

# CTLJ 誤用辭典之開發與分析

黃淑妙

國立成功大學 外國語文學系 副教授

## 摘要

本研究介紹以 CTLJ 為基礎開發之誤用辭典系統，並分析誤用辭典中各學習階段之誤用率，提供日語教學者之參考，以減少學習者之誤用。

本誤用辭典系統主要以 Perl 和 HTML 開發，提供查詢學習者之誤用字及訂正字，統計誤用出現頻率，俾便助於日語教育、日語研究、或以中文進行對照分析為目的而編纂。使用者可利用系統提供之關鍵字、學習時數、詞性與誤用型別等作為查詢條件，亦可將查詢結果下載。

CTLJ 已建有 53,865 組誤用標誌 (error tag)，誤用辭典係將上述之誤用標誌涵蓋的部分納入予以開發建置，其中包含誤用之句子數總計為 53,497，誤用之詞素累計達 73,371、個別詞素也有 9,122 例等。經分析結果得知學習時數和誤用傾向之間並無一定比例，資料顯示學習時數遞增而誤用遞減之情況僅出現於部分之學習時數。此外，誤用辭典中每個學習階段之誤用詞類樣態不一，然而，動詞誤用頻率高過其他詞類為所有學習時數之共同現象，是以減少動詞誤用為日語教學者後續優先目標。

關鍵字：CTLJ、誤用辭典、語料庫、品詞、檢索功能

受理日期：2016.09.03

通過日期：2016.10.14

# **A Japanese Misuse Dictionary System Based on CTLJ: Its Development and Analysis**

Huang Su-miao

Associate Professor, National Cheng Kung University

## **Abstract**

This paper presents a misused Japanese dictionary system based on CTLJ, and investigates how the dictionary system can be used to analyze the misuse rate of the learners of Japanese at each level of their learning.

This system was developed by using Perl and HTML with the objective of enabling the learners of Japanese to access misused words or corrections, misuse frequencies, in order to facilitate their language learning Japanese language studies, and contrastive analysis of Japanese and Chinese. This system also provides key words, learning periods, parts of speech, and misuse types. Users in addition can download Excel attachments regarding their respective search results.

Since the CTEJ already includes 53,865 error tags, the proposed dictionary system is constructed by incorporating the rich resource of the error tags. In the system, the sentence level misuses amount to 53,497, and the total number of the morphemes in misused words is 73,371, which further falls into 9,122 individual types. Analytical results reveals that misused tendency is not proportional to learning hours, and the fact that misuses decrease as learning hours increase only appears in parts of learning hours. Aside from that, misuse words and types in the dictionary system are obviously different in each learning phase; however, it is common that the misuse rate of the verb is higher than other word types for all learning hours. Therefore, the reduction of verb misuse would be the future priority goal for Japanese educators.

Key words: CTLJ, misused Japanese dictionary system, Corpus, word type, search functions

# CTLJ 誤用辞典の開発と分析

黄淑妙

国立成功大学外文系 副教授

## 要旨

本稿では CTLJ に基づく誤用辞典の開発とその機能、および上記データ中における各学習段階の誤用率に関する分析結果について報告する。

当誤用辞典は、主に Perl 及び HTML を使用して開発されたもので、誤用例の検索及び出現頻度の確認を通じて、日本語教育、日本語研究あるいは中国語との対照分析に役立つことを目的として編まれている。検索機能はキーワード、学習時間数、品詞、誤用タイプをカバーしており、利用者は検索結果をエクセルファイルで保存することも可能である。

CTLJ における修正タグの数は 53,865 セットである。CTLJ 誤用辞典はこれらの修正タグに囲まれた部分を基に開発された。CTLJ 誤用辞典における誤用を含む文が延べ 53,497 例、誤用の延べ形態素が 73,371 例、異なり形態素が 9,122 例収録されている。分析の結果を見ると、学習時間数と誤用の傾向は一定であるとは言えず、学習時間数が増えていくにつれ誤用が減少するのは学習時間数の一部に限って見られる。また、各学習時間数別の誤用状況はそれぞれ異なるが、動詞の誤用率が高くなっている点で共通しており、動詞の誤用を減らすことが日本語教育において優先的に対処すべき課題であることを示唆している。

キーワード：CTLJ、誤用辞典、コーパス、品詞、検索機能

# CTLJ 誤用辞典の開発と分析<sup>1</sup>

黄淑妙

国立成功大学外文系 副教授

## 1. はじめに

「CTLJ 誤用辞典」は、「台湾人日本語学習者コーパス (以下 CTLJ)」<sup>2</sup>を基に開発された。本稿では CTLJ 誤用辞典の開発とその機能、および上記データ中における各学習段階の誤用率に関する分析結果について報告する。

## 2. 先行研究

日本語学習者の誤用に関する研究は実に膨大な数にのぼる。黄 (2009)、伊集院 (2016) ではコーパス、学習者コーパスおよび外国語としての日本語の習得研究における先行研究について詳しくまとめた。伊集院 (2016) においては、「語彙・文法」と「第二言語習得プロセス」に関する研究のみならず、「談話・語用論」の分野における研究も紹介した。ここでは、黄 (2009) と伊集院 (2016) に収められていない研究をいくつか取り上げて見てみたい。例えば、岡田・林田 (2016) や、森山 (2015)、若生 (2012)、島田 (2011)、坂口 (2004)、迫田 (2001) などがある。

岡田・林田 (2016) は、①「あの喫茶店にこーヒーを飲む」のような誤用がどのような用法の「に」と動作場所を表す「で」の混同によるものであるかに焦点を当て、日本語学習者の「に」と「で」の習得を探った。その結果、「場所への移動がある」と判断された場合に①の「に」が産出されるという段階を経て、その後、移動先を

<sup>1</sup>本稿における研究成果は MOST 104-2410-H-006-044-「以學習者語料庫為本之日語學習發展階段研究」の助成を受けて作成した。

<sup>2</sup>「台湾人日本語学習者コーパス」(CTLJ: The Corpus of Taiwanese Learner of Japanese) <<http://corpora.flld.ncku.edu.tw/>>は台湾の13の高等教育機関で日本語を学ぶ学習者の作文1,563編(約71万字)が収録されている学習者コーパスである。詳しくは黄(2009)を参照されたい。

表す「に」と動作場所を表す「で」を正しく用いる段階に至るのではないかとしている。

森山(2015)は中国語母語話者を対象に、文脈の中で和語動詞を正確に使用できるかについて検討した。調査の結果、「日本語習熟度が高くなるにつれて文脈の中で和語動詞を正確に使用できる傾向にある」、「和語動詞に関しては理解の知識に比べ産出の知識は遅れて発達する」、「和語動詞産出の過程で(1)母語や意味からの連想による誤り、(2)日本語の漢語を経由した誤り、(3)弱い共起関係による誤り、(4)送り仮名が同じことであることによる誤り、(5)助詞を正しく理解できなかったことによる誤りが見られる」という結論を出した。

一方、坂口(2004)は、韓国語母語話者と中国語母語話者を対象に、初級から上級にかけての2年間に書かれた作文から助詞を含む文データを抽出し、学習段階による誤用の変化や誤用の種類を分析した。その結果、文法的知識の混乱や欠如による誤用に関しては、初級ではなく、混同する用法が増える中級以降で増加すると結論付けた。坂口(2004)のこの結論は森山(2015)が出した「日本語習熟度が高くなるにつれて文脈の中で和語動詞を正確に使用できる傾向にある」という結論とは対照的になっている。調査項目が異なれば、結論も異なることがありうるが、客観的な誤用の傾向を導き出すには大規模データに基づいて客観的に計算する必要がある。その視点から、本稿では CTLJ 誤用辞典を用いて各学習時間数ごとに品詞別の誤用傾向を検討する。

