We begin a new series on using Natural Language Processing tools in the writing and editing process.

Software-Assisted Ghost Writing
Measuring Readability

Six occupants of the White House have come and gone since I started using software to improve the readability of what I write. The year was 1980. I was writing a regular column for the newsletter that my employer sent to their customers.

While browsing in a bookstore, I found two paperback books by Rudolf Flesch*: The Art of Readable Writing* (1949) and *The Art of Plain Talk* (1951). Books that stay in print for over 30 years are rare, so I took a closer look.

Flesch explained "how to speak and write so that people understand what you mean." The first book presented algorithms to compute the "reading ease" and "human interest" scores for a document. The second book presented an algorithm to compute a more sophisticated "yardstick formula" for plain language. Both books contained lots of examples of readable and unreadable writing.

In addition to being a writer, I was also a hotshot programmer (or so I thought). It would be another year and a half before the IBM PC was introduced. The language that hobbyist programmers used on their Apple II's and TRS-80's was BASIC. I was an avid reader of the hobbyist computing magazines. One of those magazines included a program, written in BASIC, which computed the Flesch readability index for a text.

The programming language I used at work then was APL. In several respects, APL was years ahead of its time. APL used a unique character set, and special terminals were needed to display APL programs. In the 1990's, bit-mapped graphics displays made the characters generally available. APL was a data parallel language. Parallel hardware that could directly execute its constructs was only available on supercomputers until the late 1980's. APL was an interpreted language. It traded some loss of efficient execution for increased ease of programming. Two decades later, the inventors of Java and Python made the same trade-off.

After seeing many pages of BASIC code in the magazine article, I thought I could do much better. I decided to write a data parallel APL program that computed the Flesch index. The analysis module was rather elegant. It fit on one page and had no explicit loops. Because it was fully data parallel, it executed quickly.

After I finished my program, I enhanced it in two important ways. The Flesch index values tell you about the document as a whole. If you are going to improve your text, you need to identify the potential problems within the document. I added a feature that displayed the sentences which had more syllables than a threshold value. I also added a feature that displayed the words which had more syllables than a threshold value. With these features, I could focus on rewriting wordy sentences and replacing bulky words with simpler synonyms.

Once I made my program available, my employer decided to offer it commercially. At the time, some state governments were forcing regulated industries, like utilities and insurance companies, to produce "readable" consumer documents. We provided a service that they could use to prove that their documents achieved a certain level of readability.

Today Microsoft Word will compute the Flesch Reading Ease score and the Flesch-Kincaid Grade level score of your documents. To turn this on, click File, then Options, then Proofing. On that page, select the box labelled "Show readability statistics."

Microsoft Words computes these indices as follows:
Flesch Reading Ease score: 206.835 – (1.015 x ASL) – (84.6 x ASW)
Flesch-Kincaid Grade Level score: (.39 x ASL) + (11.8 x ASW) – 15.59
where:
ASL = average sentence length (the number of words divided by the number of sentences)
ASW = average number of syllables per word (the number of syllables divided by the number of words)

Unfortunately, the Microsoft Word implementation is rather useless for editing work. It needs the enhancements that I made to my program nearly 40 years ago.

Fortunately, there are alternatives which are much more helpful. I am a subscriber to the service provided at the website "readable.io." This service computes a large number of useful statistics related to readability, and presents the results in a more useful format.

The basic analysis uses color and underlining to highlight sentences with more than 30 syllables, words with more than 4 syllables, verbs in the passive voice, and all adverbs. It displays a table with ten metrics. This table shows the results when the website analyzes this article.

Readability Grade Levels

Formula	Grade
Flesch-Kincaid Grade Level	8.5
Gunning Fog Index	11.3
Coleman-Liau Index	10.3
SMOG Index	11.6
Automated Readability Index	8.0

```
Average Grade Level                  9.9

Readability Scores
Formula                            Score
Flesch Reading Ease                 58.2
CEFR Level                            C2
IELTS Level                           8+
Spache Score                         5.4
New Dale-Chall Score                 5.7
```

So, how does a ghostwriter use this tool? I type a text, like this article, into a word processing program like MS Word, then copy and paste it into the readability evaluator. I remove unnecessary words and phrases from long sentences that it highlights. Then I look for shorter synonyms for the long words that it highlights. Not much has changed in nearly forty years!

I applied this process to this article. In about 15 minutes, I was able to improve it by reducing the grade level a full year.

There is nothing particular to ghost writing about this process. Don’t worry, we’re just getting started. In the next issue, we will learn about computational methods a ghostwriter can use to identify the credited author’s natural style (voice).

References

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Brookline Books.
DuBay, W. H. (2014). *The Principles of Readability*. Costa Mesa, CA: Impact Information.
Flesch, R. F. (1946). *The art of plain talk.* Macmillan.
Flesch, R. F. (1949). *The art of readable writing.* Macmillan.