

論文 / 著書情報
Article / Book Information

論題(和文)	構文構造，多言語，例文の表紙および読み上げ音声を実装した日本語学習支援システム - 「あすなる」の開発
Title(English)	A Call System for Japanese Language equipped with Interfaces of Multi-functions - Development of ASUNARO
著者(和文)	仁科 喜久子, 奥村学, 八木豊, 澤谷孝志, 戸次徳久, 杉本茂樹, 傅 亮, 阿辺川武
Authors(English)	KIKUKO NISHINA, Manabu Okumura, Yutaka Yagi, TAKASHI SAWATANI, shigeki sugimoto, TAKESHI ABEKAWA
出典(和文)	特定領域研究 (A) 「高等教育改革に資するマルチメディアの高度利用に関する研究」, Vol. , No. 120, pp. 81-84
Citation(English)	, Vol. , No. 120, pp. 81-84
発行日 / Pub. date	2001,

特定領域研究（A）

「高等教育改革に資するマルチメディアの高度利用に関する研究」

領域番号 120

領域略称名 メディア教育利用

研究成果報告書

平成13年度 計画研究

A01：教育マルチメディア技術の高度化とその効果に関する研究

A02：外国語教育の高度化の研究

A03：メディア教育・情報教育の高度化の研究

平成14年3月

研究代表者 坂元 昂

(メディア教育開発センター・所長)

構文構造、多言語、例文の表示および読み上げ音声を実装した

日本語学習支援システム「あすなろ」の開発—

A Call System for Japanese Language equipped with Interfaces of
Multi-functions –Development of ASUNARO–

東京工業大学 留学生センター
International Student Center,
Tokyo Institute of Technology
仁科 喜久子
Kikuko NISHINA

東京工業大学 精密工学研究所
The Precision and Intelligence Laboratory,
Tokyo Institute of Technology
奥村 学
Manabu OKUMURA

<研究協力者>

東京工業大学 情報理工学研究所
Graduate School of Information Science and Engineering, Graduate School of Decision Science and Technology,
Tokyo Institute of Technology
八木 豊
Yutaka YAGI

東京工業大学 社会理工学研究科
Tokyo Institute of Technology
澤谷 孝志 戸次 徳久
Takashi SAWAYA Norihisa TOTSUGI

フウズラボ (有)
Fu's Lab.
杉本 茂樹 傅 亮
Shigeki SUGIMOTO Fu Ryan

大日本印刷 (株)
Dai Nippon Printing Co., Ltd.
阿辺川 武
Takeshi ABEKAWA

A reading Japanese comprehension system with multilingual dictionaries called “ASUNARO” has been developed during a two year period. We focused on interfaces of syntactic structure and multilingual displays this year. We also equipped text reading voice and a concordancer on the web. We considered what prevents learners understanding of Japanese syntax, and examined what we have to do for effective learning.

Keywords : Syntactic structures Multi-language interface Concordancer Reading voice

1. はじめに

2001年3月からウェブ上で留学生のための多言語対応日本語読解支援システム「あすなろ」を <http://hinoki.ryu.titech.ac.jp> 上で一般公開し、現在も開発継続中である。本稿では昨年度に引き続き展開した多言語、構文、例文の表示、音声実装に関して述べる。

システムは、学習者が入力した日本語文章に対し、文章中の単語の訳と文章構造を出力することを主な機能とし、日本語の構造や初歩的な単語を知りたい日本語初級者から、単に技術用語の意味を知りたい上級者までが利用できるものである。現在の日本語学習システムおよび辞書の殆どは英語訳を導くものであるが、多言語で出力できることを目標とする。システムデザインと機能については昨年度の報告書で述べたので、参照されたい(文献3)。本報告では、上記の新規に導入した機能を中心に構文表示機能、多言語表示機能、例文

表示機能、読み上げ音声実装について述べる。

2. 構文表示機能

「あすなろ」システムの特徴の一つに構文表示機能が挙げられる。KNPによって解析された構文解析結果を様々な表現法で学習者に提示できる。文の意味理解のためには、各文節の係り受け関係、文節列集合の分割位置、文節や文節列の並列関係など、構文情報の把握が有効である。しかし、そのすべてを一度に表現でき、なおかつ学習者に容易に理解できる表示法は存在しない。そこで本システムでは、以下に示す4種類の構文表示機能を実装し、学習者に提示している。

(1)KNPのデフォルト出力表示

KNPは構文解析結果のデフォルトの出力とし、罫線記号を用いたテキスト表示を出力する(図1)。文節とその係り先の文節は、罫線により結ばれる。

この表示法は各文節間の係り受け関係を把握するには有効である一方、複文の場合は、どこで文が分割されるかなど文全体の構造は把握しにくいという点がある。この構文表示図では KNP によって解析される並列構造を正しく表現することができる。

