse. There are no conflicts here any more than there's a conflict between a biologist studying the nature of bees and someone who's looking at the study of the way a colony of bees acts, building a hive or finding flowers. (Chomsky in ANDOR 2004, pp. 94-5).

Ogni programma scientifico ha i suoi strumenti specifici (corpora o introspezione, inchieste sociolinguistiche o risonanze magnetiche, ecc.), adeguati al perseguimento dei propri fini (qui uno, l'*I-language*, là molti⁴), che restano diversissimi, ma (giova il ribadirlo) che dovrebbero essere intercomunicanti: la struttura acclarata di una data lingua è materiale di *experimenta crucis* per l'uno, ma strutture cognitive universali eventualmente dimostrate esistenti possono essere un'ipotesi esplicativa di fenomeni singolari di lingue specifiche per l'altro.

A questi diversi, ma ugualmente leciti, programmi di ricerca si accompagnano anche assunzioni filosofiche ben diverse e, queste sì, tra loro mutualmente esclusive: internismo nell'un caso (lungo una linea che parte da Frege⁵, ed oltre, anche se, *allegedly*, da Cartesio) ed esternismo nell'altro (e qui il principale riferimento è a Wittgenstein oltre che all'empirismo logico). Questo è un fattore più rilevante di quanto sembri, perché spesso tali orientamenti "prelinguistici" emergono inconsciamente ed a volte inaspettatamente nelle posizioni dei vari linguisti, portando talora ad atteggiamenti non bene razionalizzabili. Entrambe le prospettive, comunque, sono in sé sostenibili, anche se naturalmente Chomsky propugna un radicale internismo, mentre altri (tra cui pongo me stesso) ritengono che l'esternismo abbia ancora molte carte da giocare ed argomenti a favore, e

³ Chiamare ugualmente *grammatica*, ad esempio ed all'estremo, entrambi i diversissimi (anche se parimenti leciti) oggetti della *Grammar Inference* (cfr. VAN ZAAANEN - ROBERTS - ATWELL 2004) e della *Generative Grammar* è pura follia, probabilmente inevitabile, ma che non può che portare su triste chine.

⁴ Che spaziano tra mille applicazioni pratiche, la *Varietätenlinguistik*, le ricerche sociolinguistiche in genere, l'analisi della conversazione, l'analisi del parlato, e che solo in quanto concernono la descrizione della *langue*, che pure è un punto centrale, possono parzialmente coincidere. Ma anche quando si fanno sostanzialmente altre cose, e quindi non c'è competizione di paradigma teorico, i confronti dei risultati ottenuti dai linguisti di corpora con quelli raggiunti dalla grammatica generativa sono preziosi quando si tratti di interpretare i dati osservati e generalizzati (FERRARI *c.p.*).

⁵ Si badi però, onde non generare inutili confusioni, come giustamente mi fa notare Giorgio Graffi, che: «la posizione di Frege è spesso qualificata da Chomsky come "esternista", nel senso che assume come nozione di base quella di "riferimento", cioè un'entità esterna al linguaggio, il cui rilievo Chomsky nega per le lingue naturali» (GRAFFI *c.p.*). Qui, meno idiosincriticamente, si intende la coppia terminologica *internismo* : *esternismo* come usuale nella tradizione filosofica, ben riassumibile nell'ottima formulazione di Voltolini: «Taken in their simplest versions, externalism and internalism are the conceptions according to which, pending on the broad vs. the narrow identification of an intentional state, the content of such a state can legitimately be conceived only either as relational or as non-relational respectively. For externalists, the representational content of an intentional state depends on a reality lying outside the subject of such a state. For internalists, no external object or event which lies or occurs outside a subject's brain (or at most its body) is relevant for the individuation of the content of an intentional state» (VOLTOLINI 1998/2002).

l'internismo questioni irrisolte⁶. Ma, appunto, di argomento filosofico e non linguistico si tratta: non deve "sporcare" inutilmente la discussione linguistica, altra è la sede in cui va posto.

Una questione, poi, strettamente legata alla precedente⁷, ma invece già anche linguistica, pure spesso non ben compresa, è l'assoluto realismo⁸ della teoria generativa (cfr. invece GRAFFI 2001, pp. 352-4), che non è un'ipotesi descrittiva ma una teoria esplicativa: è una teoria su come materialmente funziona la nostra mente (che in ciò abbia successo o meno è il gioco della scienza dimostrarlo, e non è qui pertinente) non su come la si possa al meglio descrivere. Simili rivoluzionarie⁹ assunzioni di solito non vengono fatte dai non generativisti, che siano più funzionalisti¹⁰ o strutturalisti o computazionali: e la questione (meta)linguistica ha non poco peso nella diversa valutazione soprattutto dei fattori quantitativi e statistici. Ma sgombrato il campo dagli equivoci credo ci si possa alla fine intendere: anche accettando o meno l'argomento realista in sede filosofica, in sede linguistica molto può essere mantenuto saldo.

Altra faccenda è poi che il *fair play* dichiarato nella citazione precedente, in realtà, non è stato spesso praticato dai generativisti più ortodossi e da Chomsky stesso, che anzi non è mai stato alieno dal ciurlare nel manico in generale e nella fattispecie. L'immagine bozzettistica dei due gruppi di ricerca che Fillmore dava come caricaturale *vulgata opinio* è infatti ripetuta questa volta seriamente (e ciò non giova certo alla distensione) anche nelle più recenti affermazioni di Chomsky:

Armchair linguistics does not have a good name in some linguistics circles. A caricature of the armchair linguist is something like this. He sits in a deep soft comfortable armchair, with his eyes closed and his hands clasped behind his head. Once in a while he opens his eyes, sits up abruptly shouting, "Wow, what a neat fact!", grabs his pencil, and writes something down. Then he paces around for a few hours in the excitement of having come still closer to knowing what language is really like. (There isn't anybody exactly like this, but there are some approximations).

Corpus linguistics does not have a good name in some linguistics circles. A caricature of the corpus linguist is something like this. He has all of the primary facts that he needs, in the form of a corpus of approximately one zillion running words, and he sees his job

⁶ Prima di tutte quella della natura pubblica del linguaggio e, correlatamente, della nozione di convenzione: che, se pur sollevano mille problemi, pure non sembrano in alcun modo eliminabili. Chomsky si è sempre solo limitato ad ignorare l'argomento di Wittgenstein contro il linguaggio privato (cfr. ad es. CHOMSKY 1992/2000), che purtuttavia rimane a mio parere rocciosamente ineliminabile (anche il tentativo di aggiramento di FODOR 1975, pp. 55-98 non convinceva). A mia conoscenza, l'unica proposta che "risolverebbe" soddisfacentemente la nozione di convenzione in una prospettiva naturalistica è quello abbastanza recente di Ruth Millikan (cfr. MILLIKAN 1998/2005 e 2003/05), ma Chomsky ha orgogliosamente rifiutato anche questa zattera di salvataggio (CHOMSKY 2003) con argomentazioni che non mi sono invero molto chiare (e che mi sembrano sostanzialmente ammontare alla negazione che una difficoltà esista).

⁷ Filosoficamente, infatti, come in molte (ma non tutte!) *philosophies of mind*, la mossa di base è quella della radicale naturalizzazione dell'intenzionalità; mossa che notoriamente ha i suoi problemi, che non è qui la sede di affrontare. È questa, comunque, la ragione per cui qui ed altrove preferisco parlare di *langue* piuttosto che di *competence*; mettendo però tra parentesi le forti divergenze nel fondamento dei due concetti, molto (ma non tutto, come vedremo in seguito) linguisticamente non cambia. Anche altri aspetti variamente discussi, ed a questi strettamente imbricati, come l'ipotesi innatista, non sono per il nostro ragionamento particolarmente rilevanti.

⁸ Più diffusamente sul realismo chomskyano cfr. STRAWSON 2003. Qui potrebbe bastare intendere "realista" almeno in un duplice senso: (1) gli stati mentali del linguaggio non sono più *réalités psychiques* (come per Saussure) o stati intenzionali (come da BRENTANO 1874 [cfr. pp. 115-6] in poi) ma sono "naturalizzati", come si accennava nella nota precedente, cioè ridotti ad oggetti biologici; (2) la grammatica generativa non si limita a darne una descrizione ma «constitutes an hypothesis as to how the speaker-hearer interpretes utterances» (CHOMSKY 1966, p. 75). Qui è la seconda tesi che ci riguarda. Si avverta che, peraltro, l'uso terminologico di *explanatorily adequate* in questo senso (cioè applicabile ad una grammatica generativa che «offers an explanation for the intuition of the native speaker on the basis of an empirical hypothesis», CHOMSKY 1965, pp. 26-7), e contrapposto a *descriptively adequate*, è tipico della tradizione generativa (a partire da CHOMSKY 1965), e può non risultare perspicuo fuori di questa tradizione.

