

Indici di leggibilità: Gulpease e Flesch

Un indice di leggibilità è una formula matematica che attraverso un calcolo statistico è in grado di predire la reale difficoltà di un testo in base a una scala predefinita di valori.

Per definire la formula di un indice di leggibilità si può tener conto di diverse *variabili linguistiche*, cioè della misura di alcuni parametri del testo. Le variabili linguistiche più semplici sono, per esempio, *lunghezza media delle parole* e *lunghezza media delle frasi*.

Vi sono variabili linguistiche che sono indipendenti dal contenuto del testo, come appunto le due citate, oppure variabili linguistiche legate al lessico, alla struttura del periodo, ecc.

Sono state definite molte formule per la predizione della leggibilità, ma quelle che hanno avuto maggiore successo sono quelle che considerano variabili linguistiche di facile calcolo, come per esempio la lunghezza delle parole e delle frasi.

Per lo stesso motivo, i programmi che calcolano automaticamente il valore delle formule si limitano, nella maggior parte delle applicazioni, a formule semplici.

Indice di Flesch

La formula di leggibilità che ha avuto maggior successo e diffusione nel mondo è quella di Rudolf Flesch, nota come *Formula di Flesch*. Essa considera solo due variabili linguistiche: *lunghezza media delle parole* espressa in sillabe per parola, e *lunghezza media delle frasi* espressa in parole per frase.

La *formula di Flesch*, che deve la sua diffusione proprio alla sua semplicità, ha però due inconvenienti: il primo è prodotto dal fatto che la formula è stata progettata per l'inglese ed è, quindi, tarata sulla struttura morfologica e sillabica di questa lingua; il secondo è rappresentato dal problema del conteggio delle sillabe. Infatti, questo tipo di calcolo si mostra particolarmente complesso nell'ambito della lingua italiana, poiché esso non è completamente formalizzabile mediante regole di portata generale, se non ricorrendo a stime di tipo statistico, il cui limite, purtroppo è quello di non poter descrivere e riprodurre esattamente la sillabazione delle singole parole di un testo. In particolare, nel campo della sillabazione le difficoltà maggiori sono prodotte dall'accentazione e dalla presenza di dittonghi.

Il fatto che la formula è nata per l'inglese è stato affrontato da Roberto Vacca, il quale, nel 1972, ha adattato i parametri della formula alla lingua italiana (vedi [Franchina-Vacca 1986]). Della formula esiste anche un secondo adattamento, realizzato da Vacca nel 1986, dove: **Facilità di lettura = 217 - 1,3 W - 0,6 S**. Il nuovo adattamento nasce da un'ipotesi di Vacca, autore bilingue di uno stesso testo in lingua italiana e in lingua inglese. Secondo questa ipotesi i due testi avrebbero dovuto avere lo stesso indice di leggibilità per il fatto che entrambi trattavano lo stesso argomento ed erano stati scritti dallo stesso autore. Ma in base agli esperimenti condotti dal Gruppo Universitario Linguistico-Pedagogico risulta che i dati forniti dall'applicazione della formula del 1972 sono più attendibili di quelli derivati dall'applicazione della formula del 1986. Il secondo problema, quello delle sillabe, è ancora aperto.

La formula di Flesch per la lingua italiana
adatta da Franchina – Vacca (1972) è:
Facilità di lettura = 206-0.65 S – W
Dove S=sillabe di 100 parole
W= media di parole per frase

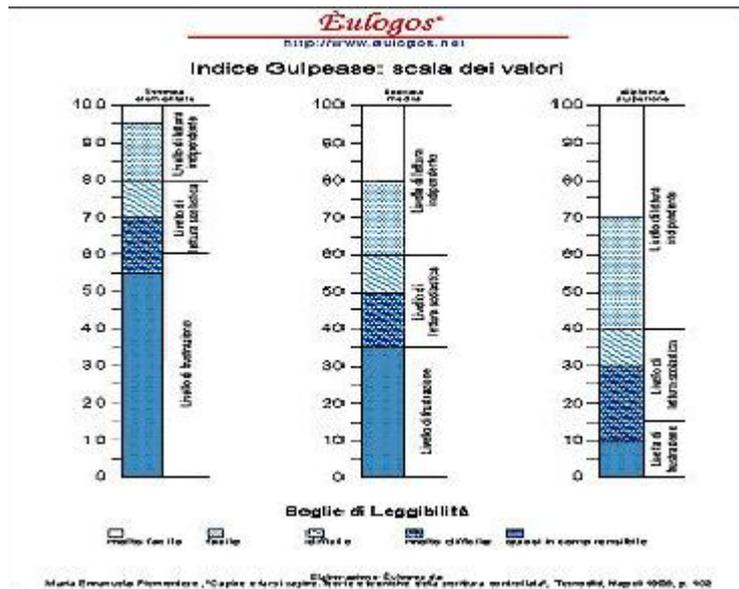
L'indice GULPEASE

Nel 1982 il GULP - *Gruppo universitario linguistico pedagogico*, presso l'Istituto di Filosofia dell'Università degli studi di Roma «La Sapienza» - ha definito una nuova formula, la formula *GULPEASE*, partendo direttamente dalla lingua italiana (vedi [Lucisano-Piemontese 1988] e [Lucisano 1992]).

La formula è stata determinata verificando con una serie di test la reale comprensibilità di un corpus di testi. La verifica è stata fatta su diversi tipi di lettore, e accanto alla determinazione della formula è stata definita una scala d'interpretazione dei valori restituiti dalla formula stessa. La scala mette in relazione i valori restituiti dalla formula con il grado di scolarizzazione del lettore. Per esempio, un testo con indice

GULPEASE 60 è: molto difficile per chi ha la licenza elementare, difficile per chi ha la licenza media, facile per chi ha un diploma superiore. La formula *GULPEASE*, oltre ad essere la prima formula di leggibilità tarata direttamente sulla lingua italiana, ha anche il vantaggio di calcolare la lunghezza delle parole in lettere, e non più in sillabe.

La formula *GULPEASE* (Lucidano – Piemontese 1988)
 Facilità di lettura = $89 - LP/10 + FR * 3$
 LP= lettere per 100/totale parole
 FR= frasi per 100/totale parole



Proprio questa caratteristica ha consentito di realizzare una versione informatizzata della formula con un buon livello di affidabilità. Si provi a tale riguardo il programma *Èulogos* SLI, all'indirizzo www.eulogos.it.

Per un approfondimento sugli indici di leggibilità sul sito sopra indicato sono reperibili un'ampia documentazione e una vasta bibliografia