

階層的句構造の表示と比較対応に基づく第二言語学習支援

Second Language Learning Environment Assisting Comparative Study of Envisioned Hierarchical Phrase Structures

中村 宏
Hiroshi NAKAMURA
一橋大学
Hitotsubashi
University

掛川 淳一
Jun'ichi KAKEGAWA
兵庫教育大学
Hyogo University of
Teacher Education

伊藤 紘二
Kohji ITOH
東京理科大学
Tokyo University
of Science

<あらまし> 言語の獲得は、音声を含めて理解と生成のいずれについても、意味制約関係を表わす階層的句構造の作り方を獲得することで行なわれるという仮説に基づいて、第二言語学習を支援するために、母語あるいは媒介語の句構造と学習対象言語の句構造を表示し、各階層において比較対応させることができるシステムを提案する。実装は、対象言語を日本語、媒介語を英語とし、テキストの対訳を用意して、その構文解析結果によって生成した xml ファイルを基に簡略化を行ない、テキストの形態素/単語をその並びのままに取り込んだパネルの階層により、句構造を視覚的に表現する。この構造に関係させて分類した表現型をテキストの中に検索、比較させて、表現の作り方を学習させ、再現作文を局所構造のチェックにより支援する。また模範のピッチ解析を組み込んで、音声学習を支援する。

<キーワード> 第二言語学習 日本語教育 句構造表示 音声言語学習 生成支援 表現獲得

1. はじめに

我々は第二言語の表現獲得に対して、音声の支援も含め、学習者と教授者を支援する総合的な学習支援システムを提案してきた[1][2]。音声を含めて理解と生成のいずれの学習も、言語の句構造に基づいて行われることに着目し、母語あるいは媒介語の句構造と学習対象言語の句構造を表示し、各階層において対応させることができる第二言語学習支援システムを現在提案しているが[3][4]、本稿では学習対象言語を日本語、媒介語を英語としたシステム構成について述べる。

2. 階層的句構造

句に主辞が置かれ、それを修飾、説明する複数の句を主辞の前あるいは後に置くことを階層的に繰り返すことによって、あらゆる言語における文は作られている[5][6]。主辞には、かかられて完結する主辞と、かかる句とかかられる主辞の関係を示す機能的な主辞がある。また、語の単位が複合したものを主辞とすることも多い。コンピュータで句構造を表示するには、主辞を

載せたパネルの上にそれにかかる句のパネルを載せることを階層的に繰り返す。

日本語の場合には、日本語係り受け解析器 CaboCha[7]による解析の仕方を基に、Japanese Phrase Structure Grammar[5]も参照して、自己完結的な主辞を「head 形態素複合」とし、結合機能的な主辞を「functional 形態素複合」として、二者の接続からなる chunk を句構造表現の単位とみなしていく。機能的な主辞が空で、完結的な主辞の活用でかかりの機能を構成することもある。構造を視覚的にパネルで表わす方法としては、かかられる完結的な主辞を載せた chunk パネルの前の方に、かかるほうの chunk を並べて乗せて表現する。日本語においては、機能的な主辞は、完結的な主辞を載せたパネルの後ろの部分に付加する形で示す。

英語では、Stanford Parser[8]による解析の仕方を基に Head-Driven Phrase Structure Grammar [6]も参照して、動詞や形容詞にかかる主語、目的語、副詞句や、名詞にかかる形容詞句のように主辞の前後の位置関係だけでかかり方が指示されている句と、前置詞や接続詞という機

