

第二部

『日本語学習者作文コーパス』を 『ひまわり』で検索する

(担当: 森篤嗣・京都外国語大)

- 『日本語学習者作文コーパス』を『ひまわり』で検索する意義
- 誤用の集計
- 本来使用されるべき正用に基づく誤用の集計

『ひまわり』における『日本語学習者作文コーパス』の利用方法



論文などで引用する際の
成果物の短縮URL

作成者：山口昌也 (国立国語研究所)
2023-01-31 (公開)

メニュー

- コーパス
 - 『日本語話し言葉コーパス』
 - 『基礎コーパス』
 - 『近代女性雑誌コーパス』
 - 辞書関連
 - 形態素解析辞書『UniDic』
 - 語彙辞書『かたりくさ』
 - 『表記統合辞書』
 - 『分類語彙表増補改訂版』
 - 用例データベース
 - 複合動詞
 - サ変動詞
 - 形容詞
 - ソフトウェア
 - 全文検索システム『ひまわり』
 - 観察支援ツール Fis hWatchr
 - 観察支援ツール Fis hWatchr Mini
 - 作文支援システム T EachOtherS
 - 『ひまわり』支援ツール
 - 『たんぽぽ』, 『プリズム』
 - 成果物の短縮URL
 - 講習会
- 最新の10件
- 2023-02-06
- 全文検索システム『ひまわり』
 - 全文検索システム『ひまわり』 / 『日本語学習者作文コーパス』の利用

1. 概要

このページでは、『日本語学習者作文コーパス』を『ひまわり』で利用する方法 (Windows環境のみ) について説明します。

- 『ひまわり』の機能を利用できる (全文検索, 単語検索, 各種集計機能など)
- 検索システム中のデータ (XMLデータ) を確認・検証・ (必要ならば) 変更できる

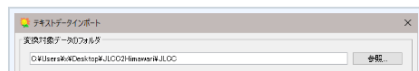
なお, 本ページの内容を利用することにより直接的, 間接的に生じたいかなる損害についても, 保証しません。

2. 準備

- 使用するPCでプログラミング言語Perlを使えるようにしてください。
 - 参考: Strawberry Perlのダウンロード及びインストール (Javadrive)
- 『ひまわり』をインストールしてください。
- 変換パッケージをダウンロードし, 展開してください。以後, 展開したフォルダを「**変換フォルダ**」とします。
 - 変換パッケージ (2023-01-31)
- 『日本語学習者作文コーパス』のHPを開き, 「利用者登録」を行ってください。
- 「使い方を見る」のページを開き, 「全文データをダウンロードする」から「data.zip」をダウンロード・展開し, 次の処理をしてください。
 - morphフォルダを変換フォルダにコピー
 - register.xlsをExcelで読み込み, [ファイル]-[名前をつけて保存]を実行し, 「Unicodeテキスト」形式で保存してください。さらに, 保存した結果 (register.txt) を変換フォルダにコピーしてください。なお, Excel以外を用いる場合は, タブ区切り・UTF-16 BOMなしとなるよう, 保存してください。

3 変換手順

- 変換フォルダにあるhenkan.batをダブルクリックしてください。コーパスがXML形式に変換され, JLCCフォルダに格納されます。
- 起動している『ひまわり』にJLCCフォルダをドラッグ&ドロップし, 『ひまわり』にインポートします。オプションは, 次のとおりです (詳細オプション, 赤枠部分はデフォルトからの変更箇所)。



[こちら](#)を参照のこと

『日本語学習者作文コーパス』の概要

- 本コーパスは、日本語学習者の作文データをコーパス化したものです。
- 初級から上級の日本語学習者304名の作文データが収録されています(語数の合計は113,554語)。
- 作文のテーマは、「外国語が上手になる方法について」(192名分)と「インターネット時代に新聞や雑誌は必要か」(112名分)です。



(<http://jhlee.sakura.ne.jp/JC/howto/>)より

『日本語学習者作文コーパス』の特徴

1. ウェブブラウザで検索可能です。
2. 作文データとともに学習者の日本語能力を測定するため、文法テスト、文字語彙テストを実施し、検索サイト上で検索オプションとして利用できるようになっています。
3. 日本語作文のほか、著者(学習者)自身による母語訳も見ることができます。
(作文意図を把握するためのもの)

本コーパスは「研究課題名：自然言語処理の技術を利用したタグ付き学習者作文コーパスの開発(課題番号:22520537)」の助成を得て開発されました。



(<http://jhlee.sakura.ne.jp/JC/howto/>) より

1. 『日本語学習者作文コーパス』を
『ひまわり』で検索する意義

ウェブブラウザで検索可能なら...

- 素朴な疑問「ウェブブラウザで検索可能なら、『ひまわり』を使う必要はないのでは？」

→『日本語学習作文コーパス』公式サイトでの検索と、『ひまわり』を用いた検索で、できることの違いを考えよう！

→森(2021)で取り上げたテル縮約形を例に挙げます



公式サイトでの検索オプション

- 「**形態素単位で検索**」or「文字列で検索」
- 「誤用のみ表示」or「正用のみ表示」or「**両方を表示**」

形態素単位で検索 文字列で検索

てる

誤用のみ表示 正用のみ表示 両方を表示

■ 検索オプション1

学習者レベル

初級 中級 上級

性別

男 女

学習環境 (収集場所基準)

国内 国外

誤用タイプ

文字 文法 文体

品詞

全て ▼

■ 検索オプション2

作文テーマ

全て ▼

主な母語

中国語 韓国語

日本語履歴

2年未満 2年以上5年未満 5年以上

テスト成績 (文字語彙)

A B C X

テスト成績 (文法)

A B C X

公式サイトでの検索結果画面

ここからファイルでもDL可能

てる： 31件の検索結果

ダウンロード ([TXT](#)/[XLS](#))

	初級	中級	上級	合計
中国語	1	3	3	7
韓国語	0	11	13	24
合計	1	14	16	31

1 (CG031) 文法 か。」「三時間。」「してるのう。先生は大学生のとき、毎日
× × × ×

学習者 原文 母語訳 (PDF)

