

# 「読書からの語彙学習」効果シミュレーションのための予備的検討

猪原 敬介<sup>†</sup>

Keisuke Inohara

<sup>†</sup>くらしき作陽大学

Kurashiki-Sakuyo University

kei.inohara@gmail.com

## 概要

読書が語彙学習に及ぼす効果について、Nagy [1]やCarver [2]の試算がこれまで提案されてきたが、関連する変数間の相互作用を仮定しないなどの点で不十分であった。本研究では、読書時間、未知語率、読み速度、学習率などについて、平均値、想定される個人差、変数間相互作用を実装する関数を、先行研究を参照することで設定し、試算を行った。結果として、標準的な読み手が年間に獲得する語彙数は、延べ数で 3,371 語であると試算した。

キーワード：語彙、読書、語彙数、計算モデル

## 1. はじめに

語彙力を高める手段の一つに、読書からの語彙学習がある。読書が語彙学習に及ぼす効果について、Nagy [1]は年間 4500 語、Carver [2]は 160 語という試算を出している（以下、それぞれ、Nagy 試算、Carver 試算と呼ぶ）。これらの試算は、計算の前提となる 1 回の読書時間、読書速度、未知語率、学習率などについて、設定の根拠が示されないものが多いばかりでなく、個人差の考慮や「未知語率が高いほど、読み速度が遅くなり、学習率が低くなる」といった変数間相互作用を仮定していない点で不十分である。

こうした多数の変数が互いに相互作用する様子を厳密に捉えるためには、計算モデルの導入が効果的である。本研究では、先行研究を踏まえたモデル化を行い、読書が語彙数増加に及ぼす効果についてシミュレーションを行う。

## 2. 方法

### 読書時間

平成 30 年度全国学力・学習状況調査報告書・調査結果資料[3]、および、猪原・上田・塩谷・小山内[4]のデータから、1 日の平均読書時間は 30 分以下が多いものの、30 分～1 時間、1～2 時間、2 時間以上にも一定割合で分布しており、個人差が大きいことが分かる。以上の先行研究から、読書時間について、10 分、20 分、30 分、45 分、1 時間、2 時間の 6 水準で検討を行うこととした。

### 未知語率・未知語率減衰係数

未知語率は、読み手の語彙とテキストに含まれる単語の難しさの相対的な関係から生じる。Carver [2]の調査では、読み手のレベルとテキストの難易度が釣り合っている際には、未知語率は約 0.7～1.9%であり、非常に簡単な場合にはほぼ 0%、非常に難しい場合には 4.3%～7.7%以上になると報告している。このことから、未知語率については、0.5%、1%、2%、3%、4%、7%、10% の 7 水準で検討する。

読書を続けることで語彙数が増えるため、同じ語彙レベルのテキストを読み続けると、未知語率は低下していくと考えるのが自然である。このことをモデルに反映するために、語彙数が増えるにつれて未知語率が低下する割合を定める未知語率減衰係数を設定する必要がある。上述の Carver [2]の Table 3 に、小学校 3～6 年生を対象とした、テキストと読み手の文章理解力の相対的な差に対する未知語率が掲載されている。このうち、相対的な差が 0（文章理解力とテキスト難易度が釣り合っている）から 7（7 学年分、文章理解力が足りていないことを意味する）の未知語率を目的変数、学年（0～7）を説明変数として単回帰分析を行うと（ $N=8$ ）、回帰係数は 0.4929 となる（ $R^2 = 0.5798$ ）。つまり、1 学年分の文章理解力が足りないことで、未知語率は 0.4929 増加するということになる。

さらに、国立国語研究所 [5]で報告されている小学 1 年生から中学 3 年生までの推定語彙数に対して、上記と同様の回帰分析を行うと、平均的な児童は、あらゆる学習源からの語彙学習効果として「年間 3330 語獲得する」と考えられる。ただしこの語数は「異なり語数」であり、試算に用いられるのは「延べ語数」である。「現代日本語書き言葉均衡コーパス (BCCWJ) [6]」における書籍（流通実態）の「延べ語数/異なり語数」比率はおおよそ 2.5[7]であることを踏まえると、年間獲得語彙数は、3330 語に 2.5 をかけた 8325 語と考えることができる。1 学年分の未知語率である 0.4929%を、年間獲得語彙数である 8325 語で除することで、「1 語獲得

