

機関番号：12603

研究種目：基盤研究(B)

研究期間：2008～2010

課題番号：20320072

研究課題名(和文) 人間の相互作用研究のための多言語会話コーパスの構築とその語用論的分析方法の開発

研究課題名(英文) The construction of multilingual corpus of natural conversations and the development of methodology of its pragmatic analysis, for the human interaction research.

研究代表者 宇佐美 まゆみ (USAMI MAYUMI)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号：90255894

研究成果の概要(和文)：

本研究の主な成果は、以下の5点である。

- (1) 自然会話分析のための文字化入力支援及び基本的な分析項目の自動集計ツールの作成
- (2) 「BTSJによる日本語話し言葉コーパス(2011年版)」の作成
- (3) BTSJによる多言語話し言葉コーパスのデータベース構築
- (4) 話し言葉コーパスの無料配布、及び、研究成果をホームページに発表することによって、他の研究者のニーズに資するようにした。
- (5) 談話研究、中間言語語用論等にかかわる基礎的研究を行い、語用論的分析方法、を精査し、「会話教育」に生かすための提言を行った。

研究成果の概要(英文)：

The main results of the present study are the following five points.

- (1) An input support system for Basic-Transcription-System-for-Japanese(BTSJ), and an automatic tabulation tool for descriptive statistics of conversation analyses were made.
- (2) A corpus of spoken Japanese by BTSJ (2011 versions) was made.
- (3) A database for constructing multilingual spoken language corpus by BTSJ was made.
- (4) A corpus of spoken Japanese by BTSJ (2011 versions) was uploaded to the homepage for contributing to other researchers' needs.
- (5) The basic research related to natural conversation analysis and interlanguage pragmatics were conducted for contributing to the Japanese conversation education.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	6,500,000	1,950,000	8,450,000
2009年度	4,400,000	1,320,000	5,720,000
2010年度	4,200,000	1,260,000	5,460,000
総計	15,100,000	4,530,000	19,630,000

研究分野：日本語教育

科研費の分科・細目：言語学・日本語教育

キーワード：(1)コーパス言語学(2)会話分析(3)語用論(4)談話研究(5)異文化間コミュニケーション(6)言語社会心理学(7)インタラクション(8)研究方法論

## 1. 研究開始当初の背景

従来の国内外の自然会話研究の多くは、少量のデータを扱う定性的なアプローチに留

まっており、会話参加者間の社会的関係、既知度、会話の場面性などの社会的要因を考慮して収集した大規模な談話データに基づく

「定量的分析」はほとんど行われていないと言っても過言ではない。また、昨今は、「言語コーパス」の作成もなされるようになってきたが、そのいずれも、語数や品詞、形態素の数を算出したり、言語形式の共起関係を抽出するなど、文法理論や自動翻訳機制作に貢献しようとする言語学的、工学的目的に即したものが多く、会話における微妙な「間」が、対人コミュニケーション上、どのような機能を持つかといったような「人間の相互作用研究」に適し、且つ、定量的分析・定性的分析の双方が可能な「自然会話コーパス」は、ほとんどないと言っても過言ではない。

このような状況を鑑み、本研究代表者らは、平成7-8年度、13-14年度、及び、平成15年度-18年度の3回、8年間に渡って心理学・統計学・情報工学を援用しながら、自然会話分析の定性的分析のみならず、定量的分析にも適した会話データの文字化と解析のシステムである「基本的な文字化の原則 (BTSJ: Basic Transcription System for Japanese)」を開発し、自然会話のデータベース化を行うとともに、データ収集や分析の方法論（自然会話分析への言語社会心理学的アプローチ）を構築し、自然会話コーパスの公開と無料配布を行ってきた

(<http://www.tufs.ac.jp/ts/personal/usamiken/kaken2007corporata.htm>)。

また、その成果は、報告書やHPにて公開してきた

(<http://www.tufs.ac.jp/ts/personal/usamiken/danwaindex.htm>)。

本研究では、これら確固たる科学的方法論に基づいて収集した既存の自然会話データをベースに、新規にもデータを収集し、より大規模で体系的な「多言語自然会話コーパス」を構築することによって、他の研究者のニーズに資することを企図する。また、人間の相互作用研究に適したコーパスを用いた科学的語用論研究の方法論を開発、確立することを目指す。それらに基づいて、本コーパスを用いた談話理論の基礎的研究を行い、その成果を「会話教育」に生かすための提言を行うなど、日本語教育との有機的統合を図る。

