

LIR - LESSICO ITALIANO RADIOFONICO (1995-2003)

La radio è fra i mezzi di comunicazione di massa quello che oggi più sta dimostrando versatilità e capacità di rinnovamento. Prove significative ne sono la costante crescita degli ascolti, soprattutto presso i giovani, e la facilità di interrelazione con Internet. Si conferma così all'inizio del terzo millennio il ruolo rilevante che la radio ha avuto nel processo di italianizzazione novecentesco e la sua centralità nell'architettura dell'italiano contemporaneo. Il *Lessico Italiano Radiofonico* (LIR) che ora si pubblica in versione elettronica è il risultato di una ricerca pluriennale, condotta presso il Centro di grammatica italiana dell'Accademia della Crusca grazie alla sensibilità per le nuove prospettive di studio sull'italiano contemporaneo del suo direttore Giovanni Nencioni. Il progetto coordinato da Nicoletta Maraschio (Dipartimento di italianistica dell'Università di Firenze) e Stefania Stefanelli (Scuola Normale Superiore di Pisa) è stato realizzato da un gruppo di ricercatori di formazione diversa (storici della lingua, linguisti, fonetisti, statistici e informatici): Gabriella Alfieri, Anna Antonini, Cinzia Avesani, Patrizia Bellucci, Marco Biffi, Stefania Buccioni, Massimo Fanfani, Nicoletta Maraschio, Massimo Moneglia, Enrico Paradisi, Eugenio Picchi, Monica Pratesi, Raffaella Setti, Stefania Stefanelli e Mario Vayra.

Il LIR si inserisce con una propria forte specificità nel quadro degli studi sull'italiano parlato che in questi ultimi anni appare particolarmente ricco e innovativo. I due DVD contengono due corpora di italiano radiofonico raccolti a quasi dieci anni di distanza, una prima volta nel 1995 e poi nel 2003. Il corpus del LIR1 (1995) raccoglie circa 64 ore di parlato radiofonico, trascritto e in voce, registrato da nove radio a diffusione nazionale (le tre RadioRai, Radio DJ, RTI 102.5, Rete 105, Italia Radio, Radio Radicale, Radio Vaticana) per un totale di 105 ore. L'insieme dei testi, etichettati in base a parametri diversi e variamente incrociabili ("emittente", "genere radiofonico", "speaker", "tipologie comunicative") può essere interrogato grazie a un potente motore di ricerca come il DBT di Eugenio Picchi, opportunamente adattato alla specificità del parlato trasmesso. Si può così conoscere la frequenza assoluta e relativa di ogni parola, ma anche quella di alcuni fenomeni morfosintattici tipici dell'italiano di oggi (dall'uso del congiuntivo/indicativo al *che* polivalente, dalla distribuzione fra *egli/lei* a quella fra *a loro*, *a lui*, *a lei* /*gli*, a fenomeni come l'esitazione, la sovrapposizione, l'autocorrezione tipici del parlato). Con il LIR2 (2003) si ripropone lo stesso tipo di corpus limitando i prelievi alla tre reti RAI: le 36 ore di registrazioni (12 per ogni rete) hanno consentito di costruire un corpus di 33 ore di parlato radiofonico perfettamente comparabile in diacronia con le registrazioni delle reti RAI del 1995. Con la parallelizzazione informatica dei due corpora, è possibile da un lato delineare un preciso profilo dell'italiano trasmesso per radio, ricorrendo a un'analisi "verticale" dei due corpora e all'impiego delle marcature di tipo sociolinguistico, e dall'altro misurare la variazione

linguistica, soprattutto in relazione a neologismi e forestierismi, tramite la comparazione “orizzontale” fra le reti RAI nel 1995 e nel 2003 che trasforma la banca dati in un preciso osservatorio linguistico.

