

# リーダビリティ研究の100年

野本 忠司<sup>1,a)</sup>

**概要:** テキストの読み易さをいかに測るかというリーダビリティの研究は今からほぼ100年前の1920年代に米国で始まった。その後、第2次世界大戦を経て、1950年頃に現在でも使われている。Flesh-Kincaid, Fox, Gunningなどの指標が誕生した。リーダビリティ研究は現在に至るまで大きなパラダイムシフトをいくつか経験しているが、本稿ではこれらがどのようなものであったのか、その時代背景と共に説明する。

**キーワード:** リーダビリティ, 歴史, 英語

## One Hundred Years of Research in Readability

TADASHI NOMOTO<sup>1,a)</sup>

**Abstract:** Research in readability began as early as the 1890s in the US, with the aim to develop an objective metric to measure how easily a given text is read. The field witnessed the burst of interest the late 1940s and early 1950s when readability indices such as Flesh-Kincaid, Fox and Gunning came into existence, which remain pretty much in use to this day. This paper traces some of the major work in readability research from its dawn to the present, focusing on how they impacted work that followed.

**Keywords:** Readability, History, English

### 1. はじめに

リーダビリティ研究は19世紀末の米国の文体研究に端を発している [6]。その後、1920年代に移民やその子弟が学校に押し寄せ、教科書がわからない児童が増えた。このため児童の能力に合った教科書を選ぶということが米教育界で大きな課題となった。また、当時「科学的な」手法を取り入れた教育法が注目されるようになったことなどと相まって、リーダビリティへの関心が高まった。

文章がリーダブル (readable) であるというのは、あるコホート (特定年齢の児童集団) が適度な早さで読むことができ、またその内容を十分に理解できる状態を指す [4]、当時のリーダビリティ研究の目的とは、ある教科書 (本) が何学年の児童に向いているかを計算する手法を見つけることであった。

Allow to Angriily					
allow	63	1b	ambition	23	3a
allowable	4		ambitious	15	4b
allowance	6		ambush	6	
allure	11	5b	amen	8	
ally	5		amend	12	5a
almanac	5		amendment	16	4a

図1 Teachers' book より抜粋

本稿ではリーダビリティ研究の流れを創成期から今日に到るまで辿りながら、その手法の変遷と特徴を概観する。なお、英語以外の研究については触れない。

### 2. 創成期

#### 2.1 Sherman, L. A. (1893)

1893年に当時ネブラスカ大学で英文学を教えていた Lucius Sherman 教授が文学研究に客観的手法を導入することを提唱し、論文を発表した [19]。この論文では、エリザベス前期、エリザベス期、ビクトリア期、それ以降と100の時代区分について文の平均的長さを計測し、時代が進むにつれ、文が短くなっていることを示した。(エリザベス前

<sup>1</sup> 国文学研究資料館・総合研究大学院大学  
NIJL, 10-3, Midori, Tachikawa, Tokyo 190-0014, Japan  
<sup>a)</sup> nomoto@acm.org

期は平均 50 語, Sherman の時代は 20 語と減少した。) さらに, どの書き手も平均的に見た場合一貫して同じ長さの文を書く, 文をより短くし, 具体性のある語彙を使うことでリーダビリティが向上する, 書き言葉は話し言葉に近づきつつあると先駆的な議論を展開した。

## 2.2 Thorndike, E. L. (1921)

1921 年に Thorndike はリーダビリティの金字塔とも呼べる研究を発表した [20]. その後 30 年近くこの分野に多大な影響を与えた. Thorndike 自身は心理学者であったが, ドイツ, ロシアの教師が語彙頻度を用いて, 児童に読ませる本を選んでいくことに注目, 頻度が高い語ほど「易しい」ことに気がついた. そして, 1921 年 (大正 10 年) に児童書, 聖書, 新聞等 41 の資料から抽出した約 4,565,000 語の中から, 頻度が最も高い 10,000 語を選び語彙頻度表を作った. これが, *Teachers' Book* の第一版 [21] である. (以降, TBW と呼ぶ.) (ちなみに完成まで 10 年の歳月を要した.) この後, 1932 年に 20,000 語 [22], 1944 年に 30,000 語 [23] に拡大された. また, この研究は後の「Zipf の法則」への道を開くきっかけとなった.

