

## 第二言語作文支援を目的とした句共起関係コーパスの構築手法

## A Construction Method of Co-Occurrence Corpus for Second Language Writing

陸 烽  
Hou RIKU

三好 康夫  
Yasuo MIYOSHI

高知大学理学部

Faculty of Science, Kochi University

あらまし：我々は先行研究において、第二言語作文支援のための句構造コーパスを用いた例文検索システムの開発を行った。本研究は、句構造情報に加えて係り受け解析した情報も利用して、書きたい文の伝えたい意図と一致する例文を見つけやすくすることが目的であり、現在、コーパスデータベースを開発中である。本発表では、句構造情報と係り受け情報から句共起関係コーパスを構築する手法について述べる。

キーワード：句共起関係，コーパス，係り受け解析，第二言語，作文支援

### 1. はじめに

本研究は第二言語学習の中でも作文に1つの文を書くことを対象とした学習支援を目指す。また、1つの言語ではなく、複数の言語でも支援できるように考えており、本研究の最終目標は、多言語に対応した第二言語学習者のための作文支援を実現することである。本稿では、英語・日本語・中国語の作文支援を目的とした句共起関係コーパスの構築手法について述べる。

### 2. 例文検索による第二言語作文支援

まず、学習者が作文するときの問題について考察する。母国語ではない言語で作文する時、自分が伝えたい意図を文として表現しても、その文が正しく表現できているかどうか、学習者は判断することができない。これは学習者自身が書いた文だけではなく、機械翻訳で作成した文であっても同様である。また第二言語作文においては、適切な語句選択が難しいという問題がある(1)。語句選択の問題は辞書の語義だけでは解決できないため、辞書の例文を参照することになるが、辞書に掲載された少ない例文だけでは語句選択の妥当性を判断することは困難である。

そこで、大量の例文をコーパスから検索し提示することで語句選択を支援する作文支援のアプローチがいくつか提案されている(2)(3)。我々は先行研究(2)において、句構造コーパスを用いた例文検索システムの開発を行った。句構造を用いることで、句の順序に基づく検索が行え、検索された例文を分類表示できることが特徴である。

### 3. 句共起関係コーパスの構築

#### 3.1. 句構造コーパスと句共起関係コーパスの違い

句構造コーパスは、言語を大規模に収集し、句構造解析により得られた句構造情報と共にデータとして保持したテキストデータである。句構造解析は構文解析の1つであり、他には係り受け構造を解析するための係り受け解析がある。句構造は意味が同じでも文の語

順が変わると、構造の形も変わる。一方、係り受け構造は、文の語順が変わった時でも構造の形は変わらないため、語順の自由度が高い日本語の構文構造表現に適しているとされている。本研究では、多言語への対応を予定しているため、係り受け構造も利用したいと考えた。そこで、句構造と係り受け構造の両方を保持した句共起関係コーパスを構築する。句共起関係を使って例文検索する際に検索条件が複雑にならないよう、使用する句共起関係パターンは次の4つに限定する。

- ① 動詞句とその主語となる名詞句
- ② 動詞句とその目的語となる名詞句
- ③ 名詞句とその補語となる句
- ④ 動詞句とその補語となる句

#### 3.2. 使用する構文解析器とパターン抽出手法

本研究で使用する構文解析器としては、現時点では日本語・英語・中国語の3か国語に対応させることを目指しているため、英語と中国語の構文解析にはStanford CoreNLP (<http://nlp.stanford.edu/software/corenlp.shtml>) を使用し、日本語の構文解析にはCaboCha (<https://code.google.com/p/cabocho/>) を使用する。

図1に示すように、Stanford CoreNLPでは、文の句構造と係り受け構造が得られる。このnsubjという共起関係子が、askedとstudentsがパターン①の共起関係であることを示している。そして句構造から、askedとstudentsを含む句の共起関係がわかる。Stanford CoreNLPを使った場合の句共起関係のパターン分類は、表1のように共起関係子に基づき行われる。

CaboChaは係り受け解析器で、句構造解析器ではないが、図2のようにchunkという句に相当する単位に文を区切ってくれるため、句の共起関係を得ることができる。そして、chunkに含まれるlink属性が共起関係を示している。ここでは、名詞句の助詞が「を」であるので、パターン②の共起関係であるとわかる。CaboChaを使った場合の句共起関係のパターン分類は、表2のようなルールに基づき行われる。

