

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

2018 年度研究成果報告

2019 年 3 月 31 日

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」 2018年度研究成果報告 — 学習者言語における従属複文分析 —

林田 理恵

(大阪大学大学院言語文化研究科)

I. 2018年度 研究の流れ

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、2年目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。

今年度は National Research University, Higher School of Economics (以下, HSE) コーパス開発部門 (ロシア連邦・モスクワ) が構築した RLC (Russian Learner Corpus, <http://web-corpora.net/RLC>) システム上に下位コーパスとして Japanese RFL Learner Corpus (以下, JRFLLC Corpus) 開設の準備を進め、以下のような JRFLLC Corpus 概要画面を設定した。

JRFLLC
RLC Subcorpus

概要

日本語を母語とするロシア語学習者コーパス — JRFLLC は、言語学・ロシア語教育を専門とする研究者、林田理恵（大阪大学）と E. ラリーヒナ（National Research University, Higher School of Economics）の共同研究によって構築されたものである。

また、本研究は日本学術振興会科学研究費助成事業による助成金（学術研究助成基金助成金・基盤研究 C—課題番号 17K02926）を受け実施されているものである。

JRFLLC は日本語を母語とするロシア語学習者のデータに基づく学習者コーパスである。本コーパスにより日本語母語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と期待する。と同時に、RLC のサブコーパスとして、学習者の各母語の違いによって、学習者言語の特性や言語ストラテジー特性にどのような異同点が観察されるか、実証的で正確な記述と分析を可能とし、類型学的研究にも将来的に資するものになると考える。

JRFLLC とは？

JEFLLC Corpus は日本の大学で、ロシア語を専攻する大学生の作文データに基づいて作成されている。これらの作文データは、15 年間に亘って収集された CEFR 基準による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、A2 (Basic level), B1 (Intermediate level) の 2 レベルの作文試験結果データで、現在、約 1400 人分、約 2800 編の作文がデータとして使用可能な状態にある。目下は、その一部に品詞情報、誤用情報タグ等が付加され、すでに公開されているが、作業は続行中で、できるだけ早い段階で全データについて、情報タグが付加された状態で公開する予定である。また、データ数そのものも、今後毎年の試験実施に伴い、増やしていくことが可能である。

メタデータについて

JRFLLC は当初、該当する教育機関の必要性によって構築が始まったが、コーパスが持つ設計基準、豊富なパラメータ上の優位性によって、今後、様々な研究・分析上の課題を解決する可能性をもたらすものと考えられる。コーパスに収録されている各テキストは以下のメタデータが付加されている。：

1. 学生名（仮名）
2. 性別
3. 学習者の母語及び学習歴を有する言語
4. 学習者の学習言語習得レベル
5. 作文作成時点での学習歴
6. 作文作成に際しての制限時間の有無及び制限時間数
7. 作文ジャンル
8. 作文のスタイル（叙述、論証等）

これらのカテゴリーは各作文の Header Identification Box (Header ID) に反映されている。このカテゴリーを使って自動的にさらなるサブコーパスを作成することも可能である。

また、JRFLLC 内の全ての語について品詞・文法性・格・アスペクトの情報をもつ形態タグが付与されている。さらに、作文内の誤りについては、誤用情報タグが付加されている。これらの情報はコーパスインターフェースにより、文法的・語彙的カテゴリー別による検索が可能である

JRFLLC の使用について

JRFLLC は web 検索ツールを介してフリーで公開する資料である。JRFLLC Corpus を用いたあらゆる研究成果物には、本コーパスを利用した旨、必ず記載すること。

— —

JRFLLC Corpus の構築が進めば、ロシア語初級レベル (A1-A2) における習得困難項目である名辞類格変化や動詞人称変化などは、本コーパスを使用して瞬時にその誤用の傾向、また学習者がその時点で作り上げている学習者言語体系の実像が明らかになると考えられる。また JRFLLC Corpus 開発プロジェクトでは、比較コーパスとして、収録作文の作成者と同世代のロシア国内大学生のデータを扱っており、トピックも類似している CoRST を利用することを検討している。ロシア語母語話者コーパスと比較することで、日本語母語学習者における語彙の過剰・過小使用など、母語話者と非母語話者の言語使用特性の差がさらに明らかになると期待している。

