形容詞に対する日英対訳用例文の収集について

白井 諭[*1] 横尾昭男[*1] 池原 悟[*2] 武智しのぶ[*3] 分部恵子[*3]

[*1]NTTコミュニケーション科学研究所 [*2]鳥取大学 工学部 [*3]㈱スバルドキュメンテック

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。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，慣用的な表現や特定の専門分野に固有の表現を除き，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[池原93]。

収集方法の問題については，市販されている人間用の和英辞書等の対訳例文や語釈からパターン対を作成する方法では十分な件数が収集できないことが挙げられる。このため，対訳コーパスから結合価パターンを収集する様々な方法が検討されている[宇津呂92] [Almuallim 94][北村95]が，網羅的にパターン対を収集できるめどは立っていない。

これに対して，筆者らは，人間用に辞書に収録されている情報よりも，人間自身が持つ知識の方が柔軟かつ多彩であると考えて，人手による結合価パターン作成用の例文作成を試みた。既に報告したように，IPAL動詞辞書[IPA87]に収録されている和語動詞861語に対しては，1.5人年の作業で10,500文が収集された[池原96]。この対訳用例文からは3,000～4,000パターン対が新たに収集できる見通しで，和語動詞に対しては網羅的なパターン対収集の見込みが得られたと考えられる[白井95]。

本稿では，和語動詞に対する対訳用例文収集の路線を形容詞に適用する。具体的には，IPAL形容詞辞書[IPA90]に収録されている形容詞136語に対する対訳用例文を収集した結果について報告する。

2 結合価パターン対とその収集について

2.1 パターン対記述の枠組み

日英翻訳システムALT-J/Eにおける意味解析用辞書の構成を図1に示す[池原93]。用言に対する結合価パターン対は，構文意味辞書として準備されるもので，2種類の意味属性（一般名詞意味属性と用言意味属性）を用いて，以下に示す基準で作成される。

① 日英の結合価パターンを対にして記述する。両者は，用言，格要素（主名詞＋助詞），副詞要素，様相情報から構成される。

② 日本語パターンでは，主名詞は，日本語の用言に対する英語の動詞が訳し分けられる最小限の深さの一般名詞意味属性を用いて記述する。

③ 主名詞を一般名詞意味属性で抽象化すると問題がある場合は名詞そのものを指定する。パターンの意味が個々の単語から推測できず，表現が固定的であれは，慣用表現パターンと呼ぶ。

④ 英語パターンには，動作動詞と状態動詞の別（進行形の可否），受身の可否，文脈処理のための用言意味属性[中岩97]を記述する。

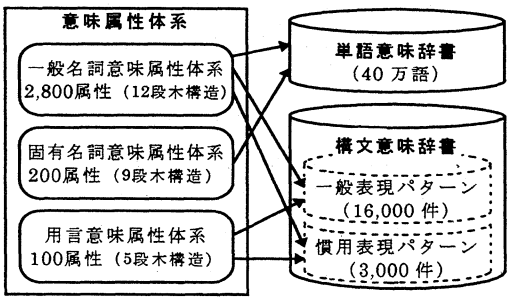

図1 ALT-J/Eにおける意味辞書の構成

2.2 パターン対収集の問題

従来，結合価パターン対は，和英辞書に記載された対訳例文に基づいて，人手により作成してきたが，翻訳実験では不足するパターン対が続出することが

問題であった。これを補うため，自動学習などの手法を応用しようとしても，単純な構造の対訳例文が大量に必要であることが問題となる。例えば，1,000和語動詞に対して1,000万文が必要であるという予想[Almuallim 94]もある。

