表示切替式読解支援における教育用「文型」に基づいた構造化

中尾桂子　　森下淳也
大妻女子大学 短期大学部　　神戸大学 国際文科学研究科

我々は教育効果を念頭に置いた読解支援システムを提案しているが，これまでのシステムでは，テキスト表示の切り替え方法に問題があり，表示箇所を決定する基準が荒くて示唆が十分ではなかった。この従来からの問題である重要箇所特定方法を再検討し，改善に向けて新たに既存の教育用文型を利用した。結果，ユーザの立場に合わせて，かつ，日本語を教えるどの教師にもわかりやすく，文章中の読解手がかりを示すことができ，表示切替の際の構造化の見直しにつながった。その表示結果，ならびに，表示結果に対する日本語教師の印象から，日本語教育用「文型」が表示箇所特定に有効であることを報告する。

Effectiveness of Educational "Sentence pattern" for display switching on comprehension support system “Text Viewer”

Keiko Nakao　　Jyun-ya Morishita
Department of Junior College　　Graduate School of Intercultural Studies
Otsuma wemen’s University　　Kobe University

We suggest a Reading Comprehension Support System by switching the text display style. In This paper, we report the effectiveness of the “Sentence Pattern”. On the conventional system, a indication method of the text included a problem. The standard rule to decide an indication point was rough, and suggestions were not enough. To this solution, we reviewed of system structure and a indicating condition to reshuffling. we reexamined a specific method and used the Sentence Pattern newly for improvement. From the impression of the teacher for the indication result and also the indication results, we report that "a sentence pattern" for Japanese education is effective for indication point identification.

１．まえがき

我々の検討している読解支援システム”Text Viewer”では，簡略化を短文化とは考えない。

文を構成する語は，大きく二種類，すなわち，述語の格関係において不可欠である必須成分と，動詞との格関係の上では必ずしも必要ではない副次的成分とに分けられる。

我々のシステムでは，この必須成分に該当するテキスト中の語を強調表示することで，他の語成分を背景化し，一元的な配列の中からスキミングすべき候補を示す。つまり，我々の「読解支援」は，文中で意味的に重要だと扱われている語をクローズアップすることで「簡略化」としているわけである。

本システムでは，一文中の副次成分と必須成分が明示的に区別された状態で学習者に明らかにされるべきであるとの観点から，故意に，副次的な部分のテキストも消さずにおき，文中の必須要素を強調表示して読者の前に前景化して，読解の手がかりを残している。

これは，読解には経験値から予測して判断する能力が必要なためであるが[1]，この経験値を養いながら，読解時に必要な要素を捉える支援を行うことで，読解の一つのプロセスをトレースするという教育的指導を念頭においている。その理由は，背景化された副次的な語成分のある通常の文章から，強調部分を拾うことで，スキミング能力を養いながら，読解を進めることができるためである。支援ではあるが，自律的な読み手を育てるという立場でシステムを設計しているのである。

したがって，文を書き換えるタイプのシステムや，辞書による語意を補填するタイプのシステムとは根本的に異なる支援方法を念頭に置いている。

本稿では，このシステムの表示マーカーとして，日本語教育用「文型」を利用した表示切替を試みた結果，さらに，改善された部分と残る課題について報告する。

２．システムの構成と課題

2.1. システムの構成

本システムは，副次的要素を切り離さずに，いつでも参照できる可逆性を考慮し，ベースの表示は，焦点を当てる箇所にマークをつけるという方法で強調している。

そのため，テキストを XML[2]で構造化しているが，この構造化は，テキスト自体とテキス

(c) Information Processing Society of Japan

トの内部に埋め込まれたものと 2 つの異なる次元での構造化である。

テキスト自体には，章節，段落，一文，一単語までを階層化できることを計画している。これは書物や文書が持つ構成上の意味関係である。なお，文書自体の構造は，入力データに必ずしも備わっているとは限らないものでもあるため，空行数で指定することができるようにしている。