今回は収集された ТРКИ 作文タスクのデータを使ったコーパスの企画であるが、同じ ТРКИ の口頭能力試験における発話データについても、現在、15 年間分が収集されており、モノログ、会話、テキスト要約などの音声データが利用可能である。将来的に作文データでの運用が軌道に乗った時点で、JRFLLC Corpus に音声データ及び文字化データを追加していくことが考えられる。作文データと並行して発話データを収録することで、「書く」活動、「話す」活動それぞれにおいて学習者の習得プロセスにどのような差が観察されるかなど、教育面のみならず言語学的にも興味深い実証研究を行うことができるようになるであろう。

JRFLLC RLC Subcorpus

Японский учебный корпус русского как иностранного, или JRFLLC, является результатом коллаборации исследователей в области лингвистики и преподавания – Хаясида Риэ (Осацкий университет) и Рахилиной Екатерины (НИУ ВШЭ).

Данный проект реализован при финансовой поддержке Японского общества содействия науке (JSPS) в форме гранта на научные исследования (№ проекта – 17K02926).

Что такое JRFLLC?

JRFLLC представляет из себя учебный корпус, содержащий образцы речи учащихся, изучающих русский язык и чьим родным языком при этом является японский. Данный корпус позволяет увидеть разнообразные языковые особенности порождения русской речи японскими учащимися. Мы также надеемся, что анализ данных особенностей поможет расширить возможности на пути к пересмотру и совершенствованию содержания обучающих материалов, методических пособий, а также учебных программ, используемых в реальной преподавательской практике в Японии. Кроме того, поскольку JRFLLC является подкорпусом Русского учебного корпуса (RLC), благодаря чему позволяет проводить точный и практически подтверждаемый анализ того, какие схожие черты и отличия наблюдаются в особенностях порождаемой русской речи и особенностях языковой стратегии учащихся, говорящих на разных языках, мы также смеем полагать, что в будущем корпус окажется полезным в том числе и с точки зрения типологических исследований.

Корпус JEFLL составлен из данных, полученных из сочинений студентов специальности «Русский язык» японских университетов. Эти данные были получены по результатам компонента «Письмо» теста по русскому языку как иностранному (The Test of Russian as a Foreign Language, TORFL) для двух уровней – A2 (Basic level), B1 (Intermediate level), который проводится Министерством образования и науки РФ на протяжении 15 лет и основывается на стандартах CEFR (Common European Framework of Reference for Languages : Learning, teaching, assessment). На сегодняшний день доступно около 2800 сочинений от порядка 1400 человек. Пока метаинформацией (часть речи, неправильное употребление и т.д.) снабжена только часть их тех сочинений, которые представлены в открытом доступе, однако работа над корпусом продолжается, и уже в самом ближайшем будущем мы планируем снабдить метаинформацией абсолютно все имеющиеся данные. Кроме того, объем самих данных впредь планируется увеличивать с каждым новым экзаменом в году.

Метаданные

Несмотря на то, что JRFLLC изначально был разработан с целью удовлетворять локальные (местные) нужды, богатство критериев, согласно которым он был спроектирован, порождает разнообразные возможности для поиска ответов на исследовательские вопросы, а также для анализа. Каждый текст снабжается следующими метаданными:

1. фамилия, имя студента (псевдоним),
2. пол,
3. период изучения языка ранее, опыт владения им,
4. языковой уровень студента,
5. временная отметка (на каком курсе было написано данное сочинение),
6. ограничение по времени, в течение которого должно было быть написано данное сочинение,
7. тип текста,
8. функция текста (повествование, аргументирование и т.д.).

Все эти категории для каждого текста в корпусе отражены в Header Identification

Box (Header ID). Возможно автоматическое создание подкорпуса на основе этих данных.

Все слова в JRFLLC снабжены морфологическими тэгами, которые содержат информацию о части речи, поле, падеже, виде. В случае, если слово было написано с ошибкой, тэг будет содержать пометку «нестандартная форма» (т.е. неправильная). Интерфейс корпуса допускает возможность поиска по грамматическим и лексическим категориям.

