DOI: 10.17469/O2106SLI000007

MARCO BIFFI, FRANCESCA CIALDINI

Banche dati per il trasmesso: il LIR e il LIT*

Il contributo ha lo scopo di descrivere due corpora dell'italiano, il LIR – *Lessico Italiano Radiofonico*, banca dati testuale e audio progettata per lo studio della lingua radiofonica, e il LIT – *Lessico Italiano Televisivo*, corpus testuale e audiovisivo sul web, che raccoglie un campione rappresentativo dell'italiano televisivo. I due corpora costituiscono un importante nucleo di base per rappresentare l'italiano trasmesso. Dopo aver ricostruito lo stato dell'arte, a partire dalla definizione di *trasmesso* dagli anni Ottanta, vengono illustrati nel dettaglio i due corpora, il metodo usato per la loro realizzazione, le modalità di interrogazione consentite dal motore di ricerca, le nuove prospettive di studio legate anche al concetto di sostenibilità.

Parole chiave: trasmesso orale, italiano radiofonico, italiano televisivo, diacronia, sostenibilità.

1. *Due corpora per l'italiano trasmesso: progetti, realizzazioni, sostenibilità*

Con il LIR (*Lessico italiano radiofonico*) e il LIT (*Lessico italiano televisivo*) la linguistica dei corpora ha rivolto la propria attenzione all'italiano trasmesso. Il LIR è stato concepito proprio quando Francesco Sabatini mise definitivamente a fuoco l'etichetta di *trasmesso*; dopo averla usata nel 1982 in un volume collettaneo dedicato all'educazione linguistica[1] e averla recuperata nella sua grammatica *La comunicazione e gli usi della lingua* nel 1984 (sottolineandone peraltro la non ortodossia con l'uso delle virgolette)[2], è nel 1994 che ne definisce i contorni e le caratteristiche, sdoganandola così per gli studi linguistici, in una relazione dal titolo, appunto, *Prove per l'italiano «trasmesso»*, al Convegno *Gli italiani trasmessi: la radio*, tenutosi a Firenze,

* Il contributo è frutto dell'elaborazione comune dei due autori, tuttavia si deve a Marco Biffi la stesura del paragrafo 1 e a Francesca Cialdini la stesura del paragrafo 2.

[1] Sabatini 1982.

[2] Sabatini 1984.

all'Accademia della Crusca, nel maggio del 1994, relazione poi pubblicata negli atti usciti nel 1997[3].

Fu proprio a seguito dei lavori di quel convegno che Nicoletta Maraschio ebbe l'idea di compilare un lessico della lingua radiofonica. E secondo la tradizione cruscante si pensò a un corpus informatizzato di partenza; un corpus che doveva però avere caratteristiche diverse da quelli visti fino a quel momento e a cui si era abituati[4]. Innanzi tutto, doveva dare accesso ai materiali autentici trascritti e informatizzati, caratteristica – si badi bene – che mancava anche al LIP (*Lessico di frequenza dell'italiano parlato*), che presentava gli indici lessicali in formato cartaceo, a stampa, e dava accesso soltanto alla trascrizione dei testi orali usati come base di partenza, peraltro a disposizione unicamente con un disco floppy da 3 pollici e mezzo allegato al volume[5]. Come è ben noto, sono successive le versioni consultabili attraverso il web, che prima hanno dato accesso diretto alla trascrizione dei testi e poi anche alla voce[6]. L'accesso al materiale autentico per il LIR era centrale: era infatti l'unica garanzia per lo studioso di poter analizzare tutte le specificità della lingua della radio, anche quelle soprasegmentali di intonazione e marcature espressive.

Quando si cominciò a progettare il LIR, a partire appunto dal 1994, gli ostacoli tecnologici erano enormi, a cominciare dalla possibilità di archiviare i file audio di un corpus di 108 ore di registrazione, che dovevano tradursi in circa 60 ore di parlato al netto della musica (un ordine di grandezza voluto perché il corpus fosse comparabile con il LIP). Si sfruttarono al massimo le tecnologie disponibili ed Eugenio Picchi realizzò una versione apposita del suo DBT (il DBT LIR), che consentiva – una volta agganciati i contesti allargati – di allineare in perfetto sincrono il file audio autentico: si cercava una forma e si arrivava ad ascoltarla nel punto esatto in cui era stata pronunciata, con

[3] Sabatini 1997.