## 2. 研究の目的

本研究の目的は、以下の通りである。

- (1) 「人間の相互作用の研究」に適した科学的語用論研究の方法論を開発する。
- (2) 「人間の相互作用の研究」に適した質の高い「話し言葉コーパス」を構築する。
- (3) より大規模で体系的な「話し言葉コーパス」を構築することによって、他の研究者のニーズに資するとともに、この分野の発展に寄与する。
- (4) コーパスを活用した談話研究、中間言語語用論等にかかわる基礎的研究を行い、そ

の成果を「会話教育」に生かすための提言を行うなどして、日本語教育との有機的統合を図り、日本語教育の体系化と、談話理論に基づく効果的な会話教育実践に貢献する。

より具体的には、以下のことを行う。

- (1) 文字化された既存のデータを再検討、修正し、データ提供者の背景的情報などの会話外の情報も組み込んだ質の高い多言語自然会話コーパスを完成させる。
- (2) 構築した「人間の相互作用研究に適した」多言語自然会話コーパスを、ホームページ上で公開し、利用希望者に無料でダウンロードできるようにすることによって、多くの研究者がさまざまな角度から分析を加える機会を拡大し、談話研究、対照談話研究、異文化コミュニケーション研究、中間言語研究等の発展に資することを目指す。
- (3) 「基本的な文字化の原則」を、ホームページ上に公開することによって、対照談話研究の発展に寄与することを目指す。
- (4) 本研究全体を通しての調査・分析結果の総合的分析を行い、日本語教育に何らかの形で提言を行う。
- (5) 多言語自然会話コーパスを用いて、談話研究、異文化コミュニケーション研究、中間言語語用論、会話教育についての研究を行う。
- (6) 上記研究の成果を広く共有化するために、本科研のホームページを作成し、多言語自然会話コーパスは、これまでと同様、利用希望者には、ホームページを通して無料でダウンロードできるようにする。他の研究成果も、適宜、HP上で公開する。

## 3. 研究の方法

研究代表者の統括の下に、研究分担者、連携研究者が中心となって、各担当部をとりまとめ、海外研究協力者との連携や、大学院生の研究協力者への連絡などを行いながら、研究を遂行していく。

本研究は、これまで研究代表者、連携研究者らが長年蓄積してきた「談話行動の基礎的研究」、「自然会話のコーパス構築」の実績と成果を発展させる形で、具体的には、以下の6点を行う。

- (1) 膨大な時間と労力がかかる自然会話データの分析の効率化、精緻化を促進するため、科学的で共有しやすく、人間の相互作用の分析に適した会話分析ツールとして開発された「基本的な文字化の原則 (BTSJ: Basic Transcription System for Japanese)」のさらなる改良と、「文字化入力の一部自動化」(一部、試作済み)を完成させる。
- (2) BTSJの改訂版に基づく、各言語の特徴を十分考慮に入れた6言語の「基本的な文

字化の原則」、すなわち、韓国語版(BTSK)、中国語版(BTSC)、英語版(BTSE)の改定、タイ語版(BTST)、スペイン語版(BTSS)、イタリア語版(BTSD)の開発を行う。

(3) BTSJの各国語版に基づいて「多言語自然会話コーパス」を構築する。上記7カ国語は、既にデータ収集してプールしてある言語である。

(この3点によって会話データの文字化作業の労力が大幅に削減されるとともに、言語を超えて、大量のデータを統一した原則によって整備・蓄積していくことが可能になり、異なる言語の談話行動の対照研究、中間言語研究にも大きく貢献することができる。)

(4) これらの多言語自然会話コーパスを活用した科学的語用論的研究の方法論を開発する。

- ① 文字化された既存のデータを再検討、修正し、データ提供者の背景の情報などの会話外の情報も組み込んだ質の高い多言語自然会話コーパスを完成させる。
- ② 構築した「人間の相互作用研究に適した」多言語自然会話コーパスを、ホームページ上で公開し、利用希望者に無料でダウンロードできるようにすることによって、多くの研究者がさまざまな角度から分析を加える機会を拡大し、談話研究、対照談話研究、異文化コミュニケーション研究、中間言語研究等の発展に資することを目指す。
- ③ 「基本的な文字化の原則」を、ホームページ上に公開することによって、対照談話研究の発展に寄与することを目指す。
- ④ 本研究全体を通しての調査・分析結果の総合的分析を行い、日本語教育に何らかの形で提言を行う。

#### 4. 研究成果

本研究の主な成果は、以下の3点である。

- (1) 自然会話分析のための文字化入力支援及び基本的な分析項目の自動集計ツールの作成
- (2) 「BTSJによる日本語話し言葉コーパス(2011年版)」の作成
- (3) BTSJによる会話コーパスのデータベース構築