2 (CG076) 文体 として勉強するの ことにしたの は母語と似てるから だ。で も、勉強すればする
ている と思ったから しかし ×

学習者 原文 母語訳 (PDF)

3 (CG106) 文体 新聞（「新」のつくりが や 雑誌 など さが 情報 が すく っ 感 て。インターが海のよ に 幅 広 な 情報
「万」のよう など せる 報 がない て じ る。ネット が 海 の よ に 幅 広 な 情報
新聞 で は と てい る 幅 広 い ×

学習者 原文 母語訳 (PDF)

4 (CG144) 文法 ただ何かがあったのかが分かつてればいいと 思います。正直な話、
分かれ × から です ×

学習者 原文 母語訳 (PDF)

5 (CG156) 人にとっても迷惑だし、読んでる人も面倒だと思っだろう。もちろん、

学習者 原文 母語訳 (PDF)

公式サイトの検査結果画面の特徴

- 原文に対して赤下線と赤字で添削を同時表示
- 学習者情報は「学習者」ボタンを押すと展開表示される

1 (CG031) か。」「三時間。」「してるのう。先生は大学生のとき、毎日
文法 x x x x 学習者 原文 母語訳 (PDF)

ID	性別	学習環境	作文テーマ	母語	日本語履歴	レベル	テスト成績 (文字語彙)	テスト成績 (文法)
CG031	男	国外	外国語がうまくなる方法について	中国語	2年以上5年未満	初級	C	C

- 「原文」ボタンで原文を表示。ただし、どの部分かは強調されない。

2 (CG076) として勉強するのは母語と似てるからだ。でも、勉強しばる
文体 ことしたの ていと思っか しか x
の るう し

学習者 原文 母語訳 (PDF)

最初、日本語を選んで外国語として勉強するのは母語と似てるからだ。でも、勉強すればするほど、母語と違いが大きくなってきた。日本語の授業は精読と文法と聴解と会話などある。日本語がうまくなるには各部分をちゃんと勉強しなければならない。いままで2年3カ月くらい日本語を勉強してきた私たちにとって、文法はそんなに難しくない。文法の一つ一つの使い方を覚えたら、普通の文章が大体読める。やっぱり、勉強の年数に従って、みんながほぼ同じ程度になった。日本語のうまい人とうまくない人の区別は覚えた単語の量と会話能力の差だと思う。つまり、基本なものだけでなく、新しくできた単語とか日本語らしい言い方とかもできるだけ理解してから、自分のものにしたほうがいい。単語の覚えは毎日一つ暗読すれば、絶対に役立つ。極簡単だけど、続けてやるのが容易ではない。そして、会話能力も日本語がうまくなるポイントだ。日本人と交流するなら、書くだけではない。会話できるかどうか、言葉が通じるかどうかは大切だ。だから、よく日本語を聞かないとヒヤリングが高まらない。よく日本語を話さないと会話能力も高めにくい。自分なりの考えでは、外国語がうまくなる方法は二つある。一つは言語の基本的なものをマスターすることである。二つは言語の能力をのぼすことである。

公式サイトの検索結果XLS

	A	B	C	D	E	F	G	H	I	J	K	L
1	learner ID	native lang	level	error	keyword_c	pos	left context	keyword	right context			
2	CG031	中国語	初級	誤	×	助動詞	「三時間。」「し	てる	のう。先生は大学生のとき、毎日			
3	CG076	中国語	中級	誤	ている	助動詞	するのとは母語と似	てる	からだ。でも、勉強すればする			
4	CG106	中国語	中級	誤	ている	助動詞	すくないって感じ	てる	。インターネットが海のように幅広い情報			
5	CG144	中国語	中級	誤	×	助動詞	ったのかが分かっ	てれ	ばいいと思います。正直な話、			
6	CG156	中国語	上級			助動詞	も迷惑だし、読ん	でる	人も面倒だと思うだろう。もちろん、			
7	CN310	中国語	上級	誤	ている	助動詞	る時、耳耳があい	てる	時、ラジオを聴きながら、すごします。			
8	CN317	中国語	上級	誤	ている	助動詞	が外国語を勉強し	てる	と重い思わないほうがいいと思います			
9	KG005	韓国語	中級	誤	、	助動詞	国人の友達を作っ	て	てがみ、メール、電話と蚊をしながら			
10	KG013	韓国語	上級			助動詞	もそも何も気にし	て	ないのにその外語がうまくなりたいた			
11	KG051	韓国語	上級			助動詞	攻しようとも思っ	て	なかった。そんな理由で私がやって			
12	KG056	韓国語	中級	誤	ている	助動詞	の学ぶ外国語を使っ	てる	人と実祭に生活に使ってる言葉を			
13	KG056	韓国語	中級	誤	ている	助動詞	実祭に生活に使っ	てる	言葉をその場ですぐに分かるし身			
14	KG076	韓国語	上級	誤	×	助動詞	しても会話はやっ	てる	機会が少ないので外国人に会って			
15	KG078	韓国語	上級			助動詞	いどの興味が持っ	てる	はずであります。それなら、その国			
16	KG092	韓国語	中級	誤	ている	助動詞	まくなって欲がっ	てる	なら、その国を理解しなければなら			
17	KG094	韓国語	中級			助動詞	、映画などを通じ	て	たくさ聞く方がいいだと思ふ。			
18	KG094	韓国語	中級			助動詞	も学ぶこともでき	て	たくさ役に立つだと思ふ。			
19	KG103	韓国語	中級			助動詞	言う目的は変わっ	て	ないです。もし新聞などのかみいんさつ物			

公式サイトの検索結果XLSの特徴

- 下記の9つを出力
 1. learner ID(学習者ID)
 2. native language(学習者の母語)
 3. Level(学習者のレベル)
 4. Error(正用／誤用)
 5. keyword_correction(添削結果形)
 6. Pos(品詞)
 7. left context(前文脈)
 8. Keyword(検索ワード)
 9. Right context(後文脈)

