

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 04 月 30 日現在

機関番号：34428

研究種目：若手研究（B）

研究期間：2010 ～ 2012

課題番号：22720225

研究課題名（和文）電子コーパス解析に基づく日英コロケーションの対応・非対応関係の分析

研究課題名（英文）Corpus-based studies of collocational equivalents between Japanese and English

研究代表者

後藤 一章（GOTO KAZUAKI）

摂南大学・外国語学部・講師

研究者番号：90397662

研究成果の概要（和文）：本研究課題では、日英パラレルコーパスを利用し、英語の頻出表現とその日本語訳が付与されたリスト、及び日本語の頻出表現とその英語訳が付与されたリストをそれぞれ作成した。自然言語処理分野の手法を援用し、コーパスにおける日本語表現の出現分布と英語表現の出現分布の比較から、対訳関係にある日英・英日表現を抽出した。得られた対訳表現リストは、日本人英語学習者が自然な英文を作成する際の有用な手がかりとなることが期待される。

研究成果の概要（英文）：This research proposes a method to create a bilingual phrasal list using NLP techniques. The similarity of each pair of English and Japanese phrases is calculated based on their frequency distributions in a bilingual corpus to extract bilingual phrasal equivalents automatically. A list of the extracted bilingual phrases can be expected to be a useful resource for Japanese learners of English.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,000,000	300,000	1,300,000
2011年度	600,000	180,000	780,000
2012年度	700,000	210,000	910,000
年度			
年度			
総計	2,300,000	690,000	2,990,000

研究分野：人文学

科研費の分科・細目：言語学・外国語教育

キーワード：教育工学・教材・教育メディア一般

1. 研究開始当初の背景

コロケーション (collocation) とは, commit a crime のような 2 語以上の単語結合のことであり, 一般的に, 日本人英語学習者にとってコロケーションの使用は困難だとされている。その原因として, 過度な母国語の干渉が指摘される。例えば, 「結論に至る」の英

語コロケーションは reach a conclusion であるが, 「発見に至る」の英語コロケーションは * reach findings ではなく, make findings である。こうした非対称性は, 西洋系言語とは大きな隔たりがある日本語を母語とする話者にとって, 直感的に理解しづらい部分と言える。つまり, 単に単語を増やし,

それを適当に組み合わせるだけでは適切なコロケーションは作成されず、コロケーションに特化した知識を意識的に増やす必要があることを意味している。こうした背景から、筆者は効果的にコロケーション学習を行えるリソース開発を行ってきた。例として、自然言語処理分野における統語解析技術を活用したコロケーション抽出手法を提案し、SVやVOなどの統語構造を特定した形での高精度なコロケーション抽出を実現した。この手法により、各コロケーションの学習優先度の推定が可能となり、効率の高いコロケーション学習リストが構築された。ただし、日本人英語学習者のコロケーション使用における問題点を改めて鑑みると、英語コロケーションのみが収録された学習リストでは不十分な可能性が指摘される。上述したように、日本語母語話者は、特に初中級レベルにおいて、英語使用時にも日本語の干渉を強く受ける。こうした母語干渉によって生じる誤りは、母語では自然なコロケーションであるという点において、その誤りに気づき難い項目である。また、問題をさらに複雑化させるのは、reach a conclusion や read one's mind (心を読む) のように、日本語と英語が逐語的に対応するコロケーションも一方で存在している点である。こうした日英コロケーションの対応・非対応関係における混乱を解消するためには、英語コロケーションと日本語コロケーションとの対応関係を体系的に整理し、その共通性や相異を学習者に明示的に示すことが肝要であると考えられる。特に、初中級レベルの学習者には、日本語に頼らず直接英文を組み立てることは難しく、ベースとなる日本語をいかにスムーズに英文化できるかが重要となる。日英コロケーション間の対応性を即座に識別できれば、よりスムーズな英作文が可能になると考えられる。

