

令和 2 年 5 月 30 日現在

機関番号：14401

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K02926

研究課題名(和文) 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

研究課題名(英文) Fundamental research on the building of RFL Learners Corpus based on the data of Japanese RFL learners.

研究代表者

林田 理恵 (Hayashida, Rie)

大阪大学・言語文化研究科(言語社会専攻、日本語・日本文化専攻)・教授

研究者番号：70185651

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：日本語を母語とするロシア語学習者(A2-B1レベル)による2800編の作文データに基づき、ロシア研究機関との協同作業によってJRFLC Corpusパイロット版を作成、オンライン公開を行った(<http://www.web-corpora.net/RLC//jrflc>)。

作成されたJRFLC(Japanese RFL Learner Corpus)パイロット版を使い、A2-B1レベル学習者言語を分析、これまで経験的にしかとらえられていなかったロシア語学習者言語の傾向と特徴を実質的データに基づいて明らかにすることに成功した。

研究成果の学術的意義や社会的意義

本コーパスJRFLC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、改良を図っていくことが可能となった。

同時に、RLC(Russian Learner Corpus)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に道を開き、類型学的研究にも将来的に資するものになると考える。

研究成果の概要(英文)： Japanese RFL Learners Corpus (JRFLC) is a collection of approximately 2800 papers (the production) written by Japanese learners of Russian as a Foreign language (CEFR:A2 to B1 levels). The corpus was built by the members of Grants-in-Aid for Scientific Research(KAKEN) in collaboration with Russian research institutes (the Linguistic Laboratory of Corpus Technologies of the National Research University Higher School of Economics), and now we have published the linguistic results online database (<http://www.web-corpora.net/RLC//jrflc>).

We have analyzed a linguistic performance of A2 to B1 learners, using a actual data, a pilot version of the created JRFLC Corpus, and scientifically proved the tendency and specific variants of Russian learners, a linguistic performance of Japanese students, which was until now only built up as hypothesis on the base of empirical theory.

研究分野：ロシア語学，ロシア語教育論

キーワード：学習者コーパス 学習者言語 ロシア語教育 アノテーション・ガイドライン

様式 C-19、F-19-1、Z-19（共通）

## 1. 研究開始当初の背景

(1) ロシア語学習者コーパスは、これまではモスクワの National Research University, Higher School of Economics（以下、HSE）が 2013 年に公開した Русский учебный корпус (Russian Learner Corpus = RLC, <http://web-corpora.net/RLC>) が唯一のものであり、日本語を母語とする学習者のデータについてはいまだ整備されていない状況であった。

(2) また、学習者コーパスにおいては、特に誤用情報の付加が重要な役割を果たすが、これまでの研究では誤用情報の分類・構成に関する明確な指針は確立されてこなかった。誤用情報には語彙レベル、形態レベル、構文レベル等、多様なタイプが混在し、これらをどのように分類し、タグとして構成していくかにコーパスの信頼度、有効性の成否が左右される。学習者言語の分析、ひいては教育面に資する有益なデータ取得に役立つコーパス作成のためには、その点の改善が急務とされていた。

(3) 研究代表者所属機関では、15 年間に亘って「外国人のためのロシア語検定試験 (The Test of Russian as a Foreign Language = 以下、TORFL)」を実施しており、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 1700 人分、約 3400 編の作文がコーパスの収録対象データとして利用可能な状態にあった。これらの大規模データは、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。このような特徴によって、客観的・実証的観察、記述、分析を可能とするコーパス構築の実現が可能となった。また、収録された作文試験結果は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であった。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ものであった。

## 2. 研究の目的

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 点を最終目標に進められた。

- (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。

## 3. 研究の方法

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に向けた理論的・技術的基盤整備に向け、

- (1) 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
- (2) HSE コーパス開発部門における研究調査
- (3) データ電子化作業
- (4) アノテーション・ガイドライン試案設計
- (5) コーパス・パイロット版の作成
- (6) パイロット版による学習者言語の試験的分析
- (7) パイロット版のオンライン公開と報告会開催  
という手順で作業を進めた。

収集整理した情報・先行研究の知見、HSE コーパス開発部門（ロシア連邦・モスクワ）におけるヒアリング調査で得た開発の経緯、具体的な技術的問題、現行の稼働状況等の情報に基づき、詳細についてを踏まえ、HSE コーパス開発部門でのヒアリング調査を踏まえ、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 Japanese RFL Learner Corpus（以下、JRFL Corpus）を開設する方向性が打ち出された。

