

レポート作成例

応用心理学 I レポート

文学テキストの量的分析：夏目漱石『こころ』を題材に

2021.06.27

学生番号 XXXXXXXX

佐藤浩輔

序論

近年、テキストを量的に分析する研究がさかんになっている。

通信技術の発展により、Web ページや SNS の投稿など大量のデータが日々生成されており、……（略）……また、技術的な側面としては、コンピュータの性能の上昇により、パワーのあるマシンを安価に利用できるようになったこと、R や MeCab など無料で公開されているソフトウェアを用いて、個人が手軽に分析環境を構築できるようになったことなどが挙げられる。

分析を始めるハードルは低くなった一方で、テキストマイニングで得られた結果が妥当なものなのか評価することは難しい。

テキストを量的に分析する方法はテキストマイニングや計量テキスト分析などと呼ばれる。前者は主に工学やマーケティング的な領域で使われ始めた用語である一方、後者は人文社会科学の内容分析の流れを汲んで使われている用語である(樋口, 2014)。

個人がテキストの量的分析を行うハードルが下がる中で、課題もある。特に入門者においては、分析から得られた結果がどの程度妥当なのかを評価しづらいという問題がある。例えば……(略)。

分析者が結果を容易に評価しやすいよう、よく知られた題材を扱うことで、テキストの量的解析のよいデモンストレーションとなると考えられる。

そこで本研究では、分析対象として夏目漱石『こころ』を題材とした。

『こころ』は、1914 に夏目漱石が朝日新聞で連載していた小説で、「先生と私」「両親と私」「先生と遺書」の三部構成となっている。新潮文庫版は発行部数 718 万部を超え、2016 年の時点で新潮文庫で最も売れている小説である(日本経済新聞, 2016)。高校の国語の教科書でも取り上げられ、……（略）。

計量的な手法で内容を可視化できるかどうかを検討するため、分析を行った。

方法

分析対象

コメントの追加 [佐藤1]: 何を目的に、なぜその研究(実験、分析、etc.)をする必要があるのか、何を明らかにするのかについて、背景とともに述べる。

コメントの追加 [佐藤2]: 分析する対象について、概要やデータの取得方法などを説明する。通常の心理学実験の場合は「実験参加者」に相当する。

レポート作成例

夏目漱石『こころ』本文を対象として分析した。テキストは青空文庫(<https://www.aozora.gr.jp/>)にて公開されているもの(新字新仮名、底本は集英社文庫版)を用いた。データの取得には石田(2017)の Aozora 関数を用い、元の HTML ファイルにあったルビ等のタグは削除した。

分析にはすべて R(4.0.2)を用いた。日本語の形態素解析には工藤ら(2004)の MeCab(ver. 0.996)を用いた。R からの操作には RMeCab パッケージ(石田, 2017)を用いた。

データ処理

生データ テキストは一文を一行のレコードとし、部・節の番号、並びに登場した順に段落・文の番号を 1 から連番の ID として付した。この ID は、第二部第一節第三段落第四文であれば 2,1,3,4 となる。このように一文を単位にしたデータをタブ区切りのファイルとし、分析に用いた。

クリーニング 誤字や脱字が含まれていないかどうかをチェックした。具体的には形態素解析の結果分割した内容を集計した頻度表を作成し、登場人物などの固有名詞が認識されているか、また、誤認識の結果が集計されていないかを確認した。確認作業の結果、うまく認識されていなかった単語については形態素解析ソフトの辞書に追加することで対応した。漢字の送り仮名などの表記ゆれについては同じ単語として集計されるよう分析ソフト上で統一した。

形態素解析 各文について形態素解析を行い、文に含まれる単語とその品詞を抜き出した。

集計 単語の頻度・バイグラム・共起について以下の方法で集計した。

頻度：出現した単語をそのまま集計した。

バイグラム：文から抽出した対象の品詞の連続をバイグラムとして数えた。例えば文「吾輩は猫である」から、名詞のみを抽出した場合、「吾輩・猫」がバイグラムとなる。

共起：文から単語を bag of words として抽出し、そのすべての単語の組み合わせを共起として数えた。ただし、同じ単語が 1 文に 2 回以上出現しても 1 単語として扱った。