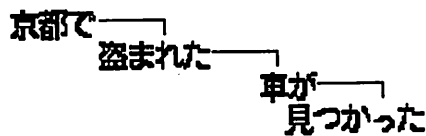


図1 KNPによる構文表示
(KNP デフォルト出力)

(2)木構造表示

文節間の関係を2分木による木構造で表現する(図2)。

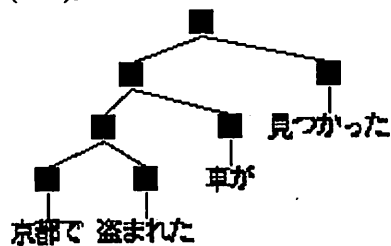


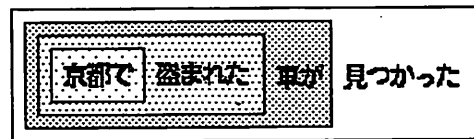
図2 木構造による構文表示

文節列間の関係が一目でわかるという特徴がある。しかし現在は2分木に限定しているため3つ以上の文節が並列関係を成すとき、それらを等位に並べることができない。また、一文が非常に多くの文節から構成される場合、木全体が大きくなり一画面に入りきらないという閲覧上の問題も存在する。

(3)入れ子ボックス表示

文節の係り受け関係を、係り元のボックスが係り先のボックスによって囲まれる入れ子構造で表現した構文表示法である(図3)。入れ子構造が何重にもなるため、学習者は表示する入れ子の階層を指定できる。木構造表示と同様、並列構造の表現に問題がある。

次の階層までを表示→1階層 2階層 3階層



次の階層までを表示→1階層 2階層 3階層

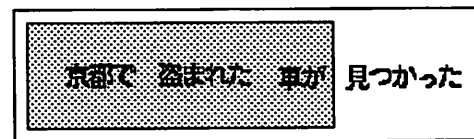


図3 入れ子型ボックスによる構文表示

(4)係り受け強調表示

係り受け関係の提示に特化した表示法である(図4)。

京都で 盗まれた 車が 見つかった



図4 係り受け強調表示

学習者が文節を選択するとその係り先の文節が強調表示される。前述の3つの表示法と異なり、文節列間の関係のような大きな関係表現することはできないが、表示法が簡便でかつ場所をとらないため、何行にも渡る通常の記事表示に埋め込むことが可能である。

以上4種の構文表示法を説明したが、いずれの表示法にも一長一短がある。また、日本語を母語とする人から見ても有用であろうと思われる表示法でも、日本語学習者にとっては、理解しにくい表示法があるかもしれない。

3. 多言語表示

日本語の意味を任意の外国語で表示することは、学習者が母国語を用いて日本語を理解するために求められる機能である。このためには、日本語と外国語を同時に一つの画面に表示する必要があるが、日本語を S-JIS/EUC/JIS の文字コードで表示した場合、異なる文字コードを持つ文字体

系(中国語簡体字、タイ語文字等)は、表示が困難となる。そこで、世界中の文字を統一コードとして扱う UNICODE を利用する。本システムでは、Web 画面表示や辞書データベースを UNICODE(UTF-8)で構成し、JUMAN や KNP の解析や結果の表示のときのみ EUC と UTF-8 との相互変換を行う処理を実装した。

日本語と外国語との変換は、日本語を EDR 辞書から概念番号に変換し、その概念番号から対応した外国語を得る。これにより、日本語-中国語や、日本語-タイ語等の各対日辞書を必要とすることなく、概念番号を含んだ各国語辞書があれば言語変換が可能であり、辞書作成作業や辞書スペースの効率化を図ることができる。

「技術」の情報

品詞：名詞

読み：ぎじゅつ

- ・ 技術
- ・ 技術, 技巧
- ・ 熟練的技术

図6 中国語簡体字表示の例

4. 考察—学習者に見やすい表示とは—

学習者にとってわかりにくい構文とは、どのようなものか、わかりにくい原因はなにか、その表示法として、どのような形式がもっとも効果的であるかの評価がこの時点で、問題になる。図で示した例文を取り上げる。

例文1 京都で盗まれた車が見つかった。

は、「京都で盗まれた車が（東京で）見つかった。」と盗まれた場所が京都という場合と、「京都で（東京で盗まれた車が）見つかった。」と盗まれたのは京都以外という場合が考えられる。実際にこの一文だけでは母語話者にも一意には決まらない曖昧文である。周辺のコンテキストにより様々な情報を集めることでいずれかの意味に決定していく。このような決定法において、母語話者と非母語話者の間にどのような違いがあるかを調査する方法を検討する必要がある。

