

# READ-IT

## *Documentazione Demo online*

ILC-CNR

ItaliaNLP Lab

[www.italianlp.it](http://www.italianlp.it)

### *Sommario*

<b>1</b>	<b>MISURAZIONE DELLA LEGGIBILITÀ DI UN TESTO</b>	<b>2</b>
<b>1.1</b>	<b>Analisi globale della leggibilità</b>	<b>3</b>
1.1.1	Valutazione globale della leggibilità del documento	3
1.1.2	Risultati del monitoraggio linguistico del testo	5
<b>2</b>	<b>PROIEZIONE DELLA LEGGIBILITÀ SUL TESTO</b>	<b>9</b>
<b>2.1</b>	<b>Identificazione dei luoghi di complessità del testo a livello di frase</b>	<b>10</b>

# 1 Misurazione della leggibilità di un testo

Gli ultimi anni hanno visto il progressivo affermarsi a livello internazionale del ricorso a tecnologie linguistico-computazionali per la misurazione automatica della leggibilità di un testo. A differenza dei metodi sino ad oggi adottati, come ad esempio la formula Flesch-Kincaid, utilizzata per la lingua inglese, o l'indice Gulpease per la lingua italiana, questa seconda generazione di misuratori di leggibilità non fa affidamento unicamente su caratteristiche generali e formali del testo, quali la lunghezza della frase e la lunghezza delle parole. L'utilizzo di strumenti di annotazione linguistica automatica permette infatti di definire la leggibilità di un testo sulla base di parametri linguistici più complessi e che fino ad ora sembravano essere inattuabili se non attraverso un accurato lavoro manuale. Tali parametri spaziano tra i vari livelli di analisi linguistica e sono rintracciati in modo automatico a partire dall'output del processo di annotazione automatica del testo. I diversi studi condotti sulla base di questo approccio avanzato alla misurazione della leggibilità di un testo differiscono tra di loro innanzitutto rispetto alla tipologia di parametri linguistici considerati. Mentre i primi studi in materia facevano affidamento su parametri lessicali, più recentemente - considerata la crescente affidabilità degli strumenti di annotazione sintattica del testo - l'attenzione si è rivolta verso l'utilizzo di parametri sintattici. Studi ancora più recenti si basano sull'uso di parametri relativi al livello di analisi del discorso e sulla combinazione di parametri di varia natura.

Per quanto riguarda la lingua italiana, il primo e al momento unico strumento avanzato basato su questi presupposti è rappresentato da READ-IT<sup>1</sup>, che è stato concepito per fornire anche un supporto alla redazione semplificata di un testo attraverso l'identificazione dei luoghi di complessità. READ-IT implementa un indice di leggibilità "avanzato" basato su analisi linguistica multi-livello del testo condotta con strumenti che rappresentano lo stato dell'arte per il trattamento automatico della lingua italiana: le tecnologie linguistiche utilizzate sono costituite da "AnIta", una piattaforma ormai consolidata e ampiamente sperimentata di metodi e strumenti per il trattamento automatico dell'italiano sviluppati presso l'Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC-CNR) che ha ricevuto ampia validazione sia nell'ambito di progetti di ricerca volti all'estrazione di informazione linguistica da corpora testuali, sia in progetti di carattere applicativo finalizzati all'estrazione di conoscenza di dominio.

READ-IT è basato su una combinazione di tratti linguistici che spaziano tra diversi livelli di descrizione linguistica: lessicale, morfo-sintattico e sintattico. Un tratto caratterizzante di READ-IT, innovativo rispetto alla letteratura internazionale in materia, consiste in una valutazione della leggibilità articolata su due livelli: il documento e la singola frase. La valutazione rispetto alla frase rappresenta un'importante novità dell'approccio sottostante a READ-IT, che riveste un ruolo centrale quando la valutazione della leggibilità sia finalizzata alla semplificazione del testo. Attraverso l'identificazione dei luoghi di complessità del testo (in termini di frasi) che necessitano di revisione e semplificazione, accompagnata da una classificazione semantica del tipo di difficoltà riscontrata (di naturale lessicale vs grammaticale), READ-IT può anche essere utilizzato come ausilio per la semplificazione del testo.

