

抽象名詞「こと」の構造と意味の解析

車井 登 池原 悟 村上 仁一

鳥取大学工学部

{kurumai,ikehara,murakami}@ike.tottori-u.ac.jp

1 はじめに

自然言語処理の最大の問題点は、表現の構造と意味に関する解釈の多様性である。なかでも抽象名詞の表す意味は多様であるために解析が困難である。

従来の抽象名詞の研究としては、「もの」の文法化に関する認知言語学的考察[1]や、「Nのコト」を選択する述語の種類の分析[2]等がある。これら抽象名詞に関する研究ではそれぞれ対象としている抽象名詞の性質を明らかにしているだけで、文中での意味的役割にはあまりふれられていない。

本研究では抽象名詞の「こと」を、日英機械翻訳への適用を考え英語表現に対応づけ可能なレベルの意味に分類する。分類のための情報としては、主として前後に出現する語の字面、品詞を用いた。

まず、2章では抽象名詞について述べる。3章では日英翻訳の観点から「こと」の文中での役割を調べ、その特徴を示す。4章では「こと」の文中での意味的役割を明らかにする。5章では分類結果の精度を示すために、テストデータに適用した結果を示す。6章では分類結果と評価結果からその考察を示す。

2 抽象名詞「こと」について

時枝文法では、具体と抽象のさまざまな認識が立体的な構造をとって存在し入子型構造であるとしている[3]。例えば、「生物」と「無生物」から「物」という名称が与えられ、さらには観念的な存在も合わせて仮名書きにした「もの」という概念が成立する。これは具体から抽象への認識の発展である。また、行動を固定してとらえ、実体化した把握を表現する名詞として「働き」、「眠り」などが生まれ、これらを合わせて抽象的な「こと」という概念が成立する。これら「もの」、「こと」を抽象名詞¹と呼ぶ。

抽象名詞は日本語のコミュニケーションにおいて、話者の意思を伝達したり、表現の意図やニュアンスを微妙に変化させる重要な機能を果たしている。それゆえに、抽象名詞の文中での意味的役割を明確にすることは自然言語処理において重要である。

ここで、抽象名詞「もの」と「こと」の出現頻度を調べた。

結果を表1に示す。結果より「もの」を含む文は小説では4.56%、新聞では1.63%であった。「こと」を含む文は小説では10.1%、新聞では5.22%であった。

¹時枝文法では「形式名詞」という名前に対して、形式だけあって内容のない語は存在しないと、「抽象名詞」と呼んでいる。

この結果から、「こと」が「もの」に比べて使用頻度の高いことがわかる。そこで、本研究では「こと」を研究対象とする。以後、抽象名詞「こと」を「コト」と表記する。

表1: 「もの」、「こと」の出現頻度

	もの	こと
小説 (約60万文)	4.56%	10.1%
新聞 (約150万文)	1.63%	5.22%

3 日英翻訳におけるコトの役割

コトの文中での意味的役割を明らかにするために、コトがどのように訳されるかを考える。

私は彼のことを知らない。

I don't know about him. (1)

この文では「彼のこと」が「him」に訳されており、コトの前方の語句によって訳し方が決まっていると考えられる。

私は以前ここへ来たことがある。

I have come here before. (2)

この文では「ことがある」が「have + 過去分詞」に訳されており、コトの後方の語句によって訳し方が決まっていると考えられる。

3.1 コトの直前に現れる語

まずコトの直前にどのような単語が出現するか調べた。2章で使用した毎日新聞記事と小説100冊を用いて調査を行う。ここでは形態素解析プログラムALT-JAWSの結果を使用した。結果を表2に示す。ここで「副用語」は「副詞・連体詞」を表す。また、「動詞以外の用言」は「形容詞・形容動詞」を表す。

結果より、助動詞は新聞で25.6%、小説では18.3%であった。また動詞は新聞で58.5%、小説では43.8%であった。どちらも動詞・助動詞が付く割合が多いことがわかった。

表 2: コトの直前に現れる語

分類	新聞	小説
格助詞	4.6%	13.2%
助動詞	25.6%	18.3%
副用語	3.3%	12.5%
動詞以外の用言	6.2%	10.3%
動詞	58.5%	43.8%
その他	1.8%	1.9%

また,それぞれの品詞を細分類した.結果を表3に示す.結果よりコトの前に格助詞がくる場合は新聞では96.6%,小説では97.0%が「の」である.副用語の場合は新聞,小説ともに99.7%が連体詞である.このことより,コトの前が格助詞の場合は「の」であり,副用語の場合は連体詞であると言える.

表 3: コトの直前に現れる語の内訳(格助詞,副用語,動詞以外の用言)

分類	区分	新聞	小説
格助詞	の	96.6%	97.0%
	その他	3.4%	3.0%
副用語	連体詞	99.7%	99.7%
	その他	0.3%	0.3%
動詞以外の用言	形容詞	63.0%	54.0%
	形容動詞	37.0%	46.0%

3.2 コトの直後に現れる語

次に同じデータによりコトの直後に現れる語を調べた.結果を表4に示す.

