

# 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

仁科喜久子<sup>1</sup> 奥村学<sup>2</sup> 八木豊<sup>3</sup> 戸次徳久<sup>4</sup> 澤谷孝志<sup>4</sup> 傅亮<sup>5</sup> 杉本茂樹<sup>5</sup> 阿辺川武<sup>6</sup>

東京工業大学留学生センター<sup>1</sup> 東京工業大学精密工学研究所<sup>2</sup> 東京工業大学情報理工学研究所<sup>3</sup>  
東京工業大学社会理工学研究所<sup>4</sup> フウズラボ<sup>5</sup> 大日本印刷株式会社<sup>6</sup>

## 1 はじめに

2001年3月から、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」をウェブ上で一般公開している<sup>1</sup>。「あすなろ」開発の主たる目的は、日本に留学している特に理工系分野を専門としている日本語学習者の専門文献読解を支援することである。

日本の大学で学ぶ留学生は論文読解も論文作成も英語である場合が多いが、研究室でのコミュニケーションは日本語が基本である。その中で研究に多くの時間が割かれ、日本語学習に時間を割くことは難しい。また、日本に留学する学生の8割は非英語圏国からであるが、日本語学習のための解説言語はほとんど英語である。非英語圏の留学生が目的の科学技術を学ぶ際、母語⇔英語⇔日本語というように、母語ではない言語を通して学習目標言語を正しく理解するには障害が多い。このような背景から、留学生の多くが専門分野の基本的な概念知識を母語で持っていることを利用して、母語から日本語が学べる読解支援システム「あすなろ」の開発をはじめた。

「あすなろ」は、学習者が入力した日本語の文章に対して、文章中の単語の訳語と構文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の意味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。また、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、英語以外にも中国語、タイ語、マレー語、インドネシア語の4ヶ国語による訳の表示が可能である。構文表示には、KNPによるデフォルト出力、一般的な木構造表示、係り受けを強調した入れ子構造、階層を変えることのできる入れ子構造という4種類の表示方法を用意しており、そのうちのどの表示方法が学習を促進するかについて検討中である。

<sup>1</sup><http://hinoki.ryu.titech.ac.jp>

本論文では、まず2節でこれまでの「あすなろ」の機能について概説する。続いて、3節で多言語表示、4節で構文表示についてそれぞれ説明し、5節で構文表示に対する考察を行う。最後に、6節でまとめを行い、今後の課題について述べる。

## 2 「あすなろ」の概略

### 2.1 システムデザインと構築

学習者はインターネットを利用して、学習したい教材を教材データベースから選ぶか、学習者自身が調べたい論文などを持ち込み、文章をシステムに入力する。システムに読み込まれた文章は、形態素解析JUMAN[3]にかけられる。プログラム処理により解析された形態素が、学習者が選択した言語の辞書データベース中の一つ対訳辞書に行き、単語項目を問い合わせる。もしも、参照したい語の単語項目があれば、単語の意味が学習者の母国語で表示される。同時に、入力した文章中の任意の一文について、KNP[4]により構文解析を行う。

上述のシステムを構成するためには、システムサーバは形態素/構文解析プログラムを実行したのち、要素の対訳を辞書データベースから検索して応答する必要がある。解析プログラムの実行と、データベースの検索はどちらも計算機の負荷が大きく、アクセスが同時に発生するような環境下においてこれらを1台のサーバで行うことには無理がある。そこで、解析プログラムを実行するサーバとデータベース問い合わせを行うデータベースサーバを分け、サーバがデータベースサーバに対訳問い合わせを行うようにシステムを構築する。すなわち、窓口のWWWサーバは解析プログラムを実行し、訳語問い合わせを他のサーバに依存して応答する。このようなシステムを構成する利点は、将来アクセスが膨大になったときにも窓口WWWサーバを

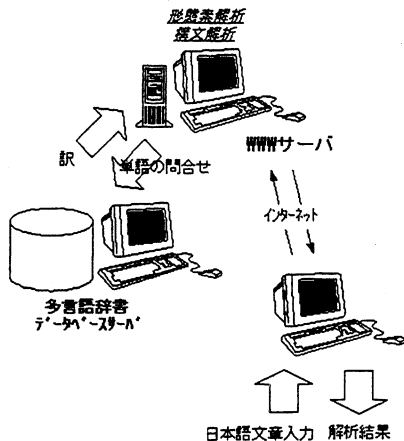


図 1: システム構成

増やすことで対応できると同時に、辞書データベースの内容更新が容易に行えることにある。WWWサーバ、データベースサーバとも、一般のDOS/V 計算機を用い、コストパフォーマンスの良いLinux ベースのOSを採用した。

