

日本語の単語と文法のレベルを用いたアニメのセリフの難易度の推定手法

Estimation of Japanese Anime Dialogues' Degrees of Difficulty using Japanese Words and Expressions

Shan, Junjie^{*1}
Shan, Junjie西原陽子
Yoko Nishihara山西良典
Ryosuke Yamanishi福本淳一^{*2}
Junichi Fukumoto^{*1} 立命館大学大学院情報理工学研究科
Graduate School of Information Science and Engineering,
Ritsumeikan University^{*2} 立命館大学情報理工学部
College of Information Science and Engineering,
Ritsumeikan University

With the spread of Japanese Anime as the representative of Japanese culture around the world recent years, more and more peoples choose to study Japanese with interesting. Due to the increasing of learners, there are two problems become to be often mentioned. First, the Japanese study materials are not always sufficient. Second, the traditional teaching materials usually not satisfied with the exercise of communication skills, which cause the listening and conversation abilities of learners abroad are usually not very well. In the research, we choose the Japanese Anime which contains plentiful voiced scene and conversations as the new Japanese study material. We analyze the degree of difficulties of dialogues in Anime and separate into different levels, in order to make utilizations for those learners abroad.

1. Introduction

More and more people choose to learn Japanese as a foreign language(JFL). We call these people the JFL learners. According to the Japan Foundation's survey [1], there were about 3.6 millions of people studied Japanese abroad at Japanese educational organizations in 2015, distributed in 137 countries. In such a large Japanese language education environment, there are two problems usually be mentioned in educational practices' feedback.

The first problem about the Inadequate Teaching Materials. This problem is always mentioned in Japanese studies abroad. The update rate of teaching materials also becomes hardly to keep up with the tendency. Traditional learning materials are becoming out-of-date and more and more difficult to satisfy with the JFL learners' growing demands.

The second one is ignorance of teaching of conversation abilities. The usual way of Japanese learning and teaching almost focuses on the Written and Reading abilities. As the practical application skills, conversation abilities might not be treated seriously in the teaching practices. The conversation ability is more important than others in language studies for learners who would like to understand in Japanese. A research conducted through JFL learners in the US showed that the students' confidence would be much enhanced when they noticed their own improvement of listening and speaking, i.e. conversation abilities [2]. Therefore, it should be urgent and important for JFL learners to take more often exercises in a real and native conversation environment.

In order to tackle these problems, we believe a new Japanese material that can be easily and quickly updated and attractive enough for handle JFL learners' attentions is necessary. Then,

連絡先:Shan Junjie, 立命館大学大学院情報理工学研究科,
gr0260hr@ed.ritsumeikan.ac.jp

this paper has taken our sights on the Japanese animations (Japanese Anime) that have more advantages than the ordinal Japanese study materials. There are at least three advantages.

At first, the quantity of anime is large and anime contains various dialogue scenes. Those scenes involve nearly every aspect of conversations with different difficulties. We can acquire various conversations from the anime and utilize them in training conversation with a real dialogue environment. Secondly, the dialogues in anime are usually spoken with standard and clear pronunciations. All the voice actors in Japanese anime are asked to participate in professional training, so that the pronunciations in anime are much more clear and standard than other normal voice recordings. And thirdly, the scenes in anime have enough attraction and interesting. Therefore, the immersions brought by the stories of anime can help the learners to keep their attentions and remain their focus on the contents.

In order to utilize those large quantities of Animes, we propose a method to estimate and separate the dialogues' degrees of difficulty of the Anime in our research. In this paper, as the first step, we analyze the dialogues' degrees of difficulty of the Anime.

2. Related Works

As one of the most famous Japanese cultures, there are many researches surround anime about using them in helping Japanese language study especially in the culture dimension [3,4].

Since 2009, Yazaki's series papers titled Anime-de Nihongo have introduced his researches about using Japanese anime in teaching Japanese [5]. In these teaching practices, the author investigated the applications of Japanese anime in developing new teaching methods and explored the effects in different types of student groups respectively. All the practices had got positive results, however, the focus of his practices were all about developing new "Teaching Strategies". Development of new

teaching materials is also required for JFL learners abroad. We try to develop new teaching materials in our study.

We have also made statistics of animes dialogues' level using the Japanese expressions with different types of story of animes [6]. Through the analysis we found that in different genre anime the level of Japanese words and expressions used in animes were different respectively. Moreover, the level of dialogue might be different according to the genres of anime. All JFL learners' levels of Japanese are not equal, but different. To use Japanese animes as learning materials, the dialogues of Japanese anime should be classified as their levels of Japanese.

