

以 Google n-gram 語料庫為本的「日語搭配詞輔助學習系統」之開發研究與成效評估

施列庭

開南大學應用日語學系副教授

摘要

學習日語之際，大量的詞彙習得是不可或缺的，然而在現今的課程安排中，並沒有規劃充分的時間給詞彙教學，在上課時進行適當的詞彙學習指導有其窒礙難行之處，因此在授課時間以外進行自主學習便成為必然的選項。針對這樣的學習型態，也有學者提案利用「基本語彙」來讓學習者進行自主學習，不過無論是否有無教師在旁指導的語彙學習，要如何使學生在增加詞彙數量的同時，不要忘記已經學習的詞彙，依舊是件困難的挑戰。

姉崎(2006)曾指出搭配詞是提高學生詞彙運用能力的最有效率指導手法。因為透過語料庫裡高頻率的搭配詞學習方式有助於詞彙區塊的習得，因此使詞彙變得容易記憶。然而在台灣，有關於透過搭配詞來學習詞彙的研究尚屬少數。

因此，本研究將著眼於提升台灣學習者的詞彙運用能力，針對如何更有效率地輔助搭配詞彙學習，提出了以 Google n-gram 語料庫為基礎的「日語搭配詞學習輔助系統」的開發研究方案。希望透過本系統除了提供學習者詞彙學習的搭配詞彙資訊外，也能夠強化其詞彙應用的能力。

關鍵詞：CALL、 搭配詞學習、 語料庫、 n-gram

受理日期：2018.08.31

通過日期：2018.11.09

Development and evaluation of 'Japanese collocation learning support system' based on Google n-gram corpus

Shih, Lieh-Ting

Associate Professor, Kainan University, Taiwan

Abstract

In the case of Japanese learning, a great amount of word acquisition is indispensable. However, there is no sufficient time given to vocabulary teaching in today's curriculum. It is difficult to carry out appropriate word teaching in class. Therefore, the type of learning outside of the time of instruction becomes an option. In view of this type of study, some scholars have proposed the use of "basic vocabulary" to assist learners to learn autonomously, but whether or not learning with or without the guidance of a teacher, how to make students increase the number of vocabulary and keep the studied vocabulary in memory at the same time, is still a great challenge.

Anesaki (2006) has pointed out that the use of collocation is the most efficient to improve the ability of students to use vocabulary. It is because learning through high frequency word in corpus facilitates the acquisition of lexical blocks, therefore it makes words easy to remember. However, in Taiwan, there is only a small number of studies on Japanese collocation study.

Therefore, this study will focus on improving the vocabulary utilization ability of Taiwanese learners, and put forward the research and development of Japanese language learning system based on Google n-gram corpus for the more efficient method of learning with vocabulary. It is hoped that the system will provide learners with related collocation information, and to strengthen students' ability on word usage.

Keywords: CALL, Collocation, Corpus, N-gram

Google n-gram コーパスに基づく「日本語コロケーション 学習支援システム」の開発研究と評価

施列庭

開南大学応用日本語学科准教授

要旨

日本語の学習に当たり、大量の語彙習得が欠かせないが、十分な時間が割り当てられていない現在のカリキュラム体制の中では、授業中に適切な語彙指導をすることは難しいから、授業以外の空き時間を使った学習形態が必然的に考えられる。こういう学習形態に応じ、「基本語彙」による学習者の自律学習を提案した学者もいたが、教師による語彙指導の有無を問わず、語彙数を増やすと同時に学習した単語を忘却しないための手法を確立することが困難なものだ。

語彙の運用力を高める指導法としてコロケーションが最も効率的であることを姉崎が指摘した(2006)。それはコーパスにおける高頻度のコロケーションによる学習法が語彙チャンクの習得に役に立つから、記憶に残りやすいのではないかと考えられる。しかし、台湾では、コロケーションによる語彙習得に関する研究はまだまだ多いとは言えない状況である。

そこで本論文では、台湾学習者の語彙運用力の向上に着目し、語彙コロケーションの習得を効率的にサポートする方法について Google n-gram コーパスに基づく日本語コロケーション学習支援システムの開発を提案した。このシステムは語彙学習のためのコロケーション情報を提供すると同時に、学習者に語彙使用に関する運用力の強化につなげていきたいと考えている。

キーワード：CALL、コロケーション、コーパス、N-gram

Google n-gram コーパスに基づく「日本語コロケーション 学習支援システム」の開発研究と評価

施列庭

開南大学応用日本語学科准教授

1. はじめに

日本語の学習に当たり、大量の語彙習得が欠かせないが、十分な時間が割り当てられていない現在のカリキュラム体制の中では、授業中に適切な語彙指導をすることは難しいから、授業以外の空き時間を使った学習形態が必然的に考えられる。こういう学習形態に応じ、「基本語彙」による学習者の自律学習を提案した学者もいたが(進藤;2002)、教師による語彙指導の有無を問わず、語彙数を増やすと同時に学習した単語を忘却しないための手法を確立することが困難なものだ。

日本語の学習において、読む・聴くといった能力を向上させるためには、できる限り多くの語彙知識を習得することが必要となる一方、語彙の運用力を高めることも重要である。なぜなら、学生にとって語彙不足のために、授業の中で自分が言い表そうとしているものは自由に表現できるかどうか不安な感じを抱え、授業についていけなくなる問題が増えてきた。

語彙の運用力を高める指導法として単語の意味提示、文による提示、コロケーションによる提示を比較した結果、コロケーションが最も効率的であることを姉崎が指摘した(2006)。それはコーパスにおける高頻度のコロケーションによる学習法が語彙チャンクの習得に役に立つから、記憶に残りやすいのではないかと考えられる。しかし、台湾の日本語教育では、コロケーションによる語彙習得に関する研究はまだまだ多いとは言えない状況である。

そこで本論文では、台湾学習者の語彙運用力の向上に着目し、語

彙コロケーションの習得を効率的にサポートする方法についてコーパスに基づく日本語コロケーション学習支援システムの開発を提案する。このシステムは語彙学習のためのコロケーション情報を提供すると同時に、学習者に語彙使用に関する運用力の強化につなげていきたいと考えている。

2. 本研究における目的

本研究は以下のような課題を解決することを目的としている。

1. 言語学習理論及び自然言語処理技術に基づき、グーグルの n-gram コーパスを利用することにより、デジタル化した言語学習支援カリキュラムの作成を行う。
2. 日本語コロケーション学習支援環境を構築し、学習者に多様な学習支援を行う。
3. 授業や教材作成の参考に、日本語語彙の使用におけるコロケーションの特徴や出現頻度などの情報を表示する日本語コロケーション予測とコロケーション検出機能を有するモジュールを開発する。
4. 教室指導と e-learning 自主学習環境との組み合わせによって、日本語コロケーションへの学習者の知覚能力が向上できるかどうかを測る。

