

## 3N-4 ナ型形容詞に対する日英対訳用例文の収集について

白井 諭<sup>\*2\*1</sup> 中岩 浩巳<sup>\*1</sup> 大山 芳史<sup>\*1</sup> 相澤 弘<sup>\*3</sup> 武智 しのぶ<sup>\*3</sup> 分部 恵子<sup>\*3</sup><sup>\*1</sup>NTT コミュニケーション科学基礎研究所 <sup>\*2</sup>ATR 音声翻訳通信研究所 <sup>\*3</sup>株式会社 ミュー

## 1 はじめに

機械翻訳における訳し分けの精度向上には、表現の共起の利用が有効である。筆者らは、日英翻訳を対象として、用言と体言の共起を記述した対訳の結合価パターン収集を進めている。その際、各パターンの適用範囲を広げるべく、文献[石綿 1983]に倣い意味属性[池原 1997]を用いて体言の最小限の抽象化を行なっている。

初期は市販の和英辞書等から一般表現 10,000 件と慣用表現 3,000 件を収集したが、高頻度用言の高頻度用法が収集されたに過ぎなかった[白井 1999]。高頻度用言の低頻度用法や低頻度用言の用法を収集するため、作成文とその訳文からパターン対を抽出する方法を開始した[白井 1995a, 1995b]。

高頻度用言の高頻度用法(動詞と形容詞の基本語[IPA1987, 1990])に該当する対訳用例文の収集を終え[白井 1997]、現在、低頻度用言(複合和語動詞、ナ型形容詞、サ変動詞)の用法を対象にした対訳用例文の収集を進めている[白井 1999]。本稿では、このうちナ型形容詞の用例文の収集結果について報告する。

## 2 対訳用例文の作成方針

これまでの例文作成の経過[白井 1995a, 1997, 1999]を踏まえ、次の方針で対訳用例文を作成することとした。

- ① 現代国語例解辞典[林 1997]に収録されているナ型形容詞を対象とし、辞典の語釈や例文を参照し、または類推することにより例文を作成する。

<備考> 自然な例文を作成できる語を対象とすることとする。まず、辞書収録語のうち、現代語として不適当と思われる語を除外する。次に、例文の作成を行ないながら、例文作成が困難なものを、例文作成者の判断で対象から除外する。

- ② 例文作成者の主観で、用言のニュアンスが異なると感じられるものを例文として広く集め、可能な限り「一般的で単純な名詞を格要素とする単文」とする。

<備考> 日本語表現の作成作業として位置付け、英訳が異なるところは要求しない。結果的に訳語が同じでも構わない。

- ③ 用言が終止形で終わる例文だけでなく、連用形や連体形の用法にも留意して収集する。

<備考> 連用形の副詞用法や連体形の限定修飾には慣用的なものがあり、それらの収集を意図する。これは、[白井 1995a]に対する橋本三奈子氏のご示唆による。

- ④ 用言 1 語当たり 2 例文を最低目標とし、ある程度考えても例文が思いつかなくなるまで作成を行なう。

<備考> これまでの経験では、 $n$  文の作成時間を  $t$  として、おおよそ  $t$  は  $2$  の  $n$  乗に比例する。具体的には、10~15 分考えても新たな例文が思いつかなければ打ち切る。

- ⑤ 収集された例文に対して、日本語原文に忠実で英語として十分通用するように、翻訳家に英訳してもらい。忠実な訳出が困難な場合は最小限度の意識は許容する。

<備考> これまでと同様に、英語を母語とする翻訳家と日本語を母語とする翻訳家の共同作業に委ねる。

## 3 収集結果

## 3.1 ナ型形容詞の整理

日本語語彙大系には、ナ型形容詞として 907 語(1,341 パターン)を収録した[池原 1997]。これに対し、例えば、学研国語大辞典には約 4,300 語ほど形容動詞と表示されている語が記載されている[白井 1999]。これらは、古語や俗語、タルト型の語なども含むが、日本語語彙大系の収録では大幅に不足していることが予想される。

そこで、まず、現代国語例解辞典[林 1997]に収録されているナ型形容詞の抽出を行なった。この際、辞書の見出し語と品詞表示に拘泥されないように努め、派生語や「～な」の用法がありそうな語を広く抽出した。また、例文作成者が自然な例文を作成できるようにするため、辞書収録語のうち現代語として不適当と思われる語を除外した。

抽出された語は 2,813 語で、319 派生語(あいしげ、やさしげ(優しげ)、等)や、名詞表示であるが「～な」が可能な 13 語(コンパクト、よき(陽気)、等)を含んでいる。このうち、「しつとう(失当)」「ゆうあく(優渥)」など 107 語は知らない語であるため、「そぞろ(漫ろ)」「とくだん(特段)」など 220 語は「～な」の可否に疑問があるため、今回の例文作成の対象外とした。「いかさま」「はばかりさま」の 2 語は形容動詞表示であるが、分類に疑問があるので除外した。俗語等と表示された 6 語(いけず、がらっぱち、ぎょうさん、なまじつか、ほんま、わや)も除外した。

なお、今回の作業の対象外であるが、タルト型形容動詞の表示がある語は 281 語であった。

### Compiling Japanese and English Corpus for Japanese Nominal Adjectives

Satoshi SHIRAI<sup>\*2\*1</sup>, Hiromi NAKAIWA<sup>\*1</sup>, Yoshifumi OYOYAMA<sup>\*1</sup>, Hiroshi AIZAWA<sup>\*3</sup>, Shinobu TAKECHI<sup>\*3</sup>, and Keiko WAKEBE<sup>\*3</sup>

<sup>\*1</sup>NTT Communication Science Laboratories, <sup>\*2</sup>ATR Interpreting Telecommunications Research Laboratories, and <sup>\*3</sup>Mu Inc.