Quanto segue documenta le informazioni restituite da READ-IT nella versione demo accessibile online.

---

<sup>1</sup> DELL'ORLETTA F., MONTEMAGNI S., VENTURI G. (2011) READ-IT: Assessing Readability of Italian Texts with a View to Text Simplification, in Proceedings of the Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2011), Edinburgh, July 30, pp. 73-83, <http://aclweb.org/anthology/W/W11/W11-2308.pdf>.

Testo da analizzare	Suddivisione in frasi	Suddivisione in token	Parti del discorso	Annotazione	Analisi globale della leggibilità	Proiezione della leggibilità sul testo
indice di leggibilità		livello di difficoltà				
READ-IT Base		73,5%				
READ-IT Lessicale		50,0%				
READ-IT Sintattico		2,2%				
READ-IT Globale		26,2%				
indice di leggibilità		livello di semplicità				
GULPEASE		49,9				
<b>[+] [-] Caratteristiche estratte dal testo</b>						
<b>[+] Profilo di base</b>						
<b>[+] Profilo lessicale</b>						
<b>[+] Profilo sintattico</b>						

Figura 1 – La scheda corrispondente alla valutazione globale della leggibilità del documento

## 1.1 Analisi globale della leggibilità

Con analisi globale della leggibilità si intende il risultato dell’analisi condotta in relazione al singolo documento. Tale risultato si articola in due sezioni distinte dedicate a:

- la valutazione della leggibilità del documento effettuata da diversi modelli di analisi basati su diversi tipi di informazione, che potremmo vedere come diversi indici di leggibilità;
- la ricostruzione del profilo linguistico del testo, condotta in relazione a un sottoinsieme dei parametri utilizzati da READ-IT per la valutazione della sua leggibilità, articolati secondo il livello di descrizione linguistica di appartenenza. Questa seconda sezione è tesa a fornire elementi di analisi utili a comprendere i risultati riportati nella prima sezione: si tratta di informazioni utili per il linguista e il linguista computazionale che permettono di monitorare il funzionamento del sistema ed eventualmente correggerlo.

La Figura 1 riporta la scheda corrispondente all’analisi globale della leggibilità del singolo documento all’interno dell’interfaccia web sviluppata, dove si distinguono chiaramente la sezione che riporta i risultati dei diversi modelli di analisi della leggibilità (in alto) da quella che fornisce informazione sulle caratteristiche estratte dal testo (in basso).

### 1.1.1 Valutazione globale della leggibilità del documento

Oggi, il paradigma dominante nei compiti di annotazione linguistica è rappresentato da sistemi basati su algoritmi di apprendimento automatico supervisionato. Questo è anche l’approccio seguito nei sistemi “avanzati” per la valutazione della leggibilità di un testo che viene così riformulata come un compito di classificazione probabilistica. Tre sono gli “ingredienti” fondamentali di questo tipo di approcci: l’insieme

delle categorie linguistiche da assegnare (i livelli di leggibilità, nel nostro caso), il corpus di apprendimento (ovvero un insieme di esempi pre-classificati a mano rispetto alle categorie (di leggibilità, in questo caso) da riconoscere automaticamente), e un insieme di tratti descrittivi accuratamente selezionati sulla base del compito di classificazione da svolgere.

Anche READ-IT si basa su questo tipo di approccio: esso opera sul testo arricchito con informazione linguistica (in particolare morfo-sintattica e sintattica) e conduce una classificazione probabilistica del testo rispetto a due classi (leggibile vs complesso) sulla base di informazione lessicale, morfo-sintattica e sintattica. In READ-IT la selezione dei parametri di analisi è stata guidata da due classi di criteri, da un lato il ruolo svolto nella valutazione della leggibilità di un testo, dall'altro la computabilità su larga scala e in modo affidabile mediante tecnologie linguistico-computazionali: per quanto il secondo aspetto possa apparire del tutto esterno e contingente, non può né deve essere sottovalutato in quanto da esso dipende l'affidabilità dei risultati raggiunti.

