

様 式 C - 1 9、F - 1 9 - 1、Z - 1 9 （共通）

科学研究費助成事業 研究成果報告書



令和 3 年 4 月 23 日現在

機関番号：11301

研究種目：研究活動スタート支援

研究期間：2019～2020

課題番号：19K23033

研究課題名（和文）学習者コーパスの量的研究と質的研究を結びつける学習者を観察単位とした分析法の開発

研究課題名（英文）Developing analytical methods which connect quantitative learner corpus research with qualitative research using the language learner as an observation unit

研究代表者

森 秀明（MORI, HIDEAKI）

東北大学・文学研究科・専門研究員

研究者番号：00847353

交付決定額（研究期間全体）：（直接経費） 800,000 円

研究成果の概要（和文）：本研究では、学習者コーパスで計量的な分析を行う場合、言語単位ではなく、学習者を観察単位とすると統計的に有効な分析ができることを明らかにした。その上で、I-JAS（『多言語母語の日本語学習者横断コーパス』）を使用し、条件表現のタラは、学習者レベルが上がるにつれ使用数が上昇するのに対し、トやバは母語話者の半数程度までしか増加しないことを解明した。また、この量的な研究を基に、タラを過剰使用する学習者の用例を質的に分析した結果、初級では誤って時間節や連体修飾節にも広くタラが使用されている実態や、上級ではトやバがふさわしい文脈でもタラが使用される傾向が強いことが明らかになった。

研究成果の学術的意義や社会的意義

本研究によって、学習者コーパスでは言語単位を観察単位としても、統計的に有効な分析が困難なことが明らかとなった。また、学習者を観察単位とすれば、質的な研究につなげやすいことも判明した。この結果は、今後の学習者コーパスの量的な分析法の見直しを迫る成果である。さらに条件表現の分析では、トやバは上級で教えればよいとする主張に対し、上級で教えても十分な成果を上げられない可能性が示唆された。

研究成果の概要（英文）：This study reveals that when conducting quantitative learner corpus analyses, statistically effective analyses could be conducted if the learner is seen as an observation unit rather than using language as an analytical unit. The International Corpus of Japanese as a second language (I-JAS) was consulted, from which it was found that although the use of the conditional expression “tara” increased as learner levels rose, but that the use of “to” and “ba” increased only up to about half the usage frequency observed in native speakers. With this quantitative study as a basis, a qualitative analysis of learners who overused “tara” revealed that in beginner Japanese classes, “tara” was widely but mistakenly used in a temporal or adnominal clause, and that in the advanced levels there was a strong tendency to use “tara” even in contexts in which “to” and “ba” were more appropriate. These results suggest that a re-examination of quantitative learner corpus analysis methods is needed.

研究分野：日本語教育学

キーワード：学習者コーパスの分析法 量的研究 質的研究 条件表現

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

現在、日本語教育学では、母語話者や学習者の言語使用実態を明らかにする量的研究と、コーパスを巨大な用例集と見なして、学習者が犯しやすい誤用等を収集する質的研究の2種類に大別されており、学習者の量的な特徴を基に、その原因を質的に解明する分析法は確立されていない。その最大の理由は、学習者コーパスにおける設計上の特性に適した分析法を独自に開発することなく、コーパス言語学において先行している均衡コーパスの分析法に倣った研究を行っているためだと考えられる。

コーパスを使用した量的な研究では、文字、単語、表現、文などの頻度を数えることが多い。均衡コーパスでは、これらの言語単位の数、研究対象となる言語の実態を反映した頻度となるように、書籍等の無作為抽出を行ってデータを集積している。しかし、学習者コーパスは、限られた教育機関の限られた学習者からデータを採取することが多く、言語単位の頻度を集計しても、学習者の代表性を持った頻度になるとは限らない。逆に、学習者コーパスでは少数の学習者によって、全体の頻度が歪められることが多い。

したがって、学習者コーパスを使用した量的研究では、これまでの均衡コーパスに倣った分析方法ではなく、学習者コーパス独自の分析法を開発していく必要がある。また、質的な研究では、現在行われているようなコーパスデータを逐一吟味して誤用例等を探す研究方法では、研究者の負担が大きく、コーパスがどんどん巨大化していけば、早晩すべての用例に目を通すことが不可能になることは想像に難くない。かといって、誤用例の一部を取り上げる研究では、探し出した誤用例等がどれほどの広がりや重要性を持っているのか、評価することが難しい。このため、量的な研究によって研究対象としている現象がどのような広がりや分布をしているのかを明らかにし、その量的な傾向を手掛かりにして、質的な分析を行う対象を絞り込んで質的研究を行う分析方法が必要とされている。

