

津軽弁の音声認識

高橋 栄 作

Sound Recognition of the *Tsugaru* Dialect

TAKAHASHI Eisaku

要 旨

世界には、消滅の危機に瀕している言語（方言）が多数あるという。言語は伝統や文化を支え継承するのに重要な役割を担っているため、そのような言語を保存記録していく必要があると考える。言語を保存する手段として、「記録保存」と「継承保存」があるが、本研究では「継承保存」を容易にするための方法を示す。方言主流社会であっても、言語使用の環境についてはさまざまであるため、継承保存を容易にするために、機械学習の手法を用いる。研究対象者の言語音を収録して、サウンド・スペクトログラムの画像を作成し、画像認識を行う。画像認識の手法は、scikit-learn algorithm cheat-sheetに従いK-meansを用いる。サウンド・スペクトログラムの画像を正しくクラスタリングできるかを考察する。結果は、K-meansによるクラスタリングで、かなりの精度で画像を認識することができた。一連の手続きを自動化できれば、今後言語（方言）の「継承保存」が容易となり多文化・多言語共生が容易におこなえると考えられる。

Abstract

It is said that many languages (dialects) are now in danger of extinction. Since languages play an essential role in preserving the relevant traditions and cultures, we need to archive those endangered languages. Out of the two means of preserving a language, recording preservation and inheritance and preservation, this study focuses on the means to facilitate inheritance and preservation of a language. Because the environment to use of a dialect varies even in a local dialect-dominated society, we use machine learning to facilitate language inheritance and preservation. We collected the voice sounds of participants to make the spectrograms for image recognition. We used K-means clustering for the image recognition based on a scikit-learn

algorithm cheat-sheet. This study examined whether K-means clustering worked well to make clusters for each spectrogram. The results showed that the procedures could make clusters of each spectrogram in a high degree of accuracy. If we can automate a series of the procedures, language inheritance and preservation will be simplified, which will facilitate multicultural and multilingual societies.

I. はじめに

木部・山田（2019）は『『いま何もしなければ』なくなってしまう』のなかで、UNESCO（国際連合教育科学文化機構）は2009年、“Atlas of the World’s Languages in Danger”（『世界消滅危機言語地図』）を公表し、世界でおよそ2500の言語が消滅の危機にあり、日本でもアイヌ語を含め8つの言語がその中に含まれているとした。さらに日本人の生活の変化などが影響し、各地の言語・方言も消滅の危機に瀕しているという。また、言語の基本は意思の伝達であるが、言語が一元化してしまった場合の懸念を、2001年UNEP閣僚級環境フォーラムで、テプファー国連環境計画事務局長が「伝統、文化の継承を支えてきたことばを失うことは、自然の貴重な教科書を失うことに等しい」と述べたことを指摘している。よって、伝統文化を継承していくためには地域言語を守る必要がある。特に「方言主流社会（佐藤他2003）」では地域言語と共通語（東京方言）の役割の違いを認識し、多文化・多言語共生のための手段が必要であるという。¹⁾ 佐藤他（2003）は、方言の果たす役割が大きい方言主流社会では、方言や共通語の価値を考える言語教育は、個人の言語生活を内省するきっかけとして重要であることを指摘している。それでは、どのように方言を継承していけば良いのか。松本（2017）は、方言主流社会といっても、家族の構成員全てが土着ではなく、家庭内の方言使用についてもさまざまであるとしている。では、そのような環境で、方言を後の世に残すためにはどのような手段があるだろうか。木部・山田（2019）は、消滅の危機に瀕している言語の話者が健全なうちに言語資料（音声や映像など）を残すことにより、世代間での言語継承が途絶えてしまったとしても復活・復興が可能であると述べている。また、「記録保存」と共に世代間継承を再開・維持するための「継承保存」にも取り組んでいるという。²⁾ そこで、本研究では「方言の継承保存」をおこなうことができる一つの方法として、近年注目されている「機械学習」により方言の音声認識をおこない、音声の弁別を容易におこなうことができるシステムの構築を目指す。機械学習は、ある情報とある情報の関係を学習するアルゴリズムを指し、機械学習を用いれば、情報の分類を自動的におこなうことができるようになる。³⁾ 音声認識のプログラムを構築し、そのプログラムに「ある音声（方言）」を与えると、その音声がどのような音声であるのかを認識することができるプログラムの構築を目指す。そうすることにより、音声認識が可能となり、言葉を失うことなく継承していくことが可能になると考える。本