[4] Sulla realizzazione del corpus LIR e sulla metodologia di ricerca in questa sede per motivi di spazio ci limitiamo a segnalare solo alcuni studi pubblicati nel corso degli anni. Si vedano almeno Maraschio *et al*. 1997; Maraschio *et al*. 2004; Alfieri & Stefanelli 2005; Biffi & Setti 2008.

[5] *LIP Lessico di frequenza dell'italiano parlato*, a cura di Tullio De Mauro, Federico Mancini, Massimo Vedovelli, Miriam Voghera, Milano, Etas, 1993.

[6] I testi sono disponibili nella *Banca Dati dell'Italiano Parlato*, consultabile all'indirizzo http://badip.uni-graz.at/it/; le registrazioni sono disponibili nel *VOLIP -Voce del LIP*: https://parlaritaliano.studiumdipsum.it/index.php/it/volip.

l'indicazione della trasmissione, del genere, della tipologia comunicativa e dello speaker (interno/esterno; maschio/femmina)[7].

Erano tempi pionieristici in cui non era ben chiara l'importanza strategica della sostenibilità. Eppure il LIR ha involontariamente precorso i tempi anche da questo punto di vista. La prima versione consultabile in locale divenne presto sempre più complicata da gestire e così è stata necessaria una prima implementazione e il trasferimento della banca dati sul web, con un nuovo software, che non garantiva l'elasticità e la potenza di ricerca della consultazione della versione in DBT, ma che però lo rendeva disponibile a tutti per scopi scientifici ma anche didattici. Tanto più che nel 2005 si erano avviati i lavori per realizzare un analogo corpus per la lingua della televisione, il LIT, che doveva avere caratteristiche analoghe al LIR con i dovuti adattamenti legati al mezzo[8]: in questo caso era necessario risalire nuovamente a un file audio-video autentico, perché per la lingua della televisione il contesto doveva essere allargato all'immagine per recuperare prossemica, postura, espressioni, combinazioni con altri codici.

Il nuovo software per il LIT, nativo web, fu adattato al LIR, tanto più che in questo modo i due corpora sarebbero stati omogenei e integrabili tra loro, come avviene effettivamente nel portale *Vivit – Vivi italiano*, all'interno della sezione *Archivi digitali*[9].

Il LIT è anche consultabile a partire dal *Portale dell'italiano televisivo*, realizzato all'interno di un PRIN a cui hanno collaborato vari gruppi di ricerca nazionali[10].

LIR e LIT si sono però dovuti nuovamente confrontare con la sostenibilità: la scelta fatta dagli informatici di appoggiarsi ad Adobe Flash Player si è dimostrata infatti esiziale quando – dopo gli annun-

[7] Cfr. Maraschio *et al.* 2004, pp. 21-34.

[8] Anche per il LIT ci limitiamo a segnalare alcuni studi pubblicati negli anni: si vedano almeno Mauroni & Piotti 2010 e Alfieri *et al.* 2016. Per la parte specificatamente informatico-linguistica cfr. Biffi 2010.

[9] *Vivit – Vivi italiano. Il portale dell'italiano nel mondo* è un archivio informatico di materiali e strumenti rivolti agli italiani all'estero, in particolare a quelli di seconda e terza generazione: https://www.viv-it.org/. Per la sezione degli *Archivi digitali* l'indirizzo diretto è il seguente: https://www.viv-it.org/schede/archivi-digitali.

[10] Il portale è il frutto del progetto PRIN 2008 *Il portale dell'italiano televisivo: corpora, generi e stili comunicativi* (unità di ricerca: Università di Firenze, Università di Catania, Università di Genova, Università di Milano, Università della Tuscia). È consultabile all'indirizzo www.italianotelevisivo.org, al quale si rimanda per approfondimenti.

ci di dismissione e la dichiarata dismissione – a partire dal primo gennaio 2021 tutto ciò che si basava su questo software ha cessato di funzionare.

Complice l'emergenza Coronavirus, quando abbiamo fatto la nostra proposta nel 2019 per il poster, LIR e LIT erano consultabili, ma quando si stava definendo il programma non lo erano già più.

Il LIR e il LIT attualmente consultabili si basano quindi su una nuova versione del programma di interrogazione (transitoria), sviluppata in questi ultimi mesi per traghettare questi due strumenti verso una nuova piattaforma che finalmente sarà realizzata con la massima attenzione alla sostenibilità.

Passiamo ora una descrizione più articolata dei due strumenti e delle loro funzionalità, illustrate anche nel divenire delle varie versioni descritte.