以下、各成果について説明する。

(1) 自然会話分析のための文字化入力支援及び基本的な分析項目の自動集計ツールの作成

「人間の相互作用の研究」に適した科学的語用論研究の方法論開発の一環として、自然

会話の文字化資料作成の効率化を図るとともに、基本的な分析項目の自動集計を行うことによって、定量的な分析を効率的かつ精緻に進めるためのツールを作成した。

このツールは、「基本的な文字化の原則

(Basic Transcription System for Japanese: BTSJ)」(以下、BTSJ)のルールに対応した文字化入力支援ツールで、汎用性を考えて、Microsoft Excelを利用して作成されており、現在のところ、Excel 2003とExcel 2007に対応している。

以下に、主な機能とその特徴、長所をまとめる。

<文字化入力支援>

① 文字化に用いる記号を入力フォームから選択して簡単に挿入できる。

発話内容には、実際に発話された内容だけでなく、発話の重複などのような音声的情報や文脈情報なども記す必要がある。これらの情報を示す記号を簡便かつ正確に入力できることは、文字化作業の効率化につながる。

② 文字化資料の入力記号の自動修正やエラーチェックを行う。

大量のデータを文字化する際には、半角・全角の混用など、多少の入力ミスは避けられないが、記号の半角・全角を自動的に統一したり、その他の記号のエラーチェックによって、入力ミスをメッセージ付きで抽出する。それによって、手動による入力ミスの修正作業の時間を格段に節約し、より完全な文字化資料を短時間で作成することができる。

③ 分析の単位である発話文の通し番号を自動付与する。

BTSJの「発話文番号」付与のルールは独特であり、手動の修正にはかなり時間がかかる。そのため、発話文番号を自動付与できるようにすることによって、文字化資料への入力ミスを防ぐとともに、文字化時間を格段に短縮した。

<記述統計のための基本的な分析項目の自動集計>

④ 各会話における「総発話文数」、話者ごとの発話文数の頻度や割合、話者交替の頻度や割合などの「会話の基本的な情報」を自動算出する。

⑤ 各会話における、分析項目ごとの頻度と割合、及び、複数の会話の分析項目の頻度や割合の平均値・最大値・最小値・標準偏差などの基本的な数値を、各項目の総計、話者ごとの総計などのいくつかの観点から算出する。

このような会話の「基本的な情報」を自動集計するとともに、ひとつの分析項目について

て多角的な観点から分析するための情報を自動的に算出することにより、定量的な分析に必要な処理を短時間で正確に行うことができる。

BTSJは総合的会話分析(宇佐美 2008)に適するようなものとして開発されたが、その名の通り「基本的な文字化の原則」である。そのため、BTSJとは異なる方法で文字化された資料も、本ツールを用いて部分的に修正を行うことで、分析項目の自動集計が可能となる。つまり、本ツールによって、既に文字化された資料を定量的な分析のための資料としても活用できるように変換し、応用することができる。

会話の分析の基本的な作業を短時間で正確に行えるツールの開発は、定性的・定量的分析の双方を含む会話の「総合的研究」の基礎を支えるものである。個人の研究の労力を格段に節約することは、個々の研究を促進し、会話の定性的分析に加えて、定量的な分析を新たに試みる研究者の助けともなるだろう。本ツールを活用して、自然会話データを扱う研究の時間と労力を節約することは、自然会話をデータとする研究のさらなる活性化を促し、この分野全体の発展にも寄与するものであると考える。

また、本ツールの作成に伴って、BTSJの記号の一部を変更したため、「改訂版：基本的な文字化の原則(Basic Transcription for Japanese: BTSJ) 2011年版」を作成した。

## (2) 「BTSJ による日本語話し言葉コーパス(2011年版)」の作成

2003年及び2007年に公開した会話コーパスを統合し、新たなデータを追加した『BTSJによる日本語話し言葉コーパス 2011年版』を作成した。コーパスに収録されている会話数は全294会話、総時間は3983分5秒(約66時間)である。

宇佐美まゆみ研究室では、多様な場面・言語(日本語、韓国語、中国語、英語など)の自然会話データを収集し、膨大な時間と労力を投入して『BTS(Basic Transcription System)による多言語話し言葉コーパス』の構築に取り組んできた。研究成果として公開していたコーパスは、以下の3つである。

- ① 『BTSによる多言語話し言葉コーパス—日本語会話1(日本語母語話者同士の会話) 2007年版』116会話、1435分54秒(約24時間)
- ② 『BTSによる多言語話し言葉コーパス—日本語会話2(日本語母語話と学習者の会話) 2007年版』37会話、691分11秒(約11時間)
- ③ 『BTSJによる日本語話し言葉コーパス—日本語会話1(初対面・友人、雑談・討論・誘い)』99会話、1604分(約27時間)

