

UN NUOVO INDICE DI LEGGIBILITÀ PER LA LINGUA ITALIANA

doi.org/ 10.15452/SR.2024.25.0006

ORCID : 0000-0001-9590-1634

Luca Pavan

Vilnius Tech

Lituania

luca.pavan@vilniustech.lt

Riassunto. Gli indici di leggibilità stabiliscono il livello di preparazione scolastica che un lettore deve avere per comprendere un testo e trovano applicazione in numerosi ambiti, dalle scuole alle aziende, ma anche negli ambienti militari. Si tratta di formule che, analizzando un testo, ne ricavano il grado di leggibilità. Occorre infatti conoscere il livello di difficoltà di un testo prima di somministrarlo a un lettore, se si vuole che esso venga compreso. Gli indici aiutano chi realizza i testi a trovare il giusto compromesso tra la leggibilità e la preparazione scolastica dei lettori. In alcuni paesi, come gli Stati Uniti, in alcune istituzioni scolastiche prima di fornire un testo agli studenti si richiede il valore di uno o più indici di leggibilità. Nelle aziende gli indici aiutano a redigere testi comprensibili a una folta platea di clienti, mentre in ambito militare da molti anni danno supporto a chi scrive manuali sull'uso degli armamenti. In questo articolo si descrive la realizzazione di un nuovo indice di leggibilità per l'italiano derivato dall'indice Rix-Pavan, a sua volta derivazione degli indici Rix e Lix. Sebbene l'IA sia in grado di stabilire la leggibilità di un testo, la frequenza di errori e inesattezze non consente, al momento, di ottenere risultati apprezzabili. Il nuovo indice, denominato Rix-Pavan-it, è molto più preciso del Rix-Pavan poiché utilizza il Nuovo Vocabolario di Base o NVdB edito da Tullio de Mauro (2016), al quale sono state aggiunte le principali forme flesse. Una nuova tabella di valutazione è stata redatta comparando i risultati dell'indice (realizzato al computer in una versione scritta in linguaggio Javascript) con una serie cospicua di testi letterari e articoli, scelti anche in base alle letture consigliate dai docenti dei vari ordini di scuola. I risultati sono molto incoraggianti e vengono qui proposti alcuni esempi di testi letterari analizzati e comparati con l'indice Rix-Pavan. Il nuovo indice consente di analizzare, nella versione al computer, testi di qualsiasi tipo, da quelli letterari agli articoli dei giornali. Può essere adottato dalle scuole e dalle pubbliche ammi-

nistrazioni al fine di creare testi che abbiano un alto grado di leggibilità. Inoltre il codice Javascript consente la creazione di un file contenente le parole del testo che non appartengono al NVdB: questa caratteristica è un valido supporto per revisionare il testo, cercando i sinonimi adatti per le parole ritenute difficili.

Parole chiave: indici di leggibilità, Rix-Pavan-it, letteratura in italiano, letteratura italiana in volgare

Abstract. A new readability index for the Italian language. Readability indices determine the level of educational background a reader must have to understand a text, and they are used in various fields, from schools to businesses, as well as in the military. These are formulas that, by analyzing a text, determine its readability level. It is essential to know the level of difficulty of a text before presenting it to a reader if one wants it to be understood. These indices help those who create texts to find the right balance between readability and the readers' educational background. In some countries, such as the United States, certain educational institutions require the value of one or more readability indices before providing a text to students. In businesses, indices help in drafting texts that are understandable to a wide audience of customers, while in the military, readability indices have long supported those writing manuals on the use of weaponry. This article describes the development of a new readability index for Italian, derived from the Rix-Pavan index, which in turn is based on the Rix and Lix indices. Although AI can assess the readability of a text, the frequency of errors and inaccuracies currently prevents obtaining significant results. The new index, named Rix-Pavan-it, is much more accurate than the Rix-Pavan because it uses the Nuovo Vocabolario di Base (New Basic Vocabulary) or NVdB, edited by Tullio De Mauro (2016), to which the main inflected forms have been added. A new evaluation table has been drafted by comparing the index's results (implemented on a computer in a version written in Javascript) with a large set of literary texts and articles, selected also based on the reading recommendations from teachers of various school levels. The results are very encouraging, and some examples of literary texts analyzed and compared with the Rix-Pavan index are presented here. The new index allows, in its computer version, the analysis of any type of text, from literary works to newspaper articles. It can be adopted by schools and public administrations in order to create texts with a high level of readability. Additionally, the Javascript code enables the creation of a file containing words from the text that do not belong to the NVdB: this feature is a valuable aid for revising the text, helping to find suitable synonyms for words deemed difficult.