一方，テキストの内部に埋め込まれたものは各単語が深層に保有するもので，テキストの表面には見えない意味的な情報である。すなわち，品詞，文字列からは特定できない漢字の読み，検索用インデックスとリンク情報であるが，これらは，テキストを変換するのに必要な種類のものである。文書が持つような表面的な構造とは異なった種類の情報であることから，テキスト部分ではなく，単語単位の構造化の XML タグに属性として埋め込むことにしている。

2.2. 利用の流れ

ブラウザで図 1 の”Text Viewer”トップページを開き，直接入力，または，アップロードしてテキストを入力する。

入力されたテキストは，Chasen[3]を利用して形態素解析を行なう。その結果得られる単語(形態素としての切れ目)，品詞情報，漢字の読みの情報，学習漢字にのみ配当学年情報を，単語単位の XML タグ内のアトリビュート属性として埋め込みながら，同時に，入力されたテキストに改行のみの空行や章節区分がある場合は，テキスト構造として，XML タグに反映する。このようにして，オリジナルテキストは，深層情報を加えて構造化した XML テキスト 1 に変換される。

次に，この XML テキスト 1 上で，下位構成要素である語同士の関係を検索し，そのつながりを反映させて，テキスト中のマークアップ箇所の特定の準備を行なう。これら，構文情報や連語や共起上のつながりといったマーク箇所特定用の情報を，さらに，XML 属性に組み込み，再度，別の XML テキスト 2 を作成する。

この 2 つ目のテキストを結果として表示するのであるが，表示の際に HTML で該当箇所をマークする。この内的処理の概略を表したものが図 2 である。

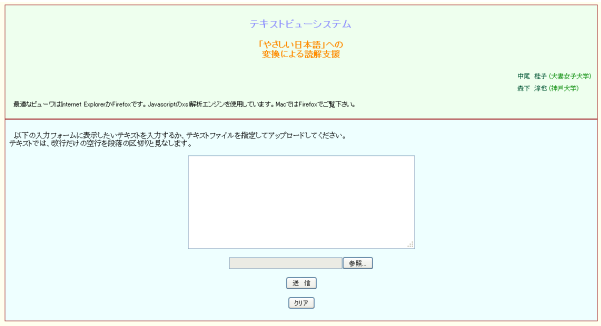

図 1: ”Text Viewer” トップページ

システムには，以上のようなベースのテキスト強調表示に加え，イメージ画像，辞書，KWIC 検索の機能を追加し，テキストに元からないが，ある方が理解促進に繋がるという情報を補填できる機会を備えている。

2.3. システムの課題

本システムは，表示の意図がわかりにくいという指摘が，従来からあった。主に，日本語教育に関係が浅い日本人からの指摘であり，教室で日本語を学習する留学生等には，読解指示の意図が伝わらないということはなく，むしろ，有益だと期待できるという声が聞かれ，日本人ほどわかり難いという声はないことから，これまであまり取り沙汰してこなかった。

しかし，専門用語がふんだんに含まれる日本語テキストを日本人が読む場合も考慮するとすれば，表示の意図を更に明確にする必要がある。

そこで，強調表示の意図を左右する，強調箇所の特定条件を見直すために，これまでの条件と，その指定方法を省みる。

これまで，マークアップの箇所を指定する条件としていたのは，①必須格，②基本文型にもとづく主述パタン，③有標性の高い「～ない」否定文，④複文マーカーと，⑤連語，共起語といった語どうしの関連性であった[4]。この条件に基づく表示結果を図 3 に示す。

この条件のうち①の必須格と②の基本文型条件が最も優勢で，たは，独立した条件であるため，この 2 つを見直すことにする。

①の必須格は，自然言語では，表層格もその数も，述語タイプにより異なる。しかし，文字列に対して機械的に処理する場合には，使用する表層格は限定されるため，単純に一般化していた。

その理由は，文は，述語の働きの対象や陳述の主，また，述語の主格といった，述語との関わりが強い語と，そうでない語があり，述語との関係が強い語は，情報として「前景」化されるもので，その他の語は付加的語句として背景的要素であるという説に基づく[5]。