また、誤用の要因については母語干渉がよく指摘される。例えば、島田(2011)は中国語母語話者の授業感想カードや日記、ブログなどの記述の誤用を分析し、そのうちの助詞、形容詞を取り上げている。助詞の誤用では「場所」を表す「に」「で」「を」の誤用が目立ち、化石化傾向も見られた。用いられていた動詞の多くは日常生活での使用度が高く、初級レベルで学習するものが多かったことから、島田は助詞と動詞を結び付けた習得が進んでおらず、別々のものとして習得してきた結果、助詞選択が曖昧なままになってしまったもの

と推測している。それに加え、一つの形態が複数の用法を担っていることも誤用の要因になっていると指摘した。このような助詞の誤用を防ぐには、日本語の習得レベルに応じて、助詞と動詞を一連の動作の流れとセットで習得させることが有効だと述べている。一方、形容詞の誤用では、連体形「な」と「の」の使い分けに関する誤用が多く、これは中国語の「的」と日本語の「な」と「の」の意味機能と形態上の区別が母語干渉になっていると指摘した。坂口(2004)でも中国語母語話者は後接する述語や文型から規則的にあてはめる傾向があると述べている。

他方、若生(2012)は、国籍・母語の異なる留学生4名に対してインタビュー調査を行い、中間言語研究の観点から留学生の日本語に見られる誤用を考察した。その結果、「引用節におけるナ形容詞と名詞文の『だ』の脱落」、「連体形+連体助詞『の』」、「場所の格助詞『に』と『で』の誤用」が見られた。ナ形容詞と名詞述語の「だ」が引用節の中で脱落する現象は、複雑な活用体系を単純化しようとする過剰般化である。また、動詞、イ形容詞の連体形に連体助詞「の」が後続する現象については、これまで中国語母語話者が母語の「的」を移転させたものだとされてきたが、ドイツ語母語話者の日本語にも確認されたことから、ユニット形成ストラテジーなども原因として考えられるのではないかとしている。

迫田(2001)は、日本語教育機関で学習している中級レベルの留学生を対象に、場所を表す格助詞「に」と「で」の使い分けに関する調査を行い、誤用の原因解明に取り組んだ。その研究結果によると、「に」と「で」の使い分けに学習者の母語の違いによる影響は見られず、誤用にも訂正例にも学習者に共通する傾向があることが分かった。また、使用例の分析や調査から、「位置を示す名詞+に」や「地名・建物を示す名詞+で」が固まりを形成している可能性が高いことが分かった。そして、これらのことから、学習者は格助詞のようなそれ自体あまり意味を持たない機能語の選択を迫られた場合に、その近くにある語の意味や機能に依存して一かたまりの固定した語

(ユニット)として扱って覚えてしまう、ユニット形成のストラテジーをとるのではないかと推測し、それが誤用を生み出す原因の一つであると結論付けた。

近年では、誤用を分析するだけでなく、誤用のデータベースである学習者コーパスの構築が増えており、例えば、「日本・韓国・台湾の大学生による日本語意見文データベース」<sup>3</sup>や「日本語学習者作文コーパス」<sup>4</sup>などが挙げられる。伊集院(2016)では、日本語学習者コーパスについて書き言葉、話し言葉、書き言葉と話し言葉に分けて表で整理している。その中で添削タグ付きのコーパスとしては、国立国語研究所の「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：作文対訳DB」<sup>5</sup>、李在鎬の「日本語学習者作文コーパス」、東京工業大学の「学習者作文コーパス『なたね』」<sup>6</sup>およびCTLJのみである。この4種類の日本語学習者コーパスについて表1で比較してみた。

表1. 添削タグ付きの日本語学習者コーパスの比較

コーパス名	作文対訳DB	日本語学習者作文コーパス	学習者作文コーパス「なたね」	CTLJおよびCTLJ誤用辞典
添削タグ付きの作文数	263編 <sup>7</sup>	304編 (学習者304名)	285編 (学習者192名)	1,563編 (学習者1,180名)

<sup>3</sup>2011年オンライン公開。

[http://www.tufs.ac.jp/ts/personal/ijuin/koukai\\_data1.html](http://www.tufs.ac.jp/ts/personal/ijuin/koukai_data1.html)

<sup>4</sup>2013年オンライン公開 <http://sakubun.jpn.org/>

<sup>5</sup>2000年CD-ROM版、2001年オンライン公開、2009年データベース改定

[http://contr-db.ninjal.ac.jp/essay\\_01.html](http://contr-db.ninjal.ac.jp/essay_01.html)

<sup>6</sup>2012年オンライン公開 <https://hinoki-project.org/natane>

<sup>7</sup>作文対訳DBにおける作文数は1,754編あるが、抽出できる添削データは263編と記載されている。( <http://contr-db.ninjal.ac.jp/cert/taiyakudb/sakubuns> による)

形態素数	未知	113,554	未知 (文字数 205,520)	CTLJ: 449,394 (文字数: 722,000) CTLJ誤用辞 典: 159,818 <sup>8</sup> (文字数: 273,048)
作文テー マ	10種類+ その他	2種類	×	21種類
母語	32種類	韓国語/中 国語	9種類	中国語/台湾語 /原住民族語
学習時間 数 (日本語 履歴)	8段階 (1ヶ月 ~5年)	2年未満/2 年以上5年 未満/5年 以上	×	12段階 (150時間未満 ~3000時間)
検索方法	全文	・形態素単 位か文字 列で検索 ・誤用/正 用/両方の 検索可能	文字列	CTLJ誤用辞典 ・文字列/POS/ 誤用タイプ/学 習時間数から 選べる ・誤用/訂正例 のどちらか
検索結果 のダウン ロード	画像 (jpg)フ ァイルま たはxml ファイル	テキスト ファイル またはエ クセルフ ァイル	テキストフ ァイル	エクセルファ イル

<sup>8</sup>159,818 は誤用と訂正例の延べ形態素数の合計である。詳しくは表 3 を参照されたい。

誤用タイプ	×	文字/文法 /文体	16種類	置き換え/挿入 /削除可/複数 提示/削除/移 動/他候補/意 味不明
品詞	×	16種類	19種類	16種類

表1に示したように、この4種類の日本語学習者コーパスにはそれぞれ特色がある。「作文対訳DB」は母語の種類が一番多いものの、特定の誤用を検索することができず、検索の結果は全文で出力される。特定の誤用を検索できるのは「日本語学習者作文コーパス」、「なたね」と「CTLJ誤用辞典」のみである。表現の方法は異なるが、この三種類のコーパスはいずれも誤用と訂正例が検索結果として文列で見られる。「日本語学習者作文コーパス」は母語と学習レベル別（初級/中級/上級）の検索語統計結果がページ上にクロステーブルで表示されており、非常に便利であるが、その中に誤用のタイプは示されていない。一方、「なたね」は誤用タイプが最も詳細に分類されており、誤用種類別で絞り込み検索もできる。「なたね」は誤用の数が6,776件あるが、実際に「行く」で検索してみると、結果は3件<sup>9</sup>しか見つからなかった。これに対して「日本語学習者作文コーパス」では41件<sup>10</sup>、「CTLJ誤用辞典」では69件の誤用が見つかった。「CTLJ誤用辞典」は誤用の数が53,865セット、誤用と訂正例の延べ形態素数が159,818あることにより、規模としては一番大きい。また、検索機能はキーワードのみならず、キーワードを入力しなくても、学習時間数、品詞、誤用タイプからも検索できる。検索結果の下に詳しい出現頻度の統計結果が見られる。学習時間数別・誤用タイプ別による統計だけでなく、出現頻度順で結果が並べられている。以下、CTLJ誤用辞典の使い方について説明する。

