

日本語学習者のインタビュー応答時における言いよどみ使用

土屋菜穂子 (青山学院大学)

An Analysis of Japanese Language Learners' Hesitations Which Appeared in the Responses to Interview Questions

Naoko Tsuchiya (Aoyamagakuin University)

1 はじめに

日本語を母語としない人々に対して行う日本語教育においては、一般的に「あの」「えーと」「まあ」「なんか」などの言いよどみ¹に関して、学習者の自然習得に任せている場面が多いと思われる。また、学習者個人によってその習得方法も当然異なるものである。

しかし例えば、日本語を自在に操り自分の専門分野についても論じられる超級レベルの話者らが使う言いよどみと、身の回りの話題に対する質問になんとか「文」で答えることのできる中級レベルの話者らが使う言いよどみとを比べた場合、「それぞれのレベル内で共通する言いよどみの使い方」はあるのであろうか。また、学習者はどのような過程を経て言いよどみを習得していくのか、これをコーパスから調査することは可能であろうか。本研究ではこの二つの点について『日本語学習者会話データベース』² (国立国語研究所, 2010年) を用いて調査・考察を行う。

2 調査の方法

2.1 今回取り上げる言いよどみ

本研究で取り上げる表現は、「あの」「まあ」「なんか」「えーと」、そして「んー」「あー」である。これらのうち「んー」「あー」を除くものは、田窪 (1995) において語彙的形式を持つ言いよどみ系感動詞とされているものである。一方「んー」「あー」は語彙的形式を持たず、特に「あー」に関しては伝統的な国文法の解釈において典型的な感動詞であると捉えることもできるが、近年のフィラー研究においては、石川 (2010)、定延 (2010) などのように、あ系感動詞もフィラーに含めることがある³。本研究においても、語彙的形式を持つ言いよどみとの比較の観点からこれらの語を取り上げる。なお、以後、本論内では便宜的に「語彙的言いよどみ」「非語彙的言いよどみ」という呼び方を使うことにする。

2.2 『日本語学習者会話データベース』の特徴と n-gram 処理の使用

データには上記の『日本語学習者会話データベース』のテキストデータを用いる。このデータは、目標言語の口頭能力レベルを測る OPI(Oral Proficiency Interview) の形式⁴で

¹ここでいう「言いよどみ」とはいわゆるフィラーに相当するものである。土屋 (2000) との整合性、また、筆者が早稲田大学日本語教育研究センターで担当しているテーマ科目名「日本語の言いよどみ・あいづち・コメント 5-6」との整合性のためにここでは「言いよどみ」という用語を用いている。フィラーと置き換えて読んでいただいても差し支えない。

²<https://dbms.ninjal.ac.jp/nknet/ndata/opi/>を参照。

³『日本語話し言葉コーパス』(国立国語研究所, 2004年)の書き起こしにおいても、あ系感動詞にはフィラーのタグが付けられている。

⁴OPIの概要については牧野, 他 (2001) を参照。

行われたインタビューのデータである。表 1 に示すように、インタビューデータは OPI のレベルの枠組みに沿って超級・上級・中級・初級の主要 4 レベル、また、超級から初級 - 下までの 10 下位レベルに分類されている。合計 339 名のインタビューデータがある。本研究においても学習者の口頭能力レベルの枠組みはこれに従うものとし、超級から初級 - 中までの主要 4 レベル・9 下位レベル、合計 338 名のインタビューデータを調査の対象とする。下位レベル単位で分析を行うため、データが 1 名のみの初級 - 下は今回は調査の対象から外している。

表 1: 『日本語学習者会話データベース』に収録されている人数と口頭能力レベル

超級	上級	中級	初級
	上級 - 上 24 名	中級 - 上 68 名	初級 - 上 21 名
超級 9 名	上級 - 中 34 名	中級 - 中 84 名	初級 - 中 10 名
	上級 - 下 52 名	中級 - 下 36 名	初級 - 下 1 名

テキストデータは以下のような漢字仮名交じり、句点なし・読点ありの書き起こしファイルになっており、インタビューアーの発話 (T) と学習者の発話 (I) が交互に示される形式になっている。発話途中に挟まれた相手のあいづちは〈 〉で示されている。

