

漫画の台詞やオノマトペにおける言語的特徴分析

内山 清子

A Study for Lexical Feature Analysis of Script and Onomatopoeia in Comics

Kiyoko Uchiyama

Abstract:

Japanese Comics, which are one of the Cool Japan promotion program, aiming to support Japanese contents business, have been attracted attention from abroad. Especially, most international students motivated to study Japanese because of an interest in pop-culture, the comics and anime. Based on this background, the purpose of this paper is focused on analyzing statistically scripts and onomatopoeia which are constituents of comics by means of natural language processing. We discussed lexical variety and balance between scripts and onomatopoeia for four categories in comics, information and knowledge obtained by those analysis and future works.

KEY WORDS : Onomatopoeia, Comics, Lexical Feature Analysis

要旨:

漫画は日本の成長戦略の一つの柱であるクールジャパンのコンテンツ産業として海外からも注目されている。特に、大多数の留学生にとって日本のアニメや漫画が日本に興味を持つきっかけとしていることから、漫画やアニメを日本語教育や文化理解に利用する動きがある。本稿では、漫画の構成要素である台詞やオノマトペに着目して、言語処理の手法を用いて言語的特徴を分析することを目的とした。4つのジャンルの作品を対象として語彙多様性や台詞、オノマトペの割合、作品における高頻度語について分析を行い、漫画の台詞・オノマトペから得られる情報や今後の方向性について考察した。

キーワード : オノマトペ、擬音語、擬態語、漫画、特徴分析

1. はじめに

漫画は、外務省のポップカルチャー外交や経済産業省のクールジャパン/クリエイティブ産業政策、文化庁によるメディア芸術祭などの政策や事業の中核として海外展開を含めたグローバルな観点からも注目されている。さらに、アニメ・ゲーム・ドラマ・映画化や、キャラクター商品の販売など、複数の異業種の分野で展開するビジネスモデルとしてのメディアミックスは、産業界にも多大な影響を及ぼしている。このように漫画は日本を代表するエンターテインメントとしての地位を確立している。

一方、学術分野においても、漫画の絵や作品及び教育用の教材としての分析・評価や、台詞などのテキスト情報に着目し、漫画の構成要素である台詞や効果音などの擬音語、擬態語（オノマトペ）に関する言語的な研究が行われている。但し、日本語学の見地からの分析が中心で言語処理技術を使った分析手法による研究はあまり行われていない。そこで本

稿では漫画のデジタルテキストを言語資源と捉え、言語処理の手法を用いて統計的に処理を行うことにより、漫画の言語的特徴について分析・考察を行う。

2. 関連研究

漫画に関連した先行研究はいくつかの種類に分類できるが、本稿では漫画に含まれる絵ではなく「言葉」に着目して実施されている研究について調べた。なお、本稿の内容に関連したオノマトペの研究についても簡単に触れているが、オノマトペの研究自体が多分野・多岐に渡っている事から、ごく一部のものをここでは取り上げている。

本稿では漫画の台詞つまり言葉を中心として分析を進めていくが、絵と言葉を組み合わせるパターンとして以下の7つの型があると言われている⁽¹⁾。まず、

(1)「言語中心型」は、文字通り言葉だけで完結し

*湘南工科大学 工学部 コンピュータ応用学科 准教授

絵はその補助的な役割を果たす、(2)「絵中心型」は絵で説明された場面に言葉は音声効果だけを加える、

(3)「両者中心型」は言語と絵が半々ずつ意味を伝える、(4)「増幅型」文字が絵のイメージを増幅させたり、その逆の効果を果たす、(5)「並列型」は絵と言葉が別々のストーリーを語っていき交差しない、

(6)「モニター型」は文字が絵の一部に取り込まれている、最後は(7)「相互依存型」で、言葉と絵が互いに協力し合い、どちらか一方では成立しないようなイメージを伝えているなどの型が存在すると分析されている。

これらの型は、同じ漫画の中であっても場面によって表現の仕方が異なっていることもあるため、一作品や著者によって上記の7つの型に分類することは難しい。本稿では主に「言語中心型」「相互依存型」を扱っている。