⁹ Ché tali nella tradizione linguistica occidentale sono, ed è vano, nonché stupido, appiattirle ridimensionandone la portata storiografica. Chomsky insiste molto sulla rivoluzione copernicana attuata dal suo programma nella storia della linguistica occidentale, ed in ciò non esagera molto.

¹⁰ Ben nota è la paradossale (e gustosa) affermazione di Hilary Putnam che «we could be made of Swiss cheese and it wouldn't matter» (PUTNAM 1975/73, p. 291).

as that of deriving secondary facts from his primary facts. At the moment he is busy determining the relative frequencies of the eleven parts of speech as the first word of a sentence versus as the second word of a sentence. (There isn't anybody exactly like this, but there are some approximations). (FILLMORE 1992, p. 35).

Corpus linguistics doesn't mean anything. It's like saying suppose a physicist decides, suppose physics and chemistry decide that instead of relying on experiments, what they're going to do is take videotapes of things happening in the world and they'll collect huge videotapes of everything that's happening and from that maybe they'll come up with some generalizations or insights. Well, you know, sciences don't do this. But maybe they're wrong. Maybe the sciences should just collect lots and lots of data and try to develop the results from them. Well if someone wants to try that, fine. They're not going to get much support in the chemistry or physics or biology department. But if they feel like trying it, well, it's a free country, try that. We'll judge it by the results that come out. So if results come from study of massive data, rather like videotaping what's happening outside the window, fine — look at the results. I don't pay much attention to it. I don't see much in the way of results. (Chomsky in ANDOR 2004, pp. 97).

La citazione precedente, però, con la sua allusione al *videotaping* (che molto assuona col famigerato “metodo audiovisivo” di tanti curricula behaviouristi) ci dovrebbe fornire anche un buon indizio sul *perché* di questi atteggiamenti: il ruolo chiave giocato dalla polemica antibehaviourista nella creazione della teoria generativa, tanto che echi di quella *querelle* continuano a risuonare anche quando il mondo della ricerca è ormai radicalmente cambiato. È una tesi che già ho sostenuto:

E non è un caso, ma anzi un importante fatto nella storia della linguistica, che la parabola linguistica di Chomsky si apra anche, a due soli anni dalle epocali *Syntactic Structures* (CHOMSKY 1957/70), con una veemente (e storicamente mortale) recensione-stroncatura del behaviourismo (CHOMSKY 1959/67), impersonato in un lavoro di Skinner: come se, appunto, fosse proprio il behaviourismo estremo il primo vero nemico con cui la nascente teoria dovesse fare i conti. E, per tracciare la storia della linguistica dell'ultimo mezzo secolo, quella lontana recensione è, a mio parere, un testo che andrebbe messo più in luce di quello che di solito non avvenga: molte delle polemiche tra linguistica empirica (per usare l'ottima etichetta di SAMPSON 2001) e linguistica generativa riproducono in parte quei vecchi schemi, ed avvengono in realtà solo tra le ali più oltranziste dei due schieramenti. (BARBERA 2008, p. 18).

Ma che non è affatto inedita, già Sampson aveva acutamente notato ciò:

The clue to the way Chomsky saddles his opponents with an obviously wrong point of view may lie in one of his earliest writings, which was a slashing review of a book about language by the American psychologist B.E Skinner. Chomsky's criticisms of Skinner were quite fair: Skinner confused the method of psychological research with its object, and suggested that because all the researcher can observe are the stimuli that impinge on a person and his behavioural responses, that is essentially all there is. Skinner scarcely seemed to believe in the reality of a complex mind that often leads a person to behave in ways that are linked only indirectly, if at all, with the stimuli that recently impinged on him. Chomsky's later writings often refer back to Skinner. But to treat Skinner's unreasonable theories as representative of the centuries-old tradition of empiricist thought is a travesty. (SAMPSON 1997, p. 50).

In altre parole: che da comportamenti possano inferirsi stati mentali non è affatto controintuitivo; inaccettabile è che *solo* da comportamenti possano inferirsi stati mentali: l'errore di Chomsky è fare

di ogni behaviourismo un fascio, e la linguistica dei corpora ne paga le penalità. La realtà¹¹ è che dai fatti di *parole* raccolti in un corpus si può risalire ai loro correlati stati di *langue*¹², anche se certamente non tutti gli elementi di una *langue* saranno contenuti in un corpus: è l'uso testimoniato dai corpora, anzi, che fonda la *langue*¹³, anche se i corpora essendo per definizione finiti¹⁴ ne rappresenteranno solo un sottoinsieme.

Che è appunto la posizione, preconizzata da FILLMORE 1992, dell'ala moderata (forse attualmente la più diffusa, ed in cui mi riconosco anch'io) della linguistica dei corpora che si suole oggi definire come *corpus based*; cui si oppone quella più intransigente, degna erede di quella testimoniata da LEECH 1991, p. 8 (cit. *infra*), e normalmente chiamata *corpus driven*¹⁵, il cui più lucido alfiere è stato John McHardy Sinclair (cfr. almeno SINCLAIR 1987 e 1991).

Ad ogni buon conto, la connessione originaria con il behaviourismo della tradizione "anglofona"¹⁶ (americana e britannica) è del resto stata appunto riconosciuta¹⁷ da un protagonista come Geoffrey Leech fin dall'inizio degli anni Novanta:

When did modern corpus linguistics begin? Should we trace it back to the era of post - Bloomfieldian structural linguistics in the USA? This was when linguists (such as Harris and Hill in the 1950s) were under the influence of a positivist and behaviourist

¹¹ Questo ed altri pochi passi sono praticamente identici con quanto ho scritto per i *Dieci anni di linguistica italiana* della SLI, ancora inedito (BARBERA 2012 *i.s.*). Non ne tento isolamenti citazionali perché non di vere citazioni si tratta: sono nati insieme e poi sono stati diversamente riadibiti in due lavori scritti pure essi in contemporanea.

¹² Un'analoga posizione è espressa da FERRARI 2007, p. 59.

¹³ Altra questione, a rigore, sarebbe per le *performances*, che non potrebbero vantare alcun ruolo fondante sulla *competence*, che anzi, al contrario le fonderebbe. Ma questo non comporta negare che si possa, *à rebours*, da stati di *performance* inferire stati di *competence*; certo questi giocherebbero perlopiù un ruolo di controllo dei dati forniti dall'introspezione. Ruolo comunque poi non così secondario, visti i molti usi ed assunzioni impropri che facciamo della nostra competenza, come la letteratura sociolinguistica ha ampiamente dimostrato.

Le due coppie oppositive *competence and performance* e *langue et parole* sono peraltro l'una derivata dall'altra, come Chomsky stesso ammette in *Aspects*, pur riuscendo ad essere, come spesso è suo stile, ingeneroso verso Saussure (per cui la *langue* non è certo «merely a systematic inventory of items»): «The distinction I am noting here [viz. *competence vs. performance*] is related to the *langue-parole* distinction of Saussure; but is necessary to reject his concept of *langue* as merely a systematic inventory of items and to return rather to the Humboldtian conception of underlying competence as a system of generative process» (CHOMSKY 1965, p. 4). Si tenga comunque conto che l'interpretazione, pur in sé sbagliata, che la *langue* fosse un puro inventario di elementi era nel 1965 ancora piuttosto diffusa: cfr. GRAFFI 2010, p. 343.

Data quindi la concezione metalinguistica (cfr. *infra*) della linguistica dei corpora ed il suo fondarsi su dei "testi" concepiti più come atti di *parole* che *performances* (se non nella limitata misura in cui i due concetti sono interscambiabili), il compito più urgente in agenda (un primo utile affondo è VENIER 2007) sarebbe appunto quello di meglio studiare natura e rapporti di *langue et parole*, sia in Saussure *lui-même*, sia *dopo* in Charles Bally (che al Saussure reale spesso si è sostituito; cfr. in prima istanza FERRARI 2007, pp. 61-65), sia *prima* nel menzionato Humboldt, ma non in modo generico, bensì specificamente per quanto riguarda il (sub)archetipo di tutte queste opposizioni (*sub-* perché l'Archetipo di tutti gli Archetipi è sempre, ovviamente, nel vecchio Aristotele), cioè la coppia humboldtiana ἐνέργεια - ἔργον (cfr. ora le acute osservazioni di VENIER 2012; su Humboldt cfr. anche il seminale e fondamentale CONTE 2010/1992).

