

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 5 日現在

機関番号：12612

研究種目：研究活動スタート支援

研究期間：2014～2015

課題番号：26880007

研究課題名(和文) 特定ジャンルの読書不足が語彙力・意味連想に与える影響：計算論アプローチによる検討

研究課題名(英文) The effect of insufficient reading amount on vocabulary and word association: A computational approach

研究代表者

猪原 敬介 (Inohara, Keisuke)

電気通信大学・情報理工学(系)研究科・研究員

研究者番号：10733967

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：本研究では、物語文、説明文、webテキストの3つのテキストタイプに焦点を絞り、特定のテキストタイプと語彙力・単語連想の関連について検討した。その結果、参加者調査では、先行研究と同様に、「物語文の読書のみが語彙力と強く関連する」という結果が得られた。一方、コーパス実験(シミュレーション)では、説明文の語彙成績が最も高くなった。この結果は、参加者調査で得られているテキストタイプによる読書効果の差異が、実はテキストそのものの違いではなく、テキスト外要因(例えば、読書動機づけなど)によるものであることを示唆している。

研究成果の概要(英文)：This study examined relationship between text-types of reading (Story-text, Explanation-text, and Web-text) and vocabulary, and word association. As a result, in participant-survey, we obtained the consistent results with previous studies, that is, only reading story-text had unique positive effect on vocabulary. On the other hand, in corpus experiment, explanation-text got the best score. These results suggested that "Story-text reading effectivity" found in the previous studies might not come from text itself, but factors outside of text, for example, reading motivation differences.

研究分野：教育認知心理学

キーワード：読書 語彙力 連想 コーパス 計算モデル

1. 研究開始当初の背景

(1) 読書不足への懸念 読書は、時間を選ばず個人ででき、図書貸出施設の活用により経済的負担を抑えることのできる、実践しやすい活動である。それに加え、読書習慣と様々な能力との密接な関連が示されており、特に語彙力との関連は頑健に認められる(メタ分析: Mol & Bus, 2011)。しかし近年では、いわゆる読書離れが起こり、その警句が叫ばれて久しい。

(2) 読書と語彙の関連 応募者はこれまで行ってきた文章理解についての研究の中で、「語彙」が文章理解に決定的な役割を果たすという結果を得てきた。そこで応募者は、学齢期以降の語彙獲得の中心である「読書」に焦点を当て、読書と語彙の関連についての本研究を開始した。

(3) 読書では言葉の「用法」が学ばれている我々は読書から様々なものを学ぶが、単語の意味を学習することについては、単語の使われ方、すなわち「用法」を学ぶことがその中心となっている(e.g., Elman, 1990)。すなわち、同じ読書でも、そこから学習できる単語の意味は、読むテキストの単語の用法に依存する。例えば、新聞では比較的「かたい」単語の使われ方が多く、小説では肩の力を抜いた「自由な」使われ方が多い。かたい用法は公的な文書作成、自由な用法は魅力的なスピーチや広告作成に必要となり、両方の用法を獲得することが実用的な語彙知識を作ることになる。このことを考えると、読書不足の問題は「特にどのようなジャンルの読書が不足しているのか」を把握すること無しには論じられないものである。

具体的には、現在「本を読んでいない」とされる人々には、「web上のブログだけならたくさん読んでいる」という人々がいる。逆に、「本を読んでいる」人の中にも、新書だけで小説は全く読まないという人もいる。これまでの研究は「読書」をひとくくりで捉えることが多かったが、ジャンルごとに読書習慣を分け、その上で読書不足の問題をより正確に捉えるアプローチが必要である。

(4) 本研究のアプローチ

こうした問題へのアプローチとして、これまでの多くの研究は「調査法」「実験法」といった心理学的手法を多用してきた。これらの手法は、検討対象である人間を直接調べる手法として重要である。一方で、人間を対象とするための制限もかかる。すなわち、純粋に「読書の影響のみ」を検討しようとしても、実際の人間は読書のみをして過ごすわけではないので、純粋な読書の影響を測定することは難しい。例えば、ある本を読書することで、書かれていた内容について興味湧き、同じテーマの映画を見たり、知り合いとその