Nella marcatura del corpus non solo sono state individuate le 4 categorie descrittive “radio”, “speaker”, “tipologia comunicativa” e “genere radiofonico”, ma all’interno di queste sono state ulteriormente specificate delle sottocategorie di forte rilevanza sociolinguistica. Le 9 radio scelte costituiscono un campione rappresentativo delle varie tipologie presenti nell’etere (le 3 radio RAI, 3 radio di flusso, 2 radio comunitarie con l’aggiunta di Radio Vaticana per la sua importanza culturale) e pertanto la marcatura delle trascrizioni in funzione delle singole emittenti consente di analizzare le specificità legate ai diversi possibili modi di fare radio. Per quanto riguarda lo speaker la marcatura prevede da un lato una distinzione fra parlanti di sesso maschile e femminile, e dall’altro tra professionisti e interlocutori esterni. Fra le tipologie comunicative una doppia griglia descrittiva classifica il parlato in base al grado di spontaneità (“parlato esecutivo”, “programmato”, “semimprovvisato”, “spontaneo”) e alla modalità di comunicazione (“monologo”, “dialogo”, “telefonata”, “monologo a più voci”, “turno frammentato”). I generi radiofonici individuati prevedono sei categorie corrispondenti a diversi gradi di programmazione (“pubblicità”, “annunci”, “letteratura”, “notizie”, “intrattenimento culturale” e “intrattenimento leggero”). Il sistema di interrogazione consente da un lato di ricercare i singoli brani corrispondenti alle varie categorie e sottocategorie, e dall’altro di definire specifici *sub-corpora* sociolinguisticamente rilevanti, definendo dei parametri di ricerca mirati attraverso l’incrocio delle categorie e delle sottocategorie previste (così, ad esempio, è possibile effettuare una ricerca limitandola soltanto al parlato radiofonico di esterni maschi in trasmissioni di intrattenimento leggero).

Particolare attenzione è stata poi dedicata alla marcatura dei forestierismi non adattati e alla possibilità di ricercare tratti tipici del parlato come le sovrapposizioni, le mende, le interruzioni.

Oltre alla possibilità di interrogare il *corpus* in modo analitico nei 2 DVD è possibile anche ascoltare per intero le registrazioni con la relativa trascrizione in una “Sala d’ascolto” virtuale.

L’opera ha grande importanza anche dal punto di vista metodologico. La messa in parallelo di suono e trascrizione per la prima volta offre allo studioso la possibilità di osservare da vicino e in modo sistematico quell’italiano particolare che è il “trasmesso radiofonico”: un italiano “difficile” che deve ubbidire a esigenze di spettacolarità, di rispetto di tempi e di ritmi, di larga comprensibilità, ma che deve apparire quanto più possibile amichevole e quindi “facile” e spontaneo.

