Thea Readability Code Notes

NB: Potentially helpful documentation: <https://www.rdocumentation.org/packages/koRpus/versions/0.11-5/topics/readability>

Many notes from from here:<https://readable.com/features/readability-formulas/>

**WE ALREADY USE**

"Flesch":

Flesch's Reading Ease Score (Flesch 1948).

*206.835 - (1.015 \* ASL) - (84.6 \* (Nsy / Nw))*

"Flesch.PSK":

The Powers-Sumner-Kearl's Variation of Flesch Reading Ease Score (Powers, Sumner and Kearl, 1958).

*(0.0078 \* ASL) + (4.55 \* Nsy / Nw) - 2.2029*

"Flesch.Kincaid":

Flesch-Kincaid Readability Score (Flesch and Kincaid 1975).

*0.39 \* ASL + 11.8 \* (NSy /Nw) - 15.59*

Notes:

The Flesch-Kincaid grade level was created by Rudolf Flesch as a way to interpret a US grade level from his Reading Ease formula. He originally created it for the US Navy to analyze their technical material. Since then, it's become a widely respected formula suitable for use with many types of text.

"FOG":

Gunning's Fog Index (Gunning 1952).

*0.4 \* (ASL + 100 \* (Nwmin3sy / Nw)*

where *n\_{wsy>=3}* = Nwmin3sy = the number of words with 3-syllables or more. The scaling by 100 arises because the original FOG index is based on just a sample of 100 words)

Notes:

The Gunning Fog Index was created by Robert Gunning, an American businessman. His consultancy, Robert Gunning Associates, aimed to help other businesses with their writing. This formula helps a writer win an audience by clearing unnecessary fog in their text

"FOG.PSK":

The Powers-Sumner-Kearl Variation of Gunning's Fog Index (Powers, Sumner and Kearl, 1958).

*3.0680 \* (0.0877 \* ASL) +(0.0984 \* 100 \* (Nwmin3sy / Nw)*

where *n\_{wsy>=3}* = Nwmin3sy = the number of words with 3-syllables or more. The scaling by 100 arises because the original FOG index is based on just a sample of 100 words)

"FOG.NRI":

The Navy's Adaptation of Gunning's Fog Index (Kincaid, Fishburne, Rogers and Chissom 1975).

*(((Nwless3sy + 3 \* Nw3sy) / (100 \* Nst / Nw))-3) / 2*

where *n\_{wsy<3}* = Nwless3sy = the number of words with *less than* 3 syllables, and *n\_{wsy=3}* = Nw3sy = the number of 3-syllable words. The scaling by 100 arises because the original FOG index is based on just a sample of 100 words)

**MIGHT BE RELEVANT**

"ARI":

Automated Readability Index (Senter and Smith 1967)

*0.5 ASL + 4.71 AWL - 21.34*

Notes:

Very useful for technical texts

<https://readable.com/blog/how-did-the-automated-readability-index-become-an-essential-tool-for-technical-writers/>

…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. 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.

"ARI.Simple":

A simplified version of Senter and Smith's (1967) Automated Readability Index.

*ASL + 9 AWL*

"Coleman":

Coleman's (1971) Readability Formula 1.

*1.29 \* (100 \* Nwsy1 / Nw) - 38.45*

where *n\_{wsy=1}* = Nwsy1 = the number of one-syllable words. The scaling by 100 in this and the other Coleman-derived measures arises because the Coleman measures are calculated on a per 100 words basis.

"Coleman.C2":

Coleman's (1971) Readability Formula 2.

*1.16 \* (100 \* Nwsy1 / Nw) + 1.48 \* (100 \* Nst / Nw) - 37.95*

"Coleman.Liau.ECP":

Coleman-Liau Estimated Cloze Percent (ECP) (Coleman and Liau 1975).

*141.8401 - (0.214590 \* 100 \* AWL) + (1.079812 \* Nst \* 100 / Nw)*

Notes:

The index doesn't involve syllable counting but instead counted word length in letters. It's widely used and has been used in the medical and translation sectors. It's also useful for the legal sector, where it's been used to explore the readability of judicial opinions.

