

日本語学習者コーパスのための 誤用タグの構築について

大 山 浩 美

要 旨

近年、第二言語学習者が産出した作文や会話文などを電子化し、そのデータを学習者コーパスとして第二言語教育に活用しようという動きが活発になっている。学習者コーパスは、教材や教授法開発、誤用分析、添削規準構築など様々な形で教育に活用できる。しかし、その貴重なデータを最大限に活用するためには学習者の誤用などにタグを付与し、様々な情報を取り出せるようにしなければならない。

本稿では関連する学習者コーパスで使用されている誤用タグについて調査し、現存するタグの利点を考察し、汎用性があるような誤用タグを試作した。さらに、国立国語研究所により収集された「作文対訳 DB」に誤用タグを付与する作業を行い、「作文対訳 DB」における助詞の誤用について頻度分析を行った。その助詞誤用頻度分析の結果、学習者コーパスにおいて興味深い現象が見られることがわかった。

【キーワード】 第二言語教育、学習者コーパス、誤用タグ、助詞の誤用頻度分析、助詞の脱落

1. はじめに

第二言語学習者の産出物を集めた学習者コーパスは、コンピュータのデータ蓄積技術の発展、検索技術の発展などにより、その種類や量を増している。学習者コーパスの利点は、学習者が何を困難に感じるかを見直すことにより教師の学生へのフィードバック、さらに教材研究、教授法の見直しなどにも役に立つことである。また、作文支援システムを開発する際にも学習者の誤りやすい点が予めわかっているならばシステム作りに大変役立つと思われる。さらに、学習者の産出物を添削する際に問題となるのは添削者間の揺れであるが、このような添削済みの学習者コーパスを用いて様々な添削者がどのように添削を施すのかを調査することができれば添削者間の一致した規準を構築することができるだろう。しかし、ただ電子化されたデータだけでは不十分であり、最終的に誤用に関する情報を引き出すためには誤用タグを構築し、付与するといったデータ整備が必須となってくる。

本稿では、国立国語研究所で収集された「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（以下、「作文対訳 DB」）⁽¹⁾」を利用し、その作文の助詞の誤りに誤用タグを付与し、助詞の誤用頻度分析を行った。本稿では、その「作文対訳 DB」へ誤用タグを付与するためのタグ設計について、さらに助詞誤用タグの頻度分析の結果について述べる。

2. 「作文対訳 DB」について

国立国語研究所で収集された「作文対訳 DB」には、様々な国籍の日本語学習者による日本語作文を中心に、その作文に対応する学習者の母語による対訳文、日本語教師による添削文、作文執筆者、添削者の日本語教育の経歴などの様々な情報が含まれている。1999年から2000年にかけての第一版では、アジア10カ国の日本語学習者の作文が集められている⁽²⁾。今回、この第一版（アジア人学生編）の作文の中で、日本語教師による添削が施されている313編について、誤用タグを付与し、助詞における誤用頻度分析を試みた。

添削が施されているファイルは、中国人日本語学習者19編、カンボジア人日本語学習者10編、韓国人日本語学習者91編、マレーシア人日本語学習者70編、モンゴル人日本語学習者10編、シンガポール人日本語学習者8編、タイ人日本語学習者59編、ベトナム人日本語学習者46編の計313編である。このファイルは、全体で文字数約20万語（191,994語）、一ファイルあたりの文字数は約600語（589.1語）、一人の学習者が400字詰め原稿用紙で約一枚半程度書いた作文だと考えてよい（表1参照）。その添削が施されている作文は、学習者の誤りの傾向を俯瞰できるだけでなく添削者が学習者の誤りに対してどのような修正を施すかを教師が考察できる貴重な資料である。次の節では現存する誤用タグについて考察する。

