

学習者に合わせた例文表示ツール

DISPLAYING EXAMPLE SENTENCES ADEQUATE TO LEARNERS' LEVEL

吉橋健治、傅亮、仁科喜久子

Kenji Yoshihashi, Liang Fu, Kikuko Nishina

東京工業大学

概要： 日本語作文支援システムの一部として、学習者に合わせた例文を生コーパスから検索し表示するツールを開発した。まず、学習者の日本語習得レベルに合わせるために、生コーパス中の各文の単語難易度、文法難易度を付与する。また、読みやすい文が検索結果の上位に含まれるように、構造的な難しさを計算し難易度を付与する。構造の難しさとして、連体修飾節、遠くにかかる文節、多段の修飾を考慮した。さらに、一般的な単語難易度とは別に、各学習者の既知の語彙を基に例文フィルタリングができる機能を設けた。これらによって、個々の学習者にとって平易でありながら豊富な例文を提示する。

キーワード： 例文、学習者レベル、文難易度、構造難易度、作文支援

1. はじめに

現在までに、第二言語としての日本語の学習を支援するシステムがいくつか開発されている (Nishina & Yoshihashi, 2007, Totsugi & Nishina, 2002, Kawamura, 2002)。第二言語を習得する上で、学習者が理解することと、適切に使用することの間にはかなりのギャップがあり (石川, 2005)、特に、作文をする場合、適切な用法辞書で意味を調べるとともに多くの例文を参照することが重要となる (Tribble & Jones, 1997)。学習支援システムを提供する立場からすると適切な例文を多く収集し、提示する必要がある。例文が適切であるためには、広範囲の文脈が無くても文意の理解が可能であることをはじめ、語彙が難しくすぎないことや、文の構造が簡単であることなど様々な条件が要求される。何を難しく感じるかには学習者間で個人差があり、全ての学習者のために、これらの条件を満たす文をあらかじめ用意すると、提示できる例文は非常に限られて少なくなってしまう。そこで、青空文庫や新聞データなどの学習者向けでない大量のテキストデータを利用し、各学習者が検索時にレベル等を指定することで、個人のレベルに適った例文をできるだけ多く提示する例文表示ツールを開発した。

2. 例文表示ツール概要

図1が、例文表示ツールの構成図である。まず、生コーパスの各文を形態素解析、構文解析する。その結果をもとに、語彙難易度や文法難易度、構造の複雑さなどを計算し、難易度情報を付加して例文データベースを構築する。また、学習者全般が知らなくても漢字圏の学習者ならば知っているはずの語彙のリストや、理工系の学生なら知っているであろう語彙のリストなどを用意する。そして、検索時に、ユーザーの母語や専門領域をもとにフィルタリングする。これにより、漢字圏の学習者がユーザーの場合には難易度の高い漢語が検索結果に含まれるが、非漢字圏の学習者の場合には含まれないというような例文検索が可能になる。

