

作文対訳データベースの多様な利用のために：「日本語教育のための言語資源及び学習内容に関する調査研究」成果報告書

著者	国立国語研究所，杉戸 清樹，宇佐美 洋，井上 優， 鏝水 兼貴，成田 高宏
ページ	1-163
発行年	2006-03-30
URL	http://doi.org/10.15084/00003706

作文対訳データベースの 多様な利用のために

「日本語教育のための言語資源及び
学習内容に関する調査研究」
成果報告書

独立行政法人

国立国語研究所

2006(平成 18)年 3 月

作文対訳データベースの 多様な利用のために

「日本語教育のための言語資源及び
学習内容に関する調査研究」
成果報告書

独立行政法人

国立国語研究所

2006(平成 18)年 3 月

刊行のことば

本報告書は、国立国語研究所で作成している「日本語学習者による日本語作文と、その母語訳との対訳データベース」（略称：作文対訳 DB）について、その設計思想を述べるとともに、具体的にデータベースを活用することによって得られた研究成果をまとめたものです。

このデータベースは、学習者の言語データを大量に集積した「学習者コーパス Learner Corpus」の一つとして位置づけられるものです。「学習者コーパス」は、言語教育、特に英語教育の世界において、学習者が使っている言語のあり方を客観的にとらえるための有力な手がかりとして、近年多大な注目を浴びています。しかし、日本語教育の世界においては、だれもが活用できる「社会共有の財産としての言語データ」というものがまだ少なく、当然そうした共有データに基づく研究も十分とはいえない状況にあります。国立国語研究所の「作文対訳 DB」は、このような状況に一石を投じ、共有言語データを用いた日本語教育研究を促進する役割を果たすものと考えています。

本データベースの CD-ROM 版は、2001 年の公開以来、国内外の研究者・教育関係者に延べ 1,000 枚以上を配布し、利用していただきました。また、オンライン版の使用登録者も、2006 年 3 月現在で約 340 名にのぼっています。本データベースを使用した研究成果も多数公表されていますが、本報告書はその中で、国立国語研究所の研究者が執筆した解説論文、研究論文を収録することとしました。

本報告書の企画・編集は、日本語教育部門 第 1 領域 主任研究員の宇佐美洋、ならびに研究補佐員の鏈水兼貴が主に担当しました。また、この 2 名のほか、第 1 領域 領域長の井上優、研究補佐員の成田高宏が論文を執筆しています。

本報告書が、対照言語学・文章論・作文教育等に関心を持つ研究者・教育関係者の皆さんの間で、広く活用されていくことを期待します。

平成 18 年 3 月

独立行政法人 国立国語研究所

所長 杉戸 清樹

作文対訳データベースの多様な利用のために

「日本語教育のための言語資源及び学習内容に関する調査研究」 報告書

研究の意義と本報告書の内容…………… 1

<概要編>

「作文対訳データベース」作成の目的とその多様な活用について
宇佐美 洋 ……………9

言語データとしての作文対訳データベース
井上 優 ……………43

「XMLによる作文添削情報表示システム」仕様の発展について
—「添削情報電子化」の真の意義を求めて—
宇佐美 洋・鎌水 兼貴 ……………53

<分析編>

フランス語母語話者の日本語作文における「意図不明表現」の分析
—母語訳との対照から見る「分かりにくさ」の理由—
宇佐美 洋 ……………81

第二言語としての日本語作文に対する記述式フィードバックと
教師ビリーフの関係
成田 高宏 ……………101

学習者日本語作文における「形態素情報」の活用
—「添削情報」との連携をめざして—
鎌水 兼貴 ……………121

学習者作文に対する教師コメントの分析
—実態の把握・分析と、そこから得られる提言—
宇佐美 洋 ……………145

「日本語教育のための言語資源及び学習内容に関する調査研究」報告書 作文対訳データベースの多様な利用のために

研究の意義と本報告書の内容

はじめに

国立国語研究所では1999年より、「日本語教育のための言語資源及び学習内容に関する調査研究」と名づけられたプロジェクトの一環として、学習者が日本語で書いた文章と、執筆者自身によるその母語訳、ならびに日本語作文の添削情報を収集し、データベースとして公開するという事業をおこなってきた。「日本語学習者による日本語作文と、その母語訳との対訳データベース」（略称：作文対訳DB）である。

このデータベースは、1990年代から特に注目を浴びるようになった「学習者コーパス Learner corpus」のひとつとして位置づけられるものである。学習者コーパスは、学習者が書いたり話したりする生の言語データに対し数量的・客観的な分析を加えることができるという点で、言語教育研究を科学的に進めていく際に極めて重要な貢献をなすものであるが、残念なことに日本語教育の世界においては、その重要性が十分に理解されてきたとは言いがたい。もちろん学習者の言語データを、数量的・客観的に分析しておこなわれた研究は多数存在するのであるが、それらの多くは教師・研究者が個別に収集したデータに基づいているため、他の研究者が研究のプロセスを再検討することができない。研究者どうし、そして研究者と教育者とが共通の基盤に基づいて対話を進めていくためには、「社会的に共有された共通のデータ」を作成し、公開していくことが必要になるだろう。そして、日本語教育の世界においてそのようなデータを整備していくことは、われわれ国立国語研究所の重要な責務のひとつであるといつてよい。

そしてわれわれは、このデータベースを作るにあたり、他の学習者コーパスにはない、ふたつの特長を持たせることにした。それは

- 1) 学習者本人による母語訳がついている
- 2) 添削情報がついている

という2点である。特に2)については(一部のデータについてのみであるが)、ひとつの作文に対し複数の添削情報を収集し、それらを比較対照できるようにしてあるところが大きな特色である。

通常、学習者コーパスは、言語データそのもの以外にさまざまな付加情報を持っている。例えば、学習者の言語歴・属性情報や、エラータグによる誤用情報などである。これらは、言語データそのものに直接付随した情報であるといえることができる。

一方でわれわれの「作文対訳DB」は、言語データそのものに対し、2つの異なった方

向から与えられた「注釈」および「評価・評定」のデータを備えている。上記 1)は、「書き手本人による注釈」であり、2)は「読み手による評価・評定」である。

この2方向からの注釈ないし評価・評定が与えられていることで、「作文対訳 DB」は従来の学習者コーパスとは異なる応用の方向性を指し示すことができている。

ひとつは、1)を使用することによって、「執筆者本人が書こうとした意図が推測できる」ということである。学習者の書いた作文に向き合うとき、われわれはしばしば、「いったい何を書こうとしているのかまるでわからない」という事態に出くわすことがある。慣れた教師であれば、文脈の情報などを駆使して学習者の意図をかなりの程度までくみ取ることが可能な場合もあるが、学習者は時に、教師がまったく予期しなかった考え違いをしている場合があり、そうした場合は書かれた言語データをいくら見てもなんら有益な知見は得られない。

そのような場合、1)の母語訳が効力を発揮する。母語訳を参照することで、執筆者本人がそこで何を書こうとしていたのかがわかり、なおかつ、執筆者がどのような思考過程を経てそのような「意味不明表現」を書いてしまったのか（母語の影響によるのか、そうでないのか）、ということを含めて、かなり正確に推測できる手がかりとなりうるのである。このことは、「どのような逸脱が起こっているか」だけでなく、「なぜそのような逸脱が起こるのか」について考察する手がかりが得られる、ということであり、これは教育上きわめて有用な情報となりうるものである。

さらに、今後の日本において有用性がさらに増していくであろうと考えられるのは、2)の「読み手による評価・評定」のデータである。

2003年、日本の大学等への外国人留学生数は10万人を超え、この数は今後もさらに増えていくことが予想される。そしてこれらの学生が日本に留学して直面する問題のうち、レポート・論文執筆など「書くこと」に関わる困難がかなり大きな部分を占めているといわれるのであるが、学習者の書いた文章に対し、教師はどのようなフィードバック（添削等）を返すのが効果的なのか、ということについての研究は十分におこなわれているとはいえない（それどころか、多くの場合作文添削は教師個人の経験と感覚に基づいておこなわれ、他の教師がどのような添削をおこなっているのかを知る機会さえないのが現状である）。ひとつの作文に対し、複数の教師からの添削情報を収集し、それを対照分析することは、「効果的添削とはどのようなものか」について考察するための重要な材料となりうるだろう。

一方で、日本に在住する「日本語を母語としない人々」がみな、日本語を学び、日本語を使って学ぶことを目的としているわけではない。今後はむしろ、何らかの別目的（それは、日本での労働であったり、地域社会での付き合いを円滑におこなうこと、であったりするだろう）を達成するために「日本語を使う」、という人々の数が増えていくことが予想される。そして日本語母語話者側も、好むと好まざるとにかかわらず、そのような人々とも日本語を使ってコミュニケーションをとり、共生していくことが求められることになる。

そのような状況において、日本語を学ぼうとする人々すべてに、日本語母語話者と同程度の日本語能力を求めていくことはできない。重要なのは、「完全ではなくても十分に意思疎通が成り立ち、なおかつ聞き手・読み手を不愉快にすることがない表現」とはどのようなものを調査するとともに、そのような表現を効率的に習得するためにはどうしたらよいか、その方法を考案して学習者側に伝えていくことであろう。

そのためには、規範から外れた表現をすべて「誤用」とひとまとめにし、「直すべきもの」というレッテルをはってしまうようなことは決して建設的な態度ではないだろう。それぞれの「規範逸脱表現」が、意思疎通においてどの程度の障害となっているのか、読み手にどの程度の違和感を与えているか、という、「分かりにくさの度合い」、「違和感の度合い」、といったものを把握し、それに適合した対応方法を考える、といったことをおこなっていかなければならない。その際、ひとつの作文に対する複数添削者の添削情報は、「規範逸脱表現に対する読み手の言語意識」を直接反映するものとして活用していくことができるのである。

本報告書の内容

本報告書は、以下の2部によって構成され、それぞれ以下のような目的を持っている。

第1部 概要編

「日本語学習者による日本語作文と、その母語訳との対訳データベース」のもつ意義を示し、本データベースの2006年3月現在における仕様について解説する。あわせて、このデータベースのその多様な応用方法を指し示すとともに、データを使用して分析をおこなっていく際の注意点についても述べる。

第2部 分析編

「日本語学習者による日本語作文と、その母語訳との対訳データベース」を用いた研究の事例を紹介する。

以下、この報告書に収録されている論文・論考の内容について略述する。

第1部の第1論文は、宇佐美洋の『作文対訳データベース』概要と作成の目的である。この論文では、本データベースの全体像を示し、従来の学習者コーパスとは異なる、このデータベースの独自の特長について述べるとともに、データベースを構成する各部門をどのような研究活動に使用していくことができるか、その多様な可能性について述べた。また、データの収集方法や、オンライン版データベースの使用法などについても触れ、データベース使用上のマニュアルとしても使用できるようにした。

第2論文は、井上優の「言語データとしての作文対訳データベース」である。この論文

は、言語の対照研究の観点から、本データベースの言語データとしての性格について考察したものである。特に、学習者本人による母語訳を言語研究のための一つのデータとして用いようとする際には、次の二点を十分にふまえる必要があることを述べた。1) 作文対訳 DB に収録された「日本語学習者による日本語作文」と「作文執筆者本人による母語訳」は、いずれも十分に推敲された文章ではない。2) 「作文執筆者本人による母語訳」は、日本語作文を翻訳した文章であり、その言語の表現として必ずしも自然とはいえない表現も含まれている。

第3論文は、宇佐美洋・鍵水兼貴の『XMLによる作文添削情報表示システム』仕様の発展について－『添削情報電子化』の真の意義をもとめて－である。この論文は、添削情報を電子化してデータベースに収録するために開発された「XMLによる作文添削情報表示システム」が、バージョン1からバージョン3にかけてその仕様をどのように発展させてきたかを紹介するとともに、仕様の発展にともない「作文情報を電子化する」ということの意味がどのように変化してきたか、ということについて論じたものである。

第2部の第1論文は、宇佐美洋の「フランス語母語話者の日本語作文における「意図不明表現」の分析－母語訳との対照から見る「分かりにくさ」の理由－」である。この論文は、フランス語母語話者が書いた日本語文章の中で、「日本語母語話者が読んでも意味が取れない部分(意図不明表現)」を取り出し、その部分に対応する母語訳と対照することによって、「フランス語母語話者が日本語で文章を書く際、どういう要因によって意図不明表現が起こりやすいのか」ということを考察した。その結果、フランス語母語話者の意図不明表現は、フランス語の抽象名詞をそのまま日本語に置き換えようとする時に最も生じやすい、ということが明らかになった。

第2論文は、成田高宏の「第二言語としての日本語作文に対する記述式フィードバックと教師ビリーフの関係」である。この論文では、1) このデータベースに収録された日本語作文に対する教師の記述式フィードバックの内容と、2) フィードバックをおこなった教師の作文教育に関するビリーフや日本語教師としてのビリーフ等に関する調査結果、の両者を付き合わせ、1)と2)の間にどのような関係があるかを分析した。その結果、文章構成に関する指導の大切さをほとんどの教師が認めているが、添削という形ではその指導がおこなわれにくいこと、コメントを記述することの重要性をすべての教師が感じていながら、実際には、コメント記述を多用する教師と行わない教師とに二分されることが分かった。これらの結果から、日本語作文教育支援者による今後の取り組みとして、①作文の添削(特に構成について)をおこない易くするツールの開発 ②教師が作文教育の知識を習得するためのリソースの開発 が求められることを述べた。

第3論文は、鐘水兼貴の「学習者日本語作文における『形態素情報』の活用—『添削情報』との連携をめざして—」である。この論文では、「対訳作文データベース」に収録された異なる種類の情報を組み合わせることによって、学習者作文の特徴を探るための方法を提示した。

まず、作文に対して形態素情報を付与した「形態素情報データベース」を活用して、「形態素の連鎖」という観点からの分析を試みた。学習者においてのみ出現頻度が高い（もしくは少ない）形態素連鎖があった場合、それは学習者にとって習得しにくい表現であったり、日本語母話者に違和感を与える表現であったりする可能性がある。

さらに、「形態素情報データベース」の量的分析によって得られた「学習者特有の形態素連鎖」が、日本語教師によってどのように修正されているかを見るために、データベース内の添削情報を使用する方法を示した。検索ツール「添削対照検索」を用いることにより、添削前の原文（「誤って使用している箇所」）からだけでなく、添削がなされた後の文章（「本来使用しなければならない箇所」）の両面から検索することが可能となった。これにより、経験のみによっては気づかれにくかった学習者作文の特徴を発見することができることを示した。

第4論文は、宇佐美洋の「学習者作文に対する教師コメントの分析—実態の把握・分析と、そこから得られる提言—」である。この論文は、作文教育経験の長い教師と短い教師の間で、学習者作文に対する「コメント」の書き方に違いがないかを比較したものである。分析の材料としては、このデータベースに収録された作文に対し教師が返したフィードバック情報を用いた。作文に対して書かれたコメントを一定の基準に基づいて分類したところ、経験の短い教師は経験の長い教師に比べ、「具体性に欠けるコメント」を書くことが明らかに多いことが分かった。一方で、「記述内容が妥当でないこと」を指摘するコメントは作文教育経験の多寡にかかわらず少なく、記述内容をクリティカルに読んでいるかどうかということと教育経験とは、あまり相関がないということが示唆された。

ここから、教師のフィードバックに当たっては、学習者の文章をクリティカルに、分析的に読もうとする意識が必要なこと、本人の言いたいことが書けていればよい、という「書き手本位の指導法」を、「書き手に読み手の存在を意識させる指導法」へと深化させていくことが必要であること、の2点を主張した。

今後は日本語教育の世界においても、学習者の言語データを大量に収集し、共有の研究資料として公開した学習者コーパスの必要性はさらに高まるものと考えられる。われわれのデータベースおよびこの報告書が、学習者コーパスの新しい活用方法について議論するためのきっかけとなれば幸いである。

概要編

「作文対訳データベース」作成の目的とその多様な活用について

宇佐美 洋

0. はじめに

本稿は、二つの部によって構成される。

第1部<設計思想編>では、国立国語研究所で作成している「日本語学習者による日本語作文と、その母語訳との対訳データベース」(略称「作文対訳DB」)の全体像を紹介し、データベース作成の基本方針を述べるとともに、データベースの具体的な活用方法と今後の発展計画について論じる。

第2部<資料編>は、データベース拡張の歴史、データ収集の方法等を記録として残すとともに、オンライン版データベースの具体的な使用方法について略述する。

第1部<設計思想編>

1. データベース全体像

このデータベースは、

- 1) 日本語学習者が書いた日本語作文(300~800字程度)
- 2) 執筆者本人による1)の母語訳(または最も楽に文章が書ける言語への翻訳)

を主たる構成要素としており、さらに参考資料として

- 3) 日本語母語話者が書いた日本語作文
- 4) 1)の添削情報(ただし一部のみ)¹
- 5) 執筆者・添削者の言語歴情報

を含んでいる。またこのほか、

- 6) 1), 3)を形態素に分割し、形態素情報を付したもの

を「派生データベース」として現在開発中である(2006年時点で未公開)。

データベースの全体像を図示すると、以下のようになる。

¹ 基本的に、1編の作文に対し1,2名の添削者のみが添削を施しているが、一部の作文データについては20~30名程度の添削者に添削を依頼しているものもある。

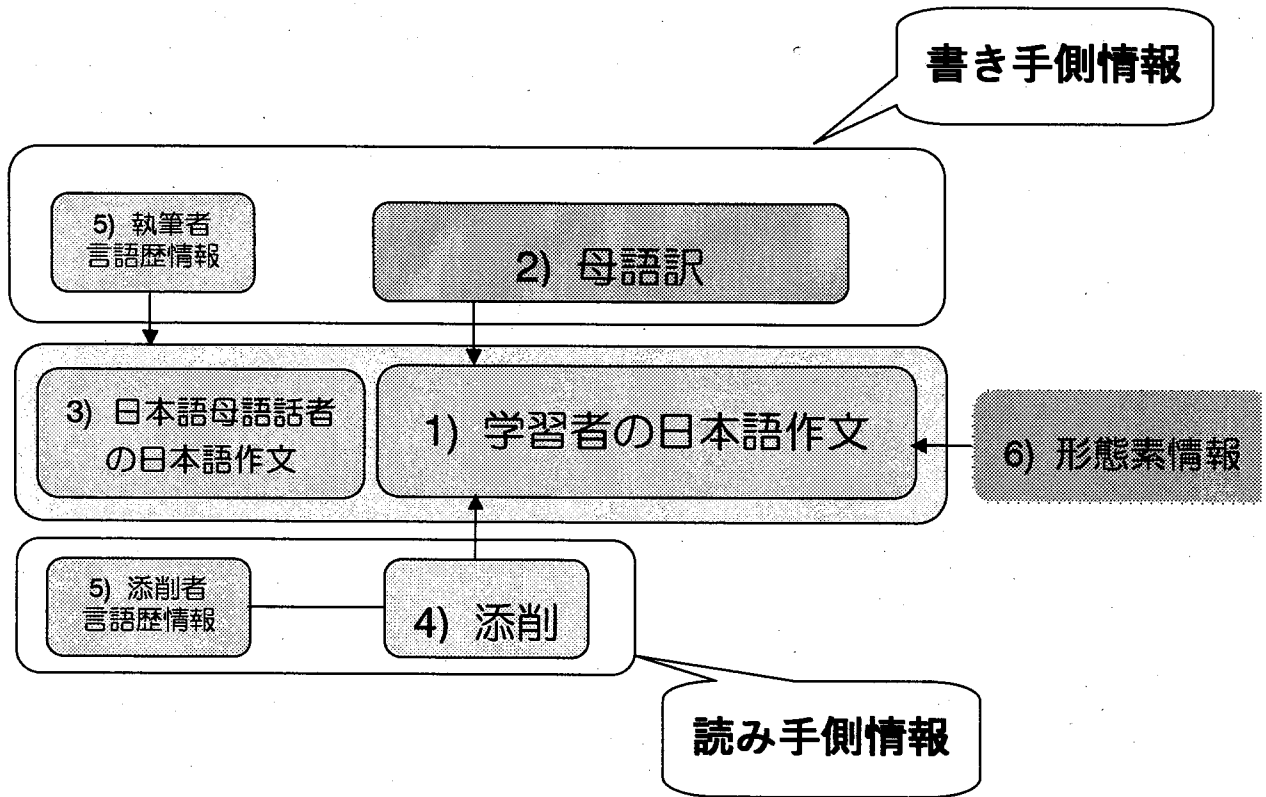


図1 データベース全体像

このデータベースの中心には、1)「学習者の書いた日本語作文」が位置する。そしてこのデータに対応する2)「執筆者本人による母語訳」は、1)に対し、

A) 書き手本人が注釈を施したものの

と考えることができるだろう。

また、4)の「添削情報」は、1)に対し、

B) 読み手から提示された反応・評価

と見ることができ、さらに5)は、

C) 言語データそのものに対する分析情報

である。つまりこのデータベースは、「学習者作文」に対し、上記3つの観点からの付加情報が付与されるという構成を持っていることになる。

では、われわれはなぜこのような形のデータベースを作成することにしたのか。データベース設計の背後にある基本思想について、他の学習者コーパスとも対比しつつ、以下で述べていくこととする。

2. 「学習者コーパス Learner Corpus」

日本語学習者に対する「書くこと」の教育改善をおこなっていく際、考察の材料として、まずは、学習者の書いたデータそのものが大量に必要であることは言うまでもない。学習者が書く日本語文章の中に、どのような問題点が見られるのか、それらの問題点について、どのように対処していくのがよいのか、また問題点を未然に防ぐにはどのような教育方法が効果的なのか。このようなことについて考えるための出発点は、すべて学習者の産出したデータそのものの中にあるとあってよい。

こうした考えに基づき、1990年代ごろから、特に英語教育の世界においては、「学習者コーパス Learner Corpus」というものが着目を浴びてきた。学習者コーパスとは、第2言語学習者が産出した文章・発話データを大量に収集し、必要な情報を付加し、さまざまな検索が可能な形で電子化したものである。学習者コーパスは、学習者が産出した大量の生データに対し、客観的・数量的な、したがってより信頼性の高い分析を可能にするとともに、主観的経験から得られる知見とはまた異なった知見を提供してくれるものとして、さまざまな方面からの期待を受けている。

英語教育の世界においては、作文データだけに限ってみても²、以下のような学習者コーパスが作成されている。

*Longman Learners' Corpus(LLC)

<http://www.longman-elt.com/dictionaries/corpus/lclearn.html>

英語学習者コーパスの先駆的存在。母語数 20, コーパスサイズは 1,000 万語。

*International Corpus of Learner English(ICLE)

<http://www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/Cecl-Projects/Icle/icle.htm>

サイズ約 200 万語, 母語数 19。品詞タグ, エラータグが付与されている。

*Cambridge Learner Corpus(CLC)

<http://www.cambridge.org/elt/corpus/clc.htm>

母語数 75, サイズは 1,500 万語。うち, 500 万語分にはエラータグが付与されているが, Cambridge 社外には非公開。

² このほか英語学習者の話しことばを対象とするものとしては、

- ・ LINDSEI(<http://www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/Cecl-Projects/Lindsei/lindsei.htm>)
- ・ The ISLE Corpus of non-native spoken English(<http://nats-www.informatik.uni-hamburg.de/~isle/speech.html>)
- ・ NICT JLE Corpus(<http://www.alc.co.jp/edusys/sst/corpus.html>)等が存在する。

*JEFLL(Japanese EFL Learner) Corpus

日本の中学1年～高校3年までの英作文データ約50万語分を収集。品詞タグ、エラータグが付与されている。小学館コーパス・ネットワーク(<http://www.corpora.jp>)で2006年春公開予定。

*HKUST Learner Corpus

香港科技大学(The Hong Kong University of Science and Technology)の入学試験および定期試験の英作文を収集したもの。コーパスサイズは2,500万語、母語は中国語。

一方日本語教育の世界においては、以下のような学習者作文コーパスが作成されているが、残念ながら学習者コーパスを用いた研究が日本語教育において十分根付いているとはいえない状況³である。

*寺村誤用データ

寺村秀夫(1990)「日本語学習者の日本語誤用例集」(科学研究費 特別推進研究「日本語の普遍性と個別性に関する理論的及び実証的研究」, 分担研究「外国人学習者の日本語誤用例集、整理及び分析」資料。執筆者は延べ339人, 執筆者の国籍は20か国にわたり, コーパスサイズは約420KB。 <http://cookie.lang.nagoya-u.ac.jp/pub/>で公開されている。

*日本語学習者の作文コーパス (名古屋大)

科学研究費 「日本語学習者の作文コーパス: 電子化による共有資源化」(96-98年度 代表者: 大曾美恵子)(未公開。ただし一部のデータのみが、寺村誤用データと同じく <http://cookie.lang.nagoya-u.ac.jp/pub/>で公開されている)

3. 本データベースの独自性

3.1. 母語訳の存在

こうした学習者コーパスは、学習者が書く文章の特徴を数量的にとらえるにあたっては極めて大きな効果を発揮する。しかしながら、単なる「特徴の分析」ととどまらず、「書くこと」の教育改善について真に実り多い成果をあげていくためには、学習者の書いた日本語データを参照するだけでは不十分である、とわれわれは考えた。そのもっとも大きな理由のひとつは、文面を読むだけでは、書き手が何を書こうとしたのか、そしてなぜそういう表現をとったのか、その意図が想像できない場合が少なくない、というところにある。

³ 日本語教育における話しことばの学習者コーパスとしては「KYコーパス」がある。このコーパスは、OPIテスターによる学習者レベル判定が付与されていることもあって、とくに言語習得に関する研究では広く用いられているものであるが、書きことばの分野において「KYコーパス」ほどに普及している学習者コーパスはまだ存在しない。

教師が、学習者の書いた文章に問題点を見つけることができたとしても、書き手の意図が分からなければ、その問題点にどのように対処していくのがよいか決めようがないし、さらにその問題点を未然に防ぐことなどではしない。

ここで問題になるのは、書き手の「真の意図」なるものをどのような手段によって知ることができるか、ということである。言語として表現される以前の「意図」は、学習者に対する綿密なインタビューによってでなければ推測できない、という考え方もあるだろう。しかし、インタビューによっては大量のデータを扱うことは困難であるため、考察に使用できるデータ数はどうしても限られたものになってしまう。

そこでわれわれは、学習者に日本語で文章を書いてもらうとともに、学習者本人に、その母語訳（またはもっとも楽に文章を書ける言語への翻訳）を作成してもらい、日本語作文と母語訳とを対照する形でデータベースを作成する、ということ考えた⁴。

もちろんこのやり方が、学習者の意図を把握するための十全の方法、または最善の方法であるというわけではない。母語訳として表現された内容は、学習者の「真の意図」を近似的に示しているに過ぎないだろう。しかし近似的であるにせよ、学習者の意図を推測するための「手がかりが得られる」というところに、研究を前進させるための契機があるとわれわれは考えた。

3.2. 添削情報の存在

さらにわれわれは、「書くこと」の教育改善に資するためには、誤用や不自然表現を多数含む学習者の文章に対し、日本語母語話者が（特に日本語教師が）、どのような評価意識を持ち、どのような反応を返すのか（どのように修正を施すのか、どういう箇所に対して、どのようなコメントを書くのか）、というデータも必要であると考えた。そこで、学習者の日本語作文に対し、教師に添削を依頼し、添削データも電子化してデータベースに収録することとした。

作文指導においては、学習者の書いた作文に対し教師が修正を施したりコメントを書いたりする、ということがおこなわれているのであるが、通常こうした添削行動は教師個人の感覚と経験に基づいておこなわれており、他の教師がどのような添削行動をおこなっているかさえ知る機会はあまりない。大量の添削情報を収集・公開し、そこに見られる添削態度を分析することは、効果的添削とはどういうものかを考える際、極めて有効な示唆をあたえることになるだろう。

またわれわれは、「教育改善」に直結する成果を得ていくためには、学習者言語データそのものに対して客観的情報を付与していくほか、学習者の作文に対し読み手が「どう反応し、どう評価するか」という、ある意味で主観的なデータも収集する必要があると考えた。

⁴ 学習者コーパスに執筆者本人による母語訳を付す、というアイディアは、国立国語研究所の故・中野洋氏が1996年から98年にかけて作成した、『日中作文コーパス』において初めて導入されたものである。われわれの『作文対訳DB』は、先行コーパスというべき『日中作文コーパス』から設計思想の多くを受け継いでいることを付記しておく。

「添削」というのはまさに、教師の主観的反応・評価のあらわれであるといえる。

従来の学習者コーパスには、「誤用情報」が付与されているものが少なくない。つまり、一定の客観的基準によって「誤用」と認定された箇所をマークし、それを正しく直した形にするとともに、それが「どういう種類の誤用であるか」という情報を示しているのである。

こうした「誤用情報」は、「添削情報」と似ているようであるが、実はかなり異質なものである、ということを最初に確認しておかねばならない。「誤用情報」は、学習者データそのものに直接付与された「一般言語学的情報」であるのに対し、「添削情報」とは、学習者データを読み手がどう解釈し、どう反応したかという、一種の「社会言語学的情報」なのである。

前述の Cambridge Learner Corpus においては、例えば以下のような形での「エラータグ」が付与されている。

In contrast <#MT>to</#MT> these advantages...

上記で、<#MT> </#MT>は「前置詞の脱落」をあらわすタグである。つまり、必要な前置詞を落としてしまったときに、書かれなかった前置詞（ここでは to）を<#MT> </#MT>で囲んで示す、というわけである。

2004年に公開された NICT JLE Corpus については、「エラータグ付与ガイドライン」が書籍⁵の一部として公開されており、これによってタグの全貌を把握することができるようになっている⁶。ガイドラインによると、エラータグは品詞によって分類され、全部で47種類が準備されている。そのうち、例えば「名詞の誤り」(ln)は、1) 活用の誤り[inf], 2) 単複の誤り[num], 3) 核の誤り[cs], 4) 可算・不可算名詞の使い分けの誤り[cnt], 5) 補部の誤り[cmp], 6) 語彙選択の誤り[lxc], の6種類に下位分類され、例えば「名詞活用の誤り」は以下のように表現される。

How many <n_inf crr="children">childs</n_inf> do you have?

タグ内の「n_inf」という部分によって「名詞の活用の誤り」であることが表示されており、またタグで囲まれた childs という文字列が修正前の形を、crr=" "の" "で囲まれた文字列(children)が修正後の形であることを示している。

こうしたタグがつけられていれば、誤用を種類別に抽出することが可能となるため、例えば学習者の言語レベルごとに起こりやすい誤用の変遷を追ったり、あるいはある種類の

⁵ 和泉絵美・内元清貴・井佐原均(2004)『日本人1200人の英語スピーキングコーパス』、アルク

⁶ NICT JLE Corpus は日本語を母語とする英語学習者の話しことばを対象とするものであるが、同じく日本語を母語とする英語学習者の書きことばを収集した JEFLL Corpus(2006年公開予定)でもほぼ同様のタグが付与されるようである。

誤用発生率が言語レベルが上がるにしたがってどのように変わっていくかを見たり、といった研究が可能となる。

しかし一方で、このようなタグ付けができることには、当然のことながら限界もある。

もっとも大きな問題点は、こうしたタグ付けによっては、語彙的・文法的な誤りしか扱うことができない、ということである。

ひとくちに誤用、といっても、誰が見てもおかしい、と感じられる「明らかな誤用 absolute errors」と、間違いとはいえないがなんとなくおかしい、と感じられる「好まれない形態 dispreferred forms」とがあるものと考えられる (Ellis & Barkhuizen 2005:59)⁷。「好まれない形態」と判定するかしないかの判断には、たぶんに判定者の主観が混じってしまうため、学習者コーパスにおいて「誤用情報」として付与されるのは、通常前者の「明らかな誤用」だけに限られることになる。

しかしそうした「明らかな誤用」が、「好まれない形態」に比べ、より深刻な誤用であるとはいえない。タグ付けが可能な「明らかな誤用」というのは、少なくとも判定者にとっては「おかしさの理由」がきちんと分析・分類できた誤用、ということになるが、コミュニケーション上、あるいは作文指導上本当に深刻なのは、「なぜおかしいのかさえ分からない誤用」なのであって、分類可能な誤用項目を扱うことは、実はそうむずかしいことではないのである。

また、いかに詳細な誤用認定マニュアルを作成したとしても、誤用の判定・分析にはどうしても個人差は付きまとう。

「名古屋大 日本語学習者の作文コーパス」では、言語習得の世界で広く用いられている「JCHAT フォーマット」を用いて、以下のような形で誤用分析例が表示されている。

```
*GAK: <八月に日本はずいぶんあつかったです> [*1] 。
%err: [1 八月に日本はずいぶんあつかったです = 八月の日本はずいぶんあつかったです];
%als: [1 格助詞の間違い。「八月」は「日本」を修飾しているので、格助詞「に」ではなく、連体助詞「の」が適切。];
```

< >[*1]で囲まれた部分が「誤用箇所」であり、%err:の行ではその訂正が、そして%als:の行では誤用の解釈が示されている。

ここで分析者は、「八月に日本は」という部分を「八月の日本は」と修正している。もちろんこの修正自体には何の問題もないが、これが唯一の修正候補というわけではない。そもそも人によっては、「八月に日本はあつかった」でも特に違和感はない、という反応さえ十分にありえよう。

また、『八月』は『日本』を修飾している」という解釈がなされているのであるが、これが唯一の解釈であるというわけでもない。むしろここでは、「八月」という名詞は単独で

⁷ Ellis, R. & G. Barkhuizen(2005) *Analysing Learner Language*, Oxford University Press.

(助詞なしで) 副詞的にも使えるにもかかわらず、一般名詞と同様に助詞を接続させてしまった、という解釈も成り立つであろう(この場合の修正は、「八月、日本はずいぶんあつかったです」となるであろう)。そのように解釈した場合、この誤用は「格助詞の間違い」としてしまっているのだろうか。むしろ、『副詞としても使用可能な名詞』の用法の間違いとしたほうが適切かもしれない。そして、いずれの解釈・いずれの修正がより適切なのかということ、を、一意に決めるすべはないのである。

もちろん、「誤用」判定基準を詳細に定義してマニュアルを策定し、そのマニュアルに基づいて判定をおこなうようにすれば、ある程度一貫性のある判定は可能になるかもしれない。しかしそもそも、何を誤用とみなすかの判断基準は個人個人によって異なるものである。われわれが独自に「誤用」判定基準を策定したとして、果たしてそれが日本語教育に携わる他の方々にとっても有用であるかどうかは保証の限りではない。

そこでわれわれは、学習者作文における規範からの逸脱に対し、統一的な基準によって誤用判定や解釈を下すことは避け、代わりに教師個人の判断による添削・コメントを、学習者作文に対し生の形で付してもらい、それをデータベースに組み込む、ということ考えた。

したがって修正の基準は添削者個人によって異なることになり、そのままの状態では比較は不可能となる。そこで一部の作文データについては、1編の作文について複数の添削者に添削を依頼し、添削情報の間での比較対象が可能となるようにした。そして、あるひとつの逸脱表現に対する「反応・評価」というものが、読み手によっていかに違うか、ということを示し、そのなかから教育に直結できる情報を得ていくことができるよう考えていくこととした。

通常添削・コメント執筆は原稿用紙に赤字で書き込む、という形でおこなわれるのであるが、添削情報を共有の研究資源として使用していくためには、当然のことながら何らかの形によって添削情報を検索可能な形で電子化しなければならない。このため、XMLを用いて添削情報を電子化し、ブラウザ上で表示するためのシステムも別途開発することとした(添削情報のXML化の使用については、本報告書に収録されている『XMLによる作文添削情報表示システム』仕様の発展について『添削情報電子化』の真の意義を求めて(宇佐美洋・鎌水兼貴, 本報告書 pp.53-78)において詳述する)。

4. データ形式

日本語作文はすべてテキスト化されているほか、学習者本人が原稿用紙に手で書いたものも画像(jpg)としても保存した。

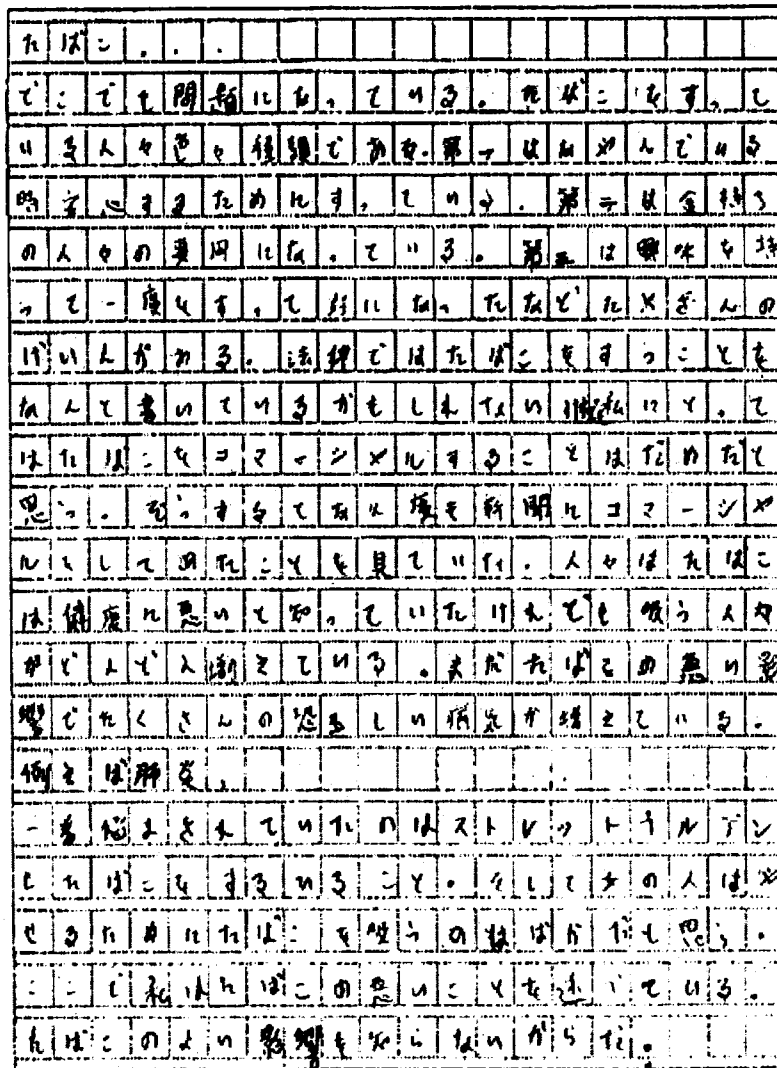
母語訳は、基本的にはUnicodeのテキストファイルとして保存されているが、公開の時点でUnicodeのテキストファイルとして保存するのが困難であった言語(カンボジア語、ヴェトナム語、モンゴル語など)については、pdfでの公開とした。またインドで収集した母語訳データ(ヒンディ語、ベンガル語、マラティ語等)については、手書き原稿の画

像ファイルが保存されている。

添削情報は、学習者が手で書いた原稿に対し、教師が直接赤字で添削情報を書き加えたものを画像として保存してあるが、前述のとおり、検索可能な形で電子化された XML ファイルも収録されている。

以下に、このデータベースに収録されているさまざまな形式のデータの実例を示す。例として示したのはモンゴル語母語話者のデータ例である。

1. 日本語手書きデータ(mn004j1.jpg)



2. 日本語テキストデータ(mn004j.txt)

たばこ…

どこでも問題になっている。たばこをすっている人々色々種類である。第一はなやんでいる時安心するためにすっている。第二は金持ちの人々の要用になっている。第三は興味を持って一度*をすって好になったなどたくさんあげいがある。法律ではたばこをすうことをなんと書いているかもしれないけれども私にとってはたばこをコマーシャルすることはだめだと思ふ。そうするとなん度も新聞*にコマーシャルをしていたことを見ていた。人々はたばこは健康に悪いと知っていたけれども吸う人々がどんどん増えている。まだたばこの悪い影響でたくさん恐ろしい病気が増えている。例えば肺*炎、一番*悩まされていたのはストレットチ*ルデンもたばこをするひること。そして女の人はやせるためにたばこを吸うのはばかだと思ふ。ここで私*はたばこの悪いことを述べている。たばこのよい影響を知らないからだ。(以下略)

3. 母語訳 pdf データ(mn004m.pdf)

Тамхи... хаа сайгүй асуудал болж байна. Тамхи татаж байгаа хүмүүс янз бүрийн шалигаантай. 1-рт сэтгэл санаагаар унасан үедээ тайвширах гэж татаж байна. 2 -рт мөнгөтэй хүмүүсийн хэрэглээ болж байна. 3 -рт сонирхож нэг удаа татаж үзээд, татах дуртай болох гэх зэрэг олон шалтгаан бий. Хуулинд тамхи татах тухай юу гэж заасныг мэдэхгүй ч миний хувьд тамхи рекламдах нь хэрэггүй гэж бодож байна. Гэтэл хэд хэдэн удаа сонинд рекламдсан байсныг харсан. Хүмүүс тамхи татах эрүүл мэндэд муу гэдгийг ч мэдсээр байж. Тамхи татах хүн улам өссөөр байна. Мөн тамхины муу нөлөөгөөр аймшигт өвчин бий болсоор байна. Жишээлбэл: уушигны үрэвсэл. Хамгийн харамсалтай нь тэнэмэл тамхи татаж байгаа явдал. Мөн эмэгтэйчүүд турхын тулд тамхи татдаг нь тэнэг юм шиг санагддаг. Энд би тамхи муу хор хөнөөлийг бичлээ. Яагаад гэвэл тамхинд сайн нөлөө байхгүй учраас. Мөн тамхины үйлдвэрлэл нь тэр улсын ганц үйлдвэрлэл эхэсэрт нь болж байгаа улс ч бий.

Муу нөлөө их байгаа ч тамхины төрөл өсөж байна.

Бас тамхи татаж байгаа хүмүүс орчноо бохирдуулж байна. Татаж дуусаад тэр газраа хаячихдаг. Үүнийгээ хогийн саванд хийвэл дээр. Тамхи татаад ажлаа мартаж явдал ч бий. Тамхины буруу нөлөөг мэдээд, тамхи татахаа больсон хүмүүс бий. Тэр хүмүүсийн нэгэн ярианаар "Би тамхи худалдаж авсан мөнгөө цуглуулсан бол одоо 2 дөвхөр байшин, машин худалдаж авч чадах байсан" гэсэн байсан. Тамхи янз бүрийн нөлөө үзүүлж байгаа нь харагдаж байна. Амьдрал ажил, эрүүл мэнд зэрэг.

Миний дуулснаар бол тамхины уралдаан байдаг гэсэн. Жнь "Хэн хурдан тамхи татах вэ?" "Хэн нэг дор хичнээн тамхи татах вэ?" зэрэг тамхины рекорд (Ертөнцийн гайхамшиг номонд) бичигдсэн байгаа.

Бүгдээрээ тамхи татахаа больж, амар тайван амьдарцгаая!

