

論文 / 著書情報
Article / Book Information

論題(和文)	多言語表示日本語読解学習支援システム「あすなる」の開発
Title(English)	Development of Japanese Reading System of Multi-Lingual Environment Asunaro
著者(和文)	阿辺川武, 八木豊, 澤谷孝志, 戸次徳久, 奥村学, 仁科喜久子, 杉本茂樹
Authors(English)	TAKESHI ABEKAWA, Yutaka Yagi, TAKASHI SAWATANI, Manabu Okumura, KIKUKO NISHINA, shigeki sugimoto
出典(和文)	第3回「日本語教育とコンピュータ」国際会議, Vol. , No. , pp. 71-74
Citation(English)	Castel/J 2002 Proceedings-The Third International Conference on Computer Assisted Systems For Teaching & Learning Japanese Computer Technology and Japanese Language Education, Vol. , No. , pp. 71-74
発行日 / Pub. date	2002, 7

多言語表示日本語読解学習支援システム「あすなろ」の開発

Development of Japanese Reading System of Multi-Lingual Environment "Asunaro"

阿辺川 武、八木 豊、澤谷 孝志、戸次 徳久、奥村 学、仁科 喜久子(東京工業大学)
杉本 茂樹、傅 亮(フウズラボ(有))

ABEKAWA Takeshi, YAGI Yutaka, SAWAYA Takashi, TOTSUGI Norihisa, OKUMURA Manabu,
NISHINA Kikuko (Tokyo Institute of Technology), SUGIMOTO Shigeki, Fu Ryan (Fu's Lab)

A Japanese reading comprehension system with multilingual dictionaries called "ASUNARO" has been developed during a two-year period. This paper focuses on interfaces of syntactic structure and multilingual displays. We examined what prevents learners' understanding of Japanese syntax, and considered methods for more effective learning. We examined the relationship of learners' understanding between syntactic structure display and their meaning comprehension.

We showed four kinds of syntactic displays to 43 testees in order to verify which display is the most understandable for them. Finally, we concluded that a tree structure is the best performance among the four displays.

Keywords: 構文表示 読解システム 学習支援 多言語 Web

1. はじめに

著者らは、日本語読解学習支援システム「あすなろ」の開発を続け、システムは、<http://hinoki.ryu.titech.ac.jp/>上で既に稼働している。本研究の最終目標は、以下に挙げる3項目である。(1)理工系留学生のためにWeb上で学習可能な科学技術日本語読解学習支援を多言語対応システムとして開発する。英語圏以外の学習者でも母語による支援により文章理解できることを目指す。(2)細分化された専門分野別、学習者の日本語能力別の学習を可能にする。一斉授業で個々の学習者が満足できる専門読解を目指すことはむずかしいが、Web上では、個別に学習者に最適な内容を選択でき、学習レベルに合わせた時間配分も可能となる。(3)自然言語処理、日本語学、第二言語習得理論(外国語学習理論)、教育学などの学際的視点から新たに各分野に新しい知見を加える。

システムの主な機能は、学習者が入力した日本語の文章に対し、文章中の単語の訳と文ごとの構文構造を出力することである。その際、Web画面表示や辞書データベースをUNICODEで構成することにより、日本語、英語、マレー語、インドネシア語の他、中国語、タイ語等の特殊な文字を含めた多言語表示ができる。

以降、2章で機能の1つである構文表示機能について説明し、3章では、どの構文表示機能が学習者に効果的であるかを調べる実験について説明する。最後に4章で実験についてのまとめを行う。

2. 構文表示機能

システムには、文ごとの構文構造を表示する機能があり、その目的は、文の理解を促進させ、読解力を高めることである。

学習者の入力した文章に対して、まず、文を形態素に分ち書きする形態素解析を行い、次に、文節の係り受けを決定する構文解析を行う。それぞれの解析には、形態素解析プログラム JUMAN、構文解析プログラム KNP を使用している。そして、その解析結果を表示するために4種類の構文構造表示機能を実装している。以下それぞれの表示機能について説明する。

a) KNP 出力表示

KNP は構文解析結果を野線記号によるテキスト表示で出力する(図2.1)。ひとつの文節とその係り先の文節は、野線により結ばれる。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が、文全体の構造を把握しにくいという点がある。

b) 木構造表示

文節間の関係を木構造で表現する(図2.2)。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、1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問題も存在する。

c) 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2.3)。入れ子構造が何重にもなり複雑にはなるが、学習者はマウスマウスカーソルを該当する入れ子の階層上に合わせることで構造を知ることができる。これも木構造表示と同様、並列をなす構造の表現をすることは難しい。

