

自然会話を素材とする共同構築型 WEB 教材を使った

「対話」と「会話」の教育

宇佐美 まゆみ
国立国語研究所

要旨

本稿では、「共同構築型多機能データベース」である「NCRB: Natural Conversation Resource Bank）」に搭載された母語場面、接触場面双方を含む自然会話を素材とする WEB 教材を紹介するとともに、対話と会話の教育を考える。この WEB 教材は、様々な母語場面、接触場面における自然な会話をリソースバンクとしてデータベース化し、研究用データを提供するとともに、共同で WEB 教材を構築していくことができるプラットフォームである。「自然会話の素材（音声・動画）の投稿」、「文字化資料のアップロード、ダウンロード」及び、「自然会話を素材とする教材」の「教材作成支援機能」を搭載しており、この機能を用いると、入力画面に設定された、場面、話者等の内容を入力してだけで、動画とその会話内容に対応した「文字化テキスト」とその解説、Q&A 等が、WEB 教材として表示される仕組みになっている。会員は、この自然会話リソースバンクのデータを用いて、「研究」と、「教材作成支援機能」を用いた「教材作成」及び「教材利用」ができる。

【キーワード】 自然会話コーパス、WEB 動画教材、自然会話教材作成、共同構築型データベース、語用論的分析

1 はじめに

はじめに「共同構築型多機能データベース」である「NCRB: Natural Conversation Resource Bank）」開発の背景・趣旨をまとめておく。「自然会話を素材とする WEB 教材」は、この NCRB に搭載されており、「共同構築型」という新しい形の WEB 教材である。

2 NCRB 開発の背景

コミュニケーションの語用論的研究のためには、いわゆる「コーパス言語学」で扱われているような大量の「テキスト」を編んだものだけではなく、比較的少量でも、話者の関係や話題等々の諸条件を統制して収集された「会話（音声・動画）」とともに、同時発話、割り込み、沈黙等の語用論的分析に必要な情報が付与された「トランスクリプト（文字化資料）」が収録されたコーパスが必須である。近年、様々な目的に基づくコーパスの構築が増えてきているが、語用論的分析には適さないコーパスがほとんどであるのが現状である。また、日本語教育に関係の深い、いわゆる「学習者コーパス」も、未だ語彙や文法項目の習得研究を主目的とするものが多いため、文字化の原則は、比較的簡素で、語用論的研究に必要なオーバーラップの情報などが付与されていないものが多い。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

な詳細な情報が付与されているものの、比較的少数の会話の定性的な分析には適していても、より数の多い会話データの定量的処理には適さない形になっている。そのため、定性的分析だけではなく、定量的分析も可能にし、研究者間で共有するためのコーパスの構築には適さない。

主に、言語使用に着目した語用論的研究のための自然会話データのコーディングは、品詞などの言語形式のみに着目する分類とは異なり、例えば、「話題導入発話」や「沈黙」等、研究者自身が研究目的や分析の観点に応じて判断しながら、行う必要がある。ただ、このようなコーディングやその集計には、膨大な時間と労力がかかるため、その時間を合理的に節約することによって、より深い質的分析も含む考察に時間がかけられるようにする必要がある。さらには、この分野の研究の発展のためには、研究者各自が労力を軽減するばかりでなく、時間と労力をかけて収集したデータを、研究者間で広く共有するための「語用論的分析に適したコーパスの構築」が必須である（宇佐美監修 2011）。しかし、個人、研究グループ、研究機関単位でさえ、コーパスを構築するには、一定の時間や労力、人件費などがかかるため、十分な予算がなければコーパスの構築も難しい。また、そのような従来式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。そのため、これからは、もっと「双方向的な参加・共有型コーパスの構築」のあり方を考える必要がある。

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」（宇佐美、2008）の方法論や、人間同士の相互作用を、定量的・定性的双方から分析するのに適する「文字化の原則」として「基本的な文字化の原則（Basic Transcription System for Japanese: BTSJ）」（宇佐美、改訂最新版 2015）を開発し、BTSJ 文字化資料と音声（一部）つきコーパスである『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』（宇佐美監修 2011）を開発した（以降 BTSJ コーパス）。また、その後、「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」を開発して自動集計機能等によって、コーパスの語用論的分析の時間と労力を格段に短縮した。その後、「BTSJ 文字化資料」と音声・動画つきの「BTSJ コーパス」を「データベース化」した。NCRB は、上記のような趣旨から生まれたものであり、元々、「総合的会話分析」（宇佐美、2008）を基本とする自然会話の語用論的分析のための、データベースであった。

