

Convegno del Gruppo di Studio sulla Comunicazione Parlata
Napoli, 23-25 febbraio 2006

*Ingresso, uso, integrazione e produttività delle parole nuove in italiano:
proposte di valutazione quantitativa e qualitativa dei dati*
Isabella Chiari

Questo contributo intende proporre ed esemplificare alcuni metodi di valutazione quantitativa dell'ingresso di nuove parole in una lingua, con particolare riferimento all'italiano. In particolare si definisce la necessità di valutare globalmente il comportamento delle unità lessicali nella lingua parlata e scritta e distinguere i livelli di *uso*, *integrazione* e *specificità*, *adattamento* e *produttività* delle unità monitorandone la presenza/assenza nel tempo e in diverse tipologie testuali.

Esiste un modo di estrarre informazioni generali sulla penetrazione di nuove unità lessicali nel sistema di una lingua, individuandone le tendenze d'uso e la diffusione? La nozione di *uso*, definita come frequenza per dispersione così come oramai diffusa nella considerazione dei fatti lessicali, è sufficiente a descrivere le dinamiche diacroniche e sincroniche dei nuovi e dei vecchi lessemi? È possibile valutare quantitativamente l'integrazione dei lessemi, la loro produttività e gli usi in modo da generare degli indici di tendenza per ciascuna unità? La nozione di dispersione è davvero quantitativa? È possibile sopperire alle limitazioni dovute alle scelte di segmentazione dei grandi corpora di riferimento in modo da ri-generare i corpora in maniera dinamica e ottenere diverse valutazioni sinottiche della dispersione e dunque dell'uso degli elementi lessicali? E una volta approntata una misura dell'integrazione delle unità qual è il livello di rumore e quanta parte ancora deve essere svolta con una verifica e distinzione qualitativa?

Con il ricorso ai dati forniti da corpora di parlato e di corpora di monitoraggio saranno esemplificati alcuni metodi che permettono di problematizzare la nozione di *uso* in relazione alla scelta dei criteri di definizione della nozione statistica di "dispersione" (cfr. Hlaváčová e Rychlý, 1999; Kromer, 2003; Dunning, 1993) e si proporrà un indice complessivo che renda conto globalmente dei diversi aspetti che caratterizzano il comportamento dei neologismi, e dei prestiti. Gli indici saranno confrontati con altri ad essi in qualche misura correlati come il *weirdness index* di K. Ahmad, 2005, l'*indice IT* di Bolasco, 2005, o il *frequency profiling* di Rayson e Garside, 2000.

- Bolasco, S. (2005), "La reperibilità statistica di tendenze diacroniche nell'uso delle parole" in De Mauro T. e I. Chiari (a cura di), *Parole e numeri. Analisi quantitative dei fatti di lingua*, Roma, Aracne, 2005, pp. 335-354.
- Dunning, T. (1993). "Accurate Methods for the Statistics of Surprise and Coincidence". *Computational Linguistics*, 19, 1, 1993, pp. 61-74.
- Gillam, L. e K. Ahmad (2005), "Pattern Mining Across Domain-Specific Text Collections", *Lecture Notes in Computer Science*, 3587, 2005, pp. 570-579.
- Hlaváčová, J. e Rychlý, P. (1999), "Dispersion of Words in a Language Corpus", *Lecture Notes in Computer Science*, 1692, 1999, pp. 321-324.
- Kromer, V. (2003), "A Usage Measure Based on Psychophysical Relations", *Journal of Quantitative Linguistics*, 10, 2, 2003, pp. 177-186.
- Rayson P. e R. Garside, (2000) "Comparing corpora using frequency profiling", *Proceedings of Workshop on comparing Corpora of ACL 2000*, 1-8, 2000, Hong Kong, 1-6.