# 学習者が犯す誤用の要因・背景からみる日本語作文支援

八木 豊（株式会社ピコラボ）[1]
ホドシチェク・ボル（東京工業大学）
阿辺川 武（国立情報学研究所）
仁科 喜久子（東京工業大学）

# Relevance of Learners' Errors in the Development of a Japanese Writing Support System

Yutaka YAGI (Picolab Co., Ltd.)
Bor Hodošček (Tokyo Institute of Technology)
Takeshi ABEKAWA (National Institute of Informatics)
Kikuko NISHINA (Tokyo Institute of Technology)

## １．はじめに

近年、国立国語研究所による「現代日本語書き言葉均衡コーパス」（以後 BCCWJ）をはじめとする日本語大規模コーパスの開発が進展し、オンラインのコーパス検索ツールとしての「中納言」、「少納言」、「NINJAL-LWP for BCCWJ」によって、特定の語の頻度や共起関係、文法的な振る舞いなどを知ることができるようになり、日本語の研究者には大きな恩恵をもたらした。また日本語教育の分野でも、日本語教育の研究者や教師によるこれらのコーパスやツールを利用した教育方法や教材開発の動きがみられるようになってきた。仁科他(2011)、Hodošček 他(2011)による日本語作文支援システム「なつめ」[2]の開発もその一つであり、文書作成時に表現したい共起語の検索と例文参照を可能にした。しかしながら、このシステムは上級レベルの一部の学習者を除いて、利用するには困難な点が多い。例えば、単語の表記を正しく習得してないと検索できない、提示される例文は学習者の日本語能力に応じたレベルに絞り込まれていないなどの問題があるためである。そこで、さらに広範囲の学習者にも容易に利用できるシステムを目指し、学習者作文コーパス「なたね」[3]を構築し、そこに見られる学習者の犯しやすい誤用を分析し、その誤用の要因や背景を知ることで、学習者が入力した文の誤用を自動的に指摘して修正案を示すシステムを最終目標とすることとした。

## ２．学習者作文コーパス「なたね」

「なたね」は、我々が独自に収集した学習者作文に対して日本語教師による添削を行った誤用タグ付きデータである。誤用タグは大きく「誤用の対象」、「誤用の内容」、「誤用の要因・背景」という 3 つの視点から構成しており、さらにそれぞれを 3 階層に細分類することで、全体として約 70 種類を定義している（曹他(2012)）。2012 年 12 月現在、大学院や大学あるいは語学学校に在籍する 192 人の日本語学習者による 285 作文（総文字数 205,520

[1] yagi@picolab.jp
[2] 日本語作文支援システム「なつめ」 http://hinoki.ryu.titech.ac.jp/natsume/
[3] 学習者作文コーパス「なたね」 http://hinoki.ryu.titech.ac.jp/natane/

表1 母語別「学習者数」

| 母語 | 男性 | 女性 | 性別未入力 | 計 |
|---|---|---|---|---|
| 中国語 | 50 | 43 | 22 | 115 |
| マラーティー語 | 6 | 23 | 7 | 36 |
| ベトナム語 | 6 | | 7 | 13 |
| 韓国語 | 6 | 1 | 4 | 11 |
| スペイン語 | 2 | | | 2 |
| マレー語 | 1 | | | 1 |
| スロベニア語 | 1 | | | 1 |
| ハンガリー語 | 1 | | | 1 |
| タイ語 | | | 1 | 1 |
| 母語未入力 | 1 | | 10 | 11 |
| 計 | 74 | 67 | 51 | 192 |

表2 母語別「作文数」

| 母語 | 男性 | 女性 | 性別未入力 | 計 |
|---|---|---|---|---|
| 中国語 | 62 | 64 | 26 | 152 |
| マラーティー語 | 6 | 23 | 7 | 36 |
| ベトナム語 | 18 | | 9 | 27 |
| 韓国語 | 24 | 3 | 7 | 34 |
| スペイン語 | 2 | | | 2 |
| マレー語 | 8 | | | 8 |
| スロベニア語 | 7 | | | 7 |
| ハンガリー語 | 1 | | | 1 |
| タイ語 | | | 1 | 1 |
| 母語未入力 | 5 | | 12 | 17 |
| 計 | 133 | 90 | 62 | 285 |

字）に含まれる約 6,500 箇所の誤用に対しておよそ 9,000 件の誤用タグを付与して公開している[4]。収集した作文は、PC 入力と手書きの区別、辞書使用の有無や時間制限などのコントロールを行っておらず、作文のテーマも、自己紹介からエッセイ風のものまで様々である。作文データそのもの以外に、性別、国籍、母語、学習歴、日本語能力（日本語能力試験のレベルや日本語教師による主観評価）といった学習者のメタ情報も可能な範囲で併せて収集しており、作文を公開するにあたっては、複数の日本語教師の協力のもとに本人の承諾を得ることができた情報のみを公開している。「なたね」における母語別の学習者数および作文数を表１、表２に示す。作文を収集できる環境が限られていることから現状では中国語を母語とする学習者が多く、全体の半分以上を占めている。

