

日本語学習者のコーパスツールの使用実態 —作文での産出に着目して—

寺嶋 弘道 (立命館アジア太平洋大学)

板井 芳江 (立命館アジア太平洋大学)

Use of Corpus Tool in Writing by Japanese Language Learners

Hiromichi Terajima (Ritsumeikan Asia Pacific University)

Yoshie Itai (Ritsumeikan Asia Pacific University)

要旨

本研究では、「ライティングにおけるコーパスツール活用モデル (寺嶋・板井, 2021)」を取り入れ、日本語学習者が作文を書く際のコーパスツールの使用実態を調査した。分析の結果、作文においてコーパスツールが使用された回数の中央値は 5 回 (最大値: 14 回、最小値: 2 回) であった。また、作文で使用された表現には、作文前に作成した語彙ネットワークから取り入れられたもの、作文を書いている間に検索されたものがあった。最も多く産出されたのは「名詞+助詞+動詞」のパターンで、中級前半レベルと中級後半レベルの言葉で構成されたコロケーションであった。さらに、その適切さを分析したところ、コロケーション、あるいはコロケーションと共に使用された文法項目が原因で誤用と判断されたもの、コロケーションが使用された節において誤用と判断されたものがあり、「混同」による誤用が多いことがわかった。

1. はじめに

日本語学習者 (以下: 学習者) のライティングでは、「私の国では車がすごく震える」という誤用¹が見られる。これらは「混同」に分類される誤用 (市川編, 2010) で、こうした誤用を減らすためには、言葉の共起関係 (以下: コロケーション²) の知識を事前に学んでおくか、何らかのリソースによって適切なコロケーションを調べる方法を身に付けておくことが必要である。例えば、コロケーションを調べられるツールの一つとして、NINJAL-LWP for TWC³というものがある。このツールは、レキシカルプロファイリングの機能を備えたコーパスツールで (赤瀬川他, 2016)、このツールを用いることで「車が揺れる」というコロケーションを見つけることが可能である。しかしながら、学習者がコーパスツールの機能を十分に活かせるとは限らない。ツールを使う場合、学習者は「車が+動詞」のパターンで表示される大量の情報の中から、「車が揺れる」という表現を探すことになる。候補が絞れておらず、読めない漢字、意味が理解できない言葉がツールに表示されれば、「揺れる」を見つけることは容易ではないだろう。

¹ 筆者が指導する日本語クラスで「自国の交通」というエッセイを書いてもらった際にフィリピン人の学習者から産出された誤用であった。

² 本研究では、実質語同士のつながりのことをコロケーションと呼ぶ。例えば、「本を読む」の場合、名詞「本」と動詞「読む」という実質語で構成されたコロケーションと考える。

³ 筑波大学で構築された 11 億語のコーパス「筑波ウェブコーパス」(Tsukuba Web Corpus: TWC) を検索するためのツールで、検索には、国立国語研究所と Lago 言語研究所が共同開発したコーパス検索システム NINJAL-LWP が使用されている。

2. 先行研究

先行研究では、学習者のコーパスツールの使用がライティングに効果があると報告がされている。Hodošček 他 (2011) は、「なつめ⁴」を使用することが作文の共起表現の産出に有効であるかを検証したところ、意味、構文、レジスターにおいて効果があったと報告している。一方、寺嶋・板井 (2021) はコーパスツールに関わる体系的な教育や支援カリキュラムにつながるような実践報告がないことを指摘し、学習者向けのコーパスツールワークショップでの実践を報告している。このワークショップは基本的なコーパスツールの使い方を学んだ後、ツールを用いたライティング活動を行うもので、「ライティングにおけるコーパスツール活用モデル」(図1)が取り入れられている。