2. 研究の目的

本研究の目的は、統計的に有効な学習者コーパスの量的研究と質的研究を有機的に結びつける分析法を開発することにある。このためには、まず、学習者コーパスに適した量的分析法を確立する必要がある。

学習者コーパス以前にも計量的な分析は数多くなされてきたが、その観察単位は学習者であった。現在は、均衡コーパスの分析法に倣って言語単位を観察単位とする分析法が主流だが、学習者コーパスの分析においては学習者を観察単位とすることで、統計的に有効な分析ができると考えられる。本研究では、言語単位を観察単位とした分析法と学習者を観察単位とした分析法を比較することによって、学習者コーパスに適した分析法とはどのようなものなのかを明らかにしていく。

学習者を観察単位とするということは、例えばある散布図を描いた際の各点が、一人一人の学習者の量的な状態を表しているということである。このため、ある分析観点から描かれた散布図を観察すれば、全体的な傾向性だけでなく、どのような特徴を持った学習者がどれくらい存在するかが明確に把握できる。これまでのように、やみくもにすべての用例を分析するのではなく、特徴的な分布をしている学習者に共通した特徴を観察することで、統計的な根拠に基づいた質的な分析を効率的に行える可能性がある。本研究では学習者を観察単位にすることによって、統計的に有効な分析を行うとともに、その分析結果の特徴を質的に解明する分析法の開発を目指す。

3. 研究の方法

(1) 学習者を観察単位とした分析法の有効性の確認：上述した通り、学習者コーパスの分析に当たっては、これまでの均衡コーパスで行われてきた言語単位を観察単位とした分析法より、学習者を観察単位とした分析法の方が、統計的にも有効であると考えられる。しかし、この研究法を広く認めてもらうためには、その意義を検証しておく必要がある。そこで、先行研究で学習者を観察単位としていた場合、これを言語単位で分析すればどのような結果となるのか、またその逆の場合はどうかなど、言語単位を観察単位とした分析法と学習者を観察単位とした分析法でこれまでの先行研究の分析結果を比較し、学習者を観察単位とした分析法の有効性を確認する。

(2) 『多言語母語の日本語学習者横断コーパス』(I-JAS)を使用した条件表現の使用実態と学習者特性との関係の解明：I-JASは、現在公開されている最大の学習者コーパスである。これを使用して代表的な条件表現であるト、バ、タラ、ナラの使用実態を明らかにする。また、I-JASでは、学習者個々の「フェイスシート一覧」が公開されているため、学習者の来日経験や日常的に会話できる母語話者の存在などの細かな学習者特性を使用した量的調査が可能になった。これを利用し、どのような学習者要因がどのような条件表現の使用に関わっているのかを解明する。

(3) 量的研究で分析対象者を絞り込んだ条件表現の質的研究：(2)の分析によって、どのような学習者がどのような条件表現を過大・過少使用しているかが明らかになる。この用例を分析し、母語話者の使用法と比較することで、そのような過大・過少使用がなぜ起きるのかを質的に明らかにする。

4. 研究成果

(1) 学習者を観察単位とした分析法の有効性の確認では、まず、学習者コーパス以前の研究で、学習者を観察単位とした Woods et al. (1986) のカイ二乗検定 (表 1) と、そのデータを基に言語単位で分析したカイ二乗検定 (表 2) を比較し、どちらが統計的に有効な分析になっているかを考察した。

表 1 メキシコ移民と母語話者の動詞時制の誤用数 (観察単位: 学習者)

Woods et al. (1986: 140) Table 9.4 より一部抜粋。効果量は、筆者が加えた。

	0	1error	2-6errors	Row total
Group A	7	7	16	30
Group B	13	11	6	30
Column total	20	18	22	60

$\chi^2 (2) = 7.22, p < .05$, Cramer's $V = .347$

表 2 メキシコ移民と母語話者の動詞時制の誤用数 (観察単位: 単語)

	誤用数	それ以外の語数	合計
Group A	71	2,929	3,000
	2.4%	97.6%	100.0%
Group B	35	2,965	3,000
	1.2%	98.8%	100.0%
合計	106	5,894	6,000

$\chi^2 (1) = 11.764, p < .01$, Cramer's $V = .044$

表 1 の場合、効果量のクラメールの連関係数は、.347 と中程度の関連性を示しているが、表 2 では、.044 とほとんど関連性は見られず、学習者を観察単位とするか誤用数という言語単位を観察単位とするかで、メキシコ移民と母語話者で動詞時制の誤用数に差があるのかないのかという結論が逆になっている。この分析を検討する際に最も重要な視点は、個体の独立性である。表 1 の場合、各学習者は互いに独立しているが、表 2 では、各単語は学習者が書いた 100 語のまとまりで強い関連性を持っており、独立性の仮定に違反している。よって統計的に有効な分析は、学習者を観察単位とした表 1 の分析だと考えられる。