"Coleman.Liau.Grade":

Coleman-Liau Grade Level (Coleman and Liau 1975).

*-27.4004 \* Coleman.Liau.ECP / 100 + 23.06395*

Notes:

Similar to the [Automated Readability Index](http://www.readabilityformulas.com/automated-readability-index.php), but unlike most of the other grade-level predictors, the **Coleman–Liau** relies on characters instead of syllables per word. Instead of using syllable/word and sentence length indices, Meri Coleman and T. L. Liau believed that computerized assessments understand characters more easily and accurately than counting syllables and sentence length

"Coleman.Liau.Short":

Coleman-Liau Index (Coleman and Liau 1975).

*5.88 \* AWL + (0.296 \* Nst / Nw) - 15.8*

"DRP":

Degrees of Reading Power.

*(1 - Bormuth.MC) \* 100*

where Bormuth.MC refers to Bormuth's (1969) Mean Cloze Formula (documented above)

Notes:

simple measure of literacy skills, meant to measure a student’s comprehension of text passages. While many other assessments test reading “skills,” like phonetics, vocabulary, and mastery of sentence structure, DRP measures a student’s overall ability to comprehend and critically absorb passages of text.

DRP scores are precise, reliable indicators of student reading comprehension. On a scale from 1-100, DRP scores are aligned to grade level entry and exit levels, so educators can use it as a screener, progress monitor, and end-of-grade growth measure.

Might be interesting

"Dale.Chall":

The New Dale-Chall Readability formula (Chall and Dale 1995).

*64 - (0.95 \* 100 \* Nwd / Nw) - (0.69 \* ASL)*

Dale and Chall created **The Dale-Chall Formula** for adults and children above 4th grade as a way to improve upon the Flesch Reading Ease Formula.  
  
The Dale-Chall Formula was unique because, unlike other formulas that use word-length to assess word difficulty, the Dale-Chall Formula uses a count of ‘hard’ words. The Dale-Chall Formula calculates the US grade level of a text sample based on sentence length and the number of ‘hard’ words. These ‘hard’ words are words that do not appear on a specially designed list of common words familiar to most 4th-grade students.

Might be interesting

"Dale.Chall.Old":

The original Dale-Chall Readability formula (Dale and Chall (1948).

*0.1579 \* 100 \* Nwd / Nw + 0.0496 \* ASL [+ 3.6365]*

The additional constant 3.6365 is only added if (Nwd / Nw) > 0.05.

"Dale.Chall.PSK":

The Powers-Sumner-Kearl Variation of the Dale and Chall Readability formula (Powers, Sumner and Kearl, 1958).

*(0.1155 \* 100 \* Nwd / Nw) + (0.0596 \* ASL) + 3.2672*

"FORCAST":

FORCAST (Simplified Version of FORCAST.RGL) (Caylor and Sticht 1973).

*20 - (Nwsy1 \* 150) / (Nw \* 10)*

where *n\_{wsy=1}* = Nwsy1 = the number of one-syllable words. The scaling by 150 arises because the original FORCAST index is based on just a sample of 150 words.

Notes:

Might be relevant for ‘incomplete’ sentences:

The FORCAST formula was the outcome of a study by HumRRO, which originally developed tools for the US army. FORCAST was developed to improve training documentation for new personnel. It doesn't rely on complete sentences for its analysis, so it's particularly useful for surveys, questionnaires, multiple question tests or any document containing lists or bullet points.

Intended for training documents: <https://readable.com/blog/how-did-the-forcast-readability-formula-help-create-readable-material-for-adults/>

"FORCAST.RGL":

FORCAST.RGL (Caylor and Sticht 1973).

*20.43 - 0.11 \* (Nwsy1 \* 150) / (Nw \* 10)*

where *n\_{wsy=1}* = Nwsy1 = the number of one-syllable words. The scaling by 150 arises because the original FORCAST index is based on just a sample of 150 words.

**MEASURES TO EXPLORE**

**CAN’T FIND EXPLANATION**

"Bormuth.MC":

Bormuth's (1969) Mean Cloze Formula.

