

読解学習支援システムにおける辞書表示方法の改善

北村 達也（静岡大学情報学部）* 川村 よし子（東京国際大学商学部）†

概要

読解学習支援システム「リーディング・チュウ太」の辞書ツールに関して、意味情報の表示方式の改善を行った。辞書ツールは、学習者が読みたい文章を入力するとその文章中の単語に意味情報を自動的にリンクするシステムである。従来の辞書ツールでは単語の基本的な意味を表す意味情報が必ずしも上位に表示されていなかったため、学習者はその単語の中心的な意味を把握しにくいという問題があった。そこで、EDR 日本語単語辞書に含まれる頻度情報を用いて意味情報の表示順序を並べ替えたところ、良好な結果が得られた。本システムはインターネット上で無料公開している。

1 はじめに

インターネットの普及により、さまざまな Web ページが作られ公開されるようになった。今やインターネットは最も新しい情報が発信されるメディアである。そこでは、日本語によるさまざまな情報も提供され、誰でもその情報にアクセスすることができる。さらに、検索エンジン（goo や Yahoo など）を使うことにより、自分が必要とする情報が載っている Web ページに簡単にアクセスできるようになった。その結果、世界中の日本語教育の現場でインターネット上の日本語資源を教材として活用することができるようになった。

このような状況を反映して、近年、インターネット上の日本語資源を日本語教育に活用するためのシステムがいくつか提案されている。Henstock[1] による **AutoGloss/J** は、学習者の作成した単語リストに英訳を付けるものである。寺ら [2] による **DL** は入力された文章を形態素解析することにより文章を単語に区切り、各単語に EDICT[3] の英訳をリンクするシステムである。北村ら [4] による **DL2** は DL に学習履歴管理機能を付加したシステムであり、北村ら [5] による **MOKO system** は DL に漢字の書き順の動画（WWKanji[6]）を表示

*kitamura@cs.inf.shizuoka.ac.jp

†kawamura@tiu.ac.jp

する機能を付加したシステムである。越智ら [7] による JUPITER は漢字学習支援システムである。仁科ら [8] による **あすなろ** は文章中の単語の意味情報に加え、その文章の構文木も表示するシステムである。あすなろはさらに多言語辞書への対応も進めている。この他、文章中の漢字の出現頻度を計測する **KanjiCounter** が小島ら [9] により開発されている。

我々はこれまで 読解学習支援システム **リーディング・チュウ太 (Reading Tutor)** を開発し、インターネット上で無料公開している¹。このシステムは、辞書ツール [10]、単語および漢字の難易度判定ツール [11][12]、読解教材バンク [13] から成る。辞書ツールは入力された文章中の単語に EDR 日英対訳辞書 [14] の辞書情報をリンクする。難易度判定ツールは入力された文章中の単語もしくは漢字の難易度を日本語能力試験の級別に表示する。そして、これらのシステムで処理した読解教材を集めたものが読解教材バンクである。

本研究では、この辞書ツールの辞書表示方式の改善を行った。従来の辞書ツールでは単語の意味情報がランダムに並んでいるため、単語の基本的な意味が必ずしも上位に表示されなかった。そのため、学習者が複数の意味情報の中から適切なものを選択するのが難しいという問題があった。そこで、EDR 日本語単語辞書 [14] に含まれる頻度情報を用いて意味情報の表示順序の並べ替えを行った。

以下ではまず辞書ツールの仕組みについて説明し、次に意味情報の表示順序を並べ変える方法とその効果について述べる。

2 辞書ツールの仕組み [10]

辞書ツールは入力された文章中の単語に読みと意味情報を自動的にリンクして WWW ブラウザに表示するシステムである。本システムは文章の入力と結果の表示に WWW ブラウザを採用しているため、Windows, Macintosh, Linux などの OS で利用できる。また、本システムは Perl というプログラミング言語で作られており、Linux 上で動作している。意味情報としては、EDR 日英対訳辞書の英訳と日本語概念（日本語による単語の説明）の 2 つを用いている。従来の辞書ツールの処理の流れを図 1 に示す。

リーディング・チュウ太の 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。

まず、茶筌 [15] 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 日英対訳辞書を検索し、意味情報の絞り込みの後、各単語（形態素）と意味情報をリンクする。意味情報の絞り込みは、形態素解析により得られた各単語の読みと品詞に EDR 日英対訳辞書の各単語の読みと品詞を照合することにより行う。リン