5. XML 化添削データ(mn004ns.xml)ソース表示

```
たばこ…<cp/>
<note value="1 マス空けるようにとの、→がある"></note>どこでも問題になっている。たばこをすっている
人々<put value="は"/>色々<put value="な理由が"/><del>種類</del>ある。第一はなやんでいる時安心する
ためにすっている。第二は金持ちの人々の<rep value="必需品">要用</rep>になっている。第三は興味を持つ
て一度*<del>を</del>すって好<put value="き"/>になったなどたくさんのげ<put value="ん"/>いんがある。
法律ではたばこをすうこと<rep value="について">を</rep><rep value="何か">なんと</rep>書い<better
value="(てある)">ている</better>かもしれないけれども私にとってはたばこ<rep value="の">を</rep>コ
マーシャル<put value="を"/>することはだめだと思う。<rep value="そういえば">そうすると</rep>なん度も
新聞*にコマーシャルをしていた<rep value="の">こと</rep>を見ていた。人々はたばこは健康に悪いと知って
い<rep value="る">た</rep>けれども吸う人々がどんどん増えている。ま<rep value="た">だ</rep>たばこの
悪い影響でたくさんの恐ろしい病気が増えている。例えば肺*炎、<cp/>
<note value="1 マス空けるようにとの、→がある"></note>一番*悩まされていたのは<doubt value="ストリー
トチルドレン?">ストレットチ*ルデン</doubt>もたばこを<rep value="すって">する</rep>ひること。そして
女の人はやせるためにたばこを吸うのはばかだと思う。ここで私*はたばこの悪いことを述べている。たばこの
よい影響を知らないからだ。<cp/>
```

6. XML 化添削データ(mn004ns.xml)ブラウザ表示

たばこ…

どこでも問題になっている。たばこをすっている人々は色々な理由が種類である。第一はな
目

やんでいる時安心するためにすっている。第二は金持ちの人々の要用 必需品 になっている。第

三は興味を持って一度*をすって好きになったなどたくさんのげんいんがある。法律ではたばこ

をすうことを なんと 書いている かもしれないけれども私にとってはたばこ
について 何か (てある)

を コマーシャルをすることはだめだと思う。そうすると なん度も新聞*にコマーシヤ
の そういえば

ルをしていたこと を見ていた。人々はたばこは健康に悪いと知っていた けれども吸う人々が
の る

どんどん増えている。また たばこの悪い影響でたくさんの恐ろしい病気が増えている。例えば
た

肺*炎、

一番*悩まされていたのはストレットチ*ルデン ストリートチルドレン? もたばこをす
目 すって

ひること。そして女の人はやせるためにたばこを吸うのはばかだと思う。ここで私*はたばこの悪

いことを述べている。たばこのよい影響を知らないからだ。

5. データベース 各部門の活用例

次に、データベースを構成する各部門がどのような意義を持っており、どのような研究に使用することが可能かについて述べるとともに、これまでにこのデータベースを用いておこなわれた、研究所内外の研究のいくつかを紹介する（そのうち、本報告書に収録されている研究所員の論文については、タイトルを太字で示した）。

*学習者の日本語作文および日本語母語話者の日本語作文

言うまでもなく、これが本データベースの中核部分となるデータである。さまざまな母語・さまざまな学習背景を持つ学習者が、それぞれに「規範からの逸脱部分」を含む文章を書き、それらが大量に収録されている。母語ごとに、また学習背景ごとに、どのような点で逸脱が起りやすいか、また起りにくいかを分析することで、日本語学習・日本語教育を円滑に進めていくための手がかりが数多く得られるであろう。また短文ではなく、まとまった長さの「文章」が収録されているところから、「分かりやすい／分かりにくい話の運び方とはどのようなものか」、「文化により、好んで使用される論理展開に差異があるか」という研究が可能となるし、また「文章の評価方法」について検討するための材料ともなるであろう。

またそうした検討をおこなっていく際、「日本語母語話者が日本語で書いた文章」を比較の対象として使用することは、学習者の書く日本語文章にどのような傾向があるかを明らかにするための重要な手がかりとなることであろう。

二通信子(2001)「アカデミック・ライティング教育の課題—日本人学生及び日本語学習者の意見文の文章構造の分析から」(北海学園大学学園論集 第110号, 61-77)は、本データベースに収録された日本人学生・中国語母語話者・韓国語母語話者の意見文129編の文章構造を比較し、文章構造におけるそれぞれの問題点を明らかにするとともに、そこから得られる「アカデミック・ライティング教育への示唆」について論じている。

伊集院郁子・高橋圭子(2004)「文末のモダリティに見られる“Writer/Reader visibility”—中国人学習者と日本語母語話者の意見文の比較—」(『日本語教育』123号, 86-95, 日本語教育学会)は、中国人学習者と日本語母語話者の日本語作文データを比較し、文末モダリティ表現の使われ方の相違について論じた。その結果、中国語母語話者は、対話性が強く、読み手に働きかける機能を持つモダリティを、日本語母語話者より多用していることを明らかにした。

*母語訳

母語訳はもっぱら、「執筆者本人の真意」を確認するための手段として用いることができる。学習者の日本語作文の中には、そこで何を書こうとしていたのか、日本語表現を読むだけでは正確に分からない場合がしばしばあるが、そのようなとき、母語訳を参照することで本人の意図が明瞭に分かることは少なくないのである。また意図が分かることにより、

学習者はなぜここでこのような逸脱表現をしてしまったのか、その理由が推測できることもある。理由は、母語の干渉であることもあるし、そうでないときもある。母語訳を参照することで、「この誤用は母語の干渉によるものではおそらくない」という判断ができることもまた、母語訳の効能のひとつである。

宮原彬(2002)「日本語作文とその母語訳から見たベトナム語母語話者の日本語学習上の問題点—母語の干渉を中心に—」(長崎大学留学生センター紀要第10号 1-24)は、本データベースに収録されたヴェトナム語母語話者の日本語作文とそのヴェトナム語訳を対照することによって、日本語作文中の誤用がどのような原因によって生まれているか、そこに母語の影響がどのような形で現れているかを分類整理して示したものである。

本報告書に収められている宇佐美洋「フランス語母語話者の日本語作文における「意図不明表現」の分析—母語訳との対照から見る「分かりにくさ」の理由—」(pp.81-99)では、フランス語母語話者が書いた日本語文章の中で、「日本語母語話者が読んでも意味が取れない部分(意図不明表現)」を取り出し、その部分に対応する母語訳と対照することによって、「フランス語母語話者が日本語で文章を書く際、どういう要因によって意図不明表現が起こりやすいのか」ということを考察した。その結果、フランス語母語話者の意図不明表現は、フランス語の抽象名詞をそのまま日本語に置き換えようとする時に最も生じやすい、ということを明らかにした。

*形態素情報

これは、日本語作文を形態素のレベルに分解し、それぞれの形態素について基本形、活用形、品詞分類などの情報を付与したものである。作文データに形態素情報が付与されることにより、表層の文字列だけでなく、形態素情報に対しても検索をかけることが可能となる。例えば、作文データの中から特定のパターンの形態素連続を抽出したり、特定の品詞(例えば副詞・接続詞)だけを抽出してその出現のヴァリエーションを調べたりすることができるようになる(ただし、形態素情報付きの作文データは、「派生データベース」としての位置づけであり、2006年時点では未公開である)。

鏈水兼貴の「学習者日本語作文における『形態素情報』の活用—『添削情報』との連携をめざして—」(pp.121-143)は、「形態素の連鎖パターン」という観点から作文を分析した論文である。日本語母語話者の作文と学習者の作文を比較することによって、学習者においてのみ出現頻度が高い(もしくは少ない)形態素連鎖パターンを見つけ出し、なぜそのような違いが生じるのかということについて、さまざまな角度から分析をおこなった。

*添削情報

添削者の属性(日本語教育歴の多寡、教育に対するビリーフ等)によって、添削時の視点はかなり異なっていることが予想される。添削者の属性と添削態度との関連を調べることは極めて興味深い調査テーマであろうし、また「多くの添削者にとって指摘しやすい誤

用、指摘しにくい誤用」とはなにか、ということを明らかにすることは、教師教育にとって有効と考えられる。さらに、添削という行動において教師がどのような態度を取っているかを詳細に見ていくことは、作文指導において効果的な添削態度とはどのようなものかを検討する際に役立つであろう。

本報告書に収録された成田高宏の「第二言語としての日本語作文に対する記述式フィードバックと教師ビリーフの関係」(pp.101-120)は、1) このデータベースに収録された日本語作文に対する教師の記述式フィードバックの内容と、2) フィードバックをおこなった教師の作文教育に関するビリーフや日本語教師としてのビリーフ等に関する調査結果、の両者を付き合わせ、1)と2)の間にどのような関係があるかを分析した。

また、宇佐美洋「学習者作文に対する教師コメントの分析—実態の把握・分析と、そこから得られる提言—」(pp.145-163)は、作文教育経験の長い教師と短い教師の間で、学習者作文に対する「コメント」の書き方に違いがないかを比較したものである。作文に対して書かれたコメントを一定の基準に基づいて分類したところ、経験の短い教師は経験の長い教師に比べ、「具体性に欠けるコメント」を書くことが明らかに多いことが分かった。一方で、「記述内容が妥当でないこと」を指摘するコメントは作文教育経験の多寡にかかわらず少なく、記述内容をクリティカルに読んでいるかどうかということと教育経験とは、あまり相関がないということが示唆された。

上記2つの論文は、添削情報を分析することで、教育改善そのものに直接貢献しようとする目的を持つものであるが、添削情報は学習者の言語データそのものをより多角的に分析するための情報としても有用であると考えられる。

学習者の書いた文章に対し、その中に含まれている「誤用」を検索することは可能であるが、「本来～と書くべきであったのに、それがうまく書けていない」という箇所を検索することは通常不可能である。しかし、添削情報も電子化し、「修正後の文字列」に対して検索をかけることで、上記のような箇所を抽出することが可能となるのである。これは、学習者の誤用に関する研究を、さらに進める契機となりうるであろう。

前項で紹介した鎌水兼貴「学習者日本語作文における『形態素情報』の活用—『添削情報』との連携をめざして—」(pp.121-143)はまた、電子化された添削情報の新しい活用方法についても論じている。この論文では、検索ツール「添削対照検索」を用いることにより、添削前の原文（「誤って使用している箇所」）からだけでなく、添削がなされた後の文章（「本来使用しなければならない箇所」）の両面から検索することが可能となり、経験のみによっては気づかれにくかった学習者作文の特徴を発見することができることを示した。

6. データベース 今後の発展

6.1. 添削タグ付きファイル専用の検索ツールの開発

このデータベースの大きな特徴のひとつとして、学習者の言語データそのものに対する言語学的な情報が付与されていることのほかに、「添削情報」、つまり、学習者の言語デー

タを読んだ時の読み手の反応・評価のデータを含んでいることが挙げられる、ということ
を述べた。われわれはこれまでのところ、こうした「添削データ」を、検索可能な形で電子
化する方法の開発とその改良を続けて来ており、2006年の段階においてその方式はとりあ
えずの完成を見たといえる（詳細は本報告書所収の宇佐美・鎌水「『XMLによる作文添削
情報表示システム』仕様の発展について－『添削情報電子化』の真の意義を求めて－」参
照）。しかし、そのデータを十二分に活用し、数量的な分析をも遂行していくためには、そ
の研究をサポートするためのツールを開発していくことが必要となる。

例えば、ひとつの作文について複数の添削情報を収集し、電子化したとしても、それら
をすべて印刷して見比べる、ということをおこなっているのではせっかく電子化した意味
が薄い。ひとつの作文の任意の箇所について、何名程度の添削者が修正を施しているのか
（言い換えれば何名の添削者が添削タグを付与しているのか）、ということが視覚的に示せ
る必要がある。なおかつ、作文の任意の箇所について、ひとりひとりの添削者が具体的に
どのような修正を施しているのか、ということが見られるように（言い換えれば、作文の
任意の箇所について複数の添削タグ付きファイルをスキャンし、添削タグが付与されてい
る箇所のみを一覧表示できるように）する必要もあるだろう。

現在われわれは、こうした「添削タグ付きファイルの検索用ツール」の開発を進めてお
り、2006年度中には試作版の完成を目指したいと考えている。

6.2. 学習者作文に対する「評定データベース」の作成

これまでわれわれは、学習者の作文に対し、日本語教師からの添削情報を収集してきた。
しかし、「研究の意義と本報告書の内容」でも述べたように、学習者が日本国内において、
「日本語母語話者と円滑なコミュニケーションをおこなっていくために必要な文章能力と
はなにか」ということを明らかにしていくためには、日本語教師の反応・評価のデータを
収集するだけでは不足である。なぜなら日本語教師が学習者データに対して返す反応と、
一般の日本語母語話者が返す反応とはかなりずれている可能性があるからである。

日本語教師は不完全な学習者データには慣れており、そこから書き手の意図を読み取る
力は一般の日本語母語話者よりかなり高い。このことにより、「分かりにくい表現」に対す
る感覚は、一般の日本語母語話者とかなりずれている可能性が考えられる。

また、同じ教師の中でも、言語教育に対しどのようなビリーフを持っているかによって、
「規範を逸脱した形」に対する許容の度合いがかなり大きくずれていることが予想される。
「規範提示」ということに重きを置く教師は、規範からの逸脱に対し、一般の日本語母語
話者よりはるかに厳しい評価を下す可能性がある一方で、「コミュニケーションが成立する
こと」を重視する教師は、コミュニケーション上支障のない逸脱表現については、逆に一
般の母語話者より寛容な反応を示す可能性があるだろう。いずれにせよ、日本語教師全体
としてみた場合、学習者の書いた文章に対する反応にはかなり大きなばらつきが出てくる
ことが予想される。日本語教師を、日本語母語話者全体を代表するサンプルとみなすこと

はできないのだ。

したがって、特に日本語教育を専門とはしないごく一般の日本語母語話者が、学習者の書いた文章に対しどのような反応・評価を示すのか、というデータを別途収集する必要があるのだが、しかし「添削」という形で反応データを収集することは困難である。それは「添削」という作業が、時間と熟練を要する、かなり精神的負担の重いものであるからである。熟練した教師でさえ、600字程度の作文の添削に1時間以上の時間を要することはめずらしくない。そのような作業を、添削経験のない人々に依頼することはかなりむずかしい。

そこでわれわれは、日本語教育を専門としない方々には「添削」をお願いすることは避け、「評定」のみをお願いすることとした。つまり、学習者の文章を読んで「おかしい」、または「意味が分からない」と思ったところに下線を引き、「おかしさ」または「分かりにくさ」の度合いを3段階で評定してもらうことにしたのである。

これによって、一般の日本語母語話者が、学習者の言語データのどういう箇所に対して「おかしさ」、「分かりにくさ」を感じるのか、その言語感覚は、日本語教師の感覚とどの程度一致するのか、どの程度ずれているのか、ということ进行分析する手がかりが得られる。また、誰もがおかしい、分かりにくい、と感じる表現を抽出することで、「コミュニケーション上真に重篤な誤用」とはなにか、ということが明らかになってくるだろう。

われわれはすでに、実験的にはあるが、一般日本語母語話者の評定データの収集を開始している。これらのデータは、「学習者と、一般の日本語母語話者が、円滑で効率的なコミュニケーションをおこなっていくことを支援する」という目的での教育をおこなっていく際には、極めて有効な考察資料となるであろう。

第2部<資料編>

1. データベース拡張の歴史

このデータベースは、以下のような経緯を経て拡張されてきた。

<1999年度>

科学研究費研究「日本語教育のためのアジア諸言語の対訳作文データの収集とコーパスの構築」(平成11・12年度基盤研究(B)(2), 課題番号国11691041)が開始。「学習者による日本語作文」「その母語訳」「日本語作文の添削情報」という3種類のデータを収集することを確定し、アジア7か国(中国・韓国・タイ・マレーシア・ヴェトナム・シンガポール・インド)の研究者にデータ収集への協力を呼びかけ、データ収集を開始した。

<2000年度>

上記7か国から収集された約800編のデータを、「日本語学習者による日本語作文と、その母語訳との対訳データベース ver.1(試用版)」として、CD-Rの形で公開。またモンゴル・カンボジアから新規にデータを収集するとともに、ヴェトナム・シンガポールから追加でデータ収集。また参考データとして、日本語母語話者の作文データも収集。年度末には、99～00年にかけて収集した約1,100編の作文データを、「日本語学習者による日本語作文と、その母語訳との対訳データベース ver.2」としてCD-ROMの形で公開。

<2001年度>

国立国語研究所の独立行政法人化に伴い、日本語学習者の作文データ収集事業は、国立国語研究所日本語教育部門のプロジェクト「日本語教育のための言語資源および学習内容に関する調査研究」の一環として位置づけられる。

欧米諸国(フランス・ドイツ・ベルギー・フィンランド・オーストリア・ブラジル)でのデータ収集を開始。

また、作文添削情報をXMLにより電子化するシステムの開発に着手。収集済みの手書き原稿の電子化を開始するとともに、ひとつの作文に対する複数添削情報の収集と、その電子化を開始。

<2002年度>

「XMLによる作文添削情報表示システム」の概要について、日本語教育学会で発表するとともに、電子化済の添削情報を、「作文添削情報データベース(モニター版)」として公開。

欧米諸国(スロヴェニア、ハンガリー、ポーランド、アメリカ)でのデータ収集をおこな

うとともに、所外研究者から、インドネシアで収集したデータの提供を受ける。

<2003 年度>

14 か国 (ブラジル, ドイツ, オーストリア, ベルギー, アメリカ, フィンランド, フランス, ハンガリー, インドネシア, ポーランド, カンボジア, モンゴル, シンガポール, ヴェトナム) から収集した 624 名分, 662 編の作文データを, 「日本語母語話者による日本語作文と, その母語訳との対訳データベース オンライン版」として試験公開開始。(URL は <http://jisl01.kokken.go.jp/taiyakuDB/>。のち <http://srasvadi.kokken.go.jp/taiyakuDB/>) に変更)。オンライン版では, 各種データの並列表示が可能となり, また, 作文・執筆者の属性により作文を抽出する機能が付与された。

<2004 年度>

「XML による作文添削表示システム」を改良して ver.2 とした。
スロヴェニアで収集したデータをオンライン版データベースに追加。データ総数は 636 名分, 674 編)となった。

<2005 年度>

「XML による作文添削表示システム」を改良して ver.3 とするとともに, ひとつの作文に対する複数添削の収集を再開。

「日本語母語話者による日本語作文と, その母語訳との対訳データベース」オンライン版と CD-ROM 版とを統合することとした。統合の結果作文データ数は 1,565 件, 添削データ数は 787 件となった。

2. データ収集方法

以下に, このデータベースを構成する各部門のデータについて, その収集方法を略述する。

* 学習者による日本語作文および母語訳

日本語作文の収集に当たっては, 国内外の日本語教育機関に協力を依頼した。主として, 日本語学科を持つ大学で収集をおこなったが, 一部の国では私立の日本語学校にも協力を依頼した。

作文を執筆する学習者の条件としては, 「日本語で, 原稿用紙 2 枚程度の分量の文章を自力で書けること」を指定したため, 結果として「大学 2 年次後半以降」の学生に執筆を依頼することとなった。執筆者本人の学習歴は調査票に記入してもらったが, 自己申告によるものなので学習時間のデータは必ずしも正確なものではない。

基本的には授業の中で作文を書いてもらうようお願いしたが, それぞれの機関の事情に

より、一部は宿題として書いてもらったデータもある。

データ収集の手順は以下のとおりである。

- 1) 国立国語研究所が指定した作文課題の中からひとつ(あるいは複数のもの)を選択し、まず日本語により作文を書く。課題の選択は、執筆者本人がおこなうこともあれば、授業運営の都合上、教師が選択・指定することもあった。執筆時には辞書を使用してもかまわないこととしたが、辞書使用の有無は申告してもらった。
- 2) 日本語作文が完成した後、執筆者本人が、その作文を見ながら母語(あるいはもっとも楽に文章が書ける言語)への翻訳をおこなった。翻訳は、日本語で書いたことをそのまま母語に移し変えるようにし、大幅な意識や大意要約にはしないよう依頼した。翻訳は、宿題として自宅でおこなってもかまわないこととした。ただし韓国のように、自宅に日本語のできる家族がいると想定できる国では、日本語の作文を添削してもらって提出するということが考えられたため、いったん日本語作文を回収してコピーをとり、そのコピーを見ながら翻訳をおこなうようお願いした。

* 日本語母語話者による日本語作文

日本国内の大学学部在籍していた日本語母語話者に執筆を依頼した。学歴・年齢的には、日本語学習者とほぼ同程度と考えられる。

ただし、作文執筆依頼時に提示した「課題文」は、学習者に提示したものとやや異なるので注意が必要である(課題文はデータベースで参照可能である)。

* 日本語作文に対する添削

公開する添削情報は、3期にわたって収集したものである。それぞれの期における添削情報収集方法は、以下のとおりである。

<第1期：2000年>

「作文対訳DB ver.2 (CD-ROM版)」に収録されている学習者の日本語作文約1,000編のうち、約300編について、日本語教師に添削を依頼した。この期においては、「日本語母語話者教師による添削」と、「日本語非母語話者教師による添削」を収集した。

「日本語非母語話者教師による添削」は、国外の日本語教育機関に勤務しており、基本的には作文を執筆した学習者を直接教える立場にある方々にお願いした。したがって学習者を直接知っており、その母語を理解することもできる方々である。

「日本語母語話者による添削」も、一部は学習者を直接教えている方々に依頼したが、それだけでは数が足りないため、日本国内の日本語教師(したがって学習者のことは直接知らない)にも依頼した(「添削者情報」の中の「現在の勤務地」が国外となっている方々が、学習者を直接知っている方々であると考えてよい)。

<第2期：2001年>

「作文対訳 DB ver.2 (CD-ROM版)」に収録されている学習者の日本語作文のうち20編に対し、23名の日本語母語話者、1名の韓国語母語話者に添削を依頼した。23名の日本語母語話者のうち、日本語教育経験者は15名、非経験者は8名であった。1名の韓国語母語話者は日本語教育経験者である。

全員が当時日本国内に在住しており、学習者のことは直接知らない立場にある。添削時には、「遠隔地に住む学習者から突然添削を依頼されたと想定し、1回限りの添削で最大の教育効果が出るように添削してほしい」と依頼した。

<第3期：2005年>

「作文対訳 DB オンライン版」に収録されている学習者の日本語作文のうち10編について、日本語教育経験者33名に添削を依頼した。このうち、作文教育歴3年以上の教師が15名、3年未満の教師が18名であった。

この期においては、添削対象の作文を「自分の意見を、分かりやすく、かつ説得力をもって伝える文章を書く」という目的の作文授業で書かれたものと想定し、その目的にそって添削してほしい、と依頼した。また前回同様、添削が可能なのは1回限りであり、1回限りの添削の中に必要と考えられる情報を可能な限り盛り込んでほしいと依頼した。

このうち、第1期と第2期に収集した添削情報は、2002年に「添削情報データベース モニター版」という形で限定配布をおこなっている。2005年度の「正式公開オンライン版」においては、第1期、第2期の添削データをオンライン版で一括公開することとした。第3期のデータについては2006年4月現在未公開だが、こちらも近日中に公開することを計画している。

3. オンライン版データベースの使用方法について

以下、「作文対訳 DB」オンライン版の使用方法について概要を述べる。

3.1. 利用登録

このデータベースは、「使用規約」に同意していただける方であれば、どなたにでも無料で使用していただくことが可能であるが、利用に先立って「ユーザ登録」をしていただく必要がある。利用登録は、「使用規約・ユーザ登録」のページ⁸の指示に従い、使用者名・Eメールアドレスを送信することによりおこなわれる。すぐに、登録されたEメールアドレスにパスワードが自動送信され、その「Eメールアドレス」と「パスワード」の組み合わせが登録者認証のために用いられる。

⁸ http://srasvadi.kokken.go.jp/taiyakuDB/kiyaku_touroku.html

3.2. 作文の抽出

3.2.1. 基本的抽出

「作文抽出・表示」のページ⁹に接続し、登録した「E メールアドレス」と「パスワード」を入力すると、以下のような「作文抽出インデックス」が開く¹⁰。

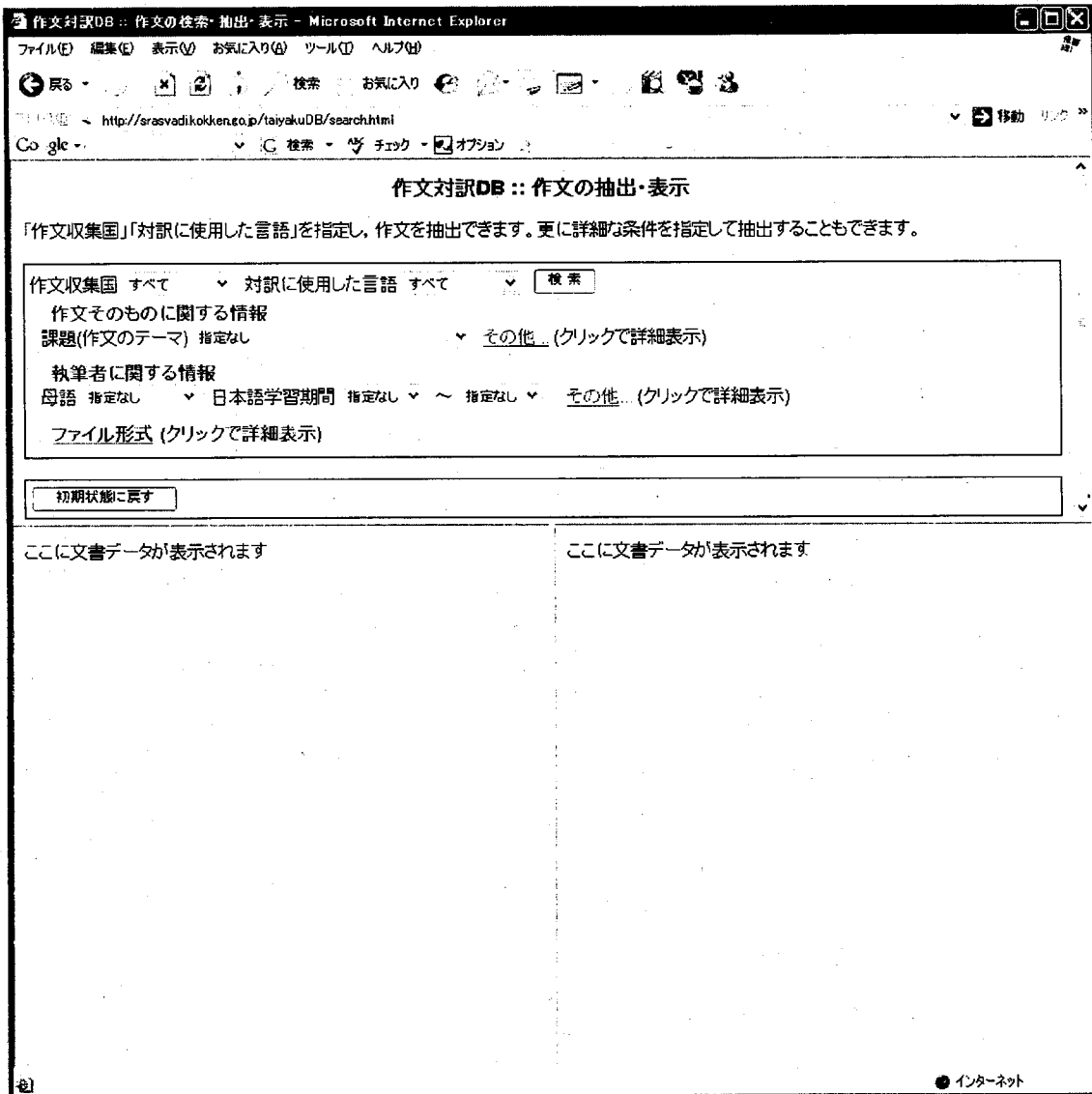


図1 「作文抽出インデックス」初期画面

その上で、「抽出したい作文の条件」「執筆者の条件」をプルダウンメニューから選択する。例えば、「カンボジア」で収集した作文で、対訳言語としては「カンボジア語」を使用しており、さらに「外国からの援助について」という課題について書かれたものを抽出し

⁹ <http://srasvadi.kokken.go.jp/taiyakuDB/search.html>

¹⁰ 以下の画面は、すべて2006年3月時点における、公開前のインデックスの状態である。操作法または使用の細部は、今後予告なく変更される可能性がある。

たい場合は、以下の画面のように条件を指定する。

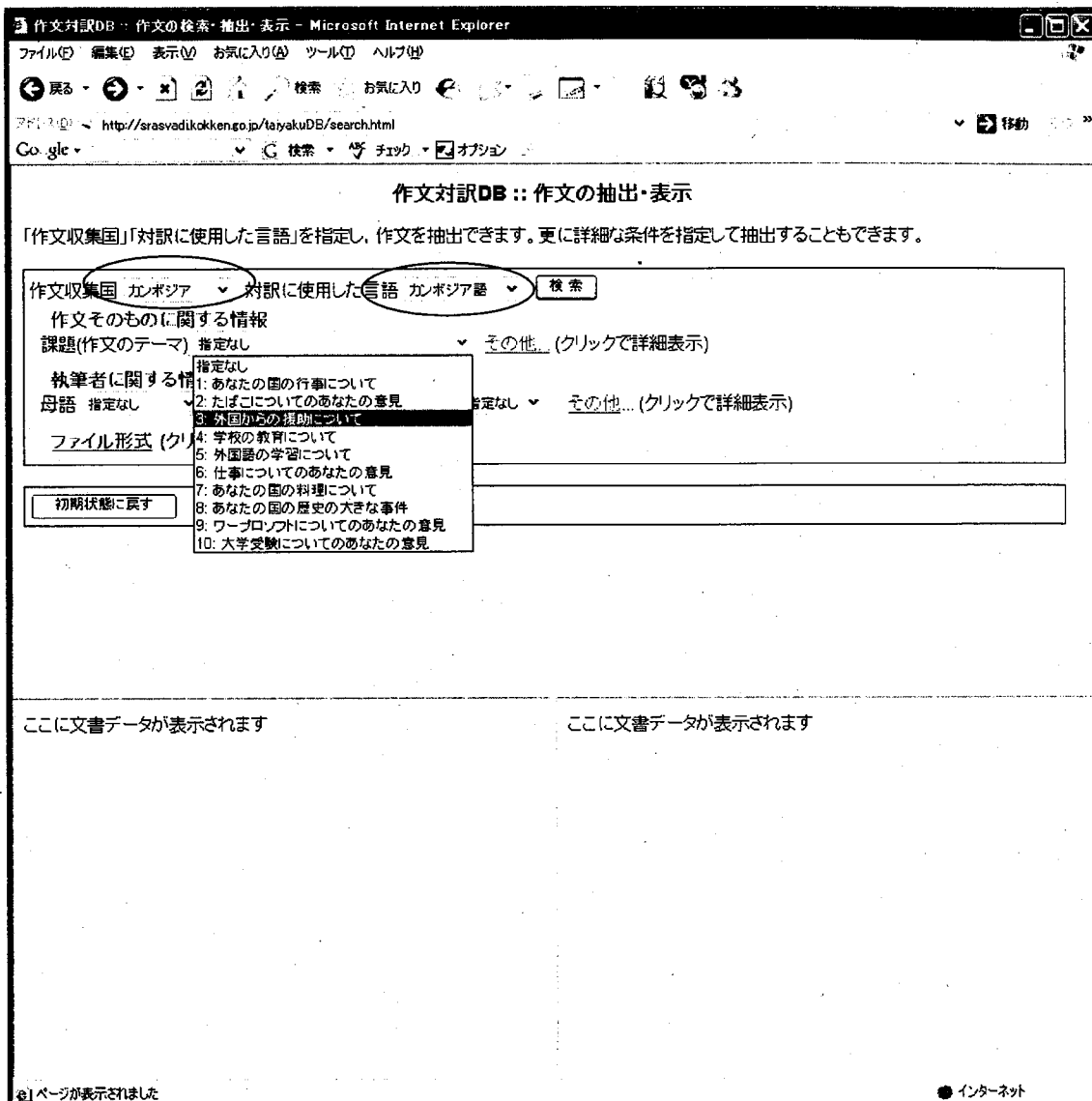


図2 「作文抽出インデックス」初期画面から、作文テーマを選択

次に「検索」ボタンを押すと、画面が以下のように切り替わる。

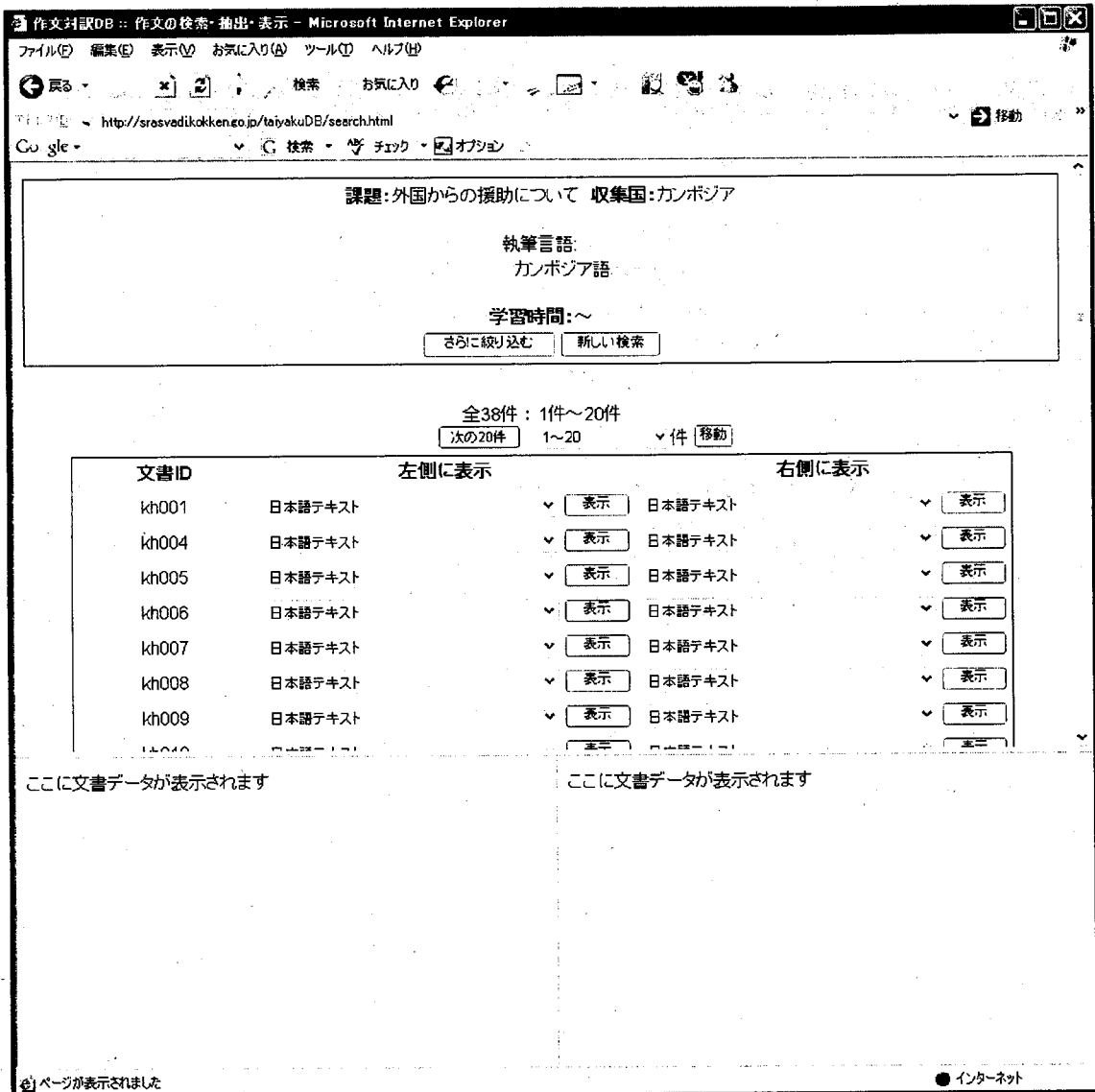


図3 課題「外国からの援助について」で書かれた作文リスト

「文書ID」は、指定された条件に合う作文のID記号である。条件に合う作文はデータベース内に38件存在することが分かる。

ここで、「左側に表示」「右側に表示」それぞれのプルダウンメニューにより、画面下部左側の窓、右側の窓に表示させたいファイルの種類を選択した上で、プルダウンメニューのすぐ右にある「表示」ボタンを押すと、下部の窓に指定したファイルが表示される。窓は左右に並んでいるため、関連する2種類のファイルを並べて表示することができる。

例えば、左側の窓に「kh010の日本語テキスト」を表示させた上で、左側の窓にその母語訳のpdfを表示させたい場合は以下のようにする（文字コードの関係でカンボジア語データはテキストファイルでは保存されていない、pdfでの公開となっている）。

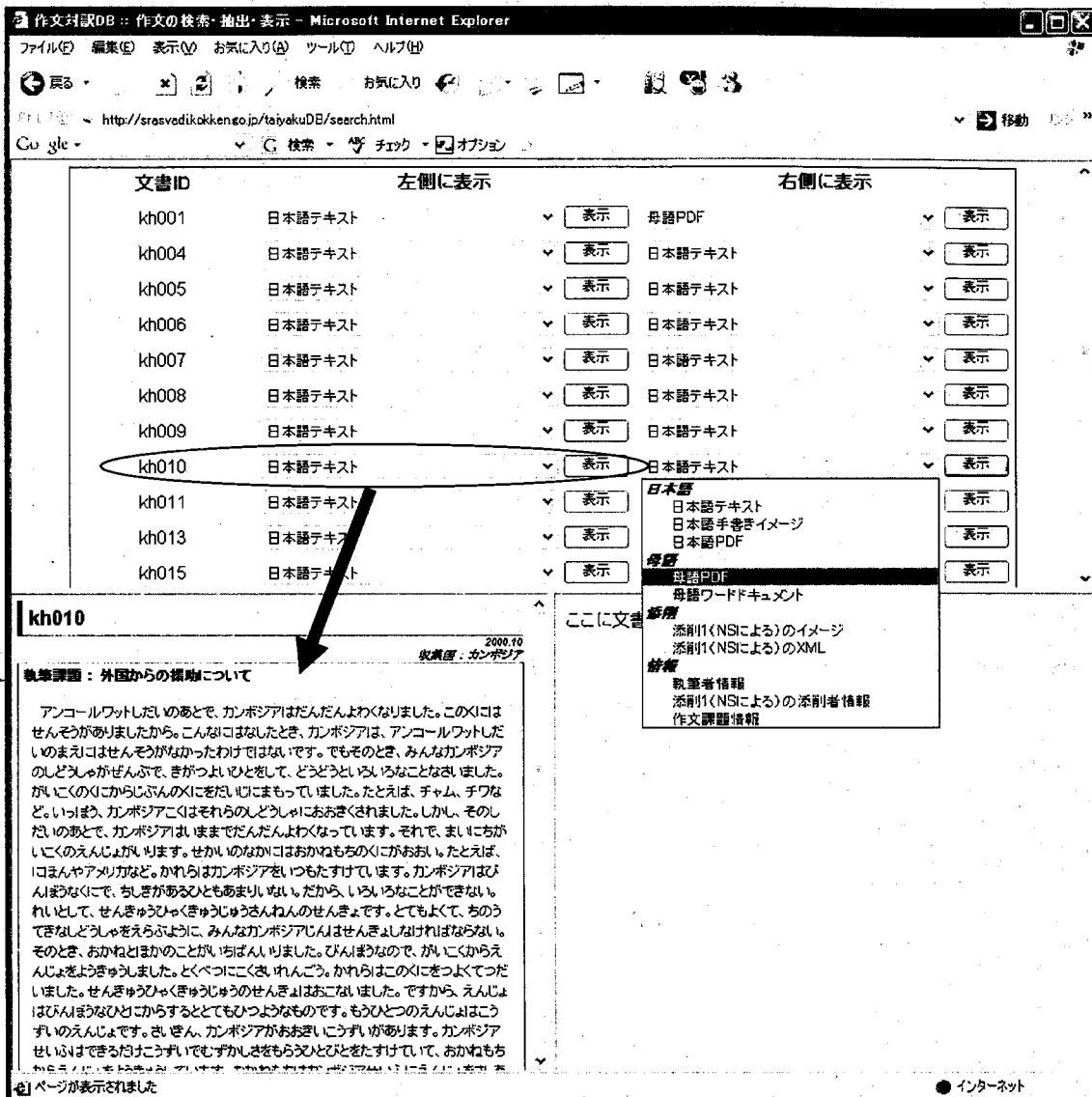


図4 画面左側に日本語テキストを表示させた上で、右側に母語 pdf を表示させようとしているところ

ここで「表示」を押すと、表示画面は以下のようになる。

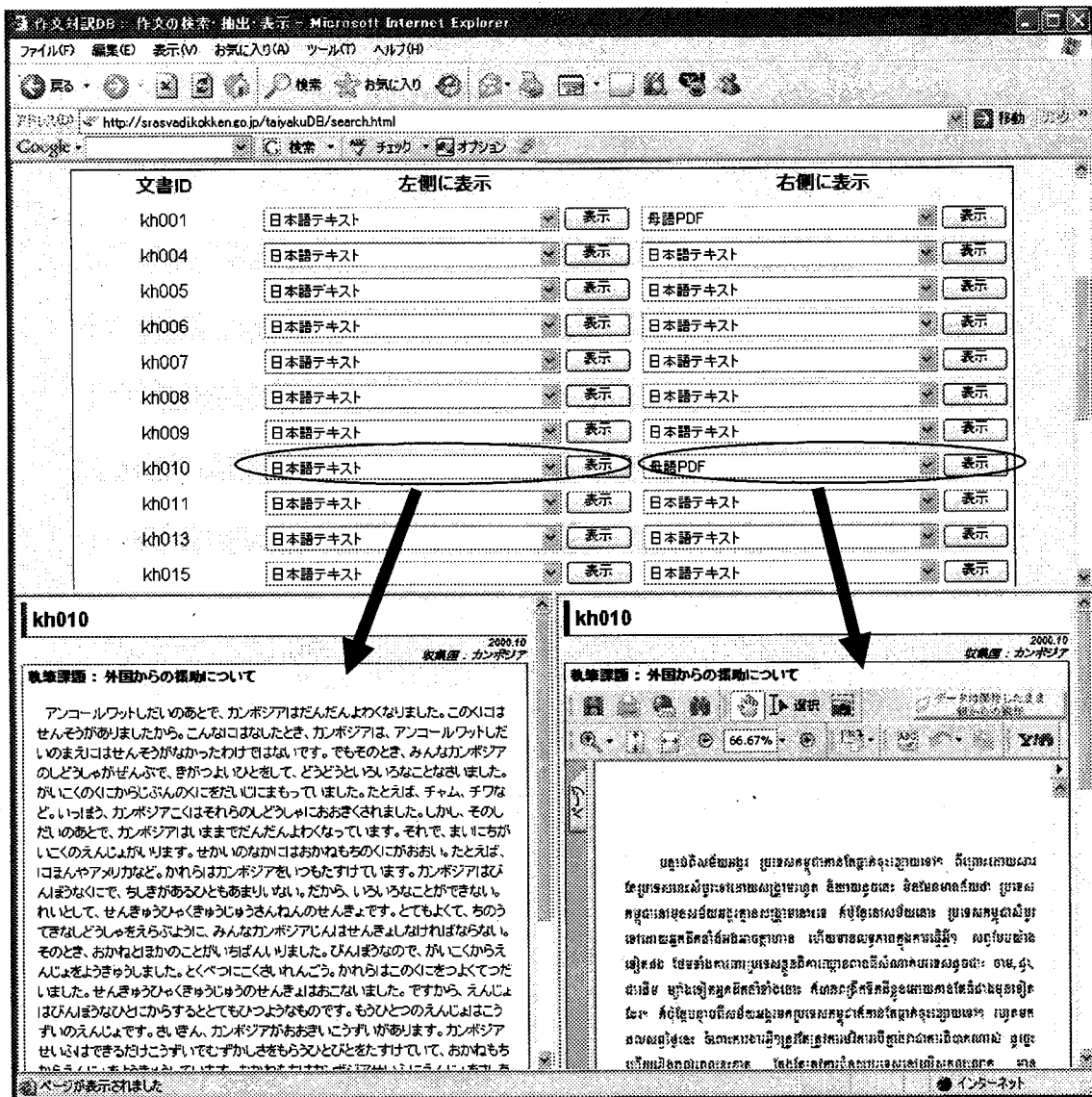


図5 画面左側に日本語テキスト、右側に母語 pdf を表示

表示されたテキストは、窓を右クリックして「すべて選択(A)」を選び、コピーした上で任意のテキストエディタなどに貼り付けることで、オフラインでも利用することができる。

右側、左側の窓に表示させる情報はプルダウンメニューによって自由に選ぶことができるため、比較対照したい任意のファイルを選んで並列表示させることが可能となる。例えば以下の画面は、左の窓に「kh010 に対する、XML 化された添削情報¹¹⁾」を、右の窓にその添削者の「言語歴情報」を表示したものである。

¹¹⁾ なお、この画面で表示される添削情報は、XML ファイルを html に変換したものであり、この画面の「ソース表示」をおこなっても添削タグは見られない。XML ファイルを XML ファイルのまま利用するためには、「ファイルのダウンロード」をおこなう必要がある。XML ファイルのダウンロード方法については p.41-42 を参照。

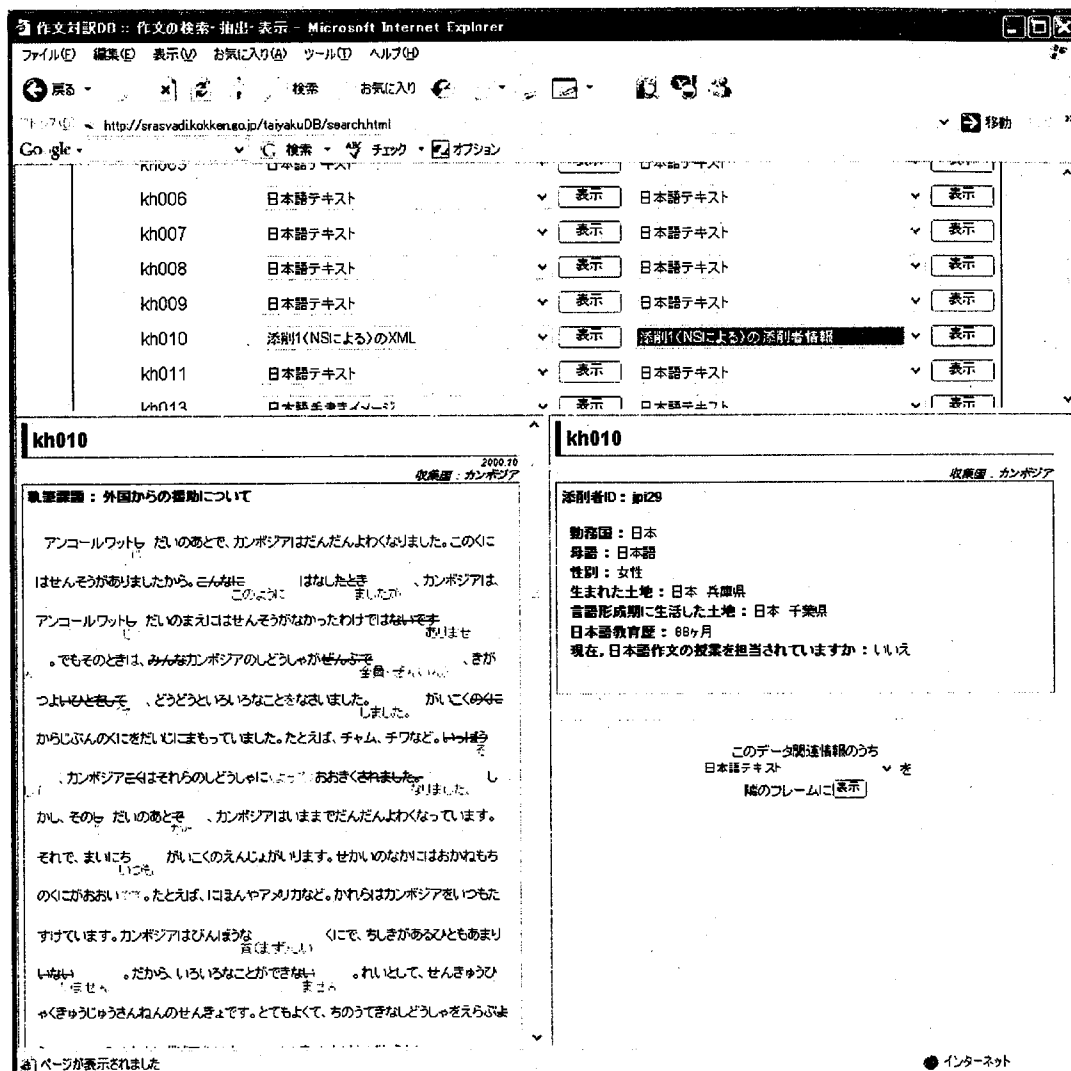


図6 画面左側に添削情報、右側に添削者情報を表示

また、条件に合う作文をすべて結合させて表示するには、上段窓の最下部に「全件表示」というメニューがあるので、ここから「日本語テキスト」または「母語テキスト」を選択すると、下段窓に条件に合う作文がすべて結合されて表示される。

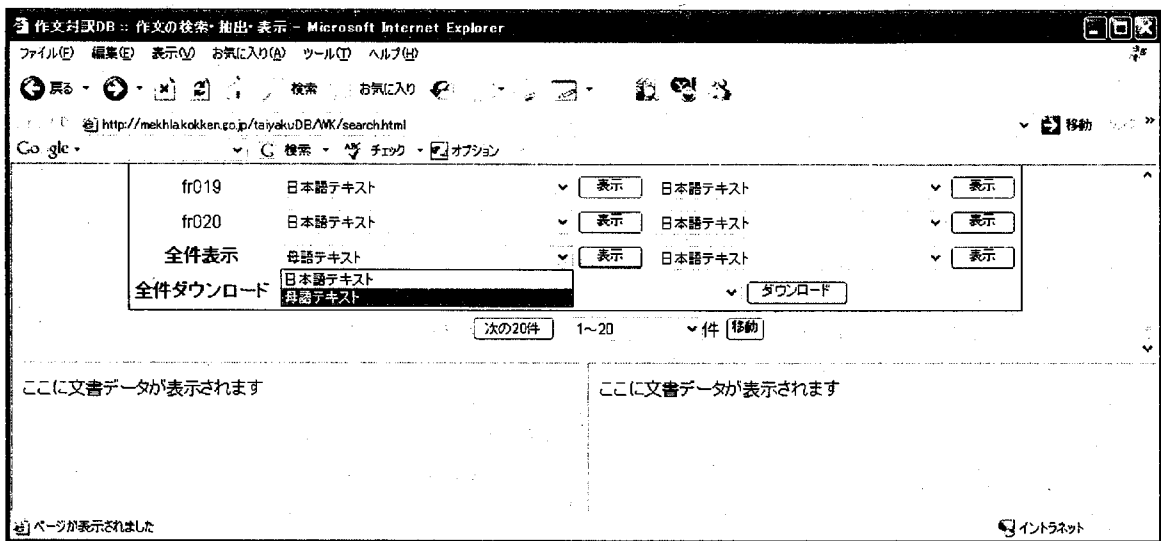


図7 画面左側に母語テキストを全件結合表示させようとしているところ

3.2.2. 「学習歴」による抽出

図1「作文抽出インデックス」の上段窓に示されているように、このデータベースでは学習者の「学習時間」の長さを指定して作文の抽出をおこなうこともできる。ただしこのデータベースに記載された「学習時間情報」は、あくまで学習者本人の自己申告に基づくものであり、国ごとの学習環境によっては、同じ期間の学習であってもその密度には大きな差があることが予想される。「学習歴」を用いた抽出をおこなう際には、そのことを十分理解のうえ、自己責任に基づいて使用するようお願いしたい。

3.2.3. その他の条件による抽出

「執筆者言語歴情報」に含まれている情報のうち、「その情報によって作文抽出をおこなうことは少ない」と考えられるものについては、デフォルトの抽出画面では非表示とされている。しかし、「作文抽出インデックス」の「その他... (クリックで詳細表示)」の部分(2箇所あり)をクリックすると、デフォルトで非表示とされていた条件が表示され、これらの条件による作文抽出も可能となる。しかしこれらの情報も学習者の自己申告に基づくものであり、これらの情報を用いた作文抽出の結果は、あくまで研究者の自己責任に基づいて使用するようお願いしたい。

また、「作文抽出インデックス」の「ファイル形式」の部分をクリックすると、「特定のファイル形式で保存されているデータ」のみを抽出することができるようになる。添削情報についてはすべての作文についてそろっているわけではないため、添削情報のあるデータのみを抽出したいときに、ファイル形式に基づく抽出をおこなうと便利である。

「ファイル抽出のために使用可能な条件」をすべて表示させると以下ようになる。

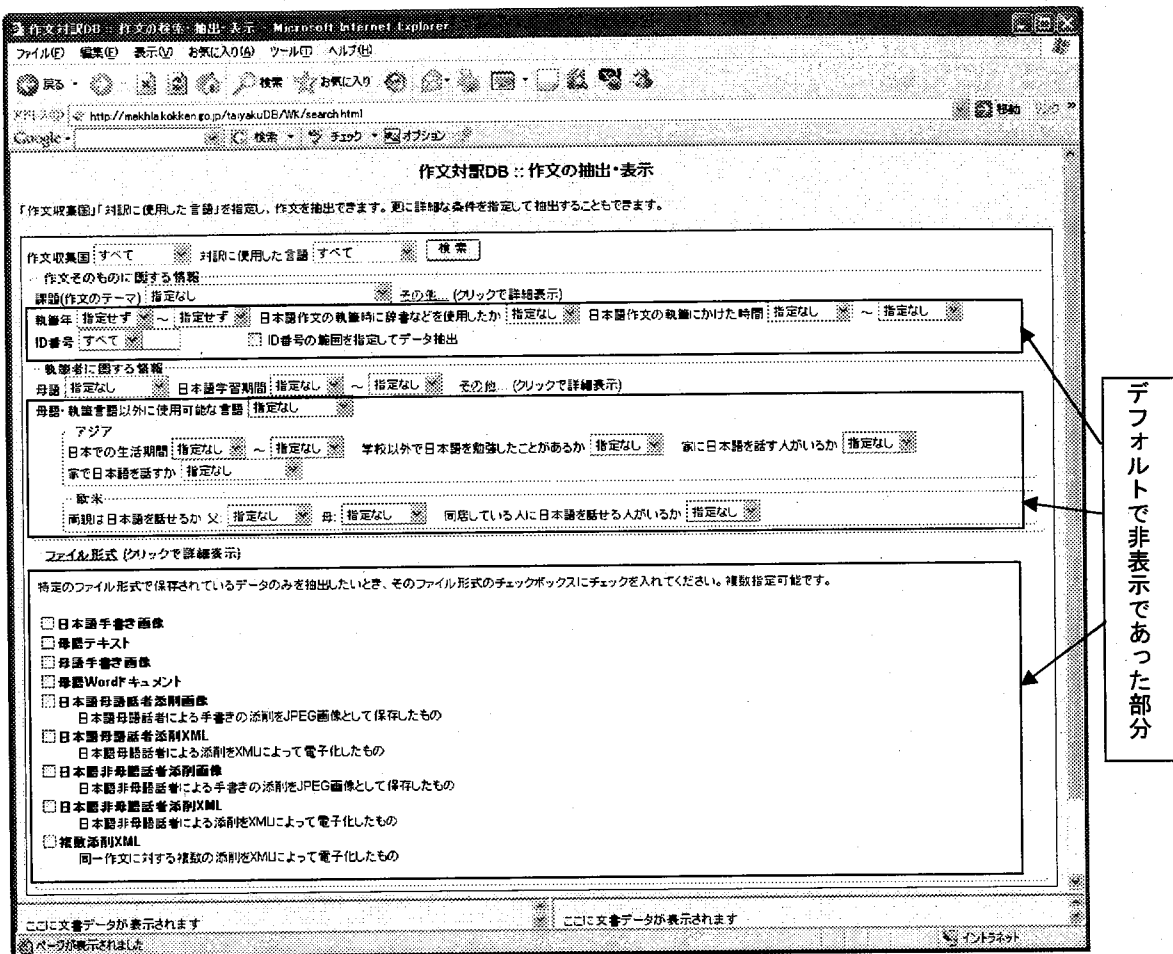


図8 ファイル形式によるデータ抽出画面

例えば、「XML化された、日本語非母語話者による添削情報」を有するデータを参照したい時は、画面下部「日本語非母語話者添削 XML」のチェックボックスにチェックを入れた後、画面上部の「検索」ボタンを押すと、「XML化された、日本語非母語話者による添削情報」を有するデータのリストが画面上部に表示されることになる。