例文2 神田からここに来るように頼まれたのできた。

この例も曖昧文であるが、「神田」が人名か地名かによって、文意が異なってくる。しかしながら、非母語話者で、「神田」が人名とも地名とも検討がつかない場合は、推量が非常に困難になる。

例文3 友達と一緒に調べた情報を新聞社に伝えた。

この例は主語省略文であるが、少数の留学生と日本人にアンケートを取ったところ、「伝えたのは誰か」という質問に対して「私」と答えるのに対して、留学生は「友達と私」あるいは「わからない」という答えが多く見られ、母語話者とは異なる傾向が見られた。

日本語においては、話者の人称を省略するのは普通であるが、非母語話者にはわかりにくいという推測が立つ。

このように常識や知識が欠ける場合でも、構文表示をみることで、文意が理解できることになれば、学習者には便利なツールとなる。現在の自然言語処理技術では、解決できない問題を多々含んでいるが、それでも可能な支援を探るのが本研究のねらいでもある。

5. 読み上げ音声の実装

「あすなろ」を実際に留学生が使用するようになり、読み上げ音声してほしいというコメントが多くあった。そこで、科学技術日本語カリキュラムにそったコースウェア版の12課分のテキスト文と語彙表の単語を日本語教師が読み上げたものをシステムに搭載し、テキスト中の文あるいは単語をクリックすると音声がかえるようにした。

6. コンコーダンスの実装

辞書の意味表示に関連して、例文を示すコンコーダンス (Concordancer) を導入した (文献2)。コンコーダンスは、聖書や古典文学の語句用法が本文中のコンテキストにそって見られる辞書として知られているが、そのコンコーダンスをコンピュータ上で機械的、自動的に表示するツールであ

る。日本語学習者が語の用法が理解しにくいときには、このシステムにより、学習レベルに適したテキストデータベースから取り出した例文を閲覧することで、理解を促進するという結果が出ており、今後本システムでの様々な利用法が考えられる。

7. むすび

以上今年度は、複数の機能を拡張することができた。

- 1) 構文表示については、さらに見やすい図について、予備実験から見通しが出て来たので、来年度は、実験を重ねて、その結果による、実装を予定している。
- 2) 多言語表示は UNICODE を利用することで WEB 上に 2 言語システム以上の文字コードを提示することが可能になった。今後は、中国語繁体字、韓国語などアジア近隣の学習者の文字を導入し、辞書機能として検索に耐える正解で十分な量を維持すること、そのために必要な辞書入力の人材を確保することも課題である。
- 3) 音声は、コースウェアの文章読み上げについての学習者の評価を行う必要があり、それにより今後の音声機能の展開を考えたい。
- 4) コンコーダンサは、読解支援として実装したが、むしろ作文支援に威力を発揮すると考えられる。読解支援の次にくる作文支援への展開を見据えなければならない。

<発表文献>

(1) 本年度の成果

1. 奥村学(2001)「自然言語処理と語学教育」 pp.9-13 専門日本語教育研究 VOL.3
2. 澤谷孝志・仁科喜久子・赤堀侃司(2001)「日本語学習者のための Web-Concordancer の開発」 pp.469-470 日本教育工学会講演論文集
3. 仁科喜久子・杉本茂樹・八木豊・阿辺川武・戸次徳久・傅亮(2001)「日本語 CALL システム開発—アジア圏理工系留学生のための多言語対応日本語学習システム研究—」 pp.53-54 特定領域研究 (A)「高等教育改革に資するマルチメディアの高度利用に関する研究」(研究代表者坂元昂) 研究成果報告書

平成 12 年度計画研究

4. 仁科喜久子・奥村学・八木豊・戸次徳久・澤谷孝志・杉本茂樹・傅亮・阿辺川武(2002)「構文表示と多言語インターフェースを備えた日本語読解学習支援システムの開発」言語処理学会論文集(印刷中ページ未定)
5. 村岡貴子・因京子・仁科喜久子・深尾百合子・加納千恵子(2001)「専門日本語教育の現状と将来の方向」PP.15-20 専門日本語教育研究 VOL.3
6. 仁科喜久子・杉本茂樹・八木豊・阿辺川武・戸次徳久・傅亮(2002)「多言語対応日本語読解支援システム「あすなろ」の開発—多言語および構文表示を中心に—」 p.30 ニュースレター「高等教育改革とマルチメディア」第 5 号

(2) その他参考文献

戸次徳久・仁科喜久子・奥村学・杉本茂樹・八木豊・阿辺川武 (2000)「構文解析を利用した構文木表示と助詞・助動詞相当句表示、および履歴利用の検討」 pp22-29 専門日本語教育研究 VOL.3

- (3) 特記事項：平成 14 年 2 月 12 日 読売新聞夕刊「マルチメディア」欄に「あすなろ」開発の記事が掲載される。