漫画の言葉を対象として研究を進めている分野として、日本語教育、日本語学、自然言語処理の3つに分類される。まず、日本語教育の分野では、留学生が日本に興味を持つようになったきっかけとして約90%の学生がサブカルチャーと報告されており、その内の75.5%が漫画やアニメであるという日本国内における調査結果がある^②。また、2009年度の国際交流基金の調査では、国内外の日本語教育機関における調査において、日本語学習の目的のうち、マンガ・アニメ等に関する知識を得るためとした割合が50.6%であったと報告している^③。

このように漫画やアニメを通して日本に興味を持つ留学生が多いことから日本語教育の教材として利用しようとする動きやニーズが高まっている。最近の動向として、関西国際センターでは海外の学習者を支援する教材開発の一環として日本語学習者のためのWebサイト「アニメ・マンガの日本語」*の開発を行ってきた^④。このサイトでは、アニメ・マンガに登場するキャラクターの特徴的な表現(フレーズ、文型・表現、呼称、発音変化)、4つのジャンル(恋愛・学校、忍者、侍)の特徴的な表現(用語、漢字、場面別フレーズ、オノマトペ)をクイズやゲームを通して学ぶ事ができるように設計されている。Webサイトの学習用コンテンツやツールの提供は、時間や場所の制約がないため、特に海外にいる日本語学習者を支援する上で重要な意義を持つ。

次に日本語学においては、漫画やアニメで使われる日本語の特徴について研究が行われてきた。伊藤(2003)は漫画とアニメの人気に伴って専門雑誌、概説書、Webサイトなどに作成された英語圏におけ

る用語集の構造を明らかにし、用語集に収録された用語や表現の特徴分析を行っている^⑤。見出し語となっている用語のほとんどが名詞であるが、最も収録回数が多いのは形容詞の「可愛い」であったと報告されている。この他に、日本語教育の教材として開発したWebサイト「アニメ・マンガの日本語」に含まれる表現のうち「ジャンル表現」の中の用語に着目して、その特徴について定量的・質的に分析をした研究がある^⑥。この研究ではジャンル用語について、日本語能力試験出題基準レベルとの対応関係、品詞の出現頻度をジャンル別に比較し、ジャンル毎に形式面、内容面について分析を行い日本語教育への応用について提案している。

また、「言葉」だけでなく、オノマトペ(擬音語、擬態語などの総称)について分析を行っている研究もある。夏目(1999)では、マンガにおける多種多様なオノマトペは独自の進化をとげマンガの中でしか使われないような表現が造られているとしている^⑦。これはマンガにおけるオノマトペが音声言語としての側面と、描かれた文字(絵としての側面をあわせもつ)、つまり音声記号と画像記号というマンガ特有の単位として扱う必要があるとして「音喩(おんゆ)」という造語を設定している。なお、本稿では、マンガには一般的な文章で用いられるオノマトペも含まれており、オノマトペと音喩を区別することは難しいという立場から音喩という用語は使用せず、オノマトペという表記で統一する。

最後に言語処理の分野では、徳久(2007)では、信頼性の高い情緒タグ付きテキスト対話コーパスを作成することを目的として、マンガの対話文を対象テキストとして、登場人物の表情を参照する方法によって情緒タグを付与した研究を行っている^⑧。マンガ「ちびまる子ちゃん」10冊の対話分を対象として「表情タグ」「情緒タグ」を暫定的に付与しながらタグの設計をし、最終的に採用した情緒タグに対する妥当性を計測し、タグの安定性を評価した。この研究は、従来のテキストコーパスでは新聞などの書き言葉が中心となっているものに対するアノテーション(注釈)を付与するものに対して、対話文コーパスとしてマンガを利用したアノテーション設計を目指したものである。

オノマトペの研究は自然言語処理の中で近年活発に行われてきているが、ブログ記事からオノマトペの用例を自動的に抽出する方法を提案した研究がある^⑨。これまでのオノマトペの用例は教科書や小説などから手作業で見つけ出すことが中心であったが、オノマトペの接続パターンをルール化し、ブログ記事から自動的に抽出する手法を提案し、適合率96.2%