3. 先行研究

3.1 台湾人日本語学習者における語彙使用問題

台湾では、ここ数年作文学習における語彙使用問題についての綿密な研究が行われてきた(馮, 2005; 陳, 2006; 葉, 2006a; 鄧, 2007)。それらの研究は、語彙の選択を苦手とする学生には語彙の誤用における作文学習の支障に悩んでいる人も少なくないことから、効果的に語彙の使用を指導する必要性があるということを明らかにした。

例えば、作文に於ける日本語母語話者と台湾人学習者の間にある

使役表現の差異を検討する際、日本語母語話者よりも、台湾人学習者のほうが他動詞の使役表現と語彙の選択に誤りが多いと馮(2005)が指摘した。こうしたエラーは学習者がどれが正しい文法形式や語彙であるかという選択に迷った結果ではないかと考えられる。

作文における初級日本語学習者の語彙学習の様子を分析したところ、陳(2006)は使用頻度の高い語彙は学習者にとって素早く習得でき、正しく習得できるカテゴリであることを発見し、形容詞、副詞、接続詞、連体詞及び連語などの使用は限られていたと指摘した。また、作文指導に教師が添削により語彙表記や文法活用などの学習者の問題を一部解消できるが、助詞や文構成の面では効果的な指導は難しいとも説明した。助詞の誤用について鄧(2007)はさらなる研究を進めた。日本語の学習時間が五年以上である学習者を対象に、作文に出た助詞の誤用について分析した結果、助詞「が」、「に」について初中級の学習者のみならず、苦手とする上級学習者も少なくない。それに、助詞「に」、「は」についても似たような問題がある。その中、「で」が一番誤用が多い助詞であると指摘された。

上述した助詞と語彙の誤用は、葉(2006a)が指摘したように、学習者が作文学習に挫折感を持つ原因の一つだと考えられる。言い換えれば、学習者自身による語彙知識の不足が作文学習を順調に進められない主な原因であると言えよう。その一方、このような結果をもたらしたのは、現今の語彙教育が意味、書き方及び発音といった語彙理解のみに焦点を当てていたことと、語彙の使い方にかかわる語彙運用の指導が非常に不足していることであろう。

こうした語彙誤用問題を解消するために、コロケーションによる学習指導が望ましいとされている。堀(2009)によると、コロケーションとは「語と語の間における語彙、意味、文法的などに関する習慣的な共起関係」を定義している。すなわち、語彙の習得に当たって文の中で語彙が語彙的、意味的、文法的な側面からどのように他の語彙と組み合わせ関係を持つかを理解しておかなければ、語彙の

運用力を活かして文章を書いたり、話したりしようとする際、多くの学習者にとっては、日本人母語話者らしい内容を産出することが大変困難であるといえる。

3.2 自然言語処理技術による日本語語彙学習支援研究

日本語学習において学習者が適切な語彙使用に悩みを抱え、結果として学習が効率的に進められないといった問題について、自然言語処理技術により作文への学習支援を提案した研究がいくつかある。例えば、赤堀・楊(1999)は論文の作成に取り組む理科系の学習者のために、自然言語処理技術を通し、学習支援システムを構築した。システムの特徴として、学習者の書いた内容から文法的に不自然な項目を抽出し、学習ヒントを提供することが挙げられる。これに対して、高林(2000)は作文を書く途中に、適切な学習指導を行う提案「Namazu」をした。「Namazu」はキーワード検索技術をもとにし、語彙や文章に使われる表現を示す学習支援システムである。一方、語用論の面から語彙の意味や関連した情報を示したりキーワード抽出による似ている表現を実例とともに提示したりする作文の学習支援システムを開発した研究者もいた(鄭・三宅・馬越・赤間, 2006; 吉川・田中, 2007)。

自然言語処理技術による語彙学習支援研究の中で大規模コーパスと単語 N-gram 統計モデルを活用した授業研究が近年盛んに行われている(高・小高・小倉, 2001; 山本・森, 2002; 橋本・安藤・竹内, 2008; Nazar&Renau, 2012; 黄・呉・許・顔・張, 2014)。単語 N-gram とはコーパス等大規模のテキストデータから学習した統計モデルの一種であり、単語間の接続確率に基づきテキストを幾つかの区切りをつけ文章の特徴を示したものである。「明日学校に行く」というテキストを bi-gram で分割処理すれば、「明日学校」、「学校に」、「に行く」という 3 つのチャンクが得られる。コーパスの大きさによってこの 3 つのチャンク(単語間の接続確率)が違うので、グーグルコーパスのような規模のテキストデータがあれば、単語間の実際の接続確率

を示す可能性が高い。

高・小高・小倉(2001)は EDR Japanese コーパスと n-gram 統計モデルを用い、日本語学習者の作文について内容を評価したり、テキスト特徴を抽出したりする研究をした。Tri-gram を基準に日本語母語話者の作文(37本) 中国出身の日本語学習者の作文(55本) を n-gram 形態分布の類似性を比較した結果、こうした類似性は日本語学習者の書いた文をどれほど母語話者の作文に似ているかを知る上で有効な指標となることがわかった。一方、単語 N-gram 統計モデルを活用し、文法誤用の分析について試みを行った研究者もいる。例えば、Google Books N-gram 大規模コーパスのテキストデータを活用し、Nazar&Renau (2012)はスペイン語の文法チェックのためのツールを開発した。スペイン語の字形変化や穴埋めなどの手法でシステムの有効性を検証した際、MS word でさえ検出できなかった語彙の不規則な複数形の誤用を検出することに成功した。その一方、bi-gram による誤用検出の成功率は予想より低かったことがわかったから、他のコーパスを利用したり、tri-gram のような長いチャンクを使ったりして成功率の向上につながると指摘した。

コーパスと単語 N-gram 統計モデルの活用例として挙げられるのは黄・呉・許・顔・張(2014)の論文作成提示における研究である。英語を母語としない学習者を対象に、学習者の入力した内容をキーワードにし、その後に来る接続可能な単語を提示する論文作成支援システム WriteAhead を開発した。例えば、“In this paper,” を入力すれば、“we present”, “we describe”, “we explore”, “we propose”, “we will”, “we show” など幾つかの接続可能な文を作成内容の参考として示された。

これらの研究から日本語の作文や語彙学習指導の際にあった問題を解くのに取り組んで、教育工学技術を導入したり様々な工夫をこらしたりする研究者の姿が伺える。しかし、教師の教育理念と学習指導を無視し、システムによる一方向的に学習支援を行おうとして

