

## 複合和語動詞に対する日英対訳用例文の収集について

5R-7

白井 諭<sup>\*1</sup> 大山 芳史<sup>\*1</sup> 武智 しのぶ 分部 恵子<sup>\*2</sup> 相澤 弘<sup>\*2</sup><sup>\*1</sup>NTTコミュニケーション科学研究所 <sup>\*2</sup>株式会社ミュー

## 1 はじめに

機械翻訳の精度向上には、用言と体言の意味的共起を記述した対訳の結合価パターンの使用が有効である。市販の和英辞書等からは十分なパターン対を収集することはできないという問題があった。

筆者らは、作例文とその訳文からパターン対を抽出する方法を提案した[白井95]。和語動詞に対する対訳用例集の作成を試みた結果、辞書に未収録の用法が多数収集され[池原96]、網羅的にパターンが収集される見通しである。形容詞に対しても同様に対訳用例集を作成している[白井97a]。

本稿では、作例文の得失について再整理するとともに、次の対象として、和英辞書には記載の少ない複合和語動詞の用例文の収集結果について報告する。

## 2 結合価パターン対の収集について

## 2.1 収集上の問題

既に報告したように[白井95,池原97]、結合価パターン対の収集は、①和英辞書から、②日本語辞書の例文から、③作例文から、という経過で、現在は③のステップを進めている。③により、結合価パターン対を網羅的に収集できる見込みである。

ところで、和英辞書からの収集が終わった時点では、a.引いた単語が載っていない、b.単語はあるが適当な用法が載っていない、が切実な問題となった。人も同様の問題に直面するが、別の表現に言い換えることによりこの問題を回避する。機械処理では言い換えることができないため、機械辞書にはこれらをあらかじめ収集しておく必要がある。

ステップ①が完了した時点で技術文書を対象として評価実験を行なったところ、異なりで、出現した用言の15%は機械辞書にエントリがなく、また、必要と思われるパターン対の23%は機械辞書に登録されていなかった。現在行なっているステップ③は、②で対象とした日本語辞書[IPA87,90]に収録されている動詞と形容詞を対象としていることから、bに該当するといえる。

Compiling Japanese and English Corpus for Compound Verbs of Japanese Origin

Satoshi SHIRAI<sup>\*1</sup>, Yoshifumi OYAMA<sup>\*1</sup>, Shinobu TAKECHI, Keiko WAKEBE<sup>\*2</sup>, and Hiroshi AIZAWA<sup>\*2</sup>

<sup>\*1</sup>NTT Communication Science Laboratories and <sup>\*2</sup>Mu. Inc.

次のステップとして、aの場合、すなわち、和英辞書には収録されていなかった用言のパターン対の収集が必要である。サンプル調査によれば、国語辞書に収録されている用言のうち、複合和語動詞、サ変動詞、ナ型形容詞の不足が目立った[池原97]。本稿では、このうち複合和語動詞を対象にする。

## 2.2 作例の得失

作例に関しては、常に恣意性が問題になる。言語コーパスに基づくのが理想的ではあろうが、比較的容易に収集されるものは機械辞書に収録済みであることが多く、現実には機械辞書に収録されていないタイプの用例を大量に収集するのは困難である。また、国語辞書にはエントリはあっても、前節のaやbでは用例の記載がない場合がほとんどで、IPAL辞書を利用した②と同様の進め方は不可能である。

ところで、③の作例文による用例収集は、一般に作例の恣意性が問題とされる場合とは若干事情が異なっていると思われる。作例の恣意性の問題の1つに、説明に合うよう作られた例文の「不自然さ」がある。③では用言を指定する以外には表現に細かな制約を設けないため、この種の恣意性は混入しにくいと思われる。また、数度唱えるうちに「不自然さ」が減少するなどの問題もあるが、時間をおいて作例文を見直すことにより恣意性を緩和することが可能である。本稿とは直接関係ないが、先に作成した和語動詞や形容詞の例文の見直しを進めている。

そこで本稿でも、用言（複合和語動詞）を指定して自由な発想により例文を作成することとした。

## 3 対訳用例文の作成

## 3.1 作成の方針

和語動詞や形容詞の例文作成[白井97b]の方針を参考にして、次の方針とした。

- ①現代国語例解辞典[林85]に収録されている複合和語動詞を対象とし、辞典の語釈や例文、または類推により例文を作成する。
- ②動詞のニュアンスが異なるものを例文として広く集めることとし、可能な限り「一般的で単純な名詞を格要素に持つ単文」とする。
- ③動詞が終止形である例文だけでなく、連用形や連体形のものも必要に応じて収集する。
- ④動詞1語あたり2文を例文作成の目標にするが、例文が思いつかなくなるまで作成を行なう。
- ⑤収集された日本語の例文を、日本語原文に忠実

でかつ英語として十分通用することを条件に、翻訳家に英訳してもらおう。忠実な訳が困難な場合は最小限度の意識は許容する。

複合和語動詞としては、和語動詞2つが組み合わさって1語化した「踏み出す」(“とりかかる”の意味の場合)のようなものを当初は想定したが、接頭語+動詞(相次ぐ, 等), 名詞+接尾語(油染みる, 等)のほか、動詞転成名詞を含むサ変動詞(一本立ちする, 受け答えする, 等)も含めることにした。