How should we classify the dialogues? To solve the problem, in this paper, we analyze the level of dialogue by comparing sentences of Japanese test of Japanese Language Proficiency.

3. Analysis Method

In order to estimate dialogues' difficulties in animes, first of all, we have made the text subtitles of anime dialogues. Then we analyze the subtitles through two aspects: words and expressions.

According to the Japanese Language Proficiency Test (JLPT), there are four levels separated by JLPT in order to check JFL learners' learning degrees, Level 4 (L4) to Level 1 (L1). L4 is the primary, L1 is the most difficult. Each level has its own words and expressions need to be grasped. We choose the JLPT as the established standard in evaluating anime dialogues' difficulties.

3.1 Japanese Words of JLPT

JLPT has formulated different word lists for each level, JFL learners need to study those words when in corresponding levels. The numbers of words in each level shows in Table 1. Examples of word in each level are shown in Table 2.

The way we analyze the words in anime dialogues is:

- 1) We use a morpheme analysis soft (Igo) to separate words in anime dialogues.
- 2) We use the separated word to match with the identical one from the JLPT word lists and count up.
- 3) We make the count up result with different anime dialogues.

3.2 Japanese Expressions of JLPT

JLPT also has formulated different expression lists for each level. Not like the words, the study of Japanese expressions is mainly started from L3. JLPT has expression lists of L1, L2 and L3. Like the analysis that we have done in the past [6], the numbers of expressions in each level are showed in Table 3. Examples of expression in each level are shown in Table 4.

Table 1: Numbers of Words in Each Level

Level	L4	L3	L2	L1
Number of Words	737	685	3636	2962

Table 2: Examples of word in each level of JLPT

Level	Examples of word
L4	青い, 行く, 後ろ, 映画館, 同じ, 学校, 可愛い, 果物
L3	挨拶, 意見, 美しい, 選ぶ, 贈り物, 会場, 季節, 比べる
L2	愛情, 医師, 失う, 液体, 応用, 片道, 記憶, 苦情
L1	愛想, 遺跡, 団扇, 英雄, 疎か, 介護, 禁じる, 空腹

Table 3: Number of expressions in each level of JLPT

Level	L3	L2	L1
# of expressions	108	177	105

Table 4: Examples of expression in each level of JLPT

Level	Examples of expression
L3	～く\になる、～だろう、～ている
L2	～かのようなだ、～恐れがある、～にしても
L1	～にあって、～ながらも、～かたわら

Table 5: Information of anime subtitle texts

Anime Name	Number of episodes	Total number of rows in subtitle texts
Hyoka	23	9221
Fate Zero	24	7394
K-ON 2	24	10405

To consider the Japanese expressions is also important. In a realistic Japanese dialogue, words and expressions that native people use depend on the dialogue situation. For example, when the dialogue happened between friends or classmates, the words and expressions peoples would choose to use will be more easy-going than the sentence happened in the service industry. So we choose to analyze words and expressions separately, in order to see the possible distinctions between them. The way we used to analyze the expressions is:

- 1) We made lists of Japanese expressions in each level with regular expression's format.
- 2) We use the regular expressions list to match with the anime dialogues and make statistics.

3.3 Previous JLPT Tests

As the standard comparison data, we choose the real JLPT tests in the past. The JLPT tests consisted of three parts: Listening, Sentence Element and Reading. The latter two are mainly examine written abilities like vocabulary accumulation and grammar usage. Because of the goal of our research is to help with improving learners' conversation abilities, we mostly pay attentions to the Listening part which owns lots difference of dialogue scenes and mainly examine the learners' conversation capabilities.

First, we collect the original texts of every levels' listening tests in every year. Second, we make statistics of the words and expressions appeared in these texts with levels. Then we use the statistics from the same aspect of anime dialogues to compare with the realistic listening tests, words and expressions. Through these two aspects of analysis, we could evaluate the degrees of difficulty in different anime dialogues.

3.4 Analyzed Subtitle Texts of Anime Dialogues

The information of subtitle texts we used in this study is shown in Table 5.

We choose these three animes because the episode numbers of them is near, but the genres of story is different. The "Hyoka" is mainly about a mystery story, "Fate Zero" owns more battle scenes, and "K-ON" is completely a story in school life. We'd like to have a look at whether the degrees of words and expressions in dialogues have differences between the genres of story.