パターン対の収集上の問題を整理すると，次の3点に集約される。

① 日英翻訳に必要なパターン対の数が不明

② それを網羅的に収集する方法が不明

③ 収集作業がコスト的に実行可能かが不明

パターン対の数はサンプル試験により推定でき，実行可能かは方法見合いであるので，網羅的に収集する方法を見いだすことが必要である。

3 対訳用例文の作成方針

3.1 基本的な考え方

人間が母語の文章を書くとき，その意味を確認するために国語辞書を参照するのはまれであり，記述内容や文脈に応じて適切な言葉を自然に選択して使用している。逆に，用例と国語辞書の語義を対応づけようとしても，容易ではない場合が少なからずある。こうしたことから，言葉の意味的用法に関する知識は，辞書に記載されている量よりも，人間の頭の中にある量の方が豊富であり，むしろ辞書にはその一部が記載されているに過ぎないと推測される。そこで，適切な刺激を与えて自分の知識を内省することにより，様々な用例を収集する方法が考えられる。その収集結果に基づいてパターン対を作成すれは，網羅的に収集されると期待される。

そこで，IPAL動詞辞書[IPA87]に収録されている和語動詞861語を対象に例文収集を試みたところ，1.5人年の作業で10,500例文が収集された[池原96]。その際の基本方針は次の2点であった。

① 日本語を中心に用法を思い浮かべ，ニュアンスの異なっているものを例文として書き出す。
＜備考＞英訳が異なるところまでは要求しない。結果的に訳語が同じになっても構わない。

②一般的な表現を集めることとし，慣用表現と意識されるものは集めない。
＜備考＞結果的に慣用表現に分類される用例が作成されても許容する。

本稿では形容詞を対象とすることから，次の条件を加えた。

③連用形と動詞，連体形と名詞が組み合わさった用法もなるべく収集する。
＜備考＞日本語の形容詞の連用形は副詞として働き，単純に英訳できない場合が少なからずある。また，連体形は被修飾名詞との組み合わせが慣用的な場合がある。

④二重主格構文のような特殊な表現も場合によっては収集する。
＜備考＞日本語の範囲ではあまり問題ではないが，英訳する上での課題表現として考える。

3.2 用例文の作成方法

具体的には，以下の方法で用例文を作成した。

① IPAL形容詞辞書に収録されている形容詞136語を対象に，IPAL例文や各種辞書などを参照しながら，類推により用例文を作成する。ただし，IPAL例文と同種と判断されるものは作成しない（IPAL例文で代表させる）。

② 作成する用例文は，可能な限り一般的かつ単純な名詞を格要素に持つ単文とする。

③ 連想が可能な限り用例文を書き出すこととするが，IPAL語義の3〜4倍を目安とし，用例文が思いつかなくなれば完了とする。

3.3 対訳文の作成について

こうして収集された日本語の用例文を翻訳家に英訳してもらう。このとき，日本語原文に忠実で，かつ，英語としても十分通用することを英訳の条件とする。ただし，忠実な訳が困難な場合は最小限度の意訳は許容する。

実は，この用例文の英訳はかなり困難な作業である。翻訳家はまとまった文章を，文の流れに沿って，意訳を織り交ぜながら英訳するのが普通で，1文ずつ独立した表現を忠実に翻訳することはまれである。このため，作業効率があまり上がらないことが問題となる。しかし，それよりも，日本語で類似した表

現が並んでいるため，日本人翻訳家は日本語のニュアンスの違いを感じてもその違いを英語では表現し切れないことがあり，逆に英米人翻訳家は日本語のニュアンスの違いを感じ取れないことがあるのが問題である。試行錯誤の結果，日本人翻訳家と英米人翻訳家がニュアンスの違いについて意見交換しながら作業を進められる場合に最もよい結果が得られることがわかった。

4 対訳用例文の収集結果

4.1 収集結果

前章で述べた方法により，IPAL形容詞辞書に収録されている形容詞136語を対象に，日本語の用例文を作成した。収集された用例文の数を表1に示す。

表1から，1形容詞あたりIPAL形容詞辞書に収録されている例文数とほぼ同じ14件の用例文が新たに作成されたことがわかる。また，これはIPAL語義の4倍弱に相当する件数である。この結果，IPAL例文と併せると3,800文あまりの対訳用例集が作成された。

