



Giuliano Merz

Anzitutto: Auguri, auguri, che l'anno nuovo sia felice e ricco di soddisfazioni. Come abbiamo visto nelle due puntate precedenti la nostra lingua può contare su varie indagini di tipo statistico, su lessici di frequenza: per il parlato è stato menzionato il LIP del De Mauro (consultabile anche in rete), per lo scritto il LIF di Bortolini et al. e il Frequency dictionary of Italian words di A. Juilland. A questi si deve aggiungere:

dell'Italiano Parlato e Scritto – diretto da Federico Albano Leoni presso il CIRASS dell'Università di Napoli "Federico II"³;

- il CoLFIS (del Laboratorio di Linguistica della Scuola Normale Superiore di Pisa), basato su un corpus - estratto da quotidiani, periodici e libri - di 3.150.075 occorrenze, che mette a disposizione gratuitamente tutte le liste lessicali⁴;
- il CORIS/CODIS, realizzato presso il CILTA dell'Univ. di Bo-

Ma questa mole di informazioni e di possibilità esplorative non si esaurisce qui, il sito della Scuola Superiore di Lingue Moderne per Interpreti e Traduttori (SSL-MIT) dell'Università di Bologna mette a disposizione dell'italianista interessato la consultazione di 10 annate complete del quotidiano "La Repubblica", con una banca dati di parole che supera i 380 milioni⁹.

A fianco di questi preziosi archivi possiamo collocare la nutrita serie di dizionari consultabili in rete: anche quest'offerta conferma un certo primato numerico, la nostra 'dizionarioite' come è stata definita in svariate occasioni, rispetto alle altre lingue europee. Fatta eccezione per il Devoto-Oli e per lo Zingarelli, tutti i vocabolari monolingui di nome sono interrogabili, gratuitamente, tramite Internet. Abbiamo a disposizione il De Mauro¹⁰, il Gabrielli 2008¹¹ (il volume era stato lanciato dal quotidiano "La Repubblica" nell'ultima settimana dell'agosto 2007 al prezzo, sensazionale per un Paese con i dizionari più costosi d'Europa, di 19.90 €), il Garzanti¹² (l'unico che richiede una registrazione ...), il Sabatini-Coletti¹³, il Treccani¹⁴ (il sito è piuttosto lento); a questo bel gruppo si aggiungono almeno due lessici, che esistono solo in forma digitale: "Parole" sul sito Virgilio/Alice¹⁵ e "Dizionario Italiano online"¹⁶. A mo' di congedo un ultimo sito: "Il Cannocchiale"¹⁷ di Fausto Raso, profondo conoscitore della lingua e coautore del pregevole "Giornalismo. Errori e orrori. Per non essere piantati in Nasso dall'italiano"¹⁸.

Grazie per l'attenzione dal vostro Giuliano Merz e-mail: giuliano.merz@uibk.ac.at

Parlando, scrivendo, dubitando La nostra lingua: parliamo di strumenti (3)

- un "Vocabolario fondamentale della lingua italiana" di A. G. Sciarone¹: l'opera basa i propri dati sia sui corpora del LIF e di Juilland sia su uno proprio di ulteriori 500.000 parole;

- il "VELI. Vocabolario elettronico della lingua italiana"², che riporta i primi 10.000 lemmi secondo i valori di uso ricavati da un corpus di ca. 27 milioni di occorrenze provenienti dall'ambito giornalistico-finanziario (: testi da Domenica del Corriere, Europeo, Il Mondo e notiziari economici dell'ANSA). Anche in questo ambito l'italiano è certamente da collocare ai livelli delle lingue maggiormente indagate (cfr. le pubblicazioni di questo tipo per il tedesco, il francese, lo spagnolo ...).

