英語表現との対応からみた抽象名詞「こと」の意味分類

車井 登　　池原 悟　　村上 仁一

鳥取大学大学院 知能情報工学研究科

〒 680-8552 鳥取市湖山町南 4-101

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

あらまし:

日本語の抽象名詞「こと」は, 用法と意味が多彩で, 機械翻訳において, 訳出困難な名詞の一つとなっている. 本稿では, 日英機械翻訳に適用することを目標に, 日本文中に使用された抽象名詞「こと」の意味を英語表現に対応させて分類する. 具体的には, まず「こと」に係る単語が 2 文節以上の節を成している場合とそれ以外の場合とに大別し, さらにそれぞれの場合で英語表現と対応づける. 結果, 得られた 31 種の分類から「こと」の翻訳形式を決定するルールを作成し, 新聞記事 182 文を対象に評価実験をしたところ 59%の正解率が得られた.

キーワード: こと, 抽象名詞, 形式名詞, 意味分類

Semantic Classification of Abstract Noun "koto" from English Expressions

Noboru KURUMAI　Satoru IKEHARA　Jin'ichi MURAKAMI

Graduate School of Information and Knowledge Engineering, Tottori University
4-101, Minami Koyamachou, Tottori city, 680-8552, Japan

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

Abstract:

In machine translation, Japanese abstract noun "koto" is one of the most difficult noun, because it has various usages and semantics. In this paper, we classify the Japanese abstract noun "koto" from English expressions for Japanese-to-English machine translation. In concrete terms, we divide "koto" roughly into lexical semantics and grammatical semantics in point of constituent of clause in the words which modifies "koto", and we classify each "koto" with English expressions. As a result we classify into 31 kinds.The rule was evaluated on newspaper with 59% accuracy.

Keywords: "koto", abstract noun, indefinite noun, semantic classification

1　はじめに

日本語の抽象名詞「こと」は, 用法と意味が多彩で, 機械翻訳において, 訳出困難な名詞の一つとなっている.

名詞の訳し分けの研究としては, 意味属性を用いることで名詞の持つ多義性解消の有効性を検証したもの [1] があるが, 一般名詞を対象にしたものである. 抽象名詞の用法の多様さを考えると一般名詞とは完全に切り離し, 個別の対応を行う必要があると考えられる. また抽象名詞を対象にした研究には,「もの」の語彙的意味と文法的意味の連続性を他の抽象名詞「こと」「ところ」と対照することで明らかにしたもの [2] や,「こと」が「名詞 + 連体助詞の + 抽象名詞コト」という環境に現われた場合に必須要素としての「こと」である場合と, そうでない場合, 中間的な場合とを述語により分類したもの [3] 等がある. これらは抽象名詞の持つ用法や意味の複雑さを分析したものであるが, 機械翻訳を対象とはしていない.

さらに, 機械翻訳を対象とした研究には,「こと」の前後の語から意味を決定したもの [4] がある. しかし,「こと」の直前と直後の語だけでは意味を決定できない場合も多くあり, また英語表現と対応づけるまでには至っていない.

本稿では抽象名詞「こと」を,「こと」に係る単語と「こと」を格要素に持つ用言に着目し, 英語表現と対応づけて意味分類する. さらに, 分類結果より抽象名詞「こと」の翻訳形式を決定するルールを作成し, 新聞記事 182 文を対象に評価実験を行う.

以下,2 章では抽象名詞「こと」の使用頻度と使用環境を調査した結果を示す.3 章では抽象名詞「こと」を語彙的意味を持つものと文法的意味を持つものとに大別し, それぞれ英語表現と対応づけて意味分類する.4 章では評価実験を行った結果を示し,5 章では考察を示す.

これより抽象名詞「こと」を「コト」と表記する.

2　コトの使用頻度と使用環境

2.1　コトの使用頻度

まず, コトの使用頻度を調べる. 比較対象として, コトと同じく代表的な抽象名詞の「もの」を取り上げる. 調査対象のデータとして, 毎日新聞記事 (約 150 万文) と小説 100 冊 (約 60 万文) を用いる. 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.

