

コーパス分析システム Co-Chu における タグ検索機能とその活用

——誤用や話し言葉にどのように対応するか——

山本裕子*・本間 妙**・川村よし子***

1. はじめに

日本語教師や日本語研究者が分析したいテキストは、文法的に正しく、形の整っている文章だけではない。学習者の作文のように誤用の含まれた文章や言い誤りや方言等を含んだ会話文も含まれているため、形態素解析がうまくいかない場合も多い。しかも、既存の学習者コーパスや会話コーパスではなく、独自のコーパスを用いて分析したいこともある。そこで、筆者らは、教師自らが集めた学習者の作文や言い誤りを含む会話文などのデータもコーパスとして分析できるツールをめざして、コーパス分析システム Co-Chu（以下、Co-Chu とする）の開発を進めてきた。

Co-Chu では、「タグ付け」をすることによって、こうしたデータに対応する。本稿では、Co-Chu でのタグ付け方法の詳細について述べるとともに、運用実験として行ったタグ検索機能を活用した話し言葉と日本語学習者の作文分析の結果を報告する。さらに、これらの実験結果をもとに、Co-Chu を日本語研究や日本語教育にどのように活用できるかについての提言を行う。

2. コーパス分析システム Co-Chu の概要

テキスト分析を適切に行うには、基礎データ、形態素解析、正確な形態素解析を可能にするタグ付けの3つを統合する必要がある。コーパス分析システム Co-Chu は、日本語テキスト分析のためのウェブアプリケーションであり、コーパス作成から分析までの【Build】

*人文学部非常勤講師・愛知淑徳大学交流文化学部 教授—日本語学・日本語教育学

**人文学部非常勤講師—日本語学・日本語教育学

***東京国際大学言語コミュニケーション学部 教授—日本語教育学・教育工学

【Import】【Edit】【Analyze】の4機能を1つのインターフェイスで使えるようにデザインされている（山本他 2018, Komori, et al. 2019 等）。

Co-Chu を使用するには、まず【Build】機能で新たに作成するコーパスの階層構造を決める。次に、【Import】機能で CSV やテキストファイルで作成したデータをシステムに取り込む。会話の文字化データのように発話者を明示する必要がある場合には、CSV ファイルを用いて図1のように発話者を示しておく。発話者の属性（母語、性別、年齢等）はデータを取り込む際に発話者のメタ情報（Speaker metadata）として入力することができる。

発話者	発話内容
NS_A	あ、そろそろごはん食べる↑
NNS_B	そうだね。

図1 発話者情報の記載例

【Edit】機能では、取り込んだデータに対して、MeCab-UniDic による形態素解析を行う。その解析結果に問題があった場合には、以下で述べるようなタグ付けによって、話し言葉や誤用を含むデータでも適切な形態素解析が行えるようになる。また、話し言葉を文字データにする際、笑い声、イントネーションの変化などを記述するために各種の記号（相槌の挿入などを示す（ ）や、上昇イントネーションを示す↑など）を用いることが多い。Co-Chu ではこうした記号が挿入されたデータを取り込んでも、形態素解析の際、これらの記号および記号内の文字を取り除いて解析することができる。【Edit】機能を用いた一連の編集作業を経て、コーパスを作成した後、【Analyze】機能によって、キーワード検索、コロケーション分析、複数データの比較等の分析を行う。

3. 形態素解析の誤解析への対応

日本語学習者の発話や作文などに限らず、日本語母語話者の話し言葉にも、言い間違い、言い直し、言い淀み、フィラー、縮約形等、不規則なものが多く含まれている。そのため、これらを形態素解析するとしばしば誤解析が生じる。誤解析が起きた場合には、タグ付け等によって対応する。その対応方法を以下に詳しく説明する。

3.1 タグ付けによる対応

収集したデータに含まれる言い間違い、言い直し、フィラー、縮約形等によって誤解析が起きた場合、Co-Chu ではデータにタグ付けをして、正しい語形を示すことで、適切な形態素解析ができるようになる。タグの種類は研究の目的に応じて自由に設定、削除が可能であ

る。例えば、話し言葉であることが誤解析の原因である場合には「話し言葉タグ」、言い間違い等誤用が誤解析の原因である場合には「誤用タグ」のように各種のタグを設定し、それぞれ「話」、「誤」といった形でタグの種別を記す。タグ付けは次のような形で行う。

| 正しい語形 | (タグ種別 実際に用いられた語形)

(例 1) 元の発話：ちょっとやでした。→ 解析結果：「や」は「や (助詞)」

タグ付け：| いや | (話 や) でした → タグ付け後の解析結果：「いや (形状詞)」

(例 2) 元の文：経験がでいた。→ 解析結果：「でいた」は、「で (助動詞)」

「い (動詞居る)」, 「た (助動詞)」

タグ付け：経験が | でき | (誤 でい) た。

→ タグ付け後の解析結果：「でき (動詞)」, 「た (助動詞)」

(例 1) では、元の発話「ちょっとやでした」では、「嫌」が「や」という話し言葉の語形で現れたため、「や」は助詞の「や」として解析されてしまう。そこで、タグによって正しい語形である「いや」を示すことで正しく解析できるようになる。「いや (形状詞)」の「形状詞」とは国文法でいう形容動詞のことである。また、この場合は話し言葉であることが誤解析の原因なので、「話し言葉タグ」(記号は「話」)を付している。(例 2) の元の作文では、「できた」と書くべきところを「でいた」と書いていた。そのため、タグを付して誤解析が生じないようにした。また、この場合は「でいた」は誤りなので、「誤用タグ」(記号は「誤」)を付している。Co-Chu では、このように誤解析が生じた原因に応じてタグの種類を自由に設定できる。

さらに、タグには、次のような形で「タグメモ」を付けることも可能である。

| 正しい語形 | (タグ種別 実際に用いられた語形：タグメモ)

(例 3) 元の発話：それ忘れなく、やっぱり次の子とか

タグ付け：それ忘れ | ないで | (誤 なく：活用)、やっぱり次の子とか

上記の例では、「忘れないで」の活用が誤っていたため、「誤」タグを付け、「タグメモ」には「活用」と記載した。このように、タグメモにどのような誤用だったかなどの情報を記載しておくことで、後の分析がスムーズに行えるようになる。

には問題点が複数の形態素にまたがり、形態素単位では対応できない場合もある。典型的な例は、次のような「ねじれ文」である。

(例 5) ヨガを習った効果は、体の状態がますますよくなっているのである。

(例 5)は、明らかに問題のある文だが、そのことを 3.1 の方法では示せない。こうした場合に、Co-Chu のメタ情報付与機能を活用する。前述したように、Co-Chu では会話のように発話者が誰であるかという情報が必要な場合には、発話者の情報を発話者のメタ情報として記した(図 1 参照)。このメタ情報欄を活用し、当該の文(あるいは発話)に関する情報を図 3 のように記すことが可能である。

