

国立国語研究所学術情報リポジトリ

『総合的会話分析』に基づく研究：『BTSJ
日本語自然会話コーパス』と『自然会話リソースバ
ンク（NCRB）』との連携に触れながら

メタデータ	言語: jpn 出版者: 公開日: 2020-10-02 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/10.15084/00003018

『総合的会話分析』に基づく研究

—『BTSJ 日本語自然会話コーパス』と『自然会話リソースバンク (NCRB)』

との連携に触れながら—

宇佐美 まゆみ
国立国語研究所

要旨

本稿では、「総合的会話分析」(宇佐美 2008)について、その趣旨と目的、方法について紹介する。またこの方法に適するように開発された文字化ルールである『基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ)』、トランスクリプト入力支援機能や分析項目の基本的記述統計量が自動集計できる『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』(講習会受講者に無償配布)、333 会話を含む『BTSJ 日本語自然会話コーパス (トランスクリプト・音声) 2018 年版』、及び、共同構築型多機能データベースである『自然会話リソースバンク (NCRB: Natural Conversation Resource Bank)』を使った「自然会話を素材とする WEB 教材」の作成方法についても紹介し、『総合的会話分析』の方法が、「会話・談話の分析」という観点からだけでなく、「日本語教育研究」と「日本語教育実践」にとって、どのような意義を持っているかを論じる。

【キーワード】 談話研究、総合的会話分析、語用論、日本語教育、自然会話コーパス
Keywords: discourse studies, integrated conversation analysis, pragmatics, Japanese language education, natural conversation corpus

1 はじめに—『総合的会話分析』の総合性—

先の『談話研究と言語教育』(宇佐美 2019)では、「談話」という用語の捉え方やそれに対するアプローチの特徴を整理し、「談話」を「談話研究」という分野横断的な、既存の分野の枠を超えた学際的研究という視野の広がりを持つものとして位置づけてきた。本稿では、広義の「談話研究」の一つとして、宇佐美 (2019) の『総合的会話分析』(宇佐美 2008)という会話を分析する方法論について、その趣旨と目的、方法について紹介する。この方法論は、(先の『談話研究と言語教育』(宇佐美 2019)において、4) 発達心理学者らによる第一言語習得研究、養育者と幼児の社会的相互作用研究の一環として行われている会話の分析、社会心理学者の研究 (Snow et al. 1990; Snow 1992; Budds, Locke & Burr 2017)と同様の実証的アプローチとして位置づけられるが、特に、「語用論的分析」に適するものとして、量的・質的、双方のアプローチを行うことと、対象とする会話それ自体だけでなく、話者間の関係や社会的属性についてもフェイスシートやフォローアップ調査などによって、分析や考察の対象に含めることを重視して開発されたものである。それが、「総合的」という用語を冠した理由の一つである。

この方法論は、『総合的会話分析』の分析手順に適するように開発された文字化のルールである「基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ)」(以降、BTSJ)、及び、トランスクリプト作成時の「入力支援機能」や、主な分析項目の基本的な記述統計量が自動集計できる機能を搭載した『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』¹ (以降、BTSJ システムセット) を用いることによって、より効率的に分析ができるようになってきている。つまり、方法論の提示に加えて、提示した方法に適した分析を効率化する「ツール」も併せて提供している。

「会話の分析」というと、言語研究の中では、質的分析を行うアプローチが多かった。その理由として、会話の相互作用や内容を量的に分析することが困難なことや、CA (会話分析) のように、もともと「言語の研究自体」を目的とするものではなかったアプローチの影響を受けているものもあるため、会話の中身 (トランスクリプト) から、相互作用を詳細に分析していくという質的手法が取られることがどうしても多くなるということがあった。

ただ、本稿では、会話の分析を行う際にも、質的アプローチと量的アプローチとの双方を行うことが必要であり、両者を有機的に融合させることは不可能ではなく、不可欠であるということを論じる。特に、言語教育の文脈では、学習者の母語による誤用の特徴を明らかにし、それに基づく指導法を工夫するなど、ある程度一般化できる研究結果を導いて実践に適用することが必要となる。つまり、個別事例の詳細な分析結果だけでは、実践に応用することは難しい。『総合的会話分析』は、ある程度、実践現場にも応用するという目的を念頭においている。そのために、会話の分析に際して、質的・量的双方からアプローチすることによって、より細やかな知見を導き、その知見を体系化できる形でまとめ、言語教育、日本語教育の現場に応用していくことを企図するアプローチである。そういう意味でも「総合的」なアプローチであるといえる。

以下では、『総合的会話分析』のより具体的な方法をまとめるとともに、この方法に適する文字化の原則である『BTSJ (Basic Transcription System for Japanese)』、トランスクリプト入力支援機能や分析項目の基本的記述統計量が自動集計できる『BTSJ システムセット』(講習会受講者に無償配布) についても併せて紹介する。また、BTSJ に基づくトランスクリプトと音声データ (一部音声なし) 333 会話を収録する『BTSJ 日本語自然会話コーパス (トランスクリプト・音声) 2018 年版』(以降、『BTSJ 自然会話コーパス』²) についても紹介し、『総合的会話分析』の方法を用いていかなる研究が可能になるかを論じる。さらには、自然会話の語用論的分析に携わる中から発想が生まれた多機能の「共同構築型自然会話データベース」で、「教材作成支援機能」も持つ「自然会話リソースバンク (Natural Conversation Resource Bank: NCRB)」(以降、NCRB) についても、簡単に紹介する。その上で、この方法が、「日本語教育研究」、及び、「日本語教育実践」にいかに関与できるかについて論じる。

また、基本的に、実証科学的アプローチをとる『総合的会話分析』は、最近の AI (人工知能) における「対話システム研究」、「ヒューマン・エージェント・インタラクション (HAI)」、「ヒューマン・ロボット・インタラクション (HRI)」における「対話研究」への一アプローチとして、情報工学分野との連携、協働が、方法論的に行いやすく、これらの分野との「学際的対話研究」にも発展する可能性が高いという利点があることについても触れたい。

