科学研究費補助金 基盤研究(C)

(課題番号 17K02926)

日本語母語学習者データに基づく ロシア語学習者コーパス構築の 基盤研究

2017 年度研究成果報告

2018 年 3 月 31 日

学習者言語を探る

― 科研費プロジェクト「日本語母語学習者データに基づくロシア語学習者コーパス構築の基盤研究」2017 年度研究成果報告総括 ―

林田　理恵
（大阪大学大学院言語文化研究科）

I. 研究の概要

本研究は，国内外で初となる日本語を母語とする学習者データに基づくロシア語学習者コーパス構築に向け，理論的・技術的な基盤整備を行うことを目的としている．作業はロシア・コーパス開発研究チームとの協同研究体制を前提に，以下の 3 点の最終目標に向け，最初の 1 年の研究調査作業を終了したところである．

1）信頼性の高いアノテーション済み学習者コーパスを作成するための，付加する情報タグの有効な分類・構成のあり方を検討，新たな付加情報タグ・ガイドラインを提案.
2）収集されている作文試験結果の電子データに基づき，アノテーション，すなわち情報タグの付加を試行，公開されている品詞タグ付けシステムを利用し，日本語母語学習者データに基づくロシア語学習者コーパスのパイロット版を作成.
3）パイロット版により試験的にコーパス分析を行い，試作したコーパスによってロシア語学習者の言語使用特徴がどの程度，明示化されるのか，提案した付加情報タグ・ガイドラインの有効性も含めて検証.

II. 研究の背景

学習者コーパスは，非誘導型・大規模データという特性によって，学習者言語における特定言語素材の過剰・過少使用や母語転移，回避方略などの客観的特徴を明らかにし得る有効な手段として注目されており，国内でも英語・日本語の学習者コーパスを中心に近年，整備が進んでいる（石川 2012:215-249，望月 2012:112-115）.

ロシア語学習者コーパスは，現時点ではモスクワの National Research University, Higher School of Economics (= 以下，HSE) 人文学部言語学コースのスタッフが中心となって 2013 年に公開した Русский учебный корпус (Russian Learner Corpus = RLC, http://web-corpora.net/RLC) が唯一のものとなっている.

筆者の所属する機関では，2000 年より 15 年間に亘って収集した学習者のロシア語作文データが存在し，それを利用したロシア語学習者コーパス（=Japanese RFL Learner Corpus 以下，JRFLL Corpus）構築が可能なことに着目，今回，科研プロジェクトとして上記の研究調査作業に取り掛かることとした．日本語母語学習者のデータを対象としたロシア語コーパス

が構築できれば，学習者のロシア語使用における多面的な言語特性が明らかになり，その分析に基づいて，日本での現行ロシア語教育における教材や指導案，カリキュラムについて抜本的な見直し，改良を図っていく可能性が広がる点が期待される．

III. 研究の特色・独創性

本科研プロジェクトの特色としてまず第 1 に挙げることができるのは，教育現場への実証的データ提供に道を開くことができるという点である．国内のロシア語教育現場では，これまでは主として経験値に依拠したカリキュラムや教材開発等がなされているが，本科研プロジェクトによるロシア語学習者コーパス開発に向けた基盤整備が進めば，

1）ロシア語学習における日本語を母語とする学習者の学習者言語の傾向や特性に関する，より客観的･実証的データに基づいた分析
2）教育現場における教材開発や指導案作成等に対する有効かつ信頼性に足る情報の還元

といったことが展望できるようになる．

第 2 には，収録を予定しているデータそのものの優位性である．コーパスで収録対象とするデータは，筆者所属機関が実施している TORFL 試験（The Test of Russian as a Foreign Language = ロシア教育科学省主催「外国人のためのロシア語検定試験」）のうち，1, 2 年次における A2 (Basic level), B1 (Intermediate level) レベルの作文試験結果，約 1400 人分，約 2800 編の作文であるが，次のような特色を持つ．

3）データが大規模であり，作文のトピックや作成年，習熟度等，テキストや学習者属性情報の内容が明確であるという特徴を持つ．そのことで，学習者言語の特性についてより実証的で正確な観察，記述，分析を可能とするコーパス構築が志向できる．
4）習得レベル別の定量比較分析や同一学習者の継時的変化を観察できる縦断的コーパスとしての活用も視野に入れることができる．
5）収録予定の作文試験結果はすべて，教員による添削済みデータであり，学習者コーパスの成否を左右する誤用情報タグの付加についても，その基礎となる情報がすでに準備された状態にある．加えて，添削は TORFL で定められた客観的評価項目に基づいて実施されたものであり，それらは，誤用情報タグの分類・構成を考察する際にも有効な情報として活用できる．

IV. 2017 年度 研究の流れ

2017 年度は，まず，他言語における既存学習者コーパスに関する情報収集と先行研究の整理を行い，その上で次の 3 点について研究調査を進めた．

1）HSE コーパス開発部門（ロシア連邦・モスクワ）における研究調査

現存する唯一のロシア語学習者コーパス RLC を構築した HSE コーパス開発部門（ロシア連邦・モスクワ）に滞在し，研究協力者である部門長 E.V. Raxilina 氏をはじめ，開発に関わった研究チームから開発の経緯，具体的な技術的問題，現行の稼働状況等，詳細についてヒ

アリングを行った．また，日本語母語学習者のデータ利用に基づくロシア語学習者コーパス･パイロット版構築に向けた協力体制，作業分担等の打ち合わせを行い，RLC システム上に下位コーパスとして JRFLL Corpus を開設することになった．HSE コーパス開発部門での打ち合わせは研究協力者である佐山豪太氏，恒任翔吾氏が担当，詳細については佐山氏，恒任氏の報告を参照されたい．

