

日本語読解学習支援システム「リーディング・チュウ太」

北村達也

甲南大学 知能情報学部 知能情報学科
神戸市東灘区岡本 8-9-1, 658-8501

概要

本稿では、著者らが1997年以来運用を続けているインターネット上の日本語読解学習支援システム「リーディング・チュウ太」について概説する。リーディング・チュウ太は「辞書ツール」、「読解教材バンク」、「レベルチェッカー」、「多言語辞書」などのWebブラウザから利用できるツールから成り、世界中の日本語教師および日本語学習者に利用され続けている。これらのシステムは、自然言語処理技術を利用したもので、その仕組み自体は非常にシンプルである。しかし、開発に日本語教師が深く関わり、長期に渡って改良を続けていることによって日本語教育の現場で本当に使えるシステムになっている。

キーワード: 日本語教育, 読解学習支援, 自然言語処理技術, インターネット, Webブラウザ

1 はじめに

日本語学習者が日本語の文を読むに際しては様々な困難が付きまとう。まず、日本語は分かち書きをしないため、単語の区切りがわかりにくい。また、膨大な数の漢字を識別し、その読みや意味を把握する必要がある。そして、これらのことから辞書を引くのが難しい。

リーディング・チュウ太の「辞書ツール」は、形態素解析システムを利用することによって、これらの困難を解決した。このシステムを利用することによって、学習者は辞書を引く手間をかけずに日本語文章を読み進めることができる。このツールは、日本語教師が独自教材を作成する際にも利用することができる。

辞書ツールの他、同じく形態素解析システムを利用して、文中の単語のレベルを色分け表示する「レベルチェッカー」も開発されている。また、辞書ツールで処理した読解教材を集めた「読解教材バンク」、約30言語の「多言語辞書」なども構築された。現在、これらのシステムは世界中の日本語教育の現場で利用されている。教育工学の研究分野では新規性が重視される傾向があるが、開発されたシステムのうちのどの程度が開発した機関以外でも長期間利用され続けているであろうか。リーディング・チュウ太のように15年以上も世界中で利用されている事例は極めてまれである。

本稿では、著者の研究の1つである、日本語読解学習支援システム「リーディング・チュウ太」の開発に至る経緯、現在提供中のツール群、そして現在進行中の研究について概説する。

2 リーディング・チュウ太に至るまで

リーディング・チュウ太の辞書ツールは、北陸先端科学技術大学院大学(当時)の寺朱美氏らの研究に由来している。寺と落水 [1] は、まず 1994 年にハイパーテキストを利用した漢字学習システムの構想を発表した。その中で、「最終的には日々の新聞記事から教材が作成でき」るようなシステムを目指すと述べている。これは、寺が日本語教師であるからこそ生まれた発想であろう。外国語教育においては、生の日本語、新鮮な話題を取り上げるため、新聞や雑誌の記事を教材にすることが少なくない。上記のアイデアは彼女の教育実践の中で生み出されたものと推察される。

当時の日本語学習システムは、映像や音声を含むマルチメディアのコースウェアや漢字や単語を覚えるためのゲームといったものが主流であった。そこではシステムとコンテンツが不可分で、利用者が独自の教材を作成する仕組みを提供するという発想はほとんどない。システムをコンテンツから独立させることによって利用者にコンテンツを選択する自由を提供しようという寺のアイデアは革新的であった。

続く 1995 年、寺らはアップルコンピュータ社のソフトウェア HyperCard を利用して上記の構想を試作したシステムを発表した [2]。このシステムでは、あらかじめ用意されたテキスト内の熟語をクリックすると、それを構成する漢字の情報が提示された。さらに、利用者が入力したテキストに対して、このようなリンクを自動作成する機能も提供した。