国名	ファイル数	文字数	1ファイル平均文字数
中国 (cn)	19	11,625	612
カンボジア (kh)	10	6,552	655
韓国 (kr)	91	57,703	634
マレーシア (ml)	70	45,355	648
モンゴル (mn)	10	5,385	539
シンガポール (sg)	8	3,670	459
タイ (th)	59	36,643	621
ベトナム (vn)	46	25,061	545
合計	313	191,994	589.1

表1「作文対訳 DB」（添削されているファイルのみ）について

3. 誤用タグ構築に関する先行研究

現存する誤用タグは、それが使用される学習者コーパスの研究目的によって異なる場合が多い。そのため、学習者コーパスの数だけ異なる誤用タグセットが存在する。統一化された誤用タグを構築することは非常に難しいことであるが、一つの規準を持つことは意義がある。誤用タグ構築において三つの主な特徴がある。一つ目は、言語学的見地（形態論、統語論、音韻論など）によってタグセットが構築されているかということである。二つ目は、目標文からの逸脱様態によってタグが構築されているかである。目標文からの逸脱様態とは、脱落 (omission) つまり本来あるべき要素が欠けている状態であるか、追加 (addition)、なくてもよい要素が付加してあるか、誤選択 (mischoice)、本来選択されるべき単語ではなく誤った要素が使用されているかのことである。三つ目は、修正後の正用語に関する情報を持つかどうかである。まず、英語学習者コーパスの誤用タグの特徴についてまとめたものを表2に表す。

ICLE (International Corpus of Learner English) コーパスは、University of Louvain-la Neuve で収集された14の学習者の母語からなる500から1000語程度の大学生の論文コーパスであり、コーパスサイズは、約200万語である⁽³⁾。このコーパスでは、2000万語中150万語に誤用タグが付与されている。このコーパスの誤用タグの例をあげる。

● 例【ICLE】

● Barons that (GVT) lived \$ had lived \$ in those (FS) castel \$ castels\$.

このタグセットでは、まず、メインカテゴリーとして誤用カテゴリー（例の“G”）、そのサブカテゴリーとして品詞（例の“V”）と誤用種類（例の“T”）とに分けられ、誤用箇所後に正用語を“\$”でくくっている。“G”は、文法的誤用(grammatical)の意味である。“V”は動詞(verb)で、“T”は時制(tense)の意味である。ここでは、“Barons that had lived in those castels.”と書くべきところを“Barons that lived in those castel.”と書いたという誤りである。タグは、オープン形式で閉じていない(Dagneaux et al., 1998)。

JEFLL コーパス (Japanese EFL Learner) は、中学・高校の日本人英語学習者約1万人以上の自由英作文のコーパスで、サイズは約70万語である⁽⁴⁾。このコーパスでは、学習者の正用と誤用を際立たせるため調査したい文法項目においての正用と誤用の両方にタグを振っている。例をあげる。

● 例【JEFLL】

● (正用) I have hardly had < ART > a < /ART > bad dream.

● (誤用) From < ER_ART > the < /ER_ART > cliff.

この例では、助詞の正用と誤用を調べるためのタグを振ってある。< ART >というタグは助詞の正用、< ER_ART >は助詞の誤用を意味している。タグは、クローズド形式で閉じるための括弧がある。

Nicholls (2003) による CLC (Cambridge Learner Corpus)⁽⁵⁾ は、75 の母語数、2 千万語から成る。うち、5 百万語分にエラータグが付与されている。下の例で、“U” は、“Unnecessary”、この“they” は不必要、それから、A は代名詞という意味である。二番目の例では、タグ中の二番目の“You hardly ever meet people ...” は誤りで、正しくは“Hardly ever do you meet people ...”だと添削してある。そして、タグは Argument Structure (項構造:AS) の誤りとなっている。タグは、クローズド形式である。

● 例【CLC】

- Lawyers, doctors, etc, < #UA > they < /#UA > hardly earn \$50,000 a year.
- < #AS > Hardly ever do you meet | You hardly ever meet < /#AS > people...

