

# LE-003 意味属性の共起による「AのB」型名詞句の翻訳規則 Semantic attribute based translation rules for Japanese noun phrase "A no B"

徳久 雅人<sup>†</sup> 守谷 有司<sup>‡</sup> 村上 仁一<sup>†</sup> 池原 悟<sup>†</sup>  
Masato TOKUHISA Yuji MORIYA Jin'ichi MURAKAMI Satoru IKEHARA

## 1. はじめに

名詞と名詞を「の」で結んだ「AのB」型の名詞句は、日本語でよく用いられる表現であり、その意味は多彩で、様々な英語表現に翻訳される。従って、「AのB」型名詞句の翻訳精度の向上が強く求められている。

この名詞句に関して、意味分類は [1, 2]、意味解析は [3, 4, 5]、用例ベースによる日英翻訳は [6] などにより研究されている。しかし、従来研究では、機械処理可能な規則の作成とその精度に関する課題が残されていた。これに対し、最近、[7] は A と B の各名詞の文法属性や意味属性の共起関係に着目し、その共起性を英語表現と対応付けることで、機械処理に適した翻訳規則が網羅的に作成可能であることを示した。

しかし、[7] では、8つの英語表現に限定し、翻訳規則を手作業で作成していた。そこで、本稿ではより豊富な英語表現について翻訳規則を半自動的に作成することを目的とする。具体的には、[7] の規則作成手法のうち、名詞の意味属性を中心的に採用する。そして、[8] の手法により意味属性を汎化し規則の適用範囲を広げる。最後に、作成した規則による英語表現の訳し分けを実験する。

## 2. 英語表現の分類

「AのB」型名詞句に対応する英語表現を調べ、本研究の対象を述べる。

### 2.1 標本の収集

「AのB」型名詞句の翻訳規則を作成するために、英語辞典等の対訳例文 33 万件より、日英名詞句対応の標本を、以下の手順で収集する。係り受け関係のある名詞句を収集するが、形式名詞、複合名詞、接頭・接尾語付き名詞、および、比喩表現については対象外とする。

- 1) 対訳例文より「AのB」型名詞句を検索
  - 2) 一般の和英辞書から名詞 A・B の各英訳語を検索し、英訳語を基に対訳から A・B 間の英語句を抜粋
  - 3) 対象外の名詞句を手で除外
  - 4) 計算機による抜粋偏りを無くすため、2) で除かれた標本から 1 万件をランダム抽出し、人手で追加収集
- 1) で約 84,000 件を得たが、2) で約 30,000 件に、そして、3) で約 27,000 件にまで絞られた。4) で約 3,000 件を追加した。以上の結果、約 3 万件の標本が得られた。

### 2.2 英語表現の分類

英語表現に着目して標本を分類すると、49 種類となった (例: 「空の星」の英語表現 “star in the sky” は「B in A」型に分類)。分類した英語表現の中には、標本全体から見て出現度数が極めて低い英語表現も存在する。学習データとして十分な標本が確保できないため、本稿では出現度数が 10 件以上である 28 種類を対象とする。これらは今回収集した標本全体の 99% に相当する。

## 3. 意味属性規則の作成

翻訳規則は、日本語語彙大系 [9] で分類されている一般名詞意味属性の共起パターンで記述する。

### 3.1 意味属性の共起パターン

一般に名詞に多義があるため、一つの名詞に対して複数の意味属性が付与される。意味属性の共起パターンを翻訳規則に用いるためには、名詞句における各名詞の意味を文意に合うものに絞る必要がある。

現在のところ、文意に合う意味属性を自動的に一意に決定させる方法がないため、収集した標本のうち 5,000 句に対して人手で意味属性の絞り込みを行う。以後この絞り込んだ標本を学習データとして使用する。また、一意に決定できない名詞もいくつか存在するため、これらの名詞については複数の意味属性を付与する。

以下に絞り込み例を示す。尚、例中の # は意味属性番号、[] は意味属性名を示しており、以後も同様である。

#### <絞り込み例>

林檎 (#674【果樹】、#854【果物】) の花 (#698【花 (本体)】、#1675【舞踊・演劇・諸芸】、#2299【興隆】)  
⇒ 林檎 (#674【果樹】) の花 (#698【花 (本体)】)

### 3.2 汎化による規則の生成

3.1 節より得られた標本を用い、規則を自動生成する。学習データの少なさによる正解率低下を防ぐため、意味属性体系の上下の階層関係を利用して、規則の汎化を行う。この汎化は、規則全体の汎化による競合の度合いが閾値以下であることを確認しながら、全体で使われる意味属性の階層を徐々に上げていく、という方法である [8]。本稿では、名詞 A・B の意味属性を交互に汎化するため、2次元に拡張している。たとえば、「店【商店】のカレー【飯】」という共起パターンのうち、【商店】と【飯】を汎化して「【店舗等】の【料理】」を得る。

作成した規則を表 1 に、英語表現ごとの意味属性規則の作成数を表 2 にそれぞれ示す。表 2 の規則数から、意味属性の階層からみて兄弟関係にあたる共起パターンが数多く作られたことが伺える。

表 1: 意味属性規則 (一部)

if A 意味属性	B 意味属性	then 英語表現
#408【図書館】	#902【出版物】	B in A
#1147【情報】	#1031【概要】	A in B
#388【場所】	#878【壁】	B of A
#919【喉】	#2583【粹】	B of A

表 2: 意味属性規則の作成数 (一部)

英語表現	規則数 (件)	英語表現	規則数 (件)
A+B	46,315 件	A of B	6,107 件
B in A	23,894 件	A on B	1,010 件
B from A	6,946 件	B as A	597 件

## 4. 補足的規則の作成

既に、[6, 7] では、ある種の「AのB」名詞句あるいは英語表現は意味属性による一般規則になじまない、と

<sup>†</sup>鳥取大学 工学部 知能情報工学科

<sup>‡</sup>平成 15 年鳥取大学大学院修了、現在 (株) シンファーム勤務

いうことを指摘している。そこで、標本のうち目立った表現については、字面、あるいは、同一文内の動詞について共起を加味した規則を作成する。

#### 4.1 字面規則の作成

「私の自転車」や「私の部屋」などの「所有格+B」の英語表現は、「私の」という字面で英語表現が一意に決定できる。従って字面に依存度の高い6種類については、翻訳規則として字面を追加記述する(表3)。登録数は、学習データにおける統計を基に、90単語を登録する。

#### 4.2 動詞規則の作成

収集した標本には、文の用言に依存する英語表現が含まれている。例えば「彼は上着の雪を払った。(He swept the snow from his coat.)」は、「払う」により英語表現が決まる。そこで、動詞に依存する6種類については、動詞の字面、および、A・Bの意味属性を規則の条件に記述する(表4)。登録は、依存度が高いと思われる動詞を学習データから人手で選定し、47個の動詞を登録する。

表3: 字面規則 (一部)

if A 字面	then 英語表現	if A 字面	then 英語表現
彼	所有格+B	学校	B at A
貴社	所有格+B	家庭	B at A

表4: 動詞規則 (一部)

if 動詞	A 意味属性	B 意味属性	then 英語表現
引く	#86【子】	#594【手】	A by B
払う	#818【衣服】	#2365【雪】	B from A

本節の規則は、[6]が文脈を必要とする場合と分類している中でも、同一文内の動詞に着目して解決できる場合を対処する。動詞規則作成の自動化は今後の課題であるが、結合価パターン[9]を利用する方法が考えられる。

### 5. 実験

作成した翻訳規則の精度を調べるため、評価実験を行う。実験データは、学習に使用しなかった標本からランダムに1,000件を抜き出して用いる。実験の際、規則は計算機に実装して行う。実験の評価基準は、「◎: 標本と訳出が一致、○: 標本と一致しないが正解、△: 文によっては正解、×: 誤った英語形式」とする。“○”は、同じ構造を持つ文において、他の言い回しが存在する場合の評価であり、例として以下のようなものが挙げられる。

#### ○の例: 患者の体温

標本: the patient's temperature(A's B)

訳出: temperature of a patient(B of A)

△, ×とした例については、後述の6.節に示す。尚、本評価では、翻訳後の英語表現のみを対象とし、各訳語の適正は考慮しない。また、形態素解析を誤った事例については評価対象外とする。実験結果を表5に示す。