しかし、このシステムは単語検出で多大な処理時間を要したため、現実的な時間内で文章の処理を完了させることが困難であった。そこで、北陸先端科学技術大学院大学在学中であった著者が、形態素解析システム JUMAN を利用することを提案し、Perl で実装した [3]。このシステム「DL」は、入力文を JUMAN で形態素解析し、各形態素に Jim Breen が作成・維持している日英辞書 EDICT の情報をリンクするものであった。JUMAN を利用したことによって処理時間は大幅に短縮された。なお、DL ではハイパーテキストを HTML で記述し、作成した教材を Web ブラウザで表示するようにした。

その後、DL は東北大学(当時)の内山潤氏の助言により Common gateway interface (CGI) 化され、一般公開された [4]。また、DL を利用した読解教材バンクの構築 [5] や学習履歴管理機能の付与 [6] も行われた。CGI 化により、利用者は Web ブラウザ上のテキストエリアにテキストを入力するだけで、自動的に単語に英訳がリンクされたハイパーテキストを得ることができるようになった。学習者は、このシステムを用いることによって、辞書を引く手間がわずらわされることなく文章を読み進めることができた。一方、教師は、授業用のテキストの単語リスト作成を一瞬にして完了させることができた。

しかし、当時はまだインターネットが本格的に普及する前であり、一般に日本語教師はあまりコンピュータに詳しくないこともあって、日本国内の日本語教育の現場で DL が利用されることはほとんどなかった。むしろ、主たる利用者は、教材不足に悩む海外の若い日本語教師や日本語学習者たちであった。

3 リーディング・チュウ太の各ツール

その後、著者は異動などを経て、1997 年、東京国際大学の川村よし子氏と東京大学の学生であった保原麗氏とともに「リーディング・チュウ太」という Web サイトを公開した。この Web サイトで

公開した「辞書ツール」は、DLに基づき保原氏が新たな機能を加え実装したものである。

“Reading Tutor”をもじった「リーディング・チュウ太」という名称は、川村氏の命名である。さらに、「チュウ太」というネズミのキャラクターを作り、そのイラストをリーディング・チュウ太のWebページに掲載した。これらは、コンピューターを敬遠する日本語教教師へのアピールに大きく貢献した。

DLのWebページは、タイトル、テキストエリア、ボタンがあるだけの淡白なものであった。良い機能を提供しさえすれば広く使ってもらえるはずという思い込みが著者にあったためである。しかし、リーディング・チュウ太の経験から、著者はツールの機能だけでなく使ってもらうための雰囲気作りも重要であることを学んだ。

その後、リーディング・チュウ太のWebサイトには「読解教材バンク」、「レベルチェッカー」、「多言語辞書」などの機能が追加され、15年以上運営され続けている。川村氏の精力的な広報活動の効果もあって、現在、リーディング・チュウ太は日本語教育システムのデファクトスタンダードの1つとなっている。

3.1 辞書ツール

辞書ツールは、Webブラウザのテキストエリアに入力された文中の単語に読み、訳語、意味(辞書情報)をリンクするシステムである。2013年9月現在、辞書ツールでは、日英辞書、日独辞書、日蘭辞書、日斯辞書、日西辞書、日本語辞書から選択し、その辞書情報を提示することができる。

辞書ツールではまず形態素解析システム茶筌 [7] を用いて入力文を形態素に分割する。一例として、「文章を読み進める。」という文を解析した結果を図1に示す。各形態素の品詞や読みや原型などの情報が行単位で表示されている。

文章	ブンショウ	文章	名詞-一般		
を	ヲ	を	助詞-格助詞-一般		
読み	ヨミ	読む	動詞-自立	五段・マ行	連用形
進める	ススメル	進める	動詞-自立	一段	基本形
.	.	.	記号-句点		
EOS					

図1: 「文章を読み進める。」を形態素解析システム茶筌 ver. 2.3.3 で分析した結果

形態素解析の後、助詞、記号、数字以外の形態素の原型を辞書で検索する。「読み」という活用した形では辞書を検索できないが、形態素解析システムにより得られる「読む」という原型であれば辞書で検索することができる。そして、形態素と辞書情報のリンクをHTMLで記述し、最後に図2のようにWebブラウザに表示する。ウィンドウ左側のフレームには入力文が、右側のフレームには辞書情報が表示されている。利用者が左側のフレームの単語をクリックすると、それに対応する辞書情報が右側のフレームの最上部に現れる。