## ３．「誤用の要因・背景」の分析

本章では、「なたね」に付与した誤用タグのうち「誤用の要因・背景」に着目して、学習者が犯しやすい誤りの傾向、学習者の母語や日本語能力といったメタ情報との関連について分析を行う。表３は「誤用の要因・背景」に含まれる誤用タグの頻度を母語別に集計した結果である。表見出しのアルファベットは学習者の母語を表しており（脚注参照）、それぞれの列がその母語における誤用タグの頻度、右端の列が「なたね」全体の頻度である。以降では、「誤用の要因・背景」に含まれる誤用タグの項目「類似」「母語干渉」「レジスター」を取り上げ、順を追って説明する。

### ３．１．類似

類似した語句との混同が要因となっている誤用が該当し、類似している内容に応じて、意味の類似、字形の類似、音の類似の 3 つに下位分類している。それぞれについて代表的な誤用例を以下に挙げる。矢印の左側の下線部が誤用箇所、矢印の右側の斜体が日本語教師による訂正例で、末尾の括弧内には学習者の母語を記した。

【意味の類似】成長についてだんだん深く<u>了解</u>→*理解*できた。（中国語）
【字形の類似】<u>公島</u>→*広島*と東京とおきなわを見たいです。（マラーティー語）
【音の類似】これは私のしょうらいの<u>ゆうめい</u>→*ゆめ*です。（マラーティー語）

意味の類似では、特に日本語で用いられるある漢語の意味が中国とは異なる意味で用い

[4] 総文字数には句読点やその他の補助記号も含む。ただし、現在もメンテナンスを継続しており、Web サイト上での表示はここで挙げた数値と一致しないことがある。

表3 誤用の要因・背景[5]

| 項目 | zh | mr | vi | ko | es | ms | sl | hu | th | 未 | 計 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 意味の類似 | 38 | 141 | 11 | 32 | | 2 | 7 | | 2 | 10 | 243 |
| 字形の類似 | 2 | 47 | 1 | 4 | | 2 | 1 | | | | 57 |
| 音の類似 | 7 | 110 | 1 | 10 | | 3 | 2 | | 1 | | 134 |
| 母語干渉 | 45 | 6 | 1 | 5 | | | | 1 | | | 58 |
| レジスター | 384 | 12 | 8 | 46 | 9 | | 2 | 4 | | 18 | 483 |
| 文体の不統一 | 411 | 21 | 10 | 10 | 9 | | | 3 | | 14 | 478 |
| その他 | 12 | 3 | 1 | 3 | | | | | | 2 | 21 |
| 計 | 899 | 340 | 33 | 110 | 18 | 7 | 12 | 8 | 3 | 44 | 1474 |

られることがしばしばある。例えば日本語で「理解」と表現する場合に、中国語では「了解」と表現することができる。このような場合、学習者は日本語のコンテクストの中に母語の意味と合致する語を挿入してしまう。日本語において「理解」と「了解」は意味的に類似してはいるが使い分けが必要であることから「(漢語の) 意味の類似」という誤用タグを付与している。この例は、中国語からの母語干渉と重なるものである。

類似に関する誤用の中で字形の類似や音の類似では、マラーティー語を母語とする学習者の誤用が著しく多くなっている。これは、マラーティー語では、作文を収集した多くが日本語レベル初級の学習者で平仮名・片仮名の読み書きも不十分であることに加えて、原則としてパソコンなどを使用せず、手書きの作文を収集したことで余計に顕著な傾向が現れたためといえる。実際は字形の類似と音の類似は相互的であり、どちらによるものかの判定は困難である。例えば、マラーティー語話者の作文中に「首で走いて→*道を歩きながら*」という誤用がある。「首」は「道」という字形の誤り、「走」は「歩」の字形の誤りである。直接学習者にインタビューできないため判定は推測によることになるが、音声では「みち」「あるいて」と認識していると思われる。上級者で日本語の音声を正確に習得していない場合があっても、漢字表記では音声習得の不正確さは顕在化しないが、非漢字圏初級学習者は、仮名表記をすることで音の類似による誤用が顕著になっている。

その他の母語については中上級の学習者で構成されており、字形の類似や音の類似による誤用はほとんど見られなくなる。意味の類似による誤用については、日本語レベルが上がっても中国語や韓国語といった漢字圏の学習者を中心に散見されることと対照的である。