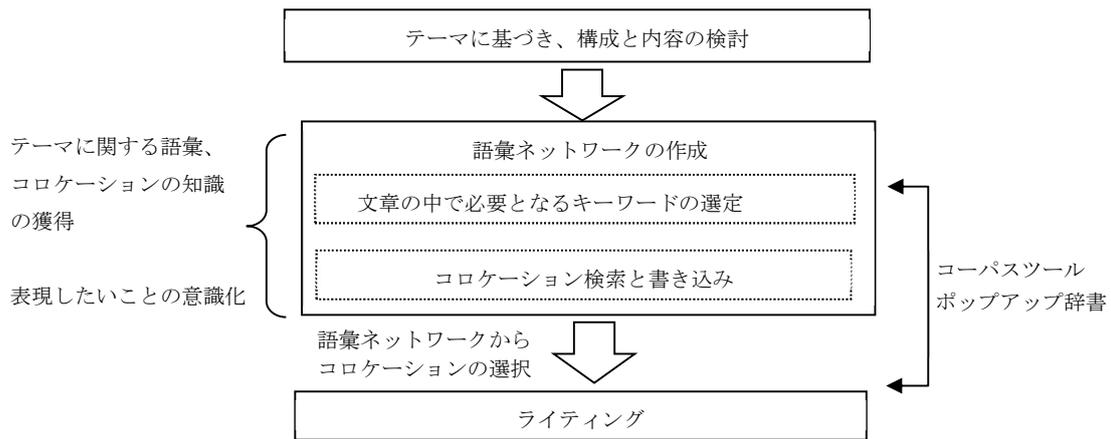


図1 ライティングにおけるコーパスツール活用モデル (寺嶋・板井, 2021)

このモデルは、①コーパスツールで表示される漢字の読み方や意味を理解するために、ツール上で使用できるポップアップ辞書を用いる、②ライティング前にテーマに関する語彙ネットワークを作成し、テーマに関する語彙・コロケーションの知識を獲得する、③語彙ネットワークの作成によって表現したいことを意識化する、④語彙ネットワークからコロケーションを選択し、ライティングを行う、⑤ライティング時にも必要に応じてコーパスツールを使用する、という特徴がある。語彙ネットワークとは、コーパスツールを用いて、キーワードの前後のコロケーション、コロケーションと共に使用される助詞などの文法項目を書き込むものである。たとえば、寺嶋他 (2021) のワークショップでは、「経験」をキーワードとし、前につながる表現として「立派な・貴重な」、後ろにつながる表現として「を積み・を活かす」という書き込みが観察されている。

これまでの先行研究を見ると、さらに調査すべき課題もあると考える。例えば、寺嶋他 (2021) では語彙ネットワークから作文に使用された表現がどのくらいあるか報告されているものの、作文を書いている間にどの程度、コーパスツールが使用されたかは明らかにされていない。また、コーパスツールを用いた場合に、どのような表現が作文で産出されるか、さらに詳しい調査が必要だろう。

3. 本研究の目的

本研究では、先行研究の課題を踏まえ、「ライティングにおけるコーパスツール活用モデル (寺嶋・板井, 2021)」を取り入れ、日本語学習者が作文を書く際のコーパスツールの

⁴ レキシカルプロファイリング機能を持ったコーパスツールの一つである。「なつめ」は日本語学習者向けのツールとして開発されている (阿辺川, 2011)。

使用実態を調査した。具体的には、学習者がどのようにコーパスツールを使用したか、作文でどのような産出をしたかを明らかにした。調査対象としたのは、コーパスツールワークショップの参加者 17 名（中級レベル 2 名、中上級レベル 15 名）であった。対象者の出身の国・地域は、ベトナム 6 名、韓国 2 名、アメリカ、インド、ウズベキスタン、フィリピン、ブラジル、マレーシア、モンゴル、中国、台湾各 1 名である。

4. コーパスツールワークショップ

2020 年 7 月 11 日、ZOOM を用いて、コーパスツールワークショップを行った。参加者は国内の大学の日本語クラスで募集した。学習者の日本語レベルは、中級～中上級レベルであった。ワークショップの目標はコーパスツールの使い方を理解し、「ライティングにおけるコーパスツール活用モデル（寺嶋・板井, 2021）」で 500 字以内の作文を書くことであった⁵。作文のテーマは、「私の将来のキャリア」であった。

ワークショップでは、まず、コーパス及びコーパスツールの説明をし、コーパスツールを用いた言葉の選択・文法の選択の練習、コーパスツールと辞書を用いて意味を理解する練習を行った。次に、語彙ネットワークの説明を行った後、一つのキーワードを提示し、その語彙ネットワークを作成する練習を行った。その後、作文のアウトラインの作成と共有、語彙ネットワークと作文の作成という流れで実施した。

今回、使用したコーパスツールは NINJAL-LWP for TWC（以下 NLT）で、ポップアップ辞書として rikaikun⁶を用いた。参加者は普段、英語で講義を受けていることから和英辞書である rikaikun でも問題がないと判断した。また、各タスクは進捗状況を把握するため、OneNote Class Notebook 上で行った。

