

# LEGGIBILITÀ E COMPRENSIBILITÀ: UN BINOMIO NECESSARIO

Debora Vena<sup>1</sup>

## 1. INTRODUZIONE

In questo contributo si presentano i primi risultati di un lavoro esplorativo di calcolo e analisi dell'indice di leggibilità GULPEASE (1988) sui testi contenuti nel *corpus* del progetto *Leggere e Comprendere* (LeCo) pubblicato sul sito <https://studiumdipsum.it/>, che raccoglie gli strumenti di *e-learning* prodotti dai docenti del Dipartimento di Studi Umanistici (DIPSUM) dell'Università di Salerno. Il *corpus* LeCo è il frutto di due progetti più ampi, L.A.M.P. (2016) e Alpha-Mente (2021), che si sono focalizzati sul recupero e il potenziamento delle abilità di lettura e comprensione per monitorare e valutare le competenze nella *literacy* degli studenti del biennio della scuola secondaria di secondo grado (cfr. Voghera *et al.* in questa monografia).

L'ipotesi di partenza è quella di misurare e valutare il legame necessario, e per nulla scontato, tra la leggibilità e la comprensibilità dei testi nel complesso processo della comprensione. L'obiettivo è porre maggiore attenzione alla reale leggibilità dei testi, presenti nel *corpus* LeCo, per renderli più accessibili agli studenti con Bisogni Educativi Speciali (BES) che per motivi fisici, biologici, fisiologici, psicologici e sociali hanno bisogno di interventi didattici mirati, specifici e personalizzati.

Sono molte le ricerche nel campo dell'educazione linguistica<sup>2</sup> che si sono occupate della comprensione del testo, dimostrando come il leggere per capire (Lorenzi, 2002) sia una delle attività cognitive umane più complesse. Si tratta di un processo attivo, dinamico e interattivo, in grado di instaurare una relazione affettiva tra il lettore e il testo, e dipende soprattutto dall'interazione trasversale di differenti variabili linguistiche, cognitive e metacognitive.

Dal punto di vista del modello cognitivista, costruito da Van Dijk e Kintsch (1983), la comprensione del testo è definita come un processo cognitivo che coinvolge la costruzione di una rappresentazione mentale coerente in una memoria episodica ed è finalizzato a cogliere il significato del testo. Ciò non implica che sia sufficiente mantenere in memoria le informazioni per comprendere, ma bisogna tener conto della relazione tra le parole, le frasi e i periodi che si susseguono in modo coerente e significativo. Il lettore in questo senso compie azioni costruttive e interattive con il testo, che portano al riconoscimento e all'implementazione di informazioni già note o nuove all'interno di strutture di conoscenza preesistenti.

Negli anni il dialogo intercorso tra la linguistica e la psicologia cognitivista ha cercato di spiegare come i processi di decodifica e comprensione del testo collaborino e interagiscano con i processi psicologici. Sono state elaborate diverse teorie che hanno portato a definire il modello cognitivo di rappresentazione della comprensione come la risultante di due processi simultanei e convergenti: il processo *bottom-up* di decifrazione e associazione e il processo *top-down* di anticipazione e previsione. Sono processi che

<sup>1</sup> Università degli Studi di Salerno.

<sup>2</sup> Vedi Quaderni del Giscel: Piemontese (1986), Lumbelli (1994), Zambelli (1994), Colombo, Romani (1996), Ferreri, 2002, Lorenzi, 2002.

diventano sempre più rapidi e indispensabili, automatizzandosi, man mano che sale il livello di competenza nella *literacy*. La sinergia cognitiva, la complementarità e la bidirezionalità di questi processi consentono l'interazione delle conoscenze e delle esperienze pregresse del lettore con i contenuti, la struttura e la lingua di un testo, rendendo molto rilevante l'idea che la decodifica del testo, ovvero la leggibilità, non si possa svincolare dalla sua comprensione, ovvero la comprensibilità.

Secondo OCSE-PISA il lettore va educato alla comprensione e guidato nell'intensa e complessa attività di recupero di conoscenze, di altre abilità e strategie in continua evoluzione, affinché possa raggiungere piena consapevolezza nella *literacy*. Pertanto, il lettore diventerà sempre più competente nelle diverse situazioni o contesti linguistici, se maggiore sarà la sua consapevolezza nel complesso processo di decodifica della comprensione del testo.

È evidente che la comprensione del testo non dipenda solo dal grado di conoscenza del lessico e/o da una maggiore competenza lessicale del lettore, ma nel processo concorrono anche altre componenti. Questi fattori sono riconducibili alla complessità delle strategie soggiacenti alla *literacy*, alle caratteristiche del testo e alle scelte metodologiche e didattiche (Cardona, 2008) come il *background* culturale ed esperienziale, le conoscenze specifiche sull'argomento, i processi di inferenza e di attivazione dell'*expectancy grammar*, la cosiddetta grammatica di anticipazione (Balboni, 2018) per facilitare la comprensione del testo.

Quando il lettore si accinge a leggere un testo, in realtà, sceglie la propria strategia di lettura sulla base dell'obiettivo che si propone e degli scopi per cui legge. Agati (1999) identifica diverse tecniche e strategie di lettura, classificandole sulla base degli obiettivi che il lettore si pone: a) lettura esplorativa (*scanning*) per individuare un'informazione specifica contenuta nel testo; b) lettura orientativa (*skimming*) per cogliere l'andamento generale del testo e valutarne la complessità; c) lettura intensiva (*intensive reading*) per scoprire cosa un testo comunica, analizzandolo in profondità; d) lettura studio (*receptive reading*) per apprendere, memorizzare e riprodurre le informazioni contenute nel testo; e) lettura di sintesi o riassuntiva per fissare i contenuti e i concetti. Per ciascuna di queste tecniche e strategie esiste un processo cognitivo e metacognitivo di acquisizione linguistica volto alla comprensione del testo, che può essere più lento o più veloce, a seconda del grado di consapevolezza nella *literacy* posseduta dal lettore. Questi livelli si diversificano e si intervallano tra due punti focali: la comprensione *estensiva*, focalizzata nel cogliere solo alcuni aspetti superficiali, la comprensione *intensiva*, che comporta un'analisi più profonda sia del testo sia degli elementi socioculturali e pragmatici che stanno oltre il testo (Balboni, 2018). Indipendentemente dalla tecnica o strategia preferita e adottata dal lettore, l'obiettivo principale del suo leggere è quello di comprendere al massimo il messaggio comunicativo del testo. Ma come fa il lettore a capire se quel testo è più leggibile o più comprensibile?

È questa la domanda a cui si cercherà di rispondere nel corso di queste pagine che si inseriscono così nel lungo dibattito che, fin dagli anni '70, ha coinvolto studiosi di educazione linguistica (vedi T. De Mauro, A. Colombo, M. E. Piemontese, L. Lumbelli, M. Ambel). Questi lavori si sono focalizzati sui bisogni dei lettori e hanno indagato le tecniche di misurazione della leggibilità dei testi scritti, le componenti della comprensibilità, la semplificazione della comunicazione in rapporto ai destinatari, al fine di dimostrare che la lingua diventa accessibile a tutti quando e se è adeguata ai destinatari e ai loro bisogni formativi.

Dopo aver definito e illustrato brevemente i principali aspetti di complessità della leggibilità e comprensibilità di un testo nel § 2 e aver presentato gli indici di leggibilità più diffusi nel § 3, verranno analizzati i dati derivanti dal calcolo dell'indice di leggibilità GULPEASE nel *corpus* LeCo nel § 4. Infine, il § 5 si soffermerà sulle possibili strategie di

semplificazione e facilitazione dei testi presenti in LeCo, in particolar modo per gli studenti BES, tracciando nuove prospettive di studio e ricerca future nel § 6.

## 2. LEGGIBILITÀ E COMPRESIBILITÀ SONO SINONIMI O CONTRARI?

Nel considerare la chiarezza espositiva di un testo la leggibilità e la comprensibilità, caratteristiche determinanti nell'agevolare o nell'ostacolare il processo cognitivo di comprensione, non sono da considerarsi sinonimi (Zambelli, 2014). Per evitare di confondere e sovrapporre i loro significati è opportuno dunque definirli e tenerli distinti, sebbene sia difficile tracciare delle linee di separazione nette.

Intuitivamente si pensa che la leggibilità di un testo dipenda dalla maggiore o minore presenza di parole difficili nel testo o dal grado di conoscenza dei significati delle parole possedute dal lettore. La leggibilità, invece, ha a che vedere con l'impianto linguistico che rende il testo, più o meno chiaro e comprensibile, a prescindere dalla complessità degli argomenti contenuti nel testo, tale per cui si crede che più un testo risulta leggibile, tanto maggiore è la probabilità che esso venga compreso facilmente. Ma conoscere i significati delle parole non implica necessariamente di averle comprese e neppure di aver compreso ciò che il testo vuole comunicare davvero.

