

換言事例の収集

一日英基本構文を対象として一

白井 諭 山本 和英

ATR 音声言語通信研究所

E-mail: {shirai,yamamoto}@slt.atr.co.jp

1 はじめに

日本語語彙大系の構文体系 [池原 97] により, 14,000 件あまりの日英基本構文を対応付けることが可能となっている。しかし, 用言の種類および用言ごとの用法の網羅性, 同じ意味を表すことが可能な表現の種類の多様性, 構文における制約条件の記述等に問題が残されていることが報告されている [白井 98]。

網羅性の問題は, 人間用辞書と機械用辞書の性格の違いに起因する。中辞典規模の和英辞書の編集方針として, 比較的使用頻度の低い語彙や用法は収録されていないと考えられる。人間が使用するとき, 目的とする語彙または用法が掲載されていないければ, 類義表現に言い換えて引き直すという形の試行錯誤を行なうことにより, 目的に準じる語彙または用法を得る。これと同様の機構を計算機処理で実現するのは極めて困難であるため, 機械用辞書では語彙や用法を網羅的に収録していることが要求される。用例の収集とその抽象化により基本構文の網羅性を高めるための検討が継続されている [白井 99]。

多様性の問題は, 機械翻訳の利点の1つである訳語の統一性に起因する。また, 初期の構築作業を急いだため, 日英基本構文の対応付けが1対1に限定された点も一因となっている。訳語の統一性は, 場合によっては訳文の単調性という欠点を招く。機械翻訳をツールとして利用する場合, 後編集の1つとして, 類語集 (シソーラス) を利用して表現を多様化することが行なわれている。名詞等を置換した場合の影響は前後の数単語程度であるのに対し, 動詞を置換した場合の影響は広範囲にわたることが多い。単語の類語集があるように, 構文に関する類義集があれば極めて有用であると考えられる。

条件記述の問題は, もともと構文体系は解析用に設計されたことと, 人手により条件記述が行なわれていることを主因とする。前者は, 例えば「犬が子犬を生む」「猫が子猫を生む」を, 意味体系に基いて「<動物>が<動物>を生む」のように名詞を抽象化して記述するため, ガ格の名詞とヲ格の名詞の相互の関連性が失われ

る。通常の解析処理における言語表現の受容には支障はないが, 誤りの検出に利用しようとする検出洩れを生じるほか, 言語生成において不適当な組合せが生成される。また, 後者は抽象化にバラつきが生じることがある。後者については支援処理により均質化を試みている [秋葉 00]。

構文体系を日本語解析や日英翻訳の基本辞書として整備していく上で, これまでは網羅性に重点を置いてきた。現状では, 低頻度の語彙や用法に関する情報を得るのに十分な量のコーパスを整備するのは容易ではないため, 人間の内省を活用するのが効果的であると考えられる。一方, 表現の多様性に関する情報は断片的なものしか存在せず, 当面は言い換えながらデータを蓄積する段階であると思われる。そこで, 日英基本構文の網羅性向上の一環として収集された日英対訳用例文 [白井 99] を題材に, 日本文と英文の意味的対応を制約として, 日本文に対する別の言い方の収集を試みたところ, 十分実行できることが分かった。また, 英文に対しても同様の結果が得られた。本稿では, 検討の概要と言い換え文の収集状況を報告する。

2 収集条件

従来は構文体系の網羅性向上を主な狙いとして内省により例文作成を進めてきた。内省による場合, 作成された例文の恣意性が問題となる場合が少なからず考えられる。しかし, 特定の場面に適合する少数の例文だけを作成するのではなく, あらゆる用法を列挙するという課題設定であるので, 例文の恣意性の問題は起こりにくいと考える。ただし, 生成された例文が自然な表現と言えるかどうかは問題となる場合はあった。これについては, 同じ作業がある程度時間が経ってから見直すか, 別の作業者と相互検査することにより排除に務めた。

