

ΑΙΟ

---

ΙΙΟ

Volume stampato con il contributo  
del Dipartimento di Studi comparati  
Università degli Studi "G. D'Annunzio" di Chieti - Pescara

Domenico Russo

**LIN**  
Lessico Italiano  
di Notorietà 2004  
*Prototipo*



Copyright © MMIV  
ARACNE EDITRICE S.r.l.

[www.aracne-editrice.it](http://www.aracne-editrice.it)  
[info@aracne-editrice.it](mailto:info@aracne-editrice.it)

00173 Roma  
via Raffaele Garofalo, 133 A/B  
(06) 72672222 – (06) 93781065  
telefax 72672233

ISBN 88-7999-898-6

*I diritti di traduzione, di memorizzazione elettronica,  
di riproduzione e di adattamento anche parziale,  
con qualsiasi mezzo, sono riservati per tutti i Paesi.*

I edizione: dicembre 2004

## Indice

<b>Introduzione</b> .....	1
§ 1. Frequenza e notorietà .....	1
§ 1.1. <i>Il Français fondamental</i> .....	1
§ 1.2. <i>Il Vocabolario di base</i> .....	2
§ 2. Il LIN 2004 .....	3
§ 2.1. <i>Il lemmario del LIN</i> .....	4
§ 2.2. <i>Il test di notorietà</i> .....	5
§ 2.3. <i>La popolazione testata</i> .....	5
§ 2.4. <i>Trattamento dei dati e natura degli indici</i> .....	6
§ 2.5. <i>Gli indici per il confronto</i> .....	7
<b>Riferimenti bibliografici</b> .....	8
Lista dei lemmi in ordine di notorietà decrescente .....	11
Lista dei lemmi in ordine alfabetico .....	133
Lista dei lemmi in ordine di Rango d'Uso Lif decrescente	255



## Introduzione

### § 1. La frequenza e la notorietà

#### § 1.1. *Il Français fondamental*

Nel corso di uno dei più importanti tentativi di stabilire il lessico fondamentale di una lingua, quello condotto da Gougenheim e dai suoi collaboratori nel corso degli anni Cinquanta e Sessanta in Francia<sup>1</sup>, emerse con tutta evidenza il fatto che le parole comprese nelle liste di frequenza elaborate statisticamente su *corpora* di lingua scritta o, come in quel caso, parlata non davano conto compiutamente del lessico che l'intuizione dei linguisti e degli insegnanti di lingua sentivano come effettivamente usato dai parlanti della comunità indagata. Nel caso del *Français fondamental* la soluzione adottata fu quella di aggiungere alla lista di frequenza anche una lista di parole *disponibili* elicitata attraverso un questionario distribuito in varie scuole di diverse regioni francesi<sup>2</sup>. Il risultato complessivo diede ragione a Gougenheim e collaboratori. L'indagine sulle parole disponibili mostrò che nella lista di frequenza mancavano parole cognitivamente e culturalmente molto importanti, mentre la loro aggiunta al lemmario fondamentale rendeva il tutto più aderente alla dinamica lessicale effettiva.

Il modo di procedere di Gougeheim contribuì a risolvere anche un problema di ordine teorico. L'indagine sulle parole disponibili infatti non solo migliorava il lemmario, ma indicava anche la strada per dare formalità scientifica a un criterio di selezione del lemmario che veniva adottato spesso, se non sempre, nei tentativi di redazione di liste di parole fondamentali, vale a dire il ricorso alla competenza linguistica dei

<sup>1</sup> Cfr. Gougenheim 1956 e *Idem* 1964.

<sup>2</sup> Cfr. Gougenheim 1956: 133 e sgg; *Idem* 1964: 137 e sgg.

redattori, criterio che in genere viene indicato come selezione per «esperienza».