研究では、方言主流社会のひとつである青森県津軽地方の方言を分析し、その音声特色を認識するシステムを構築し、多文化・多言語共生のための手段の構築を目指す。本研究の調査、分析の方法は次の通り。津軽弁の音声データを集め、音声データのサウンド・スペクトログラムを取得する。そのサウンド・スペクトログラムの画像を用いて、機械学習をおこなう。データとして得た画像は、正解情報がないため機械学習の方法としては「教師なし学習」である。そこで、プログラミング言語Pythonでscikit-learnのK-meansでクラスタリングを行い、データを分類する。機械学習により、音声の認識を容易におこなうことができれば、方言の継承保存の際に簡単に音声認識ができ、多文化・多言語共生がおこなえることを提案する。

Ⅱ. 津軽弁

分析対象の津軽弁について考察する。図1にあるように青森県は大きく2つの区分に分かれる。津軽地方は青森市から西側に位置し、本研究では、弘前、黒石、平川市（旧南津軽郡）出身の研究対象者にご協力していただいた。⁴⁾

図1 津軽の位置とその区分



(出典：佐藤他2003)

津軽弁には、次のような特徴がある。

(1) 母音の無声化

- a. あたし→あタし
- b. くし→クシ (佐藤他2003)

*カタカナの表記がささやき声になる

(2) カ・タ行子音の有声化

- a. あける→あげる (ibid.)
- b. はた→はダ (ibid.)

(3) シとスの区別がつかない

- a. シチュー→スチュー
- b. してください→すてください

さらに松本（2017）は、津軽弁の難しさを示すエピソードとして次のことを紹介している。2009年度から始まった裁判員制度では、裁判員の中には県外の人（青森県外出身者）も含まれるため、供述調書に記載されている津軽弁が理解できないというのである。

(4) じゃんこ、だせ

“金を出せ（少額のお金）”

また、弘前大学が進める「弘大×AI×津軽弁プロジェクト、あなたの津軽弁、標準語にします」というプロジェクトがある。これは、弘前大学の医療関係者もまた県外出身者が多く、津軽弁がわからないため患者とのやりとりに困るので、津軽弁を標準語に翻訳するシステムの構築を目指しているとのことである（日本経済新聞電子版 2020/ 1 /28）。例えば(5)のような表現がある。

(5) はらあんべわるい/はらがにやにやする

“お腹の調子が悪い”

Ⅲ. 津軽弁の音声認識

弘前のような方言主流社会では、佐藤他（2003）が示すように方言を使って人間関係を保っているといえ、松本（2017）は、方言使用の理由を次のように述べている。①生活や地域社会で日常を細やかに表現できる。②地域への帰属意識を表明することができる。③親しさという心の距離を表せる表現である。以上から、方言主流社会において、方言は大切であり、かつ方言は自然の貴重な大切な教科書であると言える。伝統、文化を継承するため、難解な津軽弁を「継承保存」するにはどのような方法があるのだろうか。方言主流社会といっても家族全員が津軽出身ということはないと考えられる。家族を構成する母親が県外出身であったり、祖父が県外出身だったり、家庭内の方言使用については様々な組み合わせがある。そこで、本研究では、近年注目されている「機械学習」により津軽弁の音声認識をおこなう。音声認識が手軽に実施でき、運用されれば、津軽弁を後世に語り継ぐ「継承保存」が容易に可能となると考える。

