

日英機械翻訳のための構文辞書

白井 諭*1 横尾 昭男*1 中岩 浩巳*1 池原 悟*2 宮崎 正弘*3

*1NTTコミュニケーション科学研究所 *2鳥取大学 工学部 *3新潟大学 工学部

あらし 解析ベースの機械翻訳における意味解析の精度向上には、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。筆者らは、日英翻訳での使用を目的として、和英辞書や対訳用例文からパターン対の収集を進めてきた。これまでに一般表現パターン1万3千対(6千用言)と慣用表現パターン3千対を収集しているが、全体規模としては一般表現パターン2万対(1万用言)、慣用表現パターン対は5千対が必要になると考えられる。本稿では収集の経過と現在の到達点を示すとともに、今後の課題について述べる。

キーワード 機械翻訳, 日英翻訳, 構文辞書, 結合価パターン対, 意味解析

A Semantic Valency Dictionary for Japanese-to-English Machine Translation

Satoshi SHIRAI*1, Akio YOKOO*1, Hiromi NAKAIWA*1,
[shirai, ayokoo, nakaiwa]@cslab.kecl.ntt.co.jp

Satoru IKEHARA*2, and Masahiro MIYAZAKI*3
ikehara@ike.tottori-u.ac.jp miyazaki@info.eng.niigata-u.ac.jp

*1NTT Communication Science Laboratories, *2Faculty of Engineering, Tottori University,
and *3Faculty of Engineering, Niigata University

Abstract In order to improve the quality of machine translation, the valency pattern method of analyzing the semantic relationship between verbs and nouns is known to be effective. We compiled about 12,000 general patterns (6,000 verbs) and 3,000 idiomatic patterns for a Japanese-to-English machine translation system based on Japanese-to-English dictionaries for human use and Japanese and English example sentences. We evaluated the patterns and found that at least 20,000 general patterns (10,000 verbs) and 5,000 idiomatic patterns were needed. In this paper, we introduce the semantic valency dictionary and discuss how to complete it.

Key words machine translation, Japanese-to-English translation, semantic valency dictionary, valency pattern pair, semantic analysis

1 はじめに

構文辞書は、日本語の述語が中心となって結びつく範囲をパターンとして記述し、それに対応する英語の表現を対にして持たせることにより、日英翻訳におけるいわば構文の語義を規定するために導入された。構文の辞書の作成には多くの試みがあるが、収録語数や分解能の点で実用規模に達したものはない。これに対し、約2,800の分解能を持つ一般名詞の意味属性[宮崎97]に基づいて、約40万語の大規模な単語辞書[横尾97]が構築されたことにより、構文辞書を構築することが可能になった。本稿では、構文辞書をどのような方針で作成したか、どのような文型を収録したかについて概要を報告する。

2 文型の収集の背景

言語表現の意味を担う最小単位は単語である。一般に、単語にはいくつかの語義があり、組み合わせて使用することにより相互の制約が生じている。逆に言えば、この制約のために生じる単語の共起関係を解析することにより、言語処理では単語がどの語義で使われたかを推測する。単語の語数としては、単語辞書としてまとめた約40万語がおおよそ目安として考えられる。単語辞書では、意味属性体系で定義した意味属性に基づいて、名詞に複数の意味的用法を付与することにより語義を分類した。次に名詞以外の単語の語義を分類する必要がある。

名詞以外の単語のうち、特に動詞、形容詞といった用言では格要素と述語の組み合わせによる表現が用いられる。具体的には、動詞では「何が何をどうする」「何がどうなる」などの文型、形容詞では「何がどんなだ」「何が何にどんなだ」などの文型が一般的である。この「何」のところはどういう名詞が入るかによって動詞や形容詞の意味合いが変わることがある。従って、動詞や形容詞といった単語に対して単独に語義の区分をするのではなく、名詞との組み合わせを文型としてとらえ、文型に対して語義を分類するのが現実的であると考えられる。

