

会話分析入力支援・自動集計システムの開発と活用方法

The development of a system supporting transcribing and automatic tabulation of descriptive statistics of conversations being transcribed based on Basic-Transcription-System-for-Japanese (BTSJ)

宇佐美まゆみ USAMI Mayumi、東京外国語大学大学院

木林理恵 KIBAYASHI Rie、日本学生支援機構

要旨：『BTSJ 文字化入力支援・自動集計システムセット』は、会話の定量的分析を効率的・精緻に進めるためのシステムである。このシステムは、言語使用という相互作用を通して人間関係のあり方やコミュニケーション・ダイナミクスを探る「総合的会話分析」に適するものとして開発された「基本的な文字化の原則（Basic Transcription System for Japanese: BTSJ）」（宇佐美、1997：改訂版最新 2011）に基づき、Microsoft Excel のマクロ機能を利用して作成されている。本システムは、個人の研究の労力を格段に節約するものであり、会話の定性的分析だけでなく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。

キーワード：総合的会話分析、語用論、インタラクション、コーパス言語学、研究方法論

1. はじめに

近年、自然会話をデータとする様々な観点からの研究が増えてきているが、会話の文字起こし作業には膨大な時間と労力がかかる。特に会話を定量的な側面から分析する研究の場合、文字化した会話資料全てに対し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る。本ワークショップでは、このような状況を踏まえて、自然会話の文字化資料作成、及び、定性的分析だけでなく、定量的な分析を効率的かつ精緻に進めるために開発された入力支援システムと基本統計の自動集計・表示システムを紹介する。

2. 『BTSJ 文字化入力支援・自動集計システムセット』の概要

本システムは、「基本的な文字化の原則（Basic Transcription System for Japanese: BTSJ）」（宇佐美 1997：改訂版最新 2011、以下、BTSJ）に基づいている。宇佐美（2008）は、対人コミュニケーション研究の方法論として「自然会話分析への言語社会心理学的アプローチ」（宇佐美、1999）に基づく「総合的会話分析」を提唱している。その研究目的は、言語使用という相互作用を通して人間関係のあり方やコミュニケーション・ダイナミクスを探ることに重点が置かれ、研究方法論としては定量的・定性的双方の観点からの分析を必須としている。BTSJ は、この研究方法に適するものとして開発された「文字化の原則」である。

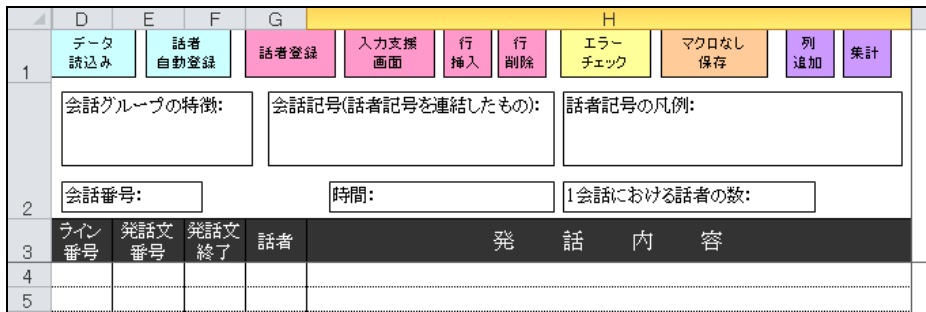
この BTSJ のルールに対応した本システムは、利用者の便宜や汎用性を考えて、Microsoft Excel のマクロ機能を利用して作成されており、「BTSJ 入力支援・自動集計システム(.xlt)」、「BTSJ 複数ファイル自動集計システム(.xls)」の 2 つのファイルからなる。文字化入力支援機能、及び、記述統計のための基本的な分析項目の自動集計機能をそなえており、現在、Windows OS、Excel 2003、2007、2010 に対応している。

3. 『BTSJ 文字化入力支援・自動集計システムセット』の主な機能と操作方法

以下、本システムの機能を具体的な操作方法とともに紹介する。

「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 シートを開くと、図 1 のような画面が現れる。1 列目に、文字化入力やコーディング項目の集計に関する様々な機能のボタンが配置されている。また、2 列目に、話者記号の凡例や会話の時間等、その会話の基本的な情報を記すようになっている。