DLと同様に、学習者は辞書ツールを利用することによって辞書を引く手間を省いて日本語の文章



図 2: 辞書ツールの処理結果

を読み進めることができる。ただし、日本語学習に対する辞書ツールの効果についての研究は未だに行われておらず、今後の課題である。

また、形態素解析の段階で誤りが生じると、処理結果にまで影響してしまうという課題もある。この問題を解決するため、辞書ツールの処理結果を利用者が編集できるシステムも開発し公開しているが [8]、あまり利用されていない。

3.2 読解教材バンク

読解教材バンク (図 3) は、自力で日本語教育向けの教材を集めるのが難しい海外の日本語学習者などを支援するため、様々なレベルのテキストに辞書ツールで辞書情報をリンクさせて提供したものである [9]。各テキストには、そこに含まれる単語の難易度に従ってレベルが示されている。英語、独語、日本語の 3 つの辞書情報を提供している。

テキストは、日本語教師であるケン・イジケビッチ氏が執筆した日本文化に関する文章の他、一般の方の海外旅行記、エッセイ、留学生の作文など 200 編を超える。辞書ツールのメリットを生かしてこれらの多様なテキストを読むことを通して、多読学習を行ってもらうのが目的である。

この読解教材バンクは、ThinkQuest@JAPAN'99 (学校インターネット教育推進協会主催) にて大学生・社会人の部最優秀賞を受賞するなど高い評価を得た。

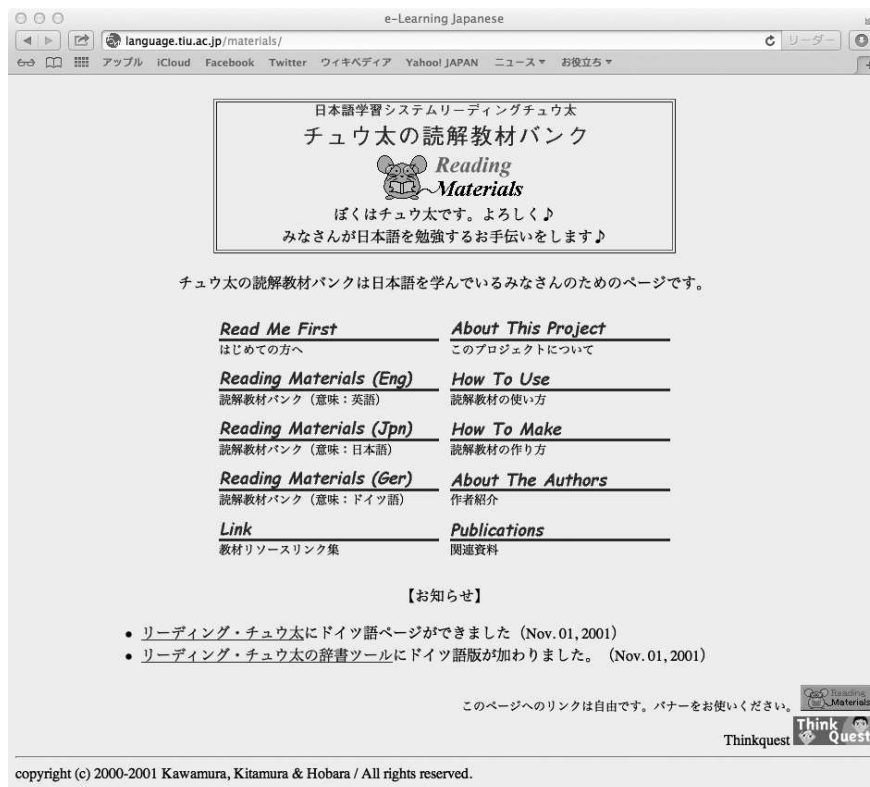


図 3: 読解教材バンクのトップページ

3.3 レベルチェッカー

レベルチェッカーは、入力文の各形態素を旧日本語能力試験出題基準などに基づいて難易度判定するシステムである [10]。現在の日本語能力試験は出題基準が公開されていないため、古い出題基準を用いている。各形態素はレベルに応じて色分け表示されるので、その難易度を一目で判断することができる。さらに、利用者が文章全体の難易度を判定するための材料として、文章全体に含まれる各レベルの単語数およびパーセンテージを表にして示している。