Keywords: readability indices, Rix-Pavan-it, Italian literature, Italian literature in the vernacular language

1. L'importanza del Vocabolario di Base e la leggibilità

Dalla sua prima edizione del 1980, il Vocabolario di Base (VdB) edito da Tullio de Mauro (De Mauro, 1980) costituisce uno strumento fondamentale per gli studi di linguistica che riguardano, in particolare, l'uso della lingua italiana da parte di un parlante di media istruzione. L'edizione del 2016, il Nuovo Vocabolario di Base (NVdB),¹ ha costituito e costituisce, con i suoi oltre 7000 lemmi, uno strumento importante nella realizzazione di prodotti informatici che aiutano ad analizzare e soprattutto a migliorare la leggibilità di un testo.² Il VdB e il NVdB sono il risultato di studi statistici sulla maggiore o minore frequenza delle parole utilizzate da un parlante medio e danno uno spaccato dell'uso effettivo della lingua nei contesti più comuni (Marello, 2018, pp. 67-68).³

Le parole contenute nel VdB coprono il 95 per cento di quelle usate quotidianamente in Italia (Chiari, 2018, p. 167). Il VdB, quindi, rappresenta “lo specchio della vita sociale e civile degli italiani” (ivi, p. 170). Il livello di istruzione che il VdB e il NVdB propongono è grosso modo quello di una licenza media (Chiari & De Mauro, 2014), ma occorre precisare che le parole conosciute dai parlanti sono molte di più, se si considerano le forme flesse. Qualsiasi strumento informatico che utilizzi il NVdB non può fare a meno di includere anche le forme flesse, *in primis* il plurale dei nomi e i tempi verbali. Il NVdB rappresenta uno strumento utile per agevolare la leggibilità di un testo. La definizione di leggibilità, nel vocabolario Treccani, è la seguente: “Il fatto d'esser leggibile, con riguardo alla scrittura, oppure alla comprensibilità”.⁴ Il meccanismo è semplice: le parole di un testo che non appartengono al NVdB vengono evidenziate e sottoposte a revisione, al fine di scegliere sinonimi che possibilmente siano di uso più comune tra i parlanti di media istruzione. A parte questo, le statistiche che riguardano la frequenza di termini non appartenenti al NVdB forniscono informazioni anche sul livello di difficoltà di un testo: un eccessivo uso di parole estranee al NVdB compromette la leggibilità per la maggioranza dei parlanti.

Molti studi sono stati condotti negli anni sulla leggibilità e vari enti governativi danno spesso indicazioni su come semplificare i testi provenienti dagli uffici delle amministrazioni pubbliche.⁵ La proliferazione degli indici di leggibilità negli Stati Uniti, ma anche negli altri paesi, è la prova che l'esigenza di chiarezza di un testo riguarda una grande moltitudine di lettori e risulta determinante nel lavoro e nelle azioni quotidiane (DuBay, 2004).

La storia degli indici di leggibilità inizia nella prima metà del Novecento, in ambito militare, negli Stati Uniti: i testi redatti per gli eserciti dovevano infatti avere un alto livello

1 Disponibile sul sito web <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana>

2 Un esempio è READ-IT, sul sito web <https://www.ilc.cnr.it/dylanlab/apps/texttools/>

3 Storicamente la prima ricerca dell'elenco di parole più diffuso in una lingua si fa risalire all'Ottocento, con il russo Nikolai Rubakin, il quale compilò una lista di 1.500 parole nella lingua russa che i parlanti erano in grado di capire (Tekfi, 1987).