2 『総合的会話分析』の趣旨と方法

『総合的会話分析』は、心理学の実証的研究手法をベースとするため、発達心理学者らによる第一言語習得研究や、社会心理学者らの方法（宇佐美 2019 参照）と同様の方法として位置づけられるだろう。ただし、『総合的会話分析』は、それに「質的分析」も重要な要素であるとして加えた方法論である。以下に、その趣旨と具体的な手順を解説する³。

2.1 『総合的会話分析』という方法論の開発の趣旨

『総合的会話分析』とは、「社会の中に生きる人間を、会話という相互作用の分析を通して探究する」ことを目的とするもので、会話自体の分析に加えて、当該の会話以外の、年齢等の社会的要因も質問紙調査やフォローアップ・アンケート等で確認し、総合的に分析するという会話の分析方法の一つである。方法論的には、「言語社会心理学的アプローチ」（宇佐美 1999）とほぼ同様の実証的手順を踏むが、「ローカル」、「グローバル」双方の観点からの分析が必須であることを強調し、一連の手順と分析の観点すべてを含む総合的な方法論として、新たに命名したものである（宇佐美 2008）。『総合的会話分析』の目的は、言語形式自体の研究ではなく、人間の相互作用に関する言語行動を分析することによって、人間関係のあり方や、対人コミュニケーションのダイナミクスや法則を明らかにすることである。ただし、言語行動の語用論的分析も含むため、言語研究にも位置づけることができる。また、量的、質的双方の分析を含むことを重視する方法論である。目的を限定するものではなく、方法の特徴と手順をまとめたものであるため、目的に応じて活用できる。教育現場で複数の学習者に適用することを目的とするのであれば、ある程度、結果が一般化できる方法論を用いる必要がある。『総合的会話分析』は、ある程度一般化できる結果を得ることを目的とし、以下のような特徴を持つ。

- (1) 条件を統制してデータを収集し、量的・質的分析を行うため、ある程度の一般化をめざす研究に適している。そのため、多くの学習者に共通する現象の分析やその指導法を検討することを目的とする日本語教育にも応用しやすい。
- (2) 語用論的分析のためには、基本的な情報としてトランスクリプトに付与されている情報に加えて、各研究者が、自身の研究目的に応じてコーディング（形式や機能の分類等）を行うことが必須である。また、語用論的分析のためのコーディングは、言語形式の分類とは異なり、研究者自身が研究目的や分析の観点に応じて明確に定義を決めた後、文脈も考慮しながら、コーディングしていくことが必須である。そのため、連動させている文字化のルールである「BTSJ」と、BTSJ方式によって文字化されたトランスクリプトに、研究者各自がコーディングを行い、その集計が行いやすくなる機能が搭載されている『BTSJ システムセット』（以下の（4）参照）というツールを併用することを推奨している。
- (3) さらに、社会的要因をグローバルな観点から集計するというような「量的分析」と、前後のコンテキストを考慮して発話文を解釈するというような「質的分析」の双方を行うことを原則とする。
- (4) そのため、『総合的会話分析』に適した文字化の原則である BTSJ による文字化入力の時間を短縮する「文字化入力支援機能」や、各研究者が行ったコーディングの集計・分析や、総発話文数に占める各話者の発話文数の頻度と割合等、「基本的な記

述統計量」の結果がワンクリックで自動表示される『BTSJ システムセット』⁴が利用できるように構成されている。

2.2 『総合的会話分析』の具体的手順

以下に、『総合的会話分析』の具体的手順を簡単にまとめる。

(1) 条件を統制してデータを収集する。

例えば、ある言語行動の男女差を比較したいのであれば、年齢などの性別以外の条件は同一にする等、目的に応じて分析対象以外の諸条件を統制するよう計画してデータを収集する。

(2) 「録音された会話」以外の部分のデータも重視する。

フェイス・シート、フォローアップ・アンケート（インタビュー）等で、必ずインフォーマントの社会的属性や、会話自体に対する感想等も収集する。その際、自由記述欄も設けて参考にはするが、基本的には、5段階評定法等を用いて、量的処理が行える形でデータを収集する。

(3) 量的分析も可能な文字化資料を作成して分析を行う。

いわゆる「会話の分析」に多い質的分析だけでなく、量的分析もしやすい形で文字化資料を作成する。この目的のために開発されたのが、文字化の原則「BTSJ」である。BTSJのルールに基づく文字化は、集計ツールである『BTSJ システムセット』上に、直接入力することも、通常のエキセルで作成した文字化トランスクリプトを読み込むこともできる。

(4) 質的分析を踏まえて分析項目の定義を明確にした上で、研究者各自が研究目的に応じたコーディング（記号化）を行い、量的処理もできるようにする。

分析項目は、特定の語彙や言語形式だけでなく、スピーチレベルの「シフト」や「沈黙」の機能など語用論的観点から重要な項目を中心に扱う。

(5) 分析項目のコーディングの結果については、2人の評定者間の評定者間信頼性係数（Cohen's Kappa：単純一致率から偶然一致する確率を差し引いたもの）を算出することによって、「信頼性」を確保する。

(6) 質的分析によって、コーディングの「妥当性」の確認を行う。

コーディングを行い、量的処理のために数値化した結果には表れにくい特徴や、コーディングの過程で見落とされている特徴がないかを、必ず、質的な分析で確認・検討する。

(7) 分析の際は、必ず、グローバルな観点からの分析によって全体的な傾向を踏まえた上で、ローカルな観点からの各発話文の意図や機能の解釈などの分析を行う。

ここで言う「ローカルな観点」とは、やりとりの中における「個々の発話の機能」を、当該の発話文の前後の文脈程度のローカルな要因を考慮に入れながら、質的に解釈することを指す。いわゆる質的な会話分析では、一般的な方法である。「グローバルな観点」とは、量的分析においては、例えば、話者の年齢や性別などの当該会話の内容以外の社会的要因を考慮に入れて、それらの要因の影響を総合的に分析するものなどを指す。

