第8回日本語実用言語学国際会議 (ICPLJ8) 発表論集 pp.128-131 (@国立国語研究所 3.22-3.23, 2014)

【分野名】語用論、会話分析、コーパス言語学

# 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・教育への活用法の一提案ー

宇佐美まゆみ
(東京外国語大学大学院総合国際学研究院)

【要旨】

本発表では、オーセンティックな相互作用としての自然会話を中心に編まれた自然会話コーパスをデータベース化するとともに、今後は、広く希望者(登録者)にも、研究と教育双方に利用してもらえるよう企図して開発している NCRB(Natural Conversation Resource Bank) (宇佐美、2014)について、開発過程、特徴や意義、その研究・教育への活用法について紹介する。さらには、このデータベースが音声・動画とともに収録する「基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ)」(宇佐美、1997～2011) に基づくトランスクリプトと、「『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』(宇佐美、2014 版) と NCRB の関係、さらには、「自然会話を素材とする教材の開発」(宇佐美、2007ab; 2012) の趣旨と NCRB との関係についても紹介する。その上で、特に、語用論的研究やコミュニケーション教育に活用するための、これからのコーパスのあり方とその構築、それに基づく日本語コミュニケーションの研究や教育のあり方について論ずる。

## 1. コーパスにかかわる問題点と課題

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、音声認識などの工学的目的のものには、音声・動画のみで、文字化資料が付与されていないものも多い。また、テキストデータがある場合も、コーパス言語学における「タグ付け」のように、形態素解析や言語形式の頻度やコロケーションの解析等の定量的処理には適していても、「談話の流れ」や「沈黙」などの情報が付与されていないものが多く、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。

しかし、特に、自然会話データのコミュニケーション教育への応用も企図する場



合、音声・動画に加え、言語使用の分析を主とする「語用論的分析に適した情報」を各研究者が研究目的に応じて付与することができ、且つ、コーディングしたデータを定量的に分析することができるような「文字化資料」に基づく「自然会話コーパス」が必須である。というのも、「言語教育への応用」という目的を主とする場合は、会話分析と言っても、少数の会話データに基づく定量的分析による特徴の記述だけでは不十分で、ある程度、一般化ができることが必要である。また、そのためには、コーパス等を活用して一定の量のデータを扱うことが必要である。

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、定性的・定量的双方から分析するのに適するように「文字化資料作成支援ツール」を開発し、その原則に基づく「文字化資料」と、音声・動画も含む「自然会話コーパス」を「データベース化」したのが、『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』と、NCRB (Natural Conversation Resource Bank) である。NCRB で検索し、ダウンロードしたデータは、既に開発済みの「文字化入力支援・自動集計、複数会話自動集計システム」を使ってコーディングや自動集計ができる。さらに、現在は、NCRB の中に、「自然会話教材作成支援ツール」を組み込んで、「自然会話を素材とする教材の作成」を支援し、一定の基準を満たす登録者で共有するシステムを開発中である。本発表では、これらの開発の流れや目的、その使い方を紹介しながら、語用論的研究のために必要な「コーパス」の特性と、ひいては、これからの「コーパス構築」のあり方、コーパスの研究・教育への活用法などについて論じる。

## 2. 本論の主張

本発表の主な主張は、以下の 7 点である。

(1) 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「会話の音声・動画」とともに、語用論的分析に必要な、同時発話、割り込み、沈黙等の情報が付与された「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。条件をある程度統制して比較した場合、コーパスの規模が小さくても、語用論的機能のみならず、使用語彙の頻度や割合も、大規模コーパスの傾向と大差ないことも報告されている (宇佐美・中俣、2013))。

(2) 語用論的分析に適したトランスクリプトの作成には、まず、その目的に応じた「文字化のルール」が必要である。(⇒「基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ)」(宇佐美、1997～2011))

(3) 語用論的分析のためには、それに適した文字化のルールとともに、そのルールに従った文字化にかかる時間を節約するツールが必要である。(⇒「BTSJ 入力支援・自動集計システムセット」)

(4) 語用論的分析のためには、基本的な情報としてトランスクリプトに付与さ

れている情報に加えて、各研究者が、自身の研究目的に応じてコーディング(形式や機能の分類等)を行うことが必須である。語用論的分析のためのコーディングは、品詞などの言語形式のみにかかわる分類とは異なり、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、入力していく必要がある。ただし、そのコーティングや集計には、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。(⇒『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』)

(5) このように、語用論的分析のためには、会話データの収集や、トランスクリプトの作成、コーティング、及び、その集計等、現状ではいずれもテクノロジーの力を借りてもすべてを自動化することは不可能なため、未だ膨大な時間と労力がかかる。そのため、この分野の研究の発展のためには、各研究者が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。(⇒『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』等)

(6) 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、今後は、もっと「双方向的な参加・共有型コーパス構築」のあり方を考える必要がある。(⇒NCRB (Natural Conversation Resource Bank))

(7) 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、2007ab、2012)が、「自然会話を素材とする教材」の作成には、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ツール」のようなものを活用して教材作成の時間と労力を削減することが必要である。(⇒NCRB の中に、「自然会話教材作成支援ツール」を組み込む。)

発表では、NCRB や開発したシステムなどのデモも交えて論じる予定である。

付記:本研究の一部は、下記の科学研究費補助金によるものである。記して感謝したい。
「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～26 年度 科学研究費補助金基盤研究 (A)-(課題番号 23242027):研究代表者(宇佐美まゆみ)

**【引用文献】**

＜邦文＞



宇佐美まゆみ(2007a )「自然会話の教材化とディスコース・ポライトネス理論 1:対人コミュニケーション論としてのディスコース・ポライトネス理論の考え方」『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. 12-25.
http://www.tufs.ac.jp/ts/personal/usamiken/gyouseki_pdf/ronbun/2007a.pdf

宇佐美まゆみ(2007b)「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. 26-38.
http://www.tufs.ac.jp/ts/personal/usamiken/gyouseki_pdf/ronbun/2007b.pdf

---------- (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:150-181. 32 頁. 2008 年 9 月.

---------- (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:63-82.

宇佐美・中俣(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:217-228.
http://www.tufs.ac.jp/ts/personal/usamiken/gyouseki_pdf/ronbun/2013a.pdf

＜英文＞

Thomas Schmidt and Kai Wörner (2009) EXMARaLDA- Creating, analyzing and sharing spoken language corpora for pragmatic research, *Pragmatics (International Pragmatics Association)* 19:4. 565-582.
http://elanguage.net/journals/pragmatics/article/viewFile/2558/2519

**【参考資料】**

NCRB (Natural Conversation Resource Bank) 宇佐美まゆみ研究室(2014)
http://www.ncrb.info/ (2014 年 4 月より一部公開予定)

「基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ)」(宇佐美まゆみ、1997～2011)
http://www.tufs.ac.jp/ts/personal/usamiken/btsj.htm

『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』宇佐美まゆみ監修(2011) (申込み者に無料配布)
http://www.tufs.ac.jp/ts/personal/usamiken/btsj_corpus_explanation.htm (改訂・大幅増補版が『BTSJ による日本語会話コーパス(トランスクリプト・音声)2014 年版』と改名して、2014 年 4 月に公開予定(申込み者に無料配布)