

## 添削・評価情報のタグ化とその集計手法

### －「評価意識」を量的に取り扱うために－

宇佐美洋・鎌水兼貴

#### 1. はじめに

国立国語研究所では、「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」(略称：作文対訳 DB)の一部門として、学習者作文による添削情報を収集してきている。収集した添削情報(第一次情報としては、一般的な赤ペンによる添削であることが多い)は、そのままスキャンして画像情報として保存しているほか、2002年以降は「XMLによる作文添削情報表示システム」により添削情報の電子化をおこなってきた(宇佐美2002)。

このシステムでは、「置換」「削除」「挿入」などという文字列操作をXMLタグで表し、修正後の文字列は属性値としてタグ内に埋め込まれる。そしてXML化された文書をブラウザで表示させることにより、手書き添削に近い形で添削情報が表示されるようになっている。

タグが挿入されたXMLファイルと、それをブラウザで表示した例(ver.1.0)を以下に挙げる。

```
<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">わる</rep>と健康<だけ<rep value="<br>が">は</rep>弱くな<rep value="るわけではありません">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></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/>
```

→人達 はよく「\*自分のこのむ?」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。けれど  
もたばこをすわる と健康\*だけが 弱くなりません 。お金もたくさん用いられま  
す 。それとともに次の世代も悪いこと を受けます。私の感じによって??が感じるのは 次の世  
代にいつもいい物を教えるのほうがいいです 。その人々は年上になって 自分の両親  
やほかの人達を例としてどしどし 「どんだん」の方が一般的 追いかけます。

われわれのタグセットは、ver.1の時代から、「明らかにおかしいから直すべきだ」という箇所だけでなく、「誤りとまではいえないが、直したほうがよくなる」という箇所、「意味がよく分からないので正確に直せない」という箇所も指摘できるようになっているところに特徴がある。上記XMLファイルの中では、<rep>タグ（置換）、<put>タグ（挿入）は「明らかにおかしいから直すべき」ということを、<better>タグは「誤りとまではいえないが、直したほうがよくなる」ということを、<doubt>タグは「意味がよく分からないので正確に直せない」ということを示している。

XMLによる添削情報の電子化方式はその後改定を続け、2007年3月現在、タグセットの最新版はver.3.0となっている（改定の経緯については宇佐美・鎌水 2006aを、最新バージョンのタグセットについては宇佐美・鎌水 2006bを参照）。

添削情報とは、添削者が学習者作文のどの部分を直すべきと考え、具体的にどう直せばよくなると考えたかを示すデータである。こうしたデータを大量に収集し、共有の研究データとして公開されることは、教師研究・評価観研究などに資するところはきわめて大きいと考えられる。しかしながら一般的な手書きの添削では、それを研究者間で共有することは極めて困難であったし、何より、「添削情報に対して検索をかけることができない」という大きな制約があった。添削情報をXMLにより電子化することは、単にデータの見ばえをよくするというのではなく、添削情報というものを機械可読的（検索可能）なデータ<sup>1</sup>とすることにほかならない。

しかしながら、手書きデータをそのままブラウザ上で表示できるようになったというだけでは研究データとしての価値は十分でない。添削情報がせつかく機械可読的なデータに変換できたのであれば、それを量的な分析にも使用できるような手法までをあわせて提案する必要があるだろう。

従来の「誤用タグが付与された学習者コーパス」(Cambridge Learner Corpus, NICT JLE Corpus など)では、誤用を認定し分類するための詳細なマニュアル<sup>2</sup>が作成され、評定者は十分なトレーニングを受けた上でタグ付けに当たっているものと考えられる。このように統一した基準によって誤用認定がおこなわれていれば、例えば学習レベルが上がっていくにつれて出現しやすい誤用の種類がどう変わっていくか、学習者の母語の違いによって起こりやすい誤用の種類がどう違うか、などの研究が可能となろう。

しかし、こうしたマニュアルによって誤用と認定され、かつ明確に分類できる表現というのは、「だれが見ても明らかな誤用」であり、評定者によって『『おかしさの理由』が明確に特定できた誤用』ということになるのではないだろうか。そして、コミュニケーション上重大な誤用というのは、「なぜおかしいのか分からないが、なんとなくおかしい誤用」であり、「ある人は許容するが、また別の人は絶対に許せないと思ってしまうような誤用」

<sup>1</sup> 鎌水(2006)は添削後の文字列を検索対象とした研究をおこなっている。従来の誤用タグによる検索は、表現として表れていない誤用については限界があった。しかし添削後の文字列からの検索を行うことで「こう書くべきであったが書かなかった」というような例を容易に調べることが可能である。

