

自律学習を可能にする日本語読解支援環境の構築

代表研究者 北村 達也 静岡大学情報学部助手

共同研究者 川村 よし子 東京国際大学商学部教授

1 はじめに

インターネットの急速な普及は語学教育の分野にも変化をもたらし、様々な教育ツールがインターネット上で提供されるようになってきた。だが、その多くは教師に代わって教えるための道具として教師の視点から開発されたもので、学ぶための道具として学習者の視点から開発されたものはまだまだ少ない。インターネットという無尽蔵ともいべき情報源を学習資源として有効に活用するには、CAI (Computer-Assisted Instruction) からCALL (Computer-Assisted Language Learning) への発想の転換が必要である。

インターネット上の情報を語学教育に活用しようという試みはすでに英語教育の分野では進められている。日本においても高校・大学等において英語学習にインターネットが利用されはじめている[10][8]。しかし、日本語教育の分野では各教育機関へのコンピュータの導入自体が遅れていることもあり、近年になってようやくインターネット上で利用可能な学習支援ツールが開発されはじめたというのが現状である。

このような背景のもと、本研究では、自然言語処理技術を活用して日本語学習者の漢字学習と文章読解を支援するシステムを構築した。これらのシステムはWebページ上で公開されているため、インターネットを介して世界中の日本語学習者が無料で利用できる。

2 マルチメディアを利用した漢字学習システム

単語や漢字を学ぶ際には1つ1つの属性を覚えるばかりでなく、文脈の中で単語や漢字を学ぶ必要もある。本研究では、自動的に文章中の単語に読みがなをリンクし、さらに単語中の漢字にマルチメディア属性データ（書き順の動画像、音訓読み、用例、音声）をリンクする MOKO (Multimedia-Oriented Kanji Online) system により、この学習法を支援することを目的とする。なお、本システムは、<http://fl176.hyper.chubu.ac.jp/moko/inputpage.html>にて公開している。

2.1 MOKO system

MOKO systemによる処理の流れを図1に示す。利用者は MOKO system のWebページにアクセスし、テキストエリアに自分の処理したいテキストを入力する。入力されたテキストはサーバに送られ、以下の処理が行われる。

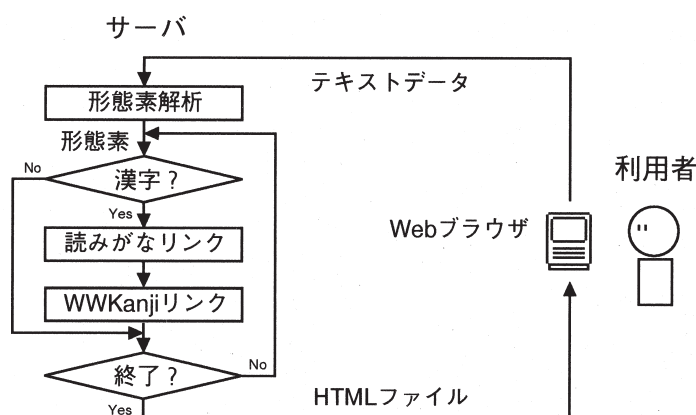


図1 MOKO systemの処理の流れ

まず、形態素解析システム茶筌[6]により形態素に分解される。これにより個々の形態素の品詞、読みがな、辞書形などの情報が得られる。

次に、形態素に漢字が含まれるか否かを調べ、含まれている場合にはその形態素と茶筌の処理により得られる読みがなとをリンクする。さらに、形態素中の漢字をWWKanjiの属性データとリンクする。WWKanjiは小森ら[5]により開発された、漢字の書き順の動画像、音訓読み、用例、音声を含むマルチメディアデータベースである。WWKanjiは全ての常用漢字を含んでいるため、一般的な文章に現れるほとんどの漢字に属性データをリンクできる。

処理の結果はHTML (Hyper-Text Markup Language) 形式で出力され、図2のように利用者側のコンピュータのWWWブラウザに表示される。WWWブラウザは文章フレーム、読みがなフレーム、漢字フレームに3分割される。文章フレームには入力したテキスト、読みがなフレームには文章中の漢字を含む形態素の読みがな、漢字フレームにはWWKanjiの属性データが表示される。