¹⁴ Anche se sono di dimensione ormai ben maggiore di quanto immaginasse Fillmore vent'anni fa: il corpus itWaC di Marco Baroni, ad esempio, conta ben 1.585.620.279 token: cfr. BARONI *et alii* 2009.

¹⁵ Per la opposizione tra *corpus-based* e *corpus-driven* cfr., sia pure con importanti differenze, TOGNINI - BONELLI 2001, pp. 65-100. Molto alla spiccia e *grosso modo*, ed un po' radicalizzando la distinzione, qui intendiamo la coppia terminologica nel modo seguente: *corpus driven* significa che tutto quello (e solo quello) che il linguista descrive deve essere attinto da un corpus (preferibilmente non etichettato e per pura inferenza statistica): un po' come avveniva per i behaviouristi stretti, skinneriani, non esisterebbero altri oggetti linguistici che quelli compresi in un corpus; *corpus based*, invece, vuol dire che gli oggetti di cui il linguista si occupa sono preferibilmente attinti dai corpora, ma possono provenire anche da altre fonti (e comunque quelli presenti in un corpus non sono gli unici oggetti linguistici esistenti), e che il modo con cui il linguista interroga il corpus non è guidato solo dalla statistica e dal corpus medesimo, ma anche da altre ragioni esterne (introspezione, ecc.) che il linguista può avere.

¹⁶ Nel prosieguo, per comodità, ci riferiremo solo a questa come *corpus linguistics*, ricorrendo invece per ogni altra estensione del termine all'italiano *linguistica dei corpora*.

¹⁷ Salvo poi disonoscere la continuità con la *corpus linguistics* contemporanea e da lui praticata, che è sostanzialmente *corpus based*. Ma gli scheletri nell'armadio restano e sono altamente diagnostici.

view of the science, and regarded the ‘corpus’ as the primary explicandum of linguistics.
(LEECH 1991, p.8)

È ragionevole dunque che la tradizione italiana, dove questo nesso non sussiste, e che anzi ha tutt’altra storia, possa sfuggire agli estremismi di questa diatriba.

Tutt’altra storia, dicevo. Infatti Sabatini ha ripetutamente argomentato (a partire da SABATINI 2006 e 2007) e solidamente dimostrato che il procedimento *corpus based* e l’idea che la norma¹⁸ si ricavi dall’uso stanno alla base della storia linguistica italiana stessa, visto che il *Dizionario* della Crusca, che di quella tradizione rappresenta un momento fondamentale, è proprio stato costruito a partire da testi: il ricorso alla voce dei testi, cioè a dei particolari atti di *parole*, è caratteristico e di fondazione per la definizione della norma stessa della nostra lingua, per cui ha avuto più peso l’uso degli scrittori delle introspezioni dei grammatici. La nostra, quindi, è una tradizione in cui l’“empirismo” anziché essere declinato all’inglese ed in versione behaviouristica è molto più simile alla linea wittgensteiniana (cfr. BARBERA - MARELLO 2008) del ricorso all’uso (*parole*) come realtà fondante di ogni lingua (*langue*); e Saussure non diceva poi cose molto diverse. Da noi non dovrebbe quindi innescarsi quella miccia (anti)behaviourista che è stata fatale negli *States* ed altrove.

Oltre ciò, io ho più volte (cfr. soprattutto BARBERA 2009, ¶ 2 pp. 15-27, e 2011) cercato di disegnare questa “linea italiana” alla linguistica dei corpora¹⁹, ravvisandovi non tanto un elemento di discontinuità e rottura come vorrebbe la tradizione anglosassone (di rivoluzione, in realtà, c’è stata solo quella generativa), quanto piuttosto di continuità con una diversa ma fondamentale tradizione: quella della linguistica filologica di fine Ottocento - primi Novecento. «Specie – dicevo in BARBERA 2009, p. 23 – con le sue manifestazioni più strutturalmente consapevoli, in alcuni casi già presaussuriane: si pensi, ad esempio, da *côté* neogrammatico, ad un Carlo Salvioni²⁰ (così moderno ma classe 1858, e che sopravviverà a Saussure sei anni soli), o da altro *côté*, ad un Hugo Schuchardt, la cui attualità e grandezza non è ancora oggi stata appieno rivendicata», se non assai recentemente, specie da VENIER 2012. E procedevo così a definirne la posizione, puntandone proprio sulla operosa “continuità”:

Ma si può, a mio parere, innovare e contribuire a costruire nuove conoscenze anche lavorando all’interno del solco di una tradizione: posizione che, se mi si consente il paragone extra-epistemologico, era stata fatta perfettamente chiara nel campo della storia della musica da Schönberg con il suo *Brahms il progressivo* del ’33, il cui intento era «dimostrare che Brahms – il classicista, l’accademico – fu un grande innovatore nella sfera del linguaggio musicale. Che, in realtà, fu un grande progressivo» (SCHÖNBERG 1933/50/60, p. 60).

Spesso si tende, infatti, a pensare il progredire di una disciplina solo nei termini di “rivoluzioni” e drastici cambi di paradigma di kuhniana memoria; ma in realtà ciò spesso avviene tramite un più lento e meno appariscente accumulo di esperienze, gradualisticamente, grazie al lento e “nascosto” lavoro fuori dalle luci della ribalta.

Il ruolo della linguistica dei corpora, almeno nella nostra accezione, è un po’ questo: innovazione nella tradizione. E se si dovesse tentare una storiografia linguistica del Novecento, è senz’altro vero che un ruolo di primo piano andrebbe assegnato alla rivoluzione generativa, ma accanto ad essa

¹⁸ Questo naturalmente non implica che il concetto di *norma* abbia a che fare con alcunché della grammatica generativa, ma solo che nella tradizione italiana finanche la costruzione della grammatica normativa è più legata a procedure empiriche, *corpus based*, di quanto normalmente non capiti.

¹⁹ Ed al di là della differente fondazione e della ancora maggiore antichità la linguistica dei corpora italiana ha ormai raggiunto una ricchezza di risultati che spesso poco ha ad invidiare alla tradizione anglofona: cfr. BARBERA 2012 *i.s.*

²⁰ E precisavo in nota: «Sulla cui produzione può oggi il lettore moderno gettare facilmente uno sguardo complessivo grazie alla meritoria edizione del centocinquantenario allestita dal Canton Ticino (SALVIONI 2008). Già CONTINI 1961/72 aveva, da par suo, definito la modernità del Salvioni, ben vedendone il suo quasi strutturalismo *avant lettre*; inquadramento recentemente arricchito ed aggiornato dall’ottimo LOPORCARO 2008. E che in questo apprezzamento non abbia parte alcun schieramento ideologico (non ci confondano in ciò le famigerate astiose querele dell’ormai senile Ascoli) lo mostrano i plausi tributati al “neogrammatico” Salvioni anche dagli “idealisti” Spitzer e Terracini (cfr. TERRACINI 1922 e SPITZER 1929-30)» (BARBERA 2009, p. 23).

esistono altre trame (linguistica storica e strutturalismo *in primis*) la cui persistenza è rilevante: non solo hanno diritto ad esistere ma possono ben rivendicare anche la loro importanza; e tra queste la *corpus linguistics*, od almeno la nostra versione di essa, potrebbe essere, appunto, il Brahms della situazione.

Il paragone schönberghiano di cui sopra può anche essere spinto più in là: la grossa contrapposizione che segna tutta la metà dell'Ottocento tra wagneriani e brahmsiani, "giovani tedeschi" rivoluzionari ed innovatori e "classicisti" conservatori e tradizionalisti, è un artefatto, una montatura polemica, non rispecchiato dalla realtà dei fatti: e che gli stessi protagonisti di quegli anni la sopportassero come tale, oggi ben sappiamo dalla pubblicazione di molti epistolari (come quello di Brahms: cfr. ad es. AVINS 1997), da studi biografici accurati come quello di Alan Walker su Liszt, dalla migliore conoscenza di figure "intermedie" come Joseph Joachim Raff, oltre che dalle argomentazioni strettamente musicali accampate per la prima volta da Schönberg. (BARBERA 2009, p. 23)

Parimenti, non c'è dubbio che la prima mossa di dialogo sia avvenuta proprio dal campo generativo italiano, con l'importante uscita di RENZI 2002, critica, certo, ma finalmente costruttiva. Non tutto è ancora a posto, e le critiche (spesso fondate) colgono soprattutto la pratica *corpus driven*²¹, non la *corpus based* da me avocata (ed oggi prevalente); ma bisogna dare atto a Renzi di incarnare una buona approssimazione al linguista che Fillmore vagheggiava. In apertura della sua *Autobiografia linguistica*, infatti, Renzi ben scriveva che «sono convinto, d'altra parte, che l'introspezione non escluda il ricorso ai corpora, come talora si crede: c'è ogni convenienza a integrare le due fonti» (RENZI 2000/2/08, p. 3); il che pienamente sottoscriviamo²². Vale dunque la pena la pena di scendere nei dettagli²³ ed esaminare più minutamente le principali obiezioni renziane.