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%

2.2　コトの使用環境

次に, コトがどのような環境で使用されているかを調査する. まず, コトにどのような語が係るのかを知るために直前の語を調査する. 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・助動詞が係る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%

次に, コトの文中での役割を知るためにコトの直後の語を調査する. 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%

結果より,「コトの直後は多くの場合が格助詞」であることがわかる.

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

3 コトの意味分類

次にコトを英語表現と対応づけて意味分類する. まず, コトの意味分類をするうえで,2 つの意味に大別する.

彼が一人で行ったことは確かです.

It is certain that he went alone. (1)

例文 (1) ではコトに「彼が一人で行った」という節が係っている. この場合のコトは節を受ける働きが強い. よって, このようなコトを「文法的役割を持つコト」とする.

私は彼のことを知らない.

I don't know about him. (2)

一方, 例文 (2) では「彼の」がコトに係っており, 節を成していない. この場合のコトはそれ自体に意味を持っている.

私は彼に, 行くことを思いとどまらせた.

I persuaded him not to go.(3)

また, 例文 (3) ではコトに「行く」が係っており,1 文節ながら節が係っている. しかし, この場合は「行く」という事象をとらえたものであり, 節を受ける役割だけではなく, コトの持つ名詞としての意味を使用していると考えることが出来る.

このように, コトに係る語が節を成していない場合, あるいは 1 文節から成る節である場合のコトを「語彙的意味を持つコト」とする.

以下では,「語彙的意味を持つコト」と「文法的意味を持つコト」をそれぞれ分類していく.

3.1 語彙的意味を持つコト

ここでは語彙的意味を持つコトの意味分類を行う. コトは性質上, 修飾語を伴わなければ用いられない. つまり, 修飾語が直接コトの意味を反映していると考えることが出来る. そこで, コトに係る語によって分類し, れぞれ英語表現と対応づける. 分類ための情報としては, 単語を分類する上で代表的な品詞情報を用いる. 分類結果を表 5に示す.

表 5: 語彙的意味を持つコトの意味分類

表記		意味	英語表現
名詞+の+コト		内容	about+名詞
		–	名詞のみの訳語
連体詞+コト	あの, あんな	指示	that
	この, こんな		this
	どんな		what
形容詞+コト		–	形容詞の訳語
形容動詞+コト		–	形容動詞+matter
動詞+コト		動作	to 動詞, 動詞+ing

ここで,「名詞+の+コト」という場合のコトについて説明しておく.

私は彼のことを知らない.

I don't know about him. (4)

上の例文における「彼のこと」は彼の職業や年齢等の「彼」に関する内容を意味している. このような場合の英語表現には名詞の訳語の前に "about" が付く.

彼女は彼のことが嫌いだ.

She does not like him. (5)

一方, 上の例文では「彼のこと」は彼自身を意味しており, 彼の職業や年齢を表しているわけではない. このような場合の英語表現は名詞の訳語のみである.

次に,「連体詞+コト」の場合であるが, 本稿では指示語の連体詞のみを分類している. また,「形容詞+コト」と「形容動詞+コト」の場合では, 英語表現が非常に単純になっている. このことについては考察で述べる.

3.2 文法的意味を持つコト

次に文法的意味を持つコトの意味分類を行う. 文法的意味を持つコトは基本的には節を受ける役割だけであるが, 機能動詞と結び付くことによって補助動詞として働く場合がある.

彼は同じ過ちを繰り返すことになる.

He will be repeating the same error. (6)

例文 (6) の場合,「なる」という機能動詞が「に」を伴ってコトと結び付いており, 補助動詞として見ることが出来る.

そこで, 補助動詞のコトとそうでないコトを分け, それぞれ英語表現と対応づける.

3.2.1 補助動詞のコト

本稿では形態素解析プログラム ALT-JAWS の結果に従い, コトを含んで補助動詞であるものを「補助動詞のコト」とする. 分類結果を次ページ表 6 に示す.