<sup>2</sup> NICT JLE Corpusの「エラータグマニュアル」は、和泉・内元・井佐原(2004)で公開されている。

なのではないだろうか (宇佐美 2006)。

そこでわれわれは、統一的な誤用認定の基準を提示するのではなく、あくまでも個人個人の言語感覚にしたがって添削・評定作業を行なってもらうことにすると同時に、ひとつの作文に対し、必ず複数の人々に添削・評定作業をお願いすることとした。もちろん添削・評定の結果は、人によって大きく異なるだろう。そうした「評定のずれ」にこそ積極的な意味があるという考えに基づき、われわれは、ひとつの作文に対する複数の添削・評定情報を一覧・集計する手法を開発し、その手法を広く公開していくことを考えた。

## 2. 本科研で収集した添削・評定データ

### 2.1. 日本語教師 (またはそれに準ずる人々) のデータ

本科研の目的のひとつとして、「同一の学習者作文であっても、それに対する反応は人によって大きくばらつく」ということを示し、そのばらつきの背後にある要因について考察する、ということがあった。この目的のため、2005年から2006年にかけて、以下のような添削データを収集した。

- 1) 「対訳作文 DB」からランダムに選択した 10 編の作文に対し、33 名の日本語教師に個別に依頼した添削
- 2) 1)と同じ作文に対し、「日本語教師を目指して勉強しているが、教育経験はない人々」10名に依頼した添削

上記 43 名の方々には、添削とは別に、作文中で「日本語としておかしい、不自然だ」と思ったところに下線を引いてもらい、その不自然さの「重篤度」を 3 段階<sup>3</sup>で判定してもらおうという「評定データ」も収集し、一定の方式に従って電子化した。添削データだけでなく、評定データも収集することにした理由は、はしがき(p.3)にも記載したとおり、「添削がおこなわれていない箇所に対して添削者は、『おかしいとは思わなかった』のか、『おかしいとは思ったが、学習レベルその他の要因を考慮して許容できると考えた』のか、あるいは『おかしいとは思ったがどう直していいか分からなかったため放置した』のかは、添削結果を見ただけでは判断できない」と考えたからであった。

以下に、「評定データ」の実例 (手書きオリジナルデータ) と、それを電子化した例を示す。

---

<sup>3</sup> 重篤度の各段階については以下のように説明した。3: 非常におかしく、必ず修正すべきである, 2: 3 と 1 の中間程度, 1: 間違っているとはいえないが、ややおかしい。

たばこを吸うのは権利ですか  
1  
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  
2  
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  
3  
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。  
2

たばこを吸う [1: のは] 権利ですか  
今、たばこのことが問題になっています。 [2: 大体] 規則を作っ  
て制限すると [3: 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 [2: が] 賛成しますが、つぎのように自分の見方  
を述べます。

上に示すように、下線が引かれた部分を [ ] で囲い、評定値を [ ] 内の冒頭に半角数字で入力し、原文テキストとの間をセミコロンで区切ってテキストファイルとして保存することとした。

## 2.2. 一般日本語母語話者からのデータ

また、日本語教師以外の人々からも学習者作文に対する評価データを収集したいと考えた。しかし添削とは極めて負担の大きい行為であるため、日本語教育の経験のない人々にいきなり添削を依頼するのは難しい。そこで日本語教育の経験のない人々からは添削データは収集せず、「評定データ」のみを収集することとした。

2006年には、113編の作文について、それぞれ10名前後の評定者（作文によって評定者の組み合わせは異なる）に評定を依頼した。このときは、「日本語として自然かどうか」という「自然度」の評定と、「意味が分かるかどうか」という「理解度」の評定を分けてお願いした。

また2007年には、2006年の調査で評価のばらつきの大きかった20編の作文について、それぞれ20名の評定者（評定者の組み合わせは同じ）に評定を依頼した。このときは評定後のインタビューに主たる目的があったため、「自然度」と「理解度」に分けて評定をお願いすることはせず、「日本語としておかしい、不自然だ」と思った箇所の重篤度のみを挙げてもらった。

以上をまとめると、今回の科研費研究では、以下の種類の評価データが収集できたことになる。

- 1) 10 編の作文に対する, 43 名の日本語教師 (または日本語教師候補者) による添削情報および評定情報 (データ数 430)
- 2) 113 編の作文に対する, それぞれ 10 名前後の一般日本語母語話者による, 「自然度」「理解度」についての評定情報 (データ数約 2,260)
- 3) 20 編の作文に対する, 20 名の一般日本語母語話者による評定情報 (データ数 400)

