

# 「基本的な文字化の原則(Basic Transcription System for Japanese: BTSJ)」文字化入力操作・集計自動化版の作成について

宇佐美まゆみ・木林理恵

宇佐美研究室では、Microsoft Excel（以下、エクセル）を用いて「基本的な文字化の原則(Basic Transcription System for Japanese: BTSJ)」（以下、BTSJ）による自然会話の文字化資料作成を行っている。自然会話分析方法の洗練化の一環として、この文字化資料作成作業を効率的かつ精緻に進めるため、エクセルを用いた BTSJ 文字化入力操作・集計自動化のプログラムの作成を、業者委託にて行った。

以下では、現段階での BTSJ 文字化入力操作・集計自動化版について簡単に紹介する。

## 1. BTSJ について

BTSJ は、以下のように表計算ソフト(Excel)で作成されている。BTSJ では、「発話文」を分析の単位としているため、「発話文番号」と「発話文終了」の列を設けている。

ライン番号	発話文番号	発話文終了	話者	発話内容	コードイシング1	コードイシング2
1	1	*	JSM02	あつ、くはじめまして<1>。	P	
2	2	*	JBM04	<くはじめまして><1>。	P	
3	3	*	JBM04	えっと、「JBM04 姓」です。	P	
4	4	*	JSM02	あ、「JBM04 姓」さんですね。	P	ね
5	5	*	JSM02	あの、こちらの大学…、<の方‘かた’ですか?><1>。	P	
6	6	*	JBM04	<あ、そうです><1>。	P	
7	7	*	JSM02	あつ、えっと、僕は、あの、「大学名 1」の、あの、大学院の、「JSM02 姓名」と申します。	P	
8	8	*	JBM04	「JSM02 姓」さん。	NM	
9	9	*	JSM02	はい、「JSM02 姓」は、はい、あの、「漢字 2 を含む言葉」の「漢字 2」って書いて。	N	
10	10-1	/	JBM04	ああ、「漢字 1 を含む言葉」に、	/	
11	11	*	JSM02	そうです。	P	
12	10-2	*	JBM04	「漢字 2」…あつ、「JSM02 姓」、ああ。	NM	
13	12	*	JSM02	はい。	NM	ね

各列について簡単に記す。

- ① 「ライン番号」はラインの通し番号である。
- ② 「発話文番号」とは「発話文」の番号のことである。発話文が終了していない場合はそれが終了するまで「10-1」「10-2」のように通し番号がつけられる。
- ③ 「発話文終了」には、発話文が終了している場合は「\*」が入力される。このとき、「発話内容」の最後には「。」が付いている。発話文が終了していない場合は「/」が入力される。このとき

「発話内容」の最後には「。」が付いている！
④ 「話者」には記号や仮名等で示された話者が入る。話者ごとにラインが色分けされている。
⑤ 「発話内容」には、話者が実際に発話したことが記されている。ここで使用される記号には、以下のものがある。
・ [全角] 1 発話文の終わりにつける。 発話文の途中に相手の発話が入った場合、前の発話文が終わっていないことをマークするためにつけて、改行して相手の発話を入力する。
、 ①[全角] 1 発話文および 1 ライン中で、日本語表記の慣例通りに読点をつける。 ②発話と発話のあいだに短い間がある場合につける。 ①複数読み方があるものを漢字で表す場合、最も一般的な読み方ではなく、特別な読み方で発せられたことを示すために、その読み方を平仮名で「」に入れて示す。 ②通常とは異なる発音がなされた場合など、音の表記だけでは意味が分かりにくい発話は、「」の中に正式な表記をする。
『』 [全角] 視覚上、区別した方が分かりやすいと思われるもの、例えば、本や映画の題名のような固有名詞や、発話者がその発話の中で漢字の読み方を説明したような部分等は、「」でくる。
“” 発話中に、話者及び話者以外の者の発話・思考・判断・知覚などの内容が引用された場合、その部分を“”でくる。
? 疑問文につける。疑問の終助詞がついた質問形式になっていても、語尾を上げるなどして、疑問の機能を持つ発話には、その部分が文末(発話文末)なら「？」をつける。倒置疑問の機能を持つものには、発話中に「？」をつける。
?? 確認などのために語尾を上げる、いわゆる「半疑問文」につける。
[↑][→][↓] イントネーションは、特記する必要のあるものを、上昇、平板、下降の略号として、[↑][→][↓]を用いる。
/少し間/ 話のテンポの流れの中で、少し「間」が感じられた際ににつける。
/沈黙 秒数/ 1 秒以上の「間」は、沈黙として、その秒数を左記のように記す。沈黙自体が何かの返答になっているような場合は 1 発話文として扱い 1 ライン取るが、基本的にには、沈黙後に誰が発話したのかを同定できるように、沈黙を破る発話のラインの冒頭に記す。
= = 改行される発話と発話の間(ま)が、当該の会話の平均的な間(ま)の長さより相対的に短いか、またくないことを示すためにつける。これは、2 つの発話(文)について、改行していくても音声的につながっていることを示すためである。その場合、最初のラインの発話の終わりに「=」をつけてから、句点「。」または英語式コンマ 2 つ「,,」をつける。そして、続行ラインの冒頭に「=」をつける。
… 文中、文末に関係なく、音声的に言いよどんだように聞こえるものにつける。
< >< > 同時発話されたものは、重なった部分双方を< >でくくり、重ねられた発話には、< >の後に、< >をつけ、そのラインの最後に句点「。」または英語式コンマ 2 つ「,,」をつける。また重ねた方の発話には、< >の後に、< >をつける。
【】 [全角] 第 1 話者の発話文が完結する前に、途中に挿入される形で、第 2 話者の発話が始まり、結果的に第 1 話者の発話が終了した場合は、「【】」をつける。結果的に終了した第 1 話者の発話文の終わりには、句点「。」の前に【】をつけ、第 2 話者の発話文の冒頭には】】をつける。