<sup>9</sup>誤用箇所のみで、部分一致で検索した。（閲覧日：2016年8月7日）

<sup>10</sup>形態素単位で、誤用のみ表示で検索した。（閲覧日：2016年8月7日）

### 3. CTLJ 誤用辞典の使い方

CTLJ 誤用辞典は、主に Perl 及び HTML を使用し開発されたものである。学生の日本語誤用の検索及び統計をとることができ、日本語専門学者の研究又は分析に役立てることができる。また、CTLJ 誤用辞典はキーワード・学習時間数・品詞・誤用タイプ別に検索でき、利用者は検索結果をエクセルファイルで保存することも可能である。以下 CTLJ 誤用辞典の機能、動作環境および動作原理について説明する。

#### 3.1 CTLJ 誤用辞典の機能

CTLJ 誤用辞典の使い方と機能を以下のように述べる。

##### (1) リンクの位置

CTLJ のトップページに CTLJ 誤用辞典のリンクがあり、一般公開の準備が整い次第、リンクからアクセスして利用できる。(図 1 参照)

## 台湾日語教育學報第27号

図 1. CTLJ における当システムへのリンク



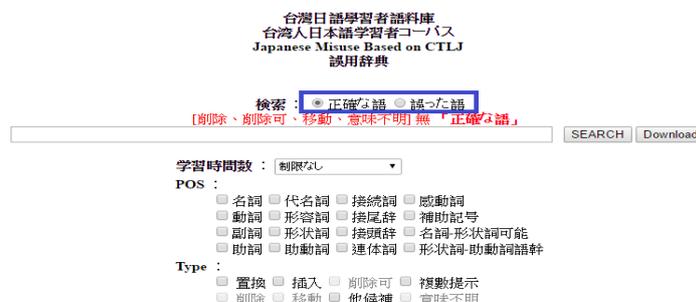
##### (2) 検索機能

キーワード・学習時間数・品詞・誤用タイプの4通りから検索条件を選択することができる。

① 「キーワード」：

正確な語（訂正例）もしくは誤った語（誤用）を検索条件に指定できる。（図 2 参照）

図 2. キーワード検索の条件指定



② 「学習時間数」：

ドロップダウンリストから学習時間数を指定できる。（図 3 参照）

図 3. 学習時間数の条件指定



③ 「品詞」：

16 種類の品詞の中から複数選択することができる。品詞は、名詞・代名詞・接続詞・感動詞・動詞・形容詞・接尾辞・補助記号・副詞・形状詞<sup>11</sup>・接頭辞・名詞-形状詞可能・助詞・助動詞・連体詞・形状

<sup>11</sup> 「形状詞」という呼び方は解析辞書 UniDic 1.3.12 による。塚脇（2012）には「UniDic では、サ変動詞語幹またはサ変名詞と呼ばれるものは名詞に、形容

詞-助動詞語幹の 16 種類に分けられる。(図 4 参照。POS は MeCab の解析結果に基づくものである。)

図 4. 品詞の条件指定

台湾日語學習者語料庫  
台湾人日本語學習者コーパス  
Japanese Misuse Based on CTLJ  
誤用辞典

検索：  正確な語  誤った語  
[削除、削除可、移動、意味不明] 無 「正確な語」

SEARCH Download

学習時間数： 制限なし

POS：

<input type="checkbox"/> 名詞	<input type="checkbox"/> 代名詞	<input type="checkbox"/> 接続詞	<input type="checkbox"/> 感動詞
<input type="checkbox"/> 動詞	<input type="checkbox"/> 形容詞	<input type="checkbox"/> 接尾辞	<input type="checkbox"/> 補助記号
<input type="checkbox"/> 副詞	<input type="checkbox"/> 形状詞	<input type="checkbox"/> 接頭辞	<input type="checkbox"/> 名詞-形状詞可能
<input type="checkbox"/> 助詞	<input type="checkbox"/> 助動詞	<input type="checkbox"/> 連体詞	<input type="checkbox"/> 形状詞-助動詞語幹

Type：

<input type="checkbox"/> 置換	<input type="checkbox"/> 挿入	<input type="checkbox"/> 削除可	<input type="checkbox"/> 複数提示
<input type="checkbox"/> 削除	<input type="checkbox"/> 移動	<input type="checkbox"/> 他候補	<input type="checkbox"/> 意味不明

④ 「誤用タイプ」：

置き換え・挿入・削除可・複数提示・comment・削除・移動・他候補・意味不明などの 9 種類に分けられる。(図 5 参照)

## 台湾日語教育學報第27号

図 5. タイプの条件指定

台湾日語學習者語料庫  
台湾人日本語學習者コーパス  
Japanese Misuse Based on CTLJ  
誤用辞典

検索：  正確な語  誤った語  
[削除、削除可、移動、意味不明] 無 「正確な語」

SEARCH Download

学習時間数： 制限なし

POS：

<input type="checkbox"/> 名詞	<input type="checkbox"/> 代名詞	<input type="checkbox"/> 接続詞	<input type="checkbox"/> 感動詞
<input type="checkbox"/> 動詞	<input type="checkbox"/> 形容詞	<input type="checkbox"/> 接尾辞	<input type="checkbox"/> 補助記号
<input type="checkbox"/> 副詞	<input type="checkbox"/> 形状詞	<input type="checkbox"/> 接頭辞	<input type="checkbox"/> 名詞-形状詞可能
<input type="checkbox"/> 助詞	<input type="checkbox"/> 助動詞	<input type="checkbox"/> 連体詞	<input type="checkbox"/> 形状詞-助動詞語幹

Type：

<input type="checkbox"/> 置換	<input type="checkbox"/> 挿入	<input type="checkbox"/> 削除可	<input type="checkbox"/> 複数提示
<input type="checkbox"/> 削除	<input type="checkbox"/> 移動	<input type="checkbox"/> 他候補	<input type="checkbox"/> 意味不明

(3) 統計機能

検索結果のページ下に、誤用語・学習時間数・誤用タイプ別に統計した誤用出現頻度が表示される。(図 6 参照)

- 「誤用語別誤用出現頻度の統計結果」：

動詞語幹と呼ばれるものは形状詞に分類されている」と記述されているように、「形状詞」は形容動詞の語幹にあたるものである。

誤用語別の誤用出現頻度を統計したもので、出現回数の延べ数と同一語の重複を数えない実数どちらも表示される。また、統計結果中から調べたい誤用語をクリックすると、同じ検索条件のまま検索した結果ページのリンク先に移動する。

● 「学習時間数別誤用出現頻度の統計結果」：

学習時間数別の誤用出現頻度を統計したものである。また、統計結果中から調べたい学習時間数をクリックすると、同じ検索条件のまま検索した結果ページのリンク先に移動する。

● 「誤用タイプ別誤用出現頻度の統計結果」：

誤用タイプ別の誤用出現頻度を統計したものである。また、統計結果中から調べたい誤用タイプをクリックすると、同じ検索条件のまま検索した結果ページのリンク先に移動する。