* <http://anime-manga.jp>

の精度を得ている。この研究は日本語学習者がオノマトベを理解するためにより多くの用例に触れる事が重要であるとする動機から自動抽出の手法の提案をしている。

以上のように、言語（日本語）としてのマンガの研究が様々な分野で盛んに行われてきたが、マンガを電子化して言語処理の手法を用いて扱っているのは、調べた中では徳久(2007)の例などわずかしかなかった。そこで本稿では、マンガのテキストを電子的に扱うことを前提とした問題に取り組み、マンガの言語分析に基づいて、今後の方向性について考察する。

3. 漫画の電子化とアノテーション

電子化はあくまでも研究利用（個人利用）を目的とする前提で、ドキュメントスキャナを用いてスキャンする方法が考えられる。しかし、今回の分析では画像ではなくテキスト中心であるため、吹き出しの中にあるテキストの文字認識や、背景と一体化した効果音としてのオノマトベはスキャンでは抽出できないと判断し、全て手作業で入力し、電子化した。

分析対象の漫画として、複数のジャンルの漫画を選定した。ジャンルとしては、バトル系、人間関係系（恋愛系）、ギャグ系、スポーツ系の4種類とした。コンピュータ応用学科の Team Project Learning (TPL)の授業の一環として、ジャンル毎にグループを作成し、電子化作業を行った。ジャンル内の作品は学生が決定した。本稿で扱うデータはその電子化作業を行ったデータから、各ジャンルを比較するために、文字数ができるだけ同程度になり、かつページの切れ目になっているところまでのテキストを対象とした。

次に、電子化するにあたってどのようなアノテーション（注釈）が必要であるかを議論した。先行研究の徳久(2007)では、絵の情報も含めた表情タグや情緒タグを取り入れているが、本稿ではできるだけ絵の情報とは独立した台詞という立場からアノテーションを考える。

新しい分野のテキストにおけるアノテーション設計は、そのテキストをどう活用するかという目的に依存するところが大きい。漫画を対象とした分析として、本稿で扱うジャンル別の語彙特徴の比較、オノマトベの分析の他、将来的にキャラクターと台詞の関係、談話構造分析などが考えられるため、その利用方法に必要な不可欠な情報の注釈を暫定的に入れることに決めた。

アノテーションの素性として、(1)台詞、(2)説明文、(3)オノマトベ、(4)ページ数、(5)補足情報として場面

説明文の5種類という枠組みを設定した。しかし、入力作業を通じていくつかの疑問点が提示された。漫画の吹き出しの中の台詞には、登場人物が実際に口に出して発言したものだけでなく、発言せずに考えている内容、回想シーン、オノマトベなど、多様な情報が含まれている事がわかった。これらについては、補足情報の中に詳細に記述することとし、台詞という枠組みに入れて扱うことにした。今後利用方法によっては、異なる枠組みが必要となる場合が出てきた時に再度アノテーションの設計を検討する。

4. 台詞の形態素解析

電子化した台詞とオノマトベを形態素解析によって、品詞に分割する処理を行った。形態素解析には Mecab*を利用した。形態素解析を実施した結果、いくつかのエラーが見られた。Mecabの解析精度は約98%と公表されており、Mecabは辞書やコーパスに依存しない汎用な設計で学習モデルとして条件付き確率場(CRF)を用いているため、これまでの形態素解析器よりも新しいジャンルのテキストに適応可能であるとしている。しかし、今回の漫画の解析ではエラーが通常より目立っていた、それは、Mecabでは新聞記事などの書き言葉のコーパスを中心にコストの推定を行っているため、話し言葉の文法には弱いとされていることが原因と考えられる。図1に対象テキストである「銀魂」の台詞における解析失敗の例を示す。

例文(1)に示すような解析失敗例があり、例文(2)では、例文(1)の失敗例を解析しやすい文に書き換えた例である。「疑いやがる」の助動詞「やがる」については正しい表記であっても失敗している。「しまいにゃ」「おめー」については、かなりくだけた口語調の表現であるため、解析が難しかったと考えられる。

