

会話属性により分類された アニメ会話をを用いた 日本語聴解テスト練習支援での 妥当性考察

第9回コミック工学研究会

立命館大学大学院情報理工学研究科

NI, YANGDI Shan, Junjie 韓 毅弘 西原陽子

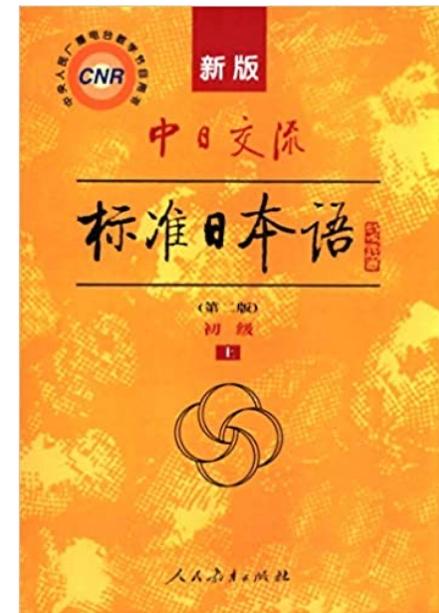
研究の背景と目的

研究
背景

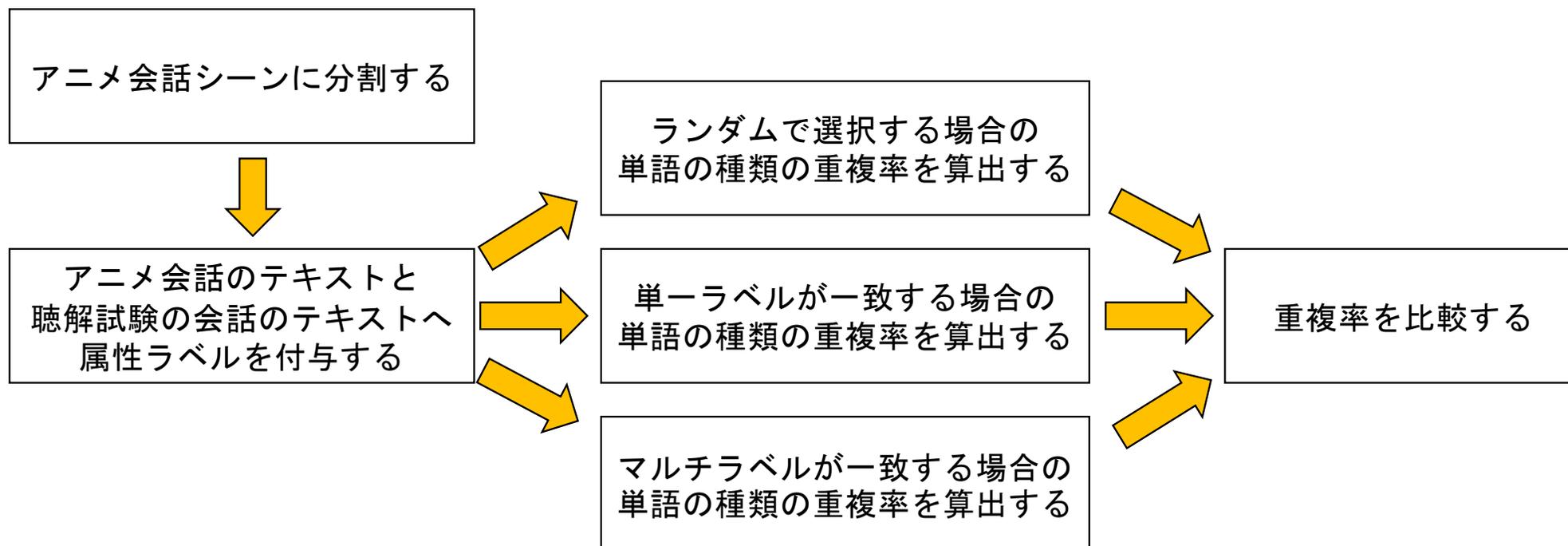
研究
目的

- 外国の日本語学習者が増えている
- 日本語の聴解テストの練習にアニメ会話会話を提供することは役に立つの既存研究がある
- 単語や表現の難易度によって、アニメ会話を提供する既存研究があったが、話者たちの関係や場所などの属性によって提供する研究は少ない。