メタ情報	対象となる文
ねじれ文	ヨガを習った効果は、体の状態がますます良くなっているのである。
倒置文	そろそろ食べようかな、ごはん。

図 3 文に関する情報の記載例

図 3 には、ねじれ文と倒置文の例を示した。ねじれ文の場合、形態素解析自体に誤りが生じているわけではないので、文中にタグを付すことはできない。そこで、左の列に「ねじれ文」というメタ情報を入れる。メタ情報は誤用に限らず、分析したい内容があれば、適宜付すことができる。例えば倒置文に着目して分析したい場合は、図 3 のように「倒置文」と入れておくことができる。このメタ情報も【Analyze】機能で検索できる。

5. 運用実験

5.1 実験の概要

Co-Chu の【Analyze】機能では、タグやメタ情報も検索できる。検索機能を実際にどのような形で活用できるのかについて、2 種類のコーパスを用いて運用実験を行った。実験 1 は話し言葉を対象としたもの、実験 2 は学習者の作文を対象としたものである。

実験 1

対象 話し言葉コーパス

目的 誤解析が生じやすい話し言葉を分析する際に、どのようにタグを付したらよいかについて検討する。

実験 2

対象 学習者の作文コーパス

目的 学習者が作文の中で犯している誤用を分析する際に、どのようにメタ情報を付したらよいかについて検討する。

なお、実験 1 は山本他 (2019) で、実験 2 は本間他 (2019) で扱った内容がもとになっている。本稿では各々の実験について概要を紹介し、タグ付けの効果や、タグ検索の活用について考察を加える。

5.2 実験 1

5.2.1 概要

【実験用コーパス】

母語話者同士の約 30 分の雑談 3 件、日本語母語話者（以下 NS）と日本語学習者（以下 NNS）の約 30 分の雑談 3 件。録音時間は合計 232 分。総短単位数 38,357。

【実験用タグ】

本実験では、以下の 6 種類のタグを付すこととした。タグの詳細については山本他（2018, 2019）を参照されたい。

- ① 話し言葉のため誤解析が生じている場合のタグ：「話」
- ② 方言のため誤解析が生じている場合のタグ：「方」
- ③ 誤用を示すタグ：「誤」
- ④ 言い淀み⁽¹⁾のため誤解析が生じている場合のタグ：「淀」
- ⑤ 上記以外の事情で誤解析が生じている場合のタグ：「他」
- ⑥ 形態素解析の単語情報に誤りがあることを示すタグ：「単」

今回の実験では、タグは原則として、形態素解析が適切に行われなかった箇所のみを付した。つまり、「話」や「方」等は、全ての「話し言葉」や「方言」に付したわけではなく、形態素解析において誤解析が生じた場合にのみ付した。ただし、助詞の誤用については、誤解析か否かにかかわらず「誤」というタグを付けることとした。助詞は誤用であっても、誤解析は起こらず、解析結果だけからはどこに誤用があるかわからない。そこで助詞に関してのみ、全ての誤用にタグ付けして、どの程度誤用があるかが把握できるようにした。

またタグのうち、「方言」は「話し言葉」のバリエーションの 1 つではあるが、今回の会話データは会話の参加者に愛知県出身者が多く、地域特有の表現の使用が特徴の 1 つとなると考え、「方言タグ」を話し言葉とは区別して設けることにした。

5.2.2 タグの検索結果

Co-Chu のタグ検索では、図 4 の検索項目画面を利用する。



図 4 タグ検索の検索項目画面

実験用タグ付きコーパスを対象に、上述した 6 種類のタグを指定して検索したところ、表 1 の結果が得られた。

表 1 タグ検索の結果

タグ ライン数	①話	②方	③誤	④淀	⑤他	⑥単	合計
5513	153	105	29 (うち助詞 14)	60	28	273	648

表 1 の結果から、実験用コーパスでは、助詞の誤用を除いて、誤解析が 634 箇所が生じていたが、そのうち①～⑤までの 361 箇所については、タグの付与により誤解析を解消することができた。③は活用等の誤り、④も「あ、あ、あのね」のように、正しい語形が産出されていない箇所のため、必然的に誤解析になっていた。そこで、それ以外の①、②、⑤、⑥についてどのような場合に誤解析が生じていたのか順に検索結果を示しながら述べる。まず、①「話」タグの検索結果を次ページの表 2 に示す。

表 2 から、誤解析が生じるのは、「コンサル」のように MeCab の辞書に登録されていないことが原因の場合もあるが、ほとんどが縮約形や長音化、促音化のような音変化が原因であることがわかる。特に本実験で用いたコーパスでは「違う」が「ちゃう」や「ちゃ」のように発音されていたケースが目立った⁽²⁾。

次に②「方」の検索結果を次ページの表 3 に示す。

太字は「方言」を話題にしている方言を説明するために用いられているもの(「おまんな」や「げん」等)である。網掛けで示したように「やん」をはじめとした、助動詞「だ」・「じゃ」が「や」に音変化した類が多く用いられていることがわかる。本来これは関西系の方言であるが、現在は全国的に若者に広く使われており、(例 6) のように「方言」というよりも実際には「若者言葉」に近い役割を果たしている。

表2 「話」タグの検索結果（頻度2回以上のもの）

頻度	用いられた語形	本来の語形	頻度	用いられた語形	本来の語形
14	ちゃ	違う	2	じい	字
13	ちゃう	違う	2	っちゃ	って
10	コンサル	コンサルタント	2	やば	やばい
8	れ	られる	2	ぜってえ	絶対
6	なんで	だ	2	あああー	ああ
5	ね	無い	2	ん	の
4	と	えーと	2	でかかった	大きい
4	あああー	ああ	2	や	嫌
3	うるさ	煩い	2	やつ	否
2	ん	ない	2	たっけー	高い

表3 「方」タグの検索結果（頻度2回以上のもの）

頻度	用いられた語形	本来の語形	頻度	用いられた語形	本来の語形
39	やん	じゃん	2	だが	だ
7	げん	の	2	車校	自動車学校
6	ん	ない	2	やん	だ
6	っち	達	2	あかん	行く
6	や	だ	2	と	てる
6	やん	「じゃん」 ⁽⁴⁾	2	ええで	良い
2	やつ	だ	2	あか	行く
2	おまんな	ござる			

(例6)

[NS_Y] だって、行くところだって、もうないんだよ、もう、あの、その、目的がないじゃん。

[NS_N] うん。

[NS_Y] おいしいものを食べる、だけだったやん。

[NS_N] うん、楽しそうやん。

(例6) では旅行の目的地について日本語母語話者同士が話している。話者の一人、[NS_Y] は「じゃん」も「やん」も用いているが、これら以外に特に方言を用いているわけではない。本実験ではこれらを「方言」とタグ付けしたが、使い方からは「若者言葉」とした方が適切であった可能性もある。このようにタグの区分には検討の余地がある。