②の基本文型として指定しているのが，ごく初歩の「基本文型」である。中尾(1998)でも報告しているように，この文型は日本語では当為

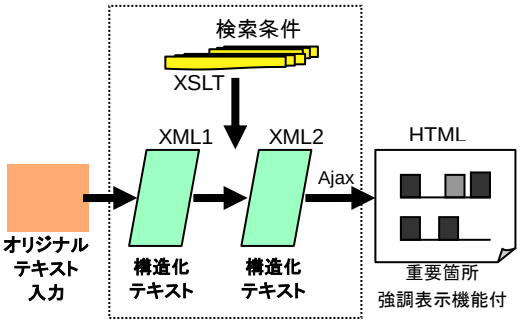

図 2: システム内の処理の流れ(xslt 検索⇒xml 変換⇒表示)

(c) Information Processing Society of Japan

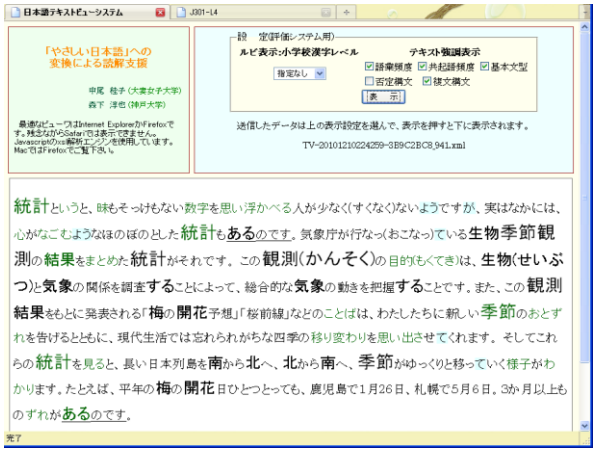

図3:旧タイプの処理で表示した結果

のものであるため，使用率が 90%を超える[6]。マークの意味が薄く，無標と変わりないのであるが，複文構造と併用することで外枠を指定するという意味が生きてくるとして利用してきた。また，談話のつながりを把握するための，文脈把握の手がかり（＝ディスコースマーカー）の一つに，係助詞「は」があるが，日本語教育で言う「基本文型」はこの「は」をベースにすることも「基本文型」を利用する理由である，

以上，①，②の利用理由を省みたが，本システムの問題点としてこれまで指摘されてきたことに，意味を理解する上で必要な内容を指す部分が，教育用「文型」の知識が不十分な利用者にはわかりにくいというものがあった。それは，強調表示の対象となる内容上重要な語の特定方法があまり厳密ではなかったことによる。

また，日本語教育で言う「基本文型」は，短文で記述されたテキストの場合は特に，ほぼ全ての語が強調表示の対象となる場合もあるように，特別視する情報として強調の対象になりにくいという欠点につながっていたのである。

さらに，格関係を表す格フレームと日本語教育用の「基本文型」は述語を中心として見るか，主題を中心として見るかという点で全く異なるものであることから，形式のみでの条件設定ではどっちつかずのものが多くなり，指定する箇所が多くなる結果，強調表示が荒くなる。

表示が荒い原因は，やはり，この①と②の 2 点にあると考えられる。

2.4. 構文情報と「文型」の併用

他の読解支援システムでは何を用いているのかというと，例えば，「あすなろ」[7]では，重要度を，自然言語処理技術を利用して係り受けスコアという構文情報で示している。そのスコアに基づき，一文の構造を階層化した図で示し，格関係を把握させるよう支援している。

確かに，文字列と品詞情報で示せる範囲では，形式的な表層格に準拠する事になるため，形態素情報以外の情報をより明確に利用する方がよい。そして，語の語義的意味ではなく，文法的な意味を利用して重要度の差別化を図るには，自然言語処理から見ると，構文情報（格フレーム）を利用するという考え方が主流である。

しかし，構文の係り受け情報は，述語とその必須格のパタンを統計的に一般化したスコアでしかない。また，語レベルの格フレームは，日本語では省略されていることも多く，一般的な文章を形成するという意味では有益ではあるが，何にでも利用できる枠組みが，個別の文章を読み進めていくのに十分な効果が見込めるのかわからない。