図 6. 出現頻度の統計結果

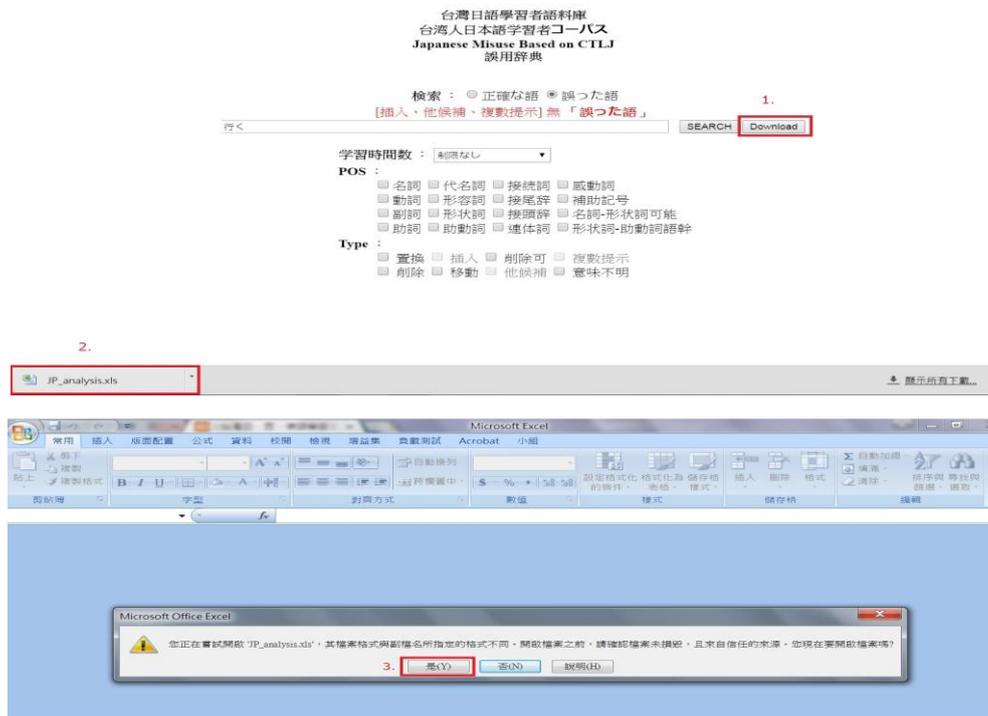
(左から順に、誤用タイプ別・学習時間数別・誤用語別・誤用出現頻度である)

類型	頻率(回数)	学習時間数(等級)	頻率(回数)	結果(正確な語)	頻率(回数)	結果断詞(補助記号)	頻率(回数)
置換	67	900-1200hr (5)	24	total: 78 total(not repeat): 47		、(補助記号)	2
挿入	9	1500-1800hr (7)	20	行く	20	補助記号 total: 2	
複数提示	2	600-900hr (4)	11	行くと	10	類型	頻率(回数)
		1200-1500hr (6)	10	行くこと	2	置換	1
		300-600hr (3)	7	行くことに	2	挿入	1
		1800-2100hr (8)	4	行くか	2		
		2100-2400hr (9)	1	行く時	2	結果断詞(形状詞)	頻率(回数)
		2400-2700hr (10)	1	に行く	1	ように(形状詞)	1
				遊びに行くことになった	1	形状詞 total: 1	
				遊びに行く	1	類型	頻率(回数)
				行くたびに	1	置換	1
				行くことにしています	1		
				行くので	1	結果断詞(補助記号)	頻率(回数)
				行くにも	1	、(補助記号)	1
				行く内に	1	補助記号 total: 1	
				行くことになりました	1	類型	頻率(回数)
				行く時でも	1	挿入	1
				に乗せられたまま行くと	1		
				行くには	1	結果断詞(名詞)	頻率(回数)
				だから、日本へ行く	1	こと(名詞)	9
				に行く時	1	時(名詞)	5
				行くようです	1	の(名詞)	3
				買いに行く	1	そう(名詞)	1
						近く(名詞)	1

#### (4) ダウンロード機能

条件を指定した後「Download」ボタンを押すと、検索結果をそのままダウンロードすることができる。図7に示した1-3のようなステップに従って実行すれば、ダウンロードができる。(図7参照)

図7. ダウンロード機能



検索結果ページには指定した検索条件が表示される。結果ページは学習時間数・誤用・結果・タイプ・訂正例の5項目に分けられる。なお、オンライン辞書のアイコンをクリックすると、Weblio辞書またはYAHOO!辞書を利用することができる。(図8参照)

図 8. 結果ページ

台湾日語教育者辞典  
台湾日本語教育者辞典  
Japanese Misuse Based on CTLJ  
誤用辞典  
 忘用誤等四字辞 **行く**  
検索類型: 正確単語  
語性: 動詞  
業種: 無指定  
学習時数: 千語以内  
**weblio** 日中辞書 **YAHOO!** 辞書 JAPAN  
辞書読み可漢字辞典  
登録結果如下

誤用 種類 数	誤用	結果	类别	正用
3	歩いて行くつと、目をそめて、ずつと上陸になって、たぐさん(船)が気付いた	行く内 行く(動詞) 内(名詞) に(助詞)	置換	歩いて行く内(内)を振り回していたようで、上陸してたぐさん(船)が気付いていることに気付いた
3	近くに地区はバス乗り場	に行く時 に(助詞) 行く(動詞) 時(名詞)	挿入	近くの地区に行く時はバスに乗ります
1	台湾と日本の違うところ日本のドラマについて、日本の学生は、学校の後で友達と遊びに行	行くそうです 行く(動詞)	置換	台湾と日本の違うところ日本のドラマによると、日本の学生は、学校の後で友達と遊びに行

誤用項目中の文をクリックすると原文の全文を閲覧することができる。置換の誤用項目中の文は黄色でハイライトされ、誤用部分は赤文字で示され、赤線で訂正されている。訂正後の内容はその直後にゴシック体の青文字で表示される。削除可、削除、意味不明の誤用項目中の文は黄色でハイライトされ、添削箇所は赤い線で引かれている。移動およびその他の誤用項目中の文は黄色でハイライトされている。(図 9 参照)

図 9. 原文の閲覧

原文如下

私の高校で時代、私は私立高校を勉強するのだ。毎日午前7時の時に、私の駅に行くスクールバスは学校に行くスクールバスは私の家の近くを通る。私の家から学校まで大体20分を要する。私の高校の生活はとても厳格で、毎晩そのまま残って自習する。毎日の試験は多いです。いつも晩にとても遅い才能に全て要って眠る。しかし私と高校の学友の感情はとてもよくて、私達は晩にみんながいつもいっしょに読書するためだ。高校の1学年の時に、私は学校の社会団体活動に参加する。私は中国の民族音楽の社に入った。私はそこに更に多くの友達を知っていた。いっしょに楽器を学ぶのがとても楽しい。高校二年の時に、学業が多いため、私は中国の民族音楽の社を脱退した。私は一心不乱の読書を図書館に行って読書する。高校二年の夏休み、私は学友と一緒に米国に行った。父がこれが1つのとてもよい機会だと思うため、私に米国に行って遊学させる、更に多くの知識を学ぶ。高校三年に着いて、私は大学の入学試験の問題に直面する。毎日すべて寮の中で本を読む。最後に私は試験して大学に着いた。両親はスベテともうれしく思う。ふだん暇がある時、私は私の高校の学友に電話をかけることができる。高校の生活は私に忘れ難くならせる。