一方、話し言葉でも(2)のような形式であれば解析可能である。このように、漫画の場合は実際に発した言葉の勢いやニュアンスを文字に反映させて表現することが多いため、そのテキスト情報をそのまま入力し、形態素解析で処理するとこのようなエラーが起きることがわかる。

漫画の台詞を扱うためには、入力段階で、台詞の表現を機械処理用に意味を変えずに書き換えるか、あるいは Mecab 自体を漫画のテキスト用に学習し直すかのどちらかの対処方法が考えられる。実際問題

* Mecab: Yet Another Part-of-Speech and Morphological Analyzer

<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html?source=navbar>

として、形態素解析の「くせ」を理解していないと文章を変換することは難しい上、変換してしまうと台詞のニュアンスが伝わりにくくなってしまいます。また、Mecab を学習させるためには適応させる漫画の文章を大量に用意する必要があるなど、漫画の台詞の形態素解析には様々な問題が残っているため、今後の課題とする。

例文(1)	しまいにゃ、おめーまで疑いやがって
し	動詞,自立,*,*,サ変・スル,未然形
まい	助動詞,*,*,*,不変化型,基本形
にゃ	助詞,特殊,*,*,*,*
,	記号,読点,*,*,*,*,*
お	接頭詞,名詞接続,*,*,*,*,*
め	名詞,一般,*,*,*,*,*
一	名詞,一般,*,*,*,*,*
まで	助詞,副助詞,*,*,*,*,*
疑い	名詞,一般,*,*,*,*,*
や	助詞,並立助詞,*,*,*,*,*
がっ	動詞,接尾,*,*,*,*,*,*,*,*,*,*,*
て	助詞,接続助詞,*,*,*,*,*,*
例文(2)	しまいには、お前まで疑うなんて
しまいに	副詞,助詞類接続,*,*,*,*,*,*
は	助詞,係助詞,*,*,*,*,*,*
,	記号,読点,*,*,*,*,*,*
お前	名詞,代名詞,一般,*,*,*,*,*
まで	助詞,副助詞,*,*,*,*,*,*
疑う	動詞,自立,*,*,*,*,*,*,*,*,*,*,*
なんて	助詞,副助詞,*,*,*,*,*,*

図 1：漫画の台詞の形態素解析失敗例と対処例

5. 台詞とオノマトペの計量的分析

5.1 台詞における語彙量

対象テキストを形態素解析して品詞ごとに分割したものに対して、語彙の豊富さ（語彙多様性）または語彙密度の指標となる TTR(Type-Token Ratio)⁽¹⁰⁾ を算出した結果を各対象テキストの文字数と共に表 1 に示す。Type は延べ語数の事を指し、テキスト中に出現する語の累計数であり、Token は異なり語数を示し、重複した語を除いた語種の数であり、 $TTR=Type/Token$ の式で表される。形態素解析のエラーが多少含まれるのと、複合語は分割されてしまうなど正確な数値とは言えないが、同じ単語を繰り返して用いられているか、異なる単語を豊富に用いて表現しているのかが TTR 比により、ある程度判別

できる指標である。

表1：対象漫画のTTR

作品名	文字数	Token	Type	TTR (%)
エヴァンゲリオン	10612	4919	1254	25.49
僕らがいた	10650	4795	974	20.31
銀魂	10692	5114	1131	22.12
Rookies	10291	5083	948	18.65

表 1 に示す通り、他のジャンルの漫画と比較してもエヴァンゲリオンが一番様々な語彙が使用されており、語彙多様性の高い作品である結果となった。しかし、今回分析した作品は対象テキストの量が少ないこともあり、作品全体を通した語彙の違いというよりも、漫画に含まれる各場面における語彙の違いという結果であると考えられる。

たとえば、僕らがいたと Rookies はジャンルが恋愛とスポーツに分かれているが、場面設定が学校であるため、ジャンル特有の特徴が出ているとは考えにくい。一方で、エヴァンゲリオンは戦闘シーン、学校生活、家での生活など場面が多数含まれ、様々な語彙が使用されていたため、このような結果になったのだと予測される。詳細な語彙については、後章において分析を行う。