3 NCRB 開発の趣旨

しかし、自然会話データは、研究のみならず、コミュニケーション教育のための教材（materials）にも成り得る（宇佐美、2012）。しかし、「自然会話を素材とする教材」の作成には、コーパス構築と同様、文字化に時間がかかる上に、さらに「教材としての解説」等も加えていかなければならない。そのため、「自然会話教材作成支援システム」のようなものを作成・活用して、教材作成の時間と労力を削減することが必須であると考えに至った。そこで自然会話コーパス、その他のデータを自然会話リソースとして、自然会話の動画と文字化資料に基づく共同構築型 WEB 教材として、「自然会話教材作成支援システム」機能と「自然会話を素材とする教材」を搭載していくことにした。それが、NCRB (Natural Conversation Resource Bank) である（宇佐美 2013）。

NCRB は、上記のような必要性を踏まえて構築された、新しいタイプの「共同構築型多機能データベース」であり、また、「共同構築型 WEB 教材リソースバンク」でもある。

4 NCRB の利用環境、基本構成、登録方法

NCRB の利用環境は、OS : Windows7, 8、ブラウザ : Internet Explorer9~11, Google Chrome で、登録可能なデータの形式は、mp4 (動画)、mp3 (音声) ファイルである。構成は、大きく、「自然会話データを使った研究」と「自然会話を素材とする教材」の2つに分かれている。それぞれ、「登録や作成」と「利用のみ」で入り口が分かれており、「自然会話を使った研究」は、「データを登録したい方はこちら」及び「データを利用したい方はこちら」、また、「自然会話を素材とする教材」は、「教材を作成したい方はこちら」及び「教材を利用したい方はこちら」と、全体として4つの入り口がある。会話の登録と教材の作成には、アクセス用の ID とパスワードが必要である。会話データ (音声・会話スクリプト、それらに関する情報) は、NCRB トップページ (図1) より、「自然会話データを使った研究」の「データを登録したい方はこちら」から登録する。作業の流れは、「音声・動画のアップロード」→「音声・動画情報の入力」→「話者情報の入力」→「会話情報の入力」→「会話スクリプトの入力」となる。各作業は「音声・動画」、「会話」、「会話グループ」という3つの画面で行い、上部のタブを切り替えることで画面を移動できる。会話スクリプトはNCRB 上でも、動画や音声を聞きながら作成していくこともできるが (図2)、「BTSJ システムセット」を用いて文字化資料を整えた後、NCRB にアップロードすることもできる (研究には、文字化入力支援機能が多く自動集計もできる後者を推奨する)。また、会話スクリプト画面では動画や音声にタイムスタンプを押すことができ、文字化や分析の際に動画や音声の頭出し再生がしやすくなっている。



図1 NCRB トップページ



図2 会話スクリプト入力画面



図3 会話グループ画面

このようにして登録した1つ1つの会話は、「会話グループ作成機能」を用いてグループ化することができる(図3)。自分のデータを1つのグループにまとめることもできるし、検索機能を用いて、「友人同士の雑談の会話」等、研究目的に応じた条件で検索し、該当する会話のみをグループ化して保存することもできる。このように、同様の条件の複数の会話を、定量的に分析することが可能である。

NCRBはBTSJシステムセットと連携している。NCRB上で決められた作業手順通りにデータ情報を入力すると、NCRB番号が自動的に付与される。このデータをダウンロードし、BTSJシステムセットを用いて文字化資料を整えることによって、通し番号の間違いや情報の入力漏れといった、データ整備の際に起きやすいケアレスミスを防ぐことができる。さらに、BTSJシステムセットでは、研究のためのコーディングや基本的記述統計の自動集計ができる。NCRBの検索機能を用いて自分の研究目的に合うデータを複数抽出し、一括ダウンロードしてBTSJシステムセットの自動集計機能等を用いることによって、会話の定量的な分析を効率よく行うことが可能になる。

