国立国語研究所学術情報リポジトリ

The significance of developing teaching materials based on an analysis of BTSJ Japanese Conversation Corpus : Focusing on teaching incomplete utterances and co-constructed utterances

メタデータ	言語: jpn 出版者: 公開日: 2020-12-18 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://doi.org/10.15084/00003074

論文１

自然会話コーパスの分析とその教材化の意義

－NCRB で教える「中途終了型発話」と「共同発話文」を中心に－

宇佐美　まゆみ
国立国語研究所

要旨

本稿では、まず、多様な場面における事前にシナリオのない自然会話を集めた『BTSJ による日本語話し言葉コーパス』（宇佐美、2011）について、簡単にその特徴を紹介する。その上で、「中途終了型発話」や「共同発話文」を取り上げ、それらが実際の会話の中でいかに使われ、どのような機能を生んでいるのかを分析する。また、日本語母語話者が無意識のうちにも、中途終了型発話や共同発話文という会話のスタイルを利用して、相手に応じた様々な働きかけをしていることを示し、欧州の日本語学習者のように、外国語として日本語を学んでいる学習者にとっては、既存の教科書だけからでは、これらの現象を学ぶことは、不可能であることを述べる。そして、このような現状の改善のために開発してきた多機能データベース「自然会話リソースバンク（NCRB：Natural Conversation Resource Bank:）」に格納された共同構築型「自然会話を素材とする WEB 教材」とその活用方法を紹介する。

【キーワード】『BTSJ による日本語話し言葉コーパス』、自然会話リソースバンク（NCRB）、中途終了型発話、共同発話文、多機能データベース

Keywords: BTSJ Japanese Conversation Corpus (BTSJ-JCC), NCRB (Natural Conversation Resource Bank), Incomplete utterances, Co-constructed utterances, Multi-functional database.

1　はじめに

「話し言葉コーパス」の構築は、録音・録画などのデータ収集や文字化作業、プライバシー保護作業等に膨大な時間と労力がかかることが一因して、書き言葉コーパスに開発の遅れをとっていた。しかし、「語用論的研究」やその日本語教育、異文化間コミュニケーション教育への応用の需要が高まるとともに、「話し言葉コーパス」の必要性も叫ばれるようになってきていた。（Romero-Trillo, 2008 等；宇佐美, 2003, 2013；宇佐美・中俣, 2013 等）。

最近は、様々な目的に基づくコーパスの構築が増えてはきているが、同時発話、割り込み、沈黙等の情報が付与されたものは皆無に等しく、語用論的分析には適さないコーパスがほとんどであると言っても過言ではない。日本語教育に関係の深い、いわゆる「学習者コーパス」も、未だ語彙や文法項目の習得研究を主目的とするものが多いため、文字化の原則は、比較的簡素で、語用論的研究に必要なオーバーラップや沈黙等の情報が付与されていないものが多い。

一方、会話分析（Conversation Analysis: CA）で使われている文字化の原則は、語用論的分析にも適用可能な詳細な情報が付与されているものの、少数の会話の定性的な分析には適していても、より数の多い会話データの定量的分析には適さない形になっている。そのため、定性的分析だけではなく、定量的分析も可能にし、研究者間で共有できる話し言葉コーパスの構築の必要性が叫ばれていた。コミュニケーションの語用論的研究のためには、いわゆる「コーパス言語学」で扱われているような大量の「テキスト」を編んだものだけではなく、比較的少量でも、話者の関係や話題などの諸条件を統制して収集された「会話（音声・動画）」とともに、同時発話、割り込み、沈黙等の語用論的分析に必要な情報が付与された「トランスクリプト（文字化資料）」が収録されたコーパスが必須である。