(超級 0010)

T: あーそうですか〈えー〉、で、えー今日本では〈はい〉何をしてらっしゃいますか

I: 今はえーと一パソコン関係の〈ん〉ウェブ関係のお仕事をしています

なお、本データベースには現段階では形態素情報は付与されていない。以上の特徴から、本データベースのテキストデータに n-gram 処理⁵を行うことが有効であると考えられる。学習者の発話から任意の長さの文字列をくまなく取り出すことで頻出度の高い表現を抽出することが可能になる。

2.3 n-gram 処理の具体的な方法

『日本語学習者会話データベース』の超級から初級 - 中レベルまでの主要 4 レベル・9 下位レベル、全 338 名のインタビューのテキストデータの学習者の発話部分に対して n-gram 処理を行う。n-gram 処理の前にはテキスト処理言語 awk で、(1) テキストファイルから学習者の発話行を抽出、(2) 抽出した発話行に挿入されているインタビューアーのあいづちを除去、(3) 非言語情報のタグを除去、(4) 時間情報のタグを除去、といった作業を行っている。話者タグ (“ I :”) および読点 (“,”) は除去していない。

以上の作業を行った各話者の発話部分のみのテキストファイルに対して n-gram をかける。抽出する文字列は各話者につき頻度 1 以上、2 グラムから 15 グラムまでである。n-gram 処理の後、下位レベルごとに n-gram の結果をマージした対照表を作成し、先に挙げた言いよどみ表現にあたる文字列の出現を手掛かりに分析していく。ただし、「えーと」には、「えーと」「えっと」「えと」「えーっと」といった複数の文字列があったため、これら 4 種類の文字列を調査の対象とし、以下ではこれらをまとめてカタカナの「エート」と表記することにする。なお、n-gram 処理には morogram (師茂樹氏 作) の Windows 実行形式である極悪氏版⁶を使用した。

⁵n-gram の人文科学への応用に関しては長尾・森 (1993)、近藤みゆき (2000)、近藤泰弘 (2000) を参照。

⁶<http://www.vector.co.jp/>から入手可能なフリーソフトである。同氏のソフトウェアは Perl 環境がなくて

3 各レベルにおける言いよどみの使用方法の調査

3.1 当該文字列の出現回数と調整頻度

まずは単純に、それぞれの下位レベル内における当該文字列の出現と、その文字列が出現した人数を見てみることにする。表2では、下位レベルごとに最も出現回数の多かった文字列から順番に表示している⁷。なお、ここで示しているのはあくまでも「出現した文字列」であり、中には言いよどみではない文字列も含まれていることに注意を要するが、その数は全体数のうちのわずかであると判断し、ここでは出現した文字列を言いよどみの語とみなす。

表 2: 当該文字列の出現回数と出現人数

超級 9名	出現回数(人数)	上級上 24名	出現回数(人数)	上級中 34名	出現回数(人数)
あの	774(9名)	あの	1778(22名)	あの	1863(32名)
まあ	462(9名)	まあ	545(23名)	なんか	1028(31名)
なんか	202(8名)	なんか	539(24名)	エート	807(17名)
エート	118(8名)	んー	389(24名)	んー	777(34名)
んー	111(9名)	あー	323(24名)	まあ	672(29名)
あー	68(9名)	エート	307(20名)	あー	528(33名)
上級下 52名	出現回数(人数)	中級上 68名	出現回数(人数)	中級中 84名	出現回数(人数)
あの	1866(45名)	あの	2805(59名)	あー	5066(84名)
んー	1679(51名)	あー	2514(67名)	んー	3267(84名)
なんか	1327(47名)	んー	1966(68名)	なんか	1237(54名)
あー	1211(52名)	なんか	1514(54名)	あの	1169(68名)
まあ	918(35名)	エート	1383(39名)	エート	953(37名)
エート	729(33名)	まあ	548(38名)	まあ	202(42名)
中級下 36名	出現回数(人数)	初級上 21名	出現回数(人数)	初級中 10名	出現回数(人数)
あー	2602(36名)	あー	2251(21名)	あー	481(10名)
んー	1213(36名)	んー	855(21名)	んー	367(10名)
あの	509(31名)	あの	196(16名)	あの	38(4名)
エート	247(19名)	エート	165(6名)	まあ	19(5名)
まあ	150(15名)	まあ	19(7名)	エート	7(3名)
なんか	93(14名)	なんか	6(5名)	なんか	3(1名)

