

Received: 2015-04-20
Revised: 2015-07-07
Accepted: 2015-07-07

Galen Medical Journal (GMJ) Title Readability Index

Afshin Borhani Haghighi¹, Morteza Khodabakhshi²✉

¹ Clinical Neurology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran.

² Islamic World Science Citation Center (ISC), Shiraz, Iran.

Abstract

Background: Title choice will undoubtedly pave the way for the informed readership to either devour a text or just spurn it. Therefore, authors are duly required to have a deep understanding of titling trends and potential inferences to be drawn by their addressees. As a case in point, choosing proper titles for medical articles is a subject of much debate among researchers to such an extent that it is postponed to be written till the finalization of a given manuscript. Titles form expectations, preconceptions, presuppositions and most importantly judgment on reader's side. Quality highly-cited research papers submitted to the internationally renowned Galen Medical Journal (GMJ) as a true index of Iranian medical journalism covering varied areas of research in both medical and interdisciplinary fields concerning health-related topics, provide a performance benchmark to firstly evaluate titling trends among Iranian authors, secondly to improve writing strategies and finally to meet international standards. **Materials and Methods:** To conduct this meta-analysis, researchers collected all articles published by GMJ. 100 titles were randomly chosen to be linguistically analyzed by Advanced Text Analyzer software. All data were fed into MC-Excel to recognize any distributional regularity in titling, subconsciously followed by GMJ authors. **Results:** The meta-analysis of randomly chosen article titles revealed a well-established trend in high Lexical Density standing above 76.67 for all titles indicating author's frozen style and academic register in their manuscripts. Moreover, Title Length hits 14.34 on average signifying author's inclination for longer titles. It might negatively impact the whole discourse due to heavy cognitive load. Gunning Fog Index ranging averagely 16.94 estimates at least 17 years of formal education needed to understand a text on first reading with no difficulty. Additionally, introducing poly-morphemic lexical items into author's title in nearly all cases, owing to their mental load, augments text difficulty. **Conclusion:** Authors of medical articles can increase their article's readability through their acquaintance with text mechanics such as lexical density, Gunning Fog Index and Readability criteria. Authors are strongly recommended to shorten title length, introduce fewer poly-morphemic words and utilize highly frequent mono-morphemic lexical items in order to increase article readability. [GMJ. 2015;4(3):100-3]

Keywords: ARI (Automated Readability Index); Gunning Fog Index; Lexical Density; Title

GMJ

©2013 Galen Medical Journal
Fax: +98 731 2227091
PO Box 7461686688
Email: info@gmj.ir



✉ **Correspondence to:**
Morteza Khodabakhshi, Islamic World Science Citation
Center (ISC), Shiraz, Iran.
Telephone Number: (+98) 71-36468420
Email Address: mortezakhodabakhshi@yahoo.com

Introduction

The concept of readability is far from new; over the past fifty years several of its facets have been examined and tested. Research has shown that readability can vary in accordance with certain specific typographic variables. However, the overwhelming majority of this research has focused upon the readability of text in medicine. This stems from the fact that most medical texts produced by scholars have not been duly analyzed in terms of linguistic features. Factors such as the effects of typeface (e.g. serif versus sans serif typefaces), letter spacing, line spacing (or leading), justification contrast, resolution, inverted text, mechanically-tinted backgrounds, size, type style, letter spacing, and page layout do impact reading ease or readability [1]. These typographic variables have been tested in order to determine various effects upon the reader. Chief among these variables are reading rate and reading comprehension. Moreover, morpho-syntactic properties of a given text determine its readability; factors like utterance length, number of poly-morphemic words, number of mono-morphemic words, lexical density, lexical variety, etc. determine readability index.

Readability metrics are formulae for evaluating the readability of texts, usually by counting syllables, words, and sentences. Readability tests are often used as an alternative to conducting an actual statistical survey of human readers of the subject text (a readability survey). Word processing applications often have readability tests built-in, which can be deployed on documents in-editing [2].