も、学習者による自発的な学習モチベーションを高めることが期待できないだろうと考えられる。

3.3 コーパスによる発見的外国語学習

外国語学習指導をする際に教師が文法規則や語彙の使い方を一方的に教えるのではなく、学習者が自己発見によりルールをまとめたり、帰納したりするように展開する学習法が多くの研究者に研究、実践されている。こうした学習法はいわゆる「発見的学習(discovery learning)」と言われ、ブルナーが提唱したもので、上述したコーパスによる外国語学習に応用した場合に、一部の既知の語彙や文法規則をベースに学習者に探索させる過程を通じ、言語知識の再構築をさせようとするものである。発見的学習は時間がかかりすぎる欠点があると指摘されているが、教師主導の授業に比べて学習者を中心とする学習法であるため、その意欲がより高く、学習者自ら獲得した知識が記憶されやすい長所がある。

学習者のニーズに合わせた指導の一方策としてコーパスによる発見的言語学習研究が近年行われている。例えば、施(2014)はDDL(Data-Driven Learning)理論と発見的学習のコンセプトに基づき、コーパスによる日本語漢字発音学習支援システムを開発した。このシステムの特徴として挙げられるのは学習者が「共同筆記」への記入活動という協同学習機能を通じて、日中漢字の音韻対応関係の促音化ルールを能動的に探索できることである。また、こうした学習者による協同学習は、学習情報の共有により学習負担を減らすだけでなく、学習モチベーションを維持する効果もあるとわかった。

また、英語語彙の間に潜んでいる文法規則を発見的に習得させることを目的として、中條・西垣・内山・山崎(2006)は日英パラレルコーパスに基づく語彙と文法の学習支援システムを開発した。後にアンケートによるシステムの有効性を検討した場合、学習者がこうした発見的学習法への満足感および学習モチベーションが向上する

ことがわかった。

上述した結果からもわかるように発見的外国語学習は時間がかかりすぎる欠点があるにもかかわらず、教師の適切な指導のもとに、授業外にコーパスによる外国語学習に応用した場合、「共同筆記」などを通じて学習負担と時間を減らすだけでなく、学習者に能動的な探索をさせることになり、言語知識の定着が良くなるだろうと考えられる。

これらの先行研究の成果を踏まえ、学習者が自ら発見的学習を行うことで、語彙における使い方の理解を促すことを目的とし、語彙応用能力の向上が可能になるように、グーグル n-gram コーパスに基づく学習支援環境の開発を行うことにした。

4. 日本語コロケーション学習支援環境の構成

本研究で提案した支援環境に使われているシステム及びサーバーは以下の表に示されるものである。学習支援環境の開発においてクライアントには学習者が一般的なパソコンを用い、操作するインタフェースには Flash を使用しているのに対して、システムの大部分を Linux で構築したデータベース・サーバー、及びウェブサーバーにより処理する。

表 1 システム開発環境

項目	名称
メインシステム	UBUNTU 12.04
データベース・サーバー	MySQL 5.3
ウェブサーバー	Apache Web Server
開発ツール	Flex builder, Mecab, ssgnc

本研究の提案した学習支援環境の利用は以下図 1 に示すように、学習者は教師の指導の元、語彙の運用規則を習得する場合、まずい

くつかの学習支援モジュールを通し具体的な語彙のコロケーションを検証しながら、共同筆記などの機能により語彙に関わる実例を記入し全ての学生と共有する。学習支援環境の外観は図2に示したものである。

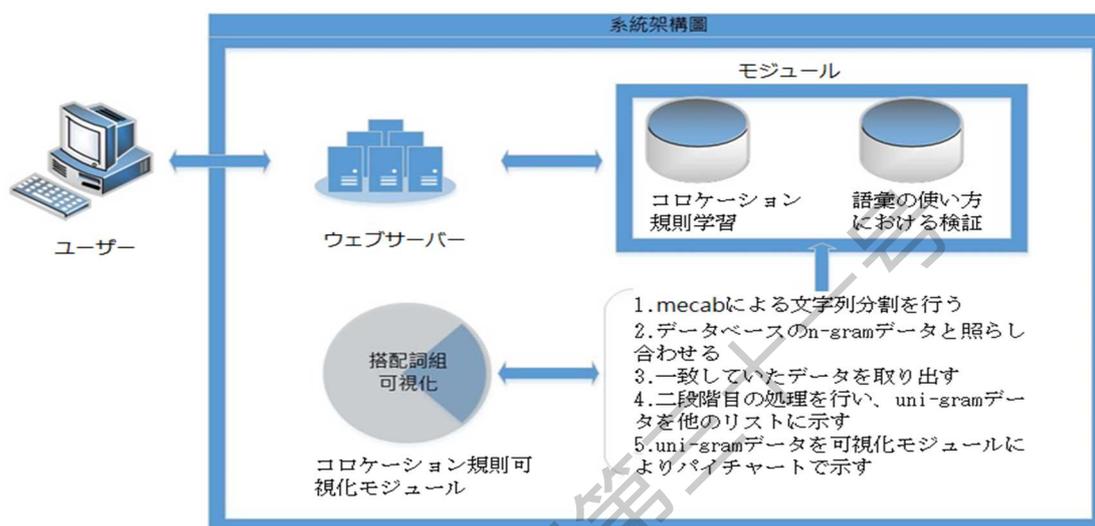


図1 学習支援環境の構成



図2 学習支援環境の外観

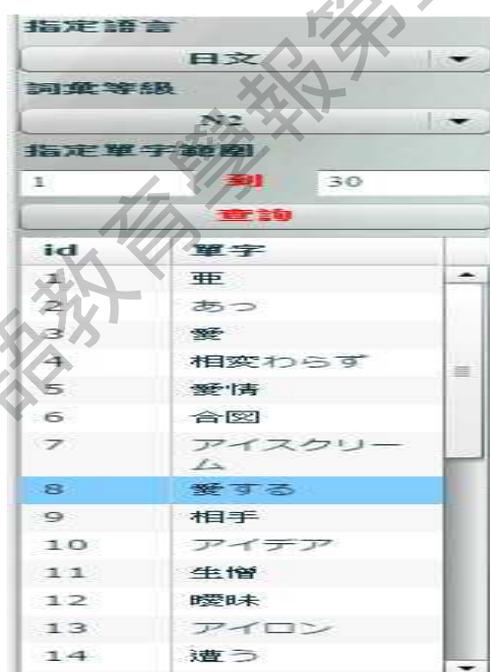
4.1 学習支援モジュール

本学習支援環境は語彙におけるコロケーション学習支援において規則学習モジュール、規則可視化モジュール、語彙の使い方におけ

る検査モジュールの3つの学習支援部からなる。学習者は自らこれらのモジュールを利用し、語彙の係り受け構造を理解し、発見的学習で語彙における運用能力を培う。以下に各モジュールの構成及び作動方法を簡単に説明する。

