

EDR 電子化辞書を活用した辞書ツールの開発*

川村よし子^{*1}・北村達也^{*2}・保原麗^{*3}

東京国際大学商学部^{*1}・静岡大学情報学部^{*2}・東京大学理学部^{*3}

日本電子化辞書研究所(EDR)の日英対訳辞書を利用した日本語読解学習支援システムを開発した。このシステムでは、形態素解析システム茶筌を利用して、入力された文章中の語を自動的にEDR日英対訳辞書とリンクして語の読みと英訳を表示する。リンクの際、茶筌により得られる読み及び品詞情報と、EDR日英対訳辞書の読み及び品詞情報とを照合させることにより、辞書情報の絞り込みを行う。加えて、EDR日英辞書に含まれる概念情報を表示する機能も実現した。学習者はこのシステムを活用して自分の読みたい文章を自由に教材化し学習することができる。

キーワード：日本語学習，教具開発，学習支援，EDR電子化辞書，インターネット

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た。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)への発想の転換が求められている。

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(山内1996)(Sugiura1998)。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(川村ほか1999a)。

こうした中であって、日本語学習者のための読解学習支援システム“リーディングチュウ太”(Reading Tutor)は、CALLシステムとして開発され、インターネット上で公開されている(川村ほか1999b)。このシステムには、学習者のためのリンク集が用意されるとともに、辞書ツール、ふりがなツール、語および漢字のレベル判定ツール等が組み込まれている。このシステムを用いることによって、学習者はインターネット上の情報を自由に選んで、自分のやり方で学ぶことができる。コンピュータは日本語を教える道具としてではなく、学習者が主体的に学ぶための道具として考えられている。

本稿ではこのシステムの主要なツールの一つである辞書ツールに焦点をあて、日本語CALLシス

*1Yoshiko Kawamura, Tatsuya Kitamura and Rei Hobar: Development of a reading tutorial system for JSL and JFL learners using the EDR Japanese-English dictionary

*1School of Business and Commerce, Tokyo International University, 1-13-1 Matobakita, Kawagoe, Saitama 350-1197 Japan

*2Faculty of Information, Shizuoka University, 3-5-1 Johoku, Hamamatsu, Shizuoka 432-8011 Japan

*3Faculty of Science, Tokyo University, 7-3-1, Hongo, Bunkyo-ku, Tokyo 133-0033 Japan

JEB0216595	今日 [キョウ]	JN4	1e850d	今日 [キョウ]	"the day that is today"	今日この日	0—"on this day"—EPP//0—"this day"—EN//0—today—EN	DATE="95/11/17"	
JEB0422837	今日 [キョウ]	JN4	3c5a5b	今日 [キョウ]	"the same date as today"	今日この日と同じ日付の日	0—"on this day"—EN//0—"this day"—EN//0—today—EN	DATE="95/2/15"	
JEB0068460	今日 [コンジツ]	JN4	0ef170	yesterday	近頃 [チカゴロ]	"the near past or present times"	近い過去から現在までの漠然とした期間	0—nowadays—EN//0—"these days"—EN//0—now—EN//0—today—EN	DATE="95/2/15"
JEB0216596	今日 [コンジツ]	JN4	1e850d	今日 [キョウ]	"the day that is today"	今日この日	0—"on this day"—EPP//0—"this day"—EN//0—today—EN	DATE="95/11/17"	
JEB0216597	今日 [コンチ]	JN4	1e850d	今日 [キョウ]	"the day that is today"	今日この日	0—"on this day"—EPP//0—"this day"—EN//0—today—EN	DATE="95/11/17"	
JEB0068461	今日 [コンニチ]	JN4	0ef170	yesterday	近頃 [チカゴロ]	"the near past or present times"	近い過去から現在までの漠然とした期間	0—"these days"—EN//0—"present day"—EN//0—today—EN	DATE="95/2/15"
JEB0145682	今日 [コンニチ]	JN4	1000c8	当代 [トウダイ]	today	今の時代	0—today—EN//0—"this day"—EN	DATE="95/2/15"	

図 1: EDR 日英辞書において「今日」を見出し語とする全レコード (7レコード)

テムという視点から、形態素解析システム茶筌と日本電子化辞書研究所 (EDR) の日英対訳辞書の辞書情報をどのように活用し、読解学習・語彙学習のためのツールとして完成させたかについて報告する。

2 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、入力された文章を形態素解析システム茶筌 2.01 (松本ほか 1999) (以下「茶筌」と略) 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 (日本電子化辞書研究所 1996) (以下「EDR 日英辞書」と略) を採用している。

本システムは、DL (寺ほか 1996a, 1996b) 及び DL2 (北村ほか 2000) 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 (Breen 1994) 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。

2.1 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 (日本電子化辞書研究所 1996)。一例として、「今日」という見出し語を持つ全レコードを図 1 に示す。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、この例のように同一の見出し語に対して複数のレコードが存在する。各レコードの第 2 フィールドが見出し語とその読みの情報、第 3 フィールド (この例では JN4 など) が品詞情報である。第 4 フィールドから第 8 フィールドまでが概念情報、第 9 フィールドが対訳情報である。

