JLPTUFS 作文コーパスの構築と今後の展望

日本語教育アーカイブ化構想への足がかりとして

鈴木智美（東京外国語大学留学生日本語教育センター）
tmsuzuki@tufs.ac.jp

１．はじめに

本発表では、東京外国語大学留学生日本語教育センター（以下「センター」とする）にて作成された「JLPTUFS 作文コーパス」について、まずその概要を報告し、その教育研究への活用事例について紹介したい。

そして、同様の作文データベースを構築していくにあたって今後考えられる視点、また今後日本語教育研究に資するデータベースをさらに構築・発展させていくことについて、「日本語教育アーカイブ化構想」というキーワードのもとに考えてみたい。

２．「JLPTUFS 作文コーパス」

2.1　作成の目的とコーパスの概要

「JLPTUFS 作文コーパス」は、広く日本語教育研究に資することを目的として、東京外国語大学「全学日本語プログラム」（JLPTUFS）[(1)]で書かれた日本語学習者の作文（期間：2009 年～2010 年）のうち、執筆者によるデータ提供の同意を得られた作文を、基本的な執筆情報とともに電子データ化したものである。

データベースは、シンプルで使いやすいことを第一に考え、情報一覧ファイル（EXCEL 形式）から、当該作文のテキストファイルおよび PDF ファイルへとリンク付けを行い、全データを 1 枚の CD に収録して配布することとした。テキストファイルは、学習者が執筆した通りに入力し、誤用の修正や誤用タグの付与は行っていない。研究目的に応じて加工可能なように、基礎的な材料をいわば「素」の形でデータとして収めたものである。

最終的に、様々な国籍・母語背景を持つ世界 55 の国・地域からの学習者の作文が、初級～超級レベルにかけて約 1,500 件集まった（内訳については 2.2 参照）。2011 年 12 月現在、関係先に約 130 部の配布が行われている。

2.2　収録作文数

「JLPTUFS 作文コーパス」に収録されている作文の内訳は以下の通りである。

表 1　「JLPTUFS 作文コーパス」収録作文データ数（レベル別）[(2)]

レベル	2009 年度春学期	2009 年度秋学期	2010 年度春学期	レベル別計
100（初級）	22	37	36	95
200（初中級）	35	40	13	88
300（中級 1）	157	119	114	390
400（中級 2）	67	136	67	270
500（中上級）	96	101	138	335
600（上級 1）	3	16	32	51
700（上級 2）	46	110	27	183
800（超級）	72	15	16	103
小計	498	574	443	1,515
計				

表 2　「JLPTUFS 作文コーパス」収録作文データ数（国・地域別）

国	データ数	国	データ数
中国	223	スロベニア	14
イタリア	107	フィリピン	14
アメリカ合衆国	89	スウェーデン	13
イギリス	86	ペルー	13
ドイツ	78	マレーシア	13
台湾	59	ハンガリー	12
スペイン	58	インド	10
韓国	56	コロンビア	10
カンボジア	49	チェコ	10
インドネシア	44	ブルガリア	9
タイ	44	メキシコ	8
ベトナム	36	モロッコ	8
トルコ	35	シリア	7
モンゴル	31	ミャンマー	7
ブラジル	30	イラン	6
フランス	30	ウクライナ	6
スイス	29	グアテマラ	6
シンガポール	27	エストニア	5
ウズベキスタン	26	クロアチア	5
ラオス	26	ニカラグア	5
中国（香港）	26	アイルランド	3
ロシア	22	アゼルバイジャン	3
カナダ	21	オランダ	2
チリ	21	スロバキア	2
オーストリア	20	ヨルダン	2
ポーランド	17	レバノン	2
オーストラリア	15	ニュージーランド	1
エジプト	14	計	1,515

３．教育研究への活用事例

「JLPTUFS 作文コーパス」を試用した研究事例として、東京外国語大学留学生日本語教育センター 鈴木・中村（編）（2011）には、研究ノート 4 編を収録している。学習者の作文をレベル別、国・地域別、テーマ別などの切り口から分け、文法・語彙、表現やコロケーション、文字表記に関する分析が試みられている。