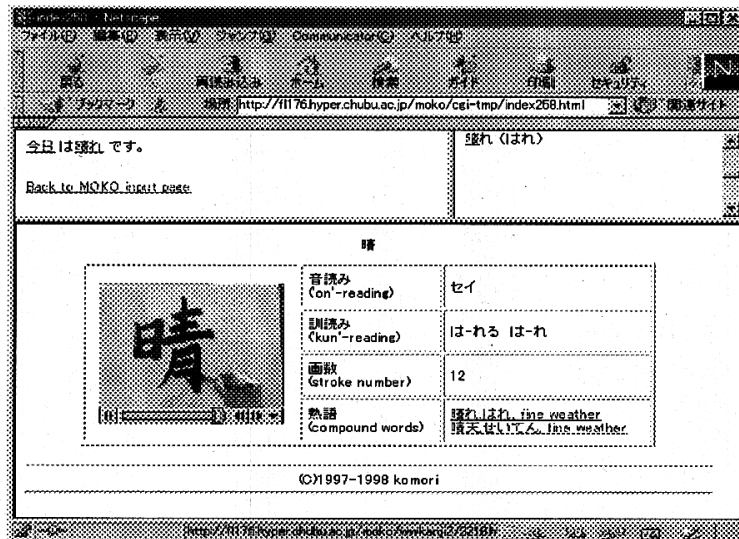


図2 MOKO systemの出力画面

利用者が文章フレーム内の単語をクリックすると読みがなフレームにその単語の読みがなが表示される。さらに読みがなフレーム内の漢字をクリックすると、WWKanjiの属性データにアクセスできる。これによって、学習者は文章を読み進めつつ個々の漢字について学ぶことができる。

2.2 漢字学習効果に関する実験

2.2.1 実験協力者

中部大学の1年生の留学生12名。実験協力者の出身国の内訳はカナダ1名、タイ1名、韓国1名、中国9名であった。留学生別科や日本語学校などで1年から1年半の学習を終え、大学入学し、1年経過した学生であり、日本語能力は中級後半である。

2.2.2 方法

12名の実験協力者を2グループに分けてMOKO systemを利用する前後にテストを行った。プレテストでは、文章が全て平仮名で書かれている紙を渡し、できるだけ漢字に直すよう指示した。それぞれのグループには別の文章を与えた。文章は1999年1月10日朝日新聞朝刊コラム私の見方(大牟田透 科学部)を用い、その前半と後半をそれぞれのグループに与えた。そして、回答に要した時間、学生が書いた漢字(熟語を含む)、漢字の正答数を計測した。

次に、この文章を漢字仮名交じり文で書いたものをMOKO systemを利用して読ませた。わからない漢字があったら、クリックして調べることができるという指示をした。その後、プレテストと同じ方法でポストテストを行った。

2.2.3 結果と考察

実験協力者は平均して25～30分程度MOKO systemで学習した。プレテストとポストテストの間では漢字の回答数、正答数、正答率が向上し、回答に要する時間も減少する傾向があった。この結果はMOKO systemの効果のみを表しているわけではないが、MOKO systemには漢字学習効果のあることが示唆される。今後、より詳細な評価実験を行う必要がある。

2.3 MOKO systemの課題

MOKO systemの課題は、WWKanjiの属性データへのアクセスの高速化である。マルチメディアデータはデータ量が大き

きいためデータ転送に時間がかかる。この問題を解消するためには国内外の教育機関と協力し各地にMOKO system のミラーサーバーを構築する必要がある。

3 EDR電子化辞書を活用した辞書ツール

辞書ツールはリーディング チュウ太のWebページ (<http://language.tiu.ac.jp/>) 上に置かれ、日本語学習のためのツールとして公開されている。

このツールは、入力された文章を茶筌[6]で解析し、その結果をもとに語と辞書情報をリンクして表示する。辞書としては、EDRの日英対訳辞書[7]（以下「EDR日英辞書」と略）を採用している。

本システムは、DL[9]及びDL2[4]の基本コンセプトを継承しているが、もととなる日英辞書として、これまでのEDICT[2]にかえて、正確かつ大規模なEDR日英辞書を採用した。また、茶筌の解析結果をもとにEDR日英辞書が持つ語の読みと品詞の情報とを活用して辞書情報を絞り込み、学習者にとって不要な情報を排除した。さらに本システムでは、各語の持つ意味の広がりに関する概念情報を日本語で提供する機能も備え、英語という媒介語の助けを借りずに語彙学習ができる仕組みも整えた。

