

## 「東工大多言語音声コーパス アイスランド語」<その1>PBtext コーパスについて (The Jensson Corpus/Jensson-malheildin)

著者	三村 竜之
雑誌名	北海道言語文化研究
巻	16
号	1
ページ	99-112
発行年	2018
URL	<a href="http://hdl.handle.net/10258/00009619">http://hdl.handle.net/10258/00009619</a>

「東工大多言語音声コーパス アイスランド語」  
<その1>  
PBtext コーパスについて  
(The Jensson Corpus/Jensson-málheildin)

三村 竜之

**Review: Icelandic in *Tokyo Institute of Technology*  
*Multilingual Speech Corpus*  
#1: On *PBtext Corpus***

**MIMURA Tatsuyuki**

**要旨** : The aim of this short report is to review an Icelandic speech database compiled by Tokyo Institute of Technology. A particular attention will be paid to *PBtext Corpus* (originally termed *The Jensson Corpus/Jensson-málheildin*), one of the three corpora in the database, which includes 256 interrogative sentences with several sentence constructions elicited from 20 native speakers (13 male/7 female) of Icelandic. The corpus will be of great significance in the linguistic research of Icelandic, particularly in the descriptive study of intonation, though it contains defects in several respects. This report shall point out those imperfect points so that those involved in phonetic/phonological studies of Icelandic may put the corpus to practical use in their research.

**キーワード** : アイスランド語 音声コーパス 利用上の問題点 疑問文 イントネーション

## 1. ねらい

「東工大多言語音声コーパス アイスランド語」(以下、「東工大アイスランド語」とする)(篠田・古井 2012)と呼ばれる読み上げ音声資料が国立情報学研究所より提供されている。アイスランド語の音声研究、特にイントネーション研究に資する資料を多数有する点で優れている一方、作成の経緯を始めとする様々な点が不明確であり、また資料自体にも種々の問題点が散見され、利用する上で注意を要する。

アイスランド語のイントネーションに関しては既に豊富な先行研究が残されており、様々な点が明らかとなつてはいるものの、殊に疑問文イントネーションに関しては方言差など未だ明らかとなっていない点は少なくない。そこで、本稿では、「東工大アイスランド語」の中でも特にPBtext コーパス(別称 The Jensson Corpus/Jensson-málheildin)と呼ばれる疑問文からなる資料を取り上げ、その詳細を示すとともに批判的な視点からコーパスの特長や

問題点並びに注意点を述べていく。

## 2. 「東工大アイランド語」について

国立情報学研究所のデータセット共同利用研究開発センターが行う事業の一つに音声資源コンソーシアム<sup>1</sup>と呼ばれるものがある。様々な音声資源・データの配布を（一部有償で）行っており、「東工大アイランド語」は（2017年12月25日時点で）無償で配布されるコーパスの内の一つである。ちなみに、東工大多言語音声コーパスには現在、アイランド語の他にインドネシア語もある。

「東工大アイランド語」はDVDの形式で配布されており、wav形式の音声ファイルと読み上げられた発話を記したテキストファイルからなるフォルダーや、コーパスの概要を記したテキストファイルからなっており、パソコン上で種々のソフトウェアを用いて簡単に利用・加工することが可能である。

コーパスの編纂は東京工業大学の篠田浩一・古井貞熙の両氏によるとのことであるが、概要を記したテキストファイルには筆頭編纂者として Arnar Thor Jensson の名がある。おそらく Jensson 氏の構築したコーパスを基盤<sup>2</sup>として「東工大アイランド語」は作成されたものと考えられる。ちなみに、この Jensson 氏はアイランド語の母語話者であり、Guðnason et al. (2012) によると 2008 年には学生（研究員か？）として東京工業大学に在籍しており、電算機による自動言語（アイランド語）認識システムに関する研究を公にしている（Jensson et al. (2008)）。おそらく「東工大アイランド語」は、この研究のために採取された録音資料が基盤となっているのではないだろうか。

「東工大アイランド語」は、News、Questions、PBtext の三つのフォルダーと readme.txt というテキストファイルからなる。PBtext コーパスに関しては後ほど詳しく扱うこととして、ここでは News コーパスと Questions コーパスの概要を示す。