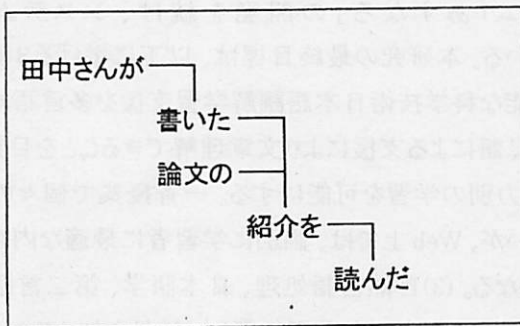


図 2.1 KNP 出力表示

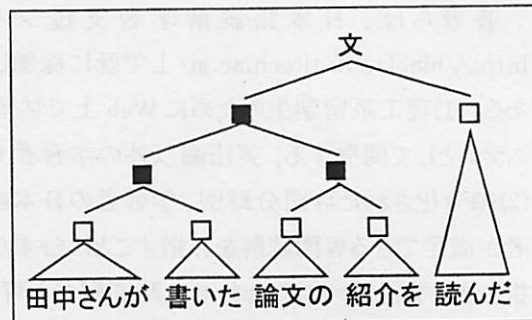


図 2.2 木構造表示

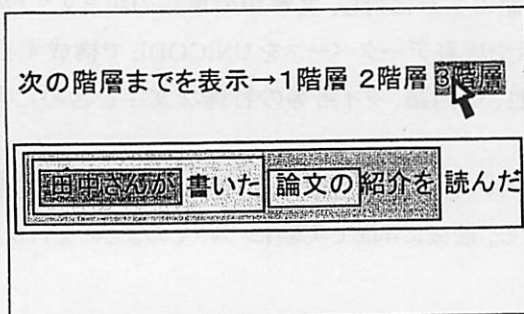


図 2.3 入れ子ボックス表示

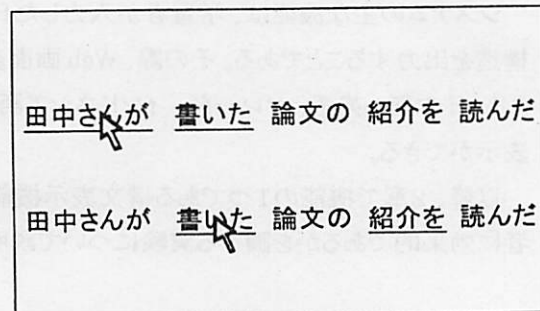


図 2.4 係り受け強調表示

d) 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2.4)。学習者が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。

3. 実験・考察

学習者にとって効果的な構文表示方法とはどのようなものかを調べるために、実際に本システムを活用する立場にある日本語学習者を対象に実験を行った。実験の概要は、最初に2通りの解釈ができる文を用意し、一方が正解の解釈となるような構文構造を作成する。各問題に対して4種の表示法のうちの1つを見せ、1分以内に2つの選択肢から選ぶというものである。

問題文として次の5つのタイプを用意した(問題例は表1を参照)。

(I) 文頭の連体修飾節の係り先が「名詞と名詞」のような並列構造となり、2通りの解釈ができる文

(II)(III) 文頭の連体修飾節の係り先が「名詞の名詞」となり、2通りの解釈ができる文

(IV) 文頭の名詞が連体修飾節内の動詞に係るか、文末の動詞に係るかの2通りの解釈ができる文

(V) 文頭の副詞が連体修飾節内の動詞に係るか、文末の動詞に係るかの2通りの解釈ができる文

(II)(III)は同じタイプであるが、(II)は遠い方に係り、(III)は近い方に係ると解釈できる問題とした。各タイプに対して語彙の異なる問題を4問ずつ作成し、合計20問を用意した。問題のタイプと4種の構文表示法の組み合わせが均等に配分されるように、最初の被験者は問題1をKNP出力表示、問題2を木構造表示、次の被験者は問題1を木構造表示、問題2を入れ子ボックス表示というように、構文表示法を順番に入れ替えて出題した。