次節以降で本ツールにおいて採用した単語難易度、文法難易度、構造難易度、ユーザー語彙フィルタリングについて説明する。

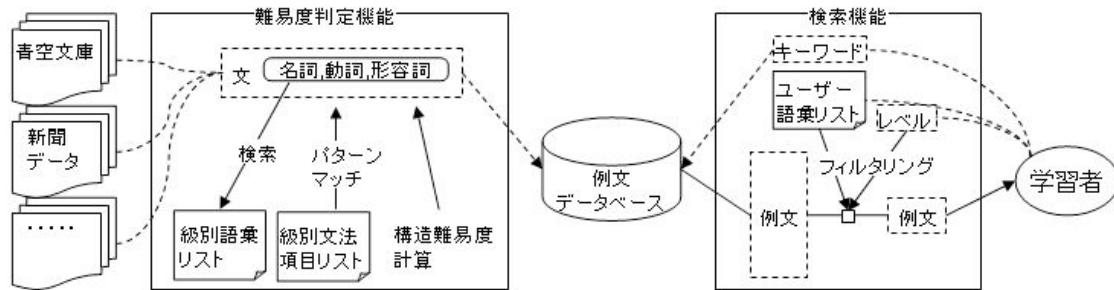


図1 例文表示ツールの構成図

3. 単語難易度

文中の全ての名詞、動詞、形容詞が日本語能力試験の何級であるかを日本語能力試験出題基準リスト(以後、「出題基準リスト」とする)を用いてチェックする(財団法人日本国際教育協会,1994)。出題基準語彙リストに含まれる語には、最も易しい4級から最も難しい1級までのいずれかの難易度が割り当てられる。出題基準リストにない語はさらに難しいという意味の0級を割り当て、文中に含まれている最高難度の語の級を文の単語難易度とする。ここで、出題基準リストの語を文字通り探してしまうと、表記のゆれが原因で語彙リストからの検索が失敗する場合がある。本ツールでは、その対応策として、出題基準リストの各語について、形態素解析ツール ChaSen の辞書に含まれるバリエーションを追加した。

4. 文法難易度

文中の機能語部分が、出題基準リストの何級の文法項目に相当するかをチェックし、語彙の場合と同様の方法で、最高難度の文法項目の級をその文の文法難易度とする。全ての文法項目をチェックしても当てはまらない機能語部分がある場合には、0級とする。

5. 構造難易度

語彙や文法が簡単でも構造が複雑で読むことが難しくなっている文は例文としてふさわしくない。小林ら(2007)は、学習者にとってよい例文を選択するにあたり、文の構造上の難しさを測ることは容易ではないとして、二つの擬似的な方法での文の難易度測定法を提案している。一つ目は、単語や品詞の接続だけによる文の難易度尺度で、大規模なコーパスから計算した単語n-gramによって文の単語あたりの確率を求め、それに応じて計算される評価値である。二つ目は、日本語文の係り受け構造の複雑さ(あるいは難しさ)に基づいた尺度で、日本語係り受け解析システムの出力結果に含まれる係り受けの有無の判断が難しかったことを示す数値を用いて推定される。本ツールで採用した尺度は、これらの尺度に比べ、より単純で、学習者にとって難解であると考えられる構造を直接数値化するものである。学習者にとって難解であると考えられる構造として以下の(1)(2)(3)を取り上げ、その難易度を定義する。なお、本ツールにおける構造難易度は、相対的に簡単なものが検索結果の上位に来るようにするために取り入れたものであり、語彙難易度や文法難易度のように、0~4の5段階評価を決

めるものではない。

- (1) 連体修飾節： 連体修飾節をなしている用言の個数を文全体の難易度とする。
- (2) 離れた文節に係る文節： 文中の係り受け関係について、係り受け距離として係り文節と受け文節の間の文節数を数える。その平均または合計を文全体の難易度とする。例えば、図2の3つの弧によって表されているような係り受け関係の場合、各係り受け距離は弧上に示した数値になる。
- (3) 多段に渡る修飾： ある文節の修飾の段数として、被修飾語を順にたどり、文末あるいは格要素に至るまでの文節数を数える。その平均または合計を文全体の難易度とする。図3の3つの弧によって示される修飾関係の場合の修飾の段数は、文節1が2、文節3が1となる。

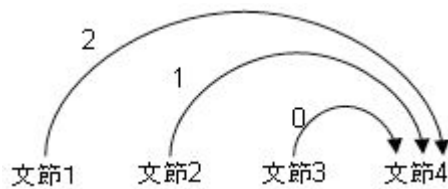


図2 係り受け距離の例



図3 修飾の段数の例

6. ユーザー語彙フィルタリング

通常、日本語コーパスは日本人が日本人のために書いた文章であり、平易な単語のみから構成されている文は少ない。例文検索結果を増やすためには、単語難易度による条件を何らかの方法で緩和する必要がある。本ツールでは、学習者個人によって既知の語彙リストが与えられた場合、それらの語彙を語彙難易度による選別時の対象外とする機能を設けた。図4にこの機能の具体例を示す。この例では、検索キーワード「権利」によって、「何を読むかは、読者の権利である」という文が検索されるが、語の下に示される級を見てみると、「読者」が1級の語であるため、ユーザーの指定した検索難易度の条件を満たさず、このままでは提示できない。もし、ユーザー語彙リスト中に「読者」があれば、その級を無視し、文の語彙難易度を2とすることで、ユーザーに提示することができる。