### (1) a. News コーパス

- i) アイランド国営放送 RÚV のニュース番組を母語話者 20 名が分担して読み上げたもの。
- ii) wav 形式の音声ファイル 400 個と資料の概要を記したテキストファイル (readme.txt)、読み上げられた文を記したテキストファイル並びに xml ファイル (transcription.txt, transcription.xml) を含む。
- iii) 話者 20 名の内訳は男性 10 名・女性 10 名。どの話者がどの音声ファイルに該当するかは readme.txt に示されているが、話者の年齢や出生地に関しては不詳。
- iv) 使用した録音機材やサンプリング周波数に関しては readme.txt に記載があるが、日時や場所、方法などに関する詳細は不明。

<sup>1</sup> 詳細は下記を参照のこと: <http://research.nii.ac.jp/src/index.html>。

<sup>2</sup> 後述する News コーパスと Questions コーパス、また本稿の対象である PBtext コーパスはそれぞれアイランド語で RÚV-málheildin, Þór-málheildin, Jensson-málheildin と名付けられており、いずれも編纂者と思われる Jensson 氏から下記の Web サイトを通じて直接入手が可能である: <http://www.malfong.is/>。

b. Questions コーパス

- i) wav 形式の音声ファイル 4462 個、資料の概要を記したテキストファイル (raedme.txt)、読み上げられた文 (とその元である英文; 後述) を記したテキストファイル並びに xml ファイル (original\_icelandic.txt, original\_icelandic.xml, original\_english.txt, original\_english.xml) を含む。
- ii) 天候に関する質問 1000 項目を母語話者 20 名が読み上げたもの。複数名の話者により全ての文が読み上げられており資料に重複が見られる。
- iii) readme.txt には天候に関する質問との記載があるものの、original\_icelandic.txt を確認すると、質問以外の発話も多数含まれている。ちなみに、読み上げられたテキストは MIT (マサチューセッツ工科大学) が作成した JUPITER Corpus と呼ばれる英文コーパスの一部をアイスランド語に翻訳し、適宜、固有名詞をアイスランドのものに置き換えたものとのことである。
- iv) 話者 20 名の内訳は男性 10 名・女性 10 名。どの話者がどの文を読み上げているか、また話者の性別と年齢に関する情報は readme.txt に記載されている。なお、News コーパスの話者との重複に関しては記載がない。
- v) News コーパスと同様に、録音日時や場所等の詳細は不明。

### 3. 「東工大アイスランド語」 PBtext コーパスについて

#### 3.1. 概要

PBtext コーパス (名前の由来は不詳) は、話者 20 名 (男性 13 名・女性 7 名) による疑問文の読み上げ音声を収めたものである。収録された疑問文は全部で 256 項目 (但し、後述するように、話者によっては項目の漏れがある)。読み上げられた疑問文に関する詳細は不明。先に触れた JUPITER corpus のような原典に関する言及がないため、おそらく Jensson 氏の作例であると思われる。

なお、PBtext フォルダに収録されている readme.txt にはその記載がないが、20 名の話者は Questions コーパスと News コーパスのいずれの話者とも重複してないとの情報が、コーパス編纂者であると思われる Jensson 氏のウェブサイト上で確認することができる (<http://www.malfong.is/index.php?lang=is&pg=jensson>)。

#### 3.2. 構成

PBtext と名付けられたフォルダの構成並びに内容は以下の通り:

(2) a. フォルダや資料の概要を記したテキストファイル (readme.txt)

各話者の性別と年齢に関する情報は記されているが、出生地や生育地に関する情報は無い。なお、2 名の女性話者が 50 歳と 49 歳であるのを除き、その他の話者は 20 代前半から 30 代前半である。

b. 読み上げられた 256 項目の原文を記したテキストファイル並びに xml ファイル:

i) original\_text\_v1.0.1.txt, original\_text\_v1.0.1.xml

ii) original\_text\_v2.0.1.txt, original\_text\_v2.0.1.xml

なお、実際の読み上げ音声に対応しているのは b)の v2.0.1 の方である。

c. etc フォルダ

HTK (The Hidden Markov Model Toolkit) と呼ばれる音声認識システム作成プログラムのためのファイルが四つ収録されている:

i) PhonemeDictionary.txt

ii) PhonemeMapText.mlf

iii) PhonemeMapWoz.mlf

iv) TriPhonemeMapText.mlf

d. 20 名の話者 (男性 13 名・女性 7 名) のそれぞれの音声資料を収めたフォルダ (f01 から m13 まで 20 個; f/m は性別を示す)