<sup>1</sup> ただし、周辺言語情報を示す[ ]に囲まれた情報が末尾にある場合、「。」および「,,」はその前に置かれる(発話内容自体と区別するため)

- [ ] 文脈的情報。その発話がなされた状況ができるだけわかりやすくなるように、音声上の特徴(アクセント、声の高さ、大小、速さ等)のうち、特記の必要があるものなどをそのラインの一番最後に記しておく。
- ( ) 短く、特別な意味を持たない「あいづち」は、相手の発話中の最も近い部分に、( )にくくって入れる。
- < > 笑いながら発話したものや笑い等は、< >の中に、「笑いながら」、「2人で笑い」などのように説明を記す。笑い自体が何かの返答になっているような場合は1発話文となるが、基本的には、笑いを含む発話中か、その発話文の最後に記し、その後に句点「。」または英語式コマ「,,」をつける。
- < > 相手の発話の途中に、相手の発話と重なって笑いが入っている場合は、短いあいづちと同様に扱って、<笑い>とする。
- # 聞き取り不能であった部分につける。その部分の推測される拍数に応じて、#マークをつける。
- 「 」 トランスクリプトを公開する際、固有名詞等、被験者のプライバシーの保護ために明記できない部分を別の表現や番号と置き換えるときに用いる。

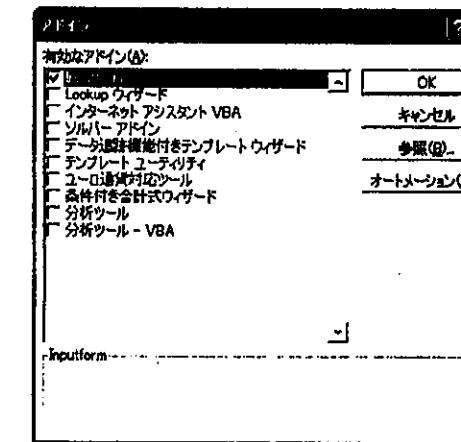
⑥ 「コーディング」には、研究目的によって様々な記号が付けられる。コーディングの列(種類)がいくつになるか、1列につき記号(項目)がいくつあるかは、研究目的ごとに設定される。

## 2. BTSJ 文字化入力操作自動化仕様について

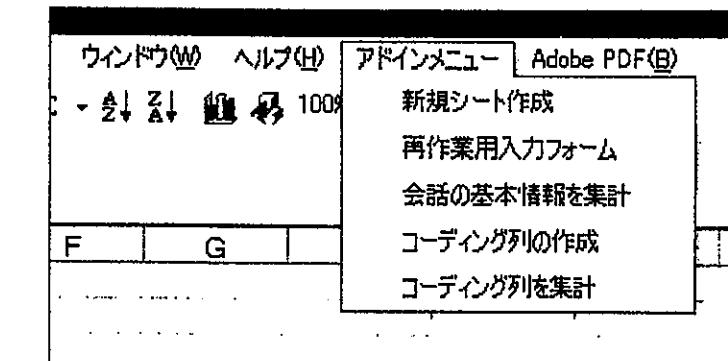
以下では、現段階での BTSJ 文字化入力操作自動化仕様について簡単に紹介する。

### 2.1 アドインについて

まず、業者が作成したプログラム「inputForm.xla<sup>2</sup>」ファイルを、  
C:\Documents and settings\ユーザー名\ApplicationData\Microsoft\AddIns に保存する。  
保存後、エクセルを起動して「ツール」メニューの中の「アドイン」を開くと、「アドイン」ダイアログボックスに「inputForm」が表示されているので、チェックを入れて OK をクリックする。