『ひまわり』の検索オプション

- 「全文(文字列検索)」or「書字形」or「語彙素」or「語彙素読み」or「品詞」or「添削」+正規表現の使用可
- 「前文脈」「後文脈」に対して「で始まる」「と一致する」「で終わる」「を含む」「正規表現」「と一致しない」「で始まらない」「で終わらない」「を含まない」「正規表現(否定)」が使用可
- 学習者の属性でフィルタをかけることが可能

検索文字列	フィルタ	コーパス	検索オプション
学習者	▼	C	で始まる ▼
文字	▼	A	で始まる ▼
文法	▼	A	で始まる ▼

『ひまわり』の検索結果画面

全文検索システムひまわり - [JLC] - config_JLC.xml

ファイル 編集 ツール ヘルプ

検索文字列 フィルタ コーパス 検索オプション

全文

前文脈

後文脈

検索

字体変換

クリア

no	前文脈	キー	後文脈	学習者	性別	学習環境	作文テーマ	母語	学習歴	日本語レ...	文字	文法	添削	添削タイプ	品詞	活用型	活用形	語彙素	語彙素読み	発音形出...	発音形基...
1	の(偏と旁が逆)を持っ	てる	。それは向を通うす	KG137	女	国外	インター...	韓国語	5年以上	上級	X	X	ている	S	助動詞	下一段-タ行	終止形-一般	てる	テル	テル	テル
2	報がすくないって感じ	てる	。インターネットが	CG106	女	国外	インター...	中国語	2年以上5...	中級	X	X	ている	S	助動詞	下一段-タ行	終止形-一般	てる	テル	テル	テル
3	勉強するのは母語と似	てる	からだ。でも、勉強	CG076	女	国外	外国語が...	中国語	5年以上	中級	A	A	ている	S	助動詞	下一段-タ行	終止形-一般	てる	テル	テル	テル
4	るに通じて、語感を育	てる	ことができる。それ	CG038	女	国外	外国語が...	中国語	2年以上5...	上級	A	A	養う	G	動詞-一般	下一段-タ行	連体形-一般	育てる	ソダテル	ソダテル	ソダテル
5	の言語に対する勘を育	てる	ことだとします。	KG070	男	国外	外国語が...	韓国語	5年以上	上級	A	A			動詞-一般	下一段-タ行	連体形-一般	育てる	ソダテル	ソダテル	ソダテル
6	が◆制を見られてもっ	てる	こともできます。で	KG127	女	国外	インター...	韓国語	2年以上5...	中級	X	X	x	S	助動詞	下一段-タ行	連体形-一般	てる	テル	テル	テル
7	ならばやく耳をそばだ	てる	だろう。 そうした	KG053	女	国外	外国語が...	韓国語	2年以上5...	中級	A	A	慣れる	G,C	動詞-一般	下一段-タ行	終止形-一般	歌てる	ソバダテル	ソバダテル	ソバダテル
8	力でよい環境をうちた	てる	とともに、十分に利用	CG055	女	国外	外国語が...	中国語	2年以上5...	中級	A	A	作る	C	動詞-一般	下一段-タ行	終止形-一般	打ち立てる	ウチタテル	ウチタテル	ウチタテル
9	聞や雑誌などは役に立	てる	と思う。「でも、P	CG134	女	国外	インター...	中国語	2年以上5...	上級	X	X	立つ	G	動詞-一般	下一段-タ行	終止形-一般	立てる	タテル	タテル	タテル
10	自分が外国語を勉強し	てる	と重い思わないほうが	CN317	女	国内	外国語が...	中国語	2年以上5...	上級	X	X	ている	S	助動詞	下一段-タ行	終止形-一般	てる	テル	テル	テル
11	がうまくなって欲がっ	てる	なら、その国を理解し	KG092	男	国外	外国語が...	韓国語	2年未満	中級	A	B	ている	G	助動詞	下一段-タ行	終止形-一般	てる	テル	テル	テル
12	」「三時間。」「し	てる	のう。先生は大学生	CG031	男	国外	外国語が...	中国語	2年以上5...	初級	C	C	x	G	助動詞	下一段-タ行	終止形-一般	てる	テル	テル	テル
13	ーっとしてかめんを見	てる	のが好きであまりくわ	KG125	男	国外	インター...	韓国語	2年未満	上級	X	X	x	G	助動詞	下一段-タ行	連体形-一般	てる	テル	テル	テル
14	るていどの興味が持っ	てる	はずであります。そ	KG078	女	国外	外国語が...	韓国語	2年未満	上級	A	A			助動詞	下一段-タ行	連体形-一般	てる	テル	テル	テル
15	分にいらぬ部分はす	てる	ようなせんたくが、よ	KG131	男	国外	インター...	韓国語	5年以上	上級	X	X			動詞-一般	下一段-タ行	連体形-一般	捨てる	ステル	ステル	ステル
16	敗したら、日本語を捨	てる	より、二度とさらに勉	CG024	女	国外	外国語が...	中国語	2年以上5...	上級	A	A	あきらめる	G	動詞-一般	下一段-タ行	終止形-一般	捨てる	ステル	ステル	ステル
17	じている。簡単に捨	てる	わけにはいかない。	CG151	女	国外	インター...	中国語	2年以上5...	中級	X	X	切り捨てる	C	動詞-一般	下一段-タ行	連体形-一般	捨てる	ステル	ステル	ステル
18	品の欠点は、すぎに捨	てる	わけにはいかないこと	CG130	女	国外	インター...	中国語	5年以上	上級	X	X			動詞-一般	下一段-タ行	連体形-一般	捨てる	ステル	ステル	ステル
19	きぎとれないからあわ	てる	人がいるんですが、き	KG042	男	国外	外国語が...	韓国語	2年未満	中級	A	A			動詞-一般	下一段-タ行	連体形-一般	慌てる	アワテル	アワテル	アワテル
20	せつ学ぶ外国語を使っ	てる	人と実祭に生活に使っ	KG056	女	国外	外国語が...	韓国語	2年以上5...	中級	A	A	ている	G	助動詞	下一段-タ行	連体形-一般	てる	テル	テル	テル
21	ある。これは家を建	てる	前に土台を建てなけれ	CG091	女	国外	外国語が...	中国語	2年以上5...	初級	B	B			動詞-一般	下一段-タ行	連体形-一般	立てる	タテル	タテル	タテル
22	います■。口変かし	てる	新しい世界で使い方	KG127	女	国外	インター...	韓国語	2年以上5...	中級	X	X	続けている	C,G	助動詞	下一段-タ行	連体形-一般	てる	テル	テル	テル
23	ている時、耳耳があい	てる	時、ラジオを聴きなが	CN310	女	国内	外国語が...	中国語	2年以上5...	上級	X	X	ている	S	助動詞	下一段-タ行	連体形-一般	てる	テル	テル	テル
24	めいしても会話はやっ	てる	機会が少ないので外国	KG076	女	国外	外国語が...	韓国語	2年以上5...	上級	A	A	x	G,S	助動詞	下一段-タ行	連体形-一般	てる	テル	テル	テル
25	人と実祭に生活に使っ	てる	言葉をその場ですぐに	KG056	女	国外	外国語が...	韓国語	2年以上5...	中級	A	A	ている	G	助動詞	下一段-タ行	連体形-一般	てる	テル	テル	テル