4 Si veda <https://www.treccani.it/vocabolario/leggibilita/>

5 Si veda, ad esempio, questa direttiva: <https://www.funzionepubblica.gov.it/articolo/dipartimento/24-10-2005/direttiva-materia-di-semplificazione-del-linguaggio>, ultimo accesso 17-10-2024

di chiarezza per evitare equivoci che avrebbero compromesso la sicurezza delle operazioni militari. Sebbene la prima formula in ambito civile sia stata realizzata nel 1923 (ibid.), dagli anni Quaranta si diffonde una formula di largo uso, quella di Rudolph Flesch (Flesch Reading Ease, 1943). Da allora in poi le formule si sono moltiplicate, arrivando al ragguardevole numero di oltre duecento. Non solo: gli indici si sono diffusi in una moltitudine di paesi, molti dei quali hanno aggiunto nuove formule adattandole alle diverse lingue.

Se nella prima metà del Novecento i calcoli per stabilire il valore di un indice venivano fatti a mano, la successiva diffusione dei computer ha semplificato enormemente il processo di calcolo. Ecco anche spiegato perché si è assistito alla straordinaria diffusione di queste formule. Molti indici sono stati quasi dimenticati, altri vengono poco usati, mentre almeno una ventina sono quelli più popolari, utilizzati largamente ancora oggi. Negli Stati Uniti, tra gli indici più frequentemente usati, ci sono l'indice Flesch, l'indice SMOG (1969) e l'indice Coleman-Liau (1975).

Stabilire il livello di comprensione di un testo, come fanno gli indici, significa decidere a quale categoria di lettori poter sottoporre la lettura di quel testo. I testi che si possono esaminare sono di tutti i tipi, da quelli letterari agli articoli dei giornali, ai documenti di aziende e apparati statali. In Italia l'indice più conosciuto è il GULPEASE, ancora largamente utilizzato (Lucisano & Piemontese, 1988). Si tratta di un indice tarato per la lingua italiana e utilizza nella sua formula il conteggio dei caratteri⁶.

La maggioranza degli indici, nelle loro formule, si basa sul conteggio delle frasi, delle sillabe o dei caratteri (Francalanci, 2019, p. 51). Il NVdB rappresenta un'alternativa valida a questo tipo di approccio. Al posto di sillabe e caratteri si utilizza in questo caso la frequenza dei termini non appartenenti al NVdB, come viene descritto qui di seguito.

2. L'indice Rix-Pavan-it

Il NVdB, per realizzare il nuovo indice, è stato esteso manualmente per includere le forme flesse, portando le entrate, da circa 7.000, a circa 86.000. Le forme flesse includono i plurali dei nomi, i tempi verbali, i superlativi dei principali aggettivi (per esempio *bellissimo*, *grandissimi*) e le principali forme enclitiche dei pronomi accoppiati ai verbi (per esempio *prenderlo*, *farmi*).⁷ L'uso di una lista di parole per generare un indice non è un'idea nuova: almeno altri due indici per l'inglese, quello di Dale e Chall (Dale & Chall, 1948) e quello di Spache (Spache, 1953), si basano su liste di parole che dovrebbero essere familiari ai parlanti di un certo livello scolastico (nel caso di Dale e Chall per esempio il livello iniziale dato dalla lista, di circa 3.000 parole, è il quarto anno della scuola elementare) (Chall & Dale, 1995). Dale e Chall danno parecchie indicazioni sul trattamento delle forme flesse, che tuttavia in

6 Conteggiare i caratteri, piuttosto che le sillabe, semplifica il calcolo automatico degli indici al computer.

7 In realtà le agglutinazioni prodotte da verbi e pronomi sono molte di più (Monteleone, 2003), ma sono state scelte solo quelle ritenute più importanti. Si è anche evitato di aggiungere alla lista quelle che interessano i pronomi combinati (ad esempio *prendendomelo*, *fartelo*, *facendomene*), considerando che sono di uso meno comune.