するたびに未知語は 0.0000592%減少する」という目安を得ることができる。そこで、未知語減衰係数として、0 (減衰なし), 0.0000003, 0.0000006, 0.0000009 の 3 水準で検討する。ただし、未知語率の下限は 0.000001 とし、これ未満にはならないように設定した。

#### どういった難易度の本を読むか：未知語率回復間隔

上述のように、読書が続けると未知語率が徐々に下がっていくが、それはあくまでも同じ難易度の本を読み続けた場合である。読み手に、徐々に難易度の高い本を選ぶ傾向があれば、下がった未知語率が回復するはずである。こうした傾向の強弱を、下がった未知語率を元の水準まで回復させる間隔(日数)として、「未知語率回復間隔」をモデルに組み込む。読み手が、より難しいテキストを読むことに意欲的であったり、適切な難易度の本を推薦してくれる人がいる場合には間隔は短く、未知語が少ない状態での読書に満足している読み手であれば間隔は長くなるだろう。先行研究がないため、恣意的な水準設定にはなるが、1 読書日(未知語率が減衰しない), 10 読書日, 30 読書日, 50 読書日, 100 読書日, 未知率回復なし, 7 水準で検討する。

#### 基本読み速度

英語の読み速度については、1 分当たりの単語数(wpm; words per minute)で表現することが一般的である。一方、日本語の読み速度については、1 分当たりの文字数で表現することが一般的である。例えば、小林・川嶋[8]では、日本人大学生 200 名の読み速度について、323 文字/分~1189 文字/分まで分布し、平均値は 653 文字/分であったと報告している。今回、語彙数についてのシミュレーションを行うに当たっては、1 分当たりの文字数をまず wpm への変換することが必要である。BCCWJ[6]の「長単位語彙表データ (v1.0)」に基づく、BCCWJには 173,794,108 文字(延べ数)が収録されており、さらに総延べ単語数は 83,309,532 単語であった。そこで、1 文字当たりの単語数を計算すると、0.48 単語となる。ここに小林・川嶋[8]のデータを掛けると、日本人大学生は平均的に 304.8wpm の読み速度となり、個人差として、155.04wpm ~ 570.72wpm であると考えられる。

ただし、上記の読み速度はあくまでもおおよその目安である。そのことも考慮し、読み速度について、100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600wpm と、11 水準で検討する。

#### 修正読み速度係数

さらに、同じ読み手であっても、読むテキストの難易度によって読み速度は大きく異なる。未知語率の高い本というのは、その読み手にとって難しい本であるため、読み速度は下がるはずである。すなわち、読み速度と未知語率にはトレードオフ関係があるはずである。

上述の、未知語率を目的変数、学年(0~7)を説明変数とした単回帰分析の結果である回帰係数 0.4929%を参考とすると、未知語率 0.5%をおよそ 1 学年分の読み手とテキストの相対的な差と捉えることができる。さらに、これも上述した Carver (1992) のデータ (Table 2) では、1 学年相当上がるごとに、14wpm 上昇することが想定されていた。そこで本研究では、修正読み速度係数を 28 とし、基本読み速度から、未知語率が 1% 上昇するごとに 28wpm 低下した値を修正読み速度として採用した。

#### 学習率：未知語によって変動する学習率関数

学習率について、Nagy 試算では 15%、Carver 試算では 5%とされていたが、前者が大学生を、後者が小学 5 年生を主に想定しているなど、どちらが正しいとはいえない。Swanborn & de Glopper [9]は偶発的単語学習を扱った 20 の実験をメタ分析した結果、学習率は平均的には 15%ほどであることを報告しているが、Swanborn & de Glopper [10]では、文章理解力の高低によって、ほぼ 0%にも 27%にもなり得ることも報告されている。このように、学習率についても、読み手のレベルとテキストの難易度との相対的な差によって変動すると考えるほうが自然である。