計算法処理向きの文型分類の先駆的なものとしては石綿と荻野による「日本語用言の結合価」(文献[水谷83]の附録2)がある。これは“用言を「体言+格助詞」との結合関係でとらえ、各々の型を体言の意味特徴と格助詞の種類によって記述した”もので、体言(名詞)の意味特徴として11分類を与え、1,154用言に対する1,775文型を示した。しかし、その序

文にあるように汎用性と規模の点で、また名詞の分類が粗すぎる点で問題がある。このような分類の趣旨を尊重しながら、単語辞書に収録されている用言性の単語1万数千語の規模で文型を体系化する必要がある。

また、日英翻訳での利用を念頭に置けば、日本語の文型が決定された段階で英語の基本構造も決定されるのが望ましい。これは日英翻訳におけるいわば構文の語義を決定しているといえる。しかし、一般名詞の意味属性だけで2,800分類されることを考えれば、これらと用言が組み合わさった文型に対する語義の記述を行なうのは極めて困難であることが予想される。そこで構文辞書としては、用言を「体言+格助詞」との結合関係でとらえ、各々の型を体言の意味特徴と格助詞の種類によって記述するとともに、体言に一般名詞の意味属性2,800分類を適用した日本語文型とそれに対応する英語文型を対にして持つことにより語義記述に代えることにした。

構文辞書では、このような文型パターン対(結合価パターン対)を網羅的に収集することを目的として構築を開始した。文型パターンの表す内容は1つの事象であると考えられる。事象には、現実世界に存在することのほか、空想世界に存在することが考えられる。空想世界の事象は、多数の人に共有されていることや特定の人は知らないこと、事実には照らして正しくないことなど種々雑多なものが考えられる。

そこで構文辞書としては、言語に関する一般的な知識の範囲であると考えられる部分をファクトベースで記述することにした。具体的には、原則として単文の肯定形を記述することとし、その際、名詞は一般名詞の意味属性を用いて英語文型の選択が可能な段階までの抽象化した。しかし、「油を売る」が〈サボる〉の意味で使われる場合のように、特定の名詞と特定の用言が結びついて個々の単語から全体の意味が導き出せないものは、名詞を個別に指定する必要がある。このように、名詞を個別指定する必要があるものを「慣用表現文型」と呼び、それに該当しないものを「一般表現文型」と呼ぶ。また、それぞれのパターンには、用言意味属性[中岩94,中岩97]を付与した。

3 パターン対の収集方法

はじめに、パターン対の収集経過の概要を述べる。

3.1 和英辞書からの収集

日英対訳の人用の辞書としては和英辞書と英和辞書がある。いずれも単語の語釈が中心的に記述され、必要に応じて用例文のほか用法や文型に関する事項が補足されている。日英翻訳の観点でこれらの辞書を比べると次のような特徴が見られる。英和辞書では英語の表現に対する日本語の表現が記述されているため、その記述には説明的なものや現実の日本語の文書では使用されない表現がしばしば使用されている。これに対して和英辞書では日本語の表現に対する英語の表現が記述されているため英和辞書のような問題はない。しかし、日本語の単語に対する英語表現しか示されていない場合が少なくなく、熟語的な表現が得られないという問題が考えられる。

以上の点を考慮した結果、第1ステップとしては和英辞書の見出しを基準にして構文辞書への収録語を選定し、その項目に記されている対訳例文から日英それぞれの基本構造を抽出することとした。例えば、「ライトハウス和英辞典(第1版, 研究社, 1984年)」には動詞「上がる」に対して5つの語義が示され、第2の語義の例文として次の対訳文が掲載されている。

彼の学校の成績が上がった。

His school record *has improved*.

