

スロベニア人初級・中級・上級日本語学習者コーパスの構築と 縦断的文法誤用分析

パウロヴィチ・ミハ（リュブリャーナ大学文学部アジア研究学科）

The Construction and longitudinal analysis of a learner corpus of Japanese on the elementary, intermediate and advanced level

Miha Pavlovič (University of Ljubljana, Department of Asian Studies)

要旨

初級、中級、上級レベルのスロベニア人日本語学習者コーパスを構築した。その中の文法誤用の分析を行った。各レベルで最も頻繁な文法誤用を判別し、レベル間の結果を比較することにより、さまざまなタイプの誤用の発生傾向の変化を分析した。その結果、237 のテキストと、1778 のマークおよび分類された文法誤用を含むコーパスが作成された。習熟度のレベルごとに、誤用の数が最も多いカテゴリーを決定した。コーパスには縦断的なデータも含まれているため、そのデータも分析し、両方の分析の結果を比較した。最後に、誤用が最も多い文法をより詳細に分析した。分析の結果は、スロベニア人日本語学習者のデータを含む初めての公的コーパスが公開された。

1. はじめに

言語的誤用は言語習得過程の重要な一部である。本研究では日本語学習者の文章に見られる誤用を扱っている。外国語学習者の言語使用で発生する誤用に関するデータは、言語習得過程のより詳しい洞察を得るのに不可欠な情報源である。学習者にとって取得しにくい言語要素についての洞察を可能とし、このデータに基づいて、教師はカリキュラムを調整し、指導を改善することができ、学習者はより効率よく言語を習得できる。また、研究者、教科書や言語教材の作成に関わる者、そして学習者自身にも役立つと言える。

そのため、言語学習者コーパスというツールが情報源として非常に有利である。最も理想的な場合に、これらのコーパスは簡単にアクセスできるように公開されており、そしてデータの作成した学習者に関するさまざまな情報（メタデータ）を備えており、さまざまな言語研究を可能にする。

本研究を行っている現時点では、日本語学習者コーパスは多数存在するが、複数のスロベニア語母語話者から取得したデータが含まれているものはまだない。従って、本研究の主な目標は、スロベニア語母語話者のデータを含む包括的な日本語学習者コーパスを構築することにより、以上述べた問題の解決策を提供することであった。もう一つの目標は、作成したコーパスの実用の試みとしてスロベニア人日本語学習者のデータを分析することであった。具体的には、個々の学習段階で見られる文法誤用を対象とする 3 つの分析のことである。

本論文では、構築されたコーパスを使用して行った 3 つの研究について紹介する。これは、三つの学習段階での最もよく見られる文法誤用の分析と、個々の段階間の比較である。分析の目的は、スロベニア人日本語学習者にとって、言語学習過程のさまざまな段階（初級・中級・上級）にて最も困難な文法要素を特定し、それらの誤用をより詳細に分析することであった。

2. 研究背景

2. 1. データ収集とコーパス構築について

本研究の一環として、三つの学習段階でスロベニア語母語話者のデータを収集した。下記学年の学生より、それぞれデータを提供いただいた結果、合計で42人の学習者から237の作文を収集した。

- ・ 初級 → 137文 → 平均280字

2016/2017年に在籍していたリュブリャナ大学日本研究コースの学部生の1年生

- ・ 中級 → 38文 → 平均500字

2017/2018年に在籍していたリュブリャナ大学日本研究コースの学部生の2年生

- ・ 上級 → 62文 → 平均600字

2020/2021・2021/2022年に在籍していたリュブリャナ大学日本研究コースの大学院生の2年生

集めたデータを上記のように学習能力の段階に応じて分類し、デジタル形式に変換した。データにて見られる誤用をアノテーションして、その分類をした。データは、簡単に使用および共有できるような形式で保存されている。詳細に説明すると最初はマイクロソフトのExcelを使用していたが、より共有しやすくするため、データをGoogle Sheetsに移動させた。また、できるだけ多くの研究を可能にすることを目的として、文章やその筆者である学習者に関する詳細な情報もアンケートで集めて、個々の文章に加えている。本当はちゃんとしたユーザーインターフェースやコンコダンサーがついているコーパスフレームワークを利用したかったが、一人でこのプロジェクトを行っていたため、バジェット（資金/使えるお金）が非常に限られていた。

