

話し言葉の分析のための文字化入力・集計支援ソフトの紹介とその使い方

- レストランでの注文場面のデータを例に -

宇佐美まゆみ
東京外国語大学

1. はじめに

近年、話し言葉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統計処理を行うといった一連の手順があり、細かく正確な作業が求められる。本発表では、このような状況を踏まえて、文字化の作業労力を最小限にとどめ、その分析の効率化を図るために開発された、話し言葉の文字化入力支援及び分析項目の自動集計システムを紹介する。

本発表では、まず、開発の目的・方針、表記法、分析の単位、改行の仕方、音声的情報や発話内容以外の周辺言語的情報などの記載といった、文字化作業に際して必要な項目を具体的に紹介する。その後、実際の会話をを用い、入力支援フォームによって発話内容の入力が簡便にできることや、エラーチェック機能によって入力ミスが抽出されることなどを実演し、このシステムによって文字化作業が効率的に行えることを示す。また、作成した文字化資料へのコーディングを行い、プログラムによる自動集計を簡単に実演することで、分析項目の集計作業が短縮化・精緻化されたことを示す。

会話の定量的な分析を行う際には、文字化・コーディング・集計といった手順において非常に緻密な作業が必要とされ、膨大な時間がかかる。このような基本的な作業を短い時間で正確に行えるシステムは、会話の定量的研究の基礎的な部分を支えるものであり、「生活日本語のプロフィシエンシー」の研究として、生活日本語を構成する様々な会話データの分析にも用いることができる。デモンストレーションでは、実際の「レストランでの注文場面」の録画データを用いて、その分析方法についても、論じたい。

2. 言語社会心理学的アプローチと「BTSJ 入力支援システム」の概要

紙幅の都合もあるので、以下には、簡潔に「BTSJ 入力支援システム」が依拠する方法論とシステム開発の趣旨をまとめる。

2.1 言語社会心理学的アプローチ

言語社会心理学 (Social Psychology of Language) とは、「言語と言語使用」の分析を通して、「人間」と人間が形作っている社会」の問題に迫る」分野と言える (宇佐美、2001: 231)。「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 :宇佐美 2007)は、自然会話分析への言語社会心理学的アプローチ「総合的会話分析 (宇佐美 1999 等)」に最も適するように開発された、定量的・定性的分析双方が行える汎用性を備えたシステムである。つまり 会話のやりとりを定性的に捉えながら、扱いたい現象をコーディング (記号化) し、定量的分析を行う 統計処理を行うことも念頭においてデータ収集等を行うことにより、目的やデータ量によっては、人間同士の相互作用の特徴の一般化も可能になる。同時に、定性的分析も行い、定量的分析では扱えなかった部分の分析も行う

「自然会話分析への言語社会心理学的アプローチ「総合的会話分析」の方法論の骨子は以下のようまとめられる。以下の6つの手順の全てを含んでこそ意味を持つものである。

条件統制してデータを収集する

フェイス・シート・フォローアップ・アンケート等で、会話の参加者や、会話収録時の状況など、会話以外の状況も、必要に応じたものを作成して確認する。

厳密、且つ、必要最低限の会話の文字化資料を作成する。

日本語プロフィエンス研究会国際シンポジウム予稿集』、日本語プロフィエンス研究会:147-149. 3 頁.
研究目的に応じたコーディングを行う

コーディングの評定者間信頼性係数 (Cohen's Kappa)を算出して信頼性を確認する。
定性的な分析によって、定量的分析だけでは明らかにできなかった点の分析や考察を行う

2.2 「BTSJ 入力・集計支援システム」の概要

「BTSJ 入力・集計支援システム」は、「文字化のルール」としての「基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ) 」(宇佐美、1997、最新版は、<http://www.tufs.ac.jp/ts/personal/usamiken/index.htm> からダウンロードできる)に基づく「入力・集計支援システム」のことである。紙幅の都合もあり、以下には、「BTSJ 入力支援システム」によって、どのようなことが便利になったのか、作業効率がいかに改善されるのかについて、箇条書きでまとめておくに留める。

文字化作業の効率化

入力補助による、発話内容の効率的な入力

入力エラーのチェック(自動修正及びエラーの抽出)

発話文」の通し番号の自動付与

分析項目の自動集計 (ひとつの分析項目の頻度と割合等を、複数の観点から算出)