この結果から、各レベルの言いよどみの使用方法を調査するためのいくつかの手掛かりを見つけることができる。例えば以下の点に注目することができる。

- それぞれの下位レベル内で出現回数が最も多いのは、超級から中級 - 上までは「あの」、中級 - 中から初級 - 中までは「あー」次いで「んー」の順番である。
- 9つの下位レベルの中間に位置する中級 - 上は、出現回数が最も多いのが「あの」、それに次ぐのが「あー」「んー」である。
- 非語彙的な言いよどみの出現回数に関して、初級 - 中から中級 - 上までは「あー」次いで「んー」の順であるのに対し、上級 - 下から超級までは「んー」の出現回数が「あー」を上回る。
- 語彙的な言いよどみの出現に関して、「なんか」の出現回数が上位に上がってくるのは中級 - 中からである。一方、「まあ」の出現回数が上位になるのは上級 - 上からである。

も n-gram 処理を行うことができる。オリジナルの morogram については <http://morogram.sourceforge.jp/> を参照。

⁷土屋 (2012) では各話者につき頻度 2 以上の文字列を抽出したため、頻度 1 以上の文字列を抽出した今回の表とは数値が若干異なっている。ただし出現回数の順番に大きな変動はない。

次に、それぞれの下位レベルでの当該文字列の出現回数を100万語あたりの調整頻度に直して示したのが表3である。調整頻度を出す際に必要となる総語数を出すにあたっては、形態素解析にMeCab、辞書はUniDicを使用した。n-gram処理をする際に作成した学習者の発話部分のテキストを形態素解析にかけ、その結果から話者タグと読点、かぎかっこ等の不要な行を除いたものを総語数として数えた。学習者が誤って発話した部分には誤解析が多少見られたが修正はしていない。このような条件下で出した下位レベル別の総語数は、超級、9名33504語、上級-上、24名85459語、上級-中、34名101273語、上級-下、52名143140語、中級-上、68名159710語、中級-中、84名171496語、中級-下、36名63508語、初級-上、21名31047語、初級-中、10名13452語である。

表 3: 100 万語あたりの出現回数

	あの	まあ	なんか	エート	んー	あー
超級	23102	13789	6029	3522	3313	2030
上級 - 上	20805	6377	6307	3592	4552	3780
上級 - 中	18360	6622	10131	7953	7657	5203
上級 - 下	13036	6413	9271	5093	11730	8460
中級 - 上	17563	3431	9480	8659	12310	15741
中級 - 中	6816	1178	7213	5557	19050	29540
中級 - 下	8015	2362	1464	3889	19100	40971
初級 - 上	6313	612	193	5315	27539	72503
初級 - 中	2825	1412	223	520	27282	35757

3.2 口頭能力レベルと当該言いよどみの出現回数との関係

これらの数値から、レベルの上昇と出現回数が比例しているのが「あの」「まあ」、反比例しているのが「んー」「あー」であることが分かる。レベルが上がるにつれ、語彙的言いよどみである「あの」「まあ」は出現回数が増えていく。調整頻度を見ると「あの」は中級-上からその出現回数が大幅に増え始めるのに対し、「まあ」は上級-下から大幅に増え始め、更に超級ではその出現回数が倍に増える。この結果から、「まあ」の習得の難しさがうかがえる。そして、レベルが上がるにつれ出現回数が減少していくのが非語彙的言いよどみである「んー」「あー」である。ただし、先にも述べたように上級-下から上位のレベルでは、「んー」の出現回数が「あー」を上回ることは注目に値する。これについては更に検討を要する。

一方、「なんか」は中級-中から大幅に増えるがそのピークは上級-中であり、上級-上からは再び減少する。同じく語彙的言いよどみであり上位レベルになるほど出現回数が増える「あの」「まあ」とは異なる使用状況が観察される。一つの解釈としては、「なんか」は中級から上級の入り口程度で多用される言いよどみであると考えられるであろう⁸。