Con il termine *leggibilità* si indicano diversi aspetti linguistici e/o grafici di tipo formale e di carattere quantitativo, che vanno dalla decifrabilità materiale del testo, come la calligrafia, il font, la qualità della grafica, l'uso del colore, il corpo tipografico per catturare l'attenzione, coinvolgere e interessare il lettore, alle caratteristiche formali del testo associate alla facilità di lettura o alla difficoltà di comprensione del testo, cioè il lessico, la sintassi, la coerenza, la coesione, l'organizzazione dei contenuti e le scelte stilistiche. Quando la leggibilità si riferisce agli aspetti logico-semantiche del testo, allora è più appropriato parlare di *comprensibilità*, come vedremo più avanti.

Ora, per esplicitare meglio la dualità e la complessità di definizione della *leggibilità* e *comprensibilità*, si fa riferimento alla terminologia usata in inglese che distingue la leggibilità in *legibility* (leggibilità fisica) e in *readability* (leggibilità linguistica), e per la comprensibilità utilizza questi termini: *comprehensibility* o *ease of understanding*, che riguardano le caratteristiche della scrittura legate alla comprensione del lettore, ovvero ai problemi di comprensibilità del testo in senso stretto (Klare, 1984).

Gli studi condotti dagli anni '80 hanno determinato quali sono i fattori di leggibilità fisica e linguistica di un testo (Tabella 1).

Tabella 1. *I fattori della leggibilità fisica e linguistica*

LEGGIBILITÀ FISICA	LEGGIBILITÀ LINGUISTICA
<b>Aspetto grafico</b>	<b>Lunghezza delle frasi</b>
Presenza di immagini, tabelle, disegni	paratassi
Font	ipotassi
Presenta di caratteri speciali per definizioni o lemmi stranieri	
<b>Organizzazione e struttura sintattica del testo</b>	<b>Lunghezza delle parole</b>
Divisione in capitoli, paragrafi, sottoparagrafi	lessico
Titolazione, spaziatura, capoversi	

La Tabella 1 mostra che la leggibilità ha una duplice natura (fisica e linguistica) ed evidenzia da un lato la complessità nel definirla, dall'altro la difficoltà di stabilire con precisione quando e se un testo sia o non sia di facile lettura. Facciamo degli esempi: un testo composto da frasi lunghe e ricche di subordinazioni risulterà per un lettore *low performer* o poco istruito, troppo complesso; così come un testo che contiene parole lunghe, con alta intensione semantica, potrebbe dar luogo a costruzioni di frasi, troppo dense di significato, e provocare la non immediatezza nella lettura nei lettori fragili o *poor comprehenders*.

Inoltre, la mancanza nella cura degli aspetti legati alla leggibilità fisica o grafica del testo potrebbe provocare la diminuzione della capacità di concentrazione del lettore e rallentare la lettura rendendola molto più faticosa, oppure incentivare un comportamento, tale per cui il lettore evita di leggere parola per parola saltando velocemente, qua e là nel testo, alla ricerca di eventuali appigli di comprensione (titoli, parole in grassetto, glosse, ecc.). Per ovviare a simili problemi, soprattutto per gli studenti BES, sarebbe necessaria una ricerca più rigorosa per scoprire se la scelta del font “giusto” sia in grado di agevolare la lettura, facilitando il riconoscimento dei caratteri ed evitando l'effetto di affollamento e di confusione tra spazi e parole.

Ad oggi, sono ancora pochi i lavori scientifici che si occupano di testare la reale efficacia della leggibilità fisica dei font e di identificare i caratteri ad alta leggibilità. La sezione di Cuneo dell'AID (Associazione Italiana Dislessia) ha svolto una ricerca in cui prende in considerazione cinque casi studio<sup>3</sup>, condotti nell'ambito degli studi comparativi su parlanti dislessici, spagnoli, inglesi, australiani, olandesi, italiani, concepiti per valutare l'efficacia tra i font tradizionali utilizzati nei word processor (*Arial, Comic Sans, Century Gothic, Times New Roman, Calibri, ecc.*) e quelli ideati *ad hoc* per i dislessici (*OpenDySlexic, Dyslexie, Easyreading, Sylexiad e ReadRegular*). Dal confronto dei risultati non si è riusciti a stabilire l'esistenza di un font specifico, che sia in grado di rendere la lettura di tutti DSA più agevole e veloce, in quanto i dati raccolti nei cinque casi studio sono poco rappresentativi e non omogenei dal punto di vista della scelta dei font da confrontare e molto spesso la preferenza di un font rispetto ad un altro dipende dal gusto estetico e soggettivo del lettore. Lo studio, però, ha messo in evidenza un'unica eccezione per quanto riguarda il font *Easyreading*, la cui leggibilità sembrerebbe superiore agli altri font tradizionali. Tuttavia, l'ipotesi di considerare il font *Easyreading* come uno strumento compensativo in grado di facilitare la lettura per gli studenti BES andrebbe approfondita: tendenzialmente si può affermare che un font “ad alta leggibilità”, dal punto di vista della leggibilità fisica o grafica, garantisce l'accessibilità al testo ai lettori con problemi visivi o neurologici, come gli ipovedenti o i dislessici. Ma siamo sicuri che la redazione di un testo con un font “ad alta leggibilità” grafica, come *Easyreading*, implica automaticamente che quel testo sia leggibile e quindi comprensibile? Questa ipotesi andrebbe verificata, perché come vedremo più avanti sia la leggibilità sia la comprensibilità possono essere o diventare ostacoli nel processo di comprensione.

A differenza della comprensibilità, la leggibilità è però oggettivamente “misurabile” grazie all'ausilio di indici specifici, che saranno illustrati nel § 3. Mentre la leggibilità ha a che vedere con l'aspetto linguistico del testo, per *comprensibilità* si intende un concetto più ampio che interessa le caratteristiche linguistiche e più qualitative, come l'organizzazione dell'informazione e il contenuto del testo (Vedovelli, 1994) e si focalizza su criteri semantici e nessi logici testuali. La comprensibilità viene definita da Lumbelli (1996) come la condizione necessaria, ma non sufficiente perché si realizzi la comprensione e riguarda gli aspetti più profondi, logico-semantici del testo: la densità delle informazioni, la loro esplicitezza, l'articolazione dei contenuti e il contesto nel quale avviene la comunicazione.

<sup>3</sup> Rello e Baeza-Yates (2013), Wery e Diliberto (2016), Maurinus (2017), Kuster (2017), Bachmann (2013).

Mentre la leggibilità è una caratteristica intrinseca del testo, la comprensibilità è una caratteristica relativa che deriva dalla relazione che si stabilisce tra il lettore e il testo (Lucisano, Piemontese, 1986). La comprensibilità di un testo dipende dalle competenze linguistiche e cognitive del lettore, ma anche dalla componente extralinguistica che ha un ruolo notevole per favorire la comprensione effettiva, in quanto un testo può essere comprensibile, ma non effettivamente compreso. Ciò implica che per trasformare la comprensibilità di un testo in comprensione effettiva devono entrare in gioco altri fattori, extralinguistici, più relazionali ed emozionali.

Al contrario della leggibilità, la comprensibilità di un testo non è misurabile quantitativamente, ma valutabile qualitativamente all'interno del complesso rapporto che si instaura tra il sistema di conoscenze, esperienze dell'emittente e quelle del destinatario del messaggio comunicativo. Ciò significa che un testo, apparentemente incoerente e insignificante per alcuni, potrebbe essere comprensibile per coloro i quali condividono un alto numero di conoscenze sullo sfondo socioculturale, linguistico, ma anche personale degli attori della comunicazione. Benché la chiarezza di un testo e la sua comprensibilità dipendano da fattori relazionali, «per essere chiari quando si scrive, bisogna sgomberare il campo dalle confusioni e dalle imprecisioni; bisogna, in altri termini, prendersi la responsabilità di ciò che si vuole comunicare» (*Manuale di stile: strumenti per semplificare il linguaggio delle amministrazioni pubbliche*, Presidenza del Consiglio dei ministri, 1997: 9). Un testo, dunque, per essere comprensibile a chi legge, deve essere compreso innanzitutto da chi lo scrive per porsi continuamente dalla parte dei destinatari. Ed ecco spiegato il motivo per cui la comprensibilità è strettamente legata agli aspetti linguistici più profondi del testo come l'architettura logico-semantica dell'intero testo e il grado di esplicitezza delle informazioni date.

In sintesi, la comprensibilità rappresenta la struttura più profonda della comprensione e nel suo definirsi tiene conto di due tipi di fattori:

1. gli *aspetti funzionali*, legati a momenti specifici dell'elaborazione o stesura del testo, come il grado di pianificazione dei contenuti e di strutturazione concettuale, gerarchizzata o organizzata in sequenza; il grado di esplicitezza delle informazioni date per evitare gli impliciti (Pietrandrea, 2021), cioè meno riferimenti impliciti, più comprensibilità del testo; il grado di esplicitezza degli scopi del testo, al fine di motivare alla comprensione e rendere il lettore più partecipe nel suo percorso di costruzione dei significati;
2. gli *aspetti logico-semantici*, come la coerenza, la coesione, le catene anaforiche, la densità delle informazioni implicite o esplicite, la vicinanza o lontananza dei contenuti alle conoscenze enciclopediche del lettore.