### 3.2 CTLJ 誤用辞典の動作環境

各ブラウザに対応している。ただし、ウェブサイトのページエンコードは Unicode に設定する必要がある。

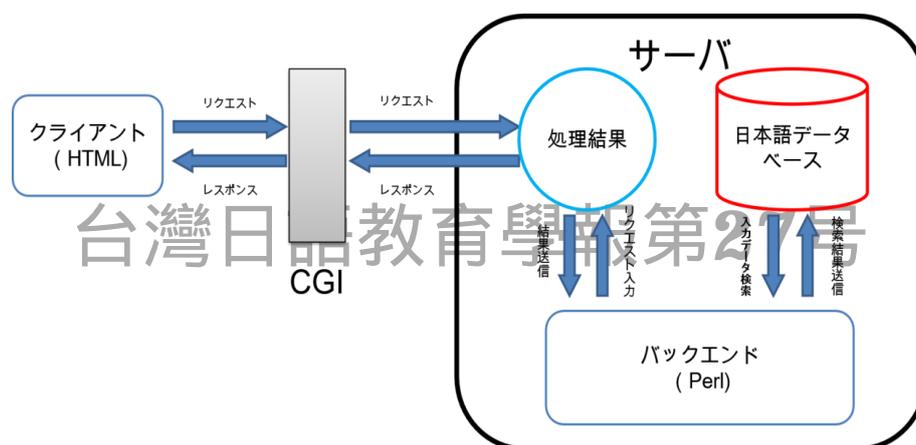
### 3.3 CTLJ 誤用辞典の動作原理

当システムは主に「クライアント」と「サーバ」に分けられる。

サーバ側は Perl を利用して情報の処理・作成を行い、クライアント側は HTML で情報の送信・表示を行う。クライアントとサーバ間は、「Common Gateway Interface(コモン・ゲートウェイ・インターフェース、CGI)」という仕組みを介して連携している。

利用者（クライアント側）がウェブサイトアクセスし検索をすると、その要求が CGI を経由してサーバ側のバックエンドに送信される。すると、バックエンドに置かれたデータベースから検索結果が呼び出され、サーバは再度 CGI を介してその情報をクライアントに返送し、最終的にはクライアント側に検索結果として表示される。（図 10 参照）

図 10. システムの動作仕様



#### 4. CTLJ 誤用辞典における品詞別の誤用率

##### 4.1 CTLJ 誤用辞典のデータ量

CTLJ における修正タグの数は 53,865 セットである。CTLJ 誤用辞典はこれらの修正タグに囲まれた部分を基に開発された。CTLJ 誤用辞典におけるデータ量は、表 2 に示したように、修正タグにおける学生の誤用を含むセンテンスの延べ数が 53,497 例で、異なり数が 20,233 例である。一方、修正タグにおいて教師の添削に使われた訂正例を含むセンテンスの延べ数が 53,481 例で、異なり数が 20,185 例である。両者を合わせば、CTLJ 誤用辞典におけるセンテンスの延べ数が 106,960 例、異なり数が 40,418 例になる。

表 2. CTLJ 誤用辞典におけるセンテンスの数

	延べ数	異なり数
誤用を含むセンテンス	53,497	20,233
訂正を含むセンテンス	53,481	20,185
合計	106,960	40,418

また、表 3 に示したように、CTLJ 誤用辞典における誤用の延べ形態素数が 73,371 例、訂正例の延べ形態素数が 86,447 例であり、合計 159,818 項目である。誤用と訂正例の異なり形態素はそれぞれ 9,122 例と 6,864 例で、合計して 15,986 項目である。

表 3. CTLJ 誤用辞典における形態素数

	述べ形態素数	異なり形態素数
誤用	73,371	9,122
訂正例	86,447	6,864
合計	159,818	15,986

各学習段階における品詞別の誤用と訂正<sup>12</sup>の合計をまとめて示すと次の表 4 になる。表 4 に示したように、CTLJ 誤用辞典における誤用の品詞について統計した結果、助詞が最も多く、その次が動詞で、名詞、助動詞が 3 番と 4 番となっており、ほかに形容詞、補助記号、副詞、代名詞、形状詞、連体詞、接続詞、接尾辞、接頭辞、感動詞、名詞-形状詞可能、形状詞-助動詞語幹などが続いている。

<sup>12</sup>本稿における誤用と訂正とは、CTLJ における修正タグに囲まれた学生の誤用および教師の添削に使われた訂正例のことである。

表 4. CTLJ 誤用辞典における品詞別の統計結果

	品詞	誤用	訂正例	誤用+訂正 例	CTLJ 原文	誤用率 (%)
1	助詞	22700	26944	49644	123930	18.3
2	動詞	15173	18015	33188	52148	29.1
3	名詞	14084	14831	28915	114148	12.3
4	助動詞	12764	18174	30938	53462	23.9
5	形容詞	2038	1972	4010	9481	21.5
6	補助記号	1628	874	2502	57501	2.8
7	副詞	1623	1552	3175	10953	14.8
8	代名詞	845	815	1660	8804	9.6
9	形状詞	723	1137	1860	6291	11.5
10	連体詞	664	685	1349	4891	13.6
11	接続詞	587	628	1215	3854	15.2
12	接尾辞	324	633	957	2234	14.5
13	接頭辞	128	120	248	1047	12.2
14	感動詞	34	31	65	250	13.6
15	名詞-形状詞可能	33	26	59	333	9.9
16	形状詞-助動詞語幹	23	10	33	67	34.3
	合計	73371	86447	159818	449394	16.3

一方、誤用数の多い助詞、動詞、名詞、助動詞、形容詞、補助記号、副詞、代名詞を例として、各学習段階における誤用と訂正例の分布状況を検証すると、図 11-18 のようになる。補助記号を除けば、いずれも学習時間数<sup>13</sup>において誤用と訂正例の数が最も多い。学習時間数 6 がその次である。ただし、代名詞の誤用は、学習時間数 7

<sup>13</sup> CTLJ における学習時間数の分け方は以下の通りである。

1 : 150 時間以下、2 : 150-300 時間、3 : 300-600 時間、4 : 600-900 時間、5 : 900-1200 時間、6 : 1200-1500 時間、7 : 1500-1800 時間、8 : 1800-2100 時間、9 : 2100-2400 時間、10 : 2400-2700 時間、11 : 2700-3000 時間、12 : 3000 時間。ただし、現在データが入っているのは学習時間数 1 から 10 までである。