以下では, まず網羅性向上を目的として実施した際の設定条件を再掲し, 次に多様性向上を狙いとして今回設定した条件を示す。

2.1 網羅的な用例の収集

最終的な目的は構文体系の網羅性の向上であるが、抽象化した状態の文型を収集するのは容易ではないため、まず様々な用法を例文という形で内省により網羅し、例文に基いて抽象化する2段階で考えることにした。例文作成の対象とする用言の選択として、現代語の用言として相応しいかどうかは個別に判定するが、1つの辞書を選択して大まかな基準として利用する方針とした。また、生成された例文が自然な表現かどうかが問題となる場合はあった。これについては、同じ作業者がいる程度時間が経ってから見直すか、別の作業者と相互検査することにより排除に務めた。一連の作業経験を踏まえ、現在は次の条件を設定している。

- (1) 現代国語例解辞典 [林 85, 林 97] 所収の用言性の語を対象とし、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した。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。例文の作成を進めながら、例文作成が困難なものを、例文作成者同士の協議により除外した。

- (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広く例文化し、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る。すなわち、対応する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許容する。

- (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のニュアンスの違いにも留意して例文を作成する。

＜備考＞連用形の副詞用法や連体形の限定用法には慣用的なものがあり、それらの収集も対象とした。

- (4) 用言1語当たりの2例文を最低目標とする。ただし、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 n^2 に比例する。10～15分考えても新たな用法が思いつかなければその用言に対する作業を打ち切ることとした。

- (5) 収集された例文に対して、可能な限り原文に忠実に、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う。(最低限度の意訳は許容)

＜備考＞経験を踏まえ、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た。

2.2 多様な用例の収集

最も直接的な動機は、1つの日本語表現に対する複数の英訳を得ることである。これは見方を変えれば英語表現を換言しているとも云える。一方、ある英語表現が必ず特定の日本語表現から生成されなければならないわけでもない。そこで、日本語の換言と英語の換言を並行して実施することにした。

また、換言事例の収集という課題は、元来は同一言語内で同義の表現を収集するべきかもしれない。しかし、例文が提示されるとそれに眩惑されて他の表現がなかなか思いつかない場合も少なくない。そこでここでは日英の対訳対の存在を前提として、その日英の文対を制約として利用しながら、換言事例を収集することとした。

ここでいう換言は、例えば英作文の際、和英辞書に載っていない語や表現に出会ったとき、日本語の別の類義表現を生成し、和英辞書を引き直すことを模したものである。従って、翻訳対象言語に精通していない単言語話者にも作業可能であると考えられる。しかし実際問題として、考えついた別の表現が和英辞書に未集録であるという状態が連続して発生すると、同義性の制約が徐々に甘く、すなわち意味のずれが拡大していく恐れがある。そこで、今回は試行ということもあり、網羅的な用例収集の作業担当者とその例文の翻訳担当者に依頼した。これは、網羅性の確保の際に微妙な日英対応の判定が容易でなかった経験に基づく。

- (1) 前節で述べた日本語の用言に対する日英の対訳例文対を対象とする。
- (2) 日本語の換言は、英文に多様な和訳をつけるつもりで行なう。逆に、英語の換言は、日本語例文に多様な英訳をつけるつもりで行なう。
- (3) 原則として、特殊な場面設定を必要としない中立的な表現を作成する。

3 収集状況と考察

構文体系の元となる構文意味辞書は、石綿らの結合価の考え方 [石綿 83] に基づき、中辞典級の和英辞書の例文を抽象化することにより構築を開始した。初期の版としては、一般文型 10,000 件と慣用文型 5,000 件が収集された。しかし、実験的評価により文型の不足が頻出することがわかり、文型を工学的に網羅する方法を検討した。現実的には、低頻度の用法を収集するのに十分な量のコーパスを獲得するのは容易ではない。そこで「内省」により様々な用法を例文として収集することとした。

表 1: 用言の種類と例文数

	該当 用言数	作成 例文数	(換言文作成なし)		換言例文数		備考
			日本語	英語	日本文	英文	
和語動詞 / IPAL	849	16,713	12,020	13,748	7,043	4,096	0(IPAL), 1(追加), 3(修正), 8(換言)
和語動詞 / 他	936	1,883			0	0	7(収集)
複合和語動詞	2,101	3,701	2,487	3,220	1,212	480	4(収集), 9(換言)
イ型形容詞 / IPAL	136	2,156	1,626	1,937	530	219	0(IPAL), 2(追加), 6(修正), 11(換言)
イ型形容詞 / 他	522	830	1	0	1,561	1,584	12(収集 & 換言)
ナ型形容詞	1,296	2,356	1,735	1,915	621	440	5(収集), 10(換言)
(サ変動詞 = 途中)	(131)	(197)	1	0	(338)	(362)	13(収集 & 換言)