なお, 1), 3)の評定対象となった作文はすべて 2)の評定対象に含まれている。1), 3)の評定対象は一部重複している。

### 3. 不自然箇所, 理解不能箇所の「範囲指定」に関する問題

ところで, 添削データにしても評定データにしても, 作文中「おかしい」または「分からない」と感じられた箇所を範囲指定し, その範囲について何らかの情報が付与されたものであるということが出来る。こうしたデータを集計し量的に扱うためには, 「どういう言語単位に対して範囲指定をおこなうか」ということが規定されなければならない。

例えば, 原文に「たばこを禁止するはずだ」という記述があり, これを評定者は「たばこを禁止するべきだ」と修正したいと考えたとしよう。この場合, 修正すべき範囲の指定方法としては, 「はず」, 「はずだ」, 「するはずだ」, 「禁止するはずだ」などさまざまな可能性がありうる。注目した箇所は同じであっても, 評定者によってそれを指定する範囲が異なっていると, 例えば「誤用・不自然箇所の数」を数えるようなときに不都合が生ずる。

国立国語研究所で作成している『日本語話し言葉コーパス(CSJ)』においては, 発話内容を文字化する際, 「発音の怠けや音の転訛, 言い間違いなどが生じた場合に付与する」タグとして「タグ(W)」というものを設定しており, そしてこのタグは原則として「短単位」と呼ばれる言語単位に付与するものと規定されている(国立国語研究所 2006: 103)。「短単位」とは, 「言語の形態的側面に着目して規定した単位」であり, 「現代語において意味を持つ最小の単位」(最小単位)を, 基準を満たす形で結合させて(または結合させないで)得られる単位のことを指す(短単位の詳細な認定基準は国立国語研究所 2006: 141-144 参照)。「短単位認定基準」によると, 例えば「室内環境」という語は, 「室内」「環境」という 2 つの短単位に分節されるため, この語が「シンナイカキョー」のように発音された場合, 以下 1) のようなタグ付与がなされ, 2)や 3)のようにはならない。

- 1) (W シンナイ; シツナイ) (W カキョー; カンキョー)<sup>4</sup>
- × 2) (W シンナイカキョー; シツナイカンキョー)
- × 3) シ(W ン; ツ)ナイカ(W ;ン)<sup>5</sup>キョー

<sup>4</sup> タグ(W)はここに示したように, 「セミコロンの左側に実際に発音された音を可能な範囲で正確に書き表わすと同時に, セミコロンの右側には丁寧に発音された場合に生じる (と予想される) 音を併記する」(国立国語研究所 2006:103)という形をとる。

本来ならば、添削や評定を依頼する際、どういう範囲の言語単位に対し「修正すべき」という指示をおこなうのかを明確にしておきたいところではある。しかしながら、「短単位」の認定基準を理解するにはある程度の文法知識がどうしても必要であり、基準に沿った単位認定をおこなうためには、評定者にそれなりのトレーニングを受けてもらわなければならない。言語の専門家だけでなく、一般の母語話者にも評定を依頼したいという目的を考えると、それは決して現実的なことではない。

また実際問題として、学習者の誤用・不自然表現の中には、一続きの言語表現を明確な単位に切り分け、「この単位に問題がある」ということを特定することが極めて困難なものが多いのである。

例えば、以下のような例を見てみよう。

**未来の希望を浄\*化するため、私たち成人は努力をつけるのは吝嗇\*ですか<sup>6</sup>。**

「短単位」を目安に誤用・不自然表現を指摘するなら、「浄\*化」、「成人は」の「は」、「つける」、「吝嗇\*」などの単位が指摘されることになるだろう。しかしながら、ここでは単に「浄化」という語の意味が分からないというより、「未来の希望を浄化する」という表現全体で何を表そうとしているのかが分からない、という見方もでき、そうすると「未来の希望を浄化する」全体を修正対象箇所とすることもできるだろう（文脈から判断すると、ここで「未来の希望を浄化する」というのは、「子どもたちがたばこを吸わないようにすること」を指しているようである。この部分をしかるべく添削するには、おそらく「浄化」という1短単位だけでなく、もっと広い範囲にわたって修正をおこなう必要があるものと思われる）。「努力をつけるのは吝嗇\*ですか」の部分についても、まったく同様のことが言えるだろう。

学習者の誤用・不自然表現にはこのように、「単位と単位との関係性がおかしい」、というものが多く、ひとつの短い単位の中に収まりきる誤用というのは実はそう多くないのである。そして「おかしさ」が複数の単位にまたがって存在している場合、「どこからどこまでがおかしいのか」ということを一意に決めるのは極めて難しい<sup>7</sup>。

そこでわれわれは、誤用・不自然表現の範囲については評定者に一切の指示をおこなわず、どこからどこまでをおかしいと感じたか、という範囲指定も評定者の判断に任せることとした。つまり、「未来の希望を浄\*化するため」全体を指摘しても、「希望を浄\*化する」を指摘しても、また「浄\*化」だけを指摘してもかまわないこととした。

