

単語の品詞とその被修飾度および 前方修飾度との関係について

1 R - 2

丸木 健次[†] 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

1. まえがき

英文科学技術抄録文を論理式へ変換する第一段階として、原子論理式の述語記号に動詞を項に名詞句をそのまま単語列としてあてる方法が考えられる¹⁾。英文科学技術文の名詞句の範囲決定のために、まず一つの被修飾名詞²⁾とその前方修飾語³⁾からなる単純名詞句を決定する。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 values of the registration parameters are automatically calculated by maximizing an integer similarity measure selected for robustness.

名詞句は、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。

単純名詞句は一つの被修飾名詞と前方修飾語からなる単語列である。著者らは、被修飾名詞と前方修飾語を決定するために、それぞれ被修飾度と前方修飾度という指標を提案している²⁾³⁾。ここでは、単語の被修飾度と前方修飾度について市販辞書に記載されている品詞との関係を示す。

2. 被修飾度

被修飾度とは単語が被修飾名詞として生起する相対頻度（確率）として定義している。被修飾度を算出するためには、被修飾度に関する大量のコーパスが必要であるがその作成には多大な労力が必要となる。そこで1984年から1993年の10年分のINSPECテ

Relation between Modificant and Preceding Modifier Degree and Part of Speech

Kenji Maruki[†], Yuji Kawasaki[†], Masayuki Takeda[†] and Fumihiko Matsuo[†]

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

ープ2,408,118文献の抄録文10,482,511文を対象に疑似的に被修飾名詞を決定し被修飾度を算出した²⁾。この疑似的な被修飾名詞の決定法を簡単に説明すると、単語を名詞と非名詞に分類し、名詞の後置語が非名詞であれば被修飾名詞であるとした。このようにして算出した被修飾度を用いると、名詞の品詞をもつ語や辞書にない語について英文科学技術文中で実際に名詞として使われる語とそうでない語とを分けることができる。

3. 前方修飾度

前方修飾度も基本的には前方修飾語として生起する相対頻度として定義しているが、被修飾度と同様に疑似的な値で代用している。実際にはtheの後置語を疑似的な前方修飾語であると考え、theの後置語として生起する相対頻度を前方修飾度として用いている。この前方修飾度を用いると、英文科学技術文中で実際に形容詞または修飾名詞として使われる語とそうでない語とを分けることができる。

4. 辞書の品詞と前方修飾度、被修飾度

ここでは、研究社英和中辞典第5版⁴⁾（以下、中辞典）における品詞について調査した。中辞典に記載されている見出し語のうち、原形の単一語は38,906語であり、品詞ごとの語数を表1に示す。次に、品詞の組合せごとの単語数を表2に示す。これらから、名詞または形容詞の品詞をもつ語が多く、多品詞語の曖昧さも名詞や形容詞を含んでいる場合が多いことがわかる。

図1に名詞の品詞をもつ語の前方修飾度と被修飾度に対する分布を示した。同様に図2は形容詞の品詞をもつ語について、図3は名詞も形容詞も含まない語について示した。これらより名詞と形容詞は他の語より前方修飾度が高いことと、名詞は形容詞より被修飾度が高いことがわかる。

表 1 中辞典の見出し語38,906語の品詞

品詞	語数
冠詞	4
間投詞	206
形容詞	10470
副詞	4386
接続詞	58
前置詞	111
動詞	6470
助動詞	26
名詞	24430
代名詞	159

表 2 中辞典の品詞の組合せ上位10

品詞	語数
名	18583
形	7345
副	3713
動名	2993
動	2724
形名	1907
形動名	456
形副	270
形副名	142
形動	107

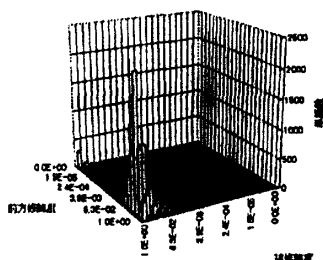


図 1 名詞の品詞をもつ語の前方修飾度と被修飾度

5. むすび

前方修飾度と被修飾度を用いることで名詞的な語と形容詞的な語を判別できることを示した。この結果は統語解析の際の品詞の決定や名詞句の範囲決定に利用できる。

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。

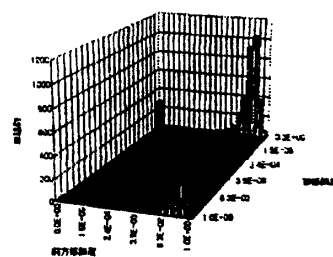


図 2 形容詞の品詞をもつ語の前方修飾度と被修飾度

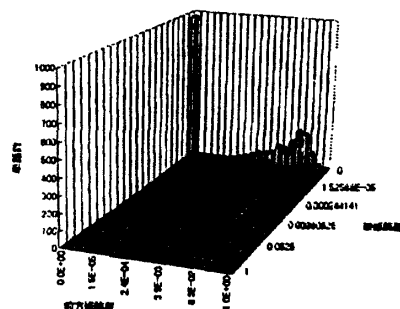


図 3 形容詞も名詞も含まない語の前方修飾度と被修飾度

参考文献

- 1) 竹田, 松尾: 英文科学技術文における単文の原子論理式への変換, 情報処理学会第49回全国大会講演論文集 (1994).
- 2) 竹田, 須田, 楠本, 松尾: 英文科学技術抄録文における名詞の決定, 情報処理学会論文誌 36(8), pp. 1828-1837 (1995).
- 3) 丸木, 柴田, 日昔, 竹田, 松尾: 英文科学技術文における単純名詞句の範囲決定, 情報処理学会第53回全国大会講演論文集 (2), pp. 23-24 (1996).
- 4) 小稲義男ほか編: 新英和中辞典, 第5版, 研究社 (1985).