4.1.1 コロケーション規則学習モジュール

学習上の利便性を確保するために、コロケーション規則学習モジュールには日本語能力試験に関するN1からN4までの語彙リストが設けられている。こうした語彙の実例リストという情報のほか、語彙コロケーション学習のために、データベース、検索方向、文の長さ及び検索範囲の指定など、様々な設定をも提供されている。



id	単語
1	車
2	あつ
3	愛
4	相突わらず
5	愛情
6	合図
7	アイスクリーム
8	覚える
9	相手
10	アイデア
11	生憎
12	曖昧
13	アイロン
14	遭う

図3 旧日本語能力試験語彙リスト

図3に示したのは、旧日本語能力試験一級から四級までの語彙リストである。学習者自身のレベルに合わせ、単語の使い方や規則について本システムによる自律学習ができるように設けられている。例えば、学習者が単語「覚えて」を選択し、検索ボタンをクリック

すると、図4のような「敢えて」に関するコロケーションのリストが示されている。リストにはそれぞれの語彙のコロケーションの頻度や比例などの情報が示され、特定語彙表現が頻出するコンビネーションを示唆する。こうしたコロケーション規則学習モジュールにより、学習者に語彙の使用知識を身に付けさせることが期待される。

ID	可能接続搭配詞組	頻率	比例
1	敢えて言えば、	6680	8.58 %
2	敢えて言うなら、	6520	8.37 %
3	敢えて言わせて	6130	7.87 %
4	敢えて書きません	4080	5.24 %
5	敢えて言おう、	3940	5.06 %
6	敢えて言うならば	3620	4.65 %
7	敢えて言おう。	2260	2.9 %
8	敢えて言いません	2220	2.85 %
9	敢えてザックリと	2190	2.81 %
10	敢えて書かせて	2090	2.68 %

図4「敢えて」に関するコロケーションのリスト

4.1.2 コロケーション規則可視化モジュール

上述したコロケーション規則学習モジュールを使ってから、学習者に語彙の使い方を素早く習得させるようにコロケーション規則発見の促進を目指して、コロケーション規則可視化モジュールには図5から6までのように、特定の単語に関する係り受け関係のある語彙リストの集計表とパイチャートによる可視化が設けられている。学習者はこれらの可視化図を通し、ある単語に係り受け関係を持つ規則について、「頻度」や「比例」の点から観察したり、確認したりすることが可能である。学習者にとってこうした具体的な可視化図による発見的学習は語彙の使用への理解を深める効果的な手段として語彙のコロケーション規則の把握に役に立つと考えられる。

彙の使い方における検証モジュールには、学習者が書いた文の適切さを確認する機能が備えられている。n-gram アルゴリズムに基づき、構築されたこのモジュールは学習者の入力した文をデータベースの n-gram データを照らし合わせながら、一致していない項目を示すことが可能である。学習者がそれらの項目を選ぶと、該当する文字列に関する共起関係のある情報がただちに提示されている。例えば図 7 のように、「生活水準が進歩する」という文が入力されていた場合、左側のリストには「生活 水準 が 進歩」、「水準 が 進歩 する」という適切さを欠くと判定された文字列が示された。「生活 水準 が 進歩」がさらに指定されていたら、右側に「生活水準が向上」、「生活水準が高い」、「生活水準が低い」などの「生活水準が」と共起しやすい文字列がリストに出力される。

こうしたシステムが提示した情報を自分の入力した文と照らし合わせて検証することにより、学習者は語彙の正しい使い方を覚えたり間違いを自己発見で確認したりすることができるために、更なる効果的な学習成果が期待できると考えられる。

図 7 語彙の使い方における検証モジュールのスクリーンショット。左側に検出結果のリスト、右側に可能接続搭配詞組の表、右上に句構分析の図が示されている。

ID	比對不符合項目	ID	可能搭配詞彙	次数
1	生活 水準が 進歩	8	上がる	1
2	水準が 進歩 する	2	上昇	1
		3	維持	1
		4	低く	1
		5	上がり	1
		6	下がる	1
		7	下がつ	1
		1	遠う	1
		9	低下	1

ID	可能接続搭配詞組	頻率	比例
1	生活 水準が 向上	1850	22.14 %
2	生活 水準が 高い	1100	13.16 %
3	生活 水準が 低い	846	10.12 %
4	生活 水準が 上がつ	774	9.26 %
5	生活 水準が 高く	610	7.3 %
6	生活 水準が 低下	525	6.28 %
7	生活 水準が 上がる	485	5.8 %
8	生活 水準が 下がつ	328	3.92 %
9	生活 水準が 下がる	328	3.92 %

図 7 語彙の使い方における検証モジュール

また、こうした検証モジュールはどのくらい学習者の語彙の誤用を検出できるかを調査するため、「日本語学習者誤用換言対コーパス¹⁾」を用いて語彙の選択が間違っている文について間違い検出テストを実施し、提案システムの有効性を検討したところ、検証モジュールは誤りの検出率がおよそ 92%となり、即ち 380 の誤用例のうち、350 例を検出できることがわかった。また、誤用例の検出率が学習者の国籍によってそれぞれ違う。誤用例個数と検出率を併せて見れば、ベトナム、中国、タイ、インドなどといった誤りの多い国の学習者の場合、提案システムの検出率が 87%から 95%までの範囲にあることがわかった。この結果、提案システムの検証モジュールにおける有効性が示唆された。

表 2 国籍別による誤用例の検出率

国籍	個数	未検出の個数	検出率 (%)
中国	75	10	86.67
カナダ	4	1	75.00
フランス	5	2	60.00
インド	20	1	95.00
マレーシア	1	1	0.00
ロシア	2	1	50.00
タイ	25	2	92.00
トルク	3	1	66.67
アメリカ	14	1	92.86
ベトナム	173	10	94.22

5. 研究方法

本研究では、e-learning でコロケーション学習支援を可能とするシステムを開発し、学習者に授業参加してもらい、共同筆記への記入をすることで、学習者のコロケーションへの知覚能力の向上促進

¹⁾ このコーパスは主として松田真希子、松田佳子は金沢大学留学生センターの受講生が提出した作文に含まれていた誤用データをまとめ、作成したものである。この研究では誤用例 2,789 事例の中から、「語彙選択」の誤りに焦点を絞り誤用例を 380 項目抽出し、提案システムの有効性の検討に用いた。

が可能かどうかを質問紙によって調査し、e-learningでのコロケーション学習の効果を検証した。

5.1 実践環境と対象

参加者は台湾北部の大学で日本語を学習中である中国語母語話者91名の大学生2～4年生である。通常の教室授業外に約30分間のe-learning自主学习を1学期にわたり、1週間に1回の割合で16回実施した。