inglese sono in numero molto inferiore rispetto all'italiano.⁸ Si danno indicazioni per escludere dal computo dell'indice anche i nomi propri, che però rende più complicata la realizzazione al computer. Le parole di un testo che non sono incluse nella lista di parole familiari vengono conteggiate all'interno di una formula, il cui risultato stabilisce il livello (all'interno di una tabella di valutazione) che il lettore dovrebbe avere per comprendere quel testo. L'indice Rix-Pavan-it funziona in modo analogo: data una lista di parole (il NVdB esteso), le parole di un testo al di fuori della lista sono conteggiate e il risultato applicato all'interno della formula dell'indice Rix-Pavan, a sua volta derivata dal Rix (Anderson, 1983). La formula originale del Rix è la seguente:

$$\text{Rix} = \text{numero di parole lunghe} / \text{numero di frasi}$$

dove per parole lunghe si intendono quelle con lunghezza maggiore di 6 caratteri. Le parole più lunghe di 6 caratteri vengono considerate difficili. Il Rix, come l'indice svedese Lix da cui deriva (Björnsson, 1968), è un indice multilingua, dunque può essere anche utilizzato per la lingua italiana. Si tratta di un indice la cui formula, che è forse la più semplice se comparata con tutti gli altri indici, fa uso del conteggio dei caratteri, considerando difficili le parole lunghe. Al contrario, nell'indice Rix-Pavan-it, che è specifico per la lingua italiana, si considerano come parole difficili quelle non incluse nel NVdB, per cui la formula diventa come segue:

$$\text{Rix-Pavan-it} = (\text{numero di parole non incluse nel NVdB} / \text{numero di frasi}) - W$$

W (superiore a una certa soglia), come nell'indice Rix-Pavan, è una variabile che descrive anche la punteggiatura: il testo è ritenuto più difficile se la punteggiatura è poco usata. In sostanza la formula è la stessa dell'indice Rix-Pavan (Pavan, 2024), ma al posto delle parole più lunghe di 6 caratteri si utilizzano le parole non appartenenti al NVdB. Per W inferiore a una certa soglia, come nel Rix-Pavan, la formula è la seguente:

$$\text{Rix-Pavan-it} = (\text{numero di parole non incluse nel NVdB} / \text{num. di frasi}) - (W + 1)$$

La formula qui proposta è simile a quella degli indici Rix e Rix-Pavan, ma il contributo della lista del NVdB è fondamentale per ottenere buoni risultati predittivi quando si analizza un testo. Dal momento che il NVdB rappresenta una sintesi dell'italiano standard, l'analisi dei testi compara la lingua in esame con quella italiana dei parlanti di media istruzione. L'indice Rix-Pavan-it, a differenza del Rix e del Rix-Pavan, non è un indice multilingua, per cui è possibile analizzare testi di qualsiasi dimensione in italiano o, al più, come vedremo in seguito, in lingua volgare.

Nella versione scritta per computer in linguaggio Javascript,⁹ per quanto riguarda i nomi propri (inclusi i toponimi), si è scelto di considerare come parole difficili tali nomi

8 Si pensi ad esempio ai tempi verbali, che in inglese è molto più semplice realizzare al computer in modo automatico rispetto all'italiano.

9 Si ricorda qui che gli indici di leggibilità, negli Stati Uniti, vengono implementati al computer dagli inizi degli anni Sessanta del secolo scorso (Klare, 1974, p. 87).

e di non escluderli dal computo considerando, dopo alcuni esperimenti, che la loro presenza nei testi (se eccessiva) complica la leggibilità.

2.1 La tabella di valutazione

La tabella di valutazione originale del Rix, non potendosi adattare al nuovo indice, è stata alterata. La nuova tabella è ottenuta in base ai risultati dell'analisi di un largo numero di testi letterari, articoli di riviste e quotidiani. I testi, perlopiù classici della letteratura italiana, sono quelli proposti dalle scuole agli studenti di diversi livelli. Si è scelto questo metodo, piuttosto che utilizzare la classica procedura del *cloze* (Bormuth, 1968). La procedura *cloze*, quando usata per determinare i livelli degli studenti, è comunque soggetta a una serie di problemi che ne influenzano i risultati (Hittleman, 1978, p. 120).