そのため，通常，日本語学習者への日本語教育では，述語の必須格をパタン化して取り出して教えることよりも，どのような個別の文章を読む場合にも利用できる知識の方，すなわち，表現上の連語や共起パタンからなる語の関係パタンの意味を教える方が優先される。つまり，どちらかといえば，必須格情報よりは，寺村(1987)の言う「表現文型」[8]など colligative effect を検索指標に応用する方が，個別の文章に対応するための基礎知識として，日本語学習者の日本語の読解には有益だということである。

したがって，日本語教育上の流れに沿った条件も検討しておく方が良く，やはり，教育用の連語パタンも，文の意味をとる上では外せないだろう。

そこで，係り受けスコアをサブとして用いながら，日本語教育用の「文型」を用いることにする。

ただし，現在，日本語教育で利用される教育用「文型」は，品詞情報，構文情報，意味論的，語用論的情報を統合した colligation 的な情報で，ある種，漠然とした構造パターンの概念的なものであることから，検索のキーワードとして指定しにくいものも多い。日本語の「文型」は，できるだけ，一貫性を維持できるようにすべきであるだろう。

３．構文情報と「文型」の併用

動詞を中心とした係り受けに基づく「構文情報」を利用すること，また，より詳細な「文型」情報を導入することで，従来の「基本文型」のみのあいまい性を補足する方法を検討する。

構文情報は KNP[9]の解析結果から，動詞に直接かかる部分を利用する。

強調表示に利用する「文型」は，学習者の利用が多い学習用文型を使うことにし，くろしお出版のご協力を得て，グループジャマシイ(1998)『日本語文型辞典』[10]の見出し語を文型に借用する。

(c) Information Processing Society of Japan

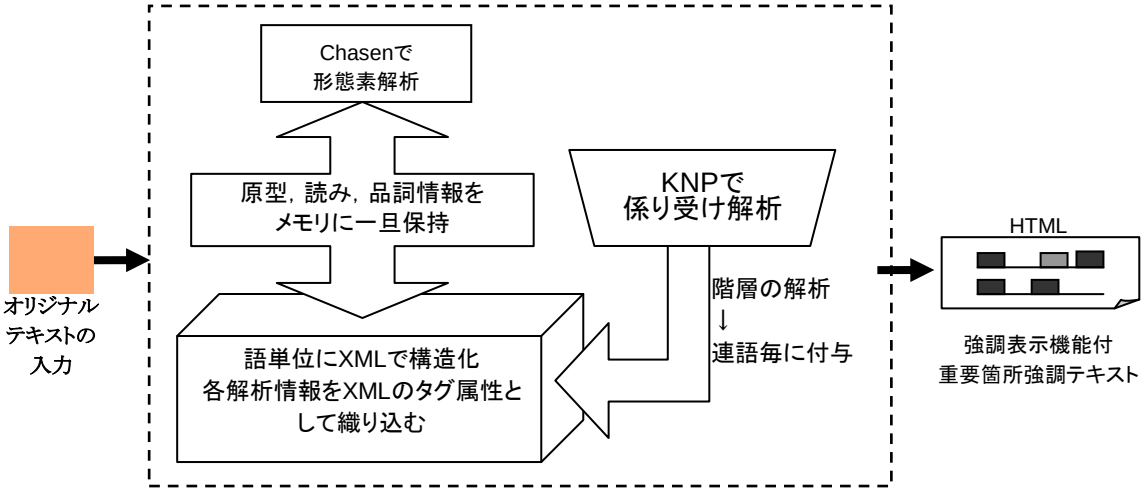

図 4：テキストへの属性付与の流れ

以下，(1)～(3)が処理であるが，その流れを図 3 に概念化する。

(1)「文型」の下処理

①『日本語文型辞典』の見出し語のテキスト部分を Chasen で解析し，1 文型，1 行の XML 文書にする

②「文型」の colligation 情報は，単語の品詞情報，「文型構成要素の組み合わせ」情報，例文を構文解析した結果から得られる「動詞に直接係る語の格助詞と品詞」情報，「文型の基本義」の 4 種とし，それぞれを，別々に下処理して，単語ごと属性や，「文型」セットの属性とし，リスト XML テキストの下位に構造化する