3.1 EDR日英対訳辞書

辞書ツールで用いているEDR日英辞書は、約23万語に及ぶ見出し語の辞書情報を有している。各見出し語には、レコード番号、見出し情報〔読み情報〕、品詞情報、(英語及び日本語による)意味の概念情報、対訳情報、管理情報が提供されている [7]。EDR日英辞書においては、同一の見出し語であっても読み及び概念が異なれば、別のレコードとして登録されている。そのため、同一の見出し語に対して複数のレコードが存在する。

3.2 辞書ツールの仕組み

辞書ツールは入力された文章中の語に読みと英訳をリンクしてWWWブラウザ上に表示するシステムである。図3に本システムの処理の流れを示す。Webページのテキストエリアから入力された文章は、サーバに送られ以下の処理が行われる。

まず、茶筌による形態素解析が行われる。解析結果をもとにして、それぞれの形態素の辞書形によってEDR日英辞書を検索し、英訳の絞り込みの後、各形態素と英訳をリンクする。EDR日英辞書の検索には、新たに開発したパトリシア木を用いた検索システムを用いている。リンクの情報はHTMLにより記述される。最後に以上の処理結果を学習者のWWWブラウザ上に表示する。

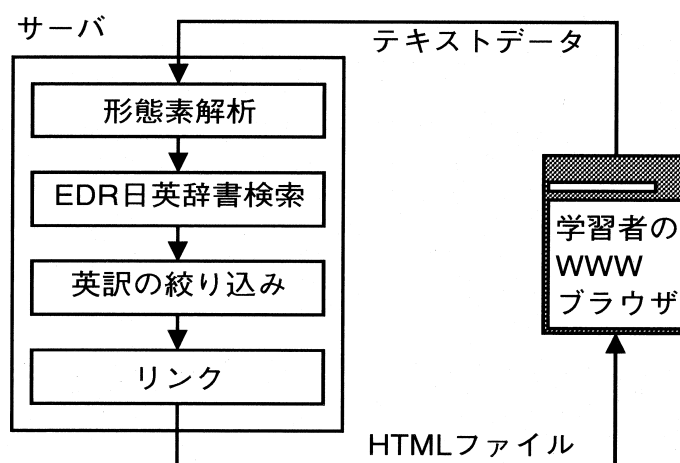


図3 辞書ツールの処理の流れ

図4が入力画面である。学習者は画面中央のテキストエリアに文章を入力（あるいはコピー&ペースト）し、利用したいツールボタン（ここでは「辞書ツール」）を押す。図5が辞書ツールの処理結果を示す出力画面である。左側のフレームには入力された文章が形態素ごとに区切られて表示されている。各々の語は、右側の辞書フレームの当該語とリンクされ、学習したい語をクリックすると、辞書フレームに読みと英訳が表示される。