2. *I due corpora: descrizione e funzionalità*

2.1 Il LIR

Il LIR è una banca dati testuale e audio progettata con l'obiettivo di studiare la varietà radiofonica di trasmesso. Il primo corpus, rappresentativo delle principali emittenti nazionali, è stato raccolto nel 1995, secondo un prelievo a scacchiera su una settimana di maggio e comprende 108 ore di trasmissioni radiofoniche di 9 emittenti nazionali (Radio1, Radio2, Radio3, Radio Deejay, Rete 105, RTL 102.5, Italia Radio, Radio Radicale, Radio Vaticana)[11]. La banca dati raccoglie, dunque, 64 ore di parlato, 650.000 occorrenze, 86.000 forme.

Il secondo corpus, relativo al 2003 e limitato alle reti RAI, comprende 32 ore di parlato, 310.000 occorrenze, 39.000 forme; come per il primo corpus, i prelievi sono stati effettuati sulla settimana di maggio, meno caratterizzata da eventi esterni come Natale, Pasqua e periodi festivi[12].

Il lavoro di costituzione del corpus si è articolato in quattro fasi principali: 1) trascrizione del materiale audio; 2) inserimento manua-

[11] Per motivi tecnici sono state aggiunte alcune registrazioni di trasmissioni di Radio Deejay e RTL 102.5 in onda in una settimana del febbraio 1996.

[12] Su questo aspetto si veda lo studio di Biffi & Setti 2008: 351.

le dei testi nella piattaforma e creazione delle trasmissioni; 3) allineamento testuale con i file mp3 forniti dalle emittenti; 4) marcatura del testo con DBT di Eugenio Picchi, che – come spiegato sopra – ha poi realizzato una specifica versione del motore di ricerca (DBT LIR) per la consultazione integrata di testi e file audio.

Il materiale è stato trascritto interamente secondo criteri che, tenendo conto di una serie di elementi, permettono di individuare i principali tratti linguistici del trasmesso radiofonico. Ne riportiamo alcuni, a titolo esemplificativo:

- si segnala la fine di enunciato dichiarativo con la doppia barra obliqua //;
- la barra semplice / indica la scansione interna dell'enunciato, come eventuali pause o il cambio di intonazione;
- le parentesi uncinate < > indicano le sovrapposizioni di turno;
- {T} indica un troncamento;
- { indica l'inizio di una esitazione e } ne indica la fine.

Nello specifico, per quanto riguarda la marcatura del corpus, sono state individuate le quattro categorie di *Radio*, *Genere*, *Tipologia comunicativa*, *Speaker*. All'interno di queste categorie sono presenti alcune sottocategorie interessanti dal punto di vista della ricerca sociolinguistica (per esempio, oltre alle emittenti radiofoniche, *Tipologia comunicativa*: monologo, dialogo, telefonata, monologo a più voci, turno frammentato, esecutivo, programmato, semi-improvvisato, spontaneo; *Genere*: pubblicità, annunci, letteratura, notizie, intrattenimento culturale, intrattenimento leggero; *Speaker*: professionista, esterno, maschio, femmina). Con la presenza di categorie e sottocategorie possiamo sia fare la ricerca di singoli brani sia individuare specifici sub-corpora linguistici. Inoltre, è possibile interrogare sia separatamente il LIR 1995 (definito LIR/1) e il LIR 2003 (definito LIR/2) sia in contemporanea, in modo da permettere, almeno per le reti RAI, una prima valutazione dei cambiamenti in diacronia nella lingua della radio[13].

[13] È anche presente la sezione *Sala d'ascolto*, una stanza virtuale in cui è possibile selezionare una trasmissione e allo stesso tempo ascoltare la registrazione e seguire la trascrizione.

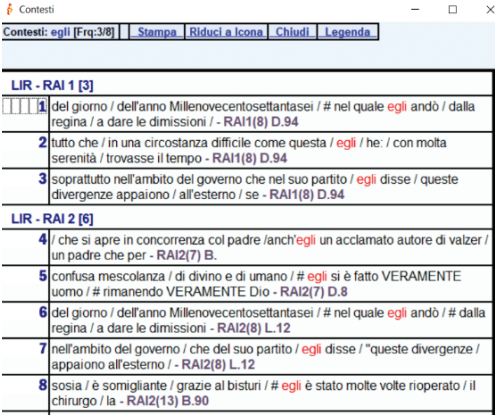
Figura 1 - *La distribuzione di* egli *alla radio nel 1995 e nel 2003*