次に⑤「他」を見る。「他」タグを付けたものは28例あったが、その中には、言い直し、

言いさしの他、発話が格助詞で始まった場合にその格助詞が誤解析されたという例があった。

(例 7) 言い直し

【元の発話】：っていう病気にかけて、かかってか、になって、で、辞めたんですが。

【タグ付け】：っていう病気 | になって | (他 かけて、かかってか、になって：言い直し)、
で、辞めたんですが。

(例 7) では「(病気) になって」と一度に言えず言い直しをしている。この場合は、「かける」「かかる」と試行錯誤の上、「なる」に単語も変更して言い直しているので、「淀 (言い淀み)」ではないと判断した。

(例 8) 言いさし

【元の発話】：俺、それ知らなかったんだけど、勝手に予約されとっ、間違えた、勝手に予約
してくれて。

【タグ付け】：俺、それ知らなかったんだけど、勝手に予約され | とって | (他 とっ)、間
違えた、勝手に予約してくれて。

(例 9) 言いさし

【元の発話】：いや、いう、それは日本に来てから。

【タグ付け】：いや、 | | (他 いう：言いさし)、それは日本に来てから。

(例 8) (例 9) では、言いかけて途中でやめ、他の内容を言っている。(例 8) では言いさした部分がコンテキストから推測可能であるので、推測した語形を「正しい語形」としてタグ付けした⁽³⁾。しかし (例 9) では「いう」が、実際には何を言おうとしているかは明白ではないが、このままでは「いう」は動詞の「言う」と解析されてしまう。そこで、「正しい語形」は空白にしてタグ付けし、誤解析を解消した。

さらに、次のように発話の冒頭に助詞が来ることによる誤解析の場合もある。

(例 10) 発話冒頭の格助詞

【元の発話】：

[NS_N] ちょっとかけた後に、そのトマト、

[NNS_J] うん。

[NS_N] で、トトマト煮みたいにするってのがすごく好きで。(発話冒頭の「で」の解析結果：
接続詞)

【タグ付け】： | トマト | (他) で、トマト煮みたいにするってのがすごく好きで。

(例10) は趣味の料理について [NS_N] が [NNS_J] に説明している場面である。[NS_N] は [NNS_J] の反応を見ながら話しているため、[NS_N] の発話の間に [NNS_J] の「うん」が挿入されている。こうした場合、3つ目の発話（[NS_N] の2つ目の発話）の冒頭の「で」は接続詞と解析される。しかしこれは1つ目の発話末の「トマト」に続く「で」なので、格助詞と解析されるべきである。この場合、実際には発話されていない「トマト」を補うことで、正しく解析できるようになるが、これは発話されていない単語なので「他」タグで示すことにした。発話の冒頭に助詞が来る現象は、近年若者を中心に広がっていることが指摘されている（西阪2008, 有田2015 他）が、本実験データでも類似の現象が多く見られた。冒頭に来るのが「は」や「に」であれば、ほとんどの場合、問題なく「助詞」と解析される。しかし、「で」や「が」の場合は、接続詞と解析される場合があるので、注意が必要だと言える。

最後に⑥の「単」タグは、そのほとんど（273のうち252例）が、(例4)のように「感動詞」を「連体詞」と誤解析した誤りであった。こうした品詞の間違いの他、「清水寺」の読みを「しみず/てら⁽⁵⁾」とするような漢字の読み方にかかわる誤りもあった。これも、当該単語あるいは読みが MeCab の辞書に登録されていないことに起因している。

以上、検索結果から、誤解析の生じやすい箇所について述べた。誤解析の要因には、MeCab の辞書を拡充すれば解決するものもあるが、音変化は常に同じ形になるわけではないのでその都度タグ付けをするしかなかろう。一方、「単」タグの大半を占める、「感動詞」が「連体詞」として解析されてしまうケースのように頻度の高いものについては、効率よくタグ付けを行う方法を今後検討していきたい。また、(例6)で述べた「方」タグの妥当性や、(例7)の「淀」と「他」の区別、(例8)(例9)で「正しい語形」として何を入れるかという判断は多分に主観的な側面がある。タグ付けにあたっては、どのようにタグを設定するか、どのタグを付すかの基準を十分に検討する必要がある。

5.2.3 タグ検索機能の活用1：誤用タグ

タグ検索機能では、タグの付いた形態素が一覧表示される。そこで、この機能を活用して、どのような誤用があったかを詳しく見ていくことにする。図5のように該当する本文に加えて、出力項目を指定することによって、タグメモに記載した誤用の内容や、発話者（話

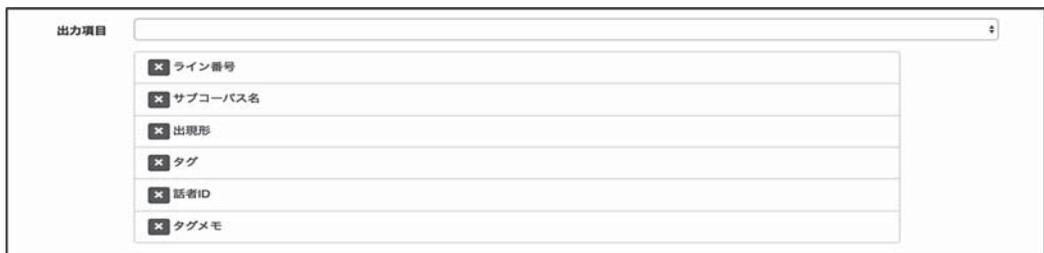


図5 出力項目選択

を、検索条件に加えて検索することができる。ここでは、タグ検索だけでなく話者に関するメタ情報も指定して検索を実行した。図7と図8はそれぞれの検索条件の指定画面である。