3.1 用言の種類と例文の収集情況

用言ごとの様々な用法が例文として収録されている IPAL 辞書 [IPA87, IPA90] に着目し、ニュアンスの異なる用言の用法を例文として追加した。次いで、国語辞書に基づいて用言の網羅性を高めることとし、用言選択の基準を現代国語例解辞典 [林 85] に求めた。IPAL 辞書に収録されていない用言を対象にして例文作成を継続中で、現在サ変動詞を作業中である。また、途中から換言作業も並行して進めている。

表 1 に 2000 年 12 月末現在の収集情況を示す。なお、備考に作業順と作業内容を示す。各項目は 1～3 人年の作業量である。付録 A～C に例文を示す。

3.2 考察と今後の課題

例文作成の方針は 2 節で述べた通りであるが、これは試行錯誤を踏まえながら設定されたものである。次のような問題が複雑に入り組んでいると思われる。

(1) 作業の習熟度

初期の作業には作成量や多様性の点で不満が残る（作業者の方が不満の程度が大きい）。IPAL 所収の動詞と形容詞を対象とした作業では見直しを行なったが、他の例文についても見直すべきかも知れない。換言についてはその必要性が高いかも知れない。また、時間において見直すことにより例文の品質は高められそうである。

(2) 動詞か形容詞か

動詞には連用形の副詞的用法（例：次いで）はあまりなく、連体形の用法と終止形の用法も意味的に差がないことが多い。これに対し、形容詞には連用形の副詞的用法だけでなく連体形の限定的用法も多様で慣用的な結び付きも少なくない。また、終止形の用法がないと思われ

ることもあり、例文作成の打ち切り判断に迷う場合が頻発した。

(3) 一般的か慣用的か

当初は一般文型の収集に重点を置き、結果的に慣用文型とすべきものを許容する形態を取った。慣用的なものを網羅的に収集するのは容易でないと考えたからである。しかし、慣用文型の中にも文字通りの解釈が可能なものも少なくなく、その逆の場合もある。換言という観点ではむしろ慣用的なものも網羅しておく必要がある。

(4) 多義的か個別的か

多義性の大きい用言では例文作成数が多くなり、例文全体を見渡ししながら用法の網羅性を検討するのは容易でない。一方、個別的な用言は自然な例文を作成すること自体が容易でないばかりか、場合によっては背景説明を必要とするなど作業効率が悪い。このような両極端の場合には何らかの作業支援を行なうことが必要である¹。

(5) 換言の程度

当初は構文体系を拡張する観点から述語部分の換言を重点的に進めてきた。しかし、格要素との組で換言すると多様な表現が成立する場合があり、徐々に条件を緩めつつある。大前提である訳文の対応の保証のみの制約でも構わないかも知れない²。あるいは、換言の結果を日英相互に交換して再検討するなど有効かも知れない。

¹ 文献 [藤田 00] では名詞の言い換えを対象とする支援環境が提案されており、今後参考にしたい。

² 筆者らは厳密な意味で同義となる換言は存在しないと考える。

4 おわりに

日本語用言の多様な用法とその換言事例作成の現状を紹介した。経験を積み重ねながら作業形態を進化させているため、初期に作成した例文には見直すべき問題も少からず残されている。また、名詞が述語となる場合はほとんど収集していないが、属性的に働く場合や、話し言葉における多用 [竹沢 01] への対応も検討していきたい。

謝辞 (株)CANNACの鳴海武史氏、武智しのぶ氏、相澤弘氏らには例文収集の実施に関して多大なご協力を賜わった。NTTコミュニケーション科学基礎研究所のFrancis Bond氏らには換言事例の収集に関してご討論頂いた。ATR音声言語通信研究所の森田千秋氏には例文集を整備して頂いている。これらの方々を始めとする関係各位に深謝する。