5. 調査結果と考察

5.1 コーパスツールの使用回数と使用方法

表 1 は、各学習者が NLT を用いて語彙ネットワークに書き込みした数（以下：「語彙ネットワークへの書き込み数」）、NLT で調べた表現を作文に使用した数（以下：「作文への使用数」）、「作文への使用数」のなかで、語彙ネットワークから使用した表現の数（以下：「語彙ネットワークからの使用数」）を示したものである⁷。

まず、「語彙ネットワークへの書き込み数」は中央値が 10 で、最大値が 35、最小値が 6、「作文への使用数」は中央値が 5 回で、最大値が 14 回、最小値が 2 回であった。また、「語彙ネットワークからの使用数」は中央値が 3 回（最大値：10 回、最小値：0 回）であった。

表 1 からわかるように、今回の調査では、JL8 を除き、最低 1 回は語彙ネットワークからの表現を使用したこと、JL16 以外の学習者は、語彙ネットワークからの表現だけではなく、作文を書いている間にコーパスツールで調べた表現を作文に取り入れたことが確認できる。

表 1 コーパスツールの使用数

学習者	語彙ネットワークへの書き込み数	作文への使用数 (語彙ネットワークからの使用数)
JL1	16	7 (6)
JL2	10	9 (4)
JL3	11	9 (5)
JL4	13	14 (10)
JL5	35	7 (5)
JL6	10	5 (1)
JL7	23	5 (4)
JL8	10	4 (0)
JL9	11	3 (2)
JL10	9	5 (1)
JL11	10	2 (1)
JL12	7	5 (2)
JL13	9	14 (2)
JL14	6	5 (2)
JL15	25	8 (6)
JL16	9	3 (3)
JL17	6	6 (4)

⁵ 語彙ネットワークの作成に 30 分、作文を書くのに 30 分の時間を取った。

⁶ Ereik Speed 氏によって開発された Google Chrome 用の和英辞書アプリで、Google Chrome の画面でカーソルを近づけると、意味がポップアップする。Chrome ウェブストアからダウンロードできる。

⁷ 学習者にコーパスツールを使用した個所に、ハイライトをしてもらった。

図2は、各学習者の「作文への使用数」と作文の文字数の関係を図で表したものである。「作文への使用数」が中央値の5回であった学習者（JL6、JL7、JL10、JL12、JL14）は、JL6を除き、200字から300字の範囲で作文を書いている。特徴的な学習者だったのは、JL4、JL13、JL15である。JL15は文字数が少ないが、「作文への使用数」が8回と多かった。また、JL4、JL13の文字数は中程度であったが、「作文への使用数」が14回と多かった。一方、「作文への使用数」が少なかったのは、JL9、JL11、JL16であった。これらの学習者は文字数も同程度であった。

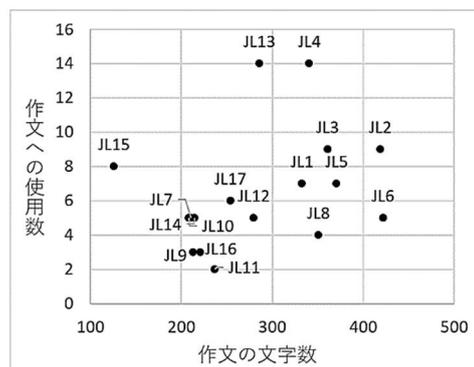


図2 「作文への使用数」と作文の文字数

5.2 使用パターンとコロケーションのレベル判定

次に、学習者がNLTを用いて、どのような表現を作文に取り入れたか分析した。表2は、NLTの分類に従い、使用パターンをカウントしたものである。最も多かったのは、「知識を習得する (JL3)」のような「名詞+助詞+動詞」で構成されたパターンで、全体の77%を占めていた。次いで、「人種差別 (JL8)」のような「名詞」と「名詞」のパターン、「幅広い交流 (JL15)」のような「形容詞」と「名詞」のパターンが多かった。

では、学習者はどの程度の難易度のコロケーションを作文に使用したのだろうか。表3はコロケーションに含まれる2語を「日本語教育語彙表 (日本語学習辞書支援グループ, 2015)」を用いてレベル判定を行い、レベルの組み合わせに基づいてカウントしたものである⁸。

表2 使用パターン

名詞+助詞+動詞	86
名詞+名詞	7
形容詞+名詞	4
形容動詞語幹+な+名詞	3
動詞+名詞	3
名詞+の+名詞	3
名詞+的+名詞	2
名詞+名詞+接尾	2
接頭辞+名詞	1
合計	111