Использование корпуса

Дальнейшее наполнение корпуса JRFLLC позволит искать и мгновенно находить общие тенденции в неправильном и сложном для «носителей» языка склонении имен (существительных и т.д.) и спряжении глаголов для уровней A1-A2, а также поможет нарисовать реальную языковую картину, которую в тот или иной момент рисует в своей голове студент, изучающий русский язык как иностранный. Кроме того, в рамках проекта по разработке корпуса JRFLLC рассматривается также вопрос использования в качестве сравнения корпуса CoRST, где собраны данные сочинений русскоговорящих учащихся-ровесников японских студентов, а темы схожи с теми, что представлены в Японии. Мы надеемся, что сравнение с данными корпуса носителей языка сделает возможным проследить разницу в особенностях использования языка его носителями и «носителями», в частности чрезмерное (и, наоборот, дефицитное) использование японскоговорящими студентами той или иной русскоязычной лексики.

Корпус изначально планировался как хранилище данных по результатам заданий «Письмо» экзамена ТРКИ, однако по факту на протяжении последних 15 лет собиралась и накапливалась также информация по устным высказываниям студентов в рамках ТРКИ (часть «Говорение»), в результате чего на сегодняшний день доступны аудиозаписи с монологами, диалогами, пересказом прочитанного текста и т.д. В ближайшем будущем, когда использование письменных текстов корпуса встанет на рельсы, мы планируем добавить к корпусу JRFLLC также аудиоданные и данные культурного характера. Мы полагаем, что размещение данных устной речи параллельно с данными письменной – поможет проводить более глубокие и подтверждаемые на практике (не только педагогические, но и лингвистические) исследования того, чем отличаются друг от друга процессы изучения студентами русского языка в обеих сферах – письмо и говорение.

Материалы JRFLLC распространяются бесплатно через поисковые системы сети интернет. Если вы использовали корпус JRFLLC в своей научной работе, вы обязаны указать это в готовом докладе, статье и т.д.

また、収録予定である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」）、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 編の作文のうち、かなりの部分のデータ電子化作業が終了し、HSE コーパス開発部門と共同でのアノテーション作業が進行中である。

II. A2-B1 レベル学習者言語における従属複文

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である。

今回、第2学年末に実施した B1 (Intermediate level) レベル受験者の作文試験結果のうち電子化されたデータの一部を使い、このレベルにおいてとりわけ習得目標として重要項目と位置付けられる従属複文の学習者の言語特性を分析した。学習者のレベルは不合格者もいるので、A2-B1 レベルとなる。

対象としたのは B1 レベル作文試験結果 4 年間分、120 名の計 240 作文（各年度 2 課題）である。これら 240 作文中に使用されていた従属複文の内訳は以下のとおりである。

表 1. TORFL B1 レベル作文試験結果における従属複文・使用例

従属複文 タイプ	接続詞・関係詞	使用総 数	誤用数（タイプ別）				複合
			総数	形態	意味	構文	
説明	что	161	18	10	5	3	
	как	24	1	1			
	какой	21	5	2	3		
	зачем	19	1	1			
	да-нет 疑問文	11	8	1			7
	сколько	10	1	1			
	чтобы	7	3				3
	когда	4	0				
	куда	4	1	(1)	(1)		1
	почему	4	0				
	где	4	0				
	каков	1	0				
	条件	если	130	28	21	2+(5)	(5)
если бы		1	1	1			
原因	потому что	116	6		1	5	
	так как	27	0				
	благодаря тому что	1	0				
定語	который	117	15	8	5	2	
	то, что ~	2	0				
	тот, кто ~	2	1	1			
	где	2	0				
	куда	1	0				
結果	поэтому	111	8	2	3	3	
	так что	6	4	2	2		
	в результате чего	3	2	2			

目的	чтобы	61	1			1	
時	когда	45	1			1	
	после того как	3	2	2			
	с тех пор как	2	0				
譲歩	хотя	4	0				
	несмотря на то, что	1	0				
計		905	107	55+(1)	21+(6)	25+(5)	6