3.2.4. 複数の添削情報の表示とダウンロード

このデータベースに収録されている作文データの中には、複数の添削者（3名以上）がそれぞれ個別に添削情報を付与したものが¹²。以下、そのデータの表示とダウンロード方法について述べる。

まず、前ページ画面のチェックボックスリストのいちばん下、「複数添削 XML」にチェックをいれ、「検索」ボタンを押すと、以下のような画面¹³が得られる。

¹² 2006年4月にオンライン版で新規公開するのは、本稿28-29ページで紹介した3期分の添削データのうち、第2期の添削データである。ひとつの作文に対し、24件の添削が付与されている。

¹³ ここには複数添削情報付きのファイルが7件しか表示されていないが、これは作業途中の仮画面であるため、公開時にはさらに多くのファイルを表示できる予定である。表示できる添削情報は今後も順次

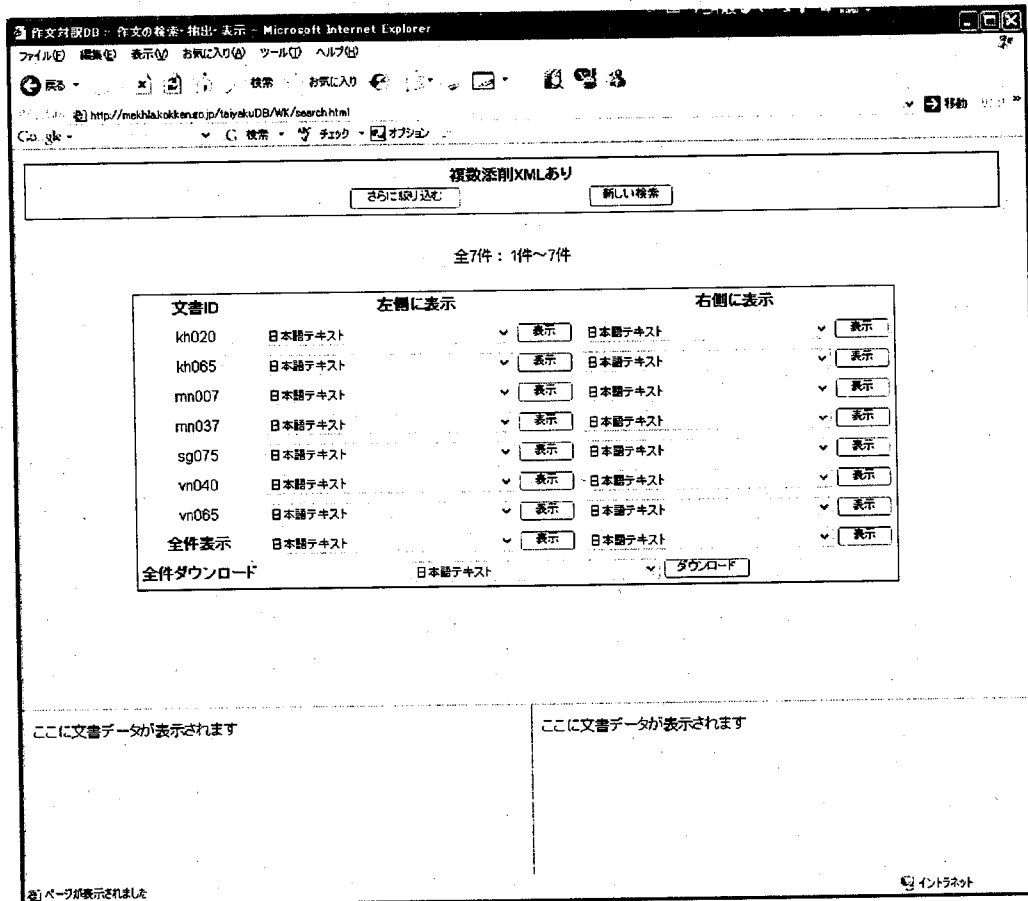


図9 複数（3件以上）の添削情報を持つ作文データのリスト

この画面で、kh065の作文について表示可能な情報リストを提示すると以下のような
 る。

増やしていく予定である。

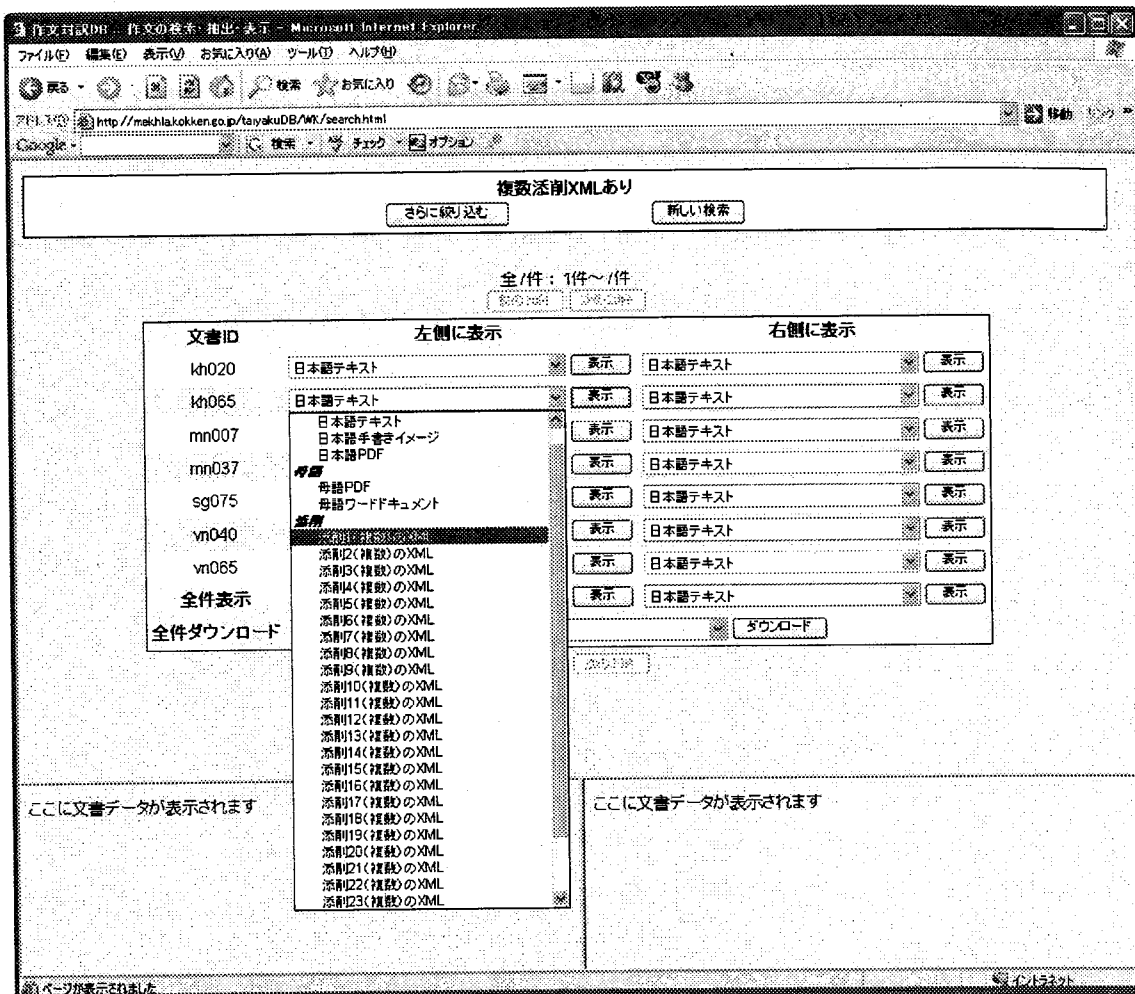


図 10 kh065 について、表示可能なデータのリスト

プルダウンメニューの中から表示したい添削情報を選択し、メニュー右の「表示」ボタンを押すと、画面に XML 化された添削情報が表示される。

なお、「全件表示」のプルダウンメニューから「複数添削の XML」という項目を選んで情報を表示させると、画面上部で表示されているファイルに対し付与された添削情報がすべて連結して表示されることになる。もし、「kh065」の日本語作文に対する添削情報のみをすべて表示したい場合は、kh065 というファイル名によってデータの絞り込みをおこなった上で「全件表示」をおこなうことになる。

具体的には以下のような手順をとる。

- 1) 画面上部「さらに絞り込む」のボタンを押すと、「図 1 「作文抽出インデックス」初期画面」が表示される。
- 2) 画面上部にある「作文そのものに関する情報」という項目から「その他...」をクリックすると、「作文そのものに関する情報」のうち、非表示となっていたものが表示される。

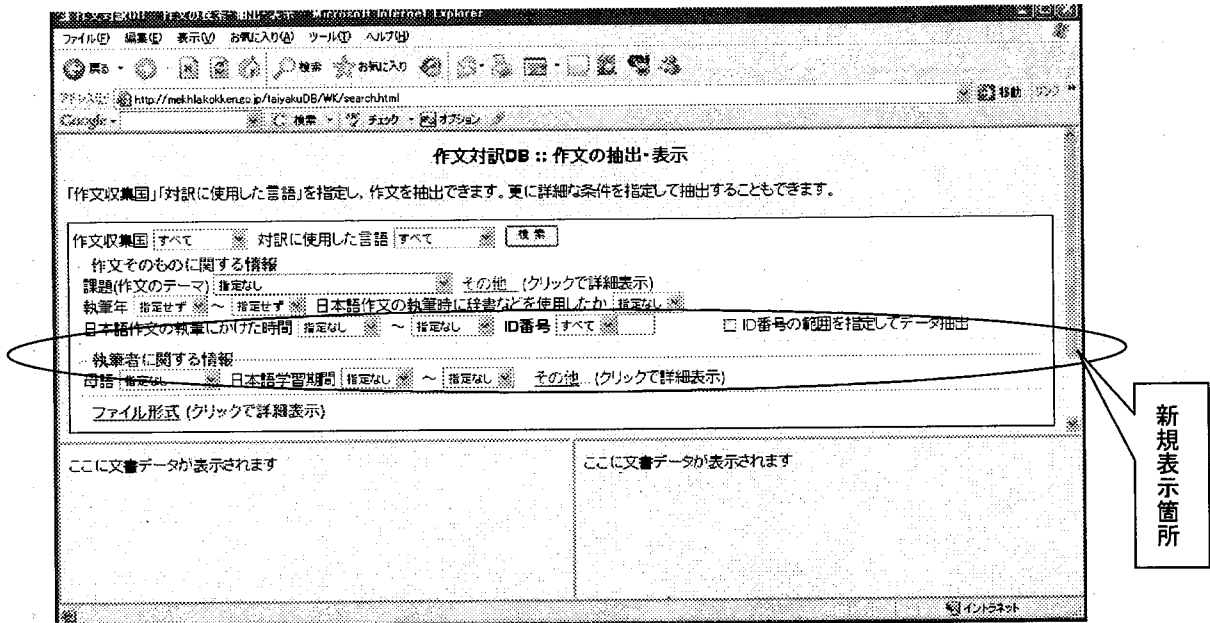


図 11 「作文そのものに関する情報」全表示

- 3) その上で、「ID 番号」右隣「すべて」と書かれている箇所のプルダウンメニューから、抽出したいファイルの国記号(ここでは kh)を選択し、さらにその右隣のボックスに ID 番号(ここでは 065)を入力する。

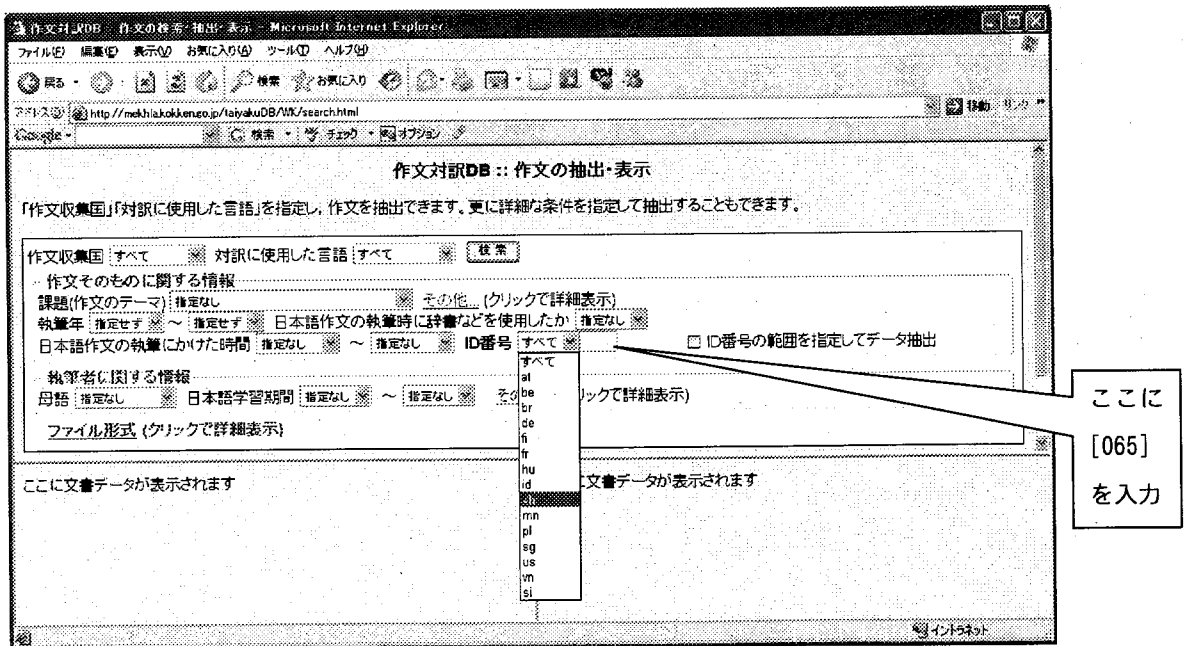


図 12 ID 番号で、国記号 kh を選択

- 4) 画面に「kh065」だけが表示される。この画面の「全件表示」プルダウンメニューから「複数添削 XML」を選び、画面上部の「検索」ボタンをクリックすると、「kh065」に

付与された複数の添削情報が連結されて表示される。

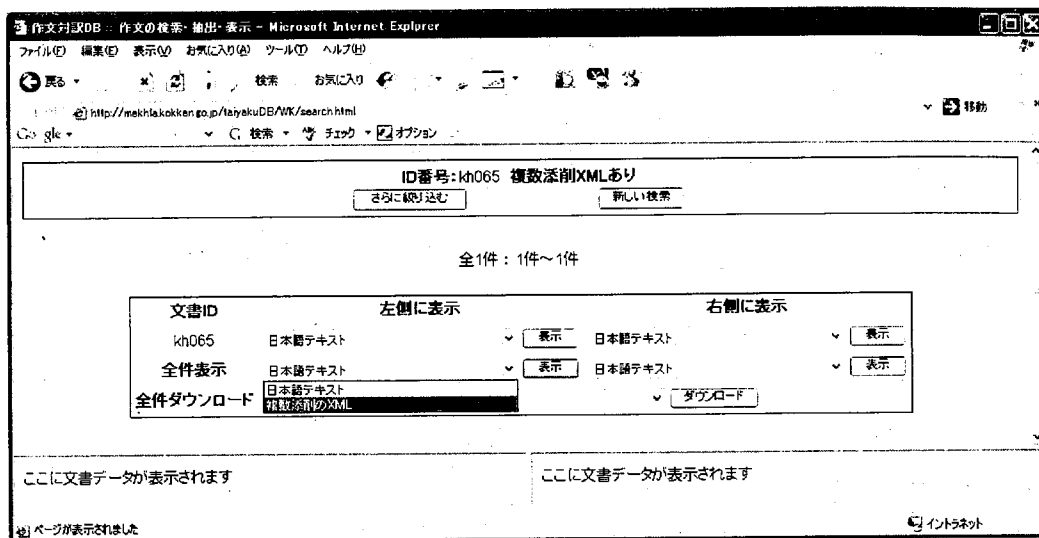


図 13 kh065 に付与された添削情報を、左側画面に全件表示しようとしているところ

ただし、ひとつの作文に対する複数の添削情報を相互比較するためには、web 上で一覧表示するだけでは意味が薄く、複数の添削情報付きファイルを一括ダウンロードし、研究者個人のコンピュータ上でさまざまな加工・検索をおこなうことが必要になる。

kh065 の作文に対して付与された、添削タグ付きの XML ファイルをすべてダウンロードするには、図 14 の下部「全件ダウンロード」から、「添削情報の XML」を選択し、右隣の「ダウンロード」ボタンを押す。

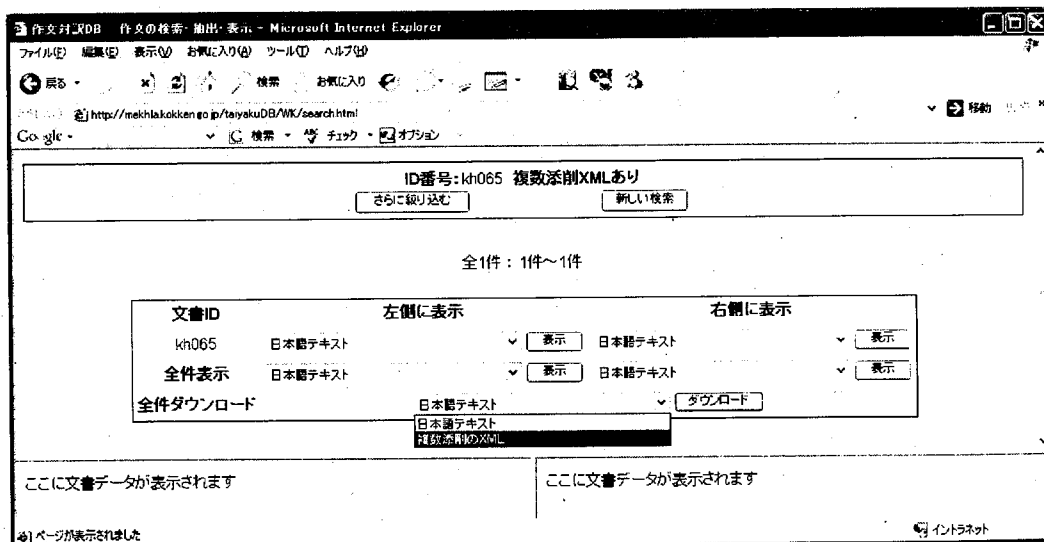


図 14 kh065 に付与された添削情報を、全件ダウンロードしようとしているところ

このようにすることで、以下のファイルを zip 形式で圧縮してひとまとめにした圧縮フ

ファイルがダウンロードされる。

- 1) kh065 に対して添削タグを付与した XML ファイル
- 2) XML ファイルをブラウザ上で表示するやり方を定義したスタイルシート correction.xsl
- 3) ブラウザ表示時に使用される若干の画像ファイル(newp.png, note.png, runon.png, sperr.png)
- 4) 添削者の言語歴情報を記録した csv ファイル rev_inf.csv

この zip ファイルを任意のフォルダに移して解凍し、任意の XML ファイルをクリックすれば、ブラウザ上で添削情報を参照することができる。ただし、XML ファイルをブラウザ上で表示させるためには、当該 XML が保存されたフォルダの中に、2)のスタイルシートならびに 3)の画像ファイルが保存されていなければならない。XML ファイルは別のフォルダに移動してもかまわないが、その場合 2)および 3)のファイルも同じ場所にコピーする必要がある。

XML ファイルをブラウザではなくテキストエディタ等で開けば、添削タグが付与された形で添削情報を参照することができ、各種の検索をかけることが可能となる。

言語データとしての作文対訳データベース

井上 優

1. はじめに

本論では、『日本語学習者による日本語作文と、その母語訳との対訳データベース』（作文対訳 DB）の言語データとしての性格について考察する。

作文対訳 DB の特徴は、学習者の日本語作文とともに、学習者本人による母語訳が付されていることである。学習者にとって、日本語は必ずしも自由に使いこなせる言語ではなく、学習者の日本語作文も、学習者が述べたかったことが完全に表現できているとは限らない。学習者本人による母語訳は、学習者が実際に日本語で何を書きたかったかを知るための手がかりを与えてくれる。また、日本語作文に誤用や不自然な点があった場合、その背景について考える手がかりも与えてくれる。

ただ、学習者本人による母語訳は、学習者の日本語作文について理解するための補助的情報としてのみ利用されるものではない。作文対訳 DB は、一種の平行ル・コーパスとしての書き言葉データベースであるから、学習者による母語訳も、単なる補助的情報としてではなく、言語研究のためのデータとして活用されることが望ましい。

以下では、対照言語学的な観点から、作文対訳 DB に収録されたデータがどのような性格の言語データであるかを考える。具体的にとりあげるデータは、『日本語学習者による日本語作文と、その母語訳との対訳データベース ver.2 (2001)』（以下「作文対訳 DB」と言う場合はこのデータベースを指すものとする）に収録された、中国語を母語とする日本語学習者による日本語作文、ならびに学習者本人による中国語訳である。

2. 作文対訳 DB に収録されたデータの基本的性格

作文対訳 DB に収録されているのは、以下のデータである。下線を施したものが、言語データといえるものである。

I : 日本語学習者による日本語作文と関連情報
a <u>日本語学習者による日本語作文</u>
b <u>作文執筆者本人による母語訳（または最も楽に文章を書ける言語への翻訳）</u>
c 日本語教師による作文の添削
d 作文執筆者・添削者の言語的履歴に関する情報
II : <u>日本語母語話者による日本語作文</u>

これらの言語データの性格について考える上で重要なのは、次の2点である。

- (1) a. 作文対訳 DB に収録された「I a 日本語学習者による日本語作文」、「I b 作文執筆者本人による母語訳」、「II 日本語母語話者による日本語作文」は、いずれ

も必ずしも十分な推敲を経ていない文章である。

- b. 「I b 作文執筆者本人による母語訳」は、日本語作文を母語に翻訳した文章であり、最初から学習者の母語で書かれた文章ではない。

通常、書き言葉研究のデータとして用いられるのは、新聞・雑誌記事、小説など、文章を書き慣れている執筆者が十分に推敲し、場合によっては第三者によるチェックも入って書き上げられる完成度の高い文章である。これに対し、作文対訳 DB に収録された「I a 日本語学習者による日本語作文」、「I b 作文執筆者本人による母語訳」、「II 日本語母語話者による日本語作文」は、いずれも、必ずしも文章を書き慣れているとはいえない書き手が、限られた時間の中で十分な推敲を経ずに書いた文章であり、母語による作文であっても、文章としての完成度は高いとはいえない。加えて、「I b 作文執筆者本人による母語訳」は、日本語作文を母語に翻訳した文章であり、最初から学習者が自分の母語で書いた文章ではない。そのため、母語で書いたとはいえ、その言語の文章としては必ずしも自然とはいえないところや、もとの日本語作文の影響を受けたと見られるところがある。少なくとも、学習者による中国語訳を見るかぎりはそのようである（詳しくは 5. で述べる）。

したがって、作文対訳 DB 収録の文章を言語データとして活用する際には、次の 2 つの点に留意する必要がある。

- (2) a. 「I a 日本語学習者による日本語作文」の特徴について考える場合は、完成度の高い日本語の文章を基準にして考えるよりは、十分な推敲を経ていない日本語話者の日本語文章（たとえば作文対訳 DB 中の「II 日本語母語話者による日本語作文」）を基準に考えるべきである。
- b. 「I b 作文執筆者本人による母語訳」がその言語の文章としてどの程度自然であるかを、第三者的な視点からチェックする必要がある。

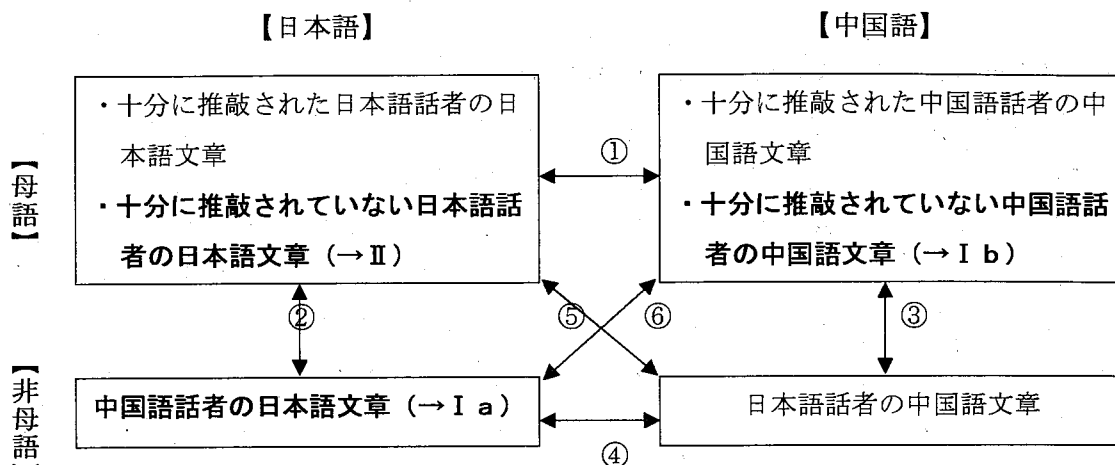
3. 作文対訳 DB で可能な対照研究

次に、日本語と中国語の対照研究という観点から、「I a 日本語学習者（中国語母語話者）による日本語作文」、「I b 作文執筆者本人による母語訳（中国語訳）」、「II 日本語母語話者による日本語作文」の位置づけについて考える。

この 3 つのデータの位置づけを大まかにまとめたものが次ページの(3)である。

対照研究において通常行われるのは、①の比較対照、すなわち日本語母語話者の日本語文章と中国語母語話者の中国語文章の比較対照である。作文対訳 DB には、「II 日本語母語話者による日本語作文」と「I b 作文執筆者本人による母語訳（中国語訳）」が含まれているので、①の比較対照（十分に推敲されていない文章の比較対照）が可能である。ただし、その際には、先にあげた(1b)、すなわち「I b 作文執筆者本人による母語訳（中国語訳）」が日本語作文を母語に翻訳した文章であり、最初から学習者の母語で書かれた文章ではないという点に留意する必要がある。

(3)



②, ③の比較対照は, 母語話者の文章と非母語話者の文章を比較して, 非母語話者の言語の特徴を把握するものである。作文対訳 DB には, 「I a 日本語学習者 (中国語母語話者) による日本語作文」とともに, 同じテーマで書かれた「II 日本語母語話者による日本語作文」が収録されているので, ②の比較対照によって, 中国語母語話者が書いた日本語作文の特徴を把握することができる。たとえば, 伊集院・高橋(2004)は, この2つのデータのうち, 意見文 (たばこについてのあなたの意見) の文末モダリティを比較し, 「II 日本語母語話者による日本語作文」に比べて, 「I a 日本語学習者 (中国語母語話者) による日本語作文」に「相手に向かって話しかけているような表現」が多く用いられていることを指摘している。

⑤, ⑥の比較対照は, 学習言語による作文と母語による作文との比較対照である。作文対訳 DB には, 「I a 日本語学習者 (中国語母語話者) による日本語作文」と「I b 作文執筆者本人による母語訳 (中国語訳)」とが含まれているので, ⑥の比較対照により, 学習者 (中国語母語話者) の日本語作文に見られる特徴が母語 (中国語) の特徴の反映であるか否かを検討する材料が提供されている。ただし, 先に(1b)として述べたように, 「I b 作文執筆者本人による母語訳 (中国語訳)」は, 日本語作文を学習者自身が中国語に翻訳した文章であるため, 中国語の特徴が日本語作文に影響を与えるというよりは, 日本語作文の方が母語訳に影響を与えていると見られる部分もある (具体例は 5. で述べる)。したがって, ⑥の比較対照を行う場合には, 先の(2b)で述べたように, 「I b 作文執筆者本人による母語訳 (中国語訳)」が中国語の文章としてどの程度自然であるかを, 第三者的な視点からチェックすることが必要である。

④の比較対照は, ②, ③の比較対照と組み合わせることにより, 日本語話者の中国語習得の様相と中国語話者の日本語習得の様相とを比較対照するものである。作文対訳 DB には, 「日本語話者による中国語文章」は収録されていないので, この種の比較対照を行うこ

とはできない。

4. 対訳資料としての作文対訳 DB の位置づけ

作文対訳 DB は一種の対訳資料である。対訳資料は、内容が同じ2つの言語の文章を比較することができるので、対照研究を行う上では便利である。

もっとも、文章の内容が同じことは、対照研究の資料として特に価値が高いということの意味するわけではない。第一に、翻訳の仕方というものは、言語学的な要因以外にも様々な要因が関与して決まるものであり、二言語間における表現の対応関係の問題として単純に処理できるものではない。第二に、ある言語の表現を別の言語の表現に置き換える作業には一定の技能が必要であり、翻訳の正確さも翻訳者の技量によって異なる。対訳資料を利用する際には、むしろ「〇〇語のこの文が××語ではこのように訳されている」ということが言語学的にどのような意味を持つかという点について、より慎重な検討が必要になる。対訳資料の意義は、そのような検討を行う材料の提供にあると言ってもよい。

李琚寧(2005)は、北京日本学研究中心(2003)『中日対訳コーパス』(日本語の小説とその中国語訳、中国語の小説とその日本語訳の平行・コーパス)を用いた研究であるが、そこでは、許可・許容を表す中国語の助動詞“可以 keyi”を用いた文が日本語に訳される場合、しばしば授受表現が用いられることを指摘している。たとえば、以下の(4)(5)の a. を日本語に翻訳した b. を見ると、原文の“可以”が「くれる」「やる」という授受表現に訳されていることが分かる。

(4) a. 但理想不等于空想。理想有科学依据，可以成为现实，也可以给人以物质力量。

(戴厚英「人啊，人」)

b. しかし、理想イコール空想じゃない。理想は科学的根拠があって、現実になり得るもの、人間に物質的な力を与えてくれるものだ。

(大石智良訳，サイマル出版会)

(5) a. “你从实招来罢，免得吃苦。我早都知道了。招了可以放你。”

(鲁迅「呐喊」)

b. 「ありていに白状するんだ。痛い目をみずにすむからな。何もかもわかっているんだぞ。白状すれば放免してやる。(竹内好訳，筑摩書房)

通常の対照研究では、“可以”と比較対照される日本語の表現は、同じく許可・許容を表す「～でもいい」「～できる」などであり、授受表現は視野に入っていない。しかし、李琚寧(2005)によれば、“可以”を含む表現が授受表現を用いて訳されることには一定の理由がある。(4)(5)で述べられていることがらは、動作の受け手の利益になる動作である。日本語文(4a)(5a)においては、当該の動作が受け手利益の動作であることが授受表現により明示的に表されている。これに対し、中国語文(4b)(5b)では、許可・許容を表す“可以”を用いて「受け手利益の事態を引き受ける・許容する」姿勢を示すことにより、間接的に利益

の受給関係を表している。例文(4)(5)は、利益関係を表す様式が日本語と中国語とで異なることを示唆するものである。このような知見は、通常の対照研究では得にくい、対訳資料ならではの知見である。

作文対訳 DB のデータを利用する場合も、「学習者は日本語作文を自分の母語でこのように訳している」ということが言語学的にどのような意味を持つかについて、十分な検討が必要である。特に、作文対訳 DB の場合は、対訳資料とは言っても、日本語作文は非母語による作文、その母語訳は母語による作文であるから、単純に両者を比較することはできない。また、日本語作文の中国語訳を行っているのは学習者自身であるから、日本語に関する何らかの誤解に基づいて翻訳を行っている可能性は通常の翻訳よりも高い。さらに、繰り返し述べているように、「I b 作文執筆者本人による母語訳 (中国語訳)」は、日本語作文を母語に翻訳した文章であり、最初から学習者が自分の母語で書いた文章ではない。それゆえ、母語で書いたとはいえ、その言語の文章としては必ずしも自然とはいえないところや、もとの日本語作文の影響を受けたと見られるところがある。したがって、作文対訳 DB のデータを対訳資料として利用する場合は、翻訳の仕方が言語学的にどのような意味を持つかを、通常の対訳資料よりもさらに慎重に検討する必要がある。

特に重要なのは、「I b 作文執筆者本人による母語訳」に見られる現象が、学習者の母語に一般に見られる現象かどうかを検討することである。学習者の日本語作文と学習者による母語訳に類似の特徴が見られた場合、その特徴が学習者の母語に一般的に見られる特徴であり、かつ日本語母語話者の作文とは異なる特徴であれば、学習者の日本語作文に見られる特徴は学習者の母語の影響によるものである可能性が出てくる。

そのような例として、文章中における疑問文の使われ方をあげることができる (井上 2005)。次の(6)は、作文対訳 DB に収録された、中国語話者 (CN015) による「たばこについてのあなたの意見」を内容とする意見文である。

(6) [CN015 : 日本語作文]

たばこを吸うのは権利ですか

今、たばこのことが問題になっています。大体規則を作って制限するとこれは権利だという話しを強調するに分けています。私は前の方が賛成しますが、つぎのように自分の見方を述べます。

まず、権利というものはかつてなことではないと思います。権利は人間の合理的な要求を合致するために出てきたものですが、ただ自分自身の権利を求めて、他人に考えていない方は完全にそれを曲解します。たとえば、殺人犯は「私は生きる権利があるよ。」と唱えたら、法律は彼に自由を与えますか。それはもちろん「×。」ですから、たばこを吸うために権利を口実にするのは、私利をはかることではないですか。[以下省略]

日本語の場合、下線部分の疑問文は、意見文で用いる疑問文としては不自然である。ま

ず、意見文のタイトルは、丁寧体ではなく、「たばこを吸うのは権利か」のように普通体を用いる。また、丁寧体を用いた意見文で問題提起を行う場合は、「法律は彼に自由を与えるでしょうか」、「私利をはかることではないでしょうか」のように、聞き手に応答を要求しない疑念表現「でしょうか」を用いる必要がある（普通体の場合は、「法律は彼に自由を与えるか」、「私利をはかることではないか」でよい）。「でしょうか」を用いずに、「法律は彼に自由を与えますか」、「私利をはかることではないですか」のように言うと、読み手に対する問いかけの文になる（この点、先に紹介した伊集院・高橋(2004)の指摘と関連する）。意見文は、読み手に一方的に話し手の考えを述べるという点で、独話に近いものである。これに対し、「ですか」は相手に質問して回答を求める疑問文であり、対話の中で用いられるものである。意見文に「ですか」がそぐわないのも、独話をベースとする文章の中に対話的な表現が現れるからである。

一方、中国語では、意見文の中で問題提起を行う場合にも、読み手を引き付けるために、相手に質問するのと同じ形の疑問文を用いることがよくある。たとえば、次の(7)は、毛沢東の「人の正しい思想はどこから来るのか」という文章の冒頭であるが、やはり相手に質問した上で自ら回答を述べるというスタイルになっている。

- (7) 人的正确思想是从哪里来的？是从天上掉下来的吗？不是。是自己的脑里固有的吗？不是。人的正确思想，只能从社会实践中来，只能从社会的生产斗争、阶级斗争和科学实验这三项实践中来。（人の正しい思想はどこから来るのか？天から降ってくるのか？そうではない。自分の脳の中にもともとあるのか？そうではない。人の正しい思想は、社会実践から、つまり社会的生産闘争、階級闘争と科学的実践の3つの実践からのみ生まれる。）

（毛沢東「人の正しい思想はどこから来るか」冒頭、『毛沢東文集』第8巻，中国：人民出版社，1999年）

日本語作文(6)の中国語訳でも、「ですか」が用いられている部分は、相手に質問するのと同じ形の疑問文に訳されている。

- (8) [CN015：母語訳]

A吸烟是一种权利吗？

目前，关于吸烟成为一个话题。大体上可以分为撰定规章进行限制和强调这本是个人权利两大阵营，我是赞成前者的，下面想就自己的意见阐述一下。

首先，我认为权利并不是一种任意的东西，权利是以人的合理要求为一致的前提下所产生的，如果仅是谋求自身权利，而不顾他人的行为是对此的一种完全的曲解，譬如，杀人犯高呼“我有生存权”，B难道法律就会赋与他自由了吗？当然是不会的。因此，为了吸烟，而把权利当作挡箭牌，C这难道不是为求一己私利的吗？

[以下省略]

学習者による中国語訳(8)における疑問文の用いられ方が、中国語の文章における疑問文

の使われ方を反映したものだとする、同じ学習者が書いた日本語作文(6)に見られる疑問文の使われ方も、母語である中国語の影響である可能性が出てくる。日本語作文(2)とその母語訳(8)を比較するだけでは、ここまでは言えない。作文対訳 DB を一種の対訳資料として活用するためには、日本語作文とその母語訳だけでなく、学習者の母語で書かれた他の文章も視野に入れるべきである。

5. 「学習者による母語訳」の性格：中国語話者による中国語訳の場合

前節での議論からも示唆されるように、作文対訳 DB に収録された言語データの活用について考える際に重要なポイントとなるのは、学習者の母語による作文がどのような文章かということである。調査する側としては、学習者にとって、母語ではない日本語で文章を書くことは難しいだろうが、母語であれば、自分が書きたかったことをより自然な形で書くことができるだろう、と考える。しかし、すでに述べたように、作文対訳 DB に収録された作文は、母語による作文であれ非母語による作文であれ、必ずしも文章を書き慣れているとはいえない書き手が、限られた時間の中で、必ずしも十分な推敲を経ずに書いた文章である。また、「I b 作文執筆者本人による母語訳」は、日本語作文を母語に翻訳した文章であり、最初から学習者の母語で書かれた文章ではない。そのため、「I b 作文執筆者本人による母語訳」がその言語の文章としてどの程度自然であるかを、第三者的な視点からチェックする必要がある。

そこで、「I b 作文執筆者本人による母語訳 (中国語訳)」を別の中国語母語話者 (黃麗華氏, 1957 年生, 中国山西省育ち, 日本語・中国語の教師経験あり) に見てもらったところ、次のようなことが少なからず見られることが分かった

- 1) 母語である中国語の文章に不自然、あるいはたどたどしい表現が散見される。
- 2) 日本語を中国語に直訳したような表現がしばしば見られる。
- 3) 同じ文章の中で書き言葉と話し言葉が混在することがある。

たとえば、次の例は、中国語話者 CN019 による日本語作文とその母語訳である (*は日本で一般に使われているものと明らかに異なる字形だが、どういう字を書こうとしているかは明らかなもの)。

(9) [CN019 : 日本語]

結婚式について

中国で結婚式は伝統的な式と西洋風な式に分ける。西洋風*の結婚式はわりに簡潔で、どの国でもあると思う。中国伝統的な式といたら、私自身自身もよく知らないが、知っていることをすこし紹介する。[省略]

しかし、今よく行われる結婚式は伝統と伝来の混合物だ。 伝統的な服は花嫁の顔が他人に見えないように「紅蓋頭」をかぶる。今そんな物は殆どいらなくなった。花嫁はウェディングドレスを着る人が多くなってきた。花婿も殆どセビロを着るのだ。

花婿は友達たちと花嫁の家へ行って、花嫁の姉妹や友達に何回も頼*んで、「紅包*」（お金が入る赤い紙封筒）を出してからさえ花嫁を連れて行ける。「花轎*」も殆どカーになった。結婚式は若者の友達*の祭りになってしまう。

それから今、式は教堂で西洋式とホ*テルで中華式と二回やることも多いようだ。

(10) [CN019: 中国語]

关于结婚仪式

在中国，结婚仪式分为传统式的和西洋式的。西洋式的比较简洁，我想每个国家都有。说到中国传统结婚仪式，我自己本身也不是特别了解，但这儿就我所知的稍作介绍。

[省略]

但，如今常_A进行的是传统物和传来物的混合物。_B传统礼服为了不让人看见新娘的脸，得盖上“红盖头”，而如今这种东西都没了。新娘多穿婚纱，新郎也一般都穿西服。新郎和朋友到新娘家，苦苦拜托新娘的朋友姐妹，并发给“红包”之后才能带走新娘，而“花轿”也都变成轿车了。结婚仪式成了青年朋友们的节日。

另外，如今，也有很多人的结婚仪式在教堂来一次基督教式的，再在饭店来一次中国传统式的。

まず、中国語訳(10)で波線を施した“这儿”，“没了”，“来一次…，再…来一次”は、局所的に話しことば的な表現，あるいはくだけた表現になっている部分である。

下線部A，Bは，日本語作文の影響を受けた可能性がある部分である。下線部Aの“进行的是传统物和传来物的混合物”は，日本語作文の下線部B「今よく行われる結婚式は伝統と伝来の混合物だ」の直訳という印象が強い。また，下線部B“传统礼服为了不让人看见新娘的脸，得盖上‘红盖头’”も，日本語作文の下線部Bを直訳したという印象が強い。日本語作文では，「伝統的な服」という主題を示し，「花嫁の顔が他人に見えないように『紅蓋頭』をかぶる」という服装に関わるコメントが後に続くという構造になっている。中国語訳の下線部Bも“传统礼服”（伝統的な礼服）の部分が主題，“为了不让人看见新娘的脸，得盖上‘红盖头’”（花嫁の顔が人に見られないように『紅蓋頭』をかぶらなければならない）の部分がコメントになっており，日本語文と同じ構造になっているが，中国語では，このような構造はねじれた文という印象が強い。

もう一つ例をあげよう。

(11) [CN006: 日本語]

みんなの権利を守るため、_A会社やレストラン、バスや電車など公共の場所ではたばこを吸わないべきですが、規則を作るのはちょっとやりすぎじゃないかと思います。私はみんなの権利はみんなですらなければなりません。たばこを吸わない人は吸う権利をたばこを吸う人にあげる、反対に、たばこを吸う人も吸わない権利を吸わない人にあげるはずです。みんなそのように考えたら、_Bたばこの問題が解決できるかなと思います。

また、たばこのコマーシャルはがい誘の。疑惑がありますので、子どもか青少年に悪い影響を与えるから、テレビで放送できないように規則を作るべきです。われわれ大人は子どもや青少年をたばこを吸わないほうへ導くべきです。

以上は私の考えですが、みんなはどうを思いますか。

(12) [CN006 : 中国語]

为保障大家的权利，A会社、レストラン、公共汽车、电车等公共场合不应该吸烟，但是要制订规则来规定，是不是做得过头了点、我想大家的权利要大家来保障，不吸烟的人要给吸烟人吸烟的权利，反过来，吸烟的人也要给不吸烟的人不吸烟的权利，如果大家都能这样想，吸烟的问题，是不是就可以解决了呢？

另外，香烟广告有诱惑人们去吸烟的。疑点，对小孩和青少年产生坏影响，所以电视里这类广告应该规定禁止播放。我们大人应该引导孩子和青少年不吸烟。

以上是我的想法，大家怎样想。

まず、中国語訳(12)で波線を施した部分は、局所的に話しことば的な表現、あるいはくだけた表現になっている部分である。

下線部Aは日本語単語をそのまま用いた例である。“公司 gongsi”(会社)，“餐厅 canting”(レストラン)という中国語単語があるにもかかわらず、「会社」「レストラン」という日本語単語を用いた理由はよく分からない。もしかすると、この学習者は、状況によっては、中国語を話すときに日本語の単語をよく使うのかもしれない。

下線部C“疑点 yidian”も、日本語の「疑い」を直訳したと見られる表現である。日本語の「…の疑いがある」は、「…の恐れ(危険性)がある」という意味で用いられることがある。日本語作文(11)では「疑惑」が用いられているが、言いたいことは「タバコの広告が人々を喫煙に走らせる恐れがある」ということだと推察される。しかし、中国語の“疑点”は、「疑わしい点」「疑問点」という意味であり、「…の恐れ(危険性)がある」という意味では通常用いられない。

下線部Bは、局所的に話しことば的になっているだけでなく、日本語作文と中国語作文とで内容にずれが生じている例である。日本語作文の下線部B，

(13) みんなそのように考えたら、たばこの問題が解決できるかなと思います。

は、「みんなそのように考えたら、たばこの問題が解決できるかもしれない」という気持ちを伴った、肯定の傾きを有する文である。これに対し、中国語訳の下線部B，

(14) 如果大家都能这样想，吸烟的问题是不是就可以解决了呢？

は、「みんなそのように考えたら、たばこの問題は解決できるのだろうか(できないのではないか)」という疑いの気持ちを伴った、否定の傾きを有する文である。“是不是 shibushi”という表現自体は、肯定の傾きを有する「…のではないか」に近い意味を表す場合もあるが、ネイティブは(14)は否定の傾きを有する文と判断し、文脈の流れに合致しない文であると言う。理由はよく分からないが、日本語作文よりも中国語訳の方が不自然

な文章になっているわけである。

6. 対訳作文 DB を言語データとして活用するために

このように、中国語話者による母語訳を見るかぎり、作文対訳 DB に収録された学習者の母語訳は、必ずしも自然な中国語の文章とは言えないところがある。「作文執筆者本人による母語訳」をもとにして学習者の母語が学習言語の作文に与える影響について考える際には、その点に注意が必要である。他の言語の話者のデータについても、ここで行ったような検討が必要である。

ただし、誤解のないように付け加えれば、このことは学習者による母語訳が言語データとしての質が低いことを意味するわけではない。我々自身も日常生活でしばしば経験するように、文章を書くことは一定の技能を要する行為であり、母語であれば何の苦もなく自然な文章が書けるわけでは決してない。また、作文対訳 DB に収録された作文のテーマは「あなたの国の行事について」、「たばこについてのあなたの意見」であり、これは母語で書く場合にも一定の技能が要求されるテーマである。さらに、学習者による母語訳は、日本語作文を母語に翻訳した文章であり、単純に母語で文章を書くのとは勝手が異なる。

これらのことを考慮に入れると、学習者による母語訳が文章としての完成度が低いことはごく自然なことだといえる。むしろ、作文対訳 DB に収録された文章は、我々が日常的に行っている「書く」活動の等身大の姿を反映したものと考えたほうがよい。従来の書き言葉の研究は、完成度の高い文章を主な研究対象にしてきたが、今後は日常的な言語活動の所産としての書き言葉を分析する方法論を確立する必要がある。また、そのためには、「『書く』とはどのような言語活動か」という観点から研究を行う必要がある。これが、作文対訳 DB に収録されたデータが我々に提起する最も重要な問題である。

参考文献

- 伊集院郁子・高橋圭子(2004)「文末のモダリティに見られる“Writer/Reader visibility”－中国人学習者と日本語母語話者の意見文の比較－」『日本語教育』123, 86-95.
- 井上優(2005)「学習者の母語を考慮した日本語教育文法」, 野田尚史編『コミュニケーションのための日本語教育文法』くろしお出版, 83-102.
- 北京日本学研究中心(2003)『中日対訳コーパス (第一版)』
- 李琚寧(2005)『為判断のモダリティの中日対照研究－“可以”と「してもいい」を中心として－』, 北京日本学研究中心博士論文

「XMLによる作文添削情報表示システム」 仕様の発展について

—「添削情報電子化」の真の意義を求めて—

宇佐美 洋・鎌水 兼貴

1. はじめに

1.1. 添削情報を電子化することの意味

2001年に公開した「日本語学習者による日本語作文と、その母語訳との対訳データベース ver.2 (CD-ROM 版)」(以下「作文対訳 DB」)には、日本語教師(日本語母語話者・非母語話者を含む)による添削情報がのべ約 300 編含まれている。これらの添削情報は、「学習者の作文に対し、日本語を母語とする教師・母語としない教師がそれぞれどのような反応を返すか」を見るために収集したものであったが、すべての添削情報は教師が原稿用紙に赤ペンで書き込んだものをスキャナで取り込んだものであり、結果として極めて使いにくいものであった。

添削情報を研究に活用できるようにするためには、何らかの形で添削情報も「検索可能な形で」電子化する必要があった。そのことはまた、大量の添削情報を研究者間で共有し、作文評価・添削方法論に関する共同研究を進めていくためにも不可欠の条件であると考えられた。

そこでわれわれは 2002 年、「XML による作文添削情報表示システム ver.1」を開発し、これによって収集した添削情報の電子化を進めるとともに、実際の作文指導や教員指導の場面においてこのシステムを活用する試みをおこなってきた。その過程において、当初のタグの仕様にはさまざまな問題点があることに気づき、それを改善するための改変を続けてきた。

本論においては、2002 年に発表された「XML による作文添削情報表示システム ver.1」から、2006 年 3 月現在の最新版である ver.3 に至るシステムの改変過程を紹介するとともに、なぜそのような改変が必要であったかについて論じる。

1.2. 添削情報電子化に関わる先行研究

「作文対訳 DB」の先行データベースである故・中野洋氏の「日中作文コーパス」¹(国立国語研究所 1998)では、添削情報は図 1 のような形で、テキストファイルとして電子化されていた。

¹ 平成 10 年度文部省(当時)科学研究費(創成的基礎研究)「国際社会における日本語についての総合的研究」研究班第 3 班：「表記・表現に関する実験的研究」(研究分担者：中野洋)において作成。

ひところ前、日本に名高[く, い, d]、国立国語研究所の学者の中野洋先生が上海にいられて、うちの大学で[講座, 講演, c]を[作って差しあげました, なさいました, d]。「中国流行歌の変化—日中流行歌の対照研究—」というテーマをめぐって、先生は流行歌の三つの特■を取り出して[私たちに, h]説明してくださいました。

<添削コード一覧>

「a」表記, 「b」文体, 「c」語彙, 「d」文法, 「e」敬語, 「f」文化、習慣, 「g」母語の干渉, 「h」表現 (間違いではないが、日本語らしくないもの)

図 1 「日中作文コーパス」における添削情報例

「日中作文コーパス」では、誤用と判定された箇所は[]で囲まれ、修正案が示されるとともに、その誤用がどのような種類のものであるかが<添削コード>として示されている。