本研究では、津軽弁の音声認識を次のようにおこなう。

- (6) 津軽弁音声データの収集、整形→音声データを画像データへ変換→画像データを数値化→学習モデルの構築→学習モデルの評価

Ⅲ（１）津軽弁音声データ収集と解析

弘前、黒石、平川市出身の研究対象者（30~40代の男性5名）にご協力していただき、津軽弁の音声データを収集した（分析資料を参照）。分析資料を音読してもらい音声をICレコーダ（RR-XS370, Panasonic）で録音した。音声データに含まれる雑音を取り除き（Audacity 2.2.2）、整形した音声を音声分析ソフトで分析した（Praat Version 6.1.09）。分析対象としたのは（7）の二つの単語であり、これらの語のサウンド・スペクトログラムを取得した。

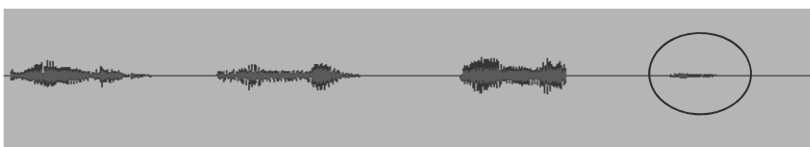
- (7) a. ハンデ
“だから”
b. ヘバ
“じゃあね/それならば”

研究対象者の発話音声をAudacityで読み込み（図2）、音声に含まれる雑音を除去した（図3）。

図2 Audacityで音声を整形



図3 雑音の除去 ○印が雑音

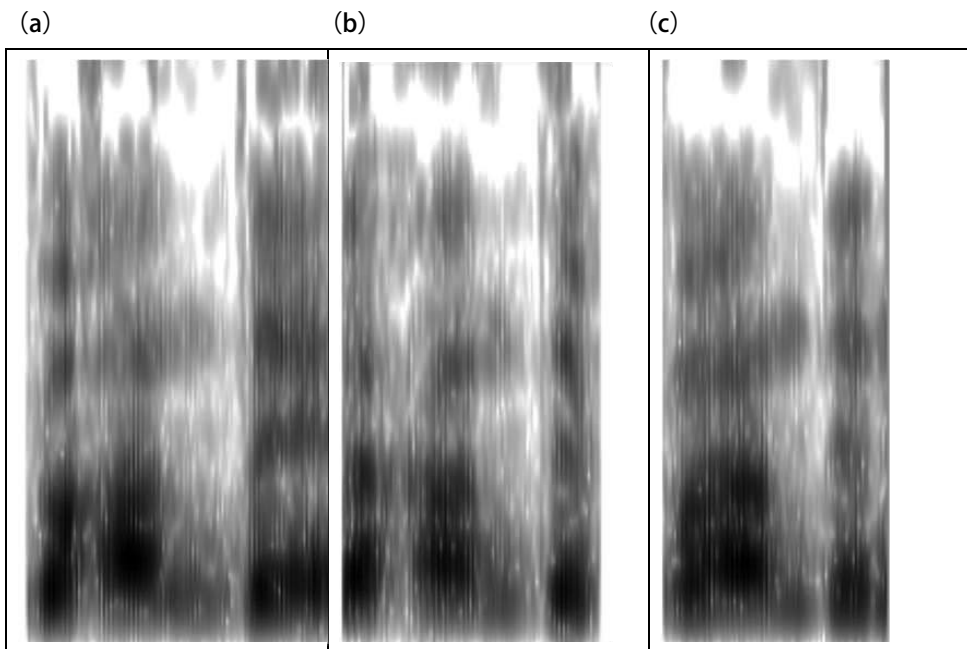


雑音を除去した音声ファイルを音声分析ソフトで分析し、サウンド・スペクトログラムを生成する(図4)。⁵⁾ 機械学習を行うため、(7)の音声の該当するサウンド・スペクトログラムを画像として切り出した(図5)。これにより得た画像は、「ハンデ」「へバ」について、それぞれ5人分の15枚の画像である(合計で30枚)。画像は、ピクセルで構成されている。ピクセルは、光の強さを0(最も暗い)~255(最も明るい)のピクセル値で表す。本研究で扱うピクセルはモノクロ(グレースケール)で構成されているが、計算を容易にするために画像を16ピクセルに整形した。