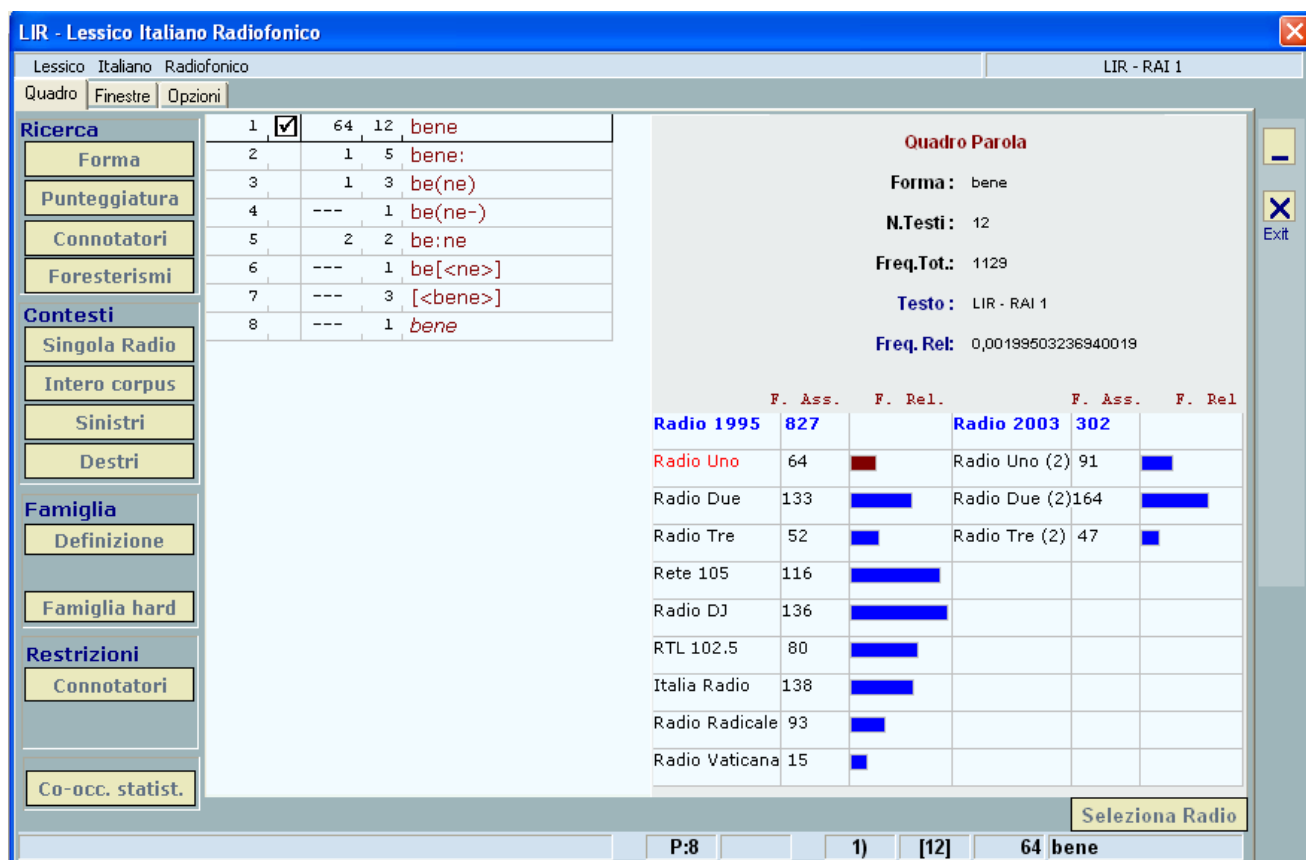


Fig. 1. Il “Quadro parola” della forma *bene*, con i dati delle frequenze assolute e gli istogrammi delle frequenze relative.

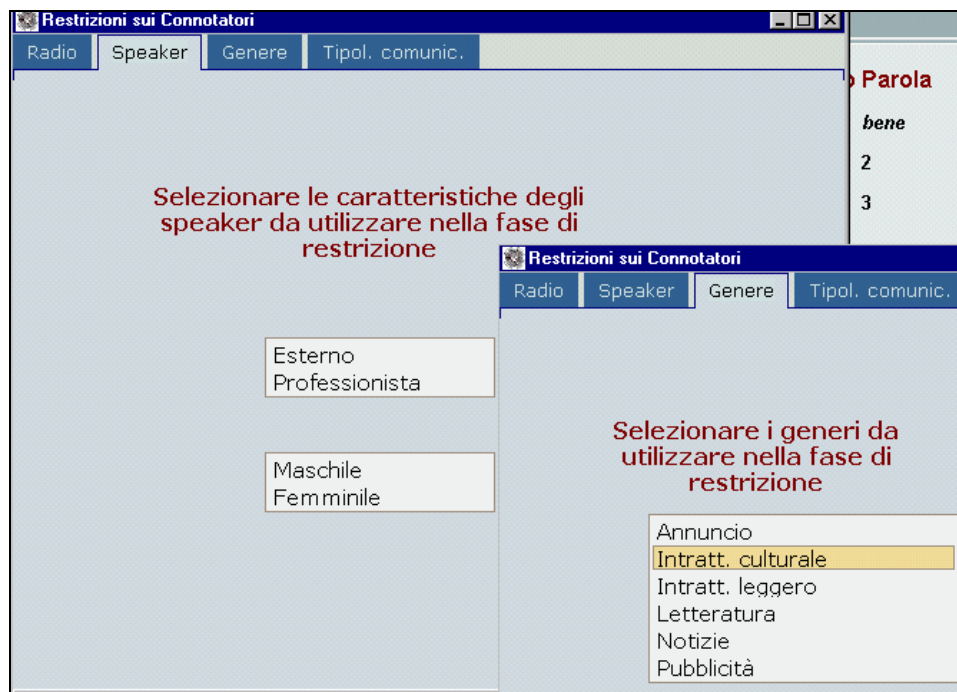


Fig. 2. La finestra di selezione dei connotatori.