I livelli di leggibilità riconosciuti da READ-IT sono due in quanto per la lingua italiana non era disponibile un corpus di testi differenziati per livelli di leggibilità di difficoltà crescente da usarsi in fase di addestramento del sistema: abbiamo quindi usato per lo sviluppo di un primo prototipo del sistema due corpora rappresentativi di testi complessi e semplificati, appartenenti allo stesso genere testuale (prosa giornalistica), costituiti rispettivamente da *La Repubblica* (Rep) e *Due Parole* (2Par), un periodico di "facile lettura" destinato a lettori caratterizzati da un basso livello di scolarizzazione e/o da lievi disagi linguistico-cognitivi<sup>2</sup>.

La valutazione globale della leggibilità del testo viene condotta sulla base di diverse configurazioni di caratteristiche del testo: i risultati sono riportati nella parte superiore della scheda. Segue una breve caratterizzazione dei diversi modelli di analisi della leggibilità menzionati:

- READ-IT BASE: in questo modello, le caratteristiche considerate sono quelle tipicamente usate nelle misure tradizionali della leggibilità di un testo, ovvero la lunghezza della frase, calcolata come numero medio di parole per frase, e la lunghezza delle parole, calcolata come numero medio di caratteri per parola. Questo modello può essere visto come un'approssimazione delle misure tradizionali di leggibilità, in particolare dell'indice Gulpease, specificamente concepito per la lingua italiana;
- READ-IT LESSICALE: questo modello si focalizza sulle caratteristiche lessicali del testo, costituite dalla composizione del vocabolario così come dalla sua ricchezza lessicale (v. sotto);
- READ-IT SINTATTICO: questo modello si basa su informazione di tipo grammaticale, ovvero sulla combinazione di tratti morfo-sintattici e sintattici desunti dai corrispondenti livelli di analisi linguistica (v. sotto);
- READ-IT GLOBALE: si tratta di un modello basato sulla combinazione di tratti di varia natura, che spaziano dalle caratteristiche generali del testo del modello READ-IT BASE a quelle lessicali e sintattiche degli altri due modelli.

Per ciascun modello, la percentuale esprime il livello di difficoltà, ovvero si riferisce alla probabilità di appartenenza del testo in esame alla classe dei testi di difficile leggibilità: la barra a fianco esprime visivamente questo valore, dove il rosso rappresenta la probabilità di appartenenza alla classe dei testi difficili e il verde a quelli di facile lettura.

---

<sup>2</sup> PIEMONTESE M.E. (1996), *Capire e farsi capire. Teorie e tecniche della scrittura controllata*, Napoli, Tecnodid.

[+] [-] Caratteristiche estratte dal testo	
<b>[+] Profilo di base</b>	
Numero totale periodi:	17
Numero totale parole:	445
Lunghezza media dei periodi (in parole):	26,2
Lunghezza media delle parole (in caratteri):	5,0

Figura 2 – La parte della scheda che riporta informazione relativa al profilo di base del documento

### 1.1.2 Risultati del monitoraggio linguistico del testo

La seconda sezione della scheda corrispondente alla valutazione globale della leggibilità del documento fornisce i risultati del monitoraggio di un sottoinsieme (selezionato come significativo) delle caratteristiche linguistiche utilizzate da READ-IT nella misurazione della leggibilità. Come anticipato in precedenza, si tratta di informazioni che non sono destinate all'utente finale del sistema.

Le riportiamo di seguito, organizzate in tre sezioni, corrispondenti ai profili di base, lessicale e sintattico del testo, corredate da una glossa esplicativa.

