

「A の B」型名詞句の日英翻訳規則について

池 原 悟[†] 村 上 仁 一[†] 宮 本 健 司^{††}

「A の B」の形式の名詞句を対象とする日英機械翻訳において、名詞の文法属性と意味属性の情報によって、どれだけの精度で英語表現形式が決定できるかについて検討した。具体的には、まず、アンカー和英辞典から取り出した 2,000 件の「A の B」型名詞句のうち、名詞句の部分に対応する英語表現が切り出せる標本約 1,100 件の英語表現形式を 57 通りに分類した。次に、その中から最もよく使用される形式 8 種類に翻訳できる場合について翻訳規則を作成した。また、翻訳規則の記述では、名詞 A、名詞 B を規定するための文法属性として、26 種類の品詞コードを、また、意味属性として、「日本語語彙大系」で定義された約 2,700 の単語意味属性を使用した。得られた翻訳規則を同じ和英辞書から取り出した別の 1,000 件の名詞句翻訳に適用して、規則の精度を評価した結果によると、規則が適用できたのは（再現率）、名詞句は 83.3%である。目標としていた翻訳規則のカバー率 90%と比較すると、若干低い結果であった。また、正解率では、模範訳と一致した訳が約 69%、一致しないが正しいと判断される例を加えると約 76%である。ほかに、文脈によっては、誤りとなるものが多数（18%）存在するが、これらは、必ずしも誤りといえず、明確な誤り（6%）は比較的少ない。これらの結果から、名詞 A と名詞 B の文法的、意味的性質に関する情報によって英語表現形式が決定できる名詞句は 80%程度であり、残りの 20%については、文中でのその名詞句の役割などを考慮する必要があること、また、その場合は、動詞との格関係が重要な判断情報となることなどが分かった。

Translation Rules for Japanese Noun Phrase “A no B”

SATORU IKEHARA,[†] JIN'ICHI MURAKAMI[†] and KENJI MIYAMOTO^{††}

Capability of the translation rules for Japanese noun phrases “A no B” described using syntactic and semantic word information was studied. First, 2,000 Japanese noun phrases and their translations were extracted from “Anchor Japanese to English Dictionary” and English translation patterns were classified into 57 formats. For 8 formats of those which frequently appear in the sentence, translation rules were constructed using the syntactic attributes (26 attributes) and semantic attributes (2,700 attributes) for Japanese nouns. The rules were evaluated applying to 1,000 examples. The results are as follows; Cover ratio of rules was 83.3% which is little lower than aimed value (90%). As for the accuracy of translation, the translation ratio which coincide to the exemplified answer was 69%. This values increases to 76% when including the examples which were judged semantically correct. There are many examples (18%) which turns wrong depending on context. These are not always incorrect so that definite error was only 6%. From these results, it is said that 80% of noun phrases can be translated correctly by the relation of syntactic and semantic information for nouns but remaining 20% requires other information such as the relation of a verb and a nouns.

1. ま え が き

日英機械翻訳において、日本語の名詞句は、表現も意味も多彩で、適切な翻訳を行うことが難しく、従来から重要な問題の 1 つとなっている。中でも、「A の B」の形式の名詞句は、出現頻度がきわめて多いのに対して、その意味は多彩で、その意味解析ができなければ日本語の意味解析はできたといえないといわれて

いる¹⁾ くらいである。

従来、名詞句の解析としては、統語的、意味的制約条件を考慮した方法²⁾ や名詞間の結合強度を利用した統語解析³⁾ のほか、名詞間の係り受け規則を自動学習する方法⁴⁾ など、多くの研究が行われてきた。名詞句の意味解析についても古くから多くの研究が行われており、特に「の型名詞句」では、名詞の意味的關係を詳細に分析した研究⁵⁾ をはじめ、意味構造を論理式で表現する方法⁶⁾、名詞の意味素性の関係から名詞句の意味を決定しようとしたもの⁷⁾、生成意味論の立場からアプローチしたもの¹⁾ などがある。

また、語彙の多義解消の観点からは、コーパスを

[†] 鳥取大学工学部

Faculty of Engineering, Tottori University

^{††} 京セラミタ株式会社

Kyousera Mita Corporation

使用した統計的な方法の研究がさかんで、多数の研究^{8)~15)}が行われており、訳語選択の方法としても、目的言語を使用する方法¹⁶⁾や単語の意味属性を使用する方法¹⁷⁾などが研究されている。これに対して、名詞句の日英翻訳を狙ったものとしては、名詞の意味素性の関係に着目して名詞句の意味を解析した発火モデルの提案¹⁸⁾があるが、機械処理可能な翻訳規則の作成とその精度に関する問題は、今後の課題となっている。

そこで、本研究では、名詞の文法属性と意味属性の双方を用いて名詞間の文法的、意味的关系を記述したとき、どれくらいの精度で日本語名詞句に対する英語表現形式が決定できるかについて検討する。