i) intro.wav

各話者が自身の年齢等について極めて簡単に自己紹介をしたもので、読み上げ音声ではない。

ii) アルファベット読み上げ音声

男性話者三名 (m01, m03, m13) を除く話者がそれぞれアイランド語のアルファベットを読み上げたもの; alphabet1.txt といった形式で通し番号をつけてファイル名が付けられている。

iii) 自然談話音声

各話者が読み上げではなく自然に発話した音声を文ごとに分節した音声ファイル。話者ごとに数は異なるが 10 から 20 個ほどが収められている。woz1.wav といった形式で通し番号をつけてファイル名が付けられている。なお、詳細は不明だが、自然発話の中から部分的に切り出して各ファイルを作っていると考えられ、そのため、woz と名前の付いたファイル同士の内容の関連や繋がりはほとんど無いと思われる。

iv) 疑問文音声

text1.wav といった形式で通し番号をつけてファイル名が付けられており、b. に示したテキストファイル (original\_text\_v2.0.1.txt) の内容にそれぞれ対応している (但し話者によっては漏れや対応にずれがある; 4.1.節を参照)。

## 4. 若干の考察

### 4.1. 整理

言うまでもなくコーパスはそのままでは生の材料にすぎず、利用する側の目的や用途に応じて整理を行うことで初めて資料として活用される。ここでは、実際のイントネーション研究において利用することを想定し、以下の 4 点に基づいて整理を行う: i) 疑問詞の有無, ii) 文構造 (単文か否か), iii) 完全な文構造であるか否か, iv) 文末音調の型 (下降調であるか

否か)。また、併せて利用上の注意を要する資料についても言及する（コーパス全体を通じての注意点に関しては 4.3 節を参照）。

なお、具体的な整理に移る前に以下の二点について留意が必要である。第一に、256 項目の全てが 20 名の話者全員に共通して利用可能というわけではない。話者によっては編纂者の判断で資料として利用するに堪えないとの判断を受け、ファイル名に“notused”と記されたものが存在する（例: text160.wav.notused）。また、話者 m13 に関しては（詳細は不明であるが）項目番号 text185 以降のファイルが全く収録されていない。以上の点を考慮すると、実際に、全ての話者を通じて比較・対象が可能な項目数は 119 項目となる<sup>3</sup>。

第二に、全話者を通じて利用可能な 119 項目のうち、text31 は疑問文ではないため、除外する必要がある（*Okei takk fyrir?* ‘Okay thank you.’；なお不可解なことに、text31 は疑問文ではないにもかかわらず original\_text\_v2.0.1.txt では疑問符が付されている）。

#### 4.1.1. 疑問詞の有無

全 256 項目から text31 を除いた 255 項目を疑問詞の有無の点から分類した結果、例えば text8 の (*Hallo*) *hvernig er veðrið?* 「(こんにちは) 天気はいかがですか。」のような疑問詞疑問文は 82 項目、また text18 の *Er veðrið gott?* 「天気はいいですか。」のような yes/no-疑問文は 139 項目となった。但し、例えば text54 の *Hvað með miðvikudaginn?* 「水曜日はどうですか? (*hvað með* は英語の *what about* に相当)」のように、疑問詞を用いながらも内容的には寧ろ yes/no-疑問文に近い、区別の困難な事例も 34 項目確認された。

#### 4.1.2. 単文か複文か

前節と同様、text31 を除く 255 項目を文構造の複雑さの点から分類した結果、単文（主述構造を一つ含む文）は 197 項目、複文は 5 項目であった。なお、4.1.3.節にて触れるが、文構造の不完全な（あるいは判定が困難な）項目もあり、ここでは考慮外としている。

#### 4.1.3. 完全な文構造を有するか否か

先に引用した text54 の *Hvað með miðvikudaginn?* 「水曜日はどうですか?」のように、主語と述語動詞（あるいはそのいずれか）を欠く疑問文はアイスランド語では決して珍しくはない。主語や述語を備えた完全な文構造を有するか否かという観点から 255 項目を分類した結果、完全な文構造を有するものは 198 項目、不完全な文構造のものは 41 項目であった。但し、text155 の *Er til té?* 「お茶はありますか。」のように、主語や述語動詞を備えているのか否か判断の困難な事例<sup>4</sup>も 16 項目確認された。

