

Can-do 型日本語学習用資源としてのアニメーション字幕の分析

大河原 龍太郎¹ 望月 源²

¹東京外国語大学 言語文化学部 ²東京外国語大学大学院総合国際学研究院
{okawara. ryutaro. s0, motizuki}@tufs. ac. jp

概要

近年, Can-do を活用した言語学習が言語教育では主流であるが, 同時に実用的な例文を増やすことが1つの課題となっている. 本研究では, 学習教材の資源としてのアニメーション字幕に焦点を当てる. 分析のため, 会話部分を人手により抽出し, AJ Can-do リストを参照し, 例文に利用可能な会話を選定する. また, 選定時に Can-do 目標からのレベル推定も行う. 抽出した Can-do 会話とレベルから, Can-do 学習に対応した会話例教材の資源としての, アニメーション字幕コーパスの利用可能性について考察する.

1 はじめに

近年, 学習者の言語能力・技能を「～できる」という形式で記述した Can-do を活用した言語学習が, 言語教育で主流となっている. Can-do 型学習では, 学習する項目に対応した会話例・例文などの教材が欠かせないが, 教科書の例だけでは十分でない. 教師が例文を作成して追加することも多いが, 手作業には量や話題の幅など限界もあり, 効率的な拡充方法が望まれる. 本研究では, 日本語アニメーションの字幕を Can-do 型学習用教材の資源として活用することを検討する[1].

日本のアニメーション作品はストーリー性が強く, キャラクター同士の会話を軸に構成されている. そのため, アニメーション字幕コーパスはさまざまな機能・目的を持った豊富な会話を含んでおり, Can-do 形式の学習に対応した質の高い日本語会話教材資源として, 活用が期待できる.

本研究では, アニメーション字幕コーパス内から会話部分を人手により抽出する. 次に, 東京外国語大学留学生日本語教育センターの AJ Can-do リスト[2]を参照し, 例文に利用可能な会話(以下, Can-do 会話)を選定する. 選定時に参照するレベル別の Can-do 目標に照らして Can-do 会話のレベル推定も行う. 抽出した Can-do 会話とレベルから, Can-do 学習に対応した会話例教材の資源としての, アニメーション字幕コーパスの利用可能性について考察する.

2 関連研究

アニメーションを学習資源として活用する取り組みは, 1990年代の「マルチメディア教材」への注目に端緒をもち[3], 特に2006年の日本語教育学会での「映画・アニメ・マンガ—日本語教育の映像素材」というシンポジウムをきっかけに活発化した[4]. アニメーションの教材利用を検討した研究の一例には, 田中と本間による研究がある. この研究は, スタジオジブリのアニメーション映画作品「耳をすませば」について, 「現実の日常会話」の出現頻度および「初級語彙・文法との対応関係」を分析した. その結果, 当作品が「未習者が『耳をすませば』で学習した場合, 初級語彙の4割程度, また, 初級教科書全体にほぼ匹敵する初級文法の8割弱がカバーできる」こと, そしてアニメーションが初級日本語学習に活用されうることを示した[5].

アニメーションは, 文法事項を広くカバーしていることに加え, 学習動機としても重要な役割を果たしている. Alsubaie and Alabbad は, ”The Effect of Japanese Animation Series on Informal Third Language Acquisition among Arabic Native Speakers”において, アニメーション鑑賞に一定の学習効果があることを報告したうえで, 学習者が無意識的に学習に取り組むという点をメリットとして挙げている[6]. また, 熊野・廣利の研究[3]をはじめとした多くの研究が, アニメ・マンガの外国人学習者の学習動機としての役割を指摘している.

3 Can-do 会話の抽出

3.1 テレビ字幕コーパス

本研究では, 東京外国語大学計算言語学研究室で収集整備している「日本語テレビ字幕コーパス」を用いる[7]. このコーパスは, 東京都の地上波デジタルテレビ放送に付与されたクロズドキャプション(字幕)データを収集し, 不要な情報を除去した後に, 字幕表示のために断片化した文を1文にまとめたうえで字幕表示のタイミング情報を付与したものであ

り、以下の特徴がある。

- 2012年12月から2021年9月の時点で518,960番組, 179,268,965文, 1,945,899,573形態素。
- 各テキスト(番組)はテレビ局の番組情報に基づいて、アニメ/特撮, スポーツ, ドキュメンタリー/教養, ドラマ, ニュース/報道, バラエティ, 映画, 音楽, 劇場/公演, 趣味/教育, 情報/ワイドショー, 福祉, および, その他の13種類に分類されている。

本研究ではこのコーパスの中から、「アニメ/特撮」のテキストをアニメーション字幕コーパス(アニメーション作品延べ41,981番組, 約7,944万語)として使用する。字幕データの例を図1に示す。

文番号	表示開始時間	表示終了時間	字幕文
18	0:04:45.44	0:04:52.45	そういえば 綾んちの喫茶店行ったことないな。
19	0:04:52.45	0:04:54.45	そうだけ。
20	0:04:52.45	0:04:56.45	ねえねえ これからお邪魔していい?
21	0:04:56.45	0:04:58.45	いいけど....