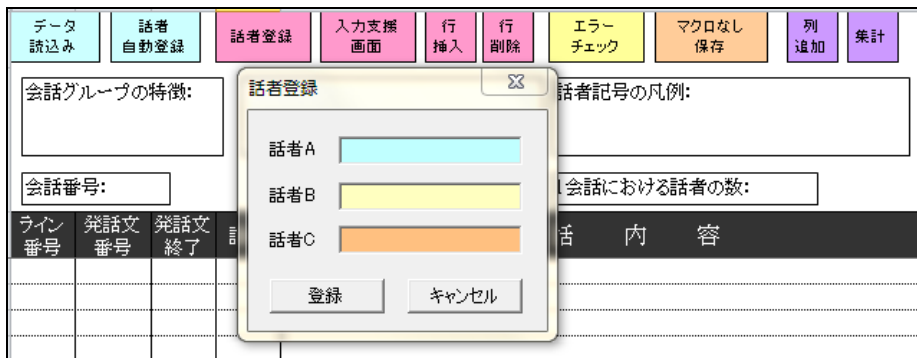
図1 「BTSJ入力支援・自動集計システム」の画面



<文字化入力支援機能について>

最初に、「話者登録」機能を用いて話者を登録する(図2)。登録できる話者数は、現バージョンでは最大3人で、登録後はプルダウンリストから話者を選ぶことができる。

図2 話者登録の画面



BTSJのルールに基づいて発話内容を入力する際には、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。入力支援画面には音声的情報や文脈情報などを示す記号のボタンがあり、対になって入力に手間がかかる記号などを入力する際は、この入力支援を利用することにより、記号の入力を迅速で効率的、かつ正確に行うことができる。(図3参照)

図3 入力支援画面を使用した例



また、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ものであるが、「エラーチェック」のボタンを押すと、記号の半角・全角を自動的に半角に統一したり、誤った入力については、エラーメッセージが出される(次ページの図4参照)。さらに、このとき、分析の単位である「発話文」の番号が、自動

付与される。このエラーチェック機能によって、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。

図4 エラーチェックによってエラーメッセージが抽出された例

| ライン番号 | 発話文番号 | 発話文終了 | 話者 | 発話内容 | |
|-------|-------|-------|------|---------------|-------------------|
| 1 | 1 | * | NS | 「NS姓」と申します。 | 文末コは「。」が「。」が必要です。 |
| 2 | | | BA05 | 「BA05姓」と申します | |
| 3 | 2 | * | NS | あ、どうぞ<2人笑>。 | |
| 4 | 3 | * | BA05 | え?、とか言って<笑い>。 | |
| 5 | 4 | * | BA05 | 日本の方ですよ?。 | |
| 6 | 5 | * | NS | はい?、あ、そうです。 | |

本システムでは、BTSJに基づく文字化資料を新規に作成するだけでなく、既存の文字化資料を本システムに読み込むことも可能である。エクセルで作成された文字化資料は「データ読み込み」ボタンを押せば本システムに読み込み、話者を自動的に登録することができる。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。つまり、本システムによって、既存の文字化資料を定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。

＜記述統計のための基本的な分析項目の自動集計機能について＞

音声文字化後は、各研究者の研究目的に応じて、「列追加」ボタンで分析項目を入力する欄を作り、コーディングを行う（図5参照）。

図5 「文末」のスピーチレベルをコーディングした例

| データ読み込み | 話者自動登録 | 話者登録 | 入力支援画面 | 行挿入 | 行削除 | エラーチェック | マクロなし保存 | 列追加 | 集計 |
|------------|--------|-------------------------------|--------|---|-----|---------|---------|-----|----|
| 会話グループの特徴: | | 会話記号(話者記号を連結したもの): BA05-NS | | 話者記号の凡例: BA…日本人ベース話者 (外国人との接触頻度高) NS…日本人対話者 | | | | | |
| 会話番号: 15 | | 時間: 0分0秒-14分45秒(終) | | 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 | | | | | 文末 |
| 1 | 1 | * | NS | 「NS姓」と申します。 | | | | | P |
| 2 | 2 | * | BA05 | 「BA05姓」と申します。 | | | | | P |
| 3 | 3 | * | NS | あ、どうぞ<2人笑>。 | | | | | NM |
| 4 | 4 | * | BA05 | え?、とか言って<笑い>。 | | | | | N |
| 5 | 5 | * | BA05 | 日本の方ですよ?。 | | | | | P |

分析項目名とコーディングに使用する記号は、Definitionシートに記入する（図6参照）。Transcriptionシートでコーディングを行う際、Definitionシートにない記号を誤って入力すると、集計時に赤字でエラーが示されるので、注意する必要があります。