『ひまわり』の検索結果画面の特徴

●公式サイトで表示した9つに加えて、下記の15種類を出力

1. 性別
2. 学習環境(国内／国外)
3. 作文テーマ
4. 学習歴
5. 文字テスト
6. 文法テスト
7. 添削タイプ
8. 活用型
9. 活用形
10. 語彙素
11. 語彙素読み
12. 発音形出現形
13. 発音形基本形
14. 書字形
15. 語種

『日本語学習者会話コーパス』を『ひまわり』で検索する意義

- 「どの部分にどのように添削がされているか」は公式サイトの方がビジュアルで見やすい
- 『日本語学習者会話コーパス』を『ひまわり』で検索する意義は...
 - 検索に正規表現が使える
 - 前後文脈のバリエーションを細かく指定できる
 - 学習者の属性のフィルタもより細かく指定できる
 - 公式サイトXLSより15種類多い出力結果でソートができる
- つまり、作文(添削)データを質的に見るなら公式サイト。量的に分析する場合も軽くなら公式サイトでOK。
- 詳細に量的分析する場合には『ひまわり』にメリットあり(「フィルタ」→「統計」が強力！文字列でなく全ての「動詞」などの検索も可能)

2. 誤用の集計

誤用コーパスの限界

- 学習者の誤用の定義は極めて困難であり、その判定にはどうしても主観が入らざるを得ない。ゆれも多い。
- データ作成者には敬意を払うべきだが、自分で判定したのではない誤用は「判定根拠が不明」としか言えない。

5 (CG156) 人にとっても迷惑だし、読んでる人も面倒だと思うだろう。もちろん、

6 (CN310) 、〇いている時、耳耳があいてる時、ラジオを聴きながら、すごします。
文体 歩い × ている

- 例えば、上記の5では「読んでる」が正用、6では「あいてる」が誤用（添削形「あいている」）と判定されている。

つまり「誤用」タグだけではNG

- つまり、「ている」の縮約形「てる(い落ち)」を検索しようとする場合、コーパスを信用して「誤用」タグだけを集計するのでは取りこぼしがあり、網羅できない。

5 (CG156) 人にとっても迷惑だし、読んでる人も面倒だと思うだろう。もちろん、

6 (CN310) 、〇いている時、耳耳があいてる時、ラジオを聴きながら、すごします。
文体 歩い x ている

- 『ひまわり』では正規表現が使えるので、様々な検索に対応可能。
- 例えば、語彙素「てる」はUniDicで語彙素として登録されているため、これが利用できる。

語彙素「てる」の検索

- 『ひまわり』で、「語彙素」で「てる」を検索すると...

全文検索システムひまわり - [JLC] - config_JLC.xml

ファイル 編集 ツール ヘルプ

no	前文脈	キー	後文脈	学習者	性別	学習環境	作文テーマ	母語	学習歴	日本語レベル	文字	文法	添削	添削タイプ	品詞	活用型	活用形	語彙素	語彙素読み
1	ているはずで	あわて	ず、あせらず、あきら	KG078	女	国外	外国語が...	韓国語	2年未満	上級	A	A			動詞-一般	下一段-タ行	未然形-一般	慌てる	アワテル
2	よくききとれないから	あわてる	人がいるんですが、き	KG042	男	国外	外国語が...	韓国語	2年未満	中級	A	A			動詞-一般	下一段-タ行	連体形-一般	慌てる	アワテル
3	自分の力でよい環境を	うちたてる	とともに、十分に利用	CG055	女	国外	外国語が...	中国語	2年以上5...	中級	A	A	作る	C	動詞-一般	下一段-タ行	終止形-一般	打ち立てる	ウチタテル
4	している時に好奇心を	かりたて	てつづいて勉強するか	CG055	女	国外	外国語が...	中国語	2年以上5...	中級	A	A			動詞-一般	下一段-タ行	連用形-一般	駆り立てる	カリタテル
5	した。それで新聞を	すて	ました■。ロインタ	KG118	女	国外	インター...	韓国語	2年以上5...	上級	X	X			動詞-一般	下一段-タ行	連用形-一般	捨てる	ステル
6	自分にいらぬ部分は	すてる	ようなせんたくが、よ	KG131	男	国外	インター...	韓国語	5年以上	上級	X	X			動詞-一般	下一段-タ行	連体形-一般	捨てる	ステル
7	まなぶならばやく耳を	そばだてる	だろう。 そうした	KG053	女	国外	外国語が...	韓国語	2年以上5...	中級	A	A	慣れる	G,C	動詞-一般	下一段-タ行	終止形-一般	歌てる	ソバダテル

- 「慌てる」「打ち立てる」「駆り立てる」などがヒットしてしまう。

語彙素「てる」を単独で検索

- 語彙素「てる」を単独で検索するには...

