3 データベース作成・分析のための道具

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)」

宇佐美 まゆみ

1．はじめに

本稿では、「言語と言語使用」における相互作用の分析を通して、社会の中に生きる人間を探究することを主目的とする「言語社会心理学」と、その中でも、人間の相互行為としての「会話」の分析に焦点を当て、その分析方法論をより具体的に提示した「総合的会話分析」(宇佐美、2008) という会話分析の方法論とそれに関する「ツール」を紹介する。また、総合的会話分析の趣旨に適するように考案された文字化の規則である「基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ、以降 BTSJ と表記)」(宇佐美、2011) と、BTSJ に基づく文字化資料の作成と分析を効率化するために開発された、エクセルに独自のマクロ機能を搭載した『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』(以降『BTSJ システムセット』と表記)、さらには、BTSJ 形式の文字化資料、音声 (一部)、話者の社会的属性情報や、データ収集条件、話題等の当該の会話以外の情報も含む『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』(以降『BTSJ 話し言葉コーパス』と表記) (宇佐美、2011) について、その特徴の概略と活用法を紹介する。

1.1　語用論的分析に適した「文字化の原則」と「文字化資料付きの会話コーパス」の必要性

近年のめざましいテクノロジーの発展を受けて、言語研究に活用できる技術やツールが多々開発されている。「コーパス言語学」で駆使されているツールや電子データなどは、そのよい例である。その結果、以前は考えられなかったような大量データの形態素解析やコロケーション研究などが盛んになり、そのような研究から重要な知見が次々と得られるようになった。

しかし、一方で、コーパス言語学が対象としているデータには、話者の年

「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)、明治書院: 132-147 16頁.

齢や性別、教育歴などの社会的属性や、上下等の話者同士の関係などの社会的要因の条件統制を行わずに大量に収集されているものが多く、ある程度サブグループに分けて分析することはできるものの、話者自身の属性だけでなく、対話相手との「関係」が言語行動に与える影響や、話者間の相互作用の分析はできないものになっている。また、「タグ付け」も、形態素解析や言語形式の頻度やコロケーションの解析などの定量的処理はできても、「談話の流れ」や「沈黙」などの語用論的分析はできないものになっている。

また、いわゆる「学習者コーパス」も、未だ文法項目の習得を研究対象とするものが多く、文字化の原則も簡素で、語用論的研究に必要な同時発話や沈黙などの情報が付与されていないものがほとんどである。つまり、人間の相互作用の語用論的分析には適していないものが、ほとんどである。

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情報が付与されているが、逆に、少数の会話の定性的分析には適していても、より数の多い会話データで特定の言語事象の頻度や割合を調べる等の定量的処理には全く適さない形になっている。

会話の分析においては、研究目的に適した「文字化の原則」が必要であることは言うまでもないが、研究目的には、言語学、語用論、社会学、第二言語習得論、対話ロボット開発の研究のため等々、様々な観点がある。また、方法論的には、定性的分析と定量的分析のどちらに適しているかという観点がある。これらを総合的に考えた上で、研究目的に適した「文字化の原則」を選択することが必要になってくる。

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」の方法論や、言語社会心理学的な関心に基づく研究に適し、人間同士の相互作用を、定性的・定量的双方から分析するのに適した「文字化の原則」の開発の必要性を痛感したのが、BTSJ 開発の動機であった。また、定量的分析を行うには、条件統制して収集された一定の数の会話が必要であることから、BTSJ という文字化の原則で統一された文字化資料付きの「BTSJ 会話コーパス」の構築にも着手した。さらには、会話の文字化と分析のための集計にかかる膨大な時間と労力を少しでも削減するために、「文字化入力支援ツール」、及び、「自動集計システム」の開発も行うことになった。