In primo luogo, è irrisa la pretesa della *corpus linguistics* di classificare la grammatica generativa «tra i vari approcci "tradizionali" alla lingua» (RENZI 2002, p. 273); ed in ciò, come dicevamo prima, concordiamo appieno: si tratta di una errata percezione storica di certa *corpus linguistics* (come quella codificata nel manuale di MCENERY - WILSON 1966), di cui avevamo anche ipotizzato le ragioni, ma non certo della linguistica dei corpora tutta, o perlomeno non di quella da noi avocata e che abbiamo, anzi, schönberghianamente, caratterizzato come "conservatrice progressista".

In secondo, si nega che la linguistica dei corpora sia una vera «teoria della lingua indotta dall'uso del corpus» (RENZI 2002, p. 273). In parte ciò è vero, ma abbisogna di alcuni distinguo. È tipico di certa linguistica *corpus driven* (dove spesso serpeggia un mal compreso mito empirista), ma la cosa è assai diffusa anche in altri settori della linguistica moderna²⁴, il pensare di poter completamente fare a meno di teorie quali si voglia, bastando collezionare fatti, neanche già i "fatti"²⁵ fossero immuni dalle costruzioni teoriche. Ciò naturalmente è falso, e quindi Renzi ha ragione. Però la mia linguistica *corpus based* ateorica certo non è (cfr. BARBERA 2009, ¶ 2, pp. 15-27), e quindi Renzi avrebbe torto. Non del tutto, però. Perché penso che, come nella tradizione generativa spesso si suole, sottintenda una teoria *esplicativa*, cioè *realista*, del funzionamento del linguaggio (cfr. *supra*, in nota); il che senz'altro nessuna linguistica dei corpora è, neanche la mia, avendone anzi rivendicato proprio il carattere rigorosamente, ed orgogliosamente, metalinguistico (cfr. BARBERA 2011/08).

²¹ Non è infatti casuale la frequenza dei riferimenti a Sinclair.

²² E la sinergia delle due imprese del *Corpus Taurinense*, cfr. BARBERA 2008, ed *ItalAnt*, cfr. RENZI - SALVI 2009, pienamente dimostrano ciò coi fatti. Quella che qui tento, infatti, non è la definizione di qualcosa che ancora non c'è, ma la giustificazione storiografica di quello che in Italia (a fronte di tante sciagure e vergogne pubbliche) è fortunatamente avvenuto e sta avvenendo.

²³ Potrà parere che mi dilunghi un po', ma penso sia utile; e poi così finalmente esaudisco una vecchia promessa che feci a Cino ormai 10 anni fa: meglio tardi che mai, soprattutto se col tempo maturano le nespole ...

²⁴ Basti pensare a tanta tipologia, il cui status epistemologico è perlomeno dubbio.

²⁵ Come usualmente ed esternisticamente intesi, non già i "concetti ingenui" di graffiana memoria (cfr. GRAFFI 1991); non si possono costruire teorie, positivisticamente, con i soli "fatti": un fatto più un fatto sono due fatti, non una teoria. E poi, come giustamente si chiede CREVATIN 2009b, p. 7, «ma quando un fatto è un fatto, quando un problema è un problema?».

In terzo luogo, su concordanze e lemmatizzazione Renzi (cit., pp. 273-274) ha poco da ridire, salvo sottolinearne la discendenza da pratiche “antiche” o comunque precomputazionali; nel che certo concorro, solo che poi attribuisco a ciò una precisa valenza storiografica, giusta la mia teoria “continuista”.

È sul tagging (e quattro), semmai, che, pur riconoscendone l'utilità, nutre più perplessità (RENZI 2002, pp. 274-5): che, in sé fondate, sono da riportare alla sua concezione realista dei fatti grammaticali; una volta accettata la posizione metalinguistica che ho dimostrato e storicizzato in BARBERA 2011/08, le cose restano tali ma prendono un'altro aspetto. Anzi dalla pratica obbligata del POS-Tagging discendono anche delle conseguenze eticamente ed epistemologicamente positive, come ho mostrato in BARBERA 2011 (specie il § 3, *Obbligatorietà nella scelta del POS-tag*), cui rimando.

In quinto luogo, è sul parsing (RENZI 2002, pp. 275-6) che la situazione si fa più critica: «se penso a qualcuno al quale venga affidato il compito di eseguire un parsing manuale, mi assalgono fortissimi dubbi. [...] In realtà, come il tagging, ma più di questo, il parsing può svolgere solo un lavoro grosso, approssimativo. Ma mentre sono convinto dell'utilità del tagging, sono scettico sull'utilità di avere un corpus provvisto di parsing» (*ibidem*, p. 276). Sulla prima questione non posso che aggiungere: *assalgono anche me*²⁶. Sulla seconda serve però qualche rettifica: la scarsa utilità linguistica²⁷ di un parsing ideale che Renzi lamenta è forse tale per l'italiano, ma molto meno per lingue come l'inglese, in cui il tagging giocoforza ha da essere più sintattico che morfologico per servire a qualcosa, per tacere di lingue ancora più isolanti come il thai od il vietnamita. Comunque, anche a prescindere da questioni interlinguistiche, resta il fatto che parsing automatici vengono comunque fatti e, così come sono, servono effettivamente ad applicazioni pratiche di NLP come il riconoscimento vocale o la traduzione automatica, che (sia pure qual più e qual meno) funzionano nel mondo reale in cui viviamo, e che quindi non giova negare. Come mai, allora? È, come spesso, questione di intendersi: un conto è pensare al parsing come un'analisi accurata (idealmente perfetta) della struttura sintagmatica dell'albero frasale, come vorrebbe un buon linguista, specie se realista generativo; un conto è pensare ad una mera segmentazione²⁸ in dei “costituenti immediati” che si è soliti chiamare, più neutramente, *chunks*, che possono anche non rivestire realtà linguistica (non essere cioè dei sintagmi) ma solo collocazionale (cioè statistica), ed essere “utili” in virtù della loro frequenza; il che ci conduce ai prossimi punti (sei e sette). Certo è che, pur attribuendo al parsing un mero valore metalinguistico, e pur messisi preventivamente

²⁶ Si sarebbe portati a credere che la nostra competenza dovrebbe sicuramente guidarci senza grossi problemi, ma, come riporta Sampson «the wrongness of this idea was established experimentally, at a workshop held in conjunction with the Association of Computational Linguistics annual conference at Berkeley, California, in 1991. Natural-language processing researchers from nine institutions were each given the same set of English sentences and asked to indicate what their respective research groups would regard as the target analyses of the sentences, and the nine sets of analyses were compared. These were not particularly complicated or messy sentences – they were drawn from real-life corpus data, but as real-life sentences go, they were rather well-behaved examples. And the comparisons were not made in terms of the labels of the constituents: the only question that was asked was how far the researchers agreed on the shapes of the trees assigned to the sentences – that is, to what extent they identified the same sub-sequences of words as grammatical constituents, irrespective of how they categorized the constituents they identified. The level of agreement was strikingly low. For instance, only the two subsequences marked by square brackets were identified as constituents by all nine participants in the following example (and results for other cases were similar): *One of those capital-gains ventures, in fact, has saddled him [with [Gore Court]]*» (SAMPSON 2001/00, p. 92). Il che ci riconduce alla nostra idea che lo stabilire un efficiente (nel senso che possa essere eseguito senza difficoltà, in modo anche automatico, non che sia quello psicologicamente vero) sistema di analisi sintattica (nel senso di *parsing*) sia una questione che non abbia a che fare tanto con la *competence* quanto con il costruire metalinguisticamente degli standard adeguati: stabilire, in questo senso, delle metriche fisse è stato il compito primario, da BLACK *et alii* 1991 a ROARKA *et alii* 2006.