ここで,「ことになる」を例にあげてコトと英語表現との対応を説明しておく.

彼は同じ過ちを繰り返すことになる.

He will be repeating the same error. (6)

例文 (6) では,「ことになる」が「予定」の意味で使われているため, 英語表現は "will" となっている.

私たち, 今度結婚することになりました.

We have decided to become married. (7)

一方, 例文 (7) では同じ「ことになる」が「事象の成立」の意味で使われており, 英語表現は "decide" となっている.

このように同じ表現でも, 意味によって英語表現が異なる場合がある.

また, 表中で明示的に時制の違いを表していない場合は, 時制は問わないものとする.

3.2.2 補助動詞でないコト

次に補助動詞でないコトを分類する. 以下に例文を示す.

勉学に励み, なおかつスポーツにも
大いに活躍することが望ましい.

It is desirable to concentrate on studies yet be active in athletic endeavors. (8)

表 6：補助動詞のコトの意味分類

表記		意味	英語表現
ことになる		予定	will
		事象の成立	decide
ことができる		可能	can(be able to)
ことにする	したことにする	偽りの事象	assume
	することにする	決定	decide to
ことが多い		短期間の反復	frequently
		傾向	tend to
		通例	usually
		頻度	often
ことがある	したことがある	過去の経験	have been
		過去の週間	used to
	することがある	可能性	there are times
		頻度	sometimes
		頻度	frequently (「よく」等の修飾を伴う)
	したいことがある	要求の存在	there is one thing
		要求の存在 (依頼の動詞を伴う)	have a favor
		要求の存在 (内容が不明瞭)	have something

この例文では,「勉学に励み, なおかつスポーツにも大いに活躍する」という節が “to ~” と訳されている. この場合, 節に主格があるならば “that ~” と訳されるので,「コトに係る節の主格の有無」は分類する為の 1 つの基準となる.

期間が一週間かかることから，
要望が強まっていた.
Because it takes one week, (9)
the demand had become strong.

また, 上のような文では,「節+コト+から」が “because ~” という英語表現と対応している. そこで,「節+コト+接続助詞」という表現を接続助詞によって分類する. 分類結果を表 3 に示す.

4 評価実験

3節での分類結果から, コトの翻訳形式を決定するルールを作成し, その評価実験を行う.

表 3: 補助動詞でないコトの分類

表記		英語表現
X+コト+格助詞+動詞	主格無し	V to X
	主格有り	V that X
X_1 は X_2 だ	主格有り	It is ~ that コト
	主格無し	It is ~ to コト
X+コト+から		because ~
X+コト+で		with ~

評価実験には 95 年度毎日新聞記事から, コトを含む 182 文を用いる. 評価は以下の 3 段階で行う.

- ○ ：正解である
- △ ：より良い翻訳方法があるが意味は通じる
- × ：間違いである

評価は翻訳家の方に依頼した. 結果, カバー率 90%, 適合率 66%であった. 適合率は評価が ○ か△ である場合を正解として計算した. 詳細を以下に示す.

表 4: 評価実験の結果

○	87 文/163 文	53%
△	21 文/163 文	13%
×	55 文/163 文	34%

5 考察

5.1 評価結果について

5.1.1 カバー率について

評価実験の結果より, カバー率 90%であった. カバー出来なかった表現には「ことも ~ 」があり,「が」と「も」を同じものとして扱うのかどうか調べる必要がある. 他には,「したことはない」等, 補助動詞のコトとして扱うべきものがあった. 本稿では ALT-JAWS が出力した補助動詞の「コト」の中で, 主に使用頻度の高い表現を分類したが, さらに多くの表現を分類する必要があると思われる.

さらに,

「 ~ (省略) ~ 指示されない」ことが ~ (10)

のように括弧の直後にコトが現れる場合もあり, この扱い方も調べる必要がある.

5.1.2 適合率について

適合率 66%はカバー率 90%に比べて非常に悪い. 間違いの原因に that 節で訳すべきところを to 不定詞で訳している場合があった.