2. 研究の目的

まず、日英の各コーパスから抽出されたコロケーションの生起頻度や用法を観察し、日英コロケーションの使用実態を解明する。特に、高頻度で特徴的に生起するコロケーションを選定し、日英重要コロケーションリストを作成する。頻繁に遭遇する項目は重要性が高いという仮定から、学習優先度を生起頻度の高さで近似させることは不合理ではなからう。さらに、本リストに基づき、表現され

る意味が同様でありながら、日英コロケーション間で語彙や構造が異なる項目を特定する。例えば、read one's mind と「心を読む」は、意味と構造の両面において対応性が高い。一方、have a dream と「夢を見る」は、統語構造は対応しているが、語彙レベルにおいて差が見られる。また、「目を覚ます」と wake up では、統語構造にも語彙対応にも相異が見られる。こうした日英コロケーションの対応関係、非対応関係を、文体論や修辞学などの点から幅広く考察し、一定の体系化を目指す。

3. 研究の方法

自然言語処理分野における日本語解析技術を活用し、日本語コーパスからコロケーションを機械的に抽出する。そのためには、現在利用可能な複数の日本語統語解析器 (CaboCha や KNP など) を比較し、最も高精度かつ高機能な解析器を選定する必要がある。日本語統語解析により、文は形態素に分解され、さらに各形態素の文中における統語的役割 (主語・述語・目的語など) が特定される。これらの統語情報が得られれば、従来から取り組んできた英語コロケーションの抽出手法が応用でき、日本語コロケーションの機械的抽出が可能になると予測される。ただし、日本語には「～する」などのサ変動詞が存在する一方、前置詞句が存在しないなど、英語コロケーション抽出における枠組みをそのまま適用することはできない。そのため、様々な日本語文法理論を検証し、機械的日本語コロケーション抽出に最適な枠組みを模索する。コーパスからのコロケーション抽出が実現できれば、日本語における各コロケーションの典型性や重要度を生起頻度に基づいて推定することが可能となり、日本語母語話者のコロケーションの使用傾向を正確に反映したコロケーションリストが構築されることになる。最後に、収集された日本語コロケーションデータと英語コロケーションデータを比較し、日英コロケーションにおける共通点と相異点を解明する。特に、日英コロケーション間における構造的対応関係と語彙的対応関係に注目する。すなわち、「心を読む (read one's mind)」のように、語彙面、構造面が共に一致していると認められる項目は日英対応コロケーションとして分類され、「目を覚ます (wake up)」のように、日本語では VO 関係にあるものが英語で

は句動詞として扱われているなど、日英において統語構造が大幅に異なる項目は構造的な非対応コロケーションとして分類される。また、「発見に至る (*make findings*)」や「夢を見る (*have a dream*)」のような、構造的にはVO関係で一致しているものの、語彙レベルにおいて日英で相異が見られる項目については語彙的非対応コロケーションとして分類する。これらの評価作業は、機械処理と手作業によって複合的に行われる。

4. 研究成果

平成 22 年度は、日本語コロケーション抽出手法の確立、及び日本語コロケーションリストの作成を行った。それと同時に、英語コーパスからコロケーションの抽出も行い、頻出英語コロケーションを特定した。頻出日本語コロケーションリスト及び、頻出英語コロケーションリストは、我々が日常的にどのようなコロケーションを使用しているかを明らかにするものであり、英語教材や学習辞書などを作成する上で有用なリソースとなる。また、英日コロケーション間の比較を行うため、英語コロケーションリストに手作業で対訳の付与を行った。しかし、英語コロケーションに手作業で日本語を付与する作業は、極めて高コストであり、信頼性にも疑問が残る結果となった。

そこで、平成 23 年度では、英語コロケーションと日本語コロケーションを機械的に対応付け、対訳コロケーションリストを自動生成する手法について取り組んだ。対訳コロケーションを自動的に取得する手法が確立されれば、言語研究や英語研究、辞書編纂など、様々な分野に貢献をもたらすことが期待される。具体的な手法としては、パラレルコーパス（翻訳コーパス）を利用し、出現頻度の分布が類似する日英コロケーションを機械的に抽出した。これは、翻訳文上に各日英コロケーションが同時に発生する確率と、個別に独立して発生する確率に基づいて両者の類似度を算出するという方法であり、意味の問題に立ち入らず、形式的特徴のみで処理できるという利点がある。その結果、80%以上の正解率で英日対訳コロケーションが抽出された。その一部を表 1 に示す。