が二位である。補助記号は学習時間数 7 において誤用と訂正例の数が最も多く、学習時間数 5 が二位である。

図 11. 助詞の誤用分布状況

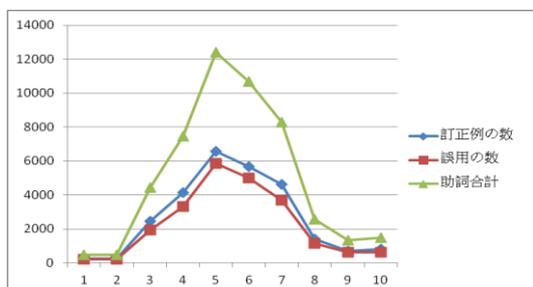


図 12. 動詞の誤用分布状況

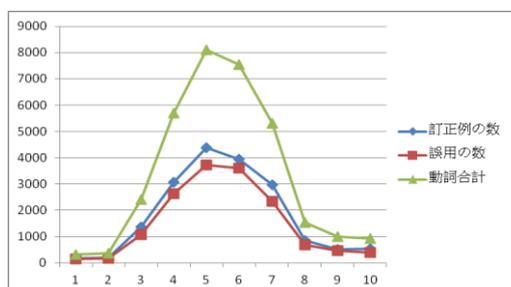


図 13. 名詞の誤用分布状況

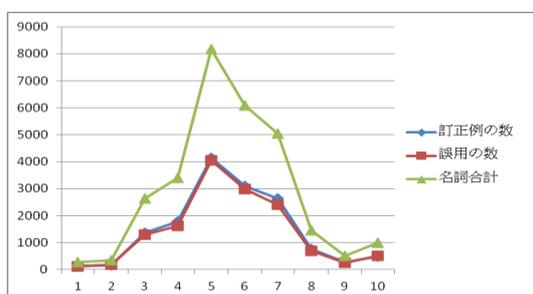


図 14. 助動詞の誤用分布状況

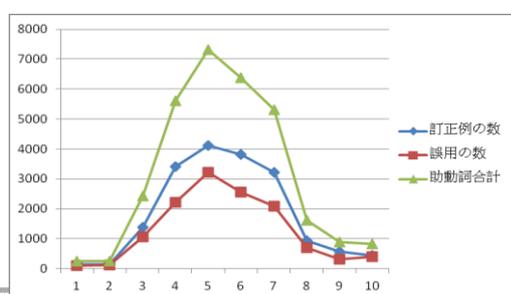


図 15. 形容詞の誤用分布状況

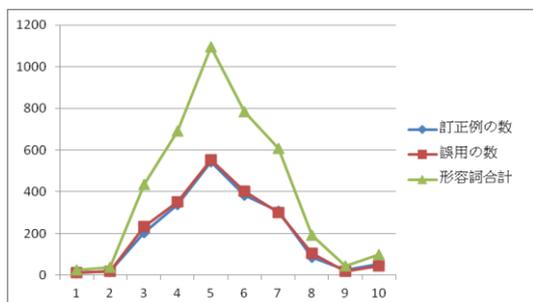


図 16. 補助記号の誤用分布状況

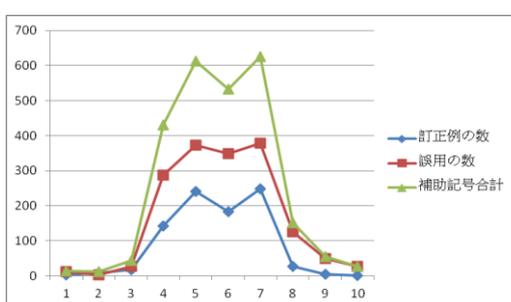


図 17. 副詞の誤用分布状況

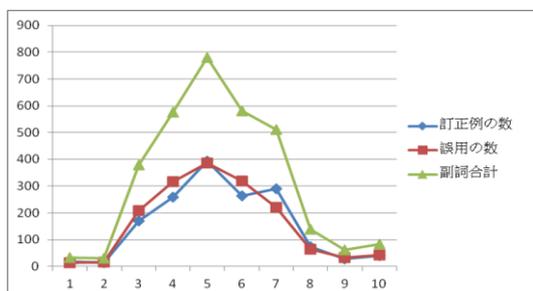
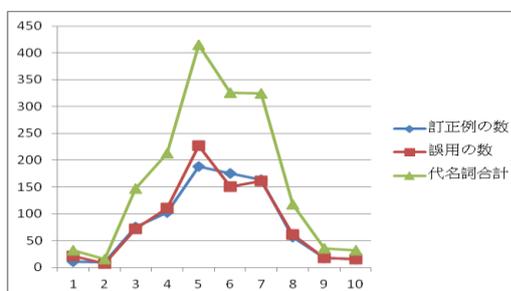


図 18. 代名詞の誤用分布状況



他方、学生の誤用と CTLJ の原文を用いて、誤用率を計算した結果、表 4 に示したように、誤用数の多い助詞、動詞、名詞、助動詞、

形容詞、補助記号、副詞、代名詞において、動詞の誤用率が 29.1%で、最も高い。助動詞が 23.9%で、その次である。形容詞が 21.5%、助詞が 18.3%で、三番と四番となっている。他に副詞が 14.8%、名詞が 12.3%、代名詞が 9.6%、補助記号が 2.8%となっている。なお、全体の誤用率としては 16.3%である。

#### 4.2 品詞別における各学習段階の誤用率

品詞別における各学習段階の誤用率を検証するため、図 19-26 に示したように、CTLJ 誤用辞典の統計をとって見たところ、助詞、動詞、助動詞においては、訂正率が誤用率を上回っていることが分かった。また、図 19-26 から学習時間数と誤用の傾向は一定になっておらず、学習時間数が増えていくにつれ誤用が減少するのは学習時間数の一部に限って見られる。学習時間数が増えていくにつれ誤用が下がっていることが観察できるのは、例えば、助詞の学習時間数 1~4、形容詞の学習時間数 2~4、補助記号の学習時間数 8~10、副詞の学習時間数 4~6、代名詞の学習時間数 1~4 と 8~10 である。

一方、各品詞における誤用率の高い学習時間数を見てみると、助詞においては学習時間数 8、動詞においては学習時間数 5、7、8 と 10、名詞においては学習時間数 2 と 10、助動詞においては学習時間数 7、8 と 10、形容詞においては学習時間数 2、7、と 8、補助記号においては学習時間数 7 と 8、副詞においては学習時間数 3 と 4、そして代名詞においては学習時間数 1 である。

これらの段階も訂正率と誤用率の合計が高い学習時間数と重なっているところが多い。例えば、助詞における学習時間数 8、動詞における学習時間数 7、名詞における学習時間数 10、助動詞における学習時間数 8、形容詞における学習時間数 2 と 7、補助記号における学習時間数 7、代名詞における学習時間数 1 である。また、ほかの品詞における訂正率と誤用率の合計が高い学習時間数として、副詞は学習時間数 7 である。