1.2　会話データの文字化資料作成のための「ルール」としての BTSJ

音声認識研究などの目的のために作られた工学系の対話データベースには、

音声だけを扱うものもあるが、会話の語用論的分析には、文字化資料は必須である。BTSJ 自体は、文字化をする際の「ルール」に相当するものを指し、いわゆる「ツール」ではない。しかし、本特集号が捉える広い意味での「ことば研究の道具」には相当すると言えるだろう。というのは、文字化の原則には目的に応じて種々のものがある中、BTSJ の文字化の原則は、定量的分析と定性的分析、双方による分析を可能にするとともに、文字化自体にかかる作業時間や労力の削減を目的とする「ツール」の開発も想定して、その規則や利用する記号を決めたからである。「会話」の分析は、その内容の文字化資料なくしては始まらないが、BTSJ は、会話の分析のためのデータを、妥当、且つ、効率的に分析することを企図して考案された「文字化のルール」であり、分析する直前の「整った会話データ」を作成するために必要な「道具」であるとも言えるからである。

1.3　文字化資料作成と分析項目の集計の労力を削減する「ツール」としての『BTSJ システムセット』

一方、『BTSJ システムセット』は、文字化資料の作成と分析項目の集計の時間と労力を格段に削減してくれる「ツール」である。BTSJ は、エクセルのような表計算ソフト上で使うことを想定してルールを考案しているが、BTSJ のルールに基づいて文字化する際に、例えば、オーバーラップされた発話とオーバーラップした発話を表す＜　＞{<} や、＜　＞{>} などの、入力に手間がかかる「対になった記号」は、当該箇所をカーソルで選択して、その記号が記されたボタンかショートカットキーを押すだけで、その箇所を囲んで正しい記号が入力されるというような「入力支援機能」や、語用論的研究のための「コーディング」が施された文字化資料については、その会話の総発話文数や、各話者の発話文数、総発話文数に占める話者ごとの発話文数の割合等の「基本情報」が、コーディング列をクリックするだけで自動算出される「自動集計機能」、及び、同じ条件でコーディングされた複数会話の場合は、分析項目の平均値や最大値、最小値、標準偏差等の基本的記述統計情報が自動的に算出され、表の形で表示される等の機能が搭載されたものである。そういう意味で、こちらは、まさに、会話の語用論的分析、特に、「総合的会話分析」のプロセスのうち、「会話自体の分析」の部分を、より効率的、効果的に行うための便利な「道具」であると言える。

1.4　『BTSJ 話し言葉コーパス』の特徴

1.1 節で説明したように、多くの既存のコーパスやデータベースは、残念ながら、会話の語用論的分析には適さないものがほとんどである。しかし、このような既存のコーパスやデータベースという道具の性質に制約されて、語用論的研究を行わないというのは本末転倒である。また、CA のように、比較的少数の会話を丁寧に分析することも一つの方法ではあるが、語用論的分析に適した形で文字化された会話データを一定量集積して、比較的多数の会話を語用論的観点から定量的に分析することも必要である。そのためには、BTSJ 形式の文字化資料を含む「自然会話コーパス」を構築することが必須であると考えた。そして、筆者の博士論文のデータ、東京外国語大学の大学院生の博論・修論・卒論データを中心に、いろいろな方々にデータ提供や協力をいただきながら、1990 年代後半から構築・改訂してきたものが、『BTSJ 話し言葉コーパス』である（http://www.tufs.ac.jp/ts/personal/usamiken/btsj_corpus_explanation.htm）。本コーパスの特徴は、BTSJ 形式の文字化資料、音声（一部）のみならず、上下、親疎などの話者間の関係や社会的属性、話題などのデータ収集条件等、語用論的分析に必要な情報が含まれている点である（詳細は、宇佐美・中俣（2013）参照）。また、後述の「BTSJ システムセット」を利用して、基本的情報や分析項目の自動集計、複数ファイルの自動集計を行うことができるという点である。

