

**日本語読解学習支援環境の再構築  
PROPOSAL FOR THREE NEW FEATURES  
IN THE JAPANESE READING TUTORIAL SYSTEM**

川村よし子, Yoshiko Kawamura, 東京国際大学  
前田ジョイス, Joyce Maeda, 東京国際大学  
宇津呂武仁, Takehito Utsuro, 筑波大学  
砂川有里子, Yuriko Sunakawa, 筑波大学  
土屋雅稔, Masatoshi Tsuchiya, 豊橋科学技術大学

**概要:** 本研究は、国際的、学際的な連携による日本語読解学習支援システム「Reading Tutor」の再構築をめざしている。辞書引きツールやレベル判定ツールを統合したこのサイトには、一日1000件以上のアクセスがあるが、よりよい学習環境の提供が求められている。現在、辞書ツールの多言語化、文型検索ツールの開発、親密度および頻度による語彙のレベル判定ツールの開発という3つのプロジェクトが進行中である。これらを、すべて同一のプラットフォームで使えるようにすることで、利用しやすい学習環境を提供することが可能になる。いずれのプロジェクトも、国際的、学際的な連携によって成り立っているものであり、ICT時代の学習環境構築の一つのあり方を示している。

**キーワード:** 日本語学習者 インターネット 辞書 文型 レベル判定

### 1. はじめに

1997年から開発を進めてきた日本語読解学習支援システム「リーディング・チュウ太(Reading Tutor)」は、日本語学習者がインターネット上の情報を用いて自由に読解学習を行うことができる自立学習支援システムである (<http://language.tiu.ac.jp>)。1999年10月に公開以来、日本語学習者のみならず多くの日本語教育関係者や研究者が利用している。このシステムには辞書ツール・レベル判定ツール・リンク集・読解教材バンク・文法クイズ等がモジュール化されて組み込まれ、有機的に統合された学習環境となっている(川村/北村2001)。辞書ツールとしては、すでに日日・日英・日独・日蘭の4種類の辞書が提供されているが、さらに多くの言語への対応が望まれている。また、文型や文法に関する情報もほしいという強い希望がある。一方、レベル判定ツールとしては、日本語能力試験出題基準(以下「出題基準」)に準拠した語彙チェッカーと漢字チェッカー(Kawamura, 2002)が用意されているが、単語の難易度判定に出題基準がどこまで有効かという点に疑問が残る。これら3種類の問題に対応するため、今回、国際的、学際的な連携を軸に日本語読解学習支援システム「Reading Tutor」の再構築をめざした取り組みを始めた。各々のプロジェクトについて以下に述べる。

### 2. 辞書ツール多言語化プロジェクト

リーディング・チュウ太の学習支援ツールのひとつである辞書ツールは、入力された文章を形態素解析し、文中の単語の辞書引き作業を一括して自動的に行うツールである(川村ほか, 2000)。形態素解析には奈良先端科学技術大学院大学松本研究室によって開発された「茶筌」(松本ほか, 2000)を用いている。現在、辞書としては日日・日英・日独・日蘭の4種類が提供されているが、さまざまな国の学習者からさらに多くの言語に対応してほしいという要望が寄せられている。そのため、2003年、辞書ツール多言語化プロジェクトがスタートし、多言語版日本語辞書編集システムが開発された(川村/前田ほか, 2006)。このシステムはインターネット上に置かれ、編集が終わった日日辞書をもとにして、世界各国の編集者が各国語版の対訳情報を入力するという形で、国際的な共同作業によって辞書の編集が進められている。この辞書編集システムについては、本大会においてワークショップを行う予定のため、詳細は別稿「多言語版日本語辞書編集システム」で述べることにする。

### 3. 文型検索ツールの開発

日本語学習者のための読解支援には、単語の読みや意味に関する情報だけでなく、文型に関する情報も不可欠である。ところが、現在の辞書ツールは、上述したように茶釜によって形態素解析した結果と辞書の見出し語とを照合して表示しているため、形態素の区切りより大きな語のまとまりの複合語、複合辞、文型等を表示することができない。また、文型には「絶対に～ない」や「～ば～ほど」のように、間にほかの語が入るものもある。そこで、こうした状況に対応可能な文型検索ツールを開発する必要がある。

土屋/宇津呂ほか (2005) は、文中の機能表現を自動的に検出するシステムを開発した。ここで言う機能表現とは、「いくつかの語が複合してひとまとまりの句となって付属語的な役割を果たしている語」のことであり、日本語教育で文型と呼ばれているものの大部分がこれにあたる。システムは「品詞体系に依存しない形式で記述された機能表現言い換えデータから自動的に機能語の検出規則と言い換え規則を生成し、生成された規則に基づいて検出と言い換えを行う」。機能表現の検出には、形態素列パターンに基づくパターンマッチングを用いている。このシステムを活用し、リーディング・チュウ太の他のツールと同一のプラットフォームで動くようにしたものが文型検索ツールである。図1はその入力画面である。