(2)テキスト構造の変更

①テキストを Chasen で形態素解析し，単語ごとに品詞と漢字の読み方を下位属性に組み込む

②テキストを構文解析し，その結果のうち，動詞に直接係る語の格助詞とその直前の語の品詞を別の情報として属性にする

以上の準備により，2 種類の XML 文書を作成し，両者の間の類似部分を X-path でサーチして，一致した部分を指標に xslt で別の文書に変換する。このフィルタリングの際，html による表示変換の指示を加えるが，強調表示は次のようなものとする。

(3)強調表示

①『日本語文型辞典』の見出し語部分，すなわち，「文型」は太字拡大表示

②動詞に直接係る部分，すなわち，必須要素である語に，色マーカーをつける

③『日本語文型辞典』の解説部分，すなわち，「文型の基本義」を別フレームで表示可能にする

④学習漢字レベルで表示切替を選択していた漢字の読み方は読みたいときだけ出す

図 3 には割愛しているが，辞書や画像検索機能は，新バージョンでも，継続して利用できるようになっている。前回，学習者にとっては別ウインドウだと面倒ではないかという意見があったことから，辞書を別の画面ではなく，同じ画面内に表示するように変更した。辞書機能の表示スタイルは図 6 に見られるような形にしている。

4．検索キーワードの具体化と文章構造

利用した条件別に結果を比較して見る。次ページ図 5 は「係り受け情報」に基づいて必須成分を表示した結果で，図 6 は『日本語文型辞典』に基づく表示結果である。

これらを見ると，多くの日本語教材に見られるような学習項目のマークアップにはなっているものの，読解のための示唆という観点から見れば，それぞれの守備範囲の違いが明らかであると言える。

つまり，構文関係に基づいた必須成分だけでも，「文型」だけでも，読み取るべき情報を示唆しているようには見えない。

そこで，必須成分，文型の 1 つ前の 1 語を表示してみた。結果は図 7 である。そうすると，少し，注目すべき箇所に近づいてきたものの，やはり，まだ，十分ではない部分が多い。

ただし，1 語前の語を含む方が情報量が増えて何がどうしたという点がとりやすくなるためによいことから，1 語前を含めた表示を行なうこととし，次に，必須成分とその前 1 語のセットと，文型とその前 1 語のセットを，図 8 のように比較してみた。

