

## 英単語に対する連語的日本語訳語の分析

白井 諭<sup>†</sup> 大山芳史<sup>†</sup> 我妻知恵<sup>‡</sup> 石崎 俊<sup>‡</sup><sup>†</sup>NTT コミュニケーション科学研究所 <sup>‡</sup>慶應義塾大学 政策・メディア研究科

## 1 はじめに

機械翻訳における意味解析では、単語の共起関係を正しく捉えることが必要である。特に原言語と目的言語で表現の基本構造を対応づけるには、用言と名詞の意味的な共起に着目した結合価パターン対の使用が有効であることが知られている。このパターン対の使用に当たっては、記述精度の問題と収集方法の問題がある。

記述精度の問題については、日英機械翻訳の場合、格要素となる名詞の意味属性を約2,000種類以上の分解精度で分類すれば、慣用表現や専門的な表現を除き、日本語の動詞を訳し分けられるような結合価パターン対が記述できることが知られている [1]。

また、収集方法の問題について、筆者らは人用の和英辞書 [2] に記載されている対訳例文、および、日本語の辞書 [3] の詳細な語義分類に基づく用例文や人の知識を内省して作成した用例文に基づいて、日英結合価パターン対の収集を進めている。現時点では17,000件あまりを収集しているが、最終的には25,000件程度にする必要があると推定している [4]。

この結合価パターン対辞書を始めとして、これまでは日本語の単語に対する英語表現を収集してきたが、日本語の単語に対して英語の単語を対応させるという方法では、訳文品質を向上させる上で基本的に限界がある。英単語の日本語訳が複数の文節に相当する場合が少なくないからである。逆に言えば、日本語の複数の文節からなる表現をうまく一つの英単語に対応づけることができれば、それだけ自然な英訳文が得られる可能性があると言える。

そこで、本稿では、日本語の複数の文節を英単語に対応づける方法の確立を目指し、その一環として、英和辞書を題材にして英単語と日本語の複数の文節が対応する表現の調査を行なう。特に、結合価パターン対辞書を充実させることを目的として、日本語の格要素と動詞が一つの英単語に対応する場合について重点的に検討する。

## 2 英単語に対する日本語訳表現の単語構成

本稿では、電子化された英和辞書（学研英和辞書データベース、約6万見出し）のデータを対象にして、英語の見出し語（単語または複合語）に付与された日本語訳を一つ一つ取り上げた。そのおのおのについて形態素解析を行な

い、その結果が複数の文節からなる日本語訳の品詞列を分類し、種類と出現件数を整理した。辞書に収録されていた英日の訳語対の総数は126,939件で、うち33,001件に複数の文節の日本語訳が付与されていた。

品詞列の分類に当たっては、まず大まかな傾向を把握することを考えた。例えば、複合名詞と単一名詞では単語列の構成は異なっているが、全体としては名詞である点で同等に扱う方が良いので、同じ品詞が連続している場合は一つにまとめた。また付属語では品詞レベルに抽象化すると特徴が失われるため、一部を除いて入力表記をそのまま使用した。その結果をソートして出現件数の多い品詞傾向を割り出した。形態素解析に当たっては、括弧書きをされた補足表現は取り除いた。集計結果を表1に示す。

表1：出現件数が上位の品詞列

No	品詞列	度数	英単語の日本語訳の例（英単語）
1	名の名	3097	イングランド人の男 (Englishman)
2	名を動	1449	向きをかえる (turn)
3	動名	853	回るおもちゃ (whirligig)
4	名に動	785	自動車に乗せていく (motor)
5	名の形	601	心のつめたい (icy)
6	静名	571	親切な行い (favor)
7	名	550	1人乗り競漕用スカル (wherry)
8	名を動名	534	筋肉を動かすこと (exercise)
9	形名	470	冒険に満ちた長い旅行 (odyssey)
10	副動	400	にっこり笑う (smile)
11	動*	386	すわらせる (sit)
12	名の動	361	湿気のある (moist)
13	名で動	312	ウインチで巻き上げる (winch)
14	連名	310	(車・機械などの) 突然の故障 (breakdown)
15	動	302	(人・国民などを) しいたげる (tyrannize)