図1 字幕コーパスの例(「ゲゲゲの鬼太郎」2020年1月5日放送より)

3.2 アニメーション作品の選定と分類

分析の対象とするアニメーション作品を以下の手順で選択しジャンルの分類をする。

- 日本国内外で人気の高い日本語アニメーション作品群を選定する。選定基準として、世界最大級のアニメレビューサイトである“MyAnimeList”¹(総ユーザー数は2021年9月時点で11,566,100人以上)の”Top Anime”および”Most Popular”のランキングのいずれかで1000位圏内とする。
- アニメーション字幕コーパスから上記リスト内にある作品を無作為に12作品選出する。
- 選出したアニメーション作品のジャンルを定める。ジャンルは、日本国内におけるアニメーション配信サービス大手であるDアニメストア²の分類に準じて決定する。
- 上記で定めたジャンルをAファンタジーとBリアリティの2グループに大別する。

今回の無作為抽出では、12作品を選出し、それぞれが7つのジャンルに割り当てられた。選出された作品とそのジャンルおよびA,Bいずれのグループに属するかを表1に示す。

表1 選定されたアニメーション作品

A: ファンタジー	B: リアリティ
A1: 七つの大罪(SF/ファンタジー)	B1: ハイキュー (スポーツ/競技)
A2: マギ 第2期(SF/ファンタジー)	B2: 映像研には手を出さな! (ドラマ/青春)
A3: 僕だけがいない街 (ホラー/サスペンス/推理)	B3: 食戟のソーマ (ドラマ/青春)
A4: PSYCHO-PASS (ホラー/サスペンス/推理)	B4: 冴えない彼女の育て方 (恋愛/ラブコメ)
A5: 僕のヒーローアカデミア(アクション/バトル)	B5: 四月は君の嘘 (恋愛/ラブコメ)
A6: 炎炎ノ消防隊 (アクション/バトル)	B6: 銀の匙 Silver Spoon (日常/ほのぼの)

3.3 コーパス内の Can-do 会話抽出

アニメーション字幕コーパス内の会話部分を次の基準で抽出する。

- 共通の話題についての、一回以上話者が交代する発話を会話
- 内容の明らかでない指示語を多く含む会話や、文脈の理解が難しい会話の選択は避ける
- 話題の転換により会話の区切りを判断する
- ただし、会話文の長さの目安を、最大20文程度、話者交代の回数を6回前後と定める
- 前の文の字幕表示終了時間と次の文の開始時間が一致している場合は会話は連続し、途切れた時点で会話が途切れたと考える

抽出した会話が Can-do 会話かどうかを次の考え方で判定する。本研究では、Can-do リストとして、東京外国語大学留学生日本語教育センターが開発した「JLPTUFS アカデミック日本語 Can-do リスト (AJ Can-do リスト) [2]」の4技能「読解」「聴解」「文章表現」「口頭表現(独話・対話)」のうち、「口頭表現(対話) (2017年3月21日版)」を用いる。AJ Can-do リストは、それぞれについて、初級1, 初級2, 中級1, 中級2, 中上級, 上級1, 上級2, 超級の8段

¹ MyAnimeList <https://myanimelist.net>

² Dアニメストア, <https://anime.dmkt-sp.jp/animestore>

階に分かれている。また、各レベルには「何ができるか」の Can-do 目標と、扱う会話のテーマ例が掲載されている。

例えば、初級1では「基本的な挨拶ができる／易しい語彙や表現を使って、簡単な質問をしたり、答えたりできる」で、テーマ・トピック例は「挨拶、買い物、注文」のように記述されている。

抽出した会話のそれぞれと AJ Can-do リストの記述を比較し、Can-do 会話とそのレベルを推定する。

今回の抽出では、対象とした全 12 作品 100 話分の字幕データから、1,600 の Can-do 会話を得た。655 会話はグループ A (ファンタジー)、945 会話はグループ B(リアリティ)に大別された。総文数 35,261 文の内、Can-do 会話の文数は 19,808 文で、全体の約 56.2%が Can-do 会話として推定された。詳細を表 2 に示す。表中で、C 文数は内 Can-do 会話の文数、C 率は Can-do 会話率、C 話数は Can-do 会話の数をそれぞれ表す。