上記と並行して、1700 人分、3400 編の手書き作文データの電子化作業を行った。

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、誤用タグ付けの手順についての説明を受け、また実際の作業プロセスに参加し、収集したデータのアノテーションに向け、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・構成についてアノテーション・ガイドライン試案の設計を進めた。

アノテーション・ガイドライン試案に基づき、すでに電子化が完了しているデータについて、テキスト及び学習者の属性情報タグ、品詞情報タグ、誤用情報タグのタグ付け作業を進め、JRFLC Corpus パイロット版を作成、オンライン公開を行った (<http://www.web-corpora.net/RLC/jrflc>)。作成された JRFLC Corpus パイロット版を使い、A2-B1 レベルでとりわけ習得に負荷がかかると思われるアスペクト、ヴォイス、さらに従属複文等における学習者言語の傾向、特徴を分析することで、JRFLC Corpus の有効性証明を試みた。

#### 4. 研究成果

##### (1) データの電子化作業とメタデータ付与

研究代表者が所属する機関で収集されている TORFL 作文試験結果 1700 人分、約 3400 編のデータの電子化作業を行い、併せて、各テキストへの以下のメタデータ付与作業を行った。

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)

##### (2) 新たな付加情報タグ・ガイドラインの提案

信頼性の高いアノテーション済み学習者コーパス作成のため、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・ガイドラインの設計を次のプロセスで行った。

(2)-1 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、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い、同機関におけるアノテーション作業に関する研修に参加、誤用タグ付けの手順などについて説明を受け、実際の作業プロセスにも参加した。

(2)-2 上記を受け、アノテーション・ガイドライン試案の設計に着手した。品詞情報の分類は、ロシア語関

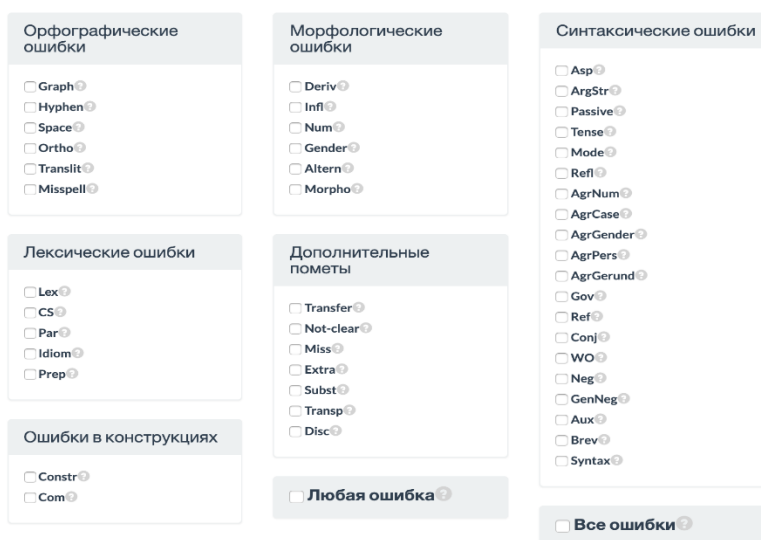


図 1

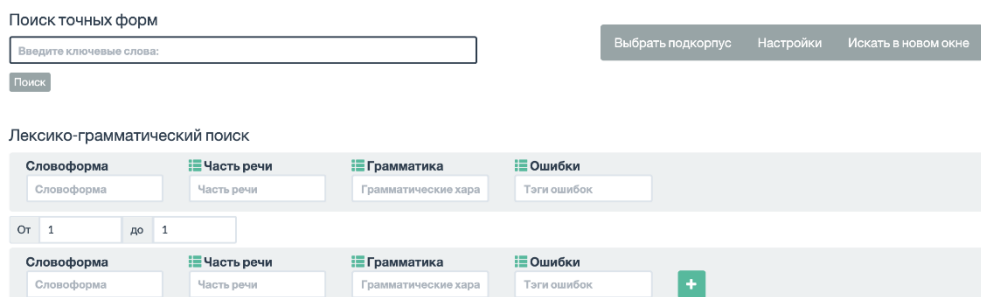


図 2

系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を採用。誤用情報タグの分類については、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、試案を設計した。その上で、RLC が採用している綴り字、語彙、構造、形態、統語、その他という大分類、さらに下位分類からなる誤用タグ (Rakhilina et al. 2016) (図 1) での検索が可能な状態へとアノテ

ション作業を進めた (図 2)。同時に、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について可能性を検討した。