	ICLE	JEFLL	CLC	Free Text	NICT JLE	MELD
年	1998	2000	2003	2003	2004	2005
代表者	Dagneaux	Tono	Nicholls	Granger	Izumi et al.	Fitzpatrick & Seegmiller
言語学的分類	○	○	○	○	○	-
目標文からの逸脱様態	○	-	○	○	-	-
正用語の表示	○	○	○	○	○	○
コーパスサイズ	200 万	70 万	2000 万	45 万	200 万	約 10 万

表2 英語学習者コーパスの誤用タグについて

Granger (2003) による Free Text Corpus では、中級から上級のフランス人英語学習者の作文が収集されている。その 45 万語のうちの 30 万語に誤用タグが付与されている。タグは、まず誤用の領域に分けられ、それから誤用の種類によって再分類される。さらに、単語の品詞情報、訂正候補も付与されている。下の例で、“F” は、形式上の誤用 (Formal) を意味する。形式上の誤用とは、表記の問題、大文字小文字の誤り、スペルミスなどをさす。“DIA” は付加記号 (diacritic) を表している。付加記号と

は、フランス語に見られる綴り字記号のことである。“NOM”は、品詞情報で名詞を表す。この例では、付加記号のない“secret”は誤りで、正しくは“secrét”であると記されている。

● 例【Free Text】

● ...qui ne sait pas garder le moindre < F >< DIA >< NOM > #secret\$ secrét
< /F >< /DIA >< /NOM >.

NICT (情報通信研究機構) の JLE (Japanese Learner English) コーパスは、日本を母語とする英語学習者 1,281 名分の英語インタビューテスト (1 名につき 15 分間) の様子を書き起こした発話文のテキストデータであり、サイズは約 2 百万語である (和泉他, 2004)。その中で 167 名の発話に誤用タグが付与されている。品詞情報、文法的誤り、語彙的誤り情報に関するタグを付与しており、下の例において、タグ内の “n_num” は、“n” で名詞 (noun) を示し、“num” で単数複数の間違ひがあることを示している。「crr=“teams”」は、修正候補を提示している。

● 例【NICT JLE】

● I belong to two baseball < n_num crr=“teams” > team < /n_num >.

MELD (Montclair Electronic Language Database) コーパスは他のタグと大きく異なり、誤用の要素と正用の要素をタグコードとしてそのまま使用している。そうすることにより、添削者同士の添削の相違、誤分類を避けるための工夫をしている (Fitzpatrick & Seegmiller, 2004)。下の例で、「School “is”」は誤りであり、修正候補として “are” が選ばれている。“0” は、脱落していたという意味である。本稿で試案した誤用タグセットは、前述の JLE コーパスのタグと MELD タグを特に参考にした。

● 例【MELD】

● School systems {is/are} since children {0/are} usually inspired becoming {a/0} good citizens.

次に日本語学習者コーパスの誤用分類の特徴について述べる。誤用タグ構築法としては、1) 特定の文法記述に基づきタグを構築している場合、2) 実際の誤用分析で抽出された誤用タイプに基づいてタグを構築する場合、とに分けられるが、英語学習者コーパスではほとんどが後者の方法をとっている。日本語学習者コーパスでは、前者の特定の文法記述に従う方法もある。この方法は、学習者の第一言語が多岐に渡る場合には有効である (清水他, 2004)。

大曾他 (1997) による名古屋大学の誤用コーパスは、上記の 1 番の方法をとり益岡・田窪文法 (益岡・田窪, 1992) の文法記述に基づいている。例を挙げる⁽⁶⁾。

	大曾他	市川	清水他
年	1997	1997	2004
言語学的分類	○	○	○
目標文からの逸脱様態	-	○	○
正用語の表示	○	○	?
タグ構築手法	1	2	2

