様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　１９　日現在

機関番号：　１４３０２
研究種目：　若手研究(B)
研究期間：　2014～2017
課題番号：　２６７７０１８０

研究課題名（和文）縦断型接触場面コーパスの構築とそれを用いた日本語教育のための談話研究

研究課題名（英文）Building longitudial contact situation conversation corpus and it application for discourse study for Japanese language education

研究代表者

中俣　尚己（Nakamata, Naoki）

京都教育大学・教育学部・准教授

研究者番号：００５９８５１８

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：東京・実践女子大学と中国・湖南大学の間で行ったSkypeを用いた日本語会話活動を文字化した『日中Skype会話コーパス』を構築、一般公開した。
コーパスの分析の結果、日本語母語話者と中国語母語話者の間で全体としては語彙数の違いは見られなかった。しかし、品詞ごとに分析をすると、副詞には大きな違いが見られ、学習者にとって難しいことがわかった。
また、本コーパスは話題が指定されていることが特徴であり、それを利用して語彙リストを作成した。その結果、時間を表す「ている」「た」は「ポップ・カルチャー」に多いなど、従来予想されていなかった機能語が話題の影響を受けている可能性が示唆された。

研究成果の概要（英文）：In this project, “Japan-China Skype Conversation Corpus” is built and published online. This corpus comprises a contact-situation conversation corpus of Japanese, containing 38 records of remote conversation activity on Skype between Jissen Women’s University in Tokyo and Hunan University in Zhangsha.
As as result of analysis, no difference is observed in size of vocabulary between native speaker and non-native speaker. However, in terms of part of speech, a large difference is observed in usage of adverbs. This shows that they are hard to acquire for learners.
An unique characteristic of this corpus is that the topic of conversation is assigned. Therefore, specific words for each topic are extracted. As a result, function words, which are not considered to depend on topic, are affected by topic actually: for example, time expression such as 'ta' or 'teiru' appear frequently in conversation of 'pop culture', not of 'eating.'

研究分野：　日本語教育

キーワード：　会話コーパス　接触場面　話題別特徴語　インテイク　副詞　TTR　LLR　機能語

１．研究開始当初の背景

研究開始当初の時点では、学習者コーパスとして、OPI などのテスト場面を録音文字化したものはあったが、学習者が自然に日本語で会話したものを収録したものは少なかった。学習者の日本語使用の実態を可視化、分析することは今後の日本語教育研究を進める上で重要であると考え、それまでに実施していた、東京・実践女子大学と、長沙・湖南大学の学生が、Skype を使って日本語会話活動を行った(中俣ほか 2013)際に録音したデータを使い、文字化することを試みた。

２．研究の目的

『日中 Skype 会話コーパス』は以下の４つの特徴を持つ。

（１）真正性がある
（２）縦断コーパスである
（３）一種の電話場面である
（４）話題が指定されている

この４つの観点から接触場面における中国語話者の談話の特徴を見出すことを目的とした。実際に重点的に行ったのは（１）と（４）である。特に話題が指定されているコーパスは稀であり、自然言語処理ならびに統計的手法を用いて、話題ごとの特徴語リストを作り上げることを目標とした。

３．研究の方法

（１）初年度はまず、録音データを業者に依頼し文字化を行った。続いて学生アルバイトを使い、文字化のチェックと個人情報の削除を行った。

（２）語数のカウントならびに特徴語抽出の前提となる単語への分割については、長岡技術科学大学の山本和英氏が開発中の「雪だるま」というソフトウェアを用いた（山本ほか 2016）。この「雪だるま」は「気が早い」のような慣用句、「かもしれない」のような複合辞、「勉強する」のようなサ変動詞、「無理だ」のような形容動詞をそれぞれ１語として出力することができる。

（３）また、話題別特徴語の抽出には先行研究でよく使用されている対数尤度比(LLR)を計算した（田中・近藤 2011）。当初は「食」のセッションとそれ以外のセッションに分割したが、この方法では話題がそれた箇所の語彙も抽出されることがわかった。そこで、人手で「ポップ・カルチャー」について話している箇所のみを確認して抽出を行ったところ、精度は飛躍的に上昇した。最終的には協力者の手を借り、コーパス全体を「ポップ・カルチャー」「家庭」「開始部」「休暇」「言語」「終了部」「小学校中学校高校」「大学」「街」「天気」「伝統」「食」「恋愛」「その他」の 14 サブコーパスに分割し、それぞれの特徴語を抽出した。