具体的には、アンカー英和辞典から取り出した「A の B」型名詞句 2,000 件のうち、名詞句の部分に対応した英語表現が存在する標本、約 1,100 件を対象とし、翻訳形式として最もよく使用される英語表現形式 8 種類に翻訳できる場合について翻訳規則を作成する。また、翻訳規則の記述では、名詞 A、名詞 B を規定するための文法属性として、26 種類の品詞コードを、また、意味属性として「日本語語彙大系」¹⁹⁾で定義された約 2,700 の単語意味属性を使用する。

以上で、作成した翻訳規則を同じ和英辞書から取り出した別の 1,000 件の名詞句翻訳に適用して、規則の精度を評価し、その結果に基づいて、文法属性と意味属性を使用した翻訳規則の可能性について考察する。

2. 日本語名詞句に対する英語表現形式

2.1 名詞句標本の収集

格助詞「の」によって名詞が結合された単純な形式の名詞句について、翻訳規則を作成することを狙って、以下の方法で検討用の標本を収集した。

アンカー和英辞典(46,108 文)より「A の B」型名詞句を含む 2,000 文を抽出し、日本語名詞が単独の名詞でないもの、もしくは、形式名詞など名詞句単位の翻訳が期待できないものを削除した。また、さらに、対訳英文を見て、該当する名詞句に対応する英語表現を取り出せないものを削除した。削除した用例の種類と例を以下に示す。

(1) 単純な名詞句でないもの

例) 「社名表示の変更」, 「新横綱の誕生」

現実には、対訳英文では、文全体の意味が意識されていることなどにより、日本語の名詞句に対応する英語表現部分が取り出せないような場合がかなり存在する。そのような場合は、文の単位で意味的な表現パターンを対応させるなどの方法が必要となるが、本研究では、そのような場合は対象とせず、与えられた日本語名詞句に対して、英語表現形式を出力できるようなサブルーチンの実現を考える。

(理由) 前方の名詞 A が複合語

例) 「この編曲の魅力」, 「大きな山の陰」

(理由) 前方の名詞 A が修飾されている

例) 「私の野球に対する興味」

(理由) 名詞 A と名詞 B に係り受け関係がない

例) 「父と母の希望」, 「私の父と母」,
「赤, 黄, 緑, 青の色」

(理由) 名詞 A または名詞 B が単独でない。

(2) 英訳文から対訳部分が取り出せないもの

例) 場合により プログラムの一部 を変更します。

There will be a change in the program due to circumstances.

(理由) 対応する英訳部分が省略

例) 彼は 京都の地理 に明るい。

He really knows Kyoto.

(理由) 英訳が意識

例) あなたのことは 主人からよく聞いています。

I have heard a lot about you from my husband.

(理由) 後方の名詞 B が形式名詞

例) 私は 鞆の中 に何が入っているか知らない。

I don't know what I have inside my briefcase.

(理由) 後方の名詞 B が位置関係を表す
参考 同種の例: 「A の上」, 「A の中」,
「A の下」, 「A の下(もと)」, 「A の外」,
「A のそば」, 「A の前」, 「A のまわり」, 「A の間」, 「A の向こう」, 「A のうしろ」

(3) 数量を示すもの

例) 3 個のリンゴ・向こう側の 4 人

以上の結果、対応する英語表現を持つ単純な名詞句は、1,104 件である。

2.2 英語表現形式の分類

上記で得られた 1,104 件の名詞句に対応する英語表現を調べたところ、英語表現は、形式上、表 1 に示す 57 種類に分類できることが分かった。

そこで、本検討では、全体でのカバー率 90% を得ることを目標に出現頻度の高く、文脈独立な翻訳が期待できる上位 8 番目までの翻訳形式となる名詞句を対象に翻訳規則を作成することとした。対象とする 8 種類の英語表現形式とその例を表 2 に示す。

表 1 英語表現形式と頻度

Table 1 English translation format and frequency.

#	形式	個数(件)	#	形式	個数(個)	#	形式	個数(個)
1	A の所有格 + B	417	20	A (動詞 ed) + B	3	39	thands to A	1
2	B of A	190	21	B by A	2	40	instead of A	1
3	A + B	113	22	A (目的格)	2	41	more + A	1
4	A の形容詞 + B	96	23	of A (所有代名詞)	2	42	some + A	1
5	A	56	24	A by B	2	43	A (動詞)	1
6	B in A	54	25	B for A (目的格)	2	44	B of A (形容詞)	1
7	単語	23	26	B (動詞) of A	2	45	among + A	1
8	B + A	20	27	B about A	2	46	as + A	1
9	B for A	19	28	A + B (動詞)	2	47	a lot of A	1
10	B(V) + A	18	29	B to A	2	48	a lot of B	1
11	B on A	14	30	A on B	2	49	lots of B	1
12	B	10	31	one's B (動詞 ing)	2	50	almost all A	1
13	B from A	8	32	about A	1	51	B (動詞) by A	1
14	B at A	5	33	due to A	1	52	B (形容詞) + A	1
15	B (動詞 ing) + A	4	34	to + A	1	53	owing to A	1
16	B + A (副詞)	3	35	during + A	1	54	these B	1
17	A (所有代名詞)	3	36	behing + A	1	55	the number of A	1
18	B (前置詞) + A	3	37	by + A	1	56	A (代名詞) + B	1
19	A of B	3	38	from A (目的格)	1	57	B under A	1

表 2 翻訳規則作成の範囲

Table 2 Cover range of translation rules.