全文検索システムひまわり - [JLC] - config_JLC.xml
ファイル 編集 ツール ヘルプ

検索文字列	フィルタ	コーパス	検索オプション
語彙素	▼	てる	
正規表現(前)		^	正規表現 ▼
正規表現(後)		\$	正規表現 ▼

検索
字体変換
クリア

- 正規表現(前)に「^」、正規表現(後)に「\$」を入力する。
- 「全文」など、下二段が「前文脈」「後文脈」の場合は一段目に「^てる\$」を入力する。

語彙素「てる」は31例

- 語彙素「てる」は31例中、誤用17例（実は公式サイトではデフォルトで出る）

no	前文脈	キー	後文脈	学習者	性別	学習環境	作文テーマ	母語	学習歴	日本語レ...	文字	文法	添削	添削タイプ
1	います■。口変化し	てる	新しい世界で使い方	KG127	女	国外	インター...	韓国語	2年以上5...	中級	X	X		
2	ットでニュースを読ん	てる	人もいるけど、時間は	KG125	男	国外	インター...	韓国語	2年未満	上級	X	X	続けている	S
3	勉強するのは母語と似	てる	からだ。でも、勉強	CG076	女	国外	外国語が...	中国語	5年以上	中級	A	A	ている	S
4	報がすくないって感じ	てる	。インターネットが	CG106	女	国外	インター...	中国語	2年以上5...	中級	X	X	ている	S
5	ている時、耳があい	てる	時、ラジオを聴きなが	CN310	女	国内	外国語が...	中国語	2年以上5...	上級	X	X	ている	S
6	自分が外国語を勉強し	てる	と重い思わないほうが	CN317	女	国内	外国語が...	中国語	2年以上5...	上級	X	X	ている	S
7	せつ学ぶ外国語を使っ	てる	人と実祭に生活に使っ	KG056	女	国外	外国語が...	韓国語	2年以上5...	中級	A	A	ている	G
8	人と実祭に生活に使っ	てる	言葉をその場ですくに	KG056	女	国外	外国語が...	韓国語	2年以上5...	中級	A	A	ている	G
9	がうまくなくて欲がっ	てる	なら、その国を理解し	KG092	男	国外	外国語が...	韓国語	2年未満	中級	A	B	ている	G
10	の偏と旁が逆)を持っ	てる	。それは向を通うす	KG137	女	国外	インター...	韓国語	5年以上	上級	X	X	ている	S
11	国のことをわらくちし	てる	たこと) ■ □次は人	KG118	女	国外	インター...	韓国語	2年以上5...	上級	X	X	い	G
12	て外国人の友達を作っ	てる	てがみ、メール、電話	KG005	女	国外	外国語が...	韓国語	2年以上5...	中級	B	A		S
13	」「三時間。」「し	てる	のう。先生は大学生	CG031	男	国外	外国語が...	中国語	2年以上5...	初級	C	C	x	G
14	があったのか分かり	てる	ばいいと思います。	CG144	女	国外	インター...	中国語	2年以上5...	中級	X	X	x	G
15	めいしても会話はやっ	てる	機会が少ないので外国	KG076	女	国外	外国語が...	韓国語	2年以上5...	上級	A	A	x	G.S
16	ーとしてがめんを見	てる	のが好きであまりくわ	KG125	男	国外	インター...	韓国語	2年未満	上級	X	X	x	G
17	か◆制を見られてもっ	てる	こともできます。で	KG127	女	国外	インター...	韓国語	2年以上5...	中級	X	X	x	S
18	っても迷惑だし、読ん	てる	人も面倒だと思うだろ	CG156	女	国外	インター...	中国語	2年以上5...	上級	X	X		
19	そもそも何も気にし	てる	ないのでその外語がう	KG013	男	国外	外国語が...	韓国語	2年以上5...	上級	A	A		
20	を専攻しようと思っ	てる	なかった。そんな理	KG051	男	国外	外国語が...	韓国語	2年以上5...	上級	A	A		
21	るていどの興味が持っ	てる	はずであります。そ	KG078	女	国外	外国語が...	韓国語	2年未満	上級	A	A		
22	文化も学ぶこともでき	てる	たくさ役に立つと思	KG094	女	国外	外国語が...	韓国語	2年以上5...	中級	A	B		
23	音楽、映画などを通じ	てる	たくさ聞く方がいいだ	KG094	女	国外	外国語が...	韓国語	2年以上5...	中級	A	B		
24	いと言う目的は変わっ	てる	ないです。もし新聞	KG103	女	国外	インター...	韓国語	2年未満	中級	X	X		
25	ットがそんなに広がっ	てる	ないので、新聞(「新	KG108	男	国外	インター...	韓国語	5年以上	上級	X	X		
26	るからいらなくなって	てる	なくなるとダメだと思	KG110	女	国外	インター...	韓国語	2年未満	中級	X	X		
27	しいじょうほうが入っ	てる	ます。そして時間を	KG110	女	国外	インター...	韓国語	2年未満	中級	X	X		
28	しの人々が何かをやっ	てる	たとか、何の衣服を着	KG144	女	国外	インター...	韓国語	2年以上5...	上級	X	X		
29	たとか、何の衣服を着	てる	たとか、すなわち歴史	KG144	女	国外	インター...	韓国語	2年以上5...	上級	X	X		
30	てつで新聞とかを読ん	てる	人々の楽しみをうばわ	KG147	女	国外	インター...	韓国語	5年以上	上級	X	X		
31	以来、まだ百年もたっ	てる	ない。それなのに人	KG152	女	国外	インター...	韓国語	2年以上5...	上級	X	X		