４．研究成果

（１）まず、当初の予定どおり 2015 年 4 月 1 日に『日中 Skype 会話コーパス』を web 公開した。利用には登録が必要である。2018 年 5 月 11 日までのダウンロード数は 120 回である。

コーパスには延べ９ペア、38 の会話を収録している。総会話時間は 46:48:35 で、１会話あたり平均 1:13:55 とまとまった長さの会話と言える。日本語解析システム「雪だるま」を使って分析した結果、総語数は 204,632 語であった(記号類を除く)。

（２）続いて、学習者と母語話者の特性について形態素解析を行って調べた。まず、本コーパスが高い真正性を有していることは表１の比較からもわかる。これまで、会話コーパスといえども「明後日」「木曜」といった語の出現は少ないことが指摘されている（北村ほか 2009）が、本コーパスはそのような語も豊富に含んでいる。

表１　日中 Skype 会話コーパスと KY コーパスの比較

語	KY コーパス	日中 Skype 会話コーパス
明後日	0	7
木曜	6	41
すごい	77	211
すごく	190	86
すげえ	0	4

（３）続いて、母語話者と学習者の語彙を比較する目的語彙の豊富さを表す指標として TTR を比較したところ、表２のようになり、差は全く見られなかった。

表２　学習者と母語話者の使用語彙量の比較

	学習者	母語話者
Token 頻度	104,156	100,325
Type 頻度	5,434	5,217
TTR	0.052	0.052

（４）そこで、品詞ごとに違いを見たところ、差が表れた。図１は母語話者の使用を１とした際に、学習者の使用がどれだけあったかを表したグラフである。これによれば、副詞、助動詞、接続詞といった機能語に大きな違いが見られる。動詞についても違いを生んでいるのは「田中という人」のように機能語として使われている箇所である。感動詞については学習者はフィラーを多用するためで特筆すべき違いではない。

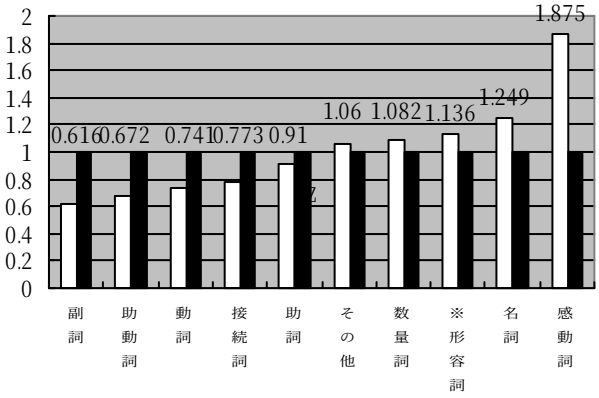

図１　品詞ごとの母語話者と学習者の使用量の違い

特に、副詞では個々の語においても違いが見られ、例えば「とても」「いろいろ」「つまり」「もし」「たぶん」は学習者に特徴的な語、「そう」「こう」「結構」「やっぱり」「なるほど」は母語話者に特徴的な語という結果が得られた。

名詞などに差が見られなかった理由としては、会話開始時点での学習者のメンタルレキシコンに存在しない語であっても、会話相手の母語話者が語を使用することでそれがインテイク、すなわち学習され、結果的に発話数は変わらないように見えると解釈される。名詞については学習者が未知の語に接し、それを取り入れている場面が多々観察されたが、副詞についてはそのような場面は見られなかった。

（５）その他、学習者と母語話者の違いは様々に観察される。例えば、否定応答表現を調査すると表３のようになる。

表３　話者と使用した否定応答表現

	いいえ	いや	いえ
学習者	32	31	44
母語話者	4	73	55

学習者は「いいえ」「いや」「いえ」をランダムに使っているように見えるが、母語話者は「いいえ」を回避している。さらに仔細に観察すると、母語話者は「ありがとう」「すみません」といった相手の感謝・謝罪の後にのみ、フェイス補償行為として「いいえ」を使用することがわかった。