## 2.3 Lively, B. and Pressey, S. (1923)

1923 年になると Lively らが Thorndike の語彙表を使った, 教科書の読み難さの指標 ‘Vocabulary Burden’ (VB) を提案した [14]. 著者らが当時の中高生向け理科の教科書に使われていた語彙調査から難解な語が多く含まれていることを発見, 教科書間の語彙比較が必要となったことが背景にある.

この研究のポイントは語彙を教科書からいかにサンプリングするかということにある. 提案されている方法は, 教科書から 5 頁毎に 1 行つづサンプリングするというもの. 無論, 本のページ数に合わせて, 2 頁毎, 10 頁毎などにサンプリング間隔を調整, できるだけ全体から均等に語彙を採取する. 次に, TBW を参照して, そこに載っていない語を見つける. 特に TBW に未登録の語を *zero value word* (ZVW) と呼ぶ, 最後に, 採取した異なり語について, その TBW 指標に従って, 値が高い (易しい) ものから順に並べ ZVW の数を 2 倍にして配置した上でメディアンを見つける. これが VB の値となる. なお, Lively & Pressey ではこの値を *weighted median index number* と呼んでいる. 図 2 に著者らによる分析結果 (引用) を示す. 縦軸が TBW 指標, 横軸が教科書を表す. 棒が二重になっているのはサンプリングを二回行ったため.

## 2.4 Vogel, M. and Washburne, C. (1928)

リーダビリティのマイルストーン的な研究が 1928 年に Vogel らによって発表された [24]. 重要な点は語彙だけではなく, 文の構造的な情報 (複文, 単文, 疑問文, 平常

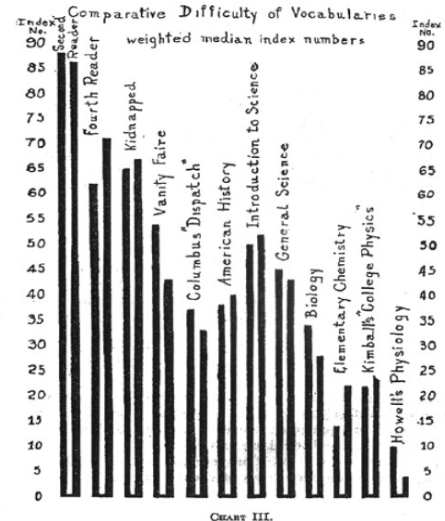


図 2 The Classic Readability Studies より引用

表 1 Vogel & Washburne (1928) より抜粋

TABLE I  
CORRELATION OF VARIOUS ELEMENTS WITH MEDIAN READING SCORE\*

Element	Correlation
Number of different words occurring in a sampling of 1,000 words .....	.770
Median index number (based on Thorndike's list) of 1,000-word sampling .....	-.704
Number of words in 1,000-word sampling not occurring in Thorndike's list ..	.674
Number of words in book .....	.592
Number of phrases in 1,000-word sampling .....	.576
Number of verbs in 1,000-word sampling .....	-.527
Number of words per paragraph .....	.518
Number of prepositions in 1,000-word sampling .....	.518
Number of phrases of all kinds in 75 sample sentences .....	.518
Number of phrases and clauses of all kinds in 75 sample sentences .....	.474
Number of adverbial phrases and clauses in 75 sample sentences .....	.467
Number of adverbial phrases and clauses in 1,000-word sampling .....	.463
Number of adjective phrases and clauses in 75 sample sentences .....	.461
Number of adverbial phrases in 75 sample sentences .....	.458
Number of words in 75 sample sentences .....	.453
Number of simple sentences in 75 sample sentences .....	-.371
Number of conjunctions in 1,000-word sampling .....	.296
Number of adverbial clauses in 75 sample sentences .....	.291
Number of nouns in 1,000-word sampling .....	-.262

文等) を導入していることに特徴がある, また, 児童の読解力テストのスコアとの相関を実際に測っているところにある. (図 1 に相関の結果を示す. ZVW, TBW との相関が高い. 前者が 0.674, 後者が -0.704. ) 昭和 3 年の論文であるが, 現代のリーダビリティ研究と比べても遜色がないレベルに達している.

