

英文科学技術文における動詞の意味的分類

3C-6

柴田 誠[†] 河崎 裕司[‡] 竹田 正幸[†] 松尾 文碩[†][†]九州大学大学院システム情報科学研究科 [‡]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ている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。そこで、動詞と前置詞の依存関係の調査を行い、compare A with BのAとBのように、実質的に目的語を2個とる動詞を調査した。その結果をもとに、動詞の分類を行った。

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Aのように動詞と前置詞の間に、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。

このうち、今回は前述のように他動詞のパターンを対象とした。また、これらの構文は、学習研究社社アンカー英和辞典第2版から抽出し、パターン数は294、動詞数は198であった。

3. INSPECテープにおける前置詞を伴う他動詞のパターンの調査

3.1 調査対象

今回は、1984年から1993年までの10年分のINSPECテープを調査対象とした。ここで、単文における

Classification of Verb in Scientific and Technical Documents
Makoto Shibata[†], Yuji Kawasaki[‡], Masayuki Takeda[†]
and Fumihiko Matsuo[†]

[†] Graduate School of Information Science and Electrical Engineering, Kyushu University, Hakozaki, Fukuoka 812-81, Japan

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

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³⁾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。

3.2 他動詞のパターンの調査

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。調査結果は表1に示す。他動詞のパターン中、1回以上生じたパターンは199であり、全体の約2/3であった。また、動詞の数は146であった。

表1 他動詞のパターンの頻度

パターン	頻度	パターン	頻度
present B to	1950	apply A to	600
require A of	1417	compare A with	493
show B to	1093	provide A with	488
present A with	747	report A to	365
examine A in	701	offer B to	365

4. 動詞の分類

4.1 TO不定詞や成句との重複

次に、1回以上生じたパターンについて、実際にパターンを構成しているか否か、高頻度のものから順に実際に文章を目で見してみた。その結果、以下の傾向が見られた。

- 前置詞がtoで、その直後に動詞の原形が来る場合は、そのほとんどがTO不定詞である。
- 前置詞が成句の一部に重複している場合は、成句であることが多い。

そこで、各パターンについて、TO不定詞や成句と重複する割合を調査した。まず、TO不定詞については、表2より、パターンによって重複の割合が大きく変わることが分かる。

表2 TO不定詞との重複の割合

パターン	割合	パターン	割合
write B to	0.818	present B to	0.479
cause B to	0.773	lead A to	0.433
help A to	0.769	address A to	0.433
recommend B to	0.714	offer B to	0.419
adjust A to	0.615	lose A to	0.400
introduce A to	0.555	direct A to	0.397
expect A of	0.540	expect A from	0.389
drive A to	0.533	report A to	0.375
afford B to	0.500	help A with	0.368
help A in	0.488	match A with	0.364

次に成句について、まず成句(研究社英和中辞典第5版より抽出)を以下の3種に分類した。

- 1) 強連結マルコフストリング²⁾の成句(1843種)
- 2) 強連結ではないがマルコフストリング²⁾ではある成句(756種)
- 3) 非マルコフストリング²⁾の成句(1995種)

1の成句から順に重複するものを除外していき、割合を調査した結果を表3に示す。この表から、強連結マルコフストリング以外の成句との重複はほとんどないことが分かる。理由としては、科学技術文ではあまり用いられないものや、more and moreのように接続詞などを含んでいるため、疑似単文では出てこないものが多いことが挙げられる。

表3 成句との重複の割合

パターン	成句との重複の割合		
	成句1	成句1,2	全ての成句
lose A to	0.400	0.400	0.400
introduce A by	0.333	0.333	0.333
express A to	0.273	0.273	0.273
estimate A at	0.263	0.263	0.263
pass B to	0.214	0.214	0.214
drive A to	0.200	0.200	0.267
start A in	0.200	0.200	0.200
show B to	0.160	0.161	0.165
incorporate A in	0.156	0.156	0.156
treat A to	0.154	0.154	0.154

4.2 パターンの分類

今回は以下の方針でパターンを構成するか否かを調査した。

- 1) 30回以上生じたパターンに限定する。
- 2) 意味的に見て科学技術文で用いられそうなものから調査する。
- 3) 前置詞がtoのパターンと、それ以外のパターンに分け、前置詞がtoの場合は、TO不定詞と重複する文を調査対象から除外する。
- 4) 強連結マルコフストリングの成句と重複する文を調査対象から除外する。
- 5) 各パターンにつき、25文づつ実際に文章を見て判定する。

調査対象のパターンは、前置詞がtoのものが20、それ以外のパターンが20で、計40パターンであった。また、動詞の数は38だった。調査結果を表4に示す。

表4 パターンか否かの判定

パターン	○	△	×	合計
前置詞がtoのパターン	10	0	10	20
前置詞がto以外のパターン	9	6	5	20

○:全てパターンだったもの

×:全てパターンではなかったもの

△:パターンの場合と、そうでない場合があったもの

5. むすび

本稿では、他動詞のパターンの生起頻度を調査し、高頻度のパターンについて、実際にパターンを構成するものと、そうでないものに分類した結果を報告した。

パターンを構成するものについては、名詞の分類にも応用できると考えられる。例えば、compare A with Bでは、AとBは比較できる名詞句である。今後は、これらの結果から名詞の分類、さらにはオントロジーの自動作成⁴⁾への応用を考えている。

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) 柴田, 辻, 竹田, 松尾: 英文科学技術文におけるオントロジー自動作成の試み, 情報処理学会第51回全国大会講演論文集(1995).