さらに目で見えてゴミを取り除く

- 語彙素「てる」31例中，誤用判定17例（うち「て(で)いる」9例）

8 (KG005) 文体 むずかしい本、そして外国人の友達を作っててがみ、メール、電話と蚊をしながら
作り、などかx

- 8は誤用判定ありで，添削結果形「、」。これは「ている」の縮約形ではないため除く。

16 (KG094) なかったらドラマと音楽、映画などを通じてたくさ聞く方がいだと思う。
や たくさん聞いた x

21 (KG110) はもうインターネットがあるからいらないって言てなくなるとダメだと思います。今
と 言っ なくすx のはよくない

- 16, 21も「ている」の縮約形ではないため除く。
- 検索結果は必ず目で見える（実は目で見える場合は公式サイトが見やすいかも）

3. 本来使用されるべき 正用に基づく誤用の集計

「ている」を使うべき誤用の種類の探索

- 本来「ている」を使うべき箇所が生じる誤用は、縮約形「てる」だけではない。

14 (CG017) 文法、外国語がたくさん ひとが学習しまし た。でも私は日本語より英語の
をたくさん の 学習 ています × × ×

15 (CG017) 文法 語より英語のほうが難しいと思いまし た。外国語がうまくなる、いろいろ 方法が
思って います なるには いろいろな

- 「ている」を使うべきところで「た」を使う誤用はよく見られる。
- 本来「ている」を使うべき箇所が生じる誤用は、どのような種類があり、どのような分布になっているのか。

→これを調べるためには、**添削結果を信用するしかない**(添削結果を信用することが、研究方法として妥当かどうかは別途要検討)。

『ひまわり』で品詞別に統計

● 検索結果235件

検索文字列 フィルタ コーパス 検索オプション

添削 てい

正規表現(前) 正規表現

正規表現(後) 正規表現

● 「てい」を検索後、「品詞」列のセルのどれかを右クリックして「統計」

全文検索システムひまわり - [JLC] - config_JLC.xml

ファイル 編集 ツール ヘルプ

検索文字列 フィルタ コーパス 検索オプション

添削 てい

正規表現(前) 正規表現

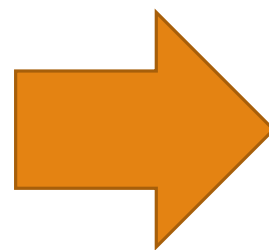
正規表現(後) 正規表現

検索 字変換 クリア

語レ...	文字	文法	添削	添削タイプ	品詞	活用型	活用形	語彙素	語彙素部
X	X	X	そこに存...	S.G	補助記号...				
B	A		日本語を...	S.G	補助記号...				
A	B		できてい	G	動詞-非自...	五段-ラ行...	連用形-促...	有る	アル
X	X		溢れている	G	動詞-一般	下一段-ラ...	連体形-一般	溢れる	アフレル
X	X		もってい	G	動詞-非自...	下一段-ラ...	連体形-一般	有る	アル
A	B		使われて...	G	動詞-非自...	コピ-	形-一般	有る	アル
A	B		持っている	G.S	動詞-非自...	コピ-(列名含む)	形-一般	有る	アル
A	B		関係して...	G	動詞-非自...	全選択	形-一般	有る	アル
A	A		持って	G	動詞-非自...	置換	形-一般	有る	アル
B	A		生きて	G	動詞-非自...	置換	形-一般	居る	イル
A	A		ぎてい	G	動詞-非自...	置換	形-一般	居る	イル
A	A		勉強して...	G	動詞-非自...	置換	形-一般	居る	イル
B	C		日本語に...	G	名詞-普通...	置換	形-一般	居る	イル

統計

検索総数: 235



品詞 頻度

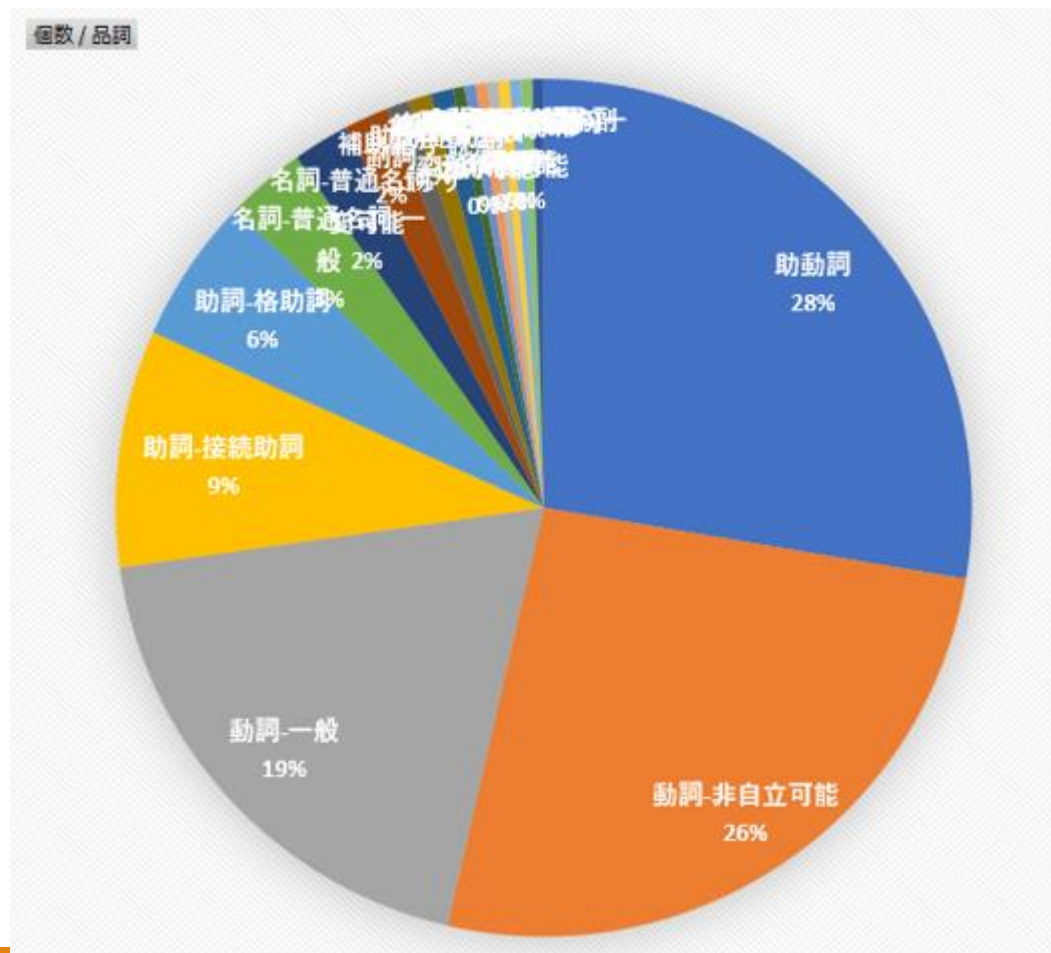
助動詞	65
動詞-非自...	61
動詞-一般	45
助詞-接続...	21
助詞-格助詞	13
名詞-普通...	7
名詞-普通...	5
副詞	4
形容詞-一般	2
補助記号...	2
助詞-副助詞	2
名詞-普通...	1
形容詞-非...	1
名詞-普通...	1
形状詞-一般	1
代名詞	1
名詞-普通...	1
助詞-係助詞	1
接尾辞-名...	1