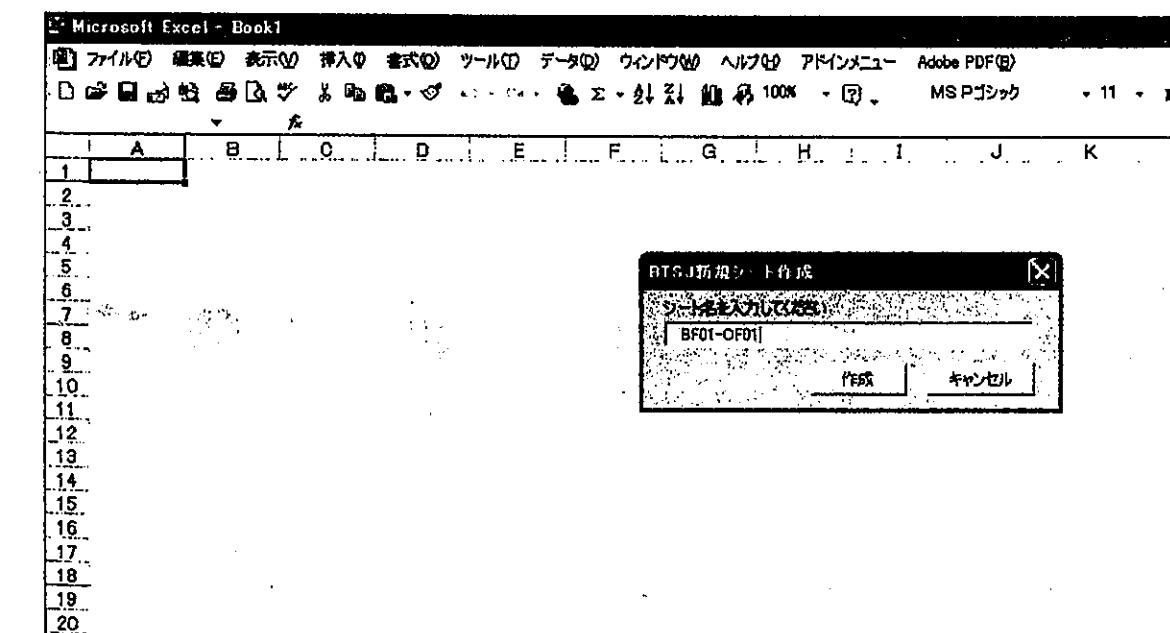


そうすると、Excel の「ヘルプ」の横に「アドインメニュー」が追加され、そこからマクロを実行することができる。



## 2.2 発話内容の入力について

最初に、シート名を記入し、作成をクリックする。



そうすると一番目のラインに「ライン番号」、「発話文番号」、「発話文終了」、「話者」、「発話内容」が記入されている基本フォームができる。

<sup>2</sup> ファイル名は、改良により変化する可能性がある。

次に、話者の記号を入力する。話者は3名まで入力することが出来る

	A	日	C	D	E
1	ライン番号	発話文番号	発話文終了	話者	発話内容
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					

「話者」セルにおける各話者（記号または仮名になっている）につき 1 つの背景色を割り当て  
当該の話者の行について、割り当てられた色を付けるようになっている。

話者を入力すると、以下のようななかたちでBTSJ入力フォームが出てくる

A	B	C	D	E	F	G
1 ライン 番号	英語文 番号	英語文 終了	話者	発話内容		
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						

「話者選択」で話者を選び、「発話内容入力」に、発話内容を入れる。つまり、話者と発話内容をラインずつ入力していく。アイコンを押すだけで記号が入力されるので、入力の煩雑さが減るばかりでなく、記号を間違える可能性も低くなる（例：括弧を閉じ忘れる）また、<>{}の後に続けて

>{<}を入れてしまうというようなミスをした場合、警告を出すようになっている。

1 ラインの入力が完了したら、完了したことを示すボタン (Enter キーの機能) を押して、次のラインに行くようになっている。

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
ライン番号	発話文番号	発話文終了	話者	発話内容														
2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
<div style="border: 1px solid black; padding: 10px;"> <p><b>BISJ入力フォーム</b></p> <p><b>話者選択</b></p> <p>話者A(青い黄色) BF01</p> <p><b>入力</b> <input type="button" value="入力"/></p> <p><b>操作</b> <input type="button" value="一つ上に戻る"/> <input type="button" value="終了"/></p> <p><b>発話内容入力</b></p> <p>はじめまして</p> <p>。 。 ‘ ’ ‘ ’ / / [ ] ( ) &lt; &gt; &lt;&gt; 「 」 「 」 &lt;&gt;[ ] &lt;&gt;D [↑] [↓] [→]</p> </div>																		