¹<http://language.tiu.ac.jp/>

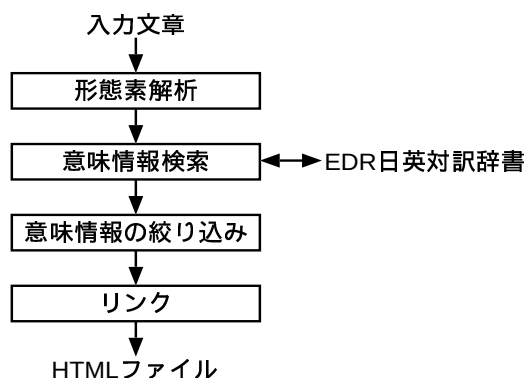


図 1: 従来の辞書ツールの処理の流れ

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 ブラウザに表示する。

辞書ツールの処理結果を示す出力画面を図 2 に示す。左側の文章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。それぞれの単語は、右側の辞書フレームの当該単語の意味情報とリンクされているため、文章フレーム内の意味を知りたい単語をクリックするだけで辞書フレームに意味情報が表示される。

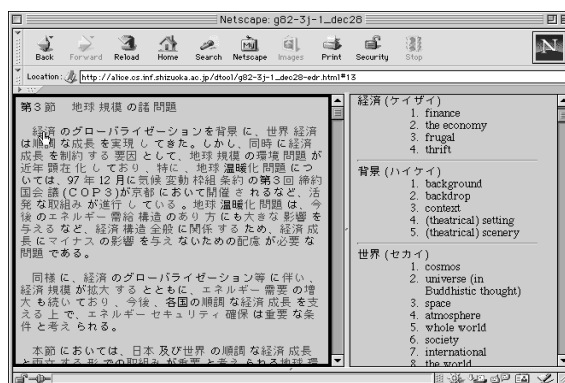


図 2: 辞書ツールの処理例

3 辞書表示方式の改善

3.1 意味情報の表示に関する問題点

以上の仕組みを持つ従来の辞書ツールでは、単語の意味を EDR 日英対訳辞書の配列順序でそのまま表示していた。EDR 日英対訳辞書の意味情報は概念番号によって整理されており、意味情報の配列もこの番号に従って並べられている。そのため、意味情報の表示において、EDR 日英対訳辞書の配列順序をそのまま用いると、単語の基本的な意味を表す意味情報が上位に表示されるとは限らないという問題があった。

例えば、「世界」という単語では、意味情報が以下の順序で表示されていた。

- 1) 仏教において、人間をふくむ全ての生き物が住む所
- 2) 宇宙全体
- 3) 芸術の創作物が作りあげている全体の場合
- 4) 人々が互いに関係しあいながら暮らす場
- 5) 地球上の全ての地域
- 6) ある物事に関する特定の範囲
- 7) 共同の目的を持ち結合した集団

また、「気」という単語では、以下の順序で表示されていた。

- 1) 関心
- 2) アルコールにおいて、香味
- 3) 気持ち
- 4) 人の心の動き
- 5) 人の性格
- 6) 心配という気持ち
- 7) 気体という物体

これらの意味情報はこの単語の説明としては全て正しい。しかし、このままの順序で表示されると、日本語学習者は各単語の中心的な意味を把握することが難しい。特に、初級・中級の日本語学習者がまず学ぶ必要のある基本的意味は、より上位に表示される必要がある。例えば、日本語学習者用に編集された『外国人のための基本語用例辞典第二版』（以下『基本語用例辞典』と記す）[17] では、「世界」の意味は次のように記述されている。

1. 地球全体。地球の上にある国全部。

2. 同じ種類のものの集まり、同じ仕事をしている人たちの集まり、

ところが、EDR 日英対訳辞書の配列順のままでは、この2つの意味が7項目中の最後の3つになってしまっている。「気」についても類似のことがいえる。『基本語用例辞典』の記述は次の通りである。

1. 心についていう、

- (a) ある人のもっている心の傾向、生まれつきの性質、
- (b) 何かしようとする、また何かしたいと思う心の働き、つもり、考え、
- (c) いろいろと考える心、心づかい、心配、えんりょ、
- (d) 心にかけて注意すること、
- (e) ある物事が心にあたえる感じ、またその心のようす、感情、気分、
- (f) 慣用的に使われているもの、