表1　収集された用例文数

形容詞	IPAL形容詞辞書			作成した用例文			例文
語数	語義数	例文数	連体	作例数	連用	連体	合計
136	485	1917	25	1909	241	261	3826
平均	3.6	14.1	0.2	14.0	1.8	1.9	28.1

注）連用・連体は例文数または作例数の内数

例文数は形容詞によるばらつきが大きい。例えば，IPAL語義数13の形容詞には「わるい」「かるい」「よわい」の3語があるが，例文合計は「わるい」125文（IPAL例文56文+作例69文），「かるい」91文（同，44+47），「よわい」79文（同，41+38）となっている。また，IPAL語義数6の17形容詞では，例文合計は31〜91文（IPAL例文15〜48文，作例15〜43文）となっている。

そこで，語義数別に集計した結果を表2に示す。表2から，平均的に見る限りは，語義数の8倍弱のペースで例文が収集されたことがわかる。従って，このばらつきはそれほど問題ではないと考える。

表2　語義数別に見た平均例文数

形容詞	IPAL形容詞辞書			作成した用例文			例文
語数	語義数	例文数	連体	作例数	連用	連体	合計
1	14	57.0	0	77.0	5.0	19.0	134.0
3	13	47.0	1.3	51.3	8.7	4.7	98.3
2	12	43.0	0.5	44.0	7.0	4.0	87.0
1	11	41.0	0	42.0	6.0	6.0	83.0
1	10	42.0	1.0	42.0	6.0	4.0	84.0
2	9	36.0	0.5	36.0	10.0	4.0	72.0
6	8	29.3	0.3	26.0	4.2	5.2	55.3
6	7	26.5	0.2	24.2	3.3	2.5	50.7
17	6	25.1	0.2	22.2	1.8	2.4	47.3
6	5	22.0	0	19.5	0.5	1.7	41.5
13	4	14.1	0.4	14.2	1.5	2.7	28.3
25	3	13.2	0.2	12.8	1.5	1.4	26.1
29	2	6.0	0.1	7.8	1.0	1.1	13.8
24	1	5.3	0.0	6.4	1.0	1.5	11.8

注）連用・連体は例文数または作例数の内数

4.2 連用用法と連体用法

連用用法として収集された用例文は，「なる」「する」に接続するものがほとんどであった。IPAL形容詞辞書では，「くなる」などと注記されている。しかし，日英翻訳の観点では次のように一体的な訳語が与えられる場合が多いので，用例文の収集は有効であったと考えられる。

1. 彼は彼女の言葉に気をよくした。
   He was encouraged by her words.
2. 舗装で道がよくなった。
   Paving improved the road.
3. 彼の言葉に彼女は気を悪くした。
   What he said upset her.
4. この牛乳は悪くなっている。
   This milk has gone off.

また，少数ではあるが，次のように「なる」「する」以外の動詞との組み合わせも収集された。

5. 彼女は彼の言葉を悪くとった。
   She took badly what he said.
6. あの教師は生徒の事を悪く言う。
   That teacher speaks badly of the students.

一方，連体用法として収集された例文は雑多である。慣用的なものが多い。以下に例を示す。

7. 彼女は娘の旅行にいい顔をしない。
   She is not willing to let her daughter travel.
8. 私は悪い予感がする。
   I feel that something bad is happening.

