

複数言語コードを含む発話転写と話者分離 : Whisper+Pyannote.audio による自動音声認識の高度化

徐 勤¹¹京都大学文学研究科
xu.qin.4f@kyoto-u.ac.jp砂岡 和子²²早稲田大学政治経済学術院
ksunaoka@gmail.com

概要

生成系 AI を取り込んだ自動音声認識技術 (Automatic Speech Recognition : ASR) は飛躍的進歩をとげつつあるが、頻繁に Code-switching (CS) が起きる自然発話の ASR はまだ課題が残る。本文は、CS 発話の典型である外国語授業の 3 つのケースを素材とし、Whisper の最新版モデル large-v3 でテキスト転写を行い、更に Pyannote.audio を用いて多人数 CS 発話の話者分離 (Speaker Diarization : SD) を行った。large-v3 適用のさいのパラメータ設定、および合成音による SD の結果、従来の手法に比べ、日本語と中国語 CS 発話の ASR 精度が大きく改善した。

1 はじめに

人の移動や情報の国際化に伴い、生活や仕事の場で多言語化が進行している。CS 発話の自然言語データの収集には、言語別・発話者別の transcription が不可欠であるが、手作業での書き起こしには膨大な時間と当該言語の専門知識を必要とする。自動でテキスト変換できるツールやサービスを使えば、CS 発話を迅速かつ低コストでテキスト化でき、多言語機械翻訳や会話分析用コーパスデータ構築の大きな支援手段となる。AI を取り込んだ自動音声認識技術は、有力な代替ツールとして期待が大きい。

本研究は CS の多い外国語授業での自然発話を分析対象とし、これを Whisper で書き起こし、Pyannote.audio を用いて話者分離することで、複数言語コードを含む音声の自動転写の課題解決を目指す。

2 Whisper の音声認識技術

Whisper は、2022 年 9 月 22 日に OpenAI が開発し

た汎用音声認識ツール^①で、68 万時間の多言語音声データで学習・訓練された。Whisper は、transcribe メソッド^②で音声ファイル全体を読み込み、最初の 30 秒間の音声チャンクで言語の種類を検出する[1]。利用可能な学習済みモデルサイズは tiny, base, small, medium, large の順に 5 種類あり、いずれも多言語の書き起こしに対応する。このうち、large モデルは最も認識精度が高いが、その分、より多くのメモリと計算時間を要する。Whisper には、2022 年 12 月に large-v2 モデルが追加され、2023 年 11 月には large-v3 が登場し、性能がさらに向上した。

2.1 Whisper large -v3 の精度

large-v3 は、large-v2 より ASR の精度と性能が向上し、多種類の言語に対し、認識の単語誤り率 (WER: word error rates)、または文字誤り率 (CER: character error rates) が全体的に低減した^③。102 言語を含むデータセットである Fleurs^④を用いた各言語の WER や CER 評価では、large-v3 モデルにおいて、スペイン語、イタリア語、英語の WER はそれぞれ 2.8%、3.0%、4.1% であり、large-v2 モデルの 3.0%、4.0%、4.2% に比べ、誤り率がおよそ 1 割~2 割低減している。一方、日本語と中国語の CER も、それぞれ large-v2 の 5.3% と 14.7% から、large-v3 の 4.9% と 7.7% に減少したが、両言語とも欧米語のそれより誤り率はやや高めの数値で推移している。これは言語種によりシステムの訓練度に差があり、日本語と中国語のデータ量は欧米語より相対的に少ないことに起因する。また両言語とも漢字を使用し、単語分ち書きをしないことも分析精度に影響を与えている。

2.2 ケース 1 : 中日 CS 発話の ASR

Whisper large -v3 の精度を、A 大学の中国語初級授

^①<https://book.st-hakky.com/data-science/whisper-transcription/>

^②<https://dev.classmethod.jp/articles/whisper-how-to/>

^③<https://github.com/openai/whisper>

^④<https://paperswithcode.com/dataset/fleurs>

業の資料で検証した。本授業は、計 52 名の学生が出席する大規模教室で行われ、録画元音はノイズが多く、日本語母語話者である教員（女性）の発話課題も散見したため、Whisper の転写精度は約 40%に留まった[2]。そこで全 90 分授業の内、31~40min 間につき、同一教員が無雑室で再録を行い、その音声を再び Whisper で転写した結果、認識精度は約 80%へと倍増した。教員の日本語と中国語の発話率は約 1:1 である。ノイズ除去と教員の発音矯正効果と考えられる[2]。本実験では、無雑音室での再録音声の内、中国語の使用が多い発話区間 (32:10~35:35 間の 3 分 25 秒) のデータを対象に、large-v2 との転写精度の比較実験を行った。