La nuova tabella va dalla quinta elementare fino all'università. Se però si volessero seguire le indicazioni di De Mauro, la tabella si potrebbe semplificare facilmente facendola iniziare dalla terza media. Se si suppone che le parole del NVdB siano conosciute dagli studenti che abbiano compiuto studi sulla grammatica di base, si è scelto di abbassare il livello minimo alla prima media (studenti di 11-12 anni d'età) oppure, per casi particolari, alla quinta elementare (studenti di 10 anni d'età). Dunque la nuova tabella appare come segue:

Indice Rix-Pavan-it	Anno scolastico
6.2 e superiore	Università
5.8 e superiore	13 (quinto anno di scuola superiore)
5.0 e superiore	12 (quarto anno di scuola superiore)
4.2 e superiore	11 (terzo anno di scuola superiore)
3.0 e superiore	10 (secondo anno di scuola superiore)
2.2 e superiore	9 (primo anno di scuola superiore)
1.8 e superiore	8 (terza media)
1.2 e superiore	7 (seconda media)
2.2 e superiore	6 (prima media)
inferiore a 0.8	5 (quinto anno di scuola elementare)

Si noti che, rispetto al Rix e al Rix-Pavan, la tabella ora include anche il quinto anno di scuola superiore, che è in genere presente nel sistema scolastico italiano, ma che non figura nella tabella originale dell'indice Rix. La nuova tabella potrebbe essere infine anche semplificata, indicando solo i vari ordini di scuola:

Indice Rix-Pavan-it	Livello
6.2 e superiore	Università
2.2 e superiore	Scuole superiori
0.8 e superiore	Scuole medie
Inferiore a 0.8	Scuole elementari (concluse)

2.2 I troncamenti

Per realizzare l'indice al computer si è anche provveduto a considerare i *troncamenti*, o *apocopi* (*esser, far, ecc.*), controllando se essi appartengano o meno al NVdB. Il programma verifica, aggiungendo una vocale finale, se la parola così ottenuta si trova nella lista del NVdB e provvede, in questo caso, ad escluderla dal computo. L'uso dei troncamenti è una prassi comune della letteratura in volgare e italiana di tutte le epoche (Meinschaefer, 2006) e, per adattare il programma anche all'analisi di testi poetici, si è considerata la loro presenza nei testi. Inoltre il programma fornisce un *file* con la lista di parole che, nei testi in esame, non appartengono al NVdB: questo può essere utile se, in caso di revisione di un testo, si vogliono cercare sinonimi per le parole ritenute difficili o quelle che ne compromettono la leggibilità.

Nella poesia del tardo Ottocento i troncamenti sono molto frequenti. Un esempio di testo poetico assegnato agli studenti delle scuole italiane è *Pianto antico* di Giosuè Carducci (1887):¹⁰

*L'albero a cui tendevi
la pargoletta mano,
il verde melograno
da' bei vermigli fior,
nel muto orto solingo
rinverdi tutto or ora
e giugno lo ristora
di luce e di calor.
Tu, fior de la mia pianta
percossa e inaridita
tu de l'inutil vita
estremo unico fior,
sei nella terra fredda
sei nella terra negra
né il sol più ti rallegra
né ti risveglia amor.*

Nel testo si può osservare una certa frequenza dei troncamenti (*fior, amor, calor*). L'analisi con il programma fornisce il file contenenti le parole non presenti nel NVdB. L'elenco è il seguente:

*pargoletta
vermigli
solingo
rinverdi
riстора
percossa
inaridita*

¹⁰ Tratto da: Carducci, G. *Giambi ed Epodi e rime nuove*. Bologna: Zanichelli, ed. 1942.

Dall'elenco non risultano parole soggette a troncamento, in quanto già presenti nel NVdB e rilevate in precedenza dal programma. Il livello che si ottiene dalla formula dell'indice è il seguente, ma si tenga conto che il calcolo include anche la punteggiatura, quasi assente nel componimento:

Rix-Pavan-it: 4.41

Anno scolastico: 11 (terzo anno di scuola superiore)

Utilizzando la stessa formula e tarando la tabella di valutazione in modo appropriato sarebbe possibile realizzare lo stesso procedimento in altre lingue, se si dispone di una lista di parole come quella del NVdB.

