

# インターネットを活用した読解教材バンクとその評価

## Development and Evaluation of a Japanese Reading Resource Bank using the Internet

東京国際大学商学部

川村 よし子

kawamura@tiu.ac.jp

### 1.はじめに

本研究の読解教材バンクは、インターネット上に置かれ、学習者が自由にアクセスできるのみならず、インターネット上の情報をそのまま教材として活用できるという「開かれたシステム」である。インターネット上には最新の情報を含む多くの情報が提供されている。しかも、検索のやり方次第では、良質で価値の高い情報を得ることができる。日本語学習者がこれらの情報をデータバンクとして活用することができれば、有用な教材となるに違いない。本システムはこうした情報の活用を支援すると同時に、日本語学習そのものも促進させる学習環境を提供するものである。ここではまずこの教材バンクの概要を説明し、その利用方法を述べる。次に、教材バンクの抱えている問題点を分析し、解決方法を提案する。

### 2. 先行研究

インターネットを利用した読解支援システムとしてはHenstock (1995) による辞書引きシステム「AutoGloss/J」<sup>1</sup>がある。AutoGloss/Jは学習者の作成した単語リストをもとに辞書引き作業を行い、単語の読みと英訳を提示するものである。また、寺ほか(1996)の開発した「DL」<sup>2</sup>は、入力されたテキストの形態素解析および辞書引き作業を自動的に行い、学習者の必要に応じて個々の単語の読みと意味を提示し読解支援を行うシステムである。北村ほか(1998)は、このDLに学習履歴管理機能をもたせたDL2<sup>3</sup>を開発し、語彙の学習がより促進できるようシステムを整えた。さらに川村ほか(2000)は、基本辞書として語彙量の豊富なEDR辞書を利用し、より多くの正確な辞書情報の提供できる「辞書ツール」<sup>4</sup>を開発した。この辞書ツールにも学習履歴管理機能が組み込まれている。一方、教材の難易度レベルを判定するシステムとしては、川村(1998/1999)の開発した「語彙チェッカー」<sup>5</sup>および「漢字チェッカー」<sup>6</sup>がある。これはテキスト中に含まれる語彙や漢字のレベル判定を自動的に行い、個々の単語や漢字の難易度を表示するシステムである。また金庭は、インターネット上の情報を読解教材として活用した「読解クイズ」を作成し、CGI (Common Gateway Interface) 化して自動採点できるようにしたものを学習者に提供している<sup>7</sup>。これらのシステムは、いずれも、学習者の語彙学習、漢字学習あるいは読解学習を個々に支援するものとして開発がすすめられ、インターネット上で公開されている。

川村ほか(1999)の読解学習支援システム「リーディング・チュウ太(Reading Tutor)」<sup>8</sup>では、これまで個別に開発されてきたシステムを一連のツール群として扱い、学習者が必要に応じて選択できる学習環境を作り上げた。このシステムには、上記の「辞書ツール」「語彙チェッカー」「漢字チェッカー」「読解クイズ」等が統合されている。このシステムもインターネット上で公開しているので、学習者はインターネットで日本語が読める環境にさえあれば、自由に利用可能である(<http://language.tiu.ac.jp>)。本研究の読解教材バンクに収められた各教材も、このシステムのツールを用いて教材化したものである。

### 3. 読解教材バンクの仕組み

#### 3.1 読解教材バンクの特徴

読解教材バンクは読解学習支援システム「リーディング・チュウ太」の構成要素の一つとして新たに開発したものである。上述のように「リーディング・チュウ太」は、インターネット上の情報を利用して学習者が自由に学ぶシステムである。ところが初級・中級の日本語学習者にとって自分のレベルにあった教材をインターネット上で探してくることは容易ではない。また、コンピュータの操作やインターネットの利用自体に慣れていない学習者に対しても、どのような学習が可能なのか実際に体験できる場を提供する必要がある。そこで今回、初級級・中級の学習者向けの読み物を集めて教材化し、「チュウ太の読解教材バンク」を構築した。

教材の収集にあたっては日本語母語話者と日本語上級学習者の双方に読解教材バンクへの教材の提供を呼びかけた。あくまでも初級・中級学習者を対象にした教材バンクをめざしたため、平易な文体で書かれていることを条件に文章を収集した。今回、読解教材バンクに収録した教材は次の通りである。