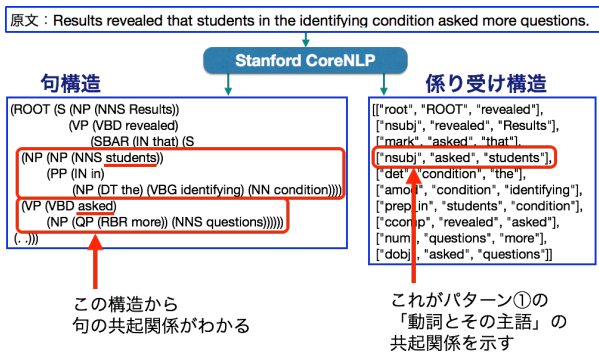


図1. Stanford CoreNLPの解析例

表1. 英語と中国語の句共起関係の4つのパターン

パターン	共起関係子(英)	共起関係子(中)
①動詞句とその主語となる名詞句	csubj, nsubj, nsubjpass	nsubj, nsubjpass
②動詞句とその目的語となる名詞句	ccomp, dobj, iobj, pobj	ccomp, dobj, pobj
③名詞句とその補語となる句	advcl, acmop, agent, mark, prep, tmod, xcomp	mark, prep, tmod
④動詞句とその補語となる句	adj, amod, appos, cop, nn, num, poss, prepc, preconj, rcmmod, xsubj	amod, cop, nn, num, prepc

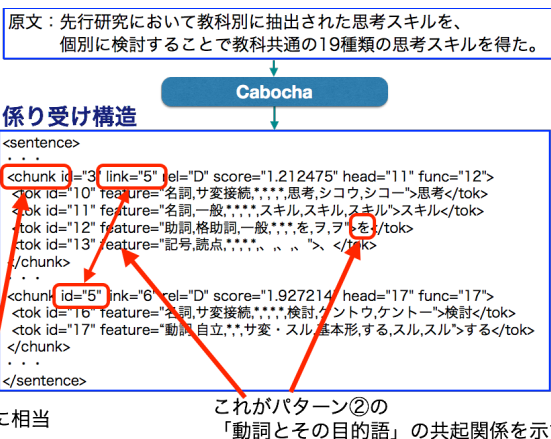


図2. Cabochaの解析例

表2. 日本語の場合句共起関係の4つのパターン

パターン	ルール
①動詞句とその主語となる名詞句	名詞のchunkのリンク先が動詞のchunkで、名詞のchunkの助詞が『は、が、では』の場合
②動詞句とその目的語となる名詞句	名詞のchunkのリンク先が動詞のchunkで、名詞のchunkの助詞が『を、へ、に』の場合
③名詞句とその補語となる句	任意の品詞のchunkのリンク先が名詞のchunkであるとき
④動詞句とその補語となる句	任意の品詞のchunkのリンク先が動詞のchunkで、①と②に当てはまらない場合

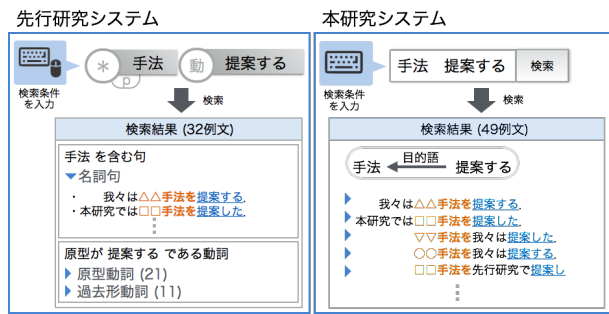


図3. 先行研究との検索インターフェースの違い

#### 4. 例文検索インターフェースのイメージ

前章のようにして得られた句共起関係を使うと、先行研究よりも伝えたい意図と一致する例文が検索しやすくなると期待できる。先行研究では、検索条件として、マウスとキーボードで検索プレートを組み立てる必要がある。図3の「手法」と「提案する」がそれぞれ検索プレートというもので、プレートの順序とつながりが検索結果に反映されるため、ユーザは検索条件を組み立てる際にはプレートの並びを考える必要がある。一方、本研究で構想している検索インターフェースでは、キーボードで2つの検索キーワード「手法 提案する」を入力するだけで検索できる。結果として、句共起関係のパターンに一致した文の一覧を見せることができる。先行研究では句の順序に基づいて検索しているため、検索結果の数にも違いが出る。この例の場合、目的語と動詞の間に主語などが入った文を先行研究の検索インターフェースのこの検索条件では見つけることができない。

#### 5. おわりに

本稿では、第二言語作文支援を目的とした例文検索ツールのためのコーパス構築手法について述べた。本研究のコーパスの特徴は、句構造情報と係り受け構造情報を用いた句共起関係を持っていることである。これにより、語順の自由度が高い日本語検索にも対応でき、書きたい文の伝えたい意図と一致する例文を見つけやすくなることが期待できる。今後は、句共起関係コーパスをデータベースに自動登録するプログラムを完成させ、例文検索インターフェースの設計と構築を行う予定である。

#### 参考文献

- (1) Someya, Y.: The State of Written Business Communication in English in the Japanese Workplace: A Questionnaire Survey, The 64th Annual Convention of the Association for Business Communication, (1999)
- (2) 三好康夫, 越智洋司, 金西計英, 岡本竜, 矢野米雄: 英作文支援における句構造情報を利用した用例検索ツール, 日本教育工学雑誌 27(3), 283-294, (2003)
- (3) 高松優, 水野淳太, 岡崎直観, 乾健太郎: 英作文支援のための用例検索システムの構築, 言語処理学会年次大会, 361-364, (2012)