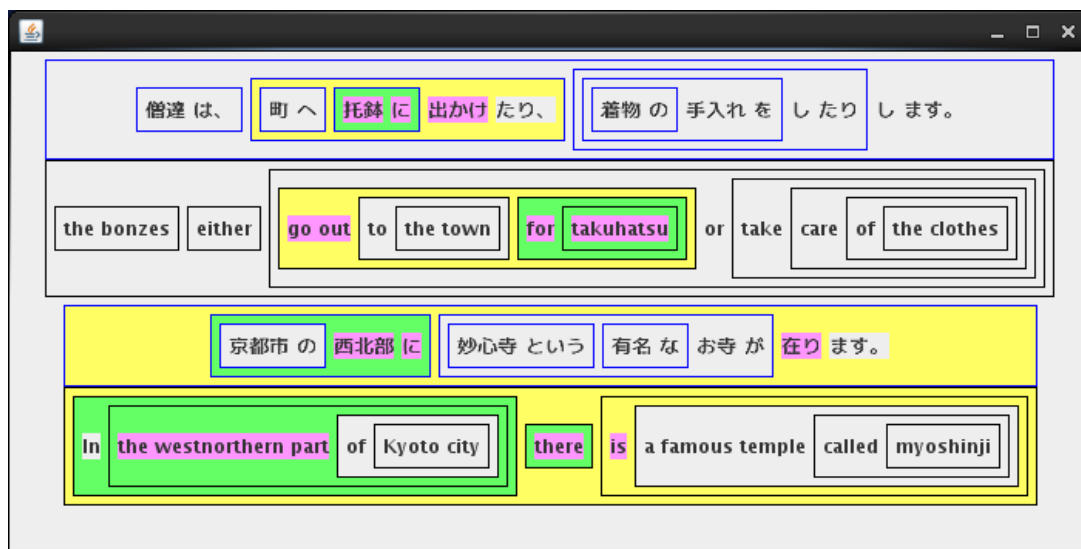


図 1 句構造を用いた表現検索結果と英文対応表現

能的主辞が置かれた句が完結的主辞を置いた句を従えて、かかり方を指示している場合がある。機能的な主辞は必ず完結的主辞を置いた句がかかる先として置かれて、独自の句を構成する。パネル表示は日本語と同様、全て主辞を載せたパネルの上にそれにかかる句のパネルを載せて表示する[4]。

3. 句構造に導かれる表現型の学習

日本語文を和英構表示の対応によって読み取らせながら、表現型の使い方を獲得させ、再現作文で学習達成度を測り定着を促すシステムを試作した。

3.1 表現検索に基づく型の使い方の学習

学習対象の日本語文に対し、英語の対訳文を予め用意し、その構文解析結果によって生成したデータを基に、簡略化を行った xml ファイルをシステムは保有する。ここから、2 で述べたように、テキストの形態素/単語をその並びのままに取り込んだパネルの階層表示によって、句構造を視覚的に提示する。

テキスト内に登場する表現型は先行研究[2]での提案を発展させて、キーとなる表層と品詞指定をふくむ句構造として分類定義し、表現型の構造を例文の構造にマッチングさせることによって自動的にアノテーションをしておく。学習者はテキスト中にある同一または関連した表現型を含む部分構造を検索し、比較して、帰納的に表現の作り方と使い分けを学習する。図 1 に「名詞-助詞」

に「動詞」の表現型検索を行なった結果とその英文対応を、図 2 には検索の際に用いた表現型ライブラリを示す。手作りで用意される媒介語との対応が、表現型の使い分けを判断する助けになる。また、使い分けの支援として、品詞の下位分類としての意味カテゴリを提供することも(たとえば、いる/ある、を生物/無生物で使い分けるなど)分かっている範囲で行なう。表現型に対応した穴埋め問題で学習の達成度を測る。

```

?xml version="1.0" encoding="UTF-8" standalone="no"?>
<expressions>
  <expression id="2">
    <phrase id="0" on="1,2" pos="動詞">
      <phrase id="1" pos="動詞">
        <func pos="助詞">
          たり
        </func>
      </phrase>
    </phrase>
    <phrase id="2" pos="動詞">
      <func pos="助詞">
        たり
      </func>
    </phrase>
    する
  </expression>
</expressions>

```

図 2 表現型ライブラリ

3.2 日本語文生成支援

句構表示を行なった教材テキストの中の一語を形態素に分解し、場合によっては余分な機能形態素も加え、それをランダムに並べて提示し、学習者は一覧から選んで正しく文章になるように並び替える。チェックボタンを押すと、学習者が並べた形態素の列の中で、原文と構造が一致す