実験は、東京工業大学に所属する日本語能力検定試験で1級ないし2級に合格者した留学生を対象に行った。43人がこの実験に参加し、それぞれの表示法に約10人ずつのデータが集まった。その結果を表2に示す。

全体の正解率は、木構造表示が一番高く、KNP出力表示が一番低いという結果となった。統計的検定は行っていないが、両者間にはある程度の差が認められることが明らかになった。KNP出力表示の正解率が低い原因は、タイプ(I)のように係り先が並列構造のとき単独の文節に係るか並列全体に係るかわかりにくいこと、そして図2.1の「書いた」と「論文」のように前の文節と次の文節の位置が上下で重なっているとき、双方の文節が係り受け関係を持っていると誤判断してしまうことである。タイプ(IV)で

表1 問題タイプと問題の例

タイプ	問題文	質問	選択肢		正解
I	昨日買ったケーキとパンを食べた。	昨日買ったのはどちらですか？	1. ケーキ	2. ケーキとパン	1
II	田中さんが書いた論文の紹介を読んだ。	田中さんが書いたのはどちらですか？	1. 論文	2. 論文の紹介	2
III	新しくできた公園のベンチで本を読んだ。	新しくできたのはどちらですか？	1. 公園	2. 公園のベンチ	1
IV	社長が有名になった会社へ行く。	有名になったのはどちらですか？	1. 社長	2. 会社	2
V	京都で盗まれた車が見つかりました。	京都でどちらが起きましたか？	1. 車が盗まれた	2. 車が見つかった	2

表2 構文表示法別正解率

	タイプⅠ	タイプⅡ	タイプⅢ	タイプⅣ	タイプⅤ	全体
KNP 出力表示	37.2%	51.2%	76.7%	93.0%	86.0%	68.8%
木構造表示	72.1%	74.4%	67.4%	90.7%	67.4%	74.4%
入れ子ボックス表示	69.8%	62.8%	79.1%	83.7%	58.1%	70.7%
係り受け強調表示	65.1%	55.8%	76.7%	95.3%	74.4%	73.5%
全体	61.0%	61.0%	75.0%	90.7%	71.5%	

は、正解の解釈が被験者の直感と一致していることから高い正解率となった。正解を逆にし、主格が存在しない解釈とした場合、構文表示法によって正解率の差が生じたかもしれない。

入れ子ボックス表示は、一見するとカラフルで綺麗に見えることから被験者に人気があった。実験後のアンケートでも一番わかりやすかったという意見も多い。しかし彼らの実際の正解率を調べるとあまりよくない。このことから被験者の実感と正解率に差があるということは、その表示法の見方を誤解しているのではないかと思われる。

4. まとめ

日本語読解システム「あすなろ」の紹介、およびシステムに実装されている4種の構文表示法の有効性を評価する実験を行った。実験の結果、構文表示法間には、ある程度の正解率の差異が見られた。また、正解率の低い問題を詳細に調査した結果、構文表示法の新たな問題点が露呈し、修正の必要性が生じた。各被験者別に見ると全問正解者が数名おり、彼らの実際の日本語レベルは様々であった。このことから、日本語の習熟度に関わらず、グラフィカルに構文構造を表現することは、文の理解の助けとなることが明らかになった。今後も学習者の文章理解には、どのような情報が必要で、どのような見せ方が効果的であるかを継続して調査する予定である。

謝辞

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等利用に関する研究」(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。

参考文献

- 仁科喜久子、奥村学、八木豊、戸次徳久、澤谷孝志、傳亮、杉本茂樹、阿辺川武、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、言語処理学会第8回年次大会発表論文集, 228-231, 2002.
- 黒橋禎夫、日本語形態素解析システム JUMAN Version3.61 使用説明書、京都大学大学院 情報学研究科, 1998.
- 黒橋禎夫、けっこうやるな KNP、情報処理学会誌, Vol.41, No.11, 2000.