この対訳文を分析し若干の情報追加を行なえば次のような結合価パターン対が得られる。

(日本語パターン)	(英語パターン)
X [成績, 能力] が	X
Y [数量] から	improve
Z [数量] まで	from Y
上がる	to Z

このようにして、中辞典クラスの和英辞書数冊から結合価パターン対を収集した。また、慣用表現を充実させるため、必要に応じて慣用表現辞書(例えば、文献[宮地 82])も利用した。この結果、5,600 用言に対して一般表現パターン対 10,000 件、慣用表現パターン対 5,000 件が収集された。その後、一般表現パターン対の中に統合できるものが含まれていたこと、また慣用表現パターン対の中に汎用化できるものが含まれていることがわかり、これらを見直した結果、一般表現パターン対 10,000 件、慣用表現パターン対 3,000 件となった。

過不足性の検証のため、これらを用いて初見の技

術文献(1,361 文)を対象に翻訳実験を行なった。

その結果、その文献に出現したパターン対は異なり 201 件で、23%にあたる 47 件が不足していることがわかった[白井 95a]。不足しているパターン対としては単語あたりの語義が多い和語動詞に対する一般表現パターン対が特に目立った。また、用言は異なり 120 語で、15%にあたる 22 語はパターン対が 1 件も収録されていないこともわかった。目的とする用言や構文が辞書に掲載されていない場合、人間であれば類似の表現に言い換えて辞書を引くことができるが、コンピュータはそれができないため、あらかじめ辞書のカバレッジを拡大しておく必要がある。そのためには、和語動詞のパターン対の充実と、未収録の用言に対するパターン対の収集が必要である。

3.2 日本語辞書の用例とその英訳からの収集

和語動詞のパターン対を充実させるには、使用例のバリエーションを数多く集めることが必要である。和語動詞に関しては、20 名あまりの日本語の言語学者が中心となって語義の分類と対応する用例文を収集分析する研究が進められ、その成果は辞書として公開されている。こうしてまとめられた IPAL 動詞辞書[IPA 87]の用例文に基づいてパターン対の収集を行なうことにした。具体的には、IPAL 動詞辞書の用例文に対し、翻訳家に日本語原文に忠実で十分通用する英訳文を作成してもらい、その対訳データからパターン対を収集することにした。この方法では、861 語の和語動詞(ただし、ひらがな表記の異なりで、漢字表記の異なりでは約 1,200 語)に対して、5,243 文(日本語 7.5 万字, 英語 4 万語)の対訳用例文が得られた。そして、パターン対収集作業の結果、新たに 1,532 パターン対が収集され、500 パターン対に対する修正情報が得られた[白井 96]。

この IPAL 動詞辞書は元来日本語処理のための語義分類であり、日英翻訳の観点から十分な語義分類となっているかどうかが問題となる。IPAL 動詞辞書の 1 つの語義に複数の例文が示され、それらが 1 つのパターン対でカバーされたケースは少なかったことから、語義あたりの用例文数を充実させるのが有効ではないかと予想される。しかし、語義に当てはまるような用例文を大量に収集するのは容易ではないという問題がある。

3.3 内省で作成した用例とその英訳からの収集

日本語用例文とパターン対の関係を観察すると、

同じ動詞を使用しているも、動詞の使われ方のニュアンスが異なる場合に別の英語表現が必要となる場合が多いことに気がつく。そこで、英語の理解できる日本人が辞書等を参考にしながら自分の知識を引き出し、日本語としてニュアンスの異なる用例文を可能な限り列挙するという方法で用例文の収集を試みた。列挙する用例文は、作業にける時間にもよるが、ある程度の時間以上考えても用例文が思い浮かばなくなるまで抽出することにした。抽出された用例文に対して、翻訳家に日本語原文に忠実で十分通用する英訳文を作成してもらい、その対訳データからパターン対を収集した。この対訳用例文の作成作業では、約 1.5 人年の作業により、IPAL 動詞辞書と同じ 861 動詞に対し、10,500 文（日本語 13 万字、英語 6.8 万語）が収集された[池原 96]。いくつかの動詞を無作為に選択してパターン対の収集を試行したところ、和英辞書や IPAL 動詞辞書から収集済みのパターン対と同程度の数のパターン対が新たに収集されることがわかった。これは、対訳例文からパターン対を収集する方法の有効性を示すとともに、人用の辞書から日英機械翻訳での使用に耐える辞書を作成するのは困難であることを示している。