Nel LIR è possibile effettuare ricerche di forme di vario tipo, a partire da quelle che riguardano la veste fonica e che consentono riflessioni sui tratti della pronuncia, fino a quelle relative a fenomeni morfosintattici tipici dell'italiano contemporaneo. Per esempio, come è stato messo in evidenza anche in altri studi, è interessante osservare la distribuzione dei pronomi di terza persona *egli*/*lui* e analizzare in particolare l'uso di *egli* tra 1995 e 2003 (36 nel LIR/1 e 6 nel LIR/2, per un totale di 42 occorrenze)[14]. Per ciascun risultato ottenuto è possibile allargare il contesto, ascoltare la porzione di testo che interessa, ricevere informazioni sui metadati e avere a disposizione l'intera trascrizione della trasmissione. Tra le ricerche possibili ricordiamo anche quella dei forestierismi, interessante anche in una prospettiva diacronica: nel LIR/1 1995, per quanto riguarda la RAI, risultano 509 forestierismi; nel LIR/2 2003 (solo RAI, come abbiamo detto) se ne contano 1050[15].

[14] Si vedano Maraschio 2005: 140-141 e Biffi & Setti 2008: 355.

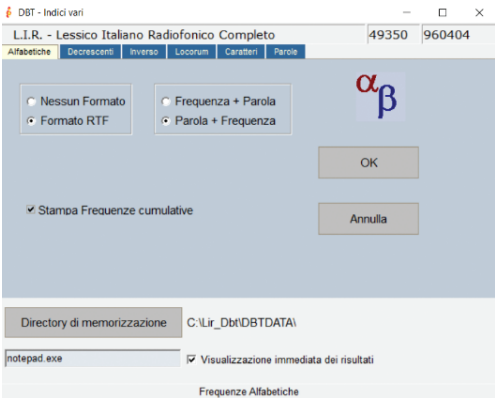
[15] Sui forestierismi nel LIR si veda gli studi di Fanfani 1997a e Fanfani 1997b. Come nota anche Maraschio 2016: 77, dagli anni Novanta a oggi è entrato in italiano un numero di forestierismi, in particolare anglismi, superiore al doppio di quello entrato nel decennio precedente.

Figura 2 - *Il contesto allargato*



Infine, la sezione "Indici" consente la generazione di indici statistici e di frequenza.

Figura 3 - *Indici vari*



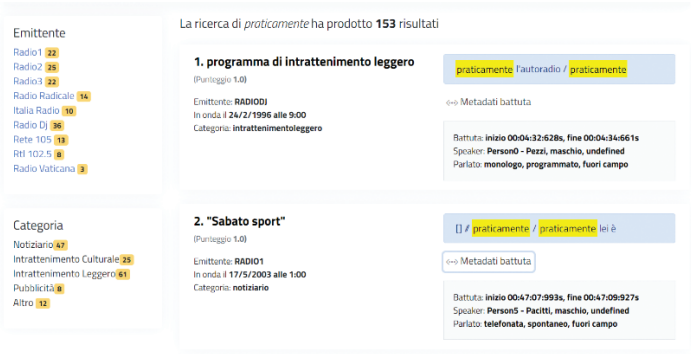
2.1.1 Il LIR web

La versione web del LIR[16] presenta un'interfaccia di interrogazione in cui sono possibili diverse tipologie di ricerche, a partire da una ricerca semplice e da una avanzata. La più immediata è quella che riguarda

[16] http://lir.accademiadellacrusca.org/lir2/.

una singola forma; per esempio, se cerchiamo *praticamente*, avverbio diffuso alla radio, otteniamo i seguenti risultati:

Figura 4 - *I risultati di ricerca*



I dati ottenuti, in primo luogo, sono relativi alle occorrenze totali (153 risultati), con le frequenze distribuite per emittente e per categoria; sono presenti inoltre i contesti immediati, che contengono le informazioni sulla trasmissione (titolo, emittente, data, categoria). È possibile anche consultare i metadati della battuta relativi a parlante, tipologia comunicativa, presenza in campo o fuori campo. Cliccando su uno dei contesti si accede al contesto allargato della trascrizione e alla registrazione audio digitale parallelizzata.

Figura 5 - *La registrazione audio parallelizzata*

Con la ricerca avanzata sono possibili ricerche più raffinate. Infatti, è possibile la ricerca di più forme attraverso l'operatore booleano AND ("Trova i risultati che contengano tutte le seguenti parole"), di una sequenza ("Trova i risultati che contengano la seguente sequenza libera", sia come "sequenza esatta", sia "sequenza ordinata" sia "sequenza non ordinata") e di forme attraverso gli operatori OR e AND NOT ("Trova i risultati che contengano una qualunque delle seguenti parole" e "Trova i risultati che non contengano le seguenti parole"). La distanza misurata in parole viene stabilita dall'utente. Per esempio, possiamo cercare la sequenza esatta *piuttosto che* (con distanza 1), per osservare la diffusione del suo uso con valore disgiuntivo in diacronia. Analizzando i contesti dei 36 risultati totali ottenuti, notiamo che le occorrenze di *piuttosto che* con valore disgiuntivo sono 5 nel corpus del 1995 e salgono a 7 nel 2003.