1.5　ツールを使うときの心構えと注意点

それぞれのツールの解説に入る前に、強調しておきたいことがある。それは、「道具」というものは、正しく使わなければ意味がないということである。また、「道具」でできることの限界が、研究自体の限界であると勘違いしないということである。道具というものは、様々な処理の時間や労力を格段に節約してくれるものではあるが、それは、研究全体の時間や労力を削減するためではなく、処理を終えた後の考察などに、より多くの時間を費やすためである。つまり、道具を使うのは「人間」であり、道具に左右されるようになってはならない。「人間」は、むしろ、道具がなかった時代以上に、様々なことを「考える」必要があるということである。

ただ現実には、例えば、e-learning やモバイル教材の開発が盛んになりつつある言語教育分野などのように、実践現場と工学系研究者、技術者の協働作業が比較的進んでいる分野において、ともすると、工学分野の最先端の技

術を使いたいということが先だって、「その実践現場で本当に必要なことを、テクノロジーを活用して効果的に行うには、今、何が最も必要か」という本来の目的に必要な研究が意識されにくくなるという現象も生じつつある。

言語研究分野においても、学際的研究が不可欠な現代では、例えば、コーパス言語学における単語の頻度調査のように、テクノロジーやツールでできることだけでなく、それらの相関的傾向のより深い解釈や、言語研究者にしか気づけない重要課題を追究し、その成果を、逆に、工学や認知科学などの隣接領域にしっかりと伝えていくことが、「言語研究者」（言語学者に限定しない）の使命になってきていると言っても過言ではないだろう。

2．総合的会話分析とは？

BTSJ は、「総合的会話分析」という研究方法論に基づく研究に適するように考案された。そのため、ここでは、まず「総合的会話分析」について説明しておきたい。

「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、年齢等の社会的要因も質問紙調査等で確認し、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）とほぼ同様の手順を踏むが、分析の際に、「ローカル分析・グローバル分析」という観点が必要であることをより強調し、一連の方法と分析の観点すべてを含むもの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」は、言語形式や言語自体の研究というよりは、人間の相互作用としての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しようとするアプローチである。ただ、言語行動の語用論的分析も含むため、言語研究でもあり得る。特にその目的を制限するものではなく、方法論の特徴をまとめたものであると言える。その特徴を、以下に簡単にまとめる。

(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に計画してデータを収集する。

(2)　「録音された会話」以外の分析も重視する。

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、定量的処理が行える形でデータを収集する。

(3)　定量的分析も可能な文字化資料を作成して分析を行う。
定性的分析だけでなく、定量的分析もしやすい形で会話の文字化資料を作成する。⇒基本的な文字化の原則（BTSJ）の開発と活用につながった。

(4)　定性的分析を踏まえてコーディングを行う。
分析項目（分類など）の定義を明確にした上で、コーディング（記号化）を行い、定量的処理もできるようにするが、分析項目は、特定の語彙や文法など言語形式や文型の頻度のようなもののみならず、語用論的分析にかかわる項目、例えば、スピーチレベルの「シフト」や沈黙の「機能」などを、定義・分類した上で、コーディングする。

(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」を確認、確保する。

(6)　定性的分析によって、コーディングの「妥当性」の確認を行う。
コーディングの過程で見落とされている特徴はないか等、定量的分析方法によって記号化、数値化した結果には表れにくい特徴を、必ず、定性的な分析で確認・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等を踏まえた上で、ローカルな観点からの分析を行う。

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、当該の発話の前後の文脈程度のローカルな要因を考慮に入れて、質的に分析、解釈することを指す。いわゆる定性的な会話分析では、一般的な方法である。「グローバルな観点」とは、定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要因の影響を考慮に入れて、それらの社会的要因の影響を総合的に分析するものなどを指す。また、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だけではなく、定性的分析においては、例えば、仮に1つの会話を分析するにしても、当該の会話の流れやその会話の結末など、当該会話のローカルなやりとり部分だけでなく、その会話の最後までの話の展開や流れ（グローバルな観点）なども考慮に入れた上で、当該のローカルな現象（ある発話の意図等）を解釈、