5 自然会話教材の作成方法

NCRBでは、自然会話教材を、基本的に、動画は1シーンにつき1分程度の長さで、3シーンでひとつのユニットを構成するように作成する。具体的な手順は、次の通りである。まず、トップページの「自然会話データを使った教材」の「教材を作成したい方はこちら」にアクセスする。自然会話の素材は、既にNCRBに登録されている会話から選ぶことができる。教材の目的に適したファイルを検索し、抽出することによって、その場面に応じた自然会話教材を各自が作成することができる。

自分で収集したデータを使用するときは、まず、「データを登録したい方はこちら」にアクセスして、データの諸情報を登録する必要がある。データ登録の際に、「会話スクリプト入力」画面で発話にタイムスタンプを押すことによって、完成した自然会話教材画面では、文字と動画が連動し、その時に発話されている内容が会話スクリプトの中で示される教材の形になる(図4)。作成は、各ユニットの名称や、各シーンの題目、場面、話題、話者同士の関係、スピーチレベルの基本状態を入力画面を通して入力した後、会話スクリプトの編集ボタンを押して、必要な発話に対して、Contents、Expressions(文法・機能・その他)、Strategies、Politeness、Cultureという5つの観点から、該当する学習項目の解説を記入していく(図5)。そして、「教材情報編集画面」から、理解を助けるための質問(選択式・記述式)とその解答、及びそれらの解説を入力していく(図6)。入力の途中や終了後は、プレビュー画面で、教材の完成版のイメージを確認することができる。それらがすべて修了したら保存して、トップページの「教材を利用したい方はこちら」に入れば、作成した教材をすぐに利用することができる。



図4 動画と会話スクリプトの連動 (完成画面の一部)