2 『BTSJ による日本語話し言葉コーパス』とは？

このような状況を鑑み、語用論的アプローチを含む「総合的会話分析」（宇佐美、2008）の方法論や、人間同士の相互作用を、定量的・定性的双方から分析するのに適する「文字化の原則」として『基本的な文字化の原則（BTSJ：Basic Transcription System for Japanese）」（宇佐美, 改訂最新版 2015, 初版は 1997）を開発し、その BTSJ による文字化資料と音声（一部）つきコーパスである『BTSJ による日本語話し言葉コーパス（トランスクリプト・音声）2011 年版』(宇佐美監修 2011）を構築した（以降『BTSJ 話し言葉コーパス』)。また、その後、『BTSJ 文字化入力支援・自動集計・複数ファイル自動集計システムセット 2015 年改訂版』（以降、『BTSJ システムセット』図 1）を開発して、少数の会話の定性的分析だけでなく、比較的多数の会話の基本的情報も、「自動集計機能」を用いることによって短時間で算出できるようにした。また、これによって、コーパスのデータの特定の形式の頻度や割合を算出した上で語用論的分析や考察を行うための時間と労力を格段に短縮した。「BTSJ 文字化資料」と音声・動画つきの『BTSJ 話し言葉コーパス』をデータベース化し、さらには、そこに、「自然会話の動画」と「文字化資料」を中心とする共同構築型「自然会話を素材とする WEB 教材」を作成するための「自然会話教材作成支援システム」機能を付与したのが、以下の 4. で紹介する『自然会話リソースバンク（NCRB：Natural Conversation Resource Bank)』である（宇佐美 2013)。

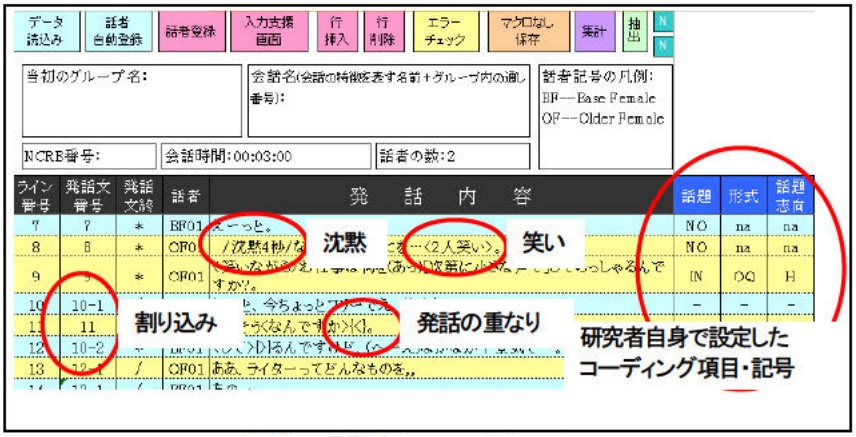

図 1　BTSJ システムセット

『BTSJ による日本語話し言葉コーパス』は、事前にシナリオのない自然な会話（1 会話約 15 分～20 分）を 294 会話収録しており、その約半数には、音声もついている。2 者間会話を中心とし、初対面、友人同士の雑談や討論場面、教師と学生の論文指導場面などが収録されている。（詳細は、http://ninjal-usamilab.info/btsj_corpus/を参照のこと）また、下記より申し込むことによって、誰でも無料で取得でき、研究や教育に利用できる。https://ninjal-usamilab.info/form/corpus/index.html (東外大 HP から移設。尚、東外大ページからも申し込み可能)

3　自然会話に見られる「中途終了型発話」や「共同発話文」

ここでは、自然会話にはよく表れているにもかかわらず、教科書などでは、あまり扱われることのなかった「中途終了型発話」と「共同発話文」の例をあげ、これらの現象は、『BTSJ 話し言葉コーパス』を用いて分析することによって、より体系的に捉えることができること

を示す。また、これらを『NCRB 自然会話教材』でいかに扱うかを 4.で説明する。

以下の例は、『BTSJ による日本語話し言葉コーパス』会話番号 52-3（52 は、通し番号。3 は、収録されているフォルダ番号）からのものである。以下の説明では、ライン番号で言及する。