L'avvento della rete e l'evoluzione dei computer ha inoltre favorito la presenza e l'accessibilità di numerosi altri corpora e lessici. In forma sintetica cercherò di dare conto dei progetti, conclusi o in fieri, a me noti. Per la lingua scritta e parlata ricorderò

- il CLIPS - Corpora e Lessici

logna e reso disponibile dal 2001, elaborato e coordinato da R. Rossini Favretti, conta attualmente 100 milioni di parole; l'aggiornamento del corpus è previsto con cadenza biennale⁵;

- i progetti LABLITA, Laboratorio Linguistico del Dipartimento di Italianistica dell'Università di Firenze: qui, tra gli altri, il corpus di parlato spontaneo di adulti, quasi 63 ore di registrazioni per un totale di 640514 forme⁶ e il corpus Stammerjohann, la prima raccolta di italiano parlato spontaneo mai realizzata, effettuata a Firenze nel 1965: ca. 41 ore di registrazione, quasi 417 mila parole⁷.

- A questi aggiungo, pur trattandosi di un sottoinsieme, il CIT. Corpus di Italiano Televisivo di Stefania Spina (Univ. di Perugia): per ora il sito offre una "demo", pari a 61.631 parole da 445' di registrazioni; è previsto un ampliamento del corpus a 500.000 parole "per rendere omogenei i raffronti con altri corpora italiani di lingua scritta (LIF) e parlata (LIP)"⁸.

¹ Sciarone, Abondio Giuseppe: Vocabolario fondamentale della lingua italiana. Perugia, Guerra Ed. 1995; 260 p. [ediz. prec.: Bergamo, Minerva Italica 1977]

² AA.VV.: VELI. Vocabolario Elettronico della Lingua Italiana. Prefazione di Tullio De Mauro. Milano, IBM Italia 1989; cofanetto con 1 vol. (397 p.) e 2 floppy disk

³ rimento a <http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=41/vers=ita>

⁴ <http://alphalinguistica.sns.it/CoLFIS/CoLFIS-Presentazione.htm> - i 136883 lemmi dell'ampio progetto sono consultabili, per infissi, prefissi e suffissi, sul mio server all'indirizzo <http://culturitalia.uibk.ac.at/wb2/morfemix2.htm>

⁵ http://corpora.dslo.unibo.it/coris_ita.html

⁶ <http://lablita.dit.unifi.it/corpora/>

descriptions/lablita/

Della collega Emanuela Cresti del Laboratorio fiorentino non posso non ricordare: E. Cresti, Corpus di italiano parlato. Firenze, Accademia della Crusca 2000; 2 voll. con CD

⁷ <http://lablita.dit.unifi.it/corpora/descriptions/stam/>

⁸ <http://www.sspina.it/cit/cit.htm>

Qui vale la pena ricordare che presso l'Accademia della Crusca è in corso il progetto di "Lessico di frequenza dell'italiano radiofonico (LIR)" a cura di Nicoletta Maraschio e Stefania Stefanelli

⁹ Per la precisione: 380.823.725; si trova all'indirizzo <http://dev.sslmit.unibo.it/corpora/corpus.php?path=&name=Repubblica>

¹⁰ <http://www.demauroparavia.it>

¹¹ <http://dizionari.repubblica.it/>

¹² <http://www.garzantilinguistica.it/>

¹³ http://dizionari.corriere.it/dizionario_italiano.shtml

¹⁴ http://www.treccani.it/site/lingua_linguaggi/consultazione.htm

¹⁵ <http://parole.alice.it/parole/vocabolario/index.html> oppure <http://parole.virgilio.it/parole/vocabolario/index.asp> oppure <http://sapere.alice.it/home/index.html>

¹⁶ <http://www.dizionario-italiano.it/> (progetto in corso di realizzazione che punta alla creazione del più esteso dizionario italiano oggi esistente...)

¹⁷ <http://faustoraso.ilcannocchiale.it/>

¹⁸ Carlo Picozza / Fausto Raso: Giornalismo errori e orrori. Per non essere piantati in Nasso dall'italiano. Roma, Gangemi Editore 2004; 126 pp.