表 1

#	英語コロケーション	日本語コロケーション	評価
1	play role	役割ヲ 果たす	○
2	make effort	努力ヲ する	○
3	take place	起きる	○
4	revise law	改正案ヲ 提出スル	×
5	take measure	措置ヲ 取る	○
6	make decision	決断ヲ 下す	○
7	take step	一歩ヲ 踏み出す	○
8	take action	行動ヲ 取る	○
9	take part	参加スル	○
#	submit bill	国会ニ 提出スル	×

平成 24 年度では、英日対訳コロケーションの抽出精度を高めると共に、日英対訳コロケーションの抽出にも取り組んだ。英日対訳コロケーション抽出手法を応用することで、日英対訳コロケーションが高精度で抽出されることが明らかとなった（表 2）。一方、対訳抽出に失敗した項目は、日英コロケーション間で語彙的・構造的に大きく異なっており、学習者にとっても直感的に把握しづらい項目であった。これらの項目は、日本人学習者が特に注意を払うべき項目として位置付けられる。本研究課題期間では、構造的な非対応関係と語彙的非対応関係の細かい分類にまでは到達しなかったが、最終的な成果として、日英対応、非対応コロケーションリストが作成された点は、日本人英語学習者のコロケーション学習において重要な意義があったと考える。

表 2

#	本語コロケーション	英語コロケーション	評価
1	方針ヲ 固める	introduce system	×
2	影響ヲ 与える	have impact	○
3	役割ヲ 果たす	play role	○
4	姿勢ヲ 示す	show stance	○
5	考えヲ 示す	tell reporter	×
6	方針ヲ 決める	hold meeting	×
7	影響ヲ 及ぼす	have effect	○
8	法案ヲ 提出スル	submit bill	○
9	方針ヲ 明らか	announce plan	○
10	改革ヲ 進める	promote reform	○

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 5 件）

①後藤一章，パラレルコーパスを用いた日英

対訳コロケーション抽出の試み, 統計数理研究所共同研究レポート, 査読無, 298号, 2013, pp. 37-47.

②後藤一章, パラレルコーパスを利用した英日バイリンガル語彙リストの構築, e-Learning 教育研究, 査読有, 6巻, 2011, pp. 25-34.

③後藤一章, パラレルコーパスに基づく英日コロケーション対応付けの試み, 統計数理研究所共同研究レポート, 査読無, 264号, 2011, pp. 15-24.

[学会発表] (計6件)

①後藤一章, 日英コロケーションの対応性と非対応性, 言語研究と統計 2013, 2013/3/27, 統計数理研究所.

②GOTO KAZUAKI, A Statistical Approach to the Development of an English-Japanese Bilingual Collocation List, eLEX2011, 2011/11/11, Bled: Slovenia.

③後藤一章, 自然言語処理を利用した外国語教育の可能性, 外国語教育メディア学会 50周年記念全国研究大会, 2010/8/5, 横浜サイエンスフロンティア高等学校.

[図書] (計1件)

①後藤一章, ひつじ書房, テキストマイニング研究論文集 (仮), 2013, 印刷中.

[産業財産権]

○出願状況 (計0件)

名称:

発明者:

権利者:

種類:

番号:

出願年月日:

国内外の別:

○取得状況 (計0件)

名称:

発明者:

権利者:

種類:

番号:

取得年月日:

国内外の別:

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

後藤 一章 (GOTO KAZUAKI)

摂南大学・外国語学部・准教授

研究者番号: 90397662

(2) 研究分担者 ()

研究者番号:

(3) 連携研究者 ()

研究者番号: