

**Università degli studi di Salerno**

Dipartimento di Scienze Politiche e della Comunicazione



Dottorato di ricerca in

Scienze del Linguaggio, della Società, della Politica e dell'Educazione

**Curriculum**

“Scienze della Comunicazione”  
XXXIV Ciclo

**Tesi di dottorato**

Comprensione e complessità testuale.  
“Un'analisi linguistico-computazionale  
della leggibilità dei testi”

Coordinatore: Prof. Filippo Fimiani

Tutor: Prof. Annibale Elia

A handwritten signature in blue ink, appearing to read 'Lorenza Melillo', is positioned to the right of the tutor's name.

Candidata: Lorenza Melillo  
Matr. 8801400081

## Sommario

INTRODUZIONE.....	5
CAPITOLO I .....	7
ALCUNE QUESTIONI PRELIMINARI .....	7
1. La letteratura per l'infanzia .....	7
2. Scelta dei testi .....	14
CAPITOLO II .....	20
COMPRESIONE E LEGGIBILITA' .....	20
1. La comprensione come problema e gli indici di leggibilità .....	20
2. Comprensione e leggibilità.....	28
CAPITOLO III .....	33
L'APPROCCIO STATISTICO LINGUISTICO.....	33
1. La statistica linguistica.....	33
2. Linguistica dei corpora e linguistica computazionale.....	40
3. Comprensione del linguaggio e disambiguazione.....	45
4. La statistica testuale.....	51
CAPITOLO IV .....	58
IL LESSICO-GRAMMATICA DELL'ITALIANO DI BASE.....	58
1. Il lessico-grammatica.....	58
2. La Grammatica .....	61
3. Il Lessico .....	95
4. Analisi lessico-grammaticale di un testo .....	141
CAPITOLO V .....	151
L'ANALISI TESTUALE .....	151
1. Vocabolario di base della lingua italiana: descrizione della risorsa .....	151
2. Il Nuovo Vocabolario di Base della Lingua Italiana .....	155
3. Analisi testuale e marche d'uso .....	157
4. Valutazione automatica della leggibilità e complessità sintattica .....	162
5. Struttura dell'analisi .....	164
6. Complessità sintattica e semplicità lessicale.....	167

7. Indice di leggibilità automatico (ILA).....	172
8. La valutazione della comprensione.....	173
9. Analisi quantitativa dei dati.....	175
Conclusioni.....	180
Appendice.....	183
Bibliografia.....	185
Allegato 1: Prova di comprensione.....	190
Allegato 2: Questionario relativo all'attività di lettura.....	191

## **Ringraziamenti**

Nel corso di questa ricerca ho avuto la fortuna di essere supportata dall'incoraggiamento, dai consigli, dalle indicazioni e dalle osservazioni critiche, oltre che del mio tutor, anche di alcuni colleghi Alessandro Maisto e Serena Pelosi a cui desidero esprimere la mia gratitudine.

Un ringraziamento particolare è riservato alla paziente e preziosa guida nei territori impervi del lessico-grammatica offertami dal mio professore Annibale Elia.

Ringrazio inoltre la maestra Carmelina Iandiorio che mi ha supportato per la buona riuscita della sperimentazione presso l'IC Don Milani.

Dedico questo lavoro a mia figlia Céline, una piccola lettrice curiosa, che lo ha ispirato, e al mio caro papà che sarebbe stato molto fiero della sua bambina.

Niente come la lettura di un libro,  
nell'apparente quiete e nel silenzio può  
dischiudere in modo imprevedibile la vista di  
nuovi orizzonti di vita.

Tullio De Mauro

## INTRODUZIONE

In questo lavoro di ricerca abbiamo analizzato testi, non strutturati, da una prospettiva lessicale e sintattica, grazie all'applicazione di un Indice di Leggibilità automatico da noi sviluppato. La prima parte del lavoro è dedicata al processo di comprensione di un testo che richiede un recupero ed una integrazione di tutte le conoscenze che il lettore possiede per la rielaborazione. Queste conoscenze sono dette enciclopediche e permettono al lettore di generare e elaborare ipotesi interpretative riguardo alle informazioni implicite del testo.

Sulla base di questa considerazione possiamo sottolineare che il lettore, da un lato deve avere disponibili ed attive nella mente delle conoscenze pregresse che possono operare come “ganci” ai quali attaccare le nuove conoscenze da apprendere, dall'altro deve essere in grado di realizzare delle inferenze.

L'estrazione dei significati avviene, quindi, attraverso processi inferenziali, ossia attraverso forme di ragionamento e di elaborazione delle informazioni non esplicite nel testo. Durante la lettura le inferenze si possono creare in modo automatico senza esserne consapevoli (Lumbelli 2009).

Per verificare il grado di leggibilità e di comprensibilità di un testo siamo ricorsi ad un approccio quantitativo che è andato ad analizzare l'aspetto lessicale e sintattico dei testi. Ciò premesso però non abbiamo sottovalutato l'approccio qualitativo della comprensione che invece ci aiuta ad evidenziare gli ostacoli legati alle difficoltà di chi legge.

Di conseguenza il lavoro tiene conto dei due aspetti, che se pur legati tra loro riguardano sfere diverse dell'approccio alla lettura.

Lo scopo della nostra ricerca è stato quello di verificare se fosse possibile analizzare la comprensibilità di un testo, attraverso l'analisi del lessico e della sintassi.

È stato Tullio De Mauro che ha realizzato la prima ricerca lessicografica finalizzata a individuare diversi sottoinsiemi del lessico della lingua italiana in base alla frequenza. De

Mauro voleva individuare il vocabolario di base dell'italiano, cioè l'insieme delle parole italiane che, in base alla loro frequenza, potessero essere considerate della più alta comprensione possibile (De Mauro 1980).

Dal punto di vista linguistico la migliore valutazione della comprensibilità di un testo è l'analisi del lessico e delle strutture sintattiche, in quanto l'uso di una parola e la sua comprensione coinvolge direttamente il grado di conoscenza della lingua posseduto dal lettore. Il livello di conoscenza della lingua, inteso come numero di parole delle quali si conosce e utilizza il significato, è dunque uno dei parametri fondamentali da valutare quando si cerca di misurare e prevedere la comprensibilità di un testo.

Il lavoro di sperimentazione che abbiamo effettuato si è basato, innanzitutto, sull'informatizzazione del NVdB (Elia et al. 2021), cioè sull'inclusione delle marche d'uso di Tullio De Mauro nei dizionari elettronici del modulo italiano del software Nooj (Vietri 2008).

L'ipotesi da cui si è partiti è stata di poter verificare la maggior presenza di parole del vocabolario fondamentale nei testi di narrativa per bambini, attraverso l'applicazione dei dizionari elettronici e delle grammatiche sintattiche. L'obiettivo di questo tipo di analisi è stato di dimostrare che esiste una correlazione tra il lessico utilizzato in un testo, la sua classificazione, fatta attraverso le marche d'uso e il livello di complessità del testo. In sintesi abbiamo voluto dimostrare, attraverso l'utilizzo della linguistica computazionale, quanto aveva già compreso Tullio de Mauro, cioè che l'"uso" rappresenta un punto di svolta per lo studio del lessico.

Alla verifica della scelta dei vocaboli in relazione all'ipotesi di De Mauro abbiamo fatto seguire una sperimentazione che si è basata sull'analisi della complessità sintattica, elaborando così un Indice di Leggibilità lessicale e sintattico.

L'applicazione del nostro Indice di Leggibilità Automatico (ILA) ci ha consentito in via predittiva, cioè prima di verificarlo attraverso l'esperienza del lettore, di controllare in che misura un testo sia comprensibile. I risultati dell'analisi automatica ottenuti sono stati verificati, poi, su un campione di lettori delle classi III, IV e V primaria dell'Istituto Comprensivo Don Milani di Manocalzati (AV).

# CAPITOLO I

## ALCUNE QUESTIONI PRELIMINARI

### 1. La letteratura per l'infanzia

Nel panorama degli studi linguistici la lingua della letteratura per l'infanzia può considerarsi come un'area che a lungo non è stata approfondita e lo hanno ben sottolineato Boero e De Luca, «accade di frequente che anche a persone di buona cultura manchi il senso stesso di letteratura per l'infanzia: la loro conoscenza è limitata alle proprie esperienze dirette, ai titoli dei libri letti negli anni della propria infanzia, e la scoperta che esiste uno specifico ambito di studi e di ricerche di letteratura per ragazzi li stupisce» (Boero/De Luca 1995).

L'indagine Istat del 2019<sup>1</sup> rivela che le opere librarie per ragazzi raggiungono il 10% dei titoli pubblicati. In termini di tiratura, le opere scolastiche e per ragazzi coprono una considerevole quota di mercato: circa quattro copie stampate su dieci sono libri per la scuola (28,4% della tiratura) ed una su sei sono libri per ragazzi (16,1%). In particolare, tra le pubblicazioni per ragazzi si ha un'elevata incidenza di opuscoli, pari al 44,8% del totale, in buona parte pubblicazioni destinate a un pubblico di bambini.

Oltre un libro di narrativa per ragazzi su tre (il 34,1%) è tradotto in una lingua straniera, con una netta prevalenza dell'inglese (17,2%).

Nel 2019 rimane stabile rispetto all'anno precedente il numero di lettori di libri: sono il 40,0% delle persone di 6 anni e più. A partire dall'anno 2000, quando la quota di lettori era al 38,6%, l'andamento è stato crescente fino a toccare il massimo nel 2010 con il 46,8% per poi diminuire di nuovo fino a tornare, nel 2016, al livello del 2001 (40,6%), stabile fino al 2019. La quota più alta di lettori continua a essere quella dei giovani: 54,1% nel 2019 tra i 15 e i 17 anni, e 56,6% tra gli 11 e i 14 anni.

In assoluto, il pubblico più affezionato alla lettura è rappresentato dalle ragazze tra gli 11 e i 19 anni (oltre il 60% ha letto almeno un libro nell'anno).

---

<sup>1</sup> Istat Produzione e lettura di libri in Italia Anno 2019 Intensa la produzione libraria, stabile il numero di lettori data di Pubblicazione 11 gennaio 2021

La lettura è fortemente influenzata dall'ambiente familiare: i bambini e i ragazzi sono certamente favoriti se i genitori hanno questa abitudine. Tra i ragazzi sotto i 18 anni legge il 77,4% di chi ha madre e padre lettori e solo il 35,4% tra coloro che hanno entrambi i genitori non lettori. In particolare i lettori più piccoli (6-10 anni) risentono maggiormente della presenza della sola madre lettrice (58,9% legge) mentre dopo i 15 anni, nonostante nessuno dei due genitori legga, ben il 40,6% di questi ragazzi lo fa.

È però un dato importante da considerare che i libri per l'infanzia sono letti anche da ragazzi che in età adulta abbandoneranno la lettura non legata a motivi professionali, si tratta di lettori spesso indotti da insegnanti e genitori.

Nonostante l'importanza che ricopre la narrativa per ragazzi in ambito editoriale poche volte i libri che vengono letti da bambini e ragazzi sono stati oggetto di studio.

E proprio su questo tipo di letteratura ha pesato a lungo il severo giudizio di Croce (1943)<sup>2</sup>:

l'arte "per bambini" (ecco la pregiudiziale) non sarà mai arte vera. [...] Lo splendido sole dell'arte pura non può essere sostenuto dall'occhio ancor debole dei bambini e dei fanciulli. [...] Basta il semplice riferimento al pubblico bambinesco, come a un dato fisso del quale faccia d'uopo tenere stretto conto, per turbare il lavoro artistico, e introdurre qualcosa ora di superfluo ora di manchevole, non ubbidiente più alla libertà e necessità interna della visione.

A tal proposito bisogna ricordare che Croce tra i libri infantili salvava solamente Pinocchio che riteneva «il più bel libro della letteratura infantile italiana», un unicum, un'eccezione che, in quanto tale, non apparteneva «alla speciale e ordinaria letteratura infantile congeniata e calcolata per bambini, perché in questo caso si tratterebbe di un più o meno abile prodotto pedagogico altrimenti pratico, privo di vita e pregio artistico» (Croce 1937)<sup>3</sup>.

Prima di proseguire mi sembra necessario fare una rapida disamina sulla letteratura per l'infanzia.

---

<sup>2</sup> Croce Benedetto, 1943, *La letteratura della nuova Italia. Saggi critici*, Bari, Laterza, vol. III.

<sup>3</sup> Croce Benedetto, 1937, *Pinocchio*, in *Id.*, *Aggiunte alla "Letteratura della nuova Italia"*, «La Critica» 35, pp. 452-454.

Di fatto, la letteratura per l'infanzia è a lungo derivata dalla tradizione orale dei racconti popolari, delle fiabe, delle leggende, dei miti, delle filastrocche e delle ninne nanne.

Fino alla fine del XVIII secolo i grandi successi per l'infanzia furono rappresentati soprattutto da libri destinati agli adulti che potevano essere validi anche per un pubblico di bambini, è il caso, per esempio, di Robinson Crusoe di Daniel Defoe e di I viaggi di Gulliver di Jonathan Swift e inoltre, la letteratura infantile aveva un ruolo quasi esclusivamente didattico doveva comunicare informazioni utili o consigli morali.

Un ruolo fondamentale nello sviluppo della letteratura per l'infanzia si è avuto a partire dal lavoro dei fratelli Grimm che ripresero fiabe della tradizione tedesca e del danese Hans Christian Andersen, che scrisse fiabe svincolate dalla tradizione come La Sirenetta, Il brutto anatroccolo, La piccola fiammiferaia.

In molti casi i confini della letteratura dedicata esclusivamente all'infanzia si confondono con quelli della narrativa d'avventura basti pensare alle opere geniali e innovative del francese Jules Verne, che con il suo Nautilus di Ventimila leghe sotto i mari anticipa di diciassette anni il primo sottomarino funzionante che appare nel 1887. Nella letteratura inglese dell'età vittoriana, d'altro canto, sono da ricordare L'isola del tesoro di Robert Louis Stevenson e Peter Pan di James Barrie, e nella letteratura italiana va ricordato un grande narratore di avventure esotiche, Emilio Salgari che scrisse circa ottanta romanzi, pensiamo soprattutto al ciclo di Sandokan.

Dello stesso periodo sono le opere di Charles Lutwidge Dodgson che scrisse con lo pseudonimo di Lewis Carroll, Alice nel paese delle meraviglie nel 1865, a cui fece seguito Oltre lo specchio (1871), due libri fondamentali della moderna letteratura per l'infanzia.

Di inizi novecento sono le storie di uomini e cani ambientate nel Grande Nord americano di Jack London Il richiamo della foresta a Zanna Bianca.

Un'attenzione particolare va data al fiorentino Carlo Lorenzini, in arte Collodi, con il suo Pinocchio, pubblicato prima a puntate sul "Giornale dei bambini" poi in volume, che come abbiamo già avuto modo di dire è stato a lungo considerato «il più bel libro della letteratura infantile italiana» (Croce 1937). Nella storia che viene raccontata "il legno è l'umanità" come ebbe a sottolineare Benedetto Croce e dietro la finzione romanzesca vengono messe in gioco

gli umori, le ossessioni e la formazione del suo autore. Tradotto in tutte le lingue (compreso il latino) i suoi contenuti hanno sempre consentito a decine di illustratori di esprimersi liberamente mentre molti dei rifacimenti hanno assunto un carattere moralistico con finalità didattiche lontano dagli obiettivi del racconto originale.

Pochi anni dopo la pubblicazione di Pinocchio, in Italia uscì un altro libro per l'infanzia destinato a uno straordinario successo, Cuore (1886) di Edmondo De Amicis. Accanto a quest'approccio melodrammatico di De Amicis si affianca chi alle finalità didattiche accosta l'umorismo: è il caso di Il giornalino di Gian Burrasca (prima fra il 1907 e il 1908 in 55 puntate sul "Giornalino della Domenica", poi in volume nel 1912) scritto da Vamba (pseudonimo di Luigi Bertelli, che dal 1906 fu direttore del "Giornalino della Domenica").

Nel 1934 la scrittrice australiana Pamela L. Travers pubblicò il primo romanzo dedicato all'istitutrice-maga Mary Poppins, mentre appartiene al periodo della seconda guerra mondiale un altro grande classico della letteratura per ragazzi Il piccolo principe (1943) dello scrittore francese Antoine de Saint-Exupéry. Anche Il Piccolo Principe tratta tematiche che si rivolgono agli adulti con una critica alla situazione politica e culturale, ma viene pian piano considerato un libro per bambini proposto in genere negli ultimi anni della scuola primaria.

Il fantastico si mescola ancora al comico nelle strane storie di Pippi Calzelunghe, personaggio creato nel 1945 dalla fantasia della scrittrice svedese Astrid Lindgren.

Una nuova fase nello sviluppo della letteratura dell'infanzia si ha nel periodo post-bellico quando oltre a combattere l'analfabetismo si dovrà combattere con quel linguaggio uscito dalla dittatura che si era nutrito di nazionalismo, razzismo e xenofobia e che doveva aprirsi ad un linguaggio di pace, libertà e giustizia. E proprio in questo contesto Italo Calvino in Il Sentiero dei nidi di ragno (1947) sostiene che - grazie alla sua "universalità di contenuti" la guerra partigiana diventa subito oggetto di una generale "smania di raccontare".

Nel ventennio successivo ci troviamo di fronte a due filoni distinti nella letteratura per l'infanzia, uno conservatore e moralista che è pervaso da insegnamenti benpensanti e buoni sentimenti e un altro che riesce a svincolarsi dal tradizionale sentimentalismo ottocentesco e tende a presentare realisticamente i problemi. Pensiamo in particolare ai racconti di Gianni Rodari e Italo Calvino. Protagonista indiscusso di questo nuovo modo di intendere la

narrativa è Marcovaldo che è alle prese con le sconfitte quotidiane ma anche con l'ostinazione di chi con determinazione va incontro ad un nuovo mondo in cui tutti i valori diventano merci. È proprio a Calvino che nel 1956 fu affidata una raccolta di fiabe della tradizione popolare di tutte le regioni italiane e come lo stesso autore tenne a dire "il libro *Fiabe italiane*<sup>4</sup> che raccoglie duecento storie del folklore di tutte le regioni italiane, può essere avvicinato alla raccolta dei fratelli Grimm". A differenza della raccolta dei fratelli Grimm però "Le Fiabe italiane" non sono riscritte per l'infanzia ma sono uno straordinario omaggio alla cultura popolare elaborata e trasmessa nei secoli. In questi anni il lavoro di Calvino e "Il canzoniere italiano"<sup>5</sup> di Pier Paolo Pasolini, uscito nel 1955, sono la risposta dei letterati più consapevoli all'importanza della dimensione fantastica nell'impegno civile.

Tra la fine degli anni Quaranta e gli inizi degli anni Cinquanta, quando Gianni Rodari esordisce come scrittore per l'infanzia, la lingua italiana non era ancora diffusa in tutta Italia, basti pensare che solo il 18% della popolazione aveva abbandonato l'uso del dialetto. In questo contesto risulta evidente che i libri scritti per l'infanzia fossero destinati ai figli delle famiglie benestanti, i contenuti d'altro canto erano ancora completamente attraversati da buoni sentimenti e moralismo.

Le opere di Rodari sono invece completamente differenti, popolate da personaggi della realtà di tutti i giorni, pompieri, stagnini, ferrovieri ecc, o da personaggi di fantasia, che permettono allo scrittore di parlare dei problemi del quotidiano, sono rivolte ad un pubblico più vasto e non a quel ristretto numero di famiglie benestanti. Ne sono un esempio Cipollino e Gelsomino nel paese dei bugiardi ambientati in un contesto contadino lontano dai problemi della realtà industriale. Ovviamente la scelta di questi lettori comporta una scelta di natura linguistica e stilistica che trasforma una lingua astratta e artificiosa in una lingua legata alla

---

<sup>4</sup> *Fiabe italiane* opera curata da Italo Calvino è pubblicata nella collezione I Millenni di Einaudi nel novembre del 1956. Si tratta di una raccolta basata sul vasto ambito degli studi etnografici e del folklore nazionale, particolarmente vivi negli anni di stesura del libro.

<sup>5</sup> Pier Paolo Pasolini raccolse in questa monumentale antologia, pubblicata per la prima volta nel 1955, le espressioni più belle e curiose di una poesia popolare ricca e varia come quella italiana. Di regione in regione, attraverso quasi 800 testi di vario genere e struttura, si passa dai canti narrativi piemontesi alle «biojghe» romagnole, dalle «vilote» venete e friulane ai rispetti toscani, dalle «canzune» abruzzesi ai canti funebri calabresi, dai «mutos» sardi agli stornelli, agli strambotti, alle ninne nanne, fino ai canti popolari delle due guerre e alle canzoni fasciste e partigiane.

quotidianità, capace di parlare delle cose di tutti i giorni. A tal proposito Tullio De Mauro ebbe a dire: “ chi parla ai bambini chi tesse favole e nonsenses non può parlare come un libro stampato, non può parlar solo di valor civile e immortalità dell’anima, ma deve parlare appunto delle mille piccole cose , con le mille comuni parole, e deve sapere fare scoccare scintille fra i diversi strati dell’esperienza e della lingua”<sup>6</sup>.

Nel 1962 nei «Libri per Ragazzi» escono le Favole al telefono con i disegni di Bruno Munari, un’opera che, come già avvenuto nelle opere precedenti di Rodari segna un profondo rinnovamento della letteratura per l’infanzia aprendo le porte alla creatività, alla fantasia e al gioco linguistico, nello stesso anno viene pubblicato Cipì di Mario Lodi.

A Milano, nel 1966, Rosellina Archinto fonda la casa editrice Emme, i suoi libri fanno conoscere i grandi illustratori stranieri e italiani: Maurice Sendak, Eric Carle, Leo Lionni, Bruno Munari. Per esempio Piccolo blu, piccolo giallo di Leo Lionni segna una radicale e profonda innovazione per i bambini piccoli: storia di un’amicizia, di un’incomprensione da parte dei genitori, in cui i protagonisti non sono esseri umani né animali ma due piccoli cerchi, blu e giallo.

Già negli anni precedenti i libri di Munari hanno un carattere innovativo infatti propongono storie brevi supportate da disegni e soprattutto da figure in rilievo, oggi questo tipo di libro viene chiamato pop-up o tridimensionale. I libri di Munari, ancora oggi disponibili, possono essere considerati innovativi perché si rivolgevano alla prima infanzia e proponevano un tipo di grafica fuori da quella tradizionale senza fate e streghe, senza castelli e principesse.

Negli anni ’60-’70 del secolo scorso l’offerta editoriale è segnata da un passaggio epocale, radicatosi per tutti gli anni ’80 e ’90: ovvero il passaggio dal libro unico alla collana.

La collana di libri per bambini e ragazzi si fa inoltre riconoscere per due altri ordini di motivi: la centralità dell’illustrazione e il formato tascabile.

L’abbassamento dei costi di produzione e la circolazione di testi tra cinema e letteratura, inoltre, si pensi a Harry Potter, ha fatto sì che il volume dei libri prodotti aumentasse, tanto che dal 1980 agli anni duemila la loro produzione è più che raddoppiata, inoltre sono nate

---

<sup>6</sup> T. De Mauro Prefazione a G. Rodari. Il Gatto viaggiatore e altre storie a cura di C. De Luca L’Unità Editori Riuniti Roma 1990.

diversi case editrici che si sono dedicate esclusivamente alla letteratura per l'infanzia. Un esempio su tutti la casa editrice romana Orecchio acerbo, il cui nome non a caso si ispira alla filastrocca<sup>7</sup> di Rodari, nella quale l'autore invita gli adulti a mantenere un "orecchio acerbo" per capire i bisogni dei bambini che spesso sono inascoltati. Partendo da questo presupposto Orecchio acerbo, nata nel 2001 ha posto l'attenzione su contenuti innovativi usando anche dei formati non convenzionali. Un esempio su tutti, Gago di Giovanni Zoppoli illustrato da Maja Celija in cui testo ed immagini si fondono in una storia senza lieto fine.

In queste poche pagine ho cercato di ripercorrere in maniera estremamente sintetica l'approccio che negli anni si è avuto nei confronti della letteratura per l'infanzia.

Un autore chiave nell'evoluzione della letteratura italiana per l'infanzia è stato Gianni Rodari di cui era un grande estimatore Tullio De Mauro che ha più volte messo in evidenza come Rodari «il primo grande scrittore per l'infanzia non toscano» fosse riuscito a sfruttare la nuova situazione linguistica italiana che iniziava ad affermarsi nel secondo dopoguerra in cui, l'italiano cominciava a diventare una lingua parlata da ampi strati di popolazione.

«Rodari attinge a questa ormai larga e comune esperienza di italiano parlato, usato davvero, scritto e letto sempre più largamente specie dalle generazioni più giovani. [...] mettendo a frutto questo patrimonio, chiamando a raccolta e a ironico conflitto tutti i possibili strati del vocabolario, parole d'ogni giorno e parole rare e in via di farsi antiquate (da ribobolo a princisbecco), parole legate alle innumerevoli tecniche del mondo industriale e parole d'una volta, Rodari ha certo potentemente influito sulla formazione linguistica delle generazioni più giovani succedutesi in Italia dagli anni cinquanta in poi nel senso di donare a esse una più sicura tranquillità nel ricorso a tutti gli strati e gli stili della lingua» (De Mauro 1990).

Tra gli interventi di Rodari ci piace ricordare quello a difesa della fiaba, che è stata a lungo contestata in quanto mezzo di trasmissione di valori borghesi, sottolineando come le critiche e l'intolleranza dovessero essere considerate "cosa vecchia", tant'è che nell'intera storia della letteratura troviamo scrittori di fama che si sono voluti cimentare nella produzione per ragazzi

---

<sup>7</sup> Un signore maturo con un orecchio acerbo G. Rosari in Parole per giocare, Manzuoli, Firenze, 1979

come lo stesso Rodari che da giornalista di “sinistra” amava esprimersi dalle pagine del Manifesto. Un altro autore che si è cimentato sulla scrittura per ragazzi è Luigi Malerba i cui libri per l’infanzia sono strettamente intrecciati con quelli per adulti “Millemosche mercenario” 1969 scritto con Tonino Guerra, “Storiette” illustrazioni di Adriano Zannini 1977, “Pinocchio con gli stivali” 2004. L’autore mette in evidenza che la differenza tra la scrittura per l’infanzia e quella per gli adulti sta soltanto in una diversa disposizione d’animo e in un diverso registro di scrittura più elementare e comunicativo per stimolare l’intelligenza e l’emotività dei più piccoli, senza nessuna pretesa pedagogica o moralistica. I libri per ragazzi sono l’espressione di un disagio e di una difficoltà di adattamento alla realtà che accomuna i giovani lettori agli scrittori. (Boero, De Luca 1995).

## 2. Scelta dei testi

Creare un *corpus* su cui svolgere un’analisi linguistica con l’utilizzo di strumenti software che impiegano specifici tools lingware non è un compito facile e richiede innanzitutto chiarezza negli obiettivi della ricerca e conoscenza del settore che si va ad analizzare.

Sulla base di questa consapevolezza oltre agli 11 testi di classici della letteratura per l’infanzia sono stati selezionati 10 libri ad alta leggibilità.

Nella scelta dei classici della letteratura per l’infanzia si è cercato di ricoprire un periodo molto ampio che va dalla fine dell’ottocento fino alla fine degli anni ‘90 del novecento.

Un *corpus* per sua definizione è una raccolta limitata di testi selezionati e organizzati per facilitare le analisi linguistiche e ogni scelta che viene fatta prevede la selezione di un testo in luogo di un altro.

Per questo *corpus* sono stati scelti classici della letteratura per l’infanzia di autori italiani e non, pertanto parte dei testi sono libri di autori stranieri tradotti in lingua italiana e testi ad alta leggibilità che sono pubblicati all’interno di collane dedicate, nello specifico il corpus sottoposto ad analisi risulta costituito da 11 testi integrali di letteratura per l’infanzia, scelti tra i libri pubblicati e tradotti tra il 1865 e il 1997 anno di pubblicazione del primo libro della serie Harry Potter di Rowling, Joanne Kathleen - Harry Potter e La Pietra Filosofale, libro

con il quale si chiude il corpus e 10 libri ad alta leggibilità tratti dalle collane Il Mulino a Vento e Attacca Parole, edite da Raffaello. I racconti selezionati all'interno di questa collana appartengono alla Serie Gialla dai 6 anni in su, alla Serie Rossa adatta a bambini dai 7 anni in su, alla serie Avventurose Scienze dai 9 anni in su che nasce con l'obiettivo di avvicinare i ragazzi al mondo delle scienze, e di stimolarne la curiosità verso il mondo che li circonda.

La fascia d'età dei bambini-lettori presi in considerazione va dai 7 agli 11 anni, e comprende quindi le classi III, IV e IV della scuola primaria.

Per decidere l'assegnazione di un testo alla fascia d'età ho tenuto conto, in primo luogo, dell'indicazione editoriale che spesso appare in quarta di copertina o nei cataloghi e in secondo luogo attraverso una bibliografia mirata sulla letteratura per l'infanzia.

L'assegnazione di un libro a una specifica fascia d'età è controversa e ha dato origine a molte discussioni. Infatti, i bambini hanno livelli di maturità, capacità di lettura e interessi differenti che influenzano le scelte dei testi e l'età di lettura, tant'è vero che in alcune circostanze lo stesso testo è stato assegnato a fasce di età differenti.

La tabella 1 elenca i testi che costituiscono il corpus preso in esame.

*Tabella 1 Testi del corpus*

	<b>TESTI</b>	<b>AUTORI</b>	<b>FASCE D'ETÀ</b>	<b>I EDIZIONE</b>	<b>TESTI IN TRADUZIONE</b>	<b>EDIZIONE</b>
1	Le avventure d' Alice nel Paese delle Meraviglie	Lewis Carroll (pseudonimo di Charles Lutwidge Dodgson)	7 - 11	1865	1 <sup>a</sup> ed. italiana, 1870	Traduzione di Teodorico Pietrocòla-Rossetti, illustrazioni di John Tenniel, Londra, Macmill an and co., 1872. Wikisource
2	L'isola del tesoro	Robert Stevenson	7 - 11	1883	1 <sup>a</sup> ed. italiana, 1883	Traduzione di Angiolo Silvio Novaro (1932)

						Newton Compton, 2013 Wikisource
3	Le Avventure di Pinocchio	Carlo Collodi (pseudonimo Carlo Lorenzini)	7 - 11	1883		Firenze R. Bemporad & Figlio 1892 Wikisource
4	Il Corsaro Nero	Emilio Salgari	7 - 11	1898		a cura di Sergio Campailla; Biblioteca economica Newton Sezione ragazzi, 37; Roma, 1996 Wikisource
5	Sandokan alla riscossa	Emilio Salgari	7 - 11	1907		Fabbri Editori, 2002 Wikisource <a href="http://www.emiliosalgari.it/testi/testionline.htm">http://www.emiliosalgari.it/testi/ testionline.htm</a>
6	Cuore	Edmondo De Amicis	7 - 11	1886		Milano Fratelli Treves, editori 1889 Wikisource
7	Il Piccolo Principe	Antoine de Saint- Exupéry	7 - 11	1943	1ª ed. italiana, 1949	Edizione bilingue: Testo originale francese di Antoine de Saint-Exupéry, versione in italiano a cura di Franco Perini. Illustrazioni di Antoine de Saint-Exupéry. Prima edizione in formato ePub 2016

8	Gelsomino nel paese dei bugiardi	Gianni Rodari	7 - 11	1959		Illustrazioni Alberto Ruggieri Editori Riuniti 2003
9	Favole al telefono	Gianni Rodari	7 - 11	1962		Illustrazioni di Francesco Altan Einaudi Ragazzi 1993
10	Marcovaldo Le stagioni in città	Italo Calvino	7 - 11	1963		(Autore, Prefazione), Domenico Scarpa Oscar Moderni 2016
11	Harry Potter e La Pietra Filosofale	Rowling, Joanne Kathleen	7 - 11	1997	1 <sup>a</sup> ed. italiana 1997	Traduttore Marina Astologo Salani Editore 2001
12	Cani, gatti, rospi e altre persone	Paola Valente	6 - 7	2019		Parole leggere - Il Mulino a Vento- Gruppo editoriale Raffaello
13	Fantasm nell'armadio	Lorenza Farina	6 - 7	2019		Attacca Parole - Gruppo editoriale Raffaello
14	Emozioni batticuore	Valeria Conti	6 - 7	2019		Serie Gialla Il Mulino a Vento- Gruppo editoriale Raffaello
15	La lezione del fenicottero	Germano Chiaverini	7 - 11	2019		Serie Rossa Il Mulino a Vento- Gruppo editoriale Raffaello
16	Il Fantasma Minestrone	Guido Quarzo	6 - 7	2018		Attacca Parole - Gruppo editoriale Raffaello
17	I tre Porcellini	narrata da Raffaella Magi	6 - 7	2019		Parole leggere - Il Mulino a Vento- Gruppo editoriale Raffaello

18	Nuvola bianca e nuvola nera	Annamaria e Nice Piccione	6 - 7	2015		Serie Gialla Il Mulino a Vento- Gruppo editoriale Raffaello
19	Oceani di plastica	Beatrice Peruffo	9 – 11	2019		Avventurose Scienze - Gruppo editoriale Raffaello
20	Uffa le femmine , uffa i maschi	Francesca Mascheroni	6 - 7	2016		Serie Gialla Il Mulino a Vento- Gruppo editoriale Raffaello
21	Alla ricerca dei colori perduti	Maria Strianese	7 – 11	2009		Serie Rossa Il Mulino a Vento- Gruppo editoriale Raffaello

Gli autori e i testi che sono entrati a far parte del *corpus* sono il frutto di un lavoro fatto a partire da articoli e libri sulla storia della letteratura per l'infanzia e dalla consultazione di articoli e libri sulla comprensione dei testi nell'infanzia, che mi hanno permesso di approfondire la conoscenza di questo settore e di individuare i testi che potevano essere considerati più significativi ai fini della mia analisi.

Nel *corpus* si trovano quindi classici (*Alice nel paese delle meraviglie*, *L'isola del tesoro*, *Il corsaro nero*, *Zanna bianca*), fiabe e favole della tradizione come *Le avventure di Pinocchio*, successi editoriali come *Harry Potter e la pietra filosofale* e testi ad alta leggibilità rivolti a lettori con difficoltà di lettura, ma non solo. Adatti anche a tutti i piccoli lettori in cerca di storie semplici e coinvolgenti. Ho inserito infine alcuni libri che possono essere considerati come rivoluzionari nel modo di intendere la narrativa per l'infanzia (*Favole al telefono*, *Gelsomino nel paese dei bugiardi*, *Marcovaldo*).

Per gli autori stranieri non ho preso in considerazione traduzioni di opere con la dicitura *libera versione di* o *adattamento di*.

I testi presi in esame sono stati scaricati in formato Pdf dal web grazie all'accesso a progetti che si pongono come scopo la pubblicazione e la diffusione gratuita di opere letterarie in formato elettronico.

## CAPITOLO II

### COMPRENSIONE E LEGGIBILITA'

#### 1. La comprensione come problema e gli indici di leggibilità

Settant'anni fa Rudolph Flesch<sup>8</sup>, lo studioso austriaco che si è occupato delle prime indagini oggettive sulla comprensione dei testi dichiarò che la leggibilità di un testo è un concetto misurabile e per far ciò ha creato un indice. Ma la domanda è: È sufficiente applicare una formula per individuare il livello di comprensibilità di un testo? La nostra risposta è no, in quanto la comprensione è un processo che coinvolge numerose abilità cognitive. Il processo di comprensione di un testo parte dalla decodifica dello stesso, e si fonda sulle abilità linguistiche di base. Ciò è particolarmente vero se ad essere coinvolti sono bambini dei primi anni della scuola primaria, periodo nel quale l'abilità di decodifica del testo è in via d'apprendimento e può pertanto risultare così impegnativo da sottrarre energie ai meccanismi di comprensione. Superata questa fase, quando cioè la decodifica diventa automatica, il lettore può con più facilità elaborare il significato.

La comprensione come integrazione di conoscenze è sicuramente un altro aspetto da considerare. Il processo di comprensione richiede un recupero e una integrazione di tutte le conoscenze che il lettore possiede per la rielaborazione del testo. Queste conoscenze sono dette enciclopediche e sono conservate nella memoria a lungo termine. L'attivazione di questo tipo di conoscenze permette al lettore di generare e elaborare ipotesi interpretative riguardo alle informazioni implicite del testo. Le conoscenze enciclopediche sono organizzate secondo due modalità: gli schemi e gli script. Lo script è la conoscenza che ciascuno ha di certi eventi nel loro svolgimento normale. È dunque la conoscenza organizzata di una serie di azioni, oggetti e eventi ricorrenti e stabili in una particolare situazione.

---

<sup>8</sup> L'indice di leggibilità (o *Reading Ease Score*, RES) secondo la formula di Flesch corrisponde a  $RES = 206,835 - (0,864 \times S) - (1,015 \times W)$ . Dove S è la lunghezza media delle parole misurata in sillabe, e W è il numero medio di parole per frase, su campioni di 100 parole (Piemontese, 1996: 96). La formula di Flesch sarà la base di partenza dell'indice Gulpease, il riferimento principale per gli studi di leggibilità sull'italiano.

Più precisamente un lettore o un ascoltatore di un discorso dal momento in cui comincia ad avere alcune informazioni inizia ad elaborarle e ad integrarle grazie agli script che possiede. Se legge “un tale si avvicinava al botteghino per fare i biglietti” può immaginare che il tale stia andando al cinema, al teatro etc; se più avanti nella storia si parla di giovani, sciarpe e bandiere il lettore potrà dedurre che si tratta di un evento sportivo. Questo significa che lo script dell’andare allo stadio per un evento sportivo funziona come una guida per decifrare la situazione oggetto del testo, quindi le conoscenze servono al lettore per integrare le informazioni ed immaginare quanto avviene nel testo.

Gli script quindi sono i primi concetti concreti di cui disponiamo e come tali sono chiamati in causa continuamente dalle nostre conoscenze linguistiche.

L’altro elemento di organizzazione delle conoscenze enciclopediche è lo schema. Lo schema è una rete di conoscenze connesse tra loro più astratto rispetto allo script. Un esempio di schema è quello del comprare. A questa azione sono legate una serie di concetti quali: qualcuno che vende, qualcuno che acquista, la sussistenza di una transazione, il valore del bene legato al suo prezzo etc.

Anche il possedere uno schema significa poter richiamare con facilità alla mente i nodi connessi, i quali permettono di integrare le conoscenze non esplicite del testo. (Cardarelo 2004).

In sintesi il processo di comprensione di un testo sia negli adulti che nei bambini è fortemente legato alle conoscenze enciclopediche. Dunque l’immaturità di queste conoscenze nei bambini, la qualità inferiore e il difficile recupero rendono più complessa la loro comprensione. Nell’applicare le conoscenze enciclopediche i bambini tendono ad utilizzare quelle più familiari e immediatamente disponibili e solo nel corso della scuola primaria, grazie allo sviluppo di diverse abilità cognitive, riusciranno a realizzare inferenze anche su concetti astratti e diventeranno progressivamente più abili nel creare legami tra informazioni lontane tra di loro (Bertolini 2012). In buona sostanza la comprensione si realizza attraverso l’interazione tra un testo e quanto il lettore già conosce. Solo apportando alle parole da comprendere il bagaglio di esperienze e conoscenze possedute e integrandole con quanto viene letto, ognuno può ricostruire i significati delle parole, delle frasi di un testo e capire

quanto sta leggendo. Se nella nostra memoria non abbiamo nulla che si colleghi a quanto stiamo leggendo non capiamo e in generale non impariamo. Comunque non impariamo in modo significativo (Ausubel 1978). Sulla base di queste considerazioni possiamo sottolineare il ruolo attivo del lettore, che in primo luogo deve avere disponibili e attive nella mente quelle conoscenze pregresse che possono operare come “ganci” ai quali possono essere appese le nuove conoscenze da apprendere, e in secondo luogo deve essere in grado di realizzare delle inferenze.

L'estrazione dei significati avviene, quindi, attraverso processi inferenziali, ossia attraverso forme di ragionamento e di elaborazione delle informazioni non esplicite nel testo. Il processo inferenziale avviene per opera della memoria grazie al recupero e all'intervento delle conoscenze pregresse utili a ricostruire il significato del testo. È grazie all'incontro tra tali conoscenze e gli elementi espliciti della storia che è possibile colmare i punti non chiari, attraverso il riconoscimento e l'integrazione delle informazioni mancanti. Durante la lettura le inferenze si possono creare in modo automatico senza esserne consapevoli e la complessità dei processi inferenziali può dipendere da diversi fattori: la mancanza di conoscenze pregresse può rendere più difficile il recupero delle informazioni sottintese, mentre il loro possesso può facilitare l'elaborazione dei significati; le inferenze che riguardano parti distanti nel testo sono più complesse da ricordare; le inferenze relative ad eventi concreti sono più facili di quelle che riguardano eventi astratti. (Lumbelli 2009). È pertanto facile comprendere che le inferenze coinvolte nella comprensione possono essere di vari tipi. In letteratura ci sono diversi tentativi di classificazione delle inferenze, di seguito intendiamo descrivere alcune categorie che a nostro avviso sono più rilevanti.

1. Le inferenze ponte o bridging inferences sono una tipologia di inferenze che permette al lettore di connettere due elementi informativi del testo grazie all'impiego di conoscenze pregresse. Attraverso l'integrazione di questi elementi il lettore costruisce un ponte tra i due elementi informativi iniziali. Questo tipo di inferenze risultano necessarie alla comprensione quando il testo è lacunoso o linguisticamente ambiguo. “La strada era gelata. Il camion correva veloce. Molte persone rimasero ferite” è un esempio di racconto semplice che per essere compreso richiede l'attivazione di

inferenze ponte. *La strada era gelata*, questa prima frase potrebbe attivare delle conoscenze pregresse legate all'informazione che il gelo rende pericolose le strade e può essere causa d'incidenti automobilistici. Il recupero di questo tipo d'informazione ci potrebbe permettere di comprendere che nel testo si parli di un incidente. Questa aspettativa viene rafforzata dalla seconda frase *il camion correva veloce*, anche qui la previsione dell'incidente viene rafforzata dalla consapevolezza che la velocità è una delle cause possibili di incidenti stradali. La terza frase *molte persone rimasero ferite* conferma questa aspettativa. Dunque attingendo alle conoscenze enciclopediche pertinenti è possibile per il lettore colmare le lacune del testo. Le inferenze ponte, dette anche connettive (Lumbelli 2007) sono normalmente attività durante la lettura e se il problema di coerenza è semplice il lettore è in grado di colmare la lacuna in maniera automatica, quando ciò non avviene entriamo nella sfera del problem solving e quindi della risoluzione non più inconsapevole della coerenza del testo.

2. Le inferenze causali consentono di comprendere il perché delle azioni avvengono nel testo, creano un legame causa effetto tra alcuni elementi testuali. In una sperimentazione con alcuni bambini sono state mostrate sequenze di immagini che presentavano una storia in maniera disordinata. È così risultato evidente che la mancanza di un collegamento tra le immagini aveva stimolato i bambini a creare una serie di inferenze che collegavano tra loro i personaggi e le azioni che venivano svolte. Quindi ciascun lettore possiederebbe una teoria ingenua creata a partire dalle esperienze pregresse che attiva durante la lettura del testo e utilizza per realizzare inferenze causali (Trabasso 1989).

I processi cognitivi che vengono messi in atto nella comprensione di un testo sono legati ai processi di elaborazione dei significati durante la lettura di un testo. Un esempio che piace molto ai diversi ricercatori che si sono occupati di questo argomento è Marcovaldo ovvero le stagioni in città (Calvino 1963) utilizzato anche per la nostra ricerca.

Nelle prime pagine di "Un bosco sull'autostrada" ci sono diversi elementi che fanno capire ad un lettore esperto che i figli di Marcovaldo, non sapendo che cosa fosse un bosco, hanno

creduto che il bosco fosse costituito dai cartelli pubblicitari presenti sui bordi dell'autostrada. Questa informazione può essere desunta da questo pezzo di testo:

“Dove l'avete preso? Esclamò indicando i resti del cartellone pubblicitario che, essendo di legno compensato, era bruciato molto in fretta. Nel bosco! Fecero i bambini.”

Un lettore meno esperto che non fosse in grado di attivare le opportune inferenze non sarebbe in grado di comprendere questo passaggio nel quale l'informazione che i bambini abbiano scambiato i cartelli pubblicitari con il bosco risulta sufficientemente chiara. In questo caso l'inferenza di collegamento dovrebbe permettere di comprendere che, se i bambini hanno portato a casa un cartello pubblicitario dicendo di averlo preso nel bosco, non sapendo che cosa fosse un bosco, hanno pensato che i cartelli pubblicitari formassero il bosco. In tal caso la comprensione del testo sarebbe deprivata della sua essenza.