図 19. 助詞の訂正/誤用率

図 20. 動詞の訂正/誤用率

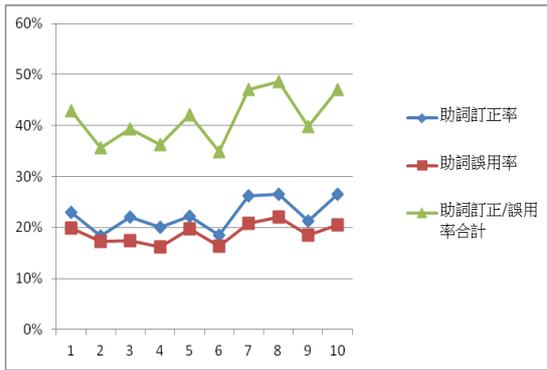


図 21. 名詞の訂正/誤用率

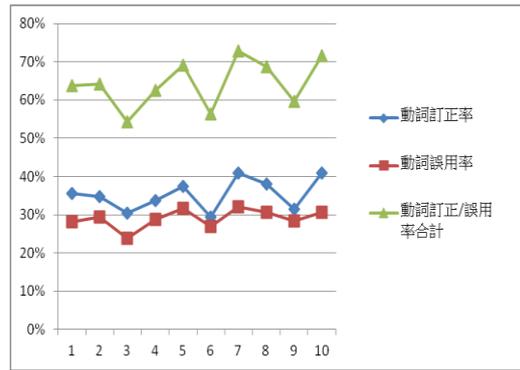


図 22. 助動詞の訂正/誤用率

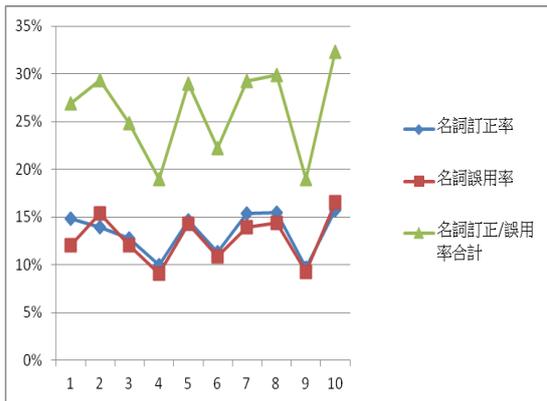


図 23. 形容詞の訂正/誤用率

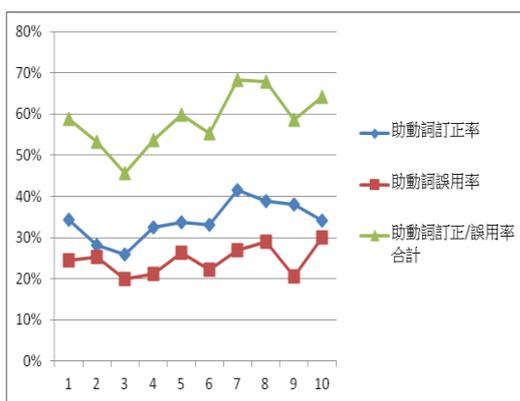


図 24. 補助記号の訂正/誤用率

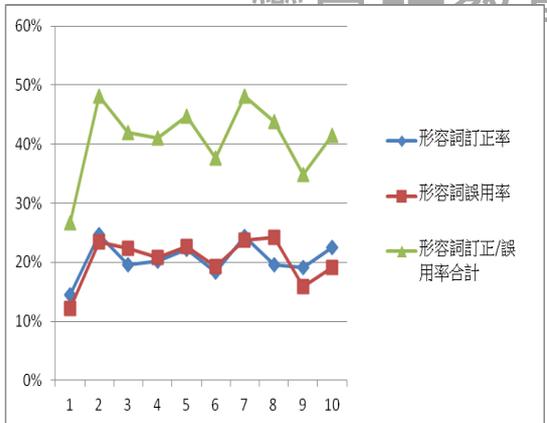


図 25. 副詞の訂正/誤用率

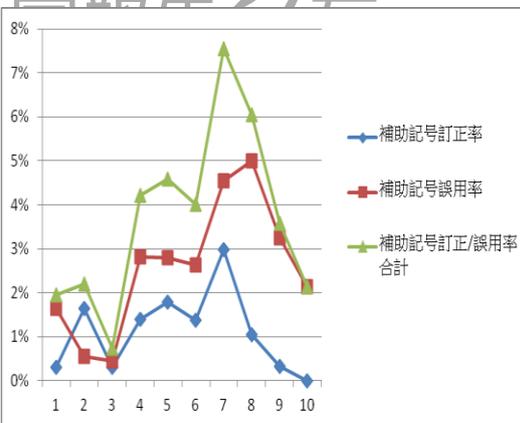
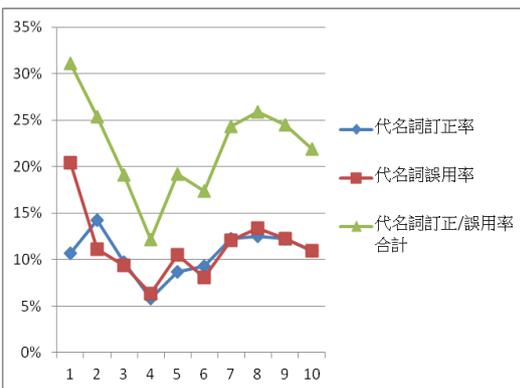
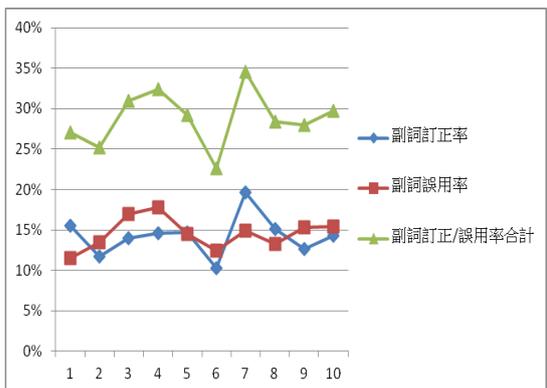


図 26. 代名詞の訂正/誤用率

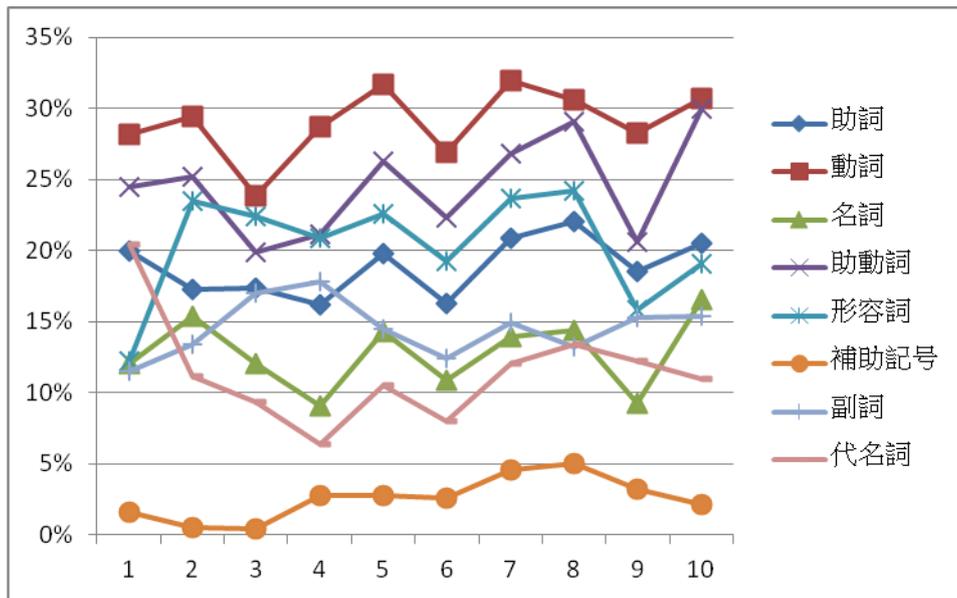


上記した品詞別における各学習段階の誤用率を表5のようにまとめることができた。また、各学習段階における品詞別の誤用率を図で示すと図27のようになった。各学習時間数における誤用がどのようになっているか図27で一目瞭然になる。学習者の中間言語の発達過程について、小柳・峯（2016：147）はKellerman（1985）が主張しているU字型の発達曲線を引用し、時には発達が後退したかのような時期もあり、多くの時間を要する複雑で累加的な発達過程をたどると言っているように、学習時間数と誤用の傾向は一定であるとは言えない。各学習時間数における誤用状況はそれぞれ異なるが、動詞の誤用率が高くなっているのは共通していることである。よって、動詞は学習者にとって習得が難しいことが窺われる。

表5. 品詞別・学習時間数別における誤用率

学習時間数	助詞	動詞	名詞	助動詞	形容詞	補助記号	副詞	代名詞
1	20%	28%	12%	24%	12%	2%	11%	20%
2	17%	29%	15%	25%	23%	1%	13%	11%
3	17%	24%	12%	20%	22%	0%	17%	9%
4	16%	29%	9%	21%	21%	3%	18%	6%
5	20%	32%	14%	26%	23%	3%	14%	11%
6	16%	27%	11%	22%	19%	3%	12%	8%
7	21%	32%	14%	27%	24%	5%	15%	12%
8	22%	31%	14%	29%	24%	5%	13%	13%
9	19%	28%	9%	21%	16%	3%	15%	12%
10	21%	31%	17%	30%	19%	2%	15%	11%