### (3) JRFLC Corpus パイロット版作成ーオンライン公開

アノテーション・ガイドライン試案に基づき、すでに電子化が完了しているデータについて、テキスト及び学習者の属性情報タグ、品詞情報タグ、誤用情報タグのタグ付け作業を進め、JRFLC Corpus パイロット版を作成、オンライン公開を行った (<http://www.web-corpora.net/RLC/jrflc>)。全データのタグ付けについては、上で述べた品詞タグ付けシステム Mystem (<https://tech.yandex.ru/mystem/>) を利用し、ロシア側研究チームが主体となり RLC での基準により誤用情報タグの追加付加作業が続行中である。

### (4) JRFLC Corpus パイロット版による学習者言語における傾向・特徴の試験的分析

作成された JRFLC Corpus パイロット版を使い、A2-B1 レベルでとりわけ習得に負荷がかかると思われるアスペクト、ヴォイス、さらに従属複文の 3 領域において、学習者言語の傾向、特徴を試験的に分析し、パイロット版におけるタグ分類・構成上の問題点や技術面での問題点を抽出した。さらには、教員が経験則から感じている学習者の典型的な誤用について、その客観的確認方法として JRFLC のデータがどの程度活用できるかについて検証した。

(4)-1 JRFLC Corpus パイロット版による試験分析によって、未来表現におけるアスペクト領域の誤用、特に完了体人称変化形に対する誤用の割合が収録データの約 30% に観察されることが分かった。データの計量的分析によって初めて、1) 未来表現のアスペクト領域での誤用出現に極端な偏りがあるという事実が明らかになり、データが示す誤用に関して、単にこの時制でのアスペクト形態・機能の未習得という点にとどまらず、2) 不完了体合成未来形の過剰使用、さらには 3) 学習者の完了体人称変化形に対する回避方略の可能性が見えてきた。

(4)-2 試験的分析結果では、設題テキスト内で使用されていた表現をそのまま書き写したものの以外で、学習者自身が自律的に産出した受動文表現はごくわずかしか観察されなかった。文法性判断テストで 60% 近い誤用率が観察される受動文表現が、B1 レベルテストにおける非統制の作文タスクでデータが出現しなかった理由として、学習者の回避方略の可能性が示唆され、受動文表現を射撃したアカデミック・ライティングの特別な訓練など、学習者の受動文回避克服のための方策の必要性が浮上する。

(4)-3 B1(Intermediate level) レベルでの重要習得項目である従属複文についても、試験的分析により学習者の言語特性を分析、その結果、次のような特徴が観察された。

- 観察された従属複文の使用総数に対し、誤用総数は約 16% という以外に低い数値を示した。
- 問題文での従属複文使用の量的・質的差が、学習者作文における従属複文使用に影響を与えていると思われる傾向が観察された。
- 誤用は言語間で論理の枠組みにズレが観察される場合に特徴的に起こっている。

上記の結果より、まず、人間に共通のより抽象度の高い論理的意味内容を表す従属複文の意味関係は、言語間での異同が少なく、成人における言語習得において従属複文はさほど負荷がかからないという分析結果が出、より大胆に学習の初期段階から複文項目を使用した教材、タスクを導入できるのではという仮説を導き出した。そのことで、学習者の知的レベル、興味関心に見合った内容のテキスト等のインプット、学習者が表現したい内容のアウトプットが可能な学習活動を組める可能性を示唆した。また、b では統語的プライミング効果 (Syntactic Priming; SP) の影響を明らかにすることができ、文法・構文等、ターゲットとなる項目を含んだ事前インプットの意識的導入の有効性を証明した。

(4)-4 動詞現在変化における過剰一般化、名詞複数形特殊変化について未習得率が高くなるといった、これまで教員の経験則からのみ言及されていた項目について、JRFLC で高い頻度で出現しているという客観的データが確認された。

### (5) 研究成果報告会の開催

2019 年 12 月にロシア語教育研究集会において研究成果報告会を開催、英語、日本語コーパス研究において豊富な経験と実績を持つ研究者を招き、「学習者コーパスの歴史：学習者の L2 使用を総体としてとらえるために」と題する講演会を実施した。講演では講師の経験をふまえた学習者コーパスの設計・構築・公開・分析にかかる諸問題について、詳細かつ具体的な事例とその対処法が展開され、フロアとの活発なやり取りが行われた。

### (6) 最終研究報告書の上梓

本研究の総括、残された課題、今後の展望をまとめた「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」研究成果最終報告を上梓した (ロシア語教育支援・就職情報サイト <http://kyoiku-ru.org/> に掲載)。