総数(延べ): 235, 異なり: ...

「品詞」での分布

品詞	頻度
助動詞	65
動詞-非自...	61
動詞-一般	45
助詞-接続...	21
助詞-格助詞	13
名詞-普通...	7
名詞-普通...	5
副詞	4
形容詞-一般	2
補助記号-...	2
助詞-副助詞	2
名詞-普通...	1
形容詞-非...	1
名詞-普通...	1
形状詞-一般	1
代名詞	1
名詞-普通...	1
助詞-係助詞	1
接尾辞-名...	1

総数(延べ): 235, 異なり: ...



『ひまわり』で助動詞を統計

1. 品詞列の列タイトルを右クリックし, リストから「助動詞」を選択
2. 「語彙素」列のセルをどれか右クリックし「統計」

全文検索システムひまわり - [JLC] - config_JLC.xml

ファイル 編集 ツール ヘルプ

検索文字列 フィルタ コーパス 検索オプション

添削 てい

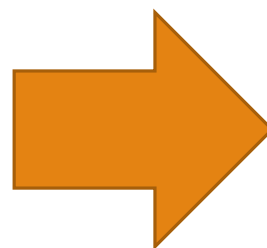
正規表現(前) 正規表現

正規表現(後) 正規表現

検索 字体変換 クリア

語し...	文字	文法	添削	添削タイプ	品詞	活用型	活用形	語彙素	語彙素
	B	A	ために勉...		名詞-普通	[文字列指定]		コト	
	X	X	している...	G	代名詞			サ	
	X	X	させている	G	副詞			サセル	
	X	X	触っている	G	助動詞			サワル	
	B	B	をもたれ...	G	動詞-非能			スル	
	A	B	してい	G	動詞-非能	助詞-係助詞		スル	
	X	X	してい	G	動詞-非能	助詞-副助詞		スル	
	X	X	してい	G	動詞-非能	助詞-接続助詞		スル	
	A	A	してい	G	動詞-非能	助詞-格助詞		スル	
	X	X	確かめてい	S	動詞-非能	助詞-一般		スル	
	B	A	してい	G	動詞-非能	動詞-非自立可能		スル	
	A	B	していた	G	助詞-副助			シカ	
	B	C	しまっ...	G	動詞-非能			シマウ	

検索総数: 235



[4] ...

ファイル 編集 ツール

語彙素	頻度
た	31
だ	10
てる	9
ます	5
れる	4
たい	2
です	1
られる	1
ない	1
させる	1

総数(延べ): 65, 異なり: 10

助動詞の誤用形式分布

語彙素	頻度
た	31
だ	10
てる	9
ます	5
れる	4
たい	2
です	1
られる	1
ない	1
させる	1

総数(延べ) : 65, 異なり : 10

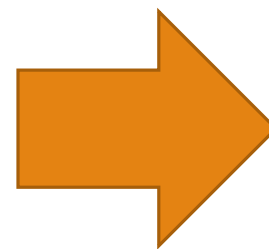


『ひまわり』で動詞を統計

1. 「フィルタ」タブで「品詞」を「動詞で始まる」に指定（「動詞-一般」と「動詞-非自立可能」の両方を対象とする場合）
2. 「語彙素」列のセルをどれか右クリックし「統計」

検索文字列 フィルタ コーパス 検索オプション

品詞	動詞	で始まる	検索 字体変換 クリア
性別		で始まる	
学習環境		で始まる	



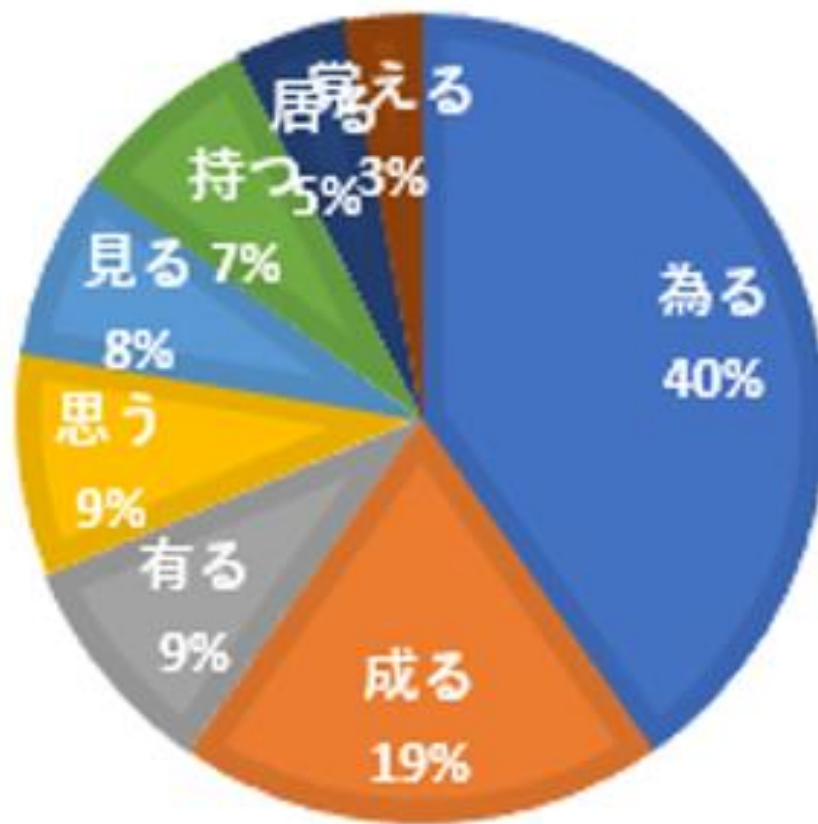
語彙素	頻度
為る	27
成る	13
思う	6
有る	6
持つ	5
見る	5
居る	3
覚える	2
生きる	1
頑張る	1
劣る	1
溢れる	1
似る	1
知る	1
来る	1