#	翻訳規則作成の対象	名詞句の例
1	A's + B 型	「私の父」 my father
2	B of A 型	「車の値段」 price of car
3	A + B 型	「梨の木」 pine tree
4	A の形容詞形+B 型	「雨の日」 rainy day
5	B in A 型	「京都の大学」 university in Kyoto
6	単語型	「月の光」 moonlight
7	B for A 型	「事故の責任」 responsibility for the incident
8	B on A 型	「海辺の石」 rock on the shore

3. 翻訳規則の検討

3.1 日本語名詞の意味的關係の記述

「A の B」型名詞句の意味解析において、横山ら⁷⁾は、名詞 A と名詞 B の意味的關係に着目して、名詞句の意味を決定することとし、名詞の意味を表すのに IPAL 基本動詞辞書²⁰⁾で定義された下記 12 種類の意味素性を使用している(表 3)。

しかし、12 種類の意味素性では、名詞句の意味を決定するのに粗すぎる場合がある。たとえば「社長の山田さん」(同格関係)と「弟の友達」(所有関係)を考えると、いずれも名詞 A と名詞 B がともに<人間>であるため、両者の意味的關係が区別できず、正しい翻訳結果が得られない。また「雨の夜」(rainy night)のように、名詞 A が形容詞訳語を持つか否かといった英単語の文法的性質で翻訳形式が決まるような場合

表 3 横山らが使用した 12 種類の意味素性

Table 3 Semantic attributes for Yokoyama, et al.

人間, 組織, 物, 動物, 植物, 生物の部分, 場所, 時間, 数量, 性質, 動作, 抽象名詞

もある。

そこで、本研究では、名詞の文法属性と意味属性の双方を使用して名詞 A と名詞 B の関係を記述することとし、文法属性としては、形態素解析プログラム ALT-JAWS で使用されている品詞体系²¹⁾から、図 1 に示すような 26 種類の名詞分類を使用することとした。

また、名詞の意味属性としては「日本語語彙大系」¹⁹⁾で定義された一般名詞意味属性 2,710 種類を使用することとした。使用する意味属性体系の上位の一部を図 2 に示す。

3.2 着目する意味的關係

本検討では、名詞 A と名詞 B の意味的關係から 8 種類の英語表現形式を決定することを狙っている。英語の表現形式が決定できるためには、まず、名詞 A と名詞 B の意味的關係が決定していることが望ましい。名詞 A と名詞 B の意味關係については、すでに詳細な意味分類が行われている⁵⁾が、この分類を計算機上で実現するのは容易でない。

しかし、機械翻訳では、対応する特定の英語表現形

式を決めることができればよいから、多少粗い意味解析であっても、目的にかなう可能性がある。そこで、本研究では、英語側の前置詞の決定で重要性が高いとみられる意味的な関係として、「同格の関係」、「場所の関係」、「時間の関係」、「所属の関係」、「部分全体の関係」、「種類の関係」、「材質の関係」の7種類の関係を考える。これらの関係を判断するための名詞 A と名詞 B の意味的な条件と名詞句の例を表 4 に示す。

以下では、これらの関係に着目して、翻訳規則を作成する。

3.3 翻訳規則作成手順

8 種類の英語表現形式に翻訳される 1,104 件の名詞句標本を分析し、前節で示した 7 種類の意味的な関係を考慮して翻訳規則を作成した。作成した翻訳規則を表 5 に示す。翻訳規則作成の手順は、以下のとおりである。

(1) 英語表現形式ごとの規則の作成

- (a) 表 5 の #1 では、まず、慣用的な訳語のある場合や 1 語で訳されるような場合 (表 2 の #6) について、辞書に登録された訳語を探す。
- (b) 次に、所有関係の場合に対して、所有者の文法的、意味的な性質によって翻訳形式を決める。
- (c) 3 番目は、時間的な関係と場所的な関係の 2 種類 (表 5 の #3 と #4) が対象であり、意味的には、表 2 の #2 と #3 に関係する。これらの場合は、訳し方は一様でなく、使用頻度の高い単語では、固有の訳し方があるなどさまざまであるので、対象の名詞の意味によって翻訳規則を定める。
- (d) 最後は、翻訳形式側からみた規則で、表 5 の #5 ~ #10 が該当する。これらは、表 4 の意味的な関係を考慮しながら、それぞれの形式に訳すための条件を拾い出したものである。翻訳形式としては、#3 と #4