<sup>3</sup> 詳細は不明であるが、話者 m13 の資料は original\_text\_v2.0.1.txt の内容と一致しない点が数多く、疑問文以外の文も含まれている。全話者を通じての比較・対象を試みるならば、この m13 は除外する必要がある。結果として総人数は 19 名になってしまうものの、比較可能な項目数は 119 項目よりもはるかに増えることになる。本稿では諸般の事情により m13 の資料も含めて共通項目を 119 項目としたが、今後は m13 を除外して新たに共通項目を絞り込み、資料の整理を行う必要がある。

<sup>4</sup> *Er til té?* の *er* は英語の *be* 動詞に相当するもので述語動詞と考えられるため、仮に *té* 「茶」を主語と捉

#### 4.1.4. 文末音調の型

全話者に共通する項目<sup>5</sup>から疑問文ではない text31 を除いた 118 項目<sup>6</sup>を文末音調が下降調（複数の音節に渡る漸次的な音調の下降も含む）であるか否かという点から分類した結果、一名のみ(m05)であるが 118 項目全てに上昇調が確認された。

また、話者によっては 2、3 例から 20 例程度、上昇調や高平調など下降調でない文末音調の現れる者が観察された（(3)の表を参照のこと）。

(3)

話者	項目数	項目番号(text No.)
f01	1	127
f03	2	11, 12
f06	21	7, 17, 18, 24, 25, 27, 28, 29, 30, 33, 37, 39, 49, 61, 62, 63, 65, 114, 153, 156, 158
f07	2	125, 136
m06	1	93
m07	3	15, 17, 18
m08	3	7, 26, 51
m09	8	5, 7, 21, 29, 38, 117, 125, 150
m11	2	5, 6
m12	3	1, 7

さらに特筆すべき点として、一名のみ(m02)であるが、一貫して上昇調で読み上げが進む中で、途中から(text161以降)一貫して下降調が現れる話者も確認された。ちなみに、この話者は text161 以降を除いては次の 4 例に下降調が観察された: text76, 77, 131, 161。

## 4.2. 長所

PBtext コーパスの長所はその多様性と量に尽きる。疑問詞疑問文と yes/no-疑問文の両方のいずれをも含み、また単文のみならず複文も収録されている。さらに、例えば、*En í Þórsmörk?* (text5)「では Þórsmörk 【地名】は？」や *Hvað með hjálminn?* (text70)「じゃあヘルメットはどうですか？」のように主語や述語を欠く構造的に不完全な疑問文も収録されており収録する項目は多岐に渡る。さらにこれらの項目を男性 13 名、女性 7 名の話者が読み上げており量的に豊富であるのに加え、話者ごとの発話テンポも様々であるため、極めて多様な資料となっている。

既に 1.2 節にて触れたように、「東工大アイスランド語」はおそらく自動音声認識システム

えれば主述構造の完全な文と解釈することも可能であるが、主語と述語動詞を入れ替えて疑問文を作るというアイスランド語の統語規則に対する例外となる。

<sup>5</sup> 疑問文の文末音調の型を話者間で比較する目的でここでは全話者に共通する項目に限定した。

<sup>6</sup> 稿末に付した付録の一覧表を参照のこと。

構築のために作成されたものであり、言語学での使用は想定してはいない。しかしながら、BBtext コーパスが提供する豊富かつ多種多様な疑問文の資料は、例えばアイスランド語研究に携わるアイスランド語を母語としない研究者にとっては大きな意味を持つ。母語話者の研究者であれば、仮に得られた資料が少量であったとしても、自身の内省を元にその質や妥当性を吟味することが可能であるが、非母語話者の研究者の場合はそうはいかない。しかし、BBtext コーパスを利用することにより、疑問文イントネーションにまつわる諸現象、とりわけ文末音調の型（下降調か否か）の分析と解釈が量的な面から保証される。この点は極めて意義が大きい。

#### 4.3. 利用上の注意点と若干の考察

前節にて述べたように多大な意義を有する一方で、PBtext コーパスは様々な点で利用上の注意を要する。尤も、既に述べたように「東工大アイスランド語」は言語学、特に記述研究の資料を目的として作成されたものではないため、以下に指摘する諸問題は不当な批判に当たるかもしれない。しかしながら、既に述べた通り、PBtext コーパスがアイスランド語研究に資するところは大きいため、不当な批判であるとの誹りを恐れず、記述言語学的研究での利用を想定した際の問題点や注意点を指摘したい。

