様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２６ 年　 ５ 月 ３１ 日現在

機関番号：１２６０３
研究種目：基盤研究(C)
研究期間：2011～2013
課題番号：２３５２０６１３

研究課題名（和文）日本語を母語としない人による日本語作文過程の分析のためのデータベース

研究課題名（英文）Database for analysis of text composing process by J.S.L. students

研究代表者
土屋　順一（TSUCHIYA, Junichi）

東京外国語大学・留学生日本語教育センター・准教授

研究者番号：１０２６２２１３

交付決定額（研究期間全体）：（直接経費）　3,200,000円、（間接経費）　960,000円

研究成果の概要（和文）：留学生が日本語の質問に対するこたえをパーソナルコンピュータのキーボードから入力する過程のディスプレイ表示を録画した。録画をチェックして、入力の変更1回ごとに1データとする日本語作文過程のデータベースを作成した。データの総数は49575、入力者の母語は80以上になった。日本語学習者の属性と入力変更の関係の分析が可能である。

研究成果の概要（英文）：In order to investigate the types of elaboration which J.S.L. students make when they type in Japanese using a keyboard, we video-recorded the display while they are typing their responses in Japanese to questions given on the computer display. By checking with video-record we made a database of elaboration by J.S.L. students. It contains 49575 data of more than 80 different language speakers.

研究分野：人文学

科研費の分科・細目：言語学　日本語教育

キーワード：キーボード入力　誤用分析　母語　日本語学習歴

１．研究開始当初の背景

研究代表者は、日本語の音韻をまちがえて記憶しているためにおこるキーボード誤入力（「授業」を「じゅうぎょう」、「中国」を「ちゅうこく」と入力するなど）に関して、母語・学習歴・学習環境などによるちがいをさぐることを目的に、学習者が日本語作文をキーボード入力する過程のディスプレイ表示を録画してデータをあつめてきた。その結果、誤入力が、母音の長短の区別と破裂音の有声無声の区別の 2 点に集中し、母語および来日前の日本語学習歴と関係がふかいことがわかった。

先行研究の過程で、最終的に誤用になっていなくても、助詞の「が」と「は」、「に」と「で」などを、作文入力の過程で、学習者がかきかえている例が発見された（これは初級の学習者がまちがえやすい事項でもある）。

たとえば、先行研究でえられた 693 名、41,845 件のデータを対象に、「に」と「で」のかきかえの検索をかけてみると、83 件ヒットした。ところが、ひとつひとつをしらべてみると、「東京でいた」→「東京にいた」のような後続動詞の制約による変更だけでなく、「3 月（で）（に）卒業した」「自分（で）（に）わかる」「ホームステイ（で）（に）いった」「一年間（で）（に）」のように、助詞「で」「に」の用法の多様性を反映した多様なかきかえ例がでてくる。先行研究で作成したデータベースでは、文字の検索はできても、文法範疇の検索はできないのが現状である。

２．研究の目的

日本語のある文章表現を完成させるのに日本語学習者がどのような言語ストラテジーをつかっているのかをあきらかにするために、先行研究ですでに「文字列」「品詞」タグがつけてあるデータに「文法範疇」タグをつけて文法面からの分析が可能なようにすることが本研究の最大の目的である。

文法範疇による分類は、たとえば、動詞の受身・使役・自他、こそあの文脈指示、条件説、連体修飾、比較、並列など音韻とくらべてはるかに多様である。すべての文法範疇のタグづけをすることは不可能であるが、ある程度タグづけができれば、文法的なかきかえを、学習者の母語・学習段階・学習環境などの観点からクロス分析することが可能になる。

それにくわえて、データの精度をあげるために、100 名程度データを追加して、データベースの規模を、800 名、50,000 件程度まで拡大する。

３．研究の方法

(1)アンケート調査票にこたえる形式で、日本語学習者に実際に日本語の自由作文をコンピュータのキーボードから入力させる。入力の過程を DVD レコーダーに録画する。入力過程を記録した映像を再生チェックして、電子テキスト化し、1 回の訂正を 1 データとしてデータベース化する。このデータベースが先行研究で 41,845 件に達している。そのデータに対して、ひとつの文法範疇に関するタグづけをおこない、文法面からの分析をおこなう。

たとえば、先行研究において、助詞のかきかえを検索する作業の過程で、ひとりの学習者に助詞のかきかえが連続している例が偶然みつかった。現時点では文法範疇にもとづいたひろい検索がかけられないので、手作業で時間軸にならべてみたところ、つぎのようになった。

「先生方が留学生と日本人学生との」
「先生方が特に留学生と日本人学生との」
「先生方が特に留学生にも日本人学生にも」
「先生方が留学生にも日本人にも」
「先生方が留学生でも日本人学生の扱い方」
「先生方が留学生でも日本人学生と同じ扱い方をされたので、」
「先生方が留学生についても日本人学生と同じ扱い方をされたので、最初速い日本語を聞き取るのがちょっと大変でした。」

最終形は誤用のない適切な文である。この学習者はひとつひとつの助詞の誤用を訂正しているのではなく、後続する部分を検討しながら、それにあわせて語彙を適切にえらび、かきかえをおこなっていると推測される。このようなかきかえに対して「比較」という機能のタグをつけて、検索できるようにして、ほかのデータと比較できるようにして、データベースの有用性をたかめる。