Figura 6 - *I risultati relativi a* piuttosto che

Inoltre, è possibile restringere la ricerca in base all'emittente, ai generi e alle tipologie e considerare le maiuscole/minuscole, ottenere l'elenco delle forme e le statistiche globali. Nella ricerca è consentito l'utilizzo dei caratteri *jolly*: il *?* sostituisce un carattere e l'asterisco *estende la ricerca a intere parti di parola. Se cerchiamo, per esempio, la forma *maxi** (*maxi-* è tra i prefissi più produttivi in LIR1/2), risulteranno 20 occorrenze. È possibile sia accedere ai contesti sia ottenere l'elenco delle forme:

Figura 7 - *L'elenco delle forme*


La ricerca di *maxi** ha prodotto **20** risultati

Forma	Num. occorrenze
maxi	15
maxiemendamento	1
maxiprocessi	1
maxirichiesta	1
maxitamponamento	1

2.2 Il LIT

Il LIT è una banca dati interrogabile che raccoglie 168 ore di trasmissioni delle reti Rai e Mediaset, prelevate nel corso del 2006 secondo una griglia statisticamente rappresentativa. Nel corpus è possibile ricercare una parola (o un gruppo di parole), accedere a dati quantitativi sulla frequenza, ai contesti immediati e al materiale autentico trascritto e marcato secondo vari parametri[17]. Come ricordato, il corpus LIT condivide le caratteristiche del LIR sia per quanto riguarda l'individuazione del campione rappresentativo di riferimento sia nelle modalità di interrogazione. Infatti, i prelievi sono stati effettuati nell'arco temporale dell'anno 2006 (dal 2 gennaio al 26 dicembre), interessante dal punto di vista linguistico per gli eventi che lo hanno caratterizzato – per esempio le Olimpiadi, i mondiali di calcio e le elezioni politiche –, nella fascia oraria serale che va dalle 19.00 alle 23.00, secondo una griglia rappresentativa dal punto di vista statistico. Per ciascuna rete è stato effettuato un prelievo di mezz'ora alla settimana a rotazione, in modo da coprire tutti i giorni della settimana nell'arco dell'anno[18].

[17] Per una prima descrizione del corpus si veda Biffi 2010. La banca dati è consultabile all'indirizzo http://lit.accademiadellacrusca.org/lit2/; come ricordato sopra, una versione precedente è consultabile nel *Portale dell'italiano televisivo* https://www.italianotelevisivo.org/ e negli *Archivi digitali* del *Vivit* https://www.viv-it.org/schede/archivi-digitali.

[18] Per esempio, lo schema di prelievo delle prime due settimane è il seguente: 1ª settimana, lunedì 19.00-19.30 Canale5 (02/01/06), martedì 19.30-20.00 Rete4 (03/01/06), mercoledì 20.00-20.30 Italia1 (04/01/06); 2ª settimana, giovedì 20.30-21.00 Canale5 (12/01/06), venerdì 21.00-21.30 Rete4 (13/01/06), sabato 21.30-22.00 Italia1 (14/01/06). Sui prelievi e sul metodo usato si veda Biffi 2010.

Dal punto di vista metodologico, il lavoro di costituzione del corpus si è articolato in quattro fasi principali: 1) trascrizione del materiale audiovisivo; 2) inserimento dei testi nella piattaforma; 3) allineamento testuale con i file video forniti da Rai e Mediaset; 4) marcatura del testo con annotazioni XML/TEI.

I criteri di trascrizione del materiale consistono in una versione semplificata dei criteri utilizzati per il LIR e permettono di individuare i principali tratti linguistici del trasmesso. Per esempio:

- si segnala la fine di enunciato dichiarativo con la doppia barra obliqua //;
- la barra semplice / indica la scansione interna dell'enunciato, come eventuali pause o il cambio di intonazione;
- le parentesi uncinate < > indicano le sovrapposizioni di turno;
- il trattino – segnala l'interruzione di parola (attaccato alla parola interrotta, es.: *veram-*);
- i due punti *:* indicano l'allungamento dovuto a esitazione;
- nel caso di parola o parte di parola incomprensibile, la forma linguistica è sostituita da *[xxx]*;
- il maiuscolo indica un fenomeno di enfasi[19].