ASR システムのパフォーマンスは、通常、文字誤り率 (CER) によって測定される[3]。本実験では、Whisper の ASR コードに verbose=True[®]を設定し、雑音低減や言語コード、ファイルサイズ調整などパラメータを変え、時間帯ごとの発話内容を高い精度で認識できるようにした。結果は付録 1 に見るように、large-v3 は、large-v2 より、認識文字総数 (句読点と記号を除く、以下同様) が約 2 割増え、中国語の認識文字数は倍増した (検出総字数 large-v2 : 168 字→large-v3 : 203 字, 中国語 52 字→100 字, 日本語: 106 字→103 字)。さらに、短時間に反復発音される単語の文字転写に関しては、large-v2 は 1 単語しか生成できないが、large-v3 は全反復回数分を書き起こした (例付録 1 : large-v3 “报纸 报纸 报纸”)。以上から、中国語授業における日中 CS 発話に対し、large-v3 の認識精度は large-v2 よりも大幅に向上したことが確認できた。

3 Pyannote. Audio による話者の分離

Whisper の ASR モデルは、誰が話したのか、話者を分離して書き起こしできない。Pyannote.audio[®]は、PyTorch 機械学習フレームワークに基づき、音声ファイル内の発音の特徴をもとに話者検出を行う。話者ごとに発話時間情報を区切り、同じグループにまとめることで[4]、話者分離 (Speaker diarization : SD) が可能になる[5]。ただし、Pyannote.audio は音声の文字転写機能がない。そこで、Whisper large-v3 の音声認識機能と Pyannote.audio の話者分離機能を組み合わせることで、どの時間帯で誰が何を話したのか、テキストとして書き起こす手法を試みた。

[®]<https://book.st-hakky.com/data-science/openai-whisper/>

3.1 ケース 2: CS 発話の話者分離 (原音)

データ 2 は、B 大学の中級中国語会話授業の冒頭で、教師 (女性、中国語母語話者) と学生 1 名 (女性、日本語母語話者) のやりとりを録音したものである。ASR 機能が破綻しないよう、発話者数を絞った録音環境とし、これを Whisper の large-v3 モデルで転写した。表 1 は冒頭 1 分半の転写結果である。本段落の発話総字数は 179 (character) で、うち日本語が 14%、中国語は 86%を占めている。Whisper の転写率は 7 割強で、発話中の学生や教師の短片語彙 (例: 嗯, 贵) や、日本語固有名詞 (例: 優月, 福井) など、認識に失敗した箇所があった。

表 1 データ 2 : 転写結果と手作業による校正

長さ: 1 分 30 秒 (総字数: 180; 中国語字数: 155; 日本語字数: 25)

好的 好的 那我们开始上课吧 同学们早上好 早上好 那我们先做一下口语练习 まずは会話の練習をやっていきましょう 中村優越(越→月) 到 中村優越(越→月) 你的毛衣真好看 你的毛衣真好看 你的衣服真好看 哪里哪里 嗯你在哪买的 我在 我在 福井 福井 啊 福井 福井ですか 我在福井买的 贵吗 贵吗 贵 贵 很贵 今天天气怎么样 今天天气很好 很好 嗯 很好 好的 中村優越(越→月) 问我一个问题 问我一个问题 噢 你 你的衣服很漂亮 谢谢 谢谢 谢谢你 你的衣服也很漂亮 谢谢

注: 下線は教師, その他は学生の発話。黒字の斜体表記は日本語, 黒字の太字表記は中国語。赤字は筆者が補充した Whisper の認識失敗箇所, 括弧内は転写正解。以下同様。

この Whisper 転写結果を、Pyannote.audio を用い、音声ファイルにおける話者分離を行った (図 1)。その結果、3 人の発話者が認識されたが、転写原文と照合すると、SPEAKER_01 と SPEAKER_02 はともに教師発話で、SPEAKER_00 は学生の発話である。また、発話が後半になるにつれ、話者認識に混乱が生じ、例えば最後の「谢谢」は学生の発話であるが、「SPEAKER_02」と誤って表示された。