図1 文型検索ツールの入力画面

画面中央のボックスに文章を入力後、下の「文型」ボタンを押すと、文中に含まれている文型が自動で検出される。文章は、直接入力しても、教材等をコピー&ペーストしてもいい。図2が結果画面である。画面左が入力された本文で、網掛けされているのが機能語である。

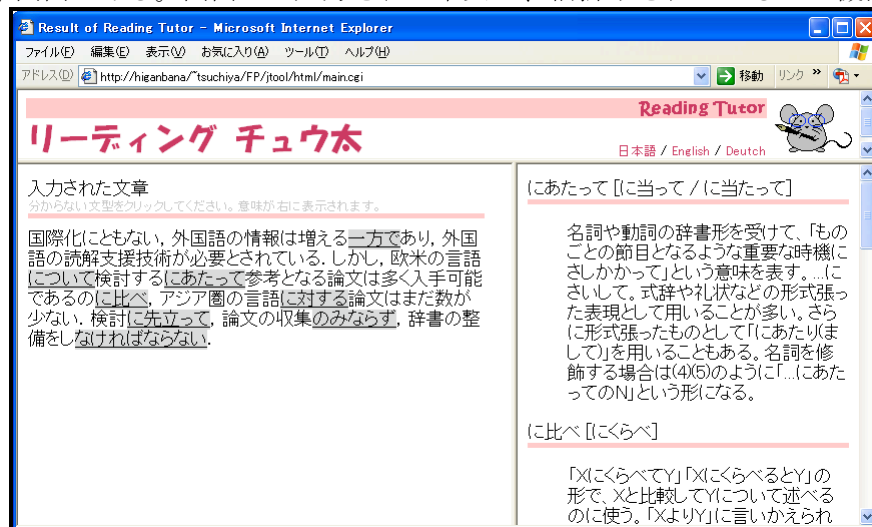


図2 文型検索ツールの結果画面

各々の機能語は画面右の辞書情報とリンクされ、各機能語をクリックすると該当する辞書項目が一番上に表示される仕組みになっている。辞書情報としては、『日本語文型辞典』（砂川ほか,1998）の辞書情報（意味や用法の説明および例文の一部）を用いている。ところが、機能語の中には、自立語と同じ形をとるものも少なくない。その場合、機能的用法か自立の用法かを判定する必要があるが、その判定精度は、現在のところ90%である。そのため、現行の文型検索ツールでは、自立語としての意味と機能語的用法としての辞書情報の双方を出す形にしている。また、機能表現の派生形の照合については、現在開発中である。

#### 4. 親密度判定ツールの開発

リーディング・チュウ太で提供されているもうひとつの読解支援ツールが、レベル判定ツールである。レベル判定ツールには、語彙チェッカーと漢字チェッカーがあり、入力された文章に含まれている単語や漢字の難易度を、日本語能力試験の出題基準に準拠して判定する。語彙チェッカーは辞書ツールと組み合わせて使うことによっていろいろな利用方法が考えられる。1999年の公開以来、日本語学習者ばかりでなく、多くの日本語教育関係者や研究者によって利用されている。また、公共機関や医療機関等、わかりやすい文章を書く必要のある機関でも活用されている。だが、いずれの場合においても、出題基準のみで難易度を判定していいのかという問題がある。一方、情報工学の分野では、難しい語彙を言い換えるための支援システムの開発が始まり、乾(2002)、鍛冶ほか(2002)、村山ほか(2003)等が文を平易に言い換える研究を進めている。だが、前2者では平易さの基準を明らかにしていない。また、後者では出題基準に準拠して平易度を決定してしまっている。いずれも、まず難易度の基準として何を用いるべきかに関する十分な吟味が必要である。

こうした背景から、川村(2006)では、難易度判定に『日本語の語彙特性』（天野ほか, 2000）の単語親密度が活用可能かどうかを調査した。その結果、平易な単語の選定には親密度が有効であること、また、日本語教育のための基本語選定においても、単語親密度を出題基準と併用する必要があることが判明した。この結果を元に、学際的共同研究によって、親密度を基準にした単語の難易度判定ツール「親密度チェッカー」の開発プロジェクトがスタートした。

親密度チェッカーは、入力された文に含まれるすべての単語の難易度を、親密度によってレベル判定するツールである。便宜上、ここでは単語親密度の6.3以上を難易度1、6.0以上を難易度2、以下0.5刻みでレベルを分け、5.0未満はすべて難易度5としている。