このシステムは広く日本語教育で活用されている。主な使われ方は、授業で使うテキストのレベルをチェックしたり、対象とする学習者のレベルに応じてテキストを書き換えたりすることである。また、読解に関する実験の際に、実験参加者に提示する文章のレベルを制御するためにも使われている。変わったところでは、公共機関の広報におけるいわゆる「お役所言葉」を訂正するためにも利用されている [11]。近年では、日本で暮らす外国人向けのやさしい日本語の作成にも使われるようになっていく。特に、災害時の情報伝達は命に関わるので極めて重要である [12]。

単語の難易度をいかに規定するかは大変難しい問題であるため、レベルチェッカーでは様々な基準を試みてきた。現在では、以下の 10 種の基準を用いた難易度判定機能を提供している (図 4)。

- 旧日本語能力試験出題基準に基づく分類 (goi)



図 4: レベルチェッカーのトップページ

- 介護用語のレベルに基づく分類 [13], [14], [15]
- goi の機能に加え，一文中の平均文字数および文章のレベルを表示
- 経済用語のレベルに基づく分類
- 筑波大学との共同研究による新基準に基づく分類
- 朝日新聞記事における頻度情報に基づく分類 (10 段階表示)
- 単語親密度 [16] に基づく分類 [17]
- 毎日新聞記事データベースから作成した IDF に基づく分類 [18]
- 読売新聞記事データベースから作成した IDF に基づく分類 [18]
- 朝日新聞記事における頻度情報に基づく分類 (6 段階表示)

3.4 多言語辞書

上述のように現在の辞書ツールでは複数言語の辞書情報を提示することができる。しかし，日本語学習者の数が多いアジア圏の言語が含まれないなどの課題があった。そのため，世界中の日本語

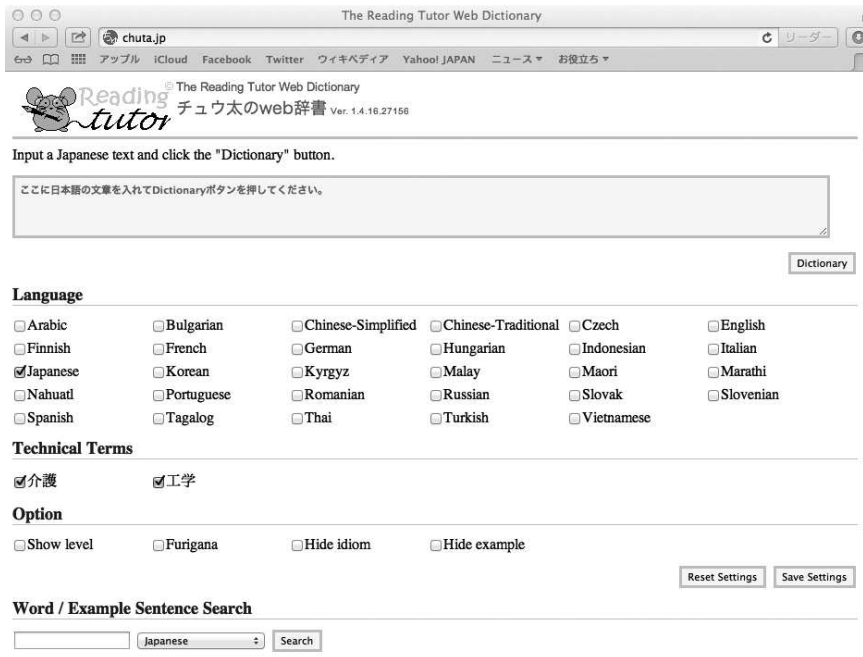


図 5: チュウ太の web 辞書のトップページ

教育関係者が協力して多言語辞書を構築するプロジェクトが進行している [19], [20]. その成果は <http://chuta.jp/> にて公開されている. 図 5 に示すように, 現在では約 30 言語の辞書が作成されている.