²⁷ Ma cfr., tuttavia, le considerazioni generali di Sampson 2003, e gli apparati teorici di NEDERHOF - SATTA 2010.

²⁸ Od anche a dei parsing “depotenziati”, cioè privati delle assunzioni forti (ed impossibili) di un parsing come lo pretenderebbe un generativista, e di fatto quasi equivalenti a dei chunking “potenziati”, come le “collezioni naturalistiche” di alberi che, a mo' di botanico tassonomista, proponeva Sampson («Essentially, I was trying to do for the English language what Linnaeus in the eighteenth century did for botany», come dice sul suo sito, <http://www.grsampson.net/Contribs.html>).

d'accordo sulla metrica della valutazione, resta che non è idealmente possibile un perfetto parsing automatico (e per forza, se è un effetto della statistica), ma al più delle approssimazioni ad esso di qualità e granularità variabili in funzione degli scopi perseguiti; approssimazioni²⁹ che però possono, appunto, già essere più o meno utili, a seconda dei casi e degli scopi. Un buon chunking, che ha minore pretese linguistiche, invece è sempre anche teoricamente possibile.

Quanto infatti, in sesto luogo, alle collocazioni (RENZI 2002, pp. 276-7), che sono storicamente una roccaforte della linguistica dei corpora e che hanno praticamente rivoluzionato la lessicografia degli ultimi due decenni, è uno dei punti in cui lo scritto renziano più abbisogna rettifiche. Se concordo con lui che, da un punto di vista strettamente linguistico, non esistono come unica categoria grammaticale, pure esistono come effetto di distribuzione statistica (cfr. BARBERA 2009, ¶ 18, soprattutto pp. 923-5, dove la questione è stata più dettagliatamente articolata). Solo alcune di esse esisteranno pertanto nel nostro lessico mentale (gli idioms, forse alcune collocazioni ristrette, ecc.) ma altre no. Non è quindi detto (ammesso che sia sempre pratico, condizione cui pure bisognerebbe pensare, specie se si è, ad esempio, dediti ad una attività “applicata” come la lessicografia) che l'introspezione sia sufficiente a cogliere tutto quel che sarebbe utile avere in un dizionario³⁰. In parte ciò è connesso alla svalutazione della *performance* (cui abbiamo già accennato) attuata dai generativisti, laddove un saussuriano esalta la *parole*, con quanto oggi pertiene alla statistica. Un'altra rettifica merita l'affermazione che nell'edificazione del *Vocabolario* della Crusca il lavoro sia stato «fatto sull'introspezione e³¹ sul corpus dell'italiano scritto» (RENZI 2002, p. 277): come abbiamo prima detto, gli studi di Sabatini (che comunque sono successivi a quello di Renzi) hanno definitivamente acclarato in ciò la precedenza del *testo* sulla *introspezione*, e quindi dell'*uso* sulla *norma* (cfr. almeno SABATINI 2006 E 2007).

Il che ci conduce al fondamentale settimo punto, quello sui rapporti tra la “tradizionale”³² linguistica qualitativa e quella quantitativa (e quindi, principalmente alla statistica), prerogativa normalmente considerata la vera *hallmark* della linguistica dei corpora³³. È in effetti questo un terreno su cui, anche teoricamente, la *corpus linguistics* si è spesso impegnata, giungendo talora a sintesi di notevole spessore come KLAVANS - RESNIK 1996. Nonostante un sagace recupero della legge di Zipf (cfr. almeno ZIPF 1929 e 1935/6; legge che però non è vero sia stata del tutto trascurata dai linguisti computazionali, cfr. ad es. l'ottima trattazione che ne dà un importante testo di riferimento, MANNING - SCHÜTZE 1999, § 1.4.3 pp. 23-29) Renzi (op. cit., pp. 277-9) non può che constatare l'irreconciliabilità della linguistica quantitativa con la generativa: e non a torto se già vent'anni fa Geoffrey Sampson (che dei metodi statistici è sempre stato un virtuoso ed un araldo) acutamente riconosceva che «One of the ideas underlying my work is that human languages, as grammatical systems, may be too different from computer languages for it to be appropriate to use the same approaches to automatic parsing» (SAMPSON 1992, p. 424): la metalinguisticità è, ossia, inerente al trattamento statistico dei corpora ed è quindi giocoforza incompatibile con una visione puramente realistica della linguistica. Incompatibile, però, non vuol dire che la metalinguistica non possa servire alla linguistica; se poi si accetta la mia idea che quelli contenuti in un corpus siano degli atti di *parole* piuttosto che delle mere *performances*, tali risultati statistici avrebbero un valore ben maggiore (sarebbero, infatti, gli elementi fondanti della *langue*: il *primum* è la pragmatica³⁴);

²⁹ Ormai abbastanza buone: stante i presupposti metrici di cui si diceva sopra, comunque ormai si passa la soglia dell'85% di correttezza: cfr. CLARCK 2010, pp. 336-7.

³⁰ O che servirebbe dire a qualcuno per insegnargli la lingua; attività nient'affatto secondaria se *insegnare una regola* è indissolubilmente legato al suo senso, come dice Wittgenstein nelle *Ricerche*: «Wo ist die Verbindung gemacht zwischen dem Sinn der Worte »Spielen wir eine Partie Schach!« und allen Regeln des Spiels? – Nun, im Regel-vezeichniss des Spiels, im Schachunterricht [corsivo mio], in der täglichen Praxis des Spielens» (I.197).

³¹ Corsivo dell'autore.

³² Ma se tale qualifica viene attribuita alla linguistica generativa, forse è vero il contrario, tanto per me come per Renzi.

³³ Per una presentazione recente dei metodi statistici attuali nella disciplina, cfr. BARONI - EVERT 2009.

³⁴ Ad un primato della pragmatica in questo senso abbiamo già accennato *supra*: l'idea (fondante per ogni pragmatica linguistica) che *la lingua sia un'azione*, è stata pure ricondotta a Humboldt (cfr. VENIER 2012), che così si troverebbe curiosamente invocato da due diverse tradizioni (generativismo e pragmatica) in consistente disaccordo tra

sennò avrebbero comunque rilevanza pratica nel descrivere una lingua (che anche quello serve), anche se non avrebbero nessuna forza esplicativa (nel senso chomskyano, ben illustrato da GRAFFI 2001, pp. 352-3) e, in tutti i casi, poco ci direbbero dell'*I-Language*. Esiste però una possibilità ben più radicale, e finora mai tentata, ma non lontana come sembrerebbe dal naturalismo chomskyano: che non solo le regole della *langue* ma addirittura quelle della *competence* abbiano un fondamento probabilistico; se proprio vogliamo naturalizzare gli oggetti mentali, perché non dovremmo poi pensare ad una sorta di meccanica dei quanti del linguaggio?

Ma questa è fantalinguistica: per tornare coi piedi sulla terra, direi che questa rassegna degli argomenti renziani potrebbe bastare per chiarificare il problema. Nell'articolo citato segue ancora una stringata critica dei principali *tenet* della "teoria standard" della *corpus linguistics*, molti dei quali abbiamo già esaminato indipendentemente qui anche noi. E visto che a quella parte ho già replicato punto per punto in BARBERA 2009, p. 24, propongo di soprassedere.

La cavalcata è stata ormai lunga, ma spero di avere così dimostrato che una convivenza pacifica delle due comunità scientifiche è ben possibile, nonostante le molte provocazioni da una parte e dall'altra, e che la "via italiana" la rende assai più facile. Un punto che resta da approfondire è la questione di *langue et parole* cui abbiamo appena accennato: ma quello sarà l'oggetto di un prossimo contributo. Qui credo di avere anche mostrato come sia possibile costruire una vera "teoria" della linguistica dei corpora senza cedere a troppe ingenuità e senza essere meramente antitetici alla grammatica generativa (oggetti, programmi e metodologie sono diversi³⁵), che, dopo Saussure, è stata l'unica vera "rivoluzione" in linguistica: "innovazione nella conservazione" deve essere a mio avviso il motto della linguistica dei corpora, e si lascino pure ad altri le rivoluzioni; indispensabili, ma attenzione che anche Robespierre poi ci ha lasciato la testa.

BIBLIOGRAFIA.

AA. VV.

- 2004 *Proceedings of the Workshop on Amazing Utility of Parallel Corpora - LREC 2004*, Lisbon, LREC, 2004.
2006 *Proceedings of the 5th International Conference on Language Resources and Evaluation - LREC 2006*, Genova, LREC, 2006.