The application of a useful readability test protocol will give a rough indication of a work's readability, with accuracy increasing when finding the average readability of a large number of works. The tests generate a score based on characteristics such as statistical average word length (which is used as an unreliable proxy for semantic difficulty) and sentence length (as an unreliable proxy for syntactic complexity) of the work.

Some readability formulas refer to a list of words graded for difficulty. These formulas attempt to overcome the fact that some words,

like "television", are well known to younger children, but have many syllables. In practice, however, the utility of simple word and sentence length measures make them more popular for readability formulas. Scores are compared with scales based on judged linguistic difficulty or reading grade level. Many readability formulas measure word length in syllables rather than letters, but only SMOG has a computerized readability program incorporating an accurate syllable counter. The Automated Readability Index (ARI) is a readability test designed to gauge the understandability of a text. It produces an approximate representation of the US grade level needed to comprehend the text.

The formula for calculating the Automated Readability Index is given below:

$$4.71 (\text{characters/words}) + 0.5 (\text{words/sentences}) - 21.43$$

Where characters is the number of letters, numbers, and punctuation marks, words is the number of spaces, and sentences is the number of sentences. Although opinion varies on its accuracy as compared to the syllables/word and complex words indices, characters/word is often faster to calculate, as the number of characters is more readily and accurately counted by computer programs than syllables. In fact, this index was designed for real-time monitoring of readability on electric typewriters.

In linguistics, the Gunning Fog Index measures the readability of English writing. The index estimates the years of formal education needed to understand the text on a first reading [3]. A fog index of 12 requires the reading level of a U.S. high school senior (around 18 years old). The test was developed by Robert Gunning, an American businessman, in 1952. The fog index is commonly used to confirm that text can be read easily by the intended audience. Texts for a wide audience generally need a fog index less than 12. Texts requiring near-universal understanding generally need an index less than 8. While fog index is a good sign of hard-to-read text, it has limits. Not all complex words are difficult. A short word can be difficult if it is not used very often by most people. The frequency with which words are in normal use affects the readability of text[4].

The complete formula is:

$$0.4 [(words/sentences) + 100 ((complex words)/words)]$$

Materials and Methods

Galen Medical Journal (GMJ) represents an Iranian research journal rich in original articles by Fasa University of Medical Sciences affiliated with Shiraz University of Medical Sciences. All technical texts produced by scholars in this journal form a unique medical discourse. To pinpoint discourse attributes of these texts, researchers took article titles as a true yardstick for comprehension starting point. It is assumed that inferences made from titles will facilitate later comprehension [7]. 100 article titles were randomly chosen within a four-year period of publication by this scholarly journal. These titles were then fed into Advanced Text Analyzer one by one. This software could be reached at <http://www.usingenglish.com>

Each title was considered a full affirmative sentence to provide this software with real input. It automatically calculates the following linguistic variables:

Title Length

Most readability formulas use the number of words in a sentence to measure its difficulty. Yet, in some cases a short sentence can be harder to read than a long one. Comprehension can sometimes be facilitated by longer sentences, especially those that contain coordinate structures. Contemporary style guides generally recommend varying the length of sentences to avoid monotony and achieve appropriate emphasis.

Hard Words

Word difficulty is usually measured by vocabulary lists or word length. In 1923, Bertha A. Lively and Sidney L. Pressey published the first reading ease formula. They had been concerned that science textbooks in junior high school had so many technical words. They felt that teachers spent all class time explaining their meaning. They argued that their formula would help to measure and reduce the "vocabulary burden" of textbooks. Their formula

used the Thorndike word list as a basis.

Lexical Density

In computational linguistics, lexical density constitutes the estimated measure of content per functional (grammatical) and lexical units (lexemes) in total. It is used in discourse analysis as a descriptive parameter which varies with register and genre.

Gunning Fog Index

It is a weighted average of the number of words per sentence, and the number of long words per word. An interpretation is that the text can be understood by someone who left full-time education at a later age than the index.