表3 日本語学習者コーパス用の誤用タグについて

例：【名古屋大学誤用コーパス】

- *GAK: <一般的に添加物は自然な物じゃなくて、<また子供達は小さいから> [*1] > [*2]、<添加物を入っている> [*3] 物をなるべく避ける。
- %err: [1 また子供達は小さいから = まだ子供達は小さいから]; [2 一般的に添加物は自然な物じゃなくて、まだ子供達は小さいから = 一般的に添加物は自然な物じゃないし、まだ子供達は小さいから]; [3 添加物を入っている = 添加物が入っている];
- %als: [1 濁音が清音の表記になっている。「た」]; [2 テ形による接続の間違い。接続助詞「し」を使ったほうがいい。]; [3 格助詞「を」と「が」の間違い。「入っている」の主体];

このタグは、本来 JCHAT と呼ばれる幼児の発話言語収集のために作られたタグを拡張したものである。誤用がある部分を<>でくり、[*] を付与している。本文の下の行の「% err」は、訂正ティアとよばれ、そこに誤用と正用が記される。「% als」は、誤り分析を記すティアである。訂正ティアと分析ティアには、インデックス番号がつけられ、本文中のどの箇所に対するものなのかがわかるようになっている。この他にも、「% com」のティアが存在し、添削者がコメントを書き入れることができる。

市川(1997, 2000)による「日本語誤用例文小辞典」の誤用分類は、言語学的見地から、モダリティ、テンス、アスペクトなどの8つの主要分類に分けられ、その次に86の項目に分けられる⁽⁷⁾。それから、脱落、付加、誤形成、混同、位置、その他の6種類に分けられる(市川, 2001)。これは、先述の「目標文からの逸脱様態」とほぼ同義である。

- 例：【市川誤用例文小辞典】
- 助詞
- 付加

- 兄弟は8人が(→φ)いて、シアトルやシカゴに住んでいる

上記の例は、助詞の「が」が不要であるという意味である。市川では、誤用文を上記の分類項目に基づいて分類しているが、タグはふられていない。

清水他 (2004) や Shimizu et al. (2005) では、大連理工大学独自の中国人学習者による日本語作文コーパス構築プロジェクトの中で、日本人中国語学習者の作文を添削した後で独自の誤用タグを設計、付与している。この誤用タグセットでは、形式的分類とタイプ別分類についての情報が付与されている。ここでの形式的分類というのは、上記の目標文からの逸脱様態とはほぼ同義である⁽⁸⁾。タイプ別分類とは、言語学的分類に近く、時制、アスペクト、助詞の過剰付加、「は」と「が」の混同などについてのタグである。タグ付与に関する実際の例は論文には明示されていない。上記の各誤用タグについて考察し、今回国研の「作文対訳 DB」への誤用タグ構築、付与という作業を行ってみた⁽⁹⁾。

4. 「作文対訳 DB」(添削済みのファイル) に付与した誤用タグについて

Granger (2003) は、誤用タグを作成する際に考慮すべき点を次のようにまとめている。それらは、タグに一貫性がある、誤用分析ができるような情報を十分に持っている、柔軟性がある、再利用可能という点である。一貫性、再利用可能かという点から言うと、前節で述べたような「目標文からの逸脱様態」(脱落、付加など)に関する誤用タグ付与の方が添削者同士による意見の不一致が少ない。しかし、タグの情報性の面で見るとそれだけでは学習者になぜ誤りを犯したのかの理由を説明できず、フィードバックとして与える情報として不十分である。James (1998) では、言語学的分類、目標文からの逸脱様態、それら二つの特徴を生かしたタグセットの構築を推奨している。