(辞典ではこの下に慣用句とその意味が列挙されている)

2. 空気・水蒸気などの気体、
3. ある物が特にもっているにおいや味、

EDR 日英対訳辞書の配列順序では最上位に「関心」があるものの、これは『基本語用例辞典』の1番目の意味範囲のごく一部にすぎない。しかも EDR 日英対訳辞書では2番目の意味として「アルコールなどにおいて、香味」があるため、学習者にとっては「気」の意味の把握が極めて難しいものとなっている。以上の例でも明らかなように、辞書ツールの表示する単語の意味情報の表示順序に関して、何らかの対応が不可欠な状態であった。

3.2 頻度情報による並べ替え

上記の問題点を解決するために、EDR 日本語単語辞書に含まれる頻度情報を用いて EDR 日英対訳辞書の意味情報の表示順序を並べ替えることを試みた。

この頻度情報は、22 万文から成る EDR コーパス [14][16] 中に含まれる単語がどの意味(概念)で用いられたかをカウントしたものである。本研究では、コーパス中の各単語はその単語の基本的な意味でより多く使われると考えて、出現頻度の高い順に意味情報を並べ替えることとした。EDR 日本語単語辞書には、概念別頻度と単語別頻度が記載されているが、本研究で用いているのは前者である。

この処理を加えた辞書ツールの処理の流れを図3に示す。意味情報の絞り込みの後、選ばれた意味情報が頻度情報によって並べ替えられる。EDR 日英対訳辞書の意味情報と EDR 日本語単語辞書の頻度情報は共通の単語コードで管理されているため、この単語コードを用いて意味情報と頻度情報を照合している。

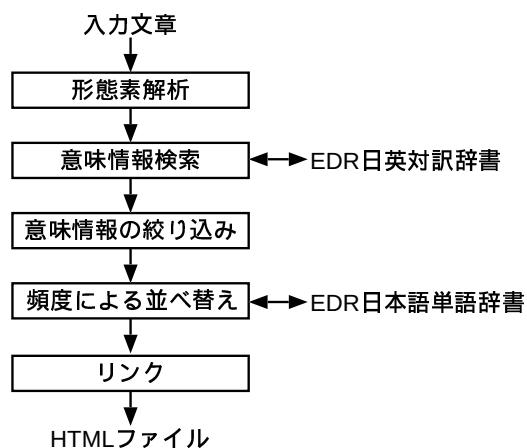


図 3: 新しい辞書ツールの処理の流れ

以上の処理の結果、上記の「世界」と「気」の意味情報は以下のように並べ替えられた。基本的な意味情報が上位に表示されていることがわかる。なお、カッコ内の数字は出現頻度である（実際のシステムでは出現頻度は表示されない）。

「世界」

- 1) 地球上の全ての地域 (1306)
- 2) ある物事に関する特定の範囲 (280)
- 3) 人々が互いに関係しあいながら暮らす場 (74)
- 4) 芸術の創作物が作りあげている全体の場合 (70)
- 5) 共同の目的を持ち結合した集団 (68)
- 6) 仏教において、人間をふくむ全ての生き物が住む所 (40)
- 7) 宇宙全体 (8)

「気」

- 1) 気持ち (435)
- 2) 人の心の動き (109)
- 3) 心配という気持ち (107)
- 4) 関心 (37)
- 5) アルコールにおいて、香味 (23)

- 6) 人の性格 (15)
- 7) 気体という物体 (3)

この結果が示すように、頻度情報による並べ替えによって、『基本語用例辞典』に記された順に近い表示順が得られた。「世界」では 1), 2) が『基本語用例辞典』の 1 番目の意味にあたり、3), 4), 5) が『基本語用例辞典』の 2 番目の意味にあたる。また、『基本語用例辞典』に含まれない意味は 6), 7) に表示されている。「気」では 1)~4) が『基本語用例辞典』の 1 番目の意味にあたり、『基本語用例辞典』の 2, 3 番目の意味は辞書ツールで 5) 以降に表示されている。

「気」では、本来は 1)~4) の直後に表示させたい「人の性格」という意味が 6) に表示されるという問題が残った。しかし、基本的な意味を上位に置くという第一の目的は達成している。頻度情報による並べ替えによって、学習者は「世界」では「地球上の全ての地域」を、「気」では「気持ち」を主要な意味と捉えることができ、各単語の持つ意味概念の把握が容易に行えるようになったといえよう。