図7 話者に関するメタ情報（母語）の指定

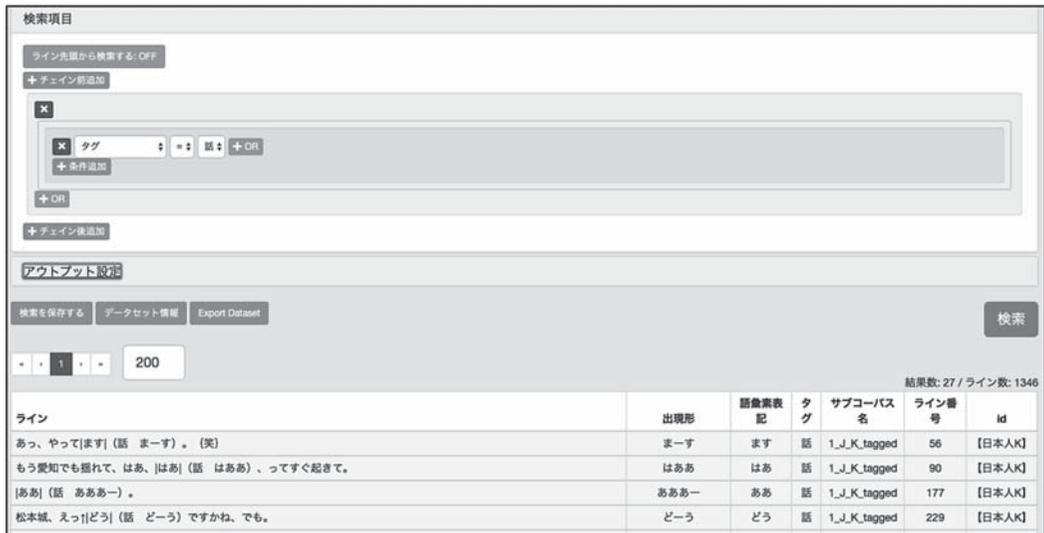


図8 「話」タグを指定した場合の検索

まず、NSとNNSの接触場面の会話⁽⁶⁾において、NS、NNSの発話にそれぞれどのくらいタグが付されたかを検索したところ、表5の結果が得られた⁽⁷⁾。

表5 接触場面でのタグ検索結果の比較

	①話	②方	③誤	④淀	⑤他	⑥単	計
NNS ライン数 1424	17	31	24	28	14	175	289
NS ライン数 1346	27	13	1	3	6	72	122
計	44	44	25	31	20	247	411

表5から「誤」「淀」「単」がNNSに多いことがわかる。この結果は、NNSの一般的な傾向と合致する。一方、「方」もNNSに多かったが、これは今回分析した会話では「方言」が話題になっていたため、NNSが方言の例をたくさん挙げていたことによる。なお、「他」もNNSに多かったが「他」には多様なものが含まれているので、ここでは除外して考える。

また、話者 ID を指定して検索し、一人の話者が相手によってどのように話し方を変容させているかを見ることもできる。そこでここでは特定の接触場面、母語場面のどちらにも参加した [NS_N] に注目し、それぞれにおけるタグの検索結果を比較した。

表 6 同一 NS の場面ごとの比較 ([NS_N] を指定して)

ライン数		①話	②方	③誤	④淀	⑤他	⑥単	計
対 NNS	1346	27	13	1	3	6	72	122
対 NS	1379	61	36	1	7	1	7	113

表 6 から、この話者が、NNS に対してわかりやすいように話そうとしていることが窺える。①「話」や②「方」が NNS に対して少なく、⑥「単」が多い。⑥「単」が多いということは、フィルターの「あの」「その」が多いということであり、考えながら話していると思われる。一方、NS に対しては①「話」や②「方」が非常に多い。ここでは全ての話し言葉や方言にタグを付けているわけではなく、誤解析の生じた箇所のみタグを付けている。5.2.2 で述べたように、「話」「方」タグが付くものは音変化のあるものが中心である。したがって、「話」「方」タグが多いということは、正確な解析がされにくい、くだけた話し方をしていることを示していると言える。

5.2.5 実験 1 のまとめ

以上、実験 1 では 6 種類のタグを設けて運用実験を行った。タグの付された箇所を分析した結果、話し言葉の場合、MeCab 辞書にない単語や、音変化の生じている箇所に誤解析が生じる傾向があることがわかった。これらに対応するには、MeCab 辞書の拡充が可能なシステムにするとともに、同種の誤解析が多く生じる場合、効率的にタグを付ける方法も検討する必要がある。ただし、音変化や言い淀みなど、常に同じ形を生じるものではないものは、その都度タグ付けするしかない。また、システムの問題ではないが、タグを付す際には、どのようなタグを設けるか、またどのようにタグを区別するかを予め明確に決定しておく必要があることは言うまでもない。

本実験では、基本的に誤解析が生じる箇所のみタグを付したが、それだけでは言語使用の実態を捉えるには十分ではないことも判明した。例えば「言い直し」は、必ずしも誤解析になるわけではない。

(例 13) 内定してくれた、なんていう、言ってくれ、まだもらってないんですけど…

下線部は「言い直し」ているが、「してくれた」「なんていう」「言ってくれ」は全て正しく解析されているので、本実験ではタグを付す対象にはしていない。こうしたものについては

解析の正誤にかかわらず、研究の目的に応じてタグを付していくことで、タグをより有効に活用できるだろう。

5.3 実験 2

5.3.1 概要

【実験コーパス】

留学生の大学学部1年生5名，大学院生2名による25件（472文）の作文コーパス。

【実験用タグ】

本実験では、学習者の誤用の全体像を、特に動詞の使用に注目して捉えるために、実験1を踏まえ、誤解析の生じる箇所のみならずタグ付けするのではなく、作文中の全ての動詞についてタグ付けを行った。タグ付けは、作文に用いられた全ての動詞に、自動詞・他動詞を区別して「自」タグか「他」タグを付し、それが正用か誤用かを記載することにした。動詞の使用が正用か誤用かを判断するには、付随する助詞、助動詞、時制など複数の要素との関わりを文レベルで見る必要がある。そのため、これら付随する要素を含めて誤用の場合に「誤」タグを付し、タグメモにどのような誤用かの情報を記載した⁽⁸⁾。以下が動詞のタグ付け方法である。

| 正しい語形 | (自/他タグ 実際に用いられた語形：タグメモ)

ただし、関わる要素が多かったり、関わる要素の誤用の種類が多い場合、タグメモには複数の情報を入れることになる。そこで、ここでは、文レベルの誤用への対応として、メタ情報を5種類に区別し、メタ情報欄に記載した⁽⁹⁾。5種類のメタ情報は次のように区分した。

- ・要素1：動詞の活用に関わるもの（受身，可能，使役，時制他）
- ・要素2：動詞に後続するもの（ている，である他，たい，てほしい他，ば，たら他，やすい，すぎる他）
- ・要素3：助詞の誤用（助詞脱落・過剰使用，「に」←「を」⁽¹⁰⁾ etc.）
- ・その他：語選択，表記，動詞脱落他
- ・コメント：動詞の正誤に関わらないもの（有対自他動詞，ねじれ文他）