(表1の補足) 品詞の表示

記号	品詞	記号	品詞
名	名詞	動	動詞
形	形容詞	静	静詞 (形容動詞)
副	副詞	連	連体詞
*	未知語	(以外)	助詞、助動詞など

表1から出現件数の多い品詞傾向を概観することができる。次節では特に頻度の高い結果が出た名詞と動詞を含むものを取り上げる。動詞を含む表現は結合価パターン対辞書 [4] への収録項目として特に重要なものである。

検討対象とした英和辞書の記述上の特徴として日本語訳にひらがな書きが多く、それが原因となって形態素解析に失敗したケースが散見された。例えば、表1のNo.11の例で「すわらせる」は「吸う」の未然形と未知語に解釈されている（この場合は「座らせる」と漢字書きすれば正しい結果が得られる）。また、日本語訳は文の形ではないにもかかわらず、単純に形態素解析を行なったため、形態素解析が入力データを文とみなしたことによる誤りもあった。例えば、「正確な」では「な」を文末表現として解析しようとしたため助詞となっている。本稿では、典型的な誤りはパターンマッチにより補正した。

今後の課題として、形態素解析の適用の仕方自体に工夫の余地がある。例えば、英語の品詞を踏まえて形態素解析の条件を調整するのが有効ではないかと考えられる。

### 3 名詞と動詞を含む表現の分類

本節では、英単語の日本語訳に名詞と動詞を含む場合を取り上げる。そのうち件数の多いものを表2のように8つに分類してそれぞれの傾向を探った。以下、各分類の上位を表3～表10に示す。

表2：英単語の日本語訳に名詞と動詞を含むもの

No	分類 (品詞列の例)	件数
1	動詞で終わるもの (ex. 名を動, 名に動て動)	4011件 →表3
2	名詞で終わるもの (ex. 名を動名, 名のように動名)	1974件 →表4
3	助動詞「た(だ)」で終わるもの (ex. 名を動た, 名の名を動た)	750件 →表5
4	助動詞「て(で)」で終わるもの (ex. 名を動て, 名から動て)	360件 →表6
5	助動詞「ない」で終わるもの (ex. 名を動ない, 名を動れない)	334件 →表7
6	助動詞「せる」で終わるもの (ex. 名を動せる, 名を動て動せる)	134件 →表8
7	形容詞で終わるもの (ex. 名に動形, 名と動て形)	51件 →表9
8	その他 (ex. 名に動れた, 名と動ての)	290件 →表10

表3では、日本語の格要素と動詞が英語の1単語に対応している。厳密には、英単語が動詞の場合は日本語の動詞は終止形、英単語が形容詞の場合は日本語の動詞は連体形という違いはあるが、いずれもこの格要素と動詞の関係をうまくつかまえることが重要であると思われる。これらは基本的には結合価パターン対辞書に収録したい。

表4には、日英翻訳の観点では英和辞典の訳語に問題がありそうなものが多い。一見して訳語が説明的であり、通常の日本語の文章には現れないような表現が多いからである。従って、これらの語を日英翻訳用の辞書に収録して

も効果は少なく、収録する場合には現実に使われそうな表現に置換するなどの措置が必要である。

表3：動詞で終わるもの (20 度数以上)

No	品詞列	度数	英単語の日本語訳の例 (英単語)
1	名を動	1449	園芸をする (garden), 武器をとる (arm), 目をさましている (wake)
2	名に動	785	隔日毎に起こる (tertian), モザイク模様にする (tessellate)
3	名の動	361	興味のある (interesting), 見晴らしのきく (commanding)
4	名で動	312	電気でうごく (electric), ちょうつがいで動く (hinge)
5	名が動	170	目がさめる (awake), (…ついて) 釈明する義務(責任)がある (accountable)
6	名動	114	ペン書きする (pen), (事隔など)2 年間続く (biennial)
7	名の名を動	65	刑の執行を猶予する (respite), 人の心を動かす (touching)
8	名と動	62	(火が) ゴーゴーと燃える (roaring), 推進体となる (propellant)
9	名から動	58	エピソードから成る (episodic), 列車から降りる (detrain)
10	名を動て動	42	帆を張って走る (sail), ふるいを通して落ちる (sift)
11	名の名に動	28	多くの目的にかなう (all-around), 個人の参加による (participatory)
12	名と動て動	25	代表として派遣する (delegate), (水が) 小滝となって落ちる (cascade)
13	名も動	21	恐れがもなく (audaciously), 人もうらやむ (enviable)