日本語学習者向けに書かれた読み物	72編
日本人読者向けに書かれた読み物	86編
日本語上級学習者が書いた意見文・報告文	7編

教材バンクにおさめられた読解教材はすべて図2のような形で提供されている。

左のフレームが本文、右が辞書フレームである。本文中の下線の引かれた単語はすべて右の辞書とリンクされている。学習者が本文を読んでいて分からない言葉、読めない文字に出会ったらその単語をマウスでクリックする。すると、右の辞書画面が自動的にスクロールし当該語句が一番上に表示される仕組みになっている。次々と新しい語句の読みや意味を知ることが、同一の語句を再度調べることができる。またこの本文や辞書フレームをそれぞれ印刷することも保存することも可能である。

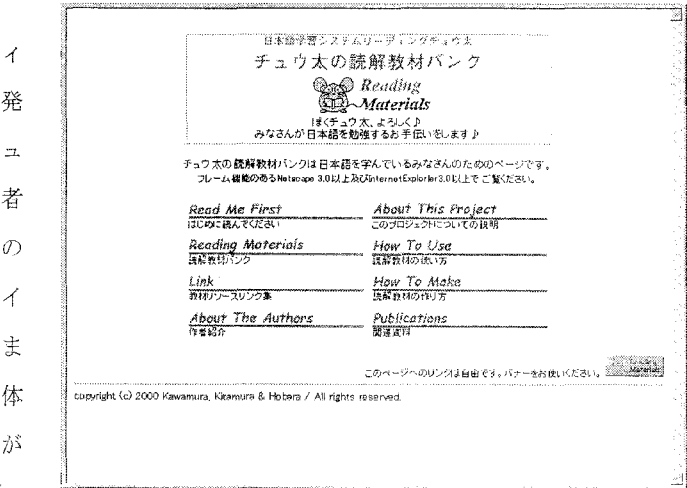


図1 読解教材バンク

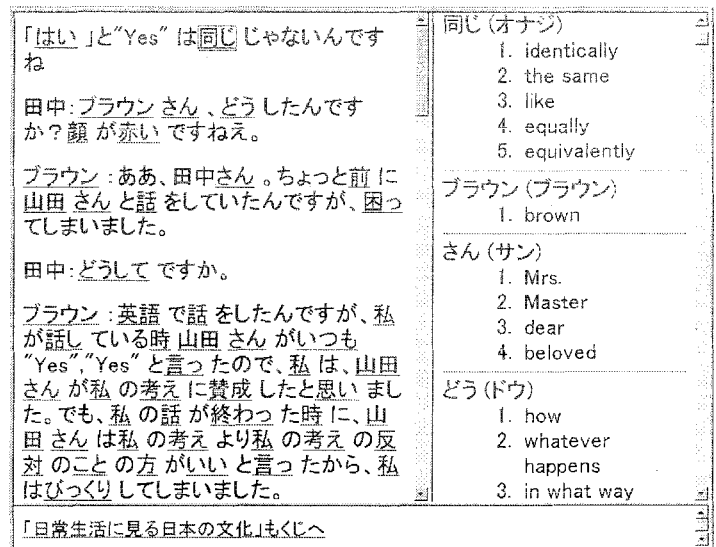


図2 教材バンクの教材

辞書にはEDR日本電子化辞書研究所の「日英対訳辞書」を用いている。この辞書は23万語が収録された大規模なもので、固有名詞をのぞき、日本語学習者にとって必要な語句はほぼ網羅されている。英訳の表示に関しては、1対1対応の翻訳ではなく、各々の単語の持つ意味概念の訳語を列挙する形を採用した。これは、「辞書ツールはあくまでも辞書の役割を担うべきであり翻訳ツールではない」という理念に基づいている。

語句の意味は文脈によって決定される。読解学習において、個々の単語の意味を文脈から読み解く (decoding) 作業は非常に重要な位置を占めるものである。辞書引き作業を軽減することによって、文脈情報と単語の意味情報の双方から文章を読み解くという読解学習が促進できるはずである。

さらにこの教材には学習履歴を表示する機能が整備されている。一連の読解作業が終了した時点で左フレームの本文下にある「list」ボタンをクリックすると、図3のようにこれまでの学習語句（読解中に辞書を参照した語句）がリストアップされる。各語句の右にある数字は学習中に当該語句をクリックした回数である。このリストの語句は右の辞書フレームとリンクされているため、単語の読みや意味がその場で再確認できる。この学習履歴管理機能を活用することで学習者の語彙の学習も促進される<sup>3</sup>。この単語リストの印刷や保存も可能である。