### 3 多言語表示

辞書データベースは、日本語・英語・中国語・タイ語・インドネシア語・マレー語の各対訳辞書と、概念辞書、日本語構文辞書などの膨大なデータをリレーション付で格納するため、商用リレーショナルデータベースを採用した。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、違う文字コードを持つ文字(中国語簡体字、タイ語、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  
• 熟練的技術

#### 「技術」の情報

品詞：名詞  
読み：ぎじゅつ  
-----  
• เทคโนโลยี

#### 「技术」の情報

品詞：名詞  
読み：ぎじゅつ  
-----  
• teknologi

図 2: 中国語・タイ語・マレー語の表示画面

### 4 構文表示

「あすなる」の特徴の一つに構文表示機能が挙げられる。これは、KNPによって解析された構文解析結果を様々な表現方法で学習者に提示する機能である。

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、なおかつ学習者が容易に理解できるように提示することは難しい。そこで、「あすなる」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て、罫線記号を用いたテキスト表示を出力する(図3)。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、複文の場合には、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。

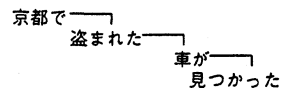


図 3: KNP 表示

#### 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4)。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画面に入りきらないという閲覧上の問題も存在する。

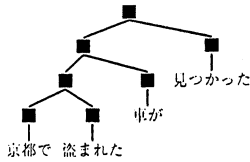


図 4: 木構造表示

### 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5)。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マウスマウスカーソルを合わせることによって指定できる。木構造表示と同様、並列構造の表現が表現しにくく、現時点では表示できない。

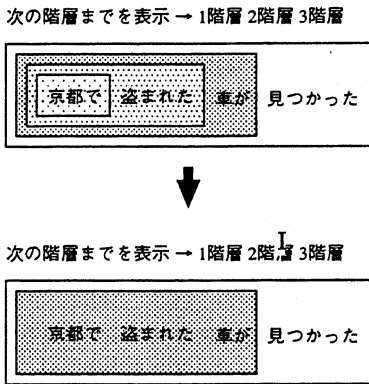


図 5: 入れ子ボックス表示

### 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6)。学習者がマウスマウスカーソルで文節を選択するとその係り元の文節列と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を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文章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

京都で 盗まれた 車が 見つかった

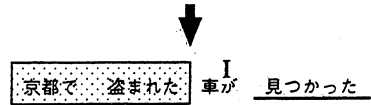


図 6: 係り受け強調表示

から見て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。

## 5 考察-学習者に見やすい表示とは-

ここで学習者すなわち非母語話者にとって分かりにくい文とはどのようなものか考える。

例1 私は大声で宮崎に行かないでくれと言いました。

例2 川口からここに来るように頼まれたので来た。

例3 友達と一緒に調べた情報を新聞社に伝えた。

例4 父親が銀行で働いている息子を探した。

例5 けさ学生の辞書がなくなったことを聞いた。

例6 歌の大好きな友達のおじさんはがんこで有名だ。

例7 電話で話していたかわいい女の子と男の子がいる。

例1~例7は曖昧文である。例1では、<「宮崎(場所)に行かないでくれ」といいました。>と<宮崎(人)に「行かないでくれ」といいました。>の解釈の可能性がある。例2では、<「川口(人)から「ここまで来るように」頼まれました。>と、<「川口(場所)からここまで来るように」頼まれました。>のように名詞句とそれを受ける述語句の係り受けによって、解釈が異なる。通常、我々はある文が存在しているディスコース中から世界知識つまり常識を活用させながら、この曖昧性を解消しようとする。「宮崎」「川口」が地名にも人名にも使われる固有名詞と分かると、この判断に迷うことになるが、文周辺の様々な情報を駆使して、どちらかに決めようとする。日本語学習者の場合は、駆使できる判断情報が少ないため、日本人母語話者に比べてさらに決定がむずかしくなる。もし構文表示が理想的に実現できれば、このような判断過程で不足する情報を支援するものとなるはずである。