上述したように、未知語率も読み手のレベルとテキストの難易度との相対的な差によって生じる。Swanborn & de Glopper [9]の論文に掲載されている、各実験の未知語率と、実験の結果として得られた学習率のデータ (Table 1 および Table 3) について、回帰分析を行った(N=20)。線形回帰では説明率が低かったため、未知語率を対数変換した上で回帰分析を行うと、

$$y = -0.149 \times \ln x - 0.4692$$

という回帰式を得た ( $R^2 = 0.2946$ )。ここで、y は学習率、x は未知語率である。そこで、シミュレーションでは、各時点の未知語率を上記の式の x に代入した学習率の値を用いる。これは、未知語率が 1%のときには学習率が 21.7%、3%のときには 5.3%、4%のときには

は 1.04%, 5%のときには-2.28%の学習率となるような式である。学習率が負の値となった場合には, 0%に変換する。

### 3. 結果および考察

#### Nagy 試算, Carver 試算, モデルケースの検討

表 1 の条件設定に基づいてシミュレーションを行った結果が, 図 1 である。年間獲得語彙数で言えば, Carver 試算が最も少なく 176 語 (Carver [2]では約 160 語と論文に掲載されていたが, 示された前提条件から計算すると 176 語となった), 次に「読書嫌い」が 329 語, 「標準」が 3,371 語, Nagy 試算が 4,500 語であり, 最も多かったのが「読書好き」で 14,489 語である。

Nagy [1] は, 読書からの語彙学習が, 年間獲得語彙数のほぼ全てを占めると主張し, Carver [2]は読書からの語彙学習は児童が必要な語彙を獲得するのにほとんど役に立たないと主張している。本研究では, Nagy 試算も Carver 試算も, 現実に即しているとは言い難い試算であることを主張し, 先行研究に準じた条件設定を行った。その結果, 「標準」のモデルケースでは 3,371 語であり, これは日本人児童の語彙数を調査した, 国立国語研究所 [5]のデータの延べ数換算である 8,325 語からすると, 約 40%ほどの語彙数となる。この点では, 本研究のシミュレーション結果は Nagy [1] と Carver [2]の間であり, 「読書からの語彙学習が約 40%を占める」と解釈することができる。

一方, 個人差の表現として「読書嫌い」と「読書好き」というモデルケースも設定した。その結果, 「読書嫌い」では 329 語, 「読書好き」では 14,489 語となっており, 「標準」のそれぞれ 0.1 倍, 4.3 倍である。このように, モデルケースによって年間獲得語彙数が大きく異なった。

本稿では結果を省略するが, 表 1 の「標準」の設定のうち, 各変数の水準のみを変化させた分析をそれぞれの変数について行った。その結果, 読書時間, 開始時の未知語率, 基本読み速度の要因が, 年間獲得語彙数に大きな影響を持つことが明らかとなった。

表 1 シミュレーションのための条件設定

	Nagy	Carver	読書嫌い	標準	読書好き
年間読書日数 (日)	200	160	100	200	300
読書時間 (分)	25	15	10	30	60
開始時の未知語率 (%)	3%	0.9%	2%	2%	2%
未知語率下限 (%)	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%
未知語率減衰係数 (%)	0	0	0.00003%	0.00006%	0.00009%
未知語率回復間隔 (日)	1	1	100	100	100
基本読み速度 (wpm)	200	163	200	300	400
修正読み速度の適用	なし	なし	あり	あり	あり
修正読み速度係数 (wpm)			28	28	28
学習率関数の適用	なし	なし	あり	あり	あり
学習率	0.15	0.05			

