

特許文の英語への訳し分けと格フレームとの関係

山形大学大学院理工学研究科教授

横山 晶一

PROFILE

1949年生。1972年東大工学部卒。同年電子技術総合研究所入所。1991年同所知能情報部自然言語研究室長。1993年4月より山形大学。現在大学院理工学研究科教授（情報科学分野）。工学博士。アジア太平洋機械翻訳協会（AAMT）Japio 特許翻訳研究会副委員長。

✉ yokoyama@yz.yamagata-u.ac.jp ☎ 0238-26-3336

1 はじめに

特許文の詳細説明や要件の文が長大で難解であり、複雑な構造を持つことはよく知られている。この構造を解明するために、主として係り受け構造を対象を絞って、解析誤りの分類 [1]、自動システムと特許文の分割 [2]、接続詞と係り受けの構造 [3] などを研究してきた。この研究は、2006年度から2008年度にかけて、科学技術研究費（基盤研究（C）課題番号18500102）のもとで行われたもので、成果全体は3冊の報告書としてまとめられている [4]。

2008年度には、格フレームの構造に着目して、格フレームの違いが、英語の訳し分けに反映されるかどうかを人手で調査した。その結果、サ変動詞の格フレーム構造の違いはある程度反映されるが、和語動詞の場合にはかなり問題があることが分かった。以下ではこの内容について詳しく述べる。

なお、本稿の内容は主として [5～7] に基づいている。

2 KNP と格フレーム

2.1 KNP による特許文の解析

KNP [8] は、汎用の係り受け解析システムで、並列構造などをよく捉えられることで知られている。形態素解析システム JUMAN の結果を入力とし、それらを文

節単位にまとめて係り受け関係を出力する。格フレームは、解析途中の段階の情報として、動詞と、同じ文の中で共起する名詞との間の関係を示すものとして参照される。

```
# ID:2004089106_PROBLEM_2-2
ゲルマイクロドロップ法及びフローサイトメトリー法を組み合わせた自然界から微生物を取得する方法を「開発」する。
.....
<Case Structure Analysis Data>
【開発／かいはいつ】 動 [1] D 方法を《ヲ》
```

図1 KNP 解析1

```
# ID:2004008046_PROBLEM_1-1
口腔疾患等に対する有効成分の探索・評価および口腔疾患予防等の開発に有効なインビトロバイオフィルムを提供すること。
.....
<Case Structure Analysis Data>
【開発／かいはいつ】 動 [3] D 剤等の《ヲ／ガ》
```

図2 KNP 解析2

図1, 2に、「開発」というサ変動詞を含む特許文の解析結果を示す。この図で、「#ID」は、入力文の特許番号とその詳細、次行が入力文、点線の下<Case Structure Analysis Data>以下に書かれたものが、入力文の格フレーム解析の結果である。

図1の「動 [1]」、図2の「動 [3]」が格フレームの分類を示す。つまり、図1と図2では、別の格フレームとして分類されることを示している。

2.2 格フレームの構造と英語への訳し分け

格フレームは、Web データベースから構成されたものの [9,10] で、次のような構造を持つ。

(例 1) {従業員、運転手、…} が {車、トラック、…} に {荷物、物資} を 積む

この例では、動詞「積む」は、「動 [3]」と分類され、直前の格として、<ヲ格>にもしくは<ニ格>を伴う。

(例 2) {選手、従業員、…} が {経験、体験} を 積む

この例では、動詞「積む」が「動 [1]」と分類され、<ヲ格>を直前の格として持つ。例 1 と 2 の違いは、動詞の意味の違いを反映している。

この結果に対して、英和または和英辞典 [11] を引くと、次のようになる。

動 [3] ⇒ 「荷を載せる」 : [load]

動 [1] ⇒ 「蓄積する」 : [acquire]

すなわち、この場合には、「格構造の違い」が「英訳の違い」に対応している。

一般的に、和英辞典の記述を調査すると、和語動詞では、意味の多義性により、多くの表現がなされている。一方、サ変動詞は、意味が限定的なために、訳し分けが比較的容易にできると考えられる。この仮説を確かめるために、次節に述べる調査を行った。

3 特許文の訳し分け

3.1 研究手順

次のような手順で調査を行った。

- (1) 特許の入力文を KNP を用いて解析する
- (2) KNP の出力結果をもとに英訳文と比較する
- (3) 比較した結果を格フレームごとに分類する
- (4) 分類の結果を評価して訳し分けを検討する

3.2 解析結果

特許情報データベース [12] の文章中から、6249 文を入力として、前節に述べた手順で解析を行った。日本語入力文に対応する英文は、入力文と同様に、上記データベースから抜き出した。

(1) サ変動詞の訳し分け

サ変動詞について解析を行った結果を表 1 の「分泌」を例にして説明する。

表 1 【分泌／ぶんびつ】における訳し分け

各フレーム	例文	文数	英訳
動 [1]	細菌株は SAM を培地中へ分泌する (the bacterium strain secretes the SAM in the culture medium)	13	secrete
	細菌は増殖時にコラゲナーゼを分泌する (Bacteria secrete the collagenase in the medium at the time of their proliferation)	1	secret
動 [2]	神経栄養因子の分泌に関わっている (participates in the secretion of neurotrophic factor)	7	secretion
動 [3]	分泌及びジスフィルド結合の形成 (Formation of a disulfide bond in secretary)	1	secretory

表 1 で、「動 [?]」は、格フレームの分類を示す。例文には、日本語文の後にカッコで人間による英訳を付してある。「分泌」を含む文は、上記 6249 文のうち、22 文が見つかった。文数の欄はその内訳を示したものの



である。右側の英訳は、実際の英訳を抜き出したものを示す。

この表から分かるように、“secret”という1例を除いて、格フレームの分類に対応していることが分かる。この1例は、人間が訳した時のスペルミスと考えられる。また、対応する英訳は、活用形を無視したものである。つまり、実際には、“secretes”、“secreted”、“secreting”などの形があるが、それらをすべて一つにまとめたものである。

(2) 和語動詞の訳し分け

表2 和語動詞の各フレーム解析結果

動詞の例	格フレーム	該当文数
【示す／しめす】	動 [4]	167
【持つ／もつ】	動 [14]	125
【異なる／ことなる】	動 [1]	100

表2に、上記6249文に含まれる和語動詞の一部の結果を示す。動詞「示す」は、1742の格フレームに分類されるが、特許文に出現する167文は、「動 [4]」というたった一つの分類に集約される。「持つ」、「異なる」も同様の結果である。つまり、この場合には、実際には多くの格フレームが存在するにもかかわらず、わずか一種類の格フレームに偏るという現象が見られる。

この場合、一つに集約されているので、訳し分けは不可能である。

4 問題点と今後の方針

本研究では、人手によって、分類・解析を行った。そのために非常に作業効率の悪いものとなった。今後は、Alignmentを自動で行うツールなどを導入して、解析の効率化をはかっていきたい。

本研究で問題となったのは、格フレームの分類の正確さである。格フレームは自動分類されているにもかかわらず、多くの要素が偏っている部分がある。とくに、

和語動詞の場合には、上記のような問題点が顕著に表れる。

今後は、格フレームそのものを見直すとともに、格フレームと考え方の類似した結合価や、述語項構造といったものを参考にしながら、さらに研究を進めていきたい。

謝辞：本研究のきっかけを与えて下さったAAMT/Japio 特許翻訳研究会（辻井潤一委員長）のメンバーの方々、守屋敏道所長、渡邊豊英部長、大塩只明主幹、埴金治課長をはじめとするJapioの方々へ感謝します。

参考文献

[1] 横山晶一：特許文解析誤りの分類と自動修正の可能性、Japio2006YEARBOOK (2006) pp.188-191

[2] 横山晶一：特許文解析誤り自動修正システムと正確な翻訳のための特許文の分割、Japio 2007YEARBOOK (2007) pp.228-233

[3] 横山晶一：特許文における接続詞と係り受けの構造、Japio2008YEARBOOK (2008) pp.68-73

[4] 横山晶一：科学技術研究費報告書(2007～2009)（本報告書には、3年間に発表したこの研究関連の論文がすべて含まれている）

[5] 奥山真澄、横山晶一：格フレームを用いた特許文の訳し分け、情報処理学会東北支部研究会 (2009) 08-6-B1-3

[6] 奥山真澄：格フレームを用いた特許文の訳し分け、山形大学工学部卒業論文 (2009)

[7] Shoichi Yokoyama, Masumi Okuyama: Translation Disambiguation of Patent Sentences using Case Frames, Machine Translation Summit XII, 3rd Workshop on Patent Translation (2009)

[8] KNP: 京都大学言語メディア研究室 (<http://www.nlp.kuee.kyoto-u.ac.jp/nl->

resource/knp.html)

- [9] 河原大輔、黒橋禎夫：高性能計算環境を用いた
Web からの大規模格フレーム構築、情報処理学会
自然言語処理研究会 171-2 (2006)
- [10] 河原大輔、黒橋禎夫：格フレーム辞書の漸次的
自動構築、自然言語処理 Vol.12, No.2 (2005)
pp.109- 131
- [11] 電子版 研究社英語大辞典 (2008)
- [12] 特許情報データベース、アジア太平洋機械翻訳協
会／日本特許情報機構 (AAMT/Japio) 特許翻訳研
究会 (2004)