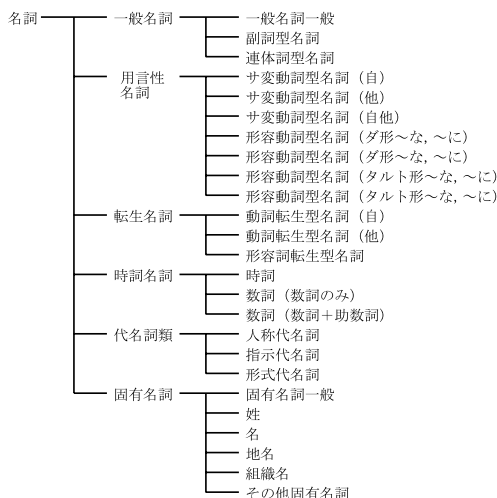


図 1 名詞の品詞コード

Fig. 1 Syntactic code for Japanese nouns.

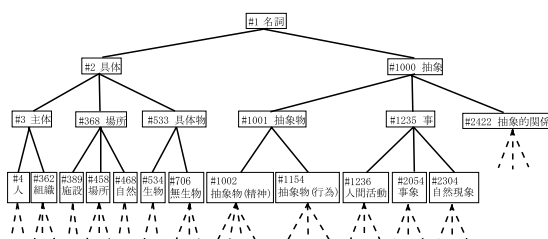


図 2 一般名詞意味属性体系の一部

Fig. 2 General noun semantic attributes system.

表 4 本検討で考慮する名詞句の意味

Table 4 Semantic information used in describing translation rules.

#	意味関係	説明	名詞 A の条件	名詞 B の条件	名詞句の例
1	同格関係	A と B が同格	品詞が < 固有名詞 (姓, 名) > でないこと, および意味属性が #4 人であること	品詞が < 固有名詞 (姓, 名) > で, 意味属性が #4 人であること	「娘の恵美」「弟の春男」, 「社長の山田さん」
		地名の同格	品詞が < 固有名詞 (地名) > でないこと, かつ意味属性が #459 地域であること	品詞が < 固有名詞 (地名) > で, かつ意味属性が #459 地域であること	「国立公園の大山」, 「選挙区の山形」
2	場所関係	A が場所を表す	名詞の意味属性が #388 場所, #865 家屋 (本体), #868 部屋		「公園の花」「京都の大学」
3	時間関係	時間的な関係を表す	品詞が < 時詞 >	品詞が < 時詞 >	「昨日の試合」「来週の火曜」
4	所属関係	B が A の所属を表す	意味属性が #362 組織	意味属性が #4 人	「学校の先生」「会社の社長」
5	部分全体関係	B の名詞が A の名詞の部分である	#673 樹木, #706 無生物	#673 樹木, または, #706 無生物の配下の意味属性	「バナナの皮」「車のドア」
6	種類関係	A が B の種類を表す	B の名詞の種類を表す		「漫画の本」「松の木」
7	材質関係	A が B の材質を表す	B の名詞の材質を表す		「水晶の指輪」「金のメダル」

表 5 「の型名詞句」の翻訳規則凡例：<文法属性>，#nnn 意味属性「単語」

Table 5 Translation rules for 「A no B」.