表 4: コトの直後に現れる語

格助詞	新聞	小説
が	34.2%	15.6%
に	15.4%	12.9%
は	10.2%	15.3%
の	1.5%	2.4%
を	20.6%	20.0%
その他	18.1%	33.8%

これより以下のことがわかる.

- コトの直後は多くの場合が格助詞

また,新聞記事と小説とを比較すると以下のことがわかる.

- 「が」の,新聞での使用頻度は小説の約2倍
- 「は」の,小説での使用頻度は新聞の約1.5倍

4 コトの用法の分類

コトの分類には日英機械翻訳機能試験文集[4]のなかで2.2節(「こと」の訳し分け)の133文を用いた.

日英機械翻訳機能試験文集は機械翻訳の性能を試験するための対訳文が約6000文収録されている.

3章での結果より,多くの種類がある前方の語句については品詞単位で分類し,種類の少ない後方の語句については字面単位で分類する.

4.1 コトの前方の語による分類

コトの前方の語による分類には品詞情報を用いる.結果を表5に示す.

「感情の強調」は「形容詞・形容動詞+コト+には」のケースを表す.「動作」はコトに係る語が動詞だけである場合を示す.

表 5: コトの前方の語による分類

表現の構造	分類
名詞+の+コト	名詞型
形容詞+コト	形容詞型
	感情の強調
形容動詞+コト	形容動詞型
	感情の強調
連体詞+コト	連体詞型
動詞・助動詞+コト	動詞
	動作
	助動詞型
	その他

4.2 コトの後方の語による分類

4.1節での分類結果から「その他」に分類されたコトについては英語との対応を明確にするために,後方の語によって細分類を行う.コトの直後に現れる語は主に格助詞である.しかし格助詞だけでは英語と対応づけることが出来ないで,さらに後方の語による分類が必要である.また,前方の動詞や助動詞も重要となってくる.そこでまず格助詞による分類を行い,さらに格助詞の後方の語と,コトの前方の動詞によって分類する.分類結果を表6に示す.

表 6: コトの後方の語による分類

区分	表現の構造	意味
コトが	動詞(過去形)+コトが+ある	過去の経験
	動詞(現在形)+コトが+あった	
	動詞(現在形)+コトが+ある	習慣
	される+コトが+ある	可能性
	コトが+できる	可能
コトに	コトに+なる	予定
	コトに+している(する)	習慣
	コトに+した	決定
	コトに+より(よって)	手段・理由
	コトに+について	動作の対象
コトで	コトで	手段・理由
コトと	コトと+する	決定
コトだ	という+コトだ	推量
コトから	コトから	理由
コトの	コトの+名詞	事象の性質・状態

4.3 動詞の意味による分類

4.1節と4.2節で英語と対応づけることが出来ないコトについて動詞の意味を用いて分類する。次に例文をあげる。

君のいうことは真実だ。

What you say is the truth. (3)

このように、「主語＋発言動詞＋コト」という環境にコトが現れる場合「内容」という意味に分類する。ここで、発言動詞とは「言う」、「提案する」、「主張する」等を指す。

4.4 構文解析による分類

さらに4.3節で英語と対応づけることが出来ないコトについては構文解析の情報を用いる。次に例文をあげる。

彼が不人気であることを私は知らなかった。

I had not known that he was unpopular. (4)

このように、コトに係る文に主語が存在する場合「概括」という意味に分類する。

4.5 慣用句的用法について

4.1節から4.4節の分類とは違う形で現れる表現に慣用句的用法がある。例文を次にあげる。

私は去年はこれといったことを何もしなかった。

I did not do anything particular last year. (5)

このように「これといったコト」は「特別」の意味である。このような用法を「慣用句型」とする。

4.6 分類のまとめ

分類の結果を表7に示す。以上より、対象となる文のコトの形式から意味を導くルールが作成できる。

5 ルールの精度の評価

4章より得られたルールの精度を評価するために実験を行った。対象とするデータには以下の2つを用いる。

- 日英機械翻訳機能試験文(133文)
- 新聞記事(560文)