4. Discussion

For now, we have counted up all the words and expressions appeared in anime dialogue with different levels content, episode by episode. Table 6 is a part of results of anime “Hyoka” about front 10 episodes in its dialogues:

The Non of words means the number of nouns that could not match any word of all levels’ list in the dialogues. They maybe the name of characters, proper nouns or some words beyond the range of JLPT test. The averaged rate of Non words was 15.1%. We try to estimate the words levels by other methods.

Through the results in Table 6, we could found either words or expressions appeared in anime dialogues were concentrated in low degrees, especially in L3 and L4 levels. The average rate of words in L4 is 42.6%. We suspect that the results were obtained because of three reasons below:

(1) The words and expressions in high levels are more used with written abilities than conversation abilities. As the conversation-based anime, dialogues in anime would more likely choose the oral usage of words or expressions in it. For example, the content in news report likes:

「もっとも、国が高等教育にかかる費用すべてを負担した場合、年間5兆円かかるとの試算がある。」

The same meaning of the sentence when people speak to each other may become:

「でも、もし国が高等教育にかかる費用を全部負担したら、一年で5兆円くらいかかると思う。」

The 「場合」 and 「試算」 are high level words that would usually used in written sentences. Those words are not often used in conversations. When in conversation the words would change to easier ones: 「したら」 and 「思う」.

Table 6: Statistics of words & expressions in Hyoka

Episode	L4	L3	L2	L1	Non	
1	Words	888	274	461	238	378
	Expressions	none	276	135	23	-
2	Words	843	203	415	157	283
	Expressions	none	267	109	15	-
3	Words	888	229	379	160	269
	Expressions	none	280	108	13	-
4	Words	839	241	452	240	292
	Expressions	none	250	121	13	-
5	Words	912	260	410	230	369
	Expressions	none	294	116	10	-
6	Words	902	318	420	163	285
	Expressions	none	340	130	14	-
7	Words	811	199	382	161	286
	Expressions	none	248	103	20	-
8	Words	674	217	311	170	239
	Expressions	none	242	100	15	-
9	Words	886	270	469	223	367
	Expressions	none	280	146	19	-
10	Words	780	275	406	208	263
	Expressions	none	241	102	20	-
Rate of Words		42.6%	12.2%	20.7%	9.15%	15.1%
Rate of Expressions		none	68%	28%	4%	-

(2) The dialogues in animes are normally about the daily life conversations, like between friends or classmates. The usage of words and expressions in this type of conversations also would not be very complicated. Here is an example dialogue appeared between tourist guide and customers:

「お客様、右手をご覧ください。マルコポーロの世家を復元した建物でございます。」

But the dialogue between friends in “K-ON 2” usually likes:

「見てみて！すごいネジ！」

The first sentence is more complicated than the second sentence because the first sentence is composed for people who may not share all contexts. The second sentence is less complicated because people in the conversation would share all contexts so they can understand the intention and the meaning even if information is lacked in the second sentence.

(3) The genre of anime also seemed to cause the different emphases in dialogues’ choosing. The anime which tells a general daily life story would more likely have a relaxed conversation environment, so that the dialogues in it also will be more easier. Besides, the anime which owns a complex background may need more contents in helping with explaining the story. For example, when the character explained what the exact background story is in “Fate Zero”, it says:

「にわかには信じがたい話だとは思いますが、あらゆる時代、あらゆる国の英雄が現代に蘇り。」

The sentences like this are rarely used in daily conversations.

As the comparison, Table 7 and Table 8 show a part of results of anime “Fate Zero” and “K-ON 2” with front 5 episodes.

Table 7: Statistics of words and expressions in Fate Zero

Episode	L4	L3	L2	L1	Non	
1	Words	1306	476	814	387	775
	Expressions	none	411	268	38	-
2	Words	513	145	300	119	223
	Expressions	none	165	85	12	-
3	Words	600	184	367	167	270
	Expressions	none	208	131	32	-
4	Words	383	115	271	140	265
	Expressions	none	136	90	22	-
5	Words	381	110	259	126	289
	Expressions	none	117	94	32	-
Rate of Words		35.2%	11.2%	22.0%	11.2%	20.1%
Rate of Expressions		none	56%	37%	7%	-

Table 8: Statistics of words and expressions in K-ON 2

Episode	L4	L3	L2	L1	Non	
1	Words	601	219	248	86	216
	Expressions	none	187	33	14	-
2	Words	796	229	318	106	255
	Expressions	none	241	85	15	-
3	Words	637	219	313	97	231
	Expressions	none	224	59	8	-
4	Words	690	159	295	90	206
	Expressions	none	170	56	11	-
5	Words	648	193	216	49	148
	Expressions	none	197	38	9	-
Rate of Words		46.9%	14.2%	18.7%	5.2%	14.8%
Rate of Expressions		none	77%	20%	3%	-

The reason why numbers of words and expressions are much larger in the first episode of Fate Zero is mainly because the duration of the first episode is much longer than the others. A normal episode of anime is about 23 minutes, but the first episode of Fate Zero has nearly 48 minutes.