#### **Profilo di base**

- *Numero totale periodi*: ovvero, il numero di periodi (o frasi) in cui si articola il testo analizzato. Si considerano separatori di periodo la punteggiatura forte (".", "?" e "!") e il ritorno a capo;
- *Numero totale parole (tokens)*: ovvero, il numero di occorrenze di parole (o tokens) in cui si articola il testo analizzato;
- *Lunghezza media dei periodi (in tokens)*: questo dato fornisce la lunghezza media dei periodi, espressa in tokens, all'interno analizzato;
- *Lunghezza media delle parole (in caratteri)*: questo dato fornisce la lunghezza media delle parole, espressa in caratteri, all'interno del testo analizzato.

Nel caso della lunghezza media dei periodi e delle parole, oltre al valore numerico, viene fornita una rappresentazione grafica che mette a confronto il dato relativo al testo oggetto dell'analisi (corrispondente alla barra azzurra) con la corrispondente informazione rilevata nei corpora di riferimento di facile (barra verde) e difficile (barra rossa) lettura. Il rettangolino a fianco fornisce una classificazione semantica del dato rilevato in relazione al testo oggetto dell'analisi.

[+] [-] Caratteristiche estratte dal testo		
<b>[+] Profilo di base</b>		
<b>[-] Profilo lessicale</b>		
<i>Composizione del vocabolario</i>		
Percentuale di lemmi appartenente al Vocabolario di Base (VdB):	76,9%	
Ripartizione dei lemmi riconducibili al VdB rispetto ai repertori d'uso:		
Fondamentale:	82,5%	
Alto uso:	16,1%	
Alta disponibilità:	1,4%	
Rapporto tipo/unità (calcolato rispetto alle prime 100 parole del testo):	0,710	
Densità Lessicale:	0,563	
<b>[+] Profilo sintattico</b>		

Figura 3 – La parte della scheda che riporta informazione relativa al profilo lessicale del documento

### Profilo lessicale

- *Composizione del vocabolario*: si tratta di un parametro complesso che riguarda la tipologia del vocabolario usato, ovvero l'insieme delle parole tipo che ricorrono all'interno del documento. Come dizionario di riferimento, abbiamo preso il *Grande Dizionario Italiano dell'uso* (GRADIT, De Mauro, 2000), e in modo particolare l'insieme dei lemmi riconducibili al cosiddetto "vocabolario di base" (d'ora in avanti, VdB). Nella prima riga, è riportata la percentuale di vocabolario del testo appartenente al VdB.

Ulteriori indicazioni in merito alla composizione del vocabolario possono essere ricavate dall'analisi della ripartizione della porzione del vocabolario del testo riconducibile al VdB rispetto ai repertori di uso (riportata nelle righe successive):

- "Fondamentale" (FO), che include circa 2000 parole conosciute e usate da coloro che hanno almeno un'istruzione elementare;
- "Alto uso" (AU), che include circa 3000 parole conosciute e usate da coloro che hanno almeno un'istruzione media (licenza media inferiore);
- "Alta disponibilità" (AD), ovvero circa 2000 parole altamente latenti, presenti all'uso che i parlanti non usano concretamente tutti i giorni, ma solo all'occorrenza.
- *Rapporto tipo/unità (calcolato rispetto alle prime 100 parole del testo)*: questa misura, conosciuta anche come "Type/Token Ratio" (o TTR), rappresenta uno dei metodi più diffusi per misurare la varietà lessicale di un testo. Questo indice mette in rapporto il numero delle occorrenze delle unità del vocabolario di un testo (al denominatore) con il numero di parole tipo (al numeratore): i valori oscillano tra 0 e 1, dove valori vicini allo 0 indicano che il vocabolario del testo è meno vario mentre valori vicini a 1 caratterizzano testi particolarmente variegati dal punto di vista lessicale. Essendo TTR un indice sensibile alla lunghezza del testo, per rendere confrontabili i dati questo indice è calcolato rispetto a campioni di testo della stessa lunghezza (nella versione corrente tale limite è stato fissato alle prime 100 parole unità di un testo).