Vogel & Washburne の大きな貢献は, *Winnetka formula* と呼ばれる以下の回帰式を開発したところにある.

$$y = 0.085x_1 + 0.101x_2 + 0.604x_3 - 0.411x_4 + 17.43 \quad (1)$$

$x_1$  は異なり語の数,  $x_2$  は前置詞の数,  $x_3$  は TBW 未登録語数,  $x_4$  は単文の数. ただし,  $X_{1-4}$  は, それぞれ, 1,000 語中のカウントを表す.  $x_1$  が必要な読解力レベルを表している. 実データ (読解力テストのスコア) との相関

表 2 スコアから学年 (grade) への換算

Formula Score	Corrected Grade Levels
4.9 and below	Grade 4 and below
5.0 to 5.9	Grades 5-6
6.0 to 6.9	Grades 7-8
7.0 to 7.9	Grades 9-10
8.0 to 8.9	Grades 11-12
9.0 to 9.9	Grades 13-15 (college)
10 and above	Grades 16 and above (college graduate)

は 0.845 であった。

## 2.5 Gray, W. and Leary, B. (1935)

それまでの研究の対象は児童であったが, Gray & Leary (G&L) [10] は初めて成人を対象にしたリーダビリティの研究を行った. G&L のユニークなところは, リーダビリティに影響しそうな属性をしらみつぶしに調べたことにある. コンテンツ, スタイル, フォーマット, 文章構成に関連した 228 属性から始めて最終的にスタイルに関連する 5 つの指標 (文の平均長, 難解語の異なり数, 人称代名詞の頻度, 異なり語数, 前置詞頻度) に絞ったモデルを提案した. 人間判定との相関は 0.645 と報告している.

## 3. 古典期

1940 年代に入るとリーダビリティに係る属性の整理, 統合が急速に進み, リーダビリティの理論的基盤が形成された.

### 3.1 Lorge, I. (1944)

G&L では 5 つの属性をもとにリーダビリティ指標を算出したが, Lorge は 3 つの属性 (文の平均長  $x_1$ , 100 語単位あたりの前置詞句の数  $x_2$ , Dale による「769 平易語リスト」[5]\*1 にない単語の数  $x_3$ ) に絞ったモデルを考案した [15].

$$y = 0.06x_1 + 0.10x_2 + 0.10x_3 + 1.99 \quad (2)$$

本件の調査対象は, 児童用教科書であったが, 大人を対象にしたテキストの評価にも利用された. また, テストデータとして, McCall & Crabbs による *Standard Test Lessons in Reading* [16] (以下, STL) を用いたが, これが以降の標準評価データとなった.

### 3.2 Flesch, R. (1948)

Flesch はもともとオーストリアで弁護士として活動していたがナチの迫害を逃れアメリカに渡り, 後にコロンビア大学教育学部で博士号を取得した. 1948 年に論文 “A New Readability yardstick” [9] を発表し, その中で以下の指標

\*1 <http://www.betterendings.org/homeschool/words/769words1.htm>

表 3 属性間の相関

TABLE I  
Intercorrelations between Four Style elements and grade Score of a Pupil Who Answered One-Half of the Questions on McCall and Crabbs

	Dale Score (3,000 List)	Flesch Affixed Morphemes†	Flesch Personal References	Lorge Hard Words (Dale 769)	Criterion $C_{50}$	Mean	Standard Deviation
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average sentence length	.5108	.4428	-.2201	.4913	.4681*	16.8037	5.3813
Dale score (words outside 3,000 list)	.....	.7932	-.4033	.7988	.6833	8.1011	6.3056
Flesch affixed morphemes†	.....	.....	-.3254	.7441	.6017	25.2819	11.0668
Flesch personal references†	.....	.....	.....	-.3422	.....	7.8245	5.5439
Lorge hard words (outside Dale 769 list)	.....	.....	.....	.....	.6148	17.4165	7.1659
Criterion ( $C_{50}$ )	.....	.....	.....	.....	.....	5.7492	1.6565

を提案した.

$$y = -1.015x_1 - 84.6x_2 + 206.835 \quad (3)$$

ここで,  $x_1$  は文の平均単語長,  $x_2$  は単語の平均音節数である. Lorge と異なり, 平易語リストのような外部知識を導入していない点に大きな特徴がある. 100 語ブロックを 1 つのサンプルとして計算する. このモデルは, Farr, Jenkins & Paterson (1951) [7] でさらに単純化された.

$$y = -1.015x_1 + 1.599x_3 - 31.517 \quad (4)$$

$x_3$  は 100 語ブロックにおける 1 音節語の数. STL を用いた  $C_{75}$ \*2 は, Flesch モデルと同等の 0.70 であった.