### 3.2 日本語例文の作成

前節の作業で抽出された語を対象に例文作成を行った。その際、例文作成が困難なものは日常使われる語として適当ではないと判断して、対象から除外した。最終的には、1,318 語に対して 2,356 例文が作成された。

表 1 に、ナ型形容詞 1 語ごとに例文が何文作成されたかの分布を示す。例文が多いのは「きれい」22 文、「ゆたか」9 文、「かんたん」「じょうぶ」「だめ」各 8 文である。

表 1 例文数の分布(全体)

例文数	22	9	8	7	6	5	4	3	2	1	計
語数	1	1	3	3	10	17	36	146	452	649	1,318

表 2 に、ナ型形容詞 1 語ごとに「～だ」の例文が何文作成されたかの分布を示す。例文が多かったのは「きれい」8 文、「だいすき」「だめ」「ゆたか」各 6 文である。ここで、0 文というのは「～だ」以外の用法、即ち、「～な」「～に」の用法の例文のみ作成されたことを示す。因みに、「～に」のみ 12 語に対し、「～な」のみ 566 語、双方の例文があるもの 40 語で、限定用法のみが圧倒的であった。

表 2 例文数の分布(「～だ」の用法)

例文数	8	6	5	4	3	2	1	0	計
語数	1	3	2	5	32	115	542	618	1,318

表 3 に、ナ型形容詞 1 語ごとに「～な」の例文が何文作成されたかの分布を示す。例文が多かったのは「きれい」「こまか」各 5 文である。ここで、0 文というのは「～な」以外の用法、即ち、「～だ」「～に」の用法の例文のみ作成されたことを示す。

表 3 例文数の分布(「～な」の用法)

例文数	5	4	3	2	1	0	計
語数	2	8	40	214	687	367	1,318

表 4 に、ナ型形容詞 1 語ごとに「～に」の例文が何文作成されたかの分布を示す。例文が多かったのは「きれい」9 文、「おおげさ」「ばかしょうじき」各 2 文である。ここで、0 文というのは「～に」以外の用法、即ち、「～だ」「～な」の用法の例文のみ作成されたことを示す。

表 4 例文数の分布(「～に」の用法)

例文数	9	4	3	2	1	0	計
語数	1	2	1	12	92	1210	1,318

### 3.3 対訳用例集の作成

前節の作業で作成された日本語例文を、なるべく日本語原文に忠実に、かつ、英語として十分通用するように、翻訳家に英訳してもらった。翻訳の際、日本語例文の解釈が問題になることがあったため、必要に応じて日本語例文の改良や例文の使用場面の注記などを行なった。現在、日英例文の意味的な対応について確認を進めている段

階で最終的な集計ではないが、日本語例文(1,318 語、2,356 文)の規模は約 35,000 字、英語訳文の規模は約 16,000 語となっている。表 5 に、これまでに収集した例文との諸元を比較する。

日本語表現は 2 文字程度長い、英語表現はイ型形容詞の場合と大差はなさそうである。ナ型形容詞の平均単語長は 3.7 であり、イ型形容詞に 3.1 より若干長めであるが、文全体の長さを説明し切れない。何が要因であるかは今後検討することとしたい。

表 5 用例文集の諸元の比較

	語数	例文数	和文平均字数	英文平均語数
ナ型形容詞	1,318	2,356	約 14.8	約 6.7
イ型形容詞	136	1,909	12.78	6.66
和語動詞	861	10,497	12.66	8.46
複合和語動詞	2,105	3,717	16.60	8.16

## 4 おわりに

機械翻訳用の結合価パターン対辞書の作成に必要な対訳例文収集の一環として行なったナ型形容詞の例文収集の概要を報告した。具体的には、現代国語例解辞典を基準としてナ型形容詞を抽出し、内省により日本語例文を作成し、忠実な英訳を付与することにより、1,318 語 2,356 文の対訳例文集を作成した。

人間が和英辞書等を使用するとき、単語自体の収録がない場合、単語はあるが適当な用法の記載がない場合が問題になる。人間は言い換えによりこの問題に対処する。しかし、計算機での使用を考えると、両者とも致命的であり、何らかの手段で収録しておくことが要請される。本稿の作業は、和英辞書からパターン対を収集した後、特に洩れている語を中心とした拡充作業の一つとして行なったもので、主に前者の問題に対応していると考えられる。

今後は、収集した対訳用例の見直しを行なった後、これらに基づいてパターン対の収集を行なう予定である。また、3.2 節で簡単に示したような用法の違いと英訳語との関係についても分析を進めたいと考えている。

### 参考文献

- [林 1997] 林巨樹(編): 現代国語例解辞典(第二版), 小学館。
- [池原 1997] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林: 日本語語彙大系, 岩波書店。
- [IPA1987] 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。
- [IPA1990] 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。
- [石綿 1983] 石綿敏雄, 荻野孝野: 結合から見た日本語文法と日本語用言の結合価, 文法と意味 I(水谷, 石綿, 荻野, 賀来, 草薙(編)), 朝倉書店。
- [白井 1995a] 白井, 池原, 横尾, 井上: 日英機械翻訳に必要な結合価パターン対の数とその収集方法, 情処研報 95-NL-110, pp.43-50。
- [白井 1995b] 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。
- [白井 1997] 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, 情処研報 97-NL-122, pp.1-6。
- [白井 1999] 白井論: 結合価パターン対の網羅的収集に向けて, 一日英機械翻訳の観点から一: 「言語資源の共有と再利用」シンポジウム, URL=http://www.etl.go.jp/etl/nl/sympo99/shirai.html。