5.2 実践

貢献者	用法	意思	搭配用法の特徴分析	単字表現
王〇輝	やっと見つけた	終於發現	後面常接多種動詞的過去式，如見つけた、終わった，表示某一現象發生或結束了	やっと～
賴〇廷	力を貸し	出力	前面常接お金、力、手、知恵等，多用於請求他人協助。	～を貸し
黃〇翔	めったにありません	幾乎沒有	頻率副詞，後面常接否定語，例如：ありません、動詞ナイ型	めったに～
林〇佑	家賃が払える	付得起房租	後面常接高い、安い等形容詞；動詞多用払える，另有下がり、上がり表房租調整。	家賃が～
吳〇鈴	痛いのを我慢して	忍耐著疼痛...	前面通常接名詞，例如：痛いの、眠いの	～を我慢
游〇萱	なかなかできない	很難以做到	後面可以加上肯定或是否定，なかなか+肯定表現，表示相當～的意思，なかなか+否定表現，表示怎麼也不～的意思。	なかなか～
鄭〇靖	春はもうすぐ	春天將要...	以現在的時間為基準點，不久的將來發生的事情或某種變化。例如，前面常接：春は、夏は、今年も	～もうすぐ
陳〇龍	休みのはず	應該是放假	前面可以接動詞原形及名詞。はず意為應該，是說話者根據主觀、過往經驗，推測某事物可能發生的結果。例如：間に合うはず、	～のはず
黃〇琪	写真を見せて	讓～看相片	前接要給他人看的事物，在給他人看的東西前面可加「B+に」，即「(A)讓B看～」。對於B來說：「不是他主動去看的」又或「B沒主權要看就看、必須由A來讓B看」	～を見せて
李〇豪	日当たりが良い/悪い	日照良好/不好	用法單純，大多只在後面接良い或悪い來形容此地日照是否良好。	日当たりが～
陳〇蓉	バカンスに行つて	去休假	後面常接行つて、出かけて、来て	バカンスに～

図8 共同筆記の一部内容

コロケーションの学習指導においては授業前、授業中、授業後という三段階で進められた。授業前に教師は、学習者全員が共有でき、クラウド内に自動的に保存されるGoogle Sheetを「共同筆記」として学習者の名前及び語彙表現のリストを記入しておいた。授業中にクラス全員に共通に学習指導を行い、「共同筆記」にある語彙表現を取り出し、実際にどのように使われるのかを説明し、語彙の接続表現に対する注意を喚起させるような提示を行い、授業後の自主学习への動機付けの効果を高めるような工夫をした。また、授業終了時

には、共同筆記にある語彙リストに基づき、特定語彙に関するコロケーションのデータをまとめるように学習者に指示を出した。授業後に学習者各自が Google アカウントで「共同筆記」(図 8)に登録し、指定された語彙表現をもとに、コロケーション学習支援システムから頻出するコロケーションを選び、特定の語彙には共起しやすい表現や特徴をまとめ、様々な実践を行ってきた。

5.3 質問紙調査と分析方法

コロケーション学習支援による実践終了後、オンラインアンケートを用いて参加者全員に語彙学習の内容、コロケーション学習支援等、計 12 項目について「そう思う」(5)～「そうは思わない」(1)の 5 段階尺度により評定させ、最後に質問 13 として自由記述による調査もあわせて行った。

アンケート調査では、学習者 91 名から得られたデータを主因子法による因子分析を行い、バリマックス回転を行った。また、自由記述については、学習者の考えを明らかにするために、KH コーダーを用いてテキストマイニングを行い、頻出語の整理、クラスター分析、対応分析及び共起ネットワークの作成により語と語の関係性を表す概念の可視化を行った。

5.3.1 因子分析

5.3.1.1 項目分析

アンケート調査の質問項目の弁別性を因子分析の前に確認するために、各項目の臨界比 (Critical Ratio) の値を項目分析により評価した。具体的な手順としてまず項目の値を再割当てし、上位 27% と下位 27% というグループ分けで独立した t 検定により 5% 水準で検討し、有意差が見られなかった項目を弁別性のないものと見なし、削除することにした。本研究における項目分析を行った結果、すべての項目には統計的有意差が認められた。

5.3.1.2 因子分析

主成分分析による因子分析を行い検討した。バリマックス法による回転で、3回の反復で収束した。因子負荷量が一因子につき0.5以上を基準に12項目から2つの因子が抽出された。結果、分散の全体における貢献度は69.639%である。

標本の妥当性について観測相関係数の大きさと偏相関係数の大きさを比較するKM0による測度は.867である。なお、Bartlettの球面性検定により変数間の関連を調べる変数同士の偏相関係数の測定結果は、相関係数の有意確率は<.000であり、独自因子は相関がないという仮説が支持され、変数の因子分析がうまくいくことが示唆される。

また、アンケート調査項目の内部一貫性を検証するために、12項目についてCronbachの α 係数を算出したところ、 $\alpha = .865$ という値が得られ、アンケートの信頼性が確認された。

5.3.1.3 因子分析の結果と考察

表3 因子分析の結果

因子/項目	負荷量	特徴値	貢献度
因子一、システムの有効性			
7. 本システムを使って、単語の使い方のまとめに役に立つと思います。	.873		
8. 本システムを使って単語の頻出表現の記憶に大いに役に立つと思います。	.837		
9. 本システムを使って孤立した単語でなく、フレーズとして単語を覚えることができますと思います。	.826		
10. 本システムを使って特定の単語と共起しやいものは何かを把握することができますと思います。	.871	4.978	55.313%
11. 本システムを使って、単語表現の特徴をまとめることにより、もっと深く単語の使い方を理解できると思います。	.893		
12. 本システムと組み合わせる「共同筆記」の使用は、学習者各自がわずかな時間でまとめたものを記入することにより全員が学習データを共有できるのはいと思います。	.803		
因子二、従来通りの単語の勉強法による問題			
4. 単語の学習は、テキストの単語表のみによるのが殆どです。時間をかけて使い方をまとめることはめ	.864	1.289	14.325

ったにありません。

5. 正しい単語の使い方について悩んだことがあります。 .846

因子1は6項目で構成されている。「本システムを使って、単語の使い方のまとめに役に立つと思います。」、「本システムを使って単語の頻出表現の記憶に大いに役に立つと思います。」など、学習支援システムの使用により単語の使い方のまとめや記憶に役に立つという内容の項目が高い負荷量を示していた。そこで「システムの有効性」因子と命名した。

因子2は2項目で構成されている。「単語の学習は、テキストの単語表のみによるのが殆どです。時間をかけて使い方をまとめることはめったにありません。」、「正しい単語の使い方について悩んだことがあります。」など従来の単語の学習に悩みを表す項目が高い負荷量を示していた。そこで「従来通りの単語の勉強法による問題」因子と命名した。

また、被験者を男性(32名)と女性(59名)に分けて、「システムの有効性」と「従来通りの単語の勉強法による問題」2因子についてそれぞれt検定で比較した結果、有意な差は見られなかった(因子1 : $t(89) = -.227, p = .821$ 、因子2 : $t(89) = -.441, p = .661$)。