図 8 の左の，必須成分とその前の語 1 語の方が，右の，「文型」とその前 1 語の方よりは，

(c) Information Processing Society of Japan

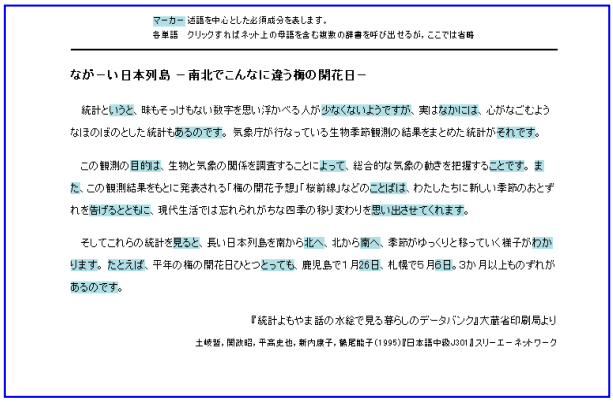

図 5:係り受け情報による必須要素の表示例

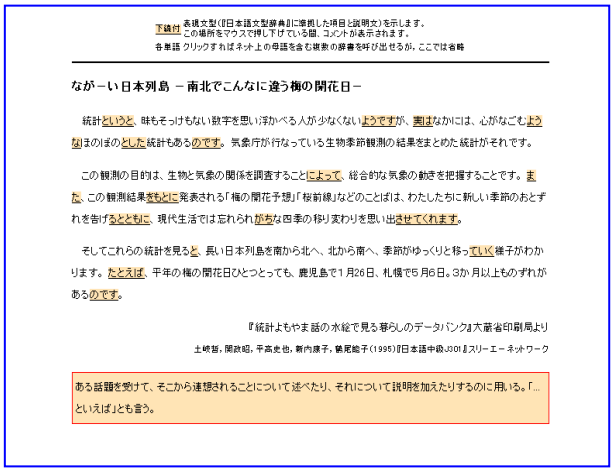

図 6:『日本語文型辞典』に基づく表示例

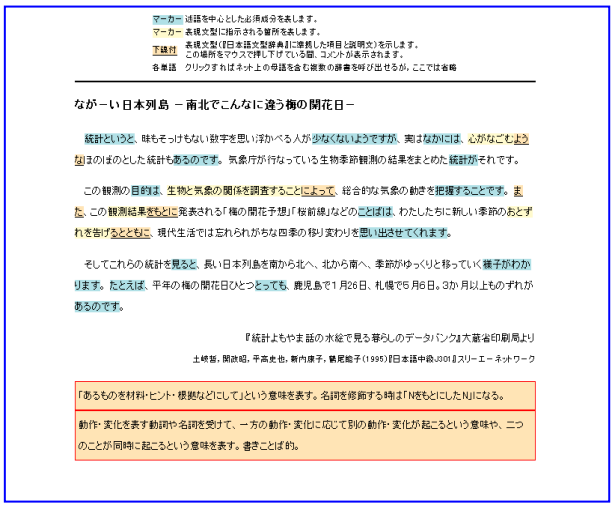

図 7:必須成分 + [文型+前 1 語]での表示例

主文が確実にマークされており，骨格として必要な部分がカバーされている。

しかし，一部分というよりは，結構な数の語が強調されているため，スキミングのための処理の支援としては少し煩雑な点が見られる。図 7 のような必須成分と図 6 の「文型」の両方合わせた折衷方法で，かつ，それらの 1 つ前の語までをマークするという方法で表示した図 8 の右の表示が強調量と強調箇所の両方ともに最も確実な表示となっている。

5．学習者の印象と教師の印象

漢字圏からの中，上級の留学生(日本語能力検定試験 2 級～1 級)4 人に，図 8 の右の必須成分と「文型」とそれらの 1 語前を強調した表示結果を見てもらい，その印象を聞いてみたところ，以下のような意見が得られた。

- （初見の文章の）だいたいの意味を自分で見るのに便利だと思います。
- 辞書を調べるのがたいへんだから，単語検索があってうれしい。
- 表示(の色)が（薄くて見えにくいために）わかりにくいですから，もうすこしがんばってください。
- いいと思う。

統制した状態での正式なアンケートではなく，ただ 4 人に聞いただけであるが，我々の狙いとしている大意把握のための支援としては有益だという印象が得られたと考えられる。

しかし，見えにくいと，表示方式に対する工夫が求められていることから，表示方法やインターフェースにも，一層の工夫が必要だということも明確になった。

また，日本語教師 3 人に，読解の支援として表示された内容はどうかと問うたところ，対象が限られることはともかくとして，以下のような印象が得られた。

- とりあえず、初見でこの文が理解できそう。
- 通常、語彙、表現（文型）と、部分を抑えて全体にかかるところ、1 面ですませられてよい。
- ページを繰って、余分な練習をしなくてもいい。
- 本文が典型的なものから離れている場合は注釈も必要だが，語検索でまあまあ（大意把握が）可能かも。

しかし，表示結果が妥当だとしながらも，「マーカーでつけた情報を見慣れるまではリテラシーの低い人には利用が難しいのではないか」というコメントなど，感覚的に，見え方の方を気にするものも多かった。

6．考察

教育用「文型」を，既存の辞書を用い，一貫した形で検索に用いたことにより，完成度の高い形式での検索が可能になった。

(c) Information Processing Society of Japan

その結果，重要度の高い必須格だけでなく，文型解釈上必要な語に，強調表示をかけられるようになり，主題に関連する重要箇所が，内容上の意図と共に，従来より，正確に特定できるようになった。