ただ、必ずしも多数のデータを扱って平均値などを算出するというようなタイプの分析だけではなく、仮に一つの会話のみの質的分析であっても、当該会話の発話文のローカルなやりとり部分だけでなく、当該の会話の流れやその会話の結末など、その会話の最後まで話の展開や流れなどのグローバルな観点も考慮に入れた上で当該の一つの発話文を解

積、考察することによって、その発話文の意図が、一般に予想されるものとは異なっていたことなどを発見できることもある（具体例については、宇佐美 2008 を参照）。このように、ローカルな発話文の解釈の選択肢を増やし、その妥当性を高めることもできる。以下に、「グローバルな分析」の3つのタイプをまとめておく。

<タイプ1> 発話のやりとりに影響する当該会話外の要因（話者の年齢・社会的地位・性等）を考慮に入れて収集したデータを元に、談話全体の特徴を量的に捉えて、それらの要因の影響の比較を行うような分析。

<タイプ2> 例えば、スピーチレベルのような談話の中の主要要素の多数の会話における平均的、典型的な分布を、「スピーチレベルの分布の基本状態（default）」（宇佐美 2003b; 2008）と捉えて、それを考慮に入れた上で、個々の発話の機能を解釈するような談話レベルの分析。

<タイプ3> 会話全体の流れや結末を把握した上で、その会話内の一つの発話文の真意や機能を解釈するような談話レベルの分析。

このように、『総合的会話分析』と名づけたのは、会話自体が中心データとなる研究ではあるが、会話を収集する際に、「条件を統制してデータを収集する」ことや、当該の録音された会話以外の社会的要因の影響も常に考慮に入れるということ、また、質的分析、量的分析の双方を用いること、さらには、グローバル、ローカル双方の観点から分析し解釈する必要があるということも方法論の手順の一つとして取り込んでいる点など、まさに、会話を「総合的に分析する」方法論であるからである。

2.3 『総合的会話分析』に適した『基本的な文字化の原則（Basic Transcription System for Japanese: BTSJ）』の開発の趣旨と活用法

1990年代初頭、当時使っていた CHILDES（MacWhinney & Snow 1985）が、大人の日本語会話の文字化には適さないことを感じたことがニーズとなって、CHILDES を参考にしながら、BTSJ の前身を考案した。その後、試行錯誤を重ねた結果、1997年に「BTSJ（Basic Transcription System for Japanese）」として公開した（宇佐美 1997）。その後、改訂を重ね、現在は、「基本的な文字化の原則(Basic Transcription System for Japanese: BTSJ) 2015年版」（宇佐美 2015）⁵として、WEB上で公開している。また、対照研究にも役立つように、BTSJ をベースに、各言語の特徴を考慮して、中国語版（Basic Transcription System for Chinese: BTSC）、韓国語版（Basic Transcription System for Korean: BTKS）、英語版（Basic Transcription System for English: BTSE）も作成されており、BTSC と BTKS については、WEB上に公開されている⁶。

BTSJ は、その名の通り、「基本的な文字化の原則」であり、汎用性を念頭において構築された文字化のルールである。会話を量的に分析する場合、文字化した会話資料に研究目的に応じてコーディングを行い、それらの分析項目を集計して統計処理を行うといった一連の手順を踏むため、細かく精緻な作業が求められ、また、膨大な時間と労力を要する。そのため、個人レベルで大量の会話データの文字化資料を作成するのは、いくら表計算ソフトを活用したとしても、限界がある。そこで、1990年代初め、「自然会話コーパス」を構築し、研究者間でデータを蓄積・共有できるようにしていくことが喫緊の課題であると考えたが、会話データを共有するためには、最低限の「共通の文字化ルール」が必要である。しかも、それは大量のデータを集積し、分析していくのに適したものである必要がある。このようなニーズのもと、内外の「文字化に関する論考」を検討した結果、BTSJ は、

以下のような条件や特徴を備えるものとなった。

- (1) 研究の視点を得るために、大量の文字化資料でも読みやすいものにする必要があることを考え、漢字仮名交じり表記にする。
- (2) 質的分析だけでなく、量的分析にも適するものにする。
- (3) コーパスとしての蓄積、データベース化が可能で、検索がしやすく、量的分析に適するようにするため、表計算ソフト（エクセル）を利用する。これは、後述する『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』へと発展した。
- (4) 発話された「文」として、一語文等も含む「発話文」を分析の単位としてコーディングできるように、改行の原則を考える。特に、日本語の場合、一文ごとの述部に相手の待遇を示すスピーチレベルが反映されるので、敬語使用の分析、文法的分析など、「文」単位でコーディングをするのが、最も合理的であると判断したことによる。
- (5) 対人機能に重要な役割を果たすと考えられる周辺言語情報や微妙な音声情報は、音声を聞かずに文字化資料のみを用いて考察する時にも、できるだけその音声が想起されやすいようにするための「研究者自身のメモ」として、ト書き的な書き方でなるべく多くを記しておく。
- (6) ポーズの長さは、物理的な測定による長さではなく、普通の人を感じた長さを相対的な観点から記述する。つまり、いつも早口であり間をあけずに話す人が、0.5 秒の間をあけると、その「間」は長く感じられるかもしれないが、いつも、とつとつと間をおきながら話す人の 0.5 秒の間は、間とも感じられないかもしれない。そういう「相対的な人間の認知の仕方」をそのまま記すようにする。
- (7) 上記に基づく文字化の信頼性は、二者間の評定の一致度から偶然一致する確率を差し引いて算出する評定者間信頼性係数によって確認する（項目にもよるが 0.7 以上あれば、問題ないと判断されることが多い）。