Una volta immesse nella piattaforma le trascrizioni del materiale audiovisivo e create le trasmissioni (complete delle informazioni di *titolo*, *rete*, *data*, *ora*), sono state individuate alcune categorie funzionali alla ricerca. Oltre alle emittenti, sono presenti le seguenti categorie relative ai generi e sottogeneri[20]: 1) *Pubblicità*; 2) *Fiction*: Film TV, Miniserie, Serie, Serial; 3) *Intrattenimento*: Cartoni animati, Varietà, Game show, Reality show, Comico/satirico, Talk show Intrattenimento; 4) *Informazione*: TG, Reportage/inchiesta, Telecronaca in diretta, Approfondimento, Talk show Informazione; 5) *Divulgazione scientifica e culturale*: Documentari, Magazine, Talk show Divulgazione.

[19] Altri criteri di trascrizione sono: punto interrogativo *?* alla fine di enunciato interrogativo; punto esclamativo *!* alla fine di enunciato esclamativo; puntini di sospensione ... per l'intonazione sospensiva; apici doppi " " per indicare titoli, citazioni, discorso diretto riportato; parentesi quadre *[ ]* per commenti paralinguistici e situazionali. Le sigle uniformate sono le seguenti: *Rai1*, *Rai2*, *Rai3*, *Canale5*, *Italia1*, *Rete4*; *Tg1*, *Tg2*, *Tg3*, *Tg4*, *Tg5*, *Studio Aperto*.

[20] Come nota Biffi 2010: 45, la categoria generi/sottogeneri è stata quella più complicata da definire, dato l'ibridismo dei generi tipico della neotelevisione.

Inoltre, sono state individuate le categorie che fanno riferimento alla tipologia comunicativa (*monologo*, *dialogo*), al tipo di parlato legato al grado di spontaneità (*esecutivo*, *programmato*, *improvvisato*), al parlante, descritto in base ad alcuni parametri sociolinguistici (*interno* se si tratta del professionista della televisione, esterno se non fa parte del mondo della tv; *uomo*/*donna*; *in campo*/*fuori campo*).

Nella fase successiva del lavoro il testo trascritto è stato allineato con i file video forniti da Rai e Mediaset e marcato con con specifiche annotazioni XML/TEI. Per ciascuna porzione di battuta è stata definita un'annotazione con marcatori XML/TEI in base alle categorie sopra descritte. In particolare, la classificazione in base al parlante, al quale vengono associate le porzioni di testo trascritto (a ciascun parlante viene attribuito un colore di riferimento, con lo scopo di facilitare l'individuazione del cambio turno), costituisce un aspetto rilevante per la ricerca[21].

Come per il LIR, anche per il LIT è possibile sia una ricerca semplice sia una più avanzata. Se cerchiamo, per esempio, un aggettivo come *straordinario*, otteniamo i seguenti risultati:

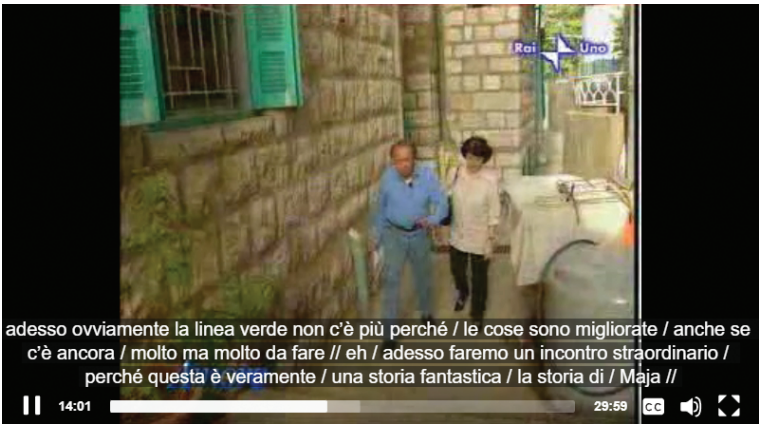
Figura 8 - *I risultati*

Le informazioni ottenute sono relative alle occorrenze totali (61 risultati), con le frequenze distribuite per emittente e per categoria. Sono presenti, inoltre, i contesti immediati, che contengono le informazioni sulla trasmissione: titolo, emittente, data, categoria. È possibile consultare anche i metadati della battuta relativi a parlante, tipologia