表4：名詞で終わるもの (20 度数以上)

No	品詞列	度数	英単語の日本語訳の例 (英単語)
1	名を動名	534	戦車を駆る人(御者) (charioteer), 詰め込み勉強をする学生 (crammer)
2	名に動名	239	相手に追いつき追い越そうとすること (catch-up)
3	名の動名	114	米国の目指す理想社会 (American dream)
4	名で動名	63	レクリエーションで過ごす日曜日 (Continental Sunday)
5	名を動た名	55	(羊の群れを先導する) 鈴をつけた雄羊 (bellwether)
6	名動名	50	一見すること (glimpse), 空飛ぶ円盤 (flying saucer)
7	名が動名	45	コンピュータがわかること (computer literacy)
8	名の名を動名	37	最新流行の服装をしている人 (fashion plate)
9	名と動名	36	中心となる物(人) (backbone), ボーン(ゴーン) という音 (bong)
10	名に動た名	34	虚栄に満ちた社会 (Vanity Fair), 馬に乗った暴力団員 (night rider)
11	名の動た名	26	毎日のきまりきった家事 (chore), 味のついたソーダ水 (soda pop)

表5では、英語の品詞は概ね形容詞であり、日本語の動詞は連体形である。表3と同様に、日本語の格要素と動詞が英語の1単語に対応する。これらも基本的には結合価パターン対辞書に収録したい。

表6は、英語の品詞は副詞であり、日本語の表現は節の形式になっている。現在の結合価パターン対の記述の枠組は単文レベルであるため、このような表現は記述できない。他の枠組での対処 [5, 6] を検討する。

表5：助動詞「た(だ)」で終わるもの(5度数以上)

No	品詞列	度数	英単語の日本語訳の例(英単語)
1	名を動た	190	腹をたてた(angry), ターバンを巻いた(turbaned), 心を傾けた(bent)
2	名に動た	183	若さにあふれた(youthful), 最初に生まれた(first-born)
3	名の動た	120	地味のこえた(fertile), 支払い期限の過ぎた(delinquent)
4	名で動た	48	油であげた(fried), 試験管内でできた(test-tube)
5	名が動た	31	感覚がまひした(anesthetic), 頭がぼんやりした(woolly-headed)
6	名動た	24	遠くへだたった(remote), 自ら認めた(confessed)
7	名を動だ	14	足を組んだ(cross-legged), つづれ織りを織り込んだ(tapestried)
8	名に動だ	14	表情に富んだ(expressive), 十分にしみ込んだ(saturated)
9	名の名を動た	14	二重の目的を持った(double-barreled), スカーフ状の物を掛けた(scarfed)
10	名から動た	12	(人名・句など)書から出た(Biblical), 上から見おろした(bird's-eye)
11	名と動た	9	いつもとちがった(unusual), 意気ようとした(triumphant)
12	名の名に動た	6	(ニットの服などが)体の線にぴったりした(full-fashioned)

表6：助詞「て(で)」で終わるもの(5度数以上)

No	品詞列	度数	英単語の日本語訳の例(英単語)
1	名を動て	119	流れを下って(downstream), 熱心さを欠いて(halfheartedly)
2	名に動て	96	農業に関して(agriculturally), 浅瀬に乗り上げて(aground)
3	名と動て	51	ふつうとちがって(unusually), 教区牧師として(vicarially)
4	名が動て	33	目がさめて(awake), 粘着性がある(stickily)
5	名に動で	10	弾力に富んで(elastically), 涙にむせて(sobbingly)
6	名動て	9	いつもきまって(invariably), 1時間続いて(hourlong)
7	名を動で	7	水銀を含んで(mercurially)
8	名から動て	5	異分子から成り立って(heterogeneously)