La scelta di distinguere l'aspetto superficiale della leggibilità dall'aspetto più profondo della comprensibilità nel processo di comprensione consente di evidenziare i diversi metodi di misurazione della leggibilità e i diversi livelli di valutazione degli ostacoli alla comprensibilità. Lumbelli (1989) riesce a distinguere questi ostacoli, di fronte ai quali possono trovarsi i lettori di un testo, in *ostacoli superficiali*, più legati alla leggibilità e alla decifrazione materiale del testo, e in *ostacoli profondi*, che hanno più a che fare con la comprensibilità del testo. Sebbene si potessero eliminare o ridurre gli ostacoli superficiali in un testo, questo meccanismo di abolizione non implica automaticamente il superamento degli ostacoli profondi alla comprensione. Infatti, non è possibile eliminare gli ostacoli profondi alla comprensione del testo senza aver prima facilitato il processo, eliminando o riducendo gli ostacoli superficiali. Esplicitiamo meglio questi concetti: considerando un testo poco leggibile, gli aspetti superficiali del testo mascherano e rappresentano un ostacolo superficiale al riconoscimento degli aspetti logico-semantici; mentre in un testo, del tutto privo di ostacoli superficiali, si possono avere problemi nella

ricostruzione degli aspetti più profondi, come ad esempio, leggere un testo di medicina per i non addetti al lavoro.

Ma allora è la leggibilità a veicolare la comprensibilità? Probabilmente sì, ma non è detto che la leggibilità porti effettivamente alla comprensione del testo. Diciamo che nell'elaborazione di un testo è importante impiegare il più possibile una strutturazione semplice e chiara per aumentare il grado di leggibilità del testo, sebbene la sua comprensibilità, come abbiamo detto, dipenda dal grado di diffusione e popolarità delle parole contenute, dal *know how* di chi legge e dalla relazione creata dal lettore tra il testo e il contenuto. Da ciò deriverebbe l'ipotesi che se un testo è più comprensibile, contiene parole sicuramente più conosciute e più frequenti, è quindi anche più leggibile.

### 3. GLI INDICI DI LEGGIBILITÀ

Gli *indici di leggibilità* sono degli indicatori che consentono di capire quanto un testo sia leggibile, chiaro e scorrevole e allo stesso tempo di facile lettura, ma non sono in grado di rendere conto né delle variabili extralinguistiche e sociolinguistiche che intervengono nella comprensione del testo, né dei nessi logici dell'impianto concettuale. Gli indici di leggibilità rappresentano gli strumenti in grado di misurare la leggibilità, in quanto definiscono una scala di valori, grazie alla quale è possibile confrontare la gradazione di comprensione di diversi tipi di testi e identificare gli utenti in grado di leggerli e comprenderli. Come tutti gli strumenti automatici, questi indici valutano il profilo quantitativo e non qualitativo del testo, cioè forniscono indicazioni utili su aspetti testuali, ma non bisogna mai perdere di vista i profili socioculturali e linguistici dei destinatari dei testi nella verifica del loro reale grado di comprensione.

A partire dagli anni '20 negli USA nascono i primi studi sulla leggibilità dei testi e si iniziano ad elaborare le prime formule matematiche in grado di misurarla. Gli obiettivi di queste prime ricerche sulla leggibilità hanno interessato principalmente gli ambiti scolastici e si sono concentrate sia sul controllo del vocabolario dei libri di testo, sia sulla comprensione dei materiali proposti agli studenti. Nonostante ciò, il primo studioso a divulgare attraverso i suoi studi il concetto di quello che oggi definiamo con *leggibilità* è l'americano Rudolf Flesch.

Nel suo libro *The Art of Plain Talk* (1946), scritto volutamente in modo "leggibile", Flesch pose le basi al suo metodo, esponendo quali fossero le caratteristiche, insite nel testo, a renderlo difficile e di non facile lettura. Flesch le distinse in difficoltà più propriamente sintattiche (ipotassi), morfologiche (quantità di prefissi, infissi e suffissi) e difficoltà più semantiche (parole astratte, nomi propri). Partendo da queste considerazioni, Flesch ideò la sua formula matematica, conosciuta con il nome di *Indice di Flesch*, che è in grado di calcolare la leggibilità per la lingua inglese.

L'Indice di Flesch tiene conto di due variabili linguistiche: la lunghezza media delle parole (S), espressa in sillabe per parola e la lunghezza media delle frasi (W), espressa in parole per frase. Utilizzando questi due valori in un'equazione, Flesch costruì la scala di punteggio del suo indice di leggibilità, che va da 1 a 100:

$$\text{EASY OF READING [facilità di lettura]} = 206,835 - (0,864 S) - (1,015 W).$$

Per ottenere il valore di S [sillabe di cento parole], ovvero l'indice di difficoltà semantica, bisogna dividere il numero di sillabe totale per il numero di parole contenuto in un testo, mentre per calcolare W [media di parole per frase], ovvero l'indice di complessità sintattica, si divide il numero di parole totali per il numero di frasi. Più il valore

dell'indice è vicino a cento, più il testo è leggibile, al contrario, più si avvicinerà allo zero, più il contenuto del testo risulterà difficile da comprendere.

La Tabella 2 riporta i valori dell'Indice di Flesch in rapporto ad una scala di facilità-difficoltà di lettore e al livello di istruzione del lettore.

Tabella 2. *I valori dell'Indice Flesch*

INDICE	INTERPRETAZIONE DEL PUNTEGGIO	PAROLE PER FRASE	SILLABE SU 100 PAROLE	LIVELLO SCOLASTICO (USA)	LIVELLO SCOLASTICO (ITALIA)
100-90	Molto facile	<8	<123	5° grado	Scuola primaria
90-80	Facile	11	131	6° grado	Scuola secondaria di primo grado
80-70	Abbastanza facile	14	139	7° grado	Scuola secondaria di primo grado
70-60	Normale	17	147	8° e 9° grado	Scuola secondaria di primo grado
60-50	Abbastanza difficile	21	155	10°-12° grado	Scuola secondaria di secondo grado
50-30	Difficile	25	167	13°-16° grado	Università
30-0	Molto difficile	>29	>192	College	Post Laurea

Dall'analisi della scala dell'indice di Flesch si evincono due elementi: 1. un testo contenente frasi lunghe risulterà sintatticamente più complesso e molto difficile da comprendere; 2. un testo di maggiore facilità di lettura presenterà parole lunghe meno frequenti. Secondo Miller (1972) le parole più frequenti sono appunto quelle più familiari, più brevi e senza affissi.

Nel 1952 un altro americano Robert Gunning in *The Technique of Clear Writing* inventa il *Fog Index* e descrive come applicare la sua formula di leggibilità destinata agli adulti. A differenza dell'Indice di Flesch, la formula di Gunning tiene conto della lunghezza media delle frasi e delle parole polisillabiche, cioè conta il numero di parole con più di due sillabe per ogni cento parole, le cosiddette parole difficili:

$$\text{READING GRADE LEVEL} = 0,4 (\text{lunghezza media delle frasi} + \% \text{ di parole difficili})$$

I valori risultanti da questo calcolo indicano il numero minimo di anni di scuola, che una persona deve aver frequentato, per leggere con facilità il testo analizzato e la Tabella 3 mostra i livelli di lettura, suddivisi per grado di istruzione, sulla base dei valori della scala del Fog Index e di alcune risorse su cui Gunning ha calcolato la leggibilità.

Il punteggio per una buona leggibilità per Gunning sarebbe quello che va tra i valori 7 e 8, in quanto qualunque testo al di sopra di 12 è troppo complesso per la maggior parte dei lettori. I valori oltre 17 vengono considerati appartenenti al livello post-laurea. Gunning applica il suo indice su diversi testi, tra cui la Bibbia, e scopre che il suo indice di leggibilità è pari a 6, cioè bastano sei anni di scuola per poterla leggere e capire.

Tabella 3. *Interpretazione dei punteggi dell'Indice Fog*

Punteggio	Grado di lettura stimato	Risorse
>17	Post-laurea	Articoli accademici
16 - 13	Università	Articoli accademici
12-9	Scuola superiore di secondo grado	Testate giornalistiche
8 - 7	Scuola superiore di primo grado	I romanzi più popolari
<6	Primaria	La Bibbia, la guida TV, ecc.

Nel 1975 è Peter Kincaid ad applicare delle modifiche sugli indici di Flesch e Gunning, dando luogo ad una nuova formula della leggibilità che valuta la complessità del testo sulla base del numero medio di sillabe per parola (S) e del numero medio di parole per frase (W):

$$GL \text{ [Grade Level]} = (0,39 \times W) + (11,8 \times S) - 15,59.$$

I risultati ottenuti sono compresi tra 0 e 12 e indicano, in misura approssimativa, il numero di anni di scuola che il lettore dovrebbe aver fatto per comprendere il contenuto del testo. I valori compresi tra 6 e 10 indicano testi che possono essere letti con facilità dalla maggior parte delle persone. In pratica, l'età del lettore equivale al grado di istruzione americano + 5.