日本語の聴解テストの練習に属性が一致するアニメ会話を
使用利点を調査する



分析手法の流れ



アニメ会話のシーンごとの分割

Shanらの方法で発話の終了時刻と次の発話の開始時刻の間の無音区間の長さを用いてアニメ会話をシーンごとに分割する

開始時刻

終了時刻

```
0:11:36.63,0:11:38.54,見たか てめえら～  
0:11:38.70,0:11:39.69,バカな...  
0:11:40.35,0:11:43.02,悪い ボールをなくしてしまった  
0:11:43.09,0:11:45.70,いいんだよ 一周回ってこい  
0:11:49.32,0:11:51.22,すごい強そうですね  
0:11:51.39,0:11:53.85,ああ さすがもと甲子園球児だ  
0:12:10.08,0:12:10.88,大丈夫？  
0:12:10.92,0:12:11.66,オッサン  
0:12:12.06,0:12:13.66,すまん 大丈夫か？  
0:12:14.02,0:12:15.42,テメェ わざとだろ  
0:12:15.46,0:12:17.73,いや 不可抗力だって  
0:12:18.42,0:12:21.45,どうする こっちの選手 誰か貸そうか？  
0:12:21.61,0:12:24.58,バカヤロウ それじゃ意味がねえ  
0:12:24.66,0:12:27.99,負けられねんだよ この町の威信にかけても  
0:12:30.01,0:12:31.59,ピッチャー交代  
...
```

- 日本語聴解テストの音声の平均時間は大体120秒くらいである
- スクリプト内の発話の終了時刻と次の発話の開始時刻の時間の差は6.5秒でアニメ会話のシーンの平均長さが120秒程度で分割する

アニメ会話と 聴解試験の会話のテキスト

- アニメは分割された「Fate Zero」, 「氷菓」, 「CLANNAD AfterStory」のテキスト1222件である
- 聴解試験は2001年から2008年までの四級のあった日本語能力試験90件である

	FateZero	氷菓	CLANNAD AfterStory
テキスト数	416	347	459

	L1	L2	L3	L4
テキスト数	30	20	20	20

怒るな怒るな
 体内の刻印虫を刺激すれば
 虫が貴様を食い潰してしまうぞ
 まあ それでもわしの見立てでは
 貴様の命はもってあと一月ほどだろうな
 充分だ
 何じゃと
 それで充分だと言ったんだ
 雁夜一年堪えた褒美じゃ
 貴様に相応しい聖遺物を見つけておいたわ
 父の親切を無にするでないぞ

アニメ「FateZero」の例

A: ねえ、経済言論の先生って、どう？
 B: あ、あの女の先生ね。素敵だよ。
 A: そういうことじゃなくて...
 B: あ、失礼。そうだね。説明は分かりやすいし、親切だよ。
 ただ、早口で、黒板の字が小さくて、時々ノートが取れないんだよね。
 A: へえ。それは困るね。そういう時はどうするの？
 B: 最初は、友達に聞いたり、自分で調べたりしてたんだけど、最近では授業の後先生に直接聞いてるよ。
 。 。 。

日本語聴解テストL1の例

アニメ会話テキストへの属性ラベルの付与

コーパスの分類方法を参考した上（日本語日常会話コーパス）、筆者一人で会話の内容、会話が行われた場所、話者たちの関係性三つの属性を付与する

- 会話の内容についてはは11種類、「説明」「深刻な話」「頼む」などがある
- 会話が行われた場所については30種類、「学校」「室外」「家」などがある
- 話者たちの関係性については15種類、「友人」「クラスメイト」「敵」などがある

ぎりぎり間に合ったではないか
聖杯に選ばれたということは
貴様もそれなりの術師として認められたということだ
ひとまずは褒めてっとるぞ雁夜
じゃがな
無様な姿よの
はら 左足はまだ動くのか