また，構文解析の結果だけで行うよりも「文型」を用いた方が，内容に即した状態で，表現上重要な箇所を示すことが確認できた。このことから，文脈上の重要情報は，必須格を中心とした係り受け主体の構文情報のみでは得られないものもあることが明確になった。

現時点では，実際の表示における HP 上への投入という処理に時間がかかるという問題があるが，目的の「読解の手がかり」である文脈上重要な箇所の強調が可能であることが確認できたことは，効率化への方向性が得られたとして評価できるだろう。

７．まとめ

日本語学習者を対象とした読解支援においては，読解時に必要なスキルを明示した教育的配慮の支援が望ましいという考えの元，手がかりを強調表示する形式でのシステム開発に取り組んでいる。

我々の，従来のシステムでは，強調の意図が不明瞭であったため，強調箇所の検索の厳密化のために，品詞情報，係り受け構文情報，意味論的，語用論的情報である文型情報を総合して文中の重要語句を強調表示した。

結果，表示をより的確につけられるようになったことから，ある程度の質的向上が実現できた。

ただし，検索方法には他にも XML の namespace を利用するなど，いくつかの方法があるため，処理速度向上に向けて，複数の検索方法を比較検討する必要がある。

また，表示方式やインターフェースのあり方など，見やすくする工夫に関しても課題が多い。

さらに，学習者の読解試験の対策を考えられるような，より詳細な情報の表示を目指した読解支援として，検索キーに対する十分な考察を探る必要がある。今後の課題としたい。

参考文献

[1] 近松暢子「外国語としての日本語の読み・読解研究」『第二言語習得研究への招待』くろしお出版 pp67-85,2003.

[2] XML:http://www.w3c.org/XML/

[3] Chasen: http://chasen-legacy.sourceforge.jp/

[4] 中尾桂子・森下淳也: 基本文型と語彙の重要度を用いた強調表示による読解支援-，人文科学とコンピュータシンポジウム論文集，Vol.2009,No.16，pp256-234，2009.

[5] Croft, W.: Syntactic Categories and Grammatical Relations, University of Chicago Press,1991.

[6] 中尾桂子:小学校検定教科書の構文調査―外国人児童の教科学習支援のための基礎研究―，小出記念日本語教育研究会論文集 7,pp41-55,1999.

[7] あすなろ: http://hinoki.ryu.titech.ac.jp/asunaro

[8] 寺村秀夫「構造文型と表現文型」『講座日本語と日本語教育』13 明治書院.,1987.

[9] KNP:http://www-lab25.kuee.kyoto-u.ac.jp/ nl-resource/knp.html

[10] グループジャマシイ『日本語文型辞典』くろしお出版, 1998.

統計というと、＿＿＿＿＿＿思い浮かべる人が少なくないようですが、実はなかには、＿＿＿＿統計もあるのです。
＿＿＿＿＿＿＿＿統計がそれです。
観測の目的は、＿＿＿調査することによって、＿＿＿＿を把握することです。
「梅の開花予想」「桜前線」などのことばは、＿＿＿＿おとずれを告げるとともに、＿＿＿＿移り変わりを思い出させてくれます。
＿＿＿＿統計を見ると、＿＿＿南から北へ、北から南へ、＿＿様子がわかります。
＿＿ひとつとっても、＿＿1月26日、5月6日。＿＿＿＿ずれがあるのです。

統計というと、＿＿＿＿＿＿少なくないようですが、なかには＿＿＿＿なごむような＿＿統計もあるのです。
＿＿＿統計がそれです。
＿＿＿目的は＿＿調査することによって＿＿＿＿＿把握することです。
また、＿＿結果をもとに＿＿ のことばは＿＿を告げるとともに、＿＿＿＿移り変わりを思い出させてくれます。

＿＿＿統計を見ると、＿＿様子がわかります。
たとえば、＿＿＿＿ひとつとっても、＿＿ずれがあるのです。

図 8: 必須成分+前 1 文節(左)　対　必須成分＋「文型」+前 1 語(右)

(c) Information Processing Society of Japan