2.4 BTSJ のルールのポイント

ここでは、BTSJ のルールで、最低限押さえておく必要があるポイントとして、「発話文」という分析の単位と、それにも関係する「改行の原則」に絞ってまとめる。オーバーラップなどの情報が付与されていない、いわゆるテキストデータに近い他の文字化のルールに比べると、BTSJ のルールや記号は、詳細で複雑なので難しいという声を聞くこともあるが、BTSJ のルール作りを始めて 20 年余りの間に、多様な会話すべての特徴の記述に対応できるようにと心がけていると、自ずとルールや記号が増えていったという経緯がある。しかし、実際には、一般的な会話を文字化するにあたり、よく使う記号は限られており、BTSJ のマニュアルを見て感じるほど複雑ではない。ここでは、正しく記述されていないと、BTSJ システムセットを用いたコーディング後の集計が正確ではなくなるため、必ず守る必要がある「改行の原則」とそれに関係する「発話文番号」の振り方の注意点のみ解説しておく。BTSJ のルールの詳細については、宇佐美 (2013, 2015) や、HP⁷を参照していただきたい。

2.4.1 改行の原則

改行の原則にかかわる以下の 2 点は、最低限、守らないと、『BTSJ システムセット』における集計が正しくできないので、注意が必要である。

基本的には、話者が交替するたびに改行する。しかし、話者が交替しなくとも、同一話者が複数の「発話文」を続けて発するとき、「発話文」ごとに改行する(例 1)。

例1 同一話者が複数の発話文を話す場合

- 1A いや、や、や、やっぱり天気の方がいいかな。
- 2B 天気の方がいいですね。
- 3B 断然天気の方がいいですね。
- 4B 洗濯物もよく乾くし。

1 発話文の途中で相手の発話が入った場合には、その途中の句末に英語式コンマ2つ「,,」をつけ、その発話文が終わっていないことをマークし、改行して相手の発話を記入する。この場合、例2で示すように、話者が交替しても同一話者によって発せられた、2-1 と 2-2 とで1つの発話文とみなす。

例2 1 発話文の途中で相手の発話が入った場合

- 1A もう、もろ駅前商店街。
- 2-1B 北浦とか、もっと、それよりもっと,,
- 3A そうそうそうそう。
- 2-2B 駅前。
- 4A あの、「店名」とかの近くです。

以下の図1に、BTSJによるトランスクリプトの例を示しておく。

ライン番号	発話文番号	発話文終了	話者	発話内容
1	1	*	TA01	###しゃべってください。
2	2	*	JNS01	えっ、何をしゃべるんですか<笑い>。
3	3-1	/	TA01	《沈黙2秒》ね、しゃべらないと,,
4	4	*	JNS01	<笑いながら>しゃべれないよ。
5	3-2	*	TA01	また私ずっとしゃべってる感じだ<笑い>と思いました。
6	5	*	TA01	あっ[[。
7	6	*	JNS01]]あっ、お菓子食べる<時に…>[[。
8	7	*	TA01	<そう>]]。
9	8	*	JNS01	かな。

図1 BTSJによるトランスクリプトの例

2.5 『総合的会話分析』における「コーディング」

『総合的会話分析』の方法論では、BTSJによる文字化資料を作成するところまでは、「データ作成」の段階であり、基本的に誰が行っても同じになるべきところである。文字化資料が整って初めて、本格的な「分析」に入る。ここからは、各研究者が、BTSJのルールとエクセルをベースとするBTSJシステムセットの機能を活用しながらも、独自の研究目的や研究対象を設定して、自由に分析を行う段階に入る。「コーディング」とは、量的な分析を行うために、会話の中に現れる形式や発話の機能などの分析項目を記号化して分類等を行っていくことである。BTSJでは、コーディングの単位を発話文としているので、各

発話文ごとに「コーディング項目」の列に、研究者が分類した記号を入力していく。総合的会話分析では、この「コーディング」の方法が適切で、妥当性を持つものであることが、最も重要だと言っても過言ではない。そのため、ここでは分析に際して、「全ての発話文をコーディングするタイプ」と「分析項目が現れた発話文だけをコーディングするタイプ」を、それぞれ「文末スピーチレベル」、「終助詞『ね』」を例として、解説する。

2.5.1 全ての発話文をコーディングする場合

例えば、文末のスピーチレベルをコーディングする場合は、細かい分類の定義は各研究者が目的に応じて決めるが、仮に、「敬体 (P)」、「常体 (N)」、「丁寧度を示すマーカがない発話文 (NM)」の3つに分類するとする。その場合、すべての発話の文末を判断して、P、N、NMのどれかを「文末」の列に記入していく。次の表1に、「全ての発話文をコーディングする場合」の例として「文末のスピーチレベル」を、また、次項で説明する「分析対象が現れた発話文だけをコーディングする場合」の例として「終助詞『ね』」のコーディング列を追加し、コーディングの例として示す。

表1 「文末のスピーチレベル」と「終助詞『ね』」のコーディングの例

ライン番号	発話文番号	発話文終了	話者	発 話 文 内 容	文末	終助詞
5	5	*	JBM03	えっと、「JBM03 姓」といいます >{<}。	P	na
6	6	*	JOM02	<「JBM03 姓」>{>}さん。	NM	na
7	7	*	JOM02	よろしくお願ひします。	P	na
8	8	*	JOM02	今日は雨が降ったりやんだり、最近…。	NM	na
9	9	*	JBM03	そうですね。	P	ね

2.5.2 分析項目が現れた発話文だけをコーディングする場合

すべての発話文に出現するとは限らないもの、例えば、終助詞「ね」の有無とその会話における機能をコーディングする場合は、終助詞が現れた発話文のみに、記号としての「ね」を入力していく。ただ、BTSJのルールでは、当該の分析対象が現れていないことを意味する空欄と入力漏れの空欄との区別を明確にするために、全ての発話文に何らかの記号を入力することを原則としている。そのため、「ね」が現れていない発話文にも、コーディングに該当するものがないことを表す「na (non applicable)」という記号を入力する。