「エート」については「なんか」ほどの特徴を数値からは見出すことはできないが、調整頻度を見る限り、やはり上級-上からその出現回数は減少している。

⁸ただし、OPIの性質がかかわっている可能性もある。OPIにおいては上級-上から超級の話者に対して社会的あるいは専門的な話題の中で話者の論理的な意見やその裏付け等を問うため、文脈的に「なんか」が出現しにくい事情がある可能性も否定できない。更に検討を要する。

3.3 当該言いよどみ同士の共起パターン

各レベルに共通した言いよどみの使用方法を更に探るために、次は当該文字列同士が共起する例を調査する。当該文字列同士が共起する例とは以下のようなパターンを指し、隣り合って出現しているものに限定する。

あの、まあ、日本語ではいろいろ（超級 0076） / あーなんか 外国人なのになんでそ（超級 0015）

これらの例も下位レベルごとに n-gram 処理の結果をマージした対照表から抽出する。仮に「あの一、まあ、なんか」という文字列があった場合には、「あの一+まあ」と「まあ+なんか」の二つの共起パターンが抽出されることになる。

表 4: 当該言いよどみ同士の共起表 超級～中級 - 上

超級 9名						
	あの	まあ	なんか	エート	ん一	あ一
あの	3名6回	8名33回	0名0回	5名8回	0名0回	0名0回
まあ	4名24回	3名5回	0名0回	0名0回	0名0回	0名0回
なんか	3名10回	2名4回	0名0回	0名0回	1名1回	1名2回
エート	2名5回	2名3回	0名0回	0名0回	0名0回	0名0回
ん一	2名2回	3名5回	1名2回	1名1回	0名0回	0名0回
あ一	1名1回	0名0回	1名1回	1名1回	0名0回	2名3回

上級 - 上 24名						
	あの	まあ	なんか	エート	ん一	あ一
あの	9名14回	13名49回	7名37回	3名4回	8名11回	4名5回
まあ	11名24回	6名6回	2名2回	2名2回	2名2回	0名0回
なんか	6名13回	2名4回	2名5回	0名0回	3名5回	4名4回
エート	3名3回	6名9回	1名1回	2名4回	2名2回	0名0回
ん一	7名7回	10名19回	5名8回	0名0回	2名2回	2名3回
あ一	5名6回	3名4回	0名0回	0名0回	2名4回	4名4回

上級 - 中 34名						
	あの	まあ	なんか	エート	ん一	あ一
あの	12名40回	7名28回	5名19回	5名7回	7名23回	5名19回
まあ	5名10回	6名8回	7名7回	2名16回	2名3回	5名8回
なんか	7名14回	6名6回	7名12回	3名3回	8名12回	3名4回
エート	2名6回	6名45回	5名5回	7名37回	5名15回	5名8回
ん一	7名10回	9名28回	15名34回	4名5回	6名10回	4名4回
あ一	5名7回	4名6回	5名8回	3名3回	6名6回	3名3回

上級 - 下 52名						
	あの	まあ	なんか	エート	ん一	あ一
あの	12名18回	8名54回	6名21回	9名37回	14名33回	9名33回
まあ	7名18回	6名15回	3名3回	3名3回	5名9回	2名4回
なんか	7名11回	2名2回	9名12回	7名9回	13名29回	10名13回
エート	9名16回	8名13回	6名8回	8名11回	9名16回	8名14回
ん一	13名20回	15名33回	14名29回	10名12回	17名36回	12名21回
あ一	13名21回	7名10回	8名12回	5名12回	10名19回	10名25回

中級 - 上 68名						
	あの	まあ	なんか	エート	ん一	あ一
あの	18名88回	3名6回	7名13回	3名3回	16名50回	15名83回
まあ	2名8回	8名12回	5名8回	1名17回	4名6回	4名6回
なんか	9名15回	3名5回	7名12回	4名41回	18名61回	10名22回
エート	2名7回	5名26回	7名21回	8名23回	11名25回	8名18回
ん一	12名25回	13名23回	17名55回	11名15回	21名54回	15名22回
あ一	10名14回	10名17回	12名32回	8名16回	22名38回	25名79回