---

<sup>5</sup> 「本来あるべきものがない」場合には、ここに示すように、セミコロン`;`の左側は「空」となる。

<sup>6</sup> 「対訳作文データベース」所収の“cn015j.txt”より。なおデータ中の「\*」は、その直前の文字の字形が誤っており、既存のフォントではその字形を表現できないことを示している。

<sup>7</sup> CSJのタグ(W)は、あくまでも「発音の怠けや音の転訛、言い間違いなど」に限って付与されるので、タグ付け範囲の認定に関する困難は生じない。

同時に、範囲指定の「入れ子」も認めることとした。例えば、

そのことばかりは昔のばなしとれきしがあります。(kh050j)

という文は、文全体の意味が分からないが、その中で「ばなし」という箇所は明らかに「はなし」と書くつもりで誤ってしまったものであろうし、「れきし」についても、漢字で書かねば違和感をもってしまう評定者もいるだろう。こうした場合、上掲の文全体を不自然表現として下線を付した上で、その中の「ばなし」「れきし」にも、不自然表現として二重に下線を付すことを認めることとした。

以下に、不自然箇所が入れ子状になった場合の電子化例を示す。

[2;そのことばかりは昔の[3;ばなし]と[1;れきし]があります。]

その上で、作文原文のテキストを文字単位に分け、「それぞれの文字について、何名の評定者が下線を引いているか」を集計し、その集計結果を視覚的に表現することとした。

以下の図1は、先ほどの「未来の希望を…」という文に対する、評定者10名の「自然度」評定結果(2006年の調査)を集計したものである。

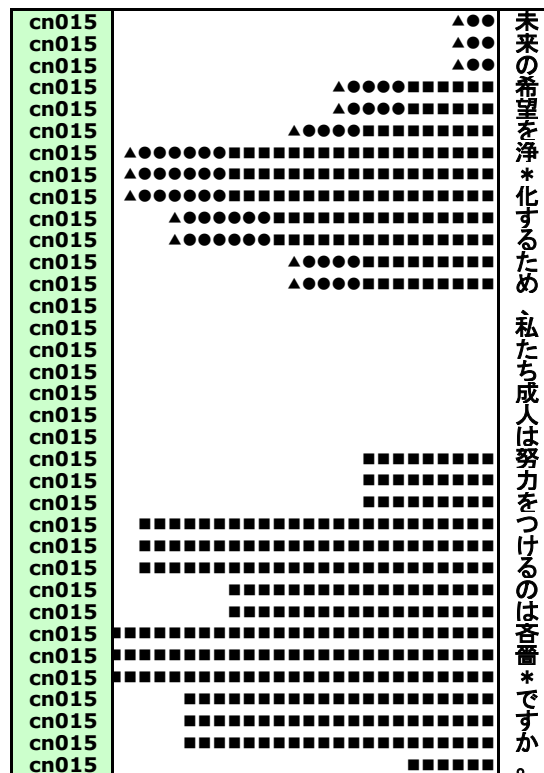


図1・評定者10名の「自然度」評定の集計結果

任意の文字について、「重篤度 3」の下線を引いた評定者が 1 名いたとき、その文字には■が 3 つ与えられる。同様に、「重篤度 2」の下線を引いた評定者が 1 名いた場合は●が 2 つ、「重篤度 1」の場合には▲が 1 つ与えられる。

「浄\*化」の部分には、■が 18, ●が 6, ▲が 1 つついている。つまり、この部分の自然さについて、「重篤度 3」と判定した評定者が 6 名、「重篤度 2」が 3 名、「重篤度 1」が 1 名いたことになる。同様に、「(努力を) つける」の部分については「重篤度 3」の判定者が 8 名、「吝嗇\*」については「重篤度 3」が 9 名いた、ということが分かる。

図 1 を見ると、「未来の希望を浄\*化するため」、「努力をつけるのは吝嗇\*ですか」という部分がそれぞれ不自然さの「チャンク」をなしていることが分かる。さらにひとつめのチャンクは、「浄\*化」を不自然さの「ピーク」とするひとつの山をなしているのに対し、ふたつめのチャンクには「つける」、「吝嗇\*」というふたつのピークを認めることができる。

このように、複数評定者の評定結果を集計することで、「チャンク」と「ピーク」という 2 種類の「数えられる単位」<sup>8</sup>が得られることになる。もし、単語・文字といった分節的な単位における誤用を研究対象とする場合は「ピーク」に、複数の単位にまたがって生じる不自然さを研究対象とする場合は「チャンク」に着目すればよい。もちろん、誤用・不自然表現の数を集計したいというときには、単純に「ピーク」「チャンク」の数を数えるのではなく、研究者自身の目視により再確認することは不可欠である。しかし、ひとりひとりの評定者が指定する誤用・不自然表現の範囲はまちまちであっても、それらを重ね合わせることによって「数えられる単位」がおのずと浮かび上がってくるのは興味深い。