なお、この作業は現在継続中であり、今後、形容詞や複合動詞などに対しても同様の方法を適用したいと考えている。

4 文型パターン の記述内容

作成を開始した時点では構文辞書の具体的な完成イメージを持っていたわけではない。まず人用に作成された和英辞書の例文を抽象化することにより機械用の構文辞書の構築を開始した。人間であれば常識的に類推したり判断したりできるような事柄は辞書には記載されていないことが多い。しかし、コンピュータにはそのような常識が期待できないため、機械用の辞書では常識的な事柄であっても一定の基準や形式に従って丹念に記述しておく必要がある。このため日本語および英語の文型パターンをどう記述するかは試行錯誤の連続となった。その試行錯誤を通して現在の形式に到達した。その過程をいくつかの観点から整理して振り返る[林 87, 奥 87]。

4.1 一般表現文型

述語を中心とする表現には、動作性の表現（何が何をどうする、何がどうなる）と状態性の表現（何がどんなだ）のほか、断定文（何が何だ）がある。

日本語では一般に、動作性の表現には動詞構文が使用され、状態性の表現には形容詞（イ型とナ型の総称）構文が使用され、断定文には名詞構文が使用される。

動詞構文や形容詞構文では、格要素と述語の結びつきが比較的強く、それぞれを単純に英訳して組み合わせても英語として通用しないことが多い。従って、一般表現文型としては、まず動詞構文や形容詞構文のように訳し分けの必要性が高いものを重点的に収集した。

名詞構文「XはYだ」は、英語でも「X be Y」のようにそれぞれを名詞として訳せばよい場合が多く、格要素と述語の結びつき比較的弱いといえる。このため、初期は文型パターン対の収集対象からは除外していた。しかし、「今日は天気だ」→「It is fine today.」のように英語が形容詞になるもの、「あなたに質問です」→「I ask you a question.」のように英語が動詞になるもの、「彼の成功は努力次第だ」→「His success depends on his effort.」のように日本語の名詞述語の複合名詞が分割されて訳されるものなども散見される。そこで、このように名詞述語の英訳が名詞とはならないものは収集対象に加えることにした[白井 94]。

(1) 動詞構文に関する補足事項

動詞構文には、「述べる」や「命じる」のように、「～と」「～よう」などで示される文相当の内容を格要素として必要とするものがある。これらに対応する英語動詞「say」や「order」も that 節や to 不定詞句を必要とする。収集対象とするのは単文であるが、これらの動詞はその性質上、複合文になるのはやむを得ないと考え、これらも一般表現文型としての収集対象とした。

動詞の中には不変化部分が共通で活用の仕方が異なるものがある。例えば、次のようなものである。

- (a) 五段活用とその転生名詞のサ行変格活用（手渡す、手渡しする）
- (b) 一段活用とその転生名詞のサ行変格活用（真似る、真似する）
- (c) サ行変格活用とその五段活用化（課する、課す）

この程度の変化形は形態素解析処理で十分対応可能であると考え、原則としてそれぞれ元の形と考えられる前者のみを登録した。ただし、「受け付ける (accept)」と「受付する (register)」では、上記の (a) に該当するが、英訳が異なるため、それぞれに対し