### 3.3 Dale, E. and Chall, J. (1948)

第二次世界大戦を背景に大衆向けのコミュニケーションの円滑化, 効率化への要請が高まりリーダビリティ研究は一段と加速した. この時期, 現在のリーダビリティモデルの基礎となる Dale-Chall formula, Flesch Reading Ease (式 3), Gunning-Fog Index などが誕生した. Dale & Chall [4] は Thorndike の TBW に対して単語の親密度を考慮していないとの立場を取り, 独自の *Dale List of 769 Easy Words* (以降, DL769) を考案した (注 \*1 参照).

$$y = 0.1579x_1 + 0.0496x_2 + 3.6365 \quad (5)$$

ここで  $x_1$  は, Dale List 上位 3,000 語 (DL769 の拡張版, 以降 DL3K) に含まれなかった語の数,  $x_2$  は文の平均単語長.  $y$  から表 2 を使って, 学年 (グレード) に換算にする.

表 3 は, Flesch 提案の属性 *affixed morphemes* (接頭辞, 接尾辞, 屈折語尾) (AF), *personal references* (代名詞の数) (PR), DL769, STL (表の  $C_{50}$ )\*3 との相関である. DL3K は, AF と DL769 と高い相関を示している. また, STL との相関も高い. AF は STL との相関が 0.6017 となっており, リーダビリティの指標として有効であるこ

\*2 STL のテストで 75%以上正答した児童の学年の平均との相関.

\*3  $C_{50}$  は STL の 376 テキストに関する読解力テストにおいて半分以上正答した被験者のスコアとの相関を表す.

表 4 属性の非線形性 (curvilinearity) (Table 3) [1]

Table 3 Tests of the linearity of the regressions of linguistic variables on comprehension difficulty at the word level of analysis

Linguistic variable	r	$\eta$	F*
SYL	395	426	12.29
LET	498	520	12.26
WOR FRE	286	334	13.42
WOR DEP	002	058	1.34

\* With 14 and 5,166 d.f. and F of 1.64,  $p < .05$ .

と示している。しかし、AF との相関が高い DL3K と STL の相関は 0.6833 であることから、リーダビリティの属性としては DL3K のほうが優れていることがわかる。なお、Dale & Chall は Flesch の AF の選び方が恣意的であり、客観性に欠くと批判した\*4。

### 3.4 Gunning, R. (1952)

1952 年になると Gunning (1952) [11] はリーダビリティを測る尺度として以下を提案した。

$$y = 0.4(100x_1 + x_2) \quad (6)$$

$x_1$  は、テキスト中の難解語 (complex words) の割合 (Dale & Chall モデル (式 5) の  $x_1$  に対応)  $x_2$  は平均音節数を表す。ただし、難解語とは 3 音節以上からなる語。ただし、固有名詞、複合語、ed, es で終わる 3 音節語を除く。明らかに Dale&Chall の垂流であるが、平易語リストを用いない点が異なる。Flesch モデルに近いと言える。(Gunning と Flesch はビジネス上の付き合いがあった。) なお、式 6 は Gunning Fog Index と呼ばれる。式の単純さから、今でも利用されている。

## 4. クローズテスト (Cloze Test) の時代

それまでの研究では多肢選択型 (multile-choice) の文章題を用いて被験者の読解力を測るのが主流であった。しかし、この方式ではテスト問題の難しさを測っているのか、テスト問題の質問の難しさを測っているのか区別できない。また、往々にして質問作成も恣意的になりがちであり、実験結果の比較が困難である。このため、クローズテスト (cloze test) と呼ばれるテスト手法が新たに考案された。

クローズテストとは課題文を構成する単語を適当な間隔で空欄に置き換え、被験者にその空欄を埋めるように求め、被験者の読解力を計るテストである。

### 4.1 Bormuth, J. (1966, 1969)

Bormuth [1, 2] によってリーダビリティ研究はその頂点を迎えたと言える。Bormuth の貢献は以下の 3 つに集約される。

\*4 この批判は Flesch の 1943 年の論文 [8] に向けられたもの。

- (1) クローズテストの読解指標としての有効性を統計的に確認した。
- (2) 属性とリーダビリティとの関係は必ずしも線形ではないことを示した。
- (3) 音韻から談話まで様々な言語レベルに関わる 100 近くの属性の影響を網羅的に調査した。