上記の表からわかるように、今回観察された従属複文の使用総数は 905 例であったが、まず特徴的なのは誤用例総数が 107 例で、使用総数の 10%強という低い数値であったことである。

他の語彙や格変化等の誤用割合について、今回は時間の関係上、正確なデータは取れていないが、このレベルの学習者では恐らくはるかに誤用数は高くなるものと思われる。これはまず第 1 に、従属複文で表現される文の論理構成が、一部の例外を除いて言語によってあまり異同が大きいとせいかと考えられる。つまり、従属複文の意味関係は語彙などとは異なり、人間に共通のより抽象度の高い論理的意味内容を表し、社会文化的な違いがあまり反映されず言語間での異同が少ないということになる。したがって成人における言語習得の場合、母語ですでに形成された抽象次元での意味的論理力は、ある程度はそのまま学習言語使用の場合にも援用でき、誤用が少なくなると思われる。

ただそうは言っても、言語間で論理の枠組みにズレが観察されるのは当然である。今回のデータで特徴的だったのは、条件節における言語間での論理枠組みの差に起因する誤用である。

Если бы вы хотите ехать дальше, я веду вас в морской берег.

Если бы вы хотите участвовать в прогулке, позвоните к нам.

Если бы я поехал в Россию, я хочу встретиться с тобой. Если хочешь, давай пойдём в японский ресторан.

上記の誤用例では、いずれも現実条件の場合に非現実条件を表す *Если бы* ～形式の接続詞・構文が使用されている。ロシア語での非現実条件節は、すでに起こっている事実に対する仮定を立てる場合（反実仮想の場合）にのみ使用が限定される。

一方、日本語では例えば「たら」を使う構文でも、「（最初はよくわからなかったけれど、）解説を読んだら解けた」「（どうしてできなかったのだ？）解説を読んだら解けた」といように、現実に起こったことなのか反実仮想なのかの区別は文脈次第という場合が多く、そのあたりの論理上の区別が構文の形に反映しにくいことがある。学習現場でも、特に、まだ起こっていない事柄で、起こる可能性があまりないと感じている場合に、日本語母語学習者が誤って *Если бы* ～形式の非現実条件節を使用するケースが多く観察される。

У первых проектов социальных сетей была цель, которая ускорить деловую переписку,

получить возможность быстро информировать ~

Участник сетей пишут *причину*, которая социальные сети в Интернете стали так популярной.

また、上記の例では「目的 (цель)」「原因 (причина)」といった抽象的な事柄を表現する語の場合に、その内容説明をする節のつなぎ方に誤用が見られる。関係詞 *который* を使用した複文は表に見るように総数 117 例が観察されるが、先行詞が抽象名詞の場合に限って意味的誤用が起こっている。

形態面でも、接続詞は比較的複雑ではなく、この点でもその他の文法項目に比べて誤用数が少なくなっていると考えられる。形態面でも誤用数が多かったのは接続詞 *если* であり、そのほぼすべてが *если* または *есль* といった、音韻面において母語で区別のつきにくい文字の取り違いによる誤用である。

次に多かったのが説明の従属文 (いわゆる間接話法文) における誤用であるが、この場合はそのほとんどが *запятая* 脱落である。

以上、収録されている一部データを使って簡単にこのレベルの学習者言語における従属複文の使用とその後用例を簡単に分析した。今回は限られたデータではあるが、それでも興味深い学習者のロシア語使用における言語特性が観察され、こういった分析に基づいて、教材や指導案、カリキュラムについての見直し、改良の可能性が示唆される。

III. 今後の研究計画・課題

本科研プロジェクトは次年度、最終年度を迎えるが、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。目下は、次年度の最大の課題として誤用タグ付け — アノテーション作業の進捗がある。また、上記従属複文データについても、年度ごとの設題と試験結果としての作文における学習者言語の傾向に、一定の関係性があるのではないかという点や、さらには従属複文タイプの頻度情報と学習者の中間言語の関りといった点についても更なる分析が必要であり、これについてもアノテーション作業の進捗に伴って、より大規模データを取り扱い、また精緻な分析が可能となる。

参照サイト

<https://ling.hse.ru/> (最終閲覧日：2019年3月18日)