Va infatti ricordato che prima o in parallelo con l'affermarsi di pratiche lessicografiche per la costituzione dei lemmari fondamentali statisticamente fondate — pratiche relativamente recenti, che richiedono, o meglio richiedevano, in anni non ancora informatizzati, l'impiego di macchine e di competenze di una certa complessità — si ricorreva spesso in questa o quella fase della redazione al parere di figure professionali che operavano nel campo della ricerca e dell'insegnamento linguistico; si ricorreva, in una parola, alla competenza lessicale individuale. L'inconveniente di questo criterio di selezione è evidente. Ricorrere alla competenza lessicale dello specialista vanifica la possibilità di avere protocolli di elicitazione replicabili, con danni evidenti sia per le possibilità di confronto intra- e inter-linguistico sul piano descrittivo e contrastivo, sia per le possibilità di falsificazione delle procedure sul piano teorico.

### § 1.2. *Il Vocabolario di base*

I problemi incontrati da Gougenheim in Francia si presentavano in Italia nello stesso torno di anni. Anche le liste di frequenza delle parole fondamentali dell'italiano manifestavano la stessa mancanza di completezza nella descrizione di quello che specialisti e parlanti avvertivano essere il lessico usato effettivamente dagli italiani mentre era generalizzato il ricorso al criterio di selezione per «esperienza» che peraltro dava vita a lavori di pregio<sup>3</sup>.

Per l'adozione di un protocollo replicabile di costruzione di un lemmario fondamentale italiano occorre arrivare alla fine degli anni Settanta con la redazione del *Vocabolario di base* di De Mauro e collaboratori<sup>4</sup>. Gemmazione di un lemmario elaborato dal CNUCE, che per altra via costituiva il LIF, il *Vocabolario di*

<sup>3</sup> Cfr. sui vocabolari di frequenza e di base dell'italiano Sgroi 1981 e Burani e Thornton 1993. In questo panorama segnalo in particolare Migliorini 1943.

<sup>4</sup> Cfr. De Mauro 1980: 102-12 e *Appendice*: 147-70 per il primo lemmario; *Idem* 1994: 115-21 per la genesi; GDU *Introduzione*: *passim* e *Postfazione*: *passim*; *Idem* 2004: 144 per la fortuna editoriale.



*base* distingue il lemmario in parole fondamentali, parole di alto uso e aggiunge una lista di parole disponibili selezionate per esperienza da un gruppo di linguisti. I criteri adottati dal *Vocabolario di base* presentano un passo in avanti rispetto ai criteri di costituzione del lemmario del *Français fondamental*. Nel caso del VdB, infatti, la lista di partenza, una volta costituita, viene sottoposta al vaglio dei parlanti attraverso appositi test di comprensione<sup>5</sup> e nel lemmario si raccolgono solo le parole che risultano note al profilo di competenza lessicale ritenuto pertinente dall'autore<sup>6</sup>.

Il tempo ha esaltato il valore dell'intelligenza selettiva e dell'aderenza all'uso linguistico effettivo del *Français fondamental* e del *Vocabolario di base*. Oggi è difficile apprezzare in positivo ricerche lessicografiche, lessicologiche o semantiche su insiemi estesi, e per così dire *speaker-oriented*, di unità lessicali che non mettano a frutto il nucleo essenziale delle lezioni di Gougenheim e di De Mauro, vale a dire il ricorso a opportune sinergie tra dati e tecniche di rilevazione statistica e dati e tecniche di elicitazione di dati psicolinguistici.

## § 2. Il LIN 2004

L'elaborazione del *Lessico italiano di notorietà*, di cui si presenta qui il prototipo al 2004, si iscrive nelle linee epistemologiche rammentate per sommi capi nel paragrafo precedente e propone di chiudere il cerchio aperto dal vocabolario disponibile di Gougenheim e tracciato dalle verifiche di comprensione di De Mauro.

Così come le liste di frequenza stabiliscono per ogni lemma il relativo indice di frequenza, d'uso e di rango, allo stesso modo il *Lessico italiano di notorietà* fornisce per ogni lemma il relativo indice di notorietà, di rango e di attrazione. Così come vocabolari fondamentali e di base presentano elementi lessicali che sfuggono, pur essendo funzionalmente rilevanti, alla quanti-

<sup>5</sup> Buona parte dei lavori di verifica si leggono in Gensini e Vedovelli 1983 (a completamento di Sgroi 1981: 288).

