

Elena Maria Pandolfi
Il lessico, l'informatica e i dizionari.
Riflettendo sugli atti di un convegno di linguisti svoltosi a Torino
"Corriere del Ticino", 20 ottobre 2006.



LA LESSICOGRAFIA è propriamente la scienza e tecnica dei dizionari, e in quest'ambito è maturata di recente una vera rivoluzione grazie all'uso dell'informatica. Se ne è parlato in un convegno svoltosi il mese scorso a Torino.

Parlare di un congresso di quattro giorni che ha prodotto due volumi di atti di quasi 1300 pagine totali è impresa assai ardua. Se è un congresso importante su un aspetto centrale del linguaggio verbale, il patrimonio lessicale delle lingue, non si può però non farne almeno qualche cenno: cosa che cercheremo di fare qui.

Si tratta del XII Congresso Internazionale di Lessicografia organizzato da Euralex, l'Associazione Europea per la Lessicografia, tenutosi a Torino dal 6 al 9 settembre 2006. Il congresso ha avuto cinque sessioni parallele, tre relazioni plenarie con nomi di punta quali Francesco Sabatini (presidente dell' Accademia della Crusca), Tullio De Mauro e Jean Aitchinson dell' Università di Oxford, due tavole rotonde, e una nutrita sezione di poster (per i non frequentatori di congressi di linguistica, i poster sono cartelloni su cui chi ha condotto una ricerca ne descrive con parole e immagini metodologie e risultati).

La lessicografia è propriamente la scienza e tecnica dei dizionari, ma necessariamente in un congresso del genere la prospettiva si allarga a molte delle questioni collegate allo studio del lessico in diverse sue dimensioni, con evidenti forti implicazioni anche per quel che riguarda la didattica delle lingue. In questo campo, è maturata in tempi recenti una vera rivoluzione tecnologica.

Lo studio lessicografico, sia esso diacronico sia sincronico, si fonda sempre più in questi ultimi anni su quella che è definita la «linguistica dei corpora» (corpus linguistics), mirante in breve all'individuazione di caratteristiche della lingua attraverso l'analisi di grosse quantità di dati empirici (testi, materiali), appunto i corpora, resi accessibili principalmente in internet e attraverso internet.

Quantità spropositate di dati linguistici sono diventate ora maneggiabili attraverso la mediazione dell'informatica applicata alla linguistica, che consente il trattamento statistico computazionale di masse di materiali inimmaginabili per linguisti abituati fino a non molti anni fa a lavorare di carta e penna su corpora artigianali. Si tratta di materia assai interessante in generale, dato che tale direzione di ricerca è destinata a prendere sempre più piede nella linguistica. Proprio sulle analisi statistiche di fenomeni linguistici, lessico ma non solo, è uscito per esempio non molti mesi fa l'importante volume curato da Tullio De Mauro e Isabella Chiari, *Parole e numeri. Analisi quantitative dei fatti di lingua* (Aracne, Roma 2005).

L'attualità di questo congresso è inoltre accresciuta dalla tempestività con cui sono apparsi gli Atti a stampa, distribuiti addirittura in apertura del congresso stesso (cosa del tutto insolita per congressi delle discipline umanistiche), grazie alla solerte attività del comitato organizzatore dell'Università di Torino guidato da Carla Marengo.

I 147 interventi di relatori provenienti da diversi paesi europei sono stati distribuiti in undici aree tematiche: Lessicografia storica e etimologia, Lessicografia e lessicologia computazionale, Il processo di formazione dei dizionari, Relazioni su progetti di lessicografia e lessicologia, Lessicografia bilingue, Dizionari dialettali e atlanti linguistici, Lessicografia per lingue speciali, L'uso di dizionari, Fraseologia e collocazioni, Temi lessicologici di rilevanza lessicografica, e la categoria tuttofare di Altri temi.