考察するということを指す。(詳細は、宇佐美（2008）を参照)

このように、「総合的会話分析」と名づけたのは、会話を収集する際に、「条件統制をしてデータを収集する」というようなことや、録音された当該の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、定性的分析、定量的分析の双方を用いるなど、まさに、会話を「総合的に分析する」方法であるというところからきている。

3．基本的な文字化の原則（BTSJ）とは？

BTSJ の前身は、1990 年代初頭に考案され、試行錯誤を重ねた結果、1997 年に「BTSJ」として公表（宇佐美、1997)、その後、何度か改訂を行い、現在は、「基本的な文字化の原則（Basic Transcription System for Japanese: BTSJ）2011 年版」として、WEB 上で公開されている（http://www.tufs.ac.jp/ts/personal/usamiken/btsj2011.pdf)。その開発の趣旨や経緯は、宇佐美（1997）に収められており、一部は、WEB 上でも公開されているので、文字化の原則の開発についての詳細は、そちらを参照されたい（http://www.tufs.ac.jp/ts/personal/usamiken/btsj.htm)。

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（BTSC)、韓国語版（BTSK)、英語版（BTSE）も作成されており、BTSC と BTSK は、上記の WEB 上で公開されている。

3.1　BTSJ 開発の趣旨

BTSJ は、その名の通り、「基本的な文字化の原則」であり、汎用性を考えて構築された文字化の原則であるので、利用者が目的に応じて、原則を追加することも可能である。会話の定量的分析では、文字化した会話資料全てに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、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、「自然会話コーパス」を構築し、研究者間でデータを蓄積し、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のが 1990 年代初めであった。

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。これらを踏まえた上で、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を採用する。

(2)　定性的分析だけでなく、定量的分析にも適するものであること。

(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、定量的分析に適するものにするため、表計算ソフト（エクセル）を利用する。これは、後述する『BTSJ 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、音声を聞かず文字化資料のみを読みながら考察する時にも、できるだけその音声が思い出される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形にする。

(6)　ポーズの長さは、物理的な測定による長さではなく、普通の人が感じた長さを相対的な観点から記述する。つまり、いつも早口であまり間をあけずに話す人が、0.5 秒の間をあけると、その「間」は相対的に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。そういう人間の相対的な認知の仕方をそのまま記すようにする。(文字化の信頼性は、評定者間信頼性係数の算出によって確認する。)

3.2　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。BTSJ のルール作りを始めて 20 年余りの間には、多岐にわたる特徴のある会話すべての記述になるべく対応できるようにと考え、ルールや記号を増やさざるを得な

かった経緯があった。しかし、現実には、一般的な会話を文字化するために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、コーディング後の集計が正確にできなくなるため、必ず守る必要がある「発話文番号」の振り方と、「改行の原則」の注意点のみ記しておく。BTSJ の詳細については、宇佐美 (2011) を参照していただきたい。

3.2.1　分析の単位としての「発話文」とその認定方法

基本的に、ひとりの話者による「文」を成していると捉えられるものを「1発話文」とする。しかし、自然会話では、いわゆる「1語文」や、述部が省略されているもの、あるいは、最後まで言い切られない「中途終了型発話」など、構造的に文が完結していない発話もある。そのような場合は、話者交替や間などを考慮した上で「1発話文」であるか否かを判断する。つまり、「発話文」の認定には、「話者交替」と「間」という2つの要素が重要になる。いわゆる「倒置」の形式に見える発話を、1発話文とするか、2発話文とするかは、発話の前半と後半の間に、明確に認知できる「間」があるかどうかによる。次の例に示すように、「間」がなければ、1発話文、間があれば、2発話文として記述する。以下の**表1**、**表2** に例を示す。