2.2 仮説

調査を実施する前に、次の仮説が立てられた。

- ・ 最も頻繁な誤用は、最も頻繁に使用される文法要素と一致するだろう。
- ・ 学習者が日本語を長く学ぶほど、誤用は減少するだろう。
- ・ 初級と上級という2つの段階の文章とは異なり、中級段階の文章はテスト場面で作成されたため、誤用が最も多く見られるのは中級段階であろう。
- ・ 上級段階では、語の活用などの形態的誤用が減少し、学習者が学んだばかりの要素でより頻繁に誤用が発生するだろう。
- ・ 縦断調査の結果は、コーパス全体の調査で最も頻繁な誤用カテゴリーの結果と一致するが、学習者間の個人差による偏差がある。

3. 分析

3.1 第1分析

まずは、最も頻繁な文法誤用タイプを、初級、中級、上級のそれぞれの段階において分析した。各段階で誤用が最も多いカテゴリーを分析し、誤用を言語的特徴（品詞等）によって15のカテゴリーに分類した。

今回、上記の分類方法の参考として、市川（1995）で使われている分類化に近い海野（2012）の『オンライン日本語誤用辞典』に使われるものをもとにして、本コーパスのニーズに応じて幾らか修正したものを使用している。

また、Google Sheets を利用して各段階で最も頻繁な誤用の結果に基づいて、最も学生に困難な誤用のカテゴリーを特定した。段階間のより信頼性の高い比較ができるように、誤用の絶対数ではなく、文章の 1000 字にあたる相対頻度という値を使用した。各段階での誤用の頻度を比較することにより、初級と中級、および中級と上級の段階に共通する誤用のタイプが特定された。結果は次の通りであった。

表 1：初級・中級・上級段階で最も誤用の多いカテゴリー

初級段階				中級段階				上級段階			
カテゴリー	誤用数	割合	頻度	カテゴリー	誤用数	割合	頻度	カテゴリー	誤用数	割合	頻度
格助詞	129	24.86%	3.21	格助詞	115	21.78%	5.47	格助詞	170	23.16%	6.66
基本文型	110	21.19%	2.73	2文の接続	107	20.27%	5.09	2文の接続	104	14.17%	4.08
修飾	82	15.80%	2.04	取り立て助詞	81	15.34%	3.85	取り立て助詞	101	13.76%	3.96
テンス・アスペクト	54	10.40%	1.34	テンス・アスペクト	39	7.39%	1.85	修飾	92	12.53%	3.61
2文の接続	50	9.63%	1.24	表現文型・接辞	39	7.39%	1.85	表現文型・接辞	57	7.77%	2.23
取り立て助詞	39	7.51%	0.97	修飾	37	7.01%	1.76	テンス・アスペクト	46	6.27%	1.80
形式名詞	14	2.70%	0.35	基本文型	29	5.49%	1.38	ヴォイス	44	5.99%	1.72
複合辞	12	2.31%	0.30	ヴォイス	24	4.55%	1.14	形式名詞	44	5.99%	1.72
ヴォイス	8	1.54%	0.20	形式名詞	23	4.36%	1.09	基本文型	30	4.09%	1.18
表現文型・接辞	8	1.54%	0.20	指示詞	14	2.65%	0.67	指示詞	28	3.81%	1.10
2語の接続	5	0.96%	0.12	複合辞	11	2.08%	0.52	複合辞	13	1.77%	0.51
指示詞	4	0.77%	0.10	2語の接続	9	1.70%	0.43	2語の接続	4	0.54%	0.16
終助詞	3	0.58%	0.07	終助詞	0	0.00%	0.00	疑問詞	1	0.14%	0.04
疑問詞	1	0.19%	0.02	疑問詞	0	0.00%	0.00	終助詞	0	0.00%	0.00
待遇表現	0	0.00%	0.00	待遇表現	0	0.00%	0.00	待遇表現	0	0.00%	0.00