[21] Su questo aspetto si veda lo studio di Biffi 2010: 45-50.

comunicativa, presenza in campo o fuori campo; cliccando su uno dei contesti si accede al contesto allargato della trascrizione e alla registrazione audio digitale parallelizzata:

Figura 9 - *Il collegamento alla registrazione audio parallelizzata*



Come nel LIR, sono possibili le ricerche con i caratteri jolly: per esempio, *straordinari** consentirà di trovare le forme di maschile plurale e di femminile singolare e plurale e la forma dell'avverbio in *-mente*. Inoltre, la ricerca avanzata consente di ottenere risultati più raffinati: è possibile, infatti, la ricerca di più forme ("Trova i risultati che contengano tutte le seguenti parole"), di una sequenza ("Trova i risultati che contengano la seguente sequenza libera", sia come "sequenza esatta", sia "sequenza ordinata" sia "sequenza non ordinata") e di forme attraverso gli operatori OR e AND NOT ("Trova i risultati che contengano una qualunque delle seguenti parole" e "Trova i risultati che non contengano le seguenti parole"). La distanza misurata in parole viene stabilita dall'utente.

Infine, è possibile restringere la ricerca in base all'emittente, ai generi (e sottogeneri), alla tipologia comunicativa, alla tipologia di parlato, al parlante (maschio/femmina, interno/esterno, in campo/fuori campo) e considerare le maiuscole/minuscole, ottenere l'elenco delle forme e le statistiche globali.

Figura 10 - *La ricerca avanzata*

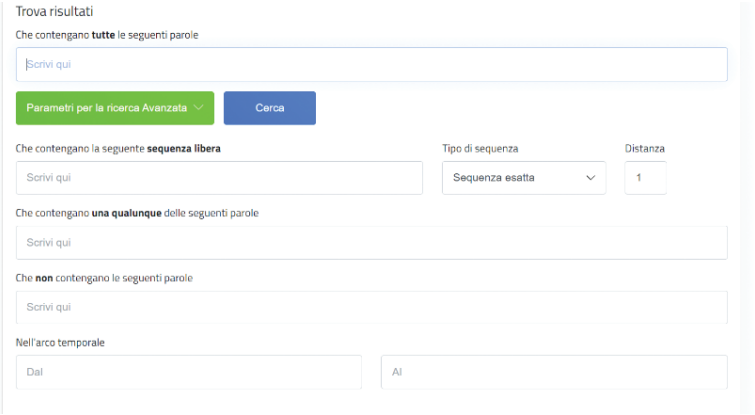
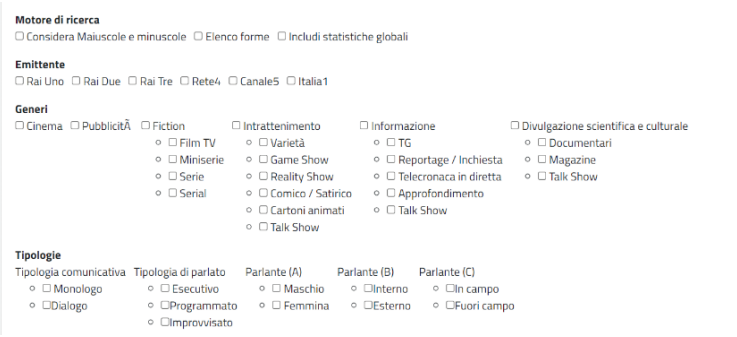


Figura 11 - *La ricerca per categorie*



Riferimenti bibliografici

Alfieri, Gabriella & Stefanelli, Stefania. 2005. Lessici dell'italiano radiofonico (LIR). In Burr, Elisabeth (a cura di), *Tradizione & innovazione. Il parlato: teoria – corpora – linguistica dei corpora. Atti del VI Convegno SILFI, Suisburg, 28 giugno-2 luglio 2000*. Firenze: Cesati. 397-412.

Alfieri, Gabriella & Biffi, Marco & Giuliano, Mariella & Motta, Daria (a cura di). 2016. *Il portale della tv, la tv dei portali, Atti del Convegno (Firenze, Accademia della Crusca, 8 marzo 2013)*. Acireale – Roma: Bonanno.

Biffi, Marco. 2010. Il LIT – Lessico Italiano Televisivo. In Mauroni, Elisabetta & Piotti, Mario (a cura di), *L'italiano televisivo 1976-2006*. Firenze: Accademia della Crusca. 35-70.