アニメ「Fate Zero」の会話シーンの例
話者たちの関係性が家族、場所が家、内容が説明

アニメ会話テキストへの属性ラベルの付与

ラベル	会話テキストの数
学校	418
室外	406
家	163
部屋	119
その他	108
街	103
店	92
神社	21
ホテル	4
図書館	3

ラベル	会話テキストの数
友人	504
学校のクラスメイト	450
敵	245
その他	198
上位関係がある人	127
親子	124
自分	120
仲間	105
恋人	72
夫婦	58

ラベル	会話テキストの数
説明	397
深刻な話	349
その他	244
依頼	225
相談	223
雑談	219
約束	184
冗談	165
紹介	151
皮肉	142

聴解テキストへの属性ラベルの付与

- 3種類の属性の上位3つのラベルを選定し、聴解テキストに付与する

A: ねえ、経済言論の先生って、どう？
B: あ、あの女の先生ね。素敵だよ。
A: そういうことじゃなくて...
B: あ、失礼。そうだね。説明は分かりやすいし、親切だよ。
ただ、早口で、黒板の字が小さくて、時々ノートが取れないんだよね。
A: へえ。それは困るね。そういう時はどうするの？
B: 最初は、友達に聞いたり、自分で調べたりしてたんだけど、最近は授業の後先生に直接聞いてるよ。そのほうが早いし、それに先生と話せるし。
A: あ、そう。実は、来年私もその授業を受けたいと思ってるんだ。
また詳しく教えて、授業の内容について。
B: うん。いいよ～。

