

1N-1

英文科学技術抄録文における 高頻度主題記述動詞の統語情報

早田龍弘 楠本典孝 竹田正幸 松尾文碩
九州大学 工学部

1. まえがき

著者らは、英文科学技術論文の抄録文理解システムの研究を行なっている。ここでの“理解”とは、抄録文を、専門語句をノードとしノード間の関係をラベル付きアークであたえる意味ネットワークとして表現することをいう。これまでの研究から、原形約1,000語の機能語についての構文・意味辞書をつくれれば、専門語句の切出しとアークのラベル付けができる可能性があることがわかっている[1]。本稿では、抄録文中に高い頻度で出現し、かつ論文の主題の記述にかかわる動詞に関する用例の調査結果を報告する。また、それらの動詞を含む抄録文については、統語情報などによって構文的な曖昧さがほとんど解消できることを示す。

2. 高頻度主題記述動詞

機能語動詞のうち抄録文中で高頻度で出現して、かつ主題を表わすと思われる動詞を高頻度主題記述動詞と呼ぶことにする。例えば表1の show, present, study,

表1 高頻度機能語(出現頻度順上位40)

the	on	from	<u>describe</u>
of	that	or	<u>study</u>
be	have	result	much
and	use	model	<u>discuss</u>
a	an	can	give
in	as	author	one
to	it	these	obtain
for	at	<u>present</u>	state
with	this	<u>show</u>	also
by	which	used	effect

Syntax about Highly Frequent Verbs for Describing Subjects
in Abstracts of Scientific and Technical Literature
Tatsuhiko SOHDA, Noritaka KUSUMOTO,
Masayuki TAKEDA and Fumihiko MATSUO
Faculty of Engineering, Kyushu University

describe, discuss などである。ここでは INSPEC テープの抄録文(110,312文)を対象にこの五つの主題記述動詞の用例について調査した。これら五つの動詞は、63.9%の抄録中に出現する。本稿では、紙面の都合上、このうち最も多様な主部をもつ show についてだけを述べる。

3. show の用例

show の活用形別出現文数を表2に示す。

表2 show の活用形別出現文数

活用形	show	shows	showed	shown	showing
文数	1,787	1,002	432	2,727	202

show, shows を含む文のうち主部をもつのは2,758文である。

表3 主部の分類(show, shows)

分類	句	割合
人	the author(s)	14.1%
	the present author(s)	
代名詞	he she	11.2%
	it they	
	this these	
結果	(the) result(s) of ...	12.5%
	(the) result(s) obtained result(s)	
解析	... analysis	3.6%
	... analyses	
	... analysis of ...	
記事	the paper this paper	1.7%
	the article this article	
	the present paper	
専門語句		56.9%

show についての調査を簡単にまとめると次のようになる。

- 表 3 に示したように主部のうち半分は専門語句である。
- 主部と動詞の間に副詞句が入る文は 12.7%あり、その語句はほとんど機能語の副詞 (also, clearly, then, however, often など) である。
- show が名詞として用いられている文は 0.2%しかない。
- showed の主部についても show, shows と同様なことがいえる。
- shown を含む文のうち半分は “it is shown that …” という形の文である。残りの 1,338 文の用法を以下の表 4 に示す。
- showing については、ほとんど分詞構文または動名詞として用いられ、現在進行形の文は 1.5%しかない。

表 4 shown の用法

動詞的用法	完了	14.5%
	受け身	81.8%
	主部が省略された副詞節	2.2%
形容詞的用法	前位用法	0.0%
	後位用法	1.5%

4. 構文的な曖昧さの解消

構文的な曖昧さの解消される過程を以下で示す。下の例では機能語は斜体で、動詞候補は数字付きの下線で表わしている。

例 1 The author *discusses*₁ different *classes*₂ of *mixed*₃ valence *compounds*₄ *studied*₅ by chemists to *show*₆ how the *rate*₇ of electron *transfer*₈ *plays*₉ a crucial role.

構文解析のためには、動詞の決定が最重要である。最初に、動詞の活用形別に動詞候補の単純な判別法を述べる。判別する動詞候補を V とする。

1. V が原形るとき

- 助動詞 + V, have to + V の場合、V は動詞候補である。
- to + V は to 不定詞である。
- 冠詞 + V, 形容詞 + V, 前置詞 (to を除く) + V の場合、V は名詞あるいは形容詞である。
- その他の場合は判別不能。

2. V が三人称単数現在形るとき

- 冠詞 + V, 形容詞 + V, 前置詞 + V の場合、V は名詞の複数形である。

- その他の場合は判別不能。

3. V が過去、過去分詞形るとき

- be + V, have + V のときは動詞候補である。
- 冠詞 + V, 形容詞 + V, 前置詞 + V の場合、V は過去分詞の形容詞的用法である。
- その他の場合は判別不能。

4. V が現在分詞形るとき

- be + V の場合は動詞候補である。
- その他の場合は動名詞、分詞構文または形容詞的用法である。

上記の動詞の活用形と品詞情報による単純な判別法により、例 1 の下線部 3,6,7 は動詞候補から除外される。残りの下線部については、この方法で判定することができない。そこで、機能語に関する統語情報を用いることになる。たとえば、study の過去分詞形が、

名詞句 + studied

の形で、名詞句を後位修飾している場合、被修飾語である名詞句を主語、studied を過去形の動詞とみなすこともできる。しかし、調査の結果より、動詞 study が能動態で用いられる場合、その主部は、ほとんどが著者を表わす語句 (the author など) に限られていることがわかっている。したがって、studied の前にそれ以外の名詞句がきた場合、この studied は過去分詞で後位修飾をしていることがわかる。このように動詞候補を除外してゆくと下線部 1,2,8,9 が動詞候補として残る。

非機能語の動詞の場合、辞書にもたないの単純な判別しか行なえない。しかし、非機能語は専門語句として使われることが多いため、機能語の動詞が動詞候補として残っている場合、動詞であることはほとんどない。よって非機能語を動詞候補から外すと下線部 1,9 が残る。

複数の機能語の動詞候補が残っている場合、一般的に高頻度主題記述動詞の方が動詞になりやすいことが今回の調査でわかった。そこで、下線部 1 の discusses が動詞であると判定する。

5. むすび

五つの主題記述動詞の用例を調査し、意味・統語情報を抽出した。また、この統語情報から構文的な曖昧さが解消される過程を述べた。この結果は、情報検索システムにおける主題検索の質の向上に役立てることもできる。今後は、引き続き機能語動詞を調査し、抄録文理解システムの実現を図る予定である。

なお、本研究は一部文部省科学研究費補助金 (重点領域「知識科学」) により行なった。

参考文献

- [1] 西村利浩：英文科学技術文献用機能語辞書の作成と評価，九大電子修士論文，1989。