Se fino ad ora abbiamo considerato i processi inferenziali che sono alla base della comprensione, arriviamo ora ad occuparci delle abilità linguistiche necessarie alla decodifica, all'uso delle parole, perché è a partire da queste che la mente crea i legami tra le conoscenze pregresse e il testo producendo significato.

Il piano dell'espressione linguistica può facilitare o rendere più complesso l'accesso al significato. È a partire da questa consapevolezza che sottolineiamo come l'articolazione e la struttura della frase siano degli elementi di accessibilità ai testi. Una parte delle ricerche che riguardano questo argomento, però, quando si riferiscono agli aspetti puramente linguistici, e intendono la sintassi e il lessico, parlano di leggibilità di un testo, preferendo lasciare il termine “comprensibilità” all'organizzazione logica e concettuale (Piemontese 2005). Il linguaggio chiaro e facile che non ostacola la comprensione a livello linguistico, dovrebbe utilizzare termini del lessico fondamentale cioè quelle parole ad altissima frequenza, usate nel 90% circa dei discorsi e dei testi secondo quanto evidenziano da Tullio De Mauro nel VdB, (De Mauro 1980) limitando l'uso di vocaboli a cui hanno accesso un minor numero di parlanti. Quindi secondo quanto sostenuto da Tullio De Mauro la questione potrebbe essere in gran parte risolta scegliendo le parole che risultano più frequentemente utilizzate nel linguaggio comune, ciò rappresenterebbe una garanzia per la comprensione.

Se nello specifico parliamo dei bambini sul piano semantico l'uso di parole poco frequenti e di non immediata comprensione può essere considerato il primo degli ostacoli, ma è chiaro che i termini difficili o nuovi possono essere compresi grazie al contesto in cui appaiono e alla capacità di creare inferenze. D'altro canto si dovrebbe usare anche una struttura sintattica semplice, cioè bisognerebbe evitare l'uso di strutture complesse con molte subordinate, in particolare di quelle implicite e delle frasi passive. Come abbiamo già accennato, la leggibilità linguistica di un testo riguarda sia la scelta dei termini da adottare sia la sintassi impiegata, in relazione a ciò l'indice di Flesch considera due variabili linguistiche nella valutazione:

- lunghezza media delle parole espressa in sillabe per parola
- lunghezza media delle frasi espressa in parole per frase.

L'indice nato per l'inglese è stato rielaborato da Roberto Vacca, il quale, nel 1972, ha adattato i parametri della formula alla lingua italiana (Franchina-Vacca 1986).

Nel 1982 il GULP - Gruppo universitario linguistico pedagogico, presso l'Istituto di Filosofia dell'Università degli Studi di Roma «La Sapienza» ha definito una nuova formula, la formula GULPEASE, partendo direttamente dalla lingua italiana (Lucisano-Piemontese 1988; Lucisano 1992). La formula è stata determinata verificando con una serie di test la reale comprensibilità di un corpus di testi. La misura della leggibilità di un testo non ha solo valore assoluto ma fa riferimento anche ai livelli di scolarizzazione. Per esempio, un testo con indice GULPEASE 60 è: molto difficile per chi ha la licenza elementare, difficile per chi ha la licenza media, facile per chi ha un diploma superiore. La formula GULPEASE, oltre ad essere la prima formula di leggibilità tarata direttamente sulla lingua italiana, ha anche il vantaggio di calcolare la lunghezza delle parole in lettere, e non più in sillabe.

La formula di Flesch serve a misurare la leggibilità (F) di un testo in inglese.

Il modo di calcolare questo indice è il seguente:

$$F = 206,835 - (0,864 * S) - (1,015 * P)$$

dove S è il numero delle sillabe, calcolato su un campione di 100 parole, P il numero medio di parole per frase, 206,835 è una costante che serve a mantenere i valori finali dell'applicazione della formula fra 0 e 100.

L'adattamento proposto da Vacca e Vedovelli per la lingua italiana è il seguente:

$$F = 206 - (0,6 * S) - P$$

La leggibilità è alta se F è superiore a 60, media se fra 50 e 60, bassa sotto 50.

L'indice Gulpease considera due variabili linguistiche: la lunghezza della parola e la lunghezza della frase rispetto al numero delle lettere. La formula per il suo calcolo è la seguente:

$$\text{Gulpease} = 89 - (LP: 10) + (3 * FR)$$

dove LP = (totale delle lettere del campione \* 100): totale delle parole, FR = (totale delle frasi del campione \* 100): totale delle parole.

### Indice Gulpease: scale dei valori

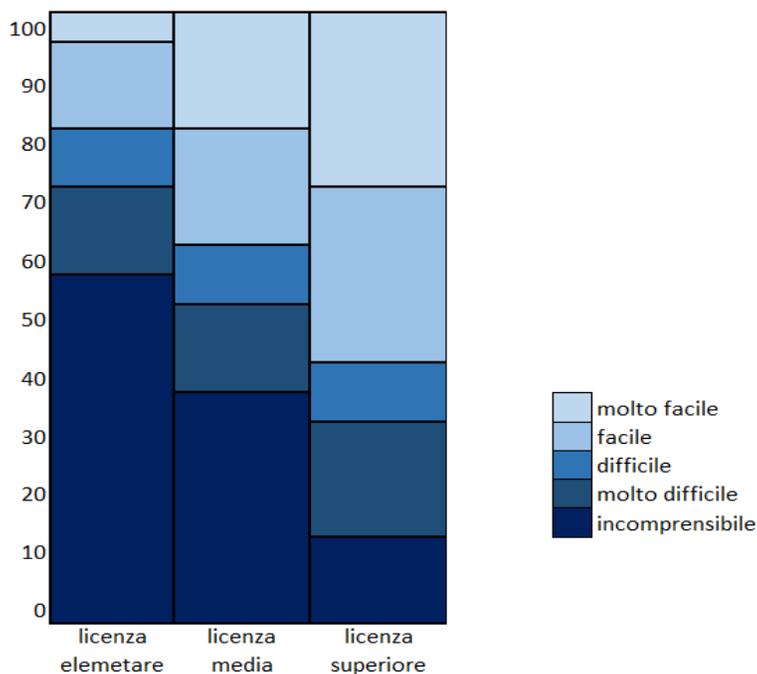


Figura 1 Indice Gulpease

Sulla base del calcolo della formula Gulpease sono stati costruiti alcuni strumenti automatici che consentono di ottenere una valutazione del livello di leggibilità del testo. Uno di questi è il servizio online Corrige!® dell'azienda Èulotech<sup>9</sup>.

Inoltre, negli anni sono state sviluppate anche tecnologie linguistico-computazionali. Per la lingua italiana in particolare è disponibile il software READ-IT, elaborato dall'Italian Natural Language Processing Laboratory (ItalianNPL Lab) dell'Istituto di Linguistica Computazionale Antonio Zampolli (ILC) del Consiglio Nazionale delle Ricerche (CNR) di Pisa. Questo software è una risorsa che si occupa della semplificazione di un testo, individuando la complessità non solo a livello lessicale e morfologica ma anche sintattica. L'analisi del testo viene effettuata mettendo a confronto i testi con due corpora di riferimento appartenenti allo stesso genere testuale (prosa giornalistica): uno composto da testi particolarmente semplici, tratti dal periodico di facile lettura Due Parole<sup>10</sup>; l'altro da testi difficili estratti dal giornale La Repubblica. Sulla base di questo confronto si orienta la valutazione. (Dell'Orletta, Montemagni, Venturi, 2011).

Tutto quanto finora evidenziato deve essere preso in considerazione quando parliamo di libri per bambini. Cercando di comprendere quale possa essere un approccio ibrido alla valutazione della comprensione facciamo riferimento, ancora una volta, a Tullio De Mauro che pur riconoscendo che

“le esperienze preliminari in ordine a un certo argomento possono facilitare la comprensione di discorsi sull'argomento stesso”

obietta che pur volendo considerare possibile un'analisi che tenga conto delle esperienze pregresse è difficile stabilire quali sono le diverse esperienze di parlanti nativi, e i criteri per prevedere i problemi di comprensione sarebbero troppo legati alla soggettività del singolo

---

<sup>9</sup> [www.corrige.it](http://www.corrige.it) Nicola Mastidoro, Eugenio Mastidoro, Fabio Pesce

<sup>10</sup> Periodico curato da Maria Emanuela Piemontese di facile lettura distribuito tra il 1989 e il 2006, prima in formato cartaceo a cura dell'Università di Roma La Sapienza, e poi online dall'associazione Parlar chiaro. Il giornale particolarmente semplice nella forma e nel contenuto, si rivolgeva ad adulti con ritardo mentale o disabilità fisiche che ne limitavano le capacità di lettura. La redazione lavorava su base volontaria e si proponeva di agevolare l'inserimento di queste persone nella società rendendo loro accessibili notizie di attualità, nozioni culturali e di vita quotidiana. Il giornale è diventato anche uno strumento didattico in alcune scuole per stranieri. (Piemontese, 1996).

lettore in quanto legati alle loro esperienze. Invece se facciamo riferimento alle parole è possibile arrivare a sapere quante ne conoscano con sicurezza le persone. (Lumbelli 2009).

## **2. Comprensione e leggibilità**

Per verificare il grado di leggibilità e di comprensibilità di un testo si può ricorrere a due tipi di approccio: uno quantitativo e uno qualitativo.

L'approccio quantitativo dà indicazioni di carattere numerico, riguarda l'aspetto linguistico del testo, e riguarda l'aspetto lessicale e sintattico. L'approccio qualitativo invece ci aiuta ad evidenziare gli ostacoli della comprensione legati alle difficoltà di chi legge, che derivano dal gap tra le operazioni linguistico-cognitive richieste dal testo e quelle che effettivamente è in grado di attivare il lettore.

Leggibilità e comprensibilità analizzano aspetti differenti: la leggibilità si riferisce agli aspetti strutturali del testo, quali lunghezza, struttura sintattica, lessico; la comprensibilità si riferisce invece agli aspetti profondi, logico-semantiche del testo quali la maggiore o minore chiarezza dello stesso, l'adeguatezza dei contenuti alle conoscenze del lettore. Di conseguenza i due aspetti pur se legati tra loro riguardano sfere diverse dell'approccio alla lettura, infatti anche se un testo risultasse del tutto privo di ostacoli formali potrebbe comunque non essere adatto alle competenze dei lettori a cui si rivolge. Questo problema ha una sua valenza specifica soprattutto nei testi scolastici.

Mentre la leggibilità è una caratteristica strutturale del testo ed è misurabile, la comprensibilità è una caratteristica relativa, legata alle conoscenze e competenze del lettore rispetto al testo, alla sua struttura, al suo contenuto.

Come abbiamo già avuto modo di evidenziare per la misurazione della leggibilità dei testi l'italiano può contare sulla formula di Gulpease (Lucisano, Piemontese 1988), che – come gli indici di leggibilità creati per la lingua inglese – si basa sull'osservazione della lunghezza delle parole e delle frasi ritenendo che parole più corte e frasi più brevi, che abbiano una struttura sintattica più semplice, permettano una più facile comprensione.

I risultati ottenuti con la formula di Gulpease sono compresi tra 0 e 100: il valore 100 indica la leggibilità più alta. La formula è anche legata al grado di istruzione. Ad esempio, un risultato  $G=52$  è quasi incomprensibile per un lettore con la licenza elementare, lo stesso risultato indica, invece, un livello di lettura difficile per il lettore con la licenza media e facile per quello con il diploma di istruzione superiore. In conclusione, nell'indice di Gulpease l'indicatore della facilità di lettura viene misurato in relazione al grado di istruzione e di preparazione del lettore.

Un altro strumento per valutare la leggibilità di un testo è il Vocabolario di Base dell'Italiano (De Mauro 1980), lo strumento realizzato da Tullio De Mauro che comprende circa 7.000 parole, che sono distinte in tre fasce (le circa 2000 parole fondamentali, che costituiscono il cuore della nostra lingua, le 3000 parole di alto uso che si pronunciano e si scrivono spesso, e le 2000 parole strategiche, usate meno di altre, ma familiari a tutti ed essenziali per la vita quotidiana), e sono le parole con maggiore frequenza statistica nella nostra lingua.

Il lavoro di ricerca che è stato svolto in questa tesi si innesta su questo filone ed ha l'obiettivo di valutare, grazie a strumenti computazionali, il livello di leggibilità del lessico dei testi della letteratura per l'infanzia, proponendo un nuovo *Indice di Leggibilità* calcolato automaticamente sulla base di feature lessicali e sintattiche.

In particolare la nostra ricerca si pone tre obiettivi:

- verificare la scelta dei vocaboli in relazione all'ipotesi De Mauro;
- sperimentare un sistema di valutazione della complessità sintattica;
- elaborare un sistema integrato automatico per produrre un *Indice di Leggibilità lessicale e sintattico*.

Nel caso della comprensione del testo nei bambini, l'indice di leggibilità è solo un primo dato, non sufficiente a stabilire se il testo sia adatto ai lettori a cui è rivolto. La comprensione non si esaurisce quindi sul piano linguistico: capire un testo non significa solo decodificarne la lingua, cioè identificare le relazioni tra le parti del discorso, ma anche riconoscere le relazioni tra i vari concetti. È necessario, cioè, attivare processi di tipo cognitivo che coinvolgono diversi piani della lingua: quello lessicale, sintattico e quello semantico.

Per ciò che concerne il piano lessicale dobbiamo prendere in considerazione il fatto che la fascia di lettori presa da noi in analisi cioè bambini dai 7 agli 11 anni sono in una fase importante di acquisizione del lessico. A questo proposito Silvana Ferreri dice:

«Anno dopo anno ogni allievo acquisisce quasi senza sforzo parole del vocabolario di base, del più ampio vocabolario comune e talora persino parole desuete, come effetto secondario dell'incontro con gli altri: parole pronunciate dagli insegnanti e dai compagni, ascoltate per radio o in televisione, lette sui giornali o nei libri; parole provate oralmente, messe per iscritto, cantate. Di questa acquisizione spontanea la scuola si avvantaggia ma non può accontentarsi. Buona parte del vocabolario di base non è noto e non lo sono certamente le molte accezioni di senso dei vocaboli più frequenti» (Ferreri 2005).

Quindi è fondamentale occuparsi dell'acquisizione del lessico sin dai primi anni della scuola primaria, prestando una particolare attenzione al significato di termini che per i bambini risultano nuovi e facendo in modo che nell'acquisizione del lessico si abbiano tre dimensioni: ampiezza cioè la consistenza del patrimonio lessicale, profondità intesa come l'insieme delle conoscenze connesse con ogni parola e automaticità cioè l'inserimento della parola nella memoria a lungo termine che permette l'utilizzazione della parola acquisita (Sobrero 2009). Dal punto di vista linguistico la migliore valutazione della leggibilità di un testo è l'analisi del lessico e delle strutture sintattiche, in quanto l'uso di una parola e la sua comprensione coinvolge direttamente il grado di conoscenza della lingua posseduto dal lettore. Il livello di conoscenza della lingua, inteso come numero di parole delle quali si conosce e utilizza il significato, è dunque un parametro fondamentale da misurare quando si cerca di valutare e prevedere la comprensibilità di un testo.

Il nostro lavoro, partito dall'analisi linguistica di testi di letteratura per l'infanzia, attraverso l'applicazione di un indice di leggibilità da noi sviluppato, vuole verificare la correlazione che c'è tra "leggibilità di un testo" e "grado di comprensione" dello stesso che può essere raggiunto da bambini che frequentano la scuola primaria.

Il target a cui ci rivolgiamo ci pone di fronte a due problemi specifici:

- il primo ci mostra la capacità di lettura come capacità di identificare e nominare lettere e parole di un testo in modo corretto e veloce (lettura strumentale e decodifica);
- il secondo la capacità di lettura vista come capacità di comprendere un testo scritto.

In letteratura questi due aspetti sono stati correlati in quanto si è ritenuto che una mancata acquisizione della lettura strumentale inficiasse la comprensione del testo. Anche se in parte questo è vero bisogna sottolineare che la difficoltà di comprensione non sempre è legata alla difficoltà di decodifica.

Negli studi di L.S. Vygotskij ( Vygotskij Lev S. 1956) l'insegnamento sistematico precede lo sviluppo cognitivo e, dunque sollecitando le risorse dell'alunno è possibile favorire lo sviluppo cognitivo; l'insegnamento interviene in quell'area che viene definita da Vygotskij area di sviluppo prossimale, l'area delle competenze non ancora sviluppate ma in formazione, che permettono all'apprendente di svolgere compiti complessi in una sorta di ponte tra le capacità di sviluppo attuali del bambino e quelle potenziali, ottenibili attraverso l'interazione con una persona più esperta. Attività come la semplificazione e la facilitazione di un testo complesso possono essere viste come il sostegno offerto dal docente all'alunno perché possa esercitare competenze non ancora del tutto mature, ma comunque in via di formazione per svolgere compiti complessi. Numerosi studi si sono focalizzati sulle abilità e sulle strategie cognitive coinvolte nel processo di lettura qui mi limiterò a presentare il modello a due vie elaborato da Coltheart e collaboratori (1993; 2001). Le due vie individuate nel modello sono:

- La via lessicale semantica diretta: questa via coinvolge le informazioni contenute nella memoria a lungo termine; in questo caso infatti la parola viene riconosciuta in quanto già presente nel lessico mentale e resa disponibile nella sua forma ortografica (la parola come è scritta), fonologica (la parola come suona) e nel suo valore semantico (la rappresentazione o il significato cui la parola rimanda).
- La via fonologica sub lessicale: questa via prevede il riconoscimento della parola, e ne permette la lettura. Ciò avviene attraverso tre operazioni conseguenti: la scomposizione o segmentazione della parola nelle singole parti costituenti o grafemi, l'associazione tra ciascun grafema e il fonema corrispondente, il riassetto in forma fonologica della parola, con l'ovvia inversione del processo (scomposizione in fonemi, corrispondenza fonema-grafema, riassetto della forma grafemico ortografica della parola) nel caso in cui la parola debba essere prodotta, cioè scritta.

La via fonologica sub lessicale consente di leggere parole nuove o inventate (non parole).

La via lessicale può essere utilizzata per leggere le parole che già si conoscono e tale via potrebbe essere attivata in maniera più o meno rapida in funzione della frequenza d'uso della parola. Inoltre, ha un ruolo importante nella lettura delle parole irregolari. Quest'aspetto è particolarmente vero, se consideriamo le lingue opache, cioè quelle lingue nelle quali la pronuncia non corrisponde a quella ottenuta dalle regole di conversione grafema-fonema, si pensi alla lingua inglese nella quale per esempio la pronuncia della parola have/haev non corrisponde a quella che si ottiene dalla conversione grafema-fonema, cosa che invece avviene per la parola save/seiv.

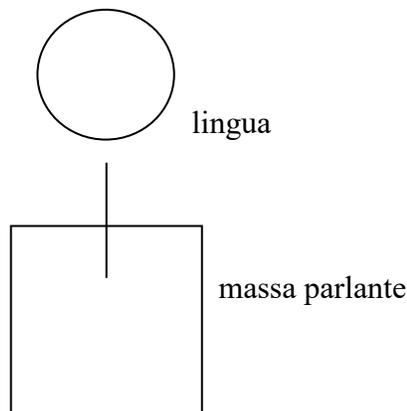
# CAPITOLO III

## L'APPROCCIO STATISTICO LINGUISTICO

### 1. La statistica linguistica

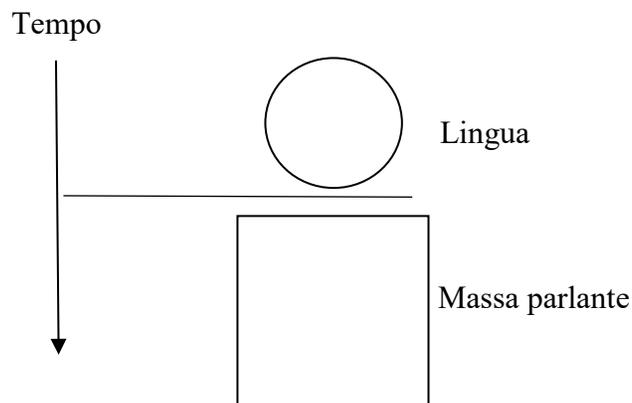
Lo studio della lingua è sempre stato visto come uno studio prevalentemente qualitativi. Detto ciò non bisogna sottovalutare che fin dagli antichi Greci era presente l'idea di uno studio quantitativo, basato sul confronto numerico delle parole (hapax legòmena – parole dette una sola volta). A partire dagli inizi del '900 con il variare delle prospettive metodologiche si ebbe l'evidenza della necessità di introdurre nello studio della linguistica la statistica. È stato osservato più volte che, usando le parole di G. Lepschy, “gli eventi linguistici si rivelano particolarmente adatti al trattamento statistico”. Proprio perché la “langue” non è un fatto matematico ma un insieme di forme e regole legate al “temps” e alla “masse parlante” (Saussure 1916).

... "occorre una massa parlante perché vi sia una lingua. Contrariamente all'apparenza, in nessun momento storico la lingua esiste fuori dal fatto sociale, perché essa è un fenomeno semiologico. La sua natura sociale è uno dei suoi caratteri interni. La sua definizione completa ci colloca dinanzi agli occhi due cose inseparabili, come mostra lo schema...



Ma in queste condizioni la lingua è vitale, non viva; noi abbiamo tenuto conto solo della realtà sociale e non del fatto storico... Poiché il segno linguistico è arbitrario, sembra che la

lingua, così definita, sia un sistema libero, organizzabile a volontà, dipendente unicamente da un principio razionale. Il suo carattere sociale, considerato in se stesso, non è in netta opposizione con questo punto di vista. (...) ciò che ci vieta di guardare alla lingua come a una semplice convenzione, modificabile a piacere dagli interessati... è ... l'azione del tempo che si combina con l'azione della forza sociale: fuori della durata, la realtà linguistica non è completa e nessuna conclusione è possibile. Se si prendesse la lingua nel tempo, senza la massa parlante... non si constaterrebbe forse nessuna alterazione: il tempo non agirebbe su di essa. Inversamente, se si considerasse la massa parlante senza il tempo, non si vedrebbe l'effetto delle forze sociali agenti sulla lingua. Per restare nella realtà bisogna dunque aggiungere al nostro primo schema un segno che indichi il cammino del tempo.



Perciò la lingua non è libera, perché il tempo permetterà alle forze sociali esercitanti su di essa di sviluppare i loro effetti, e si arriva al principio di continuità, che annulla la libertà. Ma la continuità implica necessariamente l'alterazione, lo spostamento più o meno considerevole dei rapporti.” (CLG: pp. 95-97).

Nel sistema linguistico sono presenti delle unità minime, le parole, che considerate sotto il profilo della frequenza, cioè il numero delle volte che appaiono in un testo, costituiscono un

fenomeno di massa, sottoponibile ad indagine statistica, al fine di rilevare la frequenza media del loro distribuirsi nel discorso e nel tempo. (Elia 2007)<sup>11</sup>

L'uso rappresenta un punto di svolta per lo studio del lessico, poiché su di esso è possibile basare l'utilizzo della parola e il significato ad essa attribuito.

Questo il senso del lavoro svolto da Tullio De Mauro nella direzione di una linguistica statistica che imposta lo studio del lessico sugli usi lessicali, attraverso analisi di tipo quantitativo, capaci di cogliere in qualche modo gli usi linguistici.

Le parole acquistano significato solo nella misura in cui sono usate dall'uomo, è il legame con la società che permette al significato di essere compreso.

È il comportamento linguistico degli individui, cioè il loro uso della lingua e dei suoi significati che garantisce il collegamento tra forma e significato e che consente all'uomo di trasmettere dei significati comprensibili nonostante la comunicazione non sia sempre perfetta e nonostante il fatto che non esistono due individui con lo stesso identico patrimonio linguistico e lessicale. Per analizzare gli usi lessicali dei parlanti è possibile sfruttare l'apporto che la statistica può fornire alle scienze linguistiche<sup>12</sup>.

In modo particolare è al lavoro del filologo statunitense G. K. Zipf che si deve la prima, importante sistematizzazione delle regolarità statistiche in ambito linguistico.

In precedenza i richiami al fatto che esistano degli "ordini delle parole" (De Mauro, 1980) provengono tuttavia da più fila: già presso i greci e i romani, ad esempio, filologi e grammatici osservano la diversa occorrenza delle parole nei testi, individuando voci rare o obsolete (o addirittura hapax legòmena) e parole di particolare frequenza e uso; nel Medioevo, poi, hanno origine le prime concordanze (indicazioni delle parole e delle frequenze di occorrenza) per i testi biblici e dell'antichità greco-latina così come anche i primi indices verborum (raccolta delle parole con la sola indicazione numerica dei luoghi di occorrenza).

---

<sup>11</sup> La lezione di Saussure Saggi di epistemologia linguistica a cura di Annibale Elia e Marina De Palo – Sintassi, lessico e massa parlante A. Elia 2007.

<sup>12</sup> Si veda in proposito il volume curato da Tullio De Mauro e Isabella Chiari sull'applicazione delle analisi quantitative ai fatti di lingua (De Mauro, Chiari, 2005). De Mauro è stato da sempre uno dei più convinti sostenitori nel panorama italiano dell'utilità dell'applicazione della statistica allo studio delle lingue storico-naturali, come dimostrano, tra gli altri, lavori come il contributo sulla statistica linguistica in Enciclopedia Italiana (1961).

Il maggiore interesse per questo tipo di analisi, però veniva da altri ambiti, primo fra tutti quello stenografico. Il lavoro più rappresentativo in questo campo è quello svolto da J. B. Estoup (1916), *Les Gammes sténographiques*, in cui viene stabilita la fondamentale relazione tra la frequenza (il numero di occorrenze) di un termine e il suo rango (il posto da esso occupato) in una lista ordinata per valori di frequenza decrescenti (De Mauro, Chiari, 2005). Proprio questa relazione, infatti, fornisce uno dei principali presupposti teorici della successiva sistematizzazione da parte di Zipf (1935).

Una delle principali leggi lessico-statistiche, come si è visto in Estoup, è quella che descrive la relazione tra il rango del vocabolo ( $r$ ) e la sua frequenza ( $f$ ) in una lista di parole detta anche “legge di Zipf-Estoup” o anche solamente “legge di Zipf”: il prodotto tra i due è inversamente proporzionale, per cui ad un rango più basso corrisponde una frequenza maggiore e viceversa. Al fine di verificare la relazione è necessario: contare tutte le occorrenze di parole diverse in un testo, sistemarle in ordine di frequenza decrescente, attribuendo un numero a ciascun rango e moltiplicare il numero di rango ( $r$ ) per la frequenza ( $f$ ).

Una seconda caratteristica del funzionamento lessicale individuata da Zipf è la relazione tra la frequenza della parola e il numero di fonemi che la compongono, detta anche “legge di Zipf-Guiraud”: questo è infatti direttamente proporzionale al rango della parola e cioè decresce all’aumentare della frequenza stessa.

Zipf alla base del proprio lavoro considerò il principio del minimo sforzo. Al pari di ogni altra forma di attività umana, anche il linguaggio risulta governato dal “principio del minimo sforzo”, in base al quale

«a few words occur with very high frequency while many words occur but rarely» (Zipf, 1935).

Come ha osservato Tullio De Mauro la stratificazione statistica del lessico è correlata alla forma del significante, infatti in media le parole del vocabolario fondamentale sono più brevi delle parole della fascia tecnico-specialistica (De Mauro 1961).

Vi sono poi due ultimi aspetti da considerare: da una parte la relazione tra parole più frequenti e genericità semantica, secondo cui tanto più una parola è frequente, tanto maggiore è il

numero di accezioni che possiede come suoi significati (De Mauro, 1961). Dall'altra parte c'è la relazione tra frequenza lessicale e varietà delle parole usate in un testo, secondo cui un testo è costituito per lo più da un ristretto numero di parole di alta frequenza e da numerose altre di bassa occorrenza (Chiari, 2004).

Pertanto, sono poche le parole che coprono una percentuale altissima dei testi prodotti in una data lingua, mentre è molto più ampio il numero di unità lessicali registrate nei vocabolari con occorrenza particolarmente bassa.

Si ritiene, quindi, che i 1.000 lessemi più frequenti coprano, attraverso le proprie occorrenze e ripetizioni, l'80% di ogni testo; secondo stime ulteriori di copertura testuale le 2.000 parole più frequenti coprirebbero il 90% dei testi e le 4.000 più ricorrenti il 97,5% (Guiraud, 1954; De Mauro, 1961).

Questo principio si rivela interessante da un punto di vista didattico, infatti, secondo questa logica un numero limitato di parole sembrano in grado di assicurare la massima efficacia linguistico-lessicale, vale a dire la produzione e la comprensione della maggior parte dei testi di una lingua.

A partire dal XX secolo si sono cominciate ad utilizzare le liste di frequenza per "semplificare" la comprensione e supportare l'apprendimento di persone con difficoltà linguistico-culturali e intellettive.

Numerosi dizionari di base vengono realizzati attraverso la consultazione di corpora testuali ritenuti rappresentativi della lingua esaminata, in altri termini le parole più frequenti che venivano individuate al loro interno vengono ritenute anche le più frequenti nella maggior parte dei testi.

L'elaborazione dei dizionari di frequenza hanno preceduto la preparazione dei vocabolari fondamentali, infatti i vocabolari fondamentali rappresentano oggi la zona di alta o di massima frequenza d'uso del lessico di una lingua.

Tra i primi lessici fondamentali possono essere ricordati quelli sul tedesco di Kaeding (1897), uno dei più ampi dal punto di vista dell'estensione del corpus tra i dizionari di frequenza di prima generazione, sono seguiti quelli sull'inglese di Thorndike (1921, 1931-32), di Vander Beke (1930) sul francese, di Buchanan (1927) sullo spagnolo.

Diversi lavori poco conosciuti di questo tipo sono stati svolti nell'URSS anche sulle lingue europee e soprattutto a fini pedagogici e per l'insegnamento delle lingue straniere (cfr. Alekseev, 1973). Tra i dizionari di seconda generazione invece troviamo i lavori di Josselson (1953) e Zazorina (1977) sul russo, il dizionario dell'inglese americano di Kučera e Francis (1967), la serie di dizionari di Juilland et alii (1970) sulle lingue romanze a partire da rumeno, spagnolo e francese, e il Lessico di frequenza della lingua italiana LIF Zampolli et al., (1971) per l'italiano e il più recente Lessico di frequenza dell'italiano parlato LIP De Mauro et al (1993).

Alla terza generazione appartengono invece i dizionari di frequenza costruiti a partire dai grandi corpora di riferimento delle lingue europee, come ad esempio il Leech et alii (2001) basato sul British National Corpus (De Mauro, Chiari 2005).

I lessemi considerati si basano su campioni di lingua scritta prevalentemente di tipo letterario. Solo a partire dalla metà degli anni Cinquanta cominciano a farsi strada lavori di frequenza su corpora di lingua esclusivamente parlata: il Français Fondamental (1er degré) (Gougenheim et alii, 1956; 1964) può essere considerato lo studio pionieristico di questo filone di ricerca.

L'approccio statistico ai fenomeni linguistici è stato criticato sia dai linguisti di formazione umanistica sia da quel filone di ricerca che segue una visione algebrica del funzionamento delle lingue e considera irrilevanti le analisi statistiche. Ma al primo si può far notare che, come in ogni altro campo delle scienze umane la quantificazione statistica serve a sistematizzare e precisare le intuizioni, il ruolo e le origini dei fenomeni linguistici osservati. Un esempio di applicazione dell'analisi statistica è legato all'utilizzo di programmi di lemmatizzazione automatica che ci rileva l'esistenza di una percentuale di omonimi nei testi che, in linea di massima, si aggira intorno al 50%. Questo tipo di lemmatizzazione nasce dall'analisi statistica di estesi corpora all'interno dei quali si è osservata un'ampia presenza di forme omonime e pluri-etichettabili sotto il profilo morfologico e semantico.

Le caratteristiche di un testo sottoponibili a indagine quantitativa sono molte: scelta lessicale, valutazione della leggibilità, relazione tra lunghezza delle parole o delle frasi e testo, concentrazione delle frequenze lessicali, comparazione tra lingue diverse in testi simili,

stilometria forense. I primi testi ad essere analizzati sono stati testi di tipo letterario con diverse finalità:

- a) l'individuazione delle particolarità che si trovano nelle scelte lessicali dell'autore di un testo;
- b) la determinazione delle caratteristiche linguistiche (lessicali, e morfo-sintattiche) che differenziano e/o accomunano testi di tipologie diverse;
- c) la predisposizione di strumenti stilo-statistici che permettano di risolvere problemi di attribuzione della paternità di un'opera a un autore o a un altro.

A partire da questo tipo di lavori che risalgono già agli inizi del '900 si è giunti a elaborare test di comprensione e indici di leggibilità, legando a questo tipo di analisi variabili lessicali e sintattiche. Tant'è che le ricerche che hanno portato alla definizione degli indici di leggibilità hanno messo in evidenza quanto già affermato da alcune leggi di statica linguistica e cioè che “le parole frequenti sono quelle familiari, e le parole frequenti e familiari sono brevi, e le parole frequenti, familiari e brevi hanno pochi affissi, e lo scrittore che usi parole frequenti, familiari, brevi e senza affissi presenta un basso rapporto tipo-replica” (Miller 1972).

A tal proposito già alcuni studi Volgel e Washburne 1928 avevano individuato nel rapporto tipo-replica un buon indice di leggibilità. Infatti scrivere dei testi con un alto rapporto tipo-replica significa ricorrere a parole meno comuni, più lunghe e difficili da capire rispetto a quelle comuni. (Piemontese 2005). Se andiamo a valutare l'aspetto sintattico possiamo notare che “la proporzione di frasi semplici in un periodo aveva un'alta correlazione con la lunghezza del periodo espresso in numero di parole” (Lucisano 1992).

Dal punto di vista quantitativo e statistico il lessico, uno degli elementi che a noi interessa in questo studio, pone tre tipi di problemi:

- a) quale è l'unità di popolazione o unità di analisi del lessico e quale debba essere l'ampiezza del corpus perché esso possa essere rappresentativo degli usi lessicali di una lingua o di un testo;
- b) la costruzione del vocabolario di base di una lingua in relazione alla frequenza d'uso delle parole;

- c) la valutazione articolata della cosiddetta ‘ricchezza del vocabolario’ di un testo, di un autore, ecc.

Approfondiremo questi aspetti nel paragrafo successivo sulla linguistica dei corpora.

## **2. Linguistica dei corpora e linguistica computazionale**

La linguistica dei corpora non può essere considerato un settore specifico della linguistica ma un approccio, una metodologia. Negli anni cinquanta il linguista americano Noam Chomsky ha rifiutato di considerare “la linguistica dei corpora” adeguata alla ricerca linguistica soprattutto quando si è cominciato a lavorare sulla grammatica generativa trasformazionale.

Per Chomsky una linguistica dei corpora è inadeguata perché un corpus è finito (chiuso), e incompleto ed esclude frasi possibili ma non realizzate, e per la sua incompletezza non permette di fare delle predizioni precise sulla grammaticalità delle frasi, ma ci dà solo indicazioni sulla frequenza, con la quale si manifestano determinate caratteristiche linguistiche (Chiari 2007).

La linguistica dei corpora può essere considerata un metodo di indagine sul linguaggio che prende in considerazione un insieme finito di testi che possono essere tratti dal parlato e poi trascritte o direttamente costituiti da testi scritti. Pertanto la sua limitatezza e l'impossibilità di rappresentare tutte le frasi possibili porta, molti tra cui Chomsky, a dire che lo studio della linguistica non può essere rappresentato da un insieme finito di enunciati, soprattutto in relazione alle strutture implicite e inconsce del linguaggio naturale per il quale è necessario ricostruire le regole (lessico-grammatica). Quindi la linguistica di un corpus è utile se il corpus è costituito da un campione sufficientemente rappresentativo anche se non può pretendere di esaurire la complessità dell'analisi linguistica.

Anche Lepschy è scettico nei confronti della rappresentatività dei corpora:

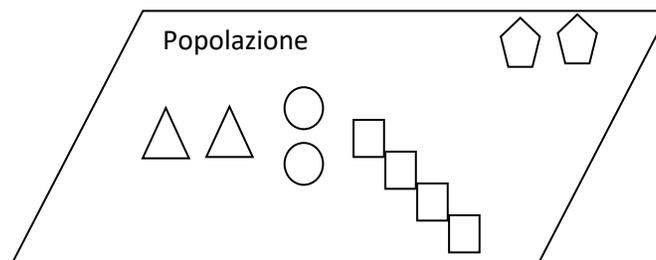
“Una lingua non è un insieme di testi; essa è piuttosto un sistema finito di regole che consentono di generare un numero infinito di testi. Non c'è raccolta che possa essere pienamente rappresentativa della lingua.” Lepschy (1978)

Pur ritenendo valida questa osservazione bisogna sottolineare due aspetti: il primo riguarda la diffusione delle tecnologie informatiche che ha reso possibile il trattamento di dati testuali molto ampi, il secondo riguarda la frequenza e l'evidenza che, pur non potendo la frequenza di una parola in un determinato corpus avere conseguenze dirette riguardo la sua frequenza nella lingua, è indubbio che i dati testuali reali, raccolti e organizzati in maniera complessa, per quanto parziali, ci permettono di estrarre dati preziosi sull'uso che i parlanti fanno della lingua.

È un dato di fatto che i corpora sono sempre campioni: il nostro obiettivo dovrebbe essere quello di estendere le osservazioni condotte sul corpus all'intera popolazione.

Partiamo qui dal significato delle parole «popolazione» e «campione» come si usano in statistica, e quindi nella linguistica dei corpora. Una popolazione è un insieme di tutte le possibili osservazioni in un dato ambito. Un campione invece è una sezione, una parte della popolazione, che include solo alcune delle possibili osservazioni.

Nel nostro caso specifico se volessimo attribuire un valore preciso all'indice di leggibilità di un testo narrativo, quale potrebbe essere "Marcovaldo" di Italo Calvino, dovremmo analizzare l'intero libro e non un solo capito dello stesso. Dal punto di vista teorico un campione dovrebbe essere scelto in modo che ogni possibile osservazione (ogni fenomeno linguistico, nel nostro caso) abbia la stessa probabilità di occorrere nel campione e nella popolazione. Il momento più delicato della costruzione di un campione risiede proprio nella scelta del materiale in modo che rappresenti la popolazione. Il campione deve, per l'aspetto che si intende studiare, essere adatto a esibire lo stesso tipo di informazioni (qualitative) con la stessa probabilità di occorrenza (quantitativa) della popolazione.



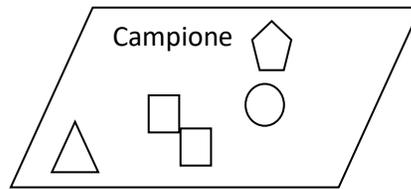


Figura 2 La rappresentatività tra campione e popolazione

Un campione per essere utile e significativo deve avere quattro caratteristiche: rappresentatività, estensione, formato elettronico e riferimento standard (McEnery e Wilson 1996).

La definizione della rappresentatività costituisce un momento cruciale nella costruzione di un corpus. La **rappresentatività**, quindi, è una caratteristica relativa che varia secondo l'aspetto linguistico che si intende studiare, pertanto un corpus rappresentativo per caratteristiche lessicali potrebbe non esserlo per caratteristiche di tipo sintattico oppure stilistico.

La rappresentatività dei corpora, nella maggior parte dei casi, si basa sulla ricchezza del vocabolario, misurata attraverso il numero di parole diverse presenti nel testo. Tuttavia un indicatore più agevole può essere considerato il numero di occorrenze (token) di parole grafiche presenti nel corpus.

Un'altra caratteristica dei corpora, che bisogna valutare con molta accuratezza, è la sua estensione, ossia la sua ampiezza.

L'**estensione** è una variabile che influenza il grado di rappresentatività di un campione testuale. Esistono diverse estensioni standard per esempio le analisi di tipo lessicale, condotte sui corpora, hanno delle soglie indicative minime per determinare un'estensione considerata sufficiente per i corpora. Nella Tabella 2 viene data un'indicazione quantitativa di massima per ricerche di tipo lessicale.

Tabella 2 Estensione di un corpus per analisi lessicale

---

<i>Corpus non rappresentativo (insuff.)</i>	<i>&lt;15.000 parole grafiche</i>
<i>Corpus piccolo</i>	<i>Da circa 15.000 a 100.000 parole</i>
<i>Corpus medio</i>	<i>Da circa 100.000 a 1 milione di parole</i>
<i>Corpus medio-grande</i>	<i>Da circa 1 milione a 50 milioni di parole</i>
<i>Corpus standard</i>	<i>Da 50 milioni a 100.000 milioni di parole</i>
<i>Corpus grande</i>	<i>Oltre i 100.000 milioni di parole</i>

---

Bisogna evidenziare però, che l'acquisizione di tecnologie di analisi automatica e l'utilizzo di applicazioni di linguistica computazionale hanno reso possibile il trattamento di corpus molto più ampi.

Un esempio di questo importante passaggio sono da un lato la lista di lemmi utilizzati per realizzare il Vocabolario di base dell'Italiano (VdB) redatto partendo dal Lessico di frequenza della lingua italiana (Lif) costituito da un corpus di 500.000 occorrenze di parole tratte da testi teatrali, periodici, romanzi, cinema e sussidiari e il corpus utilizzato per determinare la lista di lemmi del Nuovo Vocabolario di Base (NVdB) che si basa sull'analisi di un corpus appositamente costruito per l'italiano contemporaneo (scritto e parlato), di 18.000.000 parole.

Un' altro esempio di corpus molto esteso è il British National Corpus (BNC), costituito da circa 100 milioni di parole. Formato il corpus, i dati che da questo si estraggono sono considerati rappresentativi della lingua del periodo nel quale i testi sono stati raccolti. La maggior parte dei corpora è quindi statica. Lavorare su un corpus statico permette di estrarre e comparare analisi linguistiche sullo stesso materiale, inoltre può facilmente essere usato e comparato ad altri corpora statistici costruiti con criteri simili e tutte le analisi sono ripetibili. L'estensione di un corpus è tipicamente condizionata dalle risorse temporali ed economiche a disposizione per svilupparlo e fornisce una "fotografia" di una lingua attraverso i testi selezionati. Ma per avere in un corpus la rappresentazione realistica di una lingua, che è per

sua natura dinamica, sarebbe necessario avere un corpus di monitoraggio (monitor corpus) cioè una collezione di testi che muta nel tempo introducendo nuovi testi selezionati secondo gli stessi criteri usati per determinare la collezione iniziale.

A tal fine John Sinclair propone di trasformare il corpus in uno strumento di monitoraggio linguistico (Sinclair 1991). Questo tipo di corpus permette ad esempio di monitorare le dinamiche del lessico della lingua, e può essere usato in contesti lessicografici come fonte di dati per mantenere aggiornati i dizionari.

Il primo esempio di corpus di monitoraggio è la Bank of English, pensata e diretta da Sinclair: iniziato nel 1981, questo corpus ha raggiunto negli anni '90 un'estensione di circa 200 milioni di parole, per raggiungere i 450 milioni nel 2002. L'idea del corpus aperto è stata adottata anche per la costruzione di altri corpora quali il Coris per l'italiano e l'American National Corpus (ANC). Quest'ultimo è stato progettato per essere comparabile con il BNC, le cui dimensioni sono le stesse (circa 100 milioni di parole) e anche i criteri di selezione dei testi, l'ANC però comprende una sezione di monitoraggio aggiornata ogni 5 anni con l'aggiunta del 10% di nuovi testi presi da fonti e materiali digitali diversi.

In questo senso l'idea del corpus di monitoraggio si avvicina molto alle caratteristiche di dinamismo del web. A differenza del web, un corpus di monitoraggio però offre la possibilità di un controllo sul materiale selezionato per arricchire il corpus.

Un altro aspetto dell'evoluzione dei corpora da considerare è il formato elettronico (machine-readable form) che indica il trattamento informatizzato di dati testuali, archiviati digitalmente e consultabili attraverso software di interrogazione.

I primi corpora venivano conservati in forma cartacea oppure con schede perforate, oggi un corpus che non sia in un formato elettronico sarebbe inservibile. Avere a disposizione una versione machine-readable dei testi permette innanzitutto la conservazione di una quantità sempre maggiore di dati e in secondo luogo l'accesso ai testi mediante ricerche linguistiche complesse.

Per finire, il riferimento standard dà la possibilità a grandi corpora di diventare punti di riferimento per lo studio, facilitando il confronto tra corpora diversi e evitando il ripetersi di computi per analisi successive.