#	英語形式	英語形式の選択条件		英語形式
		名詞 A に関する条件	名詞 B に関する条件	
1	単語一体型	辞書検索により「A の B」という単語があるか調べる		辞書登録訳語
2	所有格	<人称代名詞>		所有格+B
3	時詞を含む場合	<固有名詞(「姓」「名」)>，#4 人，#536 動物		A's+B
		「うち」		our B
		「去年」，「先週」	#2672 季節，#2682 日，#2681 週	last+B
		「今年」，「今週」		this+B
		「来年」，「来週」		next+B
		<時詞(昨日，今日，明日，週，年，朝，夜)>	<時詞>でない	A's+B
		<時詞(日，曜日，特定の日)>		B on A
		<時詞(月，年，季節，午前，午後)>		B in A
		<時詞>		A+B
		<時詞>でない，<人称代名詞>でない， <固有名詞>でない	<時詞>	A in B
		形容詞形がある	#2701 未来	B of A
		「次」		A (形容詞形)+B
		「今回」	<時詞>でない	following+B
		「次」，「次回」	上記以外	this+B
				next+B
4	場所を含む場合	#1068 その他名称(国名)	#4 人，#389 施設	A の形容詞形+B
		#1068 その他名称(地名)	#1068 その他名称(地名)	B in A
		#465 都市	#389 施設，#4 人	B in A
			上記以外	B of A
		#388 場所	<サ変動詞>，#318 人(地位)，“#388 場所”の配下	B of A
			#4 人	B in A
		#422 区域施設，#423 庭園，#424 動植物園，#425 遊園地	#671 植物，#535 動物	B in A
		#490 岸，#2667 ふち，#493 湖岸， #418 道路，#494 堤		B on A
		#865 家屋(本体)部屋，#445 居住施設	#535 具体物	B in A
			#4 人	B of A
		#389 施設	#878 壁	B of A
			上記以外	B of A
5	for 型		#2503 才能，#2502 能力，#2488 性格，#2505 才，#1458 準備，#1469 計画，#2533 是，#1167 義務，#2048 罪， #2505 使用，#1415 償い(報い，報酬)，#382 代理	B for A
6	on 型	#2634 上		B on A
		#1005 学問一般	#1119 本，#1531 報告，#1540 報道，#1111 文章	
		#1120 目録，#1106 表，#919 帳，#1113 文書類	#1056 名称，#5 人間，#2590 値・額，#1098 文字類	
		#2552 動物(部分)	#22421 怪我	
		#878 壁	—	
		#1548 電話	#1349 声	
7	in 型	#594 手，#603 足	#2416 病気，#2421 怪我，#1242 痛み	B in A
		#462 世界，#384 社会，#1052 写真など		
8	「A + B」型	#706 無生物，#673 樹木	(#706 無生物，#673 樹木)の配下	A+B
		A と B が同格，A が B の材質		
		B の配下	<一般名詞>	
		#987 乗り物，#1749 寄与，#1705 対応	#1119 料金，#1717 契約，#2612 席，#2135 進行， #2137 走行，#986 乗り物	
		#1680 スポーツ	#1956 競争，#1680 スポーツ，#360 選手，#2672 季節	
		#1005 学問一般	#2670 時間，#1426 試験，#237 教師，#1119 本	
			#378 会，#1232 集会，#115 仲間，#119 成員，#2672 季節，#2511 事情(都合，事情)，#2091 化，#2531 不調	
9	形容詞型	<形容詞形が存在>		A の形容詞形+B
10	「B of A」型		A の配下，#871 縁側，#2625 底，#876 天井	B of A
		#4 人	#1263 感情(喜び，悩み，…)	
		“#4 人”以外	#1236 人間活動	
		#3 主体	<サ変名詞>	
			<形容詞転生名詞>	
			五感に関する語(#2355 匂い，#2527 気配，#2356 味， #2354 音，#1239 感覚，#1349 声)，#1021 法則， #1008 知識，#1037 創作物，#1220 習慣，#2591 度量衡， #2521 危険，#1065 名称，#1185 価格，#2532 是非	
		#1105 学問一般	#1424 探求，#1388 学び	
		#536 動物	#706 身体の一部	

の翻訳形式と同じものが多い。

なお、規則適用の順序の決め方については次節で述べる。

(2) 規則の適用順序の決定

与えられた標本があるカテゴリ C_i ($1 \leq i \leq n$) に属するか否かを判定するための規則 R_i があったとする。このとき、 P_i をカテゴリ C_i の標本を正しく判定する確率、 Q_i を C_i 以外の標本を誤って C_i と判定する確率とし、 X_i をカテゴリ C_i の標本が全体に占める割合とすると、全体の正解率を最大とするには、一般に、

$$R_i = \frac{X_i \cdot P_i}{Q_i} \quad (1)$$

の値が大きい順に、判定規則を適用すればよいことが証明できる。

表 5 は、得られた翻訳規則を式 (1) によって評価し、順に並べたものである。

4. 翻訳規則の精度評価

4.1 評価基準

翻訳規則の精度を評価するため、アンカー和英辞典より検討で使用した名詞句以外の名詞句 1,000 件を抽出し、翻訳規則を適用した。評価の基準は以下の 4 種類である。

印：対訳と一致している。

○印：対訳と一致していないが、間違いでない。

印：文脈によっては誤りとなる表現である。

×印：間違いである。

*印：適用対象外。

4.2 評価結果

評価結果を表 6 に示す。また、各レベルに評価された名詞句の例を示す。

例) 「わたしの事務所」

my office (規則訳), my office (正解)

例) 「あじさいの花」

○ hydrangea flower (規則訳),

hydrangea (正解)

例) 「先生の手紙」

teacher's letter (規則訳),

letter from teacher (正解)

例) 「日本のゴッホ」

× Japanese van Gogh (規則),

Van Gogh of Japan (正解)

例) 「鉄道の旅」

* 訳なし, traveling by train (正解)

表 6 から以下のことが分かる。

表 6 評価結果

Table 6 Evaluation results.