Fig. 3. La finestra “Contesti” (nel secondo caso con l’indicazione dei connotatori).



Fig. 4. La finestra “Contesto allargato” (nel secondo caso con l’indicazione dei connotatori).

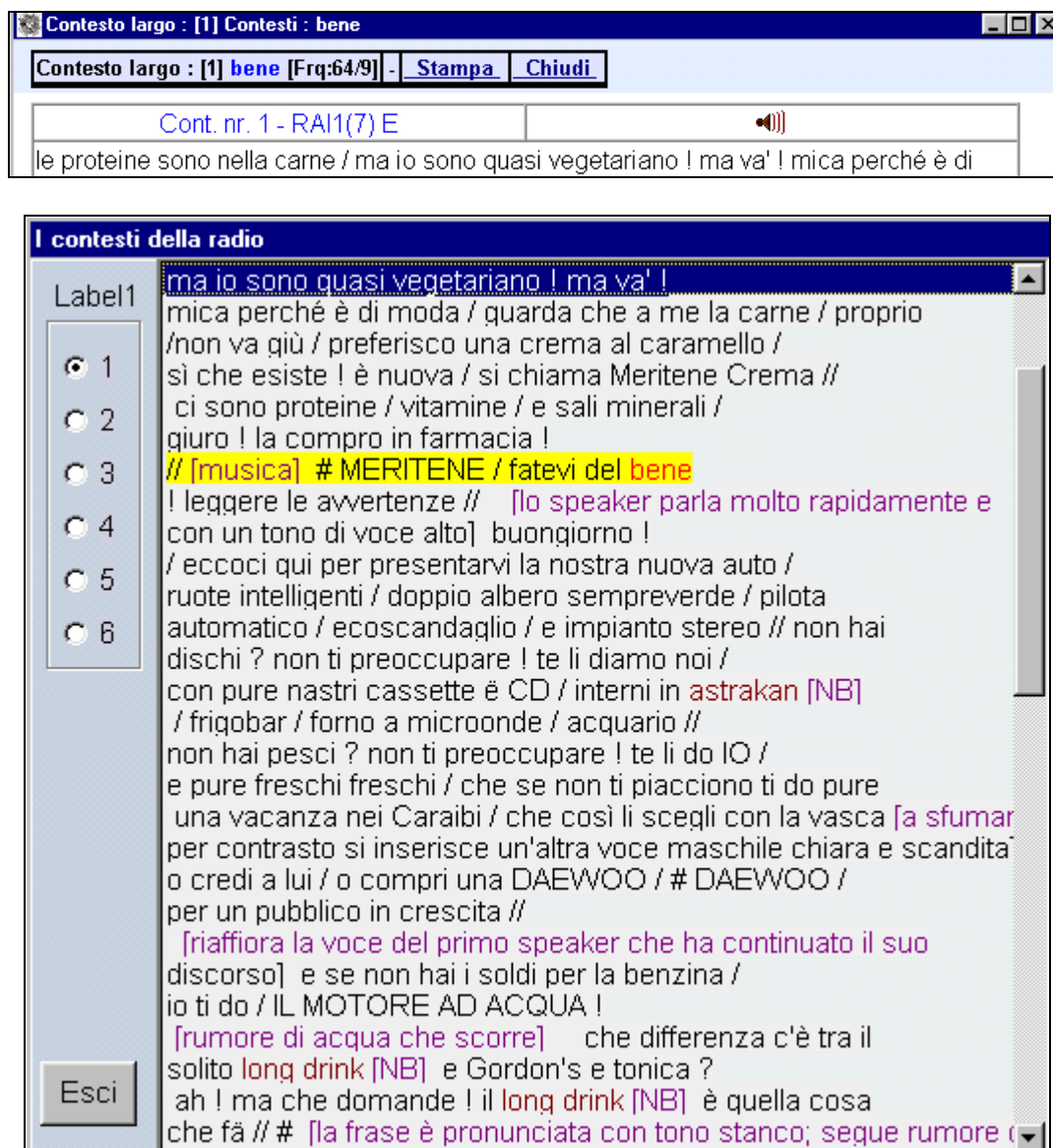


Fig. 5. Dal contesto allargato, cliccando sull'icona del megafono , si accede alla sala di ascolto del brano con la relativa trascrizione.