表4 2因子におけるT検定の結果

因子	平均値			T value	顯著性 α
	Male	Female	all		
システムの有効性	-.0325158	.0176357	-7.220e-17	-.227	.821
従来通りの単語の勉強法による問題	-.0629963	.0341675	2.680e-17	-.441	.661

本研究で見出された2つの因子には違いはあるものの、「システム

の有効性」因子の示したように、学習者が学習支援システムの使用により、単語表現の特徴をより早く把握するだけでなく、「共同筆記」における学習データの共有を通じ、膨大な単語表現の学習負担を軽減すると同時に学習効果の向上が感じられることを示唆した。一方、学習支援システムの利用をきっかけに、因子二の「従来通りの単語の勉強法による問題」の示したように、現行の単語学習では解消されていない問題点が挙げられていた。また、2因子における T 検定の結果では、性別の違いを超えて学習支援システムの有効性についての見方や共通する学習問題を抱えていることが判明した。

5.3.2 計量テキスト分析

コロケーション学習支援システムを利用した授業体験による学習者の意識変容の特徴を確認するために、下記の質問により学習者全員に自由記述の感想を書かせた。

質問 13 :

「従来通りの単語表による学習法と比較して、コロケーション学習支援システムを利用することで感じたことを教えてください。」

5.3.2.1 テキストデータの全体像

5.3.2.1.1 テキストデータ用語の頻度分析

自由記述の中で、頻度の高い語彙の抽出をテキストマイニングで行った結果は表 5 の「品詞別による頻度の高い語彙のリスト」のようなものである。表 5 に示すように、「単字」は最も多く頻出していた語であった。また、「用法」、「系統」、「方便」「比較」、「常用」、「知道」、「搭配」など、システムの実用性を示唆する語も多く抽出された。

さらに表 5 の抽出語リストの ADJ コラムを見ると、学習支援システムの利用に対して頻度が高くてポジティブな語彙である「方便」、「容易」及び「實用」の表出が見られた。「方便」に関する回答文章を抜

粹すると「可以方便瞭解單字意義熟悉單字用法.」、「能夠快速的找出單字的用法,很方便.」、「可以一次學到多個單字片語並知道其常用用法非常方便」などの記述があった。「容易」では、「多了延伸片語的方式,覺得能更容易記憶單字.不局限于課本上單一的舉例,而是更多的搭配關係.」、「使用搭配詞之後,更容易記住單字,有時也能藉由搭配詞來學到新的用法.」、「比較知道日本人平常都是怎麼用這些字,在什麼情境用這些字,對於類似單字的使用搭配方法也會比較容易辨別.」などの記述があった。また、「實用」では、「可以知道各單字的使用頻率,覺得很實用!.」、「更加方便更好背單字很實用.」などの記述があった。

以上のことから、従来の単語表による学習法に比べ、本学習支援システムによる学習は学習者には、コロケーションに直結した単語の学習を可能にしてくれるより効率的な方法であるという示唆が得られた。

表 5 品詞別による頻度の高い語彙のリスト

順位	Noun		Adj		JJ		Adv		Verb	
	語彙	回数	語彙	回数	語彙	回数	語彙	回数	語彙	回数
1	單字	61	方便	8	常用	9	比較	28	可以	45
2	用法	31	多	5	好	4	更	26	知道	22
3	系統	19	一樣	3	不同	3	很	19	搭配	19
4	詞	19	容易	3	新	2	也	15	能	17
5	片語	11	不同	2	新的	2	不	11	會	14
6	課本	11	實用	2	單一	1	但	11	使用	14
7	句子	9	快速	2	原來	1	只	8	瞭解	13
8	方法	8	有限	2	句型	1	如何	8	覺得	11
9	日文	8	清楚	2	同樣	1	而	8	看到	9
10	中文	7	不錯	1	基本	1	就	7	學習	8
11	使用	7	豐富	1	多	1	所以	6	可能	7
12	搭配	7	主動	1	大	1	可能	5	檢索	7
13	記憶	7	偏門	1	奇怪	1	容易	5	背	6
14	學習	6	全面	1	完整	1	怎麼	5	能夠	6
15	方式	5	單調	1	實際	1	只是	4	讓	6

5.3.2.2 抽出語における分析

5.3.2.2.1 抽出語におけるスラスタ分析

自由記述から表5の示したような頻出語を抽出し、テキストデータの全体像を示すだけでなく、さらに文書の中にそれぞれの語がいかに用いられているかを知るために出現パターンの似通った語を探索し、階層クラスター分析を行った。図10はクラスター分析の結果として得られる樹状図である。