3.2. 第 2 分析

個々の段階で文章を執筆した学習者が異なっているため、異なる学習者集団が結果に影響を与える可能性もある。より正確な結果を得るには、各段階で同時に収集された同じ学習者の文章が必要になる。同じ学習者から数年連続でデータを取得する必要があるため、このようなデータを取得することは非常に困難である。本研究で構築された SLJ コーパスでは、同じ学習者から 3 つの異なる時点に取得された縦断的データのサブセットも含まれているため、縦断的誤用分析が可能であった。このタイプのデータは 5 人の女子学習者から取得されており、縦断的サブコーパスを構成している。

このデータに基づいて、各段階での個々の学習者の文章の最も頻繁な誤用と、段階間の誤用の変化の傾向を分析した。次に、これらの傾向について、個々の学習者間と以上に述べたコーパス全体の分析の結果を比較した。比較を行った誤用カテゴリーは、最初の調査で最も頻繁で、SLJ コーパスのすべての誤用の 70%以上を占める 5 つの誤用カテゴリーである。そのカテゴリーは：

- ・ 取り立て助詞 (は、も、しか等)
- ・ 格助詞 (を、に、で、が等)
- ・ 基本文型 (肯定文や否定文の構造、コンピュータの使用等)
- ・ 修飾 (連体修飾、連用修飾、連体助詞「の」等)
- ・ 2文の接続 (条件節、理由節など、文の接続に関する文法)

その他のカテゴリーは、誤用の数が少ないため比較を行わなかった。以下は格助詞に関する誤用の縦断的データをまとめた表である。左側の数値は 1000 字にあたる相対頻度であり、グラフの折れ線の色はそれぞれの学習者のデータを表している。

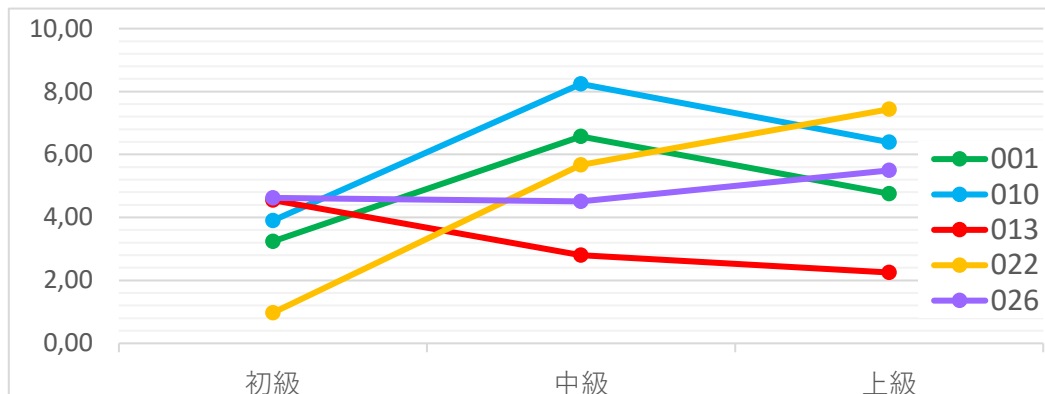


図 1：格助詞の誤用頻度の縦断的分析

3.3. 第 3 分析

第 1 調査では、助詞に関する誤用が最も頻繁であることが判明し、第 2 調査では、5 人の個々の学習者の誤用間で最大の差が見られるのも助詞の誤用であった。したがって、コーパスの誤用の最大の割合を占める 6 つの助詞をより詳細に分析することにした。その 6 つの助詞は「は」、「が」、「を」、「で」、「に」と「の」であった。これらの 6 つの助詞について、段階間の遷移での誤用の発生頻度の変化を比較し、誤用の最も頻繁な原因と、これらの助詞が最も頻繁に混同される文法要素を特定した。