当該言いよどみ同士の共起パターンを下位レベルごとに示したのが表 4, 5 である。各下位レベル内で半数以上の話者に出現した共起パターンには欄内に網掛けを施してある。例えば超級話者の表中にある「8名33回」とは、超級話者全9名中、8名の話者の発話に「あの一+まあ」という共起が合計33回出現したという意味を示す。

表 5: 当該言いよどみ同士の共起表 中級 - 中～初級 - 中
中級 - 中 84 名

	あの	まあ	なんか	エート	んー	あー
あの	11 名 27 回	1 名 1 回	6 名 17 回	1 名 1 回	13 名 38 回	7 名 58 回
まあ	1 名 1 回	9 名 14 回	0 名 0 回	1 名 1 回	3 名 6 回	4 名 4 回
なんか	3 名 4 回	3 名 4 回	6 名 6 回	0 名 0 回	16 名 58 回	15 名 25 回
エート	2 名 4 回	1 名 9 回	3 名 6 回	7 名 13 回	18 名 32 回	12 名 75 回
んー	12 名 19 回	6 名 8 回	20 名 39 回	10 名 40 回	45 名 103 回	42 名 107 回
あー	16 名 30 回	5 名 15 回	10 名 24 回	7 名 34 回	42 名 110 回	43 名 142 回

中級 - 下 36 名

	あの	まあ	なんか	エート	んー	あー
あの	4 名 12 回	1 名 1 回	0 名 0 回	0 名 0 回	3 名 8 回	4 名 28 回
まあ	0 名 0 回	5 名 7 回	1 名 2 回	0 名 0 回	1 名 1 回	1 名 1 回
なんか	0 名 0 回	1 名 5 回	0 名 0 回	0 名 0 回	3 名 4 回	1 名 2 回
エート	2 名 2 回	1 名 1 回	0 名 0 回	1 名 2 回	7 名 12 回	5 名 13 回
んー	3 名 3 回	4 名 6 回	5 名 7 回	2 名 2 回	17 名 35 回	18 名 40 回
あー	5 名 15 回	2 名 3 回	4 名 5 回	2 名 2 回	18 名 52 回	27 名 123 回

初級 - 上 21 名

	あの	まあ	なんか	エート	んー	あー
あの	2 名 4 回	0 名 0 回	0 名 0 回	0 名 0 回	4 名 12 回	5 名 5 回
まあ	0 名 0 回	2 名 2 回	0 名 0 回	0 名 0 回	0 名 0 回	0 名 0 回
なんか	0 名 0 回	0 名 0 回	0 名 0 回	0 名 0 回	0 名 0 回	0 名 0 回
エート	0 名 0 回	0 名 0 回	0 名 0 回	1 名 2 回	2 名 5 回	3 名 34 回
んー	2 名 9 回	0 名 0 回	0 名 0 回	2 名 8 回	13 名 45 回	13 名 51 回
あー	5 名 6 回	3 名 3 回	0 名 0 回	3 名 7 回	16 名 69 回	17 名 149 回

初級 - 中 10 名

	あの	まあ	なんか	エート	んー	あー
あの	1 名 1 回	0 名 0 回	0 名 0 回	1 名 1 回	2 名 2 回	2 名 4 回
まあ	0 名 0 回	1 名 1 回	0 名 0 回	0 名 0 回	0 名 0 回	0 名 0 回
なんか	0 名 0 回	0 名 0 回	0 名 0 回	0 名 0 回	0 名 0 回	0 名 0 回
エート	0 名 0 回	0 名 0 回	0 名 0 回	0 名 0 回	0 名 0 回	0 名 0 回
んー	1 名 1 回	1 名 1 回	0 名 0 回	1 名 1 回	5 名 19 回	0 名 0 回
あー	1 名 5 回	0 名 0 回	0 名 0 回	1 名 1 回	4 名 6 回	6 名 17 回

各表中には四つの大きな区切りがある。左上の区切りには、語彙的言いよどみ同士の共起（例. あの+まあ）、左下の区切りには非語彙的+語彙的言いよどみの共起（例. んー+まあ）、右上の区切りには語彙的+非語彙的言いよどみの共起（例. あの+あー）、右下の区切りには非語彙的言いよどみ同士の共起（例. んー+あー）が位置している。