3.3 並べ替えの効果

頻度情報による並べ替えの効果を調査した。調査対象は、日本の通商産業省の「平成 10 年通商白書」の「2-3-1 地球環境問題」(4600 字)である²。この文章には助詞と助動詞以外の自立語は異なり語数で 402 語含まれている。その内、意味情報が 1 つしかないもの 130 語を除いた 272 語を対象に頻度情報による並べ替えの効果を調査した。

その結果、164 語 (60.3 %) において表示順序が改善された。本研究では、基本的用法に近い意味情報がより上位に繰上がった場合に改善したとみなした。この並べ替えによって意味情報の表示順序に問題が発生した単語はなかった。この表示順序の入れ替えの結果を『基本語用例辞典』に収録されている単語について調べたところ、後述する 13 語以外はすべて、『基本語用例辞典』で上位の意味が辞書ツールの表示でも上位に表示されていた。

また、表示順序の入れ替えが行われなかった単語に関しても、頻度による並べ替えそのものができなかったわけではない。入れ替えが行われなかった単語は意味情報が最大でも 4 つ以内のものであり、大半が 2 つないし 3 つしかないものであった (4 つあったものは 2 例のみ)。しかも、いずれの単語も意味情報の並び方自体に問題は見られなかった。これらの単語は、もともとの配列順序と頻度による順序とが一致していたため表示順序に変化が現れなかったものか、あるいは単語自体の出現頻度が低く、EDR 日本語単語辞書に頻度情報が登録されていないものと考えられる。

²この文章は通商産業省の Web ページから入手できる。<http://www.miti.go.jp/>

これらを総合して考えると、頻度情報を用いた並べ替えによって辞書ツールの概念表示は日本語学習者に適した形で改良が行われたといえる。

3.4 並べ替えの結果が『基本語用例辞典』と異なる単語の分析

頻度情報による並べ替えの結果が『基本語用例辞典』の意味記述の順序と異なった単語は次の13語であった。

動詞（8語） みる、くる、出る、出す、従う、上げる、取る、組む

形容詞（1語） 高い

名詞（4語） 間、月、力、中心

この13語では動詞の占める割合が高いが、これは基本動詞の場合、現実の使用では派生的意味での用法が多いことに起因している。特にひらがな表記の動詞は、ほとんどの場合が派生的意味で用いられている。例えば「みる」の意味情報を頻度情報によって並べ替えた結果は次のようになっている。

- 1) 推定する
- 2) 判断する
- 3) 様子を見る
- 4) （はっきりしない部分を）調べる
- 5) 物の形や姿を見る

このように「推定する」「判断する」等の派生的な意味が上位になっている。これに対して『基本語用例辞典』での「見る」の意味は、

1. 目を向ける。
2. 見物する。
3. 読む。
4. 調べる。
5. 調べる。
6. 医者が病気の人のからだを調べる。

と並び、「判断する」「推定する」の意味は6番目に出てくることがわかる。

また、上記の漢字表記の動詞でも派生的な意味が上位になっていた。例えば、「出る」の意味情報を頻度情報によって並べ替えた結果は次の通りであった。

- 1) 姿を現す
- 2) (ある場所の) 内から外へ行く
- 3) (新聞や出版物が) 発行される
- 4) 品物がよく売れる

これも以下の『基本語用例辞典』の順とは大きく異なっている。

1. 内から外にうつる.
2. 出発する.
3. すがたを見せる.
4. 発生する.
5. 売れる.

これは EDR コーパスにおいて、「出る」という単語が「内から外への移動」という意味よりもむしろ「姿を現す」の意味で用いられていることが多いことによるものである。

他の品詞においても同様に派生的な意味が上位に表示されている。例えば、「高い」を頻度によって意味情報を並べ替えた結果は次のような順となる。

- 1) 値が高いさま
- 2) 下からの空間的な長さが大きいさま
- 3) 地位・格式・教養が優れているさま
- 4) 広く知れ渡っている
- 5) 山が高くそびえるさま

一方、『基本語用例辞典』で「高い」を引くと、

1. 上から下までの長さが長いようす.
2. 身分・能力・程度などがすぐれているようす.
3. 声や音が大きい様子.
4. 買うのに多くの金がいるようす.