『BTSJによる日本語話し言葉コーパス 2011年版』は、上記①から③のコーパスに、新たに44会話252分(約4時間)の文字化資料・音声データが追加されている。また、①から③では、9会話分の文字化資料にのみ音声データが付いていたが、新たに87会話について、音声データを追加、公開した。

また、本コーパスの整備にあたっては、上記(1)の、自然会話分析のための文字化入力支援及び基本的な分析項目の自動集計ツールを用いて文字化資料の校正を行い、記号などの表記を「改訂版：BTSJ2011年版」に改めた。

## (3) BTSJによる会話コーパスのデータベース構築

文字化資料及び音声などの会話データそのものだけでなく、話者の関係、会話の状況、インストラクション、データ提供者の背景の情報などもコーパスに体系的に含む統制された「質の良い」データの構築をめざし、コーパスに付与する情報構成などを検討、拡充した。

この「BTSJによる会話コーパスのデータベース」では、文字起こしや背景情報の入力などの作業中の未完成データと、一般公開を行なうことができる文字化資料として完成したデータの一元的管理を行うことができるようにし、作業効率をあげるようにした。

## 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11件)

伊集院郁子・高橋圭子(2010)「日本語の意見文に用いられる文末のモダリティー—日本・中国・韓国語母語話者の比較—」『東京外国語大学留学生日本語教育センター論集』、13-27. 査読無

宇佐美まゆみ(2009)「『伝達意図の達成度』『ポライトネスの適切性』『言語行動の洗練度』から捉えるオーラル・プロフィール」、鎌田修・山内博之・堤良一編『プロフィールと日本語教育』、ひつじ書房：33-67. 査読有

宇佐美まゆみ(2009)「視点としての日本語教育学—日本語教育学の新しいパラダイム—」『2009年度「台湾日本語教育研究」国際シンポジウム—日本語教育のジャンルのひろがり求めて—』、4-22. 査読無

宇佐美まゆみ・木林理恵(2009)「自然会話分析のための文字化入力支援及び基本的な分析項目の自動集計ツールとその使い方」『日本語学会2009年度秋季大会予

稿集』査読有  
宇佐美まゆみ(2008)「相互作用と学習—ディスコース・ポライトネス理論の観点から」『講座社会言語科学 第4巻 教育・学習』4、ひつじ書房、150-181. 査読無

[学会発表] (計 11 件)

宇佐美まゆみ(2010)「話し言葉の分析のための文字化入力・集計支援ソフトの紹介とその使い方—レストランでの注文場面のデータを例に—」『日本語プロフィシエンシー研究会、函館国際シンポジウム』、函館国際ホテル、2010年7月17日.

Usami, Mayumi (2010) Discourse Politeness Theory and Language Education Studies. Invited Lecture at School of Modern Languages and Cultures, The University of Hong Kong, The University of Hong Kong, 2010年4月30日.

宇佐美まゆみ(2009)「視点としての日本語教育学—日本語教育学の新しいパラダイム—」(基調講演)『2009年度「台湾日本語教育研究」国際シンポジウム—日本語教育のジャンルのひろがりを探求めて—』、誠宜大学、台湾、2009年12月6日.

Usami, Mayumi (2009) Japanese Women's language and politeness. Symposium on Gender Studies across Languages and Disciplines, Western Michigan University, 2009年9月21日.

Usami, Mayumi (2008b) "New Trends in Politeness Studies and Discourse Politeness Theory" Manoa. EALL Lecture Series (ハワイ大学東アジア言語文学部レクチャー・シリーズ 2008年9月11日ハワイ大学マノア校), Center for Korean, Studies Auditorium, University of Hawaii at Manoa.

Usami, Mayumi (2008a) The non-use of Shujin as an indicator of Japanese women's identity. Panel Organized by Norio Ohta, Identity and Globalization - multiple identities in the making. ACS Crossroads, University of the West Indies, Kingston, Jamaica. 2008年7月4日

宇佐美まゆみ(2008)「ディスコース・ポライトネス理論と日本語教育」特別講演会、ヨーク大学、カナダ、2008年6月3日.

[その他]

ホームページ等

<http://www.tufs.ac.jp/ts/personal/usami-ken/2011/index.html>

## 6. 研究組織

### (1) 研究代表者

宇佐美まゆみ (USAMI MAYUMI)  
東京外国語大学・大学院総合国際学研究院・教授  
研究者番号：90255894

### (2) 研究分担者

伊集院郁子 (IJUIN IKUKO)  
東京外国語大学・留学生日本語教育センター・講師  
研究者番号：20436661

### (3) 連携研究者

西郡仁朗 (NISIGORI JIRO)  
首都大学東京・オープン・ユニバーシティ・教授  
研究者番号：20228175