まずは、非語彙的言いよどみが後接するパターンの出現に注目する。超級を見てみると、「あの+まあ」などの語彙的言いよどみ同士の共起、「んー+まあ」などの非語彙的+語彙的言いよどみの共起例はいくつか見られるが、語彙的言いよどみに非語彙的言いよどみが後接するパターン（例. 「あの+あー」）や、非語彙的言いよどみが連続するパターン（例. あー+んー）はほぼ出てこない。つまり、言いよどみの内側に、非語彙的言いよどみが入り込むパターンがほとんど見られない⁹。

その後、レベルが下がるに従い、表中右上・右下の非語彙的言いよどみが後接するパターンの出現が増えていく。更に中級 - 中から下のレベルでは、表中右下の「あー+あー」や「あー+んー」のような非語彙的言いよどみの連続が半数以上の話者に現れる。

非語彙的言いよどみの連続は、多くの場合、話者にとっての言語産出の難しさが反映されているものと解釈できる。例えば以下のような非語彙的な言いよどみが三つ以上連続す

⁹現在調査中の『インタビュー形式による日本語会話データベース』（上村隆一、じんこんもん DATABASE Vol.1, 重点領域「人文科学とコンピュータ」総括班, 1998）の母語話者 50 名のデータからも同様の結果を得ている。非語彙的言いよどみの連続や、言いよどみの内側に入り込む非語彙的言いよどみの例はきわめて少ない。なお、このデータベースも OPI のインタビュー形式である。

る例は、上級 - 中から超級の話者の発話に見出すことができない。

あー、んー、んー高いー、たかー (上級 - 下 0055) / あー、んー、あー、あちよつとわ (中級 - 上 0030)

あー、んー、んー、あーいちじか (中級 - 中 0330) / あー、んー、んー、先生、は、漢 (中級 - 下 0256)

よって、口頭能力レベルが上がるほど、言いよどみに非語彙的言いよどみを後接させて発話することが減少するものと考えられる。

次に、各下位レベル内で出現する共起パターンの種類の多寡に注目する。初級 - 中から中級 - 下までは「0名0回」の欄が目立つが、中級 - 中以降、特に中級 - 上から上級 - 中まではどの言いよどみ同士も共起するという結果である。そして再び超級では「0名0回」の欄が目立つようになる。この結果に対しては以下のように解釈する。初級 - 中から中級 - 下まではそもそも「あの」「まあ」「なんか」「エート」といった語彙的言いよどみを使うのが困難であるために言いよどみの共起パターンの種類が少なく、「0名0回」の欄が目立つ。語彙的言いよどみの習得は中級 - 中程度から始まっていくものと思われるが、超級へ至るまでの過程に語彙的・非語彙的にかかわらず複数の言いよどみを組み合わせて使用する時期があり、それが中級 - 上から上級 - 下、あるいは上級 - 中までの時期に相当する。最終的に超級に至る時期には、非語彙的言いよどみを後接させる使用方法がなくなり言いよどみの使用方法が洗練され、再び「0名0回」の欄が多くなる。

4 結論・研究の応用・次への課題

「それぞれのレベル内で共通する言いよどみの使い方」はあるかという問いに対しては「ある」と答えることができる。言いよどみの習得過程を踏まえながら述べると、次のようになる。

初級から中級 - 中レベルの話者は非語彙的言いよどみを最も多く使用するが、中級 - 上レベルから語彙的言いよどみ「あの」を多く使用し始め、その後も超級に至るまで「あの」を最も多く使用することになる。ただし中級 - 上レベルは非語彙的言いよどみの使用も多く残り、過渡期的な位置を占める。また、中級 - 上から上級 - 中程度までは、語彙的・非語彙的を問わず複数の言いよどみを共起させて発話することがある。超級に至る時期には非語彙的言いよどみの後接がなくなり、言いよどみの共起パターンが減少する。これは使用方法が洗練されることであると解釈できる。