ことについて話したりする。読書そのものではないこうした行為が言語力に影響してしまうかもしれない。

そこで本研究では純粋な読書の影響を検討するために、言語コーパスに基づく計算機シミュレーションを導入する。上記の心理学的手法と、計算機シミュレーションという認知科学的手法を併用することで、これまでよりも読書の影響をクリアに検討することができると思われる。

2. 研究の目的

本研究は、現在までの「読書不足が語彙へ及ぼす悪影響」についての素朴な考えを精緻化することを目的に、ひとくくりであった「読書」を物語文、説明文、webテキストの3つのテキストタイプに分け、(1)読書が語彙力に及ぼす影響がテキストタイプによって実際に異なっているか、(2)計算機シミュレーションによってその違いを再現できるかを検討する。

3. 研究の方法

以上の目的に対して、(1)大規模web調査(参加者調査)、(2)コーパスに基づく計算機シミュレーション(コーパス実験)によって検討した。

(1) 大規模 web 調査 (参加者調査)

Web調査会社(クロス・マーケティング社)に依頼し、Web上での調査を実施した。参加者は755名であり、内訳は、専門学校に通う男性32名、女性47名、短期大学に通う男性10名、女性25名、4年制大学に通う男性331名、女性310名であった。

40問(対義語テスト20問、類義語テスト20問)で構成される語彙テストを作成した。各問題は、ある単語(問題語)の対義語もしくは類義語を4つの選択肢から選ぶ形式となっていた。

また、3つのテキストタイプ間で単語の用いられ方の違いの大きい40語を計算し、単語連想課題で用いる刺激語とした。

物語文、説明文、ブログ、web上のQ&Aサイトを読む頻度について、1(「まったくか、ほとんどない」)~5(「週(日)に数回以上」)の5件法で尋ねた。物語文はフィクション本の、説明文はノンフィクション本の、ブログとweb上のQ&Aサイトの頻度の平均値はWebテキストの読書頻度として採用した。

(2) コーパスに基づく計算機シミュレーション (コーパス実験)

BCCWJ(現代日本語書き言葉均衡コーパス、国立国語研究所、2012)を加工し、形態素数のほぼ等しい3種類のコーパスを作成した。

BCCWJの図書館流通コーパスのうち、NDC(日本十進分類法)が9番(文学)のものを物語文コーパス(7,325,446形態素,56,150異なり語,800,965文),NDCが9番と「分類なし」以外のものを説明文コーパス(7,365,504形態素,54,614異なり語,512,119文),BCCWJのうち「Yahoo!!知恵袋コーパス」と「Yahoo!!ブログコーパス」を合わせたものをWebテキストコーパス(7,356,343形態素,47,158異なり語,902,119文)とした。

コーパス内の単語出現情報を元に単語間の関連度(LSA類似度)を計算する手法の一つであるLSA(Landauer & Dumais, 1997)を用いて、参加者実験と同じ語彙テストに回答する計算機シミュレーションを行った。具体的には、単語×文の出現頻度行列に対してTF-IDF法にて重みづけを行い、特異値分解を行ったのち、異なり語数×圧縮次元数の大きさの行列として復元した。

語彙テストに回答するシミュレーションの手続きは、Landauer & Dumais (1997)の方法を採用した。すなわち、問題語と選択肢の間の4対についてLSA類似度を計算し、最も高いLSA類似度を持つ選択肢を選ぶこととした。行列の次元数は、10~1000まで10ずつ次元を操作して、最も成績の良い次元数をそのコーパスの成績とした。

4. 研究成果

上記3で述べた方法に基づいて研究を行い、以下の研究成果を得た。

(1) 大規模 web 調査 (参加者調査)