誤用分析に対する批判の一つは、研究者が誤用の絶対数だけに基づいて結論を出すことが多く、ある要素が他の要素よりも頻繁に使用されて誤用が発生しやすいことを考慮に入れていないことである。これは、本研究の SLJ コーパスでも観察できた。初級段階では、接続詞の誤用は頻繁ではないが、中級段階では、学習者が条件節、原因節等の接続詞を知っていてこれらの誤用の割合ははるかに高くなる。ある要素が実際に学習者にとって困難であるか、または単にその要素の使用数が多いため誤用も多いかを確認するには、誤用の数だけでなく、誤用率も決定する必要がある。それは具体的に、標準的な文章において予想される使用量と、実際の使用量の比率のことである。つまり使用されるべきであった回数（教師によって採点された文章の回数）と実際に学習者が使用した回数の比率である。以上述べた六つの助詞のそれぞれの誤用率を上記のように計算し、これらの結果を比較した。

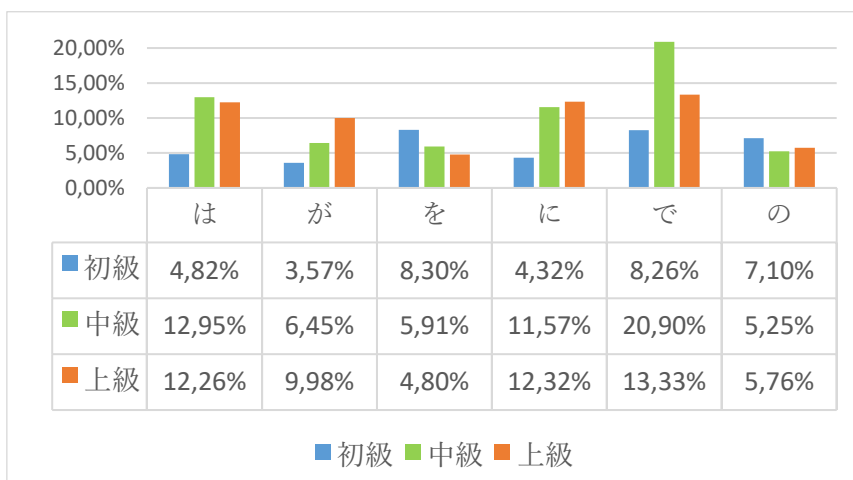


図 2：六つの最も誤用の多かった助詞の誤用率

4. 結果

研究結果からは次のことが分かった。

すべての段階で最も頻繁な誤用が助詞に関する誤用である。格助詞の「が」、「で」、「に」と「を」、取り立て助詞の「は」と連体助詞「の」の誤用が多いことが判明した。

初期段階では、基本文型、特にコピュラ「です」の使用、動詞と形容詞の活用などに関連する誤用も多く見られるが、学習時間が長いほどこれらの誤用は非常にまれになる。助詞の場合は逆である。すなわち、上記の六つのカテゴリーの誤用はある段階から次の段階に移行するにつれて誤用率が増えるのである。したがって、最も頻繁に使用される文法要素の誤用が最も頻繁であるという仮説が確認された。

それぞれの学習者で最も頻繁に発生する誤用の傾向に共通点が見られるという仮説も確認されたが、学習者間の個人差により、偏差もある。中級段階の作文状況が他の二つの段階とは異なるため、誤用の頻度が中級段階のデータで最も高く見られるという仮説は、部分的に確認された。その理由は、中級段階の誤用頻度は初期段階よりも高いものの、上級段階では、カテゴリーの半分以上の誤用頻度が中級段階よりもさらに高いためである。

より詳細に分析された六つの助詞のうち、四つの場合に初期段階での誤用割合は、上級段階での誤用割合よりも低くなっている。したがって、言語学習時間が長くなると誤用の頻度が減少するという仮説は確認されなかった。