*0.886593 - 0.03640 \* AWL + 0.161911 \* AFW - 0.21401 \* ASL - 0.000577 \* ASL^2 - 0.000005 \* ASL^3*

"Bormuth.GP":

Bormuth's (1969) Grade Placement score.

*4.275 + 12.881M - 34.934M^2 + 20.388 M^3 + 26.194 CCS - 2.046 CCS^2 - 11.767 CCS^3 - 42.285(M \* CCS) + 97.620(M \* CCS)^2 - 59.538(M \* CCS)^2*

where *M* is the Bormuth Mean Cloze Formula as in "Bormuth" above, and *CCS* is the Cloze Criterion Score (Bormuth, 1968).

"Danielson.Bryan":

Danielson-Bryan's (1963) Readability Measure 1.

*(1.0364 \* Nc / Nblank) + (0.0194 \* Nc / Nst) - 0.6059*

where *n\_{blank}* = Nblank = the number of blanks.

"Danielson.Bryan2":

Danielson-Bryan's (1963) Readability Measure 2.

*131.059 - (10.364 \* Nc / Nblank) + (0.0194 \* Nc / Nst)*

where *n\_{blank}* = Nblank = the number of blanks.

"Dickes.Steiwer":

Dickes-Steiwer Index (Dicks and Steiwer 1977).

*235.95993 - (73.021 \* AWL) - (12.56438 \* ASL) - (50.03293 \* TTR)*

where TTR is the Type-Token Ratio (see textstat\_lexdiv)

"Farr.Jenkins.Paterson":

Farr-Jenkins-Paterson's Simplification of Flesch's Reading Ease Score (Farr, Jenkins and Paterson 1951).

*-31.517 - (1.015 \* ASL) + (1.599 \* Nwsy1 / Nw)*

where *n\_{wsy=1}* = Nwsy1 = the number of one-syllable words.

**NOT STRIKINGLY RELEVANT**

"Fucks":

Fucks' (1955) Stilcharakteristik (Style Characteristic).

*AWL \* ASL*

"SMOG":

Simple Measure of Gobbledygook (SMOG) (McLaughlin 1969).

*1.043 \* sqrt(Nwmin3sy \* 30 / Nst) + 3.1291*

where *n\_{wsy>=3}* = Nwmin3sy = the number of words with 3 syllables or more. This measure is regression equation D in McLaughlin's original paper.

Notes:

SMOG is very useful for medical writing and is considered the 'gold standard' formula in healthcare.

"SMOG.C":

SMOG (Regression Equation C) (McLaughlin's 1969)

*0.9986 \* sqrt(Nwmin3sy \* (30 / Nst) + 5) + 2.8795*

where *n\_{wsy>=3}* = Nwmin3sy = the number of words with 3 syllables or more. This measure is regression equation C in McLaughlin's original paper.

"SMOG.Simple":

Simplified Version of McLaughlin's (1969) SMOG Measure.

*sqrt(Nwmin3sy \* 30 / Nst) + 3*

"SMOG.de":

Adaptation of McLaughlin's (1969) SMOG Measure for German Texts.

*sqrt(Nwmin3sy \* 30 / Nst) - 2*

**NOT RELEVANT**

"ELF":

Easy Listening Formula (Fang 1966):

*(Nwmin2sy / Nst)*

where *n\_{wsy>=2}* = Nwmin2sy = the number of words with 2 syllables or more.

"Linsear.Write":

Linsear Write (Klare 1975).

*[(100 - (100 \* Nwless3sy / Nw)) + (3 \* 100 \* Nwmin3sy / Nw)] / (100 \* Nst / Nw)*

where *n\_{wsy<3}* = Nwless3sy = the number of words with *less than* 3 syllables, and *n\_{wsy>=3}* = Nwmin3sy = the number of words with 3-syllables or more. The scaling by 100 arises because the original Linsear.Write measure is based on just a sample of 100 words)

"LIW":

Björnsson's (1968) Läsbarhetsindex (For Swedish Texts).

*ASL + (100 \* Nwmin7sy / Nw)*

where *n\_{wsy>=7}* = Nwmin7sy = the number of words with 7-syllables or more. The scaling by 100 arises because the Läsbarhetsindex index is based on just a sample of 100 words)

"nWS":

Neue Wiener Sachtextformeln 1 (Bamberger and Vanecek 1984).