3. L'analisi di testi letterari con l'indice Rix-Pavan-it

Con il nuovo indice Rix-Pavan-it sono stati analizzati alcuni testi letterari italiani appartenenti a epoche diverse, comparando i risultati con l'indice Rix-Pavan.¹¹ Come si vede dalla tabella qui di seguito, appaiono a volte differenze più o meno vistose tra i due indici:

Autore e opera	Rix-Pavan (anno scolastico)	Rix-Pavan-it (anno scolastico)
Vittorio Alfieri, <i>Saul</i>	4	7
Giuseppe G. Belli, <i>Sonetti</i>	7	12
Giovanni Boccaccio, <i>Decameron</i>	11	10
Dante Alighieri, <i>Commedia</i>	8	10
Edmondo De Amicis, <i>Cuore</i>	7	6
Ugo Foscolo, <i>Sonetti</i>	7	10
Carlo Goldoni, <i>La locandiera</i>	4	5
Alberto Moravia, <i>La ciociara</i>	10	8
Ruzante, <i>La moscheta</i>	4	10
Giacomo Leopardi, <i>Canti</i>	8	9
Alessandro Manzoni, <i>I promessi sposi</i>	8	7
Eugenio Montale, <i>Le occasioni</i>	8	9
Giovanni Pascoli, <i>Myrica</i>	5	7
Francesco Petrarca, <i>Canzoniere</i>	7	12
Carlo Porta, <i>Poesie</i>	7	Università
Giovanni Verga, <i>I Malavoglia</i>	9	9

3.1 Discussione

Dove si osservano le differenze maggiori tra i due indici è nell'analisi della letteratura in dialetto e dei testi poetici. Molte parole lunghe, di lunghezza superiore a 6 caratteri, non sono parole difficili e vengono usate in letteratura. Al contrario, molte parole brevi, che dovrebbe-

¹¹ Tutti i testi esaminati nell'articolo sono stati tratti, in forma digitale, dal sito <https://liberliber.it>

ro essere di facile comprensione, in realtà con il passare dei secoli sono cadute in disuso e al giorno d'oggi sono di difficile interpretazione. L'indice Rix-Pavan-it tiene conto delle parole attraverso il NVdB, dunque è di fatto più preciso del Rix-Pavan, che invece può contare solo sulla lunghezza delle parole stesse. Ecco quindi come si spiegano le discrepanze tra i due indici. Per quanto riguarda la letteratura dialettale appare chiaro che anche le parole brevi risultino di difficile comprensione. Alcuni autori, come Porta e Belli, utilizzavano un dialetto regionale che per molti studenti di letteratura risulta oggi assai complesso. Ecco un esempio di componimento di Carlo Porta (1815):¹²

*Col boeucc avert e cont in aria el venter
sont ch'ì a speccià quaj coss.
Sciòr mettimel denter,
gh'avaroo pussee gust con pù l'è gross.
Se poeù rivassev a impregnamm ben prest
prima che passa i fest
v'accorgiarì quanta memoria lassen
i voster benefizzi:
e s'ì che in pont de fa certi servizi
l'è mej fà regordà che regordassen.*

Se ne deduce che la lingua utilizzata è molto lontana dall'italiano standard. Tuttavia, analizzando il testo, l'indice Rix-Pavan lo interpreta leggibile per uno studente di terza media, mentre il Rix-Pavan-it lo giudica adatto a uno studente universitario. Senza un apparato critico sarebbe comunque arduo comprendere il testo anche per uno studente universitario, a meno che non si appartenga alla regione in cui si parla quel dialetto, in questo caso il milanese.¹³ Il grado di giudizio del Rix-Pavan-it è quello più vicino alla realtà. Anche la poesia in volgare presenta differenze tra i due indici, per il fatto che in volgare sono presenti, oltre a parole tronche, molti termini di lunghezza breve che oggi non vengono usati. Comunque le discrepanze non sono sempre evidenti. Se si analizza il sonetto di apertura del *Canzoniere* di Petrarca,¹⁴ si scopre che i risultati prodotti dai due indici sono quasi gli stessi. Il testo del componimento è il seguente:

*Voi ch'ascoltate in rime sparse il suono
di quei sospiri ond'io nudriva 'l core
in sul mio primo giovanile errore
quand'era in parte altr'uom da quel ch'ì sono,
del vario stile in ch'io piango et ragiono
fra le vane speranze e 'l van dolore,
ove sia chi per prova intenda amore,
spero trovar pietà, nonché perdono.*

12 Tratto da: Porta, C. *Poesie di Carlo Porta*. Milano: Mondadori, ed. 1975.

13 De Mauro, a proposito della lingua, parla di "policentrismo" di alcune città italiane, tra cui Milano (De Mauro, 2011).