最初の番号は、ライン番号、次の番号は、発話文番号、*は、発話文終了、/は、発話文未終了

例 1　4 は、「中途終了型発話」の後、相手のフォローの発話が入る形

3	3	*	JTM07	今日もらったのは。
4	4	*	JSM01	今日もらっ、あ、今日…。
5	5	*	JTM07	普通でいいんだよ<笑い><2 人笑い>。

例 2　16 は、15 の発話に「助け船」を出している。「共同発話文」の形にもなっている。

15	15	*	JSM01	あったんですけど(ん)、と、とりあえず、は、あの…。
16	16	*	JTM07	8。
17	17	*	JSM01	8 にしました(はい)、はい。

例 3　44 は、43 の発話の「助け舟」であり、「共同発話文」の形にもなっている。

43	41	*	JSM01	今日は、まず、あの…。
44	42	*	JTM07	続きね。
45	43	*	JSM01	続きですね、はい。

例 4　189 は、「中途終了型発話」と捉えて、助け舟を出したが、JSM01 が、文を自身で完結させた例。「共同発話文」の形にもなっている。

188	178-1	/	JSM01	この下の…,,
189	179	*	JTM07	下の方。
190	178-2	*	JSM01	ですか、あ。

例 1 のように、「中途終了型」の形で発話が終わる場合もあるが、例 2，3 のように、発話に「言い淀み」が出た段階で、相手の話者が、「助け舟」を出す形の発話を行い、それが、先の話者が始めた発話を別の話者が完了させるという「共同発話文」の形になることが多い。例 4 のように、聞き手の側が確認の「あいづち」的な発話をすることも多いが、それも、無意識にせよ、文の形が整うような形で行われることが多い。

会話における機能としては、「言い淀み」にすばやく反応して、助け舟を出したり、あいづち的な発話は、相手の発話に合わせたり、会話に積極的に参加していることを示しているという意味で、「ポジティブ・ポライトネス」になる。よって、このような言語行動が不十分だと、会話が盛り上がらなかったり、聞き手の反応として物足りなさや不自然さを醸し出すことにつながってしまう。このような「さりげない聞き手行動」ができると、日本語の会話として、自然なやりとりらしくなると考えられる。しかし、上記の例のような中途終了型発話や、共同発話文の例は、既存の教科書には見られず、また、実際の会話を行っている最中は、これらの現象に気づいたり、そのやり方を学んだりすることは不可能である。そのため、このような「自然会話を素材とする教材」があれば、学習者がこういう現象に着目できるよう

になり、通常の教科書では扱えないことが扱えるようになる。

4 『自然会話リソースバンク（NCRB：Natural Conversation Resource Bank）』に格納された共同構築型「自然会話を素材とする WEB 教材」について

このように、自然会話データは、研究のみならず、コミュニケーション教育のための教材（materials）にも成り得る（宇佐美, 2012）。しかし、「自然会話を素材とする教材」の作成には、文字化に時間がかかる上に、さらに「教材としての解説」等も加えていかなければならない。そのため、個々の日本語教師が、様々な「自然会話を素材とする WEB 教材」を作成しやすくするには、「自然会話教材作成支援システム」のようなものを開発して作成の一部を自動化し、教材作成の時間と労力を削減できると生産的である。NCRB は、このような必要性とアイデアを踏まえて構築された、新しいタイプの「共同構築型多機能データベース」であり、また、「共同構築型 WEB 教材リソースバンク」でもある。

本システムは、今のところ、基本的に、登録制のメンバーが皆で共同構築していく形のものを想定している。NCRB で作成した教材は、主に WEB 上での独習用を想定しているが、必要やレベルに応じて、日本語の授業中に、今回取り上げたような自然会話に多い「中途終了型発話」や「共同発話文」の他にも、自然な日本語会話の音調やテンポの学習、討論の材料などとして、多様に活用することを想定している。また、この自然会話教材は、同じ素材を、初級から超級に至るまでのすべてのレベルの学習者が、各自のレベルや興味に応じて活用できると考えている。授業で使う場合は、同じ場面におけるスピーチレベルの相手による使い分けなど、実際の場面を体験したり見たりしなければ理解しにくい「人間関係に応じたコミュニケーションの方法」について議論させるなど、様々な活用法がある。さらには、日本語の「普通の人の」自然な会話に触れさせることによって、海外の学習者の学習動機を高めることができるだろう。