<http://web-corpora.net/RLC> (最終閲覧日：2019年3月18日)

<http://web-corpora.net/RussianLearnerCorpus/search/> (最終閲覧日：2019年3月18日)

<https://tech.yandex.ru/mystem/> (最終閲覧日：2019年3月18日)

<http://www.ruscorpora.ru/search-main.html> (最終閲覧日：2019年3月18日)

Japanese Russian as Foreign Language Learner Corpus の概要と今後の展望について

佐山 豪太

(上智大学外国語学部)

1. はじめに

本稿では、まず Japanese Russian as Foreign Language Learner Corpus (日本語を母語とするロシア語学習者コーパス / 以下、JRFLLC) の概要について言及する。

コーパスは一般的に母語話者が産出したテキストを収集したものであるが、「言語学習者の言語産出に限定して収集したコーパスもあり」(石川 2008: 202)、これを学習者コーパスと呼ぶ。学習者コーパスである JRFLLC は、言語学・ロシア語教育を専門とする林田理恵 (大阪大学) と E. ラリーヒナ (National Research University, Higher School of Economics) の共同研究によって作成された⁽¹⁾。

JRFLLC は、日本語を母語とするロシア語学習者 (大学生) の作文データに基づいて構成されている。これまで日本人の英語学習者コーパスは数多く作成・整備されてきたが (cf. 赤野他 2014)、高度な研究に耐え得るロシア語学習者コーパスは存在していなかった。そのため、JRFLLC が完成した際に他の研究へもたらず寄与は計り知れない。「学習者コーパスは、外国語教育、中間言語教育、学習者研究のいずれの見地からも大いに活用しよう」(石川 2008: 202)からである。本コーパスが提示するデータを分析することにより、日本人のロシア語学習者に特有の言語特性が明らかになると考えられる。これまでは教員の経験則で行われていた教材や指導案の作成が、JRFLLC が示すデータに基づく、もしくは裏付けられることで客観性を帯びる。また、それは、カリキュラムの検討・改良へと繋がっていくであろう。実際、学習者コーパス研究は、辞書の編纂や教科書の作成を含む外国語教育の分野で大規模に利用されている (cf. マケナリー、ハーディー 2014: 124)。

2. JRFLLC の内部構造

2.1. JRFLLC を構成するテキスト

JRFLLC のテキストは、日本の大学でロシア語を専攻する大学生の作文データに基づいている。これらの作文データは、15 年間にわたって収集された CEFR (Common European

(1) 本研究は日本学術振興会 (JSPS) 科学研究費助成事業による助成金 (学術研究助成基金助成金・基盤研究 C-課題番号 17K02926) を受けている。

Framework of Reference for Languages : Learning, teaching, assessment)基準による「外国人のためのロシア語検定試験」、A2 (Basic level), B1 (Intermediate level) の2レベルの作文試験結果データである⁽²⁾。

作文の分量は、現在、約 1400 人分、約 2800 編であり、これらの作文データは使用可能な状態にある⁽³⁾。

2.2. メタデータについて

コーパスに収録されているテキストは以下のメタデータが付加されている：

1. 学生名 (仮名)
2. 性別
3. 学習者の母語及び学習歴を有する言語
4. 学習者の学習言語習得レベル
5. 作文作成時点での学習歴
6. 作文作成に際しての制限時間の有無及び制限時間数
7. 作文ジャンル
8. 作文のスタイル (叙述, 論証等)

これらのカテゴリー情報は、各作文の **Header Identification Box (Header ID)** に反映されている。このカテゴリーを使って、研究目的に沿ったサブコーパスを作成することも可能である。例えば、学習歴といった項目を検索すると、ある段階で使える表現、使えない表現の調査が可能である (cf. 赤野他 2014: 74)。

2.3. 検索インターフェイス

JRFLLC は、Higher School of Economics の Russian Learner Corpus (以下、RLC⁴) のインターフェイスを流用している。つまり、JRFLLC は、RLC のサブコーパスとしてその内部に存在している (なお、JRFLLC と同様に、他の言語の学習者によるテキストも RLC 内に含まれており、設定すればその言語でのサブコーパスも作成することが可能である)。具体的には、以下の図が示すように、サブコーパス (подкорпус) を日本語 (японский) に設定することで、JRFLLC のテキストを検索できる (つまり、японский) にチェックを入れると、検索対象が JRFLLC のテキストに限定される)。