それらの事項を念頭に誤用タグを試作し、「作文対訳 DB」の中で添削が施してある 313 名の作文中の助詞の誤用部分にタグを付与し、様々な情報を補完した。図 1 は、中国人学習者による「たばこ」についての作文にタグを付与した一例である。作文全体を < corpus > というタグでくくり、始めに作文を書いた学習者に関する情報タグを振っている。情報タグは、< id > (通し番号)、< name > (名前)、< nationality > (国籍)、< gender > (性別)、< m-lang > (母国語)、< year > (作文を作成した年) の 6 種類である。作文本体は、< text > というタグでくくっている。さらに、段落には、< p > タグ (paragraph の意)、文ごとに < s > タグ (sentence の意) を振っている。そして、誤用部分には、< goyo > タグを施し、そのタグ内に誤用タイプを示す type 属性、正用語を示す crr 属性を付加している。

```

* <?xml version="1.0" encoding="UTF-8"?>
* <!DOCTYPE corpus SYSTEM ". dtd">
* <corpus>
* <id>en001j.txt</id>
* <name></name>
* <gender></gender>
* <nationality/>
* <lang/>
* <year/>
* <text>
* <p><s>私は会社やレストランなど公共の場所では喫煙すべきと思っている。</s>
*
* <s>誰<goyo cr="か"></goyo>に、誰<goyo type="p/ni/de" cr="で"></goyo>も
* たばこを吸う権利がある。</s>
*
* <s>しかし、公共の場所では、たばこを吸うと、まわりの人に迷惑をかけるから、やめるべきと
* 思う。</s>
* <s>また、<goyo cr="たばこは、たばこを直接すわない人にも害を及ぼす。"></goyo>というい
* ろんな調査によると<goyo cr="近くでたばこをすわれることで、直接たばこをすうより倍以上の
* 悪影響を受けることがわかっている。"></goyo>たばこを吸わない人に対する悪影響は吸う人の倍以上
* だと分かった。</goyo></s>
* <s>従って、公共の場所ではたばこを吸う行為は公害と言っても過言ではない。</s>
* </p></text></corpus>

```

図1 誤用タグつきファイル例

誤用タイプタグの構成は、まず助詞を表す p (particle) が第一階層に入り、それから第二階層には、誤って使用された助詞が入り、第三階層には修正された助詞が入る。図1において、四角で囲ってある部分に着目して説明する。

- < s > ... 誰< goyo type=" p/ni/de " cr=" で " >に< /goyo >もたばこを吸う権利がある。 < /s >

上記の例で、< goyo >タグで囲まれた助詞の「に」が誤りである。添削者による修正⁽¹⁰⁾は、「に」ではなく「で」と施されていたので、誤用タイプが「p/ni/de」となる。正用語として「で」が入る。このタグの特徴として、JLE コーパスのタグのようにタイプなどの属性をタグ内に入れられるようにした。さらに、MELD コーパスのタグのように修正前と修正後についての情報をタグにそのまま生かせるようにした。言語学的理由付けは誤用タグには表現せず、できるだけ客観的で添削者による揺れがでないようなタグになるよう心がけた。しかし、上記のタグだけでは言語学的な理由付け(なぜこういう誤りを犯したのか)がなく、学習者コーパスを使った誤用分析に有益な知見を残せない。そこで、言語学的な理由付けは後付にし、理由付け表とタグを後で関連付けできるようにした。そうすることにより、各研究者によってタグではなく、理由付け表のみを変更することが可能となる。表4に、理由付け表の一例を載せる。

タグに一貫性があり、添削者誰にとっても同様の判断ができるようなタグセットであること、誤用分析ができるような情報を十分に持っていること、ある特定の研究目的だけでなく様々な研究者の目的に応じて使用できるような柔軟性、再利用可能であるということなどに配慮し、誤用タグを試案した。

は	タグ	例	理由
は→φ	p/wa/ad	私は、日本に来たばかりの時のある日[は]→ある日	日付表現の後の「は」
は→が	p/wa/ga	私は、日本にやってきた目的[は]→私が日本に	「は」と「が」の混同
に→で	p/ni/de	日本社会[に]生活するのは問題がないと言われました→日本社会で	活動をする場所の「で」
		できるだけワープロ[に]の操作を身につけたいと思います→ワープロでの	道具の「で」