同様に松田・宮永・庵 (2013: 51) や山内 (2009: 26) を使用し観察単位を替えて比較検討した場合も、いずれも学習者を観察単位とした分析法の有効性が認められた。

(2) I-JAS を利用した条件表現の使用実態では、ト、バ、タラ、ナラの合計数で見ると、初級から上級上に進むにつれて、次第に母語話者の使用分布に近づいていく傾向が見られるが (図 1)、タラは上級上では母語話者より過剰に使用されている (図 2)。一方、トやバは、上級上になっても母語話者の半数程度しか使用されていない実態が明らかとなった (図 3、4)。この現象は、日本語学習者の習得がタラに偏り、トやバの習得が十分になされていない可能性を示唆する。なお、ナラは、母語話者、学習者ともに使用数が少なく、合計値に影響を与える量ではなかった。

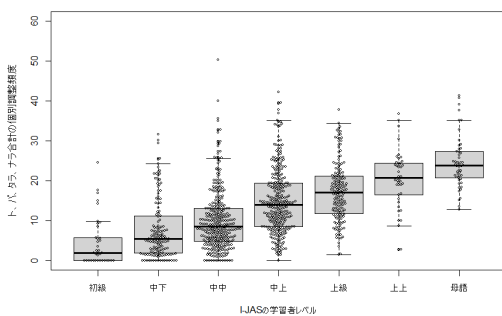


図 1 ト、バ、タラ、ナラの合計数推移

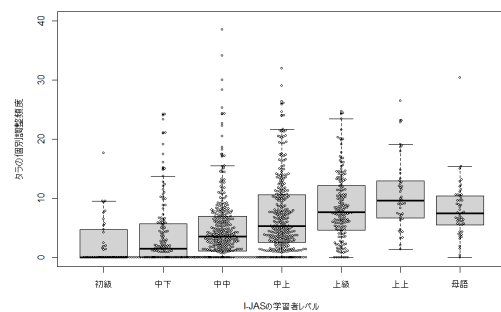


図 2 タラの頻度推移

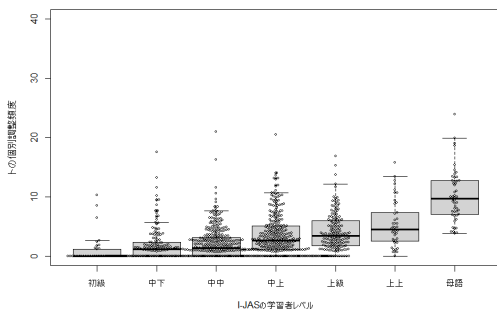


図 3 トの頻度推移

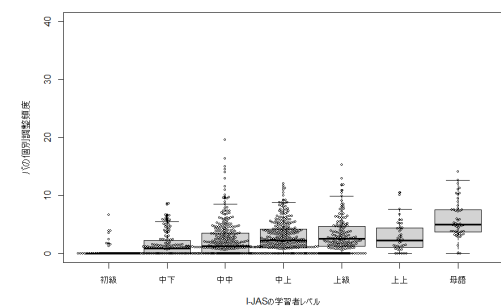


図 4 バの頻度推移

タラの使用については、日本在住の自然環境学習者や教室環境学習者に初級から過剰使用を行う傾向があることが報告されている (森, 2018: 318-319)。このことは、日常的に日本語を耳にする機会のある学習者がタラを多く使用する可能性を示唆している。そこで、海外の学習者

表3 I-JASの海外学習者におけるタラの使用頻度に影響を与える要因

	<i>b</i>	<i>S.E</i>	
定数	-3.471 ***	.606	-.004
来日の有無	1.030 **	.355	.096
母語話者の友人	1.514 ***	.348	.135
J-CATの聴解点数	.083 ***	.012	.271
J-CATの読解点数	.069 ***	.015	.178
Adjusted <i>R</i> ²	.239		
<i>n</i>	850		

* *P* < .05, ** *P* < .01, *** *P* < .001

タラの使用が変化することを意味している。他の要因の影響を除いた目的変数とのピアソンの積率相関係数を表す *r* を見ても、聴解の点数の影響度が最も高く、日本語がよく聞き取れる学習者ほど、タラを多く使用する傾向があることが明らかになった。これは日本に在住している学習者にタラの高頻度使用者が多い現象と通じる要因である。J-CAT の読解の点数もタラの頻度に影響を与える要因となっているが、これは、一定の量の文章を読んで、そこに何が書かれているかが理解できるという能力が、タラを使用した複文を理解できる能力に関係しているためだと考えられる。