Gli indici di leggibilità, appena descritti, sono quelli che hanno avuto maggiore rilevanza nel mondo delle ricerche sulla leggibilità di testi in lingua inglese. In Italia, bisogna aspettare la fine degli anni '70 per avere studi linguistici che pongono l'attenzione sui problemi legati alla leggibilità e alla comprensione del testo in lingua italiana. Infatti, il primo adattamento della Formula di Flesch alla lingua italiana risale al 1972 da parte di Roberto Vacca, ingegnere, matematico e scrittore, che divulga gli studi di Flesch e introduce per la prima volta in Italia una formula di leggibilità, l'*Indice Flesch-Vacca*:

$$IL \text{ [Indice di Leggibilità]} = 206 - 0,65 S - W,$$

dove W è il numero medio di parole per frase in un campione di cento parole; S è il numero di sillabe per cento parole; 206 è la costante applicata per mantenere i valori compresi tra 0 e 100; 0,65 è la costante riferita alla lunghezza media delle parole italiane. Secondo l'Indice Flesch-Vacca, un testo con indice, compreso tra 0 e 30, viene valutato molto difficile, mentre tra 60-70 è comprensibile per studenti con licenza media.

Negli anni successivi il gruppo GULP (*Gruppo Universitario Linguistico Pedagogico*) presso la Facoltà di Scienze dell'Educazione dell'Università della Sapienza di Roma, guidati da Tullio De Mauro, continua la propria ricerca in questo ambito con l'obiettivo di elaborare una formula di leggibilità, tarata sull'italiano, e nasce così l'*Indice di leggibilità GULPEASE*, ovvero la prima formula di leggibilità per la lingua italiana. Essa permette di calcolare un valore numerico che a seconda della competenza dell'ipotetico lettore assume una diversa interpretazione in merito alla leggibilità del testo: più è alto il valore della scala, maggiore è la leggibilità del testo che viene analizzato, ma se il lettore è altamente scolarizzato quel valore ha un diverso significato.

L'indice GULPEASE calcola la lunghezza delle parole e non delle sillabe, a differenza dell'Indice di Flesch, ed è stato formulato per verificare la reale leggibilità o gradazione di comprensione di un *corpus* di testi in diversi tipi di lettori, effettuando rilevazioni, raccolte tra il 1986 e il 1987 dalle cattedre di Filosofia del linguaggio e di Pedagogia dell'Istituto di Filosofia. Nella sua formula si considerano i seguenti fattori: il numero di lettere per parola e il numero di parole per frase (Lucisano, Piemontese, 1988):



$$\text{FACILITÀ DI LETTURA} = 89 - \text{LP}/10 + \text{FR} * 3,$$

dove LP indica le lettere per cento sul totale delle parole, FR le frasi per cento sul totale delle parole. I valori che si ottengono sono compresi in una scala da 0 a 100. Nella figura 1 si può vedere la scala dei valori dell'indice GULPEASE e capire il suo funzionamento.

Se si analizza la scala dei valori dell'Indice di leggibilità GULPEASE (Figura 1) e si considerano i livelli di lettura indipendente con soglie di leggibilità *facili*, si nota che:

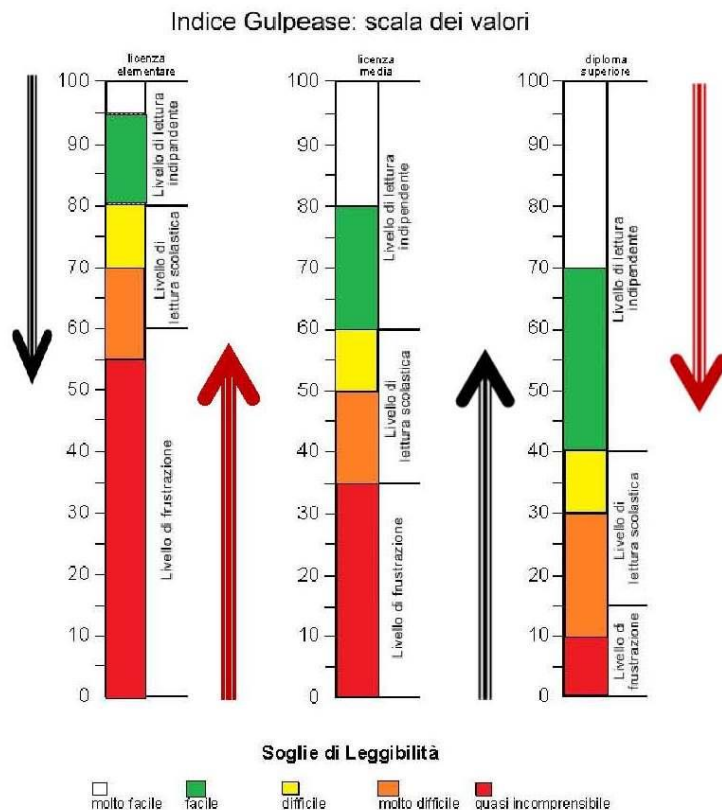
- i lettori che hanno un'istruzione elementare leggono più facilmente testi con un punteggio superiore a 80;
- i lettori con licenza media sono in grado di leggere in modo autonomo un testo con un punteggio superiore a 60;
- i lettori con un'istruzione superiore leggono testi con un punteggio superiore a 40.

Ciò significa che il livello di istruzione è inversamente proporzionale al valore dell'indice di leggibilità, come mostrano le direzioni opposte delle frecce nere, che indicano il livello di istruzione (+ o - alto) e delle frecce rosse, che indicano i valori GULPEASE (+ o - alti). Invece, se si osserva la scala dei valori sulla base delle soglie di leggibilità appartenenti al *livello di lettura scolastica* vediamo che:

- per i lettori con licenza elementare i testi possono avere valori compresi tra 55 e 80;
- per i lettori con licenza media i testi possono avere valori compresi tra 35 e 60;
- per i lettori con licenza superiore i testi possono avere valori compresi tra 15 e 40.

In questi casi i testi risultano *difficili* o *molto difficili* e i lettori *low performers* o *poor comprehenders* andrebbero guidati nel processo di comprensione del testo. Infine, le soglie di leggibilità con i valori più bassi (licenza elementare <55, licenza media <35, licenza superiore <15) evidenziano i *livelli di frustrazione* nei diversi tipi di lettori e la quasi impossibilità di leggere e comprendere questi testi, che secondo l'indice di leggibilità GULPEASE sono considerati *quasi incomprensibili*.

Figura 1. *La proporzionalità inversa dell'indice di leggibilità e il grado di istruzione*



Come abbiamo visto, la leggibilità di un testo dipende principalmente dalla struttura del testo, delle sue frasi e dalla lunghezza delle frasi e delle parole che le compongono e pertanto, grazie agli indici di leggibilità, possiamo dedurre che la leggibilità dipende esclusivamente da fattori quantitativi e oggettivamente misurabili.

Sfruttando questa specificità della leggibilità di essere misurabile, abbiamo svolto un'analisi quantitativa sul *corpus* LeCo<sup>4</sup> con un duplice obiettivo:

1. calcolare l'indice GULPEASE dei testi scritti del *corpus* LeCo e verificare se i testi scelti sono adatti ai destinatari del progetto: gli studenti del biennio di scuola secondaria superiore;
2. individuare i testi del *corpus* LeCo con indice di leggibilità  $\geq 60$ , che considereremo ad alta leggibilità e quindi idonei agli studenti BES, al fine di implementare sulla piattaforma LeCo strumenti utili a facilitare il loro processo di comprensione.

I risultati di questa analisi sul *corpus* LeCo saranno illustrati nel dettaglio nel § 4 e nel § 6.

### 3.1. L'indice di leggibilità GULPEASE: gli strumenti moderni

Oggi esistono tante applicazioni online, a pagamento e gratuite, che sono in grado di calcolare in modo automatico e automatizzato gli indici di leggibilità di un testo. Lo scopo di questi strumenti informatici, che calcolano in modo immediato la formula GULPEASE, è quello di individuare e analizzare gli indicatori che descrivono la maggiore o minore difficoltà nella leggibilità di un testo e i loro algoritmi si stanno sempre più evolvendo. In realtà, sono programmi molto semplici da usare: in molti casi è sufficiente accedere ad una delle applicazioni *open source* presenti sul web, copiare e incollare il testo da analizzare nella pagina dedicata e attendere il risultato del calcolo dell'indice di leggibilità. Esistono versioni demo gratuite, altre sono a pagamento o in abbonamento e consentono un numero minimo di tentativi di analisi sui testi, solo se registrati, e con analisi finali anche molto parziali, in quanto sono coperte da diritto d'autore e forniscono una grande mole di informazioni.

