

アニメの日本語と日本語教育 ——Co-Chuによる現実性および適切性分析——

ラニガン・マシュー

日本語教育では、アニメの教材化が期待されているが、まず教材としてのふさわしさを明らかにする必要があると考えられる。本論では、アニメが日本語教育資源としてふさわしいかどうかを測定するために、アニメの日本語の現実性と日本語教育における適切性を検討した。

アニメの日本語の分析のために、アニメを利用した日本語教育の実践例とアニメをアニメのための言語分析の先行研究を検討した。その結果、日本語教育におけるアニメの使用のふさわしさは明らかになっていないことが分かった。アニメの言語分析を行うためにコーパス日本語学の手法を利用して、コーパスシステム『Co-Chu』の開発を行うことにした。第3章では、Co-Chuのシステムについて、概要及び機能について説明した。

Co-Chuはコーパスシステムとして、コーパス日本語学の様々な作業を統一したインタフェースにしている。その作業のためのツールが様々な存在しているが、それぞれインタフェースや使い方が異なり、そのすべてを使いこなすことが非常に難しい。必要とされる知識を少なくするため、Co-Chuでは、インタフェースの統一を最も重視した。また、技術的な知識の乏しい研究者でも使いこなせるよう、最新のウェブインタフェース開発の手法に基づいて、できる限り分かりやすくした。

Co-Chuは三層アーキテクチャで構築されている。ウェブ上の三層アーキテクチャにおいて、ソフトウェアをデータベース、ビジネスロジック、ユーザインタフェースの三つに分ける。データベースはシステムの記憶装置であり、ユーザインタフェースは文字通りの意味で、人間が使用する唯一の部分である。ビジネスロジックはデータベースとユーザインタフェースを繋げるものであり、Co-Chuの場合、ウェブ上のアプリケーション・プログラミング・インタフェース（以下APIとする）である。このアーキテクチャでいくつかのメリットが考えられるが、Co-Chuに選んだ最大な理由は保守性と拡張性である。システムが役割ごとに分けられているため、両方が非常に高い。

まず、Co-Chuのデータベースとしてリレーショナルデータベースシステムである PostgreSQL

*国際人間学研究科言語文化専攻 博士前期課程

を利用している。次に、ビジネスロジックの層であるアプリケーション・プログラミング・インタフェース（API）は Ruby 言語で作られている。具体的に、ウェブの部分は Sinatra というソフトウェアを利用し、データベースアクセスは、Sequel というライブラリーを利用している。この API は、主に REST（Representational State Transfer）によるデータアクセスである。この機能で、ユーザインタフェースからデータの挿入、編集、そして削除ができる。最後に、Co-Chu のウェブ上のユーザインタフェースは AngularJS というライブラリーに基づいて、システムの全体と同じように、ユーザインタフェース自体も三層アーキテクチャになっている。このライブラリーにおいて、MVC（Model-View-Controller）アーキテクチャでウェブサイトを作成している。

具体的に、Co-Chu の主要な機能は Build、Import、Edit、Analyze という四つのモードに分けられている。Build モードでは、コーパス構築の最初の作業としてコーパスの階層木構造（Hierarchical Tree Structure）を作成する。それに加えて、メタデータの付加と発話者の管理もこのモードに含まれる。次に、Import モードでは、データを挿入する。現在対応しているデータフォーマットは BTSJ 式の Excel ファイルと Advanced SubStation Alpha の字幕データファイルの二つであるが、テキストファイルや Excel ファイルの対応も今後の課題としている。次になる Edit モードではデータ編集を行う。このモードに含まれる機能はデータ処理を①データ編集、②自動的データ処理、③エラー処理の三種類に分けられる。特に重要な機能は MeCab に基づいている形態素解析機能であり、基本的にデータ編集→形態素解析→エラー処理→他の自動的データ処理、という流れで進む。現在、形態素解析以外の自動的データ処理はコロケーション解析のみである。最後に、Analyze モードでいくつかの検索モードを利用してコーパス分析が行われる。

第4章では、アニメの日本語について述べた。まず、アニメの現実性と適切性を検討するために、アニメのスクリプトを集め、Co-Chu でコーパス化を行った。アニメを日本語の授業で利用することで第一の目的は娯楽ではないが、ほかの条件がすべて同じであれば、学習者が楽しめる作品のほうが良いだろう。そのため、アニメの選定に①できるだけ最近の人気作品であることと、②すでに終了していること、という二つの条件を付けた。人気があって終了している作品では、学習者がすでに見て内容を理解している可能性もあり、授業で比較的使用しやすいとも考えられる。また、ジャンルによる偏りを避けるため、様々なジャンルから作品を選択した。