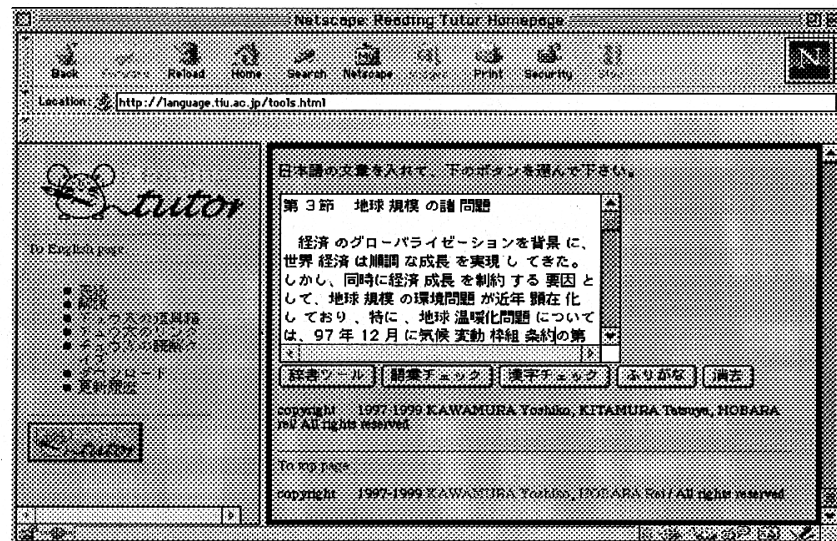


図4 辞書ツールの入力画面

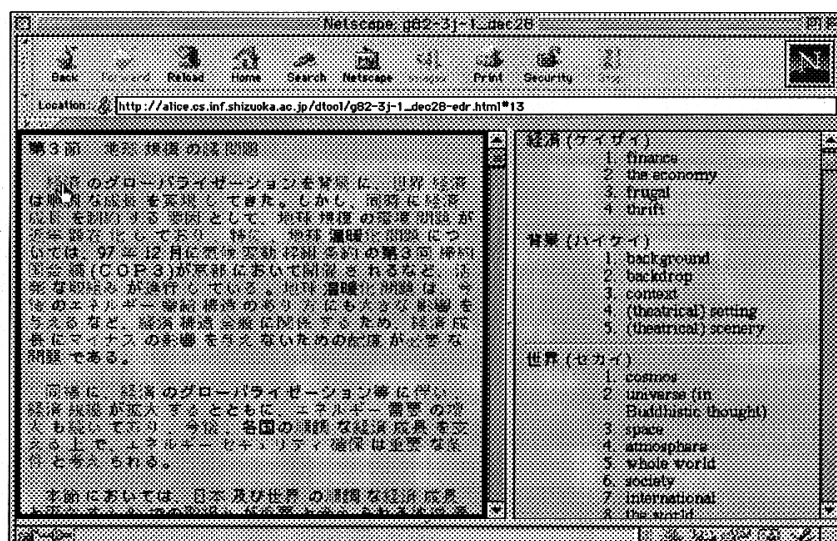


図5 辞書ツールの処理例

日本語学習者にとって、読みさえもわからない単語を辞書で引くというのは大変な作業であり、辞書引き作業のたびに読解が中断される。そのため個々の単語の意味を調べることに時間をとられて、文章全体の意味をつかむという肝心の部がなおざりにされることも多い。本システムを利用することにより、辞書引きによる中断なしに文中の単語の意味を知り、読解を進めることができる。また、辞書ツールには、学習語の知識の定着を促すように、学習者がアクセスした語のリストを文章読解後に表示する機能[4]も整備されている。

3.3 英訳情報の絞り込み

上述したように、EDR日英辞書では読みや概念の異なりごとに、語が細かく分類され、その各々に該当する英訳が列挙されている。そこで、英訳情報のうち学習者に必要なもののみを表示するように絞り込みを行うことにした。

辞書ツールでは、茶筌の解析結果とEDR日英辞書とを照合し、読み及び品詞情報が一致した英訳のみを表示することにした。また、付属語である助詞、助動詞及び数詞の英訳は表示しないことにした。さらに、英訳の重複を除去する処理も並行して行うことにした。

品詞情報に関しては、茶筌とEDR日英辞書では品詞の分類方法が微妙に異なっている。そのため、この違いを吸収する処理を行っている。代表的な例はサ変動詞と形容動詞(ナ形容詞)である。

サ変動詞とは「する」を伴って動詞として用いられる一連の名詞グループであるが、茶筌ではサ変動詞の形を持つ名詞は全てサ変動詞として分析される。ところがこれらの語の中にはサ変動詞としての意味以外に純粋名詞として別の意味を持つ語も存在している。例えば「料理」や「司会」という語は「料理する」「司会する」という形でサ変動詞として用いられるが、純粋名詞として「料理したもの」「司会をする役の人」という意味も別に持っている。このような語の場合、EDR日英辞書ではサ変動詞としての意味と純粋名詞としての意味が別々に登録されている。そのため茶筌の結果のみをもとに絞り込みを行うと純粋名詞であってもサ変動詞の英訳しか得られない。そこで辞書ツールでは名詞とサ変動詞それぞれの英訳を並記することとした。

形容動詞に関しても特別な処理が必要である。茶筌では形容動詞の形をもつ名詞は全て形容動詞として分析される。ところが例えば「健康」「元気」のように「健康な状態にある」「元気な状態にある」という形容動詞としての用法以外に「健康」「元気」という純粋名詞としての用法を別に持つ語の場合、学習者には、その意味も示す必要がある。また「現金」のように形容動詞として用いられた場合、名詞としての本来の意味とは異なった意味を持つものも存在する。そこで茶筌で形容動詞と解析された語に関しては、名詞と形容動詞の双方でEDR日英辞書を検索するようにした。

その他、接頭辞・接尾辞・助数詞等の分類基準も異なっている。個々の品詞に関して詳しく調査し、表1に示す茶筌とEDRの品詞対応表を作成した。辞書ツールはこの表にもとづいて品詞を照合し絞り込みを行っている。

