

リーダビリティの原理と速読効果 —日本人向けリーダビリティ公式開発—

田淵 龍二 (ミント音声教育研究所)

文を読むとき無意識の内に心の中で声にしている。認知心理学ではこれを文字の音韻符号化 **phonological coding** と言う。さて発声時の音声には平均 2 秒前後のまとまりがあり (湯舟・田淵, 2013), 音声を取り込む音韻ループには 2 秒前後の時間的制約があるとされる (Baddeley)。そこで読解時の内声にも 2 秒前後のまとまり (内声チャンク **Chunk of Inner Voice; CIV**) があり, この CIV 形成のしやすさと文の CIV 密度 (文に含まれる平均内声チャンク数) が読みやすさを決める主要因であるとの原理仮説 (以下 **CIV 仮説**) を立てた。この仮説を元に綴りや区切りなどの書記情報から CIV 時間長を予測し, この CIV 時間長と CIV 密度を 2 変数とする英文のリーダビリティ公式を英語母語話者向け **MGEN** と日本人英語学習者向け **MGJP** の 2 種類作成した。米国で定評のある **Flesch-Kincaid** 公式 (**FKGL**) と 2 つの新公式を比較した結果 0.9 を超える相関を得た。次に **FKGL** など 8 種の従来公式との距離 (値の近さ) を計算した結果, **MGEN** と **FKGL** がもっとも近接していることがわかった。また読解試験課題文のリーダビリティ値と試験時に測定した **WPM** の相関係数は -0.91 となり, 絶対値が 0.9 を超えた。さらに G1~9 の米語学教材 16 冊による検証で出版社指定学年との誤差 0.3 ± 0.8 を得た。以上より, 語長や文長などを変数とする従来公式も, 実は CIV 形成のしやすさを原理としていた可能性が高く, **CIV 仮説** を支持する結果となった。

1. リーダビリティの原理と, CIV 仮説による新しいリーダビリティ公式

リーダビリティ公式は 1920 年代以降の長い歴史があり, その多くは単語の親密度や単語と文の長さを変数としている。その原理は親密な単語で短い文ほど読みやすい (はず) と言うものであった。今回の新しい公式では, 呼吸活動を伴わない読解時の内声にも 2 秒前後の音のまとまりが存在すると仮定した。これは一定の分量の単語列を 2 秒前後のチャンクとなるような速さで読むこと, あるいは速く読めることが理解を促進することにも通じる考えである。

さて次に示す新公式は CIV 予測時間長と CIV 密度を 2 変数とし, 読解適応学年 (**reading grade level; GL**) を出力する。

$$\text{MGJP : } \text{GL} = 0.07496 \frac{(3S_y + 2C_n)}{Ph} + 7.926 \log \left(\frac{Ph}{Cl} \right) + 4.618$$

$$\text{MGEN : } \text{GL} = 0.07662 \frac{(3S_y + 2C_n)}{Ph} + 19.554 \log \left(\frac{Ph}{Cl} \right) - 3.141$$

where S_y : number of syllables; C_n : number of consonants; Ph : number of phrases Cl : number of clauses, a clause is approximately equal to a sentence; \log : common logarithms;

note: 式中の $(3S_y + 2C_n) / Ph$ が平均 CIV 予測時間長, (Ph / Cl) が文長 (平均 CIV 数; CIV 密度) に関わる項。音節数子音数などは **Mingle** 搭載の計測エンジンによる。英数字と空白だけの文字列をフレーズ **phrase** とし, これを CIV の単位とみなした。4 種の終止記号 (. ! ? ;) で区切られた文字記号列を節 **clause** とし, 文 **sentence** に準じた。定数は日米の語学教科書 23 冊を基準文書として重回帰分析により求めた。

新公式をフレッシュ・キンケイド公式などと比較した結果、MGJNはFKGLともっとも近接していることがわかった（表1、図1）。

表1. 相関係数

	● MGJP	▼ MGEN
△ FKGL	0.90	0.94
CLI	0.90	0.86
FORC	0.68	0.64
FryG	0.95	0.95
GFog	0.89	0.94
SMOG	0.91	0.93
LWF	0.84	0.91
ARI	0.89	0.94

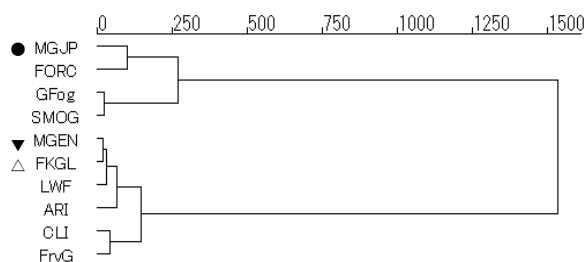


図1. リーダビリティ値の近接関係

note: 比較資料には日英の各ジャンル（教科書，検定試験，政治法律，ニュース，文学，利用規約など）から選んだ39文書を使った。デンドグラムは平方距離に基づく。

チャンク音読研究（科研24501196 神田他）で実施した読解試験（大学生81人分）の成績と8課題文のリーダビリティ

表2. 読解試験の成績とGLの相関係数

	FKGL	MGJP
wpm	-0.86	-0.91
score	-0.62	-0.39

値を解析したところ、MGJPと読みの速さWPMとは絶対値が0.9を越える負の相関となり、MGJPが読みやすさ（読みにくさ）を反映していることが示された（表2、図2）。この負の相関は、テキストの予測CIV長が長くなるほど1チャンクの音韻符号化に手間取り最適認知時間幅の2秒±1程度を超えてしまうことが理解の妨げ（読み速度の低下）につながる可能性を示唆している。

大き目のチャンク（フレーズ）はスピードを上げて（内声構音速度を上げて）処理する訓練が、読解力向上の対策として考えられる。

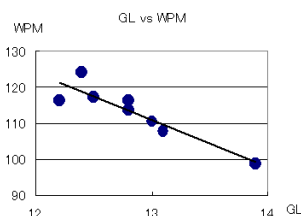


図2. WPM（読みの速さ）とGL（適応学年）は負の相関
note: 点一つが課題文。課題文は英検準2級で約300単語

2. 読解の仕組みとリーダビリティ公式の限界

読解では解字⇒内声⇒意味と進んで理解が形成される（図3）。リーダビリティ値はこの流れ（右図の上向き実線矢印）が無意識行われる場合の速度に依存する読みやすさであり、読解方略を意識的に使用する事による読みの停滞や返り読みなど（下向き点線矢印）は反映されない。

参考文献

湯舟・田淵 (2013). 「映画音声コーパスを利用した Breath Group 長の分析」. Language Education & Technology, 50, 23-41.

資料

Mingle8+2 Official Website: <http://mintap.kir.jp/public/mingle/mingle.html>

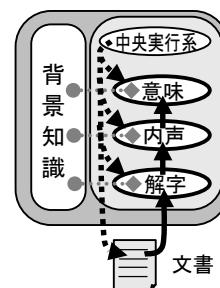


図3. 読解の仕組み reading model