- **Densità Lessicale:** questo parametro riguarda il rapporto tra parole piene (ovvero portatrici di significato) e parole funzionali all'interno di un testo, e in modo particolare la sua "densità lessicale" (abbreviata come DL) calcolata come la proporzione delle parole semanticamente "piene" (ovvero, nomi, aggettivi, verbi e avverbi) rispetto al totale delle occorrenze di parola all'interno del testo. Stando alla letteratura, valori più alti di DL corrispondono in linea di massima a maggiore leggibilità.

In tutti i casi, oltre al valore numerico del parametro viene fornita una rappresentazione grafica che mette a confronto il dato relativo al testo oggetto dell'analisi (corrispondente alla barra azzurra) con la corrispondente informazione rilevata nei corpora di riferimento di facile (barra verde) e difficile (barra rossa) lettura. Il rettangolino a fianco fornisce una classificazione semantica del dato rilevato in relazione al testo oggetto dell'analisi.

### Profilo sintattico

Il profilo sintattico di un testo si articola in due parti, a seconda che l'informazione monitorata riguardi l'analisi morfo-sintattica del testo (ovvero il livello in cui a ogni "token" del testo viene associata informazione relativa alla categoria grammaticale che la parola ha nel contesto specifico) o la struttura sintattica sottostante (basata su una descrizione della frase in termini di relazioni di dipendenza tra parole, come "soggetto", "oggetto diretto", "modificatore", etc.).

In tutti i casi, oltre al valore numerico del parametro viene fornita una rappresentazione grafica che mette a confronto il dato relativo al testo oggetto dell'analisi (corrispondente alla barra azzurra) con la corrispondente informazione rilevata nei corpora di riferimento di facile (barra verde) e difficile (barra rossa) lettura. Il rettangolino a fianco fornisce una classificazione semantica del dato rilevato in relazione al testo oggetto dell'analisi.

[+] [-] Caratteristiche estratte dal testo	
<b>[+] Profilo di base</b>	
<b>[+] Profilo lessicale</b>	
<b>[-] Profilo sintattico</b>	
<i>"Misura" delle categorie grammaticali (%)</i>	
Sostantivi:	18,7%
Nomi Propri:	8,1%
Aggettivi:	7,9%
Verbi:	14,6%
Congiunzioni:	3,8%
Coordinanti:	70,6%
Subordinanti:	29,4%

Figura 4 – La sottoparte della scheda che riporta informazione relativa al profilo sintattico del documento, dedicata alle proprietà morfo-sintattiche del testo

- “Misura” delle categorie grammaticali: che la distribuzione delle categorie grammaticali rifletta differenze e somiglianze tra diversi generi testuali così come tra la lingua scritta e parlata è un fatto ampiamente riconosciuto nella letteratura linguistica. Questo tipo di informazione è stata anche usata come parametro rilevante nella misurazione della leggibilità in lavori recenti condotti in relazione alla lingua inglese e portoghese. In questa sede, riportiamo i valori relativi a un sottoinsieme di categorie grammaticali, ovvero sostantivi (distinguendo tra nomi comuni e propri), aggettivi, verbi e congiunzioni.

Per quanto riguarda le congiunzioni, viene riportata la loro ripartizione in coordinanti e subordinanti. Questo dato fornisce una prima e molto approssimativa indicazione del rapporto tra costruzioni paratattiche e ipotattiche (queste ultime considerate in letteratura come un fattore complessità) all’interno del testo analizzato.