2.5.3 1発話文中に分析項目が複数ある場合のコーディング

研究者が扱う分析項目が、場合によっては、1発話文中に複数回現れることがある。複数の要素をコーディングするのに1つのセルしかない、コーディングができないため、このような場合には、複数の同じ項目を個々にコーディングできるように、各分析項目の後ろに「&」をつけてその直後に改行し、1つの要素につき1つのコーディング・セルが

割り当てられるようにする。つまり、「&」は、BTSJ の本来のルールでは改行される箇所ではないが、1 発話文内の複数の要素をコーディングしたいので、便宜上改行するというを示す記号である。そのことを示すために、発話文の列には、本来の、改行しない場合の「発話文番号」に、例えば、69-a、69-b のように、「-」とアルファベットを付して、ラインが 2 行にわたっていても、69-a、69-b で「1 発話文」であるということを示す（表 2）。

表 2 1 発話文中に分析項目が複数ある場合のコーディングの例－「ね」の場合

ライン番号	発話文番号	発話文終了	話者	発話内容	終助詞
73	69-a	/	F05	校長と教頭のやつはね&、	ね
74	69-b	*	F05	何事もなくね、その後なんにもフィードバックも何にもなく、〈普通の〉{<}>。	ね
75	70	*	F05	<まー、形>{<}> だけだよ。	na

3 『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』開発の趣旨と概要

ここまで、総合的会話分析の趣旨とそれに応じた文字化のルールである BTSJ について概説し、エクセルをツールとして会話の文字化資料を作成する方法、及び、量的分析のためのコーディングの仕方について説明してきた。このように、コーディングの集計自体は、プログラムが書ければ、それで集計できるし、通常のエクセルでも簡単な集計は可能である。ただ、プログラムをしないでエクセルのみの集計機能で、多数の会話の分析を行うには、膨大な時間と労力がかかるため、その時間を削減することを企図して開発したのが、『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』⁷である。本システムを利用することによって、会話の「基本的な記述統計的情報」が自動集計されるとともに、ひとつの分析項目について多角的な観点から分析するための情報が、瞬時に自動算出・表示される。このことにより、会話の量的な分析に必要な処理を短時間で正確に行うことが可能になった。

BTSJ のルールに対応した本システムは、利用者の便宜や汎用性を考えて、Microsoft Excel のマクロ機能を利用して作成されており、『BTSJ 入力支援・自動集計システム・オリジナル.xlsm』、『BTSJ 複数ファイル自動集計システム・オリジナル.xlsb』の 2 つのファイルからなっている。BTSJ のルールに関する「エラーチェック機能」や、独自のショートカットキーなどの、文字化の「入力支援機能」、及び、記述統計に必要な基本的な分析項目の「自動集計機能」を備えており、現在、Windows OS、Excel 2003、2007、2010 に対応している。エクセルに独自の機能を搭載する形で『BTSJ システムセット』を開発していったのは、ひとえに、特別のプログラムやシステムがなくても、誰でもが利用しやすいという汎用性を考えてのことである。研究者のみならず、大学院生などでも、大学や自宅で、いつでも利

用できるものにするということを優先的に考えている。また、通常のエクセルで作成した文字化資料も、BTSJの原則に従っていれば、本システムの「データ読み込み」ボタンから読み込み、話者も自動登録することが可能である。BTSJとは異なる方法で文字化された資料でも、本システムに読み込んだ後、部分的に記号などの修正を行えば、分析項目の基本的記述統計量を自動集計する機能が使えるようになる。ただし、その際は、発話文ごとに改行するという「改行の原則」を守っている必要があることは注意しておく必要がある。これらのポイントを押さえて、本システムを活用すれば、質的分析のために収集された既存の文字化資料を量的分析のための資料としても活用できるように変換し、応用することが可能になるという大きなメリットがある。

4 『総合的会話分析』に基づく研究—『BTSJ 日本語自然会話コーパス (2018年版)』を用いたもの

ここでは、簡単に『BTSJ 日本語自然会話コーパス (トランスクリプト・音声) 2018年版』(以降、『BTSJ 自然会話コーパス』と略記)、及び、共同構築型多機能データベースである『自然会話リソースバンク (NCRB: Natural Conversation Resource Bank)』を使った「自然会話を素材とするWEB教材」(以降、NCRBと略記)の作成方法についても簡単に紹介し、『総合的会話分析』による研究の例を示すとともに、『BTSJ 自然会話コーパス』と連動しているNCRBについて簡単に紹介する。

『BTSJ 自然会話コーパス』とは、会話参加者の年齢、性別、話題などを統制した上で、初対面会話、友人との会話などの自然な会話を収集し、333会話の文字化資料と音声を収録したもので、自発的会話の音声付きのコーパスとしては、世界最大規模のものである。発話の重なりや沈黙など、語用論的分析に不可欠な情報を記し、定性的分析に加えて、研究者各自が独自の観点からコーディングできるBTSJという文字化システムを使い、自動集計などの定量的分析を効率的に行うことができるツールも併せて提供している点が特徴的である。

「人間の相互作用としての会話」の対人コミュニケーション論的分析、語用論的分析に適したコーパスを企図して構築され、その目的のために、以下の3点を重視している。①「総合的会話分析」(宇佐美 2008)の方法論に基づき、会話参加者の年齢、性別、話題などを統制したデータ群を収録する。②発話の重なりや沈黙など、語用論的分析に不可欠な情報を記して細やかな定性的分析を可能にするとともに、各研究者が独自の観点から分析項目のコーディングや集計などの定量的分析を行うことができる文字化のルールである「基本的な文字化の原則」(BTSJ: Basic Transcription System for Japanese)によって文字化したトランスクリプトを収録する。③「人間の相互作用としての会話の分析」は、「会話自体」の分析のみならず、「録音された会話以外の社会的要因」の分析も重視する。そのため、各会話グループのデータ収集条件や話題、話者の年齢・性別・職業、その他の属性の情報も会話のトランスクリプト、音声とともに、データとして提供している。