表 4 は属性の非線形性を示す例である。データは、文学、歴史、地理、生物、物理の教科書から取った 275 から 300 語の文章 20 編 (1 分野 4 編) である。全体では、語数 5,181、文数 365 の規模。1 文章から 5 つのクローズテストを作成した (語を削除する間隔を変える)。被験者は米カリフォルニア州ワスコ市 (Wasco) の小学校 4 から 8 学年の児童 (被験者数は不明)。実験では児童を 5 つのグループにわけ、それぞれのグループは同一のクローズテストを与えた。

表 4 はその正答率とそれぞれの属性の関係を示している。(それぞれの属性について回帰直線を構成し、その予測と正答率の相関を表している。) SYL は一単語あたりの音節数。LET は単語あたりの文字 (アルファベット) 数、WOR FRE は TBW と Dale の語彙表をもとにした難易度インデックス、WOR DEP は構文木に基づいた語の出現位置の心理的負荷を表す指標である、前半ほど重く、後半ほど軽くなる。

$r$  はピアソンの積率相関係数 (Pearson product-moment correlation coefficient),  $\eta$  は非線形度を測る尺度、 $F^*$  は  $F$  統計量。この例では、 $F \geq 1.64$  で  $p < 0.05$ 。ただし、 $r$  と  $\eta$  の値は、小数点が脱落していると思われる。従って  $r = 395$  は  $r = 0.395$  の誤り。表によると WOR DEP 以外、すべての属性について非線形性が現れていることが分かる。

表 5 は Bormuth [2] で提案したリーダビリティモデルの 1 つである。利用形態によって (計算機用、手計算用など)、使い分けられるようないくつかのタイプを用意している。ここで LET/W は単語当たりの平均文字数、DDL/W は DL3K の登録語の平均出現頻度、W/SEN は文の平均単語長、modal v は modal verb の頻度を表す。モデルは、心理学の入門書から 330 の文章を抽出し、クローズテストを実施した上で、特定の有効水準 (35, 45, 55%) を設定し、それを超えるデータを用いて構成した。クローズテストでそれぞれの水準を満たした文章について別途読解力テスト (多肢選択問題) を施行した。そこで得たスコアを GP(35), GP(45), GP(55) と呼ぶ。表 5 のモデルはこれらの予測モデルである。Cloze Mean はクローズテストのスコア自体を予測するモデルとなっている。

このモデルの性能を表 6 に挙げた。ここで、Original Data の欄は学習データ、Cross Validation はテストデータ (サイズ 20) での結果を表す。注目すべきは、 $r$  の値であるが、学習データで 80%、テストデータで 90% 強を得ている。(因みに、stepwise polynomial regression を用いてモ

表 5 Bormuth のリーダビリティモデル (Table 15) [2]

$$\begin{aligned} \text{Cloze Mean} &= .886593 - .083640(\text{LET}/W) + .161911(\text{DLL}/W)^3 - .021401(W/\text{SEN}) + .000577(W/\text{SEN})^2 \\ &\quad - .000005(W/\text{SEN})^3 \\ \text{GP}(35) &= 3.761864 + 1.053153(\text{LET}/W) - 2.138595(\text{DLL}/W)^3 + .152832(W/\text{SEN}) - .002077(W/\text{SEN})^2 \\ \text{GP}(45) &= 3.398999 + 1.107014(\text{LET}/W) + .155327(W/\text{SEN}) - .002184(W/\text{SEN})^2 + 6.672669(\text{DLL}/W)^2 \\ &\quad - 7.523689(\text{DLL}/W)^3 - 5.266225(\text{Modal } v) \\ \text{GP}(55) &= 3.450806 + 1.094841(\text{LET}/W) + .153830(W/\text{SEN}) - .002242(W/\text{SEN})^2 + 11.478313(\text{DLL}/W)^2 \\ &\quad - 11.224816(\text{DLL}/W)^3 - 5.427013(\text{modal } v) \end{aligned}$$

表 6 Bormuth モデルの性能 (Table 15) [2]