しかし「日中作文コーパス」の方式には、いくつかの問題点がある。最大の問題点は、ある誤用をどの「添削コード」に帰属させるかについて、明確な基準を立てることが困難であることである。また、データがすべてテキストファイルとして保存され、データ入力の際特別なツールを必要としないということはデータ共有上極めて大きな利点であったが、極めて読みにくく、入力には慣れが必要になるものと思われた。さらに、オリジナルの文言と修正後の文言、それに添削コードが電子的にはすべて同じように表現されるため、例えば「修正後の文言に対して検索をかけたい」といったようなニーズに答えることは困難であった。

こうした問題点を解決するものとして、われわれは、当時普及初期であったマークアップランゲージ XML²に着目した。

添削情報をマークアップランゲージによって表示する試みは、すでに脇田他(1999)³などによっておこなわれていた。脇田らが開発したネットワーク型添削支援システム CoCoA⁴では、SGML⁵に基づいて開発されたネットワーク添削用マークアップランゲージ CCML(Communicative Correction Markup Language)により、教師が学習者作文に対し付与した添削記号をテキスト内にタグとして埋め込むことが可能となっている。タグの入力は、教師側のタグ入力支援ツール(CoCoA-Editor)によって簡便におこなうことができる。そして学習者側の表示ツール(CoCoA-Viewer)では、タグつき文章を解釈し、従来の赤字添削と同様のものを画面上に表示することができるようになっている。

脇田らのシステムは、電子メールを用いた作文添削が円滑におこなわれるように、という目的のために作成されたものであった。このため CoCoA-Editor, CoCoA-Viewer とともに

² eXtensible Markup Language (詳細情報は W3C を参照。http://www.w3.org/XML/)

³ http://oriko.cup.com/PDF/9901_JSISE.pdf

⁴ 1998年、株式会社エイペックスから「コレット ver.1」という名称で発売された後、1999年に「E-Correct」に名称変更された。現在は Web 版 E-Correct も利用できる(詳細情報は http://www.apex.jp/index.html)

⁵ Standard Generalized Markup Language (ISO8879 / JIS X 4151)

メール送受信機能を備えている。また Viewer では添削前(学習者の書いた原文)・添削後(原文に教師による取り消し線, 修正候補などが付け加えられたもの)・清書後(添削内容を反映させて原文を書き換えて表示したもの)を切り替えて表示できるなど, 電子メール上での添削を支援するツールとして極めて多様な機能を持つものであった。

われわれは, 脇田らのシステムを参考にしながらも, 「添削情報を研究者間で共有し, 作文教育・添削方法論に関する研究を進めていくにはどうすればよいか」ということを考え, 独自システムの開発を進めることとした。

1.3. XML による添削情報電子化

添削コードを独自に決め, オリジナルのテキストに付加するという点では, XML 文書は「日中作文コーパス」と変わらない。しかし, XML では, 外部から表示を制御できるスタイルシート(XSLT⁶)というファイルを利用することができる。修正前後が混在している文章であっても, スタイルシートに表示方法を定義することによって, 修正前の文言, 修正後の文言を明示的に区別して扱うことが可能となる。

表示制御が比較的自由に可能であるため, ブラウザ⁷上に「赤ペンによる添削」と非常に近い形で添削情報を表示させることもできる。さらに, そうした美しい表示を可能にすることで, データとして添削情報を活用するだけでなく, 教育現場において現実の「添削支援ツール」として使用することも可能になると考えた。

その反面, テキスト入力, 「日中作文コーパス」のように, 簡単な記号の入力ではなく, XML タグという, 統一規格にのっとりた方法で入力しなければならない。そのため直接入力することは非常に困難と思われたため, テキストエディタ⁸のマクロ機能によって入力支援をおこなうこととした。

⁶ XML Stylesheet Language Transformations

⁷ WWW 用のブラウザを用いることが想定されている。

⁸ 本研究では「秀丸エディタ」(<http://hide.maruo.co.jp/>)を用いている。

2. これまでの「XMLによる作文添削情報表示システム」

2.1. バージョン1

2.1.1. 初期の目的

添削ツールの開発においてわれわれが企図したことは、教育現場での添削支援ツールとして利用されることを目指しつつも、本質的には添削情報をデータとして大量に蓄積・共有し、添削情報を用いた研究のために用いていくことであった。可能であれば、単に添削記号をタグとして埋め込んでいくというだけでなく、添削研究に役立つような情報をテキスト中に埋め込んでいきたいと考えた。

通常われわれが添削をおこなう際に考えることとして、

- ・「この箇所は明らかにおかしいから、必ず～と直してほしい」と考えて直す場合
- ・「間違いとはいえないが、～と直すとさらによくなる」と考えて直す場合
- ・「～と直すのがいいように思えるが、自信が持てない」と迷いつつ修正する場合

などがある。このような「添削態度」とでもいうべきものをXMLのタグとして区別しておけば、例えば「誤用か正用かの判断に迷う部分」だけを抽出することができ、また同一の誤用箇所に対して、複数の添削者がどのような態度で添削を施しているかを比較対照することもできるようになると考えた。

2.1.2. バージョン1.0のタグの仕様

そこで、2002年に発表した「XMLによる作文添削情報表示システム ver.1」(宇佐美2002)では、以下のようなタグセットを用意することとした。

タグセットは大きく4つのグループからなる。

- A. 明らかな誤用箇所に対するタグ
- B. 明らかな誤用とはいえない箇所に対するタグ
- C. 段落・字下げに対するタグ
- D. 注釈等に対するタグ

それぞれについて、タグの仕様を示す。

A. 明らかな誤用箇所に対するタグ

- ・ del(削除) 書式: 削除する文字列
- ・ put タグ(挿入) 書式: <put value="挿入する文字列"/>
- ・ rep タグ(置換) 書式: <rep value="置換後の文字列">置換前の文字列</rep>
- ・ move タグ(移動)⁹
書式: <movefrom id="IDNo.">移動させたい文字列</movefrom>
<moveto id="IDNo."/>

B. 明らかな誤用とはいえない箇所に対するタグ

- ・ better (誤用とはいえないが、より適切な表現があることを指摘)
書式: <better value="より良い表現">元々の表現</better>
- ・ doubt (執筆者の意図が明らかでなく、暫定的訂正・コメントに使用)
書式: <doubt value="添削者のコメント">意図不明表現</doubt>

C. 段落・字下げに対するタグ

- ・ cp (執筆者が入れた改行位置の指定¹⁰)
書式: <cp/>
- ・ newp (改行すべき場所を指定)
書式: <newp/>
- ・ runon (追い込み: 執筆者が入れた改行を無効にする¹¹)
書式: <runon/> <cp/>
- ・ sperr (字下げエラー: 段落頭で字下げをしていないとき、または字下げをしすぎているときに使用)
書式: <sperr value="執筆者が実際に空けているマス数¹²"/>

D. 注釈等に対するタグ

- ・ remark (添削者から執筆者へのコメント・メッセージ等)
書式: <remark value="添削者コメント">コメントの対象となる表現</remark>
- ・ note タグ(入力者の注¹³)
書式: <note value="入力者注">注の対象となる部分</note>
- ・ review タグ(文章全体に対する総評)
書式: <review>添削者の総評</review>

⁹ <movefrom id="IDNo.">と</movefrom>で挟まれた文字列が、<moveto id="IDNo."/>タグのある位置に移動するよう指定されていることになる。IDNo.の部分には任意の数字(1つの作文内で重複しないように設定)が入る。

¹⁰ 添削者が新たに改行位置を指定したり、執筆者が入れた改行を無効にしたりするため、執筆者が本来入れた改行の位置をタグの形で残しておく必要がある。cpタグは添削者が挿入するものではないため、XML化する際に、ヘッダ・フッタとともに自動挿入される必要がある。

¹¹ runonタグはcpタグの前に挿入することになっている。

¹² 1字の字下げをすべきところではない場合は"0"を、2文字下げている場合は"2"を属性として入力。

¹³ 手書きの添削情報を入力する際、うまく電子化できない現象があった場合、このタグの中にその現象を言語化して説明する。例えば、漢字の間違いを訂正しているような場合、<note value="「降」という漢字の書き方の間違いを訂正している">降</note/>のような書き方にする。

2.1.3. バージョン1.0タグによるXML文書例

また、これらのタグを簡便に入力するために、テキストエディタ「秀丸」のマクロを利用したタグ入力支援ツールも作成した。以下に、タグ付きXML文書の一例を図2に示す。

```
<sperr value="2"/>人<better value="々">達</better>はよく「*<doubt value="?">自分のこのむ</doubt>」と言っています。自分の生活は自分のです。それもいいです<put value="." />けれ<rep value="ど">で</rep>もたばこをす<rep value="う">わ</rep>ると健康*だけ<rep value="が">は</rep>弱くな<rep value="る"わけ</rep>ではありません">りません</rep>。お金もたくさん<better value="使います">用いられます</better>。それとともに次の世代も悪い<better value="影響">こと</better>を受けます。私<doubt value="?">が</doubt>感じるのは">の感じによって</doubt>次の世代にいつもいい物を教える<del>の</del>方がいい<better value="ということです">です</better>。その人々は<better value="年をとって">年上になって</better>自分<put value="の">/>両親やほかの人達を例として<remark value="「どんだん」の方が一般的">どしどし</remark>追いかけます。</cp/>
```

図2 添削XMLバージョン1.0による例

こうしたタグ付きXML文書を、ブラウザ上で手書き添削と同じように表示するためのスタイルシート(correction.xsl)も作成した。

上掲のXML文書を、スタイルシートによって表示させると図3のようになる¹⁴。

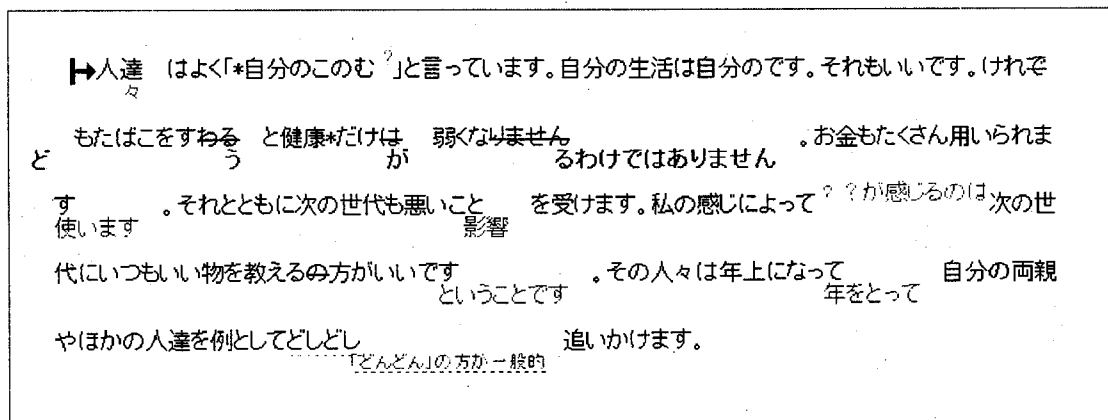


図3 スタイルシートによるXML表示例

図3は白黒での表示であるが、ブラウザ上では、「明らかな誤用」(del, put, repタグ使用箇所)に対して示された修正案は赤字によって、明らかな誤用とはいえない箇所に対する

¹⁴ ただしこのような表示を得るためには、スタイルシート correction.xsl のほかに、画像ファイル (newp.png, note.png, runon.png, sperr.png)が必要となる (例えば図3冒頭の、「字下げ」を意味する矢印マークの画像ファイルなど)。

修正案は青字によって示されることになっている。

2.2. バージョン2

2.2.1. 仕様の変更

バージョン1による、添削XMLの公開後、添削タグの仕様についての問題点を整理して、仕様を改訂することとなった。

手書き添削を電子化することを意識してタグが設定されていたために、直接コンピュータを用いた電子添削をしようとするといくつかの問題点が判明した。こうした経緯からバージョンの改訂作業に入った。

改訂は2005年3月に行われた。最大の変更は以下の2点である。

(1) タグの整理

- a. 削除(del)・挿入(put)・移動(movefrom/moveto)を置換(rep)に統合
- b. 置換(rep)タグを、修正(correct)と提案(better)の2段階でとらえる

(2) タグ挿入時にコメント属性を付与する

添削データの収集を、手書きではなく、電子添削によっておこなうことを考えたとき、手書きの添削の雰囲気を残したままのタグではなく、電子化としてより効率化するために、添削タグの簡素化が計画された。

(1)については、添削作業において修正作業にかかわるものは、文字の置換作業が基本であるとした。削除は空文字への置換であり、挿入は空文字からの置換と等価である。移動も単なる削除後同内容の挿入とみなすことで、すべての修正関係のタグを一つに統一することになった。

移動を廃止することで煩雑になることも予想されたが、1～2文字程度の短い単位での移動は、表現全体の置換として扱われる可能性があり、また、節以上の単位での移動のような文の論理構造にかかわる移動などは制限されると予想し、影響は少ないと考えた。

このように、置換タグが修正に関する全てのタグを担う結果、もともと「より良い表現への置き換え」という意味にしか用いられなかった better にも、挿入や削除の機能が与えられることとなった。本来 better についても挿入や削除が「より良い表現」として示される可能性があったので、この点についてはバージョン1の段階で仕様にはいつているべきだったといえるだろう。こうして、correct と better は2段階の同一タグとして整理されたといえる。

(2)は、作文添削において、作文の電子添削の効果を考える際に、修正を施した箇所に対する説明を添削者に求めることで、遠隔添削のような添削者と学習者とが話し合えないような状況でも、なるべく添削者側の情報を多く提示することを可能にすることが目的である。

「秀丸エディタ」のマクロへの実装においては、修正タグを入力する際、同時にコメントの入力を促すウィンドウを出すようにした。入力の必要がなければそのまま無視してよいが、常に促すことによって細かくコメントがなされることを期待した。

また、コメントが増加することによって、実際の添削画面出力においては、コメントによって画面表示が乱れないように、コメントがある箇所についてのみ示し、そこをクリックすることによって表示されるようにした。

2.2.2. バージョン 2.0 の仕様

こうして、変更されたタグバージョン 2.0 の仕様を以下に示す。改行関係のタグは基本的には変更なかった。

A. 修正関係のタグ

- 存続 • correct (rep からの名称変更)
- better
- unclear (doubt からの名称変更)
- 廃止 • put, del → correct へ統合
- movefrom/moveto → correct へ統合

統合後の修正関係のタグの使用例を示す。以下の correct は better でも同様である。

- 削除(旧 del)
書式：<correct value="">削除する文字列</correct>
- 挿入(旧 put)
書式：<correct value="挿入する文字列"/>
- 置換(旧 rep)
書式：<correct value="置換後の文字列">置換前の文字列</correct>
- 移動(旧 move; movefrom/moveto)
削除と挿入にて対応
- unclear(旧 doubt)
書式：<unclear comment="添削者のコメント">意図不明表現</unclear>

B. 改行関係のタグ

- 存続 • cp
- newparapgraph (newp より名称変更)
- runon

C. 注釈関係のタグ

存続 ・ review(文章全体に対する総評)

廃止 ・ sperr → 電子添削化するため廃止。(原稿用紙の書式に関するため)
・ remark → 廃止。タグ内の comment 属性で対応
・ note → 電子添削化するため廃止

2.2.3. バージョン 2.0 による XML 文書の例

こうして、図 2 と同じ作文例をタグバージョン 2.0 にてあらわしたものを図 4 に示す。

```
人<better value="々">達</better>はよく「*<unclear comment="?">自分のこの  
む</unclear>」と言っています。自分の生活は自分のです。それもいいです<correct  
value="。"/>けれ<correct value="ど">で</correct>もたばこをす<correct  
value="う">わる</correct>と健康*だけ<correct value="が">は</correct>弱くな  
<correct value="るわけではありません">りません</correct>。お金もたくさん  
<better value="使います">用いられます</better>。それとともに次の世代も悪い  
<better value="影響">こと</better>を受けます。私<unclear comment="?">が感じる  
のは">の感じによって</unclear>次の世代にいつもいい物を教える<correct value="">  
の</correct>方がいい<better value="ということです">です</better>。その人々は  
<better value="年をとって">年上になって</better>自分<correct value="の"/>両  
親やほかの人達を例として<unclear comment="「どんだん」の方が一般的">どしどし  
</remark>追いかけます。<cp/>
```

図 4・添削 XML バージョン 2.0 による例¹⁵

タグの機能が統合化されたことにより、rep や put, del といった表現が correct となつてしまったため、機能が分かりにくくなった。図 2 におけるバージョン 1.0 の例に比べて可読性が低下していることが分かる。

¹⁵ 囲み部の unclear タグは、バージョン 1.0 で remark タグの部分であった(本文図 2)。バージョン 2 のタグセットでは、修正を求めない単なるコメントを示すことはできない。コメントはタグに付随する属性としたため、ここでは unclear とした。執筆者の意図が明瞭ではないので、better になる可能性もある。

2.3. バージョン2.1への改訂と評定情報

2.3.1. 電子添削への以降の問題

タグが大幅に整理されたバージョン2.0は、電子添削によるデータ収集をめざして2005年3月以降テストされた¹⁶。

しかし、電子添削のテストの結果、タグが多くなることによって、XMLファイルが分かりにくくなり、入力も困難になることが分かった。テキストエディタのマクロによってタグ入力が支援されているとはいえ、XMLの仕組みに習熟しない添削者にとって、タグ入力はかなりの苦勞を伴う作業であるようだった。

ウェブサイトにおけるHTML文書を、テキストエディタで入力することがわずらわしいといわれるように、元来テキストエディタによるタグの入力には問題があるともいえる。さらに、バージョン2.0においては、タグが整理されたために逆に紛らわしさが増したことや、タグにコメント属性が追加されたことでタグの長さが長くなりすぎるということもあったと思われる。

そのため、バージョン2.0から添削データの収集を、紙面上での手書き添削から、コンピュータ上でのXML文書による電子添削へと切り替える計画をたてていたが、テキストエディタによる添削は困難と判断した。そのため、2005年8月における添削データの収集では、手書きによる添削を依頼することとした。依頼の際には「添削マニュアル」を作成し、バージョン2.0のタグ仕様で容易に電子化ができるような形での添削をお願いした。

添削データ収集作業と平行してタグ仕様の再検討もおこなったところ、「移動」という操作を「削除」＋「挿入」の組み合わせとして表現すると、削除箇所と挿入箇所の間に対応関係がある、ということが分かりにくくなる、という問題点が判明した。そこでバージョン2.1では、correct/betterタグの中にID属性を用いることで、移動機能を復活させた（詳細はp.57の脚注9を参照）。

2.3.2. 評定タグの導入

評定タグとは、作文中にどの程度修正の必要がある誤用か、という「誤りの重篤性」に関する3段階の評定¹⁷情報に関するタグである。

添削とは別に評定情報を収集した理由は、添削者が「誤りに気づいてはいたものの、教養効果に配慮したり、あるいは修正候補が思いつかなかつたり、という理由で修正をしなかった」という箇所があるのではないかと考えたからである。当初この評定情報は、添削情報とは別の派生データベースとして公開する予定であったが、同一評定者が同じ作文に対して付与した添削情報と評定情報とはやはり同じファイル上で統一的に示されたほうが

¹⁶ 2004年度国立国語研究所日本語教育短期研修(第5回)「作文添削の電子化・共有と、それを用いた応用研究の可能性」(2005年3月27日)において、バージョン2.0の公開と、改訂版秀丸マクロを用いたタグ入力のモニターテストが行われた。このほか、2005年度前半にも非公式のテストを実施した。

¹⁷ 添削作業から一定時間が経過した後、添削者に再度同じ作文を渡し、誤りと思われた箇所に下線を引いてもらい、それぞれの箇所について「誤りの重篤性」の度合いを3段階で評価してもらった。

よい、という判断に基づき、添削データベースの中に評価情報も含めることとした。

この思想はバージョン 3.0 へ引き継がれるため後述することとして、ここでは詳述しない。

2.3.3. まとめ

以上、バージョン 2.1 への変更点は、

- (1) 移動用タグの復活
- (2) 評価タグの導入

の 2 点である。バージョン 2.1 において追加された部分について以下に示す。

追加 ・ id 属性の追加

移動に用いる(correct/better どちらも)

(例)

移動元 <correct value="" id="ID No.">移動箇所</correct>

移動先 <correct id="ID No."/>

追加 ・ rating (評価)

<rating value="重篤度" s="開始位置" e="終了位置">該当箇所</rating>

重篤度： 1～3 の 3 段階で示す

開始位置・終了位置： タグが挟んだ範囲のどの部分に対する評価かを示す¹⁸

¹⁸ 添削とは別途おこなう調査であるため、添削情報と組み合わせる際には、評価タグと添削タグの範囲がずれることがある。こうした状態を回避するために、評価タグは添削タグの外側に置くようにして、囲んだ範囲の何文字目から何文字目までに対する評価かを指定できるようにした。

3. XMLによる作文添削情報表示システム バージョン3

3.1. 「添削」という行為自体の問い直し

3.1.1. 許容度の判断基準

バージョン 2.1 までの修正タグでは、「明らかにおかしいから直せ」という意味のタグ (correct)と、「誤りとはいえながより適切な表現があるから直したほうがよい」という意味のタグ (better)とが区別されてきた。つまり、学習者が産出した「逸脱表現」を、添削者はどの程度「許容できるか」が問われていたわけである。

しかし、現実に学習者作文に見られる誤用や、教師の添削行動を分析していくと、correct と better の境目は決してひとつの要因によって決まるものではなく、複数の変数が絡み合っていることが分かってくる。

その変数のひとつとして、「意味の分からなさ」というものが挙げられる。読んでいて意味が取れない箇所というのは、文章として最低の目的を果たしていないわけであるから、通常これは「許容できない」箇所となるだろう。しかし場合によっては、「明らかに規範から外れており、不自然ではあるのだが、しかし意味は明解に理解できる」という場合もある。例えば、

私は学校に勉強します。

という文があったとする。この文では、助詞「で」を使用すべきところ、誤って「に」を使ってしまっているもので、極めて初歩的な文法ミスであるということが出来る。しかし普通に日本語を理解できる読み手であれば、頭の中でたやすく「に」を「で」に置き換えて理解することが可能だろう。この誤用によってコミュニケーション上の支障が生じることはほとんどないといってよい。

こういう場合、「に」の許容度をどのように判定するか、ということについては、かなり大きな個人差が出るものと考えられる。「文章では書き手の意図が伝わるのがもっとも大切である」と考える添削者であれば、これは十分許容できる範囲の誤用ということになるだろうし、「初級で学んでいるはずの文法項目が習得できていない」ということを重く見る添削者であれば、必ず修正されるべき重篤な誤用、と判定することになるかもしれない。

もちろん、添削態度にはさまざまなものがあってよい。しかしもし、添削者が「意味は分かるけれど重篤な誤用だ」と判断したのか、「誤りではあるがコミュニケーション上問題はないから軽微な誤用だ」と判断したのが学習者にはっきり示されれば、それは学習者にとってさらに有効な情報であるといえようし、また添削者本人にとっても、「いま自分はこの『逸脱表現』に対し、どういう基準によって重篤・軽微の判断を下しているのか」に対し意識的になることは、非常に意義あることといえるであろう。

3.1.2. 非修正箇所への注目

また添削者は、許容できない箇所すべてに対し修正を施すわけではない。石橋(2002)は、添削者が作文添削をおこなう際のプロトコルを分析することにより、「添削者は、学習者の逸脱表現に気づき、留意¹⁹しても、その逸脱を『学習者の日本語習熟度では許される範囲のものである』と判断したり、その逸脱に対し適切な対案が思い浮かばなかったりした場合には、逸脱箇所を非修正のまま残すことがある」と述べている。

通常の添削では、このように「問題がありながら非修正箇所とした箇所」と、「問題がない箇所」との間には、表面上まったく区別はなされないことになる。「学習者の日本語習熟度では許される」と判断して修正しないのはよいとしても、「適切な対案が思い浮かばないために直さない」ということは、教育上問題となるかもしれない。対案は示せないまでも、「この部分は～という理由で許容できない」ということを学習者に伝える手段は用意されてしかるべきであろう。

さらに、許容できない箇所に対する修正案が思い浮かんだとしても、執筆者の執筆意図が必ずしも明確に把握できないために、その修正案が適切なものであるかどうか確信が持てない、ということもあるだろう²⁰。そのような場合、執筆者と面談できる立場にいる添削者であれば、執筆者から直接意図を聞き出すことによってより適切な修正をおこなうことが可能であるが、しかし遠隔地にいる学習者とインターネットなどを通じて指導をおこなっている教師や、学習者と直接面談する時間が取れない教師にとっては、「十分確信は持てないが、そのことを明記した上でとりあえず修正案を示す」ということも必要であることが予想される。

¹⁹ ここで「留意する」とは、学習者の「逸脱表現」に対し、「それを修正すべきものと認める」ということを意味している。

²⁰ バージョン2では unclear タグで示していた事項である。

3.2. バージョン3.0における添削タグの入力順序

そこでわれわれは、具体的には以下のようなことが可能となるように、添削情報表示システムを改訂することとした。

- 学習者の文章の中で「逸脱表現」と認定した箇所について、「意味が分からない」という理由での逸脱なのか、それ以外の要因による逸脱なのかを明示できるようにする。
- 仮に修正案が思い浮かばなくても、「逸脱」と認定した箇所は明示できるようにする。
- 修正案に確信が持てない場合も、「確信が持てない」ことを明記した上でとりあえずの修正案が示せるようにする。

こうした複雑な判断を添削者に課す上で、われわれは、添削者が入力支援ツールを用いて添削タグを挿入していく際、以下のような「一定の手順」を設定することにした。

- (1) まず、文章中の「逸脱箇所」を範囲指定し、その箇所に対する「許容度」(どの程度許してよいか)を3段階で評定する。すなわち:「必ず修正すべき」を3とし、「逸脱してはいるが、必ずしも修正しなくてもよい」を1、その中間を2とする(この数値は、修正タグ内の属性値のひとつとなる)。
- (2) その箇所に対し、対案が提示できればそれを入力し、できなければ入力しないままにする。対案が提示できない場合、範囲指定された箇所に対しては「許容度情報」のみが属性値として表現される。
- (3) 対案を提示できた場合は、その対案に対する「確信度」を2段階で指定する。すなわち:確信をもって対案を示している場合は2を、確信はないが暫定的に対案を示している、という場合は1を入力する。対案が提示できた場合、範囲指定された箇所には修正タグが付与され、(1)で指定した「許容度情報」が属性値として表現されるとともに、対案の「確信度情報」もまた属性値として表現されることになる。

この過程をフローチャートの形で表すと以下のようになる。

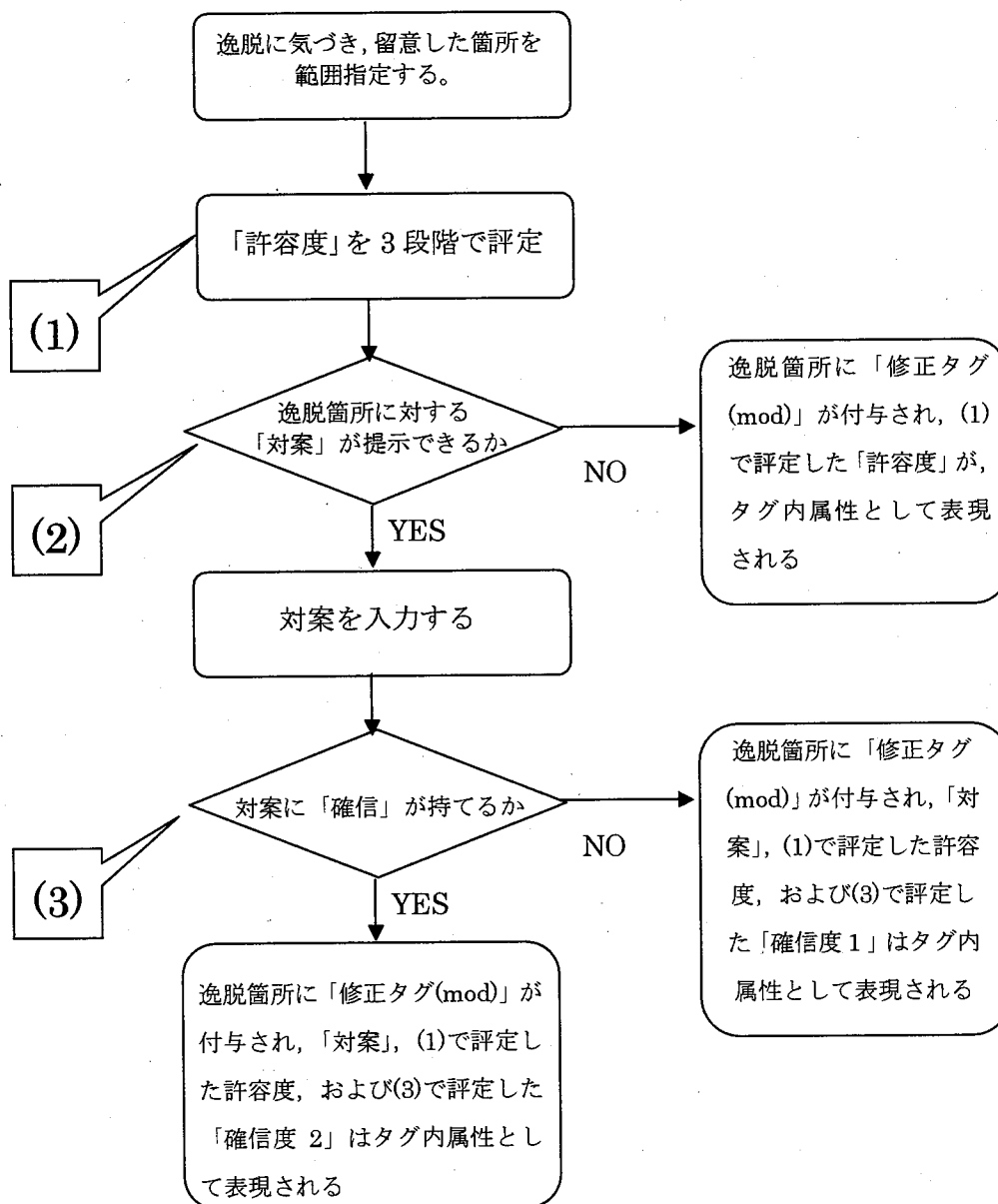


図 5 バージョン 3.0 における添削タグ入力のフローチャート

上記(1)~(3)の各ステップについて解説する。

(1) 「許容度」

添削者はまず, 全体的な「許容度」の判定を求められることになる。この段階において

は、「分かりにくさ」「不自然さ」などに関する分析的な判断はおこなわず、総合的判断としての「許せる度合い」が評定される。

(2) 「分かりにくさ」

「意味の推察が可能であるか」、つまり「分かりにくさ」の判定が求められることとなる。意味の推察が不可能なほど分かりにくい箇所であれば、当然対案提示は不可能となる。

一方で、「意味内容推察」はできるが「対案提示」は不可能、という場合もありうる。「いいたいことはよく分かるのだが、それを日本語でどのように表現していいかまったく思いつかない」、という場合である。

つまり、「対案提示不可」という場合には、「意味内容が分からないからできない」という場合と、「分かっているができない」という場合とがあることになるのだが、いずれの場合も教師として「対応不可能」という点では同じであり、学習者から見た場合、この両者を区別することにはあまり意味はないものと考えられる。そこで「対案提示」が可能か不可能か、という観点での判定だけを求めることとした。

(3) 「確信度」

対案を提示できた場合には、その対案にどの程度の確信を持っているのか、ということをも2段階で評定してもらうこととした。

手書きによる実際の添削情報を詳細に見ていくと、「執筆者の真意が十分に分からないため、対案に十分な自信が持てないのだが、それでもとりあえず対案は示しておく」という態度での添削が決して少なくないことが分かる。例えば、クエスチョンマークつきで対案を示したり、対案を示したあとで「～ということですか?」という「執筆者に対するメッセージ」を示したり、というような形での添削である。もちろん、「このようなあやふやな態度での修正はすべきでない」という考え方もありえよう。しかし実際にこういう添削行動が見られる以上、電子添削においてもそれに対応できるような準備は必要であろうし、また執筆者と直接対面できない状況での添削においては、「確信は持てないがとりあえず対案を示しておく」ということの必要性も決して否定はできない。ただしその際、対案が「確信が持てない状態で提示されたものだ」ということは、学習者に対し明確に示されるべきであろう。このような考えに基づき、「対案の確信度」を添削タグの属性値として表現することとした。

次に、こうした手順によって入力されるさまざまな添削タグの仕様について、詳細を説明する。

3.3. バージョン3.0の仕様

3.3.1. 属性仕様

バージョン3.0では、添削過程よりむしろ思考過程の整理に重点がおかれたため、これまで以上に修正関連のタグは整理され、共通属性が体系化された。

修正関連のタグは実質的に1つとなり、属性値によってその添削機能を示す形へと変化した。そのため、タグ自体の重要性が低下し、むしろ属性の解説を中心に行うこととする。

修正関連のタグがとりうる属性値は以下の5種類である。

a 属性(Acceptability) 当該箇所に対する許容の度合い

s 属性(Sureness) 添削者が、自ら提示した修正案に対し持っている確信の度合い

r 属性(Replace) 提示した置き換え文字列

c 属性(Comment) 添削者によるコメント

id 属性(Identification Data) 移動やコメント時に用いる参照用記号

(1) a 属性

【属性名】 acceptability, 許容度

【対象タグ】 mod, newp, runon

【機能】 許容度の評価

【値】 数値(3段階: 1,2,3)

a 属性は「acceptability(許容)」, すなわち、文章の「許せる度合い」に関する評価をあらわす。添削者にとって、該当箇所の「許せる度合い」を3段階で評価する。最も許容できない状態が3で、ある程度許容できる状態が1となる。

従来タグとして使用されてきた correct と better も、許容度という点以外での違いはなく、修正タグとしての動作は、挿入・削除・置換・移動とすべて同じである。そのため、タグの区別ではなく、属性による違いで示すことが可能であると思われる。

段階については、correct, better の2段階ではなく、迷った場合も考慮して3段階とした。旧タグからの変換に際しては、correct が 3, better が 1とした。

(2) s 属性

【属性名】 sureness, 確信度

【対象タグ】 mod

【機能】

【値】 数値(2段階: 1,2)

s 属性は、「sureness(確信)」, すなわち、提示した対案への確信の度合いをあらわす。確信度は2段階とし、確信の有無のみで判断する。確信がある場合が2で、ない場合が1である。そもそも対案が提示できない場合には、この属性自体が存在しないことになる。

修正候補が明確な場合(バージョン2での correct, better タグ)には、確信度については許容度にかかわらず2(確信あり)となる。

問題となるのは、バージョン1の doubt タグや、バージョン2の unclear タグのコメントとして示されていた対案である。意図が不明瞭な箇所について、添削者が文脈などから推定して修正案を提示することもある。しかしそうした添削者にとって確信のもてない修正案も、コメントに含めており、参考程度の扱いであった。確信がもてない理由は複雑で、学習者の作文側に問題がある場合もあれば、添削者が修正案の提示に自信がもてない場合もあり、通常どちらか一方に原因を求めることは難しい。添削者側も明確にどちらが原因か判断できるとは限らない。

しかし、たとえ修正案が不確実であったとしても、その提示は、置換・挿入・削除という、従来の correct や better タグと同じ働きをしていると思われる。そのため「対案提示の際の確信」という観点から、従来の correct/better と同等に扱うことにした。中間段階を設けず2段階としたのは、不明瞭な状態をさらに個人の中で分類することは困難であると想像したからだが、実際のテストでの状況次第では段階が追加されることもありうる。

(3) r 属性 – replace

【名称】 replace (置換)

【機能】 修正の提示

【値】 修正候補の文字列

【対象タグ】 mod

従来の correct, better タグにおける value 属性からの名称変更である。置換後の内容であるため、replace とした。この属性が空である場合、自動的に削除とみなされるが、もし修正候補が思いつかないまま入力を忘れてしまった場合に削除と区別がつかないことになる。教育効果を考えて、あえて対案を示さないというような場合には、削除と誤解されないようにしなければならない。

そのため、入力ツールで実装する場合には、置換の場合には、置換後の内容が入力されるまでは、空ではなく何かダミー記号を与えておく必要があると思われる。

(4) c 属性 comment

【名称】 comment (コメント)

【機能】 添削者註

【値】 文字列

【対象タグ】 mod, newp, runon, com

コメント属性には、添削者のコメントが入る。添削タグや、コメント専用のタグにおいても用いられる。修正関連のタグには全てつけられるほか、コメント専用のタグももうけた。これは特定の箇所に対する感想・メッセージ等に用いられる。そのためプラス評価の

コメントもここに属する。プラスの評価については、作文添削の効果を考える上で重要な要素になると思われるため、改良の余地があるだろう。

(5) id 属性 ID, identification data

【名称】 ID

【機能】 参照場所の指定

【値】 文字列

【対象タグ】 mod

mod タグでは、移動時に用いる。移動は、バージョン 2.1 と同様に、ID 付の削除・挿入となる。挿入時には r 属性はなしで、移動元と同じ ID 属性を示しておけば自動的に移動元でタグに挟まれた内容を挿入することができる。

移動については、長い作文の場合には、文章の構成にともなって、複数の文や、段落といった単位での移動も必要になることがある。これは文中での添削とは別のレベルの問題となるため、このような移動にどのように対応するかは今後の課題である。

3.3.2. タグ仕様

以上の属性値をとるタグについて解説する。

A. 修正関連タグ

mod

【名称】 modification, 改良

【属性】 a 属性, s 属性, r 属性, c 属性, id 属性

【使用法】

mod は、改行関係以外のすべての箇所に関する何らかの修正をおこなう場合に使用する。バージョン 2 までの correct, better, unclear の全てが mod タグに統合される。従来の correct で、

```
<correct value="置換内容" comment="コメント">該当箇所</correct>
```

と表示されていたものは、

```
<mod a="3" s="2" r="置換内容" c="コメント">該当箇所</mod>
```

へと変換される。

バージョン 2.0 における修正関連タグを mod タグで表現すると、

correct : <mod a="3" s="2" r="置換" c="コメント">

better : <mod a="1" s="2" r="置換" c="コメント">

unclear : <mod a="1" s="1" r="修正候補の部分" c="修正候補以外の部分">

unclear : <mod c="コメント">

となる。

B. 改行関連タグ

cp

【名称】 Change Paragraph, オリジナルの改行

【属性】 なし

【使用法】 <cp/>

XML では、原文テキストに多くのタグが挿入されるため、みかけは大きく変更される。タグの構造を明確にするために改行を用いてしまうと、もともと存在した改行記号との区別がなくなってしまう。そのため、表示上の改行は、別に cp として定義されている。

newp

【名称】 New Paragraph, 改行挿入

【属性】 a 属性, c 属性

【使用法】 <newp a="許容度(1-3)" c="コメント"/>

runon

【名称】 Run On, 追い込み

【属性】 a 属性, c 属性

【使用法】 <runon a="許容度(1-3)" c="コメント"><cp/></runon>

newp と runon は、改行タグを挿入・削除するという点では、通常の文字の添削行為と同様であり、mod タグで扱うことも可能である。しかし、r 属性(置換)内にタグを用いなければならないことや、対案の提示という点では、対案内容が「改行」に限定されるため、s 属性(確信度)の値は 2 に固定されてしまう。そのため現状では、従来のタグを存続することにした。

ただし runon については、cp タグを消去することを明確にするため、cp タグの前に置くのではなく、cp タグを runon タグで挟むことにした。このため、runon については、mod タグで表現可能な状態にある。この点は今後の課題である。

C. コメント関連タグ

com

【名称】 comment, コメント

【属性】 c 属性

【使用法】 <com c="コメント"/>

<com c="コメント">該当箇所</com>

コメント専用のタグである。文全体の総評を述べる場合と、文中の注のような形で述べる場合とがある。階層的には、総評タグは文全体を囲んだコメントタグと考えることができるが、現段階では別のタグとした。

コメントの場合には、必ずしも誤りであることが前提ではないが、自然さの属性をつける。プラスのコメントの場合も存在するため、独自の属性が必要になる可能性もある。

3.3.3. バージョン3.0によるXML文書例

以上が、バージョン3.0のタグセットである。図2～図4で示したのと同じ添削情報を、バージョン3.0のXML文書にした例を示す。

```
人
<mod a="1" s="2" r="々">達</mod>
はよく「*
<mod c="?">自分のこのむ</mod>
」と言っています。自分の生活は自分のです。それもいいです
<mod a="3" s="2" r="。"/>
けれ
<mod a="3" s="2" r="ど">で</mod>
もたばこをす
<mod a="3" s="2" r="う">わる</mod>
と健康*だけ
<mod a="3" s="2" r="が">は</mod>
弱くな
<mod a="3" s="2" r="るわけではありません">りません</mod>
。お金もたくさん
<mod a="1" s="2" r="使います">用いられます</mod>
。それとともに次の世代も悪い
<mod a="1" s="2" r="影響">こと</mod>を受けます。私
<mod a="1" s="1" r="が感じるのは">の感じによって</mod>
次の世代にいつもいい物を教える
<mod a="3" s="2" r="の"></mod>
方がいい
<mod a="1" s="2" r="ということです">です</mod>
。その人々は
<mod a="1" s="2" r="年をとって">年上になって</mod>
自分
<mod a="3" s="2" r="の"/>
両親やほかの人達を例として
<com c="「どんどん」の方が一般的">どしどし</com>
追いかけます。
<cp/>
```

図6 バージョン3.0によるXML文書の例

もはや XML 文書から原文や添削文を読み取ることは不可能に近い。そのため、むしろ XML 文書としての分かりやすさを優先して、タグごとに改行を挿入した。スタイルファイル(correction.xml)により、図 3 と同じ表示は得られるが、入力方法については検討が必要となる。

3.4. ソフトウェアへの実装

バージョン 3.0 の改訂により、XML ファイルは原文を読むものではなくなった。また、属性の増加や、入力順序の規則ができたため、入力方法も複雑化した。このため、電子添削はもとより、手書き添削の電子化においてもテキストエディタ(秀丸エディタ+入力マクロ)を用いた入力は困難になってきたと思われる。

このため、入力のための専用ツールが必要となる。実装方法としては、CoCoA(E-Correct)のような、WYSIWYG 方式²¹が望まれる。しかし、専用のソフトウェアは、利用者に配慮したインターフェースの設計や、使用するコンピュータの種類やバージョンに考慮したプログラミング、また、継続的な保守のコストなどの問題点がある。

また、添削システムがどのような場面で利用されるのか、なども考慮しなければならない。

以上はバージョン 3.0 のタグセットを評価する上で急務の課題である。2006 年 5 月頃までに Windows 上で動作する添削入力用プログラムの試作版を作成し、タグセットの有用性のテストや、プログラムの利用上・作成上の問題点の整理をおこなうことで、添削システム全体の改良を続けていく予定である。

²¹ 編集画面に表示されるものとほとんど同じ表示が、最終結果としても得られるということ。通常、テキスト表現で、グラフィカルな表現を生成することは困難である。そのため、入力時にグラフィカルな表現を可能にすることは、使用者に対して優しい作りになる反面、フォーマットなどの表示制御情報を隠すようにプログラムを作成する必要がある。

4. まとめ

バージョン 1.0 から現在の 3.0 までのタグの変遷を、以下に示す。

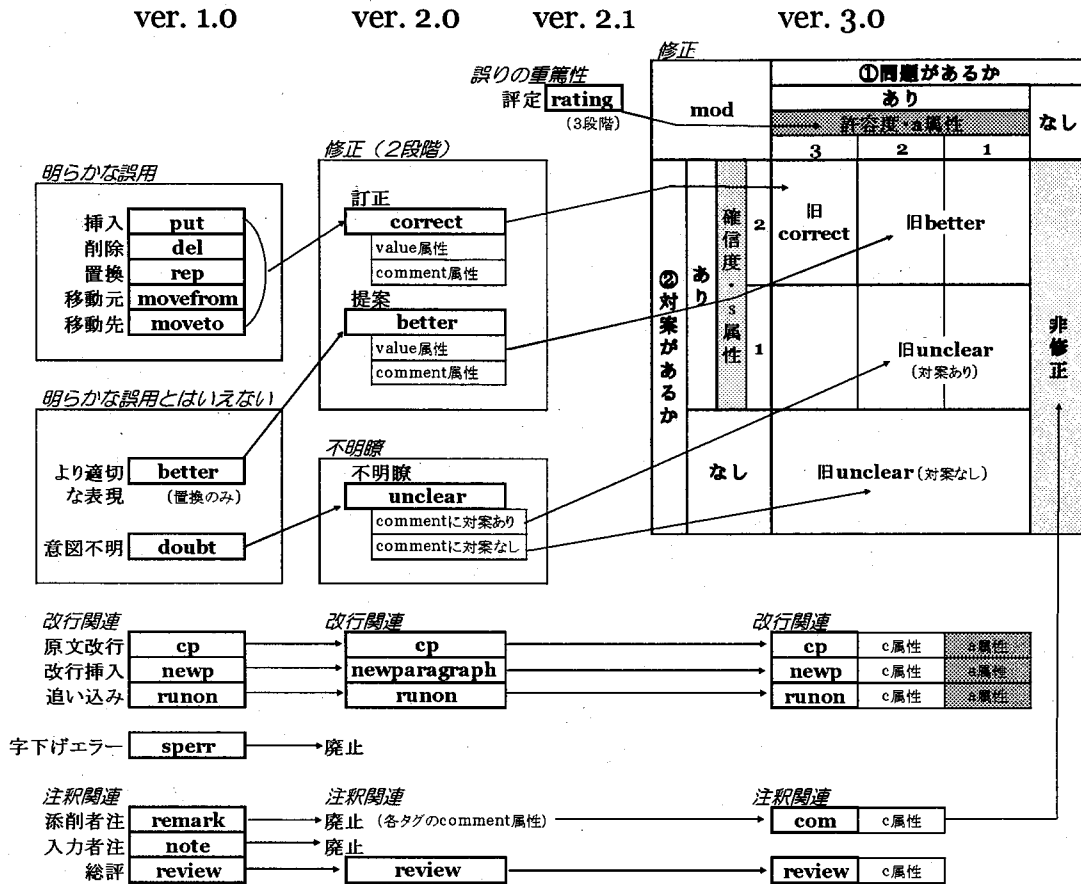


図 7 バージョン 1.0~3.0 の添削 XML タグの変遷

タグセットの変遷を時間を追ってみていくと、「逸脱表現」を大きく二つに切り分ける際の境界が少しずつずれてきていることが分かる。バージョン 1 では、`put`, `del`, `rep` といった「明らかな誤用」を表すタグ群と、`better`, `doubt` のように「それ以外」のタグ群との間に大きな境界線があった。しかし、「明らかな誤用」とは何かを決める際にはさまざまな要因が複雑に関わっていることが予想され、この境界は必ずしも明瞭なものではなかった。

そこでバージョン 2 では、「確信のある対案提示」ができる場合(`correct`, `better`²²)と、そうでない場合 (`unclear`)との間に境界を移すことになった。添削者本人にとって「確信ある対案提示」ができるかどうかは、明らかな誤用かどうか、という基準よりは明確なものであらうと思われたからであった。

²² `better` タグは、「誤りとはいえないが直したほうがよりよい」ということを示すタグであって、添削者は提示された対案に対しては確信を持っていることになる。「確信のない対案」を `better` としているわけではないことに注意。

しかしながらバージョン2においては、unclearは「対案があるが確信が持てない場合」と「対案がまったくない場合」の両方をカバーすることになっていた。対案が出せるか出せないか、という境界は添削者側にとっても学習者側にとってもかなり重要なものであり、これらを区別できないことはやや大きな問題であると思われた。

このためバージョン3では、「対案がまったくない」のか、「確信の有無にかかわらず、曲がりなりにも対案が出せる」のか、というところに大きな境界線が移動することになった。さらに、対案がある場合はその対案に確信があるのかどうか、ということが区別され、加えて、そもそも元の逸脱箇所をどの程度許せるのか、という度合いが3段階で区別されることになった。

こうしたタグセットの変遷は、「誤用・逸脱表現」というものをどうとらえるか、ということの考察が徐々に深化してきたことを示している。バージョン3にいたっては、「どの程度許容できるか」「対案が出せるか」また、「対案を出せる場合、その対案に確信が持てるか」という複数の判断軸を立て、添削者にはそれぞれについて適切に判断をしてもらうような「手順」までが指定されることになっている。その手順を追うことによって添削者は、「なんとなくおかしい」ということでなく、「どういうところが、なぜおかしいのか」ということを明確に意識しながら添削を進めていくことができるようになり、かつ「誤用・逸脱表現」に対しどのような判断を下したかということが、記録として明示的に残せるようになった。そのことは、従来の添削方式ではまったく期待できないことであった。

バージョン2までのタグシステムは、「手書きでもできることをコンピュータ上で再現できるようにしたに過ぎない」という見方もできるかもしれない。それがバージョン3に至って初めて、コンピュータ添削ならではの利点を、教育・研究の両面にわたって提示することができたものとわれわれは考えている。

参考文献

- 石橋 玲子(2002)「日本語学習者の産出作文に対する教師の修正及び非修正行動」『言語文化と日本語教育』23, 1-12, お茶の水女子大学日本語文化学会研究会
- 宇佐美 洋(2002)「XMLによる作文添削情報表示システムと、その応用」『日本語教育学会春季大会予稿集』, 231-232.
- 国立国語研究所(1998)『日中作文コーパス』,「国際社会における日本語についての総合的研究」研究班第3班:「表記・表現に関する実験的研究」文部省科学研究費 創成的基礎研究(研究分担者:中野洋)
- 脇田 里子・緒方 広明・矢野 米雄(1999)「作文教育のためのネットワーク型添削支援システム CoCoA の実践と評価」『教育システム情報学会誌』15(4), 270-275.