評価		全体での評価		適用対象範囲での評価
		件数	割合 (%)	
適用対象内	○	574 件	57.4%	68.9%
		58 件	5.8%	7.0%
		150 件	15.0%	18.0%
	×	51 件	5.1%	6.1%
適用対象外*		167 件	16.7%	—
合計		1,000 件	100.0%	100.0%

表 7 翻訳誤りの例

Table 7 Examples of translation errors.

1. 時詞を含む名詞句 (12 件)	
「あすの今ごろ」	tommorow this time (規則訳) this time tomorrow (正解)
「先週の今日」	last week today (規則訳) a week ago today (正解)
2. 地名を含む名詞句 (11 件)	
「日本の学会」	Japanese academic world (規則訳) the academic world of Japan (正解)
「エデンの園」	Garden in Eden (規則訳) Garden of Eden (正解)
「飲み屋の付け」	bill of the bar (規則訳) bill at the bar (正解)
3. 人の様態を示す名詞句 (3 件)	
「人の恨み」	human grudge (規則) grudge (正解)
「先生の姿」	teacher's figure (規則) teacher (正解)
4. 比喩的な名詞句 (2 件)	
「日本のゴッホ」	Japanese van Gogh (規則) van Gogh of Japan (正解)

(1) 作成した規則が適用できた名詞句は 83%である。

(2) そのうち、模範訳と一致した訳は約 69%で、一致しないが正しいと判断される例を加えると正解率は約 76%である。

(3) ほかに、文脈によっては、誤りとなるものが多い数 (18%) 存在するが、これらは、必ずしも誤りといえず、明確な誤り (6%) は比較的少ない。

5. 考察

5.1 適用範囲について

本研究では、カバー範囲 90%の翻訳規則を狙って、「A の B」型の名詞句に対応する英語表現 57 種類の中から、出現頻度の高い 8 種類の表現を対象に翻訳規則を作成した。これに対して、実験結果では、規則のカバー率は 83%であった。これより、狙った適用範囲に対するカバー率は、92% (=83.3%/90%) となる。

5.2 翻訳に失敗した例について

評価が × 印となった名詞句では、時や地名に関する

表 8 文脈などによって訳し分けることが必要な例

Table 8 Examples of context dependency.

#	文脈依存の場合	名詞句の例	名詞句の使用された文例
1	より細かい意味解析 がより必要なタイプ (54 件)	「先生の手紙」 teacher's letter (規則訳) letter from teacher (正解)	・先生の手紙には温かみがももっていた。 ・ The <u>letter from my teacher</u> was full of warmth.
		「良寛の書」 Ryokan's calligraphy (規則訳) calligraphy by Ryokan (正解)	・わが家では良寛の書を秘蔵している。 ・ Our family's treasure is a sample of <u>calligraphy by Ryokan</u> .
		「ドラムのテンポ」 tempo of drum (規則訳) drummer's tempo (正解)	・ドラムのテンポは曲とあまりよく合っていなかった。 ・ The <u>drummer's tempo</u> was out of beat with the music.
2	副詞的表現を使用するタイプ (78 件)	「きのうの豪雨」 yesterday's heavy rain (規則訳) heavy rain yesterday (正解)	・きのうの豪雨でその町は大きな被害を受けた。 ・ The <u>heavy rain yesterday</u> did great damage to that town.
		「きのうの上野公園」 yesterday's Ueno Park (規則訳) Ueno Park yesterday (正解)	・きのうの上野公園は 1 万人を越す人出があった。 ・ More than ten thousand people thronged to <u>Ueno Park yesterday</u> .
		「今の若者」 today's young people (規則訳) young people today (正解)	・今の若者があまり本を読まないのは、1 つにはやるが多すぎるからだ。 ・ <u>Young people today</u> don't read much because for one thing, they have too many other things to do.
3	動詞または形容詞に 訳出するタイプ (18 件)	「物理学の研究」 study of physics (規則訳) study physics (正解)	・ぼくは京都大学で物理学の研究をしたい。 ・ I want to <u>study physics</u> at Kyoto University.
		「食料品の買い出し」 shopping of food (規則訳) buy food (正解)	・母は食料品の買い出しに出かけています。 ・ My mother has gone <u>to buy food</u> .
		「火山の爆発」 eruption of the volcano (規則訳) The volcanic eruption (正解)	・火山の爆発で大地が震動した。 ・ The <u>volcanic eruption</u> made the earth shake.