### **3. Comprensione del linguaggio e disambiguazione**

Il primo passo per la costruzione di un corpus è la pianificazione della sua struttura in modo che possa essere rappresentativa dello spaccato linguistico che si vuole analizzare. Il secondo passo è l'acquisizione dei testi per il corpus. Nel nostro caso abbiamo acquisito 21 testi di cui 11 testi integrali di letteratura per l'infanzia, scelti tra i libri pubblicati e tradotti tra il 1865 e il 1997 anno di pubblicazione del primo libro della serie Harry Potter di Rowling, Joanne Kathleen - Harry Potter e La Pietra Filosofale, libro con il quale si chiude il corpus e 10 libri ad alta leggibilità tratti dalle collane Il Mulino a Vento e Attacca Parole, edite da Raffaello, con l'intento di analizzare la leggibilità lessicale e sintattica attraverso la frequenza d'uso delle parole e rilevare l'eventuale legame tra la comprensione del testo da parte dei bambini e la presenza all'interno del testo stesso di parole che fanno parte del lessico fondamentale.

Nel corso degli ultimi decenni i corpora sono diventati uno strumento importante nello sviluppo di tecnologie linguistiche quali la traduzione automatica e il riconoscimento automatico del parlato e per comprenderne l'evoluzione, più avanti, faremo una breve presentazione dei lavori più importanti che sono stati realizzati per l'italiano.

Per poter utilizzare al meglio un corpus è necessario dotarlo di informazioni specifiche. Una prima informazione utile può essere l'aggiunta della categoria grammaticale alle parole che occorrono nel corpus, la divisione in capitoli, in paragrafi, indicare il titolo del testo, l'autore, tutte informazioni che il lettore umano conosce e dà per scontate e che invece devono essere esplicitate per una interpretazione computazionale corretta. Questo processo di codifica linguistica è detta annotazione e consiste nell'attribuzione ad ogni singola parte di un testo di un tag. L'annotazione o marcatura "è una sorta di meta-codifica dei testi, costituita da un insieme di simboli per etichettarne le diverse parti e di regole per utilizzarli.

Nel 1994 vengono pubblicate le Guidelines for the electronic text encoding and interchange, note anche come TEI P3 (oggi l'ultimo standard è il TEI P4 del 2002), in particolare come determinazione preliminare le Guidelines prevedono che siano utilizzati i linguaggi SGML e XML. Esiste, anche, una versione "semplificata" di TEI, molto utile per operare

un'annotazione rapida di un testo. Si chiama TEI-Lite ed è anche tradotta in italiano ([http://www.tei-c.org/Lite/teiu5\\_it.html](http://www.tei-c.org/Lite/teiu5_it.html)). L'annotazione può riguardare l'aspetto fonetico, morfologico, sintattico e semantico ed è fondamentale per poter estrarre nel modo più semplice possibile i dati linguistici. Un esempio di annotazione è quella morfo-sintattica detta anche POS tagging, che come vedremo in seguito riguarda l'attribuzione delle categorie grammaticali alla word token (nome, verbo, aggettivo, avverbio ecc). Quanto detto finora riguarda corpora di testi scritti, nel caso dei corpora del parlato le annotazioni potranno essere fonetiche, fonologiche, prosodiche e pragmatiche.

Uno dei problemi più difficili da risolvere nell'annotazione di un testo è la disambiguazione cioè l'attribuzione della categoria grammaticale corretta. Ovviamente questo problema deriva dall'ambiguità della lingua che di solito non compromette la nostra comprensione in quanto chi legge o ascolta non valuta la singola parola ma tiene conto degli elementi precedenti e seguenti delle sequenze sintagmatiche. Il testo permette di comprendere il significato delle parole anche quando le stesse prese singolarmente sono ambigue. Tale ambiguità può derivare da significati diversi (polisemia) con pronunce identiche (omofone) e da significati diversi con grafia identica (omografe). Inoltre, ci possono essere significati che possono avere la stessa forma grafica ma appartengono a due parole diverse (omonimia) o due forme grafiche diverse che esprimono lo stesso significato (sinonimia).

Per esempio la parola mente può essere:

- a) sostantivo femminile - FO la sede in cui ha luogo l'attività del pensiero (mille pensieri mi si affollano nella mente);
- b) verbo intr. e tr. FO affermare il falso o negare il vero con deliberato proposito, alterare consapevolmente o tacere indebitamente la verità: mentire (Quell'uomo mente sapendo di mentire).

De Mauro (1994) evidenzia un tasso di omografi per l'italiano che va dal 38% al 46% rispettivamente in testi economici e nella lingua parlata, per alcuni tipi testuali si arriva al 57%. È quindi chiaro che quando un corpus viene annotato è necessario individuare e distinguere le occorrenze ambigue. L'annotazione può essere effettuata in modo automatico, semi-automatico e manuale. Per corpus di grandi dimensioni questa disambiguazione viene

effettuata attraverso applicazioni di Natural Language Processing rule based parsing (basate su regole) o statistical parsing. Nel caso del Rule Based Parsing per effettuare un'analisi all'interno di un testo, in corpus di grandi dimensioni, è necessario fare ricorso alle regole, cioè ad una formalizzazione delle strutture sintattiche di una determinata lingua, come avviene nell'approccio lessico-grammaticale inaugurato da Maurice Gross<sup>13</sup> e di cui discutiamo nel capitolo IV.

Con un approccio che utilizza tecniche di disambiguazione automatica grazie soprattutto alla possibilità di utilizzare testi in formato elettronico e grazie allo sviluppo di metodologie basate su regole o basate su statistiche, volte ad individuare le regolarità linguistiche e le strutture peculiari nell'analisi dei testi, è possibile trovare una soluzione informatica che abbia come obiettivo quello di verificare leggibilità e comprensione di un testo per finalità didattiche.

La comprensione del linguaggio naturale rimane dunque uno dei problemi aperti nel campo della Artificial Intelligence (AI). Il Natural Language Understanding (NLU), anche noto con il nome di machine reading comprehension o text understanding, è stato classificato come problema AI-Complete, ovvero uno di quei problemi che sono semplici per gli esseri umani, ma per i quali ad oggi si cerca una soluzione in termini computazionali. Il text understanding

---

<sup>13</sup> In questa nota indichiamo la bibliografia essenziale di riferimento del lessico-grammatica.

Boons Jean Paul, Guillet Alain e Leclère Christian, 1976. *La structure de la phrase simple en français. Les constructions intransitives*, Genève; Guillet, Alain, Leclère, Christian. 1992. *La structure des phrases simples en français. Constructions transitives locatives*, Droz: Genève; D'Agostino Emilio, 1984. *Les compléments de lieu comme compléments de verbe dans les constructions transitives italiennes*, in A. Guillet - N. La Fauci (eds), *Lexique et grammaire comparées des langues romanes*, Amsterdam; Elia Annibale, 1984a. *Le verbe italien. Les complétives dans les phrases à un complément*, Fasano-Paris; Elia Annibale. 1984b. *Lessico-grammatica dei verbi italiani a completiva. Tavole ed indice generale*, Napoli; Elia Annibale, 1986. *Sul lessico nella grammatica generativo-trasformativa*, in "Quaderni dell'Ist. di Linguistica dell'Università di Urbino", 4, 1986, 5-34; Elia Annibale, Martinelli Maurizio e D'Agostino Emilio, 1981. *Lessico e strutture sintattiche. Introduzione alla sintassi del verbo*, Napoli; Gross Maurice, 1975. *Méthodes en syntaxe*, Paris; Gross Maurice, 1981. *Les bases empiriques de la notion de prédicat sémantique*, "Langages", 63; Gross Maurice, 1982. *Une classification des phrases figées du français*, "Revue Québécoise de Linguistique", 11; Harris Zellig, 1976. *Notes du cours de syntaxe*, Paris; Vietri S., *Lessico-grammatica dell'italiano*, Torino, Utet 2004; Vietri, Simonetta., Elia, Annibale, D'Agostino, Emilio. 2004, *Lexicon-grammar, Electronic Dictionaries and Local Grammars in Italian*, in Laporte, Eric, Leclère, Christian, Piot, Mireille, Silberstein Max, eds.. *Syntaxe, Lexique et Lexique-Grammaire* Volume dédié à Maurice Gross. *Linguisticae Investigationes Supplementa* 24, John Benjamins Publishing : Amsterdam/Philadelphia.

consiste nel leggere un testo espresso in linguaggio naturale, determinarne il significato attribuendo un significato ai lemmi e alle frasi ed effettuare inferenze su questi. In particolare, sono state proposte molteplici metodologie, alcune delle quali accedono a basi di conoscenza pregresse (conoscenze enciclopediche), altre invece che analizzano il contenuto del corpus preso in esame senza far ricorso alla conoscenza pregressa.

In buona sostanza il processo di disambiguazione può avvenire sia attraverso il contesto del lemma che deve essere disambiguato, cioè attraverso l'analisi delle concordanze, sia attraverso l'utilizzo di informazioni esterne, quali risorse lessicali o enciclopediche, dizionari elettronici o l'apporto di conoscenza esperta.

Partendo quindi da una metodologia basata sull'utilizzo di corpora possiamo distinguere tra una linguistica corpus-based e una linguistica corpus-driven (Elena Tognini Bonelli 2001). La prima fa uso dei corpora per integrare le teorie con dati linguistici reali, naturali, prelevati da testi effettivamente prodotti. Si tratta di un approccio già attestato, per esempio, nella predisposizione di grandi dizionari come il Dictionary of the English Language (1755-1756) di Samuel Johnson.

La linguistica corpus-driven, invece, fonda la costruzione ed elaborazione di teorie e ipotesi linguistiche generali esclusivamente sui dati estratti dai corpora.

Come già abbiamo visto più sopra anche l'elaborazione di un lessico grammatica e il suo collegamento ad un sistema di dizionari morfologici elettronici può aiutare a rispondere a questi quesiti.

I dizionari elettronici altro non sono che dei database lessicali che contengono informazioni sintattiche e semantiche circa le parole in uso in una determinata lingua. È possibile costruire dizionari di diversi tipi ed includere in essi quante più informazioni possibili.

Anche nell'analisi del nostro corpus, a cui sono stati applicati i dizionari elettronici arricchiti dalle marche d'uso del NVdB, si è reso necessario individuare e distinguere le occorrenze ambigue.

Nella tabella 3 riportiamo un estratto del dizionario elettronico dei verbi di un breve passo del Cap. 5 del Piccolo Principe arricchiti con le marche d'uso estratte del NVdB:

Tabella 3 Dizionario elettronico dei verbi di un passo del *Piccolo Principe*

può/<V+CO>	vuole/<V+TS>
può/<V+FO>	vuole/<V+CO>
spuntare/<V+TS>	vuole/<V+FO>
spuntare/<V+CO>	tratta/<V+OB>
spuntare/<V+AU>	tratta/<V+BU>
spuntare/<N+TS>	tratta/<V+LE>
spuntare/<N+CO>	tratta/<V+TS>
spuntare/<N+AU>	tratta/<V+CO>
vuole/<V+BU>	tratta/<V+AU>
vuole/<V+LE>	tratta/<V+FO>

Gli studi linguistici basati su grandi testi esisteva da molto prima dell'invenzione dei computer (basti pensare alle compilazioni di dizionari del XVIII secolo che utilizzavano grandi quantità di citazioni), ma è nel 1964 che è stato realizzato il primo corpus elettronico di lingua inglese americano il Brown Corpus di circa 1.000.000 di parole. In tempi più recenti il Cobuild, poi denominato Bank of English, realizzato da John Sinclair nel 1980 ha raccolto il più grande corpus della lingua inglese. Da questo lavoro nel 1987 è nato il primo dizionario creato interamente dal computer.

Per quanto riguarda la lingua italiana lo sviluppo dei corpora è assai più recente.

I corpora maggiori propongono tipologie testuali, estensioni quantitative metodologie di trattamento molto diverse. Il primo corpus italiano dello scritto risale agli anni Settanta ed è il corpus del Lessico di frequenza dell'italiano contemporaneo (Bortolini et al. 1972), ormai non più disponibile, composto da una raccolta di diversi generi testuali (teatro, romanzi, cinema, periodici, sussidiari scolastici) per un totale di circa 500.000 occorrenze.

Successivamente è stato raccolto il corpus "Corpus e Lessico di Frequenza dell'Italiano Scritto" (COLFIS, [www.ge.ilc.cnr.it/strumenti.php](http://www.ge.ilc.cnr.it/strumenti.php)), costituito da circa 3.800.000 occorrenze appartenenti a tre tipologie diverse (quotidiani, periodici, libri). La particolarità di questo corpus e del lessico di frequenza da questo tratto è la metodologia di selezione dei

testi, basato sui dati ISTAT sulle letture degli italiani, raccolti tra il 1992 e il 1994 (Laudanna et al. 1995).

Il primo e unico corpus di italiano scritto che si proponga in senso stretto come corpus di riferimento è il Corpus di Italiano Scritto contemporaneo (<http://dslo.unibo.it/>), distribuito in forma statica (CORIS) e dinamica (CODIS) ed elaborato presso il Centro interfacoltà di linguistica teorica e applicata «Luigi Heilmann» di Bologna (CILTA) e diretto da R. Rossini Favretti (Rossini Favretti et al. 2002). Il corpus è costituito da 130 milioni di occorrenze appartenenti a diverse tipologie viene periodicamente implementato.

Un corpus più recente è il Perugia Corpus (PEC, <http://perugiacorpus.unistrapg.it/>), corpus di riferimento dell'italiano contemporaneo progettato da S. Spina, scritto e parlato, costituito da 26 milioni di parole, distribuite in 10 differenti generi testuali, con testi raccolti dal 1995 al 2011. Il corpus è lemmatizzato e dal giugno 2015, il corpus (in versione beta) è interrogabile liberamente via CQPweb.

Tra i corpora specialistici troviamo: VELI (De Mauro 1989), lista di frequenza basata su un corpus di circa 25 milioni di occorrenze (suddivise in sottocorpora per testata: «Domenica del Corriere», «Europeo», «Il Mondo», ANSA), seguito dal Corpus of Italian Newspapers, con testi del 1989 (Burr 1993), corpus tuttavia non annotato o lemmatizzato, composto da sei numeri complessivi dai quotidiani La Repubblica, Corriere della sera, Il Mattino e La Stampa, per circa 600.000 occorrenze e per finire il corpus La Repubblica (Baroni et al. 2004), corpus di circa 380 milioni di occorrenze, lemmatizzato con metodologia probabilistica, annotato per generi (news report, comment) e per argomenti church, culture, economics, education, news, politics, science, society, sport, weather), disponibile e interrogabile online (<http://sslmit.unibo.it/repubblica>), contiene le annate integrali del quotidiano dal 1985 al 2000.

Un altro aspetto non più trascurabile riguarda l'acquisizione di corpus del parlato. In particolare, nella fase preliminare del nostro lavoro, a noi ha interessato il parlato dei bambini e uno dei siti più informativi e utili per lo studio dell'acquisizione del linguaggio è <http://childes.psy.cmu.edu/> che contiene un database di trascrizioni, del software per la

codifica del linguaggio infantile (CLAN) e un manuale per la trascrizione (che utilizza il sistema CHAT).

Childes (Child Language Data Exchange System) è un archivio di trascrizioni spontanee di bambini (solitamente dai 14 mesi ai quattro anni di età) che interagiscono con adulti in varie situazioni. Generalmente ogni trascrizione si riferisce ad una conversazione di durata variabile dai 20 ai 60 minuti. Tali trascrizioni sono codificate secondo il formato standardizzato CHAT che prevede una serie di convenzioni.

#### **4. La statistica testuale**

Il corpus creato come risorsa base per l'analisi linguistica che abbiamo effettuato è costituito da 21 testi completi della narrativa classica e contemporanea per bambini di età compresa tra i 7 e gli 11 anni, frequentanti il ciclo della scuola primaria di primo grado, composto da circa 998.841 tokens. I libri per bambini hanno delle grandi varietà e delle tipologie che non hanno equivalenti nella letteratura per adulti. Esistono libri descrittivi, che sono una specie di divulgazione per i bambini. Libri sugli animali, sulla natura, libri molto semplici per i più piccoli "Il mio primo atlante illustrato" Ravensburgher, libri adatti a bambini più grandi che coltivano la passione per la storia o le scienze, la più famosa è la collana Gallimard, testi divulgativi. Il linguista che più di tutti si è occupato delle scelte lessicali come risorsa della comunicazione divulgativa fondata su dati precisi è Tullio De Mauro. Nella sua introduzione al convegno sul Linguaggio della Divulgazione (1983) De Mauro ha sostenuto che l'esistenza di un vocabolario di base formato dalle parole più frequentemente usate può essere la chiave di volta per la comprensione. Pertanto il problema della comprensione dei testi, di quelli divulgativi in particolare, potrebbe essere in gran parte risolto utilizzando le parole scientifiche che risultano appartenere al vocabolario di base e utilizzando il meno possibile quelle che non vi appartengono. A sostegno di questa tesi si è messo in evidenza che c'è un utilizzo di termini tecnico-scientifici nel linguaggio comune, due esempi sono l'ambito economico 10% dei termini, e il settore della biologia 23% dei termini. Nel nostro lavoro, invece, abbiamo concentrato l'attenzione su testi narrativi per dimostrare la complessità che

può sottendere alla comprensione di un testo letterario. In particolare vogliamo evidenziare che si trovano occasioni di complessità legate alle inferenze richieste e al lessico utilizzato. I testi scelti hanno diversa lunghezza e varia complessità. Il corpus ha una complessità crescente per soggetto (argomento), lessico e sintassi. Il lessico, in particolare, è nel suo insieme semplice perché si è resa necessaria la riduzione dello sforzo di comprensione, in modo da permettere ai lettori (bambini tra i 7 e gli 11 anni) di concentrarsi sulla comprensione dei contenuti. Tuttavia, vocaboli meno comuni di certo non disturbano il processo di comprensione, anzi il significato del singolo termine può essere compreso attraverso le integrazioni dei significati del brano. Anche la sintassi non deve essere troppo elaborata perché le frasi lunghe possono rallentare la comprensione dei significati e la loro connessione. I testi digitali oggetto della nostra analisi provengono dal web e per essere utilizzati sono stati ripuliti di tutti gli elementi che potevano alterare le elaborazioni computazionali: tabelle, figure, intestazioni e numerazioni delle pagine e in generale di tutti gli elementi utilizzati per ragioni tipografiche, infatti per poter svolgere alcuni tipi di analisi è necessario effettuare alcune operazioni di pre-processamento dei testi che, in base alle esigenze dell'analisi linguistica, possono o non possono essere applicate. Tra queste operazioni ci sono:

- La normalizzazione del testo, che consiste nel sostituire le maiuscole in modo da rendere uniforme il testo, l'eliminazione della punteggiatura, la correzione di errori di battitura o degli errori grammaticali più comuni come ad esempio i casi in cui la "è" viene scritta "È" etc.
- La rimozione delle Stop Word, che consiste nella rimozione di tutti quei token semanticamente "vuoti", ad altissima frequenza.

Per comprendere il tipo di lavoro che viene svolto sui testi prima di una elaborazione linguistica dobbiamo introdurre alcuni concetti base. La prima distinzione da evidenziare è quella tra Type e Token cioè tra tipi di parole (types) e le occorrenze di parole testuali (tokens). Una sequenza di caratteri delimitata da due separatori definisce una forma grafica, all'interno di un testo le parole possono ripetersi più volte in questo caso parliamo di

occorrenze di parole testuali (word tokens). Le parole testuali sono, quindi, le forme diverse che le parole possono avere all'interno di un testo. Ogni occorrenza di una parola all'interno di un testo è detta token. Invece le forme delle parole diverse sono i tipi di parole (word types). Per comprendere meglio il significato della differenza tra type e token osserviamo un passo tratto dal Piccolo Principe:

Allora si stira, e sospinge da principio timidamente verso il sole un bellissimo ramoscello inoffensivo. Se si tratta di un ramoscello di ravanello o di rosaio, si può lasciarlo spuntare come vuole. Ma se si tratta di una pianta cattiva, bisogna strapparla subito, appena la si è riconosciuta. C'erano dei terribili semi sul pianeta del piccolo principe: erano i semi dei baobab. Il suolo ne era infestato. Ora, un baobab, se si arriva troppo tardi, non si riesce più a sbarazzarsene. Ingombra tutto il pianeta. Lo trapassa con le sue radici. E se il pianeta è troppo piccolo e i baobab troppo numerosi, lo fanno scoppiare.

Questo breve passo contiene 105 parole (tokens) e 72 tipi di parole (types), le occorrenze riportate in tabella 4 sono:

*Tabella 4 Type e tokens tratti da un passo del Piccolo Principe*

Type	Occorrenze
si	7
il	4
di	4
se	4
pianeta	3
baobab	3
troppo	3
un	3
e	3
ramoscello	2

semi	2
ne	2
tratta	2
piccolo	2
è	2
erano	2
lo	2

I primi elementi di una lista di questo tipo, calcolata su un testo contenente ancora le Stop Words (parola sintattica o semanticamente vuota), sono nella quasi totalità dei casi, articoli e preposizioni. Per poter sfruttare a pieno la ricchezza di un corpus può essere necessario aggiungere una serie di informazioni ed effettuare delle operazioni come il POS Tagging e la Lemmatizzazione. Il POS Tagging consiste nell'attribuire ad ogni token la propria Parte del Discorso. L'acronimo POS sta per Part-of-Speech o parte del discorso, normalmente, si intende la Classe Grammaticale o Categoria Lessicale alla quale appartiene una specifica parola. Se consideriamo ancora il brano precedente, il token "si" appartiene alla POS dei "pronomi" (abbreviato PRON), "il" alla categoria "determinante" (DET) e "pianeta" appartiene alla categoria dei nomi (N). Le POS, riportate in tabella 5, che vengono normalmente considerate in linguistica computazionale sono le seguenti:

*Tabella 5 Part of Speech in linguistica computazione*

Abbreviazione	POS
<b>N</b>	Nomi
<b>V</b>	Verbi
<b>A</b>	Aggettivi
<b>AVV</b>	Avverbi
<b>DET</b>	Determinanti
<b>PRON</b>	Pronomi
<b>PREP</b>	Preposizioni
<b>CONG</b>	Congiunzioni
<b>INTER</b>	Interiezioni

Ovviamente è possibile espandere questa lista in base alle esigenze dell'analisi linguistica che si va ad effettuare, ed esempio è possibile introdurre una differenziazione tra pronomi relativi (PREL) e pronomi personali (PPERS). Nel caso della lemmatizzazione invece dobbiamo individuare il lemma che rappresenta la forma canonica delle parole, cioè la forma in cui troviamo una determinata parola nel dizionario. Per l'italiano, di norma, la forma canonica per Nomi e Aggettivi è il maschile singolare (o il femminile singolare in caso di nomi femminili), e l'infinito per i verbi. Quindi il task della Lemmatizzazione consiste nel ridurre alla forma canonica tutti i token che compongono un testo, in modo da poter ridurre il numero di type non considerando tutte le differenti forme flesse di una determinata parola ("essere", "è", "fu" e "siamo" vengono considerati come un unico type "essere").

Se volessimo calcolare le occorrenze sull'intero libro, la lista, per quanto riguarda le prime posizioni del ranking, sarebbe poco dissimile, ma con valori estremamente più alti. Nell'analisi già eliminando anticipatamente le Stop Words, riusciamo ad ottenere classifiche più significative ed informative, cioè, che ci offrono maggiori informazioni riguardo il testo che stiamo analizzando, inoltre per poter comprendere i risultati dell'analisi di un testo dobbiamo considerare la lunghezza del testo. Introduciamo, quindi, il concetto di Frequenza, o meglio di Frequenza Relativa dei type nel testo. La formula per il calcolo della frequenza è la seguente:

$$TF = \frac{n_{i,j}}{d_j}$$

dove la Term Frequency è uguale al rapporto tra il numero di occorrenze della parola  $i$  nel testo  $j$  e la dimensione del testo  $j$ .

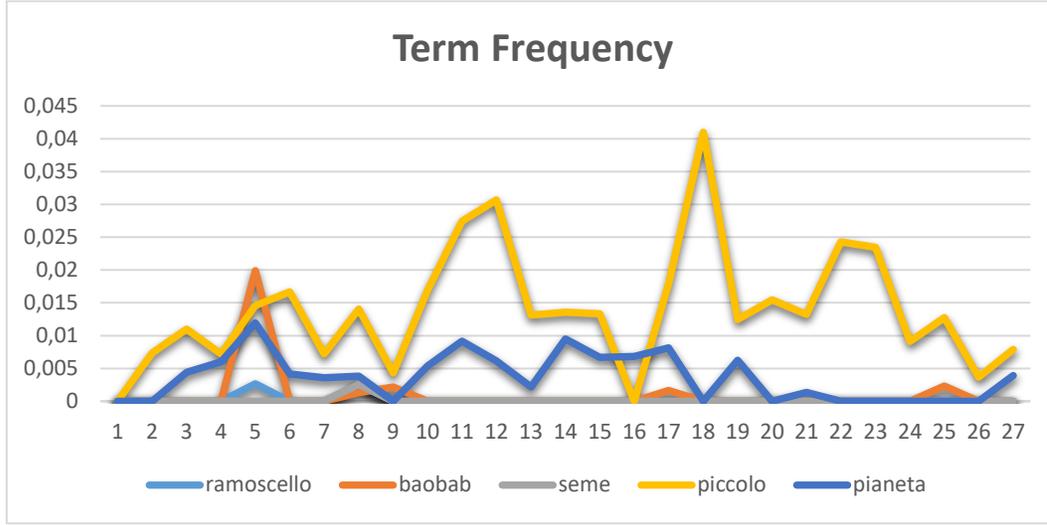


Figura 3 Term frequency  $TF = \frac{n_{i,j}}{d_j}$

La Term Frequency, tuttavia, non basta a descrivere efficacemente un testo nel caso volessimo confrontarlo con una collezione di numerosi altri testi. Se ad esempio volessimo descrivere ogni capitolo di un libro in base agli elementi di maggior “peso” per ciascun capitolo, andando dunque ad eliminare quei types che possiamo dare per scontati (come ad esempio il nome del protagonista, che descrive il libro nella sua interezza ma ci dice ben poco sul singolo capitolo) dovremo introdurre un nuovo misuratore, il Term Frequency – Inverse Document Frequency (TF-IDF), la cui formula è la seguente:

$$w_d = f_{w,d} * \log\left(\frac{|D|}{f_{w,D}}\right)$$

dove il peso di un type  $w$  per un documento  $d$  è uguale alla frequenza di  $w$  in  $d$  per il logaritmo del rapporto tra il numero di documenti in analisi e il numero  $f$  di documenti in cui  $w$  è presente.

Inoltre, è possibile rappresentare in che modo il peso di un type vari di capitolo in capitolo con una semplice rappresentazione grafica come la seguente, dove, è possibile osservare il valore di TF della parola per ogni capitolo.

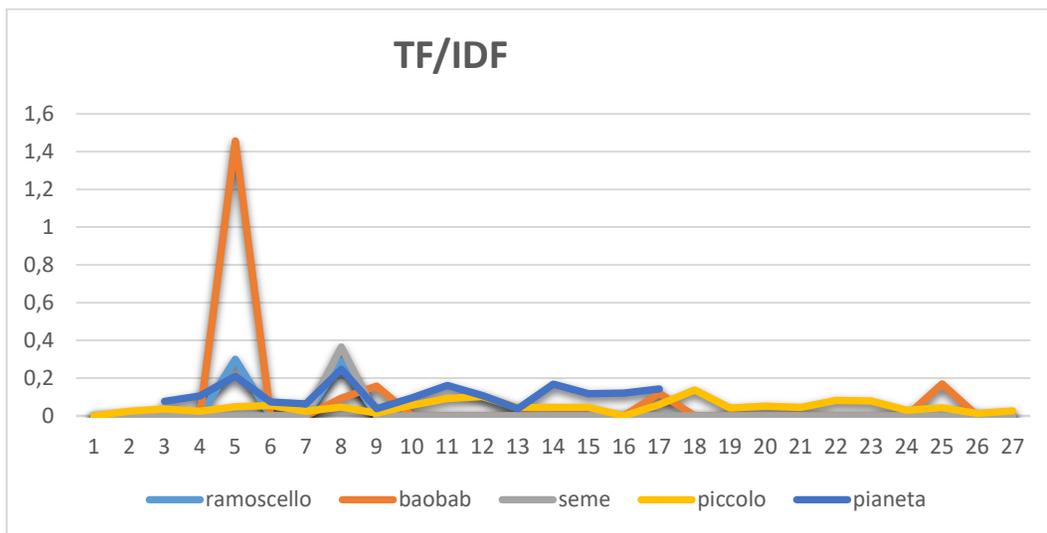


Figura 4 Term Frequency – Inverse Document Frequency (TF-IDF)

In conclusione, l'analisi statistica dei testi è un tipo di analisi particolarmente adatta a chi abbia bisogno di estrarre rapidamente informazioni poco precise dai testi. Informazioni che non ci dicono di cosa si parla nel testo, né possono rispondere a domande specifiche, ma semplicemente, nei limiti di margini di errori “statisticamente tollerabili” offrono una visione generale di un determinato testo.

Questo tipo di analisi, nonostante per alcuni aspetti rimanga superficiale, presenta il vantaggio di essere completamente indipendente dalle lingue, di essere rapida, di non presupporre competenze linguistiche avanzate e di essere a livello computazionale molto vantaggiosa.

## CAPITOLO IV

### IL LESSICO-GRAMMATICALE DELL'ITALIANO DI BASE

#### 1. Il lessico-grammaticale

In questo capitolo viene presentata l'ipotesi della creazione di un vocabolario di base dell'italiano che è basato sulla teoria lessico-grammaticale. Per la teoria lessico-grammaticale, ogni parola racchiude in sé la struttura di una frase elementare o semplice, che ne rappresenta le condizioni di uso sintattico e semantico. Le parole più facilmente rappresentative di questa posizione sono i verbi, che implicano una serie di restrizioni sulla forma, il numero e la natura dei loro complementi cosiddetti nucleari o pertinenti. I primi lavori in questa direzione sono stati realizzati da EMDA 1983, D'Agostino 1983, Elia 1984a, 1984b<sup>14</sup>. Per ogni verbo, negli anni, sono state analizzate circa 200 combinazioni sintattico-semantiche in base a una batteria di proprietà combinatorie<sup>15</sup>. Lavorando su un lessico di circa 7.000 verbi sono state individuate 63 classi in base ad almeno una proprietà che è stata indicata come definizionale della classe (EMDA1983).

Un tipo di analisi lessico grammaticale applicata al primo vocabolario di base della lingua italiana (De Mauro 1980) è stata condotta da Elia 2005<sup>16</sup>, con 553 lemmi verbali a cui sono stati aggiunti altri 134 lemmi scelti in base a valutazioni sia di disponibilità che di rappresentatività lessico-sintattica. Il risultato è stato un lessico di 1.362 usi verbali su un

---

<sup>14</sup> Per una bibliografia minima, si veda Elia, Martinelli, D'Agostino 1981, Elia 1984, D'Agostino 1992 e Vietri 2004, che rappresenta l'ultimo aggiornamento della classificazione iniziale. Nella terminologia e nel sistema di notazione del lessico-grammaticale, il lemma ha puro valore morfologico, l'uso verbale è inteso a livello sintattico e include quella che in termini lessicografici è definita accezione. Le classi sono rappresentate in tavole matriciali con righe e colonne, i simboli più usati sono: W = qualsiasi complemento, N = Gruppo nominale, Prep=varie preposizioni, Loc= preposizioni locative, cioè in, su, per, verso, tra, contro, V = uso verbale, Vsup = verbo supporto, cioè ausiliario di un nome e non di un verbo, Sim = preposizione simmetrica, cioè con, contro, E = elemento vuoto, le parentesi ( ) e il + vengono utilizzati come simboli di concatenazione polinomiale.

<sup>15</sup> Si veda ELIA, Annibale (2014). On Lexical, Semantic and Syntactic Granularity of Italian Verbs. In: (a cura di): Fryni Kakoyanni-Doa, *Penser le lexique-grammaire: perspectives actuelles*. COLLOQUES, CONGRÈS ET CONFÉRENCES SCIENCE DU LANGAGE, p. 277-288, PARIS:Honoré Champion.

<sup>16</sup> ELIA A. (2005). Lessico Grammaticale dell'italiano. In: T. DE MAURO; I. CHIARI; A CURA DI;. *Parole e numeri. Analisi quantitative dei fatti di lingua*. p. 157-170, Roma:Aracne editrice

totale di 687 lemmi, il che significa che la media di sdoppiamento delle strutture-accezioni è circa 2. La lista del vocabolario di base è una lista di frequenza in cui è stata considerata l'accezione più "semplice" del lemma. Dato un qualsiasi testo (orale o scritto), però, è ancora pressoché impossibile determinare, automaticamente, dato un lemma, quale delle sue diverse accezioni occorra in quel determinato testo. Il calcolo della frequenza e la valutazione del livello di comprensibilità sono ancora condizionati dalla mancanza di questa specifica definizione. Per quanto riguarda i verbi, per esempio, le accezioni diverse possono comportare livelli di difficoltà e di comprensibilità notevoli. Prendiamo il lemma *accomodare*, appartenente al vocabolario fondamentale, e osserviamo le diverse strutture sintattiche (con diverse accezioni) che nell'ambito della teoria lessico-grammaticale possiamo assegnargli:

- (1) Leo accomoda l'orologio *classe 20*
- (2) Mi accomoda che tu venga più tardi *classe 41*
- (3) Mi accomodai con Bice sul prezzo *classe 52*
- (4) Leo si è accomodato a fare il bidello *classe 56*
- (5) Bice si accomoda in poltrona *classe 7*
- (6) Leo si accomodò in poltrona a fumare *classe 57*

Quali di queste accezioni hanno determinato il calcolo della frequenza di *accomodare* all'interno del VdB? A livello intuitivo e in base alle valutazioni presenti nel Grande Dizionario D'uso (Grande dizionario italiano dell'uso, diretto da T. De Mauro, UTET, Torino 1999-2007), possiamo dire che l'accezione (1) e la (5) sono quelle più di base, ma fino a che punto il rapporto tra tutti i diversi usi di *accomodare* ci permette di parlare ancora di una stessa entità morfo-fonologica e fino a che punto non ci obbliga a considerare l'esistenza di diverse entità distinte? Si tratta di questioni molto complesse che non verranno prese in considerazione nel nostro lavoro di ricerca. In ogni caso, noi abbiamo escluso strutture-accezioni di livello troppo tecnico che invece sono rilevanti per un'analisi linguistica di dominio specialistico.

Un altro problema riguarda le forme "polirematiche" o "composte" o "idiomatiche", almeno in parte affrontato a livello lessicale, per quello che concerne le polirematiche, nella

redazione del NVdB (Chiari; De Mauro 2016) e di seguito affrontato mediante una classificazione parziale in ambito lessico-grammaticale.

Il verbo *vedere*, per esempio, nella frase

(7) *Eva vede rosso, quando si parla di Leo*

non è in combinazione libera con *rosso*, dal momento che il significato idiomatico di *vedere rosso* non ha niente a che vedere con il "colore rosso", ma ha relazione con "indispettirsi, essere contrariato". La frequenza d'uso non tiene ancora conto delle combinazioni idiomatiche in cui i singoli elementi non hanno lo stesso valore che hanno nelle combinazioni libere.

Nel prossimo paragrafo presentiamo la "grammatica" del lessico- grammatica, illustrando le diverse classi sintattiche secondo le proprietà che le caratterizzano e attraverso campione di frasi in cui è presente un verbo appartenente ad una determinata classe. Nel paragrafo successivo presentiamo il lessico dei verbi di base dell'italiano con l'attribuzione delle classi lessico-grammaticali.

## 2. La Grammatica

Per presentare le caratteristiche grammaticali della classificazione delle strutture di frase semplice dei verbi facciamo osservare che le classi da 1 a 10 includono le strutture di frase dei verbi cosiddetti "intransitivi" e/o preposizionali, le classi da 16 a 33 includono le strutture dei verbi cosiddetti "transitivi", le classi da 41 a 58 includono le strutture dei verbi sia "intransitivi" che "transitivi" che reggono una completiva in posizione soggetto, oggetto o secondo complemento.

Le abbreviazioni e le notazioni trasformazionali sono quelle adottate nella tradizione lessico-grammaticale.

### **CLASSE 0: N0 Vsup W (36 usi verbali)**

Gli usi verbali di questa classe sono i cosiddetti verbi supporto, cioè dei verbi a caratteristica lessicale quasi nulla, il cui scopo principale è fornire tempo, modo, aspetto a sostantivi, aggettivi e avverbi, che possono essere in rapporto morfologico con verbi a caratteristica lessicale piena. I verbi supporto sono avere, fare, essere, ma esistono diverse varianti o estensioni semantico - aspettuuali, come diventare (per essere) o effettuare (per fare). Si vedano gli esempi seguenti:

*Céline ha paura di essere sgridata*

*Leo fa una passeggiata*

*Ale è in attesa del treno*

**CLASSE 1: V****(4 usi verbali)**

Gli usi verbali di questa classe entrano nella struttura V, la più elementare dell'intero sistema della frase semplice dell'Italiano poiché caratterizzata dall'assenza di gruppi nominali in posizione N0 ed N1. Si vedano gli esempi seguenti:

*albeggia*

*annotta*

*nevica*

*piove*

**CLASSE 2: N0um V****(43 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V, in cui N0 è obbligatoriamente un N umano (a volte anche animale). Si vedano gli esempi seguenti:

*Bice sonnecchia*

*Leo tossisce*

*Leo starnutisce*

**CLASSE 3: N0-um V (38 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V, in cui N0 è obbligatoriamente un N non-umano. Si vedano gli esempi seguenti:

*l'auto cappottò*

*il pesco fiorisce*

*il suo televisore funziona*

**CLASSE 4: N0 V a N1 (23 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura in cui N1 è obbligatoriamente introdotto dalla preposizione "a"; N0 ed N1 possono essere tanto umani che non-umani. Si vedano gli esempi seguenti:

*Ale ubbidisce a Bice*

*Bice si rivolge al marito*

**CLASSE 5: N0(i) V Loc N(j) <=> N0(j) V di N1(i) (5 usi verbali)**

Gli usi verbali di questa classe entrano nelle due strutture indicate, che presentano la seguente correlazione sistematica: l'N0 della prima, indicizzato (i), assume la posizione N1

introdotto dalla preposizione "di" nella seconda, mentre l' N1 della prima, indicizzato (j) e preceduto da una preposizione Locativa che introduce un complemento "scenico" (indica cioè la "scena" entro cui si realizza il processo denotato dal verbo), assume la posizione N0 nella seconda; la freccia doppiamente orientata indica l'impossibilità di individuare una frase fonte da cui l'altra deriverebbe per trasformazione. In sintesi si tratta di costruzioni con di N che si classificano soltanto in correlazione con le costruzioni con loc N. Si vedano gli esempi seguenti:

*le stelle brillano nel cielo <=>*

*<=> il cielo brilla di stelle*

*l'ira fiammeggia sul volto di Leo <=>*

*<=> il volto di Leo fiammeggia d'ira*

*gli insetti formicolano nel giardino <=>*

*<=> il giardino formicola di insetti*

**CLASSE 6: N0 V da N1 (4 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura il cui N1 è obbligatoriamente introdotto dalla preposizione "da". Si vedano gli esempi seguenti:

*gli alunni dipendono dai docenti*

*l'uomo discende dalle scimmie*

*Ale nacque da una famiglia agiata*

**CLASSE 7: N0 V(mt) Loc N1 (66 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura in cui N0 designa un attante in rapporto di "movimento" (indicato da V) rispetto all'argomento locativo (N1) introdotto da una preposizione Locativa quale "in", "su", "da", "a", ecc. Le localizzazioni possono essere di tipo "provenienza", "destinazione" o entrambe. Si vedano gli esempi seguenti:

*Ale arriva dall'Australia*

*Bice parte da Milano*

*quel viaggiatore va da Napoli a Chiaravalle*

**CLASSE 8: N0 V Loc(st) N1 (24 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura in cui N0 entra in un rapporto di localizzazione "statica" (indicata da V) rispetto all'argomento locativo (N1) introdotto da una preposizione locativa quale "in", "su", "da", "a", ecc. Si vedano gli esempi seguenti:

*Bice abita in via Ruggi*

*la penna è sul tavolo*

*Ale resta in casa*

**CLASSE 9: N0(i) V Sim N1(j) <=> N0(j) V Sim N1(i) <=> Ni,j V (7 usi verbali)**

Gli usi verbali di questa classe entrano nelle strutture indicate, correlate sistematicamente: N(i) ed N(j) scambiano reciprocamente il ruolo sintattico di N0 ed N1 e possono altresì essere coordinati entrambi in posizione N0. Il simbolo Sim indica generalmente la preposizione "con", meno di frequente "da", "contro". Si vedano gli esempi seguenti:

*Bice combatte contro Ale <=>*

*Ale combatte contro Bice <=>*

*Bice e Ale combattono*

*il salotto comunica con la cucina <=>*

*la cucina comunica con il salotto <=>*

*il salotto e la cucina comunicano*

*Ale divorzia da Bice <=>*

*Bice divorzia da Ale <=>*

*Ale e Bice divorziano*

**CLASSE 10: N0 V prep N1 (E + prep N2)**

**(38 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura che ha una certa varietà nelle preposizioni che introducono N1 e ha la presenza eventuale di un secondo complemento preposizionale. Si vedano gli esempi seguenti:

*l'alunno conferisce alla commissione sul programma*

*Ale conta sugli amici*

*il cacciatore tira sulla preda*

**CLASSE 16: N0 V N1pl obbl**

**(2 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1, in cui N1 è obbligatoriamente plurale o realizzato da nomi indicanti "massa", "specie" o "classe". Si vedano gli esempi seguenti:

*Leo colleziona francobolli*

*Leo accumula vecchie monete*

*Leo colleziona il "National geographic"*

**CLASSE 16A: N0I V N1**

**(5 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1, in cui N0 ha il ruolo semantico di "luogo" di destinazione, di provenienza o di passaggio sia negli usi concreti che in quelli figurati. Si vedano gli esempi seguenti:

*questi pannelli immagazzinano l'energia solare*

*il vulcano eruttava lava e lapilli*

*Leo riunisce in sè molti pregi*

**CLASSE 17: N0 V N1 (E + Prep N2pl obbl)****(5 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1 che prevede l'eventualità di un secondo complemento Prep N2 anche facoltativo caratterizzato dalla restrizione N2=N plurale obbligatorio. Si vedano gli esempi seguenti:

*Leo taglia il pane (E + a fette sottili)*

*Leo divide il libro in capitoli*

*Leo distribuisce i suoi averi (E+ fra gli amici)*

**CLASSE 18: N0 V N1pc (a + di) N1c****(12 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1, in cui N1=N umano. Tale struttura è in correlazione sistematica con una struttura N0 V N1pc di N1c, per la quale esiste un rapporto di coreferenza ed inalienabilità tra N1c ed N1pc. Quest'ultima struttura entra in correlazione sistematica con la struttura N0 V N1c Loc N1pc. Si vedano gli esempi seguenti:

*La mamma carezza Céline*

*La mamma carezza il volto (a + di) Céline*

*La mamma carezza Céline (a + su + ecc.) il volto*

**CLASSE 19: N0 V N1pc (a + di) N1c <-> N1 essere Vpp (0 usi verbali)**

Gli usi verbali di questa classe si differenziano da quelli della classe 18 in base alla proprietà N1 essere Vpp. Tale proprietà nel caso della classe 19 ha interpretazione "puntuale", mentre nel caso della classe 18 ha interpretazione "durativa", come negli esempi:

*il vetriolo ha sfigurato Bice*

*il vetriolo ha sfigurato il volto (a + di) Bice*

*il vetriolo ha sfigurato Bice (a + su + ecc) il volto*

*il volto di Bice è sfigurato*

**CLASSE 20: N0 V N1"apparizione" (22 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura N0 V N1 in cui N1 presenta obbligatoriamente il ruolo semantico di "apparizione", cioè sta ad indicare che l'esistenza di quanto denotato dall'oggetto diretto ha inizio a partire dall'esecuzione dell'azione o del processo espresso dal verbo, come negli esempi:

*Ale ha edificato un grattacielo*

*Ale ha cucinato una scaloppina ai funghi*

*il sarto ha confezionato un abito a Céline*

**CLASSE 20I: N0 V V-n****(3 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura ad un solo complemento N0 V N1 che presenta la restrizione distribuzionale N1=V-n, nel senso che gli N che occupano la posizione oggetto diretto sono tutti identificabili da nomi classificatori che rappresentano V-n delle varie entrate di questa classe. Si vedano gli esempi seguenti:

*Céline balla (il tango + l'hip pop)*

*(il tango + l'hip pop) sono dei balli*

**CLASSE 20L: N0 V N1I****(9 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura ad un solo complemento e questo ha il ruolo semantico di "luogo". Le interpretazioni delle frasi possono essere a seconda dei casi, "statiche" o "dinamiche". Si vedano gli esempi seguenti:

*Céline attraversa (di corsa) la piazza*

*un ponte attraversa la città di Budapest*

**CLASSE 20NR: N0 V N1"non-ristretto"****(34 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura ad un solo complemento che non presenta particolari restrizioni distribuzionali sull'oggetto diretto. Si vedano gli esempi seguenti:

*Céline studia (geografia + scienze + ecc.)*

*Céline fotografa (Mattia + il mare + i fiori + ecc.)*

**CLASSE 20R: N0 V N1 "ristretto"**

**(77 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura ad un solo complemento che presenta forti restrizioni distribuzionali sull'oggetto diretto. In alcuni casi la distribuzione nominale può essere considerata quasi "unica", in altri casi la classe di N rientranti nella posizione N1 è identificabile da un nome classificatore. Si vedano gli esempi seguenti:

*il calzolaio ha risuolato i mocassini di Leo*

*i mocassini sono calzature con suola*

*Céline ha abbandonato (la pallavolo + l'atletica+ ecc.)*

*(la pallavolo + l'atletica + ecc.) sono attività sportive*

**CLASSE 20SP: N0 V N1  $\Leftrightarrow$  \* N1 essere Vpp da N0 (6 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura ad un solo complemento che, pur rispondendo positivamente agli altri test della transitività, non accettano il passivo, come negli esempi seguenti:

*il pacco pesa 10 Kg*

*quel vestito costa due milioni*

**CLASSE 20ST: N0 Vst N1**

**(15 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura ad un solo complemento e le interpretazioni delle frasi in cui essi occorrono, negli usi letterali come in quelli figurati, sono di tipo "statico", come negli esempi seguenti:

*dei capelli castani incorniciano il volto di Céline*

*il castello domina la vallata*

**CLASSE 20UM: N0 V N1um**

**(85 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura ad un solo complemento in cui N1=Numano obbligatorio, come negli esempi seguenti:

*Ale e Bice hanno adottato una bambina*

*un tribunale popolare ha processato il dittatore*

**CLASSE 21: N0 V N1 <=> N0 Vsup (E + Prep) V-n W (0 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura ad un solo complemento in correlazione sistematica con una struttura di nominalizzazione prBicelentemente del tipo N0 Vsup Prep V-n N1 oppure del tipo N0 Vsup V-n di N1, come negli esempi seguenti:

*gli Spagnoli hanno colonizzato le Americhe*

*gli Spagnoli hanno trasformato in colonia le Americhe*

*gli Spagnoli hanno fatto una colonia delle Americhe*

**CLASSE 21A: N0 V N1 <=> N0 Vsup V-a N1 (7 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura ad un solo complemento in correlazione sistematica con una struttura di aggettivazione del tipo N0 Vsup (E + Adv quantità) V-a N1, come negli esempi seguenti:

*il fumo ingiallisce le dita delle mani*

*il fumo rende gialle le dita delle mani*

*Ale ha ingrandito la sua casa*

*Ale ha reso più grande la sua casa*

**CLASSE 22: N0 V N1 N2****(9 usi verbali)**

Gli usi verbali di tale classe rientrano in una struttura nella quale il secondo gruppo nominale a destra del verbo, N2, è interpretabile in genere come attributo o apposizione dell'oggetto diretto, N1, come nelle frasi seguenti:

*l'assemblea ha eletto Ale rappresentante*

*Ale ha chiamato la figlia Céline*

**CLASSE 22A: N0 V N1<sub>um</sub> (da + come) N2****(3 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura a due complementi nella quale N1=N umano obbligatorio ed N2 è interpretabile in genere come attributo o apposizione dell'oggetto diretto, realizzandosi con N indicanti nomi di attività o professione, come negli esempi seguenti:

*Ale ha assunto Bice (da + come) baby sitter*

*Ale stipendia Leo (da + come) architetto*

**CLASSE 23D: N0 V N1 a N2<sub>um</sub>****(27 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura a due complementi nella quale N0 è generalmente Numano e Prep N2=a Numano obbligatorio ed è pronominalizzabile con

Ppv=gli. Le frasi che realizzano tale struttura presuppongono un "passaggio" o uno "scambio" dell'N1 tra il soggetto ed il secondo complemento, come negli esempi seguenti:

*Giulia dette un libro a Céline*

*Giulia le dette un libro*

*Ale rubò il portafoglio a Leo*

*Ale gli rubò il portafoglio*

**CLASSE 23R: N0 V N1 da N2um**

**(7 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura a due complementi nella quale N0 è generalmente Numano e Prep N2=da N2umano obbligatorio ed è pronominalizzabile con Ppv=ne. Le frasi che realizzano tale struttura presuppongono un "passaggio" o uno "scambio" dell'N1 tra il soggetto ed il secondo complemento. Si vedano gli esempi seguenti:

*Leo riceve un libro da Bice*

*Leo ne riceve un libro*

*Leo eredita un appartamento dal padre*

*Leo ne eredita un appartamento*

**CLASSE 23ND: N0 V N1 a N2um**

**(8 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura a due complementi nella quale N0 è generalmente umano e Prep N2=a N2umano obbligatorio ed è pronominalizzabile con Ppv=gli. Le frasi che realizzano tale struttura non presuppongono la nozione di "passaggio" o di "scambio" dell'N1 tra il soggetto ed il secondo complemento. Si vedano gli esempi seguenti:

*gli esperti attribuiscono il quadro a Van Gog*

*gli esperti gli attribuiscono il quadro*

*la società commissiona dei lavori a Leo*

*la società gli commissiona dei lavori*

**CLASSE 24: N0 V N1 (di + con + da) N2 -**

**(19 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura a due complementi per la quale N0 è interpretabile come soggetto causativo di movimento, il primo complemento, N1, ha il ruolo semantico di "luogo" ed il secondo complemento, Prep N2, quello di "argomento del luogo". Tale secondo complemento è generalmente pronominalizzabile con Ppv=ne. La variazione delle preposizioni (di + con) e (da) è legata all' interpretazione semantica dell'N1: le prime due se N1 è interpretato come "luogo di destinazione", l'ultima se esso riceve l'interpretazione di "luogo di provenienza". Si vedano gli esempi seguenti:

*Leo riempie l'armadio (di + con) vestiti*

*Leo ne riempie l'armadio*

*Leo pulisce i mobili dalla polvere*

*Leo ne pulisce i mobili*

**CLASSE 27: N0 V N1 da N2**

**(9 usi verbali)**

Gli usi verbali di tale classe, sia concreti che figurati, rientrano in una struttura a due complementi nella quale il secondo complemento ha la forma Prep N2=(da+contro) N2 ed è anche pronominalizzabile con Ppv=ne. Si vedano gli esempi seguenti:

*Leo difende Bice (da+ contro) i pericoli*

*la tettoia ripara dalla pioggia*

*i patrioti hanno liberato la città dagli invasori*

**CLASSE 28: N0 V N1 Loc N2"destinazione"**

**(76 usi verbali)**

Gli usi verbali di tale classe rientrano in una struttura a due complementi, nella quale il soggetto N0 è interpretabile come soggetto causativo di uno spostamento quanto indicato dall'oggetto diretto, N1, rispetto ad un luogo espresso dal complemento Prep N2=Loc N2. Tale complemento assume il ruolo semantico di "luogo di destinazione" ed ha Ppv=(ci+vi). Le preposizioni che introducono il secondo complemento sono varie, in particolare Prep=(in+dentro+su+contro+verso). Si vedano gli esempi seguenti:

*Leo mette il libro nel cassetto*

*Leo (ci+vi) mette il libro*

*Leo inserisce la spina nella presa*

*Leo (ci+vi) inserisce la spina*

**CLASSE 28SC: N0 V N1 Loc N2**

**(12 usi verbali)**

Gli usi verbali di tale classe rientrano in una struttura a due complementi dei quali il secondo assume la forma Prep N2=Loc N2 ed ha il ruolo semantico di "scenario" o di "luogo pertinente". Generalmente tale complemento è pronominalizzabile con Ppv=(ci+vi). Si vedano gli esempi seguenti:

*il regista ambienta il film in un quartiere malfamato*

*il regista vi ambienta il film*

*Pat ha perduto la fede nella sabbia*

*Pat vi ha perduto la fede*

**CLASSE 28ST: N0 Vst N1 Loc N2**

**(4 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura a due complementi dei quali il secondo assume la forma Prep N2=Loc N2. Le frasi costruite su tale struttura prevedono che l'interpretazione del verbo sia di tipo "statico" e che il ruolo semantico del secondo complemento sia quello di "luogo pertinente". Generalmente il secondo complemento è pronominalizzabile con Ppv=(ci+vi). La preposizione locativa è estremamente variabile. Si vedano gli esempi seguenti:

*un palo regge la piastra contro il soffitto*

*Leo sostiene la scala contro il muro*

*la colla fissa il manifesto alla parete*

**CLASSE 29: N0 V N1 Loc N2"provenienza" (8 usi verbali)**

Gli usi verbali di tale classe rientrano in una struttura a due complementi, nella quale il soggetto N0 è interpretabile come soggetto causativo di uno spostamento quanto indicato dall'oggetto diretto, N1, rispetto ad un luogo espresso dal complemento Prep N2=Loc N2. Tale complemento assume il ruolo di "luogo di provenienza" ed ha Ppv=(ne). La preposizione che introduce il secondo complemento è da. Si vedano gli esempi seguenti:

*Leo toglie i libri dalla libreria*

*Leo ne toglie i libri*

*Leo prende il cappello dall'appendi abiti*

*Leo ne prende il cappello*

**CLASSE 30: N0 V N1 Loc N2 Loc N3 (12 usi verbali)**

Gli usi verbali rientranti in tale classe entrano in una struttura a tre complementi, dei quali due sono interpretabili come complementi locativi, mentre il soggetto, N0, è interpretabile come soggetto causativo di uno spostamento dell'oggetto diretto, N1, rispetto ad un "luogo di provenienza", Loc N2, e ad un "luogo di destinazione", Loc N3. La sequenza Loc N2 Loc N3 normalmente assume la forma da N a N. I due complementi locativi sono in genere pronominalizzabili rispettivamente da Ppv=(ci+vi) e da Ppv=ne anche se quest'ultima spesso genera frasi di dubbia accettabilità, ma le due pronominalizzazioni non possono aver luogo simultaneamente. Si vedano gli esempi seguenti:

*Leo getta il pallone dal balcone in giardino*

*Leo vi getta il pallone dal balcone*

*? Leo ne getta il pallone in giardino*

*Leo scaraventa i nemici dalla torre nel fossato*

*Leo vi scaraventa i nemici dalla torre*

*Leo ne scaraventa i nemici nel fossato*

### **CLASSE 31: N0 V N1 Sim N2**

**(14 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura a due complementi nella quale il secondo complemento ha la forma Prep N2=Sim N2, in cui Sim=(a+da+con). Tale secondo complemento è pronominalizzabile normalmente con Ppv=(ci+ne), a seconda della forma della Prep. Tale classe è contraddistinta dalla possibilità di permutazione tra il primo ed il secondo complemento, senza variazione di significato della frase, e di realizzazione di un

oggetto coordinato plurale, dato il rapporto di "simmetria" vigente tra i due complementi. Si vedano gli esempi seguenti:

*Leo accorda il colore della cintura con il colore delle scarpe*

*Leo accorda il colore delle scarpe con il colore della cintura*

*Leo accorda il colore della cintura ed il colore delle scarpe*

**CLASSE 31A: N0 V N1 Sim N2**

**(2 usi verbali)**

Gli usi verbali di tale classe entrano in una struttura a due complementi nella quale il secondo complemento ha la forma Prep N2=Sim N2, in cui Sim=con. Tale classe è contraddistinta dalla possibilità di permutazione tra il soggetto ed il secondo complemento, senza variazione di significato della frase, e di realizzazione di un soggetto coordinato plurale, dato il rapporto di "simmetria" vigente tra i due elementi della frase. Si vedano gli esempi seguenti:

*il fruttivendolo negozia il prezzo delle mele con Bice*

*Bice negozia il prezzo delle mele con il fruttivendolo*

*il fruttivendolo e Bice negoziano il prezzo delle mele*

**CLASSE 32: 24 + 28 + 29****(1 uso verbale)**

In tale classe rientrano quegli usi verbali che accettano sia le costruzioni che entrano nelle classi 28 e 29, come le costruzioni che entrano nella classe 24. Si vedano gli esempi seguenti:

*Leo carica le arance sul camion*

*Leo carica il camion di arance*

*Céline spalma la nutella sul pane*

*Céline spalma il pane di nutella*

**CLASSE 33: N0 V N1 Prep N2****(17 usi verbali)**

Gli usi verbali rientranti in tale classe a due complementi possono essere considerati marginali o non contraddistinti da regole strutturali o di selezione particolarmente significative, come nell'esempio seguente:

*Leo accompagna Bice con la chitarra*

**CLASSE 41: Ch F V N1****(58 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1, in cui N0 è generalmente una frase infinitiva, completiva diretta o introdotta dal sintagma IL FATTO CH F/DI e N1 è un Num, cioè un sostantivo "umano", come negli esempi seguenti:

*Andare al cinema diverte Céline*

*Andare al cinema mi diverte*

*Mi angoscia che tu te ne vada*

*Il fatto che tu te ne vada mi angoscia*

**CLASSE 42: Ch F V Prep N1**

**(33 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V Prep N1, in cui N0 può essere una frase infinitiva, completiva diretta o introdotta dal sintagma IL FATTO CH F/DI e N1 un Num, cioè un sostantivo "umano", come negli esempi seguenti:

*Andare al cinema piace a Céline*

*Andare al cinema mi piace*

*Mi dispiace che tu te ne vada*

*Il fatto che tu te ne vada mi dispiace*

**CLASSE 43: N0 V Ch F**

**(118 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1, in cui N0 è generalmente un Num, cioè un sostantivo "umano" e N1 è una frase infinitiva o completiva diretta, come negli esempi seguenti:

*Leo sa che tu sei bella*

*Leo sa di essere bello*

*Bice pensa che Leo partirà*

*Bice pensa di partire*

**CLASSE 43A: N0 V il fatto Ch F**

**(11 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1, in cui N0 è generalmente un Num, cioè un sostantivo umano e N1 è una frase infinitiva o completiva introdotta da IL FATTO CH F/DI, come negli esempi seguenti:

*Leo paga il fatto che è arrogante*

*Leo paga il fatto di essere arrogante*

*Bice sfrutta il fatto che è simpatica*

*Bice sfrutta il fatto di essere simpatica*

**CLASSE 44: N0 a che Fcong**

**(20 usi verbali)**

Gli usi verbali di questa classe hanno una struttura preposizionale N0 V a N1, in cui N0 è generalmente un Num, cioè un sostantivo "umano" e N1 è una completiva (letteraria) o infinitiva preposizionale in A, come negli esempi seguenti:

*Céline tiene a che si vada al cinema*

*Céline tiene a andare al cinema*

*il vecchio professore consente a che gli si dia del tu*

*il vecchio professore consente a farsi dare del tu*

#### **CLASSE 44A: N0 V al fatto Ch F**

**(6 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V a N1, in cui N0 è generalmente un Num, cioè un sostantivo "umano" e N1 una completiva o infinitiva preposizionale introdotta solo dal sintagma AL FATTO CH F/DI, come negli esempi seguenti:

*Leo crede al fatto che tu sei malata*

*Leo crede al fatto di dover rispettare le regole*

*Bice partecipa al fatto che si deve farti un regalo*

*Bice partecipa al fatto di farti un regalo*

**CLASSE 45: N0 V (\*di) Ch F****(10 usi verbali)**

Gli usi verbali di questa classe hanno una struttura N0 V di N1, in cui N0 è generalmente un Num, cioè un sostantivo "umano" e N1 è una completiva o infinitiva preposizionale in DI; poiché in italiano la sequenza N0 V di Ch F/V-inf non è accettabile, la sequenza è N0 V Ch F /di V-inf o N0 V del fatto Ch F/di V-inf, come negli esempi seguenti:

*Leo dubita che tu parta*

*Leo dubita del fatto che tu parta*

*Leo dubita di partire*

*Leo dubita del fatto di partire*

**CLASSE 45A: N0 V del fatto Ch F****(6 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V di N1, in cui N0 è generalmente un Num, cioè un sostantivo "umano" e N1 una completiva preposizionale in DI realizzata solo dal sintagma DEL FATTO CH F, oppure è un'infinitiva in DI, come negli esempi seguenti:

*Céline si giova del fatto che è una musicista esperta*

*Céline si giova del fatto di essere una musicista esperta*

*Céline si giova di essere una musicista esperta*

**CLASSE 46: N0 V dal fatto Ch F****(10 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V da N1, in cui N0 è generalmente un N-um, cioè un sostantivo "non umano" astratto e N1 è una completiva preposizionale in DA realizzata solo dal sintagma DAL FATTO CH F, come negli esempi seguenti:

*La confusione derivò dal fatto che tu eri partito*

*I pasticci cominciarono dal fatto che tu eri partito*

**CLASSE 47: N0 V ch F a N2****(76 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1 a N2, in cui N0 è generalmente un Num, cioè un sostantivo "umano", N1 è una completiva o infinitiva diretta in Ch F/di V-inf/V-inf e N2 è un complemento generalmente Num, cioè un sostantivo "umano", come negli esempi seguenti:

*Bice disse a Leo che avBice sonno*

*Bice disse a Leo di avere sonno*

*Leo nascose a Bice che era partito*

*Leo nascose a Bice di essere partito*

**CLASSE 47A: N0 V il fatto Ch F a N2****(1 uso verbale)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1 a N2, in cui N0 è generalmente un Num, cioè un sostantivo "umano", N1 una completiva o infinitiva introdotta solo dal sintagma IL FATTO CH F/DI V-INF e N2 è un complemento generalmente Num, cioè un sostantivo "umano", come negli esempi seguenti:

*Leo addebita a Bice il fatto che non l'hanno chiamato*

*Leo addebita a Bice il fatto di non essere stato chiamato*

#### **CLASSE 48: N0 V Ch F Prep N2**

**(24 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1 Prep N2, in cui N0 è generalmente un Num, cioè un sostantivo "umano", ma può anche essere un N-um astratto, N1 è una completiva o un' infinitiva introdotta direttamente da Ch F/ di V-inf o dal sintagma IL FATTO CH F /DI V-INF e Prep N2 è un complemento preposizionale variabile in PER, IN, e altre preposizioni con N2 generalmente Num, cioè un sostantivo "umano", ma anche N-um, cioè un sostantivo concreto o astratto, come negli esempi seguenti:

*Giulia apprese dal/sul giornale che Céline era stata promossa*

*Leo apprese da Giulia che Céline era stata promossa*

*Leo apprese dal/sul giornale di essere stato promosso*

*Leo apprese dal professore di essere stato promosso*

*Céline combinò con Giulia che si andasse al cinema insieme*

*Céline combinò con Giulia di andare al cinema insieme*

**CLASSE 48A: N0 V il fatto Ch F Prep N2**

**(4 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1 Prep N2, in cui N0 è sia un Num che un N-um, cioè sia un sostantivo "umano" che "astratto", N1 è una completiva o un'infinitiva introdotta solo da un sintagma del tipo IL FATTO Ch F/DI V-INF e Prep N2 è un complemento preposizionale variabile in PER, IN, e altre preposizioni con N2 sia Num che N-um, cioè "umano", "astratto" o "concreto", come negli esempi seguenti:

*Bice collocò nel programma il fatto che Leo avrebbe cantato*

*La tua venuta complica per me il fatto di andarmene*

*Leo ha bloccato per Bice la possibilità di partire*

**CLASSE 49: N0 V N1 a che Fcong**

**(32 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1 a N2, in cui N0 e N1 sono generalmente Num, cioè sostantivi "umani", e N2 è una completiva preposizionale introdotta dalla sequenza A CHE Fcong, cioè frase al congiuntivo, nello stile letterario, oppure è un'infinitiva in A; a partire da N1, in generale, si costruiscono le strutture a verbo riflessivo, come negli esempi seguenti:

*la mamma abitua Céline a stare da sola*

*Céline si abitua a stare sola*

*Leo aiuta Bice a che riesca bene nello studio*

**CLASSE 50: N0 V N1 (\*di) Ch F**

**(15 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1 di N2, in cui N0 è generalmente un Num, cioè un sostantivo "umano", N1 è sia un Num che un N-um, cioè un sostantivo "umano", "astratto" o "concreto" e N2 è una completiva o infinitiva preposizionale in DI realizzata attraverso la sequenza No V N1 Ch F/di V-inf o grazie al sintagma DEL FATTO CH F/DI; a partire da N1 si costruiscono le strutture con il verbo riflessivo, come negli esempi seguenti:

*Bice assicura Leo che lo ama*

*Bice assicura Leo del fatto che lo ama*

*Bice assicura Leo di amarlo*

*Bice si assicura di amare Leo*

**CLASSE 51: N0 V N1 dal fatto Ch F****(6 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1 da N2, in cui N0 è sia un Num che un N-um, cioè un sostantivo "umano", "concreto", "astratto" o una frase completiva, N1 è generalmente un Num, cioè un sostantivo "umano" e N2 è una completiva o infinitiva preposizionale introdotta dal sintagma DAL FATTO CH F/DI; a partire da N1 si costruiscono le strutture con il verbo riflessivo, come negli esempi seguenti:

*Leo difese Bice dal fatto che la volevano bocciare*

*Leo difese Bice dal fatto di essere bocciata*

*Bice si difese dal fatto che la volevano bocciare*

**CLASSE 52: N0 V a che Fcong Prep N2****(4 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V a N1 Prep N2, in cui N0 è sia Num che N-um, cioè un sostantivo "umano", "astratto", "concreto" o una frase completiva, N1 è una completiva o un'infinitiva preposizionale in A e N2 è sia Num che N-um, cioè un sostantivo "umano", "astratto", "concreto" o una frase completiva, come negli esempi seguenti:

*Leo gioca a fare il dottore con Bice*

*Andare a teatro corrisponde a andare a dormire per Bice*

**CLASSE 53: N0 V (\*di) Ch F Prep N2****(13 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V di N1 Prep N2, in cui N0 è generalmente Num, cioè un sostantivo "umano", N1 è una completiva o infinitiva preposizionale in DI, realizzata da una sequenza N0 V Ch F/di V-inf o dal sintagma DEL FATTO CH F/DI V-INF e il complemento preposizionale N2 è generalmente un Num, cioè un sostantivo "umano", come negli esempi seguenti:

*Leo parla con Bice del fatto che piove*

*Céline parla a Giulia di andare al Giffoni Film Festival*

*Bice chiacchiera con Giulia del fatto che Leo parte spesso*

*Bice chiacchiera con Leo del fatto di andare a teatro*

**CLASSE 54: N0 V Ch F Prep il fatto Ch F****(15 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1 Prep N2, in cui N0 è generalmente Num, cioè un sostantivo "umano", N1 è una completiva o un'infinitiva diretta, N2 è un complemento preposizionale realizzato generalmente da una completiva o infinitiva introdotta dal sintagma PREP IL FATTO CH F/DI, come negli esempi seguenti:

*Leo si accorge che Bice è venuta dal fatto che la stanza profuma*

*Bice capisce che Leo è partito dal fatto che le finestre sono chiuse*

**CLASSE 54A: N0 V il fatto Ch F Prep il fatto Ch F (0 usi verbali)**

Gli usi verbali di questa classe hanno una struttura N0 V N1 Prep N2, in cui N0 è generalmente Num, cioè un sostantivo "umano", N1 è una completiva o un'infinitiva introdotta dal sintagma IL FATTO CH F/DI, N2 è un complemento preposizionale realizzato generalmente da una completiva o infinitiva introdotta dal sintagma PREP IL FATTO CH F/DI, come negli esempi seguenti:

*Leo riconduce il fatto che la stanza profuma al fatto che tu sei ritornata*

*Bice paragona il fatto che tu sei in casa con il fatto di essere sola*

**CLASSE 55: Ch F V N1 Prep N2 (6 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1 Prep N2, in cui N0 può essere Num e N-um, cioè un sostantivo "umano", "astratto", "concreto" o una frase completiva, N1 è generalmente N-um, cioè un sostantivo "astratto", ma anche Num, cioè un sostantivo "umano" o una frase completiva o infinitiva introdotta dal sintagma IL FATTO CH F/DI e N2 è generalmente un Num, cioè un sostantivo "umano", come negli esempi seguenti:

*Il fatto di essere stato in Francia costituì una importante esperienza per Leo*

*Il fatto di vederla felice destò molta gioia in Leo*

**CLASSE 56: N0 V (E + Prep) Vinf W (40 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura in cui reggono un'infinitiva diretta o preposizionale che non corrisponde nè a una frase completiva nè a un sintagma nominale; sono verbi definiti tradizionalmente "modali" o "servili" e "aspettuali", come negli esempi seguenti:

*Céline sa suonare*

*Céline deve suonare*

*Céline comincia a suonare*

**CLASSE 57: N0 V a Vinf W**

**(50 usi verbali)**

Gli usi verbali di questa classe hanno una struttura N0 V Loc N1, in cui N0 è generalmente Num, cioè un sostantivo "umano" e al complemento locativo Loc N1 corrisponde un'infinitiva in A che risponde alla domanda DOVE V N0? Sono verbi tradizionalmente definiti "di movimento" o "locativi", come negli esempi seguenti:

*Céline va a suonare*

*Bice corre a comprare il pane*

**CLASSE 58: N0 V N1 a Vinf W**

**(6 usi verbali)**

Gli usi verbali di questa classe entrano in una struttura N0 V N1 Loc N2, in cui N0 e N1 sono generalmente Num, cioè sostantivi "umani" e al complemento Loc N2 corrisponde un'infinitiva in A che risponde alla domanda DOVE N0 V N1? Sono verbi in genere definiti "causativi di movimento", come negli esempi seguenti:

*Leo accompagna Bice a comprare i biglietti*

*Bice avvia Leo a comprare i biglietti*

### 3. Il Lessico

Il lessico dei verbi di base dell'italiano (1.362 usi verbali), che viene presentato di seguito, è stato classificato secondo le strutture illustrate nel paragrafo precedente. A destra del numero di classe, abbiamo inserito una frase di esempio che riflette la struttura classificatoria o definizionale. Nella frase di esempio il soggetto è espresso solo se necessario per la caratterizzazione interna alla classe definizionale. L'accezione più probabilmente considerata nel VDB è scritta in grassetto, ma si tratta comunque di una nostra ipotesi personale.

<b>abbandonare</b>	20R	abbandona la partita
abbandonare	20UM	abbandona la moglie
abbandonare	28	abbandona il pacco sulla sedia
abbandonarsi	56	si abbandona a ridere
<b>abbracciare</b>	18	abbraccia la moglie
<b>abitare</b>	8	abita a Napoli
<b>accadere</b>	42	mi accade di sognare

<b>accendere</b>	0	accende il fuoco
accendere	41	un argomento così accende tutti
<b>accettare</b>	22	accetta Bice come amica
accettare	27	accetta il regalo da Bice
accettare	43	accetta (che lo trattassero così, la nomina)
<b>accogliere</b> avv	43A	accoglie bene la notizia
accogliere	20UM	accoglie gli ospiti
<b>accomodare</b>	20	accomoda l'orologio
accomodare	41	mi accomoda che tu parta ora
accomodarsi	52	si accomoda con Bice sul prezzo
accomodarsi	56	si accomoda a fare il bidello
accomodarsi	57	si accomoda. sul divano a leggere
accomodarsi	7	si accomoda sul divano
<b>accompagnare</b>	28	accompagna Bice a Napoli a vedere il film
accompagnare	31	accompagna il pesce con il vino
accompagnare	33	accompagna Bice con il flauto
accompagnare	58	a accompagna Bice a Napoli a vedere il film
<b>accorgersi</b>	54	si accorge che aveva sonno da ciò
<b>accusare</b>	20R	accusa un malessere
accusare	50	accusa Bice di aver rubato, furto
<b>acquistare</b>	10	acquista in eleganza
acquistare	23R	acquista un diamante da Leo
<b>addormentare</b>	20UM	addormenta il bimbo
addormentarsi	2	si addormenta

<b>adoperare</b>	20NR	adopera il martello
<b>adorare</b>	20UM	adora la figlia
adorare	43	adoro (che mi guardino, lo sport, Céline)
<b>affacciare</b>	28	affaccia la carabina al balcone
affacciare	8	la finestra affaccia in via Ruggi
affacciarsi	57	si affaccia a salutare
<b>affermare</b>	43	afferma (che Leo era vivo, la verità)
affermare	47	afferma a tutti che era innocente
affermarsi	2	lo scrittore si è affermato
<b>affidare</b>	23ND	affida il pacco a Bice
<b>affrontare</b>	20UM	affronta il nemico
<b>aggiungere</b>	28	aggiunge un mattone alla pila
aggiungere	43	aggiunge (che era stanco, una frase)
<b>aggiustare</b>	20	aggiusta l'orologio
aggiustare	20R	aggiusta (l'antenna, le lancette)
aggiustarsi	20R	aggiusta (la gonna, le calze)
<b>agire</b> avvm	2	agisce male
agire	10	il veleno agisce sull'organismo
<b>agitare</b>	20R	agita il liquido
agitare	41	vederti così mi agita
<b>allargare</b>	21A	allarga il tunnel
<b>allevare</b>	20R	alleva polli
<b>allontanare</b>	28	allontana la scatola dal muro
<b>allungare</b>	21A	allunga il tubo

allungare	28	allunga le gambe sul divano
allungarsi	57	si allunga sul divano a leggere
allungarsi	7	si allunga sul divano
<b>alzare</b>	20NR	alza un peso
alzare	28	alza una mansarda sul tetto
alzarsi	57	si alza dal letto a guardare il figlio
alzarsi	7	si alza dal letto
<b>amare</b>	20UM	ama Céline
amare	43	ama (parlare, lo sport, Céline)
<b>ammalarsi</b>	2	Leo si ammala
<b>ammazzare</b>	20UM	ammazza Leo
<b>ammettere</b>	20UM	ammette Bice nel club
ammettere	43	ammette (che aveva sbagliato, gli errori)
<b>andare</b>	4	la gonna va a Bice
andare	42	mi va che tu mi carezzi
andare	57	va a Roma a comprare il pane
andare	7	va a Roma
andare	8	i libri vanno sullo scaffale
andare	8	la strada va da Roma a Milano
<b>annoiare</b>	41	leggere mi annoia
<b>annunciare</b>	20UM	annuncia Leo al professore
annunciare	47	ti annuncio che partivo
<b>apparire</b>	42	mi apparì che sbagliavo
apparire	7	appare in piazza

<b>appartenere</b>	4	il libro appartiene a Céline
appartenere	42	appartiene a me giudicare
<b>applicare</b>	20R	applica il regolamento
applicare	28	applica il marchio sul pacco
applicare	23	applica Leo allo studio del latino
applicarsi	56	mi applico a studiare
<b>appoggiare</b>	28	appoggio il libro sul tavolo
<b>approfittare</b>	45	approfitto che tu non ci sei
<b>approvare</b>	43	approvo (che tu parli, Bice, il tuo gesto)
<b>aprire</b>	20	apro un varco nella parete
aprire	20NR	apro la scatola
aprire	41	le esperienze apre la mente
<b>armare</b>	24	il re arma la fortezza di cannoni
<b>arrabbiarsi</b>	2	Leo si arrabbia
<b>arrestare</b>	20ST	il muro arresta la corsa
arrestare	20UM	arresta il bandito
arrestarsi	57	si arresta in piazza a salutare
arrestarsi	7	si arresta in piazza
<b>arrivare</b>	2	quello scrittore è a.
arrivare	56	arriva a tradire
arrivare	57	arriva a Parigi a incontrare i diplomatici
arrivare	7	arriva a Parigi
<b>asciugare</b>	24	asciuga il pavimento dall'acqua
asciugare	3	il fazzoletto asciuga (presto)

<b>ascoltare</b>	43	ascolta (i ragazzi parlare, Bice)
<b>aspettare</b>	20R	Bice aspetta (un figlio)
aspettare	43	aspetta (che arrivi, Bice)
<b>assicurare</b>	27	assicura la vita (contro gli infortuni)
assicurare	33	assicura la porta con una catena
assicurare	47	assicura all'amico che sarebbe stato assolto
assicurare	50	assicura l'amico che era innocente
<b>assistere</b>	20UM	l'infermiera assicura il malato
assistere	4	assiste allo spettacolo
<b>assomigliare</b>	4	Bice assomiglia a Leo
<b>assumere</b>	22A	assume impiegati in ufficio
assumere	43	assume che X è uguale a Y
<b>attaccare</b>	20UM	attacca il castello
attaccare	23D	attacca il raffreddore a Bice
attaccare	28	attacca i manifesti al muro
attaccarsi	44	si attacca al fatto che tu la maltrattavi
<b>attendere</b>	43	attende (che Céline arrivi, Céline)
<b>attraversare</b>	20L	attraversa la strada
<b>attribuire</b>	23ND	attribuisce a se stesso l'errore
<b>aumentare</b>	10	aumenta di valore
aumentare	24	aumenta il prezzo (a,di) le pesche
<b>avanzare</b>	23R	avanza mille lire da te
avanzare	3	l'esercito avanza
avanzare	4	molte cose avanzano a Bice

<b>avere</b>	0	ha paura
avere	20NR	ha l'auto
avere	23R	ha un libro da Bice
avere	56	ha da leggere
<b>avvenire</b>	42	avvengono molte disgrazie ai poveri
<b>avvertire</b>	50	avverti Bice che aveva, avevo sbagliato
<b>avviare</b>	56	si avvia a capire
avviare	58	avvia Bice in piazza a incontrare Leo
<b>avvicinare</b>	20UM	avvicina una persona
avvicinare	28	avvicina la sedia al tavolo
avvicinarsi	57	si avvicina al bar a incontrare gli amici
avvicinarsi	7	si avvicina alla casa
<b>baciare</b>	18	bacia Bice sulla bocca
<b>badare</b>	44	bada a che non si feriscano
<b>bagnare</b>	24	bagna la faccia di acqua
bagnare	28	bagna il biscotto (nel latte + di latte)
bagnarsi	2	i vacanzieri si bagnano
<b>ballare</b>	2	balla (l'hip pop)
<b>bastare</b>	42	mi basta essere ordinato
bastare	42A	basta che non ci sia un errore
<b>battere</b>	10	batte alla porta
battere	2	il cuore batte
battere	20UM	batte il figlio
battere	20UM	batte tre concorrenti

<b>benedire</b>	20UM	il sacerdote benedice Bice
<b>bere</b>	16A	bevo il succo d'arancia
bersi	43	mi bevvi (che Bice fosse innocente, la storia)
<b>bisognare</b>	42B	bisogna che tu venga
<b>bloccare</b>	20R	blocca il meccanismo
bloccare	41	parlare di sesso mi blocca
bloccare	48	blocca per te la possibilità di essere promosso
<b>brillare</b>	20R	gli artificieri brillano la mina
brillare	5	il cielo brilla (di stelle)
<b>bruciare</b>	20NR	ho bruciato i libri
bruciare	3	il fuoco brucia
bruciare	41	mi brucia dover parlare con lei
bruciare	41	questa situazione ha bruciato il sindaco
<b>buttare</b>	3	la miniera butta (ancora)
buttare	30	butta una pietra nel mare
buttare	56	si butta a studiare
<b>cacciare</b>	20R	caccia molte lepri
cacciare	20UM	ho cacciato mia sorella
cacciare	28	caccia i libri sotto il tavolo
cacciarsi	8	il bambino si è cacciò sotto il letto
<b>cadere</b>	2	molti soldati caddero
cadere	3	il palazzo è caduto (a terra)
cadere	7	la foglia cade sulla pietra
<b>calare</b>	3	il sole cala

calare	30	calai un pacco dal 1 piano a 2 piano
calare	58	calo Leo a raccogliere i resti
calare	7	i nemici calano sulla città
<b>calmare</b>	41	fare yoga calma Bice
<b>cambiare</b>	10	cambia di casa
cambiare	20NR	cambia (auto, giacca)
cambiare	20UM	la mamma cambia il neonato
cambiare	3	(il tempo, il ragazzo) cambia
cambiare	31	cambia la mia penna con la sua
cambiare	33	cambia 100 euro in sterline
cambiare	41	fare la guerra cambia le persone
cambiarsi	10	si cambia (d'abito)
<b>camminare</b>	2	l'uomo cammina
camminare	3	l'orologio cammina
camminare	7	cammina lungo la strada
camminare	8	la strada cammina da Roma a Milano
<b>cancellare</b>	48	cancella dal programma che devi cantare
<b>cantare</b>	20I	il tenore canta una romanza
cantare	47	il bandito canta al commissario che era stato lui
cantare	47	le canta di andarsene
<b>capire</b>	43	capii (che sbagliavo, il romanzo)
capire	54	capii che eri tu da un indizio
<b>capitare</b>	42	mi capita di partire
capitare	7	capita in piazza

<b>cascare</b>	7	cascai dal letto sul pavimento
<b>cavare</b>	29	cavai la radice dal terreno
<b>cedere</b>	23D	cedette il libro a Bice
cedere	3	la trave cedette
cedere	4	cedette all'assalto
cedere	44	cedetti a travestirmi da Pierrot
<b>cercare</b>	28	cerco i guanti nel cassetto
cercare	56	cerco di capire
<b>cessare</b>	3	la pioggia cessa
cessare	56	cessa di sperare
<b>chiamare</b>	20UM	chiamo mio padre
chiamare	22	chiamo il gatto Felix
chiamare	49	chiamò i figli a fare il loro dovere
<b>chiedere</b>	47	chiesi a Céline di restare
<b>chiudere</b>	10	chiudi con la droga
chiudere	20L	chiudi la strada
chiudere	28	chiudi il pacco nella scatola
chiudere	55	ciò ti chiude ogni via d'uscita
<b>circondare</b>	24	circonda il fortino di soldati
<b>cogliere</b>	20R	colgo il fiore
cogliere	48	colgo dal tuo discorso che ti sposerai presto
cogliere	49	cogli il marito a baciare la vicina
<b>coltivare</b>	20	coltiva l'inglese
coltivare	20R	coltiva pomodori
<b>comandare</b>	20UM	il sergente comanda una squadra

comandare	28	il ministero ha comandato Bice a Milano
comandare	47	comanda ai soldati di attaccare
<b>combattere</b>	20UM	combatte i nemici
combattere	9	combatte con il nemico
<b>combinare</b>	0	combina poco in questo lavoro
combinare	31	combina l'acido con il preparato
combinare	48	combina con Bice di andare al cinema
<b>cominciare</b>	3	le lezioni sono cominciate (ieri)
cominciare	46	tutto cominciò dal fatto che partisti
cominciare	56	comincia a parlare
cominciare	8	il tragitto comincia qui
<b>commettere</b>	0	commette un delitto
<b>commuovere</b>	41	vederti malata mi commuove
<b>comparire</b>	7	compare nella stanza
<b>compiere</b>	0	compie il lavoro
<b>completare</b>	0	completa il lavoro
completare	41	il fatto di andare in India ti completerà
<b>complicare</b>	48A	ciò complica per me la possibilità di partire
<b>comporre</b>	20R	compose una lite fra Bice e Leo
comporre	20R	compose una musica
<b>comportare</b>	48	ciò comporta per me che ti devo sposare
<b>comprare</b>	23R	compra un libro dal libraio
<b>comprendere</b>	43	comprendo (che sia difficile, il discorso)
comprendere	54	compresi che eri lì dal fatto che cantavi

<b>comunicare</b>	20UM	il parroco comunica i fedeli
comunicare	47	comunicai a Bice che sarei partito
comunicare	9	comunico co Leo
comunicare	9	questa stanza comunica con il salone
<b>concedere</b>	23D	concessi un prestito a Leo
concedere	47	mi concedi di restare
<b>concludere</b>	2	quell'uomo concluse
concludere	0	concluse il suo discorso
concludere	54	conclusi che eri lì dal fatto che cantavi
concludersi	3	lo spettacolo si concluse
<b>condannare</b>	20UM	il giudice condannò Bice
condannare	43	condanno (che voi facciate ciò, Leo, quel gesto)
condannare	49	condannò il marito a restare da solo
<b>condurre</b>	20R	condussi la moto
condurre	28	condussi Céline al Luna Park
condurre	8	la strada conduce al cimitero
condurre	49	conduce la moglie a comportarsi male
condurre	58	conduce Céline a studiare il piano
condursi avv	10	si è condotta bene con Leo
<b>confermare</b>	47	confermai a Leo che il padre era partito
<b>confessare</b>	20R	ha sempre confessato la fede cristiana
confessare	20UM	il prete confessa i fedeli
confessare	47	confessai a Leo che avevo mentito
confessarsi	10	si confessa con un gesuita
<b>confondere</b>	31	confonde il rame con l'oro

confondere	41	vederti così bella mi confonde
confondersi	7	si confuse nella folla
confondersi	56	si confonde a parlare
<b>conoscere</b>	20NR	conosco il francese
conoscere	20UM	conosco quel signore
<b>conquistare</b>	20NR	conquistarono il fortino
conquistare	41	il fatto che parlasse bene mi ha conquistò
conquistare	43A	conquistò la possibilità di lavorare
<b>consegnare</b>	23D	consegnai il pacco a Leo
<b>consentire</b>	43	consento (che si parta, poche scappatelle)
consentire	44	consento a parlare
consentire	47	consento a Leo di uscire
consentire	47	le finanze non consentono a Leo di partire
consentire	54	consento con Leo sul fatto che si debba partire
<b>conservare</b>	0	conservo molte speranze
conservare	28	conservo i libri nella cassa
<b>considerare</b>	43	considero (se partirò, la proposta)
considerare	43	considero che Leo sia furbo
<b>consigliare</b>	47	consiglio a Leo di andarsene
<b>consistere</b>	10	il gruppo consiste di dieci persone
consistere	52	star bene per me consiste nel partire
<b>consolare</b>	41	mi consola vederla cadere
consolare	50	consolai Leo del fatto di essere stato bocciato
<b>consumare</b>	20NR	ho consumato le scorte
consumare	41	vederti in questo stato mi consuma

consumare	43A	ho consumato la possibilità di vederti
<b>contare</b>	10	conto fino a 10
contare	10	conto sui miei amici
contare	20NR	conto le pecore
contare	42	per me il fatto di essere sposata non conta
contare	43	conto (che partirò domani, di partire)
<b>contenere</b>	0	contieni l'ira
contenere	20ST	la scatola contiene i miei libri
<b>continuare</b>	3	il temporale continua
continuare	56	continua a lavorare
<b>controllare</b>	43	controllo (che tutto vada bene, la situazione)
<b>convenire</b>	31A	convengo il prezzo con i francesi
convenire	42	ciò non conviene a una signora
convenire	42	mi conviene partire
convenire	7	conviene a Roma
<b>convincere</b>	50	convinsi Leo (del fatto) che aveva torto, di ciò
<b>coprire</b>	20ST	la strada copre una lunga distanza
coprire	20UM	coprì il complice
coprire	24	copri il mobile con un lenzuolo
coprirsi	2	Leo (d'inverno) si copre
<b>correre</b>	20I	ha corso i cento metri
correre	3	ha corso (molto)
correre	7	è corso in camera
correre	8	l'autostrada corre da Salerno a Reggio Calabria
<b>costare</b>	20SP	questo libro costa dieci euro

costare	42	mi costa partire
<b>costituire</b>	20	costituisce uno spin off
costituire	55	andare in India costituisce per me una novità
<b>costringere</b>	49	ho costretto Céline (alla resa, a partire)
<b>costruire</b>	20	ha costruito una casa
<b>creare</b>	20	ha creato un nuovo tipo di sedia
<b>credere</b>	10	crede in Dio
credere	43	crede che tu sia stupida
credere	44A	crede al fatto che sei innocente
<b>crescere</b>	10	i girini crescono nell'acqua
crescere	3	(la confusione, il bimbo, la pianta) cresce
<b>cuocere</b>	20R	cuoci la pasta
<b>curare</b>	43	cura (che tutto sia in ordine, le piante)
curare	23	cura (la rosolia al figlio, il figlio)
curarsi	45	non si cura che tu sei malata
<b>custodire</b>	20ST	la cassa custodisce un tesoro
custodire	28SC	custodisci il bracciale nella cassa
<b>dare</b>	0	dai uno schiaffo a Leo
dare	23D	dai un libro a Leo
darsi	4	si da al vizio
<b>decidere</b>	43	decide (che parte, la partenza)
decidere	49	l'evento decise Leo a partire
decidersi	10	si decise per il vestito verde
<b>dedicare</b>	23ND	dedicò il libro al papà

dedicarsi	44	si dedicò a coltivare i campi
<b>definire</b>	43	definisci (se si deve partire, la data)
<b>deludere</b>	20UM	(Bice, lo spettacolo) deluse Leo
<b>denunciare</b>	47	denuncia a Leo di aver colto la moglie con l'amante
<b>descrivere</b>	47	descrisse a Leo che la situazione non era rosea
<b>desiderare</b>	43	desidero (che tu parta, Leo, un libro)
<b>destinare</b>	28	destinai Leo in Provenza
<b>dichiarare</b>	22	dichiaro Leo innocente
dichiarare	47	mi dichiarò che era innocente
dichiararsi	10	si dichiarò contro la droga
<b>difendere</b>	43A	difendo il fatto che si deve essere liberi
difendere	51	difendo Leo dal fatto che lo possano traviare
difendersi	2	Bice si difende (in matematica)
<b>diffondere</b>	28	diffonde il profumo nella stanza
diffondere	47	diffonde a tutti di essermi sposato
<b>dimenticare</b>	43	dimenticai (che dovevo partire, Leo, le promesse)
dimenticarsi	45	mi dimenticai che dovevo partire
<b>dimostrare</b>	47	mi dimostrò che era innocente
<b>dipingere</b>	20	dipinse un cavallo (sulla tela)
<b>dire</b>	47	gli dissi che doveva, dovevo partire
<b>dirigere</b>	28	dirige il figlio nella foresta
dirigersi	57	si dirige in piazza a comprare il pane
dirigersi	7	si dirige in piazza
<b>discendere</b>	46	ciò discende dal fatto che sei partito
discendere	6	l'uomo discende dalla scimmia