##### 4.3.1. 話者の出生地・生育地に関する情報

PBtext コーパスを記述言語学の資料、とりわけイントネーションの記述分析の資料として捉えた場合に真っ先に問題となる点は、話者の出生地並びに生育地に関する情報が欠けていることである。

アイスランド語の疑問文は、疑問詞の有無や文構造が完全である否かを問わず、文末音調は常に下降調（複数の音節に渡る漸次的な下降も含む）であることが既に指摘されているが（三村 2018 近刊<sup>7</sup>）、PBtext コーパスの話者の中には、一貫して文末音調が上昇調（複数の音節に渡る漸次的な上昇も含む）の現れる者が一名（m05）であるが存在する。4.1.4 節にて指摘した通り、散発的に上昇調が現れる話者も確認されているが、その場合は調査項目の読み間違いや後述する読み上げ（録音）方法の問題も要因として考えられるかもしれない。しかし、m05 のように一貫して上昇調が現れるとなると、その要因は話者の方言的特徴と推定する方が至極妥当であろう。

残念ながら PBtext コーパスにはこのような方言学的な考察を可能とする情報は不足しているが、仮に m05 の録音に観察されるような疑問文における上昇調の素性<sup>すじょう</sup>を探るとすれば、例えば、母語話者を通じて音声資料の評価を得ることで（単なる言い間違いではなく方言差であることの）正当性や妥当性を担保することは可能であろう。

<sup>7</sup> 本稿にて引用する拙論は全て日本学術振興会科学研究助成金による資金援助を受けている（課題番号 15K16729）。



#### 4.3.2. 録音方法等の詳細

先述の話者の出生地・生育地と甲乙つけ難いほどの重要な情報として、録音資料の採取方法に関する詳細が挙げられる。一般的な傾向性に反して、一貫して疑問文に上昇調の現れる話者(m05)が存在することは既に指摘した。その傍らで、考察対象とした 198 例中、2~3 例から話者によっては 20 例程度、上昇調や高平調など下降調でない文末音調の現れる者も確認された。また、同じく既に指摘した点であるが、一貫して上昇調で読み上げが進む中で、途中から一貫して下降調が現れる話者(m02)も確認された。

例えば英語のように、疑問詞の有無と下降調と上昇調の別が相補的に対応していれば上昇調の出現はとりたてて問題とすることではない。しかし、既に指摘した上昇調の出現は (m05 に見られる一貫した出現を除いては) 疑問詞の有無や文構造の別などを問わず散発的に観察されている。とすれば、上昇調の現れた要因は、調査項目の提示の方法も含めた録音手法に求めるのが妥当であろう。

筆者が推測するに、調査項目を一覧表にしたものを話者に与え、話者のペースで順に項目を読み上げさせたのではないだろうか。というのも、例えば、英語では“A, B, C, and D”と複数個の事物を列挙していく際、“A, B, C”の各項目には上昇調が現れるのが一般であると思われるが、同様の傾向はアイランド語にも確認されており(三村 2016: 155-156)、列挙された項目を読み上げていく際のイントネーションとして上昇調が現れた可能性は否めない。筆者も調査項目の読み上げによる文の採取を行うことはあるが、その際は、プレゼンテーションソフトを利用して調査項目をランダムに提示し、また、発話速度がなるべく自然になるよう 2 秒から 3 秒程度の間隔で調査項目が提示されるように設定している。また、一度は練習として一通り調査項目の読み上げを行ってもらい、その後、少なくとも 3 回 (3 セット) は読み上げを行ってもらうことにしている。このような方策を採っても、慣れてくるためであろうか、2 回目 (2 セット目) 以降は途中から一貫して上昇調が現れることがある。数名の話者に散発的に確認される上昇調は、このような事情が背景にあるのではないだろうか。

また、調査項目の読み上げ並びに録音の回数は一回のみではなかったかと筆者は推察する。4.3.3.節でも触れるが、調査項目の読み間違いや言い淀みのある項目や、読み上げ速度が遅く不自然な項目が少なくない。言い淀んだりつかえながらの読み上げでは偶発的に下降調以外の音調が現れることは十分に考えられる。一回しか読み上げていないとなれば、このような偶発的な音調の出現は排除できないであろう。