謝辞

このシステムは、国立国語研究所 日本語教育部門の研究補佐員・アルバイトの多大なる協力を得て開発され、発展してきました。特に井上義勝氏、中山智哉氏、傳智之氏の3氏に対し、この場で篤くお礼を申し上げます。

分析編

フランス語母語話者の日本語作文における「意図不明表現」の分析 —母語訳との対照から見る「分かりにくさ」の理由—

宇佐美 洋

1. はじめに

筆者がかつてフランス語を母語とする学習者に日本語の作文を教えていたとき、以下のような文に出会って大いに戸惑ったことがあった。

[1] 確信は人間的なものではないが…

「人間的」という語は、日本語としてあまり熟したものともいえないが、一般には「人間としての情合（ジョウアイ）があるようす」（『三省堂国語辞典 第二版』）というような、どちらかというとも人間の性格・人格を表現するために使われる語であろう。「確信」という抽象名詞に「人間としての情合がない」というのは、いったいどういうことなのであろう。

これはまさに、「学習者の産出した日本語を見るだけでは、その執筆意図が分からない例」のひとつといえよう。

この文を書いた学習者には次の授業の際、この文の「執筆意図」について詳しく聞くことができた。話を聞くと、どうやらこの学習者は、「人間にとって、確信する、ということはなかなかできないものだ」ということが書きたかったらしいことが分かった。

そこで筆者は、上記の文を

[2] 「確信する」ということは、人間にはなかなか難しいが…

と修正するとともに、[1]のような表現は、この学習者の母語であるフランス語の発想法に強く影響を受けて生まれてきたものではないかという疑いを持った。

「人間的な」という日本語をフランス語に再翻訳すると、*humain* となるだろう。この語の用法を調べてみると、複数の仏和辞書には *C'est humain./C'est une réaction humaine.* という例文が挙げられ、そこには、

[3] それは人間にはよくあることだ(大修館書店『新スタンダード仏和辞典』)

[4] それは人間なら無理もない(態度だ)(三省堂『クラウン仏和辞典』)

という訳文が添えられている。

つまりこの文脈において、フランス語の *humain* という形容詞は、『人間』という存在と結びついた(あるいは結びつきがちな)属性や特徴を広くカバーする意味領域を持ち

合わせているということが出来る¹。一方日本語の「人間的」という語は、辞書の語義説明に「情合」という語が含まれているところからも分かるように、人間の「心情」と結びついた、はるかに限定的かつ具体的なイメージを分かちがたく身にまもってしまっている。

そして読み手は、日本語の「人間性」という語によって、「情合がある」「人間味がある」という具体的な意味をいったん頭に刻み付けてしまうと、その意味では文脈には適合しないということは分かっても、刻み込まれた意味の具体性を離れ、「人間存在と結びついた属性」というもう一段階抽象的な意味へと抜け出て再解釈することはなかなか困難であるように思える。

この学習者に教えていた当時、筆者はすでにフランス語を多少知ってはいたが、それでも学習者本人に話を聞くまでは、「人間的」を「人間にはよくあること」と解釈しなおすことはできなかった。フランス語を知らない教師であれば、それはなおさらであろう。こうした「語の意味の抽象性の違い」によって生じる逸脱表現というのは、コミュニケーション上かなり厄介なものといえることができる。

学習者によって書かれた日本語の文章には、さまざまな文法的・語用論的「逸脱表現」が含まれている。こうした逸脱の中には、容易に訂正可能なものもあるが、ある種の逸脱については、例えば上述の例のように、日本語教師にさえ学習者の真意を測りかねることがある。このような逸脱は、「情報伝達」という言語の基本的機能を阻害する、という点で、特に重篤度の高いものといえることができる。教師としては、こうした「重篤度の著しく高い逸脱」の背後にどのような要因が存在するのか、少なくともこの種の逸脱が生じないようにするためには、教育上どのような工夫が必要になるのか、ということについて考察しておく必要があるだろう。

本論では、フランス語母語話者による日本語作文を対象として取り上げ、日本語教師にさえ作文の真意が測りにくい場合はどういうときであるか、そしてその「真意の測りにくさ」の背後に、母語干渉をはじめとする「言語的要因」がどの程度存在しているのかを明らかにする。

2. 使用データ

使用データとしては、『日本語学習者による日本語作文と、その母語訳との対訳データベース・オンライン版』（国立国語研究所）に収録されている、フランス語母語話者のデータ 52 編を使用した。このデータベースのオンライン版には、

A 日本語学習者によって書かれた日本語の文章

¹ *Grand Larousse de la langue française en sept volumes* における *humain* の語義説明は以下のとおり： Qui appartient à l'homme, qui est propre à l'homme en tant qu'être animé, doué de raison (par opposition aux autres espèces animales. [人間に属する～, 理性のある生物としての人間に固有の～ (他種の動物と対比させて)]

B 文章執筆者本人による、A の母語への翻訳

が収録・公開されており、これを主たる分析対象とした。

以下、それぞれのデータについて詳細に説明する。

2.1. フランス語母語話者による日本語作文

国立国語研究所が指定した数種類の作文テーマの中から、学習者に任意のものをひとつ選んでもらい、そのテーマに基づき 300 字から 800 字程度の日本語作文を書いてもらった。今回の分析では、「日本語学習において、ワープロソフトを活用することについてどう思うか」というテーマ²について書かれた作文 52 編（タイトルも 1 文と数え、746 文）を使用した。

執筆者は主としてパリの大学で日本語を学ぶ学生で、日本語学習歴は 1 年から 3 年程度。初級段階を終わり、辞書を使えばある程度まとまった分量の文章を書けるレベルには達しているが、文章中にかなりの頻度で逸脱表現・不自然表現が見られ、文章による伝達能力は決して高いとはいえない。

執筆の際には辞書等を使用してもよいものとした。辞書使用の有無、執筆にかかった時間の情報もデータベースには収録されている。

2.2. 執筆者本人によるフランス語訳

日本語作文を書き終わった後で、執筆者本人にそれをフランス語に訳してもらった。意訳や大意要約ではなく、日本語で書いたことをできるだけそのままフランス語でも書くように指示したため、フランス語の文章としてはやや不自然な点もある。若干の例外を除いて多くの場合、日本語の文とフランス語の文はほぼ 1 対 1 で対応している。

執筆順序としてはあくまで日本語→フランス語の順であるが、日本語表現の中に、フランス語表現の影響によるものではないかと思われる箇所は少なくない。これは日本語作文の執筆時にも、フランス語での表現を念頭に置きつつそれを日本語に置き換えていたことを強く思わせる。

3. 分析対象箇所の確定

前述のとおり今回の調査の目的は、

- ・フランス語母語話者の日本語作文の中に見られる各種逸脱表現の中で、「日本語教師にもその意図を測りかねるような表現」（以下、「意図不明表現」）としてはどのようなものがあるか。

² 作文課題は Appendix 参照。

- ・コミュニケーション上深刻な不全をもたらすそのような表現の背後に、どのような要因が存在するのか。
- ・それらの要因の中に、フランス語の干渉も含む「言語的要因」によるものはどの程度存在するのか。また「言語的要因」の中で、教育上特に重要と考えられるものはなにか。

を明らかにすることである。

まず、分析対象を以下のような手順で確定することとした。

1) 3名の評定者による評定

筆者を含む3名の評定者³が、フランス語母語話者による日本語作文52編全編に目を通し、

- ・前後の文脈を参照しても、表現意図がほとんど推測できない箇所
- ・前後の文脈を参照することで表現意図の推測はおおむね可能であるが、その推測に確証が持てないか、あるいは複数の推測可能性のある箇所

を抽出した。範囲指定は評定者の自由に任せた⁴。また抽出時にフランス語訳は参照しなかった。

ここではあくまで「文脈の中で文意が理解できるか」というところに焦点を当てて抽出をおこない、「規範からずれている」「初級レベルの文法項目なのにうまく使えていない」「違和感がある」というような点は考慮しなかった。つまり、規範からの逸脱があったり、文体上きわめて強い違和感があったりしても、「前後の文脈を参照すれば執筆意図が一意にかつ容易に解釈され、他の解釈はほぼありえないと考えられる箇所」は抽出対象とはしなかった。

2) 評定結果のつき合わせによる「意図不明焦点」の特定

3名の評定結果をつき合わせ、3名のうち2名以上が一致して「明瞭に文意を推測することが不可能」と判定した箇所を、今回の調査の分析対象と認定した（つまり、1名のみが「意図不明」または「意図推測困難」と判定した箇所は分析対象とはしない）。

前述のとおり、「意図不明箇所」の範囲指定は評定者の自由に任せてあるので、ひとつの「意図不明表現」に対する範囲指定は評定者によってずれており、また「意味の分かりにくさ」が生じた理由も、必ずしも指定された範囲内に存在するとは限らない。

³ 筆者を含む2名は日本語教育経験者、もう1名は、日本語教育経験を持たない日本語研究者である。

⁴ 例えば「ワープロソフトを使ったらだいじょうぶで、使わなかったらじしょをえらびます」(fr002)という文について、意図不明の範囲を一定の基準によって決めてしまうことは困難である。したがって評定者自身の判断により、文全体を選択しても、「じしょをえらびます」の部分だけを選択しても、あるいは「えらびます」だけを選択してもかまわないものとした。

そこで、「2人以上の評定者の指定範囲が重なった部分」を「意図不明焦点」として認定し、「意味の不明さ」が生じた理由は、焦点の前後を適宜参照しつつ分析することとした。この作業により、52編の作文から、93箇所(89文)の「意図不明焦点」を抽出した。

4. 「意味の不明さ」が生じた理由の分類

4.1. 教育改善に結びつく「意図不明表現」と、そうでない「意図不明表現」

上記で取り出された「意図不明焦点」を含む日本語の文と、それに対応するフランス語の翻訳文を比較対照し、「意味の不明さ」が生じた理由の分析をおこなうこととした。

ところで、ひとくちに「意図不明表現」といっても、なかには「教育の改善」に向けての有効な提言に結びつきやすいものとそうでないものがある。例えば

ア) 母語訳を参照しても理解困難な意図不明表現

については、学習者自身が考えを十分にまとめきれていなかったり、非常に個人的な思考フレームに基づいて文章が構成されていたり、という理由によって生じたことが想定され、言語学的なアプローチのみによっては手当てすることが困難である。

また、母語訳を参照することによって執筆意図は分かったとしても、

イ) (おそらくは日本語能力の絶対的な不足により) 日本語表現の中に必要な説明を十分に盛り込むことができていないために生じたと考えられる意図不明表現

についてもア) 同様、言語学的要因によっては説明のつきにくい箇所であり、また

ウ) 学習者自身の単純な勘違い等に基づいており、ひとこと指摘すれば再現性は低いと考えられる意図不明表現

は、あくまで個別的な現象であって他への応用が利かず、この種の表現を多数収集しても言語研究上資するところは薄い(吉川 1997:45)と考えられる。またこの種の「意図不明表現」は学習者の母語を問わず、また学習レベルを問わず生じうるものと考えられる(母語話者にもありうるかもしれない)。

そこで今回の調査では、

エ) 主として言語上の要因によって生じたと考えられる「意図不明表現」であって、ある程度再現性があると考えられるもの

を主たる分析対象として設定するが、ア)～ウ)も無視してしまうのではなく、それぞれの

タイプが「意図不明表現」全体の中でどの程度の頻度で現れているのかを把握しておくこととした。

4.2. 「意図不明表現」の分類

そこで今回の調査では、母語訳を参照することによって、「意図不明表現」を、以下のような手順で分類することとした。

- 1) 母語訳⁵を参照することで、執筆意図が分かるか
- 2) 執筆意図が分かった場合、「分かりにくさ」が生じた理由を言語的な要因によって説明ができ、かつ「再現性」があつて、教育上特別な手当てが必要と考えられるか

この判定作業により、「意図不明表現」を以下のように「グループA」「グループB」「グループC」の3つに分類した。

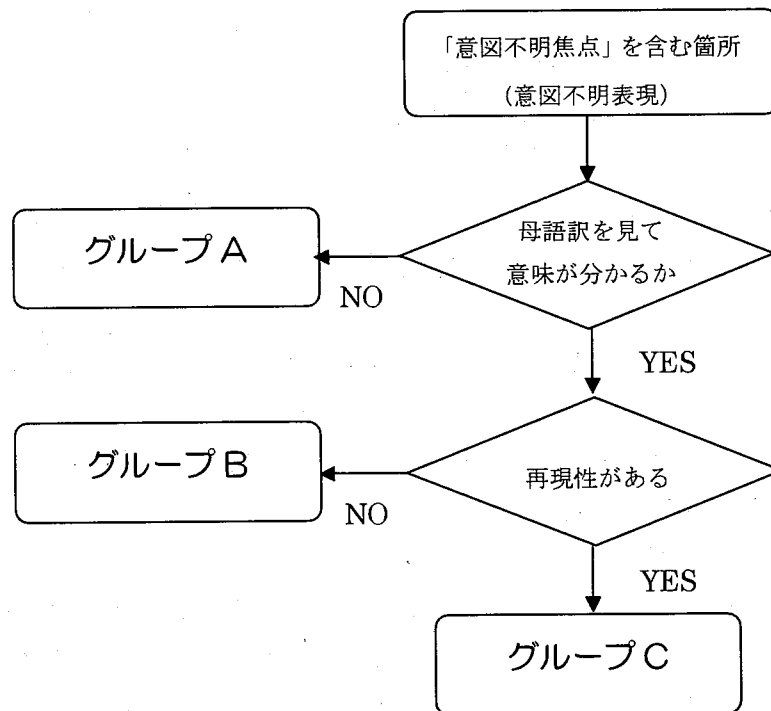


図1 「意図不明表現」分類のためのフローチャート

「グループA」は、「日本語作文を読んでも意味が分からないが、フランス語も日本語のほぼ直訳となっている、というような理由により、やはり意図不明」というものである。

例としては以下のようなものがある（なお以下の日本語例文において、「意図不明焦点」

⁵ 「対応する母語訳が存在しない」場合、「日本語作文と母語訳とが対応していない場合」は、「母語訳を参照しても意図が分からない」ものとして扱った。

として抽出された箇所には下線を引いた。

[5] 新しい各進歩は知識機構に置く石のようだと思う。(fr087)

[6] Chaque nouveau progrès est une pierre portée à l'édifice de la connaissance.

ここでは「進歩」が「石」である、という比喩が用いられているのであるが、前後の文脈を併せ読んでも、なぜ「進歩」が「石」に喩えられるのか、その理由は分からない。このような表現が生まれてくる理由は一概には決められないが、「そもそも執筆者自身が自分の考えを十分にまとめきれていない」「きわめて個人的な思考フレームに基づいて文章を書いている」ということなどが主要な理由として想定できる。これは 4.1 節で挙げたア) にあたるものである。

「グループ B」は、「フランス語訳を読むと執筆意図はよく分かるのだが、日本語の分かりにくさが、学習者自身の説明不足や、単なる勘違い等に基づいており、再現性が低いと考えられるもの」である。

例を挙げよう。

[7] 短くても手で書いた文章は個性的で、人生の大切は時には「とっておきたい」ものです。(fr052)

[8] Même courtes, les phrases écrites à la main reflètent une certaine personnalité (individualité) et bien souvent elles constituent des documents que l'on souhaite conserver, lorsqu'elles se rattachent à un événement important de la vie. [日本語への再翻訳：短くても、手書きの文章はひとの人格（個性）を反映している。人生の大切なできごとについて書かれたものであれば、それをとっておきたいと思うこともしばしばである]

母語訳によれば、執筆者はかなり複雑なことを書きたかったことが分かるが、実際にはそこまで書ききれず、日本語では非常にはしょった表現になってしまったものと推測できる。これは 4.1 節のイ) にあたるものである。

また、以下のような例もある。

[9] 第一の理由は、コンピュータやデスクトップなどは紙程残らない。(fr017)

[10] La première est que les ordinateurs et les disquettes ne durent pas aussi longtemps que le papier.

コンピュータやデスクトップが「紙程残らない」という箇所が「意図不明焦点」として抽出されたが、母語訳を見ると、「デスクトップ」とは disquettes (フロッピーディスク)

のつもりであったことが分かる。分かりにくさの理由は「紙ほど残らない」ではなく「デスクトップ」のほうにあったことが分かるのだが、これは音の類似に引きずられた単純な勘違いに過ぎず、教育上特に重視すべきものとはいえない。これは 4.1 節のウ) にあたるものである。

この例のように、「本来そこで使用すべき語を、意味的にほとんど重ならない別の語と取り違えている」ことによって生じた意図不明表現⁶は、すべてこの「グループ B」に入れた。

残る「グループ C」が、今回の調査での中心的な分析対象となるものである。これは 4.1 節で挙げた

エ) 主として言語上の要因によって生じたと考えられる「意図不明表現」であって、ある程度再現性があると考えられるもの

ということになる。

ただし、エ) に含まれる意図不明表現が、必ずしも「母語干渉によるもの」かどうかは分からない、ということをご確認ください。エ) には、単に日本語の文法項目・語彙の意味・社会言語学的な表現習慣等を正確に理解していないことによる逸脱表現が広く含まれる。もちろんその中には母語の影響が色濃く見て取れるものもあるが、個々の意図不明表現について、そこに母語の影響があるかどうかをひとつひとつ詮索することは、今回の調査ではおこなわなかった。

4.3. 「意図不明表現」の内訳

4.3.1. グループ A, B, C の内訳

意図不明箇所 93 箇所の中で、「グループ A」「グループ B」「グループ C」の内訳は以下のようなものであった。今回の調査データの中では、グループ C が 93 箇所のうちほぼ半分にあたる 42 箇所を占めていた。

⁶ 他には、「まだできない (人) (débutant) とすべきところを「まだできない」、「でも」(cependent) とすべきところを「では」としているような例である。もちろん執筆者がこのような形で「間違えて覚えてしまっている」場合、再現性があることになるのだが、間違いを指摘さえすれば簡単に修正できるものと考え、グループ B に入れた。

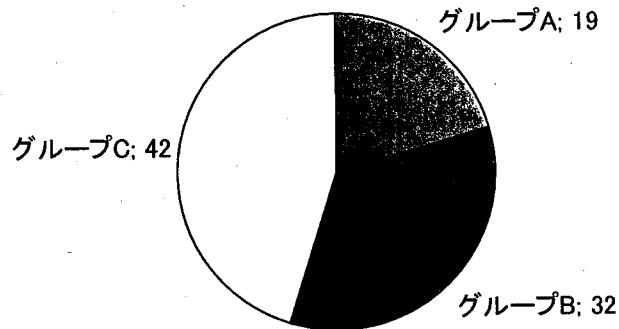


図2 「意図不明表現」グループA,B,Cの内訳

4.3.2. グループCの内訳

次に、上記の手順によって得られた「グループC」の表現について、対応するフランス語表現とを対照することにより、その「分かりにくさ」がどのような要因によって生じたかを考察し、さらに下位分類をおこなうこととした。

吉川(1997:9)では、誤用分類の要点として、

- ① 発音の誤り
- ② 表記の誤り
- ③ 語彙の誤り
- ④ 文法の誤り
- ⑤ 表現の誤り

の5カテゴリーを挙げているが、今回使用している作文データでは、①と②を区別することはできない(例えば「日本魂」とすべきところを「日本だまし」(fr082)と書いている場合、「い」を書き漏らしたのか、長音と短音の区別がついていないのかは分からない)。またそもそも、今回の分類において表記の誤りは「再現性のない誤り」としてすべて「グループB」に帰属されているので「グループC」の下位分類としては不要である。

また「語彙の誤り」の中でも、勘違いなどによって意味的にほとんど共通点のない語と取り違えた、というような例はすべて「グループB」に入れられている。「グループC」で問題となる「語彙の誤り」とは、「おそらくは意識的に語彙選択をおこなったのだが、選択した語の意味が当該文脈と十分適合するものではなかったために、執筆者の意図が読み手に伝わらなかった」というものだ、ということになる。

そこで「グループC」の下位分類としては、

- ① 語彙に関するもの
- ② 文法に関するもの
- ③ 表現に関するもの

の3つを採用することとした⁷。

ここで、ひとつの「意図不明表現」が、上記3カテゴリーのいずれかに必ず帰属されるわけでない、ということを確認しておきたい。

例えば以下のような例を見てみよう。

[11] 日本において、手書きの文字の放棄は日本が極度に推進される書道の死を意味するだろうと思う。(fr087)

[12] La disparition de l'écriture manuscrite, au Japon, signifierait la mort de la calligraphie que le Japon a porté à son plus haut niveau. [日本語への再翻訳：日本において、手で書かれたものがなくなる、ということは、日本が非常に高水準にまで発達させた(直訳は「高水準にまで持っていった」)書道の死を意味することになる]

ここでは avoir porté à son plus haut niveau(非常に高水準にまで持っていった)という内容を、「推進される」と表現してしまったわけだが、ここには、文脈に合わない語(「推進する」)を選択してしまったという問題と、その語を文法的に不適切な形で使ってしまった(使役形にするつもりで誤って受身形にし、さらに完了の「た」をつけ忘れた?)という問題の両方が絡んでいると考えられる。

こうした場合には、ひとつの「意図不明表現」について、ふたつ⁸の「意図不明要因」があると考え、上記下位分類の①、②として二重にカウントすることとした。

その上で、それぞれのカテゴリーに属する「意図不明表現」の内訳を示すと以下のようになる。特に数の多かった「① 語彙に関するもの」については、品詞の内訳も示した。

⁷ 実際の「意図不明表現」がこの3分類のいずれにあたるかを判別する際には、母語訳を参照した上で、「日本語では本来どう書くべきであったか」を推測し、「書くべきであったが書けなかったこと」が、3つの分類カテゴリーのいずれに当てはまるのかを考えることとした。例えば、表面的には日本語の時制表現がうまくできていないのであるが、日本語としては文法的手段ではなく、副詞等語彙的手段によって表現するのが最も適切、と考えられる場合は、「文法に関するもの」ではなく「語彙に関するもの」と分類した。

⁸ 「文法的誤り」として、「使役形」「アスペクト」という2種類を数えることも可能であるが、しかし誤用が生じた理由の切り分けはしばしば困難である。同一箇所での誤りについてふたつ以上の項目が関係していると思われる場合も、それらがすべて「文法」に関するものと考えられる限り、1種類の誤りとしてカウントした。

①語彙に関するもの	35
	抽象名詞 15
	一般名詞 2
	動詞 8
	形容詞 5
	副詞 3
	前置詞 1
副詞 1	
②文法に関するもの	6
③表現に関するもの	6

表1 グループCの「意図不明表現」の内訳

5. 「意図不明表現」が生じた要因の分析

以下、それぞれのカテゴリーから「意図不明表現」の実例をいくつか挙げ、その表現が生じた要因について具体的に考察してみよう。

5.1. 語彙に関するもの

語彙選択に関わる意図不明表現としては、名詞、特に「フランス語の抽象名詞を日本語に置き換える際に何らかの不都合が起こっているものが特に多かった。「語彙に関するもの」33例のうち、抽象名詞に関わるものが15例と約半数を占める。これは、すべての「意図不明表現」93箇所に対しても約16パーセントというかなりの高率を占めることになる。

実際の例を挙げよう。

[13] …コンピュータに対する本の残存物についての議論と同じ議論だと思う。(fr087)

この文は、「日本では、日本語を書く際、ワープロソフトを使うことが多くなった、という問題について論じたい」という問題提起の直後に来るものであるが、そうした文脈を併せ読んだとしても、やはりこの文で何を言いたいのかは不明であり、特に「本の残存物」の部分が分からない。ここは、3名の評定者が3名とも「まったく分からない」という判定を下した部分であった。

この箇所に対応するフランス語訳は、以下のようになっていた。

[14] Je pense que c'est le même débat que celui de la survivance du livre face à l'ordinateur….

翻訳を参照すると、「残存物」は *la survivance* という語と対応している。つまり「コンピュータに対する本の残存物の議論」とは、「コンピュータというものに直面して、(印刷物としての)本が生き残っていけるかどうかの議論」、のことであったことが分かる。

survivre (生き残る) を抽象名詞化した *la survivance* には、「生き残ること」「生き残ったもの」両様の意味があり、日本語としては両者をはっきり表現し分けなければならない。そしてここで *la survivance* は、前者の意味で使われているのであるが、執筆者は後者に対応する日本語表現である「残存物」のほうを使用してしまったのであろう。辞書を引いて、最初に見つけた表現をそのまま使ってしまった、というような可能性が想像できる。

また、以下のような表現も見受けられた。

[15] まず、読書の問題がありません。(fr030)

この文は、「ワープロソフトの使用は便利である」ということの第一の理由として提示されているものであるが、「読書」とは何のことなのか、やはりこれだけでは分からない。フランス語訳では、

[16] D'abord il n'y a pas de problème de lecture.

となっており、「読書」は *lecture* と対応している。

lecture は、「読む」という動詞的概念を名詞化するときには広く使用できる、意味範囲の広い語であり、フランス語訳を参照してもなお執筆者の真意は分かりにくいところがあるが、さらにその後の部分を読み進めると、「日本語には同音の漢字が多く、例えば『紹介』の『紹』と『招待』の『招』は、音が同じなので混同してしまうが、ワープロソフトを使うとこういう間違いは避けられる」ということが書かれている。フランス語訳に加え、さらに文脈の助けも借りて解釈をおこなうならば、[15]は、

[15] 漢字の読み方にもなう問題が起こりません。

のようにすべきであったことが分かる。

lecture は、日本語としては「読むこと、読み方、読めること、読書、朗読、読み物……」など、文脈によってさまざまな語を使い分ける必要がある。こうした使い分けを十分に習得することは、フランス語母語話者にとって非常に困難であろうことがここからうかがえる。

⁹ 「問題」は、執筆者の表記どおりである。

5.2. 文法に関するもの

文法的知識の不足によって生じたと考えられる「意図不明表現」は6例であり、語彙に関するものと比べると頻度としてはかなり少なくなる。

例が少ないため、これらの中から顕著な特徴を見出すことは難しいが、うち2例は、時制の表現がうまくいっていない、という点で共通している。

まず1例を挙げよう。

[17] しかし、それを使用する日は、私の日本語のレベルは上分だと言いますから、もう上達したくないです。(fr038)

[18] *Cependant, le jour où je l'utiliserais, c'est parce que je dirais que mon niveau de japonais sera suffisant, alors je ne voudrais plus faire de progrès.*

日本語中の「それ」とは、「ワープロソフト」のこと。「ワープロソフトを使う日には、私の日本語のレベルは十分であろうから、それ以上うまくなりたとは思わないだろう」のように書きたかったものと思われる。フランス語において動詞はすべて未来形になっており(*utiliserais, dirais, sera, voudrais*)、全体として未来の話をしているのであるが、日本語ではそのことが十分に表現されていないため、「もう上達したくない」というのがあたかも現在の話であるかのように読めてしまい、このことが分かりにくさにつながっているものと思われる。

同様の例として、

[19] それがわかるためには悪い面と有利を勉強している。(fr065)

[20] *Pour comprendre cela, nous allons étudier les avantages et inconvénients.*

というものがあつた。これは「コンピュータに対して賛否両論がある」という記述に続く文であり、フランス語では *aller* という、近接未来を示す準助動詞が使用されている。「そのこと(賛否両論)を理解するために、(コンピュータの)利点と不都合な点について検討してみよう」¹⁰ということなのであろうが、日本語では現在進行中、あるいは習慣的な動作であるように表現されてしまっている。

このように時制表現ができていないと、文章内で語られる事物に対し、書き手がいまどういう立場に立って向き合っているのかが分からなくなるため、全体としての意味理解にも支障が出やすくなるものと思われる。

その他、日本語の使役表現がうまく表現できていないことが理由と思われる例も複数あ

¹⁰ ここでも、「勉強する」という文脈に合わない動詞が使用されているということと、時制の使い方が誤っている、というふたつの「意図不明要因」が関わっているものと解釈した。

った。ひとつは[11]で挙げた例（「極度に推進される書道の死」）であり、もうひとつは以下に挙げるものである。

[21] 実際、日本の勉強をすることは時間が必要だから、記憶力はよく働く。(fr065)

[22] Effectivement, l'étude du japonais nécessitant du temps, on fait souvent fonctionner sa mémoire.

「記憶力は働く」ではなく、「記憶力を働かせなければならない」とでもすべきところであろう。

今回のデータの中からは、文法に関する意図不明表現の中では、「時制」「使役」に関わるものがそれぞれ 2 例ずつ発見された¹¹。しかし、文法に関する意味不明表現が全体で 6 例しかないこともあり、このふたつの文法項目が特に重要なものであるとは明言できない。文法項目の中で、特に深刻なコミュニケーション不全とつながりやすいものはどのようなものを明らかにするためにはさらに多くのデータにあたらなければならないが、そもそも、文法項目と関係する意味不明表現は、語彙選択に関わる意味不明表現に比べ頻度数がかなり少なかった、ということは、留意しておくべきことであろう。

5.3. 表現に関するもの

表現に関する意図不明表現としては、フランス語の「非人称構文」に関係している例を挙げる。

[23] ところが大切な宿題について、先生に敬意を表するのが問題であるとワープロソフトを使う方がいいです。(fr030)

[24] Pourtant, en ce qui concerne les devoirs importants et qu'il s'agit de les rendre au professeur, il vaut mieux utiliser un traitement de texte.

[23]に「問題である」という表現が現れたのは、フランス語の *il s'agit de* を直訳したためだということは明瞭に分かる。*il s'agit de* は、あえて日本語に訳するならば「～が問題である、～に関わることだ、～が主題だ」のようになるが、実際には非常に軽い意味で使われており、「問題である」と訳してしまうと大げさになりすぎる場合が多い。ここは、「先生に敬意を表するべきときには」くらいの表現でよかったのであるが、わざわざ律儀に「問題である」と表現してしまったので、かえって意味が分かりにくくなってしまったものと見ることができる。

¹¹ 他の不明表現の間には特に共通点は見られなかった。

[25] それで何があったのですか。(fr035)

[26] Alors, que s'est-il passé?

[25]は、「ワープロソフトを使う人は、漢字を書くことに困難を感じるようになる」という記述の後に来る文で、その後には「漢字を書く習慣がなくなり、難しい漢字も簡単な漢字も忘れてしまうようになる」という意味のことが書かれている。この文脈から考えると [25]は、

[25] それでどうなったのであろうか。

のように書くべきであったことになろう。

[26]も非人称構文であるが、[25]の理解困難性は、非人称構文に基づくものというよりは、文章中にいきなり「～ですか」という問いかけが現れてきたというその唐突さに基づくものと考えることができる。

伊集院・高橋(2004)では、やはり「作文対訳 DB」の中国語母語話者データを用い、「中国語母語話者の日本語作文には、読み手に直接語りかけるような表現が多用される」という指摘をしているが、こうした表現は、中国語母語話者ほど頻繁ではないにせよ、他の言語を母語とする学習者の作文にも現れうるということが分かる。ただ、伊集院・高橋(2004)が挙げる「語りかけ表現」の多くは、不自然ではあるとはいえ、書き手の意図がまったく分からない、というようなものではないのであるが、[25]は3名の判定者のうち2名までが「まったく分からない」と判定し、もう1名も「なんとなく分かるが確証が持てない」と判定した部分であった。同じ「語りかけ表現」であっても、「なんとなく不自然」というものから、「意味がまったく分からない」というものまでさまざまなものがありうる。「語りかけ表現」がどのような現れ方をしたときに「意図不明表現」となりうるのか、というようなことも今後の興味深いテーマのひとつとなろう。

6. 結果

今回の調査で明らかになったことは以下のとおりである。

- 1) フランス語母語話者の書いた日本語作文の中の「意図不明箇所」(日本語母語話者が見ても執筆者の意図を推測することが不可能または困難な箇所)全体のうち、母語訳を参照することによって執筆者の意図を推測することが可能な箇所(「グループ B,C」の「意図不明表現」)は、93 箇所のうち 74 箇所(79.6 パーセント)であった。これは、「作文対訳 DB」のうちフランス語母語話者のデータのみを検討した結果にすぎないが、このことは母語訳付きの作文対訳データベースが、研究上有効な資料となりうることを示しているといっていよう。

- 2) 「意図不明箇所」のうち、言語的な要因によって発生理由の説明が付き、かつ再現性があるものと考えられるもの（「グループ C」の意図不明表現）は、93 箇所のうち 42 箇所（45.2 パーセント）であった。
- 3) 「グループ C」の意図不明表現を、①語彙選択に関するもの、②文法に関するもの、③表現に関するものに分類すると、①語彙選択に関するもの、と解釈できるものが 35 箇所（74.5 パーセント）と大半を占め、②文法に関するもの、③表現に関するもの、と解釈できるものはそれぞれ 6 箇所（12.8 パーセント）ずつであった。また、語彙選択に関する意図不明表現の中では、名詞、特に抽象名詞に関するものが 15 箇所（42.9 パーセント）を占めた。

つまりフランス語母語話者の「意図不明表現」をもたらす最も重要な要因は「語彙選択」に関する問題であり、中でも抽象名詞に関わるものが重要な位置を占める、ということが数量的に示されたものといえる。

7. 考察

一般にフランス語に顕著な特徴として、「語義の抽象性」（大橋 1974）、「多義的な単語の使用」（蓮實 1977）ということがしばしば指摘される。つまり、ひとつの語が持つ意味が抽象的であり、その結果としてひとつの語がさまざまな範囲の事象を表現するのに使われる、ということである。そしてその抽象性・多義性は、17 世紀の古典主義時代に、「外延と内包という論理的な水準に従い、共通項のあるものは同一の概念で示すという姿勢によって遂行された」（蓮實 1977:189）文化政策によってもたらされたものとされる。

このため、フランス語の抽象名詞 1 語に対しては複数の日本語表現が対応することになり（例えば 5.1 節で挙げたように、フランス語の *lecture* 1 語に対し、日本語では「読むこと、読み方、読めること、読書、朗読、読み物……」などが対応する）、したがって日本語を書くときには、当該文脈に最も適合する対応表現を吟味して使用しなければならないのであるが、実際のところ学習者は必要な吟味をせず（おそらくはその手段がないため）、辞書を引いて最初に目についた訳語をそのまま使用しているということが想像される。

日本語の対応表現はどれをとってみても、フランス語の抽象名詞より具体的かつ限定的な意味範囲を持っている。したがって学習者が選択した語は、多くの場合本人が意図したよりずっと限定的かつ具体的な意味を持ってしまい、このため読み手としては、その語が身にまとっている不必要に限定的・具体的な意味に惑わされ、学習者の意図を取り損ねてしまうことになるものと考えられる。

さらにヴァルトブルク(1976)は、フランス語には「事件や行為を動詞よりもむしろ名詞で表現する」傾向があることを指摘している。この傾向は 19 世紀の間に強まったとされ、

例えばルグランの『フランス文体論』(E. Legrand, Stylistique Française)では、

[27] Ils cédèrent parce qu'on leur promet formellement qu'ils ne seraient pas punis.

「彼らは自分たちが罰せられないと正式に約束してもらったので、譲歩した」

のように、ふたつの従属節によって表現することを避け、

[28] Ils cédèrent à une promesse formelle d'impunité.

「彼らは処罰しないという正式の約束によって譲歩した」

と表現するように勧められているという(ヴァルトブルク 1976:274)。

フランス語母語話者の日本語作文を見ると、まさに上記のような発想によって作られた日本語表現に頻繁に出会うことができる。既出の例でいえば、

[11] 日本において、手書きの文字の放棄は日本が極度に推進される書道の死を意味するだろうと思う。(fr087)

[13] …コンピュータに対する本の残存物についての議論と同じ議論だと思う。(fr087)

などがそうであろう。筆者自身は 4.3.2 節において、[11]のフランス語訳([12])からの再翻訳として「手で書かれたものがなくなる、ということは…」というものを示したが、これをさらに書き換え、

[11'] 手書きの文字がなくなると、…書道は死んでしまうことになると思う

のように、名詞(「文字の放棄」「なくなる、ということ」「死」)を使うのではなく、動詞を用いてふたつの節として表現するならば、日本語としての自然さはさらに増すことになるであろう。

事件や行為を動詞ではなく名詞(多くは抽象名詞)を用いて表現する、というフランス語の表現指向それ自体は、日本語としてやや不自然な表現につながることはなっても、常に「意図不明表現」をもたらすわけではない。[11]は「意図不明表現」として抽出されたものであるが、「意味不明焦点」は「極度に推進される書道云々」の部分であって、「手書きの文字の放棄は…死を意味する」の部分については特に「意図不明」との指摘はなかった。「放棄」という抽象名詞を主語に据えたこうした表現は、確かに日本語としてやや生硬に響くが、しかし誤りでは決してなく、「翻訳調」と考えれば許容しうる範囲の表現ということもできる。

また、事件や行為を、名詞で表現するか動詞で表現するかということは、究極的には「趣

味の問題」である。一般的に、名詞で表現したほうがより簡潔な表現は可能になり、それがフランス語母語話者の趣味に合うのであるが、日本語母語話者にはそのような表現習慣は少ない、ということなのであろう。文章は可能な限り簡潔であるべきだ、と考えるフランス語母語話者が、日本語で文章を書くときにも、名詞を多用する表現を使いたがったとしても、そのことは一概に排除すべきことではない。

しかしながらこうした表現傾向は、日本語において使い方の難しい抽象名詞を過剰に出現させることにつながり、間接的に「意図不明表現」と結びつく危険性は高い。日本語による確実な情報伝達をおこなうためには、あまり使わないほうが無難な表現であるとはいえる。作文指導をおこなう教師としては、どのような態度で指導に臨むべきか、ということが問題になるだろう。

このことについて、筆者は以下のように考える：

学習者にとって日本語の抽象名詞が使いにくいからといって、その使用を避けるように指導したり、フランス語の特定の抽象名詞に対応する日本語の各種表現の間の使い分けを教え込んだりすることは、もちろん場合によっては必要なことであるが、それだけでは意味が薄い。重要なことは、学習者自身が日本語とフランス語を自ら見つめなおし、「それぞれの言語を支える基本的な発想を理解し、それぞれの言語に見られる種々の事柄をそれと関連付けて有機的に理解する」(井上 2002:10)能力を体得する、ということである。井上(2002)の用語に従えば、〈類型設定型〉の言語対照能力を身につけてもらう、ということになる。

動作や行為を表現する際に、フランス語では名詞を多用した表現が好まれるが日本語では素直に動詞によって表現されることが多いこと、フランス語の名詞は語義の抽象性が高いため、フランス語のひとつの抽象名詞に対し日本語では複数の表現が対応し、文脈に応じた使い分けが必要になること、文脈に応じた使い分けがうまくいかないと「意図不明表現」が生じる可能性が高くなること。まずはこういったことを学習者自身に知ってもらうことが必要である。その上で、コミュニケーション不全を生じやすい名詞的表現は日本語では避ける、という戦略をとるのか、日本語で抽象名詞を用いて自然な文を書くのは難しいということは承知の上で、あえてそうした表現方法に挑戦するのか。それは学習者自身が選択する、というのが望ましい方向であろう。

学習者自身にそのような能力を体得してもらうため、具体的にはどのような方策がありうるか、ということについては今後の研究に俟たねばならない。しかしその方策を考えていくにあたって、われわれはこの「作文対訳 DB」に含まれている各種データからさまざまな示唆を受け取ることができるだろう。

注記：この論文は、宇佐美洋(2004)「フランス語母語話者の日本語作文における『意図不明表現』—執筆者本人による母語訳からわかること」(『ヨーロッパ日本語教育』9, 169-174, ヨーロッパ日本語教師会) に対し、大幅な加筆修整を加えたものである。

引用文献

- 井上優(2002)「言語の対照研究」の役割と意義,『対照研究と日本語教育』, 3-20, 国立国語研究所
- 伊集院郁子・高橋圭子(2004)「文末のモダリティに見られる“Writer/Reader visibility” — 中国人学習者と日本語母語話者の意見文の比較—」,『日本語教育』123号, 86-95, 日本語教育学会
- 大橋保夫(1974)「フランス語は明晰な言語か」, 大橋他(1993)『フランス語とはどういう言語か』に再掲, 59-71, 駿河台出版社
- 蓮實重彦(1977)「明晰性の神話」,『反=日本語論』, 179-193, 筑摩書房
- 吉川武時(1997)「誤用分析 I」,『日本語誤用分析』(明治書院企画編集部編), 2-53, 明治書院
- ヴァルトブルク(1976)『フランス語の進化と構造』, 田島宏他共訳, 白水社

Appendix

作文課題

次の文を読んで、自分の意見を 400~800 字くらいの日本語で書いてください。日本語で書いたあと、その文章をあなたの母語 (mother tongue, または最も楽に文章が書ける言語) に訳してください。

最近日本では、日本語の文章を書くときにコンピュータのワープロソフト (word processing software) を使うことが多くなりました。ワープロソフトを使えば、自分では書けない漢字を使うこともできますが、一方では、ワープロソフトを使うことで漢字が手で書けなくなってしまうということも起こります。

ある人は、漢字は書けなくても読めて意味がわかればいいのだから、日本語学習でもワープロソフトをもっと使うべきだといいます。ある人は、漢字が手で書けるということはやはり大事なことなので、日本語学習ではワープロソフトは使うべきではないといいます。あなた自身はどう思いますか。理由を挙げて、あなた自身の考えを書いてください。

第二言語としての日本語作文に対する記述式フィードバックと 教師ビリーフの関係

成田 高宏

1. はじめに

本研究は、日本語学習者の日本語作文に対する教師の記述式フィードバック（添削）を調査対象とし、特にコメント型のフィードバックの内容が、教師の作文教育に関するビリーフや日本語教師としてのビリーフ、教師経験などの要素と、どのような関係があるかを分析するものである¹。その結果から、作文教育に関わる日本語教師や、作文教育を支援する立場の者の今後の取り組みについて、展望を試みたい。

2. 先行研究

記述式フィードバックに関する議論は、以前より「言語形式か、内容か」という観点からなされてきた。日本語作文に関しては坂井（1998）が、言語形式についてのフィードバックは言語形式面の改善に、内容についてのフィードバックは内容面の改善に効果があり、両者を同時に行った場合には内容の改善につながりにくいことを報告している。英語教育においては、同様に言語形式的な誤りの訂正と内容についてのコメントのどちらが作文全体の改善に効果的かという観点から、後者が効果的であることを指摘した Semke(1984)や Kepner(1991)、逆の結果を指摘した Fathman et al.(1990)などの研究がある。

また、学習者の日本語作文に対する教師の記述式フィードバックの分類を試みた研究として、上原（1997）がある。ここではコメント型のフィードバックを、①形態的要素について述べるもの、②構成、一貫性 など修辭的要素について述べるもの、③内容への応答、④評価の4種類に分けている。48名の日本語教師（うち日本語教育経験5年以上7年未満14名、3年以上5年未満12名、他24名は不明）に同一の作文を添削してもらった結果、全体としては48名のうち60.4%が何らかのコメントを記述し、また①から③の出現率には偏りがなく均等に現れたことが報告されている。④の出現率が低かったことについては、調査対象となった教師自身の学習指導の場と切り離されて行われた調査だったことや、使用した作文が1つであり比較の対象がなかったなどの理由により、評価のコメントが現れにくかったという分析がなされている。

上原（1997）の分類は、「教師が作文中に見られるどのような特徴に着目して添削を行っているのか」を探る上で有効と考える。本研究では、教師の行った記述式フィードバックのコメント部分に対し、上原（1997）に準じた分類を行った上で、さらに各種コメントの出現と教師のビリーフとの関係を探る。それにより、教師の全体的な傾向を把握すると共

¹使用した作文は国立国語研究所作成の『日本語学習者による日本語作文と、その母語訳との対訳データベース』に収録されたものである。

に、ある特定の特徴を持つ教師が、意識的あるいは無意識的に偏りがちな傾向について、示唆することが可能になるのではないかと考える。

3. 調査方法

3.1. 対象者および手順

30名の日本語教師を対象に、9件の作文（資料3）について添削をしてもらった²。作文の内訳は、日本語学習者による作文8件と、日本人による作文1件である。日本人による作文については、母語話者であることによる特別な影響が見られたわけではなかったため、この作文に対するコメントもそのままデータとして含めることとした。添削は、普段、自身の担当する学習者に対して行うのと同様の方法で行ってもらおうよう依頼した。ただし、使用する記号等は協力者間で統一する必要があるため、それについてはこちらで指定したものを使用してもらった。

次に「プロフィールに関する質問群」、「作文教育に関する質問群」（22項目、5段階評価）、「日本語教師であることに関する質問群」（16項目、5段階評価）の3種の内容からなるアンケートに回答してもらった（資料2）。以下に「作文教育に関する質問群」および「日本語教師であることに関する質問群」の項目を示す。

「作文教育に関する質問群」

- ① (1) 作文の授業では、論理的な考え方の育成を目指したほうがいい
- ① (2) 学習者の作文能力向上のためには、添削されたものを読むだけでは難しい
- ① (3) 添削時、与える情報が多いと学習者の負担が大きくなるので、できるだけ簡潔なほうがいい
- ① (4) 添削時、訂正するだけでなく、気付いた点についてコメントを記したほうがいい
- ① (5) 添削の主な目的は、正しい日本語の例を示すことである
- ① (6) 添削時、作文の内容・構成については学習者の意図を尊重し、教師が修正や指摘をしなくてもいい
- ① (7) 文末の形（だ・である、です・ます）が統一されていない場合、直したほうがいい
- ① (8) 同レベルの学習者であっても学習者の個性によって添削の仕方を変えたほうがいい
- ① (9) 学習者のレベルに合わせ、添削の仕方を変えたほうがいい
- ① (10) コメントを記す場合、学習者に考える機会を与えるよう工夫したほうがいい
- ① (11) 添削の際、誤った原稿用紙の使用は必ず直したほうがいい

² 実際には10件の作文について添削してもらったが、うち1件の作文については、内容の理解が困難なため特殊すぎる傾向が表れた。そのため外れ値としてこの作文を分析対象からはずした。

- ① (12) 学習者に作文を書かせる主な目的は、既習の知識をよりよく定着させることである
- ① (13) 学習者の作文の内容や構成面に問題があれば、直すよう指導したほうがいい
- ① (14) 添削後どう書き直すか、学習者自身が判断できるような形で添削したほうがいい
- ① (15) 添削時、正しい例を示すだけでなく、できるだけなぜそう直したか理由を付したほうがいい
- ① (16) 添削時、意図がわからない文については、無理に推測せず意味不明であることを示すだけでいい
- ① (17) 添削時、語彙や表現の不適切な使用は必ず直したほうがいい
- ① (18) 作文の授業では学習者が書き出し、結びなどの構成を考えて文章を書けるようになるよう指導したほうがいい
- ① (19) 添削をすることそのものよりも、添削内容についてのその後のフィードバックが重要だ
- ① (20) 添削時、直し方にいくつか選択肢がある場合は、1つだけでなく複数示したほうがいい
- ① (21) 添削時、文字の表記の不適切な使用は必ず直したほうがいい
- ① (22) 添削時の明示的な指示や指摘は、添削後の自主的学習の妨げになるのでできれば避けたほうがいい

「日本語教師であることに関する質問群」

- ② (1) 日本語教師のための研修会等があれば積極的に参加したい
- ② (2) 授業経験を多く積むことは教師の成長にとって大切だ
- ② (3) 教師としての勉強をしたいが時間や機会に制限があるので難しい
- ② (4) 今の自分の教え方は特に問題がないと思う
- ② (5) 所属機関の方針があるので今の授業の形を変えるのは難しい
- ② (6) 論文や研究発表は教えるためにはあまり役に立たない
- ② (7) 他の教師と授業見学をお互いにする機会がよくある
- ② (8) 他の教師と授業について話す機会がよくある
- ② (9) 学習者のためにもっと役に立ちたい
- ② (10) クラスに変わった学習者がいても対応できると思う
- ② (11) これからも教師として成長していきたい
- ② (12) 以前はよく教師としての勉強をしたが最近はしていない
- ② (13) 一般に教師はもっと自分のための時間を大切にしたい
- ② (14) 教師が一生懸命になっても報われないことが多い
- ② (15) 一般に教師はもっと学習者の自主性にまかせてもいい

② (16) 一般に教師は学習者に対して責任を感じすぎだと思ふ

3.2. 分類方法

上原(1997)では、コメント型の記述式フィードバックを、①形態的要素について述べるもの、②構成、一貫性など修辭的要素について述べるもの、③内容への応答、④評価の4種類に分けている。本稿での分類もこれに準ずることとするが、ただし④評価については、どんな点について評価しているのかという観点により①～③に含まれるよう分類することとする。たとえば、同じく評価であっても、言語形式について評価している場合には①に、内容について評価しているなら③に含める。具体的な分類方法、および、何をもちて1つのコメントと扱うかという点については、資料1を参照されたい。

4. 結果と分析

4.1. 概要

図1は、総コメント数に対する各種コメントの比率を表したものである。また、図2は、コメント数を一定範囲に区切り、何人の教師がどれだけのコメントを記したかを表したグラフである(たとえば、全9作文において記されたコメント数の合計が5～9であった教師は3人)。図3は、「作文教育に関する質問群」の各項目に対する5段階評価の回答において、4または5の肯定的反応を示した教師数の比率を示したものである。