作業過程において、複合和語動詞の語義解釈に個人差が生じる場合があることがわかった。例えば、「こづきまわす」では、極めて大きなダメージを受けると感じる人と、結果としては大したダメージではないと感じる人とがいるようである。複数の作業者が作例文を相互チェックすることにより、中立的な文になるように努めた。

### 3.2 収集結果

ステップ①により、(広義の)複合和語動詞として2,542語が抽出された。このうち、IPAL動詞辞書[IPA87]と一致する14語(あがく, 等)は対象外とし、用例が思いつかない語や現代語としてあまり使われない語(例えば, あやどる)は例文作成作業を進める過程で適宜除外した。

この結果、最終的には動詞2,105語に対して3,717用例文が作成された。動詞1語あたりの作例文数と動詞語数の関係を表1に、これを両対数グラフにプロットしたものを図1に示す。作例文数は、「うちこむ」「とびだす」の各9件、「いれかえる」「ひっかける」の各8件などと予想より多く作成された反面、1例文のみのものが1,062語となった。

表1 複合和語動詞の用例文収集

作例数	9	8	7	6	5	4	3	2	1
動詞数	2	2	3	12	25	79	247	673	1062

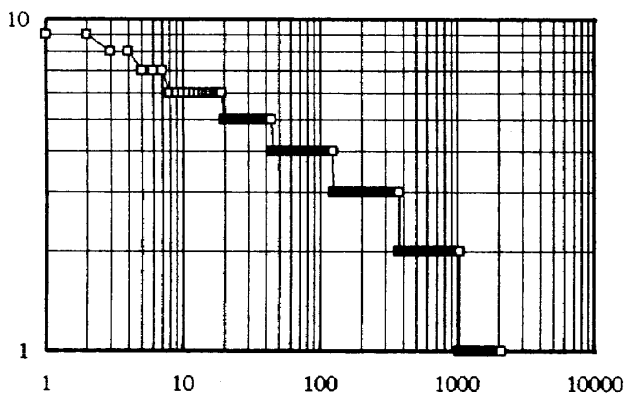


図1 作例文数(縦軸)と動詞語数(横軸)の関係

作成された用例文集全体の諸元を表2に示す。和語動詞や形容詞より文が長めになっている。その原因として、動詞自体の字数が多いこと、限定された場面で使われるため背景説明を必要とすることが挙げられる。例えば、次の文では「大きな揺れに」は

最低限度の追加説明であると言える。

J: 大きな揺れに人々は外へ飛び出した。

E: People dashed out (into the street) because of the severe shaking.

表2 用例文集の諸元の比較

種別	語数	用例文数	和文字数		英文語数	
			字/文	平均	語/文	平均
複合和語動詞	2,105	3,717	61,700	16.60	30,339	8.16
和語動詞	861	10,497	129,926	12.66	66,322	6.46
形容詞	136	1,909	24,389	12.78	12,707	6.66

なお、作例文の生産性は、最初の文の作成に要した時間  $t$  に対し、 $n$  文目は  $2^{n-1}t$  の傾向を示し、和語動詞や形容詞の場合とほぼ同様であった。

### 4 おわりに

和英辞書を使用する際、単語自体の収録がない場合と単語はあるが適当な用法の記載がない場合の2つの問題がある。機械用の辞書ではいずれの未収録も致命的であるため、何らかの手段で収録しておく必要がある。本稿では、前者の問題に対処するため、和英辞書にあまり収録されていない複合和語動詞を対象として用例文の作成を行ない、2,105語に対し3,717文が収集された。今後は収集した用例文に基づいて結合価パターン対を収集する予定である。

先に行なった和語動詞や形容詞の用例文収集は、和英辞書に適当な用法の記載がない場合に対応すると考えられ、本稿の作業と補完関係にある。本稿でも先の例文作成と同様の方法が実施可能であったことから、コーパスが収集されないという問題に対する対処法として十分有効であると考えられる。なお、作例文に対する恣意性を除去するため、作成済みの用例文の見直しも並行して進める予定である。

#### 参考文献

- [林85] 林巨樹編: 現代国語例解辞典(第一版), 小学館(1985)
- [池原96] 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)
- [池原97] 池原, 官崎, 白井, 横尾, 中岩, 小倉, 大山, 林: 日本語語彙大系, 岩波書店(1997)
- [IPA87] 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)
- [IPA90] 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形容詞辞書IPAL, 解説編 & 辞書編(1990)
- [白井95] S. Shirai, S. Ikehara, A. Yokoo & H. Inoue.: The quantity of valency pattern pairs required for Japanese to English machine translation and their compilation, NLP95, pp.443-448, Seoul (1995)
- [白井97a] 白井, 横尾, 池原, 武智, 分部: 形容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3回年次大会, D4-6, pp.469-472 (1997)
- [白井97b] 白井, 池原, 相澤, 鳴海, 横尾: 結合価パターン対作成のための日英対訳用例文の収集, 情報処理学会研究報告, 97-NL-122-1, pp.1-6 (1997)