14 Tratto da: Petrarca. *Canzoniere*, R. Bettarini (ed.). Torino: Einaudi, 2005.

*Ma ben veggio or sí come al popol tutto
favola fù gran tempo, onde sovente
di me medesmo meco mi vergogno;
et del mio vaneggiar vergogna è 'l frutto,
e 'l pentersi, e 'l conoscer chiaramente
che quanto piace al mondo è breve sogno.*

Indice Rix-Pavan: 6.90 (Anno scolastico: 12)

Indice Rix-Pavan-it: 4.91 (Anno scolastico: 11)

L'articolo *'l* non viene interpretato come parola difficile nel calcolo del Rix-Pavan-it perché il codice Javascript sostituisce gli apici con spazi vuoti. In questo modo si ha, nel file di uscita, la seguente lista di parole difficili:

*nudriva
core
giovenile
et
veggio
sí
sovente
medesmo
meco
et
vaneggiar
pentersi*

Occorre precisare che tutti gli indici di leggibilità interpretano il testo in base al livello della sua comprensione e non tengono conto del contesto e della sintassi. L'IA è in grado di analizzare anche contesto e sintassi, ma al momento gli errori prodotti sono in numero rilevante, rendendo questo mezzo ancora inaffidabile (Sun et al., 2024). I valori forniti dagli indici sono solo indicativi e quando i docenti assegnano un testo da leggere devono tener conto anche di altri fattori, in particolare gli argomenti trattati, che gli indici non possono rivelare. Si possono tuttavia usare altri strumenti per identificare l'argomento di un testo (Pavan, 2022), oltre ovviamente all'IA.

Gli indici di leggibilità sono un mezzo pratico e rapido per interpretare la difficoltà di lettura. Ancora oggi molti indici vengono usati nei più svariati settori della società: scuole, aziende, istituzioni, apparati militari ecc. Il numero così elevato di indici è la prova della loro utilità. Creare nuovi indici non è un compito troppo arduo, si tratta dell'applicazione pratica di formule. Tuttavia solo alcuni indici sopravvivono, in base alla loro precisione e alla loro maggiore diffusione.¹⁵

¹⁵ In Italia, oltre al già citato GULPEASE, c'è anche l'indice Flesch-Vacca (che risale al 1972).

4. Conclusioni

Si è accennato all'IA e dei suoi risultati promettenti. I vantaggi dell'IA sono innegabili ma, per ora, questo strumento si è rivelato ancora inaffidabile. Gli indici di leggibilità a volte sono fallaci, tuttavia la logica che seguono porta a risultati apprezzabili, rappresentando ancora oggi uno strumento valido per valutare un testo. L'indice Rix-Pavan-it, descritto in questo articolo, si rivela di ottimo aiuto quando si tratta di analizzare un testo senza aver bisogno di leggerlo. Il nuovo indice può essere usato nelle scuole, nelle aziende e anche negli uffici della pubblica amministrazione, quando si tratta di redigere documenti che dovrebbero essere compresi da un pubblico vasto. La sua precisione è dovuta all'uso del NVdB, a sua volta realizzato attraverso processi statistici che mostrano qual è la parte di termini più comuni della lingua italiana. Inoltre la formula dell'indice qui descritto può facilmente essere adattata a lingue diverse, purché si abbia come base di partenza la lista di parole più frequenti dei parlanti e che la tabella di valutazione sia tarata, ad esempio, attraverso una scelta ponderata di testi in quelle lingue.