<b>discutere</b>	53	discussi con Leo del fatto che pioveva
<b>disperare</b>	45	dispera di trovare l'amico
<b>dispiacere</b>	42	mi dispiace che parti
<b>disporre</b>	28	disposi i libri sulla mensola
disporre	43	disposi che tutti partissero
disporre	45	disposi del fatto che è il figlio
disporre	49	disposi Leo a rivedere il padre
<b>distinguere</b>	31	Leo distingue bene Bice da Sara
distinguere	54	distinguo che sei russo da un indizio
<b>distrarre</b>	41	andare al cinema mi distrae
distrarre	51	distrarre Leo dal fatto che deve farsi operare
<b>distribuire</b>	23D	distribuisce doni ai bimbi
distribuire	28	distribuisce i libri nelle scatole
<b>distruggere</b>	17	distrugge il piatto (in mille pezzi)
distruggere	41	vederti così mi distrugge
<b>disturbare</b>	41	vederti così mi disturba
<b>divenire</b>	0	Leo diviene direttore
<b>diventare</b>	0	Leo diventa direttore
<b>divertire</b>	41	vederti saltare mi diverte
<b>dividere</b>	17	divide il blocco in otto pezzi
dividere	29	divide il tronco dalla radice
<b>domandare</b>	47	mi domandò di andarmene
<b>dormire</b>	2	Bice dorme
<b>dovere</b>	47	ti devo di essere stato promosso
dovere	56	devo partire

<b>durare</b>	20SP	lo spettacolo dura (molto, poco, un'ora)
durare	3	(questo legame, l'auto) dura
<b>educare</b>	49	educa i giovani a rispettare gli anziani
<b>eleggere</b>	22	elegge Leo presidente
eleggere	49	lo elegge a rivestire una carica importante
<b>elevare</b>	20	eleva un altro piano sul tetto
elevare	41	studiare linguistica ci eleva
<b>entrare</b>	4	la gonna non entra a Bice
entrare	42	ucciderti entra nei miei piani
entrare	57	entra in casa a prendere la sedia
entrare	7	entra in casa
<b>esagerare</b>	10	Leo esagera (in tutto, con questi modi)
<b>escludere</b>	43	esclude (che si possa partire, la possibilità)
escludere	48	ciò esclude che tu devi ballare dal programma
escludere	48	ciò esclude per me il fatto di poterti vedere
<b>eseguire</b>	0	Leo esegue il restauro della cattedrale
eseguire	20	Leo esegue un disegno sul muro
<b>esercitare</b>	20R	Leo esercita la memoria
esercitare	20R	quell'avvocato esercita la professione
esercitare	49	esercita il figlio a parlare inglese
<b>esistere</b>	3	(l'uomo, la natura, il tavolo) esiste
<b>esitare</b>	56	esita a parlarne
<b>esporre</b>	20R	Leo espone i quadri in una galleria
esporre	33	espone il figlio al pericolo

esporre	47	gli esposi che volevo partire
<b>esprimere</b>	47	gli espressi che volevo partire
esprimersi avvm 2		Céline si esprime bene
<b>essere</b>	0	Leo è alto
essere	3	(l'uomo, Dio) è
essere	57	è in casa a leggere
essere	8	è in casa
<b>estendere</b>	23D	estende a tutti il diritto
estrarre	29	estrae il rame dal composto
<b>evitare</b>	20UM	evito le persone fastidiose
evitare	47	gli evitai che lo assalissero
<b>fare</b>	0	oggi fa caldo
fare	20	fai un cesto di vimini
fare	43	feci che tutti si amassero
<b>favorire</b>	20UM	Leo favorisce sempre gli amici
<b>ferire</b>	18	la ferita al braccio
ferire	41	sentirti dire ciò mi ferisce
<b>fermare</b>	20ST	l'incidente ferma la circolazione
fermarsi	57	si ferma nella stanza a piangere
fermarsi	7	si ferma nella stanza
<b>fidanzarsi</b>	9	Leo si fida con Bice
<b>fidare</b>	10	Leo fida su (di voi, la onestà)
<b>figurare</b>	8	il balletto non figura nel programma
<b>filare</b>	20R	fila la lana
filare	57	fila in piazza a comprare il pane

filare	7	fila in piazza
<b>finire</b>	56	finisce a rubare
finire	56	finisci di mangiare
finire	7	finisci per terra
<b>firmare</b>	20R	firma il documento
<b>fissare</b>	28	fissa la piastra al pavimento
fissare	43	fissai (che sarei partito, la data)
<b>fondare</b>	20	fondo un'associazione
fondare	48	fondo il fatto di essere snob su elementi emotivi
<b>formare</b>	0	i soldati formano un quadrato
formare	20	forma un vaso
formare	41	andare a scuola forma i giovani
<b>fornire</b>	23D	fornisce armi agli arabi
fornire	24	fornisce gli arabi di armi
<b>forzare</b>	20R	forza lo sbarramento
forzare	49	forza Leo a parlare
<b>fregare</b>	20R	la massaia frega le pentole
fregare	23D	frega la bicicletta a Leo
fregare	41	fare così mi frega
<b>frequentare</b>	20UM	frequenta stranieri
<b>fuggire</b>	43A	fugge il fatto di vederla
fuggire	57	fugge in stazione a prendere il treno
fuggire	7	fugge in stazione
<b>fumare</b>	20R	Leo fuma sigari

fumare	3	(lo stufato, il motore) fuma
<b>funzionare</b>	10	il sorriso funziona da difesa
funzionare	3	la macchina funziona
funzionare	48	il fatto che sei straniero funziona con le donne
<b>garantire</b>	33	garantisce il prestito per lui
garantire	47	garanti a Leo che sarebbe partito
<b>gettare</b>	30	getta il libro nel secchio
<b>giocare</b>	33	gioca 10 euro su quel cavallo
giocare	52	gioca (a poker, a fare gli indiani) con Leo
<b>girare</b>	28	gira il foglio contro la lampadina
girare	42	mi gira di andare al cinema
girare	7	l'auto gira a destra
girarsi	57	si gira a vedere Bice
<b>giudicare</b>	43	giudico (che tu sia bravo, Leo)
<b>giungere</b>	57	giunge in casa a parlare con il padre
giungere	7	giunge in casa
<b>giurare</b>	47	ti giuro che partirò
<b>giustificare</b>	48	ciò giustifica per me il fatto di averti sgridato
giustificare	50	giustifico Leo di aver picchiato Antonio
<b>godere</b>	2	gli amanti godono
godere	45	godì che ti hanno promossa
<b>gradire</b>	43	gradisco (che mi si dia del tu, un regalo)
<b>gridare</b>	47	gli gridai che partivo
<b>guadagnare</b>	33	guadagnai 10 mila euro (dall'+ nell') investimento

guadagnare	43	guadagno (che mi promuovono, la promozione)
<b>guardare</b>	27	lo guardo dai cattivi amici
guardare	43	guardo (Bice partire, Leo, la stazione)
guardare	44	guardo (a) che tutto vada bene
<b>guarire</b>	27	(il medico, il farmaco) guarisce Leo dalla tonsillite
guarire	51	ciò mi ha guarito dall'idea di andarmene
<b>guidare</b>	20R	guido l'automobile
guidare	28	guido la truppa in battaglia
<b>illudere</b>	20UM	illusi Bice
illudersi	43	si illude che il mondo sia perfetto
<b>illuminare</b>	20NR	illumina la stanza
illuminare	41	vederti così mi illumina
<b>immaginare</b>	43	immagino (che tu sia qui, la verità)
immaginare	54	immaginai che tu eri qui da un indizio
<b>immergere</b>	28	immergo la pietra nel liquido
<b>imparare</b>	43	imparai (che si deve pagare, la lezione)
imparare	56	imparai a nuotare
<b>impazzire</b>	10	Leo impazzisce per i dolci
impazzire	2	Leo impazzisce
<b>impedire</b>	47	impedii a di fuggire
<b>impegnare</b>	33	impegnai l'anello al banco dei pegni
impegnare	49	impegnati a fare il bravo
<b>imporre</b>	47	imposi a Leo di parlare
<b>importare</b>	29	l'Italia importa carbone dalla Francia

importare	42	mi importa vederti felice
<b>incominciare</b>	3	lo spettacolo incomincia
incominciare	56	incomincia a parlare
<b>incontrare</b>	28SC	incontrai un amico in piazza
incontrare	42	il fatto di ballare il valzer non incontra più
<b>indicare</b>	22A	Bice indica Leo come suo successore
indicare	47	indicai a Leo di partire
<b>indossare</b>	20R	indosso una gonna corta
<b>infilare</b>	28	infilare gli anelli negli spuntoni
infilarsi	57	si infila in camera a leggere
infilarsi	7	si infila in camera
<b>informare</b>	53	informai Leo che era stato promosso
<b>iniziare</b>	20UM	i sacerdoti iniziano i giovani guerrieri
iniziare	46	tutto inizia dal fatto che sei partito
iniziare	56	inizia a parlare
iniziare	8	il tragitto inizia qui
<b>insegnare</b>	20R	Leo insegna storia
insegnare	47	insegna a Leo che bisogna dire la verità
<b>insistere</b>	43	insiste che era innocente
insistere	56	insiste a fare lo stupido
<b>intendere</b>	43	intende (che si vada al cinema, andare al cinema)
intendere	43	intesi (che Leo era partito, un urlo)
<b>interessare</b>	41	mi interessa vederti ballare
interessare	49	interessai Leo a studiare il cinese

<b>interrompere</b>	20ST	i massi interruppero la strada
<b>intervenire</b>	10	la polizia intervenne in quell'affare
intervenire	42	nel discorso intervenne il fatto che si deve studiare
<b>introdurre</b>	28	introdussi la chiave nella toppa
introdurre	48	introduco nel programma che dovete cantare
introdursi	57	si introdusse in casa a salutare Leo
introdursi	7	si introdusse in casa
<b>invadere</b>	20L	invase l'Iraq
<b>invecchiare</b>	3	(Leo, il vino) invecchia
<b>inventare</b>	43	inventai (che ero malato, una storia)
<b>inviare</b>	23D	invio una lettera a Leo
inviare	28	inviano le truppe in Afghanistan
inviare	58	invio Leo a comprare il pane
<b>invitare</b>	20UM	invito Leo
invitare	49	invito Leo a (cena, studiare)
<b>iscrivere</b>	20UM	iscrive il figlio a scuola
<b>isolare</b>	27	il suo comportamento lo isola da tutti
<b>lamentare</b>	43	lamenta (che nessuno lo ascolta, l'indifferenza)
lamentarsi	2	si lamenta
lamentarsi	45	si lamenta che lo trattano male
<b>lanciare</b>	30	Bice lancia il pallone dal balcone in giardino
lanciare	43	gli lancia (di partire, una proposta)
<b>lasciare</b>	20UM	Leo ha lasciato il compagno
lasciare	28SC	lascia il pacco sulla sedia
lasciare	43	lascia che i bimbi corrano

<b>laureare</b>	20UM	il Dipartimento laurea molti studenti (l'anno)
laurearsi	10	gli studenti si laureano in medicina
<b>lavare</b>	24	lavo i panni
<b>lavorare</b>	2	Bice lavora
lavorare	20NR	Leo lavora il legno
lavorare	44	lavoro a che tutto sia in ordine
<b>legare</b>	31	Leo lega le fascine
legare	9	Leo lega con Bice
<b>leggere</b>	20R	Céline legge un romanzo
leggere	47	ti leggo che devi partire
leggere	48	leggo sul giornale che parti
<b>levare</b>	29	leva il libro dalla cassa
<b>liberare</b>	27	libera Bice da Leo
<b>limitare</b>	20R	limita le spese
limitarsi	56	si limita a ridere
<b>litigare</b>	9	Leo litiga con Bice
<b>lottare</b>	9	lotta (con, contro) (Leo, la droga)
<b>maledire</b>	43	m. (che voi siate qui, la vostra presenza)
<b>mancare</b>	4	(il tempo, la forza) manca a Leo
mancare	4	Leo manca all'impegno
mancare	5	la casa manca di sedie
mancare	56	non manca di sbagliare
<b>mandare</b>	23D	Leo manda un pacco a Bice
mandare	28	manda Leo in Asia
mandare	58	manda Leo a comprare il pane

<b>mangiare</b>	16A	mangia un'arancia
<b>manifestare</b>	0	Leo manifesta disappunto
manifestare	10	gli operai manifestano contro il carovita
manifestare	47	gli manifesta di amarlo
<b>mantenere</b>	20ST	il pilastro mantiene la colonna
mantenere	28ST	Bice mantiene la testa di Leo nell'acqua
mantenere	48	mantiene nel programma che dovete ballare
<b>mentire</b>	4	mente a Bice
<b>meritare</b>	43	merita (che lo uccidano, la morte)
<b>mettere</b>	0	Leo mette in angoscia Bice
mettere	28	metti il libro nella cassa
mettere	49	metti Leo a lavorare
<b>minacciare</b>	50	minacciò Leo di ucciderlo
<b>misurare</b>	20NR	misura il terreno
misurare	20SP	il terreno misura 20 ettari
<b>montare</b>	20R	il toro monta la vacca
montare	20R	monta il cavallo
montare	28	monta il perno sull'asse
montare	57	monta sulla cima a vedere il sole
montare	7	monta sulla cima
<b>morire</b>	2	gli uomini muoiono
<b>mostrare</b>	47	mostra a Leo che aveva, avevo torto
<b>muovere</b>	28	muove il libro contro il muro
<b>nascere</b>	2	il bimbo nacque
nascere	46	ciò è nasce dal fatto che sei partito

nascere	6	i mammiferi nascono dall' utero
<b>nascondere</b>	20ST	quella caverna n. un tesoro
nascondere	28	nasconde il libro nella cassa
nascondere	47	gli nascosi che avevo sete
<b>negare</b>	47	nega a Leo di aver sbagliato
negare	47	nega a Leo di partire
<b>nominare</b>	22	l'assemblea nomina Leo presidente
<b>notare</b>	43	notai (che Leo non c'era, la sua assenza)
<b>nutrire</b>	0	Bice nutre ancora speranze
nutrire	24	nutre il bimbo di omogeneizzati
<b>obbligare</b>	49	obbliga Leo a partire
<b>occorrere</b>	42	mi occorre parlare con te
<b>occupare</b>	20L	Leo occupa abusivamente la casa
occupare	20ST	questo mobile occupa troppo spazio
occuparsi	45	si occupa di cercare le cose
<b>odiare</b>	43	odia (che dicano ciò, Leo, i discorsi)
<b>offendere</b>	41	fare questo mi offende
<b>offrire</b>	23ND	offro un libro a Leo
offrire	47	ti offrii di partire
<b>operare</b>	10	opera nel settore
operare	18	opera un paziente al cuore
<b>opporre</b>	0	Leo oppone un ostacolo a Leo
opporre	31	Leo oppone Bice a Bice
<b>ordinare</b>	18	Leo ordina i suoi libri
ordinare	47	gli ordina di andare

<b>organizzare</b>	48	organizzo con Leo di partire
<b>osare</b>	56	osa parlare
<b>osservare</b>	43	osserva (che tutto è in ordine, lo scontro, Leo)
<b>ottenere</b>	23R	Bice ottiene tutto da te
ottenere	43	ottenne (che tutti partissero, la loro partenza)
ottenere	51	ottenne un vantaggio dal fatto che era ricco
<b>pagare</b>	23ND	pagai il libro a Leo
pagare	43A	pago il fatto che sono scapolo
<b>parere</b>	42	mi pare che ti sbagli
<b>parlare</b>	20R	Bice parla francese
parlare	53	parla del fatto che parte con, a Bice
<b>partecipare</b>	20R	partecipo il dolore dei parenti
partecipare	4	partecipa alla gara
partecipare	44A	partecipa al fatto di travestirsi
partecipare	45A	partecipo del fatto che sei addolorata
<b>partire</b>	57	parto per l'Asia a vendere olio
partire	7	parto per l'Asia
<b>passare</b>	23D	passo il libro a Leo
passare	28	passo la fune nel gancio
passare	56	passo a spiegare il teorema
passare	57	passo per la piazza a comprare il pane
passare	7	passo per la piazza
<b>patire</b>	43A	patisco il fatto di essere vedova
<b>pensare</b>	43	penso (che ti sbagli, una sciocchezza)
pensare	44	penso ad andarmene

<b>perdere</b>	0	Bice ha perso la pazienza
perdere	28SC	persi l'anello nella sabbia
<b>perdonare</b>	47	ti perdono di avermi tradito
<b>permettere</b>	43	permetto (che tutti mi diano del tu, ogni cosa)
permettere	47	permetto a Leo di andarsene
<b>pesare</b>	20NR	Leo pesa il pacco
pesare	20SP	il pacco pesa un chilo
pesare	42	mi pesa di parlare ora
<b>piacere</b>	42	mi piace parlare
<b>piangere</b>	2	piango
piangere	20UM	piango i miei cari
<b>piantare</b>	28	pianto l'albero nell'orto
<b>picchiare</b>	18	picchiai Céline sul sedere
<b>piegare</b>	49	piego Leo a confessare
piegare	8	la strada piego a sinistra
<b>pigliare</b>	29	piglia il libro dalla cassa
<b>piombare</b>	10	il terrore piomba su tutti
piombare	20R	piomba il pacchetto
piombare	57	piomba in casa ad arrestarlo
piombare	7	i poliziotti piombano in casa
<b>piovere</b>	1	piove
<b>porre</b>	28	poni il libro nella cassa
porre	43	poni (che tu sia colpevole, l' ipotesi)
<b>portare</b>	20NR	porta la valigia
portare	23D	Leo porta il pacco a Bice

portare	28	porta la valigia in stazione
<b>posare</b>	28	posa il libro sul tavolo
<b>possedere</b>	20NR	possiedo molto oro
<b>potere</b>	56	posso partire
<b>precipitare</b>	7	precipita nel lago
<b>preferire</b>	54	preferisco che tu parta a che tu resti
<b>pregare</b>	43	prego che tutti rinsaviscano
pregare	50	pregai Leo di andarsene
<b>prendere</b>	0	Leo prese la febbre
prendere	29	prese il libro dalla cassa
<b>preoccupare</b>	41	mi preoccupa il fatto che tu parti
<b>preparare</b>	43	preparai (che tutti fossero pagati, i premi)
preparare	49	preparai Leo a partire
<b>presentare</b>	20R	l'annunciatrice presenta i programmi della serata
presentare	20UM	Leo presenta Bice
presentare	23ND	Leo presenta il libro a Bice
<b>prestare</b>	23D	presto il libro a Leo
<b>pretendere</b>	43	pretende (che tutti tacciano, il silenzio)
pretendere	43	pretende di avere ragione
<b>prevedere</b>	43	prevedo (che tu partirai, guai)
<b>procedere</b>	7	procedo lungo il fiume
<b>procurare</b>	23ND	procuro un libro a Leo
<b>produrre</b>	20	la Fiat produce auto
<b>proibire</b>	47	ti proibisco di partire
<b>promettere</b>	47	ti prometto che me ne andrò

<b>pronunciare</b>	20R	pronuncio un proclama
<b>proporre</b>	47	ti propongo di andare via
<b>proteggere</b>	27	Bice protegge Céline dai pericoli
proteggere	51	proteggerò Céline dal fare errori
<b>provare</b>	20NR	Leo prova l'auto nuova
provare	41	fare linguistica mi prova
provare	43	ciò prova (che hai ragione, la tua innocenza)
provare	47	ti provo che sono innocente
<b>provocare</b>	20UM	Leo provoca Bice
provocare	43	ciò provocò (che tutti partirono, la partenza)
<b>provvedere</b>	24	i difensori provvedono la città di buone difese
provvedere	44	provvedo a che tutto vada bene
<b>pulire</b>	24	pulisco la casa dallo sporco
<b>puntare</b>	33	punto il fucile contro Leo
<b>raccogliere</b>	16	Leo raccoglie tutti i giornali vecchi
raccogliere	28	raccolgo l'uva nel tino
raccogliere	43	raccolgo (che sono sgradito, l'indifferenza)
raccogliere	20ST	il tino raccoglie l'uva da pigiare
<b>raccomandare</b>	47	ti raccomando di partire
<b>raccontare</b>	47	gli raccontai che ero stato in Asia
<b>ragionare</b>	53	ragiono co Leo del fatto che piove
<b>raggiungere</b>	20L	Leo ha raggiunto la piazza
<b>rappresentare</b>	20R	gli attori rappresentano il Re Lear
rappresentare	20R	Bice rappresenta la Società
rappresentare	48A	ciò rappresenta per te la possibilità di partire

<b>rassegnarsi</b>	44	si rassegna a parlare
<b>realizzare</b>	20	realizzò una scultura in oro
realizzare	43	realizzò (che tutto era finito, la fine di tutto)
<b>recitare</b>	20R	Leo recita la parte di Re Artù
recitare	47	le recitai che l'amavo
<b>regalare</b>	23D	Leo regala un libro a Bice
regalare	47	ti regalo di andare in vacanze
<b>reggere</b>	20ST	due solidi pilastri reggono il ponte
reggere	28ST	Bice regge la scala contro il muro
reggere	3	la libreria regge
<b>regolare</b>	20R	regolo il macchinario
<b>rendere</b>	0	Céline rende felice Bice
rendere	20SP	l'affitto di quel terreno rende mille euro al mese
rendere	23D	rendo il libro a Leo
<b>resistere</b>	44	resiste al fatto che piove
<b>respingere</b>	20UM	Leo ha respinto Bice
respingere	28	respingo gli ostacoli indietro
respingere	43	respingo (che mi si tratti così, la proposta)
<b>respirare</b>	2	Leo respira
respirare	20R	qui si respira aria buona
<b>restare</b>	57	resto in casa a vedere la tv
restare	8	resto in casa
<b>ribellarsi</b>	44	si ribella a che lo boccino
ricercare	20UM	la polizia ricerca il noto bandito
<b>ricevere</b>	23R	Bice riceve tutto da te

<b>richiamare</b>	20UM	l'esercito richiama i congedati
<b>richiedere</b>	47	le richiede di pagare tutto
<b>ricominciare</b>	56	ricomincia a parlare
<b>riconoscere</b>	20UM	Leo ha riconosciuto il figlio
riconoscere	20UM	riconosce l'amico
riconoscere	43	riconosci (che ha torto, la verità)
riconoscere	54	riconosce che sei tu da un indizio
<b>ricoprire</b>	20R	ricopre un incarico
ricoprire	24	ricopre di raso quei vecchi divani
<b>ricordare</b>	43	ricorda (che pioveva, il film, Leo)
ricordare	47	mi ricordò che dovevamo partire
<b>ricorrere</b>	44A	ricorre al fatto che siete nobili
<b>ridere</b>	45A	ride del fatto che piangi
ridere	53	ride con me del fatto che piove
<b>ridurre</b>	54	riduce il vivere al mangiare
<b>riempire</b>	24	riempie il vaso di vino
<b>rientrare</b>	7	rientra da Roma
<b>rifare</b>	0	oggi rifà caldo
rifare	20	rifà la casa di mattoni
<b>riferire</b>	47	mi riferì che partivi
riferire	53	mi riferì che sei partito
<b>rifiutare</b>	20NR	rifiuta il cibo
rifiutare	20UM	Leo ha rifiutato Bice
rifiutare	43	rifiuto (che mi si tratti così, Leo)
<b>riflettere</b>	20ST	lo specchio riflette l'immagine di Céline

riflettere	44	riflette al fatto di partire
<b>riguardare</b>	41	il fatto di partire ci riguarda
<b>rimandare</b>	avvf 43	rimando a dopo (che si parli di te, la partita)
rimandare	20UM	il professore ha rimandato tutti gli alunni
rimandare	28	rimando il pacco a Roma
<b>rimanere</b>	57	rimango in casa a leggere
rimanere	8	rimani in casa
<b>rimettere</b>	16A	Leo ha rimesso tutto quello che aveva mangiato
rimettere	28	rimetti il libro nella cassa
<b>ringraziare</b>	50	ti ringrazio che resti
<b>rinunciare</b>	44	rinunciò a partire
<b>riparare</b>	20NR	riparai un macchinario
riparare	27	quella tettoia ripara chi aspetta dalla pioggia
riparare	44A	riparo al fatto di aver tradito
<b>ripensare</b>	44	ripenso a partire lontano
<b>ripetere</b>	47	mi ripetesti che partivi
<b>riportare</b>	23D	ti riporto il libro
riportare	28	riporta il libro nella cassa
riportare	47	Leo mi riportò che dovevi partire
<b>riposare</b>	2	Bice riposa
riposare	41	vedere la TV mi riposa
riposarsi	2	Leo si riposa
<b>riprendere</b>	50	riprende Leo di aver gridato
<b>risalire</b>	44A	risale al fatto che partisti
<b>rischiare</b>	43	rischia (che lo lincino, la vita)

<b>riservare</b>	20R	Leo riserva due posti a teatro
<b>risolvere</b>	43	risolve di partire
<b>risparmiare</b>	47	ti risparmio di parlarmi
<b>rispettare</b>	43	rispetto (che tu sia irritato, Bice, la diversità)
<b>rispondere</b>	47	gli risposi che avevo sonno
<b>risultare</b>	42	mi risulta che parti
<b>ritenere</b>	43	ritengo che sbagli
<b>ritirare</b>	30	ritiro le truppe dalla prima linea in retrovia
<b>ritornare</b>	57	ritorno da Roma a vendere olio
ritornare	7	ritorno da Roma
<b>ritrovare</b>	28SC	Leo ha ritrovato l'anello nel cassetto
<b>riunire</b>	16	Bice riunisce tutti i suoi amici
riunire	20L	Bice riunisce molte qualità in sè
riunire	28	riunisce le truppe nella valle
<b>riuscire</b>	56	riesco a parlare
<b>rivedere</b>	20UM	rivedo un amico
<b>rivelare</b>	47	gli rivelai che ero innocente
<b>rivestire</b>	24	rivesto di raso i divani
<b>rivolgere</b>	23	ti rivolgo la parola
rivolgersi	4	Bice si rivolge al professore
<b>rompere</b>	17	Leo ha rotto il vaso in mille pezzi
rompere	28	rompo il vaso sul tavolo
<b>rovesciare</b>	16A	Leo rovescia tutto quello che aveva mangiato
rovesciare	28	rovescia il vaso sul tavolo
<b>rovinare</b>	20UM	la passione per le donne ha rovinato Leo

rovinare	41	vederti mi rovina
<b>rubare</b>	23D	ti rubo il libro
<b>salire</b>	57	salgo sul tetto a riparare la tegola
salire	7	salgo sul tetto
<b>saltare</b>	3	Céline salta
<b>salutare</b>	20UM	saluto un amico
<b>salvare</b>	27	salvo Leo dal treno
<b>sapere</b>	20R	Bice sa molte lingue
sapere	43	so (che parti, l'inglese)
sapere	56	so parlare
<b>sbagliare</b>	10	sbagli in questo
<b>sbattere</b>	20R	sbatto i tappeti
sbattere	20R	sbatto le uova
sbattere	28	sbatto il leone in gabbia
<b>sbrigare</b>	20NR	sbrigo molte pratiche
sbrigarsi	56	si sbriga a partire
<b>scappare</b>	57	scappa dalla casa
scappare	7	scappa dalla casa
<b>scattare</b>	0	Leo ha scattato una foto
scattare	7	scatta verso il fiume
<b>scegliere</b>	43	sceglie (di partire, la partenza)
<b>scendere</b>	57	scende a valle
scendere	7	scende a valle
<b>scherzare</b>	10	scherza con gli amici
<b>schiacciare</b>	28SC	schiaccia l'uva nel tino

<b>sciogliere</b>	28SC	scioglie la polvere nell'acqua
<b>scivolare</b>	7	scivolo nel secchio
<b>scomparire</b>	57	scompaio dalla casa
scomparire	7	scompare dalla casa
<b>scoppiare</b>	3	il pallone scoppia
<b>scoprire</b>	20R	Leo ha scoperto la formula di un nuovo prodotto
scoprire	20UM	Bice ha scoperto una nuova attrice
scoprire	43	scopro (che si sbaglia, gli errori)
<b>scorrere</b>	7	scorre nel secchio
<b>scrivere</b>	20R	Leo ha scritto due romanzi
scrivere	47	gli scrissi che ero innocente
<b>scusare</b>	50	ti scuso di averlo fatto
<b>seccare</b>	21A	il sole secca le piante
seccare	41	vederti mi secca
<b>sedere</b>	57	siedo in poltrona a leggere
sedere	8	siedo in poltrona
<b>segnare</b>	20R	Leo ha segnato un bel goal
segnare	48	segna nel testo che parti
<b>seguire</b>	20NR	seguo (la moda, Bice)
<b>sembrare</b>	42	mi sembra che parti
<b>seminare</b>	20UM	il bandito semina i poliziotti
seminare	28	semina i geranei nei vasetti
<b>sentire</b>	2	Leo ormai segna poco
sentire	43	sentii (che partivate, Bice, i rumori)
<b>separare</b>	31	separo Bice da Leo

<b>seppellire</b>	28	seppellisco la cassa nella terra
<b>servire</b>	20UM	Leo e Bice serve una ricca famiglia
servire	4	serve a tutto
<b>sfruttare</b>	43A	sfrutta il fatto che sei medico
<b>sfuggire</b>	20UM	gli amici ormai sfuggono Leo
sfuggire	4	sfugge a Leo
<b>significare</b>	43	ciò significa (che parti, la fine)
<b>sistemare</b>	28	sistema i libri nella cassa
<b>smettere</b>	20R	Leo ha smesso quell'abito troppo liso
smettere	56	smetti di parlare
<b>soddisfare</b>	20UM	Leo soddisfa sempre Bice
soddisfare	41	vederti mi soddisfa
<b>soffocare</b>	2	Leo soffoca
soffocare	20UM	Leo ha soffocato Bice con un cuscino
<b>soffrire</b>	2	Leo soffre
<b>sognare</b>	43	sogno (che sono re, la nascita di mio figlio)
<b>sopportare</b>	43	sopporto (che tu parti, Leo, l'inimicizia)
<b>sopravvivere</b>	4	sopravvisse all'incidente
<b>sorgere</b>	3	il sole sorge
<b>sorprendere</b>	41	cederti mi sorprende
<b>sorridere</b>	2	Leo sorride
<b>sospendere</b>	0	sospende una seduta
sospendere	20UM	mi hanno sospeso in matematica
<b>sospettare</b>	43	sospetto che tu imbrogli
<b>sostenere</b>	20ST	due pilastri sostengono il ponte

sostenere	28ST	Leo sostiene la scala contro il muro
sostenere	43	sostengo (che tu imbrogli, una tesi diversa)
<b>sostituire</b>	31	sostituisci me con te
<b>sparare</b>	4	spara al coniglio
<b>spaventare</b>	20UM	Leo spaventa il figlio
spaventare	41	vederti mi spaventa
<b>spegnere</b>	20R	spegni la luce
<b>spendere</b>	33	spesi 50 centesimi per una mela
<b>sperare</b>	43	spero che tu parta
<b>spezzare</b>	17	spezzai la stecca in due pezzi
<b>spiegare</b>	47	spiego a Céline che deve studiare
<b>spingere</b>	30	Leo spinge l'auto dal giardino in garage
spingere	49	spinsi Bice a fuggire
<b>spogliare</b>	24	spogliai Leo di tutti i vestiti
<b>sposare</b>	20UM	il sacerdote ha sposato Leo e Bice
sposare	20UM	Bice ha sposato Leo
<b>spostare</b>	30	sposta il divano dal centro verso la parete
<b>spuntare</b>	20R	spunta una matita
<b>stabilire</b>	43	stabilisco (che si parte, la data)
<b>staccare</b>	28	stacca il ramo dal tronco
staccare	31	stacca il padre dal figlio
<b>stare</b>	57	stai in casa a leggere
stare	8	stai in casa
<b>stendere</b>	28	stendi i fili lungo la parete
<b>strappare</b>	28	strappa un peletto dal mento

<b>stringere</b>	28SC	stringi i libri nella cassa
<b>studiare</b>	20NR	Céline studia matematica
studiare	43	studia (che non si deve fumare, la lezione)
<b>subire</b>	43	subisco (che mi picchino, Bice, la violenza)
<b>succedere</b>	42	mi successe di tossire
<b>suonare</b>	20R	Céline suona uno strumento
<b>superare</b>	20R	l'auto di Leo ha superato l'auto di Bice
superare	20R	Céline ha superato brillantemente gli esami
<b>supplicare</b>	50	supplicai Leo di andarsene
<b>suscitare</b>	33	suscita emozione in Leo
<b>svegliare</b>	20UM	sveglia Céline
<b>sviluppare</b>	0	Leo sviluppa male il suo ragionamento
sviluppare	20R	sviluppo una foto
<b>svolgere</b>	0	svolge un compito male
<b>tacere</b>	2	Leo tace
tacere	47	tacque che suo padre era un ladro
tacere	53	tace con Leo del fatto che Bice parte
<b>tagliare</b>	17	Leo taglia il foglio in due parti
tagliare	20R	il sarto taglia il vestito
tagliare	28	taglia il ramo dal tronco
<b>telefonare</b>	47	gli telefonai di venire subito
<b>temere</b>	20UM	Bice teme Leo
temere	43	temo (che sia tardi, Bice)
<b>tendere</b>	23D	Leo tende il pacco a Bice
tendere	28	Bice tende la corda fino al piolo

tendere	44	tende a parlare poco
<b>tenere</b>	10	tiene per l'Avellino
tenere	28SC	tiene il bimbo sulle gambe
tenere	3	il vaso tiene
tenere	44	tiene a che tu venga
tenersi	7	mi tengo al ramo
<b>tentare</b>	20UM	il diavolo tenta anche Cristo
tentare	41	il fatto che balli mi tenta
tentare	49	tento Leo a rubare
tentare	56	tenta di partire
<b>terminare</b>	56	termina di mangiare
terminare	8	il tragitto termina in questo posto
<b>tirare</b>	10	tiro contro i piccioni
tirare	30	tiro la corda da qui fino al piolo
<b>toccare</b>	18	tacco Leo sulla gamba
toccare	20NR	Céline tocca tutte le cose che vede
toccare	41	vederti angosciato mi tocca
toccare	42	tocca a me muovere
toccarsi	2	quel ragazzo si tocca
<b>togliere</b>	23D	togli il posto a Leo
togliere	29	ho tolto il baccalà dalla pentola
togliere	48A	toglie dal programma che devi ballare
togliersi	57	si toglie da quell'angolo
togliersi	7	si toglie da quell'angolo
<b>tormentare</b>	20UM	tormenta i prigionieri

tormentare	41	vederti triste mi tormenta
<b>tornare</b>	56	torna a ridere
tornare	57	torna a vedere il film
tornare	7	torna a vedere il film
<b>tradire</b>	20UM	Leo ha tradito gli amici
tradire	43	ciò tradisce (che sei un razzista, le tue idee)
<b>trascinare</b>	20UM	Leo trascina con il suo discorso il pubblico
trascinare	20UM	Leo trascina sempre tutti gli amici
trascinare	30	Leo ha trascinato la cassa dal piano fino in cantina
trascinare	49	trascina Leo a compiere azioni disoneste
trascinarsi	3	questa storia si trascina (da troppo tempo)
trascinarsi	57	si trascina fino alla finestra a chiedere aiuto
trascinarsi	7	si trascina fino alla finestra
<b>trascorrere</b>	0	ha trascorso un periodo in India
trascorrere	3	il tempo trascorre
<b>trasferire</b>	20UM	Leo ha trasferito la segretaria in un altro ufficio
trasferire	30	Leo ha trasferito il proprio denaro dall'Italia in India
<b>trasformare</b>	0	gli alchimisti hanno trasformato l'oro in rame
trasformare	41	vederti così mi trasforma
<b>trasmettere</b>	10	trasmette con l'America
trasmettere	20R	l'aria trasmette i suoni e gli odori
trasmettere	20R	la poca igiene fa sì che i virus si trasmettano
trasmettere	20R	la televisione trasmette i programmi
trasmettere	23D	Bice trasmette la missiva a Leo
trasmettere	47	trasmisi a Leo che saresti arrivato

<b>trattare</b>	10	tratto con tutti
trattare	31A	Leo tratta la liberazione di Bice con i rapitori
trattare	45A	tratta del fatto che non c'è più fede
trattarsi	42	si tratta di partire
<b>trattenere</b>	20UM	Leo trattenne gli amici fino a tardi
trattenere	28ST	Leo trattiene il denaro in mano
trattenere	51	trattenni Leo dal fare una sciocchezza
<b>tremare</b>	3	(Leo, l'ingranaggio) trema
tremare	45A	tremavo del fatto di doverti parlare
<b>trovare</b>	28SC	trovi il libro nel pacchetto
trovare	43	trovo che sei stupido
trovare	49	trovai Leo a piangere
trovarsi avvm	10	si trova bene (con Bice, in Francia)
trovarsi	8	Milano si trova in Italia
<b>ubbidire</b>	4	ubbidisce a Leo
<b>uccidere</b>	20UM	Leo ha ucciso la moglie
<b>udire</b>	43	udii (che partivi, Leo)
<b>umiliare</b>	20UM	Leo umilia sempre la moglie
umiliare	41	vederti misera mi umilia
umiliarsi	44	mi umilia a chiedere perdono
<b>unire</b>	31	Leo unisce l'oro all'argento
unire	54	unisce che sei malato al fatto che zoppichi
<b>urlare</b>	47	gli urlai di partire
<b>usare</b>	20NR	uso il martello
usare	42	in Tibet usa sputare

usare	56	usa uscire
<b>uscire</b>	3	l'ultimo libro di Bob è uscito
uscire	42	nel suo discorso uscì che non ti ama
uscire	46	ciò uscì dal fatto che non c'è amore
uscire	57	uscii a comprare il pane
uscire	57	uscii a comprare il pane
uscire	7	uscii in strada
uscirsene	43	se ne uscì che non ti voleva più vedere
<b>valere</b>	20SP	un libro vale cinque euro
valere	42	gli è valso agire così
valere	49	le sue parole valgono a convincerlo
valere	55	aver scritto Herzog vale per lui un premio nobel
<b>vedere</b>	43	vedo (che Leo esce, Leo)
<b>vendere</b>	20NR	Leo vende lamponi
<b>venire</b>	57	vieni a vedere il film
venire	57	vieni a vedere il film
venire	7	vieni a casa mia
<b>vergognarsi</b>	45	si vergogna che ha detto così
<b>versare</b>	23D	Bice, ogni mese, versa una somma a Leo
versare	28	Bice versa il liquido nel serbatoio
<b>vestire</b>	2	Leo si veste
vestire	22A	vesta la figlia da Cleopatra
vestire	24	Leo veste la figlia di seta
vestire	3	la maglia veste
<b>viaggiare</b>	2	Bice viaggia

viaggiare	7	viaggia in tutti i paesi europei
<b>vietare</b>	53	vieta a Leo di partire
<b>vincere</b>	10	vince contro Leo
vincere	20R	l'Avellino ha vinto la partita
vincere	20R	Céline ha vinto un premio
vincere	20UM	Bice ha vinto Leo a scopone
vincersi	2	Leo si è vinto
<b>visitare</b>	18	il medico ha visitato Bice al cuore
visitare	20L	Bice ha visitato la Francia
visitare	20UM	gli amici hanno visitato Bice
<b>vivere avvm</b>	2	Leo vive stentatamente
vivere	2	Bice vive
vivere	20I	Bice ha vissuto una vita allegra
vivere	8	Leo vive (in Francia, a casa dei genitori)
<b>volare</b>	2	(l'uccello, l'aereo) vola
volare	57	vola a comprare il pane
volare	7	vola a Roma
<b>volere</b>	20NR	Bice vuole sempre tutto
volere	43	voglio (che tu venga, Leo, la verità, un libro)
<b>volgere</b>	28	Bice volge lo sguardo verso Leo
volgersi	57	si volge a guardare Leo
<b>voltare</b>	28	Bice volta lo sguardo verso Leo
voltare	7	l'auto volta a destra
voltarsi	57	si volta a guardare Céline
<b>votare</b>	10	vota per quel partito

votare	20R	il parlamento ha votato la legge 180
votare	20R	Bice vota da sempre lo stesso partito
votare	43	votai (che si partisse subito, il PD)
votarsi	4	si vota alla causa della libertà

#### 4. Analisi lessico-grammaticale di un testo

Per far comprendere la portata applicativa dell'approccio lessico-grammaticale, abbiamo deciso di analizzare un testo facendo riferimento all'impianto descrittivo del lessico-grammatica. Il testo prescelto è un brano estratto da *Isidora*, una delle "Le città invisibili di Italo Calvino" CITARE. Ecco il testo:

"All'uomo che cavalcava lungamente per terreni selvatici viene desiderio d'una città. Finalmente giunge a Isidora, città dove i palazzi hanno scale a chiocciola incrostate di chioccioline marine, dove si fabbricano a regola d'arte cannocchiali e violini, dove quando il forestiero è incerto tra due donne ne incontra sempre una terza, dove le lotte dei galli degenerano in risse sanguinose tra gli scommettitori. A tutte queste cose egli pensava quando desiderava una città."

Una prima analisi è volta ad individuare i complementi *di verbo*, *di nome* e *di frase* (cf. EMDA 1983) di seguito indicati con CV, CN, CF, attraverso la trasformazione di ogni periodo in una frase semplice attribuendo le marche CV, CN, CF.

I CN, che nel testo sono molti, contengono delle frasi relative introdotte da pronomi relativi, attaccati anche a nomi che a loro volta sono CN di un N che può essere un CV.

I CF possono essere rappresentati da gruppi nominali o da frasi introdotte da una preposizione o da una congiunzione.

Per un'analisi dettagliata del testo dobbiamo ricorrere, innanzitutto, alla trasformazione di ogni periodo in una frase semplice, e successivamente passare ad analizzare le eventuali frasi incassate come se fossero frasi semplici con i loro CV, CN, CF. In alcuni casi, laddove si trova una frase passiva, ricorreremo alla trasformazione in frase attiva.

Successivamente tutte le frasi semplici verranno elencate soltanto con i propri CV e subito sarà indicata la loro struttura

No, V, N1, Prep, Loc (complementi locativi di verbo)

Partendo dall'assunto harrisiano che le preposizioni e le congiunzioni intervengono in un momento successivo possiamo predisporre la nostra analisi partendo dai verbi che operano soltanto su argomenti elementari e verbi che operano su discorsi (cioè su frasi).

Un caso particolare di operatori sono quelli che svolgono un ruolo sintattico.

E' il caso del rapporto di derivazione che lega un verbo ad un nome:

- Il chimico ha analizzato il composto
- Il chimico ha fatto l'analisi del composto

In questo caso l'N sarà riscritto V-n.

Un'altra distinzione importante all'interno di tale analisi va fatta tra verbi supporto (Vsup) e ausiliari (Aux). I Vsup costituiscono una classe molto ampia di verbi e possono essere considerati "vuoti" da un punto di vista semantico-lessicale e formano una unità lessico sintattica con V-n. Il rapporto tra Vsup e V-n può far intervenire una o più preposizioni:

- Leo ammira Bice
- Leo è in ammirazione davanti a Bice

Vsup (E+ prep) V-n può essere considerata una entrata lessicale.

Adesso scomponiamo il testo in periodi.