参考文献

- [秋葉 00] Akiba, Y., H. Nakaiwa, S. Shirai & Y. Ooyama(2000). "Interactive generation of a translation example using queries based on a semantic hierarchy". *ICTA100 (The 12th International Conference on Tools with Artificial Intelligence)*, pp. 326-332.
- [藤田 00] 藤田, 乾, 乾 (2000). "名詞言い換えコーパスの作成環境". 電子情報通信学会 技術研究報告, TL2000-32, pp. 53-60.
- [林 85] 林巨樹 (編)(1985). "現代国語例解辞典 (第一版)". 小学館.
- [林 97] 林巨樹 (編)(1997). "現代国語例解辞典 (第二版)". 小学館.
- [池原 97] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林 (編) (1997). "日本語語彙大系". 岩波書店.
- [IPA87] 情報処理振興事業協会 技術センター (編) (1987). "計算機用日本語基本動詞辞書 IPAL(Basic Verbs)", 解説編 & 辞書編.
- [IPA90] 情報処理振興事業協会 技術センター (編)(1990). "計算機用日本語基本形容詞辞書 IPAL(Basic Adjectives)", 解説編 & 辞書編.
- [石綿 83] 石綿, 荻野 (1983). "結合価から見た日本語文法" & "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 95] 白井, 池原, 横尾, 井上 (1995). "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". 情報処理学会研究報告 95-NL-110, pp. 43-50.
- [白井 98] 白井, 横尾, 中岩, 渡邊, 高橋, 関, 池原, 宮崎 (1998). "構文意味辞書と構文体系". 言語処理学会第4回年次大会, B2-2, pp. 194-197.
- [白井 99] 白井諭 (1999). "結合価パターン対の網羅的収集に向けて - 日英機械翻訳の観点から -". 「言語資源の共有と再利用」シンポジウム.
<http://www.etl.go.jp/etl/nl/sympo99/programme.html>
- [竹沢 01] 竹沢, 白井, 大山 (2001). "バイリンガル旅行会話に見られる話し言葉の特徴分析". 情報処理学会研究報告, 01-NL-141-22, pp. 137-144.

付録 A: 和語動詞の対訳換言例文 (一部)

- J0 条件が揃った。
J1 条件が満たされた。
E0 The conditions were satisfied.
E1 The conditions were met.
- J0 この真珠は大きさが揃っている。
J1 この真珠はみんな同じ大きさだ。
E0 These pearls are all of the same size.
- J0 二人は足並みが揃っている。
E0 The two are keeping step with each other.
E1 The two are keeping pace with each other.
- J0 クラブはドライバーからパターまで揃っている。
J1 クラブはドライバーからパターまで全部ある。
E0 I have all the clubs, from the driver to the putter.
- J0 全校生徒が校庭に揃った。
J1 全校生徒が校庭に集合した。
E0 All students of the school gathered on the school grounds.
E1 All students of the school assembled on the school grounds.

付録 B: イ型形容詞の対訳換言例文 (一部)

- J0 その2つの三角形の面積は等しい。
J1 その2つの三角形の面積は同じだ。
E0 The areas of those two triangles are the same.
- J0 その2つの三角形は面積が等しい。
J1 その2つの三角形は面積が同じだ。
E0 Those two triangles have the same areas.
- J0 Aの値はBの値に等しい。
E0 The value of A is equal to that of B.
- J0 今年のベースアップは無いに等しい。
J1 今年のベースアップはほとんど無い。
E0 This year's basic wage increase is almost zero.
- J0 彼の行為は詐欺に等しい。
J1 彼の行為はほとんど詐欺だ。
E0 His actions almost amount to fraud.

付録 C: ナ型形容詞の対訳換言例文 (一部)

- J0 私は彼の考えと同じだ。
E0 I have the same idea as he does.
- J0 彼女は私と同じクラスだ。
J1 彼女は私と一緒にいるクラスだ。
E0 She is in the same class as I am.
- J0 リンゴと桜は同じバラ科だ。
J1 リンゴと桜はどちらもバラ科だ。
E0 Apples and cherries both belong to the rose family.
E1 Apples and cherries are both in the rose family.
- J0 黙って借りてきたというのは盗んだのと同じだ。
J1 黙って借りてきたというのは盗んだに等しい。
E0 Borrowing without asking is tantamount to stealing.
- J0 私たちは同じ釜の飯を食った仲だ。
E0 We lived under the same roof.