### ３．２．母語干渉

中国語を母語とする学習者による熟語の誤用など、学習者の母語の影響に因ると考えられる誤用が該当する。類似の場合と同様に、代表的な誤用例を以下に挙げる。

【母語干渉】十月一日午後、わたしたちは 4 時の火車→*汽車*に乗って、…(中国語)
【母語干渉】この場合は更生された→*更生した*人間ならば例外にしたいと思う。(韓国語)

母語干渉は、コーパス全体でも 58 件と少ないうえに、そのうちのおよそ 8 割は中国語を母語とする学習者による漢字選択の誤りである。これは、中国語を母語とする学習者の割合が多いこともあるが、日本語教師が添削する際に母語干渉であると判断できる内容は、漢

[5] zh：中国語、mr：マラーティー語、vi：ベトナム語、ko：韓国語、es：スペイン語、ms：マレー語、sl：スロベニア語、hu：ハンガリー語、th：タイ語、未：母語未入力

字圏の学習者による漢字選択の誤りに限定されやすいためではないかと考える。その他は、前述の 2 つ目に挙げた誤用例のように、韓国語を母語とする学習者が自動詞に「される」をつける誤りが 2 件ほど含まれている以外に、母語干渉と一概には言えないものも含まれており、今後、タグ付けした日本語教師への確認および必要ならば修正を行う予定である。

### ３．３．レジスター

機能文法では言語表現の異なりを「社会的な拘束力をもつ言語学上の規範」における言語使用域の変異即ち「レジスター」と呼び、Halliday(2004)はレジスター機能として次の 3 項目を挙げている。

（1）コミュニケーションの目的と主題に関わる「フィールド」（Field of discourse）
（2）コミュニケーションを行うための手段に関わる「モード」（Mode of discourse）
（3）コミュニケーションパートナー同士の関係に関わる「テナー」（Tenor of discourse）

書き手と話し手がどのような関係で、どのようコンテクストのもとで発話するかによって、それぞれ異なる語彙・文法項目で記述されることを示すものである。

学習者作文においては、授業で提出するレポート内で話し言葉を使用しているなど「場」にそぐわない表現全般がレジスターの誤りに該当する。現時点でレジスターに関する誤りのタグは 483 件あるが、「話し言葉と書き言葉」の違いによるものが大部分である。類似の場合と同様に、代表的な誤用例を以下に挙げる。

【レジスター1】少子化のせいで→*ために*、これから日本人の労働者がだんだん→*次第に*少なくなります。（ベトナム語）
【レジスター2】文章を読んでいるとき、とても苦しいですね。ときどき意味はちゃんと→*十分に*理解できないこともありますよ。（中国語）
【レジスター3】女性たちも経済的に力を持ち始め、徐徐に平等に向けての運動をやり始めた→*始めた*。（韓国語）

レジスター1 の例では、理由や原因を示す接続表現に主観的な意味を含む「せい」を用いている。アカデミックな文章では判断に感情表現を含ませるのは不適切であり、レジスターの誤りと判断される。「だんだん」は話し言葉であるため、書きことばの表現に修正案が示されている。

レジスター2 の例は、初級会話で学習した終助詞が使用されている作文である。日本語の終助詞は、コミュニケーション相手の同意を求めるために有効な表現であるが、アカデミックな文章ではこの種の表現を使用しないことを習得していない例である。

レジスター3 の例は、「（運動を）やり始める」という動詞が話し言葉のなかでもくだけた表現となっている。他にも次のようなくだけた表現の作文がみられる。これらの表現は、初級教科書でも現れないものであり、日本留学後のコミュニケーションを通して教室外で習得した表現と推測される。

【レジスター4】「しかし、伝統的な習慣とか→*など*では女性が不平等な目に会うことがいまだにも多く残っている。」（韓国語）
【レジスター5】くじ引きで日本語クラスに入り日本語を勉強し始りました。うちの→*私たちの*クラスで 10 人がアメリカの大学に入学して、他の 20 人は全部日本に来ました。（中

国語）

レジスターの誤りは、前述の類似の誤りとは反対に初級の学習者であるマラーティー語母語話者にはほとんどみられなかった。これは、初級学習者がレジスターを使い分けるに至っていない点にある。初級で教えられる語彙および教材の構成からみると、おおむね話し言葉が優先的に導入される。そのため、日本語教師のほうで初級学習者に対してはそこまでチェックせず表記の誤りなどその他の添削を優先するということが、日本語教師へのインタビューから明らかになった。

レジスターが問題になるのは、このようなシラバスで学んできた学習者が中級から上級に至った段階で、レポートなどのアカデミックな文章を書く必要性が生じる場合である。アカデミックな文章では、学習者は話し言葉と書き言葉を区別して書き分けなければならないほか、作文全体を通して文体の統一も図らねばならない。次の例は、作文中での文体の不統一による誤用である。作文全体の中で文末の「真の鍵でしょう」の部分のみが丁寧体となっている。「である」の推量形がわからないために「でしょう」にしたと推測できる誤用例が、他の学習者の作文にも散見する。