て別の一般表現文型を作成した。

(2) 形容詞構文に関する補足事項

形容詞構文には、「XはYがZだ」の表現形式となる二重主格文がある。これらは次の方針で「一般表現文型」を作成した。

- (a) 「XがYをZだ」と言い換えた上で文型を記述する(彼は彼女が好きだ)
- (b) 「XにYがZだ」と言い換えた上で文型を記述する(日本は雨が多い)
- (c) 「XのYがZだ」と言い換えられるなら「YがZだ」の部分を変型として記述する(象は鼻が長い)
- (d) 言い換えができない場合は「XはYがZだ」の文型を記述する(彼は数学が得意だ)

なお、(c)に対する英語表現は“X's Y is Z”または“Y of X is Z”が一般的であるが、例えば「象は鼻が長い」では、単語辞書の一般名詞意味属性により、象の意味属性『動物』から始めて『動物』→『身体』→『頭』→『顔』→『鼻』という一連の has-a の関係がたどれるので、“Elephants have long noses(または trunks).” のように英訳することもできる。

また、形容詞構文には、「AがBよりCだ」などの比較構文がある。比較対象が、日本語では「Bより」という格要素で表現されるのに対し、英語では“A is more C than B (is C).” のように従属節となるのが特徴的である。しかし、これは一定の仕掛けを作ることにより翻訳処理の中で対応した。

(3) 複合形について

「研究開発する」「海外旅行する」「自動開閉する」「栄養豊富だ」といった複合表現は無数に生成可能であるため、異なりをすべて辞書に登録するという方針では対応しきれない。そこで「研究する＋開発する」「海外へ＋旅行する」「自動(的に)＋開閉する」「栄養が＋豊富だ」のように、用言と用言、格要素と用言などのように、解析処理で分解することにより対応することとした。ただし、別登録すると合成が困難になるなどの個別の事情があれば一般表現文型を作成した。

4.2 慣用表現文型

一般表現文型以外の構文としては慣用表現文型がある。慣用表現を厳密に定義するのは難しく、“単語の2つ以上の連結体であって、その結びつきが比較的固く、全体で決まった意味を持つ言葉だという程度のところが、一般的な共通理解になっているだ

ろう”[宮地 82]とされている。慣用表現は、述語として働くものに限っても、技術文献に3~5%出現すること[奥 87]、言語の翻訳は慣用表現から慣用表現への対応付けであると指摘する翻訳家がいること[中村 83]を考えると、辞書に収録すべき重要な表現であるといえる。

慣用表現を厳密に定義するのが難しいため、慣用表現を認定するのも容易ではない。例えば、〈サボる〉の意味の慣用表現「油を売る」では、「木陰で油を売る」は言えても「油を木陰で売る」が言えないように、語順が固定的であるのが1つの特徴といえる。また、「安い油を売る」「ガムの油を売る」のように「油」が連体修飾されると慣用表現ではない(一般表現である)。従って、慣用表現文型にはこういった慣用表現を判定する上で手がかりとなる情報(「油を売る」では「どこで」や「油」に連体修飾がないこと)を併せて記述する。しかし、この例で「ガソリンスタンドで油を売る」のように判定困難な場合も考えられる。今後、条件の見直しなどの検討が必要である。

日英翻訳の観点では、上述のような典型的な慣用的表現のほかに、英単語が日本語の連語表現に対応するため、慣用表現に準じて扱う方がいいものがある。例えば、「背が高い」は日本語では慣用表現であるとは思えない。しかし、これを文字通り、「背」→“back”，「高い」→“high”と英訳しても意味をなさないのは明らかであり、“be tall”に対応づけなければならない。このような表現も正しく訳されるか否かが大きな問題であるので、このタイプも慣用表現に準じて扱うことにする。

このほか、慣用的といえる表現としては機能動詞表現[村木 80]がある。例えば、「連絡を取る」、「影響を受ける」のように、動作名詞が実質的な意味を表し、動詞は文法的な機能を担うのが特徴である。これらは動作名詞を動詞化することにより、「連絡する」、「影響される(影響する+受身)」といった一般表現文型に還元できる。従って、日本語解析の範囲と考え、慣用表現文型としては収録しないことにする。ただし、解析の際には、「彼の影響を受ける」→「彼に影響される」、「強い影響を受ける」→「強く影響される」など動作名詞に対する修飾要素の変形が必要となるほか、機能動詞と紛らわしい表現(「注目を集める」は「注目される」と言い換えることができるが、「署名を集める」を同様に「署

