

3T6-2 日本語学習システム「あすなろ」開発の新しい展開 - 構文学習とその評価 -

阿辺川 武[†] 八木 豊[†] 戸次 徳久[†] 澤谷 孝志[†] 奥村 学[†] 仁科 喜久子[†] 杉本
茂樹^{††} 傅 亮^{††}
[†] 東京工業大学 ^{††} フーズラボ(有)

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、インターネット上で公開してきた*。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)、教育工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ことである。

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、文章中の単語の訳や文法機能語の意味を表示することである。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。

Webで利用できる同様な読解支援システムとして“リーリング・チュウ太[†]”や“理解.com[‡]”などが存在するが、本システムの特徴の1つとして入力された日本語文に対し、文節ごとの係り受け関係を表示する機能が挙げられる。

以降では、2節で係り受け表示機能を中心に本システムの概観を述べ、3節で実際の利用者による係り受

け表示機能の理解度の評価を行う。4節で本システムの今後の展開を述べ、最後に5節においてまとめを行う。

2 システムの概観

2.1 意味・読み

本システムでは、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、及び単語の読みを画面に表示する。学習者が母語で単語の意味が理解できるように、表示する訳の言語を選択できるようになっている。日本語、英語のほか、マレー語、インドネシア語、中国語、タイ語の対訳辞書が用意されており、現在語彙を拡充中である。表示の際、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、多言語表示ができる。図1にタイ語による訳の表示例を示す。

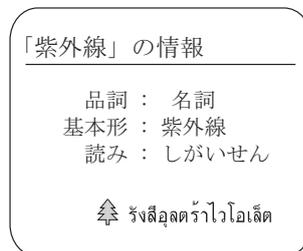


図1: タイ語表示の例

2.2 文法説明機能

学習者の入力した文章に、助動詞相当句や呼応表現のような文法項目があれば、自動的に抽出し、説明・提示する機能である。抽出する文法項目は、助詞相当句“～として”、助動詞相当句“なければならない”、呼応表現“決して～ない”、慣用的表現“～といい、～といい”などである。なお本システムでは、主として日本語能力試験1,2級の文法項目を収集し、抽出することを目標としている。

提示する内容は、意味、使用上の注意点、そして日本語能力試験の過去に出題された文を中心とした例文である。現在は、中国語による説明、例文が約100の文法項目に対して用意されている。図2に文法説明機能の例を載せる。

Development of Japanese learning system “ASUNARO” and new movement –Learning syntax and the evaluation–

[†] ABEKAWA Takeshi (abekawa@lr.pi.titech.ac.jp)

[†] YAGI Yutaka (yutaka@cl.cs.titech.ac.jp)

[†] TOTHUGI Norihisa (tothugi@ryu.titech.ac.jp)

[†] SAWAYA Takashi (sawaya@ryu.titech.ac.jp)

[†] OKUMURA Manabu (oku@pi.titech.ac.jp)

[†] NISHINA Kikuko (knishina@ryu.titech.ac.jp)

^{††} SUGIMOTO Shigeki (sige@whoselab.com)

^{††} Fu Ryan (fu@whoselab.com)

Tokyo Tech ([†])

Fu's Lab, Inc. (^{††})

* <http://hinoki.ryu.titech.ac.jp/>

[†] <http://language.tiu.ac.jp/>

[‡] <http://rikai.com/>

形式

～はおろか

説明

体言。表示A所代表的事項自不必说, 当然……, 后项B又对A进行更进一步的补充、说明, 意为更甚。相当于汉语的“别说A了, 连B都……”、“岂止……连……”。

例文

海外旅行はおろか、国内さえもほとんど見て回ったことがない。(97年1級)

図 2: 文法項目説明の例

2.3 音声

本システムでは、東京工業大学留学生センターで実際に講義で使用されている日本語テキストを掲載しており、学習者は自由に読むことができる。それぞれの日本語文に対して、日本語教師によるテキストの朗読音声を用意されている。またテキスト中の自立語に対してもその読みが音声で収録されている。

2.4 係り受け構造

本システムには、文節ごとの係り受け構造を表示する機能があり、その目的は、学習者に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、文を形態素に分か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[1]、構文解析プログラム KNP[2] を使用している。そして、構文解析結果を表示するために4種類の係り受け構造表示機能を実装している。以下それぞれの表示法について説明する。

2.4.1 KNP 出力表示

KNP は構文解析結果を罫線記号によるテキスト表示で出力する(図3)。1つの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が、文全体の構造を把握しにくいという点がある。