5 NCRB における会話データの登録とコーパスとしての活用法

NCRB の利用環境は、OS : Windows7, 8、ブラウザ : InternetExplorer9～11, Google Chrome で、登録可能なデータの形式は、mp4（動画）、mp3（音声）ファイルである。構成は、大きく、「自然会話データを使った研究」と「自然会話を素材とする教材」の 2 つに分かれている。それぞれ、「登録や作成」と「利用のみ」で入り口が分かれており、「自然会話を使った研究」は、「データを登録したい方はこちら」及び「データを利用したい方はこちら」、また、「自然会話を素材とする教材」は、「教材を作成したい方はこちら」及び「教材を利用したい方はこちら」と、全体として 4 つの入り口がある（次頁、図 2）。会話の登録と教材の作成には、利用申請により取得できるアクセス用の ID とパスワードが必要である。

会話データ（音声・会話スクリプト、それらに関する情報）は NCRB トップページ（次頁、図 2）より、「自然会話データを使った研究」の「データを登録したい方はこちら」から登録する。作業の流れは、「音声・動画のアップロード」→「音声・動画情報の入力」→「話者情報の入力」→「会話情報の入力」→「会話スクリプトの入力」となる。各作業は「音声・動画」、「会話」、「会話グループ」という 3 つの画面で行い、上部のタブを切り替えることで画面を移動できる。

会話スクリプトは NCRB 上でも、動画や音声を聞きながら入力していくこともできるが

（図 3)、「BTSJ システムセット」を用いて文字化資料を整えた後、NCRB にアップロードすることもできる（文字化入力支援機能が多い後者の方法を用いることを推奨する)。また、会話スクリプト画面では動画や音声にタイムスタンプを押すことができるため、分析の際に動画や音声を再生しやすくなる。

このようにして登録した 1 つ 1 つの会話は、「会話グループ作成機能」を用いてグループ化することもできる。自分のデータを 1 つのグループにまとめることもできるし、検索機能を用いて、「友人同士の雑談の会話」等、研究目的に応じた条件で検索し、該当する会話のみをグループ化して保存することもできる。このように、自然会話を用いて研究を行う場合は、同様の条件の複数の会話をグループ化し、定量的に分析することが可能である。

図 2　NCRB トップページ

図 3　会話スクリプト入力画面

NCRB は、2.で紹介した『BTSJ システムセット』と連携している。NCRB 上で決められた作業手順通りにデータ情報を入力すると、NCRB 番号が自動的に付与される。このデータをダウンロードし、『BTSJ システムセット』を用いて文字化資料を整えることによって、通し番号の間違いや情報の入力漏れといった、データ整備の際に起きやすいケアレスミスを防ぐことができる。さらに、『BTSJ システムセット』では、研究のためのコーディングや基本的記述統計の自動集計ができる。NCRB の検索機能を用いて自分の研究目的に合うデータを複数抽出し、一括ダウンロードして BTSJ システムセットの自動集計機能等を用いることによって、会話の定量的な分析を効率よく行うことが可能になる。

6　おわりに

これらの機能を十分に活用して、今回例にあげた「中途終了型発話」や「共同発話文」のみならず、自然会話に特徴的なあいづち、フィラー、ディスコースマーカーなどの周辺言語の使用実態を多角的に明らかにしていくことは、欧州など、海外で日本語を学ぶ学習者の「自然なコミュニケーション能力の養成」のための基礎となる。今後、益々『BTSJ 話し言葉コーパス』のみならず、「自然会話」を使った分析が日本語教育に生かされることを期待したい。

＜参考文献＞

Romero-Trillo, Jesús (ed.) (2008). *Pragmatics and Corpus Linguistics: A mutualistic entente.* Mouton Series in Pragmatics 2 . n.p: Mouton de Gruyter.