Your Unknown Vocabulary List	
<ul style="list-style-type: none"> <li>• 同じ (1)</li> <li>• 鐘 (1)</li> <li>• 困る (1)</li> <li>• 賛成 (1)</li> <li>• 反対 (1)</li> <li>• 意味 (1)</li> <li>• 晩 (1)</li> <li>• 閉まる (1)</li> <li>• 家内 (1)</li> <li>• 習慣 (1)</li> <li>• 閉める (1)</li> <li>• お辞儀 (2)</li> <li>• 挨拶 (1)</li> <li>• 握手 (1)</li> </ul>	<p>晩 (バン)</p> <ol style="list-style-type: none"> <li>1. night</li> <li>2. evening</li> <li>3. dinner</li> <li>4. supper</li> </ol> <hr/> <p>御飯 (ゴハン)</p> <ol style="list-style-type: none"> <li>1. boiled rice</li> <li>2. cooked rice</li> <li>3. food</li> <li>4. lunch</li> </ol> <hr/> <p>食べる (タベル)</p> <ol style="list-style-type: none"> <li>1. eat</li> <li>2. make a living</li> <li>3. live</li> </ol> <hr/> <p>面白い (オモシロイ)</p> <ol style="list-style-type: none"> <li>1. delightful</li> <li>2. amusing</li> </ol>

「日常生活に見る日本の文化」もくじへ

図3 学習履歴の表示画面

### 3.2 インターネット上の情報の教材化

インターネット上に置かれたこの読解教材バンクは、単に一般の利用者に対して開かれているばかりではない。インターネット上の情報をそのまま教材として利用できる「開かれたシステム」である。

トップページ(図1)の「教材リソースリンク集」をクリックすると、学習者に役立つホームページへのリンク集が開く。このリンク集は各新聞社のホームページや各種の情報検索エンジンはもちろん、日本語学習教材、日本文化、日本文学等の様々なホームページにリンクが張られている。学習者はインターネット上の情報の中から自らのニーズにあわせて教材を選択することができる。選択した情報(以下「テキスト」と呼ぶ)の教材化には「チュウ太の工具箱」の学習支援ツールを用いる。チュウ太の工具箱には次の4種類の支援ツールが組み込まれている。

日英辞書ツール テキスト中の単語の読みと、意味情報を英語で表示する

概念辞書ツール テキスト中の単語の読みと、意味情報を日本語で表示する。

語彙チェッカー 日本語能力試験の語彙レベルを基準にテキスト中の単語の難易度レベルを自動判定する

漢字チェッカー 日本語能力試験の漢字レベルを基準にテキスト中の漢字の難易度レベルを自動判定する

図4がチュウ太の工具箱の入力画面である。教材化したいテキストを、画面中央の白い枠内(テキストボックス)に入力する。このテキストボックスには、キーボードから直接入力することもできるが、インターネット上の情報をコピー&ペーストで切り張りすることも可能である。テキストの入力後に、テキストボックス下のボタンの中から、支援ツールを選択する。学習者の必要な作業は、自らが必要とする支援ツールのボタンを押す