[®]<https://github.com/pyannote/pyannote-audio>

```
print(start, end, speaker, text)
```

```
1.988 4.559 SPEAKER_02 好的,那我们从上课吧
5.146 5.916 SPEAKER_02 同学们早上好
6.665 6.949 SPEAKER_00 早上好
8.865 11.459 SPEAKER_02 那我们先做一下口语练习
11.801 11.821 SPEAKER_02 まずは会話の練習をやっていきましょう
22.798 23.34 SPEAKER_02 中村優越
28.438 28.719 SPEAKER_01 中村優越
29.424 30.498 SPEAKER_01 你的毛衣真好看
32.284 33.668 SPEAKER_01 你的衣服真好看
40.545 41.789 SPEAKER_00 你在哪买的
46.341 51.609 SPEAKER_00 我在福井买的
59.122 59.331 SPEAKER_02 贵吗
61.437 61.619 SPEAKER_02 贵吗
63.537 63.799 SPEAKER_00 很贵
64.949 66.039 SPEAKER_01 今天天气怎么样
67.81 69.299 SPEAKER_00 今天天气很好
69.625 69.875 SPEAKER_01 很好
70.506 70.931 SPEAKER_00 好的
72.047 72.374 SPEAKER_01 中村優越
73.51 74.158 SPEAKER_01 问我一个问题
75.068 75.856 SPEAKER_01 问我一个问题
81.073 83.499 SPEAKER_00 你的衣服很漂亮
85.374 85.938 SPEAKER_00 谢谢
86.14 87.118 SPEAKER_02 谢谢你
87.746 89.02 SPEAKER_02 你的衣服也很漂亮
89.222 90.379 SPEAKER_02 谢谢
```

図1 データ2：話者別の転写結果（一部）

3.1.1 ASR 精度の可視化（ケース2）

ASR 技術により，外国語授業における教師と学生の発話比率のデータを可視化できれば，教育の質向上の目安にできる．表2はケース2のデータを，Pythonで集計した結果である．授業中における学生と教師の発話文字数の比率は，Pyannote.audioと手作業による話者分離結果ともに約1対4で，発話回数の比率は約3対7となった．本授業の担当教員が，学生とのインタラクションを丁寧に行い，学生回答の繰り返しやフィードバックを与えていることが数値で可視化できる．ASRのチューニングや話者分離技術は，指標の可視化の有力な手段である．今後は手動計測に替え，自動表示プログラムの開発を目指す．

表2 データ2：話者別の集計

	Pyannote.audio		手作業校正	
	発話回数	発話総字数	発話回数	発話総字数
学生	8	33	11	40
教師	18	109	21	140
小計	26	142	32	180

3.2 ケース3: CS 発話の話者分離(合成音)

データ3は，データ2と同じ教師と別の学生（男

性）による応答の録音である．表3に Whisper と Pyannote.audio を組み合わせ，話者ごとに転写を行った．表4は表2同様，話者別の発話回数と文字数を Python で集計した結果である．

本段落の発話総字数は142で，うち日本語が13%を占めているが，学生や教師の断片的な日本語表現（例：髪の色）を，Whisper は認識しなかった．転写率は約8割で，“哦”“嗯”のようなフィラーを除き，それ以外の中国語は正しく認識された（表3）．

表3 データ3：転写結果と手動校正

長さ：1分29秒（総字数：142；中国語字数：122；日本語字数：19）

下一位 奥村 奥村 你头发的颜色真不错 髪の色
对 グッド 你头发的颜色真不错 谢谢 不客气
 你觉得今天天气怎么样 很好 很好 那昨天怎么样
 很好 哦很好 那你问我一个问题 質問 你昨天晚上
 上了什么 我昨天晚上吃了便当 你昨天晚上吃了
了什么 唐揚げ 唐揚げ 嗯 肉 肉 肉ですか 肉 肉
 肉 肉 你再说一遍 你昨天晚上上了什么 我昨天
晚上吃了肉 肉 很好 肉