ARI

It is a readability test designed to assess the understandability of a text. Like other popular readability formulas, the ARI formula outputs a number which approximates the grade level needed to comprehend the text [6]. For example, if the ARI outputs the number 10, this equates to a high school student, ages 15-16 years old; a number 3 means students in 3rd grade (ages 8-9 yrs. old) should be able to comprehend the text.

All data were inserted into MS-Excel to find distributional regularity and potential trends among Iranian authors. The mean, maximum and minimum amounts for each variable were calculated and contrasted against general trend. The uppermost figures at each end of minimum-maximum continuum indicate the best and worst titles in terms of readability.

Results

A well-established distributional regularity for a given article in GMJ enjoys the following linguistic properties:

1. The number of words per each title stands at 14.34 on average.
2. The average number of least frequently-used words (hard words) hits 4.04 per title.
3. All titles contain 6.64 poly-morphemic lexical items.

4. The average lexical density estimates 95.62
5. The Gunning Fog Index calculates 16.94
6. ARI stands at 19.77

Discussion

With a careful analysis of lexical density in all title extracts, the results reveal that almost all of the texts featured a high lexical density index featuring topmost lexical richness appropriate for native-like graduate or postgraduate levels. The Gunning Fog Index of approximate 17 is indicative of seventeen years of formal education to comprehend these medical texts without any difficulty. Unlike other internationally indexed titles, Galen lengthy titles consisting of 14.34 lexical items could impose extra cognitive load leading to constricted comprehension. Introducing 4.04 least frequent (hard) words in each title could be justified by medical frozen register. Finally, ARI score of 19.77 requires a minimum of seven academic years of education to have a deep understating of a given text. The discrepancy of Gunning Fog Index (requiring a minimum of 5 academic years) and the ARI score (requiring a minimum of 7 academic years) is largely due to the fact that students of medical

sciences in most countries might finish their theoretical courses at the end of their fifth year of general practice.

Conclusion

Would-be authors and students of medical sciences are linguistically obliged to get acquainted with morpho-syntactic attributes of technical paper writing. Titling a paper does impact its readership to the extent that judgment on the quality of an article is primarily made based upon its title [8]. Shortening title length, introducing fewer poly-morphemic words, utilizing highly-frequent mono-morphemic lexical items and familiarizing with frozen style will undoubtedly increase text readability and consequently article visibility.

Acknowledgments

Authors would like to express their heartfelt gratitude to Dr. Aliasghar Karimi and all staff members at GMJ for their cooperation and insightful comments on this draft.

Conflict of Interest

Authors had no conflict of interest when conducting this research.

References

1. Clear Writing: Ten principles of clear statement. The University of Missouri. 2006. (Accessed December 15, 2014 at <http://muextension.missouri.edu/xplor/comm/cm0201.html>)
2. Lauchman, R. Plain language: A handbook for writers in the U.S. federal government. 2001. (Retrieved April 2, 2006, from <http://www.lauchmangroup.com/PDFfiles/PLHandbook.PDF>)
3. Miles T. The fog index: a practical readability scale. West Virginia University. 1990. (Retrieved April 2, 2006, from <http://www.as.wvu.edu/~tmiles/fog.html>)
4. Roberts J, Fletcher R, Fletcher S. Effects of peer review and editing on the readability of articles published in *annals of internal medicine*. 1994. (Retrieved April 3, 2006, from http://www.amaassn.org/public/peer/7_13_94/pv3083x.html)
5. Suzuki D. The right stuff. In Brundage D, Lahey M (Eds.), *Acting on Words*. Toronto: Pearson Education Canada Inc 2008. (2nd ed., Original work published 1989), pp. 464-466.
6. Weeks W, Wallace A. Readability of British and American medical prose at the start of the 21st century. *BMJ* 2002, 325-378 (Retrieved April 3, 2006, from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=139036>)
7. Wimmer RD, Dominick JR. *Mass media research: an introduction* (8th ed.). Belmont, Calif, London: Thomson/Wadsworth. 2006.
8. Yu G. *Lexical Diversity in MELAB Writing and Speaking Task Performances*. Spaan Fell 2007.