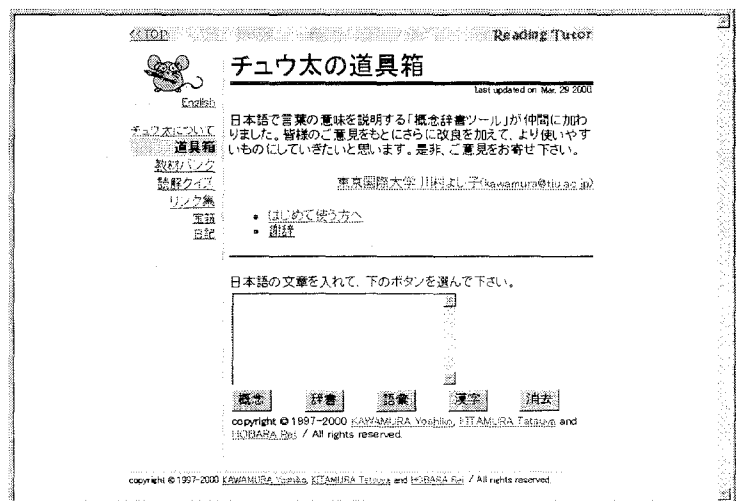


図4 チュウ太の工具箱の入力画面

ことのみである。あとはすべてコンピュータが自動的にレベル判定や辞書引き作業を行ってくれる仕組みになっている。「概念」と「辞書」のボタンは、それぞれ、日英辞書ツールと概念辞書ツールとを意味している。この日英辞書ツールと概念辞書ツールは、いずれも入力されたテキストの形態素解析と、単語の辞書引き作業を自動的に行い、各単語の読みと意味とを表示するツールである。形態素解析には形態素解析システム2.02「茶釜」<sup>9</sup>を用いている。茶釜は入力された文を自動的に形態素に区切り、各語の辞書形、読み、品詞情報等を出力するシステムである。辞書ツールはこの解析

の結果得られた語と辞書情報とをリンクして表示する。意味の表示に関しては、各単語の持つ意味概念を列挙して表示する。両者の違いは、この意味概念の表示が、日英辞書ツールでは英語で、概念辞書ツールでは日本語で行われるという点である。教材バンクに収められた読解教材はすべてこの日英辞書ツールを用いて教材化したものである。インターネット上で得られた情報を「チュウ太の工具箱」にコピー&ペーストし、「日英辞書ツール」あるいは「概念辞書ツール」のボタンを押せば、教材バンクの教材と同じ形の、辞書とリンクした教材を得ることができる。

一方、「語彙」のボタンを押すと、語彙チェッカーが選択できる。語彙チェッカーは、テキストに含まれるすべての単語の難易度を日本語能力試験の語彙レベル<sup>10</sup>をもとに自動判定するツールである。入力されたテキスト全文を形態素解析システム「茶釜」2.02<sup>9</sup>を用いて単語に区切る。語彙チェッカーは、この茶釜の分析結果を日本語能力試験の1級から4級までの語彙リストと照合し、個々の語のレベルを判定している。図5がその出力画面である。(ここでは『毎日新聞デジタル切り抜き帳』の「チェックオフ」を入力した。) 左フレームのテキストにはレベル判定の結果が色分けして表示される。

◇チェックオフ 労働組合費を会社が労働組合に代わって給与天引きで徴収し、組合に渡す制度。従業員の過半数が加入するなどの条件を満たしている労組に対して、労働基準法などで保障されている。労働省の労働協約等実態調査報告によると、条件を満たす組合を持つ企業の94.3%が採用している。同制度が禁止された場合、組合員からの滞納増加が予想され、結果として組合費の総量が減り、野党や連合の活動にも大きな影響が出る可能性がある。自民党が同制度の見直しを求める背景には、衆院選を控え、野党の支持基盤のひとつである労働組合に財政面から圧力をかける狙いがあると見られる。

総数	167(93種)
総語彙数	152(89種)
総語彙外数	15(4種)
級外	15(15種)
チェック	1
オフ	1
天引き	1
従業員	1
労組	1
に対して	1
労働省	1
協約	1
94.3	1
として	1
総量	1
自民党	1
見直し	1
衆院	1
選	1
一級	16(14種)
徴収	1
加入	1
満たす	2

「この文で使われている単語」★★★★ 少し難しい  
 総数 総語彙数 級外 一級 二級 三級 四級 特殊  

167	152	15	16	43	8	70	15
100%	91%	9.0%	9.5%	25.7%	4.8%	41.9%	9.0%
100%	100%	9.9%	10.5%	28.3%	5.3%	45.1%	9.9%

図5 語彙チェッカーの出力画面

デフォルト値としては2級終了程度の学習者を想定し、1級以上の単語を赤で表示するように設定してある。右フレームは級ごとの単語リストである。各単語の右の数字は当該の単語がテキスト中に何回現れたかを表示している。このリストによって、本文中の学ぶべき単語が一目でわかる。下フレームではテキスト全体に対する各級の単語の含有率が表示される。そして、その結果をもとに語彙レベルから見たテキストの難易度を星印で示している。