Score Predicted	Formula Validity							
	Original Data (N = 330)				Cross Validation (N = 20)			
	R	S.E.	F	Linearity F	r	S.E.	Linearity F	Mean Bias
Cloze Mean	.833	.063	147.36*	.18	.920	.04	.39	--
GP(35)	.808	.884	153.13*	1.75	.916	.57	.29	1.49*
GP(45)	.825	.852	114.40*	.00	.927	.54	.72	1.70*
GP(55)	.808	.860	101.22*	.00	.915	.52	.77	1.91*

表 7 Kincaid モデル (Table 3) (Kincaid et al. [13] から引用)

Automated Readability Index

Old:  $GL^* = 0.50 (\text{words/sentence}) + 4.71 (\text{strokes/word}) - 21.43$

New:  $GL = 0.37 (\text{words/sentence}) + 5.84 (\text{strokes/word}) - 26.01$

Simplified:  $GL = .4 (\text{words/sentence}) + 6 (\text{strokes/word}) - 27.4$

Fog Count

Old:  $GL = \frac{\text{easy words} + 3 (\text{hard words})}{2 (\text{sentences})}$ , if Average Fog Count\*\*\*  $\geq 20$

$GL = \frac{\text{easy words} + 3 (\text{hard words})}{(\text{sentences})} - 2$ , if Average Fog Count  $< 20$

New:  $GL = \frac{\text{easy words} + 3 (\text{hard words})}{(\text{sentences})} - 3$

Flesch Reading Ease Formula

Old:  $RE^{***} = 206.835 - 1.015 (\text{words/sentence}) - .836 (\text{syllables}/100 \text{ words})$

New:  $GL = .39 (\text{words/sentence}) + 11.8 (\text{syllables}/word) - 15.59$

Simplified:  $GL = .4 (\text{words/sentence}) + 12 (\text{syllables}/word) - 16$

表 8 SMOG モデル

Regression Equation	Correlation
(a) $g = 6.2380 + 0.0785 p$	<b>0.713</b>
(b) $g = 4.1952 + 0.8475 \sqrt{p}$	<b>0.709</b>
(c) $g = 2.8795 + 0.9986 \sqrt{p} + 5$	<b>0.729</b>
(d) $g = 1.0430 (3 + \sqrt{p})$ $= 3.1291 + 1.0430 \sqrt{p}$	<b>0.985</b>

デル空間の探索を行っている。) 特筆すべきは、各モデルで導入している属性である。基本的に文の長さ、音節数、平易語の数である。これは Dale&Chall とまったく同じである。結局、リーダビリティを決める要因は、表層的な単語、文の大きさと語彙の親密度だけというのが Bormuth の結論である。

#### 4.2 Kincaid, J., Fishburne, R., Rodgers, R., and Chissom, B. (1975)

1970 年代になると Kincaid, et al. [13] は米海軍の教育用マニュアルに合わせたリーダビリティ指標の提案を行った。基本的に Gunning Fog Index (節 3.4) をベースとしているが、使い勝手を考慮し若干改良している。データは Rate Training Manuals と呼ばれる兵務に関する指南書から 18 編 (平均 170 語) 収集し、それぞれについて多肢選択型の読解力テスト (Gates-MacGinitie Reading Test (GMRT)) を実施しグレードレベルを決定した上で、クローズテストを行った。そして、各グレードについて、そのグレードに属する被験者の 50%以上が対応するクローズテストで 35%以上正答した場合にのみ、そのデータを採用することとした。このようにして収集したデータを用いてモデルのパラメータを決定した。被験者は海軍訓練センター (Great Lakes Naval Training Center, The Naval Air Station (Memphis)) 所属の 569 名であった。

表 7 にモデルの詳細を載せる。3 つのモデルが提示されている。それぞれ、Old が従来モデル、New が新データでパラメータを調整したモデル、Simplified がその簡約版となっている。ここで、*stroke* は Bormuth の LET と同じくアルファベット数を指す。GL はグレード、Fog Count 中の *easy words* とは 3 音節未満の単語 (c.f. 式 6), *hard words* はそれ以上の音節を含む単語、また Average Fog Count とは  $\frac{E+3*H}{S}$ 。ただし、*E* は *easy words*, *H* は *hard words*。これらのモデルの性能については確認できなかった (モデル間の相関は 8 割を超える)。Kincaid モデルは基本的には Gunning Fox Index の焼き直しであるが、その簡便さから近年の文簡約の研究でも盛んに用いられている [25]。ただ、学習データが海軍兵務マニュアルという点で、使い方には注意が必要である。