この Web サイトでは, 単語の意味の検索機能だけでなく, 辞書ツールのように入力文の単語に辞書情報をリンクする機能も提供しており, 多くの国の日本語教育に貢献している.

4 現在進行中のプロジェクト

以下では, 甲南大学にて現在進行中のプロジェクトを紹介する.

4.1 難易度判定機能付きエディタ

レベルチェッカーは広く利用されている一方で, 利用者が単語レベルの情報を変更できないという課題があった. 例えば, ある教科書に従って授業を進めている場合, 新出単語だけ色分け表示したいという要望がある. そこで, このような個別の単語レベルリストに対応できるシステムを開発中である.

このシステムは, エディタに難易度判定機能を追加したもので, そこで使われる単語レベルリストは利用者が作成することができる [21]. このシステムを図 6 に示す. テキストエリアに文を入力し,

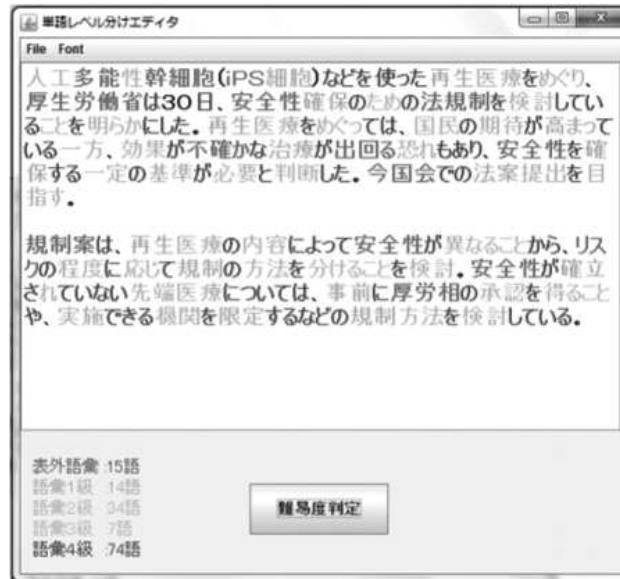


図 6: 難易度判定機能付きエディタ

ウィンドウ下部のボタンをクリックすることにより、文が形態素解析され、各単語の難易度が色分け表示される。

2013年9月現在、ある日本語教育機関にこのソフトウェアを試用していただいております、そのフィードバックを開発に反映させる予定である。

4.2 文法的要因の抽出

文の難易度の主要な要因は単語の難易度であるが、当然、文法的な要因も難易度に寄与する [22]。そこで、難易度に寄与する文法的要因の抽出も試みている。

例えば、「朝起きて、顔を洗った。」という文にはテ形中止法が含まれ、「会場へ行き、大会の準備をする。」という文には連用中止法が含まれる。これらは、日本語学習者にとって難しい構文の1つである。このような構文を形態素解析システムから得られる品詞情報と構文解析システムから得られる構文情報を利用して検出するシステムを開発した [23]。

また、日本語では「本が欲しい」などの主格を省略した文が多用されるが、これも学習者にとっては難しい。そこで、入力された文の主格が省略されているか否かを判定するシステムを開発した [24]。一見、単純に助詞を見るだけで判定できると思われがちであるが、日本語では「寿司が食べたい」という文も正しいので、このような場合に対応できる技術が必要である。そこで、インターネット上の用例を参考にして当該の単語が目的語か否かを判定する技術を開発して、高精度の主格省略検出を実現した。

4.3 スマートフォンの活用

この2, 3年でスマートフォンやタブレットが爆発的に普及した。これらを利用すれば、学習のためにPCの前に座る必要がなくなる。そこで、スマートフォンのカメラを利用して実環境に存在する印刷物の読解を支援するシステム(図7)を開発中である。