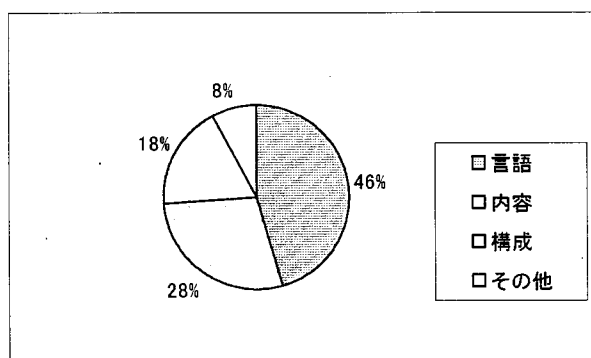


図1 各種コメントの比率

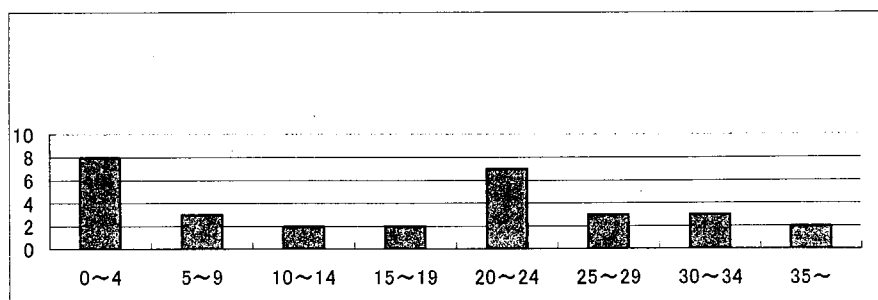


図2 コメント数と教師数

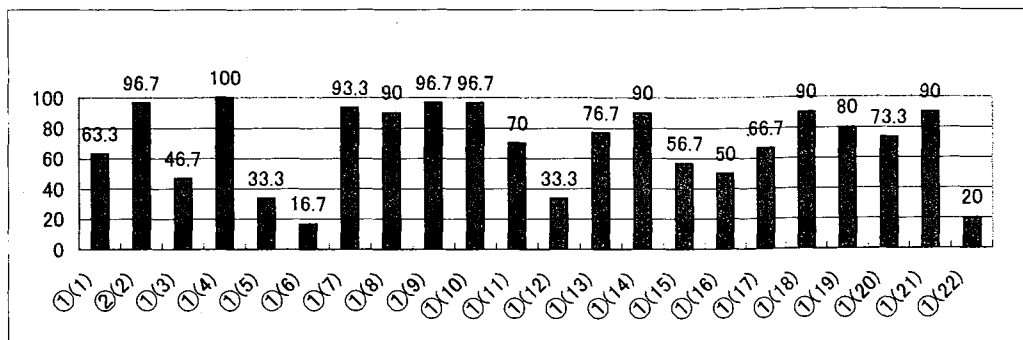


図3 肯定的反応

まず、総コメントにおける各種コメントの比率（図1）は、言語形式が46%、構成以外の内容が28%、構成が18%、その他が8%という結果となった。その他を除く3種のコメントのうち、構成に関わるコメントが相対的に少ない。これは、「作文教育に関する質問群」における、①(13)「学習者の作文の内容や構成面に問題があれば、直すよう指導したほうが良い」という質問に対し、76.7%の教師が4か5の肯定的反応を示している（図3）のとは逆の結果となっている。構成の指導について、「必要だが添削上では（あるいは添削上でなくとも）難しい」あるいは「別の方法で行うのが適当であり添削上行う必要はない」と意識されている可能性が考えられる。

次に教師が示したコメント数（図2）については、9件の作文に対し、30以上のコメントを記述した教師が5名（最大88コメント）いる一方で、コメント数が5未満の教師が8名（コメントが全くなかった2名を含む）いるというように、幅広い結果が見られた。各教師のコメント頻度を作文ごとに見ると、コメントを多く記述する傾向にある教師は、9件の作文にほぼ満遍なく記述を行い、コメントをしない教師は当然、どの作文に対してもコメントを記述していなかった。つまり、作文の違いによりコメント数に差が生じるというよりは、教師の違いによる影響が大きいということがわかった。また、図2を見ると、度数の高い部分が「0~4」「20~24」という離れた2箇所にあることがわかる。これらことから、コメントを記述することに関する考え方の差異、もしくは視点や技術上の差異が、教師間に存在することが想像される。

一方、「作文教育に関する質問群」の①(4)では、全教師が「添削時、訂正するだけでなく、気付いた点についてコメントを記したほうが良い」という項目に肯定的反応を示している（図3）。にもかかわらず、コメントを全く、あるいはほとんど記述しない教師が少なくないことは注目に値する。

4.2. 各要素の関係

「作文教育に関する質問群」（資料2，12①）への回答を基に、類似する項目同士をまとめグループ分けを施すための探索的因子分析を行った。具体的には、初期解を主因子法に

より求め、バリマックス回転をかけ、最終的に、全質問群より問①(12)および問①(16)を除外した、20項目を分析に使用した。その結果、4因子解を採用するのが妥当と解釈した。その因子負荷量を表1に示す。

	因子1	因子2	因子3	因子4	共通性
①(18)	<i>0.68</i>	0.11	0.24	0.09	0.55
①(3)	<i>-0.68</i>	0.02	0.05	0.13	0.48
①(13)	<i>0.67</i>	-0.03	0.06	0.31	0.55
①(7)	<i>0.57</i>	0.34	-0.21	0.30	0.58
①(5)	<i>0.47</i>	0.23	-0.15	0.39	0.45
①(22)	0.17	<i>0.68</i>	0.11	-0.09	0.51
①(6)	-0.28	<i>0.64</i>	-0.25	-0.13	0.57
①(9)	0.32	<i>0.63</i>	0.02	0.17	0.52
①(2)	0.15	<i>0.48</i>	0.10	-0.12	0.28
①(19)	-0.25	<i>0.46</i>	0.35	-0.05	0.40
①(1)	0.11	-0.36	<i>0.68</i>	-0.01	0.60
①(15)	0.18	0.04	<i>0.63</i>	-0.04	0.43
①(14)	-0.01	0.39	<i>0.56</i>	0.21	0.51
①(20)	-0.24	0.10	<i>0.53</i>	-0.10	0.36
①(4)	0.03	0.03	<i>0.49</i>	0.03	0.24
①(10)	-0.19	0.35	<i>0.44</i>	0.19	0.39
①(21)	0.25	-0.15	-0.12	<i>0.78</i>	0.70
①(11)	-0.03	-0.21	0.19	<i>0.64</i>	0.49
①(17)	0.27	-0.13	-0.14	<i>0.47</i>	0.33
①(8)	-0.02	0.17	0.11	<i>0.46</i>	0.25
寄与率	12.5%	12.1%	11.5%	9.9%	累積 46.07%
二乗和	2.50	2.43	2.31	1.99	

表1 「作文教育に関する質問群」因子負荷量

因子付加量の絶対値 0.4 以上を基準とし、因子1を「模範提示志向」、因子2を「授業直結志向」、因子3を「思考過程重視志向」、因子4を「教育者志向」と名付けた。それぞれの因子を成す項目を以下に列挙する。うち問①(3)に関しては、負の値であるため項目の内容を逆に解釈する必要がある。

因子1：模範提示志向

次の5項目に、「添削の目的は学習者の書いたものを基に、正しい日本語作文に直し、模

範として示すことにある」とする考え方が潜在すると解釈し、この因子を「模範提示志向」と名付ける。

- ① (3) 添削時、与える情報が多いと学習者の負担が大きくなるので、できるだけ簡潔なほうがいい (逆)
- ① (5) 添削の主な目的は、正しい日本語の例を示すことである
- ① (7) 文末の形 (だ・である、です・ます) が統一されていない場合、直したほうがいい
- ① (13) 学習者の作文の内容や構成面に問題があれば、直すよう指導したほうがいい
- ① (18) 作文の授業では学習者が書き出し、結びなどの構成を考えて文章を書けるようになるよう指導したほうがいい

因子2：授業直結志向

次の5項目に、「添削は、その前後の授業との関わりにおいてはじめて効果を発揮するものであり、したがって指導的な記述を明示する必要はない」とする考え方が潜在すると解釈し、この因子を「授業直結志向」と名付ける。

- ① (2) 学習者の作文能力向上のためには、添削されたものを読むだけでは難しい
- ① (6) 添削時、作文の内容・構成については学習者の意図を尊重し、教師が修正や指摘をしなくてもいい
- ① (9) 学習者のレベルに合わせ、添削の仕方を変えたほうがいい
- ① (19) 添削をすることそのものよりも、添削内容についてのその後のフィードバックが重要だ
- ① (22) 添削時の明示的な指示や指摘は、添削後の自主的学習の妨げになるのでできれば避けたほうがいい

因子3：思考過程重視志向

次の6項目に、「考えるプロセスを重視し、添削においては学習者に考えさせることを促進するような記述を行うことが肝要である」とする考え方が潜在すると解釈し、この因子を「思考過程重視志向」と名付ける。

- ① (1) 作文の授業では、論理的な考え方の育成を目指したほうがいい
- ① (4) 添削時、訂正するだけでなく、気付いた点についてコメントを記したほうがいい
- ① (10) コメントを記す場合、学習者に考える機会を与えるよう工夫したほうがいい
- ① (14) 添削後どう書き直すか学習者自身が判断できるような形で添削したほうがいい
- ① (15) 添削時、正しい例を示すだけでなく、できるだけなぜそう直したか理由を付したほうがいい
- ① (20) 添削時、直し方にいくつか選択肢がある場合は、1つだけでなく複数示したほうがいい

因子4：教育者志向

次の4項目に、「教育者としてのあるべき態度に関わるイメージ」が潜在していると解釈し「教育者志向」と名付ける。①(11), ①(17), ①(21)から、単に正用を重視しているとも受け取れるが、①(8)が含まれるところからこのように解釈するのが妥当であると、ここでは判断した。ただし、この因子について明らかにするためには、さらなる調査を重ねる必要があると思われる。

- ①(8) 同レベルの学習者であっても学習者の個性によって添削の仕方を変えたほうがいい
- ①(11) 添削の際、誤った原稿用紙の使用は必ず直したほうがいい
- ①(17) 添削時、語彙や表現の不適切な使用は必ず直したほうがいい
- ①(21) 添削時、文字の表記の不適切な使用は必ず直したほうがいい

さらに、「日本語教師であることに関する質問群」(資料2, 12②)についても、同様に探索的因子分析を行った。初期解を主因子法により求め、バリマックス回転をかけ、最終的に問②(1), ②(4), ②(7), ②(10)を除外した項目群より3因子解を得た。表2はその因子負荷量である。

	因子1	因子2	因子3	共通性
②(12)	0.71	-0.23	-0.18	0.60
②(16)	0.70	0.22	0.00	0.54
②(8)	-0.62	0.18	-0.08	0.42
②(3)	0.51	-0.05	0.21	0.30
②(15)	0.22	0.62	0.03	0.43
②(6)	0.16	-0.57	-0.09	0.36
②(9)	-0.17	0.53	-0.15	0.33
②(9)	-0.05	0.46	0.02	0.21
②(13)	-0.02	0.13	0.69	0.49
②(5)	0.15	-0.36	-0.65	0.57
②(2)	-0.06	0.09	-0.48	0.24
②(14)	0.15	-0.17	0.45	0.26
寄与率	15.05%	12.65%	11.93%	累積 39.63%
二乗和	1.81	1.52	1.43	

表2 「日本語教師であることに関する質問群」因子付加量

因子付加量の絶対値 0.40 上を基準とし、因子1「不適切環境意識」、因子2「教師役割探求意識」、因子3「職業非優先意識」と名付けた。それぞれの因子を構成する項目は以下の

通りである。問②(2), ②(5), ②(6), ②(8)は負の値であるため、逆の解釈が必要である。

因子1：不適切環境意識

次の4項目に、「教師という職業そのもの、あるいは現在自分の置かれている環境は、自分には適していない」と感じる心理が潜在すると解釈し、「不適切環境意識」と名付ける。

- ② (3) 教師としての勉強をしたいが時間や機会に制限があるので難しい
- ② (8) 他の教師と授業について話す機会がよくある (逆)
- ② (12) 以前はよく教師としての勉強をしたが最近はしていない
- ② (16) 一般に教師は学習者に対して責任を感じすぎだと思う

因子2：教師役割探求意識

次の4項目に、「学習者のために教師としてさらに成長していきたい、またそのための学習も進めたい」という意欲が潜在すると解釈し、「教師役割探求意識」と名付ける。

- ② (6) 論文や研究発表は教えるためにはあまり役に立たない (逆)
- ② (9) 学習者のためにもっと役に立ちたい
- ② (11) これからも教師として成長していきたい
- ② (15) 一般に教師はもっと学習者の自主性にまかせてもいい

因子3：職業非優先意識

次の4項目に、「教師としての自分は自分の一部に過ぎず、それ以外の部分を大切にしたい」と感じる心理が潜在すると考え、「職業非優先意識」と名付ける。

- ② (2) 授業経験を多く積むことは教師の成長にとって大切だ (逆)
- ② (5) 所属機関の方針があるので今の授業の形を変えるのは難しい (逆)
- ② (13) 一般に教師はもっと自分のための時間を大切にしたい方がいい
- ② (14) 教師が一生懸命になっても報われないことが多い

次に、各因子の因子得点と、5種類のコメント数（総コメント数・言語形式・構成以外の内容・構成・その他）、作文1件あたりに要した平均時間、および日本語教育歴という要素のそれぞれの相関を求めた。その結果が表3である。また、表4には因子間の相関を示す。

	コメント数	言語	内容	構成	その他	時間	教育歴
模範提示志向	-0.16	-0.03	-0.44	-0.12	0.17	-0.16	-0.16
授業直結志向	-0.02	-0.08	0.03	0.10	-0.18	-0.09	-0.41
思考過程重視志向	0.20	0.13	0.15	0.15	0.24	-0.20	0.23
教育者志向	0.02	0.11	0.00	-0.10	-0.12	0.03	0.03
不適切環境意識	0.21	0.16	0.22	0.04	0.40	-0.14	-0.07
教師役割探求意識	-0.01	-0.08	0.06	0.02	0.13	-0.25	0.19
職業非優先意識	-0.11	-0.04	-0.05	-0.10	-0.32	0.11	-0.04

表3 各因子とコメント数・プロフィール等の相関

	模範提示志向	授業直結志向	思考過程重視志	教育者志向
不適切環境意識	-0.07	-0.10	0.04	-0.20
教師役割探求意識	-0.07	0.14	0.53	0.07
職業非優先意識	-0.25	-0.04	0.10	0.02

表4 因子間の相関

総コメント数との相関が見られたのは、「思考過程重視志向」(0.20) および「不適切環境意識」(0.21)であった。言語形式についてのコメント数は、どの項目との相関も見られなかった。構成以外の内容に関わるコメント数は、「不適切環境意識」(0.22)との弱い相関が見られ、また、「模範提示志向」(-0.44)との中程度の負の相関が見られた。構成に関わるコメント数は、どの項目とも相関が見られなかった。「その他」に分類されたコメントの数については、「思考過程重視志向」(0.24)との弱い相関が見られたほか、「不適切環境意識」(0.40)との中程度の相関、および「職業非優先意識」(-0.32)との負の相関が見られた。

次に、作文1件あたりに要した平均時間についてみると、「思考過程重視志向」(-0.20)、「教師役割探求意識」(-0.25)との間にそれぞれ弱い負の相関が観察された。また、日本語教育歴に関しては、「思考過程重視志向」(0.23)との弱い相関を示す一方、「授業直結志向」(-0.41)との中程度の負の相関が見られたのが特徴的である。

「作文教育に関する質問群」から得た4因子と「日本語教師であることに関する質問群」から得た3因子の、それぞれの因子得点間において相関が見られた部分は、「模範提示志向」と「非職業優先意識」(-0.25)、「思考過程重視志向」と「教師役割探求意識」(0.53)、「正用重視志向」と「不適切環境意識」(-0.20)のそれぞれにおいてであった。特に「思考過程重視志向」と「教師役割探求意識」の間に見られた中程度の相関は、特徴的といえるだろう。

以下、中程度の相関が見られた要素について分析を行いたい。まず「模範提示志向」と「構成以外の内容」コメントとの間に、中程度の負の相関(-0.44)が見られた。ここでいう「模範提示志向」とは、作文の記述的フィードバックの主な目的が、正しい日本語を示す

ことにあるとみなす考え方である。この傾向が強い教師ほど、内容に関するコメントが少ないという結果は、このような教師が内容に関するコメントに高い価値を見出していない可能性を示唆するものととらえ得る。

また、その他のコメントと「不適切環境意識」との間に見られた中程度の相関が注目される。「不適切環境意識」とは、ここでは、日本語教師としての責任を負担に感じ、現在の自分の置かれている状況に居心地の悪さを感じる意識と解釈している。そのような意識の強い教師ほど、分類が困難な、すなわち日本語の指導上、特定の意味をなしにくいコメントを多く記述する傾向が見られたということである（資料1，1Cその他の項を参照）。日本語教師であることの負担を感じることで、指導上、重要性が高いとはいえないようなところにまで力を入れてしまうこととの関係が示唆されているという解釈も可能であろう³。

コメントとの関係ではないが、日本語教育歴と「授業直結志向」との間の中程度の負の相関も気になることである。「授業直結志向」とは、「添削は授業と結びついてはじめて効果を発揮する」という考え方と解釈しているが、逆に、添削において多くの情報を記述することについて否定的であり、その理由を授業と結びついていないことに置く考え方も解釈できる。そして、日本語教育歴の浅い教師ほど、この志向が強いという結果となっている。添削において多くの情報が記述されなくとも、実際に授業において十分なフィードバックが行われるのであれば、このような考え方も問題とはならないだろう。しかし、経験が浅いために十分な情報を記述するだけの技術や知識等が不足している場合には、このような考え方を認めながらも「では添削上ではどこまでの指導が可能なのか」という視点も持ち、実際に行ってみることも、教師としての自己の成長のために有効なのではないだろうか。

因子間の関係においては、「思考過程重視志向」と「教師役割探求意識」との間の相関が顕著であった。「思考過程重視志向」とは、添削の主な目的を学習者に考えさせることとし、その過程を大切にする考え方をさす。また、「教師役割探求意識」は、学習者のために自身が教師として成長したいと願い、そのための研究や学習が必要であるとする考え方である。その根本には、「何かを習得する際には、与えられるだけでは難しく、思考することを通じ自らのものとして取り込む過程が必要である」という信念が、共通して存在しているのである。

なお、特に表として示してはいないが、日本語教育歴と総コメント数、および日本語教育歴と各種コメントとの間に相関は見られなかった。

5. まとめと展望

記述式フィードバック（添削）のコメント部分の記述、および「作文教育に関するピリ

³ ただし、「現状に適合していないという意識が勉強不足などの状況を生み、その結果、重要性の低い箇所にか点を置いてしまうことになっている」のか、「重要な箇所だけでなく重要性の低い箇所にもまで言及しようとするために自らの負担を増やしてしまい、その結果現状に不満を感じている」のかは、判断できない。

ーフ」の調査を行った結果、構成についての指導の大切さが認められているにもかかわらず、添削上ではその指導が行われにくいことがわかった。また、添削の際に、コメントを記述することの重要性をすべての教師が感じていながら、実際には、コメントの記述を行う教師と行わない教師とに二分されることがわかった。

ここで問題となるのは、考え方と行動との間に隔たりがある教師が多数いると推測されることであり、また、その原因は何か、ということである。原因をある程度特定することにより、「考える通りの指導ができない」と感じている教師への支援がはじめて可能になる。

まず、コメントを記述しない理由として、①「知識・技術的には可能だが記述しない」場合と②「知識・技術的に難しい」場合とが考えられる。①は、①-1「時間や労力的に難しい」場合と、①-2「コメントの効果、あるいは必要性を認めていない」場合とがあり得るが、すべての教師がコメントの重要性を認める回答をしているため、①-2は考えにくい。したがって、①「知識・技術的には可能だが記述しない」場合には①-1が主な理由であると考えるのが妥当と思われる。

ここから、日本語作文教育を支援する立場の者が、研究を進めるにあたりその目的とすべきことが、2点浮かび上がる。第一に、①-1「知識・技術的には可能だが時間的、労力的に難しいため記述しない」場合に対応するための、作文の添削を行い易くする何らかの教材の開発である。第二に、②「知識・技術的に難しい」場合に対応するため、作文教育の目的や、あるいは作文教育に携わるものが知っておくべきことについて、教師が根本から学習することを可能とするような参考書や、研修の機会が必要である。

そして、これらの取り組みの際には、特に文章の構成面について重点的に考える必要があるだろう。構成に関するコメントが出にくい要因は様々あるだろうが、少なくともその重要性については多くの教師が認めているため、認識の欠如が理由ではないと思われる。理由として、1つには、添削上で構成について学習者に対しわかり易く言及することの現実的な困難さが挙げられる。これについては、コンピュータ技術を応用し視覚に訴える指導を容易にするような教材を開発するなど、改善のための様々な試みを行っていくことが望まれる。2つ目に、構成に関する視点や把握力が作文教育に独特のものであり、「作文の授業」に取り組み、試行錯誤を繰り返す以外に、それを養成する機会が得にくいことが指摘できるだろう。言語形式の指導やアウトプットの内容への言及については、作文以外の日本語授業と密接に関連しているため、日本語教師としての一般的な知識や経験が直接的に活用できると思われるが、構成に関しては、授業で作文を扱うことがなければ、考えを巡らせる機会がなかなかない。文章の構成に関わる能力の養成を目的とする教師のためのリソースや研修機会の充実が望まれる。

また、当然のことながらそれらの利用者である教師の意識と取り組みも重要である。作文の添削は、一般的に負担の大きい作業と認められているが、その主な原因は、添削者である教師が方針を決め、着眼点を絞って取り組まない限り、記述する事柄に際限がないためであろう。そこで、現実には記述する事柄の限定がなされるのだが、その方針や着眼点

の選択は、書き手の学習段階・学習目標などの情報を把握した上で、これまでに蓄積された経験や知識に基づいて行われると考えられる。すなわち、能力の高い添削者ほど多くの選択肢を持ち、また、的確な選択を行っているということができるだろう。

このような過程を円滑に経て、学習者のためになる記述を行うためには、前提となる日本語の知識や、作文教育に関する知識・経験、学習者の心理面への配慮などを含む、教師としての総合的な能力が必要となろう。そしてこれらのうち、特に見落とされがちなのは、作文教育のために独自に必要な能力である。作文教育に携わっている、あるいは将来携わる可能性のある教師には、日本語教師として一般的に求められる能力を向上させる取り組みだけでなく、作文教育独自に必要な知識を蓄積していくための取り組みが求められる。そのような取り組みは、結果的には教師自身の負担を減らすことの近道にもなると思われる。

参考文献

- 上原久美子(1997)「日本語教育における作文の『記述式フィードバック』について—『コード』による分析の試み—」『南山日本語教育』4, 135-161.
- 坂井恵美子(1998)「作文のフィードバックの方法とその効果」『関西外国語大学留学生別科日本語教育論集』8, 125-139.
- Fathman, A. K. and E. Whalley(1990) Teacher Response to Student Writing: Focus on Form versus Content. *Second Language Writing*. (Edited by Kroll, Barbara)
- Kepner, C. G.(1991) An Experiment in the Relationship of Types of Written Feedback to the Development of Second Language Writing Skills. *The Modern Language Journal* 75(3) 305-313.
- Semke H. D.(1984) Effects of the Red Pen. *Foreign Language Annals* 17(3) 195-202.

資料1：コメントの分類基準と1単位の認定方法について

1. 分類基準

A. 言語形式…文型・語句について代案を提示する，またはその使い方を説明するもの。

例「～という語は～のときに使う」

「～という表現が多いので言い方を変えた方がいい」

漢字，句読点について

B. 内容構成

a：構成以外の内容…内容に関わるが構成には関わらないもの。

b：構成…パーツ（一文以上）の順序に言及しているもの，パーツの過不足に言及していかつ場所についても言及しているもの，パーツとパーツのつながり，関係に言及しているもの，段落分けのことについて。

例「この文（この段落）はない方がいい」

C. その他…書式（言語に関わらない習慣，たとえば「タイトルは3マス開けよう」），タイトルの要求，単に「いい」「わかりやすい」「よく書けている」「まとまっている」など何に関して言及しているのかわからないもの，自分の直し方が正しいかどうか「確認してください」と要求するもの，「作文を書く前に構想を」など手順に関するものなど。ただし「～なのでわかりやすかった」「～がわかりやすかった」は，1つとカウントし「わかりやすかった」以前の部分を用いて分類。「わかりやすかった」の部分をもとにカウントしない。

例「～さんの意見がわかりやすかった」（→内容×1）

2. 切り方（1単位の認定）のルール

①1つのコメント内に異なる種類の言及がいくつかある場合（特に最後に多い）それぞれをカウントする。

例 構成を考えてうまくまとめよう （→構成×1）

・ 第一段落と第二の関係がわかりにくい （→構成×1）

・ 第三段落はなくてもいい （→構成×1）（計3）

逆に，全体で1つのことを言おうとしているために分離しにくいものは1つにカウントする（例を挙げている場合もこれに当たる）。

例（1）「自動詞と他動詞の違いに気を付けましょう。 例)災害(さいがいで)人口(じんこう)(が)減る。不景気(ふけいき)なので，社員(を)減らす。」（→形式×1）

例（2）「て形を使った表現でわかりにくいところがありました。て形を使

った表現をもう一度整理してみてください。」(→形式×1)

例(3)「たばこを禁止した方が良い理由として、大きく二つ例を挙げています。一つは体に悪い、一つは子どもに悪い影響を与える ですが、それぞれの理由にテレビの影響のことが書かれているので一つの理由にまとめた方がいい と思いました。」(→構成×1)

②1箇所を指して「～ですか？それとも～ですか？」と聞いているものは1つとカウント。

③「全体的に…」のような総評も、形式に言及している部分があればそこを形式に、内容に言及している部分があればそこを内容にカウント。

④文になっていなくとも、単なる書き換え案以上の情報を示していればコメントと解釈。逆に、単なる書き換え案でも、一文になっていればコメントと解釈。箇条書きも。

3. 区分が難しい場合の補足ルール

①形式と構成以外の内容との区分

- ・語句について触れているものでも、「わからないのでもっと説明を」のように、文中での説明を求めるものは構成以外の内容とする。
- ・「～という表現はつかわないほうがいい」という時、その理由によって構成以外の内容か形式か分ける。文型や語句の機能を理由としている場合は形式。「印象が悪い」などは構成以外の内容。
- ・同じく語句や文型の代替をすすめていても、内容との関連を理由にしているものは構成以外の内容に。たとえば「この語を使うとあとで述べられていることとつじつまがあわなくなる」など。逆に「意見を強く言うために」など機能が理由になっている場合は形式。

②形式と構成との区分

- ・同じ語句の繰り返しをやめるようにすすめるものは、形式。語句ではなく、パーツ(一文以上)について「繰り返しが多いため簡潔にまとめた方がいい」というコメントは構成。

③構成とその他

- ・「作文を書く前に考えよう」など手順に関するコメントはその他に。同時に構成に言及している場合はその部分について構成にカウント。

資料2：調査票

作文添削に関する調査票

1. 氏名：

2. 所属：

(以上 2 つの情報はデータ整理のためにお書きいただくもので、データベースには載りません。また、以下の情報はデータベースに載りますが、公開を希望しない情報については記載しなくても結構です。ただし、8. 日本語教育歴と 9. 作文教育歴 は必ず記載してください)

3. 性別：[男・女]

4. 生年月： 年 月

5. 出生地（市町村名まで）：

6. その後の主な居住地：

(5. 6. のなかで、あなたご自身の母語習得にとってもっとも重要であったと思われる地点には「 」をつけてください)

7. 職業：[日本語教師・日本語研究者・学生・その他]

8. 日本語教育歴： 年 ヶ月（「国語教育歴」は含みません）

9. 作文の授業を担当した経験年数： 年 ヶ月（概数でかまいません）

10. 作文の授業を担当したことはないが添削の経験はあるという場合は、どのような場面でのように添削されたかを簡潔にお書きください。

11. 日本語教育能力検定試験合格の有無：[有・無]（有の場合は合格年度： 年）

12. 以下の質問にお答えください。回答は、ご自身のお考えや現状について、(5…非常にあてはまる 4…ややあてはまる 3…どちらともいえない 2…ややあてはまらない 1…全くあてはまらない) の 5 段階で答えてください。類似した質問がありますが、統計的に処理するために必要なものですので、お手数ですが全ての質問にお答えください。

① 日本語学習者に対する、添削を含めた作文指導について

(1) 作文の授業では、論理的な考え方の育成を目指したほうがいい

(5 4 3 2 1)

(2) 学習者の作文能力向上のためには、添削されたものを読むだけでは難しい

(5 4 3 2 1)

(3) 添削時、与える情報が多いと学習者の負担が大きくなるので、できるだけ簡潔なほうがいい

(5 4 3 2 1)

(4) 添削時、訂正するだけでなく、気付いた点についてコメントを記したほうがいい

(5 4 3 2 1)

(5) 添削の主な目的は、正しい日本語の例を示すことである

(5 4 3 2 1)

- (6) 添削時、作文の内容・構成については学習者の意図を尊重し、教師が修正や指摘をしなくてもいい (5 4 3 2 1)
- (7) 文末の形 (だ・である、です・ます) が統一されていない場合、直したほうがいい (5 4 3 2 1)
- (8) 同レベルの学習者であっても学習者の個性によって添削の仕方を変えたほうがいい (5 4 3 2 1)
- (9) 学習者のレベルに合わせ、添削の仕方を変えたほうがいい (5 4 3 2 1)
- (10) コメントを記す場合、学習者に考える機会を与えるよう工夫したほうがいい (5 4 3 2 1)
- (11) 添削の際、誤った原稿用紙の使用は必ず直したほうがいい (5 4 3 2 1)
- (12) 学習者に作文を書かせる主な目的は、既習の知識をよりよく定着させることである (5 4 3 2 1)
- (13) 学習者の作文の内容や構成面に問題があれば、直すよう指導したほうがいい (5 4 3 2 1)
- (14) 添削後どう書き直すか、学習者自身が判断できるような形で添削したほうがいい (5 4 3 2 1)
- (15) 添削時、正しい例を示すだけでなく、できるだけなぜそう直したか理由を付したほうがいい (5 4 3 2 1)
- (16) 添削時、意図がわからない文については、無理に推測せず意味不明であることを示すだけでいい (5 4 3 2 1)
- (17) 添削時、語彙や表現の不適切な使用は必ず直したほうがいい (5 4 3 2 1)
- (18) 作文の授業では学習者が書き出し、結びなどの構成を考えて文章を書けるようになるよう指導したほうがいい (5 4 3 2 1)
- (19) 添削をすることそのものよりも、添削内容についてのその後のフィードバックが重要だ (5 4 3 2 1)
- (20) 添削時、直し方にいくつか選択肢がある場合は、1つだけでなく複数示したほうがいい (5 4 3 2 1)
- (21) 添削時、文字の表記の不適切な使用は必ず直したほうがいい (5 4 3 2 1)
- (22) 添削時の明示的な指示や指摘は、添削後の自主的学習の妨げになるのでできれば避けたほうがいい (5 4 3 2 1)

② 日本語教師一般について

- (1) 日本語教師のための研修会等があれば積極的に参加したい (5 4 3 2 1)
- (2) 授業経験を多く積むことは教師の成長にとって大切だ (5 4 3 2 1)
- (3) 教師としての勉強をしたいが時間や機会に制限があるので難しい

- (5 4 3 2 1)
- (4) 今の自分の教え方は特に問題がないと思う (5 4 3 2 1)
- (5) 所属機関の方針があるので今の授業の形を変えるのは難しい (5 4 3 2 1)
- (6) 論文や研究発表は教えるためにはあまり役に立たない (5 4 3 2 1)
- (7) 他の教師と授業見学をお互いにする機会がよくある (5 4 3 2 1)
- (8) 他の教師と授業について話す機会がよくある (5 4 3 2 1)
- (9) 学習者のためにもっと役に立ちたい (5 4 3 2 1)
- (10) クラスに変わった学習者がいても対応できると思う (5 4 3 2 1)
- (11) これからも教師として成長していきたい (5 4 3 2 1)
- (12) 以前はよく教師としての勉強をしたが最近はしていない (5 4 3 2 1)
- (13) 一般に教師はもっと自分のための時間を大切にされた方がいい (5 4 3 2 1)
- (14) 教師が一生懸命になっても報われないことが多い (5 4 3 2 1)
- (15) 一般に教師はもっと学習者の自主性にまかせてもいい (5 4 3 2 1)
- (16) 一般に教師は学習者に対して責任を感じすぎだと思う (5 4 3 2 1)

13. あなたにとって「よい文章」とはどのようなものでしょうか。「よい文章」に必要な条件について、あなたのお考えをご自由にお書きください。「～の場合には～が重要である」というように、場合によって異なるとお考えであれば、それもわかるようにお書きください。

資料3：作文

<fr001j.txt>

ワープロソフト

ワープロソフトを使い方を知ること大事だと思います。会社でみなさんはワープロソフトで書くしよるいを利用するからです。だから、学校で学生は使い方を習うべきだと思います。

それでも、手で漢字を書き方も大事です。いつもワープロソフトを使えば、その方をわすれてなりますから、小学校と中学校とき学生は手で書き方だけ習うべきです。その後で、高校の学生は手でもワープロソフトでも漢字を書くことを知っているはずで。

でも、ある人は書かなければ習う漢字の書き方をわすれると言っています。ほんとうですが、ワープロソフトを使ったらだじょうぶで、使わなかったらじしょをえらびます。

<us004bj.txt>

きょう、日本にコンピュータのワープロソフトはありふれたものになる。インターネット、電子メール、けいたい電話などぎじゅつで情報は早く伝える。しかし、このぎじゅつは新しい問題をひき起こした。ワープロソフトは、一方では便利だり、使*いやすいであるが他方では漢字を書き方が忘れやすいだから、日本語の教しは日本語学習にワープロソフトを使わせるかどうか分らない問題である。

私にとって先生は、私にワープロソフトを使わせなかったからたくさん経験がないが使ったときに便利と思った。じしょで漢字の書き方をあおぎ見ないでワープロソフトで作文を書いた。しかし、ふつうの場合は手で日本語の作文と漢字を書いており、漢字の意味、読み方、書き方も覚えて手伝いたと思う。それだから、日本語学習の教育のために教しは生とワープロソフトと手で書くこととともに作文を書けばいいと思う。それはめいめいにび点と弱点があるかはい合のほうがいいと思う。

<us006bj.txt>

現在、世界中、コンピューター使う人が多くなってきました。コンピューターの便利所がたくさんです。これらの内で、ワープロソフトがとても大切ですが、どんな国でも、ワープロソフトが母語の全体の流暢に困っています。例えば、英語のワープロソフトは自動的につづりを書き直す「スペルチェック」という特徴があります。ですから、ワープロをよく使ったら、だんだん、正しいつづりが本当に分かる言葉が減少します。日本語の仮字が英語のアルファベットより複雑なので日本では、私にとって、その問題がもっと大きいです。つづり問題だけではなく、書く問題であります。漢字が読めて意味が分かればいいですが、十分ではなくて流暢ではありません。日本語ではひらがかなのつづりが同じで漢字が違う単語が多いです。ですから、手で書きなければならなかったら、ひらがなばかりできることは不十分です。その上、今や日本でワープロを使う人は学校で漢字の意味と読むことと書くことを学びました。ですので、ワープロが使っても、この人は書かない人より日本語が了解できます。二字の漢字が同類だったら、よく識別できます。さらに、日本文化は仮字に関係します。ワープロが漢字の大切さを減少させると、日本文化も損害すると思います。

したがって、日本語の流暢は書けることに依存します。

<vn065j.txt>

ベトナムで毎年交通事故の統計について死者の数は増えづつつけています。だから、政府は全体国民に2001年から、100CC以上のバイクを使う人はヘルメットをかぶらなければなりません。

ヘルメットをかぶるを聞いて、皆の多数は反対します。なぜなら、私にとって第一理由はベトナムでヘルメットを生産することはできません。外国から輸入しなければなりません。値段は300,000~400,000ドン一個。ヘルメットの形は丸くて、重さは500gr以上なので使ってから、置く所が必要です。それは第二理由。第三理由はベトナムの交通はルールあまりがないので、道をバイクを使う時、目の速さが必要だ。ヘルメットをかぶる時周りの様子を見にくい。第四理由は女性の方です。家へ出る前に顔をかざらなければならないので、ヘルメットをかぶることは好きではありません。以上の四つの理由のヘルメットの欠点です。皆の考えることはよく考えればヘルメットをかぶることは利点もあります。大切な役割は皆の頭を守ることです。交通事故があった時、頭の方の傷が多くなります。そのうちに若物の事故が大きな割合をしめています。他の利点はヘルメットをかぶる時空気の汚染も避けます。ヘルメットをかぶる時私達はマスクとサングラスのかけることは必要ではなくて、ヘルメットをかぶることは利点が少なくても、私の意見はバイクを使う人はそれをかぶることが必要だと思います。

私にとって、人の命が一番大切です。財産はなくなっても、また働いて、お金で買うことももちろんできます。人の命が亡くなって山ほどお金がたくさんあっても買うことができません。だから、ヘルメットの大きな利点を承智しなければなりません。皆さん、政府の要球を従わなければなりません。

<jp002j.txt>

私は公共の場所における喫煙は規制してもよいと考えています。私自身、タバコは吸わないので、すぐ側で吸われることは、煙を自分も吸ってしまうし、嗅*いもあって嫌です。これがまず1番の理由でしょう。しかし、それでは、喫煙者の気持ちを全然考えていない、分かっていないと言われてしまうかもしれません。私は、喫煙すること自体は個人の自由だと思うし、実際周囲の人で喫煙者はとても多いし、彼らからすると、私達が食事をするのと同様にタバコを吸うのであろうと思います。日常生活の自分の体のリズムに、喫煙が溶け込んでしまっているため、喫煙者には、公共の場所での喫煙に対する違和感が感じられないのだと思います。だからといって許していいとは思いません。やはり、タバコは体に害を及ぼすものであり、本人よりも他人の煙の方が害が大きい以上、吸わない人への十分な配慮が必要だと思います。小さい子供達は、今、周囲からの煙を吸って害が体に蓄積され、大人になると今度は、大人を見て来た通りに自分自身が喫煙し、害をさらに蓄積するのです。しかも現状としては、法律で20才未満の喫煙は禁止されていますが、実際には10代中半から吸う人が多いように思いますが、これから、肺がんの発がん者の人数は、どんどん増加していくのではないのでしょうか。私には医学の知識などありませんが、健康を害するものです。それを、公然と公共の場で吸うことを、規制することに疑問はな

と思います。喫煙所を設けている所は多くあるわけだし、吸うことはできるのだから、やはり規制するところは、きちんとする必要があると思います。

<cn044j.txt>

今、日本ではなく、中国ではたばこのことも問題になる。

歴史についての資料によって、たばこを吸ったことは明時代から始まったそうだ。最初*、人人はたばこを吸うたら、体を調節できると思う。恐らくこれはたばこが作られた原因だろう。

しかし、人人は最近健康意識を増強*するとともに、たばこは人間に損害があるという点に気がついた。そして、たばこを吸った人よりもそのはたけを吸わされた人の方が損害を受けがちなものだ。それにしても、ある人たちはたばこを吸い続ける。彼たちはたばこを吸う損害がわかるのに、公共の場所でたばこを吸う。それはすごい。

私はやく中学二年生になった時に、たばこを吸い始めた。と言う原因はおとなの生活の憧れと好奇心からだ。そうして、いつもクラスメートとたばこを吸った。最近、「会社やレストラン、バスや電車などの公共の場所ではたばこを吸えない」と言う規則が見える同時に、たばこを吸わないような宣伝が始まる。それによって、たばこを吸うなら、癌を起りがちだそうだ。

生命はきれいで、ありがたくないか。だから悪い習慣を矯めなければならない。たばこがその一部分だ。もし、個人な悪い習慣を保つために、生命のようなありがたいのさえ失うたら、ばかではないか。

<cn085j.txt>

たばこを禁止するのは必要ですか。

今の世界には、たばこは一番重要な問題になるようです。大部分の国にはこの問題があります。最近日本にこの問題についての激しい討論があります。ある人はたばこを禁止するはずだと思っています。ところがある人はだれもたばこを吸う権利があるはずだと言います。そして、どのほうがもっともいいですか。私は一番目の意見を支持しています。

まず、たばこは体に悪い物なんです。たばこの中にたくさんの有毒な成分があります。この物は人の肺の癌を引起原因の中で一番重要です。特別子どもによってこれはもっとすごいです。子どもの体は生長してとても弱です。それでたばこは彼たちの肺に悪い影響を酷くしています。子どもは大部分強い好奇心をもっているのです。彼たちは外の人やテレビで放送などの影響を受けやすいです。子どもはおとなの影響を受けてたばこを吸う可能性はとても大きいです。子だちは道理をわかっていて時手おくれになったほうです。そこで、たばこを禁止することはとても必要なようです。

そのうえ、たばこを吸う時に、たくさんの有毒の煙が出てきます。それはたいへんですね。その気体は環境に汚染しています。ある日、私たちの空は青くなくなって、吸っている空気は不潔で木はなくなります。それは怖しいね。だれかそんな地球で生活をしていて欲しいのですか。

だから、たばこを吸うことを禁止したほうがいいです。だれも公共の場所でたばこを吸えなければならぬはずだ。公共の場所はみんなので一人のじゃない。もしある人はたばこを吸いたくて、彼は専門の地方へ行くはずだ。そしてテレビは公共のさいでんなんだから、たばこについての放送を禁止するはずだ。

中国もたばこについての問題があります。今はよく

解決していません。早く解決したほうがいい。たばこを吸っている人に「吸わない」お願いします。

<kr174j.txt>

私はたばこを公共の場所で吸うことにたいしてぜったい反代します。もちろんたばこを吸うことは自分の権利だが、化人にひがいをすべきではありません。私もたばこを公共の場所で吸いますが、化の人にひがいをすることを思えば、いい気分になくなります。

そういう立ち場から見ると、たばこのこうこくもきんしするべきです。たばこのコマーシャルが人にちよくせつてきなえいきょうをおよんではありませんが、大ぜいの人々にたばこを吸いたくなるからです。今たばこを吸っている多くの人々は実は青少年の時からたばこの吸うことはじめました。青少年はたばこがいくらあぶないのかよく分かりませ。つまり青少年はたばこを吸うことをきまるけんりを持ってないです。そのような青少年にたいしてこうこくをすることはどうとくできではないです。

もちろんたばこを公共の場所で吸えないという規則はたばこを吸っている大ぜいの人々にたいして大きい不便です。しかしこういう不便なことが化の人のしよめいをみじかくすることより大事なことはありません。たばこはできれば自身一りで、あるいはたばこを吸う人も吸うべきです。

さいきんたはこのねだんがどんどんふえることもたばこのよくないえいきょうを下がるたもの方法の一つですが、いまよりつよいきそくをつくる必要があります。そうしてたばこのためにくるしている大ぜいの人々をすこしでもへるべきです。

たばこを吸うことはぜったいよいことではないですが、たばこを吸う人のけんりを防ぐはずはないです。しかし化の人のためにでも公共の場所ではたばこを吸わないようにしましょう。

<kr203j.txt>

私はたばこを吸うことについて反対する方です。

一番の理由はからだに障るためです。もちろん、ストレス解消を図って吸うことですが一本を吸えば寿命を何時間ずつちぢめるとして問題になることいせんか。二番の理由は環境の問題です。たばこを吸う人々がたばこの吸い殻をごみ箱にすてれば問題になることはありません。しかし、ほとんどたばこの吸い殻がちいさいのでなんの気もなしにすてて問題になります。一人が吸いませんとおおぜいの人々に被害を与えない方です。さいごにたばこを吸えない人にはたばこの煙が頭をいたいて吐き気をする程度にいやです。ところで、自分のストレス解消を図ると人に被害を与えるこんな行動は本当に文化市民の姿がないだと思います。少数の人々ので多数の人々が被害をこうむることはあることがないことだと思います。たばこを吸う権利を主張すれば人に被害を与えるはいはんに吸ったとすることが私の意見です。

そして、これはべつとの問題のには女がたばこを吸うことについてりかいをできません。理由は皆さんも思うように流産や奇形児などが出産できるかくりつが高いためです。もちろん、こどもをつくりたい女性もいるようです。しかし、みんなが自分のかんがえのとおりにできるのはないだと思います。たばこが与える悪影響をもういちど考えてみて行動するといいいと思います。

学習者日本語作文における「形態素情報」の活用

－「添削情報」との連携をめざして－

鎌水 兼貴

1. はじめに

本稿では、「作文対訳データベース」の作文に対して、形態素情報を付与した「形態素情報データベース」を紹介する。そして利用例の一つとして形態素の連鎖の集計から学習者の作文の傾向を探る方法を示す。また、実際の用例からの検索を容易にするために、「作文対訳データベース」の一部をなす「添削データベース」を利用する。形態素情報でみられた特徴的な例を、「添削データベース」から検索することで、学習者の誤用について調べることができる。本稿では検索のためのツール「添削対照検索」を紹介する。

「形態素情報データベース」「添削対照検索」ともに開発作業中であり、本稿では2006年3月現在の内容に基づいている。

2. 「形態素情報データベース」

2.1. 形態素情報の必要性

「作文対訳データベース」の構成(→p.10)において、「形態素情報データベース(以下『形態素DB』)」は「派生データベース」に位置づけられる。「母語訳」が書き手側の意図を汲み取る情報で、「添削」が読み手側の反応を示す情報と考え、**「形態素情報」は作文そのものに関する情報**ということができる。

「作文対訳データベース」の作文に形態素情報を付加する構想は、峯(2002)によって提案された。峯は、学習者の日本語の習得過程を模索する第二言語研究として、二つのタイプ分けを行った。一つは学習者の産出する不自然な表現の情報を提供し、日本語教授の際の手助けとするために学習者の誤用傾向を記述するタイプであり、もう一つは、正用誤用を問わず学習者の産出する日本語を発達過程の一段階と捉え、中間言語体系を記述するタイプであるとした。

峯は特に後者に注目し、中間言語の研究において学習者の作文に形態素情報を付加することは不可欠であるとした。そして、学習者作文データベースの利用者が日本語習得研究に必要な形態素情報を指定して検索を行おうとするとき、データベースに形態素情報がない場合には、どの利用者が分析の前段階として形態素解析という同じ手順を踏む可能性がある**と指摘した**。当時公開されていた国立国語研究所(2001)の「作文対訳データベース」には、そうした形態素情報が存在しなかった。そのため峯は分析者の共通処理部分となる**形態素解析の情報を付与した拡張データベースの作成を提唱した**。

2.2. 概要

前述の峯(2002)の形態素付与の構想では、形態素分析ツール「茶筌」¹を用い、習得研究上重要な情報が「茶筌」の出力で得られない場合には人手で修正するとしていた。しかし情報の修正作業は煩雑かつ膨大であることや、「研究者間でその分類に異論のない形態素情報で」という客観的基準の作成の困難さなどが予想されたため、現時点では区切りミスの修正や誤用箇所の情報²の付与にとどめた。

作業としては、まず「作文対訳データベース」の作文ファイル(テキスト形式)を「茶筌」を用いて形態素解析を行う。作文は形態素ごとに各行に分けられ、各行は、元の「表層(出現形)」の形態素に続いて、「表層の読み」「基本形(辞書形)」「基本形の読み」「品詞」「活用型」「活用形」の6種類の情報が付加されて出力される。その出力をMicrosoft Excelなどのツールに読み込んで人手によって修正を行った。そしてその修正等の補足情報を付加し、1行あたり合計8つのフィールドからなる表形式のデータを作成した。

ファイル形式はMicrosoft Excel形式ではなく、処理上の利便を考慮してCSV形式ファイル(カンマ区切りテキストファイル)を採用した。図1はMicrosoft Excelに読み込んで表示したものである。

	A	B	C	D	E	F	G	H
16	韓国	カンコク	韓国	カンコク	名詞-固有名詞-地域-国			
17	の	ノ	の	ノ	助詞-連体化			
18	港空	コウクウ	航空	コウクウ	名詞-一般			誤表記
19	会社	ガイシャ	会社	ガイシャ	名詞-一般			
20	の	ノ	の	ノ	助詞-連体化			
21	ばあい	バアイ	ばあい	バアイ	名詞-副詞可能			
22	は	ハ	は	ハ	助詞-係助詞			
23					記号-読点			
24	たばこ	タバコ	たばこ	タバコ	名詞-一般			
25	を	ヲ	を	ヲ	助詞-格助詞-一般			
26	吸う	スウ	吸う	スウ	動詞-自立	五段・ワ行促音便	基本形	
27	の	ノ	の	ノ	助詞-連体化			誤接続
28	人	ヒト	人	ヒト	名詞-一般			
29	は	ハ	は	ハ	助詞-係助詞			
30	昇進	ショウシン	昇進	ショウシン	名詞-サ変接続			
31	する	スル	する	スル	動詞-自立	サ変・スル	基本形	
32	こと	コト	こと	コト	名詞-非自立-一般			
33	が	ガ	が	ガ	助詞-格助詞-一般			
34	不可能	フカンフ	不可能	フカンフ	名詞-形容動詞語幹			
35	な	ナ	な	ナ	助動詞	特殊・ダ	体言接続	
36	規則	キンク	規則	キンク	名詞-一般			
37	が	ガ	が	ガ	助詞-格助詞-一般			誤接続
38	作っ	ツクッ	作る	ツクル	動詞-自立	五段・ラ行	連用タ接続	誤表記
39	て	テ	て	テ	助詞-接続助詞			
40	適用	テキヨウ	適用	テキヨウ	名詞-サ変接続			
41	なっ	ナッ	なる	ナル	動詞-自立	五段・ラ行	連用タ接続	誤接続
42	て	テ	て	テ	助詞-接続助詞			
43	い	イル	いる	イル	動詞-非自立	一段	連用形	
44	ます	マス	ます	マス	助動詞	特殊・マス	基本形	
45	。	。	。	。	記号-句点			

図1 データファイル例(CSV形式ファイルをMicrosoft Excelで表示)