「。」を入力し、「入力」ボタンを押すと、以下のように発話内容がエクセルに送られ、ライン番号・発話文番号が自動的に付される。発話内容に「。」あるいは「,,」が含まれていないと、発話内容入力をエクセルに送ることができないようになっている。このため、「。」や「,,」の漏れや、発話文終了セルとの記号の不一致といったミスを防ぐことができる。

A	B	C	D	E	F
ライン番号	発話文番号	発話文終了	話者	発話内容	
1	1	*	BF01	はじめまして。	
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					

再入力が必要な場合は、「一つ上に戻る」を利用する。BTSJ 入力フォームを消し手しまった場合は、アドインメニューから「再作業用入力フォーム」を出すことができる。

エクセルに送られた「発話内容は」、記号凡例における「半角」「全角」という指定に合わせて、「半角」なら「半角」に、「全角」なら「全角」に自動的に調整できるようになっており、例えば、間違えて「全角」で入力

しても、Enter を押した後「半角」に修正される。アラビア数字とアルファベットも、自動的に半角になるようにしてある。

入力した文字化資料は、EXCEL 形式で保存される。

### 2.3 集計機能について

以下のように、会話の基本情報およびコーディングの集計を行うことができる。

The screenshot shows the software's main menu at the top with options like ファイル(E), 編集(E), 表示(S), 挿入(I), 書式(O), ツール(T), データ(D), ウィンドウ(W), ヘルプ(H). A sub-menu "アドインメニュー" is open, showing "新規シート作成", "著作権用入力フォーム", "会話の基本情報集計", "コーディング列の作成", and "コーディング列を集計". Below the menu is a table with columns "話者" (Speaker) and "発話内容" (Speech Content). The table contains several rows of data, each with a code like OF01 or BF01, followed by a Japanese sentence and a category code like P, NM, or S.

会話の基本情報については、アドインメニューから「会話の基本情報の集計」をクリックすると1会話につき、例1のように、以下の情報を自動的に算出することができる。

- ライン数(話者ごとの頻度及び合計)
- 発話文数(話者ごとの頻度及び合計)
- 全発話文数における各話者の発話文数の頻度と割合(小数点第2位まで)
- 相手の発話中のあいづち: 相手の発話文数
- 相手の発話中の笑い: 相手の発話文数
- 話者交替の頻度の集計

例1) 会話の基本情報

話者	ライン 数	ライン数 の割合	発話文数 の頻度	発話文数 の割合	相手の発話中の あいづち		相手の発話中 の笑い	
					頻度	割合	頻度	割合
A	8	50.00%	8	50.00%	3	37.50%	0	0.00%
B	8	50.00%	8	50.00%	2	25.00%	0	0.00%
合計	16	100.00%	16	100.00%	5	62.50%	0	0.00%
話者交替								
13								

コーディング項目の集計については、アドインメニューから「コーディング列を集計」を選んで、「コーディング」の列ごとに、各項目の頻度と割合を例2のように集計できる。集計結果は、会話資料とは別のシートに出てくる。また、「コーディング」の列が1列増えるごとに、シートもひとつ増える。集計対象は以下の通りである。

- 総発話文数における、各項目の頻度と割合(分母は、総発話文数)
- 各項目の総計における、各項目の頻度と割合(分母は、項目の総計)
- 各項目の総計における、話者別の頻度と割合(分母は、項目別の総計)

例えば、会話における文末のスピーチレベルを以下のようにコーディングしたとき、集計結果は例2のようになってくる。

#### コーディング:

「P」発話文末が「ですます」体やその活用形等である発話文。

「N」発話文末が「だ/である」体やその活用形等である発話文。

「NM」発話文末に丁寧度を示すマーカーのない発話文。典型的なものでは、あいづち詞、応答詞、中途終了型発話等がこれに当たる。

#### 例2) コーディングの集計

The screenshot shows the software's main menu at the top with options like ファイル(E), 編集(E), 表示(S), 挿入(I), 書式(O), ツール(T), データ(D), ウィンドウ(W), ヘルプ(H). A sub-menu "アドインメニュー" is open, showing "新規シート作成", "著作権用入力フォーム", "会話の基本情報集計", "コーディング列の作成", and "コーディング列を集計". Below the menu is a table with various sections and data rows. The sections include "総発話文数における、各項目の頻度と割合", "各項目の総計における、各項目の頻度と割合", and "各項目の総計における、話者別の頻度と割合". The data rows show counts and percentages for speakers A and B across categories P, S, and NM.

以上。