これらに関して、正誤を判断した上でメタ情報として記載した。図9は、5種類のメタ情報の記載例である。

要素1	要素2	要素3	その他	コメント
			対あり	私は2017年7月7日 (誤 : 語釈) に 奇縁の (誤 キセキノ : 表記) 手羽先という居酒屋でアルバイトを 始め (他 始め : 正, 対あり) た。
(中期)				
	てくれる○			研修中でも 楽しく (誤 楽しく : 活用) 教え (他 教え : 正) てくれる のが (誤 : 助詞脱落) 特徴だった。
	でもらう○、ている×			アルバイト研修の1日目、最初は皆 で (誤 と : 助詞) 自己紹介を し (他 し : 正)、店長に制服の着替えについて 教え (他 教え : 誤, アスペクト不要) てもらっ た (誤 てた : アスペクト不要)。
	てくれる×			カード行商、手洗いやお店の色々 な (誤 : 活用) ルールなども丁寧に 教え (他 教え : 誤, アスペクト) て もらっ (誤 くださ : アスペクト) た。
可能○				漢字が 認め (自 認め : 正) なくてメニューを 覚える (他 覚える : 正) の が (誤 : 助詞脱落) 非常に大変だった。
可能○	でもらう○	助詞脱落、助詞脱落、助詞をーが		メニュー を (誤 : 助詞脱落) 1週間ぐらいで 覚え (他 覚え : 誤, 助詞脱落) てから、 食べ (誤 食べ : 表記) の 作り方 を (誤 : 助詞脱落) 覚え (他 覚え : 誤, 助詞脱落) てもらい、研修に接客の仕事 が (自 有 : 助詞) できる (自 できる : 誤, 助詞) ようになった。最初は注文を 取る (他 取る : 正) とときや器物を 作る (他 作る : 正) ととき何回も 間違っ (他 間違っ : 正) たことも あっ (誤 ある : 時制) た。
	でもらう○、てしまう○			一か月ぐらいで、レジメ 覚える (他 覚える : 正) ことになったが、 覚える (他 覚える : 正) ことが多くて、前 教 (他 教え : 正) てもらったことを 忘れ (他 忘れ : 正) て まう (他 まう : 正) こともあった。
				ようやく一か月半で研修終了に なっ (自 なっ : 正) た。
	ば○、ていく○		対あり	研修中 に (誤 で : 助詞) 私 は (誤 が : 助詞) 覚え (他 覚え : 正) れば、 覚える (他 覚える : 正) ほど、 覚える (他 覚える : 正) ことが急速に 増え (自 増え : 正, 対あり) ていくことに 気づき (自 気づき : 正) ました。
可能○可能○、可能○	ている○	助詞脱落	話し言葉	しかし、研修中のとき 楽しく (誤 楽しく : 活用) 感じ (他 感じ : 誤, 話し言葉) て い (誤 てな : 話し言葉) かった私 は (誤 : 誤用)、なんでも でき (自 できる : 正) ようになってから、アルバイト先の仲間と色々 話 (自 話 : 正) て 楽しく (誤 楽しく : 誤用) アルバイト が (誤 : 助詞脱落) でき (自 でき : 誤, 助詞脱落) た と思 (自 思う : 正) う。
可能○	ている○		話し言葉	お客様とお話し する (自 する : 正) ことも多くあり、お客様とも仲良く でき (自 でき : 誤, 話し言葉) て い (誤 て : 話し言葉) た。
	たら○、たい○			奇縁の (誤 キセキノ : 表記) 手羽先は自分の有名なお店であるし、手羽先の後も7つ ある (自 ある : 正) から大分に 行く (自 いく : 正) ことに なっ (自 なっ : 正) たらぜひ一度 食べ (他 食べ : 正) たいことをおすす し (他 し : 正) たい と思 (自 思う : 正) う。

図9 5種類のメタ情報の記載例

次に実際のタグ付け方法を(例14)に基づいて説明する。

(例14) 元の文：そこで先生は私たちに留学生の説明会に参加させる。

この文には以下のようにタグ付けする。

そこで先生は私たち | を | (誤 に : 助詞) 留学生の説明会に参加 | させ | (他 させ : 誤, 助詞, 時制) | た | (誤 る : 時制)。

(例14) では、他動詞「(参加)させ」が用いられているので、「させ」に「他」タグを付す。また、「させ」に付随する助詞と時制が誤用であるため⁽¹¹⁾「誤」タグを付し、それぞれタグメモに誤用の種類を記載した。また、「させ」の「他」タグのタグメモには、付随する助詞と時制が誤用であることを簡単に記載した。そして、詳細な情報はメタ情報として図10のように記載した。他動詞「させ」は、使役形として正しく用いられているが、時制が誤っている。また助詞は、本来「を」にすべきところに「に」を用いている。

要素1	要素2	要素3	その他	コメント
使役○、時制×		に←を		そこで先生は私たち を (誤 に : 助詞) 留学生の説明会に参加 させ (他 させ : 誤, 助詞, 時制) た (誤 る : 時制)。

図10 (例14)のタグとメタ情報の記載の仕方

このような形で、コーパス中の全ての動詞にタグ付けを行い、付随する要素を含めて誤用がない場合は「正」タグを、誤用がある場合は「誤」タグを付すとともに誤用の種類をメタ情報として記載した。

タグおよびメタ情報が付されたコーパスをもとに、タグ検索機能を活用することでどのよ

うな分析が可能になるかを以下に検討していく。

5.3.2 タグ検索機能の活用 3：使用動詞の自他の区別と正誤

実際に学習者が動詞をどのように使用しているのかを見るため、実験コーパス内の自動詞、他動詞の各使用数とその正用数、誤用数を検索し、誤用率を割り出した。検索項目において、タグを「他」、タグメモを「正」と指定して検索した結果を図 11 に示す。

結果数: 306 / ライン数: 472				
ライン	タグ	タグメモ	サブコーパス名	ライン番号
勉強不足だから、自信を 持つ (他 持つ: 正) ていない。	他	正	Go・ベトナム	4
日本語学校 が (誤 で: 助詞) 4:20PM に (誤 が: 助詞) 終わっ (自 終わっ: 誤、助詞間違い、対あり) たが、いつも 残っ (自 残っ: 正、対あり) て、19:00PMまで勉強 し (他 し: 正) た。	他	正	Go・ベトナム	7
こんな私 が (誤 は: 助詞) 奨学金を もらう (他 もらう: 正) なんて思いも よら (自 よら: 正) ないだろう。	他	正	Go・ベトナム	5
宿題をちゃんと 出し (他 出し: 正、対あり)、授業で 分から (自 分から: 正) ないことが あっ (自 あっ: 正) たらすぐ先生に質問 し (他 し: 正) た。	他	正、対あり	Go・ベトナム	8
宿題をちゃんと 出し (他 出し: 正、対あり)、授業で 分から (自 分から: 正) ないことが あっ (自 あっ: 正) たらすぐ先生に質問 し (他 し: 正) た。	他	正	Go・ベトナム	8
留学生と いえ (自 いえ: 正) ば、バイト し (他 し: 正) ながら、学校 に (誤 を: 助詞) 通う (自 通う: 誤、助詞間違い) 学生が多い。	他	正	Go・ベトナム	10

図 11 他動詞の正用の検索結果 (一部)

タグ検索を行うと、図 11 のように、詳しいメタ情報も同時に表示されるため、実際に誰 (サブコーパス名) がどの作文の何行目 (ライン番号) に、何の動詞をどのように使用したか (ライン) などを見ることができる。

さらに、他動詞の誤用、自動詞の正用、誤用についても検索したところ、表 7 のような結果が得られた。

表 7 使用動詞の自他の区別と正誤

	総使用数	正用数	誤用数	誤用率
自動詞	483	375	108	22.4%
他動詞	385	306	79	20.5%

総使用数および正誤用数は、いずれも自動詞の方が多いが、誤用率で見ると大差がないこ

とがわかる。このようにコーパス全体を俯瞰することによって、学習者の誤用の傾向を捉えることが可能になる。

また、特定の動詞について、正用と誤用のどちらが多いかも容易に見ることができる。例えば「考える」について、どのような使い方をしているかを検索すると、図 12 のようになる。