(2) 機能のタグをつける作業をくりかえしながら、タグの種類（並列、連体修飾など）をふやしていく。

(3)先行研究のデータ収集時にはあまりみられなかった、つぎのような口語先行型の学習形態の学習者のデータを、年間 30 名程度のペースであらたに採集して追加する。

①幼少時に情操教育として日本語を学習し、その後高等教育まで継続して学習した者
②アニメ、マンガなどを通じて日本語を独習し、その後高等教育まで継続して学習した者

(4)研究成果を発表する。

４．研究成果

(1)並列表現

留学生による日本語作文キーボード入力における並列表現を分析した結果、以下のような傾向がわかった。

① 並列表現に「、」が多用される。名詞数が 2 個の場合にも「、」と「と」がほぼ同数で、3 個以上の場合は「、」が圧倒的多数である。
② 「、」の連続は回避されないが、おなじ助詞の連続は回避される傾向がある。母語の「A,B,C and D」という並列表現の影響があるとおもわれる。
③ 「とか」の使用はすくない。はなしことばの調査結果とはことなる。

④ かきかえのすくない表現「と」「や」。最初に並列する名詞を確定してからタイプすることがおおい。
⑤ かきかえのおおい表現「とか」「、」。名詞をおもいだしながらタイプしていくことがおおい。
⑥ 「中国語話者は「～も～も V」の構造を回避する」という説を検証したところ、その表現は、母語に関係なく使用頻度がひくかった。
⑦ 「「～も～も V」の構文は習得が困難な項目である」という説を検証したところ、数すくない使用者の日本語レベルは比較的ひくく、逆に上級者はこの構文をつかっていない。習得が困難だからつかわれないとはいえない。

(2)格助詞＋「の」
留学生による日本語作文キーボード入力における格助詞＋「の」のかきかえを分析した結果、つぎの 3 種に分類できた。
① かきかえのおおい助詞
『場所の「での」』
『手段の「での」』
『対象の「への」』
『比較の「との」』
「での」は連体修飾を完成する過程で推敲されることがおおいようで、かきかえがおおい。場所の「での」は総数もおおく、中級者もよくつかうが、手段の「での」はレベルがたかく、かきかえもおおい。興味ぶかいことに、マレー語話者は『場所の「での」』の使用が非常におおく、『手段の「での」』の使用がすくない。
比較の「との」はかきかえがおおく、レベルがたかい。
② かきかえのすくない助詞
『相手の「との」』
「からの」
「までの」
相手の「との」は中級者もつかい、用例数がおおいが、かきかえはすくない。「からの」「までの」のかきかえはすくなく、ひとつの助詞のようにセットとしてつかわれる傾向がみられる。特に「からの」の用例数は「での」「との」についでおおいが、かきかえはすくない。
③ 使用が回避される助詞
『方向の「への」』
方向の「への」は 2 例だけで、どちらもモンゴル語話者である。誤用「にの」が 1 例もなかったこととあわせてかんがえると、方向をあらわす助詞と「の」の連結は、回避される傾向があるといえる。

(3)口語先行型の学習者
あらたにデータの収集を目指した口語先行型の日本語学習者のデータは目標どおりには収集できなかった。
先行研究においても、日本語能力がたかいにもかかわらずキーボード入力ができないため調査を依頼できない、という学習者はいた。それは、
① パソコン環境がなく、もっぱら手がきである。
② パソコンで欧文の入力はするが、日本語環境がないので、日本語入力はしたことがない。
というものであった。ところが、本研究がスタートしてからであった口語先行型の日本語学習者の中には、QWERTY 配列のキーボードをつかったことがない、という者がふえてきた。数千字におよぶレポートもスマートフォンのキーボードから入力する（もちろん母語でも日本語でも）というのであった。QWERTY 配列のキーボードに、はじめてむかって「キーをさがしながらタイプすることはできますが、大文字と小文字のきりかえはどうやってしますか？」という学習者さえいた。Shift キーさえしらないのである。
タブレットからの入力では、手がき入力も頻繁につかっているようで「スマホ」の時代には、あたらしいデータ収集の方法が必要であることを実感した。

(4)データベースの仕様
資料提供者：795 名
未習来日者：260 名
大学で 3 年以上日本語専攻後来日者：142 名
資料提供者の母語上位 10：
北京語　97 名
マレー語　82 名
朝鮮語　80 名
モンゴル語　39 名
ベトナム語　32 名
モンゴル語・北京語　31 名
タイ語　30 名
台湾語・北京語　27 名
朝鮮語・北京語　27 名
ウイグル語・北京語　21 名

ソフトウェア：File Maker Pro
データ数：49,575 件
総入力時間：340 時間 36 分
総文字数：315,680 字
消去文字数：104,957 字

消去されて、最終的な作文にのこっていない約 10 万字（とはいっても大半が単純なタイプミスだが）が、このデータベースの基底である。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　１　件）
土屋順一、留学生による日本語作文における並列表現のかきかえの傾向、日本語教育連絡会議論文集、査読無、vol.25、2013、7-11

〔学会発表〕（計　3　件）
①土屋順一、モンゴル人学習者による日本語キーボード入力の特徴、モンゴル国立大学国際言語文化学部教員研究会シンポジウム、2013.09.14、モンゴル国立大学

②土屋順一、留学生による日本語作文入力過程データベースは文法研究につかえるか、東京外国語大学国際日本研究センター『外国語と日本語との対照言語学的研究』第10回研究会、2013.07.13、東京外国語大学

③土屋順一、留学生による日本語作文における格助詞＋「の」のかきかえの傾向、第9回国際日本語教育・日本研究シンポジウム、2012.11.24、香港城市大学

６．研究組織
(1)研究代表者
土屋　順一（TSUCHIYA Junichi）
東京外国語大学・留学生日本語教育センター・准教授
研究者番号：10262213