表3 コロケーションのレベル判定

中級前半・中級後半	33
中級後半・中級後半	15
初級後半・中級前半	12
中級前半・上級前半	12
中級前半・中級前半	9
中級後半・上級前半	6
初級前半・初級後半	5
初級後半・中級後半	5
初級後半・上級前半	4
初級前半・中級前半	3
初級前半・中級後半	3
初級前半・上級前半	2
初級後半・初級後半	1
上級前半・上級前半	1
合計	111

最も多かったのは、中級前半・中級後半 (33) のコロケーションの組み合わせで、例えば、「経験を生かす (JL7)」、「文化を尊重する (JL13)」、「関係が築ける (JL1)」といったものが見られた⁹。次いで多かった中級後半・中級後半 (15) の組み合わせでは、「急速な発展 (JL15)」、「発展を目指す (JL15)」、「財産を管理する (JL2)」、中級前半・上級前半 (12) の組み合わせでは、「分析を行う (JL10)」、「将来を見据える (JL16)」、「学術を研究する (JL1)」のようなものが見られた。

⁸ 日本語教育語彙表にない場合は、筆者らでレベル判定を行った。

⁹ 下線部はNLTで検索された言葉

また、初級レベルの簡単なコロケーションも検索されており、初級後半・中級前半 (12) の場合は、「試験に合格する (JL6)」、「良い計画 (JL2)」などがある。最も簡単なコロケーションの組み合わせだと思われる初級前半・初級後半 (5) の場合は、「会社に入る (JL13)」、「勉強を頑張る (JL17)」¹⁰などが見られたが、これらのコロケーションは全て作文を書いている最中に検索されたものであった。

5.3 語用分析

次に作文の中で NLT が使用された部分が適切であるかを分析した。適切さの分析では、NLT を用いて取り入れたコロケーション、あるいはコロケーションと共に使用された文法項目が原因で誤用と判断されたもの (以下:「コロケーションでの誤用」)、コロケーションが使用された節において誤用と判断されたもの (以下:「節での誤用」) に注目した。誤用については、市川編 (2010) の「脱落、付加、誤形成、混同、位置、その他」の分類を用いた。また、節については丸山他 (2016) で示された「主節、従属節:連用節、補足節、連体節」という分類を参考にした。例えば、JL12 は「日本人と交流したくて英語も使える仕事をしたいため、日本で国際的な企業を働きたい。」という文の主節において、NLT で調べた「国際的な企業」という表現を用いている。この場合、「国際的な企業」という表現は適切だが、「日本の国際的な企業で働きたい。」と修正することができる。したがって、「コロケーションでの誤用」はなく、「節での誤用」があると判断した。

今回、「作文への使用数」は全体で 111 あったが、そのうち 40 が「コロケーションでの誤用」、あるいは「節での誤用」が起きていると判断した。ただし、1 つの使用場面でも複数の誤用が起きている場合もあるため、その場合は、複数の誤用としてカウントした。表 4 はタイプごとの誤用数をカウントしたものの、表 5 は実際の誤用例である。

表 4 NLT 使用時における誤用数

コロケーションでの誤用					節での誤用				総数
脱落	付加	誤形成	混同	その他	脱落	付加	誤形成	混同	
4	4	5	14	1	6	2	2	12	50

表 5 NLT 使用時における誤用例

タイプ	誤用例		誤用の原因
コロケーションでの誤用	脱落	十分経験積んだ後で、～ (JL10)	「を」の脱落
	付加	根深い問題について詳細に調べて、～ (JL5)	「の」の付加
	誤形成	日本語に興味を抱きて、(JL7) グローバルなビジネスに入りたい (JL12)	「抱いて」の誤形成 グローバルの誤表記
	混同	ビジネスアナリシスはニーズを定義し、(JL11) 興味を湧くようになれば～ (JL7)	「把握」と混同 「が」との混同
	その他	職場の雰囲気をよく感じて、(JL1)	文脈と合わない表現
節での誤用	脱落	私グローバルに活躍し (JL12)	「は」の脱落
	付加	日にほんごで苦手がなくなる (JL17)	「に」の付加
	誤形成	多いな勉強量が要求される。 (JL6)	「多くの」の誤形成
	混同	グローバルなビジネスに入りたい (JL12)	「携わりたい」との混同