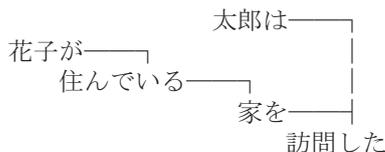


図 3: KNP 出力表示

2.4.2 木構造表示

文節間の関係を木構造で表現する(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、1文が非常に多くの文節から構成されるとき、木構造全体が複雑になり、全体を把握しづらくなる場合がある。その場合、図4下部のように節点をマウスでクリックすることにより、文節をまとめあげ木構造を簡単化することができる。

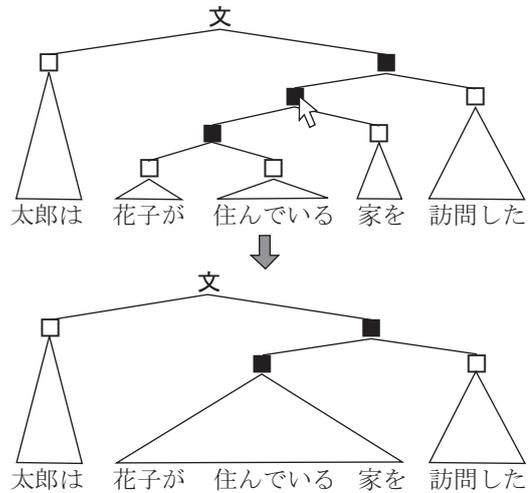
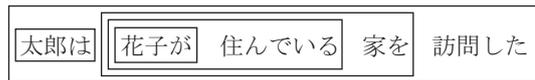


図 4: 木構造表示

2.4.3 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係り受け表示法である。入れ子構造が何重にもなり複雑にはなるが、学習者は図5下部のように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より大まかな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。

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

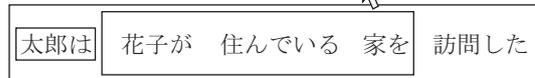


図 5: 入れ子 BOX 表示

2.4.4 係り受け強調表示

2 文節間の係り受け関係の提示に特化した表示法である(図 6)。学習者がマウスカーソルで文節を選択するとその係り元の文節列がボックス表示され、係り先の文節が下線表示される。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の表示と並列して埋め込むことが可能である。



図 6: 係り受け強調表示

3 係り受け構造表示の評価

3.1 実験方法

前節で 4 種類の係り受け構造表示法を説明したが、学習者にとって効果的な係り受け構造表示方法とはどのようなものであろうか。この疑問を解決するために、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係り受け構造を作成する。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形式である。

問題文として次の 5 つのタイプを用意した(問題例は図 7 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) 文頭の副詞節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、正解の選択肢を (II) では遠い方に係り、(III) で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 種の係り受け表示法の組み合わせが均等に配分され

タイプ(I)

問題文: 古くなった時計とカメラを持って行った。

質問: 古くなったのはどちらですか?

選択肢: 1. 時計 2. 時計とカメラ

正解: 1.

タイプ(II)(III)

問題文: 田中さんが書いた論文の紹介を読んだ。

質問: 田中さんが書いたのはどちらですか?

選択肢: 1. 論文 2. 論文の紹介

正解: 2. (タイプII) 1. (タイプIII)

タイプ(IV)

問題文: 社長が有名になった会社へ行く。

質問: 有名になったのはどちらですか?

選択肢: 1. 社長 2. 会社

正解: 2.

タイプ(V)

問題文: 京都で盗まれた車が見つかりました。

質問: 京都でどちらが起きましたか?

選択肢: 1. 車が盗まれた 2. 車が見つかった

正解: 2.

図 7: 問題例(正解の選択肢になるような係り受け構造を被験者に提示する)

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、係り受け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留学生で、日本語能力検定試験で 1 級ないし 2 級に合格者した者を対象に行った。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。

3.2 結果

係り受け表示法別、問題タイプ別の正答率の平均値を表 1 に示す。各平均値は、正答率上位と下位の被験者をそれぞれ 3 人ずつ省いて求めたものである[§]。

3.2.1 係り受け表示法別

係り受け表示法別の正答率は、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。両表示法を有意水準 5% のもとで t 検定を行ってみたが、2 群の母平均に差があるとはいえなかった ($P=0.0605$)。ただし、もう少し被験者を増やせば、各表示法の差を検出できるかもしれない。

KNP 出力表示の正答率が低い原因として、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がわかりにくいことが考えられる。また、KNP 出力表示は、2 つの文節が上下で重なり合っ