4.3 生産性

日本語の用例文作成には約3カ月を要したので，1時間あたり約2.8文の用例文を作成したことになる。和語動詞の場合は1時間あたり約5文であった[池原96]ので，生産性は落ちている。連用用法や連体用法等を収集対象に加えたのが主な原因と思われる。しかし，辞書からは収集できない用法が多数得られたことを考えれば，十分満足できる範囲である。

また，例文作成をIPAL語義数との関係で見れば，IPAL語義数 n に対して，IPAL例文とは異なる n 文を作成するのに要した時間を t とすると，$2n$ 文の作成時間は $2t$，$3n$ 文では $3t \sim 4t$，$4n$ 文では $6t \sim 7t$ と和語動詞の場合とほぼ同じ傾向となった。

4.4 問題点

例えば，「あまい」に対しIPALは「1.味に甘みが多いと感じる」「2.飲食物などの味に甘みが多い」など10語義に分類している。この1と2はかなり微妙で，用例文としての違いを区別するのは困難である。

そこで，IPAL形容詞辞書の語義分類の特徴を簡単に考察した。語義数順の形容詞語数の分布を図2に示す。IPAL動詞辞書の語義分類の特徴[白井96]に比べると語数順位5～50あたりが細かく分類されているような印象があり，これが語義数を基準にした例文作成に影響を与えているかもしれない。

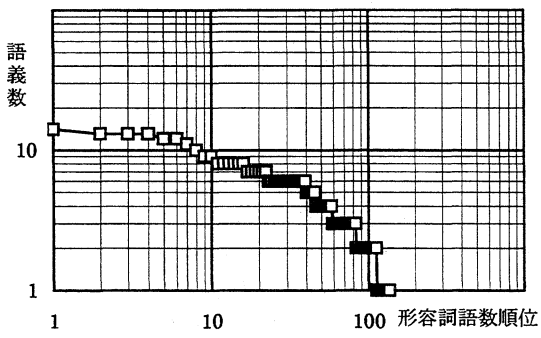

図2　IPAL形容詞辞書の語義分類の特徴

5 おわりに

日本語の形容詞の意味的用法について，辞書に記載されている情報よりも，人間の知識の方が柔軟かつ多彩であると考えて，人間の知識を内省する方法で形容詞の用例文作成を行なった。その結果，IPAL形容詞辞書に収録されている形容詞136語に対して，1,909文が収集され，IPAL例文と併せると3,826文の対訳用例集を作成することができた。

内省による例文作成の妥当性については，和語動詞の例文作成で2人が独立に例文を作成してみると，8割以上は同様の文を作成したことから，言語知識の個人差はそれほど大きくないと考える。客観性を追及する立場からは，コーパスから収集することが考えられるが，大規模コーパスの構築は容易ではないことに加え，有意性の高い表現は辞書に収録されていることが多いので，網羅性の点では疑問がある。以上から，現在の条件下では内省による例文作成が網羅性を追及でき，かつ，実施可能な方法である。

謝辞　対訳用例文収集にご協力下さった鳴海武史氏を始めとする㈱スバルドキュメンテックの関係各位に感謝する。

参考文献

[Almuallim 94] H. Almuallim, Y. Akiba, T. Yamazaki, A. Yokoo & S. Kaneda: A tool for the acquisition of Japanese to English machine translation rules using inductive leaning techniques, CAIA94, pp.194-201, San Antonio (1994)

[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 (1993)

[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)

[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 (Basic Verbs), 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 (Basic Adjectives), 解説編 & 辞書編 (1990)

[北村95] 北村,松本: 対訳テキストからの翻訳知識の獲得と機械翻訳システムへの応用, 言語処理学会第1回年次大会, C2-7, pp.289-292 (1995)

[中岩97] 中岩,池原: 日英の構文的対応関係に着目した日本語用言意味属性の分類, 情報処理学会論文誌, Vol.38, No.2, pp.215-225 (1997)

[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, NLPRS 95, pp.443-448, Seoul (1995).

[白井96] 白井,井上,小出,井田倉,横尾: 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, 情報処理学会第53回全国大会, 4L-4, pp.2－59-60 (1996)

[宇津呂92] 宇津呂,松本,長尾: 二言語対訳コーパスからの動詞の格フレーム獲得, 情報処理学会論文誌, Vol34, No.5, pp.913-924 (1993)