Through these two results, the rates of word and expressions in each level were different to each other. The rates of words in each level were 35.2%(L4), 11.2%(L3), 22.0%(L2), and 11.2%(L1) in "Fate Zero." The rates in "K-ON 2" were 46.9%(L4), 14.2%(L3), 18.7%(L2), and 5.2%(L1). The overall difficulty of K-ON is lower, because the conversations in K-ON are almost about relaxed daily dialogues.

The statistics also show the anime dialogues contained every level's words and expressions. This phenomenon may tell, even in one same episode, the degrees of difficulty of different scenes also could have large differences. Therefore, we also need to separate the scenes and consider their own degrees. If we want to utilize the animes in helping JFL learners with training of conversation abilities in the future, the estimate of scenes' degrees also would be important.

On the other hand, in the JLPT listening tests, there are also some types of conversation scenes in the dialogue. Such as asking a question, explaining something or just making greetings to other peoples. When JFL learners fail or have weakness in some types of practical dialogue scenes, they can make specialized trainings through all the same types of scenes appeared in the animes. The classification for each anime scenes also would be a challenge.

In fact, the word lists and expressions lists were formulated by JLPT tests are just for the standard Japanese language study. There are also lots of oral-flexible expressions and usages in the daily conversations of anime dialogues. To distinguish them from the anime scenes or just judge them from the dialogues also could help the JFL learners to understand the Japanese conversations more deeply. For the same reason, although we have matched the words or expressions in exact JLPT levels, the degrees of conversations content would not be easily confirmed. Not like the reading contents or standard Japanese sentences or sections. The difficulties for them would be more fixable. That's also why we need to consider the past tests' content as comparison data.

5. Conclusion

In this research, we put forward a method for estimate the Japanese anime dialogues' degrees of difficulty, and match with the existing proficiency tests' content. The estimation method through two aspects of investigation, words and expressions. The work includes three phases of content:

(1) The words and expressions appeared in anime dialogues were counted up separately. Words have four levels lists and expressions have three levels lists.

(2) The past JLPT listening tests were chosen as the standard comparison data. The goal of this research is to provide a corresponding level degree with the JLPT test for different anime dialogues inputted.

(3) The different scene of dialogues also need to be subdivided. In order to match with the JLPT tests content more exactly and clearly, the dialogue scenes in animes will be divided into several

fragments before estimate the level degrees of different fragment scenes. Also we could find out the similar scenes with JLPT tests content from the anime dialogues.

For now, we have done the first step, and have made a uniform analysis programs in doing words and expressions' statistics. We could use this method to analyze any new dialogue content in animes and to make statistics for JLPT tests' texts. Next, we will continue to arrange the content of past JLPT listening tests and give the comparison reference, to finish the evaluation of anime dialogues' level degree.

References

- [1] JAPAN FOUNDATION, Survey Report on Japanese-Language Education Abroad 2015, November, 10, 2016, URL: <https://www.jpf.go.jp/j/about/press/2016/dl/2016-057-1.pdf>
- [2] Matsumoto H. Peak learning experiences and language learning: A study of American learners of Japanese. *Language, Culture and Curriculum*, 20(3): 195-208, 2007.
- [3] Kumano N, Hirokaga M. An Examination of "ANIME & MANGA": Regional Information Overseas and Japanese-Language Materials. *JAPAN FOUNDATION Japanese Education Bulletin*, 2008 (4): 55-69 (in Japanese).
- [4] Manion A. Discovering Japan: Anime and Learning Japanese Culture. Unpublished master's thesis. East Asian Studies Center, University of Southern California, Los Angeles, CA. URL:<http://www.chanpon.org/archive/manionthesis.pdf>,2005.
- [5] Yazaki M. Development of "Anime-de Nihongo" Japanese Learning Activities with Anime: Examples from the Classes for Young JAL Learners. *Bulletin of the Shizuoka University International Center*, 5: 57-74, 2011 (in Japanese).
- [6] Junjie S, Nishihara Y. Analysis of Japanese Animation Dialogue Level using Japanese Expressions. *The Fifth Asian Conference on Information Systems*, Oct, 2016.