Il programma *Èngolos* è uno di questi sistemi automatizzati che offre servizi parziali e in abbonamento per il calcolo dell'indice GULPEASE. Esso è stato impiegato per l'analisi del *corpus* delle prime cinque annate (1989-1994) del periodico di facile lettura *Due Parole* (Piemontese, 1996) ed è in grado di condurre sia l'analisi della leggibilità di un testo sia quella statistico-lessicale, più qualitativa incentrata sulla comprensibilità, che approfondiremo meglio nel § 5. Il suo *Censor* (1996-2006) offriva un servizio gratuito, tramite e-mail, ma oggi in disuso. Oltre a calcolare l'indice di leggibilità GULPEASE, *Censor* metteva a confronto l'indice con il *Vocabolario di Base* di Tullio de Mauro (VdB). Questo tipo di analisi permetteva agli insegnanti di capire le difficoltà con i testi disciplinari da parte di studenti italiani e stranieri, appena inseriti nella realtà scolastica. Esso è stato sostituito da un altro servizio automatico, più evoluto, in versione demo e a pagamento: *Corrige! Leggibilità* (2016) che impiega una tecnologia sicuramente più avanzata di trattamento automatico del linguaggio, in quanto è capace di generare un documento in formato HTML con tutti i dati di leggibilità e di lessico. Esso analizza la leggibilità del testo frase per frase utilizzando l'indice GULPEASE e confronta le parole del testo con il VdB grazie alla lemmatizzazione automatica. Mette inoltre in evidenza le parole ritenute più difficili per il lettore e comprensibili solo da lettori specializzati in ambiti professionali specifici.

<sup>4</sup> *LeCo* (*Leggere e Comprendere*) è un progetto sul recupero e il potenziamento delle competenze nella *literacy* degli studenti del biennio di scuola secondaria superiore (cfr. Voghera *et al.* in questa monografia).

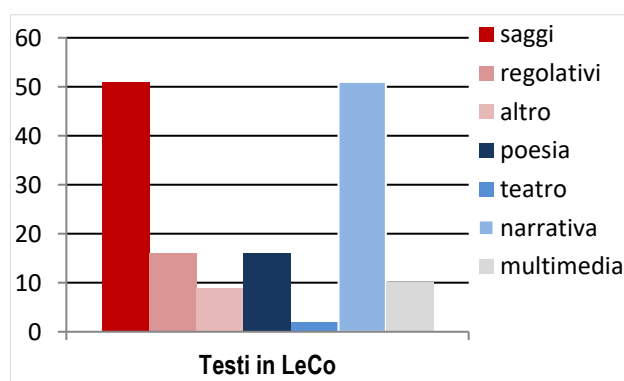
Il calcolo dell'indice di leggibilità sui testi scritti del *corpus* LeCo è stato effettuato grazie ad un'applicazione *open source*, gratuita, frutto del progetto *Farfalla Project* (2013) di Andrea Mangiatordi, un ricercatore dell'Università di Milano-Bicocca, che ha dedicato i suoi studi all'accessibilità dei siti web e allo sviluppo di soluzioni di programmi gratuiti. Questo progetto ha come scopo l'implementazione di strumenti web accessibili e inclusivi, in grado di personalizzare la lettura e la navigazione delle pagine web, mediata dal computer, per renderle più leggibili ai bambini con disabilità e bisogni speciali e rispondere così ai loro bisogni di inclusione sociale. Lo strumento del *Farfalla Project* permette di calcolare l'indice di leggibilità GULPEASE del testo, il numero totale delle parole e di individuare i destinatari, divisi in tre categorie (licenza elementare, licenza media, diploma superiore), sulla base della difficoltà di lettura del testo analizzato.

Si è scelto questo strumento per calcolare gli indici di leggibilità GULPEASE dei testi scritti nel *corpus* LeCo (§ 4), verificare l'idoneità dei testi scritti presenti in LeCo per gli studenti del biennio di scuola superiore e individuare i testi ad alta leggibilità che garantiscano l'accessibilità agli studenti BES. Questo studio della leggibilità nel progetto LeCo, che non è stato pensato per studenti con bisogni educativi speciali, deriva dalla necessità di rispondere ad un bisogno formativo, espresso da alcuni docenti nella fase di sperimentazione didattica, al fine di individuare un percorso adeguato e inclusivo anche per gli studenti BES.

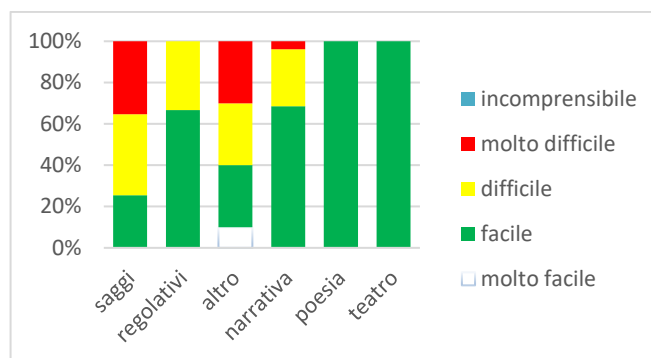
#### 4. LA LEGGIBILITÀ IN LeCO

Il *corpus* LeCo è composto da 155 testi, suddivisi per tipi testuali: funzionali (saggi, regolativi e altro), letterari (poesia, narrativa, teatro) e multimediali (orali). Nella figura 2 si osserva che la saggistica e la narrativa sono bilanciati nel numero di testi (51) rispetto agli altri tipi di testo, così come i testi regolativi (15) e le poesie (16), e che i testi multimediali (10), essendo orali, non sono stati considerati nel calcolo dell'indice di leggibilità GULPEASE.

Figura 2. *Il corpus LeCo: i generi testuali*



Pertanto, se analizziamo i 145 testi scritti di LeCo sulla base dei valori degli indici di leggibilità GULPEASE e li classifichiamo tenendo conto delle loro soglie di leggibilità, osserviamo nella figura 3 i livelli di analisi di lettura indipendente (facile e molto facile), di lettura scolastica (difficile e molto difficile), di livello di frustrazione (incomprensibile), calibrati sulla scala dei valori dell'indice di leggibilità GULPEASE (vd. Figura 1 del § 3).

Figura 3. *Le soglie di leggibilità dei testi in LeCo*

Alla luce di questi dati si può affermare che: a) i testi presenti in LeCo sono idonei e adeguati agli studenti del biennio di scuola superiore secondaria; b) i testi funzionali rispetto ai testi letterari sono leggermente più difficili; c) non sono presenti testi incomprensibili.

Invece, per calcolare l'indice di leggibilità GULPEASE dei testi scritti del *corpus* LeCo e individuare quelli che potessero essere idonei agli studenti BES abbiamo lavorato per fasi e gradi:

1. Calcolo dell'indice di leggibilità GULPEASE solo su testi funzionali di livello iniziale.
2. Calcolo dell'indice di leggibilità GULPEASE su tutti i generi di testi e per tutti i livelli di difficoltà.
3. Analisi comparativa degli indici di leggibilità in base ai livelli di difficoltà tra saggistica e narrativa.
4. Identificazione dei testi ad alta leggibilità per gli studenti BES con indice di leggibilità GULPEASE  $\geq 60$ .

Nella fase 1 il calcolo dell'indice di leggibilità ha interessato solo testi funzionali di livello iniziale per verificare se il metadato *iniziale*, cioè l'etichetta scelta per indicare la difficoltà di livello del testo e del questionario corredato, corrispondesse nella scala di valori dell'indice GULPEASE a testi con indice di leggibilità  $\geq 60$ , ovvero il requisito fondamentale per un testo idoneo ad un lettore in possesso di licenza media. La tabella 4 mostra i dati risultanti da questa prima analisi sul calcolo dell'indice di leggibilità GULPEASE sui testi funzionali di livello iniziale presenti in LeCo.

Tabella 4. *Dati analisi calcolo indice GULPEASE su testi funzionali di livello iniziale*

N° Testi in LeCo	Genere testuale	Livello	Indice di GULPEASE				Alta leggibilità
			$\geq 80$	$\geq 60$	$\geq 56$	$< 56$	
9	Saggi	Iniziale	0	1	1	7	2
5	Regolativi	Iniziale	0	3	0	2	3
1	Altro	Iniziale	0	0	0	1	0
<b>Tot. 15</b>			<b>0</b>	<b>4</b>	<b>1</b>	<b>10</b>	<b>5</b>

Su 15 testi funzionali di livello iniziale solo 4 testi presentano un indice di leggibilità GULPEASE  $\geq 60$ , un solo testo ha un valore  $\geq 56$  (e quindi approssimabile, per eccesso, a 60) e tutti gli altri testi hanno valori compresi tra 46 e 52. Questi 10 testi, considerando la

scala di valori dell'indice GULPEASE, sono adatti a lettori con diploma superiore. I testi da ritenersi ad alta leggibilità e idonei quindi a studenti del biennio della scuola secondaria superiore sono 5: due saggi e tre testi regolativi per il livello iniziale. Questi dati hanno evidenziato la necessità di proseguire nella ricerca, in quanto il metadato di LeCo con il livello di difficoltà *iniziale*, che si basa su criteri diversi di attribuzione<sup>5</sup>, non corrisponde in modo automatico a indici di leggibilità alti. Ed ecco perché nella fase 2 abbiamo esteso il calcolo dell'indice della leggibilità GULPEASE a tutti i livelli di difficoltà e a tutti i generi testuali, funzionali (Tabelle 5, 6, 7) e letterari (Tabella 8) presenti nel *corpus* LeCo.