表4 言語学的理由付け表

5. 「作文対訳 DB」の助詞の誤用について

「作文対訳 DB」の助詞の誤り全てに誤用タグを付与し、それらを集計した。助詞全体の誤りタグの合計が 3037 で、異なりタグの合計が 791 であった(表 5)。助詞の誤りの中で最も頻度が多かった誤用タイプについて国別にグラフにした(図 2)。「作文対訳 DB」の助詞の誤り全体で、om (脱落 omission) タイプの誤りが圧倒的に多く、次に、ga (が) に関する誤り、wa (は) に関する誤り、wo (を)、no (の)、ni (に)、de (で)、mo (も)、toka (とか)、to (と) に続く(図 2)。

さらに、脱落タイプの誤りの中で最も多いのは、no (の) を書き損ねているという誤りである(タグは、“p/om/no”、図 3 参照)。no (の) の脱落の誤りは、助数詞を使用する場合に「の」がなかったり、漢語同士の間に「の」がなかったりする場合が多かった。no (の) の脱落の誤りの中で助数詞に関する誤りの例をあげる。

- cn069j.txt
- < s >二匹< goyo type= “p/om/no” crr= “の” >< /goyo >黄牛と一緒にお互い< goyo type= “p/ni/wo” crr= “を” >に< /goyo >頼り合いました。< /s >

国名	助詞タグ延べ数	助詞タグ異なり数
中国 (cn)	179	64
カンボジア (kh)	92	53
韓国 (kr)	722	171
マレーシア (ml)	905	182
モンゴル (mn)	80	39
シンガポール (sg)	67	35
タイ (th)	579	112
ベトナム (vn)	413	135
合計	3037	791

表5 「作文対訳 DB」助詞誤用について

これは、中国人学習者の書いた作文であるが、助数詞の前後で「の」の助詞を書かないという誤りである。添削者は、「二匹黄牛」ではなく「二匹の黄牛」の方が正しいと添削している。「作文対訳 DB」においてこのタイプの誤りが非常に多いことがわかった。それから、次に多いのは「は」の脱落である。助詞の「は」は、助詞の中でも非常に習得が難しい項目であり、誤りが多かったのもうなずける。

● krl85j.txt

● < s >他人に< goyo type="" crr="被害を与えている" >がいます< /goyo >
 と< goyo type="p/om/wa" crr="は" >< /goyo >おもいません。< /s >
 今回多く見受けられたのが、「は」を追加した「～とは思いません」、「～とは言いませぬ」への修正である。「～とは思いません」という表現でも十分文法的であり間違いではないが、添削者による修正が施されている。この「は」の使い方は、かなり高度な表現であると思われる。

しかし、添削者がこの表現に遭遇した時に「～とは言いませぬ」の方が自分の意見を述べる上で主張の強さを感じられ、修正したくなる気持ちは理解できる。こういった誤り、修正の仕方は、学習者コーパスの全体を俯瞰して初めて注目できる誤用だと思われ、学習者コーパス構築と誤用タグの整備の必要性を強く感じる。