#### 4.3 McLaughlin, G. (1969)

McLaughlin [17] は Gunning Fox Index をさらに簡便に

表 9 Collins-Thompson & Callan (2005) の結果

DATA	SIZE	GL	SU	UNK	TYPES	MLK	FK
DRS	17	1.4 - 5.5	0.67	0.72	<b>0.93</b>	0.50	<b>0.93</b>
AZ	228	1.0 - 6.0	0.63	0.78	<b>0.86</b>	0.49	0.30
W1	250	1.0 - 6.0	<b>0.69</b>	0.38	0.26	0.36	0.25
W2	550	1.0 - 12	<b>0.79</b>	0.63	0.38	0.47	0.47

した SMOG モデルを発表した。データは STL から 390 のテキストを選び、それぞれから 30 文を抽出。対応する学年は 100%正答した被験者の学年を利用。試行錯誤の末、表 8 を得たと述べている。ここで、*p* は 30 文中の 3 音節以上含む単語の数を表す (Dale&Chall の *complex words* に対応する)。モデルの精度は (学習データとの相関) は表の右に提示されている通りである。McLaughlin の研究のポイントは、(1) テキスト全体の統計量を測る必要はなく、30 文程度のサンプリングでよい、(2) 単語の難易度は音節数で予想できると主張している点にある。

## 5. AI の時代

近年のリーダビリティ研究は大量データへのアクセスが格段に容易になったという時代的な背景もあり、統計や機械学習の技術的成果を積極的に取り入れるというアプローチが主流である。ただ、学習モデルの構成のみに注意が集中し、リーダビリティを阻害する要因は何かという根源的な問いへの探求が軽視されていることが憂慮される。

この意味で見るとべき研究があまり存在しない。例えば、Pilter [18] は、PTB (Penn Tree Bank) のテキスト (新聞データ) を用いた大学生被験者のリーダビリティ判定を SVM で学習させる方法を提案しているが、被験者の層、利用したデータのジャンルの偏りが強い。このためリーダビリティ研究への貢献は極めて小さいと言わざるを得ない。

唯一注目されるのが、Collins-Thomson & Callan [3] であろう。著者らはいわゆる Flesch Kincaid モデル (表 7 の Flesch Reading Ease の改良版 (New) に相当) がウェブ文書のリーダビリティ予測に使えないということに気づき、言語モデルをベースにしたアプローチを展開した。

$$L(T | G_i) = \sum_w C(w) \log P(w | G_i) \quad (7)$$

ここで *w* は単語トークン、 $P(w | G_i)$  は学年  $G_i$  のもとで *w* が発生する確率である。但し、*P* は kernel smoothing 法 [12] を使って補正する (学年によって単語分布が異なるという事実を反映することが目的)。

彼らの実験結果を表 9 に載せる。DRS は Spache Diagnostic Reading Scales (小学低学年用文章読解力テスト)、AZ は READINGA-Z.COM<sup>\*5</sup>、W1 と W2 はインターネットから採取したオンライン教材のコパスを表す。GL は学

\*5 <https://www.readinga-z.com>

年レベル (grade level). SU は式 7, UNK は DL3K に未登録語の割合, MLF は British National Corpus の平均対数頻度, TYPE は 100 語ブロック中のタイプ数をそれぞれ指標にしたモデル. FK が Flesch Kincaid モデル. モデルの下の数字は, データとの相関を示す. SU が W1-2 で性能がよいことが分かる. FK は DRS で成績がよい. これは FK の想定している利用条件に DRS が近いと推察される. DL3K ベースの UNK は W1 で相関が低くなるが, 概ね健闘している点は興味深い.

## 6. おわりに

この約 100 年に及ぶ研究で分かってきたのは, リーダビリティの問題は基本的に語彙に還元できるという点であろう. Dale&Chall リストが誕生から 50 年を経ても, 健在であるのは驚きに値する. しかし, なぜ語彙がリーダビリティを左右するのかという問いは非常に興味深い. 学校教育を通して特定の語彙に対して親密度が上がるように人々が訓練されているからかもしれない. つまり, リーダビリティとは, 人々の生得的な認知現象というより, 訓練を受けた結果生ずるのではないかという考え方である. 残念ながら, 現在このような認知現象としてのリーダビリティの研究はあまり進んでいないが, 将来このような視点での研究が発展することを期待したい.