ライン	語彙表表記	タグ	サブコーパス名	タグメモ	ライン番号
改めて「考える」(自 考え: 誤、表記)で「みる」(誤 見る: 表記)と、日本へ「来」(自 来: 正)たとき何も「でき」(自 でき: 正)なかった私は「(誤 : 主題の八脱落)料理も」できる「(自 できる: 正) ようになって、アルバイトも楽しく「でき」(自 でき: 正) で日本語も「話せる」(自 話せる: 正) ようになった。	考える	自	Ba・ネパール	誤、表記	17
改めて「考える」(自 考える: 正)と、以上のような経験から「(他 見ると: 誤、不要)、失敗「する」(自 する: 正)「こと」(誤 : 接続)こそ、やる気「と」(誤 があり: 不要、ねじれ)最後の目標まで「あきらめ」(他 あきらめ: 正)ずに「頑張れる」(自 頑張れる: 正) 勇気「を与え」(誤 : 重文述部、ねじれ)、「それ」(誤 : 指示詞、重文主語、ねじれ)が今後人生の宝物「になる」(誤 である: 自動詞脱落、語選択、ねじれ)と「思う」(自 思う: 正、ねじれ)。	考える	自	院Se・モンゴル	正	16
しかし、ある学校は文系しか「なくて」(自 ないで: 誤、活用)、私は理系の学校へ「行き」(自 行き: 誤、テンス)「たかった」(誤 たい: 時制)「から」(誤 なら: 助詞)、「考え」(他 考え: 誤、時制)「なかった」(誤 ない: 時制)。	考える	他	Sa・中国	誤、時制	13
ある学校は英語の試験が必要なので、私はそれで全然「考え」(他 考え: 誤、時制)「たなかった」(誤 たくない: 時制)。	考える	他	Sa・中国	誤、時制	14
「それで」(誤 当然: 語選択)、将来のことを「考え」(他 考え: 正)てみたい。	考える	他	Go・ベトナム	正	3
もし、国へ「戻っ」(自 戻っ: 正、対あり)たら今のホーチミンと10年前のホーチミンは「どちら」(誤 どうち: 表記)が「住み」(自 住み: 正)やすいか、何が「違っ」(自 違っ: 正)のかを「考え」(他 考え: 正)てみたい。	考える	他	Go・ベトナム	正	4
本当に便利だが、今の生活は「早く」(誤 早い: 活用)、人々はお金を「稼ぐ」(他 稼ぐ: 正)ことを「考える」(他 考える: 正)だけで、周りの人々を「忘れ」(他 忘れ: 正)てしまっている。	考える	他	Go・ベトナム	正	17
今日は何を「食べる」(他 食べる: 正)か、このような問題を「考え」(他 考え: 誤、時制)「なかった」(誤 ない: 時制)。	考える	他	Sa・中国	誤、時制	6
私はハノイで「生まれ」(自 生まれ: 正)たが、現在のハノイと20年前「と」(誤 : 助詞脱落)、何が「違っ」(自 違っ: 正)のか「考え」(他 考え: 誤、助詞脱落)「て」(誤 : 助詞脱落)みたい。	考える	他	So・ベトナム	誤、助詞脱落	2
自分が使用「し」(他 し: 誤、話し言葉)「てい」(誤 て: 話し言葉)たセルフォンと今普及「し」(自 され: 誤、受身不要)ているスマートフォンの利点や不利点などについて比較「し」(他 し: 正)、何が「違っ」(自 違っ: 正)のかを「考え」(他 考え: 正)てみたい。	考える	他	Ba・ネパール	正	2
経済的に「考える」(他 考える: 正)とセルフォンの値段は「(誤 が: 助詞)安い。	考える	他	Ba・ネパール	正	6

図 12 「考える」の検索結果画面

図 12 から、「考える」は 11 回使用されており、そのうち正用は 6 回、誤用は 5 回であり、「時制」の誤用が 3 回、表記と助詞脱落がそれぞれ 1 回であることがわかる。このように、誤用の有無にかかわらず、全ての動詞にタグを付すことによって、動詞使用の実態を一覧することができる。

次に、自動詞、他動詞の対があるものについて分析する。まず、有対自動詞・有対他動詞それぞれの使用数および正誤を検索した結果を表 8 に示す。

表8 有対自他動詞の使用数および正誤数

	総使用数	正用数	誤用数（誤用の動詞）	誤用率
有対自動詞	39	32	7（終わる，伝わる，増える，足りる，探し始まる，決まる，入る）	17.9%
有対他動詞	48	44	4（助ける，温める，叶える，始める）	8.3%

動詞の総使用数 868 のうち、有対自他動詞の使用数は 87 であり、全体の 10.0%であった。一方、誤用については、動詞の総誤用数 187 に対し、有対自他動詞の誤用は 11 であり、特に有対自他動詞の誤用が多いとは言えない。しかし、有対自他動詞内で、自他動詞のどちらかに誤用が多いかを比較すると、自動詞の誤用率が 17.9%、他動詞の誤用率が 8.3%と、誤用率にかなり開きがある。実験コーパスの規模が限られているため、今後より詳しい検証が必要ではあるが、有対自他動詞では自動詞の方に誤用が起きやすいという可能性が考えられる。例えば、有対の「入れる」と「入る」を検索してみると、表9のような結果が得られた。使用数はどちらも7回であり同数だったが、自動詞「入る」のみ誤用が起きていた。

表9 「入れる」「入る」の正誤数

	正用	誤用	計
入れる	7	0	7
入る	4	3	7

さらにデータを増やし、他の有対自他動詞に関しても分析を進めることによって、学習者の誤用の傾向や、習得が困難なものを明らかにすることができるであろう。

5.3.3 タグ検索機能の活用4：誤用の多い要素

次に、メタ情報として記載した要素1～3の中で誤用が多いものが何かを見た。先にも述べた通り、要素1～3は以下のように区分している。

要素1：動詞の活用に関わるもの（受身，可能，使役，時制他）

要素2：動詞に後続するもの（ている，である他，たい，てほしい他，
ば，たら他，やすい，すぎる他）

要素3：助詞の誤用（助詞脱落・過剰使用，「に」←「を」etc.）

これら動詞に付随する要素の誤用について、図13のような形で、メタ情報としてそれぞれの要素を指定し、どの要素に問題が多いかを調べた。結果を多い順に表10に示す。



図 13 「可能形」の正用を検索する場合のメタ情報検索項目の設定

表 10 誤用の多い要素

	正用 ⁽¹³⁾	誤用
助詞 (要素 3)	—	89
時制 (要素 1)	—	21
～ている (要素 2)	39	15
活用 (要素 1)	13	15
受身 (要素 1)	17	5
可能 (要素 1)	28	5
使役 (要素 1)	2	2