Tabella 5. *Dati analisi calcolo indice GULPEASE su testi funzionali di livello intermedio*

N° Testi in LeCo	Genere testuale	Livello	Indice di GULPEASE				Alta leggibilità
			≥80	≥60	≥ 56	< 56	
28	Saggi	Intermedio	0	3	4	21	7
9	Regolativi	Intermedio	0	4	3	2	7
3	Altro	Intermedio	0	0	1	2	1
<b>Tot. 40</b>			<b>0</b>	<b>7</b>	<b>8</b>	<b>24</b>	<b>15</b>

Su 40 testi funzionali di livello intermedio solo 7 presentano un indice di leggibilità GULPEASE  $\geq 60$ , 8 testi hanno un valore  $\geq 56$  (e quindi approssimabile, per eccesso, a 60) e tutti gli altri testi hanno valori che si attestano tra 39 e 55 e secondo la scala di valori dell'indice GULPEASE sono da considerarsi idonei a lettori con diploma superiore. I testi ad alta leggibilità per gli studenti BES, appartenenti al livello intermedio, sono 15: sette saggi, sette testi regolativi e un testo di altra funzione linguistica.

Tabella 6. *Dati analisi calcolo indice GULPEASE su testi funzionali di livello avanzato*

N° Testi In LeCo	Genere testuale	Livello	Indice di GULPEASE				Alta leggibilità
			≥80	≥60	≥ 56	< 56	
14	Saggi	Avanzato	0	2	2	10	4
1	Regolativi	Avanzato	0	0	0	1	0
6	Altro	Avanzato	1	2	0	3	3
<b>Tot. 21</b>			<b>1</b>	<b>4</b>	<b>2</b>	<b>14</b>	<b>7</b>

Su 21 testi funzionali di livello avanzato risulta che solo un testo ha un indice di leggibilità GULPEASE  $\geq 80$ , 4 testi presentano un valore  $\geq 60$ , 2 testi hanno valori  $\geq 56$  (e quindi approssimabile, per eccesso, a 60) e i restanti hanno valori compresi tra 41 e 54. Questi ultimi testi sono adatti a lettori con diploma superiore. Invece, i testi per il livello avanzato da ritenersi ad alta leggibilità e presumibilmente idonei agli studenti BES sono 7 e più precisamente quattro saggi e tre testi di altra funzione linguistica.

Si è proceduto parallelamente con l'analisi e il calcolo dell'indice di leggibilità GULPEASE anche per i testi letterari di LeCo, considerandoli sempre sulla base della tipologia testuale sia dei livelli di appartenenza dei singoli testi, come mostrato nella Tabella 7 riassuntiva.

<sup>5</sup> Vd. Voghera *et al.* in questa monografia.

Tabella 7. *Dati analisi calcolo indice GULPEASE su testi letterari in LeCo*

Testi letterari in LeCo							
N° Testi in LeCo	Genere testuale	Livello	Indice di GULPEASE				Alta leggibilità
			≥80	≥60	≥ 56	< 56	
20	Narrativa	Iniziale	1	10	2	7	13
16		Intermedio	0	13	1	2	14
15		Avanzato	0	11	3	1	14
<b>Tot. 51</b>			<b>1</b>	<b>34</b>	<b>6</b>	<b>10</b>	<b>41</b>
5	Poesia	Iniziale	5	0	0	0	5
7		Intermedio	7	0	0	0	7
4		Avanzato	4	0	0	0	4
<b>Tot. 16</b>			<b>16</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>16</b>
0	Teatro	Iniziale	0	0	0	0	0
1		Intermedio	1	0	0	0	1
1		Avanzato	1	0	0	0	1
<b>Tot. 2</b>			<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>

Il dato rilevante che emerge da questa analisi è che i testi letterari risultano più leggibili, in quanto, tranne in pochi casi, i valori dell'indice di leggibilità GULPEASE si attestano sul  $\geq 60$ . Un'altra interessante osservazione derivante dai dati presenti nella Tabella 7 è che il genere poesia e il genere teatro, poco rappresentati nel *corpus* LeCo, presentano indici di leggibilità GULPEASE  $\geq 80$  e quindi leggibili da tutti gli studenti.

Per quanto riguarda, invece, la narrativa il discorso si complica, in quanto i valori degli indici di leggibilità si diversificano a seconda del diverso livello di difficoltà del testo. Si ha, infatti, solo un testo con indice  $\geq 80$  per il livello iniziale, 34 testi con indice  $\geq 60$  per i tre livelli difficoltà, 6 con indici  $\geq 56$ , approssimabili per eccesso a  $\geq 60$ , e solo 10 testi presentano valori inferiori a 56. Nello specifico i 6 testi appartenenti al livello iniziale hanno indici compresi tra 52 e 55; i 2 testi di livello intermedio hanno valori compresi tra 50 e 51 e l'indice di leggibilità dell'unico testo di livello avanzato con il valore più basso è 45.

Ricapitolando, i dati mostrati nella Tabella 8 risultano molto interessanti, in quanto ci dicono che 86 sono i testi che attestano un indice di leggibilità GULPEASE con valori  $\geq 60$ . Essi sono più della metà dei testi presenti nel *corpus* LeCo e diversificati sia nel genere sia nei livelli di difficoltà.

Tabella 8. *Testi in LeCo ad alta leggibilità*

GENERE TESTUALE	GULPEASE $\geq 60$	GULPEASE $< 60$	GULPEASE $< 35$
Saggi	13	20	18
Regolativi	10	5	0
Altro	4	3	3
Narrativa	41	12	2
Poesia	16	0	0
Teatro	2	0	0
<b>Totale testi ad alta leggibilità</b>	<b>86</b>		

Secondo la scala dei valori dell'indice di leggibilità GULPEASE i testi con indice GULPEASE  $\geq 60$  non creano livelli di frustrazione tali da mettere in difficoltà gli studenti del biennio, ma questa constatazione è valida anche per gli studenti BES? È sufficiente calcolare l'indice di leggibilità GULPEASE dei testi, verificare che il suo valore sia idoneo ai destinatari per favorire e promuovere il loro processo di comprensione?

Nella fase 3 si è cercato di stabilire proprio il difficile rapporto tra testo e destinatario, promuovendo un'analisi comparativa degli indici di leggibilità GULPEASE tra la saggistica e la narrativa, tenendo come variabile costante il livello di difficoltà dei testi, stesso metadato nel *corpus* LeCo, per evidenziare le differenze in termini di numero di testi che possono essere ritenuti "ad alta leggibilità".

Tabella 9. *Analisi comparativa degli indici di leggibilità della narrativa e dei saggi*

Testi letterari				Testi funzionali				BES
Genere	Livello	$\geq 80$	$\geq 60$	Genere	Livello	$\geq 80$	$\geq 60$	$\geq 60$
Narrativa	Iniziale	1	10	Saggio	Iniziale	0	5	16
	Intermedio	0	13		Intermedio	0	7	20
	Avanzato	0	11		Avanzato	1	6	18
<b>Totale</b>		<b>1</b>	<b>34</b>	<b>Totale</b>		<b>1</b>	<b>18</b>	<b>54</b>

Dal confronto dei dati risulta che i testi narrativi con indici di leggibilità GULPEASE  $\geq 60$  sono il doppio dei saggi e che i testi totali da ritenersi ad alta leggibilità sono 54, cioè più di 1/3 dell'intero *corpus* LeCo.

L'identificazione di questi testi ad alta leggibilità per gli studenti BES si è tradotta nella fase 4 di questa ricerca nell'implementazione di:

1. strumenti di facilitazione della comprensione del testo, come l'audiolettura dei testi e dei questionari correlati per favorire e garantire un maggior grado di comprensibilità, e non solo di leggibilità, per gli studenti che hanno dimostrato maggiori difficoltà nella comprensione, come i BES e in particolar modo i dislessici (DSA);
2. nuovi metadati sulla piattaforma LeCo come il calcolo dell'indice di leggibilità GULPEASE, il conteggio delle parole totali del testo e l'identificazione di testi non continui e complessi contenenti immagini, grafici e tabelle che ostacolano la riproduzione orale del testo (vd. Figura 4).

Il calcolo dell'indice di leggibilità GULPEASE ci ha permesso di apportare interventi migliorativi alla piattaforma LeCo, ponendo maggiore attenzione ai diversi tipi di destinatari e implementando i metadati già presenti (vd. Voghera *et al.* in questa monografia) con altri: il valore dell'indice di leggibilità GULPEASE, il numero di parole totali del testo e l'audiolettura. In questo modo abbiamo cercato di rendere la piattaforma LeCo più inclusiva e accessibile agli studenti BES, e allo stesso tempo allargato la nostra prospettiva di ricerca come vedremo meglio nel § 6.

