

Elena Maria Pandolfi
Osservatorio linguistico della Svizzera italiana, V. Franscini 30a, 6500 Bellinzona
tel. 091 814 15 10
empandolfi@heuristics.ch

LIPSI. Lessico di frequenza dell'italiano parlato nella Svizzera italiana

Con il presente contributo si intendono proporre gli aspetti metodologici e qualche primo risultato della ricerca in corso presso l'Osservatorio linguistico della Svizzera italiana: "LIPSI, Lessico di frequenza dell'italiano parlato nella Svizzera italiana". L'elaborazione di un *corpus* di italiano parlato nella Svizzera italiana (Canton Ticino e valli italofone dei Grigioni Italiani) composto da circa 400.000 parole grafiche, attraverso le fasi della raccolta del materiale parlato, della trascrizione, della indicizzazione e della lemmatizzazione, ha permesso di costruire liste di frequenza di lemmi e forme ordinate per rango decrescente. Le finalità sono principalmente descrittive della varietà IRSI (italiano regionale della Svizzera italiana) e contrastive nei confronti di altre liste di frequenza elaborate per l'italiano d'Italia: il LIP (Lessico di frequenza dell'italiano parlato, De Mauro et alii, 1993) e il C-Oral-Rom (*Integrated Reference Corpora for Spoken Romance Languages*, a cura di Emanuela Cresti e Massimo Moneglia, 2005), ma anche con il Vocabolario di base della lingua italiana. Non mancano però anche possibili prospettive legate alla didattica dell'italiano come lingua seconda (si potrà eventualmente valutare quale lessico insegnare, o tener conto dei risultati del LIPSI in relazione a manuali per la didattica dell'italiano in contesto svizzero). Verranno brevemente descritte la metodologia adottata nell'allestimento del *corpus* e le scelte di lemmatizzazione in casi particolari che coinvolgono problemi di morfologia lessicale, per poi passare alla presentazione di alcuni risultati dell'applicazione di procedimenti elementari di statistica linguistica. Verranno trattati in particolare la distribuzione statistica dei lessemi, dalla frequenza 1 (hapax) alla frequenza massima (22.530 occorrenze dell'articolo *il*), alcune applicazioni della legge di Zipf (che mette in relazione la frequenza e il rango), il valore della densità lessicale (Halliday, 1995), la frequenza e la natura delle polilessicali (o polirematiche). Verranno inoltre presentati alcuni risultati derivanti dal confronto con il LIP e il C-Oral-Rom e con il Vocabolario di base per poi passare a qualche prima considerazione su tratti dell'IRSI parlato che emergano dai dati.

Bibliografia essenziale:

- Biber Douglas, Susan Conrad, Randi Reppen, 1998, *Corpus linguistics. Investigating language structure and use*, Cambridge, Cambridge University Press.
- Cresti Emanuela, Massimo Moneglia (eds.), 2005, *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam/Philadelphia, John Benjamins.
- De Mauro Tullio, Federico Mancini, Massimo Vedovelli, Miriam Voghera, 1993, *Lessico di frequenza dell'italiano parlato*, Milano, Etaslibri.
- De Mauro Tullio, Isabella Chiari (a cura di), 2005, *Parole e numeri. Analisi quantitative dei fatti di lingua*, Aracne, Roma.
- Halliday Michael Alexander Kirkwood, 1995, *Lingua parlata e lingua scritta*, Scandicci (FI), La Nuova Italia.
- Lenci Alessandro, Simonetta Montemagni, Vito Pirrelli, 2005, *Testo e computer. Elementi di linguistica computazionale*, Roma, Carocci.
- Pandolfi Elena Maria, 2006, *Misurare la regionalità. Uno studio quantitativo su regionalismi e forestierismi nell'italiano parlato nel Canton Ticino*, Osservatorio linguistico della Svizzera italiana, Locarno, Daddò.
- Thornton Anna Maria, Claudio Iacobini, Cristina Burani, 1997, *VDVDB -Una base di dati sul Vocabolario di base della lingua italiana*, (con floppy disc), Roma, Bulzoni Editore.