(2) ロシアでは CEFR への対策の一環としてロシア連邦教育科学省の認定により外国語としてのロシア語検定 (英: Test of Russian as a Foreign Language) が制定された (中澤 2012: 155)。

(3) なお、JRFLLC は web 検索ツールを介してフリーで使用できるため、すでに外部から検索可能な状態にある。ただ、JRFLLC は JRFLLC Corpus を用いたあらゆる研究成果物には本コーパスを利用した旨を記載する必要がある。

(4) RLC に関しては Рахилина (2016) や <http://www.web-corpora.net/RLC/> を参照されたし。

Подкорпус

Период от <input type="text" value="2014"/> до <input type="text" value="2015"/>	Пол автора <input type="radio"/> м <input type="radio"/> ж <input checked="" type="radio"/> любой
Тип текста <input type="radio"/> устный <input type="radio"/> письменный <input checked="" type="radio"/> любой	Русский язык <input type="radio"/> иностранный <input type="radio"/> эритажный <input checked="" type="radio"/> любой
Доминантный язык	
<input type="checkbox"/> Выделить все	<input type="checkbox"/> корейский
<input type="checkbox"/> английский	<input type="checkbox"/> французский
<input type="checkbox"/> итальянский	<input type="checkbox"/> немецкий
<input type="checkbox"/> норвежский	<input type="checkbox"/> сербский
<input checked="" type="checkbox"/> японский	<input type="checkbox"/> китайский
<input type="checkbox"/> казахский	<input type="checkbox"/> нидерландский
<input type="checkbox"/> шведский	<input type="checkbox"/> финский
<input type="checkbox"/> эстонский	<input type="checkbox"/> дагестанский
<input type="checkbox"/> португальский	<input type="checkbox"/> таджикский

図 1. RLC におけるサブコーパスの設定画面

このような仕組みを採用している狙いは、まず一からコーパスの検索インターフェイスを作成する工数の削減にある。また、図が示すように、他の言語のサブコーパスも RLC の内部に存在している。RLC 内に本研究の JRFLLC が含まれていることにより、今後、同じ検索インターフェイス、同じ検索基準で複数の言語を分析対象とした横断的な研究が可能となるであろう。

2.4. 誤用タグ

学習者コーパスのテキストは、誤用情報タグがテキストに付加されている。その情報によって、学習者がどのような間違いをするかを把握できる。本研究のテキストへのタグ付け作業は、現在進行中である。

上述の通り、JRFLLC は RLC の内部にサブコーパスとして存在している。したがって、RLC が採用している誤用タグ (Rakhilina et al. 2016) で検索することが可能である。

図 2. RLC で採用している誤用タグ一覧

誤用タグには綴り字、語彙、構造、形態、統語、その他といった大分類が用意されており、さらに、そこから下位区分が設定されている。

RLC の誤用タグは、これまでの HSE 側の研究の成果として出来上がった一つの形であるため、現段階ではそれに基づいて本研究のテキストにタグ付けをしていく。ただし、今後、本研究で新たに誤用タグを作成するべきかという議論はしていく（つまり、日本人学習者に特有の間違いなどに対応した誤用タグを追加することは可能であり、検討していく）。

3. JRFLLC の活用例

赤野他 (2014: 79-80)が述べているように、学習者コーパスを用いた研究は仮設検証型と探索型の 2 つのタイプに大別される。ここでは、JRFLLC のテキストを実際に使用し、仮説として想定される、日本人学習者に典型的な文法の間違いを検討する。

ロシア語は、完了体動詞と不完了体動詞で未来体の作り方が異なっている。不完了体動詞は、**БЫТЬ** の未来形と動詞の不定形、一方、完了体動詞は、動詞を人称変化させた形が未来時制を表す。学習者はよく両者を混同しており、教員の経験則から、**БЫТЬ** の未来形と完了体動詞の不定形を結びつけてしまう場合が多いと推測される。これを客観的データに基づいて考察できるところに学習者コーパスの真価があると考えられる。では、以下で実際に **БЫТЬ** の 1 人称単数未来形 **буду** と完了体動詞 **поехать** 「出かける」といった語結合をキ