これらは学習者の中間言語の分析のほんの一例であり、このコーパスが多くの研究者に利用されることで、今後も様々な知見をもたらすと考えられる。

（６）　次に、話題ごとの特徴語の抽出作業を行った。まずは、「食」についてのセッションのファイルを対象コーパス、それ以外のファイルを参照コーパスとし、ＬＬＲを計算したところ名詞 190 語、動詞 35 語、形容詞 11 語が抽出され、それぞれ実際の文脈で「食」に関連して使われていた割合は 83.7%、80.0%、90.9%であった。しかし、語抽出は基本的に話題がそれた箇所で起こっており、全体としてはよく特徴語を抽出できていることがわかった。また、特筆すべきは「もの」という一見話題と関連しているとは思えない語が抽出されたことである。詳しく分析を行うと、「ハンバーグをパンではさんだもの」のような「ＮをＶしたもの」という文型が「食」の話題には頻出し、またこの文型は他の話題では出現しないことがわかった。これは食べ物の説明には非常に便利な文型であるが、直観ではなかなか思いつかない表現である、コーパスからの特徴語抽出の強みを確認できた。

（７）次に、調査方法を改善し、調査協力者とともに、「実際に「ポップ・カルチャーについて話している箇所」と「それ以外」にコーパスを分割し、特徴語抽出を試みた。結果は以下の通り、合計 252 語が抽出され、全体として 96%が実際にポップ・カルチャーについて話している場面で使われていることがわかった。

表４　ポップ・カルチャー特徴語

品詞	語数	例
一般名詞	93	アニメ・映画・ドラマ・歌
固有名詞	86	嵐・ホタルノヒカリ・木村拓哉
動詞	19	聴く・知る・出る・読む
形容詞	17	人気・好き・面白い・カッコいい
副詞	7	とても・いろいろ・最近・去年
機能語	3	ている・た・の

特筆すべきは、テンス・アスペクトマーカーである「ている」「た」が特徴語とされたこと、そしてそれと呼応するように「去年」「最近」といった時間の副詞が抽出されたことである。これは「食」の話題と比べて、ポップ・カルチャーが時間が問題になることが多いということを意味する。例えば、「あの鯛焼き屋はおいしいです」と「あの鯛焼き屋はおいしかったです」はもちろん意味は異なるが、聞き手の行動は変わらない。しかし、「今やっている映画が面白い」と「昔見た映画が面白い」では完全に聞き手の行動は異なってくるということである。もちろん「ている」や「た」がある話題にしか出現しないという意味ではないが、相性が良い話題・悪い話題が

存在するということの示唆を計量的に示すことができたことに意義がある。

（8）そこで、最終段階として『日中 Skype 会話コーパス』全文を協力者の力を借りて話題に分割した。当初設定した話題以外に「恋愛」「小学校・中学校・高校」「天気」など頻出する話題も加えた。そして、特徴語の抽出を行ったが、特に機能語について以下に取り上げ、報告する。

表５　機能語の話題別特徴語

話題	特徴語
食	られる（可能）・の（準体助詞）・かな・そう（様態）
家庭	ころ・いつ・時・と（同伴）・た・なくちゃいけない・も
言語	に対する・といえば・場合・を・って・ても・時・と思う・ば・と（接続助詞）・
ポップカルチャー	誰・ている・た・の（格助詞）
伝統	の（格助詞）・ながら・では（場所）・を・になると・と（同伴）・よう（様態）・たり・みたい
町	が・たい・し・の（格助詞）
大学	まだ・たり・くらい・に比べて・ます・の（格助詞）・ために
休暇	まで・くらい・から（格助詞）・に・たい・です
恋愛	といい（勧め）・です
小中高	とか・時・ちゃ・てく
天気	です・くらい・以上・時に・なぜ・は・な
開始部	ません・ちょっと・ます・た・か・です・ない
終了部	じゃあ・また・ございます・から（格助詞）・にする・まで・までに・こそ・ので

従来の研究では機能語は話題に依存しないとされてきたが、いくつかの機能語が抽出された。その理由としては以下のものが考えられる。
①コロケーションによるもの。例えば「られる」は「食べられる」という形で使われることが多かった。これは「食べる」の影響を受けるため、間接的に「食」の話題に多くなる。
②文タイプによるもの。例えば主格マーカー「が」は「町」に多いが、これは「～には～にがあります」という存在文がこの話題に多く使われるためである。「天気」の話題に主題マーカー「は」とコピュラ「です」が多いのも、この話題には単純な名詞文・形容詞文が多く含まれ、「～は～です」という構造をとるからである。
③述語タイプによるもの。ポップ・カルチャーに時間に関わる「た」「ている」が多い。
また、機能と関わって多く使われる語もあった。格助詞の「の」は複数の話題で特徴語とされたが、コロケーションは「日本の」「中国の」が多く、すなわち「比較」の場面で多く用いられると言える。また、日本語の理由表現としては「から」と「ので」の２つが存在するが、「から」は全ての話題に頻出し、結果特徴語としては抽出されなかった。一方で、「ので」は終了部の特徴語として抽出された。これは従来言われている通り、「ので」の方が待遇度が高く、すなわち次のセッションの約束をする際に、相手への配慮を表しながら理由を述べるさいに特徴的に使われている。