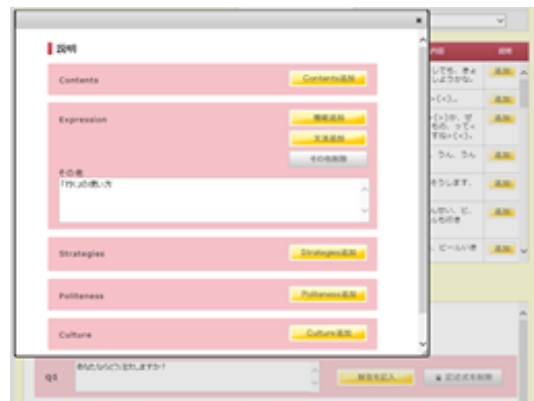


図5 学習項目や解説の入力画面

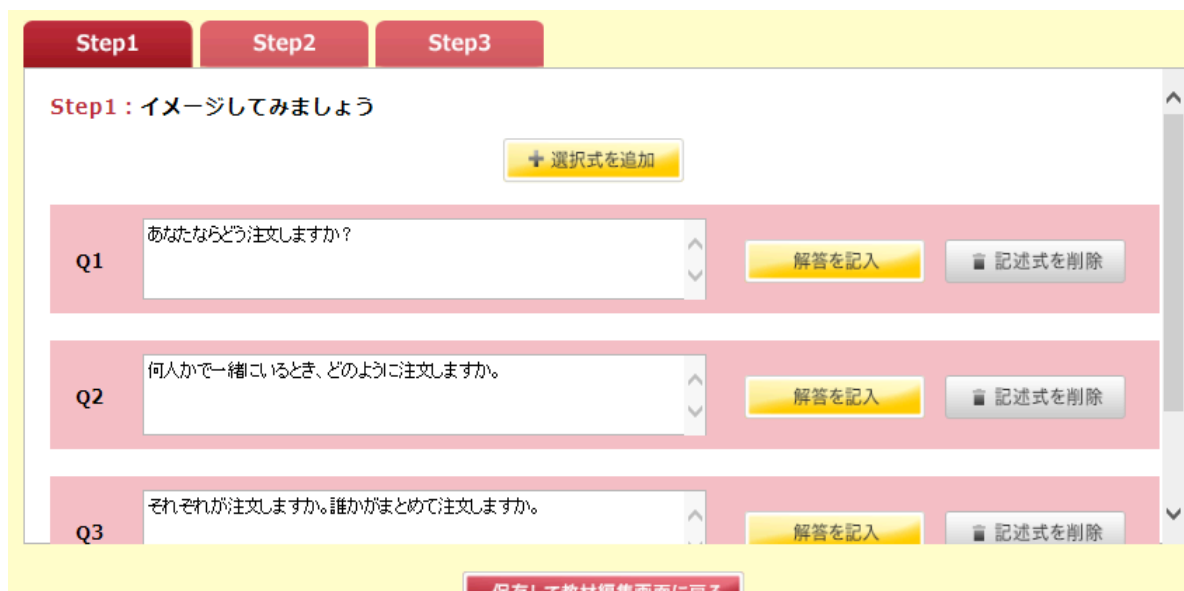


図6 Q&A の入力画面

6 対話と会話の教育について

昨今、厳密には、「会話」と「対話」は、異なるものであることが認識されつつある。「会話」とは、「価値観や生活習慣なども近い、親しい人同士のおしゃべり」を指し、「対話」とは、「あまり親しくない人同士の価値や情報の交換、あるいは親しい人同士でも価値観が異なるときに起こる価値のすり合わせなど」と定義されている（平田、2012）。母語場面、接触場面、双方を含むNCRB データバンクの自然のやりとりのデータから、まずは、場面による「対話」と「会話」の出現状況や特徴の分析をすることが重要である。該当するデータを教材作成支援機能を用いて「WEB 教材化」していく過程の中から、その特徴や教育についての重要な視点が見えてくるからである。

7 今後の課題

まず、現実的な問題として、本データベースをより広く一般に公開していくためには、

サーバーの管理、運営方法やその原則などのルール作りをする必要がある。また、「データ提供者」各自が、登録データの肖像権や著作権に適切に対処した上で登録することを徹底させる必要がある。また、データ提供者とデータ利用者の役割が一方的、固定的にならないように、なんらかのルールを設けていく必要がある。一方で、より多くの人々がデータ提供への意欲を高めてくれるよう工夫する必要がある。

また、NCRB の「自然会話教材作成支援システム」によって作成した教材の WEB 上での利用状況の把握や分析、クラス内での利用について、学習者の反応や利用効果を何らかの形で検証していくというような「評価」も行っていく必要がある。そのためにも、データ提供者、兼、利用者として、多くの方々に利用してもらえようようにしたいと考えている。

付記：宇佐美まゆみ研究室「NCRB (Natural Conversation Resource Bank)」は、『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 23-26 年度科学研究費補助金基盤研究 (A) (研究代表者：宇佐美まゆみ) -(課題番号 23242027) による研究成果の一部である。

<引用文献>

- 平田オリザ (2012). 日本語教育と国語教育をつなぐ「対話」、『対話とプロフィシエンシー』、凡人社。
- 宇佐美まゆみ (2007) 「自然会話の教材化とディスコース・ポライトネス理論 2:教材としての自然会話の価値」『第一回ルーマニア日本語教師会日本語教育・日本語学シンポジウム報告書』ルーマニア日本語教師会. Avrin Press. pp.26-38.
- 宇佐美まゆみ(2008)「相互作用と学習ーディスコース・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、pp.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」<http://www.tufs.ac.jp/ts/personal/usamiken/btsjcorpus explanation.htm> から申請すれば無償で得られる。
- 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、pp.63-82.
- 宇佐美まゆみ(2013)「NCRB (Natural Conversation Resource Bank) の開発とその意義についてーこれからのコーパスのあり方とその研究・教育への活用法への一提案ー」『第8回日本語実用言語学国際会議 (ICPLJ8) Conference Handbook』国立国語研究所、pp.128-131.
- 宇佐美まゆみ(2015a)「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 2015 年改訂版」『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 23-26 年度科学研究費補助金基盤研究(A)(研究代表者：宇佐美まゆみ) (課題番号 23242027)
- 宇佐美まゆみ(2015b)「改訂版：基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ) 2015 年度版」<http://www.tufs.ac.jp/ts/personal/usamiken/btsj.htm> からダウンロードできる。
- 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 言語資源研究系・コーパス開発センター、pp.217-228. Marke Zine 2014/04/07, <http://markezine.jp/article/detail/19622>