ABEILLÉ

- 2003 *Treebanks: Building and Using Parsed Corpora*, edited by Anne Abeillé, Dordrecht - Boston - London, Kluwer Academic Publishers, 2003.

AIJMER - ALTENBERG

- 1991 *English Corpus Linguistics. Studies in Honor of Jan Svartvik*, edited by Karin Aijmer, Bengt Altenberg, London - New York, Longman, 1991.

loro (per Chomsky, cfr. qui soprattutto la nota 13). Naturalmente, alla base di tutto c'è sempre Platone col *Cratilo* («τὸ λέγειν μία τις τῶν πράξεων ἐστίν» 'il dire è una qualche azione', 387b) e nel Novecento moltissimo conta l'eredità di Wittgenstein (cfr. nelle *Ricerche* «Eine Sprache verstehen, heißt, eine Technik beherrschen» I.199, «Darum ist ›der Regeln folgen‹ eine Praxis» I.202, ecc.), sia pure con le mediazioni Bühleriane del caso; d'altronde, il ruolo di *Grund* giocato dalle *Philosophische Untersuchungen* era già stato ben visto in anni lontani da Maria-Elisabeth Conte (cfr. ad es. CONTE 2010/1983, p. 161/95).

³⁵ Il che non significa che a volte alcune delle metodologie sviluppate dall'uno non possano essere usate dall'altro, purché sfrondate delle implicazioni teoriche, direi metafisiche, eventualmente inconciliabili, su cui si fondano (o dicono di fondarsi); anzi, alcune di queste metodologie sono ormai irrinunciabili per la linguistica *tout court*.

ANDOR

2004 Andor József, *The Master and his Performance: An Interview with Noam Chomsky*, in «Intercultural Pragmatics» I (2004)₁ 93–111.

ANTONY - HORNSTEIN

2003 *Chomsky and his Critics*, edited by Louise M. Antony and Norbert Hornstein, Malden (MA) - Oxford, Blackwell Pub., 2003 “Philosophers and their critics” 10.

AVINS

1997 *Johannes Brahms' Life and Letters*, selected and annotated by Styra Avins, Translations by Josef Eisinger and Styra Avins, Oxford - New York, Oxford University Press, 1997.

BALLY

1937/09 Charles Bally, *Traité de stylistique française*, 2. édition, Carl Winter - C. Klincksieck, Heidelberg - Paris, 1937₂ [1909₁; 1951₃, Genève, Georg & cie].

1944/32 Charles Bally, *Linguistique générale et linguistique française*, Paris, Ernest Leroux, 1932; poi Bern, Francke, 1944. Traduzione italiana: *Linguistica generale e linguistica francese*, introduzione e appendice di Cesare Segre, traduzione di Giovanni Caravaggi, Milano, Il Saggiatore, 1963.

BARBERA

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. Ri-elaborazione 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.

2012 Manuel Barbera, *Il neo-Corpus Taurinense e l’arte della query*, comunicazione al *Seminario: sintassi dell’italiano antico e sintassi di Dante. Pisa 14-15 ottobre 1911*, ora in TAVONI 2012, pp. 61-79.

2012 *i.s.* Manuel Barbera, *Linguistica dei corpora*, in *La linguistica italiana all’alba del terzo millennio (1997-2010)*, a cura di Gabriele Iannaccaro, Roma, Bulzoni, in corso di stampa “SLI Società di linguistica italiana”.

BARBERA - CORINO - ONESTI

2007 *Corpora e linguistica in rete*, a cura di Manuel Barbera, Elisa Corino, Cristina Onesti, Perugia, Guerra Edizioni, 2007 “L’officina della lingua. Strumenti” 1.

BARBERA - MARELLO

2008 Manuel Barbera - Carla Marello, *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.

BARONI *et alii*

2009 Marco Baroni, Silvia Bernardini, Adriano Ferraresi and Eros Zanchetta, *The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-crawled Corpora*, in «Journal of Language Resources and Evaluation» XLIII (2009)₃ 209-226.

BARONI - EVERT

2009 Marco Baroni and Stefan Evert, *Statistical Methods for Corpus Exploitation*, in LÜDELING - KYTO 2009, vol. 2., pp. 777-802.

BLACK *et alii*

- 1991 E[zra] Black, S[teve] Abney, D[an] Flickenger, C[laudia] Gdaniec, R[alph] Grishman, P[hil] Harrison, D[an] Hindle, B[ob] Ingria, F[rederick] Jelinek, J[udith] Klavans, M[ark] Liberman, M[itch] Marcus, S[alim] Roukos, B[eatrice] Santorini, and T[omek] Strzalkowski, [*ParsEval*]. *A procedure for quantitatively comparing syntactic coverage of English grammars*, in *HLT '91 Proceedings of the 4th DARPA Workshop on Speech and Natural Language*, Stroudsburg (PA), Association for Computational Linguistics, 1991, pp. 306–311.

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.

BRENTANO

- 1874/1924 Franz [Clemens Honoratus Hermann] Brentano, *Psychologie vom empirischen Standpunkt*, in zwei Banden, Leipzig, Verlag von Duncker & Humblot, 1874. Poi anche 1911 ed infine: *Psychologie vom empirischen Standpunkt*, mit ausführlicher Einleitung, Anmerkungen und Register herausgegeben von Oskar Kraus, Leipzig, Meiner, 1924 e 1928 “Philosophische Bibliothek”. Traduzione inglese (basata sulla 2a ed. di Leipzig, Meiner, 1924-28): *Psychology from an empirical standpoint*, edited by Oskar Kraus, English edition edited by Linda L. McAllister, translated by Antos C. Rancurello, D.B. Terrell and Linda L. McAllister, London, Routledge and Kegan Paul, 1973 “International Library of Philosophy and Scientific method”.

BÜHLER

- 1965/34 Karl Bühler, *Sprachtheorie; die Darstellungsfunktion der Sprache*, Jena, G. Fischer, 1934; poi Stuttgart, Gustav Fischer Verlag, 1965. Cfr. anche l’ed. inglese *Theory of Language: the Representational Function of Language*, translated by Donald Fraser Goodwin, Amsterdam - Philadelphia, J. Benjamins Pub. Co., 1990.

CHOMSKY

- 1957/70 Noam Chomsky, *Syntactic Structure*, Mouton, The Hague, 1957. Versione italiana: *Le strutture della sintassi*, introduzione [traduzione e note] di Francesco Antinucci, Roma - Bari, Laterza, 1974₂ [1970₁] “Universale Laterza” 129.
- 1959/67 Noam Chomsky, *Review of B[urrhus] Frederik Skinner, Verbal Behaviour* (New York, Appleton - Century - Crofts, 1957), in «Language» XXXV (1959) 26-58; poi, con una nuova prefazione (pp. 142-143) anche in JAKOBOVITS - MIRON 1967, pp. 142-sgg. Disponibile anche online alla pagina <http://cogprints.org/1148/0/chomsky.htm>.
- 1962/58 Noam Chomsky, *A Transformational Approach to Syntax*, presentato alla *3rd Texas Conference on Problems of Linguistic Analysis in English*, poi in HILL 1962, pp. 124-158. Anche: *Discussion*.
- 1965/70 Noam Chomsky, *Aspects of the Theory of Syntax*, Cambridge (MA), The MIT Press, 1965; trad. it.: *Aspetti della teoria della sintassi*, in *Saggi linguistici*, vol. II, Torino, Boringhieri, 1970, pp. 39-258.
- 1966 Noam Chomsky, *Cartesian Linguistics. A Chapter in the History of Rationalist Thought*, New York, Harper & Row, 1966; ristampa: Lanham (MD) - New York (NY) - London (EN), University Press of America, 1983.
- 1992/2000 Noam Chomsky, *Explaining Language Use*, in «Philosophical Topics» XX (1992) 205-31, poi in CHOMSKY 2000, pp. 19-45.
- 2000 Noam Chomsky, *New Horizons in the Study of Language and Mind*, Cambridge, Cambridge University Press, 2000.
- 2003 Noam Chomsky, *Reply to Millikan*, in ANTONY - HORNSTEIN 2003, pp. 308-315.

CINI - REGIS

2000/2 *Che cosa ne pensa oggi Chiaffredo Roux? Percorsi di dialettologia perzezionale all'alba del nuovo millennio. Atti del Convegno internazionale (Bardonecchia, 25-27 maggio 2000)*, a cura di Monica Cini e Riccardo Regis, Alessandria, Edizioni dell'Orso, 2002.