Biffi, Marco. 2016. Il portale dell'italiano televisivo: corpora, generi, stili comunicativi. In Alfieri, Gabriella & Biffi, Marco & Giuliano, Mariella & Motta, Daria (a cura di), *Il portale della tv, la tv dei portali. Atti del Convegno (Firenze, Accademia della Crusca, 8 marzo 2013)*, 11-30. Acireale – Roma: Bonanno.

Biffi, Marco & Setti, Raffaella. 2008. Dieci anni di italiano parlato alla radio: corpora LIR 1995/ LIR 2003 a confronto. In Pettorino, Massimo (a cura di), *La comunicazione parlata, Atti del Congresso Internazionale (Napoli 23-25 febbraio 2006)*, 361-398. Napoli: Liguori.

Bonomi, Ilaria & Maraschio, Nicoletta. 2016. *Giornali, radio e tv: la lingua dei media* (Collana "l'Italiano. Conoscere e usare una lingua formidabile", a cura dell'Accademia della Crusca e Repubblica, n. 8) [poi ristampato nel 2017]. Roma: Gruppo Editoriale L'Espresso.

Cialdini, Francesca. 2016. L'aggiornamento della banca dati LIT e il DIALIT. In Alfieri, Gabriella & Biffi, Marco & Giuliano, Mariella & Motta, Daria (a cura di), *Il portale della tv, la tv dei portali. Atti del Convegno (Firenze, Accademia della Crusca, 8 marzo 2013)*, 31-47. Acireale – Roma: Bonanno.

Fanfani, Massimo. 1997a. Forestierismi alla radio. In *Gli italiani trasmessi: la radio*, 729-788. Firenze: Accademia della Crusca.

Fanfani, Massimo. 1997b. I programmi radiofonici della RAI. *Bollettino d'informazioni* VII (1-2). 73-77.

LIP *Lessico di frequenza dell'italiano parlato*, a cura di Tullio De Mauro, Federico Mancini, Massimo Vedovelli, Miriam Voghera, Milano, Etas, 1993.

Maraschio, Nicoletta. 2005. La Radio. In Lo Piparo, Franco & Ruffino, Giovanni (a cura di), *Gli italiani e la lingua*, 135-146. Palermo: Sellerio Editore.

Maraschio, Nicoletta. 2016. La radio. In Bonomi, Ilaria & Maraschio, Nicoletta (a cura di), *Giornali, radio e tv: la lingua dei media* (Collana "l'Italiano. Conoscere e usare una lingua formidabile", a cura dell'Accademia della Crusca e Repubblica, n. 8) [poi ristampato nel 2017], 53-98. Roma: Gruppo Editoriale L'Espresso.

Maraschio, Nicoletta & Antonini, Anna & Bellucci, Patrizia & Fanfani, Massimo & Stefanelli, Stefania & Avesani, Cinzia & Pratesi, Monica. 1997. Il progetto LIR. I lessici di frequenza dell'italiano radiofonico. *Bollettino d'informazioni* VII (1-2). 53-94.

Maraschio, Nicoletta & Stefanelli, Stefania & Buccioni, Stefania & Biffi, Marco. 2004. Dal corpus LIR: prove e confronti lessicali. In Albano

Leoni, Federico & Cutugno, Francesco & Pettorino, Massimo & Savy Renata (a cura di), *Il Parlato Italiano, Atti del Convegno Nazionale "Il Parlato Italiano", Napoli 13-15 febbraio 2003*, 1-36. Napoli: M. D'Auria Editore.

Sabatini, Francesco. 1982. La comunicazione orale, scritta, trasmessa: la diversità del mezzo, della lingua e delle funzioni. In Boccafurni, Anna Maria & Serromani, Simonetta (a cura di), *Educazione linguistica nella scuola superiore. Sei argomenti per un curricolo*, 105-127. Roma: Istituto di Psicologia CNR.

Sabatini, Francesco. 1984. *La comunicazione e gli usi della lingua: pratica, analisi e storia della lingua italiana.* Torino: Loescher.

Sabatini, Francesco. 1997. *Prove per l'italiano «trasmesso»*, In *Gli italiani trasmessi: la radio*, 11-30. Firenze: Accademia della Crusca.

Setti, Raffaella. 2011. Interrogando il LIT. Il lessico televisivo contemporaneo tra spettacolarità e stereotipia. Lo spettacolo delle parole. In Caffarelli, Enzo & Fanfani, Massimo (a cura di), *Lo spettacolo delle parole. Studi di storia linguistica e onomastica in onore di Sergio Raffaelli*, 167-182. Roma: Società Editrice Romana.