---

<sup>8</sup> 「ピーク」を作ることで「数えられる単位」を得ようとする考えは、音韻論において「音節」の認定をおこなう際、単語内の「聞こえ度(sonority)」のピークの数をその単語の音節数とする、という考えからヒントを得ていることを付記しておく(窪田・本間 2002)。



## 4. 分析ツールの作成

### 4.1. 概要

以上のような評定結果を集計するためにツールを作成した。ツールは WWW サーバー上での動作をする CGI スクリプトである<sup>9</sup>。スクリプト言語は Perl を用いている。

このプログラムは、タグ付きの添削・評定のテキストの集計をおこなうためのツールである。入力するタグは、添削で用いた XML 形式のタグと、評定で用いた簡易タグの両方のタグが集計可能となっている。

添削・評定のタグの書式は、調査ごとに異なっているため、そのまま比較することには注意が必要である。しかし、ある表現に対してなんらかの「気づき」があった場合にタグが付与されている、という点では一致している。

### 4.2. 集計方法

前節で少しふれたが、集計方法には複数の方法がある。重篤度の加算方法については、一つは重篤度の得点に応じた加算方法であり、もう一つは重篤度にかかわらず人数を集計する方法である。

また、タグが「入れ子」になっている場合の集計方法については、一つは、複数のタグがある中で重篤度が最も重いものを採用する方法であり、もう一つは複数の重篤度を全て合計する方法である。前者は得点の最大値が一定なので得点の比較には優れている反面、重い誤用の中に軽い誤用がある場合には誤用が埋もれてしまう可能性がある。後者は「ピーク」の認定には優れるが、最大値の意味するものがわかりにくくなるので、分析時には注意が必要である。

加算方法については、後述するファイルにて指定することができる。「入れ子」の集計方法については、現在は最も重い重篤度を採用する方法をとっている。

また、文字単位の集計ではあるが、挿入に関する添削・評定には、文字と文字の間に行われるため、実際に集計対象となる箇所は、作文の文字数の2倍となる。文字間の処理は、タグの終了判定においても必要である。つまり同じタグが連続して付与されているときに、文字だけだとタグの切れ目の判定ができなくなるためである。

### 4.3. 「山」の出力

#### 4.3.1. 文字のずれのチェック

添削の XML タグや、評定の簡易タグは、テキストエディタによって付与される。すなわちオリジナル文書を手で加工するため、誤って文字を消してしまったり、不要な文字を追加してしまったりする可能性がある。

そのため、ツールにおいては、タグの集計だけでなく、文字列のずれがないかどうかについてチェックする機能も存在する。これは全てのタグを取り払って元の文章を比較する

---

<sup>9</sup> このツールは、「作文対訳データベース」の web ページから利用できるようにする予定である。

ものである。図 2 に、文字のずれをチェックする画面の例を示す。

文字のずれが生じている部分に色をつけて表示している。図では、左から 4 番目のファイルの「中国は」の次の文字のところが他のファイルとずれていることがわかる。タグ挿入時に誤って関係のない空白を挿入してしまったため、文字がずれていることがわかる。

