

女兒向けコミック雑誌テキストの潜在意味分析

久保千尋[†] 岡安千明[†] 河合麻里奈[†] 楠えりか[†] 小澤美穂[†] 小林碧[†] 清水彩葉[†] 千葉理央子[†] 高田明典[†]
フェリス女学院大学文学部[†]

1. はじめに

おおむね 9 歳から 14 歳程度の女子児童・女子生徒における価値観の形成に関して、少女向けコミック・少女漫画は少なからぬ役割を果たしていると考えられる。筆者らはこれまで、物語構造分析の手法を用いてアニメ・小説・CM・コミックなどの訴求構造分析を行ってきたが、これまでの分析手法のもとでは、多数の物語に共有されている価値観を包括的にとらえることは、不可能とは言えないまでも容易ではない。また、大量のテキストから物語構造を抽出するという作業は、テキストが数十万文字となるような場合には現実的ではない。本研究においては、テキスト分析の手法の一つである潜在意味分析を用いることによって、包括的な「意味構造」の抽出を通して、そこに表現されている価値観の推定を試みた。分析対象としては、現在発行されている主要な女兒向けコミック雑誌とし、その文字表現にのみ着目した分析を行った。

2. 分析方法

分析にあたっては、主として潜在意味分析を用いる。潜在意味分析とは、対象とする文の単位（文オブジェクト）にどのような単語が含まれているかを示す「文オブジェクト-単語共起行列」に対して特異値分解を施し、次元縮約を行うことによってテキストの主たる意味構造を把握する主成分分析的手法である。この手法においては、物語の流れを捨象して単語の構成のみを分析対象にすることによって、いくつかの利点が生じる。その一つには、今回対象としたコミック雑誌のように、（連載途中のため）物語として完結していないものでも分析対象とすることができる点があげられる。また、物語の話素の抽出に際しては、比較的よく訓練を受けた分析者が必要であるが、潜在意味分析ではそのような作業が必要なく、単にテキストを入力し、必要な前処理を施すという作業のみで分析にかけることができるという簡便性も特長の一つとしてあげることができる。また、分析手法に際しての客観性をあまりに求めることは、必ずしも当該分野の研究の質の向上にはつながらないものではないと考えるが、物語構造分析において指摘されがちな分析手法の客観性に関しても、ある程度担保されると考えられる。

3. 分析対象

2009 年 11 月から 12 月にかけて発行された女子児童・

An Analysis of Comic Books for Girls with LSA (Latent Semantic Analysis)

[†] KUBO Chihiro, OKAYAU Chiaki, KAWAI Marina, KUSUNOKI Erika, KOZAWA Miho, KOBAYASHI Midori, SIMIZU Ayana, CHIBA Rioko, TAKADA Akinori.

[†] Faculty of Letters, Ferris University.

生徒向けコミック雑誌を選択した。選択にあたっては発行部数をもとに、下記の 7 誌を対象とした。

ちゃお 2009 年 12 月号
Betsucomi (ベツコミ) 2009 年 12 月号
別冊花とゆめ 2009 年 12 月号
マーガレット 2009 年 11/20 号
別冊 マーガレット 2009 年 12 月号
なかよし 2009 年 12 月号
りぼん 2009 年 12 月号

入力した作品数は計 90 タイトル、合計 195461 文字 / 13484 行であった。また、形態素解析の結果、119463 単語（記号を含む）であった。

4. 分析の手順

4. 1. 文オブジェクトの抽出

分析対象とするテキストは、コミックの本編中の文字列のみとし、また、原則としてセリフのみを分析対象とした。つまり、状況の説明やト書きに該当する部分はすべて捨象し、原則として「吹出し」の中に記述されているもののみを対象とした。ただし、吹出しの中に記載されておらず、図中にあるものであっても、明らかに登場人物の内言や内心の表現であると思われるものや、セリフであることが明らかであるものに関しては、分析対象とした。一つの作品を一つの文オブジェクトとし、合計 90 の文オブジェクトが構成された。

4. 2. 文オブジェクト-単語の共起行列の作成

4. 1. で示した方法によって抽出された文オブジェクトのデータ列から単語を抽出し、文オブジェクト-単語の共起行列を作成した。

形態素解析は MeCab を用いて行い、文オブジェクト-単語の共起行列の作成は RMeCab を用いて行った。抽出する単語は、名詞・形容詞・動詞とし、形容動詞・助詞・助動詞・数詞・接続詞などは分析対象からは除外した。共起行列の作成にあたっては、2 つ以上のタイトルに共起している単語のみを対象とし、重み付けは行わず、一つのタイトル中に同じ単語が何度出現しても得点を 1 とした。ただし、分析の内容に応じて、IDF などの大域的な重み付けや、文書の長さによる正規化重み付けを使用しているが、本稿で後述する分析においては重み付けを行っていない。

4. 3. 特異値分解

4. 2. で作成された文オブジェクト-単語の共起行列を特異値分解し、文オブジェクトの主成分得点および単語の主成分得点を得た。

5. 結果および考察

ここでは、名詞のみを分析対象として行った潜在意味分析の結果のみを例示した。4.2 で述べたように重み付けは行わなかった。また、また、90 タイトルのうち、60% (54 タイトル) を越えるものに出現する「それ」「ここ」「何」「人」など 18 個の単語は共起行列から除外した。