このシステムでは、まずスマートフォンのカメラで印刷物の写真を撮り、それをサーバーに送信する。その写真をサーバーにて文字認識し、辞書情報をリンクしたXMLファイルをスマートフォンに返信する。このシステムは、PC内のテキストデータに限られていた辞書ツールを実環境まで拡張させるシステムといえる。ただし、Googleが提供する翻訳システムがすでにカメラからの入力に対応しているので、差別化していく必要がある。

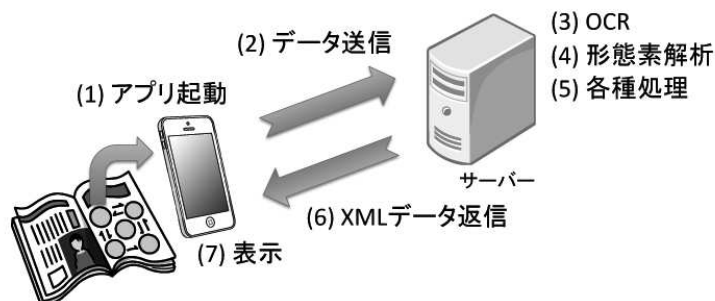


図7: スマートフォンのカメラを利用した読解支援システムの構想

5 おわりに

本稿では、著者の研究の1つである、日本語読解学習支援システムリーディング・チュウ太について概説した。当初、このWebサイトは日本語教師と学習者を支援する目的で始められたが、時代とともにその役割を拡大しつつある。その1つが、日本で生活する非日本語母語話者の支援である。

現在、様々な分野で非日本語母語話者が日本語を使って仕事をする場面が増加している。超少子高齢化社会の日本では、今後この傾向が増えることは間違いない。また、今後も様々な災害の発生が予想されている。このような場面でのコミュニケーションに寄与できるよう、引き続きこのWebサイトを維持する予定である。

加えて、非日本語母語話者との円滑なコミュニケーションのために、習得しやすく理解しやすい「ユニバーサルな日本語」が一種の標準語のような形で求められるようになるであろう。そのような日本語の設計にも活用できるシステムを視野に入れつつ、今後も研究を続けていく。

謝辞

本研究の一部は科研費基盤研究(B)(24320096, 21320095, 18320083)、東京国際大学特別研究助成(2008年度)、JSTシーズ発掘試験(2009年度)、甲南大学総合研究所(2009-2010年度)、及び私立大学