表 2 Can-do 会話

作品	話数	総文数	C 文数	C 率(%)	C 話数
A1	8	3,566	1,634	45.8	139
A2	8	3,132	1,400	44.7	102
A3	8	1,777	1,204	67.8	111
A4	8	2,529	1,246	49.3	119
A5	8	2,760	853	30.9	68
A6	8	2,586	1,233	47.7	116
小計	48	16,350	7,570	46.3	655
B1	8	3,438	1,782	51.8	122
B2	8	3,722	2,724	73.2	191
B3	9	2,685	1,438	53.6	115
B4	8	2,569	1,632	63.5	132
B5	8	2,280	1,165	51.1	85
B6	11	4,217	3,497	82.9	300
小計	52	18,911	12,238	64.7	945
合計	100	35,261	19,808	56.2	1,600

4 Can-do 会話の分析

3で抽出した Can-do 会話について、会話率、レベル分布の観点から利用可能性について分析をする。

4.1 Can-do 会話率

表 2 の「C 率」欄に、各作品の Can-do 会話率が示

されている。Can-do 会話率は、作品中の Can-do 会話文数を作品中の総文数で割ったもので、1 つのテレビ番組内にどのくらい Can-do 会話が含まれているかの割合を示している。最も高い B6(銀の匙 Silver Spoon)では 82.9%、最も低い A5(僕のヒーローアカデミア)においても 30.9%の Can-do 会話率が得られた。また全作品を総合した Can-do 会話率は 56.2%であった。これらの結果より、人気作品群には学習教材資源として十分な量の Can-do 会話が含まれていると考えられる。

グループ A とグループ B での Can-do 会話率を比較すると、A が 46.3%であるのに対し、B では 64.7%であり、全字幕に対する Can-do 会話の割合は B の方が高かった。グループ A (ファンタジー) はサブジャンルが「SF/ファンタジー」「ホラー/サスペンス/推理」「アクション/バトル」であり、戦闘アクションなどの描写が多い。一方、グループ B(リアリティ)のサブジャンルは「スポーツ/競技」「ドラマ/青春」「恋愛/ラブコメ」「日常/ほのぼの」であり、一般的な会話がより多かったものと推測される。

4.2 Can-do レベル分布

Can-do 会話として利用可能かどうかを判定する際に、Can-do 目標の記述を参照し、レベルの推測もおこなった。各レベルごとの Can-do 会話の分布を表 3 に示す。

表 3 Can-do レベルの推定結果

レベル	グループ A	グループ B	合計
初級 1	11	37	48
初級 2	131	153	284
中級 1	81	80	161
中級 2	285	705	990
中上級	401	347	748
上級 1	63	88	151
上級 2	13	3	16
超級	0	0	0
合計	985	1,413	2,398

表 3 に示す今回の調査で、アニメーション字幕コーパス内の Can-do 会話は、初級 1 から上級 2 までの幅広いレベルをカバーしていた。なかでも、中級 2 と中上級の Can-do レベルについて、豊富な会話例を含んでいた。一方、上級以上と推測される Can-do 会話は少なく、最上位の「超級」に該当する Can-do

会話は抽出されなかった。この結果より、人気作品群のアニメーション字幕コーパスは、比較的広い範囲の Can-do レベルの会話を含んでおり、特に中級2と中上級の Can-do レベルについて、豊富な会話例を含んでいると推測される。

次に、グループ A とグループ B 間の Can-do レベル分布を比較する。事例が 0 だった超級を除いた 7 レベルでの結果に対して、帰無仮説「A・B グループ間に Can-do レベル分布の差はない」とした χ^2 検定を行ったところ、帰無仮説は棄却され、2 グループの Can-do レベル分布に有意な差がみられた ($\chi^2 = 136.21, p < 2.2e-16$)。続いて残差分析を行った。調整済み残差の結果を表 4 に示す。

表 4 残差分析における調整済み残差

	初級 1**	初級 2	中級 1*	中級 2***
A	-2.583	1.842	2.466	-10.256
B	2.583	-1.842	-2.466	10.256
	中上級***	上級 1	上級 2**	超級
A	8.399	0.166	3.277	-
B	-8.399	-0.166	-3.277	-

表 4 で*は 5%水準、**は 1%水準、***は 0.1%水準で差が有意であったことを表す。統計的に見て、Can-do 会話は中級 2 ではグループ B、中上級ではグループ A に非常に高く、初級 1 ではグループ B、上級 2 ではグループ A に高く、中級 1 ではグループ A にやや高い割合で出現しているといえる。