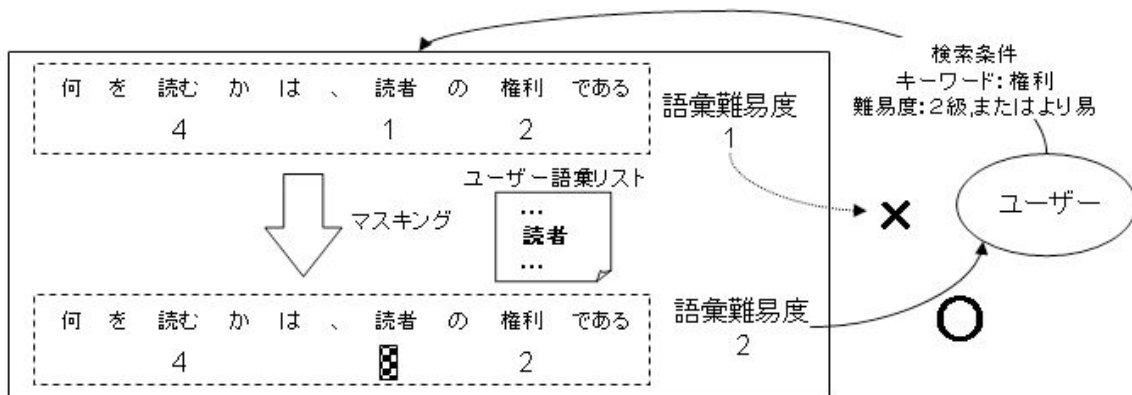


図4 ユーザー語彙フィルタリングの例

7. まとめ

日本語学習者が作文をする上で参照できる適切な例文を提示するツールについて説明した。このツールの手法では、まず、学習者のレベルに合わせて提示するために、語彙や文法的な難しさ、文の複雑さの三点を、あらかじめ数値として評価し、その評価値を含む例文のデータベースを構築しておく。そして、検索時に各学習者がレベルを設定することで、個人のレベルに合った例文の提示を可能とする。

8. 今後の課題

現在は利用しているコーパスが限られているが、さらにコーパスを入手し、ジャンル別、目的別に求められる例文を提示可能とする方法を検討する。構文難易度の計算法に関して、妥当性の検証や、よりよい計算法の検討が必要である。

謝辞

本研究は21世紀 COE プログラム「大規模知識資源の体系化と活用基盤構築」(代表者古井貞熙)および18年度科学研究費補助金基盤研究(B)「理工系日本語学習者のためのe-Learning 統合学習支援システム開発と評価」の一部補助によって行われた。

参考文献

Nishina, K. and Yoshihashi, K. (2007) *Japanese Composition Support System Displaying Co-occurrences and Example Sentences, Symposium on Large-scale Knowledge Resources (LKR2007)*, pp.119-122.

Totsugi, N. and Nishina, K. (2002) *Development of a System for Composition in Japanese by Utilizing the Dependency Structure Analyzer -Focusing on Adjectives*, 3rd *International Conference on Computer Assisted Systems for Teaching & Learning /Japanese*, pp. 67-70.

Kawamura, Y. (2002) *Implementation and Evaluation of a Computer-Based Reading Comprehension Program for Students from Non-Kanji Backgrounds, CASTEL/J'2002 Proceedings*, pp.147-150

石川慎一郎(2005)大学生英語学習者の受容語彙力と発表語彙力の関係-語彙サイズテストおよびエッセイ・コーパス分析に基づくアプローチ-中部地区英語教育学会紀要 pp.337-344

Tribble, C. and Jones, G. (1997) *Concordances in the Classroom*, Houston: Athelstan

独立行政法人国際交流基金, 財団法人日本国際教育協会(1994). *日本語能力試験出題基準*. 東京: 凡人社

奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座 (2007), ChaSen, from <http://chasen-legacy.sourceforge.jp/>

小林朋幸, 大山浩美, 坂田浩亮, 谷口雄作, 太田ふみ, Evans, N. 浅原正幸, 松本裕治 (2007), 日本語読解支援のための語義毎の用例抽出機能について, 言語処理学会第13回年次大会, pp.787-788.