参加者を、物語優位群(n=176;他の2つのテキストタイプよりも物語文の読書頻度が高い者)、説明優位群(n=74;他よりも説明文の読書頻度が高い者)、Webテキスト優位群(n=322;他よりもWebテキストの読書頻度が高い者)、その他群(n=183;以上のどの群にも当てはまらない者)に分けた。

群ごとに語彙テストの平均正答率(SD)を算出したところ、それぞれ79.5%(15.3%)、75.0%(17.1%)、70.7%(18.1%)、75.8%(16.8%)であった。

参加者と材料を同時にランダム要因とした一般化線形混合モデル(従属変数の分布には二項分布を仮定)により、物語優位群と他の群との平均値についての検定を行った結果、他のすべての群よりも物語優位群の語彙テストの正答率が有意に高かった(それぞれ、 $z_s = -2.146$ ($p < .05$), -5.979 ($p < .01$), -2.267 ($p < .05$))。

(2) コーパスに基づく計算機シミュレーション(コーパス実験)

語彙テストの平均正答率は、物語文コーパス(110次元)では67.5%、説明文コーパス

(150次元)では82.5%、Webテキストコーパス(150次元)では62.5%であり、説明文コーパスが最も高かった。

(3) まとめと今後の展望

参加者調査においては、物語優位群の語彙テスト正答率成績が他の群よりも高くなっており、先行研究(McGeown, Duncan, Griffiths, & Stothard, 2015; OECD, 2010; Spear-Swerling, Brucker, & Alfano, 2010)と一致する結果が得られた。これに対してコーパス実験では、説明文コーパスに基づくシミュレーション結果における語彙テスト正答率が最も高くなっており、参加者実験の結果と食い違った。

以上の結果から、先行研究及び本研究における参加者実験の結果に見られる物語文読書優位の現象については、読書そのもの(テキスト内要因)では説明されないことが示唆された。

ただし、以上の結論は暫定的なものである。コーパス実験で用いたシミュレーションが、テキストタイプ間のテキスト内要因の差異を十分に捉えきれていないために、参加者調査とコーパス実験の結果が一致しなかった可能性がある。また、参加者調査の結果は1時点のデータであるため、因果関係を反映したものではない可能性があり、このことも結果の食い違いを産む要因となり得る。参加者調査を縦断調査や介入実験に変更すること、シミュレーションの設定を網羅的に変更することが、今後の課題である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計1件)

[1] 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*, **63**, pp.254-266. <http://doi.org/10.5926/jjep.63.254> 査読有

[学会発表](計3件)

[1] 猪原敬介・上田紋佳・塩谷京子・小山内秀和(2015). 読書頻度が読字スキルに及ぼす発達の影響:小学校低学年児童への縦断調査による発達の検討(1). 日本教育心理学会第57回総会, 朱鷺メッセ(新潟コンベンションセンター), 2015年8月26日.

[2] 上田紋佳・猪原敬介・塩谷京子・小山内秀和(2015). 読んだ本のジャンルが言語力

に及ぼす影響 小学校低学年児童への縦断調査による発達の検討(2) . 日本教育心理学会第 57 回総会, 朱鷺メッセ(新潟コンベンションセンター), 2015 年 8 月 26 日.

[3] 猪原敬介・上田紋佳・塩谷京子・小山内秀和 (2015). 読字スキルの発達に及ぼす読書の影響: 小学校 1 年生から 3 年生にかけての縦断調査による検討. 第 59 回日本読書学会大会, 全林野会館(プラザ・フォレスト), 2015 年 8 月 2 日.

〔図書〕(計 1 件)

[1] 猪原敬介 (2016). 読書と言語能力 言葉の「用法」がもたらす学習効果. 京都大学学術出版会. 300 頁.

〔その他〕

ホームページ URL :

<https://sites.google.com/site/keisukeinoharaweb/>

6 . 研究組織

(1)研究代表者

猪原 敬介 (INOHARA, Keisuke)

電気通信大学・情報理工学(系)研究科・

研究員

研究者番号 : 10733967