<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.
- 林田 理恵 2017. 「ロシア語学習者コーパス構築の可能性と意義」『言語文化研究』43号, 大阪大学言語文化研究科.
- 林田 理恵 2018. 「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 —」『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』([http://kyoiku-ru.org/news\\_all/kaken20190327-13/](http://kyoiku-ru.org/news_all/kaken20190327-13/)).
- 林田 理恵 2019. 「ロシア語の学習者言語を探る—A2-B1 レベル学習者の従属複文使用」『ロシア語教育研究』第10号.
- Исаченко, А.В. 1960. Грамматический строй русского языка в сопоставлении со словацким. Морфология., Ч.2, Братислава: Издательство Словацкой Академии наук.
- 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.
- 石川慎一郎 2017. 『ベーシック応用言語学』, 東京: ひつじ書房.
- 小池生夫, 河野守夫(編). 2003. 『応用言語学事典』, 東京: 研究社.
- マケナリー, T., ハーディー, A. 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.
- Рахилина Е. В. 2016. “О новых инструментах описания русской грамматики: корпус ошибок”, Русский язык за рубежом. № 3. С. 20-25.
- Selinker, L. (1972) “Interlanguage. “International Review of Applied Linguistics”, Language Teaching, 10(3), pp.209-231.

<参照サイト>

- <https://ling.hse.ru/> (最終閲覧日：2020年5月20日)
- <http://web-corpora.net/RLC> (最終閲覧日：2020年5月20日)
- <http://web-corpora.net/RussianLearnerCorpus/search/> (最終閲覧日：2020年5月20日)
- <https://tech.yandex.ru/mystem/> (最終閲覧日：2020年5月20日)
- <http://www.ruscorpora.ru/search-main.html> (最終閲覧日：2020年5月20日)

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 8件）

1. 著者名 林田理恵	4. 巻 2
2. 論文標題 学習者言語における従属複文分析 - 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	5. 発行年 2019年
3. 雑誌名 「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	6. 最初と最後の頁 1-8
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 佐山豪太	4. 巻 2
2. 論文標題 Japanese Russian as Foreign Language Learner Corpus の概要と今後の展望について	5. 発行年 2019年
3. 雑誌名 「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	6. 最初と最後の頁 9-14
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 林田理恵	4. 巻 1
2. 論文標題 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	5. 発行年 2018年
3. 雑誌名 「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	6. 最初と最後の頁 1-4
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 恒任翔吾	4. 巻 1
2. 論文標題 RLCの仕組みと誤用タグ付けの実際	5. 発行年 2018年
3. 雑誌名 「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	6. 最初と最後の頁 5-9
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 佐山豪太	4. 巻 1
2. 論文標題 学習者コーパスの研究利用と今後の方針	5. 発行年 2018年
3. 雑誌名 「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	6. 最初と最後の頁 10-12
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 林田理恵	4. 巻 10
2. 論文標題 ロシア語の学習者言語を探る-A2-B1レベル学習者の従属複文使用	5. 発行年 2019年
3. 雑誌名 ロシア語教育研究	6. 最初と最後の頁 15-28
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 林田理恵	4. 巻 3
2. 論文標題 学習者言語分析の可能性と将来的展望	5. 発行年 2020年
3. 雑誌名 「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」研究成果最終報告	6. 最初と最後の頁 1-10
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 佐山豪太	4. 巻 3
2. 論文標題 Japanese Russian as Foreign Language Learner Corpus の外国語教育への活用	5. 発行年 2020年
3. 雑誌名 「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」研究成果最終報告	6. 最初と最後の頁 11-16
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 林田理恵
2. 発表標題 学習者言語の傾向を探る - ロシア語学習者コーパス構築の基盤研究 -
3. 学会等名 日本ロシア語教育研究会東日本地区2017年度研究例会
4. 発表年 2017年

1. 発表者名 林田理恵, 佐山豪太
2. 発表標題 JapaneseRussianForeignLanguageLearnerCorpusの概要と今後の展望について
3. 学会等名 日本ロシア語教育研究会東日本地区2019年度研究例会
4. 発表年 2019年

1. 発表者名 石川慎一郎(林田理恵企画講演)
2. 発表標題 学習者コーパスの歴史:学習者の L2 使用を総体としてとらえるために
3. 学会等名 ロシア語教育研究集会2019 (招待講演)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

Russian Learner Corpus <a href="http://www.web-corpora.net/RLC/">http://www.web-corpora.net/RLC/</a> ロシア語教育支援・就職情報サイト <a href="http://kyoiku-ru.org/">http://kyoiku-ru.org/</a>
--



## 6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	佐山 豪太  (Sayama Gota)  (60824480)	上智大学・外国語学部・講師     (32621)	