

英文科学技術文における成句について

7B-5

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

1. まえがき

英文科学技術抄録文においてイディオムの形で生起する単語列が、実際にイディオムとして用いられているかどうかを INSPEC テープの抄録文を対象に調査した結果について述べる。

2. 調査対象

研究社英和辞典第5版上に記載されている成句のうち、ここではその第1語目が動詞でないものをイディオム、それらイディオムと同一な単語列をイディオム性単語列と呼ぶことにする。10年分の INSPEC テープから次のように選択したイディオム性単語列について実際にイディオムかどうかを調査した。

- 1) as ... as possible や, in a person's absence のような途中に ... や, a person's などの不特定部分を含まないもの。そのようなイディオム性単語列は 4,722 種類あった。
- 2) 異なり単語についての生起確率と条件付き確率を INSPEC テープ 10 年分から求め、それに基づき強連結マルコフストリング¹⁾であると判明したもの。そのようなイディオム性単語列は 1,918 種類あった。非マルコフストリングや、弱連結マルコフストリングであるものは、イディオムをなさないか、生起頻度が低いため、調査対象とはしなかった。
- 3) 累積相対頻度が 8 割程度となる高頻度のもの。イディオム性単語列の頻度の順位と累積相対頻度の関係を図 1 に示す。

その結果、調査対象となった単語列は 70 種類であった。

Idiomatic Phrase in Scientific and Technical Documents
 Kazunari Ooki[†], Makoto Nishimura[‡], Masayuki Takeda[†]
 and Fumihiko Matsuo[†]

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

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

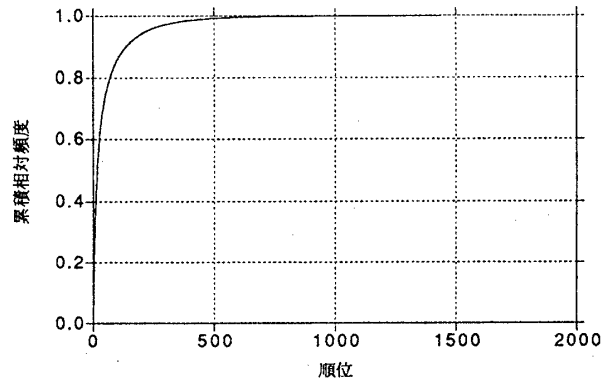


図 1 イディオムの累積相対頻度

その中で高頻度のものを表 1 に示す。ただし、表中の compared with 及び compared to は第 1 語目が動詞であるが、辞書上の成句 (as) compared with [to] から展開されるものであるため、調査対象に含めた。

3. 判定手順

10 年分の INSPEC テープ 2,408,118 文献の抄録文 10,482,511 より、対象である 70 種類のイディオム性単語列それぞれに対してそれが生起する抄録文を 30 文抽出し、イディオムかどうかを判定した。ただし、次のような場合は除外した。

- 複数の調査対象がお互いに一部重なり合っている場合。例えば、compared with respect to は、調査対象である compared with, with respect to のいずれからも除外する。
- ある調査対象が別の調査対象の一部となっている場合。例えば、調査対象の as well as は、同じく調査対象の as well から除外する。

4. 調査結果

前章で抽出した抄録文における単語列がイディオムであるかどうかを判定した結果、すべての 30 文がイディオムであると認められた単語列は調査対象の 70

表 1 高頻度上位 35 の強連結マルコフストリングの
イディオム

イディオム	生起頻度	累積相対頻度
due to	149091	0.059260
as well	125017	0.108950
as well as	116334	0.155190
up to	93874	0.192502
in terms of	89999	0.228275
in order	89183	0.263722
compared with	88492	0.298896
in order to do	84341	0.332419
in addition	56603	0.354917
by means of	54928	0.376749
a number of	51446	0.397198
in particular	46375	0.415631
that is	45239	0.433612
with respect to	44495	0.451297
compared to	42007	0.467994
could be	38847	0.483435
because of	35404	0.497507
together with	33954	0.511003
more than	33883	0.524470
according to	33669	0.537853
so that	31297	0.550292
as a result	28695	0.561698
in detail	26257	0.572134
close to	25582	0.582303
of all	25375	0.592388
for all	25257	0.602427
at least	23445	0.611746
independent of	22777	0.620799
in addition to	20184	0.628822
in general	19907	0.636734
in all	17843	0.643827
first time	16776	0.650495
the first time	16595	0.657091
for the first time	15827	0.663381
for example	15012	0.669348

種類のうち、50種類(71%)であった。特に3単語以上の単語列については対象となった21種類のうち、19種類(90%)までがその中に含まれていた。これらのイディオム性単語列は抽出しなかった抄録文においてもほとんどイディオムであると思われる。

一方イディオムとは認められなかったイディオム性単語列には、次のような傾向があった。

- 口語イディオム。
表2に示す5例があった。口語イディオムと同一の単語列は、科学技術文にはイディオムとしては、まず生起しないのではないかとと思われる。
- 最後尾の単語に形容詞語義が存在する。
表3に示す8例があった。表の()内は、最後尾

表 2 口語イディオム

イディオム	意味
could be	たぶん, そうかも
of all	数ある... の中で
and how	非常に, とても
kind of	ほとんど, 大体, やや
the end	ひどいもの[こと], 忍耐の限界

表 3 最後尾の単語が形容詞語義を持つイディオム

イディオム	意味
in general (名)	一般に, 一般の
or more (副)	或いはそれ以上, 少なくとも
in parallel (名)	平行に, 同時に
at all (代)	少しも... でない, 少しでも
for all (形) *	... にもかかわらず
of all (形) *	数ある... の中で
in all (代) *	全部で, 合計で
in one (名) *	一致して

の単語の品詞名を表わす。

これらは、最後尾の単語が形容詞となってその直後に名詞や名詞句が続くことが多く、イディオムでなくなる傾向が強かった。例えば、in parallel の場合 ... in parallel processing systems. という文があり、この parallel は明らかに形容詞である。

特に、表において * を付した単語列は全くイディオムとして生起していなかった。

したがってイディオムかどうかを判定する際に、その単語列の直後の構造を確認して文法的なつながりを明らかにする必要がある。

5. むすび

本稿では、強連結マルコフストリングであるイディオム性単語列について、実際それがイディオムとして生起しているかどうかを判定調査した結果を報告した。

今後の課題としては、イディオムの中に現れる単語の生起確率と条件付き確率の比とその単語の品詞との関係を明らかにすることである。

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) Takeda, M., Matsuo, F.: Markov String Grammar, Memoirs of the Faculty of Engineering, Kyushu University 55(3), 279-284 (1995).