図6 Definitionシートへの記入例

| | A | B | C | D | E | F | G |
|----|---------------------------------------|----|---|---|---|---|---|
| 1 | Definitionシート | | | | | | |
| 2 | | | | | | | |
| 3 | 集計結果の表の中で表示したい順に、コーディングの記号を縦に入れてください。 | | | | | | |
| 4 | 最後には、必ず「x」と「#」をこの順番に入れてください。 | | | | | | |
| 5 | | | | | | | |
| 6 | コーディング項目名 | 文末 | | | | | |
| 7 | 記号1 | P | | | | | |
| 8 | 記号2 | N | | | | | |
| 9 | 記号3 | NM | | | | | |
| 10 | 記号n | x | | | | | |
| 11 | x | # | | | | | |
| 12 | # | | | | | | |

Transcriptionシートで、集計したいコーディング項目列を選び「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」を自動算出した結果、及び、分析項目ごとの頻度と割合を各項目の総計・話者ごとの総計などのいくつかの観点から自動算出した結果が、新しいシートに表示される。

個別の会話の分析項目を集計したのち、「BTSJ複数ファイル自動集計システム(xls)」を利用して、複数会話の分析

項目の結果を処理する。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き、「BTSJ 複数ファイル自動集計システム(xls)」をクリックして、そのフォルダを選ぶと、複数の会話の分析項目の頻度や割合、その平均値・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した結果が自動表示される(図7参照)。

図7 BTSJ 複数ファイル自動集計システムによって分析項目を処理した例

| 4-7 会話の記録詳細 | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|-----------|-------------|-------------|-------------|------------|------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 基本情報 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 録音文 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 話者 | 開始の 時刻 | 終了の 時刻 | 開始の 標準偏差 | 終了の 標準偏差 | 開始の 平均値 | 終了の 平均値 | | | | | | | | | | | | | | | | | | | | |
| JAMA | 10 | 11 | 1.00 | 0.00 | 0.00 | 0.00 | | | | | | | | | | | | | | | | | | | | |
| JAMR | 10 | 14 | 1.00 | 0.00 | 0.00 | 0.00 | | | | | | | | | | | | | | | | | | | | |
| 合計 | 0 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | | | | | | | | | | | | | | | | | | | | |
| 基本文型 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 開始の 時刻 | 終了の 時刻 | 開始の 標準偏差 | 終了の 標準偏差 | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0 | 1.00 | 0.00 | | | | | | | | | | | | | | | | | | | | | | | |
| コーディング項目別記録詳細 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 表1 各項目の頻度と、録音文数に占める各項目の割合 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 話者 | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | |
| JAMA | 10 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| JAMR | 10 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 合計 | 20 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| コーディング不明の項目(0) コーディング不明の項目(0) (割合は0.00%) | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 表2 各項目の頻度と、録音文数に占める各項目の割合(聞き手側の英語文と、コーディングの対象でない英語文を除く) | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 話者 | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | |
| JAMA | 10 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| JAMR | 10 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 合計 | 20 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| コーディング不明の項目(0) コーディング不明の項目(0) (割合は0.00%) | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 表3 各項目の頻度と、話者ごとの項目の総計に占める各項目の割合 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 話者 | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | |
| JAMA | 10 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| JAMR | 10 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 合計 | 20 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| コーディング不明の項目(0) コーディング不明の項目(0) (割合は0.00%) | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 表4 各項目の頻度と、各項目の総計に占める聞き手側の割合 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 話者 | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | | M | |
| JAMA | 10 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| JAMR | 10 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| 合計 | 20 | 0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| コーディング不明の項目(0) コーディング不明の項目(0) (割合は0.00%) | | | | | | | | | | | | | | | | | | | | | | | | | | |

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自動的に算出・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、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、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。

4. おわりに

会話の分析の基本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」の基礎を支えるものである。本システムを活用して、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、個々の研究を促進するのみならず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られる。

引用文献

宇佐美まゆみ、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キプト・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、1997 年
 宇佐美まゆみ、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 vol.18, no.11、40-56、1999 年
 宇佐美まゆみ、相互作用と学習、講座社会言語科学4 教育、ひつじ書房、150-181、2008 年
 宇佐美まゆみ、改訂版: 基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ) 2011 年度版、from <http://www.tufs.ac.jp/ts/personal/usamiken/btsj.htm>