総発話文数と話者交替数の算出

複数の会話の基本分析項目の合計や平均の自動算出

2.3 「BTSJ 入力・集計支援システム」の文字化や分析項目の表示例 (イメージのみ)

話者 自動登録	話者登録	行 挿入	行 削除	入力支援 画面	入力記号 チェック	エラー チェック	列 追加	集計	保存
会話グループ名:	会話記号:	記号凡例:							
会話番号:	時間:	1会話における話者の数:							

ライン番号	発話文番号	発話文終了	話者	発話内容	語彙
1	1	*	NSF08	はじめまして<笑い>。	P
2	2	*	NNSF08	はじめまして。	P
3	3	*	NNSF08	「NNSF08 名」と申します。	S
4	4	*	NSF08	あ、で、「NNSF08 名」さん?。	P

【参考文献】

宇佐美まゆみ 1997a 「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 1997b 「自然会話分析法への言語社会心理学的アプローチ」『日本人の談話行動のスクリプ

- 日本語プロフィエンス研究会国際シンポジウム予稿集』、日本語プロフィエンス研究会:147-149. 3 頁.
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: 1-5.
- 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 2001 「21 世紀の社会と日本語 -ポライトネスのゆくえを中心に-」『月刊言語』30(1) (1 月号特集「21 世紀の日本語」)、大修館書店: 20-28.
- 宇佐美まゆみ、木山幸子 2007a 「人間の相互作用研究の基盤となる文字化システム :会話教材作成への示唆」談話研究と日本語教育の有機的統合のための基礎的研究とマルチメディア教材の試作』平成 15-18 年度科学研究費補助金基盤研究 B(2)研究成果報告書(課題番号 15320064):7-16 .
- 宇佐美まゆみ、木林理恵 2007d 「基本的な文字化の原則(Basic Transcription for Japanese:BTSJ)』文字化入力操作 集計自動化版の作成について」談話研究と日本語教育の有機的統合のための基礎的研究とマルチメディア教材の試作』平成 15-18 年度科学研究費補助金基盤研究 B(2)研究成果報告書(課題番号 15320064):40-47 .
- 宇佐美まゆみ、李恩美、鄭榮美、金銀美 2007e 「基本的な文字化の原則(Basic Transcription system for Japanese: BTSJ)』の韓国語への応用について」談話研究と日本語教育の有機的統合のための基礎的研究とマルチメディア教材の試作』平成 15-18 年度科学研究費補助金基盤研究 B(2)研究成果報告書(課題番号 15320064):48-82.
- 宇佐美まゆみ、肖? ?、戴琦、高娃、李宇霞、仇曉? 2007f 「基本的文字化の原則(Basic Transcription system for Japanese: BTSJ)の中国語への応用について」談話研究と日本語教育の有機的統合のための基礎的研究とマルチメディア教材の試作』平成 15-18 年度科学研究費補助金基盤研究 B(2)研究成果報告書(課題番号 15320064):83-103 .
- 宇佐美まゆみ、鈴木卓 2007g 「基本的な文字化の原則(Basic Transcription system for Japanese: BTSJ)の英語への応用について」談話研究と日本語教育の有機的統合のための基礎的研究とマルチメディア教材の試作』平成 15-18 年度科学研究費補助金基盤研究 B(2)研究成果報告書(課題番号 15320064):104-112 .
- Bakeman, R. & Gottman, J.M. (1986). Observing interaction: an introduction to sequential analysis. Cambridge University Press.

【BTSJ (基本的な文字化の原則 :文字化入力のルール)】

以下のサイトより 最新版がダウンロードできる。

宇佐美まゆみ研究室

<http://www.tufs.ac.jp/ts/personal/usamiken/index.htm>

【公開コーパス】

以下のサイトにある利用申請書の提出によって、無料でダウンロードできる。

宇佐美まゆみ監修 (2007)

『BTS による多言語話し言葉コーパス- 日本語会話 1 (日本語母語話者同士の会話)』

『BTS による多言語話し言葉コーパス- 日本語会話 2 (日本語母語話者と学習者の会話)』

<http://www.tufs.ac.jp/ts/personal/usamiken/corpora2007.htm>

宇佐美まゆみ監修 (2007)

『BTSJ による日本語話し言葉コーパス1 (初対面・友人、雑談・討論・誘い)』

一部音声付

<http://www.tufs.ac.jp/ts/personal/usamiken/kaken2007corpora.htm>