<sup>6</sup> Profilo che nel caso del VdB (cfr. De Mauro 1980: 148) viene calibrato attorno alla competenza lessicale di uno studente di terza media inferiore.

ficazione statistica, anche il *Lessico italiano di notorietà* include parole disponibili di cui stabilisce i relativi indici. In questo modo il LIN genera un lemmario di natura unitaria, indicizzato, replicabile, i cui indici dipendono in modo diretto dal tempo della rilevazione e dalla massa parlante intervistata.

Un lessico con le caratteristiche del LIN permette una pluralità di applicazioni e di sviluppi, ma è utile soprattutto in due casi. Il primo è relativo allo studio delle configurazioni lessicali di natura individuale, di gruppo o comunitarie. Il secondo è relativo allo studio delle evoluzioni lessicali, individuali, di gruppo o di comunità, relativamente a questa o quella parola, a questo o a quel gruppo o insieme di parole. Entrambi i settori citati, com'è noto, alimentano con dati e temi di riflessione la ricerca teorica di base e il versante applicativo delle linguistica (primo tra tutti l'educazione linguistica).

### § 2.1. Il lemmario del LIN

Il lemmario del LIN è costituito da 5801 lemmi che comprendono 3812 (pari al 71% ca) dei 5356 lemmi del LIF e dai 1989 lemmi del vocabolario AD del GDU<sup>7</sup>. I lemmi LIF sono scelti in base alla divisione dell'intero lemmario in 6 intervalli di rango d'uso composti rispettivamente il primo di 500 lemmi, il secondo e il terzo di 750 e i restanti tre di circa 1000 lemmi ciascuno. I lemmi selezionati dal primo intervallo sono 497 (esclusi *essere* e *avere* ausiliari e *volta*), dagli altri intervalli sono stati prelevati i primi 663 lemmi di ognuno (con l'esclusione di *coda* dal terzo e di *serbare* dal quinto). La tabella 1 riporta le cifre in dettaglio:

Intervallo	Numero dei lemmi	Intervallo del rango d'uso	Lemmi prelevati	Intervallo di rango prelevato
1	500	1.0-500.0	497	1.0-500.0
2	750	501.0-1250.0	663	501.0-1169.0
3	751	1251.5-1999.0	663	1251.5-1920.0
4	989	2002.0-2981.5	663	2002.0-2671.5
5	968	2992.5-3955.0	663	2992.5-3678.5
6	1043	4013.5-4973.0	663	4013.5-4630.0

Tabella 1 - Numero dei lemmi e valori dei ranghi d'uso degli intervalli considerati

<sup>7</sup> Estratti dalla prima edizione, 1999.

### § 2.2. Il test di notorietà

Il principio su cui si basa il test di notorietà usato per stabilire il LIN riprende e rimodella una procedura introdotta da De Mauro e dai suoi collaboratori nello studio del lemmario del VdB<sup>8</sup> che consiste, nell'essenziale, nel correlare in opportune unità di rilevazione lemmi di cui è noto il valore d'uso, nel nostro caso i lemmi LIF, e lemmi che non hanno valori di frequenza d'uso come sono appunto le parole disponibili, nel nostro caso il vocabolario AD del GDU.

Le unità di rilevazione del LIN sono 663<sup>9</sup> serie di dieci lemmi ciascuna. In ogni serie sono presenti 6 lemmi LIF, uno per ognuno dei 6 intervalli di valore d'uso stabiliti, 3 lemmi AD e un lemma il cui grado di notorietà e di frequenza d'uso è esterno sia al vocabolario AD che ai valori LIF.

Agli intervistati si chiede di attribuire un punteggio da 1 a 10 alle parole in base alla percezione di notorietà soggettiva che ogni lemma suscita alla loro attenzione metalinguistica<sup>10</sup>. L'attribuzione dei punteggi deve avvenire in cinque turni di assegnazione successivi, in ognuno dei quali l'intervistato deve selezionare una coppia di parole composta dalla parola più nota e da quella meno nota tra quelle in esame, attribuendo il punteggio maggiore alla parola nota e il minore alla meno nota<sup>11</sup>.

### § 2.3. La popolazione testata

I dati del LIN 2004 risultano da 20 rilevazioni di bambini di una classe di seconda media, popolazione che si ritiene pertinente ai fini dell'individuazione delle linee di demarcazione

<sup>8</sup> Cfr. in particolare Riso 1986 e relativa tesi di laurea.