ト、バ、ナラについても同様の重回帰分析を行ったが、統計的に有意な *b* 値を示したのはトの J-CAT の聴解で.034、文法で.029、バの文法で.021、ナラの語彙で.019 であった。これらの数値はタラと比べるといずれも低く、使用回数に対する重要な要因にはなっていない。

(3) 上記の量的な分析を基に、外れ値を示す日本在住の自然環境学習者の用例分析という質的調査を行った。この結果、従来条件表現には顕著な誤用はないとされていたが(庵, 2017: 18)、初級の学習者にはタラの誤用が多く、「マリ,さん,と,ケン,さん,は,地図,を,見たら,犬,が,バスケット,を,入り,まし,た。」(ID: JJN40-SW1、連番: 200)、「ケン,さん,は,家,に,着いたら,鍵,を,忘れ,まし,た, (JJN40-SW2、60)」のように、「見ているときに」「家に着いたとき」といった時間節の代わりに、タラを使用している例や、「一番,の,一,電車,一,乗ったら,あと,が,【地名2】の, うん,、,ないたら,後,が,、,ちょっと,歩いて,<C>あー,<K>【語学教室名1】が,、,着い,た(JJN40-I、1670)」のように、「乗った後」「なかった後」のような連体修飾節を作る意図での誤用が見られた。ここからは、初級段階では耳なじみのあるタラが、条件節以外にも広く使用されている様子がうかがえる。一方、習得レベルが上がるにつれ、このような誤用はなくなっていくが、タラ以外のト、バの使用は増えていかず、「で,も一,私,料理したら,一,あの一,だい,、,だいたい,、,あの一,ふいりぴ,、,フィリピン,の,料理,です, (JJN42-I、6970)」のように、「料理すると」となる方が自然な場合でも、タラが使用されていることが分かった。これらを総合して考えると、日本語の会話でトやバより多用されるタラという表現が、使い勝手の良い従属節を作る助詞として過剰般化され、誤用が淘汰された後、トやバを使用した方がふさわしい場面でも、多用され続けるのだと考えられる。これまで、トやバは、上級以下ではあまり十分に使いこなせていないため、初級からではなく、上級からしっかり教えるのが良い(山内, 2009: 39)とする考え方もあったが、タラ一辺倒で日本語を習得してきた学習者が、上級でトやバを教わったからといって、本当にそれらを使用できるようになるのかについては疑問が残る。この問題については、初級段階からトやバを正しく使用する学習者とはどのような学習者なのかを明らかにすることによって、さらに追及していきたい。

<参考文献>

- Woods, A., Fletcher, P. and Hughes, A. (1986) *Statistics in language studies. Cambridge Textbooks in Linguistics*. New York: Cambridge University Press.
- 庵功雄(2017)『一歩進んだ日本語文法の教え方1』くろしお出版。
- 松田真希子・宮永愛子・庵功雄(2013)「超級日本語話者の談話特性 テキストマイニングを用いた分析」『国立国語研究所論集』5:43-63.
- 森秀明(2018)「学習者コーパスを使用したレベル別頻度比較の方法」『Learner Corpus Studies in Asia and the World』3:303-322.
- 山内博之(2009)『プロフィシエンシーから見た日本語教育文法』ひつじ書房。

850 名に対し、来日経験、身近な友人に日本人がいるか、日本語の授業以外で日本語を話すかといった学習者特性と、J-CAT (Japanese Computerized Adaptive Test) の聴解、語彙、文法、読解を説明変数、タラの頻度を目的変数とする重回帰分析を行った(表3)。この結果、来日の経験があれば、タラの使用頻度が約1回、身近な友人に日本人がいれば約1.5回増加することが分かった。この他、J-CAT の得点の聴解と読解の点数に、有意な関連性が認められた。J-CAT は、各項目が100点満点であるため、*b* の偏回帰係数は.083と小さく見えるが、10点の違いで.83回の

5．主な発表論文等

〔雑誌論文〕 計2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 2件）

1．著者名 森秀明	4．巻 第84巻第1・2号
2．論文標題 個体を観察単位としたコーパス分析の有効性	5．発行年 2020年
3．雑誌名 文化	6．最初と最後の頁 10-22
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 森秀明	4．巻 32巻7号
2．論文標題 学習者コーパスに適した分析法の検討	5．発行年 2020年
3．雑誌名 計量国語学	6．最初と最後の頁 436-446
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 -

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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
--	---------------------------	-----------------------	----

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------