*(19.35 \* Nwmin3sy / Nw) + (0.1672 \* ASL) + (12.97 \* Nwmin6char / Nw) - (3.27 \* Nw1sy / Nw)- 0.875*

where *n\_{wsy>=3}* = Nwmin3sy = the number of words with 3 syllables or more, *n\_{wchar>=6}* = Nwmin6char = the number of words with 6 characters or more, and *n\_{wsy=1}* = Nwsy1 = the number of one-syllable words.

"nWS.2":

Neue Wiener Sachtextformeln 2 (Bamberger and Vanecek 1984).

*(20.07 \* Nwmin3sy / Nw) + (0.1682 \* ASL) + (13.73 \* Nwmin6char / Nw) - 2.779*

where *n\_{wsy>=3}* = Nwmin3sy = the number of words with 3 syllables or more, and *n\_{wchar>=6}* = Nwmin6char = the number of words with 6 characters or more.

"nWS.3":

Neue Wiener Sachtextformeln 3 (Bamberger and Vanecek 1984).

*(29.63 \* Nwmin3sy / Nw) + (0.1905 \* ASL) - 1.1144*

where *n\_{wsy>=3}* = Nwmin3sy = the number of words with 3 syllables or more.

"nWS.4":

Neue Wiener Sachtextformeln 4 (Bamberger and Vanecek 1984).

*(27.44 \* Nwmin3sy / Nw) + (0.2656 \* ASL) - 1.693*

where *n\_{wsy>=3}* = Nwmin3sy = the number of words with 3 syllables or more.

"RIX":

Anderson's (1983) Readability Index.

*Nwmin7sy / Nst*

where *n\_{wsy>=7}* = Nwmin7sy = the number of words with 7-syllables or more.

"Scrabble":

Scrabble Measure.

*Mean Scrabble Letter Values of All Words*

. Scrabble values are for English. There is no reference for this, as we created it experimentally. It's not part of any accepted readability index!

"Spache":

Spache's (1952) Readability Measure.

*0.121 \* ASL + 0.082 \* Nwnotinspache / Nw) + 0.659*

where *n\_{wnotinspache}* = Nwnotinspache = number of unique words not in the Spache word list.

"Spache.old":

Spache's (1952) Readability Measure (Old).

*0.141 \* ASL + 0.086 \* (Nwnotinspache/ Nw) + 0.839*

where *n\_{wnotinspache}* = Nwnotinspache = number of unique words not in the Spache word list.

"Strain":

Strain Index (Solomon 2006).

*Nsy / (Nst / 3) / 10*

The scaling by 3 arises because the original Strain index is based on just the first 3 sentences.

"Traenkle.Bailer":

Tränkle & Bailer's (1984) Readability Measure 1.

*224.6814 - (79.8304 \* AWL) + (12.24032 \* ASL) - (1.292857 \* 100 \* Nprep / Nw)*

where *n\_{prep}* = Nprep = the number of prepositions. The scaling by 100 arises because the original Tränkle & Bailer index is based on just a sample of 100 words.

"Traenkle.Bailer2":

Tränkle & Bailer's (1984) Readability Measure 2.

*234.1063 - 96.11069 \* AWL - 2.05444 \* 100 \* (Nprep / Nw) - 1.02805 \* 100 \* (Nconj / Nw).*

where *n\_{prep}* = Nprep = the number of prepositions, *n\_{conj}* = Nconj = the number of conjunctions, The scaling by 100 arises because the original Tränkle & Bailer index is based on just a sample of 100 words)

"Wheeler.Smith":

Wheeler & Smith's (1954) Readability Measure.

*ASL \* 10 \* (Nwmin2sy / Nw)*

where *n\_{wsy>=2}* = Nwmin2sy = the number of words with 2 syllables or more.

"meanSentenceLength":

Average Sentence Length (ASL).

*Nw / Nst*

"meanWordSyllables":

Average Word Syllables (AWL).

*Nsy / Nw*