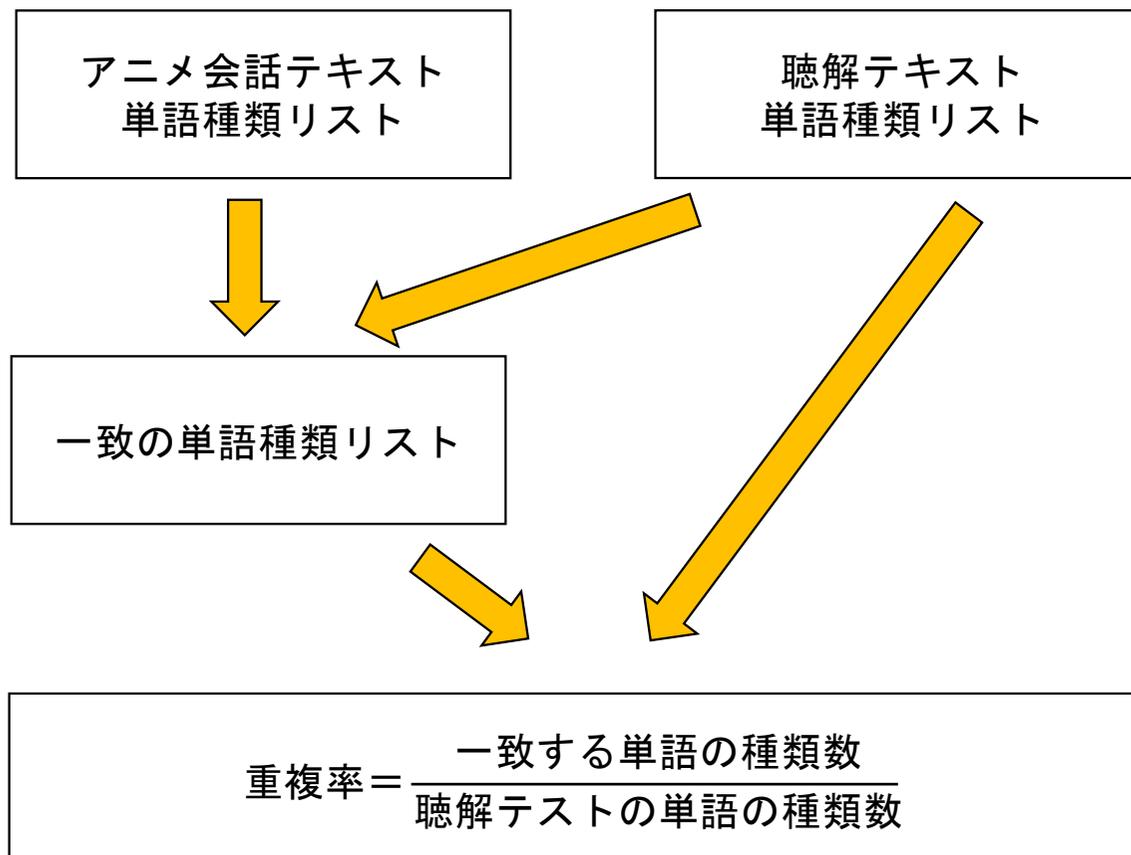
属性は3種類

- 話者たちの関係性
(今回は「友人」、「クラスメイト」、「敵」)
- 場所
(今回は「学校」、「家」、「室外」)
- 内容
(今回は「説明」、「頼む」、「相談」)

日本語聴解テストL1の例
話者たちの関係性が友人とクラスメイト、場所が学校、内容が相談

単語の種類重複率の算出

1. 形態素解析して、聴解テキストの単語種類数を計算する
2. 属性が一致するすべてのアニメ会話テキストの単語種類数を計算する。中に聴解テキスト一致する単語の種類数を計算する
3. 2.と同じ数のアニメ会話テキストをランダムで選択し、単語種類数を計算する中に聴解テキスト一致する単語の種類数を計算する
4. 重複率 = $\frac{\text{一致する単語の種類数}}{\text{聴解テストの単語の種類数}}$ の公式で属性が一致とランダムの場合の単語種類の重複率を計算する



重複率計算具体例

アニメ対話 : 「お嬢様とか、やめてくれよ。」

↓ 形態素解析

単語リスト : [お嬢, 様, とか, やめる, くれる]

聴解テスト会話 : やめてくれるといいけど。

↓ 形態素解析

単語リスト : [やめる, くれる, いい, けど] (合計 4 単語)

重複率 = $2 / 4 = 0.5$

単一ラベルの分析結果

ラベル	アニメ	聴解	重複単語	単語重複率	ランダムな場合の重複率
説明	6168	282	217	0.769	0.727
頼む	4503	187	148	0.791	0.770
相談	4323	341	246	0.721	0.712
友人	6020	255	203	0.796	0.784
クラスメイト	5535	169	148	0.875	0.840
学校	5342	197	172	0.873	0.862
室外	4519	220	167	0.759	0.822
家	1639	212	143	0.674	0.698
平均	4756	233	181	0.783	0.777

- 単一ラベルが一致するアニメ会話テキストと聴解テキストの平均単語重複率は**0.783**
- ランダムアニメ会話テキストを選択する場合は平均単語重複率は**0.777**

マルチラベルの分析結果

マルチラベル	アニメ	聴解	重複単語	単語重複率	ランダムな場合の重複率
説明+友人	4783	204	169	0.828	0.779
説明+学校	4540	84	68	0.809	0.761
友人+学校	5155	214	183	0.855	0.817
学校+クラスメイト	5177	143	125	0.874	0.818
友人+相談	3695	253	200	0.790	0.747
友人+室外	1605	184	120	0.652	0.608
学校+相談	3467	224	185	0.825	0.794
平均	4073	187	150	0.805	0.761

- マルチラベルが一致するアニメ会話テキストと聴解テキストの平均単語重複率は**0.805**
- ランダムアニメ会話テキストを選択する場合は平均単語重複率は**0.761**

考察

	平均単語重複率	ランダムな場合の平均重複率	上がる程度
単一ラベル	0.783	0.777	0.005
マルチラベル	0.805	0.761	0.044

- 属性が一致するアニメ会話テキストの方が、ランダムに選択されたアニメ会話テキストよりも単語の重複率が高くなる
- 一致する属性の数が1つから2つに増えるほど、重複率が高くなる

本研究の結論

- 日本語の聴解試験の練習にアニメ会話を使用することの妥当性を分析した
- **ランダム**にアニメ会話を提供するよりも、**属性が一致**するアニメ会話を提供する方が単語が重複する会話になる
- **単一ラベル**が一致するよりも、**マルチラベル**が一致する方が単語が重複する会話になる。

今後の課題

- 他の論文を参考し、さらに合理的な会話テキストの分類方法を探す
- 属性が一致するアニメ会話シーンの分類を行う機械学習器を構築する