いずれにしても、残念ながら PBtext コーパスには録音手法に関する情報は収められていないため詳細は不明である。文末音調が下降調ではない資料は解釈や扱いに関して細心の注意が必要である。

#### 4.3.3. 言い淀みや調査項目の読み間違い

既に指摘したように、音声ファイルの中には編纂者から資料として適していないとの判定を受けて notused と記されたものがある。しかしながら、notused と記されていない、いわば資料として適格と判断された資料の中にも、調査項目の読み間違いや言い淀みを含むものが

多数存在する。

例えば、f05 の読み上げた text153 は *Hvar til bú?* であるが、original\_text\_v2.0.1.text では *Er til bú?* である；直後の text154 が *Hvar er Jón Bó?* のため、引きずられて *hvar* と言ってしまったのではないだろうか（ここから、おそらく調査項目は一覧表の形で話者に提示されていたのではないかと推察される）。残念ながら、各話者のフォルダーに収録されているテキストファイル original\_text\_v2.0.1.tex を逐一確認したわけではないため、読み間違いなのか、はたまた話者によって読み上げる文をわずかに変更していたのかは定かではないが（尤も、後者の可能性は低いと思われる）、話者間での資料の比較・対照を試みる際には注意が必要である。なお、現時点で筆者が確認している読み間違いと思われるものは以下の通り：f01 text177; f02 text134, f05 text153, text161 text177。

調査項目の読み間違いと推察される事例に比して、言い淀みや読み上げ速度が極端に遅い事例は相当数に上る。言い淀みや読み上げ速度の低下の原因として次の三つが考えられる。第一に、前節でも触れたが、読み上げと録音の回数が一回きりだったのではないかと推察される。複数回に渡り録音を行ってれば言い淀みやつかえを含む資料は除外することができたのではないだろうか。第二に、調査項目の長さが考えられる。調査項目の中には、読み上げ調査には不適切と思われる長さの項目が幾つか存在する（例: text117 *Andblæs tilnefnd leikhópsins augu útgefanda safnrita og dagblada?*）。第一点目とも関連するが、複数回に渡り読み上げを行わない限りは、このような長い文をスムーズに読み上げることは極めて困難であろう。最後に、調査項目に誤りや見慣れない語句・固有名詞が含まれていた可能性がある（例: text132 の *Vúlundur* は *Völundur* の誤りではないか<sup>8</sup>）。この種の誤りや見慣れない語句を含むと考えられる項目は、話者がその内容を確認しながら読み上げるため自ずと速度が遅くなり、全体的に変動の少ない高く平坦な音調を伴う可能性もある（例えば m06 の text93 はこれに該当するのではないだろうか）。イントネーション研究の資料として扱う際には細心の注意が必要であろう。

#### 4.3.4. その他

既に 4.1 節にて触れた通り、コーパス編纂者の意向から資料として使用に堪えないと判断されたファイルが存在する。しかし、これら notused とファイル名に記された資料がなぜそのような判断を受けているかの理由は readme.txt にも記されておらず、詳細は不明である。フリーウェアの音声分析ソフト Praat (Boersma and Weenink 2017) であればそのまま読み込むことが可能であるし、その他の音声再生ソフトウェアであっても、ファイル名から “notused” の部分を削除して拡張子を “.wav” とすれば再生可能なはずである。実際に内容を確認してみると分かるが、音質が劣悪でもなければ極度な言い淀みや調査項目の読み間違いがあるわけでもなく、なぜ notused の扱いとなっているのかに関しては全くもって不明である。むしろ、既に述べた通り、notused の扱いとなっていないファイルの方が言い淀みや誤りを含んで

<sup>8</sup> 現時点で筆者が確認している範囲では、本文にて言及した text132 と text159 に誤りがあるのではないかと考えられる（text159 の *Busslast* は何か他の語の誤りではないか）。

いることもあり、その扱いの基準が何であるか皆目不明である。少なくとも筆者の感想としては、notused の扱いを受けたファイルは記述音声学や音韻論の資料としては十分使用に堪えるものであるし、是非とも利用すべきである。

## 5. 結語

以上、本稿では「東工大アイスランド語」に収録されている PBtext コーパスに関して、特にイントネーションの記述研究における利用を想定して、批判的な視点からその特長や注意点を述べてきた。確かに種々の問題点はあるものの、アイスランド語研究における PBtext コーパスの意義と有用性は決して小さくはない。筆者も含め、一人でも多くの研究者が同コーパスを活用し、アイスランド語研究の発展に貢献することを心から切望する。また、本稿がコーパス活用の際の一助となれば幸いである。