名される」と言い換えることはできない)に注意することが必要である。

4.3 記述する修飾要素の範囲

述語に対する連用修飾要素には格要素と副詞要素がある。このうち格要素は、助詞表現を伴うものと、時間表現や数量表現など助詞を伴わないものがある。この時間表現や数量表現は副詞的に働く場合が多い。

初期の日本語文型では「格要素+述語」が文型の骨格をなすという意識が強かったため、副詞要素を除外し、格要素のみをパターンに記述するという方針を立てた。英語には主語や目的語のように構文的に必須となる要素があるが、日本語にはそのような意味での必須性は希薄であるため、どの格要素を記述するかは判定が容易でない場合がある。日英翻訳への適用を目的とすることから、格要素の記述の要否は、必要に応じて英語表現と対比することにより決定した。すなわち、英語の主語、目的語、補語や、英語文型を特徴づける前置詞句に対応する日本語の格要素を記述する。そして、時間、場所、原因理由などを表す副詞的要素は原則として記述しない。

しかし、助詞を伴わない数量表現が英語の目的語と対応する表現文型が存在する場合や日本語の副詞と動詞が英語の動詞1語に対応する場合も少なくないため、副詞的格要素をすべて除外するのは問題がある。そこで、これらに対応するため、時間表現や数量表現などの副詞的格要素や副詞要素も必要に応じて日本語文型に記述するように方針を変更した。そして、その表現文型に特徴的と考えられる要素には、必須性を指定するフラグを付与した。

4.4 格要素の制約条件

格要素は名詞句と助詞表現とからなるため、名詞句に対する制約条件と、助詞表現に対する制約条件が考えられる。

名詞句に対する制約条件は、初期は名詞句の中心要素として働く名詞に対する意味的制約条件として、一般名詞意味属性を用いて抽象化することにより、該当する属性、該当してはならない属性を指定することにより記述した。このようにすれば、適合する文型パターンの選択をある程度までは排他的に制御できるからである。

また、助詞表現に対する制約条件は、標準的に使用されると思われる格助詞1語により指定した。助詞表現には様々なバリエーションがあるが、日本語

解析で標準化する。例えば、「彼は彼女に対して挨拶する」では「彼が彼女に挨拶する」と「彼を彼女に挨拶する」の2つに標準化し、日本語文型との照合により助詞を最終的に決定する。ただし、文型パターンに特徴的な助詞表現や格助詞への還元が困難な場合は特殊な助詞表現を許容した。なお、助詞を伴わない格要素は、「助詞を伴わないこと」が制約条件である。

中心名詞に対する一般名詞意味属性による制約条件の指定には次のような問題がある。多用される文型パターンでは名詞のバリエーションの範囲がすぐにわかるため、一般名詞意味属性による抽象化は容易で、最適な制約条件の記述が可能であるのに対し、特殊な文型などでは名詞のバリエーションの範囲がわからないため、適切な抽象化は困難な場合があり、文型パターンの排他制御ができない恐れがある。

そこで、名詞句に対する制約条件として、中心名詞を一般名詞意味属性で指定する以外に、中心名詞の字面そのものの指定、そのほか、複合語や連体修飾など、名詞句の構成要素の個別指定なども許容することにした。このように字面による指定を認めたことにより、一般表現文型と慣用表現文型の垣根が低くなる効果が生じた。先に述べたように、慣用表現を厳密に識別するのは難しい場合や、名詞句の制約条件の設定に自信がない場合に、とりあえず字面を指定する。この指定では、同じ字面の名詞にしか文型は適合しないため、汎用性はないが副作用の恐れもない。後に、類似用例が見つかるなどして、条件設定の汎用化が可能になった段階で、改めて条件設定を見直すようにする[白井 94]。