5.2 台詞の品詞構成

次に、品詞構成を調べる事により、テキストの特性を見てみる。表 2 に台詞の品詞構成を示す。どのジャンルにおいても品詞の構成割合はほぼ同じで、名詞が 35%以上を占め、続いて助詞と動詞が多く使われていることがわかる。今回は同じ漫画の対話中心のテキストにおけるジャンル別の特徴が出なかったが、書き言葉や一般的な話し言葉のテキストと比較することにより、漫画特有のテキスト特性を見つけてることができる。書き言葉のテキストを扱う時には、感動詞やフィラーなどの品詞はほとんど出現しないため、これは漫画特有の現象だと考えられる。

5.3 オノマトペの割合

次に台詞に付随するオノマトペについて比較を行った。オノマトペは台詞の中にも含まれている事がある。たとえば「チャホヤされてええ気になってんちゃうわ (エヴァンゲリオン)」の「チャホヤ」はオノマトペであるが、本稿ではこのような台詞の中で用いられているオノマトペは分析対象外として、台詞の吹き出しの中以外の背景で表現されているものを対象とした。表 3 にオノマトペの文字数と語彙量

の統計調査の結果を示す。

表 2：主要品詞構成比 (%)

品詞	エヴァン ゲリオン	僕らがいた	銀魂	Rookies
名詞	36.84	34.37	35.35	37.71
助詞	27.83	26.76	28.35	26.22
動詞	13.97	14.91	16.29	14.42
助動詞	9.84	10.45	9.29	10.45
副詞	3.54	3.15	2.99	3.50
感動詞	2.44	2.52	1.37	2.05
形容詞	2.11	2.90	1.78	1.69
フィラー	0.85	1.48	0.39	1.51
接続詞	0.51	1.42	0.98	0.81

このオノマトペは分析対象の台詞が記述されていた範囲内で出現したものである。ここでは形態素解析を行わず、表記通りに算出している。台詞の文字数はだいたい同程度に合わせたが、オノマトペは「エヴァンゲリオン」が突出して多い。台詞とオノマトペの Token(述ベ語数)の合計に対するオノマトペの割合は、「エヴァンゲリオン」が一番多く 6.59、「銀魂」が 3.58、「rookies」が 1.24、「僕らがいた」が 1.04 と差が出た。「エヴァンゲリオン」は台詞の語彙多様性が高い原因として様々な場面設定が考えられたが、全体的にオノマトペを好んで使っている漫画であり、「僕らがいた」や「Rookies」と同じ学校生活の場面と比較しても、オノマトペを多用する傾向があった。

表 3：オノマトペの文字数と TTR

作品名	文字数	Token	Type	TTR(%)
エヴァンゲリオン	1815	324	241	74.38
僕らがいた	231	50	41	82.00
銀魂	778	183	120	65.57
Rookies	279	63	50	79.37

6. 台詞とオノマトペの質的分析

5章で語彙統計に基づいて分析を行ったが、本章では具体的な内容を見て行く。表 4 に各作品における名詞の高頻度語上位 10 語を示す。名詞には代名詞、固有名詞等も含まれるがここでは一般に分類される名詞を抽出した。

各ジャンルで 1 作品しか扱っていないため、作品の特徴的な言葉が上位に出てくる事がわかる。「僕らがいた」と「Rookies」は同じ学校生活に関する場面であったが、友人や仲間を中心とした前者と、学校という権力に立ち向かう後者とでは使用される語彙

が異なる事がわかる。また、「銀魂」では、歴史 SF ギャグバトルという複数の要素を持った内容であるため、「侍」「幕府」などの言葉が多く用いられている。特に注目すべき点としてエヴァンゲリオンでは、「使徒」「シンクロ」といった一般で用いられる用語がこの作品では一種の専門用語となっていることである。その専門用語が頻繁に使われていることが特徴的である。