Fra gli argomenti che hanno avuto notevole presenza anche trasversale vorremmo citare in particolare la questione delle «collocazioni», sequenze di due o più parole con forti legami di associazione reciproca. Tra le collocazioni possiamo inserire come caso particolare anche i gruppi di parole che vengono chiamate tecnicamente «polirematiche» o «unità plurilessicali»: esempi per l'italiano che mostrano le diverse manifestazioni della categoria possono essere «mettere in piedi», «a vista d'occhio», «essere all'altezza», «casa editrice», «donna di servizio», «ordine del giorno». Le collocazioni possono però non essere espressioni idiomatiche, ma appunto semplici sequenze di parole che hanno una forte carica associativa in relazione al tipo di testo in cui compaiono. Se per esempio parliamo (o scriviamo) di giardinaggio avremo facilmente sequenze di parole come piantare fiori, seminare il prato o semina dell'orto: statisticamente si può prevedere che certe parole siano più facilmente vicine ad altre semanticamente complementari. Lo studio delle collocazioni ci fornisce quindi indicazioni interessanti sulla struttura lessicale di un testo, e sulla possibilità di avere sequenze di parole statisticamente più probabili. Appare chiaro che una delle possibili e produttive applicazioni di tali analisi è la didattica di una lingua seconda.

Margarita Alonso Ramos (Università della Coruña) nel suo contributo fa per esempio il punto dell'avanzamento degli studi in questo settore e descrive il *Diccionario de colocaciones del español* come un passo avanti in una nuova metodologia nella didattica della lingua seconda, lo studio appunto di probabilità statistiche di combinabilità delle parole, che integri l'approccio tradizionale nella didattica del lessico e della sintassi.

Le collocazioni stanno anche acquistando sempre maggiore spazio nei dizionari bilingui come significativo arricchimento della descrizione del singolo lemma; e partendo dalla prospettiva delle collocazioni si può addirittura arrivare a concezioni in parte originali della lingua, come quella proposta da Luisa Giacoma (Università di Torino) e Susanne Kolb (Università di Firenze), autrici di un nuovo dizionario tedesco-italiano/italiano-tedesco uscito nel 2001, che definiscono la lingua come «un gigantesco puzzle nel quale ogni tessera corrisponde a una parola e può essere combinata alle altre in modo non totalmente libero, ma secondo certe regole e limitazioni» (p. 967).

Un altro tema di particolare interesse ben rappresentato in questi Atti e che costituisce uno sviluppo relativamente nuovo negli studi del settore è l'applicazione di metodi computazionali ai lessici dialettali e agli atlanti linguistici. Attraverso opportuni calcoli statistico-computazionali è possibile tracciare mappe linguistiche dettagliate delle varietà dialettali, spesso accessibili e interrogabili in rete (e questa è una grande novità, destinata a riscuotere il più vivo apprezzamento da tutti coloro che fanno ricerca nell'ambito della variazione linguistica).

Va inoltre ricordato, fra gli aspetti più significativi della ricerca lessicografica contemporanea così ampiamente testimoniata in questi Atti, che in tutto questo fiorire di applicazioni computazionali non è stata affatto trascurata la dimensione storica dei fatti linguistici, con trattazioni che vanno per esempio dall'inglese del Settecento, al greco e al latino.

Ciò naturalmente non toglie che la maggior parte dei contributi, com'è lecito aspettarsi, si riferisca direttamente o indirettamente ai problemi e ai metodi della costruzione dei dizionari, e che la parte del leone la facciano qui i dizionari bilingui, campo in cui gli strumenti elettronici hanno consentito progressi veramente eccezionali. E i due volumi che recensiamo appaiono destinati a soddisfare a questo proposito anche gli appetiti che potremmo dire più curiosi, spaziando i contributi davvero in tutte le aree dello scibile lessicografico, dal dizionario di base per la scuola cubana (R. Miyares) al dizionario di falsi anglismi in italiano (C. Furiassi) al lessico etimologico di latinismi e romanismi in sloveno (A. Segal), e via dicendo.

ELISA CORINO, CARLA MARELLO, CRISTINA ONESTI (a cura di), *Atti del XII Congresso Internazionale di Lessicografia, Torino 6-9 settembre 2006*, 2 voll., Ed. Dell'Orso, 1288 pagine, 100 euro.