Bibliografia

- » Anderson, J. (1983). Lix and Rix: Variations on a Little-known Readability Index. *Journal of Reading* 26(6), 490-496.
- » Björnsson, C. H. (1968). *Läsbarhet*. Bokförlaget Liber.
- » Bormuth, J. R. (1968). The Cloze Readability Procedure. *Elementary English*, 45(4), 429-436.
- » Carducci, G. (1942). *Giambi ed Epodi e rime nuove*. Zanichelli.
- » Chall, J.S., & Dale, E. (1995). *Readability Revisited - The New Dale-Chall Readability Formula*. Brookline Books.
- » Chiari, I. (2018). Il vocabolario di base dell'italiano e la società civile. In S. Gensini, et al., *Tullio De Mauro. Un intellettuale italiano* (pp. 165-172). Sapienza Università Editrice.
- » Chiari, I., & De Mauro, T. (2014). The New Basic Vocabulary of Italian as a linguistic resource (pp. 113-116). In R. Basili & A. Lenci & B. Magnini (a cura di), *First Italian Conference on Computational Linguistics CLiC-it 2014*. Pisa University Press.
- » Dale, E., & Chall, J. S. (1948). A Formula for Predicting Readability: Instructions. *Educational Research Bulletin*, 27(2), 37-54.
- » De Mauro, T. (1980). Guida all'uso delle parole. In *Libri di base 3* (pp. 149-183). Editori Riuniti.
- » De Mauro, T. (2011). Per la storia linguistica dell'Italia repubblicana. *Italica*, 88(1), 40-58.
- » DuBay, W. H. (2004). *The Principles of Readability*. Impact Information.
- » Francalanci, L. (2019). *Proposta di un metodo di valutazione automatica della leggibilità di pagine web in lingua italiana*. Tesi di dottorato, Università degli Studi di Firenze.
- » Hittleman, D. R. (1978). Readability, Readability Formulas, and Cloze: Selecting Instructional Materials. *Journal of Reading*, 22(2), 117-122.
- » Il Nuovo vocabolario di base della lingua italiana (2016). <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana>
- » Klare, G. R. (1974). Assessing Readability. *Reading Research Quarterly*, 10(1), 62-102. <https://doi.org/10.2307/747086>

- » Lucisano, P., & Piemontese, M. E. (1988). GULPEASE: una formula per la predizione della leggibilità di testi in lingua italiana. *Scuola e città*, 110-124.
- » Marello, C. (2018). Tullio De Mauro e la lessicografia. *Sull'attualità di Tullio De Mauro* (pp. 67-86). il Mulino.
- » Meinschaefer, J. (2006). Troncamento and the phonological phrase in Florentine Italian. *Rivista di Linguistica*, 18(2), 295-324.
- » Monteleone, M. (2003). *Lessicografia e Dizionari Elettronici. Dagli Usi Linguistici alle Basi di Dati Lessicali*. Fiorentino & New Technology.
- » Pavan, L. (2022). Detecting Main Topics using Dictionary-based Topic Analysis. *International Journal of Linguistics, Literature and Translation*, 5(12), 48-52. <https://doi.org/10.32996/ijlilt.2022.5.12.6>
- » Pavan, L. (2024). Una variante dell'indice di leggibilità Rix che include la punteggiatura. *L'italiano dei giovani per i giovani* (pp. 40-51). Vytauto Didžiojo Universitetas.
- » <https://ebooks.vdu.lt/einfo/3197/l'italiano-dei-giovani-per-i-giovani-glottodidattica-dell'italiano-ls/>
- » Petrarca, F. (2005). *Canzoniere* (ed. R. Bettarini). Einaudi.
- » Porta, C. (1975). *Poesie di Carlo Porta*. Mondadori.
- » Spache, G. D. (1953). A New Readability Formula for Primary-Grade Reading Materials. *Elementary School Journal*, 53, 410-413. <https://doi.org/10.1086/458513>
- » Sun, Y., & Sheng, D., & Zhou, Z. (2024). AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications*, 11, Article number: 1278, 1-14. <https://doi.org/10.1057/s41599-024-03811-x>
- » Tekfi, C. (1987). Readability Formulas: An Overview. *Journal of Documentation*, 43(3), 261-273. <https://doi.org/10.1108/eb026811>

Luca Pavan

Vilnius Tech

Užsienio kalbų skyrius

Kūrybinių industrijų fakultetas

Saulėtekio al. 11

10223 VILNIUS

Lituania