（9）日中 Skype 会話コーパスはこれまでにない特徴を持つコーパスであり、特に中国語母語話者と日本語母語話者のある語の使い方を比較する時には便利である。
さらに、話題別に特徴語を抽出することで、実質語のみならず機能語も話題によって偏るのではないかという知見を得ることができた。これはこれまでほとんど考慮されてこなかった視点であるが、この関係が明らかになれば、教材開発や教室活動の検討に大いに貢献できる。しかしながら、『日中 Skype 会話コーパス』は厳密に話題が統制されているとはいい難く、また、学習者コーパスならではの特徴も存在する。母語話者の会話においても、機能語は話題の影響を受けているのかを解明することは急務であるといえ、平成 30 年度からは新たな研究体制で新しいコーパスを作り、本科研で得た仮説を検証していく。

＜引用文献＞
①中俣尚己・漆田彩・小野真依子・北見友香・竹原英里、Skype を活用した日中会話交流プログラム、実践国文学、83 巻、2013、132(25)-109(48)
②山本和英、高橋寛治、椙澤優希、西山浩気、日本語解析システム「雪だるま」第 2 報～進捗報告と活用形態素の導入～、信学技報、Vol.116No.213,、2016、63-68
③田中牧郎・近藤明日子、教科書コーパス語彙表、言語政策に役立つ、コーパスを用いた語彙表・漢字表等の作成と活用、2011、55-63
④北村達也・冨岡洋介・川村よし子、IDF を用いた単語レベル判定システムの構築と検証、日本語教育方法研究会誌 16 巻 1 号、2009、pp.52-53

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　４　件）
① NAKAMATA, Naoki 、 Vocabulary Depends on Topic, and So Does Grammar 、 Journal of Japanese Linguistics、査読有、Vol.45、in printing
②陳建明・中俣尚己、程度を表す副詞の日

中対照と日本語学習者コーパスの分析―話し言葉と書き言葉の違いに注目して―、庵功雄・杉村泰・建石始・中俣尚己・劉志偉(編)『中国語話者のための日本語教育文法をもとめて』、日中言語文化出版社、査読なし、2017、pp.95-124
③中俣尚己、コーパスから見る日本語教科書、吉田英幸・本田弘之(編)『日本語教材研究の視点』、くろしお出版、査読なし、2016、pp.92-114
④中俣尚己、学習者と母語話者の使用語彙の違い―『日中 Skype 会話コーパス』を用いて―、日本語／日本語教育研究、査読有、７巻、2016、pp.21-34

〔学会発表〕（計　６　件）
①中俣尚己、学習者コーパスから見るとりたて表現の使用状況―母語話者と比較して―、NINJAL シンポジウム、2017
②中俣尚己、母語話者の自動詞他動詞の使用状況、公開シンポジウム「日本語自動詞・他動詞を考える」、2017
③中俣尚己、語彙は話題に従属する、文法も話題に従属する、第 10 回実用日本語言語学国際会議、2017
④中俣尚己、真正性のある接触場面会話コーパスを用いた話題別特徴語の抽出－ポップ・カルチャーの場合－、日本語教育学会 2016 年度春季大会、2016
⑤中俣尚己、接触場面における学習者と母語話者の語彙はどこが異なるのか？―「日中 Skype 会話コーパス」の分析―、日本語／日本語教育研究会　第７回大会、2015
⑥中俣尚己、「日中 Skype 会話コーパス」を用いた話題別語彙の抽出―「食」の場合―、第８回コーパス日本語学ワークショップ、2015

〔図書〕（計　１　件）
森篤嗣(編)、『日本語教育への応用』、朝倉書店、印刷中　（「学習者話し言葉コーパス分析」の章を担当）

〔その他〕
『日中 Skype 会話コーパス』の配布サイト
http://nakamata.info/database.html

６．研究組織
(1)研究代表者
中俣　尚己・(NAKAMATA, Naoki)
京都教育大学・教育学部・准教授
　研究者番号：００５９８５１８