また、「ターン」を分析の単位としたい場合は、「発話内容」の右に新しく例を作り、ターン番号を記入していく形で行うことができる。

表1　1発話文と認定した場合

ライン番号	発話文番号	発話文終了	話者	発話内容
1	1	*	A	けっこうおもしろかったですよー、小説とか読んだりして。

表2　2発話文と認定した場合

ライン番号	発話文番号	発話文終了	話者	発話内容
1	1	*	A	けっこうおもしろかったですよー（うん）《少し間》。
2	2	*	A	小説とか読んだりして。

3.2.2　BTSJ の改行の原則

BTSJ における「改行」は、以下の3つのルールによって行う。以下の**図1**に「BTSJ による文字化資料の例」を示し、ポイントを説明する。

(1)　話者が交替するたびに改行する。

(2)　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は、「発話文」ごとに改行する。（図1のライン番号4-5）

(3)　話者の発話途中で相手の発話が入った場合（図1のライン番号7）

話者の発話の途中で相手の発話が入って話者が一旦交替した場合は、改行し、相手の発話を記入する。そして、最初の話者の発話の続きは、相手の発話の後に改行して続ける。つまり、1発話文でも複数のラインにわたって記される場合がある。このような1発話文が1ラインで終了せず、複数ラインにわたる場合も、同一話者によって発せられた「1文」を成していると捉えられるものは、複数のラインにまたがる発話をまとめて「1発話文」とする。そして、**図1**の「発話文番号」の列における「7-1」、「7-2」のように、異なるラインにまたがっていても同じ発話文であることがわかるように同じ番号をつけ、その後に「-」をつけて発話された順番を記す。

「発話文終了」の列には、発話が完結していることを示す「*」を入力する。また、完結していない先行発話には、「発話文終了」の欄に「/」を記す。1会話の「発話文数」は、「発話文番号」が表すとともに、左から3行目の「発話文終了」の列の「*」を数えてもわかるようになっている。また、BTSJ の

会話グループの特徴: 日本人会話
会話記号(話者記号を連結したもの): JJ14-JJ15
話者記号の凡例: JF: 日本人女性
会話番号: JJ 10-2
時間: 6'02
1会話における話者の数: 2

ライン番号	発話文番号	発話文終了	話者	発話内容
1	1	*	JF14	おはよ。
2	2	*	JF15	おはよう。
3	3	*	JF14	この前心理学あるか<とう>(<)。
4	4	*	JF15	<あー>▷K笑い>。
5	5	*	JF15	大丈夫だった?。
6	6	*	JF14	え、多分単位はとれたと思<笑い>)。
7	7-1	/	JF14	<卒業><し。
8	8	*	JF15	<###>▷とね<笑いながら>。
9	7-2	*	JF14	できるかな。
10	9	*	JF15	そう、そうだよねく笑いながら>。

図1　BTSJ による文字化資料の例

ルールでは「発話内容」の列における「。」も、発話文の完結を意味するため、質問発話で文末に「?」があっても、文が完結している場合は、「?。」と、必ず最後に「。」をつける。そのため、「発話文番号」と、「*」と「。」の数はすべて同じになるはずである。通常のエクセルで作業した場合でも、それぞれの数を数えて一致を確認することによって、文字化の漏れやミスを発見したり、発話文数の検算ができるように工夫してあるのである。

4．BTSJ による文字化資料を用いた研究方法（コーディングの仕方）

「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「分析」に入る。ここからは、各研究者が、BTSJ の原則とエクセル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、自由な分析を行う段階に入る。「コーディング」とは、定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の分類等を記号化していくことである。BTSJ では、コーディングの単位を発話文としているので、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「コーティング」の方法が適切で、妥当性を持つものであることが、最も重要と言っても過言ではない。そのため、ここでは個々の研究目的に応じて分析項目をコーディングする方法を説明する。そのため、以下の表 3 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、「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』」のそれぞれのコーディング列を追加したものを、コーディングの例として示す。

表 3　「文末スピーチレベル」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	よろしくお願いします。	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

4.1　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ことであるが、仮に、「敬体(P)」、「常体(N)」、「丁寧度を示すマーカーがない発話（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。