2.2 茶筌

辞書ツールにおいて文章解析に用いている茶筌は、入力された文章を形態素に分解し、その辞書形、読み、品詞情報等を出力するシステムである。形態素とは「意味をもつ最小の言語単位」をいう (長尾ほか 1990)。一例として、「学校に行きま

す。」という文章の解析結果を図2に示す。各行の第1フィールドが形態素、第2フィールドが読み、第3フィールドが辞書形、第4フィールドが品詞および活用情報を表している。

学校	ガッコウ	学校	2
に	ニ	に	59
行き	イキ	行く	45
ます	マス	ます	71
.	.	.	75
EOS			

図2: 茶筌 2.01 による「学校に行きます。」の解析結果(実行コマンド chasen -F %m \\t %y \\t %M \\t %h \\n)

2.3 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 CGI (Common Gateway Interface) により以下の処理が行われる。

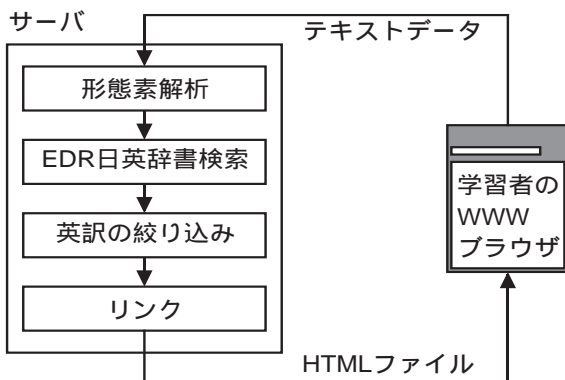


図3: 辞書ツールの処理の流れ

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 (Hyper Text Markup Language) 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。

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



図4: 辞書ツールの入力画面



図5: 辞書ツールの処理例

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分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、読解をすすめることができる。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(北村ほか 2000)も整備されている。

2.4 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。例えば図 1 が示すように、「今日」という見出し語で検索した場合、7レコードが該当する。読みは「キョウ」「コンジツ」「コンチ」「コンニチ」の4種類が含まれている。また意味概念は「キョウ」「コンジツ」「コンニチ」に対しては、それぞれ2つの異なった概念が収められている。そのため茶釜を用いて得られた形態素の辞書形のみをもとに EDR 日英辞書を検索すると、必要以上に多くの英訳が得られてしまう(異なり語数で7、延べ語数では21得られることになる。)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、茶釜の解析結果と 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、茶釜と 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、茶釜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。そのため茶釜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。茶釜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。そこで茶釜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 に示す茶釜と 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。

2.5 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。茶釜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、茶釜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

3 概念情報の表示

表 1: 茶釜 2.01 と EDR 日英辞書の品詞情報の対応表 .

品詞	茶釜	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

2.1 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。この日本語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: 日本語概念情報の提示画面

いるため、茶釜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」(ベルリン/マン/デート)のみであった。辞書ツールの開発過程で用いていた茶釜 2.0b8 では「締約国会議」を「締約/国会/議」と区切ってしまっていたが、茶釜 2.01 ではその部分も正しく解析できている。

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択することは難しい。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。

図 6 が日本語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。

この日本語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むずかしい，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。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。

4 おわりに

本稿では，EDR 日英辞書を利用した日本語読解支援システムについて報告した。本システムは，形態素解析システム茶筌と EDR 日英辞書の読み及び品詞情報を照合させることにより，辞書情報の絞り込みを行っている。加えて概念情報を表示する機能も実現した。

今後は，頻度情報や共起情報（日本電子化辞書研究所 1996）の利用や親密度（天野など 1999）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 を提供していただきました。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。なお，本研究の一部は（財）電気通信普及財団及び JALT Research Grant の支援により行われまし

た。ここに記して感謝の意を表します。

参考文献

- Breen, J. (1994), "EDICT", ftp://ftp.cc.monash.edu.au/pub/nihongo/edict.doc
- 天野成昭, 近藤公久 (1999), "NTT データベース シリーズ 日本語の語彙特性", 三省堂
- 川村よし子, 金庭久美子 (1999a), "インターネットを用いた日本語の学習支援システムの構築", 情報教育方法研究, 2(1):49-54
- 川村よし子, 金庭久美子, 北村達也 (1999b), "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:63-68
- 北村達也, 川村よし子, 内山潤, 寺朱美, 奥村学 (採録決定), "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
- 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 浅原正幸 (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
- 長尾真, 石田晴久, 稲垣康善, 田中英彦, 辻井潤一, 所真理雄, 中田育男, 米澤明憲 (1990), "岩波情報科学辞典", 岩波書店
- 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
- Sugiura, M. (1998), "Web Based Dynamic Reference System for English as a Foreign Language Learners", CALICO Annual Symposium 1998, San Diego, CA, U.S.A.
- 寺朱美, 北村達也, 落水浩一郎 (1996a), "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, 3(1):10-11
- 寺朱美, 北村達也, 落水浩一郎 (1996b), "日本語読解支援システム dictlinker", 平成 8 年度日本語教育学会秋季大会予稿集:43-48
- 山内豊 (1996), "インターネットを利用した英語授業", NTT 出版