ーワードとして JRFLLC 内を検索する。

1. Письмо о стипендии, ЯСТРКИ15-18 | jap | FL | None
Я буду поехать в Россию зимой, в январе. <...->
Я буду поехать в Россию зимой, в январе.
2. Велосипедная прогулка, ЯСТРКИ12-2 | jap | FL | None
Завтра я буду поехать на велосипедную прогулку. <...->
Завтра я буду поехать на велосипедную прогулку.
3. Письмо в Россию, ЯСТРКИ11-38 | jap | FL | None
Итак я буду поехать в России весной. <...->
Итак я буду поехать в России весной.
4. ? 🗨 Показать / ✖ Скрыть
5. Рассказ о себе, ЯСТРКИ10-3 | jap | FL | None
У меня план, что я буду поехать учиться в России. <...->
У меня план, что я буду поехать учиться в России.
6. Рассказ о себе, ЯСТРКИ2005-60 | jap | FL | None
Поэтому я буду поехать в Россию. <...->
Поэтому я буду поехать в Россию.

図 3. JRFLLC における буду поехать の検索結果

同様に、быть の 1 人称複数形 будем と完了体動詞 поехать といった語結合の検索結果を以下に挙げる。

1. Объявление о поиске компаньона, ЯСТРКИ14-05 | jap | FL | None
В следующий воскресенье мы будем поехать в Москву. <...->
В следующий воскресенье мы будем поехать в Москву.
2. Объявление о поиске компаньона, ЯСТРКИ14-05 | jap | FL | None
Мы будем поехать туда на поезде, потому что это очень удобный. <...->
Мы будем поехать туда на поезде, потому что это очень удобный.
3. Киото, ЯСТРКИ12-4 | jap | FL | None
Потом, мы будем поехать в японское кофе, и отдохнуть там. <...->
Потом, мы будем поехать в японское кофе, и отдохнуть там.

図 4. JRFLLC における будем поехать の検索結果

図 3 と 4 が示す通り、JRFLLC 内には未来形の作り方を間違えた例が確認される。仮にこのような誤用が他の動詞との語結合に散見されるようであれば、教材などで今以上の注意喚起をするとといった改善が可能である。

4. 今後の展望

現段階では、JRFLLC のテキストに対する誤用タグの付与は完了していないため、包括的なデータの検索は実施できない（まず仮説に基づいて буду поехать, будем поехать といった検索ワードを想定し、1 つ 1 つの語形を打ち込んで情報を得るしかない。これは極め

て非効率であるし、必要な情報の見落としも起こり得る)。

実際の研究に使用することを考えた場合、誤用タグ付けが最大の問題である。今後、誤用タグ付けが進めば（例えば、上述の未来系の間違いは tense 「時制」、aspect 「動詞の体」と、**быть** の未来形といった指示があれば）、入手したい情報を包括的に得ることができる。これは次年度の課題としたい。

(さやまごうた)

参考文献

- 赤野一郎, 投野由紀夫, 堀正広 (2014) 『英語教師のためのコーパス活用ガイド』, 東京: 大修館書店.
- 石川慎一郎 (2008) 『英語コーパスと言語教育: データとしてのテキスト』, 東京: 大修館書店.
- マケナリー, T., A., ハーディー (2014) 『概説コーパス言語学: 手法・理論・実践』, 東京: ひつじ書房.
- Rakhilina, E., Vyrenkova, A., Mustakimova, E., Ladygina, A. and Smirnov, I. 2016. “Building a learner corpus for Russian”, *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition at SLTC 2016*.
- Андрюшина, Н.П. (ред.). 2014. *Лексический минимум по русскому языку как иностранному. I сертификационный уровень. Общее владение, 7-е изд*, СПб: Златоуст.
- Рахилина Е. В.2016. “О новых инструментах описания русской грамматики: корпус ошибок”, *Русский язык за рубежом*. № 3. С. 20-25.