今回選択した作品は、①『GATE 自衛隊 彼の地にて、斯く戦えり』、②『アルドノア・ゼロ』、③『干物妹！ うまるちゃん』、④『けいおん！』、⑤『赤髪の白雪姫』の五つである。Co-Chu に取り込んで、形態素解析を行った結果、語数は 15 万語となった。また、比較対象となるコーパス『BTSJ による日本語話し言葉コーパス 2011 年版』は約 115 万で、アニメコーパスの 7 倍以上である。

以上述べたように、アニメコーパスを対象に、現実と適切性の二種類の分析を行った。本研究において、現実性は実際に日常生活で使われている日本語との一致度を表し、適切性は日本語教育の目的を果たせるかどうかを示す。次に、それぞれの分析方法と結果について述べる。

まず、現実性を検討するために、語彙使用分析とコロケーション分析を行った。実際に使われ

ている日本語と一致するかどうかを検討するために、それぞれ『BTSJ』による日本語話し言葉コーパス 2011 年版』（以下 BTSJ）と比較した。語彙使用分析では、「イ形容詞」「ナ形容詞」「動詞」という三つの品詞を調べ、各品詞の使用頻度順と用法の割合を比較した。

イ形容詞では、アニメ（ $n=3166$ ）に最も使用されている語は「無い」「良い」「すごい」という順であり、BTSJ（ $n=21497$ ）と一致していることが分かった。さらに、この両方のコーパスにおいてこの3語はイ形容詞の延べ数の大半を占めている。「すごい」の頻度（アニメ=4.23%、BTSJ=12.66%）に差が見られたが、コーパスの性質（BTSJには雑談が多いことなど）に原因があると考えられる。さらにイ形容詞の使用を検討するために語彙用法を調べた。具体的に、用法を「非過去+肯定」「非過去+否定」「過去+肯定」「過去+否定」の四つに分けて分析した。その結果、アニメの使用（非過去+肯定=91.63%、非過去+否定=1.86%、過去+肯定=6.48%、過去+否定=0.03%）は BTSJ（非過去+肯定=92.57%、非過去+否定=2.35%、過去+肯定=4.97%、過去+否定=0.11%）とほとんど一致していることを明らかにした。四つの分類の中で、アニメと BTSJ では「過去・肯定」（例えば「寒かった（です）」）にのみ有意な差が見られた（ $p<0.001$ ）。以上の結果から、イ形容詞の使用はほとんど現実的であるといえるだろう。

ナ形容詞を検討するためにイ形容詞と同様の分析を試みた。しかし、語彙頻度表を抽出した結果、ナ形容詞と認められない語（例えば、「よう」「みたい」「普通」など）も多く含まれていることが分かった。だが、Co-Chu で使用されている電子化辞書 UniDic に基づく事情により、Co-Chu でナ形容詞のみを抽出することができないというのが現状である。また、本来ナ形容詞はイ形容詞と同様に用法を四つに分けることができると考えられるが、上記の事情に加え、用法の抽出にも困難があることが分かった。特に、各用法の文型が多く存在しているからである。例えば、「過去・否定」の用法では「～ではない」「～ではありません」「～じゃない」など、様々なパターンが考えられ、分析が難しい。以上の問題点があるため、ナ形容詞の現実性の分析方法を再検討することを今後の課題としたい。

動詞も以上と同様にまず使用頻度表を抽出した。その結果、アニメ（ $n=16446$ ）と BTSJ（ $n=81813$ ）の上位3位に違いがあった。アニメは「する」「いる」「いう」の順であり、BTSJ は「いう」「する」「ある」の順である。しかし、「いる」は補助動詞としての使用（～てV）が含まれていると考えられるため、「～てV」の形のものを除いて抽出してみた。そうすれば、アニメ（「する」=16.94%、「いう」=4.64%、「ある」=4.10%）と BTSJ（「いう」=14.93%、「する」=12.39%、「ある」=6.54%）の上位3位は順番が異なるが、出る語彙が一致するようになる。補助動詞に関しては、アニメにおいて、「～ている」が非常に多く、動詞の16446回中、475回出現している（2.89%）。一方、BTSJ では、「～ている」は81813回動詞が出現している中、910回（1.11%）しか使用されていない。このことから、補助動詞の使用についてはさらに検討する必要がある。また、動詞の用法については、アニメ（非過去+肯定=77.81%、非過去+否定=7.07%、過去+肯定=14.49%、過去+否定=0.63%）と BTSJ（非過去+肯定=79.67%、非過去+否定=8.34%、過去+肯定=11.42%、過去+否定=0.58%）の一部は一致しているが、「非過去+否定」と「過