Struttura sintattica a dipendenze		
Articolazione interna del periodo:		
Numero medio di proposizioni per periodo:	3,588	
Proposizioni principali vs coordinate (%):		
Principali:	53,8%	
Subordinate:	46,2%	
Articolazione interna della proposizione:		
Numero medio di parole per proposizione:	7,295	
Numero medio di dipendenti per testa verbale:	1,754	
"Misura" della profondità dell'albero sintattico:		
Media delle altezze massime:	7,118	
Profondità media di strutture nominali complesse:	1,314	
Profondità media di "catene" di subordinazione:	1,000	
"Misura" della lunghezza delle relazioni di dipendenza (calcolata come distanza in parole tra testa e dipendente):		
Lunghezza media:	2,544	
Media delle lunghezze massime:	13,941	

Figura 5 – La sottoparte della scheda che riporta informazione relativa al profilo sintattico del documento, dedicata alle proprietà della struttura a dipendenze del testo

- *Articolazione interna del periodo*: questo costituisce un parametro complesso volto a caratterizzare l’organizzazione interna del periodo. Esso include informazioni come:
  - a. il numero medio di proposizioni per periodo: si tratta di un dato elementare, ma già significativo, costituito dal rapporto tra proposizioni e periodi. Chiaramente, con l’aumentare di questo valore cresce la complessità sintattica del testo (Rep registra una media di 2,41 proposizioni per periodo contro l’1,64 di 2Par): questo dato, però, non dice ancora nulla su come le diverse proposizioni si rapportino l’una con l’altra all’interno del periodo;

- b. proposizioni principali vs subordinate: questo dato registra la proporzione di principali e subordinate; si noti che un altro uso di costruzioni subordinate contribuisce in modo significativo alla complessità grammaticale del testo;
- *Articolazione interna della proposizione*: è descritta nei termini di
  - a. numero medio di parole per proposizione;
  - b. numero medio di dipendenti per testa verbale;
- *“Misura” della profondità dell’albero sintattico*: un altro aspetto rilevante per la misura della complessità del testo riguarda i livelli di incassamento gerarchico: in presenza di più di una proposizione subordinata all’interno dello stesso periodo, diventa cruciale ricostruire quale tipo di rapporto sussista tra di esse, ovvero se siano ricorsivamente incassate l’una all’interno dell’altra.

Una prima e approssimativa indicazione dei livelli di incassamento gerarchico all’interno della struttura sintattica può essere ricostruita a partire dall’altezza massima dell’albero, che misura la massima distanza che intercorre tra una foglia (rappresentata da parole del testo senza dipendenti) e la radice dell’albero, espressa come numero di archi (ovvero relazioni di dipendenza) attraversati nel cammino foglia-radice. Il parametro “media delle altezze massime” riporta il valore medio delle altezze massime degli alberi a dipendenza registrate all’interno del testo analizzato.

Questa misura viene raffinata ulteriormente focalizzandosi su particolari tipi di costrutti sintattici:

- a. la ricorrenza di strutture nominali complesse costituite da una testa nominale modificata da aggettivi e/o complementi preposizionali: un’alta frequenza di “catene” particolarmente “pesanti” di modificatori nominali rappresenta senza dubbio un fattore di complessità sintattica (questo dato è riportato sotto la voce “Profondità media di strutture nominali complesse”, che registra la media delle profondità di strutture nominali con modificatori);
  - b. la ricorrenza di proposizioni subordinate ricorsivamente incassate (questo dato è riportato come “Profondità media di ‘catene’ di subordinazione”).
- *“Misura” della lunghezza delle relazioni di dipendenza*: è noto che la contiguità di elementi semanticamente e/o sintatticamente ‘vicini’ permette una più immediata recuperabilità e accessibilità dei rapporti sussistenti tra le parole. La “lunghezza” delle relazioni di dipendenza, calcolata come la distanza in parole tra la testa e il dipendente, rappresenta dunque un fattore di complessità ampiamente riconosciuto nella letteratura linguistica, psicolinguistica e linguistico-computazionale. Questo aspetto della struttura sintattica viene monitorato attraverso due parametri, corrispondenti alla media della lunghezza di tutte le relazioni di dipendenza (con esclusione dei legami riguardanti al punteggiatura), e alla media dei legami di dipendenza più lunghi per ciascuna frase.

## 2 Proiezione della leggibilità sul testo

READ-IT fornisce un duplice supporto nell’identificazione e classificazione semantica dei luoghi di potenziale complessità del testo: in prima battuta, viene condotta la valutazione della leggibilità delle singole frasi sulla base di un modello statistico.