次に、例3は、<(だれかが)友達と一緒に調べた情報を(だれかと友達がいっしょに)新聞社に伝えた。>

や<(だれかが=私が)友達と調べた情報を(私が一人で)新聞社に伝えた.>などの複数の解釈ができる。コンテキストからの判断が難しい場合、常識がはたらく。留学生の場合、文頭に「<〇〇>が」とあると、文末述語がと呼応すると短絡的に考える傾向がみられる。例4は、<(父親が銀行で働いている)息子を(だれかが)探している.>と<父親が(銀行で働いている)息子を探している.>の2つの解釈が考えられる。日本語母語話者の場合、「△△(私)は、〇〇が～」と文頭に文主語を補って考えることも試みるであろう。一方、日本語学習者の場合は格助詞「が」があると文主語だと考える傾向がある。曖昧性解消の方略として隠れた文主語を探そうとする日本語母語話者と異なるなど、構文構造に対する認識習慣の違いがあると考えられるが、曖昧性解消手続きの上ではかえって混乱を来たすことになる。例5は、「けさ～なくなった」か「けさ～聞いた」かこの文だけからは分からない。この文は、日本語母語話者にとっても判断しにくい。「学生の辞書がなくなったことをけさ聞いた」と語順が変われば曖昧性は解消できる。現実には必ずしも注意深く書かれていない文も読解いしなければならず、日本語学習者にも母語話者にも同じように判断しにくい文が存在している。

例6、例7は並列修飾の例である。例6は「歌が好きなのはおじいさん」である場合と、「歌が好きなのは友達」の場合がある。例7は「かわいい男の子とかわいい女の子」の場合と「かわいい女の子と(ふつうの)男の子」の場合がある。先に現システムではこの提示法ができていないことを述べた。この種のものも、日本語母語話者にもわかりにくいものである。これらの場合、システムによって正確な構文表示が与えられれば、留学生のハンディキャップを支援することになるが、現段階ではディスコース情報を扱うのはむずかしい。しかしながら、このような構文情報提供の方向性は、読解支援の一助になるとと思われる。

## 6 まとめ

多言語表示は UNICODE を利用することでウェブ上に2言語以上の文字コードを提示することが可能となった。今後は、中国語繁体字、韓国語などアジア近隣の学習者の母語文字を導入し、辞書機能としての検索に耐えるものにしなければならない。構文表示機能については、日本語母語話者とは異なる学習者の文法認識過程が考えられた。これらの予測を実証するため

に、近い将来、学習者の構文認識に関する実験を行う予定である。今回言及しなかったが、この他に例文提示システム(コンコーダンサ)も開発中である[2]。このシステムは、語の用法を検索するもので、ウェブ上の文章から検索できるものであり、意味用法を具体的に知ると同時に、作文にも強力な支援をされると考えられる。自然言語処理技術の応用として、日本語学習支援システムのための工夫は発展的に考えられる。読解支援の他に作文支援、会話支援など様々なツールの拡張も見据えて今後も改良を重ねることを課題とする。

## 参考文献

- [1] 奥村学. 自然言語処理と語学教育. 専門日本語教育研究, Vol. 3, pp. 9-13, 2002.
- [2] 澤谷孝志, 仁科喜久子, 赤堀侃司. 日本語学習者のための Web-Concordancer の開発. 日本教育工学会第17回大会講演論文集, pp. 469-470, 2001.
- [3] 黒橋禎夫. 日本語形態素解析システム JUMAN version 3.61 使用説明書. 京都大学大学院 情報学研究所, 1998.
- [4] 黒橋禎夫. 日本語構文解析システム KNP version 2.0b6 使用説明書. 京都大学大学院 情報学研究所, 1998.
- [5] 戸次徳久, 仁科喜久子, 奥村学, 杉本茂樹, 八木豊, 阿辺川武. 構文解析を利用した構文表示と助詞・助動詞相当句表示, および履歴利用の検討. 専門日本語教育研究, Vol. 2, No. 2, pp. 22-29, 2000.

なお、本研究は科学研究費補助金特定領域研究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 研究分担者 仁科喜久子・奥村学)、同基盤研究B2「アジア圏理工系留学生のための多言語対応辞書を装備した日本語学習システム開発」(研究代表者仁科喜久子)、同基盤研究C「概念知識に基づく母語からの理工系専門日本語習得方法に関する研究」によって行われた。