### PERIODO 1

All'uomo che cavalcava lungamente per terreni selvatici viene desiderio d'una città

1. (All'uomo CV) (che cavalcava lungamente per terreni selvatici CN) (viene desiderio Vsup Vn) (d'una città CV)

2. (L'uomo CV) (cavalcava V) (lungamente CF) (per terreni (selvatici CN) CV)

Come abbiamo evidenziato in precedenza, laddove troviamo una frase passiva, ricorremo alla trasformazione in frase attiva, pertanto **venire desiderio di N a N** è una frase passiva a verbo supporto (*venire*) che va trasformata in attiva:

1A. (L'uomo CV) (ha desiderio Vsup Vn) (di una città CV)

**N<sub>0</sub> Vsup Vn di N<sub>1</sub>**

Nel caso invece della frase 2 possiamo vedere che:

2A. L'uomo cavalcava per terreni (selvatici) → risponde alla domanda **dove?**

Quindi **per** = **Loc** (preposizione locativa)

**N<sub>0</sub> V Loc N<sub>1</sub>**

## PERIODO 2

Finalmente giunge a Isidora, città dove i palazzi hanno scale a chiocciola incrostate di chioccioline marine, dove si fabbricano a regola d'arte cannocchiali e violini, dove quando il forestiero è incerto tra due donne ne incontra sempre una terza, dove le lotte dei galli degenerano in risse sanguinose tra gli scommettitori.

1. (L'uomo CV sottinteso) (giunge V) (finalmente CF)
2. (a Isidora CV)
3. [(Isidora) città dove i palazzi hanno scale a chiocciola incrostate di chioccioline marine, dove si fabbricano a regola d'arte cannocchiali e violini, dove quando il forestiero è incerto tra due donne ne incontra sempre una terza, dove le lotte dei galli degenerano in risse sanguinose tra gli scommettitori CN]

1A. L'uomo giunge (a Isidora città dove)

**N0 V Loc N1**

[(Isidora) città dove **CN1 → CN2\***]

*\*[CN2: dove F (in cui F = frase relativa)]*

1.2 (i palazzi CV) (hanno V) ((**scale a chiocciola CN3**) CV) {scale a chiocciola = polirematica nominale}

2A. I palazzi hanno scale a chiocciola

**N0 avere N1**

(i palazzi CV) (hanno V)

[(**scale a chiocciola\*** **CN3\*\***) CV]

*\*{scale a chiocciola = polirematica nominale}*

*\*\*[CN3: Vpp incrostate: F (in cui F = frase relativa)] che sono state incrostate \_ passiva, volta all'attivo:*

1.3 (N0 sconosciuto CV) (incrosta V) (**scale a chiocciola CV**) {di [chioccioline (marine CN)] CV}

3A. Scale a chiocciola incrostate di chioccioline marine

## **N0 V N1 di/con N2**

N0 incrosta scale a chiocciola di chioccioline (*di N2 = con N2*)

### 4A. (città dove) Si fabbricano a regola d'arte cannocchiali e violini

Nell'analisi della frase vediamo che c'è una frase relativa introdotta da *dove* ridotta in forma SI impersonale/passiva

N fabbrica cannocchiali e violini →

Cannocchiali e violini sono fabbricati da N →

Si fabbricano cannocchiali e violini

con la COORDINAZIONE.

1.4 a) (N0 SI impersonale CV) (fabbrica V) (cannocchiali CV)

(a regola d'arte\* CF)

*\*{a regola d'arte = polirematica avverbiale}*

COORDINAZIONE: e

1.4 b) (N0 SI impersonale CV) (fabbrica V) (violini CV)

(a regola d'arte CF)

### 4A. a/b N0 fabbrica cannocchiali, violini

## **N0 V N1**

Come risulta evidente la nostra analisi parte dal discorso minimo, cioè dalla frase semplice e in seguito mostreremo una sequenza di tutte le frasi semplici presenti nel nostro testo. Per poter eseguire questo tipo di studio si procede all'eliminazione di quanti più elementi è possibile e si valuta fino a che punto la frase sia accettabile, cioè fino a quando la frase abbia un senso. Dunque possiamo dire che ogni discorso è scomponibile in discorsi più piccoli e che questi possono essere ridotti a frasi semplici che siano accettabili e che diano la possibilità di studiare la struttura formale del discorso, cioè della comunicazione, e solo in seguito passare alle *aggiunzioni*. Attraverso la sintassi possiamo analizzare le strutture elementari delle frasi semplici e successivamente verificarne le combinazioni. Il

meccanismo di combinazione delle frasi è creativo in quanto genera continuamente nuove frasi, frasi che un parlante nativo può produrre e comprendere senza mai averle udite prima. La ricorsività è il meccanismo che permette al parlante di produrre frasi sempre nuove, mediante l'inserimento di una frase in un'altra. Dei meccanismi ricorsivi sono presenti nello stralcio di testo estratto da "Le città invisibili di Italo Calvino" che stiamo analizzando. In particolare abbiamo visto:

- La relativazione

"città dove i palazzi hanno scale a chiocciola incrostate di chioccioline marine"

Date più frasi  $F_1 =$  città dove i palazzi  $F_2 =$  hanno scale a chiocciola  $F_3 =$  incrostate di  
chioccioline

Isidora è una città che ha palazzi che hanno scale a chiocciola che sono incrostate di  
chioccioline marine

$F_1$  Pro-Rel  $F_2$  Pro-Rel  $F_3$

- Coordinazione

"Si fabbricano a regola d'arte cannocchiali e violini"

Date due frasi  $F_1 =$  si fabbricano cannocchiali e  $F_2 =$  si fabbricano violini

$F_1$  Cong coord  $F_2 =$  si fabbricano cannocchiali e violini

Se teniamo conto che questi meccanismi possono combinarsi tra di loro avremo una idea della creatività sintattica basata sulla ricorsività. A questi meccanismi, che sono di ricorsività infinita dobbiamo porre un limite empirico legato alla memoria e alla comprensione umana. La ricorsività è solo uno degli aspetti della creatività del linguaggio umano, un secondo è quello della co-occorrenza e selezione. I discorsi prodotti dai parlanti nativi sono dotati di senso ed è proprio a proposito di questo senso che possiede diversi gradi, che viene introdotto il concetto di accettabilità che Maurice Gross definisce come la combinazione di parole ammissibili o non ammissibili in una data lingua e dice (Gross 1975 p. 23) che "l'accettabilità è in effetti una nozione molto complessa che comporta delle intuizioni di forma e di senso, e che dipende da numerosi fattori culturali". Dunque, oltre ad una corretta costruzione esiste l'accettabilità che è necessario evincere in relazione alla combinazione

delle entrate lessicali tra loro. La co-occorrenza è proprio questo, due o più elementi che occorrono insieme in una determinata struttura.

5A dove quando il forestiero è incerto tra due donne ne incontra sempre una terza

**(dove quando CN5)\***

*\*[CN5: dove F (in cui FB = frase relativa)]*

*dopo l'inserimento di una frase (FA) introdotta dalla congiunzione QUANDO con valore temporale ma anche con valore ipotetico (periodo ipotetico della realtà), come se fosse SE: SE il forestiero è incerto tra due donne (ALLORA) il forestiero incontra una terza donna*

1.5 a) FA (il forestiero CV) (è incerto Vsup Agg) (tra (due DETnum) donne CV)

1.5 b) FB (il forestiero CV) (incontra V) (una (terza PRONum) (donna) CV) (sempre CF)

NOTA: **due** è determinante numerale cardinale, **terza** è pronome numerale ordinale

5A.1 Il forestiero è incerto tra (due) donne

**N0 Vsup Agg Prep N2**

5B.1 Il forestiero incontra una terza (donna)

**N0 V N1**

**(Il forestiero CN6)\***

*\*[CN6: dove F (in cui F = frase relativa)]*

6A (le lotte (dei galli CN)\* CV) (degenerano v) (in (risse (sanguinose CN) (tra gli scommettitori CN)CV)

6A. 1 Le lotte degenerano in risse

**N0 V in N1**

CN: **dei galli** = complemento di nome, che può essere interpretato anche come **tra i galli**, che si può sciogliere con una ulteriore frase relativa:

*le lotte che fanno i galli (tra loro)*

**(dei galli CN7)\***

*\*[CN7: dove F (in cui F = frase relativa)]*

che ci porta a un'altra frase semplice:

I galli fanno lotte con i galli = i galli lottano con i galli

I galli fanno lotte tra di loro = i galli lottano tra di loro

→ Le lotte tra i galli

→ Le lotte dei galli

Quindi abbiamo un'altra frase da considerare

6.A.2 (I galli CV) (fanno lotte Vsup V-n) (con i galli (tra di loro) CV= SIM)

6.B. 1. I galli fanno lotte con/tra galli/di loro

**N0 V Sim N1**

CN: **tra gli scommettitori**= complemento di nome, che può essere interpretato, analogamente al caso precedente, facendo riferimento alla frase relativa:

**(gli scommettitori CN8)\***

*\*[CN8: dove F (in cui F = frase relativa)]*

... risse che gli scommettitori fanno con gli scommettitori (tra di loro)

Pertanto abbiamo la frase semplice

Gli scommettitori fanno risse con gli scommettitori

**fare risse** = nessun verbo

Gli scommettitori fanno risse tra di loro

Quindi abbiamo un'altra frase da considerare

6.C. (Gli scommettitori CV) (fanno risse Vsup N) (con gli scommettitori (tra di loro) CV= SIM)

6C.1. Gli scommettitori fanno risse con/tra scommettitori/di loro

**N0 V Sim N1**

### PERIODO 3

A tutte queste cose egli pensava quando desiderava una città

1.7 (Egli CV) (pensava V) (a (tutte queste DETcomp) cose CV) (quando desiderava una città CF)

1A. Egli pensava a (tutte queste) cose

**pensare** però accetta anche una frase completiva con **il fatto che f**

**N0 V a N1, al fatto che F**

QUANDO= congiunzione → frase temporale

1.7.1 (Egli sottinteso CV) (desiderava V) (una città CV)

NOTA: Desiderare che **F → N0 V Ch F**

2A. Egli desiderava una città

**desiderare** però accetta anche una frase completiva<sup>17</sup> quindi

**N0 V N1, Ch F**

### Le frasi semplici totali con classe LG e presenza in VDB (FO)

- |                                     |                             |
|-------------------------------------|-----------------------------|
| 1. L'uomo ha desiderio di una città | Classe 0 Vsup desiderio VDB |
| 2. L'uomo cavalcava per terreni     | Classe 7                    |
| 3. L'uomo giunge a Isidora          | Classe 7 VDB                |

---

<sup>17</sup> Ci sono verbi che selezionano o un complemento di nome o un complemento di frase per esempio:

1. Céline dice (una bugia, che Giulia parla troppo, di stare zitto) a Wisdom
2. Céline prova (la tua innocenza, che tu sei innocente, di essere innocenze) al commissario

I verbi dire e provare selezionano in posizione N<sub>1</sub> sia un elemento nominale sia una frase introdotta dalla congiunzione che e da una eventuale proposizione *che*, nelle frasi (1) e (2) è *di*. Questo tipo di frase è denominata **completiva** (o frase complemento di verbo o complemento di frase) e i verbi che, in alternativa a un elemento nominale, selezionano una frase come complemento di verbo sono detti **verbi a completiva** (Vcomp). Questo tipo di frase prende il nome di completiva se è a tempo finito e introdotta dalla congiunzione *che*, invece è denominata **infinitiva** se è a tempo infinito e introdotta o meno da una preposizione. La notazione formale per le completive è **Ch F** mentre per le infinitive è **V-Inf** (Vietri 2004).

4. I palazzi hanno scale a chiocciola	<i>Classe 20 NR VDB</i>
5. N <sub>0</sub> incrosta scale a chiocciola di chiocciole	<i>Classe 24</i>
6. N <sub>0</sub> fabbrica cannocchiali, violini	<i>Classe 16A</i>
7. Il forestiero è incerto tra donne	<i>Classe 0 Vsup</i>
8. Il forestiero incontra una terza	<i>Classe 28SC VDB</i>
9. Le lotte degenerano in risse	<i>Classe 42</i>
10. I galli fanno lotte con/tra galli/di loro	<i>Classe 0 Vsup lotta VDB</i>
11. Gli scommettitori fanno risse tra di loro	<i>Classe 0 Vsup</i>
12. Egli pensava a cose	<i>Classe 44 VDB</i>
13. Egli desiderava una città	<i>Classe 43 VDB</i>

Osserviamo che su 13 frasi 7 hanno verbi o nomi che occorrono con Vsup nel VDB.

### **Il testo dopo l'analisi**

#### **Testo con solo frasi semplici senza CN né CF**

“All'uomo viene desiderio d'una città. Giunge a Isidora. A cose egli pensava”

Rivediamo il testo in cui in corsivo sono i verbi presenti nelle frasi incassate come complementi di nome o di frase, mentre sottolineati sono quelli principali.

“All'uomo che *cavalcava* lungamente per terreni selvatici viene desiderio d'una città. Finalmente giunge a Isidora, città dove i palazzi *hanno* scale a chiocciola *incrostate* di chiocciole marine, dove si *fabbricano* a regola d'arte cannocchiali e violini, dove quando il forestiero *è incerto* tra due donne *ne incontra* sempre una terza, dove *le lotte* dei galli *degenerano* in *risse* sanguinose tra gli scommettitori. A tutte queste cose egli pensava quando *desiderava* una città.”

**Verbi principali = 3:** avere desiderio, giungere, pensare

**CV= 6, CN= 15, CF= 3**

**Verbi in frasi incassate = 10:** cavalcare, avere, incrostare, fabbricare, essere incerto, incontrare, fare lotta, degenerare, fare rissa, desiderare

**CV= 16, CN= 6, CF= 3**

**Frase semplici totali = 13**

**Complementi di verbo totali = 22**

**Complementi di nome totali =21**

**Complementi di frase totali = 6**

Quello che possiamo osservare è che il testo di Italo Calvino contiene solo tre frasi semplici cosiddette “principali” e che la maggior parte delle informazioni che ne fanno un testo ricco di suggestioni è contenuta nelle componenti aggiuntive di queste frasi. Frasi relative, temporali, pseudo-concessive, aggettivi, apposizioni e avverbi rappresentano in modo preponderante il messaggio narrativo dello scrittore. L’analisi lessico-grammaticale ne ha messo in luce la straordinaria complessità.

Siamo ben lontani dalla possibilità di tradurre questa complessità in un sistema automatico di analisi, ma, a nostro avviso, è possibile proporre un primo livello di automatismo abbastanza soddisfacente.

## **CAPITOLO V**

### **L'ANALISI TESTUALE**

#### **1. Vocabolario di base della lingua italiana: descrizione della risorsa**

Il lavoro lessicografico di Tullio De Mauro comincia nel 1980, anno in cui pubblica Guida all'uso delle parole che in appendice contiene la prima versione del vocabolario di base della lingua italiana, che è già una realtà in altre lingue come l'inglese e il francese.

Lavori precedenti in quest'ambito, diversi comunque dal VdB di De Mauro per metodologia, obiettivi e numero di lemmi, sono: Thompson (1927), una tesi di laurea magistrale contenente un elenco di 500 parole; Knease (1933), una selezione di circa 2.000 parole da testi letterari; Skinner (1935), un elenco di 3.000 parole estratto dal libro di testo in italiano per stranieri; Russo (1947), un elenco combinato di 3.137 parole; Migliorini (1943), un elenco di 1.500 lemmi selezionato in base all'intuizione del linguista e non da un'analisi di tipo statistico dei dati. Per finire negli anni '70 (Juilland e Traversa, 1973; Sciarone, 1977) hanno lavorato rispettivamente ad un dizionario di frequenza per l'italiano e al vocabolario fondamentale della lingua italiana. Il primo lavoro presenta le 5014 parole più frequenti emerse dall'analisi, condotta su computer, di 500.000 occorrenze, il secondo propone una lista di 2500 parole ottenute dall'analisi di un corpus di 500.000 occorrenze. Un altro studio da prendere in considerazione, data la caratterizzazione di questo lavoro è il Lessico Elementare (Marconi Lucia *et alii*, 1994). Lo scopo di questo lavoro è stato quello di creare un dizionario di frequenza aggiornato e completo che fosse un punto di partenza per le ricerche linguistiche e uno strumento operativo di lavoro per chi si occupa di educazione, di bambini e di lingua scritta.

Il Lessico Elementare è un lessico di frequenza della lingua scritta, cioè un dizionario che contiene una lista di lemmi tratti da testi scritti, che sono stati scelti in base alla frequenza d'uso. Il volume comprende tre liste di frequenza:

- il Lessico Elementare di base: la lista delle 6095 parole più frequentemente usate nei testi scritti da e per i bambini che deriva dall'analisi di circa 1.000.000 di parole;
- il Lessico di Lettura: la lista delle parole che i bambini leggono di più;
- il Lessico di Scrittura: la lista delle parole che i bambini scrivono di più.

Il Lessico di Lettura è stato tratto dalla letteratura per l'infanzia cioè libri di lettura e di fiabe, giornalini e fumetti e testi scolastici mentre il Lessico di Scrittura è stato tratto da testi scritti da alunni delle cinque classi della scuola primaria. I testi scritti dagli alunni sono composti da circa 500.000 parole si riferiscono al triennio 1987 – 1989 e provengono da tutto il territorio nazionale.

De Mauro con il VdB apre la strada ad uno studio della linguistica statistica, che comincia a mettere in evidenza gli aspetti quantitativi e numerici della descrizione linguistica, creando in questo modo una rottura epistemologica con l'approccio tradizionale allo studio lessicale. Si afferma in tal modo l'importanza della statistica linguistica che studia la frequenza delle parole e la loro dispersione nei testi.

Tale studio per l'italiano nasce sulle orme degli studi lessicografici francesi che avevano già ampiamente utilizzato metodi statistici basati sulla frequenza per analizzare e descrivere il Français Elementaire (FF1) e individuare il lessico di alta disponibilità cioè quel lessico fortemente disponibile per i parlanti<sup>18</sup>.

Il gruppo di ricerca francese, dunque, pone l'attenzione al problema del lessico di frequenza partendo dallo studio sul vocabolario disponibile.

Un mot disponible è infatti una parola che, pur essendo utilizzato e ricorrente, si trova raramente nei testi parlati (Gougenheim et alii, 1956 [1964]): diversamente dall'aver una probabilità di occorrenza statisticamente assegnabile, essa "viene in mente" nel momento in cui risulta necessaria. Con il termine disponibilità lessicale si intende infatti la proprietà di una parola di essere più o meno immediatamente richiamata in associazioni mentali.

Lo studio sul lessico disponibile svolto dal FF1 rivolge la sua attenzione alla dimensione mentale – verso l'esprit – dei parlanti. L'indagine consiste infatti nell'individuazione di una serie di parole riferite a oggetti concreti che vengono più frequentemente associati dai parlanti ad ambiti tematici vicini al parlante stesso o a centri d'interesse (centres d'intérêts). Nel caso specifico, il campione sulla base del quale è stato indagato il lessico disponibile del francese è stato composto da studenti di scuola di età compresa fra i nove e i dodici anni, provenienti da quattro diverse aree geografiche del territorio francese.

Il vocabolario di base raccoglie due categorie di vocaboli:

- a) i vocaboli di maggior uso nei testi di una lingua in un dato momento, estraibili dai dizionari di frequenza delle varie lingue;
- b) i vocaboli che anche se poco usati nella lingua scritta e parlata sono percepiti come ad alta disponibilità perfino più dei vocaboli di maggior uso.

---

<sup>18</sup> Per un approfondimento bibliografico Gougenheim (1954).

I vocaboli di maggior uso sono ricavati dall'analisi statistica dei testi mentre i vocaboli di maggiore disponibilità sono invece ricavabili soltanto dal parlato.

Il primo vocabolario di base dell'italiano, d'ora in poi VdB, fu elaborato alla fine degli anni settanta e pubblicato come appendice al libro di Tullio De Mauro Guida all'uso delle parole, n.3 dei "Libri di base", 1ª edizione, Editori Riuniti, Roma 1980, pagine 149-183. Successivamente è stato utilizzato in diverse opere lessicografiche come il Gradit (Grande dizionario italiano della lingua dell'uso, 2ª edizione, 8 volumi, Utet, Torino 2007). (Grande dizionario italiano dell'uso, diretto da T. De Mauro, UTET, Torino 1999-2007.)

L'elenco di vocaboli redatto nel 1980 si fondava:

- sui dati di frequenza (numero di occorrenze di una parola nelle sue forme coniugate e declinate),
- sui dati di uso delle parole tratte dal LIF.

Il Lessico di frequenza della lingua italiana contemporanea, pubblicato dalla Ibm a Pisa nel 1970, elaborato dall'équipe del Cnuce di Pisa (Centro universitario di calcolo elettronico) formato da Uberta Bortolini, Carlo Tagliavini e Alberto Zampolli.

Il LIF costituito da 500 000 parole della lingua italiana contemporanea nell'uso scritto, portava all'individuazione di 5000 lemmi. La principale caratteristica del LIF era che i testi che costituivano il corpus erano caratterizzati da sincronicità e attualità. Infatti furono presi in carico testi di un periodo compreso tra il 1945 e il 1968, organizzati in cinque gruppi: teatro, romanzi, cinema, periodici, sussidiari. In secondo luogo, nella sua definizione venivano evidenziati i problemi di fondo della compilazione dei dizionari di frequenza. Oltre al LIF (Bortolini et al., 1972) il VdB ha beneficiato di un elenco di frequenze di parlato Italiano, LIP (De Mauro et al., 1993).

All'interno del VdB le parole sono suddivise in parole di maggiore uso, le prime 2000, che costituiscono il vocabolario fondamentale, così definito anche nei dizionari di frequenza di altre lingue e parole ad alto uso altre 3000. Il vocabolario fondamentale copre circa il 90% delle occorrenze di parole sia in testi scritti che nel parlato, mentre il vocabolario di alto uso copre circa il 6% delle occorrenze, anche se di uso molto superiore delle 50.000 parole comuni e del restante lessico.

Le 3000 parole ad alto uso del LIF prese in considerazione dal VdB sono state indagate attraverso un test di comprensibilità (curato da Massimo Vedovelli) e risultano accolte nel VdB solo quelle 2.750 comprese da almeno la metà di alunni e alunne di terza media di varie

regioni italiane. Solo quelle costituiscono il vocabolario di alto uso. Il vocabolario di alta disponibilità, infine, includeva circa 2.300 parole.<sup>19</sup>

L'insieme del VdB elencava dunque circa 7.050 parole delle tre categorie: fondamentale, d'alto uso, di alta disponibilità, d'ora in poi FO, AU, AD: erano le parole che, nella seconda metà del novecento, ai parlanti l'italiano risultavano le più conosciute o per il loro grandissimo uso rispetto al restante vocabolario o perché le più sentite dai parlanti stessi.

Nel corso degli anni il VdB è stato utilizzato per marcare le parole e in alcuni dizionari come il GRADIT è stato utilizzato nelle definizioni del significato dei lemmi per avere la certezza di una maggiore chiarezza per il maggior numero di persone possibili. In particolare nel GRADIT la marca d'uso non indica il lemma, ma ogni singola accezione: FO fondamentale, AU di alto uso, AD di alta disponibilità (FO, AU e AD rappresentano le parole che coprono circa il 98% dei discorsi), CO comune, TS tecnico-specialistico (seguita dell'ambito di riferimento), LE di uso solo letterario, RE regionale, DI dialettale, ES esotismo (forestierismo), BU di basso uso, OB obsoleto.

L'inserimento del VdB nel GRADIT segna una svolta lessicografica nella redazione dei dizionari, infatti a partire dalla seconda metà degli anni '90 i principali dizionari d'uso della lingua italiana hanno utilizzato le marche d'uso, sebbene con differenze nell'etichettatura.

L'utilizzo del VdB non si è limitato a questo infatti la casa editrice Giunti lo ha impiegato per indirizzare la redazione dei libri di testi per la scuola primaria al fine di migliorarne la comprensibilità, come del resto successivamente Editori Riuniti se ne è servito per ottenere che autori e autrici utilizzassero il più possibile parole del vocabolario di base. Alcuni editori stranieri, come Langenscheidt, si sono serviti del VdB per dizionari Bilingui.

A dare continuità a questi studi ci ha pensato il Dipartimento di Scienze del Linguaggio della Sapienza che con Emanuela Piemontese ha pubblicato un mensile di facile lettura, *Due parole*, che aveva l'obiettivo di supportare le persone con difficoltà linguistico-culturali o intellettive attraverso l'utilizzo sistematico di parole del vocabolario di base nella trattazione di ogni genere di argomento.

---

<sup>19</sup> Si veda anche Gougenheim *et al.* (1954). La nozione di disponibilità è stata introdotta a partire da uno spoglio lessico-statistico di campioni di lingua parlata, che aveva evidenziato come tra le unità lessicali più frequenti mancassero le unità legate ai contesti quotidiani e concreti di vita. A fronte di tale assenza, i vocaboli di alta disponibilità sono stati individuati tramite un'inchiesta che chiedeva a un campione di studenti delle scuole francesi di indicare le 20 parole che ritenevano maggiormente utili a proposito di 16 centri di interesse (tra cui il corpo umano, la casa, i pasti, ecc.) e rivedendo e integrando in una seconda fase questo elenco in seno a una commissione di esperti.

## 2. Il Nuovo Vocabolario di Base della Lingua Italiana

Dopo anni di ricerche nel 2016 esce la versione rinnovata del VdB il Nuovo Vocabolario di Base della lingua italiana (NVdB) (Chiari, De Mauro 2016) con 7.500 parole.

Tale esigenza nasce dall'evoluzione storico-linguistica del paese. I processi di unificazione e omogeneizzazione linguistici sono andati aumentando con il passare degli anni e lo stesso De Mauro sottolinea come i valori percentuali della popolazione che parla esclusivamente italiano sia aumentata e soprattutto la maggioranza della popolazione parla e scrive italiano e, allo stesso tempo, parla un dialetto. Nella evoluzione dell'italiano si sono consolidate caratteristiche che in origine erano proprie del parlato. Nel Nuovo vocabolario di base diverse centinaia di parole che non sono più usate, sono state eliminate e diverse centinaia che in passato erano assenti sono state inserite.

Il NVdB è formato da circa 2000 lemmi del Lessico Fondamentale (FO), 3000 lemmi di Alto Uso (AU) e 2500 lemmi di Alta Disponibilità (AD). Dalla osservazione dei lemmi presenti nel NVDB si può rilevare che il 73,3 % del vecchio lessico FO è presente in questa stessa categoria e così anche il 47% del lessico di AU. (Chiari, De Mauro 2014) Mentre il 15% dei nuovi vocaboli del lessico FO appartenevano nel VDB al lessico di AU, per esempio adulto, anziano, assenza, camion, buco, cassa, codice, concerto, individuo, insegnante, lavoratore, letteratura, maggioranza, paziente, procedura, reagire, ruolo, ritmo, strumento, telefonata, turno. Escono dal lessico FO vocaboli come aggiustare, agricoltura, carità, ebbene, fidanzato, mela, patria, pietà, volgare. Inoltre, molti vocaboli che prima appartenevano al lessico comune dell'italiano sono ora entrati nel lessico FO, come ad esempio auto, foto, chiuso, interessante, nascosto, evento, villaggio (Chiari e De Mauro 2012).

Inoltre, l'acquisizione di tecnologie di analisi automatica e l'utilizzo di applicazioni di linguistica computazionale hanno reso possibile l'analisi e la verifica dei lemmi su un corpus molto più ampio.

Mentre nella redazione del VdB si era partiti dal Lif basando l'analisi su un corpus di 500.000 occorrenze di parole raccolte in cinque categorie, il NVdB si basa sull'analisi di un corpus appositamente costruito per l'italiano contemporaneo (scritto e parlato), di 18.000.000 parole. Il corpus è stato organizzato in 6 sub-corpora di dimensioni simili, normalizzato: stampa (giornali e periodici), letteratura (romanzi, racconti, poesie), saggistica (libri di testo, saggi, enciclopedia), intrattenimento (teatro, cinema, canzoni e programmi TV), comunicazione mediata dal computer (forum, newsgroup, blog, chat e social network), registrazioni di parlato. L'intervallo cronologico dei testi va dal 2000 al 2012 e sono stati

scelti prevalentemente testi non eccessivamente lunghi di autori contemporanei, in modo da includere il testo integralmente. (De Mauro, Chiari 2012).

I testi sono stati ripuliti escludendo nomi propri, numeri, simboli, ideogrammi e icone, e lasciando come oggetto di analisi soltanto le parole. Le occorrenze di parole sono state lemmatizzate, cioè ricondotte a unità lessicali (vocaboli) presenti come lemmi già nel Gradit e sono state lasciate da parte le parole di frequenza minima, al di sopra della quale sono stati individuati 33.000 vocaboli.

Una delle principali novità del NVdB è l'inserimento di espressioni multiword (idiomi, espressioni fisse) nell'elenco dei lemmi.

Quindi, rispetto alle 260.000 del Gradit (Grande dizionario italiano dell'uso 1999 1a), De Mauro ha selezionato quindi le 7500 parole che possono essere considerate di base, cioè comprensibili a un vasto pubblico, anche con livelli di scolarizzazione non elevati.

Il NVdB si suddivide tra parole del vocabolario fondamentale (FO: 2000), vocabolario di alto uso (AU: 3000) e vocabolario di alta disponibilità (AD: 2500). Il NVdB marca ogni parola con l'abbreviazione dei diversi livelli.

Le duemila parole fondamentali coprono circa il 90 per cento delle occorrenze, le tremila parole d'alto uso coprono un ulteriore 6 per cento, il restante 8 per cento è occupato da occorrenze delle restanti ventottomila parole incontrate nei testi. Tra queste ultime alcune parole di alta disponibilità hanno notoriamente un comportamento instabile e spesso non attestato nei corpora scritti e parlati per via della loro concretezza e del loro legame con la vita quotidiana.

Queste tre tipologie non sono omogenee cioè il vocabolario fondamentale e di alto uso si basano su criteri lessico-statistici, mentre l'alta disponibilità si basa su una valutazione sociolinguistica, si tratta infatti di parole che i parlanti nativi hanno l'impressione di usare costantemente ma che in realtà hanno una frequenza molto bassa. Sono parole come pantofola, forchetta, padella legate ad azioni quotidiane e proprio per questo non hanno bisogno di essere pronunciate o scritte.

Con il passare degli anni le diverse edizioni di *Guida all'uso delle parole* hanno permesso l'implementazione dei dati metalinguistici. Nell'edizione del 1983, De Mauro inserisce le categorie grammaticali (classe, genere, numero) che aiutano a distinguere gli omografi (Ufficiale s.m., Ufficiale agg.; Deserto s.m., Deserto agg.). Nel 1989 viene aggiunto tra virgolette il significato per disambiguare i casi dubbi, e per finire nel 2003 si inseriscono i significati per distinguere due lemmi omografi (Riso "cibo" s.m. Riso "ridere" s.m.), si selezionano i significati da inserire nel VdB (Pianeta "corpo celeste" s.m.) e si indicano i

significati dei lemmi polisemici (Calcio “pedata, sport”). Nonostante questo lavoro di perfezionamento restano ancora tanti i casi di ambiguità, in quanto la polisemia è direttamente proporzionale alla frequenza d’uso di un lemma. Ciò significa che la maggior parte delle parole ha più significati e se volessimo fare un’analisi statistica della lingua potremmo vedere che più una parola è usata, più sono i suoi significati. La disambiguazione può avvenire spostandosi a destra e a sinistra del contesto in cui la parola è inserita.

Un altro elemento nuovo del NVdB è l’indicazione delle marche d’uso non per lemma, ma per singola accezione.

Gli elementi che rendono possibile lo sviluppo della nostra sperimentazione sono l’estrazione delle marche d’uso del NVdB e l’utilizzo di un tool di linguistica computazionale che ci permette di analizzare e confrontare i dati derivanti dall’applicazione delle marche d’uso.

### **3. Analisi testuale e marche d’uso**

Il lavoro di sperimentazione che è stato portato avanti, in un primo momento si è basato sull’informatizzazione del NVdB (Elia et al. 2021), cioè sull’inclusione delle marche d’uso di Tullio De Mauro nei dizionari elettronici del modulo italiano del software Nooj (Vietri 2008) che include dispositivi computazionali per la descrizione formale di fenomeni linguistici.

Dal corpus dei 21 testi analizzati per la nostra ricerca sono stati estratti 2 testi:

1. “Il piccolo principe” di Antoine de Saint-Exupéry
2. “Marcovaldo” di Italo Calvino

sono stati confrontati con 5 articoli scientifici che hanno come oggetto il Coronavirus- Covid 19 di vari autori.

L’ipotesi è stata di poter verificare la maggior presenza di parole del vocabolario fondamentale nei due testi di narrativa per bambini. I nostri testi sono stati suddivisi in capitoli, normalizzati e salvati in formato testo, formato di codifica UTF-8 in modo tale che la mancanza di formattazione li rendesse semplici e pronti per l’elaborazione linguistica. Il software che è stato utilizzato è Nooj, programma per il trattamento automatico del linguaggio naturale, che permette la costruzione di dizionari elettronici e di grammatiche

locali e la loro applicazione a corpora di grandi dimensioni. Il software NOOJ utilizza due specifici tools lingware:

- i dizionari elettronici, sviluppati dal Dipartimento di Scienze della Comunicazione dell'Università di Salerno, che integrano al momento cinque dizionari elettronici (sistema DELA)
- le grammatiche locali.

I DELA sono basi di dati lessicali formalizzate, strutturate omogeneamente in cui le caratteristiche morfo-grammaticali delle entrate (genere, numero e flessione) sono indicate da etichette alfanumeriche univoche e non ambigue. I DELA hanno la funzione di motori linguistici per i software di analisi testuale automatica, e sono di due tipi:

- di parole semplici (denominati DELAS-DELAFA), ed includono tutte le parole composte da sequenze di lettere non interrotte e delimitate da spazi bianchi (ad esempio la parola barca);
- di parole composte (denominati DELAC-DELACF), ed includono tutte quelle sequenze formate da due o più parole e che costruiscono congiuntamente singole unità di significato (ad esempio, le sequenze barca a vela) (Vietri 2001).

Il software Nooj gestisce al suo interno un unico dizionario elettronico contenente sia parole semplici che composte. Il secondo dei tools lingware, la grammatica locale, include istruzioni di carattere lessicale e morfo-grammaticale e viene usato dal software per leggere, comprendere ed etichettare sequenze di parole all'interno di testi in formato elettronico. Le grammatiche locali sono, quindi, dei tools per la semplice analisi testuale automatica, ovvero per l'information retrieval, e per effettuare operazioni di parsing. La loro denominazione deriva dal fatto che ciascun algoritmo viene elaborato per descrivere un solo aspetto della grammatica di una data lingua. Le grammatiche possono essere di tre tipi:

- grammatiche flessive e derivazionali,
- grammatiche morfologiche
- grammatiche sintattiche.

Le grammatiche flessive hanno come finalità quella di riconoscere le forme flesse delle parole semplici e delle parole composte. Le grammatiche morfologiche permettono la descrizione dei fenomeni derivazionali, le grammatiche sintattiche il riconoscimento di pattern specifici. Grazie a questi strumenti è possibile aprire un corpus e, in base alle risorse

linguistiche disponibili, effettuare un'analisi linguistica morfo-lessicale e sintattica; localizzare pattern mediante espressioni regolari e grammatiche; costruire le concordanze in base agli elementi presi in considerazione.

In questo caso, all'interno del modulo italiano di Nooj (a cura di Simonetta Vietri) sono stati integrate due risorse ad hoc:

1. Dizionari elettronici delle marche d'uso estratte dal Nuovo Vocabolario di Base della Lingua Italiana di Tullio De Mauro che includessero **FO** fondamentale, **AU** di alto uso, **AD** di alta disponibilità (FO, AU e AD rappresentano le parole che coprono circa il 98% dei discorsi), **CO** comune, **TS** tecnico-specialistico (seguita dell'ambito di riferimento), **LE** di uso solo letterario, **RE** regionale, **DI** dialettale, **ES** esotismo (forestierismo), **BU** di basso uso, **OB** obsoleto, utilizzate da Tullio De Mauro per la creazione del NVdB, il vocabolario che ha come obiettivo l'identificazione delle parole che costituiscono il nucleo centrale della lingua italiana. La figura 5 mostra come funziona il dizionario applicato a un testo. Qui si può osservare che ogni parola è etichettata con più marche d'uso, a causa della forte presenza di omografi nel Dizionario Italiano.

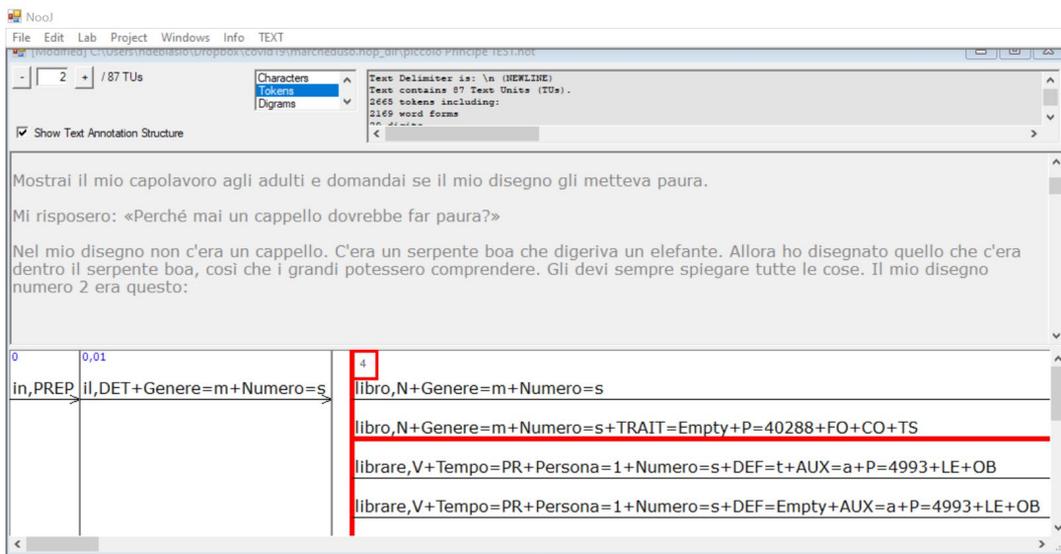


Figura 5 Esempio di una parte di testo analizzata con il dizionario delle marche d'uso del NVdB

2. Grammatiche sintattiche per il riconoscimento e l'estrazione dei tag associati alle marche d'uso. La grammatica riconosce tutti i verbi, i nomi, gli aggettivi e gli avverbi contrassegnati con le marche d'uso del NVdB nel dizionario. Quindi estrae sia le

marche d'uso che la Part-Of-Speech (POS), in modo da produrre, concordanze esportabili relative al rispettivo POS. La figura 6 mostra una Grammatica Sintattica.

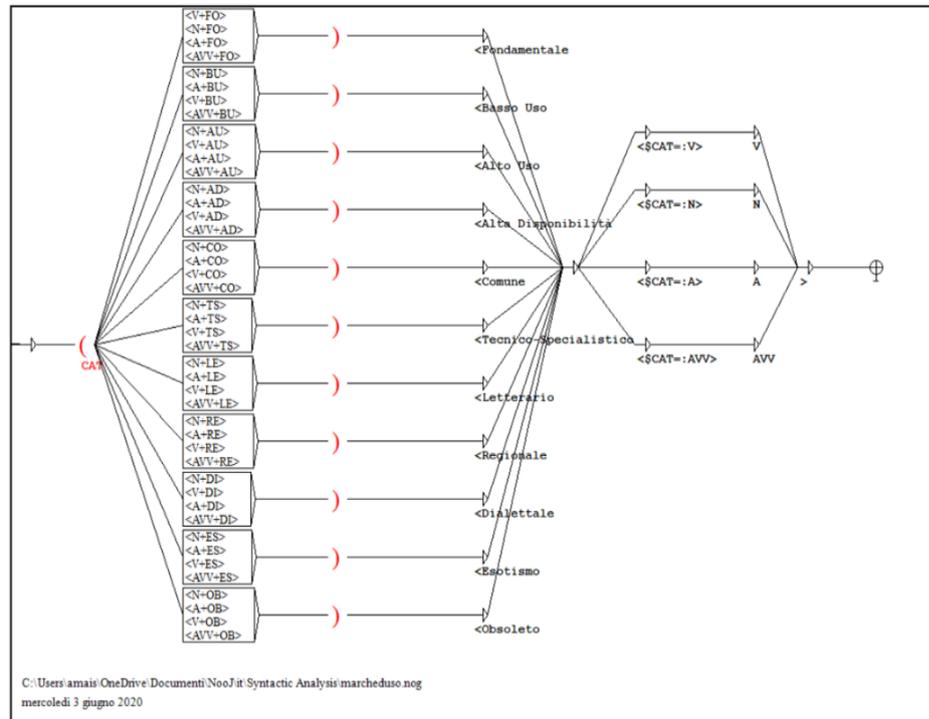


Figura 6 Grammatica sintattica per il riconoscimento e l'estrazione delle marche d'uso del NVdB

I dizionari creati e le risorse linguistiche sono state aggiunte all'interno di Nooj nelle "preferenze" ("Preferences"), per rendere possibile l'analisi linguistica ("Linguistic Analysis") su tutto il corpus, sulla base delle risorse inserite. Per effettuare la tokenizzazione del testo, il corpus inserito è stato preliminarmente messo nel formato "Noc".

I dizionari elettronici e le grammatiche locali, del modulo italiano di Nooj, arricchiti con le marche d'uso sono stati applicati al corpus. In questa prima analisi le marche d'uso utilizzate sono state **FO** fondamentale, **AU** di alto uso, **AD** di alta disponibilità che rappresentano le parole che coprono circa il 90% dei discorsi.

L'obiettivo di questo tipo di analisi è di dimostrare che esiste una correlazione tra il lessico utilizzato in un testo, la sua classificazione, fatta attraverso le marche d'uso e il livello di complessità del testo. In sintesi vogliamo dimostrare, attraverso l'utilizzo della linguistica computazionale, quanto aveva già compreso Tullio de Mauro, cioè che l'"uso" rappresenta un punto di svolta per lo studio del lessico, andando così nella direzione di una linguistica statistica, che attraverso l'analisi di tipo quantitativo cerca di cogliere gli usi linguistici. I risultati della prima fase sono mostrati nella tabella 6 e riassunti nella figura 7.

Tabella 6 Frequenza delle marche d'uso FO, AU e AD nel corpus

Marche	Articoli Covid (22020 tokens)		Il Piccolo Principe (17916 tokens)		Marcovaldo (41345 Tokens)	
	Frequency	%	Frequency	%	Frequency	%
FO	330	1.50	5300	29.58	10888	26.33
AU	216	0.98	1166	6.51	3533	8.55
AD	102	0.46	350	1.95	1609	3.89

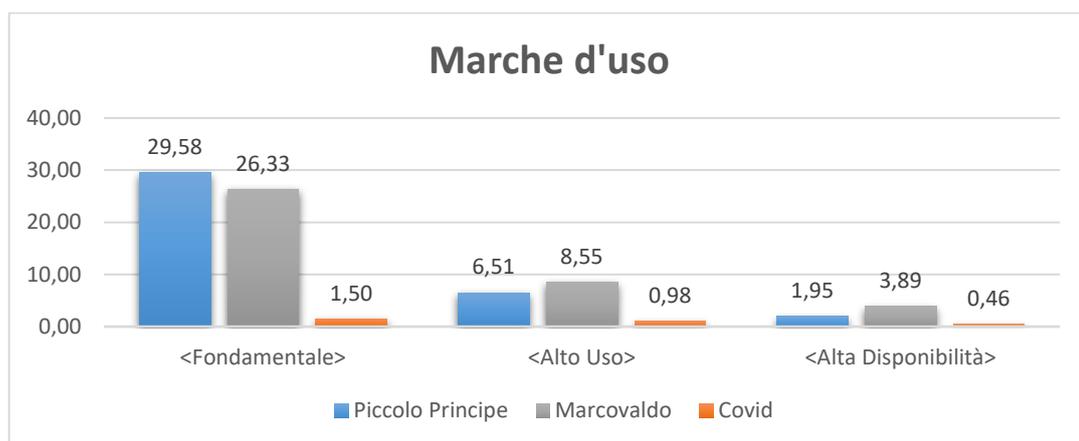


Figura 7 Diagramma delle frequenze delle marche d'uso FO, AU e AD nei testi

Il secondo step del nostro esperimento ha riguardato l'estrazione delle marche d'uso insieme al relativo POS, in modo da confrontare i tre testi anche per distribuzione delle marche d'uso sulla categoria grammaticale delle parole selezionate. La tabella 7 e figura 8 mostrano i risultati della seconda fase.