表7：助動詞「ない」で終わるもの(5度数以上)

No	品詞列	度数	英単語の日本語訳の例(英単語)
1	名を動ない	106	おそれを知らない(fearless), 上着を着ていない(shirtsleeve)
2	名の動ない	66	(草木の)実のならない(barren), 言い訳の立たない(indefensible)
3	名に動ない	50	伝統にとらわれない(Bohemian), 法廷に出頭しない(default)
4	名が動ない	30	気が置けない(affable), (ひな鳥が)羽が生えそろうていない(unfledged)
5	名に動れない	6	人に知られていない(out-of-the-way), 規則に縛られない(go-as-you-please)
6	名を動れない	5	責任を問われない(irresponsible), 心を動かされない(unaffected)

表8：助動詞「せる」で終わるもの(2度数以上)

No	品詞列	度数	英単語の日本語訳の例(英単語)
1	名を動せる	86	波をたてる(wave), 親分風を吹かせる(bossy)
2	名に動せる	11	兵役につかせる(enlist), (人を)社会生活に順応させる(socialize)
3	名と動せる	5	フッ素と化合させる(fluorinate)
4	名で動せる	5	(人を)難問で当惑させる(pose)
5	名の名を動せる	4	人の心を感動させる(eloquent)
6	名動せる	3	(人に)一杯くわせる(cod)
7	名を動て記動せる	3	人をだまして…させる(beguile)
8	名を動て動せる	3	パン種を入れてふくらませる(leaven)
9	名に名を動せる	2	手に汗をにぎらせる(exciting)

表7では、表5と同様に、英語の品詞は形容詞であり、日本語の助動詞「ない」は連体形である。表3、表5と同様に、日本語の格要素と動詞が英語の1単語に対応する。これらも基本的には結合価パターン対辞書に収録したい。

表8は、表3と同様に、日本語の格要素と動詞句が英語の1単語に対応する場合である。助動詞「せる」の付加は、使役表現というよりは自動詞を他動詞化している場合が多いようである。また厳密に言えば、英単語の品詞が動詞の場合には日本語の助動詞「せる」は終止形であり、英語の品詞が形容詞の場合には日本語の助動詞「せる」は連体形であるという違いはあるが、いずれも結合価パターン対辞書に収録したい。

表9において、表面的には英単語の品詞は形容詞と副詞であり、対応する日本語の表現は連体形と連用形(特に、副詞形)である。格要素と動詞からなる表現に様相が付加されたもの(No.1の「人目につきやすい」=人目につく+やすい)のうち、末尾の形容詞が連体形の場合は結合価パターン対辞書に収録したいタイプであるが、同じタイプで連用形のものは表6と同様の問題があるほか、No.6

の従属節を伴う表現や No.7 の複文表現などは現在のパターン対の枠組みでは記述できない。これらへの対応については今後の検討課題である。

表 9：形容詞で終わるもの（2 度数以上）

No	品詞列	度数	英単語の日本語訳の例（英単語）
1	名に動形	12	人目につきやすい (conspicuous), 大目に見てやってもよい (forgivable)
2	名を動形	6	事故を起こしやすい (accident-prone), 人を信じやすい (confiding)
3	名が動形	3	気分が変わりやすい (moody), 地震が起きやすい (seismically)
4	名の動形	2	気の変りやすい (mutable), 地震の起きやすい (seismic)
5	名を 動て形	2	(髪が) くしを入れてない (unkempt), ニスを塗ってない (unvarnished)
6	名と 動て形	2	長としてふさわしい (masterful), 長としてふさわしく (masterfully)
7	名を 動名が形	2	度を失うことがない (equable), (場所が) 風雨を避ける所がない (inhospitable)

表 10 を概観すると、表 3 から表 9 までは述べた考え方を準用できる場合が多いと思われる。例えば、項番 1, 2, 5 などは表 5 の助動詞「た (だ)」で終わるものと同様である。しかし、個別には複雑な単語構成のものもあり、日本語表現としてはかなり不自然であるため、これらの扱いについても今後の検討課題である。