表現，人間に関する表現，比喩の表現などが目立った．その例を表 7 に示す．

時詞や場所を表す名詞を含む名詞句では，英語において固定的な言い方をする場合が多く，一般則になじまない場合がある．そのようなものの頻度は比較的多いが種類は多くないので，辞書登録するなど別扱いとすることが期待される．

「人の心」や「人の気持」などの表現は，英語表現では「人」の部分が省略された表現となる場合が多い．本研究で作成した規則では，ほとんどの場合「human + B」と訳すことにしているため誤りとなっている．また「A の姿」などの様態を表す表現も英語では「姿」の部分が省略される場合が多く，評価は×印となっている．

また，比喩的な表現で名詞 A が国名の場合，その形容詞形を使用した規則が適用され「A (形容詞) + B」の形式に訳されるため，評価は×印となった．これについても別の枠組みが必要と考えられる．

5.3 文脈などに依存する場合について

評価が「より細かい意味解析の必要なタイプ」，「副詞的表現を使用するタイプ」，「動詞ま

たは形容詞に訳すタイプ」の 3 種類に分類した．各タイプの例と標本数を表 8 に示す．

第 1 のタイプで最も多かったのは「先生の手書き」のようなタイプの名詞句であった．これらの名詞句を訳し分けるには，動詞との関係などの情報が役立ちそうである．たとえば，表 8 の例文では，格助詞と動詞の関係から「先生の手紙」は「先生が所有している手紙」ではなくて「先生が書いた手紙」，もしくは「先生から届いた手紙」であることが判断できる．

全体で最も多いのが第 2 のタイプである．このタイプの例では，時詞を副詞に訳す点に特徴があり，翻訳結果が必ずしも誤りとはいえないような例を多く含んでいる．すでに述べたように，翻訳の失敗例では，時詞と場所を示す語を持つ名詞句が問題となったが，この種の名詞句は，別に検討する必要がある．

抽象度の高い意味を表す名詞（「もの」「こと」「の」「ところ」「とき」「わけ」）については，意味的に用法が多彩であり，すでに，それらを含み表現の意味解析が研究されている²²⁾．また，時間と場所に関する表現についても研究例^{23),24)}がある．それらの研究と同様，これらの単語を含む名詞句は，名詞句翻訳の一般則と切り離して扱うことが適切と考えられる．

第3のタイプでは、多くが動詞との組合せで訳される。特に「を格」をとる名詞句で、名詞Bがサ変動詞型(研究、掃除など)、かつ、動詞が「する」「行う」など機能動詞の場合、英語では、名詞Bは動詞に訳されることが多い。結合価文法などを使用して、対象とする名詞句と動詞の格関係を調べる必要があると考えられる。

6. 結 論

本研究では、日本語名詞句の中でも、形式は単純だが使用頻度が高く、意味的にも多彩な「A の B」型名詞句を採り上げ、日英機械翻訳規則について検討した。

具体的には、まず、アンカー和英辞書から「A の B」型名詞句を有する日英対訳文 2,000 件を取り出し、A、B が単純な名詞で、対応する英語表現を有する名詞句 1,104 件の対訳例を分析して、英語の翻訳形式を 57 通りに分類した。カバー率 90% の翻訳規則の作成を狙って、その中から、出現頻度の高い 8 種類の表現形式を対象に、日英翻訳規則を作成した。翻訳規則の記述では、26 文法カテゴリからなる文法属性体系と 2710 種類の意味属性種類からなる一般名詞意味属性体系を使用した。

同上の辞書から得られた名詞句標本のうち、用例検討に使用しなかった名詞句標本 1,000 文件を対象に翻訳規則を適用した結果では、規則が適用できたの(再現率)は、名詞句は 83% である。目標としていた翻訳規則のカバー率 90% と比較すると、得られた規則のカバー率は、若干低い結果となった。

正解率では、模範訳と一致した訳が約 69% で、一致しないが正しいと判断される例を加えると約 76% である。ほかに、文脈によっては、誤りとなるものが多数(18%)存在するが、これらは、必ずしも誤りといえず、明らかな誤り(6%)は比較的少ない。

本検討では、名詞 A と名詞 B の文法属性と意味属性の情報から英語翻訳形式を決定することを試みたが、検討結果では、この範囲の情報で翻訳できる名詞句は 80% 程度であり、あとの 20% については、文中でのその名詞句の役割などを考慮する必要がある。今後は、動詞との格関係を考慮した方法について検討していきたい。