なお、頻度の高い語のうち、「もの」「こと」などのように、内容の理解に寄与しない単語群をストップワードとして集計から除外した。除外したストップワードについて表 1 にまとめた。また、登場人物の「K」についてはそのままでは記号として扱われてしまうことから、人名として扱うよう辞書に登録した。

表 1 ストップワードとして除外した語

事, の, よう, それ, もの, 人, 何, 一, ん, 方, 二, 前, 氣, 中, 上, 今, ため, 時, そこ, どこ, これ, そう

コメントの追加 [佐藤3]: 使用したソフトウェアとそのバージョンを記述する。

コメントの追加 [佐藤4]: 分析の手続きについて述べる。通常の心理学実験では「実験手続き」に相当する。

コメントの追加 [佐藤5]: データをクリーニングしたかどうかについて記述する。

コメントの追加 [佐藤6]: ストップワードや辞書について言及する。

レポート作成例

結果

『こころ』本文に出現する名詞のうち上位 30 位の頻出語を示した（図 1）。「私」「先生」「K」「奥さん」など、登場人物に関する語が上位を占めており、……（略）

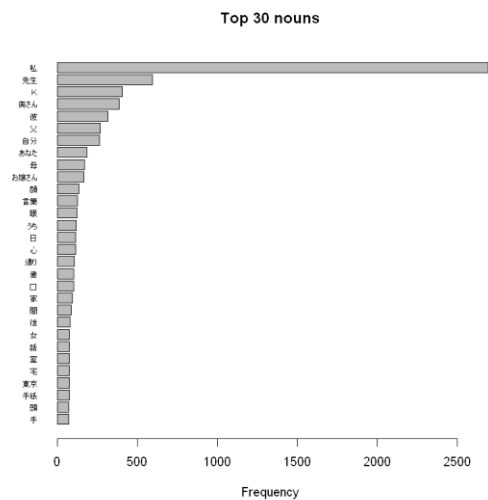


図 1 出現頻度上位 30 位の名詞

（略）

考察

本研究では、○○するために、夏目漱石『こころ』を題材に分析を行った。

その結果、……（略）このことから……（略）。

以上より、デモンストレーションするという目的は果たされたと考えられる。

本研究の限界として、……（略）

最後に、……（略）

コメントの追加 [佐藤7]: 結果を記述する。

図表や数値の示すところを解説するセクション。心理学の場合、結果とその解釈を分離させる傾向があるのでここではデータから事実としていえることを書く。結果の解釈については考察で議論する。

コメントの追加 [佐藤8]: 結果のまとめに入る前に研究の目的に言及があるとよい。

コメントの追加 [佐藤9]: 簡潔に結果をまとめる。

コメントの追加 [佐藤10]: 結果から考察できることを書く。

コメントの追加 [佐藤11]: 序論で言及した目的に対して結果がどういう意味を持つかを書き、結論付ける。

コメントの追加 [佐藤12]: 研究デザインやデータ、分析、結果あるいは考察に関して、研究で得られた知見を制約するような（潜在的な）問題点があれば言及しておく。

その問題点が解決可能なら当然解決しておいてしかるべきなので、解決不能な問題点のみ記載する。

レポート作成例

参考文献

石田 基広 (2017). R によるテキストマイニング入門 第2版 森北出版

日本経済新聞 (2016). 漱石没後 100 年、人気衰えず 書店で文庫フェア :日本経済新聞.
Retrieved from https://www.nikkei.com/article/DGXLASDG08H0C_U6A211C1CR0000/
(July 9, 2019)

樋口 耕一 (2014). 社会調査のための計量テキスト分析 内容分析の継承と発展を目指して
ナカニシヤ出版

コメントの追加 [佐藤13]: 文中で引用した文献を挙げる。

心理学のレポートの場合、特に指定がなければ日本心理学会『執筆・投稿の手びき』を参考にする。