2）データの電子化作業の開始

収集されている TORFL 作文試験結果について，データの電子化作業を HSE 側の助力を得て開始した．現在は一部がすでに RLC 本体のデータとしてアップロードされている．詳細は佐山氏の報告を参照のこと．

3）HSE コーパス開発部門でのアノテーション作業に関する研修に参加

この点の詳細についても研修に参加した恒任氏の報告を参照されたい．

V. 今後の研究計画・課題

本科研プロジェクトは最終年度の目標として 1) コーパス・パイロット版の作成，2) コーパス・パイロット版による試験分析，3) コーパス・パイロット版のオンライン公開を予定しているが，それに向け，次年度の課題としてアノテーション・ガイドライン試案の設計がある．これについては今年度行った HSE でのヒアリング及び研修で得た知見を基に，データテキスト及び学習者の属性情報タグ，品詞情報タグ，誤用情報タグの分類・構成について試案設計を行わなければならない．

品詞情報の分類は，ロシア語関係のコーパスにおいて信頼度が高いものとして広く使用されている Russian National Corpus (http://www.ruscorpora.ru/index. html) における品詞タグ分類(1) を採用することを予定している．

一方，誤用情報タグの付加は一連の作業の中でもとりわけ重要な作業となる．学習者の言語使用特徴の言語学的分析，ひいてはカリキュラム・指導案開発や教材作成などに有効なデータを提供し得るかどうかが，アノテーション済みコーパスとしての信頼性に直結する．誤用情報タグの分類については，ロシア側の研究チームとの意見交換を踏まえ，RLC での基準を参考に，a) 日本語を母語とする学習者における特徴的な言語使用特性や b) 誤りの頻度が多く観察される言語領域 等を考慮した独自の分類・構成を検討し，試案を設計することが求められる．

さらに，現行 RLC はかなり緻密な誤用タグ分類，構成となっているが，あくまで個々の誤用項目についてのタグであり，1 度の検索では複合的な誤用情報が引き出せない．また，そもそも誤用情報のみに絞った検索ができず，そのままでは誤りの頻度データを得ることもできない．JRFLL Corpus ではこれらの点に特に留意した設定を考慮することが必要となっている．

(1) この分類はロシアで開発された MyStem (https://tech.yandex.ru/mystem/) という品詞タグ付けシステムを使ったものである．このシステムについては恒任氏の報告に詳しい．

上記のカテゴリ分類に加え，誤り情報のタグ付けで重要な役割を担うのが，誤りの程度を示す情報の付加である．TORFL の評価基準にある「コミュニケーションに障害をもたらすかどうか」という項目は，学習者コーパス作成に際しても重要なタグ情報として採用すべきである．これらの情報は今回のデータにはすでに教員の添削内容に情報として盛り込まれており，タグ情報として利用が可能な状態にある．

誤りの程度を示す情報の付加という点は，学習者言語をどのような視点から評価していくのかという，今回のコーパス開発の基本理念にも関わってくる内容である．すなわち「母語話者」基準から逸脱するものをすべて一律の誤りとして括るのではなく，「コミュニケーションに障害をもたらすかどうか」というレベル設定を付加することで，学習者の国際語としてのロシア語能力を分析データとすることが可能になる．そこには，語学教師にいまだ根強い意識としてある「ネイティブ信仰」(鎌田 2005: 322) からの脱却，「母語話者レベル」(2)を 100% として，そこから学習者の能力を測定していくのではなく，あくまで複数言語・複数文化が共存する社会で生きていく力として，学習者が何をどこまでできるのかを見ていこうという姿勢がある．

こういった点を考慮した独自のアノテーション・ガイドラインを作成できるかどうか，次年度の課題は本科研プロジェクトの成否を決める内容になるとも言えよう．

（はやしだ　りえ）

参考文献

石川 慎一郎 (2012)『ベーシックコーパス言語学』，ひつじ書房.

鎌田 修 (2005)「OPI の意義と異議－接触場面研究の必要性－」，『言語教育の新展開』，ひつじ書房，311-332.

駒井 裕子 (2003) 「日本語母語話者の OPI－より明確な超級判定のために－」 ソウル OPI 国際シンポジウム（於韓国ソウル建国大学校）ハンドアウト.

望月 通子 (2012) 「日本語教育における学習者コーパスの構築と ICLEAJ」，『関西大学外国語学部紀要』7 号，111-119.

参照サイト

https://ling.hse.ru/ (最終閲覧日：2018 年 3 月 25 日)

http://web-corpora.net/RLC (最終閲覧日：2018 年 3 月 25 日)

http://web-corpora.net/RussianLearnerCorpus/search/ (最終閲覧日：2018 年 3 月 25 日)

https://tech.yandex.ru/mystem/ (最終閲覧日：2018 年 3 月 25 日)

http://www.ruscorpora.ru/search-main.html (最終閲覧日：2018 年 3 月 25 日)

(2) そもそも「母語話者」と言っても，それぞれのネイティブによって言語能力には，当然，差があり，「母語話者レベル」がどのようなレベルを指すのかという点にも疑問が残る。駒井 (2003) で，日本語母語話者 11 名に OPI (Oral Proficiency Interview) を行った結果，ACTFL ガイドライン（アメリカ外国語教育協会会話能力評価基準）で Advanced-Mid/-High 程度の能力（4 段階中 3 の中―上レベル）であったという報告がなされている。