図4 Praatでの音声分析



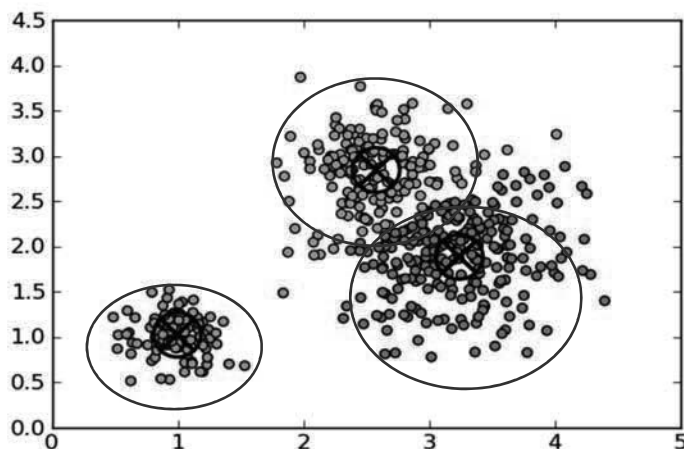
図5 サウンド・スペクトログラム
ある研究協力者の3つの「ハンデ」のサウンド・スペクトログラム



IV. 分析

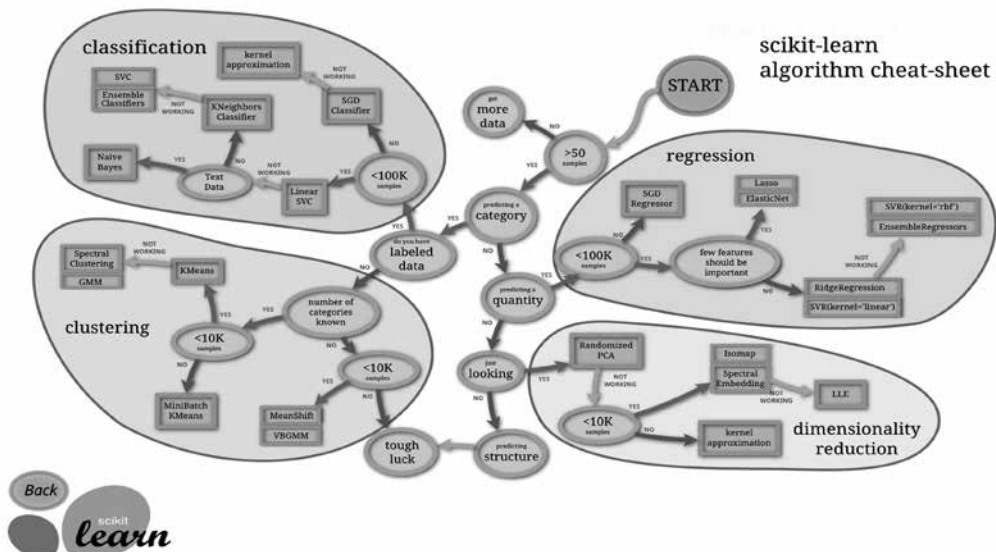
分析に利用する画像には正解となるラベルがつけられていないので、機械学習の際には「教師なし学習」を用いる。教師なし学習では、訓練データに対する正解ラベルがなく、訓練データ自身からそれが持つ規則性や、構造を見つけることを目的とし、この手法には、入力されたデータの類似性に基づきデータをグループ分けする「クラスタリング」がある（松田2019）。そこで、クラスタリングの手法はいくつかあるが、本研究ではK-meansを取り上げる。K-meansはクラスタ数を予め指定し、斬新的にクラスタ化を進めていく方法である（鶴他2016）。クラスタ化の手順は次の通り、① 各データを何らかの手段でクラスタに割り振る。② クラスタ毎の中心を計算する。③ 各データからクラスタ中心への距離を求め、もしデータが最も近いクラスタ以外に属しているようなら、データの所属を最も近いクラスタに変更する。②③を繰り返しクラスタを形成する（鶴他2016）（図6）。K-meansを用いる際にscikit-learn algorithm cheat-sheetを参照する（図7）。scikit-learnとは、プログラム言語Python向けの機械学習フレームワークで、機械学習で使われる様々なアルゴリズムに対応し、Pythonでよく使われる数値計算ライブラリーと連携しやすい（川島2019、松田2019）。本研究で利用するアルゴリズムは、scikit-learn algorithm cheat-sheetを参考にK-meansを利用する。scikit-learn algorithm cheat-sheetは、どのアルゴリズムを使うべきか示したもので、それに従い、STARTから、データが50以上、本研究で利用するデータはカテゴリーデータ、ラベルなしデータ、カテゴリーの数がわかっていて、データが10,000以下なのでK-meansを利用する。ここで、サウンド・スペクトログラムの画像が、50以下と少な