次に、グループ A とグループ B の Can-do 会話の総数に占める、各レベルの Can-do 会話数の割合を表 5 に示す。

表 5 Can-do 会話の各レベルでの割合

	初級 1	初級 2	中級 1	中級 2
A	1.11	13.29	8.22	28.98
B	2.61	10.82	5.66	49.89
	中上級	上級 1	上級 2	超級
A	40.71	6.39	1.31	0.00
B	24.55	6.22	0.21	0.00

表 5 に示すとおり、グループ A においては、中上級、中級 2 が多く、それぞれ 40.71%と 28.98%だった。以下、初級 2、中級 1、上級 1 と続き、初級 1 と上級 2 は非常に少なかった。一方グループ B においては、中級 2 が最も多く 49.89%であり、次に中上級が 24.55%だった。以下、初級 2、上級 1、中級 1、初級 1 と続き、上級 2 は非常に少なかった。

4.3 分析のまとめ

以上の分析から次のことがいえる。

- 今回調査した人気作品群のアニメーション字幕コーパスには、Can-do ベースの会話例教材資源として十分な量の会話が含まれる。
- 字幕内の Can-do 会話率は作品のジャンルによって異なり、ファンタジーに分類されるグループ A よりもリアリティに分類されるグループ B が高い傾向が見られた。
- 字幕コーパス内の会話は初級から上級までの広い Can-do レベルをカバーしており、特に、中級 2 と中上級の割合が多かった。また、今回の調査では最上位の超級の Can-do 会話は抽出されなかった。
- 字幕コーパス内の Can-do レベル分布では、平均 Can-do 会話率が低いグループ A の方が、比較的難度の高い中上級の会話をグループ B より含んでいた。これは、A の「ファンタジー」作品は、よりシリアスな話題についての会話が多いこと、一方で B の「リアリティ」作品は日常シーンなどを多く含み、日常会話レベルの Can-do 会話がより多く含まれることなどが理由として推測される。

5 おわりに

本研究では、日本語アニメーションの字幕を Can-do 型学習用教材の資源として活用することを検討した。アニメーション字幕コーパス内から会話部分を人手により抽出し、AJ Can-do リストの各レベルの Can-do 目標を参照し、例文に利用可能な Can-do 会話の選定とレベル推定を行った。調査・分析の結果から、アニメーション字幕コーパスは、さまざまな機能・目的を持った豊富な会話を含んでおり、Can-do ベースの会話例教材の資源として適していると考えられる。本調査はアニメーションを学習教材に利用するという観点からは、比較的規模が大きいと言えるが、今後さらに対象数を増やした詳細な分析を行う必要がある。また、今回の著者による Can-do レベル推定に加えて、今後日本語教育者も含めた複数での推定を行いレベル推定の精度を高める必要がある。さらに、今後の展望として、本研究で抽出した Can-do 会話を Can-do レベル付きの教師データとして用いることで、字幕コーパスからの Can-do 会話自動抽出への応用が考えられる。

謝辞

本研究は JSPS 科研費 JP19H04224, JP20H00096 の助成を受けたものです。

参考文献

- [1] 大河原龍太郎, Can-do 型日本語学習用資源としてのアニメーション字幕の分析, 卒業論文, 東京外国語大学, 3. 2022.
- [2] 東京外国語大学留学生日本語教育センター, JLPTUFS アカデミック日本語 Can-do リスト (AJ Can-do リスト), https://tufts-jlc-kyoten.jp/kyoten/development/ajcan-do/cando_list, 2022-01 閲覧.
- [3] 熊野七絵, 廣利正代, 「アニメ・マンガ」調査研究--地域事情と日本語教材 An examination of "ANIME & MANGA": regional information overseas and Japanese-Language materials, 国際交流基金日本語教育紀 4, pp. 55-69, 国際交流基金, 2008.
- [4] 臼井直也, 日本語教育におけるアニメーション研究のこれまでとこれから, 日本研究教育年報 21, pp. 103-117, 東京外国語大学, 2017.
- [5] 田中里実, 本間淳子, 初級語彙・文型による「耳をすませば」スクリプトの分析: 日本語学習資源としてのアニメーション映画の可能性, 北海道大学留学生センター紀要 13, pp. 98-117, 北海道大学留学生センター, 2009.
- [6] Alsubaie, Sara S.; Alabbad, Abbad M., The Effect of Japanese Animation Series on Informal Third Language Acquisition among Arabic Native Speakers. *English Language Teaching*, v13 n8, pp. 91-119, 2020.
- [7] Hajime Mochizuki and Kohji Shibano, Building Very Large Corpus Containing Useful Rich Materials for Language Learning from Closed Caption TV, *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, Volume 2014, No. 1, pp. 1381-1389, Association for the Advancement of Computing in Education (AACE), 10. 2014.