一方、工具箱の「漢字」のボタンを押せば、漢字チェッカーが選択できる。漢字チェッカーはテキストに含まれるすべての漢字の難易度を日本語能力試験の漢字レベル<sup>10</sup>をもとに自動判定するツールである。表示画面や表示方法は語彙チェッカーと全く同じ形式になっている。学習者は自分の読もうとしている文章の中にどのレベルの漢字がどの程度含まれているのかを知るとともに、学ぶべき漢字のリストがその場で得られる仕組みになっている。

学習者はこれらのツールを利用して、自分の選んだ文章を自由に教材化して学ぶことになる。この読解教材バンクは単に教材を提供するだけでなく、インターネット上のあらゆる情報を教材として活用するための学習環境を提供しているのである。

#### 4. 教材バンクの評価

「チュウ太の読解教材バンク」に対する学習者の評価は高い。「教材が面白い」「日本語の勉強ばかりでなく、日本を知るのに役に立つ」「日本人の考え方を知ることが出来た」といった教材そのものへの評価に加え、「単語の読みや意味がすぐにわかるので便利だ」「いちいち辞書を調べなくて済むのでどんどん読める」と、システムへの評価も高い。また、「これを使って、毎日、新聞を読むことにした」「日本人の友人からのメールを読むのにも利用できる」と、積極的に自分なりの活用をしている学習者もいる。

ところが、この教材バンクでは、入力した本文をコンピュータが自動的に形態素解析し、その解析結果をもとに辞書引きを行っている。そのため、解析ミスによって誤った辞書情報が提供される可能性は否定できない。現在の所、解析の精度は98.5%程度である。肝心の辞書情報が誤っていれば学習を妨げてしまうことも起こりうる。

特に初級・中級学習者に対しては十分な配慮が必要である。

前述したように読解教材バンクに収められている教材も、すべて日英辞書ツールを用いて教材化したものである。著作権の問題もあり、提供してもらった読み物をそのままの形で辞書ツールにかけて教材化している。その結果、茶筌が解析しにくい文章もあり、単語を正確に切り出せず、誤った辞書情報が表示されている箇所が各教材ごとに数カ所発見された。以下、どのような場合に問題が発生したかをまとめるとともに、それをどのようにすれば回避できるかを明らかにする。

a. ひらがなの多い文

【問題点】ひらがなが長く続くと解析ミスが発生しやすい。例えば、「おにいさん」は「お・にいさん」と分析され、間投詞としての「お」の意味が表示されてしまう。「頃」「誰」等もひらがなだと、「ごろつき」「だれる」のような意味が表示されてしまう。「あいづち」も「あ+いづち」になる。また「かんたん」「しつもん」等の単語では、語の切り出しは正確に行えるが、辞書にひらがなでは登録されていないので、辞書表示が出来ない。

【解決方法】通常漢字で書かれる言葉は漢字に直す。

b. 複数の読み方のある単語

【問題点】茶筌は、基本的な構文解析によってより頻度の高い読みを優先して表示する仕組みにはなっているが、「御」は「ゴ」、「今日」は「キョウ」という読みが常に表示される。また、「行った」「通った」はそれぞれ「オコナウ+タ」「トオル+タ」と分析される。

【解決方法】読み誤りの生じてしまう語句に関しては、ひらがなで表記し、正しい解析結果が得られるようにする。ただし、「行った」に関しては「いった」と表記すると今度は「イウ+タ」と分析されてしまうという問題が残るため、辞書ツールそのものに手を入れ、「いう」「いく」に関しては解析結果を表示しないことにした。

c. ひらがな表記の同音異義語

【問題点】ひらがなで書かれた同音異義語の分析は機械では極めて難しい。さらに動詞では、「ある／あう」「いく／いう」「くる／きる」等、ひらがな表記ではテ形・タ形が同じになってしまう活用形もある。

【解決方法】機械での処理を考えると、これらの語句は可能な限り漢字表記が望ましい。ただし、補助動詞として用いられているものについてはひらがな表記が一般的なため、分析のミスが発生する可能性の高いものは辞書情報を表示しないという形で辞書ツールを改良することにした。

d. 接辞的に用いられることのある単語

【問題点】「人」「本」「物」等の語句は他の名詞と組合わさって「～人」「～本」「～物」と言う形で用いられることがある。そのため、こうした語句が名詞の直後にあると解析ミスが生じやすい。