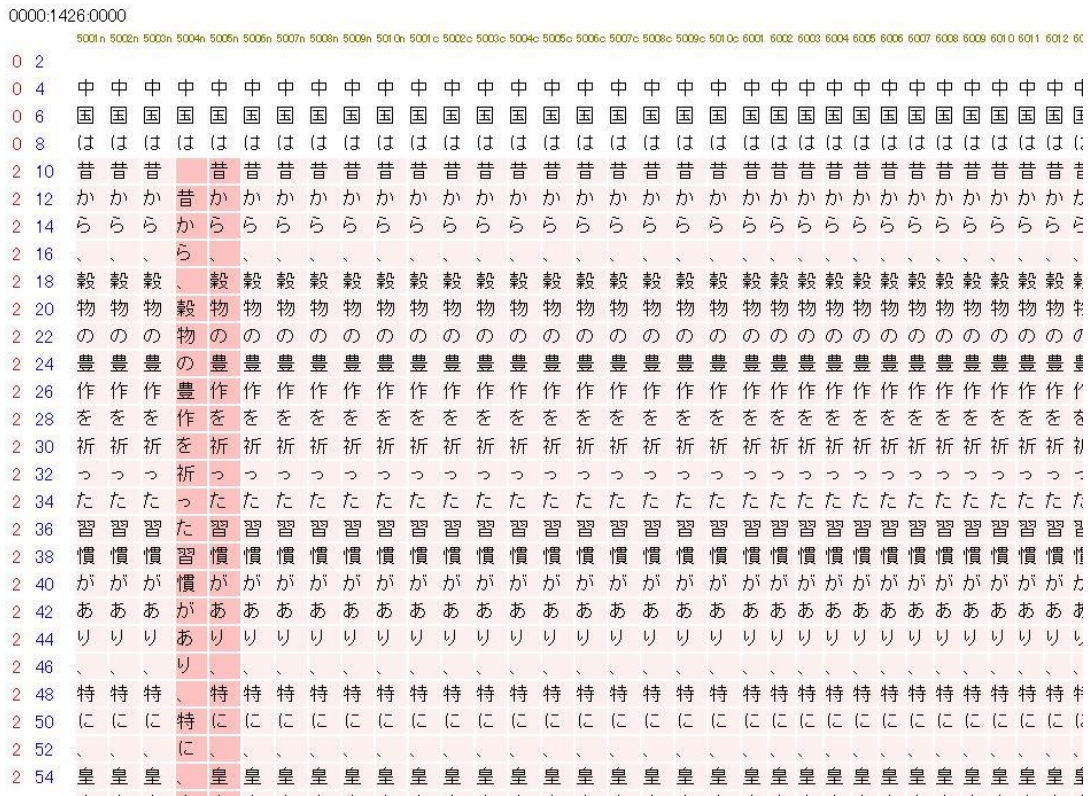


図 2・文字チェック画面

#### 4.3.2. 重み付けの決定

添削や評定では、誤りの箇所に対して、何らかの重み付けがなされている。しかし、その重み付けの方法は、添削・評定の方針によっても、タグの名称やタグの属性といった、タグの仕様によっても異なっている。

例えば、添削 ver.2.0 タグにおいては、重篤度が高い修正には **correct** タグ、やや低い修正には **better** タグと、タグの名前によって重篤性の重み付けがなされている。これが、ver.3.0 タグにおいては、**mod** タグに統一されており、**mod** タグの **a** 属性として、**a=3** と **a=2** に分けられる。また、新たに **a=1** の場合も存在する。

こうしたタグの重み付けの対応については、一律に定められないため、得点とタグ・属性の対応表を、事前に別途ファイルで指定しておく必要がある。なお、このツールは、基本的に XML ファイルを集計するように作成されているので、簡易タグに対しても XML 形式として指定しなければならない。

簡易タグは、

[ 評定値 / コメント ; 評定対象となる文字列 ]

という書式であるが、これは、

<rating value=" 評定値" comment=" コメント" >文字列</rating>

という XML タグとみなされる。なお、このツールでは、簡易タグから上記 XML タグへの変換は、ツール内で自動的に行うため、簡易タグのファイルを修正する必要はない。以下に、タグ名・属性値と得点との対応表の作成例を示す。

得点	タグ名	属性名	属性値
3	correct		
2	better		
1	unclear		
3	mod	a	3
2	mod	b	2
1	mod	c	1

たとえば、集計において誤用の重篤性よりも、誤用の場所を重要視したい場合には、得点を一律に 1 にすればよい。また、一部のタグを集計から外したい場合には、リストから外せばよい。分析に応じて、タグ名・属性値と得点の対応関係を定めるとよいだろう。

こうして作文の各文字ごとに与えられた得点を集計し、山の形で出力される。得点のままでは、添削・評定者の数が調査によって異なり比較ができない。そのため、得点は比率の形式に統一され、一定の段階数で出力される。

図 3 はグラフの出力例である。グラフは縦表示・横表示のどちらでも可能である。図における作文 cn044 は、1つの作文に対し、以下 6 種類の添削・評定情報がそろっている。

- 1) 2000 年収集の、日本語母語話者による添削情報 (添削者数 23 名)<sup>10</sup>
- 2) 05-06 年に収集した、日本語教師 (および教師になろうとしている人々) による添削情報 (添削者数 43 名)
- 3) 2) と同一の添削者が施した評定情報 (評定者数 43 名)

---

<sup>10</sup> このデータは本科研費研究期間中に収集したものではない。「作文対訳コーパス」からランダムに抽出した 20 編の作文に対し、24 名の添削者 (日本語教育経験者も未経験者も含む) に添削を依頼し、その添削結果を ver.1 のタグセットを用いて実験的に電子化したものである。

- 4) 06年に収集した、一般日本語母語話者による「自然度」評価情報（評価者数10名）
- 5) 06年に収集した、一般日本語母語話者による「理解度」評価情報（評価者数10名）
- 6) 07年に収集した、一般日本語母語話者による評価情報（評価者数20名）