表4 データ3：話者別の集計

	pyannote.audio		手作業校正	
	発話回数	発話総字数	発話回数	発話総字数
学生	4	14	10	43
教師	18	106	15	99
小計	22	120	25	142

```
print(start, end, speaker, text)
```

```
0.667 1.677 SPEAKER_00 下一位,奥村。
6.724 9.116 SPEAKER_00 奥村,你头发的颜色真不错。
20.503 22.416 SPEAKER_00 你头发的颜色真不错。
24.163 24.327 SPEAKER_01 谢谢。
25.0 25.163 SPEAKER_00 不客气。
26.262 28.658 SPEAKER_00 你觉得今天天气怎么样?
31.01 31.253 SPEAKER_01 很好。
32.061 33.879 SPEAKER_00 很好。那昨天怎么样?
37.224 37.633 SPEAKER_01 很好。
38.202 40.0 SPEAKER_00 很好。那你问我一个问题。
47.784 50.457 SPEAKER_00 我昨天晚上吃了便当。
52.705 54.477 SPEAKER_00 你昨天晚上吃了什么?
60.0 73.658 SPEAKER_01 我昨天晚上吃了肉。
75.347 75.367 SPEAKER_00 肉。
76.694 76.714 SPEAKER_00 肉。
77.735 77.755 SPEAKER_00 肉。
78.041 78.061 SPEAKER_00 肉。
79.082 79.102 SPEAKER_00 肉。
80.02 80.041 SPEAKER_00 肉。
81.0 83.658 SPEAKER_00 你再说一遍。你昨天晚上吃了什么?
84.383 86.96 SPEAKER_00 我昨天晚上吃了肉。
87.506 88.134 SPEAKER_00 肉。很好。肉。
```

図2 データ3：話者別の転写結果

次に、Pyannote.audio を使用して話者ごとに発話内容を転写した結果(図 2), SPEAKER_00(教師発話)と、SPEAKER_01(学生発話)の 2 人の発話者が認識された。前半部分の話者は正しく認識されたが、後半部分では 5 箇所話者認識が混乱を呈し、自動話者分離の精度は約 77%であった。このため、話者分離によって得られた学生対教師の発話文字数の比率 1:9 は、実際の発話文字数の比率 3:7 よりも低くなった(表 4)。SD 数値が実際と乖離した原因の一つに、原音声のノイズ混入が考えられる。そこで、Python の音声合成ライブラリーである Edge-TTS^①を使用して、表 3 の校正内容を元に、女性と男性の 2 種類の音声を設定することで、無雑音の対話音声ファイルを合成した。その上で、Whisper と Pyannote.audio を組み合わせて、ケース 3 の合成音声ファイルの書き起こしを行った(図 3)。

```
print(start, end, speaker, text)
0.303 1.495 SPEAKER_00 下一位, 澳村
2.501 4.937 SPEAKER_00 澳村, 你头发的颜色真不错!
6.022 6.042 SPEAKER_01 - 髮色?
8.624 8.645 SPEAKER_00 对!
9.843 10.232 SPEAKER_01 - 很好!
11.941 13.576 SPEAKER_00 你头发的颜色真不错!
14.845 15.09 SPEAKER_00 - 谢谢!
17.143 17.713 SPEAKER_00 不客气!
18.701 20.617 SPEAKER_00 你觉得今天天气怎么样?
21.861 22.269 SPEAKER_01 - 很好!
24.223 24.57 SPEAKER_00 很好!
25.542 26.495 SPEAKER_00 那昨天怎么样?
27.743 28.151 SPEAKER_01 - 很好!
30.122 31.014 SPEAKER_00 哦, 很好!
31.922 33.056 SPEAKER_00 那你问我一个问题!
34.529 34.549 SPEAKER_01 - 请问!
36.962 38.577 SPEAKER_01 你昨天晚上吃了什么?
40.621 42.237 SPEAKER_00 我昨天晚上吃了便当。
43.221 44.797 SPEAKER_00 你昨天晚上饭吃了什么?
45.882 46.086 SPEAKER_01 - 炸鸡!
48.202 48.406 SPEAKER_01 - 炸鸡!
50.723 50.743 SPEAKER_01 嗯!
52.203 52.224 SPEAKER_01 肉!
53.743 53.763 SPEAKER_01 肉!
56.122 56.346 SPEAKER_01 肉吗?
58.305 58.326 SPEAKER_01 肉!
59.665 59.685 SPEAKER_01 肉!
61.703 61.724 SPEAKER_01 肉!
63.022 63.043 SPEAKER_00 肉!
65.402 66.255 SPEAKER_00 你再说一遍!
67.161 68.677 SPEAKER_00 你昨天晚上吃了什么?
69.921 71.496 SPEAKER_01 我昨天晚上吃了肉!
73.501 74.455 SPEAKER_00 肉, 很好!
75.424 75.445 SPEAKER_00 肉!
```