漫画における専門用語は、通常使われる一般的な意味を持つものに、漫画特有の定義を当てはめている場合と、造語で表現する場合の 2 通りがある。造語の場合は、辞書に登録されていないため未登録語として抽出することが可能であり、検出が比較的簡単である。一方で、本来の意味とは違う意味付けをした用語に関しては、出現する文脈の前後で説明を入れているため、そのパターンを識別することによって抽出が可能になるが、現段階では自動的に抽出することが難しい事例である。先行研究にもある通り、国内外で漫画に使われる特有の用語集というものを作成しているサイトが多く存在することから、Web サイトから用語の見出し語を収集し、検出する方法も今後の課題と言える。

表 4：高頻度名詞上位 10 語

エヴァンゲリオン	僕らがいた	銀魂	Rookies
使徒	人	侍	先生
人類	クラス	姉	野球
初号	ゆかた	国	教師
パイロット	女子	剣	ネクタイ
ロボット	男子	幕府	夢
父	悩み	万事	教頭
フィールド	気	問題	学校
シンクロ	議長	人	人間
戦闘	名前	天人	生徒
兵器	うそ	皇子	校長

次にオノマトペの分析を行ったが、オノマトペについては、何を一つの単位と考えるかという基準を設定していなかったために、統計を取る事が難しかった。今回調べたオノマトペはほとんどが擬音語であった。擬態語で用いられているものはむかついている時の「むっかあ」、精神的にショックを受けた時の「がーん」といったものしかなかった。「がーん」については、何かがぶつかった時の擬音語として用いられていることもあり、曖昧性を持つオノマトペである。擬音語と擬態語としての曖昧性もあるが、同じ音の表現でも異なる場面で用いられることもあるため、どのような状況で使われているオノマ

トペであるかも比較すると面白いかもしれない。漫画では擬態語は台詞の中に含まれているか、あるいは心情を説明する文章で表現するなどあまり用いられていないようであった。文字情報だけで表現するようなブログや小説の方が多く用いられるのではないかと予測される。

7. おわりに

本稿では、漫画を電子化し、そのテキストを言語処理の手法を用いて定量的・質的に分析を行った。今回は電子化データを用いて、どのような分析が可能であるか、どのような現象が見えてくるのかの調査が主な目的であった。実際、データが少なすぎたため、比較分析が困難なレベルであったが、電子化に伴うアノテーション方法や形態素解析失敗などの問題点をいくつか発見できた。

ジャンルによって出現語彙の種類が異なると安易に考えていたが、ジャンルが違っていても場面によって使用される語彙というのはある程度特定されるため、ジャンルというよりも場面という細分化された単位も必要ではないかと考える。

今後は更に漫画のテキストデータ量を拡大すると同時に、談話としてのコーパスと捉え、談話解析やコミュニケーションの事例集としての応用を検討していく。

参考文献

- (1) スコット・マクラウド, 岡田斗司夫訳, マンガ学, (1998), 161-163, 美術出版社
- (2) 熊野七絵, 日本語学習者とアニメ・マンガ~聞き取り調査結果から見える現状とニーズ~, 広島大学留学生センター紀要, 20, (2010), 89-103.
- (3) 国際交流基金, 海外の日本語教育の現状日本語教育機関調査, (2009), 凡人社
- (4) 熊野七絵, 川嶋恵子, 「アニメ・マンガの日本語」Web サイト開発, 日本語学, (2010.04), 103-117.
- (5) 伊藤雅光, 英語圏における日本の漫画・アニメ用語集の構造と基本用語の特徴, 日本語学, 22, 12, (2003.11), 66-74.
- (6) 熊野七絵, アニメ・マンガの日本語~ジャンル用語の特徴をめぐって~, 広島大学国際センター紀要, 1, (2011,10), 35-49.
- (7) 夏目房之介, マンガの力, (1999), 84-85, 昌文社
- (8) 徳久雅人, 村上仁一, 池原悟, 漫画における表情

に着目した情緒タグ付きテキスト対話コーパスの構築, 自然言語処理, 14, 3, (2007.04), 193-217.

(9) 内田ゆず, 荒木健治, 米山淳, ブログ記事からのオノマトペ用例文の自動抽出方法, 知能と情報(日本知能情報ファジィ学会誌), 24, 3 (2012.06), 811-820.

(10) 石川慎一郎, ベーシックコーパス言語学, (2012), 275, ひつじ書房