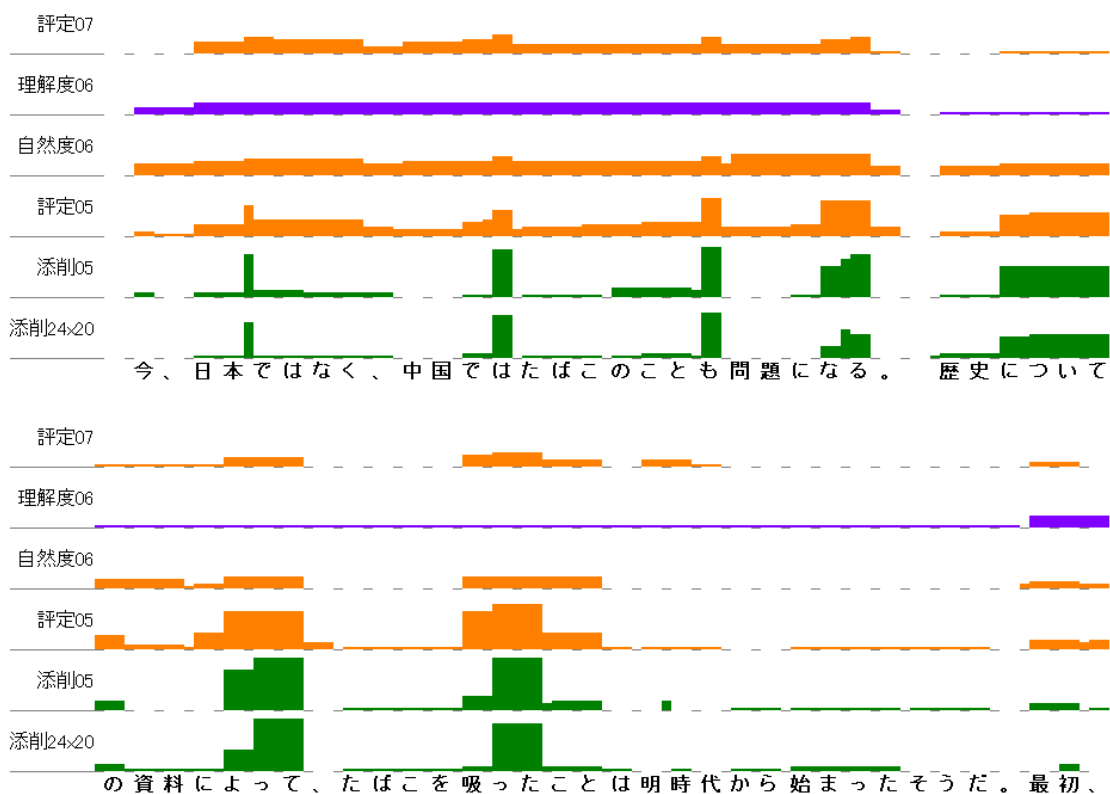


図3・添削・評価の山の例

#### 4.3.3. 形態素解析との接続

さらにわれわれは、峯(2003)に基づき、作文データを形態素単位に分割し、それぞれに対し形態素情報を付加するという作業もおこなっている<sup>11</sup>。この形態素情報を添削データと組み合わせることによって、品詞情報から添削情報を分析することが可能となる。同時に、前述の単位と添削・評価の関係を考える上でも重要である。図4に出力例を示す。こちらは縦表示の例である。

形態素情報がまだ付与されていない作文データについては、分析者自身がそのデータを「茶釜」にかけて解析し、その結果を添削情報と組み合わせることも可能である。しかし学習者の日本語（だけでなく、校正されていない日本語の多くは文法上の誤りを多く含む）