なお, 本稿は Thomas François 氏 (Université Catholique de Louvain) の Natural Language Generation Summer School 2015 での資料 (*Readability: a one-hundred-year-old field still in his teens*) を参考に作成した. この場を借りて感謝する.

## 参考文献

- [1] Bormuth, J. R.: Readability: A new approach, *Reading Research Quarterly*, Vol. 1, pp. 79–132 (1966).
- [2] Bormuth, J. R.: Development of Readability Analyses, Final Report, Project No. 7-0052, U.S. Department of Health, Education and Welfare (1969).
- [3] Collins-Thompson, K. and Callan, J.: Predicting reading difficulty with statistical language models, *Journal of the American Society for Information Science and Technology*, Vol. 56, No. 15, pp. 1448–1462 (2005).
- [4] Dale, E. and Chall, J.: A formula for predicting readability, *Educational research bulletin*, Vol. 27, No. 1, pp. 11–28 (1948).
- [5] Dale, E.: A Comparison of Two Word Lists, *Educational Research Bulletin*, Vol. 10, No. 18, pp. 484–489 (online), available from (<http://www.jstor.org/stable/1470966>) (1931).
- [6] DuBay, W. H.(ed.): *The Classic Readability Studies*, Impact Information (2006).
- [7] Farr, J. N., Jenkins, J. J. and Paterson, D. G.: Simplification of the Flesch Reading Ease Formula, *Journal of applied psychology*, Vol. 35, No. 5, pp. 333–357 (1951).
- [8] Flesch, R.: *Marks of a Readable Style*, A Study in Adult Education, No. 897, Teachers College, Columbia University (1943).
- [9] Flesch, R.: A new readability yardstick, *Journal of Applied Psychology*, Vol. 32, No. 2, pp. 221–233 (1948).
- [10] Gray, W. S. and Leary, B. E.: *What makes a book readable - with special reference to adults of limited reading ability*, The University of Chicago Press (1935).
- [11] Gunning, R.: *The Technique of Clear Writing*, McGraw-Hill, New York (1952).
- [12] Hastie, T., Tibshirani, R. and Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer (2013).
- [13] Kincaid, J. P., Fishburne, R. P., Rogers, R. L. and Chissom, B. S.: Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel, Research branch report 8-75, Naval Technical Training Command (1975).
- [14] Lively, B. A. and Pressey, S. L.: A Method for Measuring the “Vocabulary Burden” of Textbooks, *Educational Administration and Supervision*, Vol. 9, pp. 389–398 (1923).
- [15] Lorge, I.: Predicting Readability, *Teachers College Record*, Vol. 45, pp. 404–419 (1944).
- [16] McCall, W. A. and Crabbs, L. M.: *Standard Test Lessons in Reading. Books II, III, IV, and V*, Bureau of Publications, Teachers College, Columbia University, New York (1938).
- [17] McLaughlin, G. H.: SMOG grading: A new readability formula, *Journal of Reading*, Vol. 12, No. 8, pp. 639–646 (1969).
- [18] Pitler, E. and Nenkova, A.: Revisiting readability : A unified framework for predicting text quality, *Proceedings of the the Conference on Empirical Methods in Natural Language Processing*, pp. 186–195 (2008).
- [19] Sherman, L. A.: *Analytcs of Literature: A Manual for the objective study of English prose and poetry*, Ginn and Company (1893).
- [20] Thorndike, E.: *A Teacher’s Word Book of Twenty Thousand Words*, Bureau of Publications, Teachers College Columbia University (1923).
- [21] Thorndike, E. L.: *The Teacher’s Word Book*, Teachers’ College, Columbia University, New York City (1921).
- [22] Thorndike, E. L. and Lorge, I.: *The Teacher’s Word Book of Twenty Thousand Words*, Teachers’ College, Columbia University, New York City (1932).
- [23] Thorndike, E. L. and Lorge, I.: *The Teacher’s Word Book of 30,000 Words*, Teachers’ College, Columbia University, New York City (1943).
- [24] Vogel, M. and Washburne, C.: An Objective Method of Determining Grade Placement of Children’s Reading Material, *ELementary school journal*, Vol. 28, pp. 373–381 (1928).
- [25] Woodsend, K. and Lapata, M.: Learning to simplify sentences with quasi-synchronous grammar and integer programming, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, ACL, pp. 409–420 (2011).