Figura 4. L'ambiente LeCo e i nuovi metadati

The screenshot shows the LeCo web interface. On the left is a sidebar menu with options: Profilo personale, Visualizza documenti (selected), Gestione documenti, Abilitazione test, and Storico dei Risultati. The main content area is titled 'VISUALIZZA DOCUMENTI' and shows a search bar with 'saç' entered. Below the search bar is a table of documents:

# Testo	Titolo	Genere	Difficoltà	Indice Gulpease	Num Parole	Audio	Operazione
84	La Sachertorte	Altro	Iniziale	60	959		Visualizza

Below the table, it indicates 'Mostrando da 1 a 1 di 1 risultati (filtered from 155 total entries)' and navigation buttons for 'Precedente' and 'Successiva'.

## 5. LA COMPRESIBILITÀ IN LeCO

La leggibilità del testo, valutata attraverso l'utilizzo di indici di leggibilità, è utile ma non sufficiente a garantire la reale fruibilità dei testi; per cui l'analisi della leggibilità va osservata più da vicino e più in profondità, focalizzando l'attenzione su aspetti più profondi: lessicali, morfosintattici e sintattici del testo. Non si può, infatti, parlare di leggibilità senza tralasciare il concetto di *comprensibilità*. Seppur riguardano sfere diverse, entrambe sono fondamentali nel processo della comprensione del testo.

Zambelli (2014) afferma che un testo è più leggibile quando presenta un lessico appartenente al VdB, il linguaggio utilizzato non è figurato, la struttura sintattica è lineare e i periodi sono brevi, non ci sono tante riprese anaforiche implicite, ci sono poche nominalizzazioni, i connettivi sono consueti e prevale la coordinazione. Mentre un testo è comprensibile quando è ben tematizzato, non è denso di informazioni, le inferenze sono adeguate alle conoscenze del destinatario e sono chiare e ben segnalate le intenzioni comunicative. Tutti questi fattori hanno a che vedere non solo con la leggibilità linguistica, ma anche con la comprensibilità, più legata al contenuto del testo, alla sua coerenza e coesione, alla sintassi, alla morfologia, al registro e quindi allo stile.

Uno dei percorsi possibili per controllare il grado di comprensibilità di un testo è quello di prendere come punto di riferimento le proposte di facilitazione che si sono compiute, in vari ambiti, per rendere la lingua italiana scritta, più diretta e più chiara. Si pensi, per esempio, ai criteri per la redazione di testi comprensibili nei settori legislativo e burocratico (Presidenza del Consiglio dei ministri, 1993); oppure all'esperienza di scrittura di testi facilitati del periodico *Due Parole. Mensile di facile lettura*, realizzato dall'Università di Roma "La Sapienza" e rivolto agli studenti che, per diversi motivi, hanno una capacità limitata di trarre informazioni e dimostrano difficoltà nella comprensione del testo. Molti sono gli studi e le ricerche, compiute da diversi linguisti, in questa direzione negli anni. Tutti questi studi hanno in comune un aspetto: il tener conto dei destinatari con differenze individuali di vario genere e con cause endogene, esogene, temporanee o permanenti, imputabili a difficoltà legate alla comprensione. Inoltre, questi studi hanno evidenziato che è possibile rintracciare particolari contesti scolastici in cui si evidenzia il grado di difficoltà che gli studenti hanno nella comprensione del testo quando:



- a) l'approccio al testo è passivo e le attività di comprensione proposte non specificano l'obiettivo in modo chiaro;
- b) la focalizzazione è sulla decodifica delle parole del testo, anziché sulla comprensione del significato che si sta decodificando;
- c) il fare inferenze e/o il formulare ipotesi sul testo risulta complesso;
- d) la scelta ricade su una sola modalità di fruizione del contenuto da comprendere, ad esempio la sola lettura del testo, di norma lineare e sequenziale;
- e) gli indizi linguistici, contestuali e paratestuali non vengono riconosciuti nelle catene anaforiche;
- f) scarsa è la consapevolezza del livello di comprensione e non ci si accorge di cosa sta o non sta realmente comprendendo lo studente;
- g) l'organizzazione, la rielaborazione e la sintesi diventano attività difficili e complesse da svolgere per chi ha livelli bassi di competenze di comprensione.

Gli esempi precedenti evidenziano che la non comprensione effettiva del testo può derivare anche da problematiche più serie, non sottovalutabili e assimilabili ai disturbi della lettura e, nei casi più gravi, ai limiti nel funzionamento cognitivo del processo di comprensione. Tener conto di questi aspetti ed essere in grado di individuarli il prima possibile è molto importante, non solo per quanto riguarda la leggibilità fisica e linguistica del testo, ma anche per la sua comprensibilità che non può essere assolutamente trascurata. Quella comprensibilità che, come abbiamo detto più volte, non è né misurabile, né quantificabile in termini numerici, ma considera gli aspetti lessicali e morfo-sintattici del testo, analizzandoli da un punto di vista più qualitativo e sotto articolato.

Nel *corpus* LeCo l'analisi qualitativa sulla comprensibilità dei testi è in fase di elaborazione ed è oggetto di uno studio in *fieri*. L'idea di partenza è quella di utilizzare strumenti di annotazione linguistica automatica del testo per ricavare indicatori del grado di leggibilità e comprensibilità di un testo più complessi e trarre conclusioni più solide. L'obiettivo è appunto quello di suggerire un approccio metodologico per definire un indice di qualità linguistica della leggibilità e della comprensibilità, più efficace, e in grado di abbassare il livello di incomprensibilità del testo e allo stesso tempo di migliorare il grado di esplicitzza e la struttura sintattica del testo.

### 5.1. READ-IT e LeCo: studio in *fieri*

Per valutare la comprensibilità dei testi del *corpus* LeCo, osservandoli sulla base di parametri linguistici più complessi, ci siamo serviti di READ-IT<sup>6</sup>.

READ-IT è uno strumento automatico per l'analisi della leggibilità di un testo, appartenente alla seconda generazione di misuratori di leggibilità, che non fanno affidamento unicamente alle caratteristiche generali e formali del testo, come il *Farfalla Project*, ma considera altri livelli di analisi più complessi e classificati nei seguenti modelli:

1. MODELLO BASE che valuta le caratteristiche, tradizionalmente usate nelle formule di leggibilità come GULPEASE, ovvero la lunghezza della frase e la lunghezza delle parole, che corrispondono al testo non elaborato.
2. MODELLO LESSICALE che consente una valutazione combinata tra caratteristiche di base con quelle propriamente lessicali, che misurano la composizione del

<sup>6</sup> È uno strumento sviluppato dall'*Italian Natural Language Processing Laboratory* (ItaliaNLP Lab) dell'Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) del CNR di Pisa.

vocabolario. Il modello lessicale prende come riferimento il *Grande Dizionario Italiano dell'uso – GRADIT* e calcola: a) le percentuali di parole appartenenti al VdB e la loro distribuzione sulla base del vocabolario fondamentale, di alto uso e di alta disponibilità; b) il rapporto *type/token*, che misura la ricchezza lessicale del testo e la densità lessicale, ovvero la proporzione di parole piene sul totale delle occorrenze.

3. MODELLO MORFOSINTATTICO, le cui caratteristiche morfosintattiche riguardano il modello statistico del linguaggio basato sulle probabilità di distribuzione delle categorie grammaticali nel testo e la distribuzione dei modi verbali.
4. MODELLO SINTATTICO che combina tutti i tipi di caratteristiche precedenti con quelle sintattiche e che misura la probabilità incondizionata dei diversi tipi di dipendenze sintattiche, le caratteristiche della profondità dell'albero sintattico (parametro rilevante per valutare la complessità di un testo), la percentuale di proposizioni subordinate rispetto alle principali e la posizione delle subordinate rispetto alla principale, le caratteristiche dei predicati verbali, la lunghezza delle relazioni di dipendenza, calcolata come la distanza in parole tra la testa e il dipendente.

La complessità di questi modelli e la gradazione della qualità della loro analisi linguistica aumenta man mano che l'analisi della leggibilità si sposta su aspetti legati alla comprensibilità.

Infatti, il programma READ-IT consente di valutare la leggibilità, non solo del testo intero, ma anche di ogni singola frase contenuta nel testo e di identificare in modo automatico i punti di complessità e le difficoltà delle singole frasi, in termini di comprensibilità. Ciò è importante anche alla luce dei possibili sviluppi della ricerca, come vedremo nel § 6, in quanto ci consente di ricalcolare la leggibilità del *corpus* LeCo:

1. a livello testuale, per misurare la comprensibilità nei quattro modelli di analisi linguistica di READ-IT;
2. a livello frasale, per individuare i punti critici, che inficiano la comprensione o rendono il testo difficile, e operare azioni di semplificazione e di riscrittura.

Questi due passaggi, se applicati ai testi di LeCo con indici di leggibilità GULPEASE bassi, potrebbero migliorare la leggibilità e allo stesso tempo la loro comprensibilità.