中級以上の段階で最も頻繁な誤用は、学習者が学んだばかりの文法要素であるという仮説は、部分的に確認された。中級段階では、初級段階の文章の執筆時に学習者がまだ知らなかった接続詞に関する誤用が中級段階では増加すると同時に、かなり早い時点から学ぶ格助詞の誤用がすべての段階で最も頻繁で、上級段階では、学んだばかりの文法要素の誤用割合に増加は見られないからである。もちろん、学習者がそれらの要素を避けているために誤用が発生しない可能性もあるが、その仮説を確認するにはさらなる調査が必要である。

5. おわりに

このように、本研究では、スロベニア語母語話者の言語使用データから成る日本語学習者コーパスを作成し、コーパスのデータを元にして三つの言語分析を行い、スロベニア人日本語学習者の文章における文法誤用の発生傾向を個々の段階において分析した。作成したコーパスはオンライン公開されているため、それに基づいて他の一連のコーパス分析を実行することも可能としている。学習者の文章内の語彙やスタイルの誤用の比較、または、他の学習者コーパスとの比較、他の母語話者のコーパスとの比較などが可能である。本コーパスは、今後も発展させ、できるだけ多くのユーザーに利用できるよう継続していく予定である。

文 献

- Ellis, Rod (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Granger, Sylviane in Magali Paquot (2017). »Towards standardization of metadata for L2 corpora.« Presentation at the CLARIN workshop on Interoperability of Second Language Resources and Tools, 2018. <https://sweclarin.se/swe/workshop-interoperability-l2-resources-and-tools>
- Meunier, Fanny (2016). »Introduction to the LONGDALE Project«. In *Studies in learner corpus linguistics. Research and applications for foreign language teaching and assessment*, eds. Erik Castello, Katherine Ackerley in Francesca Coccetta, pp. 123-126. Berlin: Peter Lang.
- Volodina, Elena, Lena Granstedt, Arild Matsson, Beáta Megyesi, Ildikó Pilán, Julia Prentice, Rosén

Dan, et al. (2019) "The SweLL Language Learner Corpus: From Design to Annotation". Northern European Journal of Language Technology (NEJLT). pp. 67-104.

Pavlovič, Miha (2018) Izgradnja in analiza korpusa usvajanja japonščine kot tujega jezika na osnovni in srednji ravni (bachelor's thesis), Ljubljana: University of Ljubljana.

市川保子 (1997) 『日本語誤用例文小辞典』 凡人社.

海野多恵・他 (2012) 「オンライン日本語誤用辞典」 東京外国語大学

<http://cblle.tufs.ac.jp/llc/ja/index.php?menulang=en>

川口良 (1995) 「中上級日本語学習者の作文にみる誤用の一例」 言語文化と日本語教育, pp. 178-188.

佐治圭三 (1992) 『外国人が間違えやすい日本語の表現お』 ひつじ書房.

Serrano, Nestor・入野野みはる (2018) 『「誤用」から日本語学習の「成果」をはかる試み――中級学習者による「旅」の作文を用いて』 JLTANE 2018, Dartmouth, 2018年8月18日.

寺村秀夫 (1990) 『外国人日本語学習者の日本語誤用例集』 寺村誤用例集データベース. <http://teramuradb.ninjal.ac.jp/teramura.goyoureishu.pdf>.

原沢伊都夫. (2012) 「日本語初中級学習者の作文指導：学習者の誤用分析をもとに」 静岡大学国際交流センター紀要, 6, pp. 79-92.

丸山岳彦・田野村忠温 (2007) 「コーパス日本語言語学の射程」 日本語科学. 22, pp. 5-12. 国立国語研究所

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, pp. 111-119. 関西大学外国語学部.

野田尚史・迫田久美子 (2019) 『学習者コーパスと日本語教育研究』 くろしお出版.

吉田桃子 (2016) 「イタリア語話者の日本語作文の誤用分析から分かること」 日本語教育シンポジウム報告・発表論文集.