動詞の活用に関わる要素 1 の誤用が多いが、最も多かったのは要素 3 の助詞に関わるものである。そこで、さらに助詞の誤用について詳細を見た。要素 3 に含まれる助詞の誤用には「助詞が入るべきところになく脱落しているもの (助詞脱落)」「助詞が不必要なのに過剰に使用されているもの (助詞不要)」「違う助詞を入れたもの (誤用←正用)」の 3 種類がある。要素 3 にはこれらのメタ情報が動詞に付随した助詞の誤用として記載してある。なお、要素 3 に記載されているのは動詞に付随した助詞の誤用であるが、動詞に関わらない助詞の誤用もある。各々の誤用には「誤」タグを付し、タグメモに助詞、助詞脱落等と記載してあるので、タグメモを活用して検索することができる。表 11 にそれも含めた助詞の誤用の検索結果を示した。

表 11 助詞の誤用の詳細

	誤←正	助詞不要	助詞脱落	計
(動詞に付随しない) 助詞の誤用	64	1	49	114
動詞に付随する助詞の誤用	61	3	25	89
助詞に関する誤用総数	125	4	74	203

助詞に関する誤用総数 203 のうち、違う助詞を入れた誤用数 125 が最も多く誤用の 61.6% を占める。助詞脱落は、動詞に付随しないところでも多発しているが、違う助詞を入れた誤用については、ほぼ半数が動詞に付随して生じている。そこで、動詞に付随して生じたこの

表 12 違う助詞を入れた誤用が多い（3回以上）助詞と誤用例

誤←正	誤用数	誤用例
を←が	11	専門知識が足りないのを分かった。
を←に	10	このような発表会を初めて参加し、
で←に	6	夏は暑いので、快適な高い所で建て、
が←を	5	そのことが決めてから、
に←を	3	チャンスがあれば、日本語学校に訪ねて、
に←が	3	私は東尋坊の美しさに気に入る、
を←で	3	スピーチ大会を優勝した。

誤用についてどのような誤用が多いのかを見たものが表 12 である。

表 12 から、網掛けをした助詞「を」「に」「が」に関わる誤用が多いことが窺える。また、検索結果画面では、それぞれの誤用が一覧できる形で示されるので、例えば「を←に」の誤用の場合、動詞の「参加する」に付随する場合に誤用が起きやすいことがわかった。「を←に」の誤用のうち、「参加する」に付随した場合が 4 件あり、他にも同じ意味で用いられた「出る」、または「慣れる」「通う」など限られた動詞において同様の誤用が見られた。また、この誤用は特定の学習者に偏ることなく起きていた。このように、メタ情報を活用することによって、誤用の詳細を知ることができる。また、コーパスを個人に特定して検索すれば、個々の学習者の助詞の誤用の傾向を見ることもできる。ただし、メタ情報は一旦データとして取り込んでしまうと、検索画面では確認できず、編集もできない。そのため、効率よく検索するには、事前に十分な研究計画を立てた上で記載項目を決定する必要がある。

5.3.4 タグ検索機能の活用 5：「～ている」の誤用の数と内訳

最後に、要素 2 から誤用の多かった補助動詞「～ている」の誤用について見ることにする。次のページの図 14 に検索結果を示す。

図 14 に見られるように、タグメモに「アспект不要」などの情報を入れておくと、どのような誤用であるかを見ることができる。表 13 に「～ている」の誤用の数と内訳を示す。

表 13 「～ている」の誤用の数と内訳

どのような誤用か	誤用の数（計 15）
「～ている」が不要なのに使用している	11
「～ている」が必要なのに使用していない	2
「～ている」の時制の間違い	2

総計 15 の誤用のうち、「～ている」が不要であるのに使用しているものが最も多い。例えば次のような文である。

結果数: 14 / ライン数: 15				
ライン	タグ	タグメモ	サブコーパス名	ライン番号
他の友達はN3が 受かつ (自 受かつ: 正, 対あり) たのに、自分はN3すらまだ勉強 し (他 し: 誤, アスペクト) てい (誤 : 補助動詞必要) なかった。			Go・ベトナム	3
			Go・ベトナム	3
難しい (誤 むずかし: 表記) 漢字 が (誤 を: 助詞) あった (自 あつてる: 誤, 助詞間違い, アスペクト不要, 時制) 時、辞書を 調べ (他 しらべ: 正) て、何回も 書け (他 書く: 誤, 活用) ば (誤 から: 副詞節) 覚え (他 覚えし: 様態) やすいと 思い (自 思い: 正) ます。	誤	表記	So・ベトナム	9
			So・ベトナム	9
毎日料理 し (自 し: 正) て (誤 てて: アスペクト不要)、味付けも できる (自 できる: 正) ようになった。			Ba・ネパール	10
			Ba・ネパール	10
日本語が 分から (自 分から: 正) なくて、アルバイト先でオーダ ーが 分から (自 分から: 正) なかったり、物の名前が 分から (自 分から: 正) なくて非常に 困っ (自 困っ: 正) た (誤 てた: アスペクト不要)。			Ba・ネパール	14
			Ba・ネパール	14
その時私は親が一生懸命 働い (自 働い: 正) て教育 し (他 させ: 使役不要) たことも 思いだし (他 思い出し: 正) た (誤 てた: アスペクト不要)。			Ba・ネパール	15
			Ba・ネパール	15
東尋坊が 酔っ払っ (自 酔っ払っ: 正) て、 眠っ (自 ねむ っ: 正) ていたときに彼を海に 突き落とし (他 突き落とし: 誤, アスペクト不要) た (誤 ていた: アスペクト不要)。			Ap・タイ	13
			Ap・タイ	13
ベクトルと数列の問題はむずかしかったため、長時間を かけ (他 かけ: 誤, アスペクト不要) た (誤 ていた: アスペクト不要)。			Ap・タイ	10
			Ap・タイ	10
しかし、 行き (自 行き: 正) たい学校は まだまだ (誤 また また: 表記) 決め (他 決め: 誤, アスペクト不要, 時制) なか った (誤 てない: アスペクト, 時制)			Sa・中国	16
	自	正	Sa・中国	16

図 14 「～ている」の誤用の検索結果 (一部)

(例 15) アルバイト研修の 1 日目、皆で自己紹介をし、店長に制服の着替えについて教えてもらっていた。

(例 15) では、「教えてもらった」とすべきところを、「もらっていた」としており、「～ている」が不要なのに使用されている。

このような誤用が「～ている」の誤用に多いことから、学習者は「～ている」を適切に使えていないことがわかる。また、図 14 のサブコーパス名を見ると、学習者「Ba・ネパール」が「～ている」の誤用を最も多く犯していることもわかる。さらにメタ情報の要素 2 を利用して「Ba・ネパール」の他の補助動詞の使用について検索すると、「Ba・ネパール」は、他の学習者に比べ、より積極的に補助動詞を使用しており、「～ている」以外の補助動詞は適切に使用されていることがわかった。つまり「Ba・ネパール」は特に「～ている」の使用に問題が多いと言える。