図6 K-meansによるクラスタリング
3つのクラスターがある（一部改変）



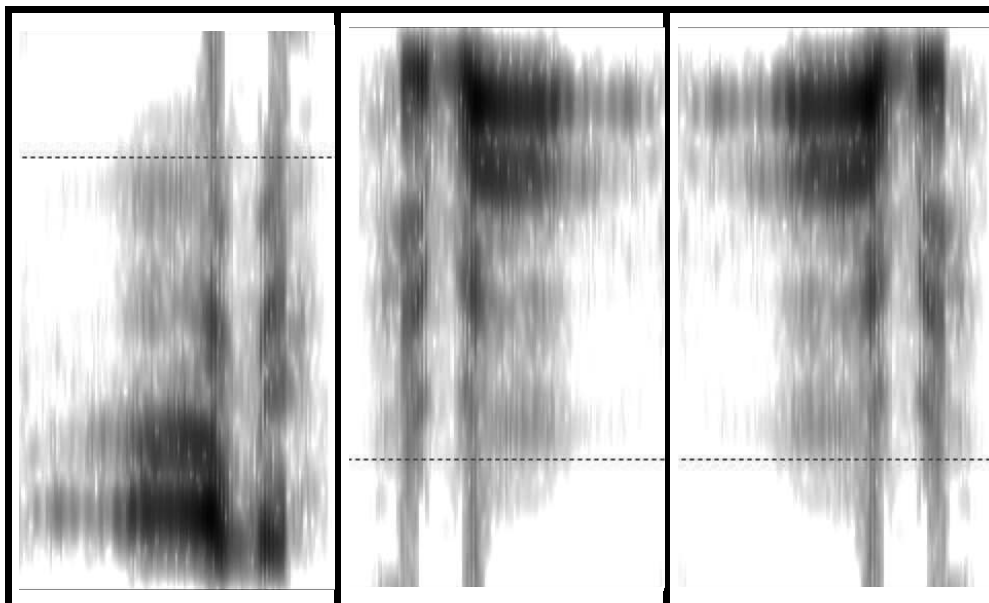
(Retrieved November 13, 2020 from <https://bluewatersql.wordpress.com/2016/03/31/data-science-day-1-machine-learning-for-the-rest-of-us/>)

図7 scikit-learn algorithm cheat-sheet



(Retrieved November 15, 2020 from https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html)