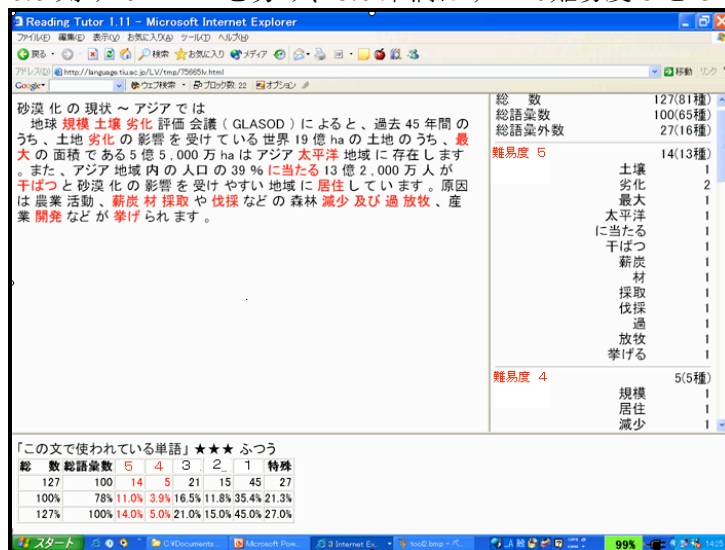


図3 親密度チェッカーの結果画面（図はイメージ）

図3は親密度チェッカーの結果画面である。画面左の本文の赤字部分は難易度4以上の単語

である。画面右は難易度別の単語リストで、本文の単語とリンクしている。単語の右の数字はその単語の本文中に表れた頻度である。また、画面下の表は文中に含まれる単語を親密度別に延べ語数でカウントしたものである。一番上の欄は実数、次が記号等の特殊文字を含めた場合の難易度ごとの単語の比率、下が特殊文字を除いた場合の比率である。

また、同様の仕様で単語の使用頻度情報を基にしたレベル判定ツールの開発も計画している。語彙チェッカーや漢字チェッカーに加えて、これらのツールを組み合わせる利用することによって、将来的には、単語の難易度判定の基準を決定し、それに従って単語の難易度判定を自動的に行えるシステムを完成することを目指している。

## 5. おわりに

以上、現在進行中の各プロジェクトについて概説してきたが、これらのツールをすべて同一のプラットフォームで使えるようにすることで、学習者にとって利用しやすい学習環境の構築が可能である。また、いずれのプロジェクトも国際的、学際的な連携があつてこそ、なりたつものである。今後も国内外の諸研究機関との連携によって、世界の日本語学習者によりよい学習環境を提供していく予定である。ICT時代の学習環境構築にはこうした連携が不可欠であり、世界各国の日本語教育関係諸機関からの協力を期待している。

## 謝辞

本研究は、科学研究費（課題番号 18320083）の研究助成を受けている。また、文型検索ツールおよび親密度チェッカーの開発にあたっては、それぞれ、くろしお出版、NTT コミュニケーション科学基礎研究所の天野成昭氏の許諾を得た。ここに記して感謝の意を表したい。

## 参考文献

- 天野成昭/近藤公久, NTT データベースシリーズ日本語の語彙特性(第1期), 三省堂, 2000.
- 乾健太郎, コミュニケーション支援のための言い換え, 言語処理学会第 8 回年次大会ワークショップ「言い換え/パラフレーズの自動化」, 71-76, 2002.
- 鍛冶伸裕/河原大輔/黒橋禎夫/佐藤理史, 国語辞典とコーパスを用いた用言の言い換え規則の学習, 言語処理学会第 8 回年次大会, 331-334, 2002.
- 川村よし子, 日本語学習者のための基本語選定の一試案, *Japanese Language Education in Europe, Vol.11*, 72-78, 2006.
- Kawamura, Yoshiko (2002). *Two New Tools for Analyzing Japanese Textbooks: Vocabulary and Kanji Level Checker*, in Nakajima, Kazuko(ed.). *Learning Japanese in the Network Society*, Calgary: University of Calgary Press, 71-88.
- 川村よし子/前田ジョイス/金庭久美子/植木正裕/川村ヒサオ/根津誠/保原麗/Hans Coppens /Jonathan Bunt/Kristina S. Hmeljak/Saleh Adel, 多言語版日本語辞書編集システムの開発と運用実験, *Japanese Language Education in Europe, Vol.10*, 146-151, 2006.
- 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, 第6号, 241-255, 2001.
- 川村よし子/北村達也/保原麗, 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, 日本語教育工学会論文誌, 24, 7-12, 2000.
- 国際交流基金・日本国際教育協会, 日本語能力試験出題基準【改訂版】, 凡人社, 2002.
- 砂川有里子/駒田聡/下田美津子/鈴木睦/筒井佐代/蓮沼昭子/ベケシュ・アンドレイ/森本順子, 日本語文型辞典, くろしお出版, 1998.
- 土屋雅稔/宇津呂武仁/佐藤理史/中川聖一, 形態素情報を用いた日本語機能表現の検出, 言語処理学会第 11 回年次大会 C3-1, 2005.
- 松本裕治/北内啓/山下達雄/平野善隆/松田寛/高岡一馬/浅原 正幸, 形態素解析システム『茶筌』 version 2.2.1 使用説明書, 2000, from <http://chasen.naist.jp/hiki/ChaSen/>.
- 村山賢洋/麻岡正洋/土屋雅稔/佐藤理史, 語彙の規格化とそれに基づく用言の言い換え支援, 言語処理学会第 9 回年次大会, 85-88, 2003.