図 27. 各学習段階における品詞別の誤用率



先行研究で述べたように、森山(2015)では「日本語習熟度が高くなるにつれて文脈の中で和語動詞を正確に使用できる傾向にある」としているが、図 27 の誤用状況によると、学習時間数が高くなるにつれ、誤用率が下がっていくという傾向は見られない。動詞の種類や用法によって誤用の傾向も違ってくる可能性があり、学習時間数別のデータを出力して細やかに検証する必要がある。

例えば、陳(2016)の研究によれば、「テイル」の五つの用法の習得に関して、「動作の持続」と「単なる状態」の2つの用法が最も習得されやすく、「結果の状態」と「繰り返し」の2つの用法が同程度で続き、「パーフェクト」の用法が最も習得されにくいという結論を出した。また、「テイル形」の習得では「テイルの用法」よりも「動詞タイプ」から強い影響を受けていることが分かった。一方、「テイタ形」の習得に関しては、「動詞タイプ」と「日本語学習期間」の違いは学習者の「テイタ形」の習得に影響しないという結論を出している(陳 2016: 327-328)。従って、学習時間数と誤用との関係は、学習時間数が高くなるにつれ、誤用率が下がっていくという言い方に対しては慎重な検証が必要だと思われる。習得過程に関して、小柳・峯(2016: 142)は特定の文法カテゴリーに焦点を絞った、記述

的な習得研究や誤用研究も重要な研究であると述べたうえ、その研究を行うために人間の認知的な側面に目を向け、言語処理のプロセスから見た発達段階の精緻化を図っていくことが必要だと主張している。

いずれにせよ、誤用に関する研究を進める際、客観的な研究結果を出すためには、検索機能が備わった大規模データに基づいて検証を行う必要がある。CTLJ 誤用辞典の完成により、日本語の誤用研究や教育資源の共用に大きく貢献できることを望んでいる。また、特定の品詞や学習時間数に関する研究を行う際、本稿の分析結果が参考になれば、幸いである。

### 4.3 まとめ

CTLJ における修正タグの数は 53,865 セットである。CTLJ 誤用辞典はこれらの修正タグに囲まれた部分を基に開発された。CTLJ 誤用辞典における誤用を含む文が延べ 53,497 例、誤用の延べ形態素が 73,371 例、異なり形態素が 9,122 例である。CTLJ 誤用辞典における品詞の統計結果は、助詞が最も多く、その次が動詞で、名詞、助動詞が 3 番と 4 番となっており、ほかに形容詞、補助記号、副詞、代名詞などが続いている。

学生の誤用と CTLJ の原文を用いて、誤用率を計算した結果、動詞の誤用率が 29.1% で、最も高い。助動詞が 23.9% で、その次である。形容詞が 21.5%、助詞が 18.3% で、三番と四番となっている。他に副詞が 14.8%、名詞が 12.3%、代名詞が 9.6%、補助記号が 2.8% となっている。なお、全体の誤用率としては 16.3% である。学習時間数と誤用の傾向は一定になっておらず、学習時間数が増えていくにつれ誤用が減少するのは学習時間数の一部に限って見られる。また、各学習時間数における誤用状況はそれぞれ異なるが、動詞の誤用率が高くなっているのは共通しており、動詞は学習者にとって習得が難しいことが窺われる。動詞の誤用を減すことは日本語教育において優先に対処すべき課題だと思われる。

## 5. おわりに

CTLJ 誤用辞典における検索機能は、キーワード、学習時間数、品詞、誤用タイプをカバーしており、利用者は検索結果をエクセルファイルで保存することも可能である。本稿では CTLJ 誤用辞典の開発とその機能、および上記データ中における各学習段階の誤用率に関する分析結果について報告した。各学習段階における誤用の傾向および詳しい分析は今後の課題である。

誤用を客観的に検証するには、検索機能が完備した大規模データが不可欠である。CTLJ 誤用辞典の完成により、日本語教育、日本語研究あるいは中国語との対照分析に役立てることを期待している。

### <付記>

本稿はバリ島で行われた ICJLE2016 におけるポスター発表した後、大幅に修正・加筆したものである。

## <謝辞> 台湾日本語教育學報第27号

本稿におけるシステムの開発は成功大学情報工学系盧文祥先生、およびその研究室の大学院生侯凱仁氏、黃彥軒氏、周洵氏、洪千越氏、鄭宇軒氏等の協力を受けており、ここで深く感謝の意を表します。システムのメンテナンスも同研究室の協力を受けています。また、CTLJ 誤用辞典の編集や校正は主に研究助手の坪内由里佳さんと蕭凡鈞さんに協力していただきました。そして、誤用例と訂正例の校正はほかに吉本宇佑さんと宮平百々子さんにも大変お世話になりました。

### 参考文献

伊集院郁子(2016)「学習者要因の分析②—コーパスに基づく研究—」

徐敏民・近藤安月子(編)『日語教学研究』北京:外語教学与研究出版社, 385-406.

岡田美穂・林田実(2016)「中国語を母語とする中級レベルの日本語

- 学習者の移動先を表す「に」と動作場所を表す「で」の習得」  
『日本語教育』163, 東京：日本語教育学会, 48-63.
- 黄淑妙 (2016) 「台湾人日本語学習者コーパス誤用辞典の開発」  
『International Conference on Japanese Language Education  
(ICJLE) 2016』バリ島：BNDCC.
- 黄淑妙 (2009) 『日本語習得の達成度分析—「台湾人日本語学習者  
コーパス」(CTLJ)の構築と分析を中心に—』台北：致良出版社
- 小柳かおる・峯布由紀 (2016) 『認知的アプローチから見た第二言  
語習得—日本語の文法習得と教室指導の効果—』東京：くろし  
お出版
- 坂口昌子 (2004) 「日本語学習者が生成する格助詞「が」・「を」の誤  
用とその修正について—作文データからみた母語別誤用傾向  
—」『研究論叢』63, 京都：京都外国語大学, 65-75.
- 迫田久美子 (2001) 「学習者の誤用を生み出す言語処理のストラテ  
ジー(1)—場所を表す「に」と「で」の場合—」『広島大学日本  
語教育研究』11, 広島：広島大学, 17-22.
- 島田かおり (2011) 「日本語学習者の誤用分析—中国語母語話者の事  
例研究」『山口国文』34, 山口大学人文学部国語国文学会,  
96-106.
- 陳建璋 (2016) 『台湾人日本語学習者による日本語のAspect形式  
「テイル」の習得について』名古屋大学大学院国際言語文化研  
究科日本語文化専攻博士学位論文
- 塚脇幸代 (2012) 「“名詞らしさ”と品詞の概念」言語処理学会第18  
回年次大会発表論文集  
[http://www.anlp.jp/proceedings/annual\\_meeting/2012/pdf\\_dir/P1-1.pdf](http://www.anlp.jp/proceedings/annual_meeting/2012/pdf_dir/P1-1.pdf) (閲覧日：2016年8月7日)
- 前川喜久雄監修 (2014) 『講座日本語コーパス2.書き言葉コーパス—  
設計と構築—』東京：朝倉書店
- 森山仁美 (2015) 「文脈における和語動詞語彙の産出」『日本語教育』  
161, 東京：日本語教育学会, 2-14.

若生正和 (2012) 「留学生の使用する日本語に見られる諸特徴：中間言語分析の観点から」『大阪教育大学紀要 第 I 部門 人文科学』61(1), 大阪：大阪教育大学, 35-40.

Kellerman, E. (1985). If at first you do succeed. In S. M. Gass, & C. Madden (Eds.), *Input in second language acquisition*. Rowley, MA: Newbury House. 345-355.

## 台灣日語教育學報第27号