図8 水増し画像
左から左右逆、上下逆、上下逆をさらに左右逆



いので、データの水増しをおこなう。川島（2019）によると「機械学習や深層学習の学習データを準備する際に、学習用データが足りない場合にデータの個数を増やす手段として、画像データの水増しという方法がある」という。水増しはデータの数が少ない時、機械的にデータを生成し増やす処理である（松田2019）。そこで本研究では、Shafique（2019）のプログラムを参考に画像の水増しを行った（図8）。このプログラムによれば、一枚の画像から左右反転、上下反転、さらに上下反転したものを左右反転して3枚のファイルを作成し合計4枚とした。参加者5人分のファイルから、(7a)「ハンデ」と(7b)「へば」それぞれ合計60枚のデータを作成した。(7a)「ハンデ」と(7b)「へば」の画像データ60枚をそれぞれ数値化し、画像のタイプがオリジナル、左右逆、上下逆、上下左右逆と4種類の画像があるので、K-meansで4つのクラスタを作成し、オリジナル画像はオリジナル画像同士、左右逆画像は左右逆画像同士と、それぞれの属性にあったクラスタに分類できるか機械学習をおこなった。

V. 結果

(7a)「ハンデ」と(7b)「へば」それぞれのクラスタに含まれるオリジナル画像の含有率は以下表1、2の通りである。また、「ハンデ」のクラスタ散布図を以下に示す。

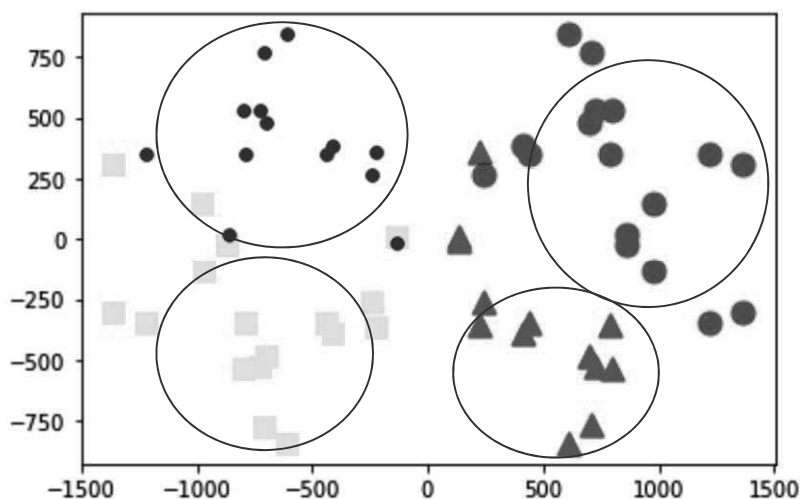
クラスタ	オリジナル画像	全画像	オリジナル画像の含有率
1	1	13	0.077
2	2	13	0.154
3	12	16	0.750
4	0	17	0.000

表1「ハンデ」オリジナル画像の含有率

クラスタ	オリジナル画像	全画像	オリジナル画像の含有率
1	0	12	0.000
2	7	18	0.399
3	8	16	0.500
4	0	14	0.000

表2「へば」オリジナル画像の含有率

図9「ハンデ」のクラスタ散布図



Ⅵ. 考察

表1より、(7a)「ハンデ」のクラスタ3のオリジナル含有率が75%と高いことがわかる。表2より(7b)「へバ」のクラスタ3は50%のオリジナル含有率である。また、「ハンデ」のクラスタ4と「へバ」のクラスタ1と4のオリジナル含有率が0%である。「ハンデ」のクラスタ1も7.7%とオリジナルの含有率は低いことがわかる。「ハンデ」については、クラスタ化がある程度成功していることがわかる。しかし、「へバ」はクラスタ2がおよそ40%のオリジナル含有率であり、「ハンデ」に比べると低い結果であった。

Ⅶ. まとめ

音声ファイルを画像化し、その画像をクラスタに分け、正しく分類できるか検討した。結果に示すように、ある程度の精度はあることがわかった。一連の手続きを自動化し運用することができれば「継承保存」が容易におこなえ、方言の「継承保存」の際に容易に音声の比較ができ、多文化・多言語共生が簡単におこなうことができると考えられる。津軽弁だけでなく、他の方言・言語などへの適応も可能と予測される。