そのため、当コーパスに収録された会話は、研究目的に応じて、「話者の属性(年齢、性別等)」や「対話相手との関係」などの、話者の話し方に大きな影響を与える社会的要因を考慮に入れた分析が可能になっている。これが、本コーパスの最大の特徴であると言っても過言ではない。

本コーパスの公開の最大の目的は、未だ質的分析のみに留まっている、「言語運用」に

重きをおいた「語用論的研究」の妥当性や信頼性を高めるために、より多くの条件統制されたデータでこれまでの知見を量的にも検証できるようにし、人間の相互作用、言語運用に重きをおく「語用論的研究」の幅を広げ、大量データの形態素解析などのような言語形式の機械的な分析だけではなく、話者間の上下、親疎関係などの実際の言語運用と人間関係の構築に極めて重要な「対人関係の情報」や、文レベルを超えた談話の流れ（文脈）を十分に考慮した分析を促すことによって、自然会話をデータとする言語運用、人間の相互作用の研究の発展を促進することである。本コーパスは、下記 URL から申請することによって、誰でも利用することができる (<https://ninjal-usamilab.info/form/corpus/index.html>)。すべてを網羅できておらず、また、各論文の詳細な研究方法まで確認はしていないが、本コーパスを用いるという時点で、既に、ある程度、『総合的会話分析』の手順と方法を踏みやすくなっていると考えられる。また、本コーパスを利用したこれまでの論文は、網羅的ではないが、下記の「BTSJ 使用論文リスト」から参照することができる (https://ninjal-usamilab.info/lab/btsj_list/)。また、BTSJ か BTSJ 自然会話コーパスを使用した場合は、下記の「BTSJ 使用報告はこちら」から、各自が登録することもできる (https://ninjal-usamilab.info/lab/btsj_form/) ので、是非、活用していただきたい。

5 『自然会話リソースバンク (Natural Conversation Resource Bank: NCRB)』と『BTSJ システムセット』との連携機能について

『総合的会話分析』の方法論自体とは、直接関係はないが、この方法論に適するように開発された『BTSJ』という文字化のルールと、「BTSJ で文字化されたトランスクリプト」に、総合的会話分析の方法に基づいて分析項目をコーディングしたものの「基本的な記述統計量」を自動集計してくれるツールである『BTSJ システムセット』が、『BTSJ 自然会話コーパス』とセットになっていること、さらには、『BTSJ 自然会話コーパス』に収録されている会話データ (333 会話) のみならず、各自が『BTSJ』に基づいて文字化した会話の文字化データをエクセルで持っていれば、その文字化資料を、自然会話のリソースバンクである NCRB にアップロードできるということが、『BTSJ 自然会話コーパス』と『NCRB』の連携機能である。『NCRB』とは、研究用と教材用の双方の自然会話データリソースを、一元化して格納してある「自然会話リソースバンク」の「オンライン上のプラットフォーム」である。

5.1 「自然会話リソースバンク (Natural Conversation Resource Bank: NCRB)」－日本語教育における「コミュニケーション能力」養成への応用－

『総合的会話分析』に基づく様々な語用論的研究から導き出される知見は、日本語教育における「自然なコミュニケーション能力」を養成することに生かせる知見の宝庫である。『総合的会話分析』の手法で、自然会話を分析しているうちにその教材化の必要性に気づき、実験的に構築し、辿り着いたのが、『NCRB (Natural Conversation Resource Bank)』である。これは、新しいタイプの「共同構築型多機能データベース」であり、「自然会話データ」の登録・利用の二つの入り口と、「自然会話を素材とする WEB 教材」の登録・利用の二つの入り口があり、合計 4 つのセクションから構成されている。日本語教育の実践に関係が深いのは、NCRB の「教材作成支援機能」を用いて「WEB 教材」を作成するという『自然会話リソースバンク』の中の「教材作成支援機能」の活用と、それを利用して「作

成した教材」を学生に紹介して独習してもらったり、時には、授業で使ったりするという実践である。コミュニケーションの語用論的研究のために、自然会話コーパスを作成してきたが、自然会話のデータは、研究のみならず、コミュニケーション教育のためのよき教材にも成ると確信した。ただ、「自然会話を素材とする教材」の作成は、文字化に時間がかかる上に、「教材としての解説」なども加えていかなければならないため、単独、或いは、単独グループで作成するにも、膨大な時間と労力がかかる。そこで、「自然会話 WEB 教材作成支援システム」を開発して作成過程の一部を自動化し、教材作成の時間と労力を削減するように考案した。

『NCRB』は、このような必要性とアイデアを踏まえて、「様々な場面における事前にシナリオのない自発的な会話」を録音・録画し、それらをリソースバンクとして蓄積し、データベース化して構築された。本システムで作成した教材は、主に WEB 上での独習を想定しており、初級から超級に至るまでのすべてのレベルの学習者が、同じ素材を各自のレベルや興味に応じて活用することを推奨している。必要やレベルに応じて、日本語会話の音調、間合いなどの学習や討論の材料として、授業中に活用することもできる。授業で使う場合は、スピーチレベルのシフトなどの、場面や人間関係に応じたコミュニケーションの方法について気づきを促し、議論をさせる等、様々な活用法が考えられる。

ここでは、図 2 に NCRB のトップページを示すことで NCRB がイメージできるようにするに留める。研究者・教育者の方々には、是非、データや教材を登録していただきたく、協力を呼びかけたい。



図2 NCRB トップページ

発話内容を解説する

ライン番号	発話文番号	話者	発話内容	説明
0001	1	NS01	うーん、わたしでも、きょう、カッパにしようかな。	追加
0002	2-1	NS02	ああ、<先ス>(<)>。	追加
0003	3	NS01	<##動に動>(<)>か、せんぱいなんがうちが、つて<かんじて、まっすね>(<)>。	追加
0004	2-2	NS02	<うん、うん、うん、うん>(<)>。	追加
0005	4	NS01	<じゃあもうそろそろ、ね>(<)>。	追加
0006	5	NS02	<じゃあ、せんぱい、じ、>(<)>、ビールも好きだよ。	追加
0007	6	NS01	あつ、そうね、ビールい好きだよな。	追加