また、「野菜の水を切る」では格要素の中心要素は「水」であるので、日本語文型には「水」と「を」を制約条件として記述すればよいように思われる。しかし、対応する英語は、例えば“drain water off vegetables”であり、「野菜」と「水」がそれぞれ“drain”に支配されている。これに対応するため、上述の連体修飾の指定を必要に応じて分離することができるよう、日本語の記述形式を改良した。

4.5 文型パターンへの付加要素

述語に対して、使役、受身、可能などの語尾表現が付加されると、文型パターンは変化する。この変化の仕方はいくつかの場合分けが可能であるので、初期は変化形の収集は対象外としていた。すなわち、

述語は動詞や形容詞の終止形を記述した。

ただし、「花を持たせる」は表面的には「花を持つ+使役(せる)」であるが、「せる」を伴って初めて慣用表現として働く文型である。このため、慣用表現文型については、慣用表現を特徴づける付加要素をパターン対に記述した。

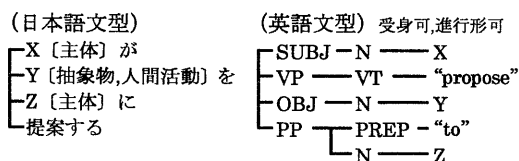
パターン対を特徴づけるという観点からは、「回転させる」→“rotate”も「回転する+使役(せる)」というよりも、自動詞「回転する」の他動詞化であると考えられる。また、「乾燥している」→“be dry”も「ている」を伴って初めて英語との対応付けが可能となる。パターン対の収集が進むに伴い、このような付加要素も積極的に記述するようにした。

「～し出す」は英語の“begin to ～”に対応する様相表現として処理するため、通常は文型パターンの作成対象とはしない。しかし、「編み出す」には〈編み始める(begin to knit)〉のほかに〈新しい方法を考え出す(devise)〉の解釈が可能であり、字面上は識別するのが困難である。このような場合は、例外的に両者を登録した。

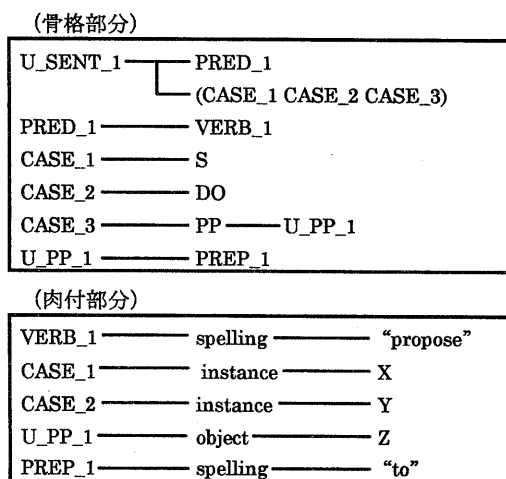
4.6 英語文型パターンの記述

英語には5文型のような決まった表現形式があり、主語の人称や数と定動詞の屈折変化には相関がある。日本語文型では、使役、受身、可能などをパターン対への付加要素として扱うことを述べたが、この付加要素を英語に反映するには主語や目的語、動詞といった文法機能がわかっている方が便利である。そこで、日本語文型と英語文型の対応づけでは、単に格要素がどの位置に来るかだけでなく、それがどのような文法機能を持つかを併せて記述することにした。文法機能としては、主語(SUBJ)、動詞句(VP)、目的語(OBJ)、補語(CO)、副詞句(ADVP)、前置詞句(PP)などを上位概念として設け、下位概念として具体的な単語の品詞などを指定した。また、文型を変形するための情報として、受身の可否、進行形の可否を指定した。