第 1 軸において主成分得点の絶対値が大きかった単語は、「好き」「みんな」「たち」「一緒」「中」(いずれも負値) などであり、第 2 軸では「わけ」「教室」「ウソ」「ムリ」「気持ち」「クラス」「告白」(いずれも正値)、「君」「街」「世界」「危険」「問題」「言葉」(いずれも負値) などが見られた。

これらの結果から、分析対象としたコミックの意味構造として「孤独—凝集」「私的世界—公的世界」を抽出した。

これら二つの軸によってコミック (文書オブジェクト) を配置した例を図 1 に示す。図中、H_L、B_Lなどの記号列は以下の意味を示す。

- C_L : ちゃお 2009 年 12 月号
- B_L : Betsucomi (ベツコミ) 2009 年 12 月号
- H_L : 別冊花とゆめ 2009 年 12 月号

- M_L : マーガレット 2009 年 11/20 号
- BM_L : 別冊 マーガレット 2009 年 12 月号
- N_L : なかよし 2009 年 12 月号
- R_L : りぼん 2009 年 12 月号

また、タイトル名の記載は冒頭 2 文字に省略されている。

文献

- [1] 石田基広 : R によるテキストマイニング入門, 森北出版, 2008.
- [2] 和多太樹, 関隆宏, 田中省作, 廣川佐千男 : 単語の出現頻度に着目した病院評判情報の分析, 情報処理学会研究報告. SLP, 音声言語情報処理, 2005 (50), pp. 15-20.
- [3] 豊田秀樹 : データマイニング入門, 東京図書, 2008.
- [4] 重久礼美, 高田明典 : 映像作品の物語構造分析の自動化に関する一研究, 多文化・共生コミュニケーション論叢, (2), pp. 11-22, 2007.

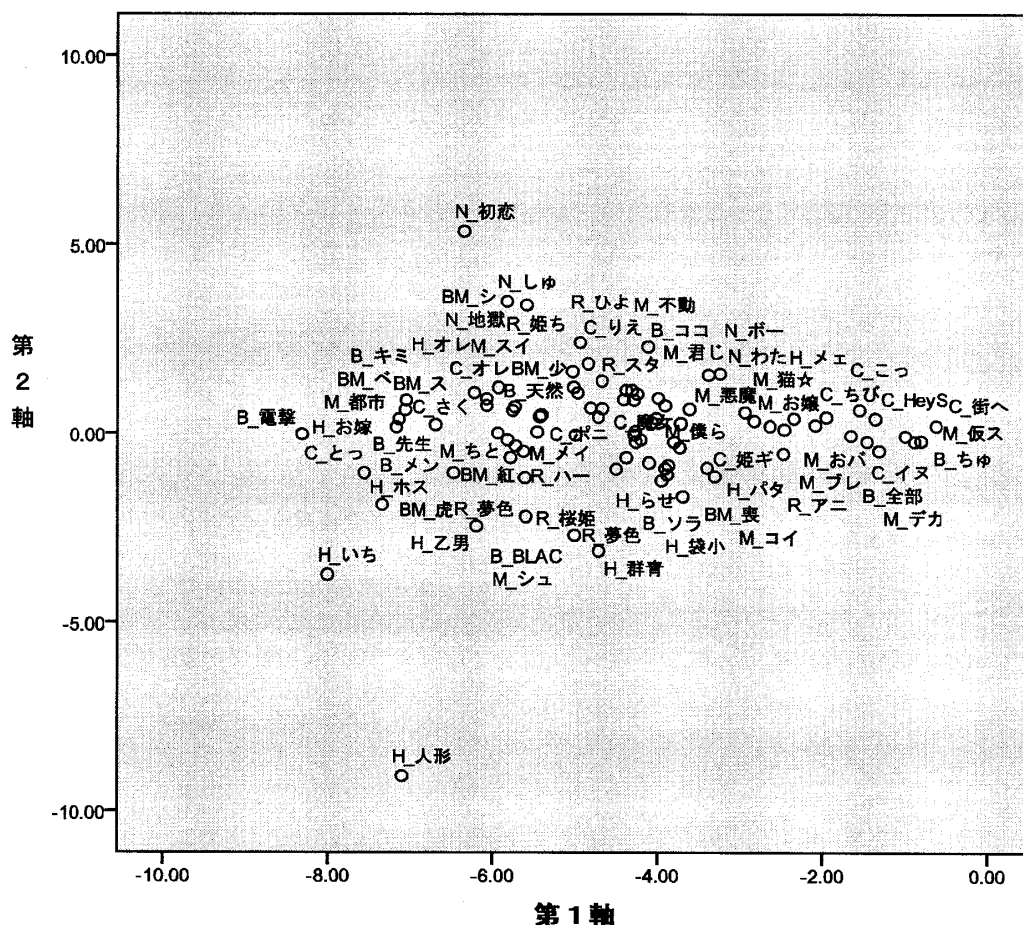


図 1 少女コミック雑誌 7 誌 90 タイトルの潜在意味分析 : タイトルの関係