¹⁰ 作文を書いている間に検索されたものであったため、どちらが検索語か判断できなかった。

表4からわかるように、「コロケーションでの誤用」において、最も多かったのは混同である。混同は表5のような言葉の混同（例：定義→把握）や助詞の混同（例：興味を湧く→興味が湧く）以外にも、文体、活用の混同等があった。次いで、多かったのは誤形成で、活用の誤用や誤表記が原因となっていた。一方、「節での誤用」でも混同が多く、誤用の原因も「コロケーションでの誤用」と類似していた。次いで多かったのは脱落で、助詞や形式名詞等の脱落が見られた。

5.4 語彙ネットワークと作文

次に、学習者がどのような語彙ネットワークを作成し、作文でどのような産出をしたかを見てみる。最も「作文への使用数」が多かった一人がJL4であった。図3はJL4が作成した語彙ネットワークで、図4はJL4の作文である。図4の_____（一重下線）は語彙ネットワークから使用された表現で、_____（二重下線）は作文を書いている間にコーパスツールを使い、調べた表現である。JL4の場合、「語彙ネットワークへの書き込み数」は13だが、そのうち10を作文で使用しており、効率的に語彙ネットワークの表現から作文に取り入れたことがわかる。そのため、語彙ネットワークの段階から、作文で何を書くかということが明確であったと考えられる。

	贅沢	贅沢品（良い意味ではないと感じるので、ハイエンドに変わる）
	ハイエンド	ハイエンド製品 ハイエンド市場
	賞金	を稼ぐ earn awards を獲得する achieve 100万円の賞金
	自習	に取り組む make effort に明け暮れ all the time, morning and evening
優れた 取材した	ルポルターージュ	活動
	キャリア	を積み上げる を磨く

図3 JL4の語彙ネットワーク

私が積み上げたい将来のキャリアはコンテンツストラテジストとして活動することである。具体的には、女性のファッション雑誌にハイエンド勢品の広告を載せるという仕事に携わりたいと思う。ゼミに入ってから、日本の雑誌についての研究を始める予定である。卒業後、日本のハイエンド市場で経験を十分に積んでから、ベトナムに帰国したいと思っている。なぜ贅沢品のコンテンツストラテジストになりたいかというと、キラキラなものを見ることが大好きからだ。_____に入学した後、様々な賞金を稼ぎ、選者の前に自分のキャリアについて発表したり、エッセイで書いたりした。最近、100万円の賞金を獲得し、将来のキャリアに進んだ。今後キャリアを磨くため、自習に明け暮れ、自分でハイエンド製品に関するルポルターージュ活動を行いたいと思う。

図4 JL4の作文

一方、語彙ネットワークからの使用がゼロだったのがJL8である。図5を見ると、JL8は2つのキーワードについて様々な表現を書き込んでいることがわかる。しかし、語彙ネットワークに2つのキーワードしかなかったこと、作文前に語彙ネットワークの表現をどのように使用するか十分にイメージしていなかったことから、作文を書いている間に使用する機会が見つけられなかったと考えられる。

苦労した、くろうした、 troubles・hardship 幅広い、はばひろい、 extensive	経験	が浅い、があさい have little experience が乏しい、がとぼしい、meagre が不可欠だ、がふかけつた、 indispensable
誤った、あやまった、 to misunderstand 優れた、すぐれた、 excellent	理解	を深める、をふかめる to have a better understanding を促進する、をそくしんする、 to encourage を促す、をうながす、 to prompt

図5 JL8の語彙ネットワーク

将来、政府で働き、外交官になりたい。なぜ外交官になりたいかというと、中学生からずっとネットで様々な国の人と話していた。その人々から様々な文化や見方を習い、非常に面白く、楽しかった。「なぜその人をそのように考えるか」「なぜそのようなことがわるい」な考え方をできおうした。異文化や異なる意見を理解することが初めてでき、自分が手に届く世界が広がった。その後、旅行の興味も引き付けた。美味しいな食べ物を食べたり、美しい景色を見たり、伝統的な服を着たり、歴史的な場所を行ったりすれば、ある国の文化や習慣を理解ことをできる可能性があるはず。しかしながら、世界中で戦争や人種差別などがあり、楽しく旅行と異文化を習うことが時々ない。そのため、卒業後、アメリカの政府の中にインターンシップし、様々な経験を積み重ねるつもりだ。