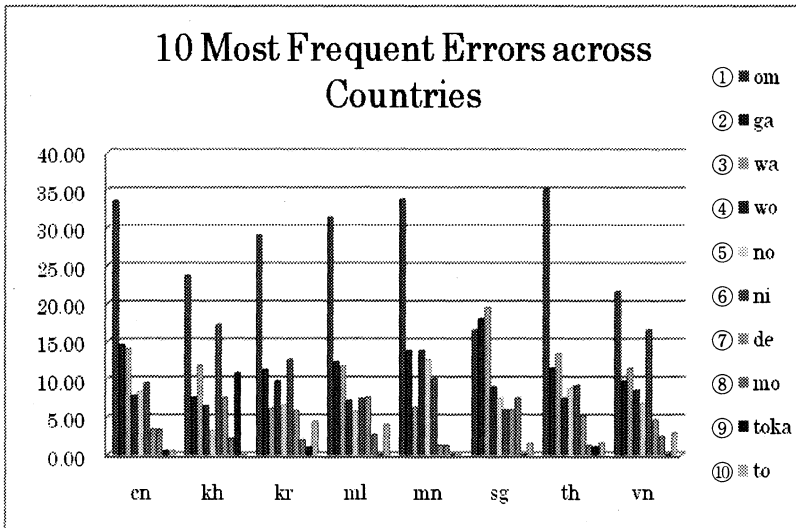


図2 助詞誤用の上位10件

※グラフ内、左から①～⑩の順に表記。

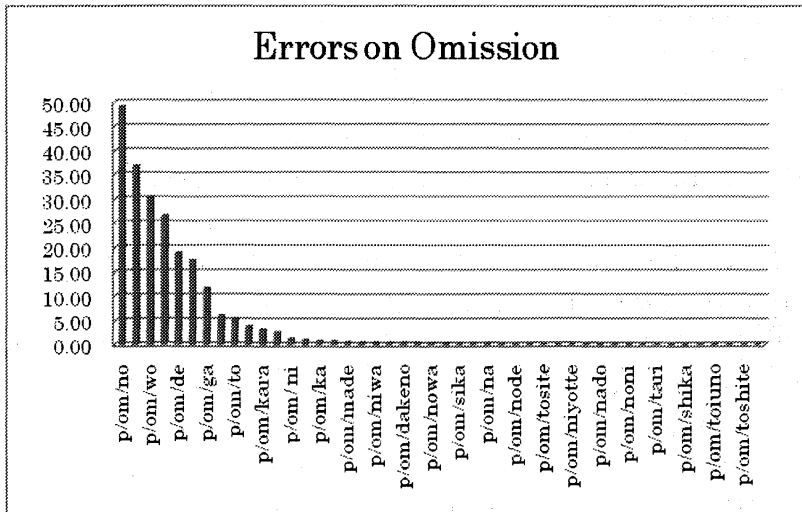


図3 脱落タイプの助詞誤用頻度

6. まとめと今後の課題

国研により収集された「作文対訳 DB」に誤用タグを付与する作業を行った。その作業を行う際に、関連する学習者コーパスとそれぞれのコーパスで使用されている誤用タグについても調査した。現存する誤用タグを元にその長所短所を考慮し、誤用タグに必要な特徴（一貫性があることなど）を持つように工夫し、タグ設計を行い、タグ付与作業を行った。その誤用タグを利用し、「作文対訳 DB」における助詞の誤用頻度分析を行った。その結果、助詞を脱落させる誤りが最も多いことがわかった。さらに、助詞「の」や「は」に関する習得の難しさがわかった。今回は助詞誤用のみの分析を行ったが、現在その他の誤りについても誤用タグ付与作業を行っている。今後、助詞以外の誤用についても誤用分析を試みるつもりである。

注

- (1) <http://www2.kokken.go.jp/eag/wiki.cgi?page=taiyakuDBn%2Ftop>
- (2) 2007年3月公開分は10カ国分（アメリカ、ブラジルなど）が追加され20カ国分となっている。詳しくは <http://www2.kokken.go.jp/eag/wiki.cgi?page=taiyakuDBn%2Ftop>
- (3) <http://www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/Cecl-Projects/Icle/icle.htm>