¹ <http://chasen.aist-nara.ac.jp/>

² 誤表記や誤接続といった情報が備考として付記される。誰からみても明らかな誤用に限って示しているため、すべての誤用にわたって検討しているわけではない。データベース作成時の参考情報として使用し、この情報を用いた集計処理等は目的としていない。

形態素情報を付与する作文数は、一つの母語あたり 60 程度を目標にした。現時点(2006 年 3 月)で作業がほぼ終了し、現在チェック段階にあるデータを使用して分析を行った。以下に、本稿で使用したデータの各母語話者別の日本語作文数を示す。

日本	韓国	モンゴル	中国	フランス
66編	60編	42編	60編	68編

表 1 「形態素 DB」における母語別作文数

学習者のレベルは日本語で 400 字以上の長さの文章を独力で書ける程度としたため、結果として大学 2 年次後半以降となった。また、日本語母語話者は日本人の大学生が主である。

2.3. 形態素の連鎖について

形態素情報の利用例として、形態素連鎖を取り上げる。自然言語処理において「N-gram」として知られ(長尾他 1998), N 個の文字列(単語)が隣接して共起する頻度を調べ、連鎖の確率を用いて文の自動解析や生成などに活用されている。日本語教育の分野では作文データから学習者のコロケーション情報の自動抽出を目指した大曾(2004)における一連の研究などがある。本稿では連鎖する単位を形態素として、作文における連続する形態素の共起の頻度やパターンなどから執筆者の文章の特徴を探ることを試みる。

形態素の連鎖を考える場合に、品詞をどの程度細かく集計するか、連鎖数をいくつにするかについて考えなければならないが、本稿における形態素分類は、なるべく形式的な基準による、おおまかな分類から考えることにした。

本稿では「茶筌」による品詞分類の最上位の階層のみを扱った。「名詞、動詞、形容詞、副詞、助動詞、助詞、接続詞、連体詞」という非常に広い分類である³。表 2 の「品詞(連鎖)」の欄が本稿における品詞の取り扱いの例である。学習者の作文の特徴をつかむために、まずは粗い分類から分析することが必要と考えた。

表層	韓国	の	卒業式	は
品詞(茶筌)	名詞-固有名詞 -地域-国	助詞-連体化	名詞-一般	助詞-係助詞
品詞(連鎖)	名詞	助詞	名詞	助詞

表 2 本稿の形態素連鎖における品詞の取り扱い

³ 浅原正幸・松本裕治(2003)による。「レル・ラレル」「セル・サセル」について『茶筌』では「動詞-接尾」と動詞扱いになってしまう。他にも茶筌特有の解析結果があるため「動詞-接尾」のみ下位分類を残すことは問題となる。しかし学習上重要な助動詞であることを考慮し、試験的に「動詞-接尾」のまま残した。

N 個の形態素が連続する場合、「N 連鎖」と呼ぶ。図 2 の例では、形態素の 3 連鎖を 12 形態素の文から得る場合を表している。本稿では文をまたがる連鎖は使用しない。そのため文末の後ろから 2 形態素が連鎖の頭になることはなく、この文からは 3 連鎖は 10 個得られることになる。一般化すると、m 個の形態素からなる文において、n 連鎖(=n 個の形態素の連鎖)の数は、m-n+1 個となる。

連鎖数を短くすると、連鎖の種類は品詞の分類が粗いために少なくなり、一つの連鎖に多様な情報が含まれるため分析は困難になる。一方、連鎖を長くすると一つの連鎖中にまとまりを持った短い連鎖が複数含まれることが多くなり傾向がわかりにくくなる。

原文	公共	の	場所	で	は	禁煙	す	べき	と	思っ	て	いる
1	公共	の	場所									
2		の	場所	で								
3			場所	で	は							
4				で	は	禁煙						
5					は	禁煙	す					
6						禁煙	す	べき				
7							す	べき	と			
8								べき	と	思っ		
9									と	思っ	て	
10										思っ	て	いる

図 2 形態素連鎖の例

このため何連鎖で集計したらよいか問題となる。表 3 は、形態素データベースにおける形態素と句読点の集計である⁴。

まず複文や重文も含んだままではあるが、句点をひとつの単位とする。「平均 1」は、単純に形態素数を句点数で割った値である。日本語母語話者が約 22 形態素ともっとも長く、韓国は日本に近い値を示している。以下、モンゴル、中国と続き、フランス語母語話者は 16~17 形態素ともっとも短い。次に「平均 2」は、文の中の区切りを仮に読点に求め、形態素数を句読点数で割ったものである。日本語母語話者は読点が多いことがわかる。

以上から、8~12 形態素をひとつのまとまりと仮定すると、集計する単位はそれより短い連鎖数が適当と考える。そして 3~7 連鎖で試行した結果、本稿では 5 連鎖で集計することにした⁵。

	形態素数	句点	読点	平均1	平均2
日本	23929	1092	1462	21.91	9.37
韓国	20281	980	666	20.69	12.32
モンゴル	12862	708	404	18.17	11.57
中国	20264	1163	1318	17.42	8.17
フランス	16211	975	623	16.63	10.14

平均 1 : 形態素数/句点数

平均 2 : 形態素数/(句点数+読点数)

表 3 母語別形態素数一覧

⁴ 句読点は茶釜では記号として処理され、記号には空白や括弧なども含まれるが、ここでは無視している。

⁵ 本稿における連鎖数は暫定的なものである。連鎖数の数量的決定については今後の課題としたい。

表 4 は、母語別の品詞の出現頻度を、日本語母語話者における使用頻度順に並べたものである⁶。

もっとも多いのは助詞であり、続いてほとんど同程度の頻度で名詞が続いている。少し離れて動詞が続き、やや少ない頻度で助動詞が続いている。この 4 品詞で全体の 9 割を占め、各母語話者ともに順位も同じである。他の品詞も多少の使用率の違いはあるものの、ほとんど差がない⁷。

連鎖 ¹	日本	韓国	モンゴル	中国	フランス
助詞	35.07% 1	34.74% 1	34.33% 1	34.19% 1	33.24% 1
名詞	30.73% 2	32.37% 2	29.76% 2	33.88% 2	32.08% 2
動詞	15.77% 3	14.61% 3	16.36% 3	13.89% 3	14.50% 3
助動詞	11.28% 4	10.68% 4	10.62% 4	10.48% 4	11.65% 4
副詞	2.18% 5	2.31% 5	2.74% 6	2.19% 5	2.66% 5
形容詞	1.66% 6	2.30% 6	2.79% 5	1.89% 6	2.19% 6
動詞:接尾	1.16% 7	0.67% 9	0.44% 9	0.76% 9	0.57% 9
接続詞	1.08% 8	1.14% 7	1.90% 7	1.21% 7	1.47% 7
連体詞	0.99% 9	1.11% 8	0.93% 8	1.02% 8	1.17% 8

表 4 母語別品詞使用頻度

3. 形態素 5 連鎖からみた作文の特徴

3.1. 形態素連鎖頻度一覧

表 5 は、形態素 5 連鎖の母語別出現頻度を日本語母語話者における頻度順によって並べたものである。日本語母語話者の上位 50 位にあわせて示した。パーセンテージは連鎖総数中の割合であり、隣の数字は出現順位である。上位 20 位は順位の部分を濃く表示している。図 2 で示したように連鎖はそれぞれ重なりあっているため、ある表現の形態素数が集計する形態素連鎖数より長い場合には、複数箇所に分かれる。

概観すると、上位には「名詞+助詞」の連鎖が多い。主語や目的語といった役割を語順ではなく助詞によって表す日本語の特徴があらわれているといえる。さらに「動詞+助動詞」という連鎖もみられる。これらの 4 品詞は、表 4 の上位 4 品詞を占めており、「名詞+助詞」の連鎖から「動詞(+助動詞)」とつながる文は、日本語の典型的な文といえることができる。

⁶ 集計上は「1 連鎖」と同様である。

⁷ 唯一頻度に差があるように思われるものが、試験的に下位分類まで残した「動詞・接尾」である。これは「レル・ラレル」「セル・サセル」という受身や使役を作る助動詞であり、学習者には使いこなすににくい面があると思われる。このように文法項目をより具体的にすれば頻度の違いは現れてくると思われるが、今回は全体の傾向の把握が第一と考え、下位分類を対象とした研究は今後の課題としたい。

連鎖1	連鎖2	連鎖3	連鎖4	連鎖5	日本	韓国	モンゴル	中国	フランス					
助詞	名詞	助詞	名詞	助詞	3.736%	1	4.374%	2	3.436%	2	5.340%	2	4.065%	2
名詞	助詞	名詞	助詞	名詞	3.540%	2	4.556%	1	3.592%	1	5.641%	1	4.281%	1
名詞	助詞	名詞	助詞	動詞	2.963%	3	3.057%	3	2.986%	3	3.509%	3	3.245%	3
名詞	助詞	動詞	助詞	動詞	1.580%	4	1.228%	6	1.850%	4	0.947%	11	1.271%	6
名詞	助詞	動詞	名詞	助詞	1.567%	5	1.598%	4	1.780%	5	1.609%	4	1.696%	4
助詞	名詞	助詞	動詞	助詞	1.475%	6	1.548%	5	1.602%	6	1.431%	5	1.376%	5
助詞	名詞	助詞	動詞	名詞	1.245%	7	1.164%	7	1.112%	8	1.066%	8	1.129%	8
名詞	助詞	助詞	名詞	助詞	1.120%	8	0.799%	14	0.606%	20	0.676%	16	0.691%	15
助詞	動詞	名詞	助詞	名詞	1.024%	9	0.986%	11	1.228%	7	0.987%	10	0.888%	10
助詞	動詞	名詞	助詞	動詞	0.978%	10	0.725%	15	0.754%	17	0.676%	17	0.722%	14
助詞	名詞	助詞	動詞	助動詞	0.869%	11	1.055%	9	1.088%	9	1.283%	6	1.234%	7
動詞	名詞	助詞	名詞	助詞	0.815%	12	0.843%	12	0.910%	10	0.799%	13	0.691%	16
助詞	動詞	助詞	名詞	助詞	0.807%	13	1.031%	10	0.832%	13	1.026%	9	0.876%	11
名詞	助詞	名詞	助詞	助詞	0.786%	14	0.666%	17	0.505%	23	0.543%	22	0.413%	27
助詞	動詞	助詞	動詞	名詞	0.786%	15	0.424%	31	0.816%	14	0.271%	51	0.370%	35
動詞	助詞	動詞	名詞	助詞	0.690%	16	0.429%	30	0.777%	16	0.217%	60	0.302%	41
助詞	名詞	助詞	助詞	名詞	0.669%	17	0.547%	25	0.381%	37	0.434%	31	0.296%	43
名詞	助詞	動詞	助詞	名詞	0.664%	18	1.055%	8	0.669%	19	1.076%	7	0.839%	12
助詞	名詞	助詞	名詞	動詞	0.660%	19	0.631%	21	0.459%	29	0.686%	15	0.561%	20
助詞	助詞	名詞	助詞	名詞	0.648%	20	0.488%	27	0.295%	49	0.365%	36	0.302%	42
助詞	動詞	助動詞	名詞	助詞	0.631%	21	0.705%	16	0.840%	12	0.498%	26	0.500%	23
助詞	動詞	助詞	動詞	助動詞	0.610%	22	0.661%	19	0.886%	11	0.563%	21	0.642%	17
助動詞	名詞	助詞	名詞	助詞	0.606%	23	0.611%	22	0.505%	24	0.577%	20	0.376%	33
名詞	助詞	動詞	助動詞	名詞	0.560%	24	0.661%	18	0.801%	15	0.646%	19	0.518%	21
助詞	助詞	名詞	助詞	動詞	0.501%	25	0.335%	45	0.280%	50	0.345%	40	0.352%	36
助詞	動詞	助詞	動詞	助詞	0.497%	26	0.276%	56	0.459%	30	0.168%	74	0.247%	53
助詞	名詞	助詞	名詞	助動詞	0.468%	27	0.547%	24	0.404%	35	0.528%	24	0.426%	25
名詞	名詞	助詞	名詞	助詞	0.456%	28	0.804%	13	0.575%	21	0.883%	12	0.907%	9
名詞	助詞	名詞	名詞	助詞	0.456%	29	0.651%	20	0.466%	28	0.730%	14	0.771%	13
動詞	助詞	名詞	助詞	名詞	0.426%	30	0.606%	23	0.412%	34	0.661%	18	0.512%	22
名詞	助詞	動詞	名詞	助動詞	0.422%	31	0.256%	57	0.218%	69	0.242%	55	0.222%	61
動詞	名詞	助詞	動詞	助動詞	0.389%	32	0.335%	44	0.443%	32	0.321%	43	0.345%	37
動詞	名詞	助詞	動詞	助詞	0.389%	33	0.291%	52	0.257%	55	0.212%	61	0.259%	51
動詞	助動詞	名詞	助詞	名詞	0.384%	34	0.390%	35	0.451%	31	0.326%	42	0.253%	52
名詞	助詞	名詞	助詞	名詞	0.359%	35	0.340%	43	0.225%	64	0.242%	53	0.179%	74
名詞	助詞	名詞	助動詞	名詞	0.359%	36	0.414%	32	0.233%	61	0.454%	29	0.395%	29
動詞	助詞	名詞	助詞	動詞	0.359%	37	0.463%	28	0.474%	27	0.449%	30	0.395%	30
助詞	名詞	名詞	助詞	名詞	0.351%	38	0.488%	26	0.319%	46	0.498%	25	0.623%	18
名詞	助詞	名詞	動詞	助詞	0.347%	39	0.330%	46	0.389%	36	0.286%	49	0.234%	55
助動詞	助詞	名詞	助詞	名詞	0.339%	40	0.247%	59	0.202%	74	0.276%	50	0.278%	46
助詞	名詞	動詞	名詞	助詞	0.339%	41	0.365%	39	0.249%	60	0.237%	57	0.154%	87
連体詞	名詞	助詞	名詞	助詞	0.334%	42	0.325%	47	0.365%	38	0.360%	37	0.413%	26
助詞	名詞	助詞	動詞	動詞:接尾	0.334%	43	0.148%	88	0.086%	154	0.306%	45	0.142%	97
名詞	助詞	動詞	助動詞	助詞	0.330%	44	0.399%	33	0.474%	26	0.380%	34	0.580%	19
名詞	助詞	名詞	助詞	形容詞	0.309%	45	0.345%	41	0.575%	22	0.370%	35	0.210%	66
接続詞	名詞	助詞	名詞	助詞	0.297%	46	0.399%	34	0.676%	18	0.528%	23	0.401%	28
助詞	名詞	助動詞	名詞	助詞	0.297%	47	0.434%	29	0.264%	53	0.424%	32	0.234%	57
動詞	助詞	動詞	名詞	助動詞	0.293%	48	0.118%	108	0.124%	113	0.069%	164	0.167%	81
副詞	名詞	助詞	名詞	助詞	0.284%	49	0.345%	42	0.225%	66	0.291%	48	0.290%	44
助詞	名詞	動詞	助詞	動詞	0.284%	50	0.202%	72	0.319%	47	0.133%	92	0.148%	90

表 5 母語別形態素連鎖頻度 (5連鎖; 日本語母語話者での頻度順)

3.2. 形態素 5 連鎖の例

3.2.1. 「～である」

母語別の使用頻度に差がある形態素連鎖を選んで、いくつか示す。まずは、日本語母語話者の 54 位である「名詞-助詞-名詞-助動詞-助動詞」を挙げる(表 6)。50 位までを挙げた表 5 の外にあるが、日本語母語話者には比較的頻度の高い連鎖のように思われる。

日本	韓国	モンゴル	中国	フランス
54	147	156	79	108
0.272%	0.089%	0.078%	0.158%	0.117%

表 6 「名詞-助詞-名詞-助動詞-助動詞」連鎖の母語別出現頻度

日本では 0.272% を占めるのに対して、他の母語話者での出現頻度は低い。中国、フランスで日本の半分前後であり、韓国、モンゴルは日本の三分の一以下である。この連鎖は全部で 144 例あった。以下に一部を示す。

あらゆる所で 日本 一番盛んな 直接だめと言う	それ	は	主食	じゃ	なく	て副食で どうでしょう。 と思う。 たが 。 。
	禁煙	が	可能	だっ	たら	
	それ	は	四季	で	ある	
	神道	が	主	で	あっ	
	独自	の	習慣	で	ある	
	行事	は	春節	で	ある	
の	は	無理	でしょ	う		

上記でもわかるように、多くが、

「名詞」+「助詞」+「名詞」+「で」+「ある(あった, あり…)」

という形式で、全体の 75.0%(108 例)を占めている。「名詞+助詞」の連鎖が続いたのちに文が終わる形式が多いと思われるが、文末において「である」の出現頻度が日本語母語話者だけ高いことになる。実際、他の母語話者の作文の例をみると、「です・ます」という丁寧体を用いていることがわかる。

日本語学習においては、丁寧体の会話から学習することもあり、また、会話場面は文章場面よりも多いと考えられるため、作文においても普通体(だ体)は用いにくいと思われる。逆に、日本人は、高等教育では作文では普通体で書くように教えられており、丁寧体を用いることはまれである。そのため学習者で普通体の作文が少ないことと、日本語母語話者において丁寧体の作文が少ないことが重なって、両者に差が出たものと思われる。

3.2.2. 「よく」

前節とは逆に、学習者に多く日本語母語話者に少ない連鎖をみる。ここでは日本語母語話者 199 位の「名詞-助詞-副詞-動詞-助動詞」を取り上げる。

日本	韓国	モンゴル	中国	フランス
199	135	94	91	63
0.075%	0.099%	0.155%	0.138%	0.222%

表 7 「名詞-助詞-副詞-動詞-助動詞」連鎖の母語別出現頻度

例をみると、

それに対して	漢字	を	よく	知っ	たら	ワプーロを
タバコを吸う	場合	が	よく	あり	ます	
けれど実際	それ	は	なかなか	でき	ない	ことで
喫煙する	こと	は	一向に	構わ	ない	が
心づかい	思いやり	を	いつも	忘れ	なく	て
	私	は	ちょっと	偏っ	た	考えかも
	タバコ	を	まったく	すわ	ない	人も
コマーシャル	広告	は	いっぱい	ある	だろ	う。

と、「名詞-助詞」の連鎖が続き、副詞をともなって動詞(+助動詞)がくるパターンと考えられる。日本語母語話者以外でも出現頻度はあまり高くないが、フランス語母語話者だけが 0.222% と突出している。さらに詳しくみると、フランス語母語話者では動詞の前の副詞において「よく」の使用が多く、全 122 例中、副詞部分が「よく」となる例は 27 例(22.1%)だが、その半数以上(15 例, 55.6%)がフランス語母語話者によって占められている。

「よく」は早い段階で学習する副詞であるため使用しやすいことが考えられるが、原因については今後検討するものとする。本稿では、副詞の種類という点に着目する。

表 8 は、この連鎖で出た副詞の母語別一覧である。たしかにフランス語母語話者は「よく」を多く用いていることがわかる。それに比べて他の国は副詞にバリエーションがあるように思える。バリエーションがあるといっても、修飾する動詞の種類を無視することはできないが、ここでは副詞の難易度のバリエーションという視点から考える。

日本	あまり(全て1)、いっこうに、おそらく、きょくりよく、これから、じっさい、しばしば、できるだけ、どう、とうてい、なかなか、べつに、ほとんど、まず、よく
韓国	ぜんぜん(2)、ちょっと(2)、ほとんど、よく、いちどに、いつも、こう、すこし、どうして、ときとき、なるべく、なんと、まだ、もう、もちろん、もっと
モンゴル	ぜんぜん(4)、よく(3)、あまり(2)、ちょっと、ずいぶん、そう、そんなに、のんびり、まったく
中国	ぜんぜん(6)、よく(5)、もう(2)、すこし(2)、いっぱい(2)、もっと、ほとんど、こう、できるだけ、なかなか、いろいろ、しっかり、たくさん、とても、なにもかも、はっきり
フランス	よく(15)、あまり(6)、ぜんぜん(3)、もう(2)、もっと(2)、すぐ(2)、ほとんど、もちろん、しばしば、どう、ぜひ、どうも、また

表 8 「助詞-名詞-助詞-副詞-動詞」連鎖時に出現した母語別副詞一覧

副詞の難易度は、日本語能力検定試験の級数を用いることにする。級数判定には、東京国際大のウェブサイトで開催されている「チュウ太の道具箱—レベル判定ツール」⁸という、文を入力すると文中の単語の級数を出力するツールを用いた。なお「級外」と判定された語は「1級以上」とみなし便宜的に「0級」として処理した。表9に母語別の平均級数を示す。

母語	平均値
日本	2.267
中国	3.250
モンゴル	3.267
韓国	3.389
フランス	3.784

表9 日本語能力検定試験における表8の副詞の平均級数

これまでは「よく」の多用という点からフランス語母語話者と他の母語話者との違いを強調してきたが、平均級数という観点からみた場合、むしろ日本語母語話者と学習者との違いのほうが際立っていることがわかる。

たしかにフランス語母語話者は、日本語能力検定試験4級レベルである「よく」という平易な単語を多用したために最も低い級数になってはいるが、日本語母語話者からみた場合には、フランス語母語話者も他の非日本語母語話者も、「平易な副詞を使用している」という点では同じである。

富谷(2004)は、会話において用いられる単語の日本語能力検定試験の級数は、日本語母語話者と学習者との間にあまり差がないとしている。本稿は副詞だけの例ではあるが、作文という書きことばの場面においては、難易度の高い単語を用いていることがわかる。

作文も会話と同様、意味が伝わることが重要である。しかし「書きことばらしさ」を持つためには、より高度な語彙も必要になるということを、表9は示しているといえるだろう。

⁸ <http://language.tiu.ac.jp/tools.html>

3.2.3. 「形態素 DB」における注意点

「作文対訳データベース」は、できるだけ同じ条件の作文の比較を目的としているために、文章の内容を課題設定によって統制している⁹。しかしその反面、作文が課題文の表現に左右されてしまう部分がある。たとえば、「助詞-形容詞-名詞-助詞-動詞」という連鎖は、学習者では日本人母語話者の2倍以上の出現頻度である(表 10)。

日本	韓国	モンゴル	中国	フランス
137	73	42	46	72
0.104%	0.202%	0.342%	0.296%	0.191%

表 10 「助詞-形容詞-名詞-助詞-動詞」連鎖の母語別出現頻度

しかし連鎖の内訳を見ると、「に」+「悪い」+「影響」+「を」+「与える」という連鎖が多いことがわかった。これは、学習者用の課題文において、

現在日本では、会社やレストラン、バスや電車など公共の場所でたばこを吸ってはいけない、というような規則はありません。ある人は言います。「日本でも、公共の場所ではたばこを吸えないよう、規則を作るべきだ。また、たばこのコマーシャルは子どもに悪い影響を与えるから、テレビで放送できないようにするべきだ」。

一方日本には、次のように言う人もいます。「規則を作って禁止するのはおかしい。だれにもたばこを吸う権利があるはずだ」。

あなたはどのように思いますか。日本ではこれからどうしたらいいか、あなたの意見を書いてください。

と、上記の囲み箇所の「に悪い影響を与える」という文が影響していることが予想される。しかし、これだけでは日本語母語話者において少ない理由がわからない。当初は日本語母語話者だけが課題文の表現を自ら変えて用いるから、といった理由も考えたのだが、実際には単に課題文において、

喫煙を規制するかどうかには賛否両論があります。喫煙は百害あって一利ないものであるから、公共の場所ではたばこを吸えないよう法律で規制すべきだ、またたばこのコマーシャルは子どもに悪い影響を与えるから、テレビ等での放送も厳しく制限すべきだ、という意見がある一方、喫煙者にも喫煙の権利があるはずだから、規則で一律に禁止するのは不当である、という意見もあります。この件に関するあなた自身の考えを、規制反対か賛成か必ずどちらかの立場に立ったうえで、日本語で論じてください。

と「に悪い影響を与える」に変わっていることが原因であると思われる。

以上、作文全てに影響を及ぼすわけではないが、課題文が作文に与える影響というものも注意しておく必要があるだろう。

⁹ 課題は一つではなく、複数から選択式にしている。

4. 添削データベースとの連携

4.1. 「添削対照検索」について

4.1.1. 概要

ある形態素連鎖において、学習者と日本語母語話者との間で使用頻度が大きく異なる場合、学習者がその形態素連鎖に多くみられる用法について、

- (1) 用法は正しく、単に使用頻度が異なる
- (2) 用法は正しいが、誤った別の表現を用いることが多いため、使用頻度が少なくなる
- (3) 誤りを含んだ用例が多いため、使用頻度が多くなる

などが考えられる。(1)の例については、前述したとおり、丁寧体・普通体といった文体の差異や、「よく」をはじめとする簡単な副詞の多用であるというように説明することができる。(2)(3)の例の考察には、「形態素 DB」の誤用情報だけでは不十分と思われる。単語や文法の誤りといった誤用に関する判断自体、明確な基準をもつものではなく、不自然さの度合いは連続的と考えられるからである。さらに判断においては個人差も存在する。

このような場合に利用できるのが、「対訳データベース」におけるもう一つの柱である、「添削データベース」である。上記の二つの可能性について、「添削データベース」から推測するには、以下のような方法をとる。

- (2) → 修正された結果、当該語句になった例を検索する
- (3) → 当該語句を用いたものの修正された例を検索する

添削データは XML によって記述されているため、添削タグを除去した上での検索(添削前の原文)と、添削タグを反映した上での検索(添削後の文)を行うための二つの文の生成は、タグ解析を行うことで容易にできる。現状でも、スタイルシートの切り替えによって、添削前、添削後の文を別々に検索することは可能だが、「添削されたかどうか」をキーとして添削前後を対照して検索をすることはできない。

このツールの目的は、前述のように、添削されている箇所を、添削前の形からでも、添削後の形からでも、どちらからでも検索できるようにすることである。

表示方法は、KWIC 形式とした。タグ解析は検索の度には行わず、事前にタグ解析をして添削前→後、添削後→前という2つの対照用ファイルを作成した。そのため、添削文が追加されたらその都度検索用ファイルを作らねばならない。

プログラムには Perl スクリプトを用いた。Web のフォーム機能を用いて入力し、HTML 形式で出力するために、サーバー上の CGI として動作させた。

4.1.2. 使用データ

データは、XML 添削タグのバージョン 1.0 に基づき公開された、第 1 回添削¹⁰と第 2 回添削¹¹のデータを用いた。表 11 に添削文数を示す。

	作文数	一作文あたりの添削数	母語話者*	非母語話者	計
第1回	256	1	258	49	307
第2回	20	24	460	20	480
計			718	69	787

*一作文の添削数が2人の作文が2つある

表 11 「添削対照検索」に収録した添削

対象となるのは日本語母語話者であるため、非母語話者をのぞく 718 添削文を用いた。一部の作文は複数の添削がなされており、少数の人しか誤用とみなさないような箇所についても拾い出すことができる。ただし添削者による判断が同じ場合には、同質の箇所が複数拾われることになるため、用例数をカウントする際には注意しなければならない。

母語別の添削数は以下ようになる。

韓国	モンゴル	中国	インドネシア	カンボジア	マレーシア	シンガポール	タイ	ベトナム
154	58	68	48	58	146	62	107	86

表 12 「添削対照検索」で用いる作文の母語別本数

現状では、第 1 回と第 2 回の添削データを元にしてしているため、ヨーロッパの添削情報は入っていない。本稿における 4 カ国のうちフランスについてはデータがないことになる。この点は母語別の特徴を考察する場合には問題になるが、どのような例に違和感を覚えるかを調べる場合や、学習者に共通の問題を考えるような場合には、活用できるであろう。

¹⁰ 国立国語研究所(2001) CD-ROM 版・オンライン版

¹¹ 国立国語研究所(2002) モニター版

4.1.3. 検索画面

画面を図 3 に示す¹²。「検索条件」のところに入力欄があり、オプションとして、検索対象を添削前にするか添削後にするか、KWIC 表示において前後何文字にするか、などが指定できるようになっている。今後はさらに、添削前(後)の語を検索する際に、添削後(前)の文字列からの検索も可能にする予定である。

検索方法の説明の詳細は、図 3 の画面表示例の中にある説明をもって省略する。また、実際に検索を行った結果の例については、次節以降の分析中に示す。

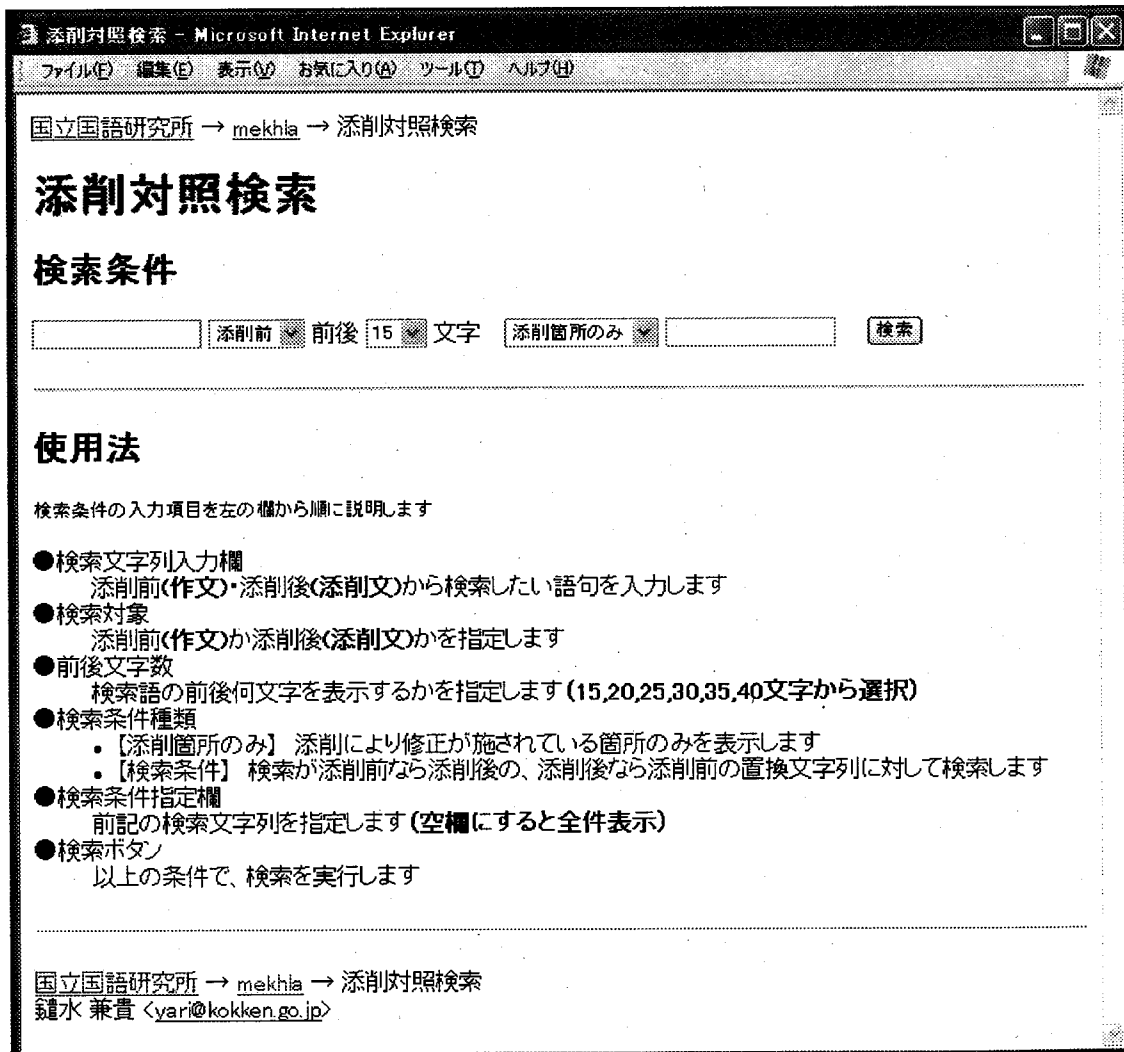


図 3 「添削対照検索」の画面

¹² 現時点(2006年3月)では試作段階であり、今後変更される可能性がある。

4.2. 形態素連鎖と組み合わせた例

前述した形態素連鎖と、「検索対照検索」を組み合わせた例を以下に示す。

4.2.1. 「～ている」

4.2.1.1. 出現頻度の違い

表 5 の日本語母語話者における形態素 5 連鎖の出現頻度上位 50 位では、国別で差がみられるものの、どの国においても用例が多いため、特に学習者において誤りや違和感を含む例が多いとはいえない。

分析する必要があるが、ここでは一例として、日本 15 位(0.786%)の「助詞-動詞-助詞-動詞-名詞」をとりあげる(表 13)。

日本	韓国	モンゴル	中国	フランス
15	31	14	51	35
0.786%	0.424%	0.816%	0.271%	0.370%

表 13 「助詞-動詞-助詞-動詞-名詞」連鎖の母語別出現頻度

「助詞-動詞」が連鎖して名詞に続くため、名詞節を形成しているものと思われる。そのため、例をみると、

形だけになっ	て	き	て	いる	よう	にも感じますが の隣の人達が がいとそこをさけて がこもっています。 です。 です。 です。 と呼ばれる
たばこ	を	吸っ	て	いる	人	
近く	に	吸っ	て	いる	人	
どの年も余裕	が	ある	と	いう	気持	
タバコの害	を	受ける	と	いう	こと	
道端などにポイッ	と	投げ捨て	て	しまう	こと	
それに合った衣装	を	着	て	踊る	の	
習朝の日の出	を	見	に	行く	初日の出	

のように、以下の二つの連鎖が多いことがわかった。

- (1) 「動詞連用形」 + 「接続助詞て」 + 「補助動詞いる」
- (2) 「格助詞と」 + 「補助動詞いう」

全 494 例中、(1)の「～ている」が 287 例(58.1%)を占める。他でも「動詞+助詞」が「動詞」に接続する場合には多い組み合わせである。また(2)の「～という」も 77 例(15.6%)と多い。文や句を構成する「助詞+動詞」を助詞が受けているため多くなるものと思われる。

「～ている」の例が多いが、前述の「～である」のように圧倒的とはいえない。そのため「～ている」だけを再集計して表 14 に示す。

日本、モンゴルが多く、中国が少ないことは前記と変わらないが、日本とモンゴルの差がひらいており、モンゴル語母語話者で「～ている」の頻度が高いことがわかる。

日本	韓国	モンゴル	中国	フランス
0.397%	0.247%	0.638%	0.138%	0.197%
95例	50例	82例	28例	32例

表 14 「～ている」の母語別出現頻度

こうした差は、母語におけるアスペクト表現の影響がみられる可能性が高い。しかしここでは、「添削対照検索」を用いて、別の側面から使用の違いについて考えてみる。

4.2.1.2. 「間違いかた」の違い

検索条件では、「～ている」の「添削前」と「添削後」について用例を検索する。まず「添削前」をみる。これは日本語母語話者の添削によって修正された箇所であり、「本来『～ている』にすべきだった箇所」と考えられる。ただし「より適切な例」として挙げられた場合¹³もあり、必ずしも誤りではないが違和感があるような場合についても示される。

また、一人の添削結果だけが唯一ではないため、誤用データベースに比べて、添削結果の日本語のゆれは大きい。すなわち、誤用かどうか判断がわかるような部分まで含めた広範囲の検索となる¹⁴。

図 4 は、「添削対照検索」の実行例である。画面では、添削後の結果が「ている」になった箇所、すなわち「本来『ている』になるべきだった箇所」を検索している。全部で 394 件あるが、修正箇所の一部に「ている」の文字が含まれていればよいため、「ている→ています」や、単語自体の修正などがある場合など、必ずしも「ている」そのものが修正内容になるわけではない。なお、修正後の文章から「ている」を検索すると 1355 件であった。

一方で、図 5 は、「添削前」すなわち、原文で「ている」と書いたにもかかわらず修正の対象となった箇所を検索した結果の一部である。修正した結果「ている」ではなくなった箇所が 228 件であるため、原文における「ている」の数は 1189(=1355-394+228)件だったことになる。

¹³ タグバージョン 1.0 では better タグ。

¹⁴ この検索システムではタグの種類は指定できないが、better タグの場合には表示上の区別をしている。タグシステムは改良されており、今後の改良では新しいタグシステムにも対応した検索を可能にする予定である。

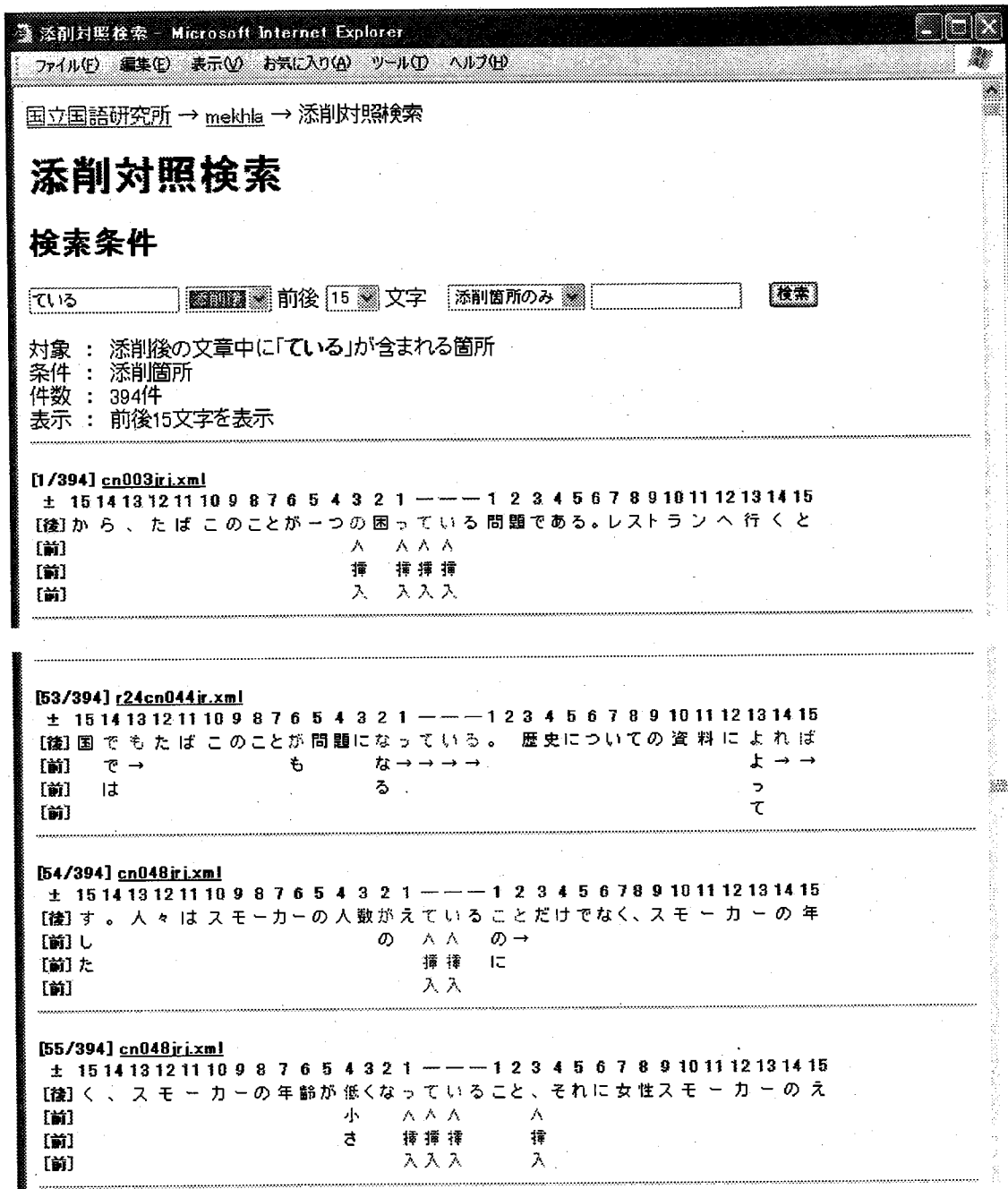


図 4 「添削対照検索」における「添削後が『～ている』である箇所」の検索

国立国語研究所 → mekha → 添削対照検索

添削対照検索

検索条件

ている 添削前 前後 15 文字 添削箇所のみ

対象 : 添削前の文章中に「ている」が含まれる箇所
 条件 : 添削箇所
 件数 : 228件
 表示 : 前後15文字を表示

[1/228] th073iri.xml
 ± 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 — — — 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
 [前] もあって、自分もきれいだと思っでいる人はヨシリストに参加します。
 [後] * * * * *
 [後] く →
 [後] 自
 [後] 分
 [後] 。

[98/228] m1083iri.xml
 ± 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 — — — 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
 [前] すえば 2 秒ぐらい寿命が短くなっでいるといわれています。でも私にとっ
 [後] * * * * * × * * * * × * * * * × * * * * × * * * * × * * * * × * * * *
 [後] 削除 削除 削除 削除 削除 削除 削除 削除
 [後] 除除除 除除除 除除除

[99/228] m1088iri.xml
 ± 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 — — — 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
 [前] なりつつあります。技術が発展していきるとともに親が仕事を大切にします
 [後] す →
 [後] は

図 5 「添削対照検索」における「原文(=添削前)が『~ている』だった箇所」の検索

結果のうち、形態素連鎖にある母語について集計した結果が表 15 である。表の件数の「人」の欄は、修正対象となった作文筆者の人数をあらわす。モンゴルの下の「60人換算」とは、韓国・中国が60人なのに対して、モンゴルが42人しかいないため、60人になおした値である。

母語	原文		修正前		修正後	
	件	人	件	人	件	人
韓国	358	35	60	16	72	22
モンゴル	188	11	39	7	28	7
(60人換算)	269	16	56	10	40	10
中国	38	11	1	1	59	8

表 15 「～ている」に対する添削数一覧

添削が複数人数いるため、ある一人の作文が大勢の添削者から修正対象となった場合には、件数が大きくなることになる。例えば、ある作文において24人の添削者全員に修正されるような誤りをした場合には、件数は24件になってしまう。そのため、そのまま比較することは危険であるが、表15の原文の欄をみとめることにする。

まず、「原文 (=オリジナルの作文中の使用数)」からわかることは、中国ではもともと「～ている」の使用数が少ないということである。それに対して韓国・モンゴルは多くの「～ている」を用いていることがわかる。特に一人当たりの数を見ると、モンゴル語母語話者は韓国語母語話者の1.7倍であることから、形態素連鎖におけるモンゴル語母語話者の「～ている」の使用者の多さと一致する。

つぎに「修正前 (=「～ている」と書いたが修正された数)」をみると、こちらも中国ではほとんど修正された例がないことがわかる。これも前述のモンゴル語母語話者同様、形態素連鎖における中国語母語話者の「～ている」の少なさと一致しているといえる。

そして「修正後 (=修正結果が「～ている」となった数)」をみると、こちらでは中国の値は高くなっている。つまり、本来「～ている」を使わなければならない部分で中国語母語話者は使っていないということになる。一方で、モンゴル語母語話者や韓国語母語話者は、「修正が必要な『～ている』」、すなわち「～ている」の誤用が多いと考えることができる。

以上、「～ている」の誤用についてまとめると、

韓国・モンゴル語母語話者 : 使って間違う
中国語母語話者 : 使わずに間違う

という違いがあることがわかる。この違いは「～ている」を教える際には重要な意味を持つと思われる。

さらに用例からの検討も必要だが、「添削対照検索」における用例出現数の違いから、誤用のタイプの違いを推測し、気づきにくい学習者の日本語使用の特徴を発見できたといえるだろう。

4.2.2. 「～ほうがいい」

「～ている」とは逆に、学習者で多く日本語母語話者で少ない例を探してみる。本稿では品詞の分類が粗いため、低頻度の連鎖で出現頻度が大きく異なる例を探すことは困難である。また前述したように、課題文の影響によって学習者が「悪い影響」という表現が多く使用したため、「形容詞-名詞」を含む低頻度の連鎖において、日本語母語話者が学習者より少なくなる例が目立ってしまった。

ここでは、日本で258位(0.054%)である「動詞-助動詞-名詞-助詞-形容詞」の連鎖に注目する。表16に母語別出現頻度を示す。

日本	韓国	モンゴル	中国	フランス
258	143	58	162	98
0.054%	0.094%	0.249%	0.069%	0.136%

表16 「動詞-助動詞-名詞-助詞-形容詞」連鎖の母語別出現頻度

この連鎖の用例は全部で100例である。内訳をみると、

今まで	考え	た	こと	も	なかっ	た。
たばこを	吸わ	ない	お年寄り	や	幼い	子供にまで
するのは	やめ	た	ほう	が	いい	です。
ですから適当に	使っ	た	ほう	が	いい	です。
考え方をしっかり	持っ	た	ほう	が	いい	かもしれません。
自分の意志に	まかせ	た	方	が	いい	と思う。
規制	し	た	方	が	良い	という考えに
字をやさしく	かえ	た	方	が	いい	です。

のように、多くが、

「動詞未然形+ない」 + 「ほう(方)」 + 「が」 + 「い(良)い」
 「動詞連用形+た」

という連鎖で、全体の61例(61%)を占めている。このうち「ほう(方)が良い」となっている5例は「よい」という形式とみなして除外すると、56例のうち、「ほう」の文字による違いは、「ほうがいい」26例、「方がいい」26例と同数であった。日本語母語話者は3例しかなく、53例が学習者の用例であった。

「ほうがいい」は、アドバイスを表す表現として用いられる。表現としては平易な単語で構成されていることもあり使いやすいと予想される。しかし話しことば的な表現であり、書きことばで用いる場合に、たとえ「いい」を「よ(良)い」に変えたとしても、くだけた印象は残るであろう。

また原義の「方」の意味が示すように、忠告内容のほかに選択の余地がないような場合において「ほうがいい」を用いると違和感が生じると思われる。

「添削対照検索」にて、「学習者が『ほうがいい』『方がいい』と書いたが、添削の結果修正された」箇所を検索する。図 6 は検索画面の例である。

すると、「ほうがいい」は、17 件の修正が加えられていた。全出現数は 135 件であるため 12.6% に修正が加えられたことになる。また「方がいい」での検索結果は 6 件¹⁵であった。全出現数 76 件のうち 7.9% が修正されていたことになる。「ほうがいい」が「方がいい」より修正率が高いのは、「ほう」から「方」への文字の修正があるからであろう。

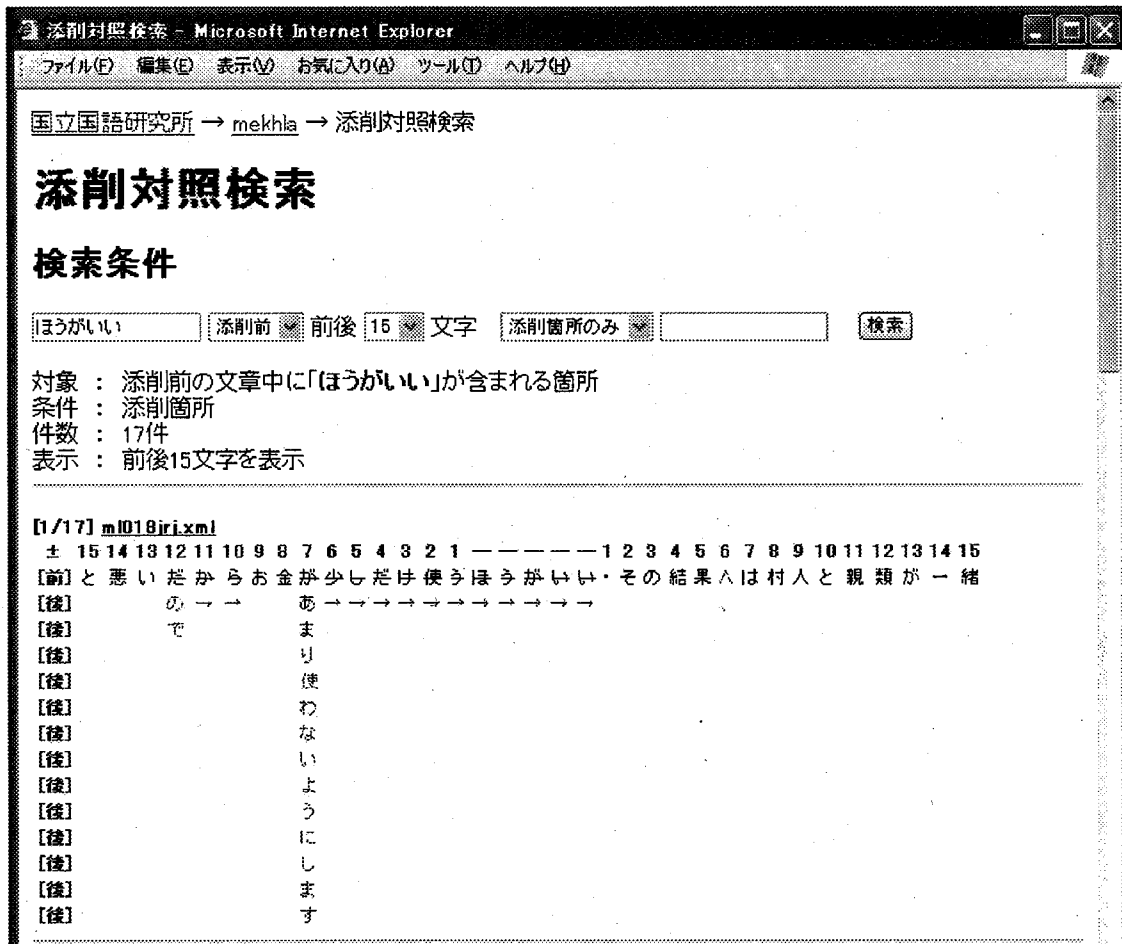


図 6 添削前が「ほうがいい」の検索例

¹⁵ 修正箇所は 7 件あったが、1 件は正用であった。

さらに、検索された添削例を示して検討する。

- ①私の感じによって 次の世代にいつもいい物を教えるの方がいいです。
は 物の 伝えてほしいと思う。
- ②必要とって 子供たちに悪影響あったら無意ですから適当に使ったほうがいいです。
も いけません うべきです。
- ③たばこ すう人と吸わない人のために分別的なルームをしたほうがいいです。
を 部屋 作ったら と思います
- ④でも体が病気になりたい人はたばこを吸ったほうがいいよ。
えばいいでしょう
- ⑤私はたばこを吸えないよう規則を作らないほうがいい、と思っています。
ることに反対です