今後は、「東工大アイスランド語」に収録されている他のコーパスに関しても、記述言語学での利用を視野に入れた批判的な視点からその特長や問題点を考察していきたい。

## 参考文献

- Boersma, Paul, David Weenink (2017). *Praat: doing phonetics by computer* (version 6.0.36). <http://www.fon.hum.uva.nl/praat/> (2017年12月25日アクセス).
- Guðnason, Jón, Oddur Kjartansson, Jökull Jóhannsson, Elín Carstensdóttir, Hannes Högni Vilhjálmsson, Hrafn Loftsson, Sigrún Helgadóttir, Kristín M. Jóhannsdóttir, Eiríkur Rögnvaldsson (2012). “Almannarómur: An open Icelandic speech corpus.” The 3rd International Workshop on Spoken Language Technologies for Under-Resourced Languages.
- Jensson, Arnar Thor, Koji Iwano, Sadaoki Furui (2008). “Language Model Adaptation Using Machine-Translated Text for Resource-Deficient Languages.” *EURASIP Journal on Audio, Speech, and Music Proceeding 2008*.
- 三村竜之 (2016). 「アイスランド語における文音調 (イントネーション) の記述に向けて」. 『北海道言語文化研究』14, pp. 147-158.
- 三村竜之 (2018 近刊). 「アイスランド語疑問文イントネーションの諸相」. 『室蘭工業大学紀要』67.
- 篠田浩一、古井貞熙 (2012). 『東工大多言語音声コーパス アイスランド語 (TITML-ISL)』【CD-ROM】. 東京: 国立情報学研究所音声資源コンソーシアム.

## 付録: PBtext 全話者共通項目

---

text1	Hvernig er útlitið á laugardaginn í kring um Hengilinn?
text2	Verður sól við Mývatn?
text4	En hvernig verður hitinn?
text5	Hvernig er spáin fyrir helgina?

- text6 En spáin fyrir Austurland?
- text7 Hvernig er spáin næsta mánudag?
- text8 Halló, hvernig er veðrið?
- text11 Okei en verður hlýtt?
- text12 En verður rok í Eyjum?
- text13 En í Þórsmörk?
- text14 En ekki á morgun heldur hinn?
- text15 Er rigning í kvöld?
- text16 Hvenær byrjar að rigna?
- text17 Er sól á morgun?
- text18 Er veðrið gott?
- text19 Getur þú sagt mér hvað verður heitt á morgun?
- text20 Er skýjað um morguninn?
- text21 Er grillveður núna fyrir Norðan?
- text24 Hvert er skyggnið í Reykjavík núna?
- text25 Hversu mikill vindur er á Akureyri?
- text27 Hvernig er spáin fyrir morgundaginn?
- text28 Hver er spáin á Fimmvörðuhálsi næstu daga?
- text29 Verður snjór í Hlíðarfjalli?
- text30 Verður rigning sautjándá júní?
- text33 Verður frost í nótt?
- text34 Hvernig lítur morgundagurinn út í Reykjavík?
- text35 Hvernig er spáin fyrir Klaustur á morgun?
- text37 Hvenær kemur rigning á Austurland?
- text38 Verður svipað veður næstu daga?
- text39 Verður skýjað á Snæfellsnesi?
- text42 Hvað áttu mörg skópör?
- text43 Gætir þú sagt mér hvort það er stormur eða hálka á Keflavíkurveginum?
- text44 Hvert er hitastigið á Reykjanesbraut?
- text46 Er hálka og skafrenningur á leiðinni til Akureyrar?
- text47 Hver er sjávarhiti í Faxaflóa?
- text48 Hver er langtímaspáin fyrir Suðurland?
- text49 Hver er snjódýptin í Bláfjöllum?
- text51 Er súld á Vestfjörðum?
- text53 En á fimmtudag?
- text55 En hvernig er spáin fyrir þriðjudag?
- text56 Var brúðkaupsferðin góð?

