

英文科学技術文における前置詞を伴う動詞の統語構造

7B-3

西村 真[†] 大木 一成[‡] 竹田 正幸[‡] 松尾 文碩[‡]
[†]九州大学大学院工学研究科 [‡]九州大学工学部

1. まえがき

著者らは、英文科学技術抄録文の主動詞を97.7%の確率で決定する方法を開発した¹⁾。

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関与するのは次の三つの場合である。

- 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.

- 前置詞句が文の動詞を修飾する場合。

例文

They describe this algorithm in detail.

- 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
- 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

Syntactic Structure on Verb with Preposition in Scientific and Technical Documents

Makoto Nishimura[†], Kazunari Ooki[‡], Masayuki Takeda[†] and Fumihiro Matsuo[‡]

[†] Division of Engineering, Graduate School, Kyushu University 36, Hakozaki, Fukuoka 812-81, Japan

[‡] Faculty of Engineering, Kyushu University 36, Hakozaki, Fukuoka 812-81, Japan

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

10年分のINSPECテープ(10,482,511文)において、前置詞を伴う自動詞のパターンが、それぞれのどの程度現れるか生起頻度を調べた。

生起の調査は、単語列のマッチングに基づいて行い、意味的な判断は行っていない。例えば、パターン go

表2 高頻度上位10の自動詞のパターン

自動詞のパターン $w_1 w_2$	頻度	$p(w_2)$	$p(w_2 w_1)$
lead to	96719	0.02193	0.76888
compare with	93303	0.01037	0.47129
apply to	86511	0.02193	0.48981
consist of	85414	0.05501	0.96582
depend on	77424	0.00740	0.89625
result in	67101	0.02274	0.08866
associate with	64635	0.01037	0.72890
relate to	59980	0.02193	0.62250
deal with	58854	0.01037	0.93842
correspond to	49520	0.02193	0.52265

表3 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	動詞の頻度	前置詞の確率比			
			距離1	距離2以内	距離3以内	距離1以上
allow A to B	14149	28608	6.03339	7.23307	7.63056	7.85484
provide A for B	10989	43215	4.34596	5.10545	5.39690	5.56033
give A to B	7308	29543	3.05992	3.55288	3.77759	3.92865
require A of B	7308	16502	1.67427	2.17260	2.35833	2.48589
make A of B	6424	22985	0.63790	1.14489	1.38862	1.56885
present A to B	6385	45050	1.22083	1.73131	1.99254	2.25095
have A on B	6424	71800	3.12881	3.41758	3.54932	3.63930
describe A to B	5552	47846	0.80229	1.37156	1.62549	1.84291
apply A to B	5548	15520	4.46982	5.27516	5.52997	5.67733
show A to B	5259	79162	0.49474	0.81213	0.95156	1.05508

to は, go to, goes to, going to, went to, gone to と展開してマッチングを行い, それぞれの頻度を総計してパターン go to の頻度としている。

また, 動詞と前置詞の単語の生起確率, 条件付き確率を10年分のINSPECテープから計算し, 単純マルコフ連鎖であるかどうかの調査も行った。生起の高頻度順に並べた調査結果の一部を表2に示す。表に示す上位10では, すべてのパターンがマルコフストリング²⁾であり, 上位100でも, 99のパターンがマルコフストリングであることがわかった。マルコフストリングでなかった場合は, model in のみである。

3.2 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, その単語列が基本名詞句³⁾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, 二つの基本名詞句から構成された名詞句である。動詞 compare に二つの基本名詞句が生起したあと, 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 の前の基本名詞句は一つである。

ここでは, 動詞のあとに生起する基本名詞句の個数を距離と定義して, 前置詞の現れやすさと距離の関係を, INSPECテープ5年分で調査した。and / or を含む文は連言問題に関係して扱いが困難であり, ここでの調査対象から除外している。なお, 基本名詞句は著者らの研究室で現在研究中の, 基本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3に示す。表中の, 動詞の頻度は, その動詞が能動態で出現した頻度を表し, パターンの頻度は, 動詞が能動態で出現したときに対応する前置詞が距離1以上で出現した頻度を表す。 p_1 を, 動詞が出現したときに対応する前置詞が距離1以上 n 以内に出現する確率, p_2 を, 動詞句と基本名詞句以外の単語内で, その前置詞が出現する確率であるとする, 前置詞の確率比は, p_1/p_2 である。ある同じ距離までについて着目した場合, 前置詞の確率比が高いパターンの方がその距離内において, より動詞と前置詞の結び付きが強いと考えられる。

4. むすび

本稿では, 前置詞を伴う動詞の統語構造を, 自動詞のパターンと他動詞のパターンに分けて生起頻度と, 動詞と前置詞の結び付きを調査した結果を報告した。

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).