Il LIR in numeri:

- 9 radio a livello nazionale
- 144 ore di registrazioni
- 90 ore di parlato trasmesso
- 952.487 occorrenze
- 125.479 forme
- 1 giga di dati in formato digitale

Indici di frequenza decrescente di *LIF* e *LIR*: le 20 forme più usate

LIF = *Lessico di frequenza della lingua italiana contemporanea* a cura di U. Bortolini, C. Tagliavini e A. Zampolli, Milano, Garzanti, 1971.

LIP = *Lessico di frequenza dell'italiano parlato*, a cura di T. De Mauro, F. Mancini, M. Vedovelli e M. Voghera, Milano, Etaslibri, 1993.

<i>LIF</i>	<i>LIP</i>
1. il	1. il
2. di	2. di
3. egli	3. essere (v.)
4. a	4. uno (art.)
5. essere (v.)	5. a
6. e	6. e
7. uno	7. egli
8. in	8. non
9. che (pron.)	9. in
10. non	10. che (pron.)
11. essere (v. aus.)	11. io
12. io	12. che (cong.)
13. avere (v. aus.)	13. avere (v. aus.)
14. da	14. per
15. che (cong.)	15. fare
16. fare	16. dire
17. con	17. essere (v. aus.)
18. per	18. ci (pron.)
19. ma	19. da
20. avere (v.)	20. questo (agg.)

Schema riassuntivo tratti da: Valeria Della Valle, *Dizionari italiani: storia, tipi, struttura*, Roma, Carocci, 2005, p. 76.

Indici di frequenza decrescente del *LIR*: le

50 forme più usate

LIR 1995

1) di	18551
2) che	13705
3) e	13561
4) è	11003
5) il	10000
6) la	10000
7) in	8407
8) a	8120
9) un	7506
10) per	6883
11) non	6418
12) l'	5328
13) una	4907
14) del	4474
15) si	4310
16) i	4224
17) con	4200
18) della	3827
19) le	3813
20) da	3408
21) ma	3326
22) sono	3169
23) anche	3087
24) ha	3075
25) questo	2885
26) he	2670
27) al	2613
28) come	2431
29) sì	2363
30) dei	2355
31) perché	2352
32) più	2326
33) io	2175
34) ci	2113
35) no	1973
36) lo	1940
37) c'	1914
38) questa	1876
39) se	1810
40) dell'	1803
41) alla	1702
42) mi	1608
43) gli	1590
44) he:	1567
45) delle	1556
46) nel	1497
47) poi	1443
48) due	1355
49) eh	1323
50) su	1257

LIR 2003

1) di	8683
2) e	6635
3) che	6210
4) è	5281
5) il	5053
6) la	4765
7) a	4239
8) in	4108
9) un	3738
10) per	3517
11) non	2755
12) l'	2668
13) una	2239
14) con	2168
15) del	2127
16) si	1857
17) da	1829
18) della	1748
19) i	1719
20) le	1696
21) anche	1571
22) ma	1544
23) sono	1520
24) ha	1495
25) al	1414
26) sì	1238
27) questo	1230
28) più	1198
29) he	1180
30) come	1149
31) perché	1055
32) ci	992
33) he:	955
34) dei	933
35) lo	931
36) c'	908
37) no	892
38) dell'	871
39) tra	855
40) questa	781
41) gli	780
42) alla	777
43) nel	769
44) se	736
45) poi	709
46) due	698
47) io	696
48) mi	684
49) delle	677
50) quindi	660