表 10：その他

No	品詞列	度数	英単語の日本語訳の例（英単語）
1	名に動れた	32	(…に) あっけにとられた (agape)
2	名で動れた	26	ストライキで閉鎖された (strikebound)
3	名を動ずに	10	手段を選ばずに (catch-as-catch-can)
4	名に動れる	8	定期に刊行される (periodical)
5	名を動れた	7	(法律上の) 権利を奪われた (aggrieved)
6	名と動ての	7	市民としての (civic)
7	名に動られた	6	雪に閉じ込められた (snowbound)
8	名が動ずに	6	予測がつかずに (unpredictably)
9	名を動す	5	版權をおかす (pirate)
10	名の名に 動れた	5	作者の死後に出版 (発表) された (posthumous)
11	名に動ずに	5	比較にならずに (incommensurably)
12	名で動れる	5	(コーヒーが) パーコレーターでろ過さ れる (percolate)
13	(4 種類)	各 4	[名を動られた] ハンディキャップを つけられた (handicapped)
14	(7 種類)	各 3	[名動ぬ] いつもかわらぬ (regular)
15	(20 種類)	各 2	[名の動ないほど] 言い訳の立たない ほど (inexcusably)
16	(91 種類)	各 1	[名形動ないくらいに] これ以上小さ くできないくらいに (irreducibly)

## 4 おわりに

英和辞書全体を概観して、英単語に対応する日本語の表現が複数文節から構成される場合、格要素と動詞を含んでいるものが相当多いことが分かった。特に助詞「を」を含むものが 1/3 以上あると思われる。

このように熟語訳になっているものは、機械翻訳の品質向上を図る上で大きなカギとなる。ルールベースの機械翻訳ではこのような組み合わせ訳となるものを充実させていく必要があると考えられる。これらは結合価パターン対として辞書に収録することが望まれる。

以前、日本語例文に対する英訳文を翻訳家に作成してもらったところ、日英の文要素が対応しにくいものが数多くあることを報告した [7]。例えば次のようなものである。

1. 彼は樽の栓をあけた。  
He tapped the barrel.
2. 彼は護衛の任についた。  
He was assigned as a bodyguard.

これらの文において無理をして文要素を対応づけると、1 では「の栓をあける」→ “tap”，2 では「の任につく」→ “be assigned” のようになる。いずれにおいても単語対単語の翻訳を基本に考えていたのでは適切な訳を得ることができない例である。

これらはいわば偶然見つかったものであり、このようなデータをどのようにして効率よく得るかが課題であった。個別の例文を提示すれば翻訳家は適訳を与えてくれるが、網羅的にルール化するのとは不可能であった。本稿の検討により、英和辞書の日本語表現には通常日本語文書では使用されないような説明的な表現が多い。その扱いについては今後検討していく予定である。

## 参考文献

- [1] 池原, 宮崎, 横尾: 日英機械翻訳のための意味解析用の知識とその分解能, 情報処理学会論文誌, Vol. 34, No. 8, pp. 1692-1704 (1993).
- [2] 研究社: ライトハウス和英辞典 (1984) など.
- [3] 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, 解説編 & 辞書編 (1987).
- [4] Shirai, S., Ikehara, S., Yokoo, A. and Inoue, H.: The quantity of valency pattern pairs required for Japanese to English machine translation and their compilation, NLP RS 95, pp. 443-448 (1995).
- [5] 白井, 池原, 河岡, 中村: 日英機械翻訳における原文自動書き替え型翻訳方式とその効果, 情報処理学会論文誌, Vol. 36, No. 1, pp. 12-21 (1995).
- [6] Matsuo, Y., Shirai, S. and Ikehara, S.: Changing syntactic classes in transfer-based machine translation, NLP RS 95, pp. 432-437 (1995).
- [7] 池原, 白井, 相沢: 和語動詞に対する日英対訳用例文の収集について, 言語処理学会第 2 回年次大会 B6-3, pp. 253-256 (1996).