上記の例をみると、忠告内容について、読み手に対して選択の余地がないと思われる場合、アドバイスとしては適当でないと添削者が判断しているものと思われる。①は希望表現に、②は義務表現に置き換えられている。③、④の場合も忠告内容以外の選択肢がない。⑤の場合には、学習者はここまで明確な判断はしていないようにもみえるが、これは意見文の場合には判断を明確にすべきだという添削態度があらわれているのであろう。

ただし「ほうがいい」自体は、文法的に誤用とはいえないため、同じような状況で日本語母語話者がどのような表現をしているのかを調べるとよいが、形式面のみによる検索では難しい。

そのため、逆に「どのような表現を日本語母語話者が『ほうがいい』『方がいい』に修正するのか」という視点からみしてみる。「添削参照検索」による結果は、「ほうがいい」が修正後が5件(全123件)、「方がいい」は修正後が14件¹⁶⁾(全85件)である。

ここでも文字の修正が多く、「方」から「ほう」への修正が5件であり、ちょうど「ほう」から「方」への修正(6件)に匹敵する。「いい」から「よ(良)い」への修正が4件であり、逆は1件のみであった。

この程度の件数では量的な判断はできないが、推測として、「ほう」か「方」かの選択は添削者の個人差の反映であり、「いい」か「よい」かの選択については、「よい」のほうが「いい」よりも文章語的であるという判断が関係しているのではないかと思われる。

¹⁶⁾ 修正箇所は16件。うち2件は正用であった。

⑥たばこをすわる前に たびたびその悪さ を思わなければならない。
う たばこが体に悪いということ 思い出した方がいい と思います

⑦しかし他の人のためにでも公共の場所ではたばこを吸わないようにしましょう。
他 したほうがいいと思います

⑥は判断の余地を残す表現に置き換えているといえる。⑥の場合には、たばこを吸う人が常に「たばこの悪さ」を思いながら吸うとは考えにくいので、「～なければならない」のような選択肢のない表現は難しいためであろう。

⑦の場合は、聞き手に対する呼びかけという文章的でない表現からの置き換えである。「吸わないで欲しい」「吸うべきではない」ではなく、「吸わないほうがいい」としているのは、「吸わないようにしましょう」に、読み手の都合を考えない、すなわち他の選択肢がないという意識が働いた可能性もある。

以上、形態素連鎖から得られた「ほうがいい」の例について、作文添削データベースの用例から、日本語母語話者である添削者がどのような点に違和感を持つのかを考察した。学習者が「ほうがいい」を多用するからといって、そのまま誤用が多いことにはつながらない。しかし学習者の使用場面が、日本語母語話者の使用場面よりも広い場合には、違和感を生じさせることになる。添削データを検討することで、学習者が用いる「ほうがいい」の意味特徴をつかむことにより、どのような点に注意することで教育現場に活かすことができるかを考えることができると思われる。

5. まとめ

5.1. 本稿の意義

「作文対訳データベース」の派生データベースである「形態素 DB」の分析例として、形態素連鎖という量的な集計から学習者に特徴的な表現を探し出し、さらに「添削データベース」の検索を組み合わせることで、学習者の表現に対して添削者がどのように反応しているのかを考察する例を示した。

形態素連鎖の視点は、本稿では試験的なものに過ぎないが、学習者の文章の生成のパターンを探る上で有効なものと思われる。これまでいわれてきた現象の検証だけでなく、新しい現象の発見にもつながるであろう。

また、試作ツールである「添削対照検索」は、添削前と添削後の両面から用例を検索することが可能であるため、日本語母語話者の表現と学習者の表現が対照されることになる。このことは、学習者の日本語習得過程の分析に役立つものと思われる。

特に「使うべきなのだが使っていない」という形で誤用の例を検索できることがこの検索ツールの特徴である。教育においても「使って間違えた」という部分に対する指摘はしやすいが、「本来使うべきだった」という部分は見えにくく指摘が難しい部分であったと思われる。この点は、教育現場での活用が期待できる部分といえよう。

5.2. 今後の課題

今後は、峯(2002)の計画に沿う形で「形態素 DB」を単なるテキストデータから、XML 文書化して、より汎用の検索ツール¹⁷⁾によって利用できるようにする予定である。検索ツール「添削参照検索」についても試作段階であるため、今後も改良を続ける必要がある。

現段階では「形態素 DB」は CSV 形式でしか作成されていないため、「添削データベース」とは直接はつながっていない。「添削データベース」には形態素情報が付加されていないため、形態素から添削情報を検索することはできない。添削タグとの関係があるため、検索ツール側で対処する可能性もある。また添削後の文章についても、多少は誤用を含むと思われるが、「茶筌」による形態素解析は可能であると考えている。

これにより「添削対照検索」の検索方法においても形態素情報を用いた検索ができるように改良する必要がある。現状では、形態素連鎖が品詞のみの集計であるのに対して、形態素情報は逆に表層の形でしか検索できなかった。今後は、たとえば「動詞+ている+名詞」のように、品詞と表層の組み合わせで検索をできるようにしたほうが、より効率よく分析ができると思われる。

¹⁷⁾ すでに開発されているものとして、国立国語研究所で開発された XML 文書の全文検索システムである、『ひまわり』(<http://www.kokken.go.jp/lrc/index.php>) がある。KWIC 表示をはじめ、資料に適した形での閲覧を可能にする。

参考文献

- 浅原正幸・松本裕治(2003)「ipadic version 2.6.3 ユーザーズマニュアル」
<http://chasen.aist-nara.ac.jp/chasen/doc/ipadic-2.6.3-j.pdf>
(奈良先端科学技術大学院大学 情報科学研究科 自然言語処理学講座)
- 大曾美恵子(2004)『日本語学習辞書編纂に向けた電子化コーパス利用によるコロケーション研究』文部科学省科学研究費 報告論文集
- 川村よし子・北村達也「チュウ太の道具箱—レベル判定ツール」
<http://language.tiu.ac.jp/tools.html> (東京国際大学)
- 国立国語研究所(2001)「作文対訳データベース」CD-ROM
- 国立国語研究所(2002)「作文添削データベース・モニター版」CD-ROM
- 国立国語研究所(2002)『作文教育改善のためのデータベース・ツール活用』2002年度作文教育研究委員会拡大研究会報告書
- 富谷玲子(2004)「大学生のスピーチにおける日本語使用実態の分析—日本語非母語話者と日本語母語話者の比較—」『2004年度日本語教育学会秋季大会予稿集』日本語教育学会
- 長尾真・黒橋禎夫・佐藤理史・池原悟・中野洋(1998)『岩波講座言語の科学 9 言語情報処理』岩波書店
- 峯布由紀(2002)「形態素情報つき日本語作文コーパスデータベース設計の試み—第二言語としての日本語の習得研究へむけて—」『作文教育改善のためのデータベース・ツール活用』国立国語研究所 2002

謝辞

「形態素 DB」の計画・作成に携わった元非常勤研究員の峯布由紀氏に感謝の意を表したい。このほか開発作業に従事した研究補佐員やアルバイトの皆様にもこの場にて御礼申し上げます。

学習者作文に対する教師コメントの分析

—実態の把握・分析と、そこから得られる提言—

宇佐美 洋

1. 学習者作文への教師フィードバックは有用なのか？

1.1. 教師フィードバックに対する否定的見解

教師が、学習者の書いた文章に対し返すフィードバック¹は、果たして教育上役に立つのかどうか。こうした問題意識に基づく研究は、80年代以降、主としてアメリカでの英語教育(L1話者に対するものもL2話者に対するものも含む)の場において盛んにおこなわれてきた。全体的には(特に80年代は)、以下に挙げるような否定的見解が目立っている。

- ・教師はしばしば誤用を見逃すことがある。些細な誤用は直すくせに、重大なあいまいさを引き起こす問題点は放置している。内容に対するコメントや、どう書き直せばいいかというストラテジーの提示は少ない(Zamel1985)
- ・教師のフィードバックは個々の文法的・意味的誤りに対して学習者の注意を喚起することに焦点が置かれ、結果として個々の単語や短いフレーズしか扱っていない(Pica1986)
- ・教師のコメントは学生の作文に役立つという証拠はない(Leki1990)

同様に日本においても、例えば以下のような否定的見解が提示されている。

- ・子どもの作文に対する教師添削は、子どもに達成感を与えず、思考育成にとってマイナスになる場合も(内田1990)
- ・教師は学習者が自力で直せるところにも手を入れている(小宮1991)

こうした批判は、教師がもっぱら「できあがった文章に対する判定者」(特に言語形式の「正しさ」についての判定者)として振舞っており、学習者が自ら考えを深め、自ら向上していくための手助けをほとんどしていない、というところに向けられているといえる。

教育現場からも、教師フィードバックに対する疑問の声はしばしば聞かれる。例えば、手間をかけてフィードバックをおこなっても学習者は注意を向けてくれない、文章を見るだけでは学習者の真意が分からないので面談しながらでない適切なフィードバックはできないが、そのような時間はなかなか取れない、というような声である。

¹ 本論において「教師フィードバック」とは、「教師が教育的配慮に基づき、学習者の作文に対して提示するあらゆる種類の言及のうち、文字によって書き記されたもの」のことを指す。いわゆる「添削」(不適切箇所を指し示し、必要あれば代案を提示する行為)、「コメント」(学習者の作文の一部または全体に対し、基本的には文の形で提示される言及)などをすべて含むことになる。

1.2. フィードバックによらない作文指導法

そこで、できあがってしまった作文に対し教師がフィードバックをおこなうというのではなく、「学習者が文章を書くプロセスに対し、教師が手助けをする」という思想に基づく指導法が提唱されている。いわゆる「プロセスアプローチ」である。この指導法においては、学習者がアウトラインを作り、執筆し、大幅に書き直し、という執筆プロセスを繰り返すことによって自らの思考を深めることに主目的が置かれ、教師にはそれを「判定する」のではなく、「援助する」という役割が期待されることになる（岡崎・岡崎 2001）。

さらに、この「プロセスアプローチ」の流れに位置づけられる学習活動で、近年特に注目を浴びているものとして「ピア・レスポンス」が挙げられる。これは、「作文プロセスの中で学習者同士の少人数グループ（ペア、あるいはグループ）でお互いの作文について書き手と読み手の立場を交換しながら検討し合う作文学習活動」（池田 2002）という活動である。この活動は「教師不在の作文活動」（The teacherless writing class）とも呼ばれ、活動において教師が（少なくとも明示的に）果たす役割はさらに小さいものとなっているのであるが、ピア・レスポンスは作文プロダクトの質の向上に対し、教師フィードバックに劣らぬ効果を示していることが明らかにされている。特に内容面に関しては、教師フィードバック以上の効果が確認されているという（池田 2002）。

1.3. 教師フィードバックに対する肯定的見解

一方で、教師フィードバックに対する肯定的見解も存在しないわけではない。Fathman and Whalley(1990)のように、「内容、言語形式それぞれについて訂正やコメントを受けると、書き直しのとき内容面でも言語形式面でも向上が見られる」ということを指摘する研究も存在する。また Ferris(1995)、石橋(2001)は、フィードバックを受ける学習者側の意識を調査し、大半の学習者は教師フィードバックを有効と認識していることを示した。学習者側には、教師フィードバックに対するニーズは確かに存在しているのである。Leki(1990)は、「教師のコメントは学生の作文に役立つという証拠はない」と述べているが、「教師コメントが不要である」と言い切ってしまうだけの根拠も、見つけることは難しいように思われる。

1.4. プロセスアプローチに対する批判

また、プロセスアプローチ等に対しても批判は存在する。プロセスアプローチでは、学習者が、自分が書きたいことは何か、ということをも自分自身で発見し、その内容を文章として表現していくことに重点が置かれている。つまり「書き手本位」であるというところにその主眼が置かれているのであるが、例えば学術論文を書くときなどは、「その専門領域で期待されること」が何かを知った上で、それに応える形で文章を書いていくことが必要となる。文章に対する「期待」がはっきりしている分野においては、学習者自身の中での思考の深まりを待つプロセスアプローチはあまり向かない、というのである（岡崎・岡崎

2001)。ピア・レスポンスにおいても、読み手の「協働学習者」がその専門領域での「期待」を熟知しているとは言えない場合、必ずしも十分な効果が得られないことは十分考えられる。こうした場面においては、専門領域で期待されていることを熟知した教師が、学習者の文章を適切に読み、適切にフィードバックを返せる（もちろん、形式面だけでなく内容・構成面に関して）、ということも、極めて重要な意味を持っているだろう。

1.5. 筆者の主張

このようなことを踏まえた上で、筆者は以下3つの主張をしたいと考える。

- 1) 作文指導の方法に唯一絶対のものはない。学習条件・学習目的にあわせ、指導スタイルも適宜使い分けられるべきである。

筆者自身は、プロセスアプローチ、ピア・レスポンスの効果について疑念を持っているわけではない。教育において、教師と学習者・学習者同士の対面でのコミュニケーションが有用であり、かつ学習者自身による「気づき」が重要であることは言うまでもない。しかし作文指導というものが、「常にそういう形態でおこなわれなければならない」とは思わない。場合によっては対面での説明より、記録として残せる記述的フィードバックのほうがより効果的、ということもありうるだろうし、10人からのあいまいなコメントより、プロの読み手による正鵠を得たたったひとつのコメントのほうがはるかに有用、ということもあるはずである。どの方法が最も優れているか、という考え方ではなく、それぞれの指導スタイルの特色と利点を知り、場面に応じてそれぞれのスタイルの利点を最大限に活用する、ということを考えるべきであろう。

- 2) プロセスを見ることはもちろん大切。しかし、「プロダクトの重要性」も軽視されてはならない。

学習者自身が思考を深めていくのを援助していくためには、執筆のプロセスに着目しなければならないのは言うまでもない。しかし一般に文章の読み手は、完成した結果としてのプロダクトしか見てくれない。文章を書く、という行為は、基本的には「プロダクトで勝負する」ということなのである。

そこで筆者は、「書き手本位の指導法」から、「書き手に読み手の存在を意識させる指導法」へとさらに歩を進めることを提唱したい。それはつまり、「自分のプロダクトが読者にとってどのように受け取られるのか」という視点を書き手自身に持たせるということに他ならない。こうした視点を持たせることは、学術論文を書くなど「アカデミックライティング」の分野において特にその必要性は高いだろう。

このために教師に必要とされるのは、「プロダクトから問題点を確実に読み取り、必要あ

ればその問題点に対して適切なフィードバックを返せる」という能力である。もちろん、読み取った問題点をすべて学習者に伝えなければならない、というわけではない。状況に応じて、何を伝え、何をあえて伝えないかという取捨選択は必要となろう。しかし、まず手元にあるプロダクトから問題点を確実に読み取っておけるということ、それは、プロセスアプローチやピア・レスポンスといった教育活動を円滑に進めていくために必要不可欠な能力であろうと考える。

したがって筆者は、

3) 「効果的フィードバック」に関する調査研究を進める必要は、依然として存在する

ということを、ここで改めて主張しておきたいと思う。

もちろんその際のフィードバックは、教師から学習者に対し、「ここは正しい」「ここは間違っている」という判定結果を一方向的に知らせるようなものであっては意味が薄い。「学習支援のきっかけ」となりうるようなフィードバック、学習者のアイデアをふくらませることのできるフィードバックとはどういうものなのか、ということについて、科学的な調査に基づいた研究が必要であると筆者は考える。

1.6. リサーチクエスチョン

上記のような主張に基づき、筆者は以下のことを明らかにするための調査をおこなうことにした。

1) 日本語教師は、学習者の作文のどのような点に着目して、どのようなコメントを書いているか。

学習者の作文学習にとって、どのようなコメントが有効なのかを明らかにするためには、実際の教育場面において学習者にさまざまな形態のコメントを返し、それによって学習者の作文能力がどのように向上したか、ということを見ていく必要がある。しかしそのような調査に入る前に、まずは教師のコメント執筆にどのような特徴があるか、実態把握をおこなう必要があると考えた。

2) 作文教育経験の多寡によって、コメントの書き方に違いがあるか。その違いから、「効果的コメントの書き方」についての示唆が得られないか。

一般に教師は、作文添削の仕方やコメントの書き方について、体系的に学ぶ機会はあまり持たないものと考えられる。ただ、教育経験を積み重ねることによって、自ら添削能力・コメント執筆能力を向上させていくという可能性は考えられる。石橋(2002)においても、

経験の長い教師と短い教師とでは、添削時の行動にかなりの差異がある、ということが指摘されている。

今回の調査では、教育経験の長い教師と短い教師の間に、「コメントの書き方」について何か違いが見られるか、その違いはコメント執筆能力の向上によるものと考えていいか、また、教育経験を積み重ねることによっても変化があまり見られない側面があるとしたらそれはどういう点か、ということについて考察を進める。

2. 調査方法

調査は、以下のような手順でおこなった。

2.1. フィードバック対象作文

「日本語学習者による日本語作文と、その母語訳との対訳データベース」(以下、「作文対訳DB」)に収録された作文のうち10編を選び、日本語学校に勤務する日本語教師に添削およびコメント執筆を依頼した。

「作文対訳DB」には、「意見文」および「叙述文」が収録されているが、今回フィードバックの対象として選んだのはすべて「意見文」であった。意見文の執筆課題は、例えば「たばこを法律で規制することについてどう思うか」「日本語学習においてワープロソフトを使用することについてどう思うか」というようなものであった。1編あたりの文字数は300～800字程度である。

10編の作文について、執筆者の母語の内訳は以下のとおりである：フランス語2、英語2、ベトナム語1、中国語2、韓国語2、日本語1。

2.2. 教師の属性

添削・コメント執筆は、首都圏の日本語教育機関（主として日本語学校）に勤務する現役の日本語教員計33名にお願いした。

教師の属性の内訳は以下のとおりである。

*作文教育歴豊富群（以下「豊富群」）：15名

作文教育歴3年以上²

教育歴平均9.5年、最長16.5年、最短3年、標準偏差4.6

*作文教育歴寡少群（以下「寡少群」）：18名

作文教育歴3年未満

教育歴平均0.8年、最長2.42年、最短0年、標準偏差0.7

² 作文教育経験の長い経験と短い教師について、フィードバック過程の違いを分析した石橋(2002)では、作文教育歴3年以上の教師と3年未満の教師を分けて分析している。本研究でも「3年以上・3年未満」という基準を踏襲することとした。

いずれも、日本語教育歴は「1年以上」の方をお願いした(平均8.0年、標準偏差5.7)。また上記で「作文教育歴」というのは、「作文指導を主たる目的とする、恒常的な授業における指導の年数」のみを指すこととし、個人的・散発的な作文指導はこの年数には含めないこととした。結果として、日本語教育歴は極めて長いが「作文指導を主たる目的とする授業」を担当した経験はなく、このため「作文教育歴寡少群」のほうに分類される方もいた。

所属機関によって作文指導の方針が異なっていることを想定し、調査結果に偏りが出ないように、フィードバックをお願いする先生の数はひとつの日本語教育機関から最大6名を限度とすることとした。

2.3. フィードバック時の依頼内容

協力者の先生方には、以下のような条件の下でフィードバックをおこなうよう依頼した。

- *フィードバック対象の作文は、「自分の意見を、分かりやすく、かつ説得力をもって伝える文章を書く」という目的の作文授業で書かれたものと想定し、その目的にそって添削をお願いしたい。
- *フィードバックが可能なのは1回限りであり、その1回限りのフィードバックの中で、必要と考えられる情報を可能な限り盛り込んでいただきたい。つまり、フィードバック済の作文を執筆者に返し、書き直し・再提出させた上でもう一度フィードバックをおこなったり、直接執筆者と面談して、フィードバックの意図について説明したりすることはできない。
- *執筆者に対するコメントや質問、メッセージなどは、適宜余白に書いてほしい(ただし、「必ずコメントを書いてほしい」という依頼はしなかった)。
- *執筆者の母語・日本語レベル等は特に知らせない。日本語レベル等は書かれた作文の内容から推測し、それにあわせてフィードバックをしてほしい。

「フィードバックは1度限り」という条件を設定したのは、「教師は、学習者のプロダクトだけを見たとき、そこから何を読み取り、それに対してどのような反応を返すのか」を調べるためであった。

日本語学校などにおける一般の作文指導においては、学習者との面談等によって学習者の真意を確認したり、フィードバックの意図について学習者に直接説明したりする機会がある。その点で、最も一般的な作文指導のやり方とはやや異なっているが、「通信添削」のような場面においては十分ありうることであり、特に不自然な状況設定ではないと判断した。

また作文ごとのフィードバックにかかった時間を、作文ごとにそれぞれ記録するよう依頼した。

2.4. 分析に使用したデータ

フィードバックを依頼した 10 編の作文のうち、フランス語母語話者による作文 1 編は極めて意味がとりにくいものであった。このため、途中でフィードバックを放棄してしまった方がいる反面、3 時間かけて極めて詳細なコメントを書かれた方もいるなど、コメントの書き方に非常に大きなばらつきが見られ、単純な比較は困難であると考えられた。そこでこの作文については今回の分析対象からはずし、残りの 9 編に対して書かれたコメント³を分析対象とすることにした。

コメントは、文章全体に対して書かれたものも、文章の一部に対して書かれたものも同様に分析対象とした。

3. コメントの分類

3.1. 先行研究

学習者作文に対する教師フィードバックを分類する試みは、例えば以下のような先行研究においておこなわれている。

上原(1997)

- I. 形態的要素についての指摘
- II. 形態的要素についてのコメント
- III. 修辭的要素についてのコメント
- IV. 内容への応答コメント
- V. 評価に関わるコメント
- VI. 評価

石橋(2002)⁴

- I. 表層レベル修整
形態上の修整
意味保存の修正
- II. テキストレベル修整
マイクロ修整
マクロ修整

しかし上原(1997)の分類は「コメント」と「添削」の両方を、石橋(2002)の分類は「添削」のみを扱うものであり、「コメント」のみを分類する基準ではない。そこで今回筆者は、特に「コメント」のみを扱うための分類基準を新たに設定することとした。

3.2. CC (コメントカテゴリ) の設定

まずは、コメント分類のカテゴリを設定した。

1 人あたり 9 編、33 名分の計 297 編のフィードバック情報付き作文から、「コメント」を抜き出し、性質の似たもの同士でグループ化することにより、以下 5 種類 (下位分類を含め 6 種類) の「コメントカテゴリ (以下「CC」) を得た。

1. 評価 (教師の心的判断を明示的に表示するもの)

³ 今回は「添削」の内容については分析対象とはしなかった。

⁴ 石橋(2002)の分類は Faigley & Witte(1981) に基づくものである。

1-1 肯定的評価

1-2 否定的評価

2. 指示（執筆者に対し何らかの注文を出すもの）
3. 解説（執筆者に対し実態を指摘したり説明を与えたりするもの）
4. 文章の内容に関する質問（意味不明の点について情報を求めたり，内容に疑義があるときに執筆者の真意を確認したりするもの）
5. その他（感想，執筆者に対するメッセージ，文章の内容に関係しない質問など，上記のカテゴリのいずれにも当てはまらないもの。意味不明のコメントも含む）

分類にあたっては，コメント文中に以下のような表現が含まれていることを手がかりとした。

1-1 肯定的評価

よく／うまく／上手に書けている，論理的だ，すばらしい，説得力がある，明快だ，～ができています，しっかりしている，読みやすい…

1-2 否定的評価

分かりにくい，分かりません，～ができていません，～が気になります，変です，不十分です，説得力がない，ふさわしくない，記述内容が不適切，内容に疑問がある…

2. 指示

～してください，～しよう，～すること，～したほうがいい，～ないほうがいい，～ないといけない，～するともっとよくなる，必要です，必要ありません，～に注意…

3. 解説

あなたの作文は～となっている（実態指摘），一般に日本語では～するものだ（常識の説明），～と添削したが，その理由は～だ（添削理由の説明）…

4. 文章の内容に関する質問

なぜですか？ どういう意味ですか？（意味不明の箇所について質問），～ということですか？（自分の読みが正しいかどうかを確認），本当ですか？ これでいいのですか？（記事内容の正しさ・信頼性について確認）…

上記 CC のいずれにも当てはまらない「5. その他」の例としては，例えば以下のようなものがあつた。

もったいない…

漢字はにがてですか？⁵

今はけいたい電話でメールができますから、とても便利になりました。でもけいたい電話の漢字もときどき正しくありません。ですから、手で書く練習もしなければなりませんと思います。

上記の分類のうち、「評価」と「解説」の間でやや判定に迷うものがあった。例えば「漢字がたくさん書けています」というようなコメントである。

これは、表面的には「実態の指摘」（あなたの作文はこうなっている）であるが、コメントを書いた教師としては「たくさん書けている、だからよい」という「肯定的評価」を意図していたのかもしれない。

しかし、コメント執筆時の教師の意図を正確に推測することは不可能である。このような場合は、あくまでも文の表面的な意味を重視して「3. 解説」と分類することとした。一方、「漢字がたくさんかけているところが非常にいいです」のように、教師の心的判断が明示的に表現されていれば、「1-1 肯定的評価」とした。

3.3. CU（コメントユニット）の設定

次に、コメントの数を数えるために、どこからどこまでをひとつの「コメントユニット（以下「CU）」として認定するかを決めなければならない。

コメント分割にあたっては、以下2つの基準に拠ることとした。

<基準1>

コメント1文の中に、異なるCCに属する内容が含まれる場合は、同一のCCに属すると考えられる部分をそれぞれ1CUとする。

例えば、「中国におけるたばこの問題について具体的に説明し、〇〇さんの意見も明解です」というコメントは、以下のように分割する。

中国におけるたばこの問題について具体的に説明し、…「3. 解説（実態指摘）」
〇〇さんの意見も明解です…「1.1 肯定的評価」

<基準2>

複数の連続するコメントが、作文内の同一箇所について、同一の視点からなされており、なおかつ全体が同一のCCに属すると考えられる場合は、それら全体を1CUとする。

⁵ 形式的には「質問文」であるが、文章の内容に関する質問ではないので「その他」になる。

例えば、「中学二年生のときにたばこを吸ったことについてどう思いますか。同じように、若者がたばこを吸っていますが、どう思いますか」というコメントは、2つの文によって構成されているが、全体で「4. 質問」とする。

4. 分析結果

4.1. CC ごとの分布の比較

豊富群、寡少群それぞれについて、CUの総数を調べたところ、結果は以下のとおりであった。

豊富群 (15名) : 376 ユニット
 1名あたり 25.1 ユニット
 寡少群 (18名) : 361 ユニット
 1名あたり 20.1 ユニット

1名あたりのCU数はやや豊富群のほうが多いようであるが、t検定で検討した結果、統計的有意差は出なかった ($t(31)=0.67, p=0.50$)。

また、CCごとのCU頻度数の分布は以下のものであった。

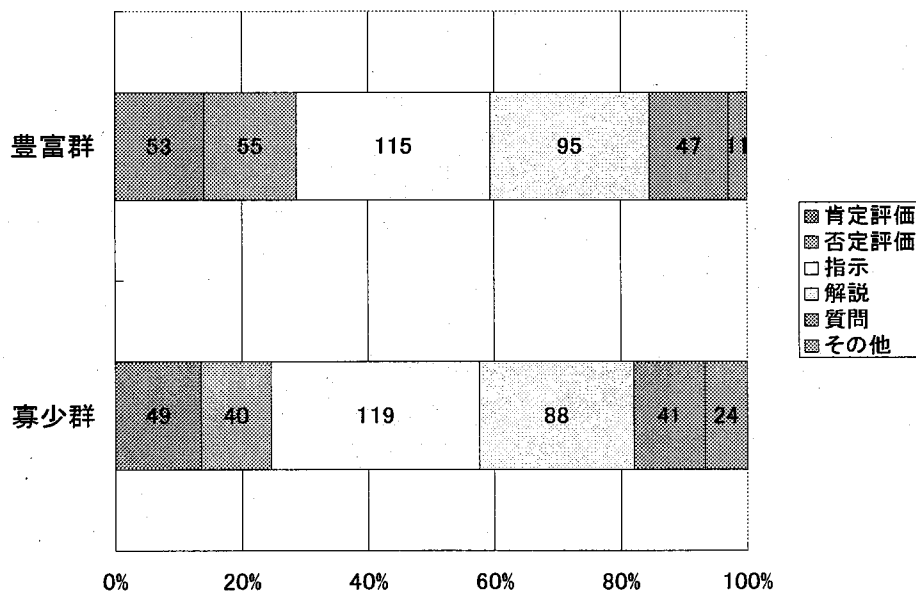


図1 豊富群・寡少群のCCごとのCU分布

豊富群・寡少群ともによく似た分布をしており、 χ^2 検定で検討した結果、豊富群・寡少群それぞれのCU分布に有意差は出なかった ($\chi^2(5, N=33)=7.80, p=0.17$)。

4.2. 具体的コメントと非・具体的コメント

4.2.1. 判定基準

3.2 節で設定した CC については、豊富群・寡少群の間に CU 分布の違いは見られなかった。しかし、コメントの内容をさらに詳細に検討してみると、「評価」および「指示」という「執筆者に対し何らかの『積極的働きかけ』をおこなう行為」について、その働きかけが「具体的なものであるか、そうでないか」という観点で再分類をおこなったとき、豊富群と寡少群の間はかなり顕著な差が見出されることが分かった。つまり、豊富群のコメントは具体性に富んでおり、何を、どのように改善すればいいか、ということがコメントを受ける学習者に伝わりやすいと思われるのに対し、寡少群のコメントは包括的であいまいであり、学習者にとって何をどうすればいいのかという指針が、必ずしも明瞭に示されているとは限らない、ということである。

このことを数量的に確認するためには、「具体的コメント」と「非・具体的コメント」とを、一定の基準によって明確に切り分けることができなければならない。

そこで筆者は、コメントが具体的かそうでないかを判定するための根拠として、以下 3 つの基準を設定することとした。

<基準 1>

文章中のどの箇所に対するコメントであるかが明示されているか

<基準 2>

文章中のどういう側面・どういう項目に関するコメントであるかが明示されているか

<基準 3>

評価・指示の理由が明示されているか

この 3 基準のうち、少なくともひとつ以上を満たしているコメントを「具体的コメント」とし、基準をひとつも満たしていないコメントを「非・具体的コメント」と判定することとした。

以下に、判定例を示す。

[1] 「3 段落目は言っていることが分かりません」

[2] 「意味の分かりにくいところがあります」

[1]のようなコメントは、「3 段落目は」の部分で文章中の特定の箇所を指し示し、この箇所についての言及である、ということを示しているが、[2]にはそのような場所指定が

なく、どこを指しているのかが不明である。したがって[1]のようなコメントは、「具体的・否定的評価」、[2]は「非具体的・否定的評価」と判定されることになる。

[3] 「「べき」「はず」の使い方に注意」

[4] 「語彙の意味に注意」

[3]では、文法項目の中でも特に「べき」「はず」の使い分け、という特定の事項について学習者の注意を喚起しているが、[4]では「語彙の意味」といった、極めて漠然とした事項⁶にしか言及しておらず、具体的な注意喚起の機能は果たしていないものと考えられる。[3]は「具体的・指示」、[4]は「非具体的・指示」と判定される。

[5] 「(今まで英語の話でしたが、ここから日本語の話が始まりますから) 行を変えた方がいいです」

[6] 「意味を考えて段落を分けてください」

[5]に関しては、「行を変えよ」という指示をおこなうにあたってその理由も示されているため「具体的・指示」となるが、[6]はそのような理由も、場所特定も、特定の事項に対する注意喚起もないため「非具体的・指示」となる。

4.2.2. 具体的コメント、非・具体的コメントの分布の比較

「評価」「指示」に分類されたコメントについて、上述のような基準によって「具体的」「非具体的」に再分類をおこなうと、豊富群・寡少群の間には以下のような顕著な差が見出された。

⁶ 注意喚起の範囲を、「具体的言語要素を特定できるまで絞り込んでいる場合」を「具体的」と判定することとした。例えば、「助詞に注意」というコメントは、「文法に注意」などというコメントよりは絞り込みがおこなわれているが、「どの助詞なのか」という、特定の言語要素までの絞り込みはなされていないと考え「非具体的」と解釈する。ただし、「助詞に注意：例えば「は」「が」のように、実例が挙げられていれば「具体的」とした。

⁷ このコメントの前半部分は、「実態指摘」をしていると考えられるため「解説」に分類される。「解説」に関しては、具体的・非具体的の分類はおこなわない。

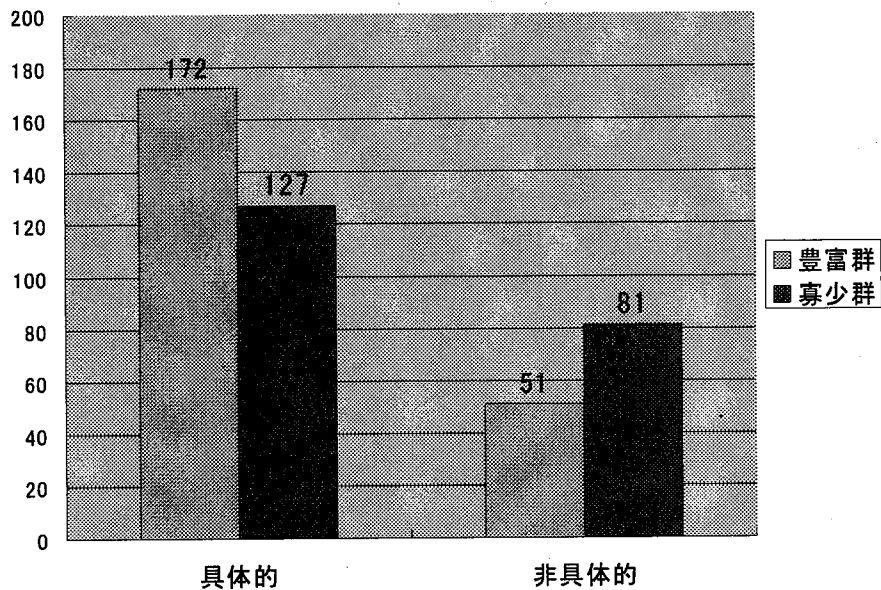


図2 具体的コメント・非具体的コメントの出現頻度 (全体)

豊富群には具体的コメントが、寡少群には非具体的コメントが相対的に多い、ということが明らかに見て取れる。具体的・非具体的コメントの分布について、豊富群と寡少群の違いを χ^2 検定で検討したところ、極めて高い水準での有意差が認められた ($\chi^2(1, N=33)=12.34, p=0.0004$)。コメントの具体性と、作文教育歴との間にはかなりはっきりとした相関があるといえそうである。

さらに、CCごとに具体的コメント・非具体的コメントの分布を示すと以下のようなになる。

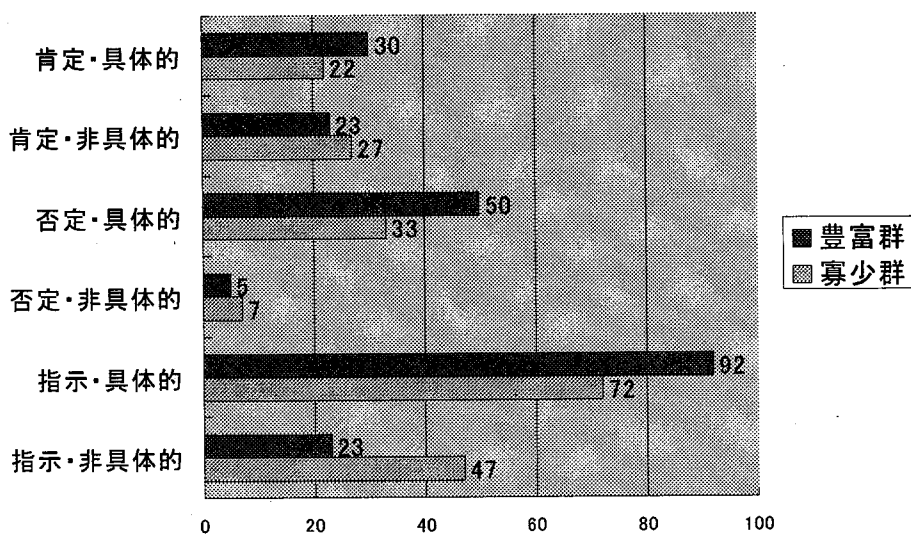


図3 具体的コメント・非具体的コメントの出現頻度 (CCごと)

このグラフを見ると、特に「指示」という行為をおこなう際に、豊富群と寡少群の間でコメントの具体性について差が出やすいことが分かる。

さらに興味深いのは、豊富群も寡少群も、「否定的評価」に言及する場合は「非具体的」なコメントが非常に少なくなる、という点である。「否定的評価」をおこなうということは、学習者に対してはかなり大きな精神的ストレスとなるものである。このため豊富群も寡少群も、否定的評価をおこなう際には学習者に対し十分な配慮をおこなおうとし、結果として非具体的なコメントが少なくなったものと考えられよう。

4.3. フィードバックにかける時間の比較

次に、豊富群と寡少群それぞれが、フィードバックという行為に対してかけた時間を比較してみる。

1編の作文フィードバックにかけた平均時間は、豊富群が21.9分、寡少群は18.6分であり、豊富群のほうが平均にして3分程度長い時間をかけていることになる。t検定により平均の差の検討をおこなったところ、この差は0.5%水準で有意であることが明らかになった ($t(292)=2.86, p=0.0045$)。これは、先行研究とは異なる結果であった。

石橋(2002)は、作文教育歴3年以上の「豊富群」と3年未満の「寡少群」それぞれがおこなったフィードバックの平均所要時間を計ったところ、豊富群の教師のほうが短時間でフィードバックを終えていたという(豊富群24.0分、寡少群33.0分)。そしてこの結果に対し石橋は、「第2言語の作文指導経験豊富群の教師にも寡少群の教師と異なる、より明確な学習者の能力別テキストスキーマがあり、調整への意思決定に関与している」(石橋2002:8)という解釈を示している。石橋のこの解釈は、大筋において首肯できるものだといってよいだろう。

では今回の調査で、むしろ豊富群のほうがフィードバックに長い時間をかけていたのはなぜなのであろうか。石橋の調査と今回の調査とで最も大きく違う点は、今回の調査においては、フィードバックが可能なのは1回限り、という条件が指定されていたということである。このため教師は、プロダクトから最大限の情報を読み取り、必要と考えられるコメントを1回限りのフィードバックの中に盛り込むよう、暗示的に求められていたことになる。

このような状況に置かれたとき、豊富群の教師は寡少群の教師に比べより深く、より分析的に作文を読み込もうとし、そのことがフィードバック時間の長さにつながってくるものと解釈できるかもしれない。具体性に富んだコメントを書くためにはそれなりの深い読みが必要になるものと考えことから、そのような推測が成り立ちうるといえるだろう。

もちろん今回の調査では、添削時間とコメント執筆時間とを切り分けることができないため、豊富群の教師が何のために時間を使ったのかを正確に突き止めることは不可能である。上記の推測について検証することは、今後の興味深い研究テーマとなりうるだろう。

4.4. 学習者作文の「内容の妥当性」に関するコメント

最後に、フィードバック時に必要となるさまざまな視点の中で、作文教育経験の長さだけによっては獲得しがたいものもあるのではないかと、ということを指摘しておきたい。

今回フィードバックを依頼した作文の中には、文章の内容として妥当性にやや欠けると判断される箇所が散見された。例えば以下のような箇所である。

[7] 今の世界には、たばこは一番重要な問題になるようです。(cn085j)

常識的に考えて、今の世界でたばこが「一番重要な問題」であるとはいえない。もっと深刻な問題はいくらかでも存在する。せめて「一番重要な問題のひとつ」とするなど、過剰になり過ぎない書き方にすべきである。

[8] (たばこを吸うと有毒な煙が出る、という話に続けて) ある日、私たちの空は青くなくなつて、吸っている空気は不潔で木はなくなります。(cn085j)

たばこの煙によって空が青くなくなったり、木が枯れたり、ということは事実としてない。「たばこを禁止すべき」という主張をサポートする根拠を挙げようとして、つい事実と反する根拠を挙げてしまったものと推測される。

[9] 私は医学の知識などありませんが、(たばこは)健康を害するものです。(jp002j)

「医学の知識がない」のであれば、「たばこは健康を害する」と断言するのは矛盾。「健康を害する」ということを言いたいのであれば、前半部分は説得力を弱めるため不要。もし書くなら、「医学の特別な知識がない自分にとってさえも、たばこが健康を害するということは明らかだ」というような書き方にすべき。

このような「非妥当箇所」について、何らかのコメントまたは添削がなされているかどうかを調査した。それぞれの箇所について、「非妥当性」を解消するための手当てをおこなっている教師数は、豊富群 15 名、寡少群 18 名のうち以下の人数であった。

[7]: 豊富群 1 名 寡少群 3 名

[8]: 豊富群 3 名 寡少群 4 名

[9]: 豊富群 2 名 寡少群 5 名

全体としてみると、上記のような箇所についてコメントまたは添削をおこなう教師は 1

割から2割, という結果であった。

また, 作文教育歴の長さ, フィードバック時に内容の妥当性に注目しているかということの間に, 明確な関係を見出すこともできないようである。そもそもコメントまたは添削をおこなった教師の数が少ないために明確なことは言えないのであるが, 作文教育経験が長くなったからといって, 内容の妥当性について配慮するようになるわけではなさそうである。

5. 結論

5.1. 調査で明らかになったこと

今回の調査から, 以下のことが明らかになった。

- 1) 作文教育歴豊富群の教師と寡少群の教師とに, 学習者作文に対する添削・コメント執筆を依頼したところ, 豊富群の教師のほうが寡少群の教師より, 具体性に富んだ内容のコメントを書く傾向があり, 作文教育経験の長さ, コメントの具体性とは相関があることが分かった。特に「指示」をおこなう際, 豊富群と寡少群との間でコメントの具体性に差が出ることが分かった。
- 2) 「否定的評価」をおこなう際には, 豊富群・寡少群ともに, 具体性のないコメントはあまり書かないことが分かった。「否定的評価」は学習者に対して心理的ストレスを与えるものであるため, 教師は可能な限り具体的なコメントを書き, ストレスを少しでも和らげようとしているものと考えられる。
- 3) 今回の調査においては, 豊富群のほうが寡少群よりフィードバックに多くの時間をかけていた。コメントを書くよう暗示的に求められるような状況においては, 豊富群の教師のほうがより深く, 分析に作文を読もうとしているものと推測されるが, このことについては今後の検証が必要である。
- 4) 文章の内容が妥当性を欠いているような箇所について, 何らかの手当てをおこなう教師は全体として1,2割であった。そして, 妥当性の欠如に対する修正・コメントの有無と教育経験との間に関係は見出せなかった。

5.2. 提言

これらの結果を踏まえ, 筆者は作文指導におけるフィードバックに対し, 以下のような提言をおこないたいと考える。

***フィードバックにおいては「分析的な視点」を持つことが重要**

今回の調査では、「否定的評価」をおこなう際には、教育経験の多寡を問わず、教師は具体的なコメントを書こうとしていることが分かった。その理由としては、「コメントを受ける学習者への配慮」ということが考えられるだろう。

こうした配慮は、「肯定的評価」や、「指示」をおこなう際にも必要なものといえるのではないだろうか。

「肯定的評価」に分類されたコメントの中には、単に「よく書けています」というだけの「判で押したような(rubber-stamped)」ものがしばしば見受けられた。もちろんこの種のコメントは、学習者の努力を認め、encourage するために書かれたものであろうと思われる。しかし、どういうところが、どういう理由でよかったのか、ということの詳細に書くほうが、学習者を encourage する効果ははるかに高いものといえるだろう。同様に「指示」をおこなう際にも、漠然と注意点を指し示すのではなく、具体的にどういう箇所を、どのように注意すればいいのか、明示的に指定したほうが、教育効果ははるかに高まるものと考えられる。

そのようなコメントを返していくためには、単に「ここがよい、ここがよくない」という直感的な判断をおこなうだけでなく、「自分はなぜこの部分がいいと思ったのか、なぜこの箇所がよくないと思ったのか」ということを、「分析的に考える習慣を持つ」ということが重要になるだろう。

***書き手の主張そのものは批判すべきでない。しかし「主張の述べ方」については、もっとクリティカルにコメントしていくべきではないか**

作文教育に携わる教師に対し、「どのような文章がいい文章だと考えるか」という質問をすると、しばしば、「執筆者本人の言いたいことがきちんと書けているのがいい文章である」という回答が得られる。これはおそらく、「学習者本位」「書き手本位」という基本思想に基づいているものと思われる。

しかしながら文章というものは、書き手だけでなく、読み手の存在があって初めて成立するものであろう。「本人の言いたいことが表現できる」というのは、実は文章の最低条件にしか過ぎない。その条件がクリアできたならば、次には「本人の言いたいことが、反論を受けることなく読み手に届く」ためにはどうすればいいのか、ということを考えなければならない。

そのためには文章の中に、事実と反する記述や、論理的な矛盾などが残っていることは非常に不利となる。そういう箇所に対して教師は、「執筆者本人がそう言いたかったのであるから」と見逃すのではなく、もっとクリティカルにコメントしていくべきではないかと考える。

もちろん、この種の問題点をすべてコメントとして伝えなければならない、というわけではない。学習者の言語レベルや学習目的によっては、そこまで伝える必要がないという

こともあるだろうし、学習者自身に気づいてもらうため「あえて言わない」という選択肢もあってよい。

しかしこの種の問題点に対し、教師が「確実に気づいている」ということは非常に大切であると思う。気づいていてあえて指摘しないことと、気づいていないこととの間には極めて大きな違いがあるだろう。

こうした「気づき」は、「経験」のみによってはなかなか獲得しにくいものであるのかもしれない。こうした「気づき」を得るために必要なものは、第一に前述の「分析的視点」を持つことであり、第二には、1.5節で述べたように、「書き手本位の指導法」を「書き手に読み手の存在を意識させる指導法」へと深化させる、という意識を持つことであると筆者は考える。

引用文献

- 池田玲子(2002)「第二言語教育でのピア・レスポンス研究：ESLから日本語教育に向けて」『第二言語習得・教育の研究最前線－明日の日本語教育への道しるべー』289-310, お茶の水女子大学日本言語文化学会
- 石橋玲子(2001)「産出作文に対する教師のフィードバックー日本語学習者の認識と対応からー」『拓殖大学日本語紀要』11, 89-98.
- 石橋玲子(2002)「日本語学習者の産出作文に対する教師の修正及び非修正行動」『言語文化と日本語教育』23, 1-12, お茶の水女子大学日本言語文化学会
- 上原久美子(1997)「作文の「記述式フィードバックコード」の必要性についてー日本語教師への実態調査からー」『平成9年度日本語教育学会春季大会予稿集』183-188.
- 内田伸子(1990)『子どもの文章 書くこと考えること』東京大学出版会
- 岡崎眸・岡崎敏雄(2001)『日本語教育における学習の分析とデザイン 言語習得過程の視点から見た日本語教育』凡人社
- 小宮千鶴子(1991)「推敲による作文指導の可能性ー学習者の能力を生かした訂正ー」『日本語教育』75号, 124-134.
- Faigley, L. & S. Witte(1981) "Analyzing Revision", *College Composition and Communication* 32, 400-414.
- Fathman A.K.& E.Whalley(1990) "Teacher response to student writing: Focus on form versus content", In B. Kroll (Ed.), *Second language writing: research insights for the classroom*, 178-190, Cambridge University Press.
- Ferris D.R.(1995) "Student reactions to teacher response in multiple-draft composition classrooms", *TESOL Quarterly* 29(1), 33-53.
- Leki I.(1990) "Coaching from the margins: issues in written response", In B. Kroll (Ed.), *Second language writing: research insights for the classroom*, 57-68,

Cambridge University Press.

Pica, T. (1986) "An interactional approach to the teaching of writing", *English Teaching Forum* 24(3), 6-10.

Zamel, V. (1985) "Responding to student writing", *TESOL Quarterly* 19, 79-101.

謝辞

今回の調査に添削者として協力して下さった先生方、またその先生方を紹介して下さった日本語教育機関代表者の方々に、篤くお礼を申し上げます。

***企画・編集**

宇佐美 洋（国立国語研究所日本語教育部門第1領域 主任研究員）

鎌水 兼貴（国立国語研究所日本語教育部門第1領域 研究補佐員）

***執筆**

井上 優（国立国語研究所日本語教育部門第1領域 領域長）

宇佐美 洋（国立国語研究所日本語教育部門第1領域 主任研究員）

成田 高宏（国立国語研究所日本語教育部門第1領域 研究補佐員）

鎌水 兼貴（国立国語研究所日本語教育部門第1領域 研究補佐員）

***刊行物検討委員会**

朝日 祥之（国立国語研究所情報資料部門第2領域 研究員）

井上 優（国立国語研究所日本語教育部門第1領域 領域長）

宇佐美 洋（国立国語研究所日本語教育部門第1領域 主任研究員）

丸山 岳彦（国立国語研究所研究開発部門第1領域 研究員）

柳澤 好昭（委員長：国立国語研究所日本語教育部門 部門長）

作文対訳データベースの多様な利用のために

「日本語教育のための言語資源及び学習内容に関する調査研究」
報告書

2006(平成18)年3月30日発行

独立行政法人 国立国語研究所

〒190-8561 東京都立川市緑町 3591-2

TEL: 042-540-4300(代) FAX: 042-540-4333(代)

URL <http://www.kokken.go.jp/>