CLARCK

2010 Stephen Clark, *Statistical Parsing*, in CLARK - FOX - LAPPIN 2010, pp. 333-363.

CLARK - FOX - LAPPIN

2010 *The Handbook of Computational Linguistics and Natural Language Processing*, edited by Alexander Clark, Chris Fox and Shalom Lappin, Malden (MA) - Oxford (UK), Wiley - Blackwell, 2010 "Blackwell Handbook in linguistics".

CONTINI

1961/72 Gianfranco Contini, *Modernità e storicità di Carlo Salvioni*, in «Archivio storico ticinese» V (1968) 209-18. Poi in CONTINI 1972, pp. 369-386.

1972 Gianfranco Contini, *Altri esercizi (1942-1971)*, Giulio Einaudi Editore, 1972.

CONTE

2010/1983 Maria-Elisabeth Conte, *La pragmatica linguistica*, in SEGRE 1983, pp. 94-128; poi in FILIPPONIO 2000, pp. 11-57; ed infine, col titolo *Pragmatica linguistica* in CONTE 2010, pp. 161-197.

2010/1992 Maria-Elisabeth Conte, *Frammenti di pragmatica humboldtiana*, in «Lingua e Stile» XXVII (1992) 505-521. Poi in CONTE 2010, pp. 323-341.

2010 Maria-Elisabeth Conte, *Vettori del testo. Pragmatica e semantica fra storia e innovazione*, a cura di Federica Venier e Domenico Proietti, Roma Carocci, 2010 "Università" 820.

CORINO - MARELLO - ONESTI

2006 *Atti del XII Congresso Internazionale di Lessicografia, Torino, 6-9 settembre 2006 | Proceedings of the XII EURALEX International Congress. Torino, Italia, 6th-9th September 2006*, a cura di Elisa Corino, Carla Marello e Cristina Onesti, 2 voll., Alessandria, Edizioni dell'Orso, 2006.

COTTICELLI KURRAS

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

CREVATIN

2009a *I luoghi della mediazione. Confini, scambi, saperi*, a cura di Franco Crevatin, Trieste, Deputazione di storia patria per la Venezia Giulia, 2009 "Fonti e studi per la storia della Venezia Giulia. Serie seconda: studi" 18.

2009b *Mediazione, identità, confini*, in CREVATIN 2009a, pp. 7-40.

DE CESARE - FERRARIS

2007 Anna-Maria De Cesare - Angela Ferraris, *Lessico, grammatica, testualità*, Basel, Universität Basel, 2007 = «Acta Romanica Basiliensa» XVIII (2007). Il volume è basato sui contributi presentati al *Convegno Lessico, grammatica e testualità, nell'italiano scritto e parlato, Basilea 17 e il 18 febbraio 2006*.

DOMOKOS - SALVI

2002 *Lingue romanze nel Medioevo. Atti del convegno, Piliscsaba, 22-23 marzo 2002*, a cura di Domokos György e Giampaolo Salvi, in «Verbum. Analecta Neolatina» IV (2002)² 267-526.

- FERRARI
2007 Angela Ferrari, *Grammatica, testo e stylistique de la langue*, in DE CESARE - FERRARI 2007, pp. 53-73.
- FILIPPONIO
2000 *Ricerche praxeologiche*, a cura di Angela Filipponio, Bari, Adriatica Editrice, 2000.
- FILLMORE
1992 Charles J. Fillmore, “*Corpus Linguistics*” or “*Computer-aided Armchair Linguistics*”, in SVARTVIK 1992, pp. 35-60.
- FODOR
1975 Jerry A. Fodor, *The Language of Thought*, New York, Thomas Y. Crowell Company, 1975.
- GRAFFI
1991 Giorgio Graffi, *Concetti ‘ingenui’ e concetti ‘teorici’ in sintassi*, in «Lingua e stile» XXVI (1991) 347-363.
2001 Giorgio Graffi, *200 Years of Syntax. A Critical Survey*, Amsterdam - Philadelphia, John Benjamins Publishing Company, 2001 “Amsterdam Studies in the Theory and History of Linguistic Science. Series III. Studies in the History of the Language Science” 98.
2010 Giorgio Graffi, *Due secoli di pensiero linguistico. Dai primi dell’Ottocento a oggi*, Roma, Carocci, 2010 “Frece”.
- HILL
1962 *Third Texas Conference on Problems of Linguistic Analysis in English: May 9-12, 1958*, edited by A[rchibald] A. Hill, Austin, University of Texas, 1962 “Studies in American English”.
- JAKOBOVITS - MIRON
1967 *Readings in the Psychology of Language*, edited by Leon A. Jakobovits and Murray S. Miron, [Upper Saddle River, NJ], Prentice-Hall Inc., 1967.
- KERTÉSZ - RÁKOSI
2008 Kertész András - Rákosi Csilla, *Conservatism vs. Innovation in the (Un)grammaticality Debate*, in *New Approaches to Linguistic Evidence. Pilot Studies / Neue Ansätze zu linguistischer Evidenz. Pilotstudien*, Frankfurt am Main - etc., Lang, 2008, pp. 61-84.
2012 Kertész András - Rákosi Csilla, *Data and Evidence in Linguistics: A Plausible Argumentation Model*, Cambridge University Press, 2012.
- KLAVANS - RESNIK
1996 *The Balancing Act. Combining Symbolic and Statistical Approaches to Language*, edited by Judith L. Klavans and Philip Resnik, Cambridge (Mass.) - London (England), MIT Press, 1996.
- LEECH
1991 Geoffrey Leech, *The State or the Art in Corpus Linguistics*, in AIJMER -ALTENBERG 1991, pp. 8-29.
- LOPORCARO
2008 Michele Loporcaro, *Carlo Salvioni linguista*, in SALVIONI 2008, vol. V, pp. 45-113.
- LÜDELING - KYTO
2009 *Corpus Linguistics, An International Handbook*, edited by Anke Lüdeling and Merja Kyto, Berlin: Mouton de Gruyter, 2009, 2 volumi.

MANNING - SCHÜTZE

1999 Christopher D. Manning - Hinrich Schütze, *Foundations of Statistical Natural Language Processing*, Cambridge (Massachusetts) - London (England), The MIT Press, 2000³ [1999₁].

MCENERY - WILSON

2001/1996 Tony McEnery - Andrew Wilson, *Corpus Linguistics. An Introduction*, Edinburgh, Edinburgh University Press, 2001₂ [1996₁, 2005¹] “Edinburgh Textbooks in Empirical Linguistics”.

MEIJS

1987 *Corpus Linguistics and Beyond. Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora [ICAME 7]*, edited by Willem Meijs, Amsterdam, Rodopi, 1987.

MILLIKAN

1998/2005 Ruth Garrett Millikan, *Language Conventions Made Simple*, in «Journal of Philosophy» XCV (1998)⁴ 161-180, poi in MILLIKAN 2005, pp. 1-23.

2003/05 Ruth Garrett Millikan, *In defense of Public Language*, in ANTONY - HORNSTEIN 2003, pp. 215-37, poi in MILLIKAN 2005, pp. 24-52.

2005 Ruth Garrett Millikan, *Language: A Biological Model*, Oxford, Clarendon Press, 2005.

NEDERHOF - SATTA

2010 Marc-Jan Nederhof - Georg Satta, *Theory of Parsing*, in CLARK - FOX - LAPPIN 2010, pp. 105-30.

PINKER

1994/95 Steven Pinker, *The Language Instinct. The New Science of Language and Mind*, New York, William Morrow, 1994, poi London - ecc., Penguin Books, 1995.

POPPER

1935/59 Karl R[aimund] Popper, *Logik der Forschung; zur Erkenntnistheorie der modernen Naturwissenschaft*, Wien, J. Springer, 1935 “Schriften zur wissenschaftlichen Welt-auffassung” 9. Poi *The logic of Scientific Discovery*, London, Hutchinson, 1959.

1975/72 Karl R[aimund] Popper, *Conoscenza oggettiva. Un punto di vista evoluzionistico*, Introduzione traduzione e commento di Arcangelo Rossi, Roma, Armando Armando Editore, 1975. Edizione originale: *Objective Knowledge. An Evolutionary Approach*, Oxford, Clarendon Press, 1972.

PUTNAM

1975/73 Hilary Putnam, *Philosophy and Our Mental life*, paper presented at a *Foerster Symposium on Computers and the Mind* at the University of California (Berkeley) in October 1973. Poi edito in PUTNAM 1975, pp. 291-303.