- text57 Hafið þið þekkst lengi?
- text58 Ertu fúl?
- text60 Afhverju bjó amma hér?
- text61 Heyrðir þú talað um draugagang?
- text62 Fannst eitthvað úr gamla bænum?
- text63 Áttu mörg pör af skóm?
- text64 Hvað eru þær að tala um?
- text65 Ætlarðu að reyna við efnin í fyrramálið?
- text67 Vinnur mamma þín lengi?
- text70 Hvað með hjálminn?
- text72 Er nöglin svört á Laufeyu?
- text73 Eru komnir ungar í mjúka dúninn?
- text74 Tekur skák langan tíma?
- text75 Hvað er hægt að gera við drykkfelda syni?
- text76 Jáá, er þúkinn hann Baddi með vandræði?
- text77 Eru þetta ekki fínar kökur?
- text82 Hringir Lambkár í sjöunda óeirðarmann?
- text83 Hvert er nauðsynjaerindi siðprúða skapdeildarmannsins?
- text84 Er áeggjun, auvirðing í stafnbúum?
- text85 Er bláeygðum puntubótum sæhafandi?
- text92 Hvað er óorða og gætiára?
- text93 Er Halldóra sögð ráfa um í flokki förukellinga, á snöpum á milli bæja?
- text96 Var haustþing Þingbrekku ljúft?
- text97 Hvar dræpi jáorðum óvart í doktorsgráðu?
- text99 Voru menn samþykkir landflæmingu ríkdóms Hálfðánar?
- text101 Var hestaatið og hólmgöngurnar hæversk ánauð?
- text104 Jæja, gleymdi pabbi að læra að þrjóna?
- text107 Er kjallaraíbúðin með húsdraug og páfagauk?
- text109 Kom Bjartmar seinn og tómhentur til læk dönsku frumkvöðlanna?
- text112 Fæddist stúdent Steinar í Þrándheimi?
- text114 Dvaldist Hrói með Franskbrauð á laun?
- text115 Inniheldur upphaf Morgunblaðsins myndlist í ágúst?
- text116 Sáu Killian og Gloría Vilhelm í peysu?
- text117 Andblæs tilnefnd leikhópsins augu útgefanda safnríta og dagblaða?
- text121 Er Djöflaeyjan eftir chileönsku skáldkonuna Evu?
- text125 Eru heyannir í dagrenningu?
- text127 Hvað með dagdrauma?

- text128 Ææ, mygglaði brauðið?
- text131 öö er þetta putti, ígulker eða íæti?
- text132 Hvað ertu að pæla, Vúlundur?
- text134 En hvað er dagauga, maus og íhatur?
- text135 Veistu um íörvandi og íausandi gíg?
- text136 Má ég fá íís?
- text137 Hvað er að vera íeyrir, úvinn og næba?
- text140 Er Eyvindur Jakobsson ljóshærður?
- text142 Fer fólk í bíó samfellt á hátíðardag?
- text144 Sendi Jóhann Búbba happþæga með flugpósti?
- text146 Geyjaði gúmmí súkkulaðiögnin?
- text147 Er söngvarinn óöruggur?
- text148 Passa peysuermarnar?
- text150 Reið skjálftinn yfir þau?
- text152 Er vaun öfund?
- text153 Er til bú?
- text154 Hvar er Jón Bö?
- text155 Er til te?
- text156 Er Tómas á bæ?
- text158 Er fall fararheill?
- text159 Busslast ufsinn af öllu afli?
- text161 Iii, er yddarinn egghvass?
- text162 Ertu með ibbing?
- text163 Hvar er dageldið?
- text164 Hvað með ugghausinn?
- text166 Er sígengið veggfast?
- text167 Þekkirðu konuna mína?
- text168 Hafa þeir þá bara ekki tönnina?
- text169 Hafði Auður nýrun í röð?
- text171 Hver veit leiðina heim?
- text172 Hvað meinar hann með ljótu höftunum og bönnunum?
- text174 Sagði hann eitthvað um félagsmál?
- text175 Er búíð í húsinu?
- text176 Fá fuglapörin að borða?
- text177 Kostar jógúrt mikla peninga?
- text178 Fékstu kúlu á hausinn og blóðnasir?
- text179 Jánkaði hann við spilinu?

text182 Trúir þú á álfa?

text183 Vaknar Sæunn við dauflegt hljóð?

text184 Veistu hvort þetta er einmitt svona?

---

執筆者紹介

氏名：三村竜之（みむら・たつゆき）

所属：室蘭工業大学大学院工学研究科ひと文化系領域

Email：m76tatsu@gmail.com