

## 英文科学技術文における前置詞を伴う動詞句の決定

2L-6

柴田 誠<sup>†</sup> 丸木 健次<sup>†</sup> 西村 真<sup>‡</sup> 竹田 正幸<sup>†</sup> 松尾 文碩<sup>†</sup>  
<sup>†</sup>九州大学大学院システム情報科学研究科 <sup>‡</sup>松下通信工業

## 1. まえがき

著者らは、英文科学技術文を対象に動詞句の構造決定と名詞句の範囲決定の研究を行っている。動詞句の構造決定において、主動詞を97.7%の確率で決定する方法を開発した<sup>1)</sup>。

文中の動詞を決定した後の文の統語構造を決定する上での重要な問題の一つとして、前置詞句の係り受けの問題がある。前置詞句が文の統語構造に参与するのは次の三つの場合である。

- 1) 前置詞句が名詞句の一部を構成する場合。

例文

The authors find no evidence for such a twist.

- 2) 前置詞句が文の動詞を修飾する場合。

例文

They describe this algorithm in detail.

- 3) 前置詞が動詞の統語構造の一部を構成する場合。

次の例文は、compare A with B という構文を構成して、AとBとを比較することを意味する。

The author briefly compares Quick BASIC with Turbo BASIC.

本稿では3の場合の、実際の文章における生起と、その動詞と前置詞の依存関係を調査した結果を報告する。

## 2. 前置詞を伴う動詞のパターンの抽出

前置詞を伴う動詞のパターンは、次の二つに大別することができる。

- compare A with B のように動詞と前置詞の間に、不定な目的語を含む他動詞のパターン

Decision on Verb Phrase with Preposition in Scientific and Technical Documents

Makoto Shibata<sup>†</sup>, Kenji Maruki<sup>†</sup>, Makoto Nishimura<sup>†</sup>, Masayuki Takeda<sup>†</sup> and Fumihiro Matsuo<sup>†</sup>

<sup>†</sup> Graduate School of Information Science and Electrical Engineering, Kyushu University, Hakozaki, Fukuoka 812-81, Japan

<sup>‡</sup> Matsushita Communication Industrial, Minatokita, Yokohama 223, Japan

- differ from A のように動詞と前置詞の間に、不定な目的語を含まない自動詞のパターン

これらの構文を、研究社英和中辞典第5版から、他動詞の〔+目+前+(代)名〕と記述があるものを前置詞を伴う他動詞のパターンとして、自動詞の〔+前+(代)名〕と記述があるものを前置詞を伴う自動詞のパターンとした。表1に、パターン、動詞、前置詞の数をそれぞれ示す。

表1 前置詞を伴う動詞のパターン

	自動詞のパターン	他動詞のパターン
パターンの数	2854	3361
動詞の数	1412	1724
前置詞の数	48	47

## 3. INSPEC テープにおける前置詞を伴う動詞のパターンの調査

## 3.1 前置詞を伴う自動詞のパターン

前置詞を伴う自動詞のパターンについては、10年分のINSPECテープ(10,482,511文)を対象に、単語列のマッチングを行った<sup>4)</sup>。その結果、高頻度上位100のパターン中、99のパターンがマルコフストリング<sup>2)</sup>であることがわかった。

## 3.2 単純名詞句

前置詞を伴う他動詞のパターンについての調査結果の前に、単純名詞句<sup>3)</sup>について簡単に説明する。

単純名詞句とは、被修飾名詞<sup>3)</sup>と、その前方修飾語のみからなる名詞句であり、以下の例文における下線を施した単語列が単純名詞句である。

The paper describes a language to specify such models.

表 2 高頻度上位 10 の他動詞のパターン

他動詞のパターン	パターンの頻度	単純名詞句数ごとのパターンの頻度							
		1	2	3	4	5	6	7	8
present A to	1950	1286	379	177	63	26	13	4	2
make A of	1517	1148	324	44	1	0	0	0	0
allow A to	1479	1241	171	40	19	7	1	0	0
describe A to	1474	733	540	144	35	13	6	3	0
require A of	1417	1225	166	22	3	1	0	0	0
give A to	1401	1121	179	70	23	6	2	0	0
show A to	1091	579	326	134	43	4	3	2	0
have A on	1072	984	64	17	7	0	0	0	0
have A with	815	542	225	40	5	3	0	0	0
play A in	801	779	19	3	0	0	0	0	0

名詞句とは、単純名詞句が前置詞、接続詞などによって結合したものであると考えることができ、単純名詞句は名詞句の基本単位であると考えられる。

### 3.3 前置詞を伴う他動詞のパターン

次に示す例文は、compare と with が他動詞のパターンを作っている。下線は、その単語列が単純名詞句であることを示す。

They compare the estimates derived from LINPAC with those derived from LINMAP.

目的語は、the estimates derived from LINPACであり、二つの単純名詞句から構成された名詞句である。動詞 compare に二つの単純名詞句が生じたあと、前置詞 with が出現している。一方、第 1 節 3) で示した例文では、with の前の単純名詞句は一つである。

そこで、動詞と前置詞との間に生起する単純名詞句の個数を INSPEC テープ 10 年分を対象に調査した。ここで、単文における動詞が高い確度で決定できることから、対象を単文に絞るために、科学技術抄録文において重文・複文をつくる、接続詞、疑問詞、セミコロン、コロンを含まない文を疑似単文として抽出した。この疑似単文のうち、単純名詞句の決定において問題となる現在・過去分詞形を含まない約 34 万文を調査対象とした。なお、単純名詞句は著者らの研究室で現在研究中の、単純名詞句の範囲決定の手続きで擬似的に決定したものを利用し、動詞も、研究室で開発した動詞決定法で擬似的に決定したものを利用した。

他動詞のパターンの調査結果を表 2 に示す。表中の、単純名詞句数ごとの頻度は、各パターンについて、それぞれ動詞と前置詞との間に含まれる単純名詞句の数ごとのパターンの頻度を示す。今回の調査では、動詞

の受動態を除外しているため、動詞と前置詞との間に単純名詞句を全く含まない場合は、自動詞のパターンを形成していると考えられる。そこで、今回は動詞と前置詞との間に単純名詞句を 1 個以上含むものについてのみ調査を行った。他動詞のパターン中、1 回以上生じたパターンは、842 であり、全体の約 1/4 であった。また、動詞の数は 593 であった。

## 4. むすび

本稿では、前置詞を伴う動詞の統語構造について、他動詞のパターンの生起頻度と、動詞と前置詞の結び付きを調査した結果を報告した。今後は、実際の意味に基づいたパターンの検証を行っていく予定である。

なお、本研究は、一部文部省科学研究費補助金 (# 07558162) の援助により行った。

### 参考文献

- 1) Nishimura, M., et al. : Determination of Verb Phrase in Scientific and Technical Documents, *Proc. Natural Language Processing Pacific Rim Symposium '95*, pp. 95-100 (1995).
- 2) Takeda, M., Matsuo, F. : Markov String Grammar, *Memoirs of the Faculty of Engineering, Kyushu University* 55(3), pp 279-284 (1995).
- 3) Himukashi, Y., et al. : Identification of Noun Phrase in Scientific and Technical Documents, *Proc. Natural Language Processing Pacific Rim Symposium '95*, pp. 492-497 (1995).
- 4) 西村, 大木, 竹田, 松尾: 英文科学技術文における前置詞を伴う動詞の統語構造, 情報処理学会第 52 回全国大会講演論文集 (1996).