表5: 実験結果 (1,000文)

規則	評価			
	◎	○	△	×
意味属性規則	42.1%	18.0%	19.3%	20.6%
意味属性規則+相補規則	53.3%	16.0%	13.7%	17.0%

### 6. 考察

5.節の実験において評価が×, △となった名詞句について、誤った英語表現を出力する原因について考察する。

#### (1) ×になる原因

##### i. 単語に複数の意味属性がある場合

#### 例1: 屋上の望遠鏡

標本: the telescope on the roof

訳出: the telescope in the roof

意味属性による規則は、3.2節の学習の結果、A・Bを制約する意味属性を1つずつとした。屋上は【上】と【家屋】の両方の意味を持つ。このような単語は、1つの意味属性では性質が表現できないため誤訳となった。今後、複数の意味属性で制約する規則を作成する必要がある。

#### ii. 類似する意味関係に様々な英語表現が存在する場合

##### 例2: 列車の車掌

##### 例3: タクシーの運転手

例2, 3は、意味的に類似しているが、各々「conductors on this train(B on A)」と「taxi driver(A + B)」という英語表現を持つ。規則では、両者とも“A + B”の英語表現であり、例2では誤りとなる。この場合、意味属性で訳し分けをするのは困難であると考えられる。

#### (2) △になる原因

##### 例4: 彼の指示を受けるべきだ。

標本: You must receive instruction from him.

訳出: his instruction

例4では、名詞句「彼の指示」のみを見ると、“his instruction”という翻訳になるが、「受ける」という用言があるため、「彼からの指示」という意味合いを持つ。これは、動詞規則の不足に起因する。従って今後、更にカバー率の高い動詞規則を作成する必要がある。

### 7. おわりに

本研究では、「AのB」型名詞句の英語表現28種類について、意味属性の共起パターンを用いて翻訳規則を半自動的に作成した。計算機による実験を行ったところ、意味属性のみの規則による翻訳精度は60.1%であり、字面・動詞規則を併用した翻訳規則の精度は69.3%であった。文脈によっては正解となるものがあるため、明確な誤りとなったものが各々20.6%と17.0%になった。[7]に比べて明確な誤りが増加したが、意味属性を用いた規則の自動作成に向けて、一つの目安が示せたといえる。

規則の自動作成は、名詞句レベルで文を局所的に解析すれば訳出できる条件において実現できたといえる。今後は、構文構造や文法属性も考慮した規則の自動作成が望まれる。また、意味属性に加え、物理的構造概念など知識の補強や、名詞の格という考え[5]などによる意味属性の自動絞り込みなども課題に挙げられる。

### 参考文献

- [1] 島津, 内藤, 野村: “助詞「の」が結ぶ名詞の意味関係の subcategorization”, 情処研報, **NL-53-1**, pp. 1-8 (1986).
- [2] 平井, 北橋: “日本語文における「の」と連体修飾の分類と解析”, 情処研報, **NL-58-1**, pp. 1-8 (1986).
- [3] 富浦, 中村, 日高: “名詞句「NPのNP」の意味構造”, 情報処理学会論文誌, **36**, 6, pp. 1441-1448 (1995).
- [4] 佐藤, 西原, 鎌田, 横山: “フレーム構造理論を用いた名詞句「AのB」の意味解析”, 信学技報, **NLC97-3**, pp. 17-24 (1997).
- [5] 黒橋, 酒井: “国語辞典を用いた名詞句「AのB」の意味解析”, 情処研報, **NL-129-16**, pp. 109-116 (1999).
- [6] 飯盛, 佐川, 大西: “「の」を含む名詞句の日英翻訳に対する用例ベースアプローチ”, 情処研報, **NL-112-8**, pp. 53-59 (1996).
- [7] 池原, 村上, 宮本: “「AのB」型名詞句の日英翻訳規則について”, 情報処理学会論文誌, **43**, 7, pp. 2300-2308 (2002).
- [8] 中井, 池原, 村上: “階層構造を持つ属性の組とクラスで与えられる構造規則のクラス分類”, 言語処理学会大会, pp. 277-280 (1999).
- [9] 池原, 宮崎, 白井 他: “日本語語彙大系”, 岩波書店 (1997).