

IPAL 動詞辞書の用例文に基づく日英翻訳用結合価パターン対の収集

4L-4 白井 諭^{*1} 井上 浩子^{*2} 小出 ひとみ^{*2} 井田倉 紀子^{*2} 横尾 昭男^{*1}^{*1}NTT コミュニケーション科学研究所^{*2}NTT アドバンステクノロジー(株)

1 はじめに

機械翻訳における意味解析には、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。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、慣用表現を除き、日本語の動詞を訳し分けられるようなパターン対が記述できることが知られている[池原 93]。

収集方法の問題に対して、筆者らは、人用の辞書に記載されている情報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。その過程で、和英辞書から抽出したパターン対だけでは現実の文書に含まれる用言の訳し分けには不十分であること[白井 94]、それを補完するには IPAL 動詞辞書[IPA87]のような精密な語義分類に基づく用例文とその英訳文の利用が有効であること[白井 95]を報告した。

本稿では、IPAL 動詞辞書に記載されている日本語の用例文(5,243 文)とそれらの文の翻訳家による英訳文とから抽出されたパターン対の数とその種類について概観する。また、IPAL 動詞辞書に記載された語義数とパターン対の数の関係についても検討する。

2 パターン対の収集結果

IPAL 動詞辞書[IPA 87]には、和語動詞 861 語(ひらがな表記の異なり数、漢字かな交じりにすると約 1,200 語)の精密な語義や語義毎の用例文等が記載されている。この用例文(5,243 文、約 7.5 万字)に対して、日本語原文に忠実で、かつ、英語としても十分通用する英訳文(約 4 万語)を翻訳家に作成してもらった。その対訳データからアナリストがパター

ン対を抽出したところ、対象動詞に対する既存のパターン対 2,603 件に対して、新規に 1,532 件が作成され、既存の約 500 件の修正情報が得られた。

作業過程では、複数の対訳データが 1 つのパターン対に集約可能な場合に注意した。次の例からは、別のパターン対を作ることでもできるが、共通に使えるパターン対を作ることにより汎用化をねらう。

1. 彼は頭に手拭いを当てた。

He put a towel on his head.

2. 彼女は傷口に消毒用ガーゼを当てた。

She placed a sterile gauze on the injury.

日: N1(主体)が N2(布, ガーゼ)を N3(*)に 当てる
英: N1 put N2 on N3

また、予想された範囲ではあるが、次のような問題もあり、個別に対応を検討した。②の問題は見逃す恐れがあるため、特に注意を払った。

①日本文と英文で格要素の対応が取りづらい

3. 彼女はさっと野菜の水を切った。

She shook the water off the vegetables.

対処: 日本語の格要素を英語側で分割使用する。

日: N1(人)が N2[N3(具体)の | N2(水)]を 切る
英: N1 shake N2 off N3

4. 運営委員会は Aチームを初戦で Bチームに当てた。

The steering committee made Team A compete with Team B in the first game.

対処: 保留(英語は使役構文を使用している)。

②英文の動詞が汎用的に使えない可能性がある

5. 彼女は濡れた衣服を風に当てている。

She is drying her clothes in the wind.

対処: この文の訳としては何ら問題ないが、単に風を当てる場合には使えないので、一般的な hang を使ってパターン化する。

③英文では句として訳出された場合

6. 子供がサンタクロースに宛てて手紙を書いた。

Children wrote to Santa Claus.

対処: 保留するが、他処理[松尾 95]で救済予定。

Extraction of Japanese-to-English Valency Pattern Pairs from Example Sentences in the IPAL Basic Verb Dictionary

Satoshi SHIRAI^{*1} Hiroko INOUE^{*2} Hitomi KOIDE^{*2},
Noriko ITAKURA^{*2} and Akio YOKOO^{*1}

^{*1}NTT Communication Science Laboratories and
^{*2}NTT Advanced Technology Corporation

3 語義数とパターン対の数の関係

語義数との関係を見るため、IPAL 動詞辞書の見出し項目ごとにパターン対の数を集計した。ただし、動詞の漢字表記が他の見出し項目のものと重なっている場合は、表面にはその識別が困難であることから、IPAL の見出し項目を合併させた。具体的には、次の8つである。

- あげる 2→1, うまる/うずまる(埋まる) 2→1,
- うめる/うずめる(埋める) 2→1, かわる 2→1,
- さす 3→1, する(五段活用のもの) 2→1,
- はいる/いる(入る) 2→1, ふく 3→2