完成画面

Step 1 : イメージしてみましょう

図3 教材作成画面の一部(左)と完成画面(イメージ)(右)

6 『総合的会話分析』に基づく研究の今後の展開

以上、『総合的会話分析』の趣旨と方法を概観した。量的分析と質的分析の併用という、相容れないのではないかと思う人もいるかもしれない。しかし、実際に文字化資料をデータとする「会話の分析」だからこそ、条件を統制して複数の会話を収集し、量的に処理するとともに、記号に変換されて得られた結果のみに満足することなく、実際のトランスクリプト、ひいては、「音声」に立ち返りながら、実際の発話の内容とやりとりを分析、考察するという質的分析も行う必要がある。『総合的会話分析』というのは、まさに、量的分析と質的分析の必然的融合とも言えるだろう。

実証的方法をベースとするこの方法論は、何らかの現象の原則や規則性を見出すことを目的とする研究に適している。日本語教育研究においても、個別事例から何かを発見するタイプではなく、例えば、「中国人学習者の語用論的誤用とその指導法」を研究する等、多数の学習者に共通する特徴を明らかにすることによって、彼らに適する指導法を考えるとこのようなことを目的とする研究は多い。このような目的がある場合などは、量的分析と質的分析を行い、ある程度の一般化を可能にする方法論である『総合的会話分析』が有効であると言えるだろう。また、昨今、日本語教育においても、「コミュニケーション能力」の育成の重要性が叫ばれているが、そのためには、「語用論的研究」の発展が必須である。『総合的会話分析』は、そもそもが、語用論的分析をより実証的にして進化させることを目的の一つとして開発されたものである。そういう意味で、この方法が日本語教育研究に貢献できるところは、多々あると考える（宇佐美 2013）。

最後に、会話データを中心に扱う語用論的研究分野のさらなる発展のために必要なことをまとめる。各研究者が時間と労力をかけて収集した会話データを、研究者間で広く共有して活用するためには、「語用論的分析に適したコーパスの構築」が必須である。それを目指したのが、『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』（その後 2015 年に一部改訂）（宇佐美 2011）であり、その後、さらにデータを拡充整備して、2018 年 8 月 31 日に国立国語研究所日本語教育研究領域より公開した『BTSJ 日本語自然会話コーパス（トランスクリプト・音声）2018 年版』（宇佐美 2018）である。個人、研究グループ、研究機関単位でさえ、コーパス構築には、時間、労力、経費がかかるため、十分な予算がなければコーパスの構築と維持は難しい。また、従来のコーパス構築の方法では、「データ提供側とその利用者」というように役割が一方的で固定されてしまうため、相互の発展に結びつかない。今後は、もっと「双方向的な参加・共有型コーパス構築」のあり方を考える必要がある。このような考えに基づいて開発されたのが、「多機能共同構築型データベース」としての「自然会話リソースバンク（Natural Conversation Resource Bank: NCRB）」⁸である。ここに蓄積された会話データは、『総合的会話分析』の方法論で行う研究に適するものが多いので、是非、活用されたい。

7 おわりに

本稿では、「『総合的会話分析』に基づく研究—『BTSJ 日本語自然会話コーパス』と『自然会話リソースバンク（NCRB）』との連携に触れながら—」と題して、現在、盛んに行われている「会話の分析」の中から、『総合的会話分析』の方法論を中心に解説した。また、『BTSJ 自然会話コーパス』、『BTSJ システムセット』そして、自然会話の日本語教育への

応用のために構築された「共同構築型 WEB コミュニケーション教材」を格納する『自然会話リソースバンク (Natural Conversation Resource Bank: NCRB)』を紹介した。これらは、すべて連動している一連の方法論やデータベース、コーパス、コーパス分析ツール、オンライン教材というように総合的に考案されたものである。そういう点で、まさに 21 世紀の研究と言語教育教材のあり方を先駆的に示すものと考えてもいいのではないかと考えている。すべてに十分に触れることはできなかったが、それぞれについては、参考文献や参考サイトを参照するとともに、今後の展開にも注目していただけたら、幸いである。

<謝辞>

本稿は、2018 年 8 月 3 日 (金)、2018 年日本語教育国際研究大会 (ICJLE2018) おいて行った、『国立国語研究所連続講義』2 に基づくものである。

また、本稿で扱った『BTSJ 日本語自然会話コーパス (略称)』、『Natural Conversation Resource Bank: NCRB』は、国立国語研究所の機関拠点型基幹研究プロジェクト「日本語学習者のコミュニケーションの多角的解明」(平成 28 年 4 月～) のサブプロジェクト「日本語学習者の日本語使用の解明」(リーダー：宇佐美まゆみ) <https://ninjal-usamilab.info>、および JSPS 科研費 18H03581 「語用論的分析のための日本語 1000 人自然会話コーパスの構築とその多角的研究」(研究代表者：宇佐美まゆみ)、JSPS 科研費 18K18685 「コミュニケーション能力を高める自然会話教材の高度共有化—共同体の構築に向けて—」(研究代表者：宇佐美まゆみ) の成果の一部である。

本稿の編集に際しては、担当の加藤さやか氏、岩崎典子氏に、大変にお世話になった。記して心から感謝したい。

¹ 現在は、本稿でも論じてきたような『総合的会話分析』の目的や趣旨を正しく理解した上で活用していただきたいということから、不定期に開催している「BTSJ 活用方法講習会」のすべてを受講した本人に限って、無償で配布する形にしている。

² オンライン申請者に無償配布している。URL: <https://ninjal-usamilab.info/form/corpus/index.html>

³ 扱うテーマの関係上、宇佐美 (2013、2015 等) と一部重なる内容や記述もあることをお断りしておく。ただ、より詳しく知りたい方は、それらの論文を参照されたい。

⁴ 『BTSJ システムセット (略称)』には、TSV ファイルへ出力する機能もあり、NCRB との連動や形態素解析の分析等、興味に応じて様々な形でデータを扱うことができる。