鈴木（2010）では、辞書使用可という条件のもとに書かれた初中級レベルの作文に注目し、そこに観察される不自然な表現について見た。語の多義的意味の理解や、類義語の選択、適切な連語の形成などに関わる諸問題とともに、辞書の使用に関係すると思われる点として、不自然な漢語表現や句単位の表現の使用が観察され、母語からのいわば“直訳”の発想が壁となっていると思われる点が浮かび上がってきた。この考察は、鈴木（印刷中）の研究へと発展的につながっている。

また、髙野（印刷中）では、文体の観点から接続詞「だから」に着目し、その使用状況をレベル別に詳細に見ている。また、本学外国語学部日本課程１年次主専攻「日本語教育学入門」（講師：髙野愛子, 2011 年度）の授業では、実際に学習者の作文を添削するという作業を通じて、日本語の表記・文法・文体について考え、さらに作文の内容に基づき学習者の異文化理解や文化摩擦について考えるという目的のもと、このコーパスが活用されている。

４．今後の展望

4.1　学習者日本語の総合的データベース

このような作文コーパスを、今後より「作文」そのもののデータ的価値という側面に着目して発展させていくならば、例えば作文条件の統一（各レベルで同一テーマの作文課題を課すなど）などの工夫を加えていくことなどが考えられる。学習者日本語の総合データベースということを考えるならば、文章表現のみならず、口頭表現における音声データなども当然対象となるだろう。

4.2　日本語教育アーカイブ化

一方、作文コーパスを、広く現場の日本語教育活動のありようをデータベース化する構想の一環としてとらえるならば、例えば文章表現等のクラスを担当した教員が、どのように当該データの作文を添削指導したかという点なども、コーパスを利用する日本語教師や日本語教育を学ぶ者から見ると興味を引かれる点である。「JLPTUFS 作文コーパス」は、実際に「全学日本語プログラム」で行われている日々の教育活動のありようを、学習者の作文という点からデータとして蓄積し、残していくことを主眼として考えたものである。

このような方向で教育研究に資するデータベースの構築を進めていくならば、各レベル・各技能の日本語教育の授業の実際を「日本語教授法アーカイブ」としてデータベース化していくことなども、その資料的価値だけでなく、日本語教師養成の観点などから見ても検討課題の一つとなるのではないだろうか。「全学日本語プログラム」など、各レベル・技能のクラスが充実した日本語プログラムにおいて、そのようなデータベース構築がモデルケースとして進めていくことができればと思う。

注

(1) 東京外国語大学「全学日本語プログラム」（JLPTUFS: Japanese Language Program, Tokyo University of Foreign Studies）は、交流協定校からの交換留学生、日本語・日本文化研修留学生、教員研修留学生、国費・私費の研究生、予備教育課程の国費研究留学生など、非正規の留学生を対象として 2004 年春学期より全学的に開かれることとなった日本語プログラムである。

(2)「初級」等の名称は、2011 年度秋学期に改訂されたもので、作文を収集した時期（2009～2010 年）とは若干の異なりがある。初級（100）レベルは１学期間（15 週間）で初級を一通り学習し終える集中コース、初中級（200）は初級後半から中級前期までをカバーする集中コースである。中級１（300）は中級前期、中級２（400）は中級の中程、中上級（500）は中級後半で、500 レベル終了時に日本語能力試験 N2 合格が目安となる。上級 1（600）を経て、上級 2（700）レベルでは日本語能力試験Ｎ１合格が目安となり、超級（800）は既に日本語能力試験 N1 に合格し、大学の学部・大学院などの授業も十分に受講可能なレベルである。

参考文献

鈴木智美（2010）「辞書の使用が引き起こす学習者の不自然な表現―『JLPTUFS 作文コーパス』の作文から見えてくること―」『2010 世界日語教育大会論文集』（DVD 版）1436-0-1436-9

鈴木智美（印刷中）「留学生の辞書使用についての実態調査―東京外国語大学で学ぶ留学生へのアンケート調査の結果と分析―」『東京外国語大学留学生日本語教育センター論集』第 38 号

髙野愛子（印刷中）「接続詞『だから』をめぐる文体差」『東京外国語大学留学生日本語教育センター論集』第 38 号

東京外国語大学留学生日本語教育センター 鈴木・中村（編）（2011）『「JLPTUFS 作文コーパス」の構築』東京外国語大学留学生日本語教育センター教育研究開発プロジェクト「JLPTUFS 作文コーパス」報告書（別添 CD 付き）