IPAL 語義は日英翻訳のための分類ではないが、多義語ほど細かく訳し分ける必要があるため、IPAL の詳細な語義分類が参考になる可能性がある。そこで、語義数とパターン対数の関係について調べてみることにした。図1に語義数と既存(IPAL 用例からのパターン対抽出前)のパターン対数(2,603件)との関係、図2に IPAL 語義数と IPAL 用例から新規に抽出されたパターン対数(1,532件)との関係、図3に抽出後の合計パターン対数(4,135件)との関係を示す。ただし、語義数とパターン対数が同じ動詞は図中の同じ点を占めている。

図からわかるように、IPAL 用例からパターン対を抽出した後では、語義数とパターン対数の相関関係がかなり高くなっている。これは、語義対応に収録されている用例を用いたため、ある意味では当然といえるかもしれない。しかし、IPAL 用例からパターン対を抽出する前の段階から、既に語義数とパターン対数の間には一定の相関関係が認められる。従って、IPAL 語義は日本語としての語義分類であり、日英翻訳における訳し分けを意図しているわけではないが、日英翻訳を考える上でも一定の手がかりを与えているといえる。

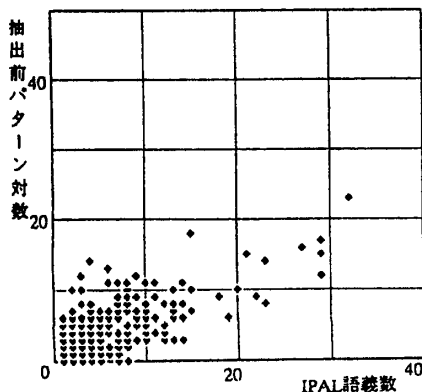


図1 語義数と抽出前パターン対数

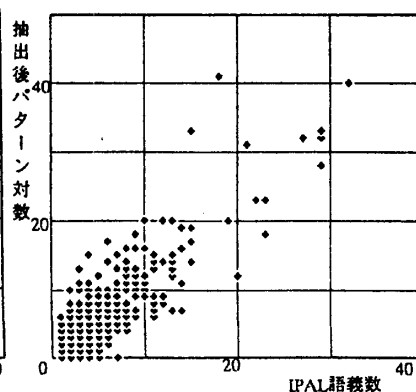


図2 語義数と抽出後パターン対数

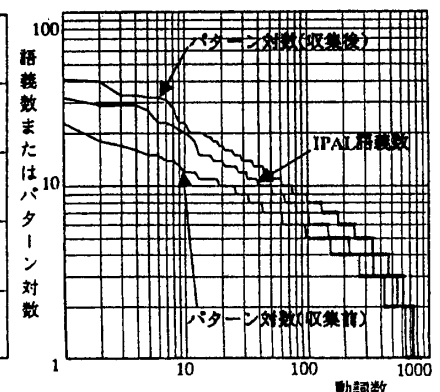


図3 語義数またはパターン対数の分布

別の観点からの考察として、動詞ごとのパターン対数または IPAL 語義数を多い順に配列した結果を図3に示す。図から、パターン対数順と IPAL 語義数順とも Zipf の法則のような右下がりの関係が見られる。この前提に立てば、収集前パターン対数の傾きの変化が少ないのに比べ、IPAL 語義数、収集後パターン対数とも、動詞数で見て 1~10 の範囲の傾きが 10~100 の範囲の傾きに比べ緩やかであり、パターン対の一層の拡充の必要性が予想される。

4 おわりに

本稿では、日英機械翻訳に有効な結合価パターン対を、IPAL 動詞辞書に記載されている用例文とそれらの文に対する翻訳家による英訳文とから収集した。その結果、IPAL の語義分類は日本語の語義分類であって日英翻訳が目的ではないが、5,243 文から 1,532 件の新パターン対が得られ、語義数とパターン対数には一定の相関関係があることなどがわかった。

動詞ごとのパターン対数の分布状況を見ると、多義語のパターン対はかなり不足していると予想される。これについては、並行して作成している用例文[池原 96]からパターン対を収集することにより、不足分を補っていく予定である。

参考文献

[池原 93] 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
 [池原 96] 池原,白井,相澤: 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
 [IPAS7] 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
 [松尾 95] Matsuo, Y., Shirai, S. & Ikehara, S.: Changing syntactic classes in transfer-based machine translation, *NLPRS 95*, pp.432-437
 [白井 94] 白井,横尾,池原,井上: 日英対訳用構文意味辞書の記述精度の向上と作成支援, 情報処理学会第 48 回全国大会 6Q-9, pp.3-141-142
 [白井 95] Shirai, S., Ikehara, S., Yokoo, A. & Inoue, H.: The quantity of valency pattern pairs required for Japanese to English machine translation and their compilation, *NLPRS 95*, pp.443-448