(たかはし えいさく・高崎経済大学地域政策学部教授)

註

- 1) 方言主流社会とは、佐藤他（2003）によれば、共通語（東京方言）はもちろん通じるが、方言を使って生活した方が、地域社会内での人間関係をうまく保てる社会のことを指す。方言の言語的価値や存在意義を日常の生活に見出しにくくなった地域社会を共通語中心社会と言う。
- 2) 木部・山田（2009）によれば「継承保存」とは、世代間継承を再開・維持することと述べている。
- 3) 川島（2019）は「機械学習（ML: Machine Learning）とは、人間が明示的にプログラミングせずに、コンピュータがデータから学習するためのプログラミングの手法です」と述べている。MLには次の3つのタイプがある。教師あり学習、教師なし学習、強化学習
- 4) 音声データを提供いただいた研究協力者にはこの場を借りて、御礼申し上げる。データは、本研究のみに利用する。
- 5) 図4の上半分は音声波形を示し、下半分はスペクトログラムを表す。スペクトログラムは、横軸が時間を示し、縦軸が周波数を表す。各周波数成分の音の強さが黒の濃淡で表示されている。

付記

本論文は、日本比較文化学会第42回全国大会 2020年度国際学術大会で発表した内容に、加筆・修正を加えたものである。

プログラミングに関する問題解決をサポートするサイトteratailでの匿名の助言者にここで謝意を示す。teratail: <https://teratail.com/>

参考文献

- 内山千早. (2007). 『今すぐ話せる津軽弁』. 株式会社フィーラーステーション: 青森
- 鶴英雄他. 株式会社システム計画研究所 (編). (2016). 『Pythonによる機械学習入門』. オーム社: 東京
- 川島賢. (2019). 『今すぐ試したい! 機械学習・深層学習 (ディープラーニング) 画像認識プログラミングレシピ』. 秀和システム: 東京
- 木部暢子・山田真寛. (2019, 3月). 「いま何もしなければ」なくなってしまう 『ことばの波止場』. Vol. 5. (Retrieved from <https://kotobaken.jp/catalogue-digest/05/>)
- 佐藤和之編・平山輝男 編集委員代表. (2003). 『青森県のことば』. 明治書院: 東京
- 松田晃一. (2019). 『Python ライブラリの使い方 手軽に応用プログラミング』. カットシステム: 東京
- 松本敏治. (2017). 『自閉症は津軽弁を話さない』. 福村出版: 東京
- 山田伸哉. (2020, 1月). 弘前大、津軽弁をAIで共通語に翻訳. 『日本経済新聞』 (Retrieved from <https://www.nikkei.com/article/DGXMZO54940000Y 0A120C 2L01000>)
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay (2011) "Scikit-learn: Machine Learning in Python"

Software

- Audacity. (Version 2.2.2). [Computer software]
- Praat: doing phonetics by computer. (Version 6.1.09). [Computer software]. Boersma, Paul and Weenink, David (2020)
- Python. (Version 3.7.6). [Computer software]

分析に用いた資料

1 旅人と

そごのけんどまっすぐいって、つぎあたり左。
んだでばな。わがった。へば、わつれでいくはんで。

2 風呂で

今日はさびかったはんで ふろたぐが
ちゃんとぬぐだまねば かじえひぐはんでな
たげたがな？
よし。へばちゃんとかましてっと

3 先生と生徒

先生：はい、へば国語がらやるがな
生徒：先生、教科書わすいできたじゃ。

4 ナンボ カチャマシ ヘヤダガサ。

5 キョウノシアイ ドンデッタ/ガライモヨ

6 イッシュカンブリデ ユサハイッタキャ サツパド シタジャ

7 ハヤグ マガナレ/ジカンネ

8 ワイハ タマゲタジャ

出典： 佐藤他（2003）、「今すぐ話せる津軽弁」（株）フィーラーステーション2007）