また,「決定」の意味と対応している decide が間違いである場合も比較的多かった.

~ (省略) ~ を引き受けて, 三菱銀の子会社として経営再建を図る<u>ことになった</u>. (11)

これは, コトの意味として「決定」が選択されていながら,decide とはならない場合で, さらに微妙なニュアンスをとらえて意味分類する必要があると思われる.

他に, △ と評価された中には 1 文だけでは決定できないというものもあった. 現時点で解決法は見当たらないが, 今後の問題の 1 つであると考えている.

5.2 新聞と小説の違いについて

2.2節でコトに係る語を調査した結果, 新聞では 84.1%, 小説では 62.1%が動詞・助動詞であった. 実際に動詞が係って「語彙的意味のコト」となる場合は非常に少なく, コトが主に文法的意味として使われていると考えられる.

また, コトの直後の語では, 新聞と小説の間に「が」と「は」に関する大きな違いが見られた. 格助詞の違いは訳語に影響して来ると考えられる. 本稿では主に新聞記事を対象に分類しているが, 翻訳対象によっては分類を修正する必要があるかもしれない.

5.3 語彙的意味のコトについて

「語彙的意味のコト」については,「連体詞+コト」,「形容詞+コト」,「形容動詞+コト」の分類が非常に簡単になっている. この理由として以下に 2 つの例文をあげて説明する.

あの人は相当な年なのに, まだ子供っぽいことばかり言う.
That person is quite old, but he speaks in a childish manner. (12)

この例文を見ると「子供っぽいこと」が “childish manner” と訳されている. 実際に例文全体を見てもコトを「態度」や「挙動」と解釈するのは自然である. よって他の訳語である “matter” や “incident” は当てはまらないことが分かる.

嫌なことはこれっきりにしてほしい.
I would like to have unpleasant happenings end with this one. (13)

この例文では「嫌なこと」が “unpleasant happenings” と訳されている. このようにコ

トを “happening” と訳してあるのは分類に用いたコーパスでも 1 文だけであった.

例文を 2 つ挙げて分かるように, 語彙的意味のコトは非常に複雑な意味を持っており, コトに係る語だけではなく文全体をとらえる必要があると思われる. よって本稿では, 代表的な訳し方のみを取り出して分類した.

5.4 主格の省略について

「 文法的意味のコト 」を英語表現と対応づける場合, 主格の有無が一つの基準となっている.

永井荷風の日記「 断腸亭日乗 」を読むと, 友人から卵, 牛肉, 浅草のりなどを度々もらっていたことが分かる. (14)

この例文に本稿でのルールを適用する. すると, コトに係る節に主格が無く「 分かる 」が他動詞なので,to 不定詞が出力される. しかし, 翻訳家の方によると, that 節で訳すのが正解である. このことから, 主格の表現方法を検討する必要があると思われる.

6 おわりに

本稿では, 日英機械翻訳に適用することを目標に, 日本文中に使用された抽象名詞「 こと 」の意味を英語表現に対応させて分類した. 分類にあたっては, コトに係る単語が 2 文節以上の節を成している場合とそれ以外の場合とに大別し, それぞれ「 文法的意味のコト 」と「 語彙的意味のコト 」とした. さらに,「 文法的意味のコト 」は「 補助動詞のコト 」とそうでないものとに分類し, 英語表現と対応づけた.「 語彙的意味を持つコト 」はコトに係る語の品詞に着目し, 英語表現と対応づけて分類した.

分類結果よりコトの翻訳形式を決定するルールを作成し新聞記事に適用した結果, カバー率 90%, 適合率 66%を得た.

参考文献

[1] 桐澤 洋, 池原 悟, 村上 仁一 : “名詞の訳語選択における意味属性の有効性”, 電子情報通信学会技術研究報告,NLC 99-5 (1997)

[2] 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)

[3] 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)

[4] 車井 登, 池原 悟, 村上 仁一: “抽象名詞「 こと 」の構造と意味の解析”, 言語処理学会第 5 回年次大会 (1999)