図 3 データ 3: 合成音声の転写結果

転写の結果, 2 人の話者(SPEAKER_00 は教師発

話, SPEAKER_01 は学生発話)を区別し, 学生と教師の発話回数はそれぞれ 16 回と 18 回で, 発話内容の総字数は学生が 39 字, 教師が 97 字となった。学生と教師の発話文字数の比率は, 人間の自然発話および合成音声ともに, 学生対教師が約 3 対 7 となり, 合成音声により ASR の精度が向上したことを裏付けよう。後半部分では 3 箇所話者認識が混乱したが, 他の部分では話者は正しく認識され, 話者分離の精度も約 91%に向上した。いっぽう Whisper は, 断片的な日本語表現を中国語と間違え, 勝手に自動翻訳を行った(例表 3 と図 3: 日本語の「唐揚げ」を, 「炸鸡」と中国語に自動翻訳)。Whisper を始め対話型 AI の過剰サービスを回避することは難しい。

4 まとめと課題

本研究は外国語授業の CS 発話を素材とし, Whisper と Pyannote.audio を組み合わせて, 複数言語コードを含む発話音声の自動転写と話者分離を行い, 得られた発話内容に対し, Python でコードを書き, 話者ごとの CS 言語の使い分けと発話量の集計を行った(図 1, 2 参照)。さらに Edge-TTS を使用し, 無雑音の対話音声ファイルを合成し, 同様に ASR を実施することで, 精度の向上を検証することができた(図 3 参照)。各ツールの機能はまだ改善の余地があるが, Whisper の書き起こし機能と Pyannote.audio の話者認識機能を組み合わせれば, CS 発話における話者別の発話内容, および発話量の分析がかなりの精度で可能であることが分かった。外国語教師が, 授業内の CS 言語の配分や自身の発話量を把握するツールとして, 授業改善に役立てるなど, 具体的応用事例が考えられる。ASR 転写結果と比較することで, 発話者の発音の不明瞭や不正確な箇所を瞬時に検出することも可能になる。

今後の課題として, Whisper の複数言語コードの認識機能の更なる向上以外に, Pyannote.audio の発話者数増加に伴う話者分離機能の混乱, およびデータ処理能力の時間的減衰を如何に防ぐのか, システムの技術的課題の解決が望まれる。本研究はこれら予備的実験を基に, 今後言語別と発話者別の CS 発話分布, および発話テキスト変換精度を自動表示するプログラムの開発を目指す。

①<http://www.otupy.net/archives/43836875.html>

謝辞

本研究は JSPS 科研費基盤研究 C (21K00773) 「ハイフレックス型授業の相互行為検証と PC シミュレーションに基づく実践応用」研究代表者：砂岡和子の助成を受けている。

参考文献

1. RADFORD, Alec, et al. Robust speech recognition via large-scale weak supervision. *In: International Conference on Machine Learning*. PMLR, 2023. p. 28492-28518.
2. 砂岡和子, 徐勤. “生成系 AI と初級外国語学習者の Code Switching 発話「聴取」パフォーマンス”, 電子情報通信学会技術研究報告 (ET2023-23), pp.33-37, 2023.
3. ZHAO, Yun, et al. BART based semantic correction for Mandarin automatic speech recognition system. *arXiv preprint arXiv:2104.05507*, 2021.
4. STOLCKE, Andreas. Improving Diarization Robustness using Diversification, Randomization and the DOVER Algorithm. *arXiv preprint arXiv:1910.11691*, 2019.
5. BREDIN, Hervé, et al. Pyannote. audio: neural building blocks for speaker diarization. *In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020. p. 7124-7128.