また、ルールの適用順序を以下のように決定した。

1. 慣用句型の抽出
2. 前方の語による分類
3. 後方の語による分類
4. 動詞の意味による分類
5. 構文解析による分類

慣用句型は字面による情報だけで意味が特定できるため、最初に適用する。

また、各ルールの適用は人手により行う。

5.1 機能試験文による評価

分類に用いた日英機械翻訳機能試験文(133文)を用いて評価を行った。評価方法を以下に示す。

$$\text{カバー率} = \frac{\text{ルール適用データ数}}{\text{全データ数}} \quad (6)$$

$$\text{適合率} = \frac{\text{分類が英訳に有効であると判断した数}}{\text{ルール適用データ数}} \quad (7)$$

表7: 分類のまとめ

前方の語	後方の語		動詞の意味	構文解析	分類
名詞＋＋コト	-		-	-	名詞型
形容詞＋コト	-		-	-	形容詞型
形容動詞＋コト	-		-	-	感情の強調
	-		-	-	形容動詞型
動詞・助動詞＋コト	-		-	-	感情の強調
	が	動詞(過去形)＋コトが＋ある	-	-	動作
		動詞(現在形)＋コトが＋あった	-	-	過去の経験
		動詞(現在形)＋コトが＋ある	-	-	習慣
		される＋コトが＋ある	-	-	可能性
	に	コトが＋できる	-	-	可能
		コトに＋なる	-	-	予定
		コトに＋している(する)	-	-	習慣
		コトに＋した	-	-	決定
		コトに＋より(よって)	-	-	手段・理由
		コトに＋ついて	-	-	動作の対象
		で	-	-	手段・理由
		と	-	-	決定
		だ	-	-	推量
		から	-	-	理由
		の	-	-	事象の性質・状態
	-		主語＋発言動詞＋コト	-	内容
	-		-	コトに係る文に主語有り	概括
	-		-	-	慣用句型
これといったコト	-		-	-	

適合率は人手による判断である。評価結果を表8に示す。

表8: 機能試験文へのルール適用結果

カバー率	適合率	正解率
82%	98%	80%

5.2 新聞記事による評価

テストデータとして毎日新聞記事(1995年度1年分)からコトを含む560文を用いて評価を行った。評価方法を以下に示す。

$$\text{カバー率} = \frac{\text{ルール適用データ数}}{\text{全データ数}} \quad (8)$$

$$\text{適合率} = \frac{\text{分類が日本文と対応していると判断した数}}{\text{ルール適用データ数}} \quad (9)$$

5.1節と同じく、適合率は人手による判断である。評価結果を表9に示す。

表9: 新聞データへのルール適用結果

カバー率	適合率	正解率
63.4%	86.5%	54.8%

5.3 机上評価のまとめ

机上評価のまとめを以下に示す。

- 機能試験文でカバー率82%、適合率98%、正解率80%を得られた。
- 新聞記事でカバー率63.4%、適合率86.5%、正解率54.8%を得られた。

6 考察

6.1 評価結果について

機能試験文による評価結果のカバー率は82%であった。分類法の検討で使用したデータであるにもかかわらず、この値が十分な精度を示しているとはいえない。本手法では主に字面と品詞の情報をを用いたが、さらに単語の意味を用いることで新たなルールの作成が可能であると考えられる。

また、機能試験文と新聞記事においてカバー率に大きな違いがある。この原因として、機能試験文にないパターンのコトが新聞で現れることが考えられる。今後、より多くの表現を調査することで改善されることが考えられる。

6.2 コトに係る用言の意味分類

本手法での「形容詞型」「形容動詞型」から英語を導くためには、係っている用言の意味分類が必要であると考えられる。次に例文をあげる

大きいことは良いことだ。

To be big is good. (10)

このように同じ「形容詞型」でも訳が異なっている。この原因として「大きい」と「良い」の意味の違いが考えられる。

6.3 コトの係り先について

次の段階の解析手法として、コトの係り先の考慮があげられる。次に例文をあげる。

輸出を増大させることが大切だ。

It is essential to expand exports. (11)

*to expand*を導くためにはコトの係り先である「大切だ」に注目する必要があることがわかる。

6.4 コトに係る格について

また、例文(11)に主格を与えると訳が変化する。

日本が輸出を増大させることが大切だ。

It is essential that Japan expand exports. (12)

したがって、コトに係っている格の影響も考慮する必要がある。

7 あとがき

本研究では抽象名詞「こと」の意味的役割を明確にする手法を提案した。分類にあたっては日英機械翻訳への適用を考え、英語表現に対応づけ可能なレベルの意味に分類することを狙った。また、分類のための情報としては、主として前後に出現する語の字面、品詞を用いた。これらを考慮した結果、21種類に分類した。

分類結果から得られたルールを日英機械翻訳機能試験文133文と新聞記事560文に適用した。結果、機能試験文ではカバー率82%適合率98%、新聞記事ではカバー率67%適合率86%を得た。

これらの結果から抽象名詞「こと」の前後の語だけである程度解析できるが、それ以上の精度を求めるためには係り先や格の影響など、他の情報も必要になることがわかった。

参考文献

- [1] 佐々 祐子, 堀江 薫: “形式名詞「もの」の文法化に関する認知言語学的考察”, 言語処理学会第3回年次大会発表論文集, pp.565-568 (1997)
- [2] 笹栗 淳子, 金城 由美子: “現代日本語「Nのコト」を選択する述語の種類: コーパスに基づく分析”, 言語処理学会第4回年次大会発表論文集, pp.338-341 (1998)
- [3] 三浦 つとむ: “日本語の文法”, 勁草書房, 第3版 (1978)
- [4] 池原 悟, 白井 諭, 小倉 健太郎: “言語表現体系の違いに着目した日英機械翻訳機能試験項目の構成”, 人工知能学会誌, Vol.9, No.4, pp.569-579 (1994)