5.3.5 実験 2 のまとめ

実験 2 では、学習者の作文中の誤用の実態を見るため、誤解析が生じた箇所だけにタグ付

けするのではなく、使用実態を見たい全ての項目にタグを付すことにした。ここでは、使用された動詞に注目し、全ての動詞に自動詞、他動詞の区別をするタグを付し、その動詞に付随する要素を含め、使用の正誤をタグメモに記載した。また、文レベルの誤用への対応として、メタ情報を5種類に区別し、正誤を判断した上で文の左の列に付した。これらのタグとタグメモ、およびメタ情報を活用し、学習者の自動詞、他動詞の誤用率と有対自他動詞の使用実態、誤用の多い要素の中から助詞の誤用と補助動詞「～ている」の誤用実態を見た。これにより、研究の目的に応じて付したタグが、見たい箇所分析に有効であることがわかった。また、どのような誤用かを見る場合にタグメモの情報や文レベルのメタ情報を指定して取り出すことができるため、幅広い分析が可能になることが判明した。

ただし、現在のシステムでは、タグやタグメモについては【Edit】機能による修正が可能だが、メタ情報は編集画面に表示されないため、修正できない。検索結果を閲覧した際に、タグやメタ情報の記載漏れを発見することはままある。そのため、データ取り込み後もメタ情報欄の閲覧、編集が可能になるようにシステムを改善する必要がある。また、実験1の場合と同様、検索結果に漏れがないように、研究目的に合わせたタグメモやメタ情報の記載方法の吟味と統一が重要である。

6. おわりに

本稿では、Co-Chuでのタグ付け方法の詳細について述べるとともに、タグ検索機能を活用した話し言葉と日本語学習者の作文分析の運用実験の結果を報告した。

実験1では、話し言葉を対象に、誤解析の生じる箇所にタグ付けを行い、不規則な音変化が含まれる発話であっても、適切に言語分析をすることができることを示した。また、実験1では誤解析の生じる箇所に注目してタグを付したが、実験2では、作文に用いられた全ての「動詞」にタグを付し、「動詞をどのように使っているか」を見た。これにより、正用と誤用が同時に検索でき、学習者の動詞使用の全体像を捉えることが可能であることを示した。

このように、Co-Chuは研究や教育の目的に合わせて、タグやメタ情報を付すことで、日本語研究や日本語教育に活用できる。日本語教育においては、実際の言語使用に基づいた教育が容易に行えるようになる。様々なタイプの用例を簡単に提示することが可能になり、また学習者は誤用の実態をよりリアリティのあるものとして捉えられる。さらに、同一学習者の作文や会話データを蓄積していけば、縦断的に誤用の傾向を見ることもできる。本実験では雑談と作文をデータとして用いたが、今後、さらにデータを拡充し、タグ付け方法の工夫を重ね、Co-Chuのさらなる活用方法の検討を進めていきたい。

謝辞 本研究の一部は、科学研究費基盤研究（C）18K00723の助成を受けた。また、シス

テムの開発には、LendingHome のラニガン・マシュー氏の協力を得ている。

注

- (1) 「ぼ、ぼ、僕」のような、発音がスムーズになされないものをここでは「言い淀み」とする。
- (2) 「ちゃう」は必ず誤解析になるわけではない。環境によって「違う」と正しく解析される場合もある。本データでは 82 の「違う」のうち、15 が「ちゃう」であったので、タグを付けなくても正しく解析された「ちゃう」が 2 回あったことになる。
- (3) この場合、「(予約され) とって」は方言であるが、「予約されとって」と発話されていれば誤解析にはならない。ここでの誤解析の原因は「とっ」と言いさしになっているためなので、「方」タグではなく、「他」タグを付した。
- (4) この「やん」は「岐阜県の人をよく『やん』を使いますね」のように方言を話題にしている中で用いられたものを指す。
- (5) / は形態素の区切りを示す。
- (6) 実験コーパスのうちの 3 件のサブコーパスが該当する。
- (7) ただし、この中で「淀」タグは、会話中に現れた全ての該当する箇所が付しているが、他のタグは全ての該当する表現を拾っているわけではないので、注意が必要である。
- (8) 「誤」タグの付け方は 3.1 を参照。
- (9) 「4. 文レベルの問題への対応」を参照。
- (10) 「に」←「を」は、本来「を」にすべきところを「に」にしてしまったということを表す。つまり、「誤用」←「正用」ということである。
- (11) 例 14 の元の文中の下線部分「に」と「る」が誤っている。
- (12) 「探し始まる」という語はないため、これは厳密には有対動詞とは言えないが、有対動詞「始まる」「始まる」が影響した誤用であると考えられるので、ここに加えた。
- (13) その形式が用いられた場合に、正しく用いられていれば正用、誤りがあれば誤用となる。ただし「助詞」と「時制」は、用いられた動詞に付随して誤用が見られた場合のみ、誤用であることを要素欄に記載しているため、誤用のみが検索可能となる。

引用文献・引用 URL

- 有田節子 (2015) 「日本語疑問文の応答の冒頭に現れる『は』について—係助詞から感動詞へ—」『国立国語研究所論集』9, 1-22, 国立国語研究所.
- Komori, S., Yamamoto, H., Homma, T., & Lanigan, M. (2019) Co-Chu: Japanese Corpus Analysis System and Two Analyses for Language Teaching. In E. Zimmerman & A. McMeekin (Eds). *Technology-Supported Learning In and Out of the Japanese Language Classroom*. Bristol UK: Multilingual, Matters, 171-202.
- 西阪仰 (2008) 「発話順番内において分散する文—相互行為の焦点としての反応機会場—」『社会言語科学』10-2, 83-95, 社会言語科学会.

- 本間妙・山本裕子・川村よし子・小森早江子（2019）「コーパス分析システム Co-Chu の作文指導への活用」『日本語教育方法研究会誌』Vol. 26, No. 1, 62-63, 日本語教育方法研究会.
- 山本裕子・川村よし子・小森早江子・本間妙（2018）「話し言葉や誤用の含まれたテキストに対応可能なコーパス分析システムの開発」『2018 年度日本語教育学会秋季大会予稿集』295-300, 日本語教育学会.
- 山本裕子・川村よし子・ラニガン・マシュー・小森早江子・本間妙（2019）「誤用や話し言葉に対応可能なコーパス分析システムにおけるタグ検索機能」『CASTEL/J 2019 Proceedings』8th International Conference on CASTEL/J, 179-182.
- MeCab <http://taku910.github.io/mecab/>（2019 年 3 月 31 日閲覧）
- UniDic <https://unidic.ninjal.ac.jp>（2019 年 3 月 31 日閲覧）

（2019 年 10 月 21 日 受理）