Tabella 7 Marche d'uso con distribuzione dei POS all'interno dei testi

Marche	Articoli Covid (22020 tokens)		Il Piccolo Principe (17916 tokens)		Marcovaldo (41345 Tokens)	
	Frequency	%	Frequency	%	Frequency	%
FO-V	101	0.46	2110	<b>11.78</b>	3806	<b>9.21</b>
FO-N	257	<b>1.17</b>	1214	6.78	3222	7.79
FO-A	186	0.84	851	4.75	1290	3.12
FO_AVV	28	0.13	587	3.28	1076	2.60
AU-V	72	0.33	348	1.94	893	2.16
AU-N	130	0.59	504	2.81	1591	3.85
AU-A	138	0.59	189	1.05	586	1.42
AU-AVV	0	0.00	32	0.18	75	0.18
AD-V	1	0.00	109	0.61	402	0.97
AD-N	83	0.38	135	0.75	831	2.01
AD-A	18	0.08	67	0.37	303	0.73
AD-AVV	3	0.01	39	0.22	73	0.18

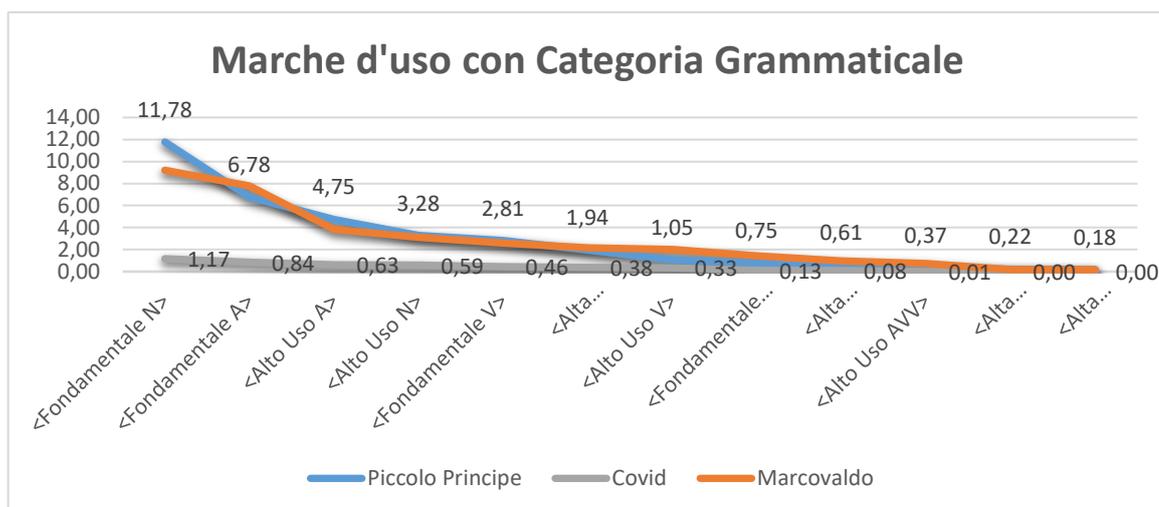


Figura 8 Diagramma delle frequenze delle marche d'uso con POS per testo

I dati che vengono fuori da questa prima analisi ci mostrano come i testi ci danno un feedback sulla frequenza delle parole del vocabolario fondamentale (FO) completamente coerente con la nostra ipotesi. Il Piccolo Principe, che è da considerarsi narrativa per l'infanzia a tutti gli effetti contiene, in percentuale, il numero maggiore di parole appartenenti alla marca d'uso FO. Anche Marcovaldo mantiene un'analogia distribuzione delle gamme lessicali. Al contrario, i documenti scientifici su COVID 19, destinati agli specialisti del dominio, contengono solo una piccola parte di parole FO.

#### 4. Valutazione automatica della leggibilità e complessità sintattica

In un secondo momento il lavoro portato avanti si è posto tre obiettivi:

1. Verificare la scelta dei vocaboli in relazione all'ipotesi di De Mauro già presentata nella sperimentazione descritta nel paragrafo precedente;
2. Sperimentare un sistema di valutazione della complessità sintattica;
3. Elaborare un sistema integrato automatico per produrre un Indice di Leggibilità lessicale e sintattico.

La metodologia proposta prevede il matching dei termini del Nuovo Vocabolario di Base (Chiari e De Mauro, 2016) informatizzati in Elia et al. (2021). Il conteggio dei termini, suddivisi in categorie grammaticali, ha permesso di effettuare un'analisi della complessità lessicale dei testi.

La complessità sintattica, al contrario, si è giovata di una fase di *parsing* che ha preso in esame diverse feature sintattiche quali relazioni di subordinazione o linearità dei periodi. Come è evidente il nostro lavoro si basa sull'assunto che la leggibilità di un testo è un concetto misurabile. Già in passato attraverso l'utilizzazione di formule si sono intrapresi studi sulla leggibilità dedicati a testi italiani. Un lavoro su tutti va menzionato: la collana dei Libri di base degli Editori Riuniti, avviata nel 1980 sotto la direzione di Tullio De Mauro, che aveva l'obiettivo di raggiungere un pubblico molto vasto, con almeno la licenza della scuola secondaria di primo grado. Lo scopo fu quello di far nascere una collana di alta comprensione, che in seguito portò alla pubblicazione del Vocabolario di base. In questo lavoro la leggibilità dei testi fu misurata con l'applicazione della formula di Flesch, adattata da Vacca.

Successivamente, partendo proprio dal Vocabolario di base (De Mauro 1980), sono stati realizzati altri due lavori che hanno permesso di creare una versione informatizzata della formula Gulpease (Piemontese 2005), che oltre a contemplare la lunghezza delle frasi e la lunghezza delle parole, includevano un'analisi lessicale del testo a partire dal Vocabolario di base (Amizzoni 1991; Mastidoro 1991)<sup>20</sup>.

Il concetto di leggibilità a cui ci siamo ispirati per la realizzazione del nostro lavoro si basa sulla facilità o difficoltà di lettura e comprensione di un testo da parte di bambini normodotati che frequentano la scuola primaria. Dunque, ci siamo concentrati sull'attribuzione automatica di un Indice di Leggibilità originale basato su due diversi valori: la comprensibilità lessicale e la complessità sintattica.

L'applicazione di un Indice di Leggibilità automatico consente in maniera predittiva, cioè prima di verificarlo attraverso l'esperienza del lettore, di controllare in che misura un testo sia facile da capire. Questo tipo di strumenti potrebbero essere molto utili per stabilire il grado di difficoltà che i testi scolastici, ad esempio, possono creare ad un giovane lettore.

I risultati dell'analisi automatica effettuata è stata verificata su un campione di lettori delle classi III, IV e V primaria dell'Istituto Comprensivo Don Milani di Manocalzati (AV).

Questa fase di sperimentazione ci è sembrata fondamentale per la validazione dei risultati ottenuti in quanto la comprensione, non è solo legata alla complessità di un testo ma, come abbiamo sottolineato nella parte introduttiva della tesi, è legata al recupero e all'integrazione

---

<sup>20</sup> Dagli studi di Maurizio Amizzoni (Calcolo automatico della leggibilità: l'indice Gulpease, 1991) e di Nicola Mastidoro (Rilevamento automatico del tasso di vocabolario di base, 1991) ha preso il via il lavoro che ha portato alla creazione di un algoritmo che è alla base del programma Eulogos Censor (<http://www.eulogos.net/it/censor/>). Il programma analizza la leggibilità di un testo secondo l'indice Gulpease e confronta le parole del testo con i vocaboli del VdB.

di tutte le conoscenze che il lettore possiede per la rielaborazione di un testo. Infatti, il processo di comprensione sia negli adulti che nei bambini è fortemente legato alle conoscenze enciclopediche. Dunque l'immaturità di queste conoscenze nei bambini e la difficoltà di creare inferenze rendono più complessa la comprensione.

## 5. Struttura dell'analisi

Le variabili linguistiche prese in considerazione nel nostro Indice di Leggibilità Automatico (ILA) sono il lessico e la sintassi (Elia 2007).

Per comprendere il valore del lessico partiamo dall'assunto di Tullio de Mauro secondo il quale il linguaggio che non ostacola la comprensione a livello linguistico, dovrebbe utilizzare termini del lessico fondamentale cioè quelle parole ad altissima frequenza, usate nel 90% circa dei discorsi e dei testi. Ciò però non significa che un testo ad alta leggibilità non possa utilizzare altre parole, ma che quando si usa un lemma non di base bisogna spiegarne il significato, e per farlo bisogna usare lemmi presenti nel VdB.

“Un discorso costruito tenendo d'occhio le parole dell'elenco, le ‘parole di base’, può anche contenere parecchie parole fuori elenco, purché spiegate con le parole di base” (De Mauro 1980: 138).

Un testo ad alta leggibilità non ha solo questa caratteristica ma ha frasi brevi e ben collegate fra loro sia dal punto di vista linguistico che concettuale, evitando metafore e astrazioni. Inoltre, ci sono anche dei requisiti redazionali attinenti alla grafica quali caratteri grandi, illustrazioni esplicative del testo, disposizioni dei contenuti (riquadri di spiegazione, testo a nuova riga per ogni frase, ecc.). Quanto detto vale anche per i testi ad alta leggibilità presi in considerazione nel nostro corpus.

Nel nostro ILA, nella valutazione della comprensione di un testo dal punto di vista lessicale, sono stati selezionati i termini Fondamentali (FO), di Alto Uso (AU) e di Alta Disponibilità (AD) del Nuovo Vocabolario di Base NVdB (Chiari, De Mauro 2006), suddivisi per categoria grammaticale (Aggettivi, Avverbi, Nomi e Verbi) andando a formare 12 classi di termini.

Questi termini sono stati individuati all'interno dei testi ed è stato attribuito un peso differente ai termini delle diverse classi, ovvero un valore doppio è stato assegnato alle parole marcate con FO.

L'utilizzazione del NVdB, come riferimento fondamentale per il controllo del lessico di testi scritti in italiano, e la sua informatizzazione (Elia et al. 2021) ci hanno permesso di creare uno strumento che consente di valutare la probabilità che un testo sia compreso da un'ampia fascia di lettori. Infatti, l'applicazione sistematica del NVdB ci ha consentito di classificare il grado di comprensione dei testi a livello lessicale.

Alla fase di *matching* dei termini appartenenti al NVdB basata sulla sua informatizzazione è seguita una fase di pre-processing, comprendente POS Tagging, Lemmatizzazione e Parsing. Per il pre-processing è stato utilizzato **Tint** (The Italian NLP Tool), un software per l'elaborazione del linguaggio naturale (NLP) in italiano. Tint (The Italian NLP Tool) è una pipeline open source scritta in linguaggio Java sviluppata dal laboratorio di Digital Humanities della Fondazione Bruno Kessler. La risorsa contiene oltre al modulo di parsing tutte le funzioni tipiche di un analizzatore sintattico (pos-tagging, lemmatizzazione e analisi morfologica).

Il Part-of-Speech Tagging (o PoS Tagging) utilizzato ha attribuito ad ogni parola appartenente al nostro corpus la corretta categoria grammaticale disambiguando, inoltre, i termini che potevano appartenere a diverse categorie grammaticali.

Con la lemmatizzazione siamo risaliti alla parola base del lemma partendo dalla forma che si è usata nel testo.

Il parsing, che è il processo attraverso cui si realizza una rappresentazione strutturale d'una sequenza di parole, ha prodotto una struttura ad albero contenente informazioni di tipo sintattico, che rappresenta la struttura della frase analizzata. Di seguito alcune delle etichette del parser stabilite dallo standard di Stanford (Universal Dependencies Representation):

root - root

dep - dependent

aux - auxiliary

auxpass - passive auxiliary

cop - copula

arg - argument

agent - agent

comp - complement

acomp - adjectival complement

**ccomp - clausal complement with internal subject**

**xcomp - clausal complement with external subject**

obj - object

dobj - direct object

iobj - indirect object

pobj - object of preposition

subj - subject

nsubj - nominal subject

nsubjpass - passive nominal subject

**csubj - clausal subject**

csubjpass - passive clausal subject

cc - coordination

conj - conjunct

expl - expletive (expletive “there”)

mod - modifier

amod - adjectival modifier

appos - appositional modifier

**advcl - adverbial clause modifier**

det - determiner

predet - predeterminer

preconj - preconjunct

vmod - reduced, non-finite verbal modifier

mwe - multi-word expression modifier

mark - marker (word introducing an advcl or ccomp)

advmod - adverbial modifier

neg - negation modifier

**rmod - relative clause modifier**

quantmod - quantifier modifier

nn - noun compound modifier

npadvmod - noun phrase adverbial modifier

tmod - temporal modifier

num - numeric modifier

number - element of compound number

prep - prepositional modifier

poss - possession modifier

possessive - possessive modifier (’s)

prt - phrasal verb particle

parataxis - parataxis

goeswith - goes with

punct - punctuation

ref - referent

sdep - semantic dependent

xsubj - controlling subject

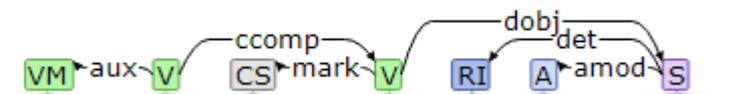


Figura 9 Esempio di rappresentazione sintattica di Tint

## 6. Complessità sintattica e semplicità lessicale

Nella valutazione della leggibilità dal punto di vista sintattico tramite l'analisi del risultato del parsing sono stati prese in esame diverse feature sintattiche come la presenza di subordinate, la linearità dei periodi, la presenza di modificatori ed incisi.

Nel nostro parser gli indicatori sintattici presi in esame sono stati i seguenti:

1.  $V_{mod}$  che include frasi subordinate (soggettive, preposizionali, infinitive, relative, avverbiali),
2.  $V_{punt}$  che evidenzia la presenza di coppie di segni di punteggiatura interne alla frase,
3.  $V_{ordine}$  che prevede un ordine di frase diverso da Soggetto-Verbo-Complemento ed evidenzia la presenza di soggetti non riconducibili ad un Sostantivo.

I valori presi in considerazione in  $V_{mod}$  sono 2:

- Il valore 1 che comprende Xcomp: clausal complement with external subject ovvero un complemento clausale di un verbo o di un aggettivo che è un complemento predicativo senza un proprio soggetto. Il riferimento del soggetto è necessariamente determinato da un argomento esterno al "xcomp".

Il valore 2 che comprende:

- Csubj: clausal subject ovvero una frase soggettiva;
- Ccomp: clausal complement with internal subject ovvero un complemento clausale di un verbo o di un aggettivo è una proposizione dipendente che è un argomento centrale. Cioè, funziona come un oggetto del verbo o dell'aggettivo;
  - Antonio dice [che tu nuoti]
  - La cosa importante è [mantenere la calma]
  - Antonio dice [che nuoti], ma non Antonio dice [che nuota].
- Rcmmod (**acl:relel**): relative clause modifier ovvero una frase relativa
- Advcl: adverbial clause modifier ovvero una frase avverbiale

Per  $V_{punt}$  che considera la punteggiatura sono state valutate:

- Virgole: +1 per coppia di virgole presenti nella frase
- Punto e virgola: +1 per ogni coppia di Punto e Virgola presente nella frase.

Per  $V_{ordine}$  che prevede un ordine di frase diverso da Soggetto-Verbo-Complemento ed evidenzia la presenza di soggetti non riconducibili ad un sostantivo è stato considerato l'ordine della frase:

- Soggetto: +1 se il soggetto appare DOPO il verbo
- Complementi: +1 se un complemento che non è il soggetto appare PRIMA del verbo.

Il punteggio complessivo per il calcolo della complessità sintattica viene effettuato attraverso la formula:

$$S_t = \frac{\sum S_{ft}}{N_{ft}}$$

Dunque la complessità sintattica del testo è data dalla sommatoria delle complessità sintattiche di ogni frase del testo diviso per il numero di frasi del testo. Di seguito un esempio dell'applicazione della formula della complessità sintattica a due frasi tratte da due tesi del nostro corpus.

Quando Gelsomino **nacque (advcl)** la gente del paese si alzò nel cuor della notte, **credendo (advcl)** di **aver udito (xcomp)** le sirene delle fabbriche che **chiamavano (acl:relcl)** al lavoro: era soltanto Gelsomino che **piangeva (acl:relcl)** per **provare (advcl)** la voce, **come fanno tutti (advcl)** i bambini appena nati. (*Gelsomino nel paese dei bugiardi*, G. Rodari).

- Lunghezza: 45 elementi
- V\_mod: 7
- V\_punt : 2
- V\_ordine: 2
- SL: 0.837, CS: 2.398: ILA: -1.56

Alle otto e mezzo, il signor Dursley prese la sua valigetta ventiquattr'ore, **sfiorò (advcl)** con le labbra la guancia della moglie, e tentò **di dare un bacio (xcomp)** a Dudley, ma lo mancò perché, in quel momento, in preda a un furioso capriccio, il pupo stava **scagliando (advcl)** i suoi fiocchi d'avena contro il muro. (*Harry Potter e la pietra filosofale*, J. Rowling).

- Lunghezza: 62 elementi
- V\_mod: 3
- V\_punt : 7

- V\_ordine: 0
- SL: 0.931, CS: 1.872: ILA: -0.941

Il secondo esempio riportato di seguito rende ancora più evidente il funzionamento della nostra formula di complessità sintattico. Abbiamo trasformato delle frasi semplici di un testo in frasi subordinate e l'indicatore di complessità è aumentata in modo significativo.

La mia vita è monotona. Vado a caccia di polli, gli uomini cacciano me. Tutti i polli si somigliano, e tutti gli uomini si somigliano. Dunque mi annoio un po'. Ma se tu mi addomestichi, nella mia vita ci sarà un sole. Riconoscerò un rumore di passi che sarà differente da qualsiasi altro. Gli altri passi mi faranno nascondere sotto terra, il tuo mi chiamerà fuori dalla tana, come una musica.

- Semplicità Lessicale: 0.642
- Complessità Sintattica: 0.651
- Totale: **-0,008**

La mia vita è monotona perché vado a caccia di polli e gli uomini cacciano me, ma tutti i polli si somigliano, e tutti gli uomini si somigliano, dunque mi annoio un po'. Ma se tu mi addomestichi, nella mia vita ci sarà un sole e riconoscerò un rumore di passi che sarà differente da qualsiasi altro perché gli altri passi mi faranno nascondere sotto terra, mentre il tuo mi chiamerà fuori dalla tana, come una musica.

- Semplicità Lessicale: 0.642
- Complessità Sintattica: 2.014
- Totale: **-1.373**

Per lo sviluppo di questa seconda parte dell'indice, che riguarda la semplicità lessicale, si è partiti dal matching di Nomi, Aggettivi, Avverbi e Verbi inclusi nel NVdB all'interno di un testo.

Nel caso in cui una parola è ambigua o appartiene a più categorie, viene applicata la seguente scala di priorità per determinare a quale marca appartenga:

- FO - AU – AD

Per ogni testo, la **semplicità lessicale** è determinata dalla seguente formula:

$$L_t = 2 * \log \frac{(2 * A_{FO} + N_{FO} + AVV_{FO} + V_{FO}) + A_{AU} + N_{AU} + AVV_{AU} + V_{AU} + A_{AD} + N_{AD} + AVV_{AD} + V_{AD}}{NVdB_t}$$

La formula di semplicità lessicale è uguale al doppio del logaritmo delle parole FO del testo per due, più le parole AU e AD del testo, fratto il totale delle parole del NVdB trovate nel testo. Nella formula viene data un valore doppio alle parole che sono marcate come fondamentali per creare un equilibrio tra il peso attribuito all'analisi lessicale e il peso attribuito all'analisi sintattica.

Il punteggio massimo ottenibile è :  $2 * \log(2) = 1.386295294$ .

L'algoritmo, che comprende i termini Fondamentali (FO), di Alto Uso (AU) e di Alta Disponibilità (AD) del Nuovo Vocabolario di Base NVdB, suddivisi per categoria grammaticale (Aggettivi, Avverbi, Nomi e Verbi) che sono stati applicati al corpus, ci mostra che esiste una correlazione tra il lessico utilizzato in un testo, la sua classificazione, fatta attraverso le marche d'uso, e la complessità del testo.

Tabella 8 Frequenza delle marche d'uso con distribuzione dei POS all'interno dei testi

	Coruore	Favole al telefono	Gelsomino nei paesi dei bugiardi	Harry Potter e la pietra filosofale	Il corsaro nero	Il Piccolo Principe	L'isola del tesoro	Le avventure di Alice nel paese delle meraviglie	Le avventure di Pinocchio	Marcovaldo	Sandokan alla riscossa	Cani, gatti, rospi e altre persone battucore	Emozioni del fenicottero	La lezione del minestrone	Il fantastico porcellino	I tre porcellini	Nuvola bianca e nera	Oceani di plastica	Uffa le femmine	Alla ricerca dei colori perduti	Fantasma nell'armadio
Tokens	99.346	84.543	40.085	102.507	117.761	16.923	78.748	30.733	49.819	160.551	122.153	11.997	12.067	20.084	982	856	1.630	31.308	1.438	14.128	982
Agg FO	3.753	2.636	1.253	2.963	3.845	720	2.644	753	1.442	4.355	4.163	383	434	683	34	56	39	1.177	49	488	35
Agg AU	585	518	171	550	1.102	90	464	83	156	1.063	1.023	112	74	73	5	1	9	216	5	83	7
Agg AD	122	106	39	102	170	16	95	30	62	250	162	18	14	13	2	3	3	37	1	18	5
Nom FO	11.013	8.373	4.098	8.847	9.685	1.537	7.676	1.892	4.233	15.774	9.225	1.367	955	1.533	42	100	136	3.200	116	1.329	104
Nom AU	2.586	2.541	1.128	2.384	4.019	308	1.874	536	1.075	4.808	3.865	407	292	440	61	28	45	950	41	334	46
Nom AD	934	1.177	471	842	1.208	109	850	215	736	3.233	1.237	173	109	176	40	2	15	284	6	112	11
Avv FO	3.902	3.444	1.821	4.862	3.950	327	3.170	1.515	2.279	5.545	5.065	490	579	1.203	52	30	92	1.395	85	693	35
Avv AU	124	121	42	159	265	25	164	48	84	136	354	24	9	35	-	-	1	55	4	12	2
Avv AD	142	100	35	134	193	31	114	13	14	202	170	11	8	29	3	4	59	1	16	-	
Ver FO	13.339	11.831	6.142	16.282	15.018	2.590	9.834	4.543	6.902	18.679	16.248	1.699	1.961	3.570	138	124	252	4.323	249	2.385	123
Ver AU	2.245	2.091	1.005	2.673	4.055	348	2.735	795	1.129	3.990	3.561	363	293	355	18	20	60	904	21	279	38
Ver AD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Tabella 9 Frequenza in termini % delle marche d'uso con distribuzione dei POS all'interno dei testi

	Coruore	Favole al telefono	Gelsomino nei paesi dei bugiardi	Harry Potter e la pietra filosofale	Il corsaro nero	Il Piccolo Principe	L'isola del tesoro	Le avventure di Alice nel paese delle meraviglie	Le avventure di Pinocchio	Marcovaldo	Sandokan alla riscossa	Cani, gatti, rospi e altre persone battucore	Emozioni del fenicottero	La lezione del minestrone	Il fantastico porcellino	I tre porcellini	Nuvola bianca e nera	Oceani di plastica	Uffa le femmine	Alla ricerca dei colori perduti	Fantasma nell'armadio
Tokens	99.546	84.543	40.085	102.507	117.761	16.923	78.748	30.733	49.819	160.551	122.153	11.997	12.067	20.084	982	856	1.630	31.308	1.438	14.128	982
Agg FO	9,29%	7,78%	7,53%	7,22%	8,30%	9,96%	8,38%	7,01%	7,67%	7,29%	8,42%	7,19%	8,78%	8,08%	8,27%	<b>12,47%</b>	5,85%	8,95%	8,35%	8,03%	8,31%
Agg AU	1,45%	1,53%	1,03%	1,34%	2,38%	1,24%	1,47%	0,76%	0,88%	1,78%	2,06%	2,10%	1,50%	0,86%	1,22%	0,22%	1,35%	1,64%	0,85%	1,42%	1,66%
Agg AD	0,30%	0,31%	0,17%	0,25%	0,37%	0,22%	0,30%	0,28%	0,33%	0,42%	0,33%	0,34%	0,28%	0,15%	0,49%	0,00%	0,45%	0,21%	0,17%	0,31%	1,19%
Nom FO	<b>27,28%</b>	24,71%	24,62%	21,56%	20,90%	22,64%	24,33%	17,60%	22,50%	26,39%	18,94%	25,67%	19,32%	18,13%	10,46%	22,27%	20,39%	24,33%	19,76%	22,80%	24,70%
Nom AU	6,40%	7,50%	6,78%	5,81%	6,67%	4,26%	5,94%	4,99%	5,72%	8,05%	7,94%	7,64%	5,91%	5,31%	14,84%	6,24%	6,75%	7,22%	6,98%	5,73%	10,93%
Nom AD	2,31%	3,47%	2,83%	2,05%	2,61%	1,51%	2,69%	2,00%	3,91%	3,89%	2,54%	3,25%	2,21%	2,08%	9,73%	0,45%	2,23%	2,16%	1,02%	1,92%	2,61%
Avv FO	9,66%	10,16%	10,94%	11,85%	8,54%	12,83%	10,05%	14,10%	12,13%	9,28%	10,40%	9,20%	11,71%	14,23%	12,65%	6,68%	13,79%	10,61%	14,48%	11,89%	8,31%
Avv AU	0,31%	0,36%	0,25%	0,39%	0,57%	0,35%	0,52%	0,45%	0,45%	0,23%	0,73%	0,45%	0,18%	0,41%	0,00%	0,00%	0,15%	0,42%	0,68%	0,21%	0,48%
Avv AD	0,35%	0,30%	0,21%	0,33%	0,42%	0,43%	0,36%	0,12%	0,07%	0,34%	0,35%	0,21%	0,16%	0,34%	0,73%	0,00%	0,60%	0,45%	0,17%	0,27%	0,20%
Ver FO	33,04%	34,92%	36,90%	39,68%	32,40%	37,21%	31,17%	<b>42,27%</b>	36,69%	31,26%	33,36%	31,90%	39,67%	<b>42,22%</b>	33,58%	27,62%	37,78%	32,87%	<b>42,42%</b>	40,92%	29,22%
Ver AU	5,56%	6,17%	6,04%	6,51%	8,84%	4,81%	8,67%	7,03%	6,00%	6,68%	8,01%	6,82%	5,93%	4,20%	4,38%	4,45%	9,00%	6,87%	3,58%	4,79%	9,03%
Ver AD	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%

Tabella 10 Frequenza delle marche d'uso FO, AU e AD nel corpus

	Cuore	Favole al telefono	Gelsomino nel paese dei bugiardi	Harry Potter e la pietra filosofale	Il corsaro nero	Il Piccolo Principe	L'isola del tesoro	Le avventure di Alice nel paese delle meraviglie	Le avventure di Pinocchio	Marcovaldo	Sandokan alla riscossa	Cani, gatti, rospi e altre persone	Emozioni batticuore	La lezione del fenicottero	Il fantastico minestrone	I tre porcellini	Nuvola bianca e nera	Oceani di plastica	Uffa le femmine	Alla ricerca dei colori perduti	Fantasma nell'armadio
FO	79,27%	77,58%	79,99%	80,32%	70,14%	82,64%	73,93%	80,98%	78,98%	74,22%	71,13%	73,96%	79,49%	82,65%	64,96%	69,04%	77,81%	76,76%	85,01%	83,65%	70,55%
AU	13,72%	15,56%	14,09%	14,05%	20,46%	10,67%	16,60%	13,22%	13,05%	16,73%	18,73%	17,01%	13,51%	10,79%	20,44%	10,91%	17,24%	16,16%	12,10%	12,15%	22,09%
AD	2,97%	4,08%	3,21%	2,63%	3,39%	2,16%	3,36%	2,40%	4,32%	4,64%	3,22%	3,79%	2,65%	2,58%	0,45%	3,30%	2,81%	1,36%	2,51%	3,80%	

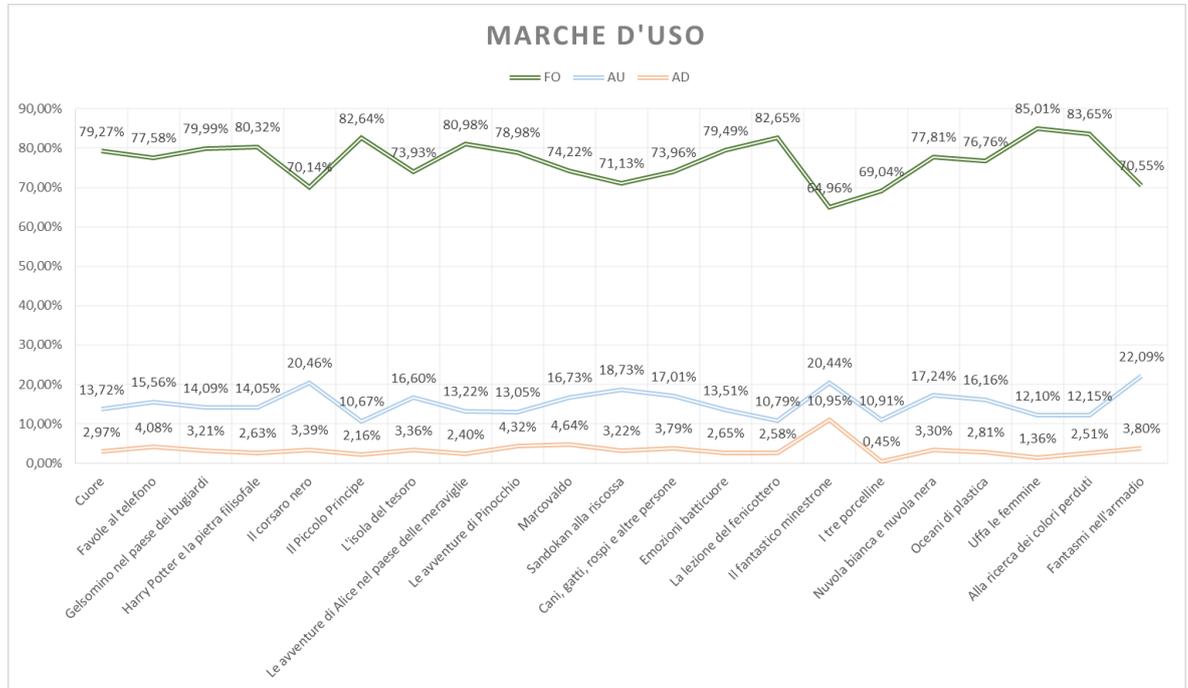


Figura 10 Diagramma delle frequenze delle marche d'uso FO, AU e AD nei testi

## 7. Indice di leggibilità automatico (ILA)

I valori ottenuti dall'applicazione delle due formule, complessità sintattica e semplicità lessicale, sono stati composti in un unico *Indice di Leggibilità Automatico* del testo, sperimentato sugli undici testi di letteratura per l'infanzia e sui dieci testi ad alta leggibilità. I risultati della sperimentazione hanno evidenziato valori coerenti con i risultati attesi. Alcuni libri sono risultati particolarmente complessi soprattutto in relazione all'aspetto sintattico (ad esempio, *Marcovaldo*) ed in generale ai testi considerati ad alta leggibilità è stato attribuito un punteggio di leggibilità particolarmente alto.

### Calcolo dell'ILA

$$L_t - S_t$$

Tabella 11 Indice di Leggibilità Automatico

	Cuore al telefono	Favole del paese dei bugiardi	Gelsomino nel paese Harry Potter e la pietra filosofale	Il corsaro nero	Il Piccolo Principe	L'isola del tesoro	Le avventure di Alice nel paese delle meraviglie	Le avventure di Pinocchio	Marcovaldo	Sandokan alla riscossa	Cani, gatti, rospi e altre persone batticuoere	Emozioni del fenicottero	La lezione del minestrone	Il fantastico porcellino	Nuvola bianca e nuvola nera	Oceani di plastica	Uffa le femmine perdute alla ricerca dei colori	Fantasmidi nell'armadio			
Lexical	0,86	0,75	0,86	0,74	0,59	0,94	0,72	0,65	0,73	0,52	0,64	0,83	0,74	0,83	0,54	0,89	0,66	0,74	0,79	0,88	0,67
Sintactic	1,04	0,75	0,74	0,71	0,82	0,47	0,98	1,06	0,90	0,94	0,54	0,58	0,42	0,58	0,64	0,04	0,21	0,71	0,54	0,32	0,50
ILA	-0,18	-0,01	0,12	0,03	-0,23	0,46	-0,25	-0,41	-0,17	-0,42	0,11	0,25	0,32	0,25	-0,10	0,85	0,45	0,03	0,26	0,56	0,17

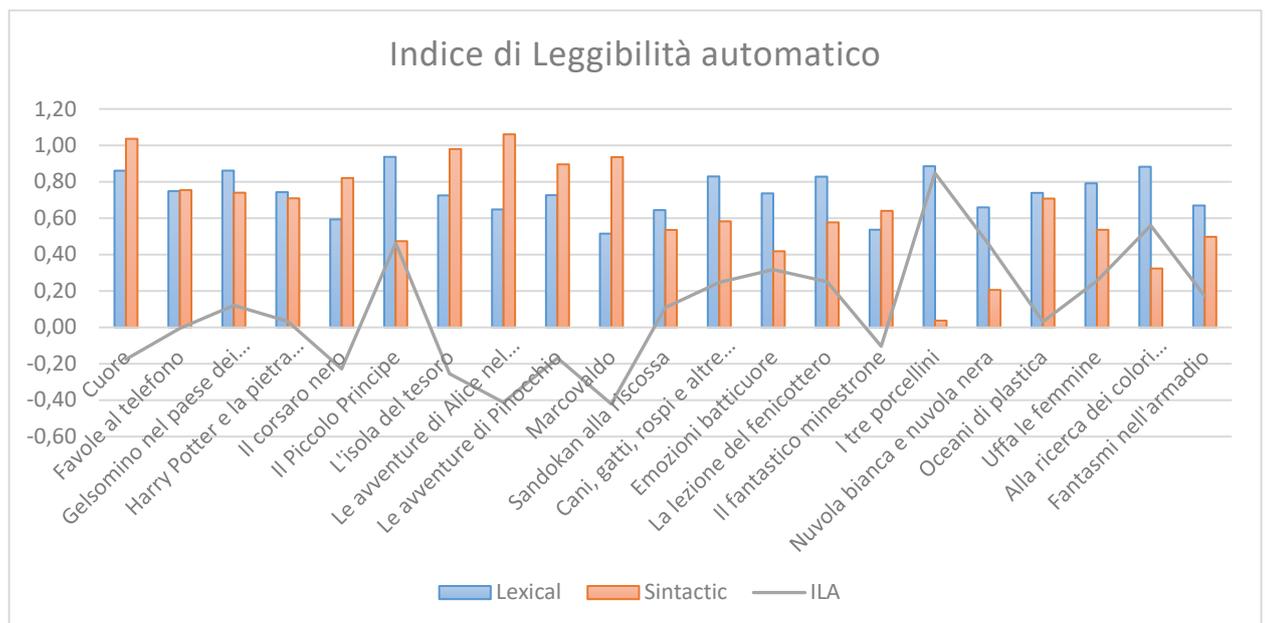


Figura 11 Diagramma dell'Indice di Leggibilità Automatica

I risultati ottenuti sono coerenti con l'ipotesi che l'alta frequenza delle parole del vocabolario fondamentale (FO) possa essere considerato un indicatore di elevata comprensibilità dei testi. In effetti, il corpus considerato per la nostra sperimentazione è formato da testi di narrativa per bambini tra i 7 e gli 11 anni ed è ipotizzabile che abbiano un'alta comprensibilità. Inoltre i dati ci permettono di evidenziare che la complessità sintattica non sia associabile a mutamenti della lingua sull'asse diacronico, ma piuttosto a differenze funzionali.

## **8. La valutazione della comprensione**

Il processo di comprensione è particolarmente complesso, come abbiamo ampiamente spiegato nel cap. II, pertanto le variabili da considerare nella valutazione di tale processo sono molteplici.

Nella nostra prova di valutazione abbiamo scelto di sottoporre agli alunni delle classi 3, 4 e 5 A della scuola primaria Don Milani di Manocalzati (AV) quattro testi da leggere, proponendo poi delle domande di comprensione.

I testi sono:

Marcovaldo Le stagioni in città Italo Calvino

1. Il bosco sull'autostrada Primavera
2. Funghi in città

Alla ricerca dei colori perduti Maria Strianese

3. Il fiume porta brutte notizie \_Sulla spiaggia non si può giocare

Gelsomino nel paese dei bugiardi

4. Gelsomino risponde all'appello, segna una rete; poi viene il bello Gianni Rodari

Ai bambini è stato permesso di rileggere il testo in modo da superare eventuali problemi di comprensione legati alla decodifica dello stesso.

Per la raccolta dei dati si è scelto di utilizzare la forma del questionario anonimo. Nel questionario veniva inoltre chiesto di indicare: la scuola e la classe di appartenenza, il sesso e l'età. La richiesta di indicare l'età serviva soltanto a confermare l'appartenenza alla classe e infatti nell'elaborazione dei dati si è tenuto conto solo della classe. I bambini intervistati avevano dunque un'età compresa tra i sette e gli undici anni.

Tra i mesi di dicembre 2021 e gennaio 2022 sono stati somministrati quattro questionari a risposta multipla a 3 classi. La scelta di domande a risposta multipla, pur essendo con molta

probabilità la forma più affidabile di item a risposta chiusa, è meno soddisfacente della risposta aperta perché limita i processi di comprensione e suggerisce una risposta al lettore che per suo conto non sarebbe stato, probabilmente, in grado di dare. Nonostante ciò abbiamo ritenuto che fosse adeguata alle nostre esigenze per due motivi fondamentali:

- la risposta aperta avrebbe potuto penalizzare quei bambini che hanno difficoltà nel linguaggio espressivo,
- la scelta multipla permette di attribuire un punteggio univoco.

L'ampiezza dell'intervento purtroppo è abbastanza limitato, a causa dei lunghi periodi di chiusura delle scuole in Regione Campania, legato alla scelta di tenere i bambini in Dad durante l'ultimo anno e mezzo a causa del COVID 19.

L'Istituto comprensivo Don Milani si trova in provincia di Avellino. I dati raccolti dunque provengono da una realtà piccola e omogenea.

Ogni classe ha avuto a disposizione un'ora per procedere alla lettura del testo e alla compilazione del questionario, tuttavia, la maggior parte degli alunni ha impiegato meno del tempo a disposizione per completare il compito. Della distribuzione dei testi, dei questionari e della spiegazione di come procedere nello svolgimento dell'attività si sono occupate le maestre prevalenti delle singole classi.

Ai bambini è stato spiegato che la prova aveva l'unico scopo di fotografare una situazione e non giudicare chi era "più bravo" o "meno bravo" e che i risultati avrebbero avuto come fine quello di validare un lavoro di ricerca avviato dall'Università degli Studi di Salerno.

Durante le somministrazioni a scuola gli atteggiamenti dei bambini sono stati di due tipi. La maggior parte ha risposto ai quesiti tranquillamente e serenamente, mentre certi hanno vissuto la prova come se si trattasse di una sorta di "verifica"; alcuni bambini, infatti, hanno sottolineato che alcuni testi erano "difficili". Per quanto riguarda le insegnanti sono state molto disponibili ed interessate alla ricerca.

Durante la compilazione le maestre sono rimaste a disposizione degli alunni per chiarimenti circa il modo di completare le parti del questionario ma non potevano intervenire a supporto degli alunni che mostravano difficoltà nella comprensione.

La prova è stata svolta da tutti i bambini, anche da due bambini diversamente abili della classe V che sono stati supportati dall'insegnante di sostegno nell'esecuzione del compito assegnato.

Tra i bambini del campione sono inoltre presenti alcuni di origine straniera, numericamente però poco rilevanti (2 in totale) e provenienti da famiglie che vivono in Italia da molti anni.

In totale sono stati raccolti 154 questionari così suddivisi: 51 per la classe terza, 36 per la classe quarta e 67 per la classe quinta.

Il grosso limite di questa sperimentazione è che la raccolta dei dati è avvenuta in un'area limitata e socialmente omogenea. Pertanto riteniamo che se la sperimentazione venisse fatta in altri contesti potremmo ottenere risultati differenti. Inoltre, il campione di 44 bambini non è sufficiente per consentire una generalizzazione dei risultati ottenuti.

È nostra intenzione proseguire con la ricerca e portare la sperimentazione innanzitutto in altre realtà geografiche aumentando il numero degli alunni coinvolti per allargare il campione valutato e verificare se ci siano o meno differenze.

## **9. Analisi quantitativa dei dati**

Dai dati raccolti tramite la prova è emersa una effettiva coerenza con i risultati ottenuti dall'analisi automatica della leggibilità effettuata con l'algoritmo da noi sviluppato.

L'analisi dei dati evidenzia una rispondenza tra il grado di comprensione dei testi sottoposti ai bambini e i risultati ottenuti dall'applicazione dell'Indice di Leggibilità Automatico al corpus. Ovvero tra i quattro brani selezionati sono risultati ostici i due tratti da "Marcovaldo e le stagioni in città".

Il nostro Indice di Leggibilità Automatico aveva individuato, infatti, in Marcovaldo il testo che con un "ILA" di  $-0,42$  risultava essere il più complesso. Se passiamo all'analisi specifica dei risultati delle prove, relativi ai due brani presentati ai bambini "Il Bosco sull'autostrada e Funghi in Città", possiamo vedere che le tre classi hanno trovato le stesse difficoltà di comprensione.

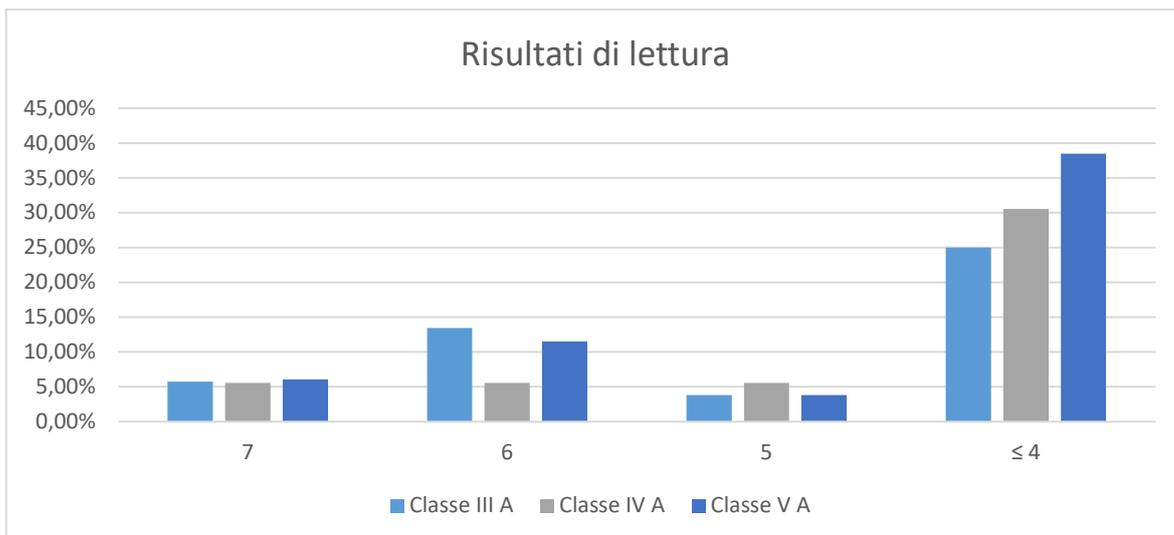


Figura 12 Diagramma dei risultati di lettura Marcovaldo “Il Bosco sull’autostrada e Funghi in Città”

I questionari delle tre classi con tutte le risposte corrette, relativi ai due brani di Marcovaldo, sono 9 (pari ad una media del 5,8%), mentre i questionari che hanno un numero di risposte corrette  $\leq 4$  sono 44 (pari ad una media del 31,34%). I questionari con 6 risposte corrette sono 15 e con 5 risposte corrette sono 6. Dalla rilevazione risulta che l’andamento dei maschi e quello delle femmine è analogo non ci sono quindi differenze legate al sesso.

La stessa coerenza risulta evidente dalla lettura dei risultati relativi ai due testi con un indice di leggibilità pari a 0,12 per “Gelsomino nel paese dei bugiardi” di Gianni Rodari e 0,55 per “Alla ricerca dei colori perduti” di Maria Strianese.