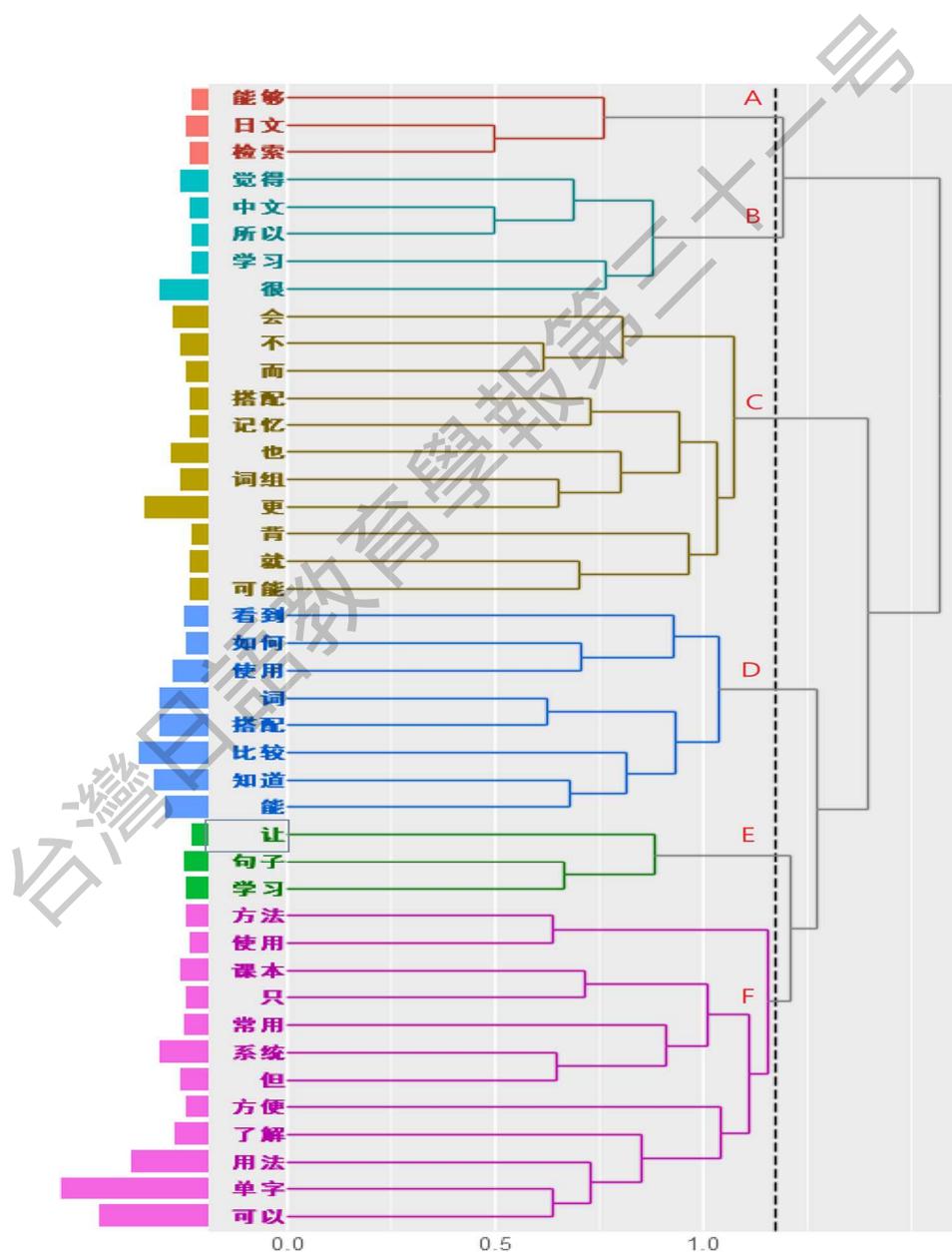


図9 クラスタ分析の結果

この樹状図には似通っていた出現パターンの語をさらに A から F の 7 グループに分割された。例えば、「能够」、「検索」、「日文」のグループ A や「覺得」、「中文」、「學習」、「所以」、「很」²は出現パターンが似通っていたことがうかがえる。

5.3.2.2.2 抽出語における対応分析

また、図 9 のクラスター分析の結果を参考にして、語の共起関係を探索する方法として、対応分析を行い出現パターンの似通った語を探索し、可視化した結果は図 10 である。対応分析で抽出された語群について、A から F の 7 グループに分割した結果は図 10 の示したものとなった。図 10 に示すとおり対応分析の結果では、

「但」、「詞」、「更」のように、中心である原点 (0, 0) の近くに布置されている語は、各文書に平均的に出現している語であるのに対して、原点から離れている語が第 1 固有値 (成分 1) の軸と第 2 固有値 (成分 2) の軸に「方便」、「常用」、「了解」などは特徴的な語として配置される。

さらにグループの成分を見てみると、原点 (0, 0) から遠くに布置されているグループ B の「學習」、「中文」、「很」やグループ F の「方便」、「常用」の語は、回答文章を抜粋すると「有很客觀的資料可以知道這個詞是怎麼使用的. 而且我們得根據其顯示的結果, 自己去歸納搭配詞的特性. 不像課本只有寫中文解釋. 所以有比較深刻的學習歷程. 覺得受益良多」、「搭配詞檢索系統可以直接整理出單字的用法, 覺得比起看未必翻得出單字原味的中文意思, 直接看單字怎麼在日文中被使用比較能夠掌握意義.」、「可以一次學到多個單字片語並知道其常用用法非常方便.」などの記述があった。中国語の翻訳だけに頼っては習得しがたい従来の単語学習法と比較して、学習支援システムの利点を特徴づける語のグループであると解釈できる。

² クラスター分析と共起ネットワーク分析では、KH コーダーにある Stanford POS parser が簡体字しか処理できないので、原文のままに示すことにした。

その一方、原点付近に布置されているグループ D は「知道」、「搭配」、「詞」、や「使用」などが、回答文章を抜粋すると「可以知道與單字常用的搭配詞或助詞。」や「可以一次學到多個單字詞組並知道其常用用法非常方便。」などの記述があり、本システムはコロケーションの使い方の習得の手助けとして挙げられるものである。また、原点の近くにプロットされていることから、自由記述の文章に平均的に出現しているものであるとしては特徴的なグループでもあると理解できる。

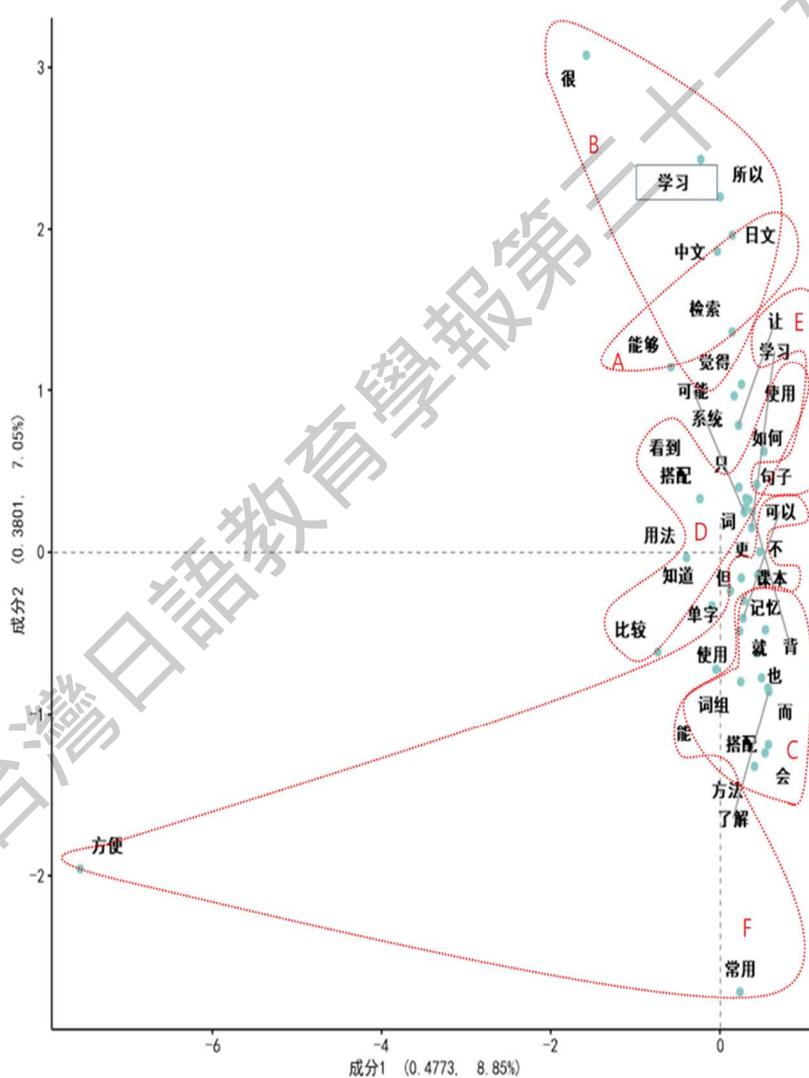


図 10 対応分析の結果

6. おわりに

本研究は、台湾人日本語学習者を対象に語彙運用の問題を解決するため、先行研究の成果と理論を背景に持ち、自律学習のためのツールとして学習支援環境の開発に取り組んだ。その結果、「コロケーション規則学習モジュール」、「コロケーション規則可視化モジュール」、「語彙の使い方における検証モジュール」といった3つの学習支援機能が開発されてきた。