る部分について合体させて構造表示を行なう。こうして出来たマクロな部品の並び替えを階層的に行なわせることにより、再現作文を支援する。この支援は、表現獲得の評価としても使うことができる[9]。

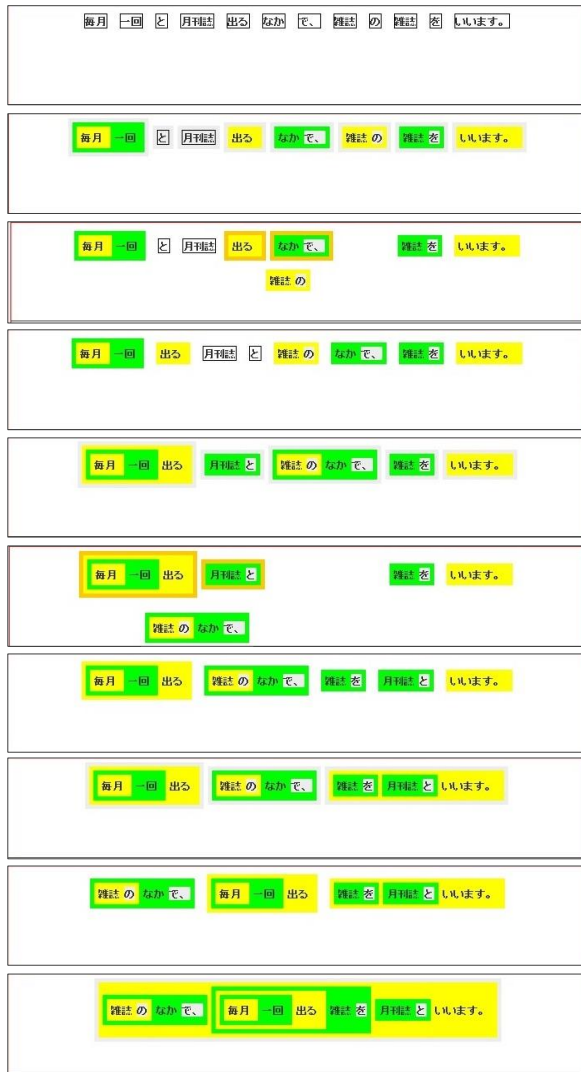


図3 再現作文の流れ

4. 音声言語学習支援

表現型の使い方の学習や再現作文においてシステムでは、学習テキストに対応した模範音声再生を行ない階層構造単位で学習者に聞かせて、表現獲得を支援する。再生音声に同期して、予め分析された韻律（高低アクセントとポーズ）を、句構造に関係させてハイライト表示することにより、韻律自体の学習を支援すると共に、韻律特徴

の獲得を表現獲得に結びつける。

模範音声の音声分析と音声認識に基づいて得られたタイムスタンプとピッチ情報を、CaboCha[7]によるかかりうけ解析結果のXMLファイルのchunkタグに、属性として埋め込む。読み上げテキストを与えての音声認識であるから、認識誤りは少ない。図4に示すようなXMLファイルに基づき、モーラ単位のピッチ表示を階層的句構造に埋め込んだパネル表示を生成する。そして、音声の再生に合わせて、タイムスタンプで同期を取りながら、句構造とモーラピッチのハイライト表示を行なう。図5に表示例を示す。

```

</sentence>
</paragraph>
- <paragraph lid="1">
  <sentence lid="0">
    <chunk lid="0" score="1.17228" rel="D" p="1" link="1" id="0" head="0" func="0" end="5540"
      begin="5280">
      <tok lid="0" id="0" read="コノ" pos="連体詞" ne="0" ctype="" cform="" base="この">
        <mora p="1" end="5400" begin="5280" pitch="249.3485500822916"></mora>
        <mora p="0" end="5540" begin="5400" pitch="292.07022251928123"></mora>
      </tok>
    </chunk>
    <chunk lid="1" score="3.60965" rel="D" p="1" link="4" id="1" head="1" func="2" end="6380"
      begin="5540">
      <tok lid="0" id="1" read="ザツン" pos="名詞一般" ne="0" ctype="" cform="" base="雑誌">
        <mora p="0" end="5720" begin="5540" pitch="304.8462324180303"></mora>
        <mora p="0" end="5930" begin="5720" pitch="259.7766263889131"></mora>
      </tok>
      <tok lid="1" id="2" read="ハ" pos="助詞 係助詞" ne="0" ctype="" cform="" base="は">
        <mora p="0" end="6380" begin="5930" pitch="256.2853111938282"></mora>
      </tok>
    </chunk>
    <chunk lid="2" score="0.961198" rel="D" p="1" link="3" id="2" head="3" func="3"
      end="6980" begin="6420">
      <tok lid="0" id="3" read="マインキ" pos="名詞 普通名詞可能" ne="0" ctype="" cform="" base="毎
  