【文体の不統一】現状がかえない、どうしも真の先進国にならない。女性の社会進出は先進国に真の鍵でしょう→*であろう*。（中国語）

以上のようなレジスターの不整合としての誤用例は、話し言葉による会話場面を中心とする初級の教材での学習内容を習得した後で、文章を書く段階に入って、書き言葉のレジスターの知識が不足しているためと考えられる。現時点では、このような区別をレジスターの異なりとして体系的に教える教材はほとんどなく、アカデミックな表現が必要な上級レベルの学習者に対する教材やコースウエアへの対応が十分でないと推測できる。

## ４．まとめと今後の課題

本稿では、作文を支援するシステムを上級者のみでなく広範囲の学習者にも容易に利用できるシステムを目指し、学習者作文コーパス「なたね」を構築し、自動校正システムを最終目標として、そこに見られる学習者の犯しやすい誤用を分析し、その誤用の要因や背景を考察した。

誤用の要因と背景を分析するために、「なたね」に収録されている「意味の類似」「字形の類似」「音の類似」「母語干渉」「レジスター」の誤用例を観察し、考察した結果、以下のような結論を得た。

(1) 「字形の類似」「音の類似」による誤りは、非漢字圏初級学習者の例に多く見られた。語の表記と音声理解は相互的なものであり、どの母語の学習者にも誤った理解はあるが、特に非漢字圏初級学習者は漢字表記にハンディキャップがあるため、仮名表記を使用することで音声理解の誤りが顕在化していると考えられる。

(2) 「意味の類似」による誤りの中で漢字圏学習者によるものは、母語における漢語の意味と日本語における意味の異同によって誤ることがあり、母語干渉の影響もあると考えられる。

(3) 「母語干渉」は、語の意味の類似によるものが多く見られ、構文的なものもわずかであるが見られた。

(4) 「レジスター」の誤用は、初級レベルではほとんどタグが付けられていない。その理由は、初級学習者の語彙、表現の学習範囲が話し言葉中心であり、レジスターの違いを示すバリエーションがないことから、誤用としてタグを付けられないためである。一方、上級者では、話し言葉によって学んだ日本語の知識で、アカデミックな文章を書く段階になって、レジスターの知識が不十分であるために、不適切な表現が散見されることになる。文体の不統一についても文法的な知識の不足が影響している部分があると考えられる。

上級学習者は初級で学んだ話し言葉に加えて、アカデミックな書き言葉、さらに高度なフォーマルな話し言葉、手紙などのフォーマルな書き言葉表現など様々なバリエーションを習得する必要が生じてくる。これらの表現を教室の授業だけで学ぶには、時間的制限もあり、習熟することは困難である。

我々の今後の課題としては、さらに学習者データを追加し、不適切な表現を分析することで、学習者に必要な適切な文章表現の提示を可能にするシステムを目指す必要がある。

## 謝辞

本研究は、文部科学省科学研究費補助金基盤研究（C）「日本語作文支援システムで考慮すべき学習者属性情報と提示項目の分析研究」(研究代表者：阿辺川武、研究期間：2012 年 4 月～2015 年 3 月）および同補助金挑戦的萌芽研究「日本語学習者誤用コーパスを利用した作文システムの開発」（研究代表者：仁科喜久子、研究期間：2010 年 4 月～2013 年 3 月）による助成を得て実施しています。

## 参考文献

仁科喜久子、村岡貴子、因京子、Joyce Terence Andrew、鎌田美千子、阿辺川武（2011）「バランス・コーパス利用による日本語作文支援システム『なつめ』の構築と評価」特定領域研究日本語コーパス平成 22 年度公開ワークショップ（研究成果報告会）予稿集、pp.215-224.

Hodošček Bor、阿辺川武、Bekeš Andrej、仁科喜久子（2011）「レポート作成のための共起表現産出支援―作文支援ツール「なつめ」の使用効果―」専門日本語教育研究 13 号、pp.33-40.

曹紅荃、八木豊、黒田史彦、仁科喜久子（2012）「学習者コーパス「なたね」の構築と応用の可能性」第 5 回「日本語教育とコンピュータ」国際会議（Castel/J）

Halliday M.A.K. and C.M.I.M. Matthiessen (2004). An Introduction to Functional Grammar. 3rd ed. London: Arnold

仁科喜久子監修（2012）「日本語学習支援の構築　言語教育・コーパス・システム開発」凡人社

八木豊、ホドシチェク・ボル、仁科喜久子（2012）「BCCWJ と学習者作文コーパスを利用した日本語作文支援－表記と共起に関する誤用添削プロトタイプ構築－」第 1 回コーパス日本語学ワークショップ予稿集、pp.315-320.