I questionari delle tre classi con tutte le risposte corrette, relativi al primo capitolo di Gelsomino nel paese dei bugiardi e di Il fiume porta brutte notizie e Sulla spiaggia non si può giocare tratti da “Alla ricerca dei colori perduti”, sono 41 (pari ad una media del 50,2%), mentre i questionari che hanno un numero di risposte corrette  $\leq 4$  sono 9 (pari ad una media del 16,67%). I questionari con 6 risposte corrette sono 18 e con 5 risposte corrette sono 13.

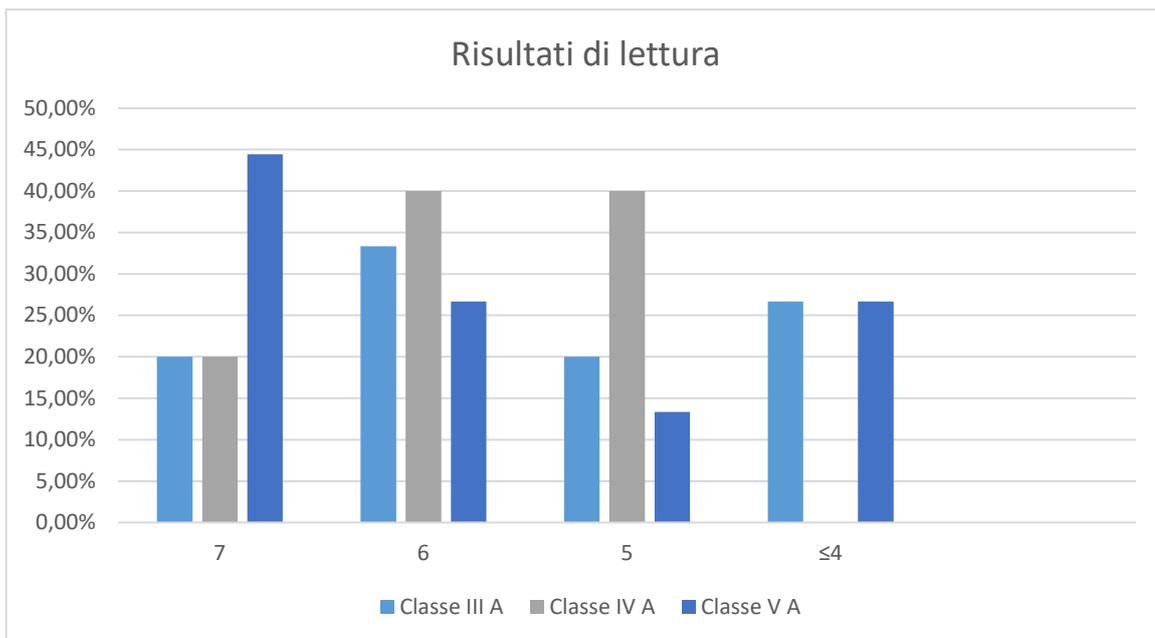


Figura 13 Diagramma dei risultati di lettura "Gelsomino nel paese dei bugiardi e di Il fiume porta brutte notizie"

Per ottenere una visione complementare dell'andamento dell'applicazione dell'indice di leggibilità automatica e della comprensione dei testi nei bambini, riportiamo due grafici che danno la misura precisa della congruità dei risultati ottenuti.

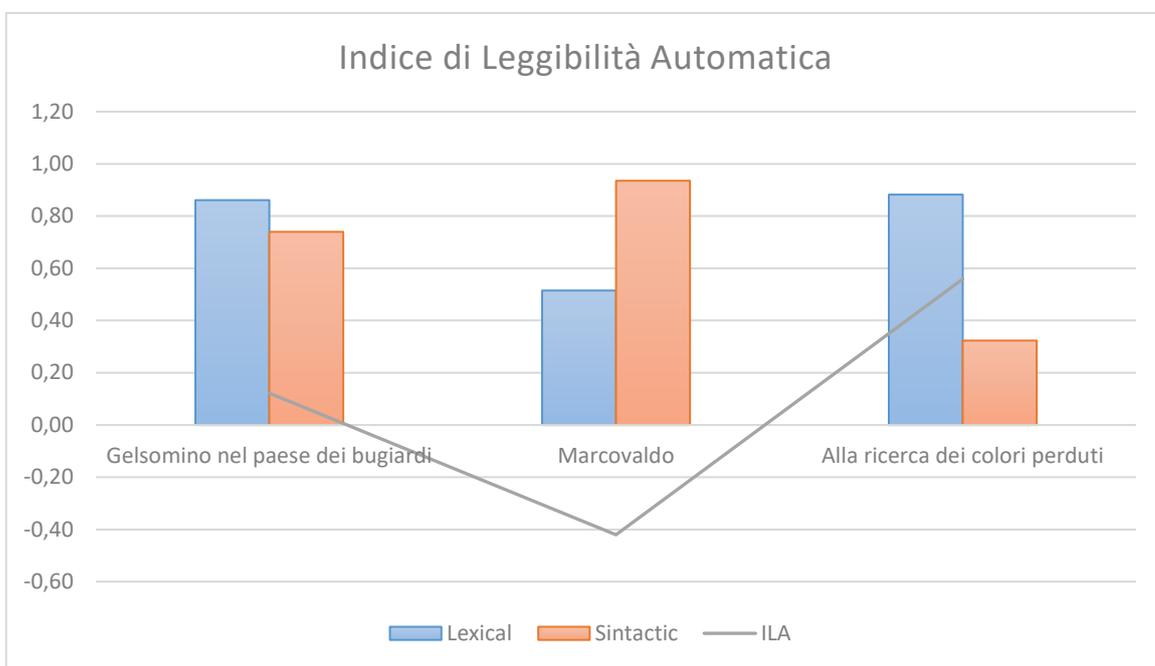


Figura 14 Diagramma dell'Indice di Leggibilità Automatica

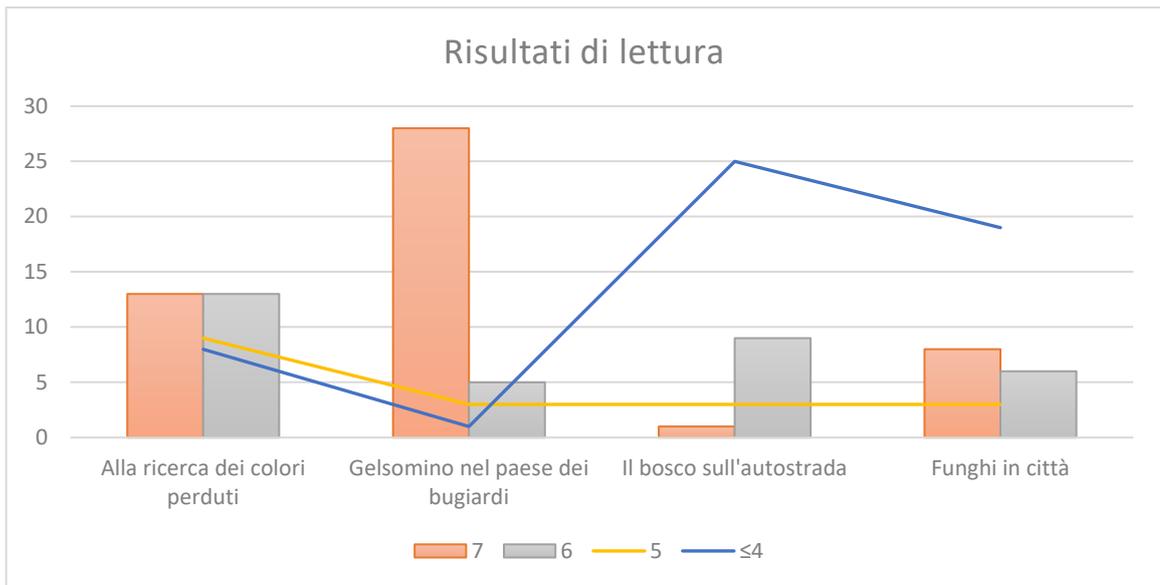


Figura 15 Diagramma dei risultati di lettura

Prendendo in considerazione i valori medi registrati nelle risposte ai quattro brani, la situazione è quella delineata dalla figura 15 che rappresenta l'andamento delle risposte ai questionari. I risultati sono evidentemente negativi per i brani tratti da Marcovaldo. Lo stesso andamento è rilevabile nella figura 14 dove l'Indice di Leggibilità Automatico è negativo nell'analisi di Marcovaldo.

Vogliamo concludere la valutazione mettendo a confronto l'analisi lessicale con l'analisi sintattica. Come si vede dalla figura 16 i tre testi presi in esame hanno un elevatissimo numero di parole con marca d'uso FO. Infatti risulta piuttosto evidente che a rendere i brani tratti da Marcovaldo più complessi non è il lessico ma la sintassi come si vede in figura 17, perché in tutti i brani somministrati la marca d'uso (FO) supera il 70% (fig. 16).

Partendo quindi da quanto affermato dallo stesso Tullio De Mauro, cioè che le parole del vocabolario fondamentale dovrebbero far parte del patrimonio lessicale di ogni persona che ha completato le scuole elementari (oggi primarie), le differenze tra le tre classi potrebbero dipendere da un maggior o minor numero di parole acquisite.

Questa affermazione è particolarmente vera se andiamo a confrontare i risultati ottenuti sui brani tratti da "Alla ricerca dei colori perduti" di Maria Strianese.

Partendo dal presupposto che questo testo fa parte dei 10 testi del corpus ad alta leggibilità e che le parole marcate FO al suo interno sono pari all'83,65% e che dall'analisi sintattica il testo ha un grado di complessità basso, pari a 0,32 le differenze tra le tre classi potrebbero essere legate all'acquisizione del lessico da parte dei bambini. Inoltre, all'interno del testo si trovano occasioni di complessità legate alle inferenze.

I bambini di terza che hanno risposto correttamente a tutte le domande sono 3 su 15, i bambini di quarta sono 2 su 10 e i bambini di quinta sono 8 su 19.

Dai dati raccolti emerge dunque che tra la classe terza e la quinta c'è un divario, a conferma del fatto che all'aumentare dell'età aumentano conoscenze lessicali e che nell'applicare le conoscenze enciclopediche i bambini tendono ad utilizzare quelle più familiari e immediatamente disponibili e solo nel corso della scuola primaria diventano progressivamente più abili nel creare inferenze (Bertolini 2012).

Tutto ciò ci fa presumere che nonostante le tre classi abbiano fatto rilevare nel complesso una omogeneità nella comprensione dei testi resta la difficoltà della riflessione metalinguistica che è un processo complesso, che si sviluppa con tempistiche diverse in ogni bambino e comunque molto dopo l'acquisizione linguistica vera e propria.

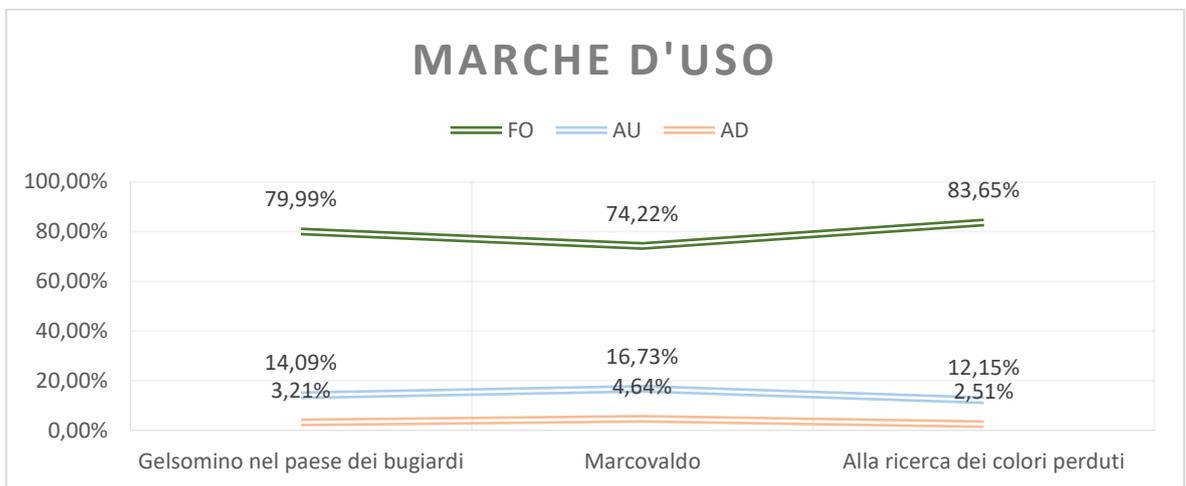


Figura 16 Marche d'uso FO AU AD

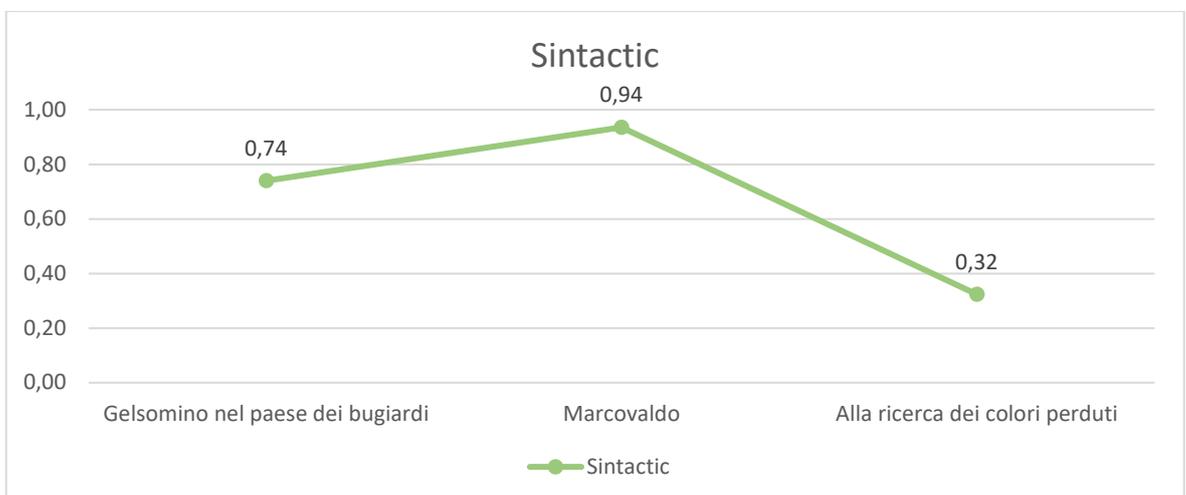


Figura 17 Analisi sintattica

## Conclusioni

Dall'indagine condotta emergono per lo più conferme delle premesse teoriche da cui siamo partiti, infatti i libri analizzati hanno una leggibilità proporzionale alla percentuale di termini appartenenti al lessico fondamentale, o comunque rientranti nel NVdB. I dati che vengono fuori da questa analisi ci mostrano come il corpus ci dia un feedback sulla frequenza delle parole del vocabolario fondamentale (FO) completamente coerente con la nostra ipotesi di partenza.

È stato verificato il carattere incrementale del lessico: nei tre livelli scolastici presi in considerazione si è assistito ad un aumento della comprensione almeno per quello che riguarda i testi ad alta leggibilità. Questo conferma che all'aumentare dell'età aumentano anche le competenze lessicali dei bambini ed in particolare lo stacco è maggiore tra la terza e la quinta primaria. Tuttavia, almeno per ciò che concerne la sperimentazione in classe, tutto quanto rilevato va considerato come una tendenza non generalizzabile poiché la ricerca è stata condotta per un numero limitato di bambini testati, appartenenti ad un unico contesto scolastico.

A noi sembra doveroso sottolineare che è necessario assicurarsi che la conoscenza del nucleo di parole che formano la struttura portante del 98% degli enunciati sia acquisita dai parlanti nativi, cioè che siano acquisite le 7500 parole del NVdB che possono essere considerate di base, cioè comprensibili a un vasto pubblico, anche con livelli di scolarizzazione non elevati. È innanzitutto a questa fascia di lessico, e in particolare alle parole fondamentali, che va dedicata la massima attenzione, anche e soprattutto per creare progetti di apprendimento.

... "occorre una massa parlante perché vi sia una lingua.

Contrariamente all'apparenza, in nessun momento storico la lingua esiste fuori dal fatto sociale....

È dunque fondamentale lavorare sull'acquisizione del lessico sin dalla scuola primaria, lavorando sulle tre dimensioni: ampiezza, profondità e automaticità.

Non solo gli insegnanti di italiano devono porre tra i loro obiettivi il lessico, ma anche quelli che insegnano le altre discipline, perché ogni argomento offre nuove parole. Conoscere le parole del NVdB è importante non solo per comprendere il significato delle singole parole, ma anche perché attraverso le parole presenti nel NVdB è possibile spiegare il significato di parole più complesse.

Dalle considerazioni fatte in questo lavoro emerge il ruolo di grande rilievo che il lessico riveste all'interno della lingua ed è per questo che ci sembra auspicabili l'utilizzazione di sistemi di misurazione lessicale che supportino le case editrici nella redazione di testi scolastici e non.

Fra i fattori che concorrono alla leggibilità/comprendibilità di un testo, quindi, alcuni riguardano il lettore, altri riguardano il testo (lessico e sintassi).

Lucisano mette in evidenza come le formule siano insufficienti a mettere in risalto le abilità del lettore e in questa direzione si muovono altri studiosi come Lucia Lumbelli con le sue ricerche sulla comprensibilità dei testi.

Altri indici, nel corso degli anni hanno considerato semplicità lessicale e complessità sintattica e tal proposito Maria Emanuela Piemontese afferma che chi produce un testo deve:

- avere chiari gli obiettivi e i destinatari del testo;
- preferire parole di uso comune (vocabolario di base);
- usare una sintassi semplice, priva di proposizioni subordinate e a incastro;
- scrivere frasi brevi (15-20 parole);
- improntare le scelte grafiche e tipografiche alla massima leggibilità del testo.

Nella produzione del nostro indice ci siamo attenuti a questi criteri infatti è stata individuata una fascia d'età precisa nella scelta dei lettori a cui indirizzarsi, l'analisi lessicale ha posto le sue basi sull'applicazione di dizionari elettronici arricchiti con le marche d'uso dell'NVdB e l'analisi sintattica ha preso in considerazione alcune feature sintattiche quali la presenza di subordinate, la linearità dei periodi, la presenza di modificatori ed incisi.

L'applicazione degli indicatori sintattici riportati di seguenti:

1.  $V_{mod}$  che include frasi subordinate (soggettive, preposizionali, infinitive, relative, avverbiali),
2.  $V_{punt}$  che evidenzia la presenza di coppie di segni di punteggiatura interne alla frase,
3.  $V_{ordine}$  che prevede un ordine di frase diverso da Soggetto-Verbo-Complemento ed evidenzia la presenza di soggetti non riconducibili ad un Sostantivo;

ci ha permesso di evidenziare come prendendo in esame un testo composto da un elevato numero di frasi subordinate ( $V_{mod}$ ) l'indicatore di complessità aumenti in modo significativo.

Possiamo concludere pertanto che un testo con un punteggio di leggibilità alto dato dal nostro Indice di leggibilità automatico (ILA), derivante dal rapporto tra la semplicità lessicale e la complessità sintattica, è un testo con una buona comprensibilità.

## Appendice

TESTI	AUTORI	I ED	TRAD.	EDIZIONE
Le avventure d' Alice nel Paese delle Meraviglie	Lewis Carroll (pseudonimo di Charles Lutwidge Dodgson)	1865	1870	Traduzione di Teodorico Pietrocòla-Rossetti, illustrazioni di John Tenniel, Londra, Macmillan and co., 1872. Wikisource
L'isola del tesoro	Robert Stevenson	1883	1883	Traduzione di Angiolo Silvio Novaro (1932) Newton Compton, 2013 Wikisource
Le Avventure di Pinocchio	Carlo Collodi (pseudonimo Carlo Lorenzini)	1883		Firenze R. Bemporad & Figlio 1892 Wikisource
Il Corsaro Nero	Emilio Salgari	1898		a cura di Sergio Campailla; Biblioteca economica Newton Sezione ragazzi, 37; Roma, 1996 Wikisource
Sandokan alla riscossa	Emilio Salgari	1907		Fabbri Editori, 2002 Wikisource <a href="http://www.emiliosalgari.it/testi/testionline.htm">http://www.emiliosalgari.it/testi/testionline.htm</a>
Cuore	Edmondo De Amicis	1886		Milano Fratelli Treves, editori 1889 Wikisource
Il Piccolo Principe	Antoine de Saint-Exupéry	1943	1949	Edizione bilingue: Testo originale francese di Antoine de Saint-Exupéry, versione in italiano a cura di Franco Perini. Illustrazioni di Antoine de Saint-Exupéry. Prima edizione in formato ePub 2016
Gelsomino nel paese dei bugiardi	Gianni Rodari	1959		Illustrazioni Alberto Ruggieri Editori Riuniti 2003
Favole al telefono	Gianni Rodari	1962		Illustrazioni di Francesco Altan

				Einaudi Ragazzi 1993
Marcovaldo Le stagioni in città	Italo Calvino	1963		(Autore, Prefazione), Domenico Scarpa Oscar Moderni 2016
Harry Potter e La Pietra Filosofale	Rowling, Joanne Kathleen	1997	1997	Traduttore Marina Astologo Salani Editore 2001
Cani, gatti, rospi e altre persone	Paola Valente	2019		Parole leggere - Il Mulino a Vento- Gruppo editoriale Raffaello
Fantasmii nell'armadio	Lorenza Farina	2019		Attacca Parole - Gruppo editoriale Raffaello
Emozioni batticuore	Valeria Conti	2019		Serie Gialla Il Mulino a Vento- Gruppo editoriale Raffaello
La lezione del fenicottero	Germano Chiaverini	2019		Serie Rossa Il Mulino a Vento- Gruppo editoriale Raffaello
Il Fantasma Minestrone	Guido Quarzo	2018		Attacca Parole - Gruppo editoriale Raffaello
I tre Porcellini	narrata da Raffaella Magi	2019		Parole leggere - Il Mulino a Vento- Gruppo editoriale Raffaello
Nuvola bianca e nuvola nera	Annamaria e Nice Piccione	2015		Serie Gialla Il Mulino a Vento- Gruppo editoriale Raffaello
Oceani di plastica	Beatrice Peruffo	2019		Avventurose Scienze - Gruppo editoriale Raffaello
Uffa le femmine , uffa i maschi	Francesca Mascheroni	2016		Serie Gialla Il Mulino a Vento- Gruppo editoriale Raffaello
Alla ricerca dei colori perduti	Maria Strianese	2009		Serie Rossa Il Mulino a Vento- Gruppo editoriale Raffaello

## Bibliografia

- Ausubel D.P. (1978) In Defense of Advance Organizers: A Reply to the Critics Review of Educational Research Spring 1978, Vol. 48, No. 2, Pp. 251-257.
- Barbagli A., Lucisano P., Dell'Orletta F., Montemagni S. e Venturi G. (2015) "Il ruolo delle tecnologie del linguaggio nel monitoraggio dell'evoluzione delle abilità di scrittura: primi risultati", Italian Journal of Computational Linguistics, Accademia University Press, Torino, pag. 99-117.
- Barbagli A., Lucisano P., Dell'Orletta F., Montemagni S. e Venturi G. (2015) Tecnologie del linguaggio e monitoraggio dell'evoluzione delle abilità di scrittura nella scuola secondaria di primo grado Il Sileno / Filosofi(e)Semiotiche 2, 2015
- Bertolini C. (2012) Leggere insieme per comprendere un testo: descrivere e valutare i processi comunicativi in Giornale Italiano della Ricerca Educativa pp. 13-24.
- Bertolini C. (2012) Senza parole. Promuovere la comprensione del testo fin dalla scuola dell'infanzia Editore: Junior Collana: Ricerche.
- Bertolini C., Cardarello R. (2009) L'abilità inferenziale nel processo di comprensione di testi verbali e di testi visivi in Giornale Italiano della Ricerca Educativa pp. 9-17.
- Bisconti V. (2012), La svolta lessicografica di Tullio De Mauro e i dizionari contemporanei Chroniques Italiennes, série web, n° 23/2.
- Boero P., De Luca C. (1995) La letteratura per l'infanzia Editori Laterza
- Bolasco S. (2003) L'analisi statistica dei dati testuali: intrecci problematiche e prospettive. Giornata di studio Applicazioni di analisi testuale.
- Bolasco S. (2005) Statistica linguistica e text mining: alcuni paradigmi applicativi Quaderni di statistica Liguori Vol. 7.
- Bosc F., Poesia e lessico nella scuola primaria: primi risultati di un progetto Italiano LinguaDue, n.1. 2019.
- Bottero E. (2014) *Il metodo di insegnamento. I problemi della didattica nella scuola di base*, Franco Angeli, Milano.
- Bruner J. S. (1987) Il linguaggio del bambino. Come il bambino impara a usare il linguaggio Armando Editore
- Cardarello R., Contini R. (2012) Parole immagini metafore. Per una didattica della comprensione Editore: Junior Collana: Ricerche.
- Cardarello R., Lumbelli L. (2019) La comprensione del testo Attività su brani d'autore per le classi IV e V primaria Giunti Editore Srl
- Cardona M. (2008) L'abilità di lettura e lo sviluppo della competenza lessicale Studi di Glottodidattica 2008, 2, 10-36.

Carloni F. (2000) Le relazioni statistiche tra frequenza e significato delle parole nella lingua italiana Source: Italica, Vol. 77, No. 4, Linguistics and Pedagogy , pp. 523-534 Published by: American Association of Teachers of Italian.

Cornoldi C., Colpo G. Carretti B. (2017) Prove MT – KIT Scuola Classe 3 – 4 – 5 primaria Giunti EDU

Chiari I. (2005) Linguistica e informatica: la linguistica dei corpora in Italia Bollettino di Italianistica - Roma Carrocci pag. 101-118

Chiari I. (2007) Introduzione alla linguistica computazionale Edizione La terza.

Chiari I. (2008) “La chiave probabilistica delle lingue: teoria linguistica e applicazioni computazionali, in L. Fulci e E. Sciubba (a cura di), Linguaggio, Mente e Società, EuRoma-La Goliardica, Roma, pp. 55-79.

Chiari I. (2012) The new basic vocabulary of Italian: problems and methods in Statistica Applicata (Rocco Curto, Napoli) pp. 21-35

Chiari I. (2018) Economia e ridondanza: il contributo demauriano nel quadro della riflessione novecentesca Bollettino di italianistica n.s., anno XV, n. 2.

Chiosso G. (2018) La pedagogia contemporanea Editrice La scuola

Chomsky N. (1980) Regole e rappresentazioni Dalai Editore.

Chomsky N., Berwick R. C. (2016) Perché solo noi. Linguaggio ed evoluzione, Bollati Boringhieri

D’Agostino E., Elia A., Martinelli M. (1981) Lessico e strutture sintattiche Introduzione alla sintassi del verbo italiano Liguori editore

D’Agostino E. (1998), Il lessico di frequenza dell’italiano parlato e la didattica dell’italiano Quaderns d’Italia 3.

De Mauro T. (1961), “Statistica linguistica”, in Enciclopedia italiana. Appendice III, 2 vol., Roma: Istituto della Enciclopedia Italiana, II, pp. 820-821.

De Mauro T. (1980), “Il vocabolario di base della lingua italiana”, in Guida all’uso delle parole, Roma: Editori Riuniti, pp. 146-172.

T. De Mauro Prefazione a G. Rodari. Il Gatto viaggiatore e altre storie a cura di C. De Luca L’Unità Editori Riuniti Roma 1990.

De Mauro T. (1994), Capire le parole, Roma-Bari: Laterza (2a ed. 2002, citeremo da questa edizione).

De Mauro T., Chiari I. (2005) Parole e numeri. Analisi quantitative dei fatti di lingua Aracne

De Mauro T., Chiari I. (2014) The New Basic Vocabulary of Italian as a linguistic resource First Italian Conference on Computational Linguistics CLiC-it 2014

- De Mauro T., Mancini F., Vedovelli M. & Voghera M. (1993), *Lessico di frequenza dell'italiano parlato (LIP)*, Milano: Etaslibri.
- Dell'Orletta F., Montemagni S., Venturi G. (2011), *READ-IT: Assessing Readability of Italian Texts with a View to Text Simplification Proceedings* pag. 73–83, Edinburgh, Scotland, UK, July 30, 2011, Association for Computational Linguistics.
- Denes G. (2009) *Parlare con la testa. Le basi neurologiche e la struttura del linguaggio* Zanichelli
- Elia A. (1984) *Le verbe italiani*, Schena Editore
- Elia A. (2005) *Lessico Grammatica dell'italiano*. In T. De Mauro; I. Chiari; a Cura di Parole e numeri. *Analisi quantitative dei fatti di lingua* Pag.157-170 Roma Aracne editrice.
- Elia A. (2012) *Lessico e sintassi tra tempo e massa parlante*. In: *Il lessico nella teoria e nella storia linguistica* Pisa Il Calamo Pag.15-47
- Elia A., Maisto A., Melillo L., Pelosi S. (2021) *Lexical complexity and basic vocabulary of the Italian language* In *Formalising Natural Languages: Applications to Natural Language Processing and Digital Humanities 14th International Conference, NooJ 2020, Zagreb, Croatia, June 5–7, 2020, Revised Selected Papers*
- Elia A; Monteleone M.; Vellutino D.; Bocchino F.; Langella A.M. (2010) *Grammatiche locali per il riconoscimento automatico e la classificazione delle FAQ sull'Informazione Comunitaria Europea*. In Bolasco, S., Chiari, I., Giuliano, L. *Statistical Analysis of Textual Data, Proceedings of 10th International Conferences "Journées D'Analyse Statistique des Données Textuelles"* .
- Ferreri, S. *L'alfabetizzazione lessicale. Studi di linguistica educativa*, Roma, Aracne, 2005 pag. 1-40
- Ferretti F. (2014) *I sassi nello stagno di Noam Chomsky il Manifesto Cultura Edizione del 19 Maggio 2019*
- Gallina F. (2017) *Il Nuovo Vocabolario di Base e il Lessico degli Stranieri: Una Prima Ricognizione Quantitativa. Studi Italiani di Linguistica Teorica e Applicata*, anno XLVI, 2017, numero 1337.
- Graffi G. (2008) *Che cos'è la grammatica generativa* Carocci Le bussole
- Graffi G. (2019) *Origin of language and origin of languages* John Benjamins Publishing Company pp 6 - 23
- Graffi G. (2019) *The Piaget-Chomsky debate, forty years later. A retrospective evaluation and some open issues* Il Mulino – Rivisteweb Fascicolo 1, gennaio-aprile 2019 pp 53 - 71
- Jafrancesco E. (2002) *L'abilità di lettura: leggibilità di un testo e proposte di facilitazione* Didattica & Classe Plurilingue, numero 1, aprile-maggio 2002
- Lavinio C. (2000), *Tipi testuali e processi cognitivi*, in *Didattica ed educazione linguistica*, a cura di F. Camponovo & A. Moretti, Firenze, La Nuova Italia, pp. 125-144.
- Laudanna A., Voghera M. *Il linguaggio Strutture linguistiche e processi cognitivi* 2006, Gius. Laterza & Figli

- Lumbelli L. (2009) *La comprensione come problema. Il punto di vista cognitivo* Laterza
- Lumbelli L. (2012) *Condizioni cognitive di una lettura autonomamente motivata* ANNO 8/ N. 20 - Ottobre 2012 - *La lettura durante l'intero arco della vita*
- Marconi D. (2009) *Linguaggio e Scienza Cognitiva* Treccani
- Marconi L., Ott M., Pesenti E., Ratti D., Tavella M. (1993) *Lessico Elementare Dati Statistici sull'Italiano scritto e letto dai bambini delle elementari* - Zanichelli
- Marzano A., Vegliante R. (2014) *La comprensione del testo orale: percorsi didattici e di stimolazione con la lavagna interattiva (LIM)* *Giornale Italiano della Ricerca Educativa* –anno VII – numero 13 pp 151 – 165
- Massa E. (2016) *Sull'unità di significato della lingua parlata: alcune osservazioni su base lessico-statistica* In *Parlare le lingue romanze* Camilla Bardel and Anna De Meo. Naples
- Mastidoro N., Amizzoni M. (1993), *Linguistica applicata alla leggibilità: considerazioni teoriche e applicazioni* *Bollettino della Società Filosofica Italiana*, 1993.
- Mocci S., Penna M.P. (2012), *Una prospettiva vygotkijana per un modello unificato di comprensione del testo* *Rivista italiana di filosofia del linguaggio* vol. 6 n.2 pag. 137-148.
- Montessori M. (2017) *Psico-grammatica. Dattiloscritto inedito* Curatore: C. Tornar, G. Honegger Fresco Franco Angeli - *Ricerche sperimentali-Monografie educ.*
- Montessori M. (1916) *L'autoeducazione* Garzanti editore
- Piaget J. (1974) *Dove va l'educazione* a cura di A. Granese Armando Editore
- Piemontese M.E (1991) *Scrittura e leggibilità: «Due parole»*. In Michele A. Cortelazzo (a cura di) *Scrivere nella scuola dell'obbligo*, Quaderni del Giscel, La Nuova Italia, Firenze, 1991, pp. 151-167.
- Piemontese M.E (1995) *Il Lip: uno strumento per la didattica della lingua italiana in Italia e all'estero* *Italica* Vol. 72, No. 4, pp. 474-487 In: *American Association of Teachers of Italian*
- Piemontese M.E (2005) *Misurazioni quantitative degli stili verbali e indici di leggibilità* . In T. De Mauro; I. Chiari; a Cura di *Parole e numeri. Analisi quantitative dei fatti di lingua* (pag. 377-398). Roma Aracne editrice.
- Pinker S. (1998) *L'istinto del linguaggio* Mondadori - Oscar saggi
- Quintarelli E. (2015) *Comprensione del testo con le sequenze temporali. Storie e attività per bambini da 6 a 8 anni* Erickson.
- Sigman M. (2017) *La vita segreta della mente* Dea Planeta libri
- Sinclair J. (1991), *Corpus, Concordance, Collocation*, Oxford University Press, Oxford, UK.
- Sobrero A. (2009), *L'incremento della competenza lessicale, con particolare riferimento ai linguaggi scientifici*, *Italiano LinguaDue*, 1, 2009.

- Tognini Bonelli E. (2001) *Corpus linguistics at work* John Benjamins Publishing, 2001 - Language Arts & Disciplines
- Varisco, B.M. (1995). Paradigmi psicologici e pratiche didattiche con il computer. In *Tecnologie Didattiche*, n. 7.
- Venturi G., Dell'Orletta F., Montemagni S., Flore E. e Bellandi T. (2017) "La qualità dei consensi informati. Un'analisi linguistico-computazionale della leggibilità dei testi", *Salute e territorio*, ETS, Pisa, vol. 212, pag. 35-39.
- Vietri S. (2001) *Navigare nei testi. Applicazioni in linguistica computazionale* Edizione scientifica italiana
- Vietri S. (2004) *Lessico grammatica dell'italiano* Utet Libreria
- Vietri S. (2008) *Dizionari elettronici e grammatiche a stati finiti. Metodi di analisi formale della lingua italiana* Editore Plectica.
- Vietri S., Elia A. (2001b). *Analisi automatica dei testi e dizionari elettronici*. In Burattini, E. and Cordeschi, R. (2001). *Intelligenza artificiale*. Roma: Carocci.
- Vygotskij Lev S. (1956), *Pensiero e linguaggio* Giunti Editore - Psicologia.
- Zambelli L. (2014), *Semplificare i testi di studio: quando, come* Italiano LinguaDue, 1. 2014.
- Zipf J. K. (1949), *Human Behaviour and the Principle of Least-Effort, An Introduction to Human Ecology*, Cambridge, MA: Addison-Wesley Press.

## Allegato 1: Prova di comprensione

Età \_\_\_\_\_ Sesso F

M

Scuola \_\_\_\_\_ Classe \_\_\_\_\_

Data

Prova \_\_\_\_\_

### Sintesi della prova

Numero di risposte

corrette \_\_\_\_\_

Fascia di

prestazione \_\_\_\_\_

## Allegato 2: Questionario relativo all'attività di lettura

Marcovaldo Le stagioni in città

Italo Calvino

Inverno Il bosco sull'autostrada

### Domande

1. Quando Marcovaldo uscì di casa mandava fruscii cartacei ad ogni passo. A che cosa erano dovuti i fruscii cartacei?
  - A. A quattro o cinque giornali infilati sotto la giacca.
  - B. A un mazzo di fogli che ha in mano.
  - C. Alla sega che ha sotto al cappotto.
  - D. Ai colpi d'aria.
2. Perché gli alberi del bosco trovato dai bambini di Marcovaldo sono giudicati strani?
  - A. Perché i bambini non avevano mai visto un bosco.
  - B. Perché quegli alberi si trovavano vicino ad una autostrada.
  - C. Perché i bambini avevano molta fantasia.
  - D. Perché i rami erano a forma di dentifricio, di bottiglia o di pneumatico.
3. Marcovaldo chiede ai figli: "Dove lo avete preso?" A che cosa si riferisce?
  - A. Ai pezzi di alberello a forma di primula che bruciavano nella stufa.
  - B. Ai pezzi di alberello a forma di primula che non erano ancora stati bruciati.

- C. Al fuoco che riscalda la stanza.  
D. Al fascio di rami che bruciava nella stanza.
4. E' veramente un bosco quello che trovano i figli di Marcovaldo?
- A. Si  
B. No
5. Quello che i figli di Marcovaldo trovano ai lati dell'autostrada che cosa era in realtà?
- A. Un vero bosco.  
B. Un bosco di abeti.  
C. Un insieme di cartelloni pubblicitari.  
D. Uno strano bosco.
6. Chi sta buttando giù i cartelloni pubblicitari quando Astolfo parte per l'ispezione?
- A. Un gruppo di bambini.  
B. I figli di Marcovaldo.  
C. Marcovaldo.  
D. Un signore con gli occhiali.
7. Come mai Astolfo scambia Marcovaldo con un cartellone pubblicitario?
- A. Astolfo aveva bisogno degli occhiali ma non lo diceva.  
B. Astolfo era abbagliato dalla luce.  
C. Astolfo voleva evitare di ripetere l'errore che aveva già fatto due volte.  
D. Marcovaldo era riuscito a restare perfettamente immobile.

Marcovaldo Le stagioni in città

Italo Calvino

Primavera Funghi in città

Domande

1. Che cosa nota Marcovaldo un giorno vicino alla fermata del tram?
  - A. Delle primule gialle.
  - B. Dei funghi.
  - C. La base degli alberi.
  - D. Dei bernoccoli.
2. Come mai solo Marcovaldo si accorge dei funghi che spuntano alla fermata del tram?
  - A. A differenza degli altri cittadini era sensibile alle cose naturali.
  - B. Erano poco visibili, si vedevano solo dei piccoli bernoccoli.
  - C. I funghi erano ancora in gran parte sotterranei.
  - D. Gli altri cittadini guardavano verso il tram che stava arrivando.
3. Perché Marcovaldo non vuole dire ai bambini dove ha visto i funghi?
  - A. I bambini avrebbero potuto danneggiarli giocandoci.
  - B. Voleva mostrare loro personalmente la bellezza dei funghi che aveva trovato.

- C. Non voleva che altri venissero a saperlo e gli portassero via i funghi.
- D. Temeva che ai bambini capitasse qualche incidente.
4. Sabato durante la notte piove e Marcovaldo si sveglia e balza di gioia. Perché?
- A. Perché la pioggia farà crescere i funghi.
- B. Perché la pioggia fa bene al terreno.
- C. Perché respiro l'odore di polvere bagnata e muffa fresca.
- D. Perché a Marcovaldo piace la pioggia.
5. Perché lo spazzino Amadigi dice che, se Marcovaldo raccoglie i funghi, vuol dire che sono buoni da mangiare?
- A. Sa che Marcovaldo ha una particolare preparazione in fatto di funghi.
- B. Sa che Marcovaldo conosce bene le cose naturali.
- C. Sa che Marcovaldo ci vede molto meglio di lui.
- D. Sa di essere molto ignorante in generale.
6. Mentre Marcovaldo e i bambini raccolgono i funghi chi incontrano?
- A. Amadigi anche lui con un cesto pieno di funghi sotto il braccio.
- B. I parenti di Amadigi.
- C. La gente che a quell'ora stava aspettando il tram.
- D. Gente con l'ombrello appeso al braccio
7. Anche in questo brano Marcovaldo mostra un cambiamento di stato d'animo. Quale?

- A. Sull'entusiasmo per i funghi prevale lo stupore nel vedere che altre persone li avevano trovati.
- B. Passa dall'ammirazione per i funghi che aveva trovato all'ostilità per Amadigi.
- C. Passa dalla smania di possesso alla generosità.
- D. In un primo tempo propone un pranzo collettivo, ma poi preferisce mangiare a casa sua.

## Gelsomino nel paese dei bugiardi

Gelsomino risponde all'appello, regna una rete; poi viene il bello

Gianni Rodari

Capitolo Primo

### Domande

1. Che cosa fa il narratore per difendersi dalla voce di Gelsomino?
  - A. Mette nelle orecchie un mezzo chilogrammo di bambagia.
  - B. Si tappa le orecchie con le mani.
  - C. Ascolta gli aeroplani a reazione per abituarsi.
  - D. Non ha problemi ad ascoltarlo.
2. Per quale motivo quando Gelsomino nacque la gente del paese si alzò nel cuore della notte?
  - A. La gente era felice perché era nato un nuovo bambino e tutti volevano festeggiare.
  - B. Il pianto di Gelsomino era così forte che la gente pensò che fosse il suono delle sirene delle fabbriche.
  - C. Le sirene delle fabbriche cominciarono a suonare tutte insieme per segnalare la nascita di un nuovo bambino.
  - D. Per festeggiare una nuova nascita, la gente andava a lavorare molto prima del solito.

3. A che ora Gelsomino dava il primo Strillo al mattino?
- A. Il primo strillo squillava alle sette in punto.
  - B. Strillava per l'intera notte.
  - C. Alle otto del mattino.
  - D. Non strillava mai.
4. Perché, quando il maestro ha fatto l'appello, la lavagna è andata in mille pezzi e i vetri della finestra sono caduti nel cortile?
- A. Qualcuno da fuori la classe ha tirato delle pietre.
  - B. Il maestro ha colpito per errore la lavagna e la finestra con la bacchetta.
  - C. La classe è vecchia e andrebbe riparata dai bidelli.
  - D. La voce di Gelsomino è così forte che rompe le cose.
5. Gelsomino, tornato in sé, capì di averla fatta grossa: che cosa ha capito Gelsomino?
- A. Che, scaldandosi per la partita, aveva fatto il tifo per la squadra avversaria.
  - B. Che si è distratto proprio mentre si svolgeva un'azione importante.
  - C. Che la squadra del paese sta vincendo meritatamente.
  - D. Che è stata la potenza della sua voce a spingere la palla in rete.

Alla ricerca dei colori perduti

Maria Strianese

Il fiume porta brutte notizie

Sulla spiaggia non si può giocare

### Domande

1. Che cosa emergeva dall'acqua bassa del fiume?
  - A. La carcassa arrugginita di una lavatrice, un grande bidone di plastica.
  - B. Il Gatto Armonico.
  - C. Pesci quizzanti.
  - D. Alberi, arbusti e fiori.
2. Su cosa indagavano Luca e Silvia?
  - A. Su chi ha buttato i rifiuti nel fiume.
  - B. Su cosa ha reso il fiume sporco e giallo.
  - C. Sul mal di pancia dei pesciolini.
  - D. Sull'uso del sapone.
3. Come si può risolvere il problema del fiume?
  - A. Buttando i rifiuti nell'acqua.
  - B. Usando molto sapone per lavarsi e fare il bucato.
  - C. Versando l'olio nel lavandino.
  - D. Evitando di usare tanto sapone e non gettando liquidi pericolosi nel lavandino.
4. Perché la spiaggia non piace a Silvia?

- A. Perché fa troppo caldo.
  - B. Perché la sabbia era troppo bagnata.
  - C. Perché la spiaggia era coperta di immondizia e il mare aveva un odore malsano.
  - D. Perché correva il rischio di inciampare.
5. Che cosa crea il catrame che arriva sulla spiaggia?
- A. Le navi petroliere che scaricano in mare il residuo del carico trasportato
  - B. La spazzatura lasciata sulla spiaggia.
  - C. La spazzatura gettata in mare e spinta a riva dalle onde.
  - D. Le navi dei pescatori.
6. Quale è il comportamento giusto per proteggere la spiaggia?
- A. Non buttare bottiglie, lattine e cartacce dove capita.
  - B. Portare con sé bicchieri, piatti e posate di plastica.
  - C. Usare sacchetti di plastica.
  - D. Sporcando solo dove ci sono i bagnini che puliscono
7. I delfini e i gabbiani sono danneggiati dal comportamento sbagliato degli uomini?
- A. Sì.
  - B. No.
  - C. A volte.
  - D. No perché viene danneggiato solo l'ambiente.