```

図4 韻律表示xmlファイル

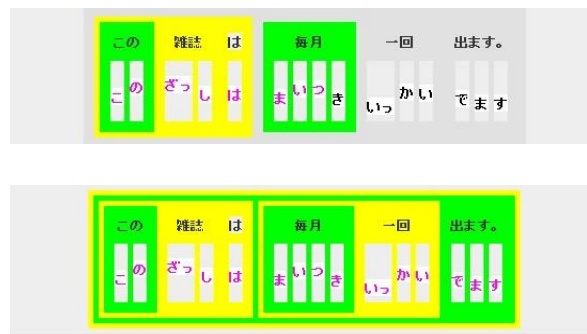


図5 音声再生に伴うハイライト表示例

5. まとめと今後の課題

日本語の学習環境として、日本語のテキストと学習者の母語あるいは媒介語によるその対訳を、同様な階層的句構造で並列表示し、読解を支援すると共に、学習者の要求した箇所の表現型に応じた和文・媒介語構造の提示と、同じあるいは関連した表現型の検索結果の表示から、表現の作り方

と使い分けを帰納的に獲得させる仕組みを試作した。また、表現検索で支援しながら、再現作文を、部分的な句構造の正否をフィードバックして支援する仕組みを試作した。

また、再生音声に同期して、韻律を句構造と対応させて表示し、音声言語学習を表現獲得につなげることを試みている。学習者の発話を音声認識エンジンによって評定してフィードバックするシステムも検討したい。なお、提案した支援に対する評価実験を行なう予定としている。

参考文献

- [1] 中村宏, 小松崎聖, 掛川淳一, 伊丹誠, 伊藤紘二: “音声再生に合わせて文字列と韻律の視覚提示を行なう第二言語表現獲得支援システム” 教育システム情報学会誌, 24 (4), pp.410-416, 2007.
- [2] 掛川淳一, 中村宏, 関谷政則, 伊丹誠, 伊藤紘二: “自然言語処理を用いて日本語教育のための例文検索を支援するシステム” 日本教育工学会論文誌, 25 (2), pp.85-94, 2001
- [3] 中村宏, 掛川淳一, 伊藤紘二: “句構造表示に導かれた第二言語学習支援環境” 日本教育工学会第 32 回全国大会, 2a-B108-01, 2016.
- [4] 伊藤紘二, 中村宏, 掛川淳一, 佐々木諒: “句構造の提示に基づく第二言語習得支援”, 日本第二言語習得学会第 15 回年次大会, P2-4, 2015.
- [5] T.Gunji: “Japanese Phrase Structure Grammar: A Unification-based Approach,” Springer Netherlands, 1987.
- [6] C.Pollard, I.A.Sag: “Head-Driven Phrase Structure Grammar,” University of Chicago Press, Chicago, 1994.
- [7] 工藤拓, 松本裕治: “チャンキングの段階適用による日本語係り受け解析” 情報処理学会論文誌, 43 (6), pp.1834-1842, 2002
- [8] D.Klein, C.D. Manning: “Accurate Unlexicalized Parsing,” Proc. ACL, 2003.
- [9] 掛川淳一, 神田久幸, 藤岡英太郎, 伊丹誠, 伊藤紘二, “日本語学習支援システムにおける作文診断処理系の提案と試作”, 電子情報通信学会論文誌,D-I, Vol.J83-D-I No.6 pp.693-701 2000年6月