表1 茶筌2.01とEDR日英辞書の品詞情報の対応表

品詞	茶筌	EDR 日英辞書
名詞	1-16	JN1-JN4, JAM
サ変動詞	17	JN1, JVE
形容動詞	18	JN1, JAM
数詞	19	リンクせず
形式名詞	20-26	JN1, JB1
接尾辞	27-32	JN4, JB1, JUN
助数詞	33	JUN, JN6
名詞 (補助的用法)	34-37	JB1
接頭辞	38-42	JT1-JT4, JN5, JNM
前置助数詞	43	JT1-JT4, JN5
動詞	44-45	JVE
補助用言	46-47	リンクせず
形容詞	48-51	JAJ
副詞	52-54	JD1-JD2
連体詞	55	JNM
接続詞	56	JC1-JC3
助詞	57-70	リンクせず
助動詞	71	リンクせず
感動詞	72	JIT

3.4 絞り込みの効果

前節で述べた英訳情報の絞り込みの効果を調査した。『平成10年度通商白書』の「2-3-1 地球環境問題」(3975字)を入力文章として、絞り込みにより不要な英訳がどれだけ減少するかを調べた。茶筌による解析によって得られた延べ語数2323語のうち助詞、助動詞、数詞及び重複を除去した単語数は487語である。これをもとにEDR日英辞書を検索するとレコード総数は3567となる。品詞および読みによる絞り込みの結果、品詞によって2102レコード、また、読みによって487レコードの不要な英訳を排除することができた。

以上の絞り込み作業は、茶筌による読みと品詞情報の解析が正確にできていることを前提としているため、茶筌の性能が絞り込みの成否を大きく左右する。今回の文章の解析結果において、解析の誤りは「ベルリンマンドート(注:地名)」(ベルリン/マン/デート)のみであった。

また、読みに問題のあるものは「本節（ホンブシ）」「92年（トシ）」「議定（ギジョウ）書」等、8種類あった。「本節」に関しては茶釜およびEDR日英辞書への単語登録によって修正可能であり、その他はすべて、複数の読み候補を出力するという茶釜の機能を利用することで解決可能なものである。ただし、複数の読みを併記した場合、学習者がその中から正しい読みを選び出すことは難しい。また、英訳情報も増えてしまう。いずれの形を採用すべきかは今後より詳しい調査を行った上で決定したい。

4 概念情報の表示

3.1節で述べたようにEDR日英辞書には語の概念情報も収められている。辞書ツールでは、英訳表示機能とは別に、日本語による概念説明を表示させる機能を備えることにした。この概念表示機能は特に上級レベルの学習者向けに開発した。

一般に辞書というと、まず学習者の母語との対訳辞書が考えられる。だが上級学習者にとっては日本語による説明で十分理解可能なはずである。また、日本語学習者の中には漢字圏出身者も多い。漢字の助けを借りることができれば、中級の学習者であっても大半の説明文は理解可能である。

図6が概念表示機能を用いた出力画面である。基本設計は辞書ツールの英訳表示機能と同一である。左フレームの単語をクリックすると右フレームに単語の読みと日本語による概念説明が表示される。