【解決方法】直前の名詞との間に「、」を挿入することでこの解析ミスは完全に防ぐことができる。

e. 副詞および副詞相当語句

【問題点】日本語において副詞の位置は比較的自由であるが、機械にとっては区切りの判断が難しくなる。例えば「昼間人に会った時」と言う文は「ヒルマ・ジン・ニ・アウ・タ・トキ」と分析されてしまう。

【解決方法】副詞の直後に「、」の挿入によって完全に防ぐことはできるが、また、その副詞を動詞の直前に持ってくることもよっても解析ミスの大半は防ぐことができる。

f. 複文や複雑な構造の文

【問題点】文の構造が複雑になればなるほど解析ミスが生じる可能性は増えてくる。例えば、「日本ではみんなで注文してお酒を注いだり注いでもらったりしながら飲みます。」という文では「注い・だり・注・い・で・もらっ・たり」となり、2番目の「注ぐ」を解析しきれない。また、「お土産をもらったらアメリカではどうしますか」という文では「もらったら」の「ら」が接尾辞として分析されてしまう。

【解決方法】いずれの場合も「、」の挿入によって解析ミスを防ぐことができる。上記の例でも「注文して、お

酒を注いだり、注いでもらったりしながら」「お土産をもらったら、アメリカではどうしますか」とするだけできれいに解析される。

以上のように、解析ミスのおお半が表記方法の工夫と句読点の挿入によって回避できることがわかった。その他の解析の誤りに関しても、文型を変える、語句を変える等の工夫によってほとんど回避できる。また茶釜の辞書にその語句を付け加える必要があるものに関しては、今後も整備をすすめていく予定である。

終わりに。

教材バンクに収録した教材に関しては、原作者の許可を得て、上記の修正作業を行った改訂版を提供することにした。しかし、利用者がインターネット上のテキストをそのまま教材化する場合には上記の問題が生じてしまう可能性は常に存在する。上級学習者であれば、それなりの対応が可能だろうが、初級・中級の学習者に対しては、できるだけ正しい辞書情報が表示されるようあらかじめ手を入れた教材を提供していく必要がある。

この読解教材バンクもインターネット上で公開している (<http://language.tiu.ac.jp>) ので、インターネットで日本語が読める環境にさえあれば、世界のどこからでも利用可能である。一人でも多くの方々に利用していただき、いろいろな方のご意見・ご助言をもとにさらによりよいものにしていきたいと考えている。

#### 謝辞

本システムは静岡大学北村達也氏との共同研究によるものです。開発には東京大学理学部の保原麗氏から多大な協力を得ました。奈良先端科学技術大学院大学の松本裕治先生からは一般公開前の「茶釜」2.0b8を提供していただきました。また、EDRの酒井佐芳氏からはEDR日英対訳辞書を用いた辞書ツールの公開許可をいただきました。さらに、氏家研一氏、中島玲子氏をはじめとして多くの方々が教材をご提供くださいました。ここに記して感謝の意を表します。

#### 参考文献

1. Henstock, Peter V. "Applications in Computer Aided Language Learning for the Student and Instructor", M. A. thesis, Purdue University, 1995.
2. 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.
3. 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」『日本教育工学会論文誌』23(3), pp.127-133, 1999.
4. 川村よし子・北村達也・保原麗「日本語CALLシステムにおけるEDR日英対訳辞書の活用」『日本教育工学会報告集』JET2000-1, pp. 7-12, 2000.
5. 川村よし子「読解のためのレベル判定システムの構築 ―語彙チェッカーの開発と活用」『日本語教育方法研究会誌』Vol.5, No.2, pp. 10-11, 1998.
6. 川村よし子「漢字の難易度判定システム『漢字チェッカー』を用いたテキストの分析」『東京国際大学論叢』第59号, pp. 73-87, 1999.
7. 川村よし子・金庭久美子・北村達也「インターネットを活用した読解学習システムの構築とその評価」, 『日本語教育学会秋季大会予稿集』, pp.63-68, 1999.
8. ㈱日本電子化辞書研究所『EDR電子化辞書仕様説明書』1996
9. 松本裕治・北内啓・山下達雄・平野善隆・松田寛・浅原正幸「日本語形態素解析システム『茶釜』ver 2.0 使用説明書第二版」, NAIST Technical Report, NAIST-IS-TR99012, 1999.
10. 日本語能力試験企画小委員会編『日本語能力試験出題基準』国際交流基金・日本国際教育協会, 1993.