去+肯定」の結果に見られるように、一致しないところもあることが分かった。

現実性をさらに検討するために、アニメと BTSJ に現れるコロケーションを調べた。コロケーションという連語は、構成の語と一緒に出現する確率が高いものである。例えば、日本語では「役割」といえば、「～を果たす」で続く可能性が非常に高い。各コーパスの2～6グラムのコロケーションを抽出した結果、アニメから1599種類、BTSJ から6617種類取り出した。各グラム数の上位10位（つまり、アニメ5作品の各コーパスから50項目）をみると、5件しか重なっていない。それは「ている」「てくる」「ほうが良い」「じゃないですか」と「よろしくお願いします」の五つである。それぞれ重要なものではあるが、それらしか一致していないことではコロケーションの面でアニメは現実性があるとは言えないだろう。さらに、それぞれのコーパスの全コロケーションを見ても、予想外に一致しない。アニメコーパスと BTSJ の両方に現れるコロケーションは297件しか存在しない。

しかし、この問題はアニメが現実的ではないからではないことも十分考えられる。それより、アニメコーパスの規模に影響を受けている恐れがある。例えば、アニメコーパスでは6グラムに上れば頻度が非常に下がっており、結果の質も下がってしまうことが見られる。今後はさらに多くのアニメのスク립トをコーパスに取り込み、アニメコーパスの規模を大きくすれば、この問題は自然に解決される可能性がある。

以上の結果から、「イ形容詞」と「動詞」の使用はある程度現実的といえるだろう。また、コロケーションの面は現実的とは言えない。ただし、データの量と質にも原因があると考えられるため、さらに検討する必要がある。

次に、適切性分析の分析方法と結果について述べる。上述したように、本研究でいう適切性はアニメが日本語教育の目的を果たせるかどうかを示す。そのために、評価基準として一般的であると考えられる旧日本語能力試験の出題基準をベースにして対応分析を行うことにした。様々な分析が考えられるが、今回は一つの試みとして旧能力試験の観点からアニメの適切性分析を行う。本来、適切性分析は一つの作品に対して行うものであるが、作品ごとの分析を行い、結果が非常に近いことが分かったので、ここではまとめて述べる。

現実性分析と同様、適切性分析の対象品詞をイ形容詞、ナ形容詞、動詞の三つにした。ただし、「ナ形容詞」だけではなく、ほかのものも含まれていることを改めて注意したい。その結果、異なり語数では3・4級語彙が比較的少なく、2級以上が多い。一方、延べ数で見ると、3・4級語彙は基本的に大半を占めており、2級以上の語彙の使用が少ない。動詞に焦点を当てると、「する」は4級語彙が多い原因であると考えられる。本来、「する」自体ではなく、その前に来る語（例えば、「練習する」の場合、「練習」）のレベルの方が重要であると考えられる。本分析では、「する」のみが分析対象と期待している。

また、レベル割に加えて、旧日本語能力試験に対するカバー率分析を行った。その結果、3・4級語彙は8割以上カバーされており、それ以上レベルが上がるとカバー率が下がる（2級＝約6割、1級＝約3割）ということが分かった。この結果から、アニメは上級語彙資料として適切で

はないといえるだろう。しかし、これはアニメの問題ではなく、発話の種類による現象であるという可能性が高い。そのため、BTSJを対象にして同じ分析を試みた。その結果、BTSJ、つまり日常会話に、非常に近い傾向が見られることが分かった。このことから、旧日本露能力試験が上がれば上がるほど、話し言葉が少なくなり、書き言葉が多くなるということが分かる。したがって、アニメを上級語彙資料として使用するとすれば、ほかの資料を補助的に使用する必要があると考えられる。ただし、具体的な使用方法などについては、さらに検討する必要がある。

アニメのスク립トを集め、Co-Chuでコーパス化し、分析を行った結果、アニメの日本語にははっきり現実的であるといえる部分は指定できたが、一致しない部分も多かった。その原因はデータに残っている形態素解析のエラー、またはコーパスが小規模であることが考えられる。また、適切性の分析からアニメが日本語教育において様々な使用方法が考えられることを指摘することができた。しかし、今回の分析は一つの試みであり、現実性と適切性の両方の測定方法をさらに検討する必要がある。特に、今回利用した分析方法は語彙の分析に限っており、現実性と適切性を見出すために、文法や語用論的分析など、ほかの分析を行う必要がある。

本研究を進めるために、① Co-Chuをさらに充実したシステムにする、②より多くアニメの発話データのコーパス化を行う、③現実性と適切性の測定方法を再検討することを今後の課題としたい。