1975 Hilary Putnam, *Mind, Language and Reality. Philosophical Papers II*, Cambridge (UK), Cambridge University Press, 1975.

1997 Hilary Putnam, *A Half Century of Philosophy. Viewed from Within*, in «Dedalus» CXXVI (1997)1 175-208.

RENZI

2000/2/08 Lorenzo Renzi, *L'autobiografia linguistica in generale, e quella dell'autore in particolare, con un saggio di quest'ultima*, in CINI - REGIS 2000/2, pp. 329-339, poi in RENZI 2008, pp. 3-16.

2002 Lorenzo Renzi, *Il progetto ItalAnt e la “grammatica del corpus”*, in DOMOKOS - SALVI 2002, pp. 271-294.

2008 Lorenzo Renzi, *Le piccole strutture. Linguistica, poetica, letteratura*, a cura di Alvise Andreose, Alvaro Barbieri, Dan Octavian Cepraga, con la collaborazione di Marina Doni, Bologna, Società editrice il Mulino, 2008.

RENZI - SALVI

2010 *Grammatica dell'italiano antico*, a cura di Giampaolo Salvi e Lorenzo Renzi, 2 voll., Bologna, il Mulino, 2010.

ROARKA *et alii*

2006 Brian Roarka, Mary Harper, Eugene Charniak, Bonnie Dorr, Mark Johnson, Jeremy G. Kahn, Yang Liuf, Mari Ostendorf, John Hale, Anna Krasnyanskaya, Matthew Lease, Izhak Shafran, Matthew Snover, Robin Stewart, Lisa Yung, *SParsEval: Evaluation Metrics for Parsing Speech*, in AA. VV. 2006, pp. 333-338.

SABATINI

2006 Francesco Sabatini, *La storia dell'italiano nella prospettiva della corpus linguistics*, in CORINO - MARELLO - ONESTI 2006, pp. 31-37.

2007 Francesco Sabatini, *Storia della lingua italiana e grandi corpora. Un capitolo di storia della linguistica*, in BARBERA - CORINO - ONESTI 2007a, pp. xij - xvj.

SALVIONI

2008 Carlo Salvioni, *Scritti linguistici*, a cura di Michele Loporcaro, Lorenza Pescia, Romano Broggin, Paola Vecchio, [Locarno], Edizioni dello Stato del Canton Ticino, 2008. ¶ Vol. I. *Saggi sulle varietà della Svizzera italiana e dell'Alta Italia*. ¶ Vol. II. *Dialettologia e linguistica storica*. ¶ Vol. III. *Testi antichi e dialettali*. ¶ Vol. IV. *Etimologia e lessico*. ¶ Vol. V. *Introduzione e Indici*.

SAMPSON

1987/2001 Geoffrey Sampson, *Evidence against the grammatical / ungrammatical distinction*, in MEIJS 1987, pp. 219-227; poi in SAMPSON 2001, § 10 pp. 165-179.

1992 Geoffrey Sampson, *Probabilistic Parsing*, in SVARTVIK 1992, pp. 425-447.

1997 Geoffrey Sampson, *Educating Eve. The "Language Instinct" Debate*, London - New York, Cassel, 1997.

2001/00 Geoffrey Sampson, *The Role of Taxonomy*, in SAMPSON 2001, § 6 pp. 74-93. Based on a paper delivered to the *Royal Society - British Academy Discussion Meeting on Computers, Language and Speech, September 1999*, then in «Philosophical Transactions of the Royal Society» series A, CCCLVIII (2000) 1339-55.

2001 Geoffrey Sampson, *Empirical Linguistics*, London - New York, Continuum, 2001.

2003 Geoffrey Sampson, *Thoughts on two decades of drawing trees*, in ABEILLÉ 2003, pp. 23-41.

SAUSSURE

1916/67/95 Ferdinand de Saussure, *Cours de linguistique générale*, publié par Charles Bailly et Albert Séchehaye, avec la collaboration de Albert Riedingler, édition critique préparée par Tullio de Mauro, postface de Louis-Jean Calvet, Paris, Payot, 2001^r [1995₃, 1972₁] "Grande bibliothèque Payot". Edizione originaria: *ibidem*, 1916. Edizione italiana: *Corso di linguistica generale*, introduzione traduzione e commento di Tullio De Mauro, Roma - Bari, Laterza, 1967₁.

SCHÖNBERG

1933/50/60 Arnold Schönberg, *Brahms il progressivo*, conferenza tenuta il 12 febbraio 1933 e poi riprodotta rielaborata in SCHÖNBERG 1950/60, pp. 56-104.

1950/60 Arnold Schönberg, *Style and Idea*, New York, Philosophical Library, 1950. Traduzione italiana di Maria Giovanna Moretti e Luigi Pestalozza: *Stile e idea*, con un saggio di Luigi Pestalozza, Milano, Feltrinelli, 1980₃ [1975₂, 1960₁] "I fatti e le idee. Saggi e biografie" 293.

- SEGRE
1983 *Intorno alla linguistica*, a cura di Cesare Segre, Milano, Feltrinelli, 1983.
- SINCLAIR
1987 *Looking up: an Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*, edited by J[ohn] M[cHardy] Sinclair, London - Glasgow, Collins ELT, 1987.
1991 John [McHardy] Sinclair, *Corpus, Concordance, Collocation*, Oxford, Oxford University Press, 1991.
- SKINNER
1957 B[urhus] Frederik Skinner, *Verbal Behaviour*, New York, Appleton - Century - Crofts, [1957] = London, Methuen, [1957].
- SPITZER
1929-30 *Meisterwerke der romanischen Sprachwissenschaften*, herausgegeben von Leo Spitzer, München, Hueber, Band I. 1929, Band 2. 1930.
- STRAWSON
2003 Galen Strawson, *Real Materialism*, in ANTONY - HORNSTEIN 2003, pp. 49-88,
- SVARTVIK
1992 *Directions in Corpus Linguistics. Proceedings of the Nobel Symposium 82. Stockholm, 4-8 August 1991*, edited by Jan Svartvik, Berlin, Mouton de Gruyter, 1992 "Trends in Linguistics. Studies and Monographs" 65.
- TAVONI
2012 *Sintassi dell'Italiano antico e sintassi di Dante*, a cura di Mirko Tavoni, Pisa, Felici Editore, 2012.
- TERRACINI
1922 Benvenuto [Aaron] Terracini, *Carlo Salvioni*, in «Archivio glottologico italiano» XVIII (1914-1918-1922)¹⁹²² 586-600.
- TOGNINI-BONELLI
2001 Elena Tognini-Bonelli, *Corpus Linguistics at Work*, Amsterdam - Philadelphia, John Benjamins Publishing Company, 2001 "Studies in Corpus Linguistics" 6.
- VAN ZAAANEN - ROBERTS - ATWELL
2004 Menno van Zaanen - Andrew Roberts - Eric Atwell, *A Multilingual Parsed Corpus as Gold Standard for Grammatical Inference Evaluation*, in AA. VV. 2004.
- VENIER
2007 Federica Venier, *Per un superamento della dicotomia langue/parole: sentieri paralleli e intersezioni di retorica, linguistica testuale e pragmatica*, in DE CESARE - FERRARI 2007, pp. 9-52.
2012 Federica Venier, *La corrente di Humboldt. Una lettura di La Lingua franca di Hugo Schuchardt*, Roma, Carocci editore, 2012 "Lingue e letterature".
- VOLTOLINI
1998/2002 Alberto Voltolini, *Internalism & Externalism [Second Draft]*, online paper, 1998 last updated 2002: <http://host.uniroma3.it/progetti/kant/field/voltolini.html>.
- WITTGENSTEIN
1941-7/53/67 Ludwig Wittgenstein, *Philosophische Untersuchungen*, ds., 1941-47, poi Oxford, Basil Blackwell, 1953. Traduzione italiana: *Ricerche filosofiche*, edizione italiana a cura di Mario Trinchero, Torino, Einaudi, 1983⁵ "Paperbacks" 148 [1967]₁.

ZIPF

- 1929 George Kingsley Zipf, *Relative Frequency as a Determinant of Phonetic Change*, in «Harvard Studies in Classical Philology» XL (1929) 1-95.
- 1935/65 George Kingsley Zipf, *The Psycho-biology of Language. An Introduction to Dynamic Philology*, Introduction by George A. Miller, Cambridge (MA), The M.I.T. Press, 1965. Edizione originale Boston (MA), Houghton Mifflin, 1935.

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