⁵ <http://ninjal-usamilab.info/pdf/btsj/btsj2015.pdf>

⁶ 中国語版 (BTSC) <http://ninjal-usamilab.info/pdf/btsj/btsc.pdf>、韓国語版 (BTSK) <http://ninjal-usamilab.info/pdf/btsj/btsk.pdf>

⁷ 『基本的な文字化の原則(Basic Transcription System for Japanese: BTSJ) 2015 年改訂版』
<http://ninjal-usamilab.info/pdf/btsj/btsj2015.pdf>

⁸ 『自然会話リソースバンク (NCRB)』は、会話分析研究のための自然会話コーパス、自然会話を素材とする会話教材のためのデータバンク、及び、会話データ検索、文字化資料へのタイムスタンプ付与、自然会話教材作成支援システムなど、多機能を併せ持つデータベースで、質的分析はもちろんのこと、量的分析を行う際にも便利である。

<参考文献>

Budds, K., Locke, A. and Burr, V. (2017) ‘Using Discourse Analysis in Social Psychology’ in Brooks,

- J. and King, N. (eds.). *Applied Qualitative Research in Psychology*. Palgrave Macmillan.
- MacWhinney, B. and Snow, C. E. (1985) The child language data exchange system. *Journal of Child Language* 12: 271-296.
- Snow, C. E. (1992) Perspectives on Second-Language Development: Implications for Bilingual Education. *Educational Researcher* 21 (2): 16-19.
- Snow, C. E., Perlmann, R., Gleason, J. B. and Hooshyar, H. (1990) Developmental perspectives on politeness: Sources of children's knowledge. *Journal of Pragmatics*. Volume 14 (2): 289-305.
- 宇佐美まゆみ (1999) 「談話の定量的分析 —言語社会心理学的アプローチ—」, 『日本語学』 18(11), 40-56.
- 宇佐美まゆみ (2000) 「談話研究と言語教育」, 『AJALT』 23, 22-26.
- 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』, ひつじ書房: 150-181.
- 宇佐美まゆみ監修 (2011) 『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』 (2015年に一部改訂).
- 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ョン研究』, くろしお出版: 63-82.
- 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』 32(14), 132-147.
- 宇佐美まゆみ (2015) 「『総合的会話分析』の趣旨と方法—量的分析と質的分析の必然的融合—」 (特集: 日本語教育の研究手法—「会話・談話の分析」という切り口から—), 『日本語教育』 162, 34-49.
- 宇佐美まゆみ (2017) 「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—」 (パネル「『BTSJ 日本語会話コーパス』を活用した教材作成への提案 —ヨーロッパにおける自然なコミュニケーション教育のために—」) 『ヨーロッパ日本語教育』 第22号, 279-285.
- 宇佐美まゆみ監修 (2018) 『BTSJ 日本語自然会話コーパス (トランスクリプト・音声) 2018年版』, 国立国語研究所. (https://ninjal-usamilab.info/lab/btsj_corpus/)
- 宇佐美まゆみ (2019) 「談話研究と言語教育—1960年代から現在までの流れ—」 『ヨーロッパ日本語教育』 第23号.
- 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』, 217-228.
- 宇佐美まゆみ・山崎誠 (2018) 「『BTSJ 日本語自然会話コーパス (トランスクリプト・音声) 2018年版』の紹介と『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』を用いた分析法」, 計量国語学会: 1-6.
- 川口義一 (2017) 「自然会話の「文脈」から見た文法教育の改善」 (パネル「『BTSJ 日本語会話コーパス』を活用した教材作成への提案 —ヨーロッパにおける自然なコミュニケーション教育のために—」) 『ヨーロッパ日本語教育』 第22号, 285-290.
- 萩原 孝恵 (2017) 「教科書会話 vs. 自然会話 —カラとノデの語用論—」 (パネル「『BTSJ 日本語会話コーパス』を活用した教材作成への提案 —ヨーロッパにおける自然なコミュニケーション教育のために—」) 『ヨーロッパ日本語教育』 第22号, 290-296.

<参考サイト>

宇佐美まゆみ研究室 HP <https://ninjal-usamilab.info/lab/>

「基本的な文字化の原則(Basic Transcription System for Japanese: BTSJ)について」

https://ninjal-usamilab.info/about_btsj/

「基本的な文字化の原則(Basic Transcription System for Japanese: BTSJ) 2015 年改訂版」

<http://ninjal-usamilab.info/pdf/btsj/btsj2015.pdf>

「BTSJ 日本語自然会話コーパスについて」 https://ninjal-usamilab.info/lab/btsj_corpus/

「日本語学習者の日本語使用の解明」(リーダー: 宇佐美まゆみ) の HP, 国立国語研究所
基幹型プロジェクト『日本語学習者のコミュニケーションの多角的解明』(平成 28 年 4
月～) <https://ninjal-usamilab.info/>

「BTSJ 使用論文リスト」 https://ninjal-usamilab.info/lab/btsj_list/

「BTSJ 使用報告」 https://ninjal-usamilab.info/btsj_use-report/

BTSJ, 及び、『BTSJ 日本語自然会話コーパス (2018 年版, それ以前の版)』を利用した論文の書誌情報を報告すると、自動でリスト化されて HP に掲載される。また、いろいろな研究者の「BTSJ 使用論文」の書誌情報が閲覧できるようになる。『基本的な文字化の原則 (BTSJ)』や、『BTSJ 日本語自然会話コーパス』を利用した論文であれば、どなたでも『宇佐美まゆみ研究室 HP』から利用報告を入力することができるので、是非活用されたい。

『自然会話リソースバンク (Natural Conversation Resource Bank : NCRB)』

<http://ncrb.ninjal.ac.jp/>

研究用に自然会話データを登録・ダウンロードするための入り口と、教材用として、自然会話を素材とする教材を登録・利用するための入り口がある。(2019 年度中に一般公開予定)