宇佐美まゆみ（2003）「改訂版: 基本的な文字化の原則 (Basic Transcription System for

Japanese: BTSJ)」『多文化共生社会における異文化コミュニケーション教育のための基礎的研究』平成 13-14 年度科学研究費補助金基盤研究 C(2) (課題番号: 13680351)(研究代表者: 宇佐美まゆみ) 研究成果報告書: 4-21. 2015 年改定版は、下記からダウンロードできる。(http://www.tufs.ac.jp/ts/personal/usamiken/btsj.htm)

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」西原鈴子・西郡仁朗編『講座社会言語科学 第 4 巻 教育・学習』pp. 150-181 ひつじ書房.

宇佐美まゆみ (2012)「母語話者には意識できない日本語コミュニケーション」野田尚史編『日本語教育のためのコミュニケーシヨン研究』pp. 63-82 くろしお出版

宇佐美まゆみ (2013)「会話データの作成・分析―「総合的会話分析」と「基本的な文字化の原則(Basic Transcription System for Japanese: BTSJ)」」『日本語学』第 32 号, 第 14 巻, pp.132-147, 明治書院

宇佐美まゆみ・中俣尚己 (2013)「『BTSJ による日本語話し言葉コーパス（トランスクリプト・音声）2011 年版』の設計と特性について」『第 3 回 コーパス日本語学ワークショップ予稿集』, pp. 217-228, 国立国語研究所 言語資源研究系・コーパス開発センター.

宇佐美まゆみ (2013)「NCRB（Natural Conversation Resource Bank）の開発とその意義についてーこれからのコーパスのあり方とその研究・教育への活用法への一提案ー」『第８回日本語実用言語学国際会議（ICPLJ8）Conference Handbook』, pp. 128-131, 国立国語研究所.

宇佐美まゆみ (2015)「BTSJ 文字化入力支援・自動集計・複数ファイル自動集計システムセット 2015 年改訂版」『自然会話リソースバンク構築による世界的教材共有ネットワーク実現のための総合的研究』平成 23-26 年度科学研究費補助金基盤研究(A)(研究代表者：宇佐美まゆみ) (課題番号 23242027)　研究成果.

The significance of developing teaching materials based on an analysis of BTSJ-Japanese Conversation Corpus:

Focusing on teaching *incomplete utterances* and *co-constructed utterances*

Mayumi USAMI
National Institute for Japanese Language and Linguistics

Abstract

The present study analyses the functions of *incomplete utterances* and *co-constructed utterances* in natural conversations taken from BTSJ Japanese Conversation Corpus (BTSJ-JCC). Although these phenomena are commonly perceived in natural conversations, they are not found in the *textbooks*; therefore, it is difficult to integrate them into the Japanese language education. However, these two utterance types have various functions and play an important role in smooth and naturalistic communications; (i) many *incomplete utterances* have *hedging functions* (negative politeness), (ii) they trigger the addressee to complete the speaker's unfinished utterance (*co-constructed utterance*), (iii) the addressee indicates *understanding* by finishing the unfinished utterance. *Incomplete utterances* trigger addressee's positive politeness in this manner, and thus, they facilitate smooth conversation.

Since native speakers of Japanese employ these strategies according to the social factors of the addressees, it is almost impossible for the learners of Japanese in Europe to acquire them sorely from the existing Japanese textbooks. Based on these situation, a multifunctional database *Natural Conversation Resource Bank: NCRB*, a platform for collaboratively-constructed *WEB teaching materials using natural conversations* has been developed. As the NCRB has *support functions for developing teaching materials* for natural communication, the users can develop *WEB teaching materials* with explanations of these conversation strategies as well as those of grammatical explanations by following the simple illustrations built in the program.

Finally, I insist that the platform such as NCRB for utilizing teaching natural conversation resources is essential to improve the Japanese language education in Europe.