図6 JL8の作文

JL4 や JL8 以外にも特徴的な学習者がいる。例えば、JL13 の場合、語彙ネットワークの使用が 2 回と少ないが、作文を書いている間に調べたものが 12 回あった。JL13 が語彙ネットワークから用いたのは、「文化を尊重する」、「文化を取り入れる」といった中級前半から中級後半の表現であったのに対し、作文を書いている間に調べたものは、「店を開く」、「会社に入る」、「菓子を売る」などで、初級レベルの簡単な表現であった。一方、作文の文字数が 126 文字と少ないが、8 回使用したのが JL15 であった。JL15 の場合、「幅広い交流」、「交流を促進する」、「急速な発展」、「発展を目指す」、「貢献を果たす」など、中級後半から上級レベルの表現が多く使用されていた。また、これらの表現は全て、語彙ネットワークから取り入れられたものであった。

全体を見ると、作文を書いている間にコーパスツールを使用した場合、初級レベルの簡単な表現を探す傾向が強かったわけではない。また、簡単な表現でも、学習者にとって必要な検索であれば、ツールの機能が活かされていると考える。しかし、JL13 のように語彙ネットワークを作文に活かせないという問題については、改善方法を検討するべきだろう。

6. まとめと今後の課題

本研究では「ライティングにおけるコーパスツール活用モデル(寺嶋・板井,2021)」を用いた場合、学習者によって差異があるものの、作文で使用された表現の中に、語彙ネットワークから取り入れられたもの、作文を書いている間に検索されたものがあることが確認できた。また、作文で最も使用されたのは、「名詞+助詞+動詞」のパターンで、中級前半レベルと中級後半レベルの言葉で構成されたコロケーションであった。さらに、その適切さを分析したところ、コロケーション、あるいはコロケーションと共に使用された文法項目が原因で誤用と判断されたもの、コロケーションが使用された節において誤用と判断されたものがあり、「混同」による誤用が多いことがわかった。

学習者の誤用を見ると、コーパスツール上で表記や活用、助詞、さらには用例を確認する習慣を身に付けておくことで誤用を避けられるものもあっただろうと考える。学習者がそうした観察や分析が自然にできるようになれば、より自律的な学習が可能になる。そのためには、継続的に使用し、コーパスツールの使用に慣れることが必要だろう。

また、学習者の中には語彙ネットワークに書き込んだ表現をうまく取り入れられる者とそうではない者が見られた。このモデルの有効性をさらに高めるためには、語彙ネットワークの段階で作文がイメージできるようになる必要がある。例えば、語彙ネットワークから取り入れた表現を使った短文を作り、その短文を作文に使用するというプロセスを加えることで、より語彙ネットワークからの表現を取り入れられる可能性があるのではないだろうか。今後はそうした改善を加えたモデルについて検証をしていきたい。

謝 辞

本研究は JSPS 科研費(18K00701「コーパスツールユーザーを育成するための教材開発とその検証」)の助成を受けたものである。

文 献

- 阿辺川武・ホドシチェク ボル・仁科喜久子 (2011) 「語の共起を効率的に検索できる日本語作文支援システム『なつめ』の紹介」言語処理学会 第17回年次大会 発表論文集, pp. 595-598.
- 赤瀬川史朗・プラシヤント パルデシ・今井新悟 (2016) 『日本語コーパス活用入門』大修館書店.
- Hodošček Bor・阿辺川武・Bekeš Andrej・仁科喜久子 (2011) 「レポート作成のための共起表現産出支援－作文支援ツール『なつめ』の使用効果－」専門日本語教育研究 13号, pp. 33-40.
- 市川保子 (編) (2010) 『日本語誤用辞典 外国人学習者の誤用から学ぶ日本語の意味用法と指導のポイント』スリーエーネットワーク.
- 丸山岳彦・佐藤理史・夏目和子 (2016) 「現代日本語における節の分類体系について」言語処理学会第22回年次大会発表論文集, pp.1113-1116.
- 日本語学習辞書支援グループ (2015) 「日本語教育語彙表 Ver 1.0 (<http://jhlee.sakura.ne.jp/JEV/>よりダウンロード可能)
- 寺嶋弘道・板井芳江 (2021) 「ライティング力の向上を目指したコーパスツールワークショップの試みとその考察」APU 言語研究論叢 6号, pp. 49-65.

関連 URL

- Erek Speed 『rikaikun』 <https://chrome.google.com/webstore/detail/rikaikun/jipdnfibhldikgcjhfnomkfpcebammhp?hl=ja>
- 筑波大学・国立国語研究所・Lago 言語研究所 『NINJAL-LWP for TWC』 <https://tsukubawebcorpus.jp>