[§] 正答率上位の被験者 3 人はいずれも全問正解者である。

表 1: 実験結果

表示法	正答率	問題タイプ	正答率
KNP 出力	67.6%	I	61.5%
木構造	76.2%	II	59.5%
入れ子ボックス	72.4%	III	76.4%
係り受け強調	73.0%	IV	92.6%
		V	71.6%
全体	72.3%	全体	72.3%

て表示されることがあり、双方の文節に実際には係り受け関係が存在しないのに、係り受け関係が存在すると誤判断してしまうことなどが考えられる。

今回の実験では、被験者は各表示法については初見であり、それぞれの見方をきちんと把握しているとはいえない。各表示法の見方をしっかり教えた上で実験を行った場合、結果が異なる可能性もあり、これは今後の課題としたい。

3.2.2 問題タイプ別

次に問題タイプ別の結果であるが、タイプ (II) が正答率が悪い。タイプ (II) と (III) は、同じ曖昧性を持つ問題で正解の選択肢だけが異なる。差が出る原因として、被験者が各種表示法の見方を理解していない場合と、係り受け構造を見ないで解答している場合が考えられる。どちらの場合も被験者は、文意や単語の並びなどから、ある種の直感を用いて解答していると思われる。文節が近い文節に係るという直感と正解が一致するのはタイプ (III) の方である。これは実験結果と一致する。

また、タイプ (IV) がとりわけ正答率が高い。これは正解が被験者の直感と一致したためであると思われる。文頭の主格が必ず文全体の主格となる言語を母語としている人は、タイプ (IV) において正解が「1. 社長」となるような、文全体の主格が存在しない構文は、なかなか理解しにくい(図7参照)。今回の問題ではタイプ (IV) の正解はすべて、主格が存在する構文としているため正解率が高くなった。主格が存在しない構文を正解とした問題の方が、正解率に差が生じ、正しく係り受け構造表示を見ているかを測定できたと思われる。

以上から断定は出来ないが、今回の実験では、係り受け構造を十分理解せず、直感で解答している被験者も多数存在するのではないかと考えられる。今後、同様な実験を行うならば、被験者が実際に係り受け表示を見て解答しているかを確認できるような実験方法を考案する必要がある。

4 今後の展開

本システムでは主に日本語読解に対する支援機能を提供している。今後、利用者の意見を取り入れながらシステムの改良を施す予定であるが、更なる「あすなろ」の拡張として、日本語作文支援機能を提供する予定である。作文支援を担う機能としていくつかの技術が既に発表されている。熟語の読みがわからない時、個々の漢字の読みから最も適した熟語を推測し提示するシステム“FOKS”[3]、学習者の入力した修飾表現が適切であるかを統計的に判断する手法[4]、指定された単語を含む例文を前後の文脈と共に表示するコンコーダンス[5]など種々の技術を用いて、日本語作文支援機能を実装をしていきたい。なお既存の読解支援機能と新規の作文支援機能は別々に存在するのではなく、相互に補完し合うものである。

また、ある程度多様な機能が実装された時点で、利用者がシステムを利用した形跡をログとして保存し、分析を行う。ログからシステムの機能ごとの利用形態がわかり、機能改善の重要な指針となり得る。さらに利用者が我々の想定外の使い方をする可能性もあり、日本語学習と学習支援システムの関係性のあり方を考えるきっかけとなるかもしれない。

5 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、係り受け表示法間には、ある程度の正答率の差異が見られた。しかし実験結果から、係り受け構造の提示が有効に文理解の促進の助けとなっているか測定することは難しいことがわかった。実験の方法については、係り受け表示のみから推論できるような問題への改善につとめ、評価方法の精度を高めることを今後の課題とする。

参考文献

- [1] 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.
- [2] 黒橋禎夫, けっこうやるな KNP, 情報処理学会誌, Vol.41, No.11, 2000.
- [3] Slaven Bilac, Timothy Baldwin, Hozumi Tanaka, Construction of a Japanese learner-friendly dictionary interface, 言語処理学会第8回年次大会発表論文集, 460-463, 2002.
- [4] 戸次徳久, 仁科喜久子, 構文解析器を利用した作文支援システムの開発 形容詞表現に関して, 第3回「日本語教育とコンピュータ」国際会議予稿集, 67-70, 2002.
- [5] 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 の開発と評価, 第3回「日本語教育とコンピュータ」国際会議予稿集, 75-78, 2002.