また、言いよどみの語に焦点を当てれば以下のように述べることができる。

語彙的言いよどみ「あの」「まあ」の出現回数はレベルの上昇と比例関係にあり、逆に非語彙的言いよどみ「んー」「あー」の出現回数はレベルの上昇と反比例の関係にある。ただしそれぞれの語には出現状況に特徴があり、「あの」は中級 - 上から、「まあ」は上級 - 下から大幅に増え始める。「んー」と「あー」では初級 - 中から中級 - 上までは「あー」の方が出現回数が多いが、上級 - 下から超級にかけては出現回数の多寡が逆転する。また、「なんか」は中級 - 中程度から上級 - 中の間で多用される。「エート」については不明である。

以上、『日本語学習者会話データベース』のテキストデータと n-gram 処理を用いて調査・考察を行った。冒頭に述べた「学習者はどのような過程を経て言いよどみを習得していくのか、これをコーパスから調査することは可能か」という問いに対しても、ある程度可能であると答えることができる。

本研究の発展は、教育実践の場に応用可能である。学習者の口頭能力レベルに応じた言いよどみの習得を促すことに役立てることができる。ただし今回の調査では取り上げる語を限定しており、かつテキストデータのみを用いたため課題も多い。以下に、次への課題を挙げる。

- 音声データの利用。特に非語彙的言いよどみ「んー」「あー」の音質の調査。初級から超級に至るまで、音質が異なっているものが一律に表記されている可能性がある。
- 「えー」「その」等、他の言いよどみの調査。
- OPI形式を持つ他のインタビューデータとの結果比較。『KY コーパス』¹⁰（鎌田修・山内博之，1999年）や『インタビュー形式による日本語会話データベース』の母語話者のインタビューデータ等。
- OPI形式を持たない他のインタビューデータとの結果比較。『日本語話し言葉コーパス』の模擬講演インタビュー等。

付記

この研究は、国立国語研究所のプロジェクト『日本語教育データベースの構築-日本語学習者会話データベース』を利用して行われたものである。

参考文献

- 石川創 (2010) 「あいづちとの比較によるフィラーの機能分析」『早稲田日本語研究』19, 早稲田大学日本語学会, pp.61-72
- 近藤みゆき (2000) 「n グラム統計処理を用いた文字列分析による日本古典文学の研究」『千葉大学人文研究』第29号, 千葉大学文学部, pp.187-238
- 近藤泰弘 (2000) 「《文化資源》としてのデジタルテキスト—国語学と国文学の共通の課題として—」『国語と国文学』第77巻11号, 東京大学国語国文学会, pp.127-139
- 近藤泰弘・近藤みゆき (2001) 「N-gram の手法による言語テキストの分析方法—現代語対話表現の自動抽出に及ぶ—」『漢字文献情報処理研究』第2号, 漢字文献情報処理研究会, pp.50-55
- 定延利之 (2010) 「会話においてフィラーを発するということ」『音声研究』第14巻第3号, 日本音声学会, pp.27-39
- 田窪行則 (1995) 「音声言語の言語学的モデルをめざして—音声対話管理標識を中心に—」『情報処理』第36巻第11号, 情報処理学会, pp.1020-1026
- 土屋菜穂子 (2000) 「対話コーパスを用いた言い淀みの統語論的考察」『青山語文』第30号, 青山学院大学日本文学会, pp.13-26
- 土屋菜穂子 (2012) 「日本語学習者の口頭能力レベル別言いよどみ使用—『日本語学習者会話データベース』の n-gram 分析をもとにして—」『日本語教育学会秋季大会予稿集』日本語教育学会, pp.253-254
- 長尾眞・森信介 (1993) 「大規模日本語テキストの n グラム統計の作り方と語句の自動抽出」『情報処理学会研究報告. 自然言語処理研究会報告』93(61), 情報処理学会, pp.1-8
- 牧野成一, 鎌田修, 山内博之, 他 (2001) 『ACTFL-OPI 入門 日本語学習者の「話す力」を客観的に測る』, アルク
- 山内博之 (2004) 「語彙習得研究の方法—茶釜と N グラム統計—」『第二言語としての日本語の習得研究』7号, 第二言語習得研究会, pp.141-161
- 山内博之 (2009) 『プロフィシェンシーから見た日本語教育文法』, ひつじ書房

¹⁰http://opi.jp/shiryo/ky_corp.html を参照。