- (4) <http://jefll.corpuscobo.net/>
- (5) <http://www.cambridge.org/elt/corpus/clc.htm>
- (6) この例は、<http://cookie.nagoya-u.ac.jp/pub/goyooman.html> より採用したが、現在はアクセス不可である。<http://lang.nagoya-u.ac.jp/~sugiura/CHILDES/goyooCHILDESformat.html> に移行したと思われるが、説明に少し変更がある。
- (7) ムードの下位項目として 20 項目、テンス・アスペクト 10 項目、自動詞・他動詞・ヴォイス 5 項目、やりもらい 3 項目、取立て助詞 3 項目、格助詞・連体助詞・複合助詞 10 項目、連用修飾・連体修飾 2 項目、従属節 33 項目
- (8) 「付加 (Addition)」、「脱落 (Omission)」、「混同 (Confusion)」、「位置 (Misordering)」、「誤形成 (Misformation)」、「転移 (Transfer)」の 6 種による分類である
- (9) 緒方 (2000); 中山・佐野 (2000) では、「作文対訳 DB」の添削済のファイルに作文添削情報を電子化して付与するタスクが試みられ、目標文からの逸脱様態による誤用分類と類似した、Insert (挿入)、Replace (訂正)、Delete (削除)、Separate (改行)、Join (追い込み - 二つの段落を一つにする)、Movefrom-Moveto (移動) 等のタグが提案されている。しかし、その後の進捗は見当たらなかった。
- (10) ここでの修正は、国立国語研究所で作成された「作文対訳 DB」の修正に全て基づいている。

参考文献

- (1) Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided Error Analysis. *System*, 26, 163-174.
- (2) Fitzpatrick, E., & Seegmiller, M. (2004). The montclair electronic language database project. In U. Connor, & T. Upton (Eds.), *Applied Corpus Linguistics. A Multidimensional Perspective.*: Rodopi. 223-237.
- (3) Granger, S. (2003). Error-tagged learner corpus and CALL: A promising synergy. *CALICO*, 20, 465-480.
- (4) James, C. (1998). *Errors in Language Learning and Use: Exploring Error Analysis.*: Longman.
- (5) Nicholls, D. (2003). The Cambridge Learner Corpus-error coding and analysis for lexicography and ELT. In A. et al. (Ed.), *Proceedings of the Corpus Linguistics 2003 Conference (CL2003)*. 572-581.
- (6) Shimizu, M., Du, F., & Dantsuji, M. (2005). A project to construct the Chinese

learners' parallel corpus of Japanese and develop Dutch corpus linguistics tools.
*Proceedings of the 6th International Symposium on Applied Linguistics and
Language Learning*, 585-595.

- (7) 益岡隆志・田窪行則 (1992) 『基礎日本語文法改訂版』 くろしお出版.
- (8) 市川保子 (1997) 『日本語誤用例文小辞典』 凡人社.
- (9) 市川保子 (2000) 『続・日本語誤用例文小辞典—接続詞・副詞—』 凡人社.
- (10) 市川保子 (2001) 「日本語の誤用研究」 国際交流基金日本語グループ (編) 『日本語教育通信』 第 40 版 国際交流基金 14-15.
- (11) 緒方広明 (2000) 「マルチメディア文書に対する共同添削支援システムの研究開発」
(<http://www.yano.is.tokushima-u.ac.jp/cocoaj/download/paper.pdf>)
- (12) 清水政明・宋 哲・孟 慶榮・杜 鳳剛・壇辻正剛 (2004) 「中国人日本語学習者
対訳作文コーパスの構築における誤りタグの構築について」 『日本語言文化教学与研
究国際学術研討会』.
- (13) 大曾美恵子・杉浦正利・市川保子・奥村 学・小森早江子・白井英俊・滝沢直宏・
外池俊幸 (1997) 「日本語学習者の作文コーパス：電子化による共有資源化」 『言語処
理学会第3回年次大会論文集』.
- (14) 中山智哉・佐野 洋 (2000) 「多言語作文コーパスデータベース - 設計と実装につい
て」 『第 8 回国立国語研究所国際シンポジウム「日本語とアジア諸言語との作文対訳
コーパス：対照言語学・日本語教育への応用』.
- (15) 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.