<sup>11</sup> 学習者作文を形態素解析ソフトウェア「茶釜」にかけたあと、作業者の目視により誤解析を修正したうえでデータベースに搭載している。

であるために、どうしても解析ミスが多く現れる。その点は注意する必要があるだろう。

形態素データベースとの組み合わせについては、現時点では正式には実装されておらず試験的な機能である。

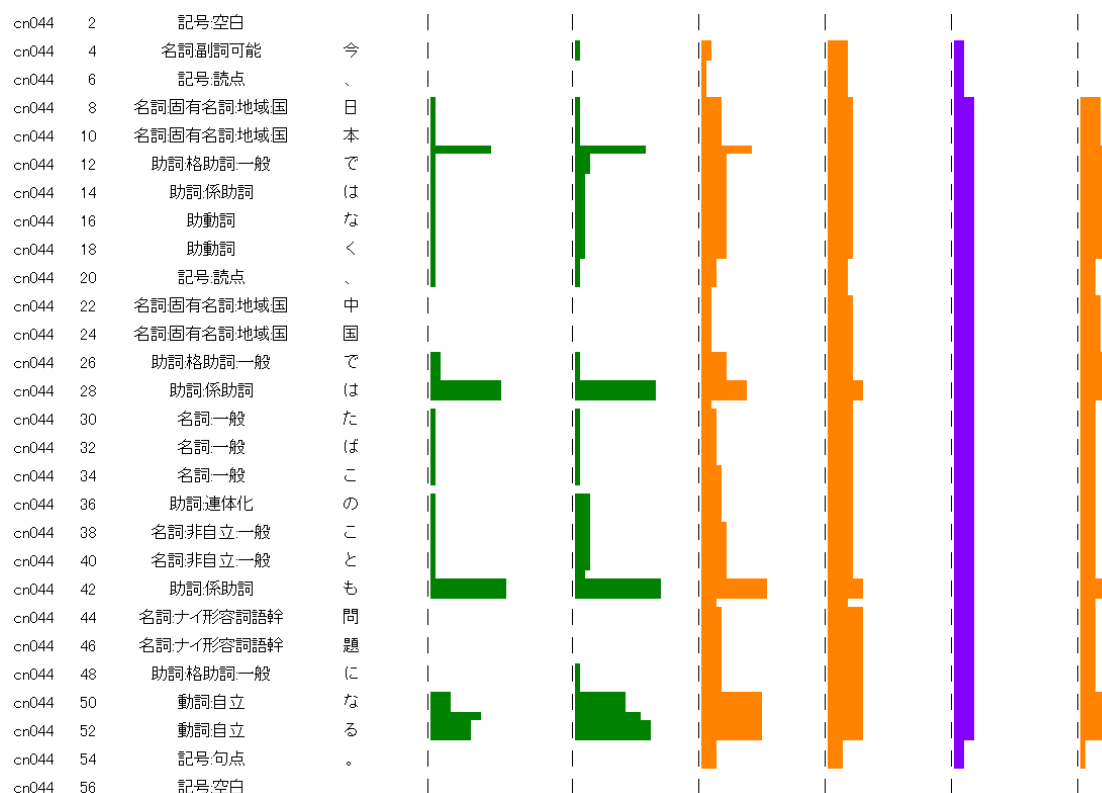


図 4・形態素情報を組み合わせた例

## 5. おわりに

以上、同一作文に対する複数の添削・評定情報を電子化・集計する方法を示すとともに、その分析ツールについて紹介した。

「評価観」の研究をおこなっていくためには、宇佐美(2007)がおこなっているように、個々の添削者・評定者がどのように学習者の作文を読み、考え、評価をおこなったか、という思考の過程をミクロに見ていくことも重要である。同時に、多くの人々の添削・評定データを集めることによって、添削者・評定者全体に共通する傾向をマクロにとらえていくことも必要であろう。

本ツールは、「評価観」をマクロにとらえていくためのものとして位置づけることができる。宇佐美(2007)では、評価観研究においては質的研究・量的研究の両側面が必要であることを述べているが、本ツールは「量的研究」に対し、さまざまな応用の可能性を持っているといえるであろう。

## 参考文献

- 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
- 宇佐美洋(2002)「XML による作文添削情報表示システムの開発と, その応用」『2002 年度日本語教育学会春季大会予稿集』 231-232
- 宇佐美洋(2006)「「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」『作文対訳データベースの多様な活用のためにー「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』, 9-42
- 宇佐美洋(2007)「学習者作文の評価における観点の多様性ー一般日本語母語話者の場合ー」『日本語学習者の書き言葉に対する対照言語学的・文章論的研究』 課題番号 17520360 平成 17-18 年度 科学研究費補助金基盤研究(C)(2) 研究成果報告書 研究代表者: 宇佐美洋, 29-67
- 宇佐美洋・鍵水兼貴(2006a)「「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ー「添削情報電子化」の真の意義を求めてー」『作文対訳データベースの多様な活用のためにー「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』, 53-78
- 宇佐美洋・鍵水兼貴(2006b)「添削時の思考過程を考慮した, XML による作文添削情報表示システム」『2006 年度日本語教育学会春季大会予稿集』 269-270
- 窪菌晴夫・本間猛(2002)『音節とモーラ』, 研究社
- 国立国語研究所(2006)『日本語話し言葉コーパスの構築法』国立国語研究所報告 124
- 峯布由紀(2003)「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ー第二言語としての日本語の習得研究へむけてー」『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 89-107
- 鍵水兼貴(2006)「学習者日本語作文における「形態素情報」の活用ー「添削情報」との連携を目指してー」『作文対訳データベースの多様な活用のためにー「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』, 121-144