4.2　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。BTSJ のルールでは、当該の分析対象が現れない場合も、空欄と入力漏れ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、分析対象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。

4.3　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 つのセルしかないと、1 つずつの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「&」をつけてその直後で改行し、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するために、便宜上改行するということを示す記号である。そのことを示すために、発話文番号の列には、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 は、続けて発話された「1 発話文」であるということを示す。表 4 に例を示す。

表4　「ね」のコーディングを例として

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

5.『BTSJ 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、定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。コーディングの集計は、通常のエクセルでも可能である。ただ、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、それを格段に削減することを企図して開発されたのが、『BTSJ システムセット』である。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、エクセル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」と「BTSJ複数ファイル自動集計システム」の2つのファイルからなる。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、誰もが利用しやすい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考えた。

しかし、現在のところ、本システムは、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、無料で配布する形にしている。講習会は、これまで国内5か所、海外3か所で開催している。今後の講習会の情報は、随時、「宇佐美まゆみ研究室」HP、社会言語科学会のML、第二言語習得研究会のML(JASLA)、linguistics-jpのMLなどで配信する予定である。

5.1『BTSJ システムセット』の主な機能

5.1.1「BTSJ 入力支援・自動集計システム」の機能

「BTSJ入力支援・自動集計システム(.xlt)」には、Transcriptionシートと Definitionシートの2つのシートがある。Transcriptionシートのほうで、会話の内容を入力するが、図2に示したように「入力支援画面」を利用することによって、入力の時間と労力を節約できる。また、エラーチェック機能によって、一部は自動修正ができる。

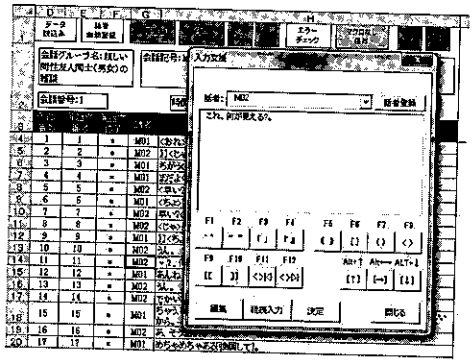

図2　入力支援画面を使用した例

また、図3のDefinitionシートにコーディング項目名を記入した後、Transcriptionシートに戻り、集計したいコーディング項目列を選択して、「集計」ボタンを押すと、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」が自動算出され、その結果が、新しいシートに瞬時に表の形で表示される。

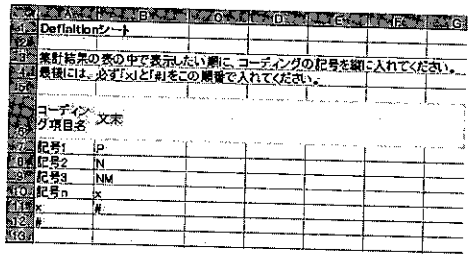

図3　Definition シートへの記入例

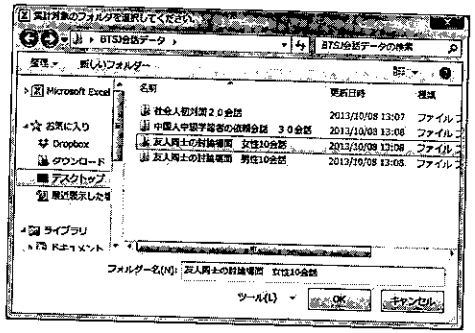

図4　BTSJ 複数ファイル自動集計システムの画面の例

5.1.2「BTSJ 複数ファイル自動集計システム」の機能

個々の会話の分析項目を集計した後は、処理したい複数の会話を1つの

フォルダにまとめて保存しておく。「BTSJ 複数ファイル自動集計システム(.xls)」をクリックして、図4に示したように、そのフォルダを選択するだけで、複数の会話の分析項目の頻度や割合、平均値・最大値・最小値・標準偏差などの基本的な記述統計値が自動算出され、瞬時に表示される(図5参照)。