<sup>9</sup> Cfr. Russo 2004.

<sup>10</sup> La nozione di «notorietà» adottata dal LIN è la sommatoria della varietà dei fatti che determinano un parlante a sostenere che conosce o non conosce e quanto nell'uno e nell'altro caso una parola e si basa sulla doppia assunzione che la dizione *conoscere una parola* coincida con l'intero insieme dei possibili usi linguistici e che, se interrogato in proposito, un parlante è sempre in grado di pronunciarsi in merito.

<sup>11</sup> Cfr., per i dettagli e le ragioni delle scelte citate, Russo in c. di s. e *Idem* 2004: 3-8.

principali della configurazione lessicale complessiva richiesta al prototipo così come per l'individuazione di dati che possano essere proiettati sul piano dei grandi numeri.

Le rilevazioni sono state curate in modo valente e magistrale da Daniela Campitelli dal settembre 2003 al maggio 2004 presso l'Istituto «Benedetto Croce» di Pescara, dove risiedono tutti e venti gli intervistati selezionati come campione<sup>12</sup>.

#### § 2.4. *Trattamento dei dati e natura degli indici*

Gli intervistati hanno espresso i loro giudizi su supporto cartaceo successivamente informatizzato. Somma e media aritmetica sulla somma sono le uniche operazioni compiute sui coefficienti espressi.

In base alla procedura di raccolta dei dati e ai risultati ottenuti il LIN mette a disposizione i cinque indici numerici che seguono.

1. *Indice di notorietà*. Quest'indice coincide con la media aritmetica calcolata sul totale dei coefficienti attribuiti e viene contrassegnato dalla lettera *N*.

2. *Grado di notorietà*. Quest'indice contrassegnato dal simbolo *Gn* è stabilito in analogia alla struttura statistica del lemmario LIF in base alla tabella che segue:

RdU	Gn
10.00-8.00	1
7.99-7.89	2
7.88-7.47	3
7.46-6.56	4
6.55-4.56	5
4.55-3.00	6
2.99-1.00	7

Tabella 2 – *RdU = Rango d'Uso; Gn = Grado di notorietà*

<sup>12</sup> E' interessante aggiungere qui che l'intera rilevazione è corsa parallela allo svolgersi del programma didattico annuale ed è stata realizzata a costo zero e senza monte ore dedicato, ma come momento dell'insegnamento lessicale svolto nel quadro della programmazione di educazione linguistica prevista per la classe.

3. *Rango di notorietà*. Quest'indice si evince dall'ordine discendente delle medie, viene espresso sempre in cifra tonda e si ripete per i lemmi di media uguale. Per esigenze di spazio e di facilità di lettura dei dati, il rango di notorietà forma la prima parte (quella a sinistra del punto fermo) di ciò che qui si definisce come *indice di salienza funzionale*. Il simbolo di quest'indice è *Sf*.

4. *L'indice di salienza funzionale*. Quest'indice, simboleggiato da *Sf*, è composto dal rango di notorietà (a sinistra del punto fermo) e dal numero dei lemmi compresi in quel punto di rango (a destra del punto fermo).

5. *L'indice di attrazione media*. Quest'indice risulta dalla media degli indici di attrazione relativa (positivi per le parole note, negativi per le non note) connessi ai dieci coefficienti usati nel test<sup>13</sup>. Il simbolo impiegato per quest'indice è *Am*.

#### § 2.5. *Gli indici per il confronto*

Nel testo tutti gli indici elaborati sulla base della rilevazione vengono presentati a destra del lemma nell'ordine in cui sono stati presentati nel § precedente. Per rendere immediati gli opportuni confronti, il LIN riporta a sinistra dei lemmi il valore del rango d'uso LIF indicato dalla sigla *RdU* e il relativo intervallo di valori di rango d'uso adottato dalla rilevazione, indicato dalla lettera *I*<sup>14</sup>.

<sup>13</sup> Cfr. per i dettagli di quest'aspetto della rilevazione Russo in c. di s.

<sup>14</sup> L'*RdU* 8888.0 e l'*I* 8 sono convenzionali. Rinviano infatti alle parole di alta disponibilità.