付録

付録1 手作業および Whisper large-v2 と large-v3 の転写結果

黒字の斜体表記は日本語, 黒字の太字表記は中国語

<p>large-v2 (検出総字数: 168 字, 中国語: 52 字, 日本語: 106 字) はい 很好 那下面是第三課 第三課すごく少ないですね なんか 第一課より少ないですね はい 第三課 我們念一下 はい いきます 報紙 電子辞典 豆漿 法語 留学生 手機 謝謝 眼鏡 はい では次の人 第三課 ね 念的是 報紙 新聞ね 豆漿 あなるほど 豆漿も読んだ 確かに 豆漿 もう一個間に長いやつ 電子辞典 電子辞書ですね はい 電子辞典 法語 うん 法語 三声三声だから 発音するとき 二声三声で読まれます のでね</p>
<p>large-v3 (検出総字数: 203 字, 中国語: 100 字, 日本語: 103 字) はい 很好 那下面是第三課 第三課すごく少ないですね なんか 第一課より少ないですね はい 第三課 我們念一下 はい いきます 报纸 报纸 报纸 电子词典 电子词典 电子词典 豆浆 豆浆 豆浆 法语 法语 法语 留学生 留学生 留学生 手机 手机 手机 谢谢 谢谢 谢谢 眼镜 眼睛 眼镜 では次の人 第三課ね 念的是 报纸 新聞ね 豆漿 なるほど 豆漿も読んだ 確かに 豆漿 もう一個間に長いやつ 電子辞典 電子 辞書ですね はい 電子辞典 法語 うん 法語 三世三世だから 発音するとき 二世三世で読まれますので ね</p>
<p>手動転写 (総字数: 212 字, 中国語: 104 字, 日本語: 108 字) はい 很好 那下面是第三課 第三課すごく少ないですね なんか 第一課より少ないですね はい 第三課 我們念一下啊 はい いきます 报纸 报纸 报纸 电子词典 电子词典 电子词典 豆浆 豆浆 豆浆 法语 法语 法语 留学生 留学生 留学生 手机 手机 手机 谢谢 谢谢 谢谢 眼镜 眼镜 眼镜 はい では次の人 第三 課呢 念的是 报纸 はい 报纸 新聞ね 豆漿 あなるほど 豆漿も読んだ 確かに 豆漿 もう一個 あいだに 長いやつ 電子词典 電子辞書ですね はい 電子词典 法語 うん 法語 三声三声だから 発音する時 二声 三声で読まれますのでね</p>

付録1 の日本語訳 (Whisper の ASR 結果ママ)

はい とてもいいです 次は第3課です 第三課すごく少ないですね なんか 第一課より少ないですね はい
 第3課 読みましょう はい いきます 新聞 新聞 新聞 電子辞書 電子辞書 電子辞書 豆乳 豆乳 豆乳 フ
 ランス語 フランス語 フランス語 留学生 留学生 留学生 スマホ スマホ スマホ ありがとう ありがとう
 ありがとう メガネ メガネ メガネ はい では次の人 第3課ね 読んだのは 新聞 はい 新聞 新聞ね 豆
 乳 あなるほど 豆乳も読んだ 確かに 豆乳 もう一個 あいだに 長いやつ 電子辞書 電子辞書ですね
 はい 電子辞書 フランス語 うん フランス語 三声三声だから 発音する時 二声三声で読まれますのでね

付録2 データ2の日本語訳 (下線は教師; その他は学生の発話)

はい はい 授業を始めましょう 皆さん おはようございます おはようございます まずスピーキングの練
 習をしましょう まずは会話の練習をやっていきましょう 中村優月さん はい 中村優月さん あなたのセ
 ーターすてきですね! あなたのセーターすてきですね! あなたの服とてもきれい! いえいえ うん ど
こで買ったの? 私は 私は 福井 福井 あ 福井 福井ですか 福井で買いました 高かった? 高いですか
 高い 高かった とても高いのね 今日の天気はどう? いい天気です はい うん よくできました では 中
 村優月さん 質問してください 質問して ああ あなた 先生の服すてきです ありがとう ありがとう あな
 たの服もすてきです ありがとうございます

付録3 データ3の日本語訳 (下線は教師; その他は学生の発話)

次の人 奥村さん 奥村さん 髪の毛の色いいね 髪の色? そう グッド? いい色だね ありがとうございます
 す いやほんとに 今日の天気はどう? いい天気です そうだね 昨日はどうだった? とても良い天気で
 した ああ そうだね では質問して 質問? 先生は昨日の夜, 何を食べましたか? 昨日の夜は弁当を食べ
 たな では君は, 昨日の夕食, 何食べた? 唐揚げ 唐揚げ うん 肉 肉 肉ですか 肉 肉 肉 肉 もう一度言っ
 て 昨日の夜は何食べた? 昨日の夜は肉を食べました 肉 いいですね 肉