また、翻訳失敗例の多くは、時詞、地名、人の様態などに関連する名詞句である。今後、これらの名詞句を対象に個別的でより精密な規則を検討することが大切である。

参 考 文 献

- 1) 菊池隆典, 白井英俊: 語彙意味情報に基づく日本語名詞句の意味解析, 科学研究費補助金成果報告書, No.09480067(研究代表者: 郡司隆男), pp.67-94 (2000).
- 2) 穴倉, 宮崎正弘: 構成要素の統語・意味的制約を利用した日本語名詞句解析, 信学技報, NLC94-49, pp.41-48 (1995).
- 3) 江尻秀彰, 宮崎正弘: 名詞間の接続強度と「の」型名詞句の用例を利用した日本語名詞句構造解析法, 情報処理第 56 回全国大会, 1Q-2 (1998).
- 4) 池原 悟, 中井慎司, 村上仁一: 多義解消のための構造規則の生成方法と日本語名詞句への適用, 自然言語処理論文誌, Vol.6, No.1, pp.143-174 (2001).
- 5) 島津 明, 内藤昭三, 野村浩郷: 助詞「の」が結ぶ名詞の意味的関係の解析, 計量国語学, Vol.15, No.7, pp.247-266 (1986).
- 6) 富浦洋一, 中村貞吾, 日高 達: 名詞句「NP の NP」の意味構造, 情報処理学会論文誌, Vol.36, No.6, pp.1441-1448 (1995).
- 7) 横山昌一, 加藤貴子, 廣重拓司: 格助詞「の」の分類と解析, 情報処理学会第 1 回年次大会, pp.137-140 (1995).
- 8) Brown, P., Della, S., Della, V. and Mercer, R: Word-sense disambiguation using statistical method, *Annual Meeting of ACL*, pp.264-270 (1991).
- 9) Yarowsky, D.: Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora, *COLING-92*, pp.454-460 (1992).
- 10) Dagan, I. and Itai, A.: Word-sense disambiguation using a second language monolingual corpus, *Computational Linguistics*, Vol.20, No.4, pp.563-596 (1994).
- 11) Agirre, E. and Riguu, G.: Word Sense Disambiguation using Conceptual Density, *COLING-96*, pp.16-22 (1996).
- 12) 角田達彦, 田中英彦: 英語名詞の多義性解消における文脈としての場面情報の評価, 自然言語処理, Vol.3, No.1, pp.4-27 (1996).
- 13) Barriere, C.: Redundancy: helping semantic disambiguation, *COLING-ACL '98*, pp.103-109 (1998).
- 14) Wilks, Y. and Stevenson, M.: Word Sense Disambiguation using Optimised Combinations of Knowledge Sources, *COLING-ACL '98*, pp.1398-1402 (1998).
- 15) Li, H. and Abe, N.: Word Clustering and Disambiguation Based on Co-occurrence, *COLING-ACL '98*, pp.749-755 (1998).
- 16) 野美山浩: 目的言語の知識を用いた訳語選択と

その学習性, 情報処理学会自然言語処理研究会, 91-NL-86 (1991).

- 17) 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, 信学技報, NLC-55-5, pp.29-36 (1999).
- 18) 島津 明, 内藤昭三, 野村浩郷: 助詞「が」が結ぶ名詞の意味関係の subcategorization, 情報処理学会言語処理研究会, 86-NL-53, pp.1-8, (1986).
- 19) 池原 悟, 宮崎正弘, 白井 諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林 良彦 (共編) 「日本語語彙大系」全5巻, 岩波書店 (1997).
- 20) 情報処理振興事業協会技術センター: 計算機用日本語基本動詞辞書 IPAL—辞書編 (1987).
- 21) 宮崎正弘, 白井 諭, 池原 悟: 言語過程説に基づく日本語品詞の体系化とその効用, 自然言語処理学会論文誌, Vol.2, No.3, pp.3-26 (1995).
- 22) 車井 登, 池原 悟, 村上仁一: 英語表現から見た抽象名詞「こと」の意味分類, 情報処理学会自然言語処理研究会報告, 99-NL-133, pp.13-19 (1999).
- 23) 的場和幸, 池原 悟, 村上仁一: 日英時間表現の意味と対応関係の解析, 平成12年度自然言語処理学会年次大会, A4-7, pp.245-248 (2001.3).
- 24) 廣瀬文男, 池原 悟, 村上仁一: 日本語の位置的関係表現の意味的対応関係, 平成12年度自然言語処理学会年次大会, A6-2, pp.490-493 (2001.3).

(平成13年6月28日受付)

(平成14年4月16日採録)



池原 悟 (正会員)

1967年大阪大学基礎工学部電気工学科卒業。1969年同大学院修士課程修了。同年日本電信電話公社に入社。数式処理, トラフィック理論, 自然言語処理の研究に従事。1996年スタンフォード大学客員教授。現在, 鳥取大学工学部教授。工学博士。1982年情報処理学会論文賞, 1993年同研究賞, 1995年日本科学技術情報センター賞(学術賞), 同年人工知能学会論文賞, 2002年電気通信普及財団賞(テレコム・システム技術賞)受賞。電子情報通信学会, 人工知能学会, 言語処理学会, 機械翻訳協会各会員。



村上 仁一

1984年筑波大学第3学群基礎工学類卒業。1986年同大学院修士課程理工学研究科理工学専攻修了。1986年NTTに入社。NTT情報通信処理研究所に勤務。1991年国際通信基礎研究所(ATR)自動翻訳電話研究所に出向。1997年鳥取大学工学部知能情報工学科に転職。現在に至る。主に音声認識のための言語処理の研究に従事。電子情報通信学会, 日本音響学会, 言語処理学会各会員。



宮本 健司

2001年鳥取大学工学部知能情報学科卒業。京セラミタ株式会社に入社。現在に至る。