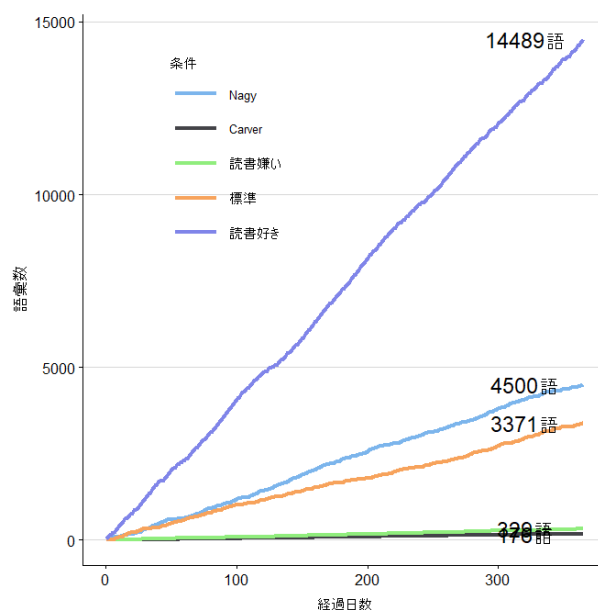


図 1 Nagy 試算, Carver 試算, モデルケースの語彙数

### 4. 本研究から得られる示唆

#### 学術的示唆

読書が語彙力に及ぼす影響についての多くの調査研究では, 読書日数や読書時間の測定を行っている。しかし本研究の結果からは, これらに加えて読むテキストの未知語率と基本読み速度が, 獲得される語彙数へ大きな影響を及ぼすことが明らかになった。「どのような難易度の本をどのような速度で読んでいるか」を把握することが, 読書が語彙に及ぼす影響を正確に知るために, 今後は必要となってくると考えられる。

また, 今回の結果では, 読書からの語彙学習によって年間獲得語彙数の約 40%がカバーされることを述べたが, では残りの 60%はどこから来るのかが不明である。今回の試算では, 獲得語彙の「出現頻度」は全く考慮していなかった。一般に, 語彙数測定に出題され

る語彙の出現頻度は低い語彙が多くなるため、今後は、出現頻度の低い語彙の獲得について、シミュレーションを行う必要がある。

### 教育的示唆

表1の「標準」の設定のうち、各変数の水準のみを変化させた分析では、開始時の未知語率が1%と2%のときに年間獲得語彙数が最大であり、0.5%と3%がそれに続いていた。4%以上となると、ほぼ語彙が獲得できなくなった。

語彙を育てる目的で読書を進める場合には、読む本の未知語率、すなわち、読み手に対するテキストの難易度を意識する必要がある。未知語率が4%を超えてしまうような難しすぎる本を読ませるよりは、未知語率が0.5%程度のやや易しい本を読ませる方が良い。ただし、未知語率が0%になってしまっただけでは語彙学習ができないので、あまりにも易しすぎる本は避け、「やや易しい」～「ちょうど良い」難易度の本を選ぶ意識が、効果的な語彙学習を成立させるのではないかと考えられる。

## 5. 文献

- [1] Nagy, W. E., (1997) "On the role of context in first- and second-language learning.", in *Vocabulary: Description, Acquisition and Pedagogy*, N. Schmitt and M. McCarthy, Editors. Cambridge University Press: Cambridge. pp. 64-83.
- [2] Carver, R.P., (1994) "Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction." *Journal of reading behavior*, Vol. 26, pp. 413-437.
- [3] 国立教育政策研究所, (2018) "平成30年度全国学力・学習状況調査報告書・調査結果資料." Available from: <http://www.nier.go.jp/18chousakekkahoukoku/index.html>.
- [4] 猪原敬介・上田紋佳・塩谷京子・小山内秀和, (2015) "複数の読書量推定指標と語彙力・文章理解力との関係：日本人小学校児童への横断的調査による検討." *教育心理学研究*, Vol. 63, pp. 254-266.
- [5] 国立国語研究所, (2009) "教育基本語彙の基本的研究 増補改訂版." 明治書院.
- [6] Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Hanae Koiso, H., Yamaguchi, M., Tanaka, M., Den, Y., (2014) "Balanced Corpus of Contemporary Written Japanese." *Language Resources and Evaluation*, Vol. 48, pp. 345-371.
- [7] 山崎誠, (2010) "語の平均使用度数に現れるテキストの特徴." 特定領域研究「日本語コーパス」平成21年度公開ワークショップ(研究成果報告会) 予稿集, pp. 5-14.
- [8] 小林潤平・川嶋稔夫, (2018) "日本語文章の読み速度の個人差をもたらす眼球運動." *映像情報メディア学会誌*, Vol. 72, pp. 154-159.
- [9] Swanborn, M. S. L. & de Glosper, K., (1999) "Incidental word learning while reading: a meta-analysis." *Review of*

- Educational Research, Vol. 69, pp. 261-285.  
 [10] Swanborn, M. S. L. & de Glosper, K., (2002) "Impact of reading purpose on incidental word learning from context." *Language Learning*, Vol. 52, pp. 95-117.