本研究は、「日本語学習者誤用換言対コーパス」を利用して、「語彙誤用」の点について検証モジュールの有効性を確認したりしたほか、学習支援システム利用による効果において、一学期にわたるブレンド型授業を行ない、システム利用の有効性や単語学習問題についてアンケート調査を用いて意識変容に影響を与える要因の分析を試みた。その結果、「システムの有効性」と「従来通りの単語の勉強法による問題」など二つの因子が抽出され、単語運用の学習に関しては、学習支援システムの優位性が確認された。また、現今の学習法で単語に関する使い方の記憶を苦手とする者がいることがされてきていた。

テキストマイニングによる計量テキスト分析では、コロケーションに直結した単語の学習において、それぞれに学習支援システムの実用性を示唆する言葉に特徴があることが明らかになった。また、これらの言葉の共起関係から特定のパターンがあることがクラスター分析や共起ネットワークで明らかになった。

上述した分析の結果から、本研究で提案した学習支援システムは日本語学習者にもっと効率的かつ柔軟性のある学習ツールを提供し、その結果、単語運用力の向上傾向や学習内容への理解の深化といった情意面の変容等の効果が見られた。

教師にとって自身の授業計画に取り入れて授業後の宿題として学習者に本システムを利用させるといったブランデッド式授業を進めることができ、発見学習による学習者のモチベーションの向上を実

現するだけでなく、積極的に学習活動への参加を促進することも可能となるだろう。

本研究を契機として、今後さらに、この学習支援環境の開発経験をベースにして、学習支援環境の利用は学習者の語彙使用から文法習得、聴解能力にどのように影響するかについて、教育実践を通じてコンピュータによる支援学習の可能性を探求し、更に研究を進めて台湾の日本語教育の一助になることを期待したい。

参考文献

- 赤堀侃司、楊接期 (1999) 文章の結束関係を用いた科学技術日本語テキストの作成支援システム. 第15回日本教育工学会大会.
- 姉崎達夫 (2006) 中学生における英単語認知と意味アクセス. 上越英語研究. (7) 49
- 高建斌、小高知宏、小倉久和 (2001). "外国人日本語学習者による日本語作文のn-gramモデルを用いた特徴抽出と作文評価." 福井大学工学部研究報告 49(2): 217-225.
- 黄冠誠、吳鑑城、許湘翎、顏孜曦、張俊盛 (2014). "學術論文簡介的自動文部分析與寫作提示." Computational Linguistics and Chinese Language Processing 19(4): 29-46.
- 高林哲 (2000) Namazu: 全文検索で文書の山に立ち向かう. 情報処理学会論文誌, 41(11), 1227-1232.
- 田中久美子、吉川拓哉. (2007) 類似表現検索による作文支援. 言語処理学会第13回年次大会.
- 施列庭 (2014). コーパスに基づいた漢字発音学習支援環境の開発—漢字の促音化を例に. 台湾日語教育學報(22): 59-87.
- 進藤三雄(2002). <論説>英語語彙学習のためのCALLソフトウェア開発 アドミニストレーション(8)3 181-200
- 陳淑娟 (2006) 作文に置ける語彙習得についての一考察—使用語数と語類の変化を中心に—. 東吳日語教育學報(29), 29-63.
- 鄭在玲、三宅真紀、馬越庸恭、赤間啓之 (2006) グラフ理論を用いた日本語連想作文支援システムの開発と評価 PC カンファレンス.
- 鄧美華 (2007) 作文データの中に見られる助詞の誤用分析—日本語学習歴5年(含む)以上の台湾人学習者のデータを使って— 2007年台大日本語文創新國際學術研討會.
- 中條清美、西垣知佳子、内山将夫、山崎淳史 (2006) 「初級英語学習者を対象としたコーパス利用学習の試み」『日本大学生産工学部研究報告B』 (39) 29-50
- 中條清美、アンソニー ローレンス、西垣知佳子、内堀朝子(2008)日英パラレルコーパスを利用した文法発見学習(大会テーマ:グロー

- バルな英語コミュニケーション能力とは-英語教育再考). JACET全国大会要綱. 47. 105-106
- 橋本喜代太、安藤秀明、竹内和広 (2008). 複数コーパスからの特徴表現抽出を利用した文体統一性の評価とその活用について. 言語処理学会第14回年次大会発表論文集.
- 樋口耕一 (2014) 『社会調査のための計量テキスト分析 —内容分析の継承と発展を目指して—』 ナカニシヤ出版
- 堀正広 (2009) 『英語コロケーション研究入門』 東京：研究社.
- 馮寶珠 (2005) 作文における使役表現の分析—中国人日本語学習者と日本語母語話者の作文の比較. 日語教育與日本文化研究國際學術會議.
- 山本幹雄、森輝彦 (2002). “日本語学習者向け日本語スペルチェック.” 日本語教育方法研究会誌 9(1): 18-19.
- 葉淑華 (2006a) ネットコースのデザインと実践をみる—「日本語作文」授業を例に—. 日語教育與日本文化研究國際學術研討會.
- Nazar Rogelio, Renau Irene. (2012). Google Books N-gram Corpus used as a Grammar Checker. Proceedings of the EACL 2012 Workshop on Computational Linguistics and Writing, FAvignon, France.

付録：日本語学習者誤用換言対コーパスのデータ（一部）

国籍	誤用例
VN	「高校物理」から物理に属する術語を学ぶことができる。
VN	21世紀は科学技術を強く開発する時代である。
VN	2らいしゅうのにちようび
CH	40年以上に渡り曲折を経験した北陸新幹線
VN	505の学生を対象をして
VN	Steve Jobsが凄いスピーチをあげた
VN	アニメからひとことを勉強した。
VN	アベンチャー旅行なども発展できる可能性が高いである。
VN	アボガドのつぶをてきしゅつします。
CH	アメリカ人と日本人は両方も会社内で良好な人間関係を維持することは必要であると思うが
VN	ある大学生はできるだけアルバイトがある。
TH	いい人間関係がある
VN	いい将来のために
PR	イオンの近くにのミスタードーナツに行っています。
CH	いくつの欠点がある
VN	いちばん意義が多い彼の言葉は～
CH	いつ行って見たい
CH	イノベーションを造り
VN	いろいろな良い機会が達するはずだ。
VN	いろいろな機会があろうと思っている。
TH	インターネットで買う方もあまりわかりませんから
CH	インターフェースである
IN	おおきいしょくよくがあります
IN	おみやげがほしいなら、はまべの見つかったシェルをもってかえってもいいです。