総数(延べ): 106, 異なり: 47

動詞の誤用形式分布(頻度2以上)

語彙素	頻度
為る	27
成る	13
思う	6
有る	6
持つ	5
見る	5
居る	3
覚える	2
生きる	1
頑張る	1
劣る	1
溢れる	1
似る	1
知る	1
来る	1

総数(延べ) : 106, 異なり : 47



「ている」を使うべき誤用の種類の探索

- 「添削」で「てい」を含む誤用は235例
- そのうち、助動詞が65例、動詞（非自立可能61＋一般45）106例

14 (CG017) 文法、外国語がたくさん ひとが学習しまし た。でも私は日本語より英語の
をたくさん 学習 ています × × ×

15 (CG017) 文法 語より英語のほうが難しいと思ひました た。外国語がうまくなる、いろいろ方法が
思っ て いま す なる には いろ いろ な

- 助動詞65例の48%が「た」。
- 続いて「だ」10例。

(CG118) 4 文法/文 体 ますが、新聞や雑誌より携帯性が 不足 だ という点もあります。いくら インタ ネ ット
の面で劣っ て い る インタ ネ ット

- 先ほど調べた「てる」9例が続く。

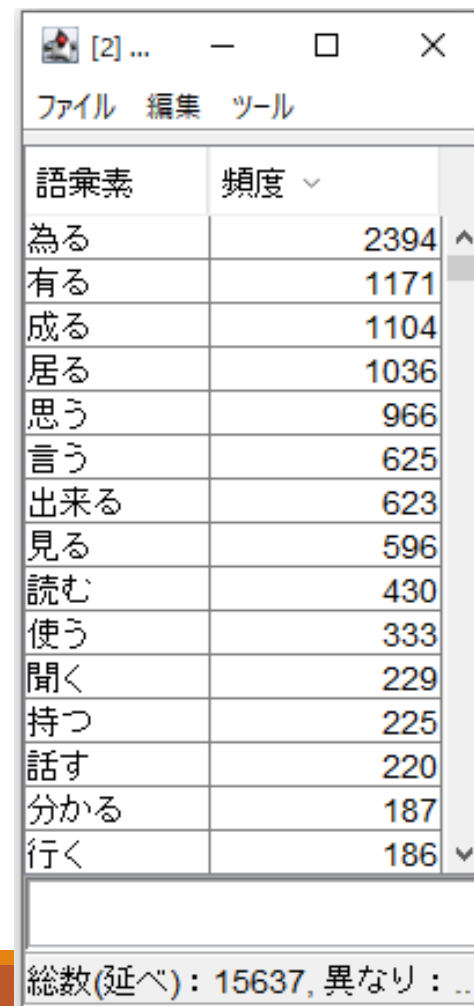
「ている」を使うべき誤用の種類の探索

- 「添削」で「てい」を含む誤用は235例
- そのうち、助動詞が65例、動詞(非自立可能61+一般45)106例
- 動詞106例の40%が「する」。続いて「なる」ある。
→これはそもそもコーパス全体での使用そのものが多いことによる偏りを疑うべき

→「品詞」で「^動詞」を検索 cf.「助動詞」「感動詞」

- 「する」 $27/2,394 = 1.13\%$
- 「なる」 $13/1,104 = 1.18\%$
- 「ある」 $6/1,171 = 0.05\%$

品詞指定の検索は公式
サイトではできない



語彙素	頻度
為る	2394
有る	1171
成る	1104
居る	1036
思う	966
言う	625
出来る	623
見る	596
読む	430
使う	333
聞く	229
持つ	225
話す	220
分かる	187
行く	186

総数(延べ): 15637, 異なり: ...

まとめ

- 『日本語学習者作文コーパス』を『ひまわり』で検索する意義

→作文(添削)データを質的に見るなら公式サイト。量的に分析する場合も軽くなら公式サイトでOK。詳細に量的分析する場合には『ひまわり』にメリットあり(「フィルタ」→「統計」が強力！文字列でなく全ての「動詞」などの検索も可能)

- 誤用の集計

→誤用を検索しようとする場合、コーパスを信用して「誤用」タグだけを集計するのでは取りこぼしがあり、網羅できない。検索結果も必ず目で見ろ。

- 本来使用されるべき正用に基づく誤用の集計

→これを調べるためには、添削結果を信用するしかない(添削結果を信用することが、研究方法として妥当かどうかは別途要検討)。

参考文献

- 森篤嗣2021「書き言葉におけるテル縮約形と日本語教育」金澤裕之・川端元子・森篤嗣(編)『日本語の乱れか変化かーこれまでの日本語、これからの日本語』ひつじ書房, 199-217.
- 山口昌也2014「全文検索システム『ひまわり』を用いた既存言語資料の活用方法の検討」『第6回日本語コーパスワークショップ予稿集』国立国語研究所, 151-156.
- 林炫情・李在鎬・宮岡弥生・柴崎秀子・趙垺熙(2013)「言語処理技術を利用した日本語学習者作文コーパスの開発」『日本文化学報』韓国日本文化学会, 56, 129-142.