と記述されている。

「高い」という単語は、EDR コーパスでは本義の「上から下までの長さが長いようす」という意味より、むしろ「値段が高い」という意味で多く用いられていたわけである。これは実際に新聞・雑誌等でこの単語が用いられる状況を考えても違和感のないものである。

つまり、この結果は現実社会での各単語の用法をかなりの程度で反映していると思なすことができる。

また、学習者の日常生活でも「高い」を「値段が高い」の意味で用いることは少なくない。この例は、頻度情報という基準で行われる並べ替えの特徴をよく表している。

以上のように、頻度情報による並べ替えの結果が『基本語用例辞典』の意味の順序とは異なった13語は、いずれもその表示順序はそれなりに意味を持つものであり、日本語学習者にとって不要な意味、わかりにくい意味が上位に表示されてしまったものは全くなかった。

4 おわりに

読解学習支援システム「リーディング・チュウ太」の辞書ツールに関する意味情報の表示順序の改善について報告した。新しい辞書ツールでは、EDR 日英対訳辞書により得られる意味情報の順序を、EDR 日本語単語辞書に含まれる頻度情報を用いて並べ替えた。その結果、大半の単語の意味情報が『基本語用例辞典』に近い順序に並べ替えることができた。また、『基本語用例辞典』の意味の順序と異なる単語についても、現実社会での各単語の用いられ方をそれなりに反映したもので学習者にとって意味のあるものであった。

本システムは、インターネット上で公開可能な電子辞書が準備されれば英訳や日本語概念以外の形で意味情報も公開できる。韓国語を母語とする日本語学習者のためには、韓国語での意味情報の表示が求められるが、そのためにはインターネット上で公開可能な日韓辞書の整備が必要である。韓日両国の協力により早期に日韓辞書が整備されることを期待したい。

謝辞 本システムの開発には東京大学大学院理学系研究科の保原麗氏の協力を得た。日本電子化辞書研究所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いた。また、本研究の一部は（財）電気通信普及財団からの支援を受けた。ここに記して感謝の意を表す。

参考文献

- [1] Henstock, “Applications in Computer Aided Language Learning for the Student and Instructor”, M. A. Thesis. Purdue University (1995)
- [2] 寺, 北村, 落水, “日本語読解支援システム dictliker”, 日本語教育学会秋季大会予稿集 pp. 43-48 (1996)

- [3] Breen, “EDICT”, <ftp://ftp.cc.monash.edu.au/pub/nihongo/edict.doc>
- [4] 北村, 川村, 内山, 寺, 奥村, “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 (1999)
- [5] 北村, 小森, “マルチメディアデータを利用した読解支援システム”, CASTEL/J'99 pp. 173–184 (1999)
- [6] 小森, 上田, 難波, “オンラインデジタル動画学習支援システム”, 人文科学とコンピュータ 97-CH-36, Vol. 97, No. 108, pp. 49–54 (1997)
- [7] 越智, 矢野, 林, “電子化された日本語文書を教材とした漢字学習システム”, 教育工学関連学協会連合第5回全国大会予稿集 pp. 213–214 (1998)
- [8] 仁科, 奥村, “「やさしい科学技術日本語読解入門」—多言語対応オンライン科学技術日本語学習支援読解教材としての利用法—”, 日本語教育方法研究会誌, Vol. 7, No. 1, pp. 16–17 (2000)
- [9] 小島, 仁科, “日本語学習支援システムの運用 (1)”, 日本語教育方法研究会誌 Vol. 4, No. 2, pp. 10–11 (1997)
- [10] 川村, 北村, 保原, “EDR 電子化辞書を活用した日本語教育用辞書ツールの開発”, 日本教育工学会誌 (採録決定)
- [11] 川村, “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)
- [12] 川村, “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)
- [13] 川村, 北村, “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 (採録決定)
- [14] 日本電子化辞書研究所, “EDR 電子化辞書仕様説明書” (1996)
- [15] 松本, 北内, 山下, 平野, 松田, 浅原, “日本語形態素解析システム『茶釜』 version 2.0 使用説明書 第二版”, NAIST Technical Report NAIST-IS-TR99012 (1999)
- [16] 横山, 荻野, “言語データと言語処理”, bit Vol. 32, No. 2, pp. 12–20 (2000)
- [17] 文化庁, “外国人のための基本語用例辞典第二版” (1978)