このように、本システムを利用することによって、会話の「基本的な情報」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。これらにより、会話の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可能になる。

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、本システムに読込んで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には注意されたい。これらのポイントを押さえて、本システムを活用すれば、既存の文字化資料も定量的分析のための資料としても活用できるようになり、研究や応用の範囲が広がるという大きなメリットがある。

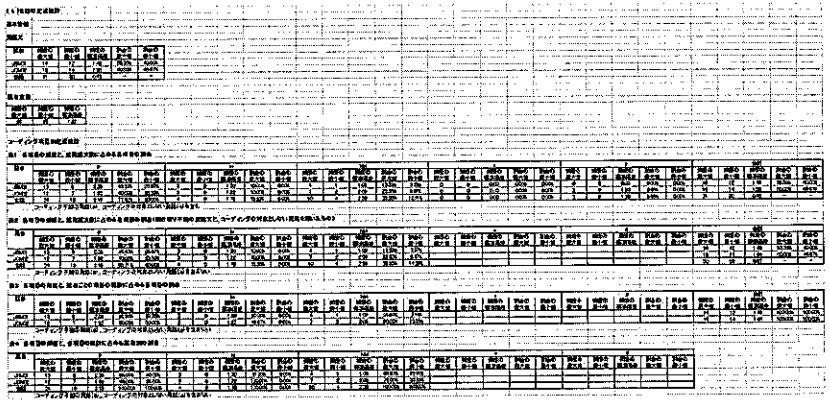

図5　BTSJ 複数ファイル自動集計システムの自動集計結果の例

さらには、マクロ機能を外して保存したり、CSV ファイルへ出力したりする機能もあり、形態素解析の分析等、幅広い興味に応じて、様々な形でデータを扱うことができるようにもしてある。

6．おわりに

会話の分析の基礎的な作業を短時間で正確に行えるシステムの開発は、定性的・定量的分析の双方を含む会話の「総合的研究」としての「総合的会話分析」の基礎を支えるものである。本システムは、自然会話データを扱う研究における「文字化作業」、「集計作業」など、基礎的作業の時間と労力を少しでも削減し、効率化したいという切なる望みから生まれた。しかしながら、最後に改めて強調しておきたいことは、基礎的作業の時間と労力を効率化したかったのは、決して研究自体にかける時間と労力を削減するためではないということである。基礎的作業で節約できた時間を、少しでも多く自分自身の頭で「考える」ことに充てたかったからである。便利なツールのおかげで、比較的短時間で整理されて表示される数々の数値。それらが何を示しているのか、何を語っているのかを、実際の会話の文字化資料や音声の定性的分析によって、じっくり確認していく。そこには、生きた人間の相互行為がある。

本システムが、個々の研究のより深い考察を促し、自然会話をデータとする語用論的なアプローチ全体のさらなる活性化に少しでも貢献することができれば幸いである。

引用文献

宇佐美まゆみ（1997）「基本的な文字化の原則（Basic Transcription System for Japanese: BTSJ）の開発について」、『日本人の談話行動のスクリプト・ストラテジーの研究とマルチメディア教材の試作』、平成 7-8 年度文部省科学研究費基盤研究 C (2)（課題番号：07680312）（研究代表者：西郡仁朗）、研究成果報告書：12-26.

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（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（2011）「基本的な文字化の原則（Basic Transcription System for Japanese: BTSJ）2011 年版」http://www.tufs.ac.jp/ts/personal/usamiken/btsj2011.pdf

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』、国立国語研究所：217-228. http://www.ninjal.ac.jp/event/specialists/project-meeting/files/JCLWorkshop_no3_papers/JCLWorkshop_No3_28.pdf

付記　本研究は、科学研究費補助金基盤研究(A)「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 年度～平成 26 年度、研究代表者：宇佐美まゆみ）による補助を得ている。記して感謝したい。

（うさみ・まゆみ　東京外国語大学大学院教授）