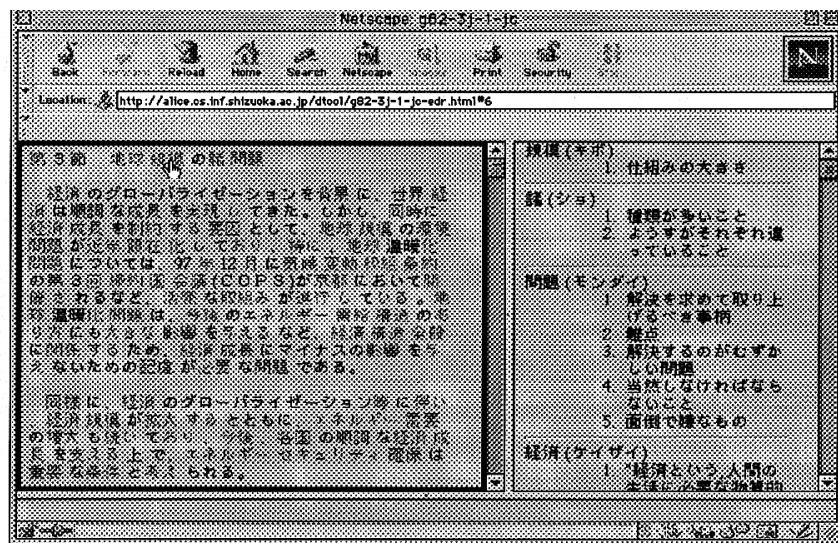


図6 日本語概念情報の提示画面

この概念表示機能を早稲田大学の上級および研修（超上級）クラスの学習者に試用させたところ、2クラス計35名の学習者全員から「役に立つ」という肯定的評価を得た。特に研修クラスの学習者からは「日本語の単語の意味は、日本語で説明された方がわかりやすい」「訳語があると、その言語に頭のスイッチを切り替えなければならない」「日本語で日本語をどう説明するか、概念の説明文自体も日本語の勉強になる」等のコメントがあった。通常、上級ともなれば、新出語彙も日本語で説明され、母語の助けはほとんどなしに学習がすすめられる。インターネットを用いた読解学習においても上級以上の学習者にとっては日本語で日本語を理解するツールが求められていると言えよう。

この日本語による概念表示機能に関しては、概念説明自体が日本語学習者用に書かれたものではないため説明に用いられている表現が難しいことや、概念説明の中に当該単語が含まれ同語反復になっているものもあるといった問題がある。従って、学習ツールとしての完成度を高めるには、説明文の修正作業が不可欠である。また、概念説明によって、単語の意味がどの程度把握できるのか、母語による訳語を示した場合と日本語による概念説明を示した場合とで語彙学習にどのような違いが生じるか等に関しても調査する必要がある。

5 おわりに

本稿では、マルチメディアを利用した漢字学習システム及びEDR日英辞書を利用した読解学習支援システムについて報告した。前者は、形態素解析技術と漢字属性データを利用して、文章の中で漢字を学ぶことを支援する。後者は、EDR日英・概念辞書を利用して、文章の読解学習を支援することができる。

今後は、読解学習支援システムの高度化について検討を進める。頻度情報や共起情報[7]の利用や親密度[1]による重み付け等によって、さらに辞書情報の絞り込みを行い、より効率の良い読解学習支援ツールにしていく予定である。また、概念表示機能に関しては、例文の表示等を含め、学習者にわかりやすい概念説明を提供できるよう改良していく予定である。

謝辞

辞書ツールの開発にあたり、奈良先端科学技術大学院大学の松本裕治先生から一般公開前の茶筌2.0b8を提供いただきました。また、EDRの酒井佐芳氏からはEDR電子化辞書を用いた辞書ツールの公開許可をいただきました。(財)電気通信普及財団には本研究を支援していただきました。ここに記して感謝の意を表します。

参考文献

- [1] 天野、近藤 (1999), “NTTデータベースシリーズ 日本語の語彙特性”, 三省堂
- [2] Breen (1994), “EDICT”, <ftp://ftp.cc.monash.edu.au/pub/nihongo/edict.doc>
- [3] 川村、金庭、北村 (1999), “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
- [4] 北村、川村、内山、寺、奥村 (1999), “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”, 日本教育工学会誌 Vol. 23, No. 3, pp. 127-133
- [5] 小森、上田、難波 (1997), “オンラインデジタル動画漢字学習支援システム”, 人文科学とコンピュータ, 97-CH-36, Vol. 97, No. 108, pp. 49-54
- [6] 松本、北内、山下、平野、松田、浅原 (1999), “日本語形態素解析システム『茶筌』version 2.0 使用説明書 第二版”, NAIST Technical Report NAIST-IS-TR99012
- [7] 日本電子化辞書研究所 (1996), “EDR電子化辞書仕様説明書”, 日本電子化辞書研究所
- [8] Sugiura (1998), “Web Based Dynamic Reference System for English as a Foreign Language Learners”, CALICO Annual Symposium 1998, San Diego, CA, U.S.A.
- [9] 寺、北村、落水 (1996), “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
- [10] 山内豊 (1996), “インターネットを利用した英語授業”, NTT出版

<発表資料>

題 名	掲載誌・学会名等	発表年月
CGI-DL2:The Use of Learning Process Records in a Japanese Reading Support System	12th World Congress of Applied Linguistics	1999年8月
マルチメディアデータを利用した読解支援システム	The second International Conference on Computer Technology and Japanese Language Education	1999年8月
日本語CALLシステムにおけるEPR日英対訳辞書の活用	日本教育工学会	2000年1月
インターネットを活用した日本語読解教材の作成方法—学習者個々のレベルやニーズに対応した教材作成—	JALT The Language Teacher	2000年2月
概念辞書ツールを活用した語彙学習の提案	日本語教育方法研究会	2000年3月