## 2.1 Identificazione dei luoghi di complessità del testo a livello di frase

La scheda “Proiezione della leggibilità sul testo” è stata concepita per fornire un supporto al redattore del testo e guidarlo – ove necessario - nel processo di revisione e semplificazione. In particolare, all’interno di questa scheda è possibile identificare i periodi che necessitano di revisione e il tipo di difficoltà riscontrata in relazione ad essi. Come si può osservare in Figura 6, per ciascun periodo viene riportato il livello di “difficoltà lessicale” e “difficoltà grammaticale” in due colonne distinte, computato da due modelli di analisi della leggibilità distinti, rispettivamente basati su caratteristiche lessicali e sintattiche del testo: questi modelli rappresentano la versione corrispondente ai modelli di leggibilità globale denominati “READ-IT Lessicale” e “READ-IT Sintattico” (cfr. sezione 1.1.1).

Il livello di difficoltà è rappresentato cromaticamente mediante colori che vanno dal verde (testo leggibile) al rosso (testo particolarmente difficile): il rosso così come sfumature giallo-arancioni marcano periodi che necessitano di revisione. Quando la difficoltà si situa al livello lessicale (prima colonna), viene richiesta una revisione delle parole usate nel testo: l’identificazione delle parole incriminate avviene al passo successivo (vedi sotto). Quando invece la difficoltà riguarda il livello grammaticale (seconda colonna), il redattore è invitato a riformulare il periodo facendo uso di strutture grammaticali più semplici (ad esempio, evitando l’uso ripetuto di strutture subordinate oppure di “catene” di modificazione eccessivamente lunghe).

Testo da analizzare	Suddivisione in frasi	Suddivisione in token	Parti del discorso	Annotazione	Analisi globale della leggibilità	Proiezione della leggibilità sul testo			
SID	frase				base	less.	sint.	glob.	
1.	Art.1								
2.	L'Italia è una Repubblica democratica, fondata sul lavoro.								
3.	La sovranità appartiene al popolo, che la esercita nelle forme e nei limiti della Costituzione.								
4.	Art.2								
5.	La Repubblica riconosce e garantisce i diritti inviolabili dell'uomo, sia come singolo, sia nelle formazioni sociali ove si svolge la sua personalità, e richiede l'adempimento dei doveri inderogabili di solidarietà politica, economica e sociale.								
6.	Art.3								
7.	Tutti i cittadini hanno pari dignità sociale e sono eguali davanti alla legge, senza distinzione di sesso, di razza, di lingua, di religione, di opinioni politiche, di condizioni personali e sociali.								
8.	È compito della Repubblica rimuovere gli ostacoli di ordine economico e sociale, che, limitando di fatto la libertà e l'eguaglianza dei cittadini, impediscono il pieno sviluppo della persona umana e l'effettiva partecipazione di tutti i lavoratori all'organizzazione politica, economica e sociale del Paese.								
9.	Art.4								
10.	La Repubblica riconosce a tutti i cittadini il diritto al lavoro e promuove le condizioni che rendano effettivo questo diritto.								
11.	Ogni cittadino ha il dovere di svolgere, secondo le proprie possibilità e la propria scelta, un'attività o una funzione che concorra al progresso materiale o spirituale della società.								
12.	Art.5								
13.	La Repubblica, una e indivisibile, riconosce e promuove le autonomie locali; attua nei servizi che dipendono dallo Stato il più ampio decentramento amministrativo; adegua i principi ed i metodi della sua legislazione alle esigenze dell'autonomia e del decentramento.								
14.	Art.6								
15.	La Repubblica tutela con apposite norme le minoranze linguistiche.								
16.	Art.7								
17.	Lo Stato e la Chiesa cattolica sono, ciascuno nel proprio ordine, indipendenti e								

Figura 6 – La scheda “Proiezione della leggibilità sul testo”