等経常費補助金の支援を得て行われた。辞書作成，教材提供，システム開発などにご協力いただいた皆様に感謝します。

参考文献

- [1] 寺朱美, 落水浩一郎, “ハイパーカードを利用した漢字学習支援システム,” 日本語教育方法研究会誌, vol. 1, no. 3, pp. 36–37, 1994.
- [2] 寺朱美, 桑山正彦, 落水浩一郎, 加納千恵子, “ハイパーカードを利用した漢字学習支援システム,” 日本語教育方法研究会誌, vol. 2, no. 1, pp. 2–3, 1995.
- [3] 寺朱美, 北村達也, 落水浩一郎, “WWW ブラウザを利用した日本語読解支援システム,” 日本語教育方法研究会誌, vol. 3, no. 1, pp. 10–11, 1996.
- [4] 寺朱美, 北村達也, 落水浩一郎, Tomoko Graham, Anne R. Lavin, “日本語読解支援システムの検証: MITにおけるアンケート結果の報告,” 日本語教育方法研究会誌, vol. 4, no. 1, pp. 26–27, 1997.
- [5] 北村達也, 寺朱美, 奥村学, “DL を利用した読解教材バンクの構築,” 日本語教育方法研究会誌, vol. 6, no. 1, pp. 12–13, 1999.
- [6] 北村達也, 川村よし子, 内山潤, 寺朱美, 奥村学, “学習履歴管理機能を持つ日本語読解支援システムの開発とその評価,” 日本教育工学雑誌, vol. 23, no. 3, pp. 127–133, 1999.
- [7] 松本裕治, “形態素解析システム「茶釜」,” 情報処理, vol. 41, no. 11, pp. 1208–1214, 2000.
- [8] 北村達也, 荊木亜里沙, 森川結花, 永須実香, 川村よし子, 前田ジョイス, 斉木美紀, 金善子, “Web 上の読解教材作成を支援するツールの開発及び活用法,” 甲南大学情報教育研究センター紀要, vol. 10, CD-ROM, 8 pages, 2011.
- [9] 川村よし子, 北村達也, “インターネットを活用した読解教材バンクの構築,” 世界の日本語教育: 日本語教育事情報告編, vol. 6, pp. 241–255, 2001.
- [10] 川村よし子, “語彙チェッカーを用いた読解テキストの分析,” 講座日本語教育, 第 34 分冊, pp. 1–22, 1998.
- [11] 小田順子, “文章表現の適切さが自治体ウェブサイトにも与える影響に関する一考察: 広報文改善ガイドラインの構築と書き換えによる実験を通して,” 放送大学大学院修士論文, 2012.
- [12] 佐藤和之, “市民が支える外国人への情報伝達と「やさしい日本語」化支援,” 地方自治職員研修, vol. 44, no. 9, pp. 49–257, 2011.
- [13] 野村愛, 川村よし子, “介護福祉士候補者の自律学習支援のための語彙リスト作成,” 日本語教育方法研究会誌, vol. 18, no. 1, pp. 14–15, 2011.
- [14] 野村愛, 川村よし子, 斉木美紀, 金庭 久美子, “単語難易度と出題頻度に配慮した介護福祉士候補生のための語彙リスト作成,” 日本語教育方法研究会誌, vol. 18, no. 2, pp. 12–13, 2011.

- [15] 川村よし子, 野村愛, 名藤杏子, 金庭久美子, 斉木美紀, 北村達也, “介護福祉士候補生のための国家試験に向けた教材の開発,” 日本教育工学会第 27 回全国大会講演論文集, pp. 625–626, 2011.
- [16] 天野成昭, 近藤公久, 日本語の語彙特性 第 1 期. 三省堂, 東京, 1999.
- [17] 川村よし子, 北村達也, “文章の難易度判定のための単語親密度チェッカーの開発,” 日本語教育方法研究会誌, vol. 15, no. 2, pp. 24–25, 2008.
- [18] 北村達也, 富岡洋介, 川村よし子, “文章の難易度判定のための単語親密度チェッカーの開発,” 日本語教育方法研究会誌, vol. 15, no. 2, pp. 24–25, 2008.
- [19] 金庭久美子, 川村よし子, “多言語版日本語辞書における用例作成の諸問題,” 日本語教育方法研究会誌, vol. 15, no. 1, pp. 14–15, 2008.
- [20] 三輪譲二, 川村よし子, “「チュウ太の多言語版 Web 辞書」の利用状況調査報告,” 日本語教育方法研究会誌, vol. 19, no. 1, pp. 62–63, 2012.
- [21] 柴田大介, 村田進, 北村達也, 川村よし子, “単語難易度判定機能を有するエディタの開発,” 日本語教育方法研究会誌, vol. 20, no. 1, pp. 70–71, 2013.
- [22] 川村よし子, 前田ジョイス, 保原麗, 川村ヒサオ, “文章の難易度判定システム構築のための基礎調査,” ヨーロッパ日本語教育, vol. 15, pp. 171–178, 2011.
- [23] 水嶋博志, 内田聖也, 北村達也, 川村よし子, “学習者にとって難解な構文の自動検出,” 日本語教育方法研究会誌, vol. 18, no. 1, pp. 64–65, 2011.
- [24] 北村達也, 中村慶太, 川村よし子, “日本語文章における主格省略の自動検出,” 甲南大学紀要知能情報学編, vol. 5, no. 2, pp. 129–142, 2012.