## 6. I NOSTRI SVILUPPI FUTURI: GLI STUDENTI BES

Cercando soluzioni, strumenti e strategie didattiche nuove e più inclusive per facilitare gli studenti BES e inaugurare un percorso didattico adeguato a tutti gli studenti con difficoltà nella comprensione del testo, l'analisi e il calcolo degli indici di leggibilità GULPEASE in LeCo hanno evidenziato nuove azioni di studio e di ricerca:

1. Semplificazione dei testi che hanno indici di leggibilità bassi con un livello di frustrazione alto (<35);
2. Calcolo della leggibilità delle domande e delle risposte dei questionari, creati per valutare e misurare le competenze della *literacy* degli studenti;
3. Riformulazione dei questionari con indici di leggibilità e comprensibilità bassi;
4. Scelta del font *Easyreading* per la redazione dei testi e dei questionari per migliorare la leggibilità grafica, non solo linguistica del *corpus* LeCo;
5. Creare la versione orale di LeCo: *Ascoltare e Comprendere* (A&C).

Introdurre la modalità di ascolto in progetti incentrati sulla comprensione testuale come LeCo, significa da un alto considerare l'ascolto un facilitatore del comprendere,

dall'altro costruire progetti come A&C più orientati a una didattica dell'ascolto. Adottare strumenti multimodali (audiolettura, sintesi vocale con lettura ad effetto karaoke, audiolibri, *podcast*) inoltre in ausilio al testo scritto favorisce il processo di comprensione. La multimodalità garantisce infatti il riconoscimento più immediato di appigli visivi o auditivi in grado di esplicitare meglio i contenuti e influisce positivamente sul grado di leggibilità e comprensibilità dei contenuti del testo sia scritto, sia orale, conferendo allo studente una maggiore consapevolezza nell'abilità trasversale della comprensione.

Le attività di comprensione ipotizzate nel progetto A&C saranno diversificate a seconda degli scopi comunicativi, suddivisi in diversi livelli di difficoltà sulla base dei criteri di competenza dell'ascolto e delle sotto-abilità, senza perdere mai di vista il livello di acquisizione dello studente e il grado di comprensione. Il nostro auspicio è che il progetto A&C possa migliorare e potenziare ancora di più le prestazioni degli studenti nella *literacy* e educare ad un altro modo di affrontare la comprensione del testo attraverso l'ascolto. Un esempio di tale consapevolezza potrebbe derivare dalla scelta come ipotetica attività, in A&C, della tecnica della lettura rallentata, fatta da un esperto o da un docente, con l'obiettivo di far riflettere lo studente sull'organizzazione logico-concettuale del testo in ascolto e sulle diverse operazioni cognitive simultanee, necessarie e messe in gioco nel comprendere.

Alla luce di queste osservazioni il progetto di ricerca A&C si inserisce nell'ambito dell'educazione linguistica, con lo scopo di individuare strategie innovative didattiche, multimodali, che possano rendere il complesso processo della comprensione più accessibile e sempre più inclusivo nella scuola di oggi, al fine di dare maggiore consapevolezza alla promozione di una didattica per l'educazione alla comprensione, che non sia più incentrata solo sullo scritto, ma che preveda anche interventi didattici mirati sul parlato e sull'ascolto (Voghera, Sammarco, 2021).

## 7. CONCLUSIONI

La lettura e la comprensione sono abilità fondamentali e importanti per accedere ai nuovi saperi, per promuovere la crescita personale e per interagire in modo adeguato nell'ambiente sociale. Tuttavia, nella competenza della *literacy* si possono incontrare difficoltà che inficiano il processo di comprensione. Per evitare che queste difficoltà permangano negli adulti, come mostrano gli ultimi rapporti PIAAC/OCSE<sup>7</sup>, sarebbe necessario affrontarle in tempo e durante il percorso formativo scolastico. Al fine di migliorare questi risultati, pertanto, sarebbe opportuno promuovere una didattica della lettura e della comprensione, più consapevole ed esplicita, nell'ambito dell'educazione linguistica.

In queste pagine si è cercato di dimostrare che la leggibilità e la comprensibilità sono due facce della stessa medaglia e per questo motivo tra gli strumenti che possono contribuire ad una migliore didattica della comprensione vi sono gli indici di leggibilità. Essi permettono di valutare il livello di complessità o difficoltà nella leggibilità e comprensibilità dei testi (vd *Farfalla Project* e READ-IT), di individuare i testi ad alta leggibilità e di segnalare le parti di testo su cui operare processi di semplificazione per renderle più accessibili.

Questi strumenti potrebbero essere utili anche ai docenti di altre materie per renderli maggiormente consapevoli della trasversalità della competenza di comprensione, e si potrebbero ideare percorsi formativi, in grado di trasformare i docenti in "facilitatori" e mediatori nello sviluppo della competenza *literacy* a tutti i livelli e gradi di scuola. In tal

<sup>7</sup> Vd. Voghera in questa monografia.

modo i docenti potrebbero selezionare con cura gli *input* e definire gli obiettivi didattici curriculari in modo chiaro e diversificato, graduando le difficoltà nella comprensione sulla base della leggibilità e comprensibilità. Gli insegnanti acquisirebbero in questo modo gli strumenti adatti a monitorare e affrontare i problemi di svantaggio linguistico sia degli studenti BES sia degli studenti più deboli in *literacy*, e di volta in volta sarebbero in grado di modificare la propria strategia didattica per adattarla ai sempre nuovi bisogni educativi.

## RIFERIMENTI BIBLIOGRAFICI

- Ambel M. (2006), *Quel che ho capito*, Carocci, Roma.
- Agati A. (1999), *Abilità di lettura*, Paravia, Torino.
- Balboni P. (2018), *Fare educazione linguistica. Insegnare italiano, lingue straniere e classiche*, UTET Università, Torino.
- Basile G. (2016), *Descrittori di competenze (A-E). Selezione di attività di sviluppo di determinate abilità*, Manoscritto, Università di Salerno.
- Cardinale M., Giachino G. (1981), *La lettura*, Zanichelli, Bologna.
- Cardona M. (2008), “L’abilità di lettura e lo sviluppo della competenza lessicale”, in *Studi di Glottodidattica*, 2, pp.10-36.
- Colombo A., Romani W. (a cura di) (1996), “«È la lingua che ci fa uguali». Lo svantaggio linguistico: problemi di definizione e di intervento”, Quaderni del Giscel, La Nuova Italia, Firenze.
- Colombo A. (2002), *Capire e non capire*, Zanichelli, Bologna.
- Della Casa M. (1991), *La comprensione dei testi. Modelli e proposte per l’insegnamento*, FrancoAngeli, Milano.
- Dell’Orletta F., Montemagni S., Venturi G. (2011), “READ-IT: asseggging readability of Italian texts with a view to text simplification”, in *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2011)*, Edimburgh, 30 luglio, pp. 73-83.
- Ferreri S. (2002), *Non uno di meno. Strategie didattiche per leggere e comprendere*, Quaderni del Giscel, La Nuova Italia, Firenze.
- De Mauro T., Piemontese M. E., Vedovelli, M. (a cura di) (1986), *Leggibilità e comprensione*, Atti dell’incontro di studio, Roma, Istituto di Filosofia, Villa Mirafiori, 26-27 giugno, in *Linguaggi*, III, 3.
- De Mauro T. (1994), *Capire le parole*, Laterza, Bari.
- De Mauro T. (2016), *Il nuovo vocabolario di base della lingua italiana*, in *Internazionale*: <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana>.
- Klare G. R. (1984), *Assessing Readability*, Pearson.
- Lorenzi A. (2002), “La varietà delle operazioni di comprensione”, in Ferreri S. (a cura di), *Non uno di meno. Strategie didattiche per leggere e comprendere*, Quaderni del Giscel, La Nuova Italia, Firenze, pp. 131-151.
- Lucisano P., Piemontese M. E. (1988), “Gulpease. Una formula per la predizione della difficoltà dei testi in lingua italiana”, in *Scuola e Città*, 3, pp. 57-68.
- Lumbelli L. (1996), “Il problema della soglia tra comprensione e incomprensione: linguistica e psicologia cognitivista”, in T. De Mauro *et al.* (a cura di), *Leggibilità e comprensione*, Atti dell’incontro di studio, Roma, Istituto di Filosofia, Villa Mirafiori, 26-27 giugno, in *Linguaggi*, III, 3, pp. 17-25.
- Miller G. A. (1972), *Linguaggio e Comunicazione*, La Nuova Italia, Firenze.

- Piemontese M. E. (1996), *Capire e farsi capire. Teoria e tecniche di scrittura controllata*, Tecnodid, Napoli.
- Pietrandrea P. (2021), *Comunicazione, dibattito pubblico, social media. Come orientarsi con la linguistica*, Carocci, Roma.
- Van Dijk T. A., Kintsch W. (1983), *Strategies of discourse comprehension*, Academic Press, New York.
- Vedovelli M. (1994), *La lingua italiana d'uso. Morfosintassi del parlato e dello scritto*, Progetto MILIA, modulo 10, Ministero della Pubblica Istruzione, Direzione Generale.
- Voghera M., Sammarco C. (2021), *Ascoltare e parlare. Idee per la didattica*, Franco Cesati Editore, Firenze.
- Zambelli M. L. (2014), "Semplificare i testi di studio: quando, come", in *Italiano LinguaDue*, 6, 1, pp. 327-341:  
<https://riviste.unimi.it/index.php/promoitals/article/view/4236>.