例えば、「XがYをZに提案する」→“X propose Y to Z”では次のようになる。



初期は、このように文法機能の上位概念と下位概念を一体的に記述した。このため、複雑な構造のとき全体の構造を捉えにくくなったり、同じ構造であっても文法機能を毎回書く必要があることが問題となった。そこで、英語文型の骨格部分と肉付部分を分けて記述するように改めた[横尾 94]。上の例をこの方式で記述すると次のようになる。



この骨格構造と英語文型の関係を見ると、英語文型の70%までが骨格構造9個でカバーされ、以下、80%までが18個、90%までが51個、95%までが131個など、600個ほどの骨格構造ですべての英語文型が表現される。

5 現状の到達点と今後の展望

収集したパターン対は1996年末時点では16,000件で、内訳は一般表現パターン13,000件、慣用表現パターン3,000件である。このうち、制約条件の汎用化作業中のもの1,200件を除き、格要素の中心名詞を字面指定したもの(4.4節参照)を慣用表現パターンにして集計すると表1のようになる。

表1 構文辞書に収集済みのパターン対

項目	一般表現パターン対		慣用表現パターン対	
	異なり用言	パターン対	異なり用言	パターン対
和語動詞	1,244	4,298	517	2,648
複合和語動詞	516	908	33	54
サ変動詞	3,075	4,375	24	34
形容詞(イ型)	266	569	50	564
形容詞(ナ型)	902	1,341	1	3
名詞	24	25	0	0
合計	6,027	11,516	625	3,303

直観的には、複合和語動詞、サ変動詞、ナ型形容詞の用言数が少ないようなので、今後はこれらを補充していく必要がある。例えば、複合和語動詞は、「現代国語例解辞典（第1版、小学館、1985年）」には約2,200語ほど記載されている。

また、以前、パターン対の収集可能な量を見積もったところ、一般表現パターン対 20,000件、慣用表現パターン対 5,000件であった[白井 95a, 白井 95b]。今後は、用例文をどれだけ効率的に集めるかが特に重要な課題となる。

人手による用例文作成では個人差の問題が考えられるが、和語動詞の例文作成実験では用言の用法に関する言語知識の個人差はそれほど大きくないと判断される[池原 96, 白井 97a]。また、客観性の観点では言語コーパスの使用が考えられるが、大規模な対訳コーパスの作成は難しく、抽出される有意な表現は辞書に収録済みのものが多く、網羅性、効率とも問題がある。このため、現在の条件下では、内省による用例文作成とそれに基づくパターン対収集が、網羅性と実現性の両方を満たすと考えられる。

英和辞書は表現が説明的であるなどの理由でこれまではパターン対の収集対象とはしてこなかった。しかし、連語的な表現を1つの英単語に対応づけるなど、明快な英語を作成する上で有効なパターン対が収集できる可能性が高い[白井 97b]。例えば、次の例である。

彼は樽の栓をあけた。
He *tapped* the barrel.

今後はこのような可能性についても検討する。

6 おわりに

結合価パターン対を体系化した構文辞書の収集経過を振り返り、記述内容について概要を述べるとともに、現状の到達点をまとめた。実際の文型パターンを収集することにより、初めて具体的な記述形式を決めることができたといえる。

今後は、訳し分けが問題となる和語動詞やイ型形容詞のパターン対を充実させるとともに、複合和語動詞、サ変動詞、ナ型形容詞など、現時点で収録語数の少ない用言のパターン対を重点的に収集する。また、連語的な表現を英単語1語に対応づけるためのパターン対の収集も試みる。これらにより、構文辞書の整備を進めていく。

謝辞 本辞書は、林良彦氏・奥雅博氏（現在は NTT 情報通信研究所）を中心に設計を開始し、元(株)ケーピーエスの細井純子氏・八木晶子氏、NTT アドバンステクノロジーの井上浩子氏・渡辺いづみ氏らを始めとするアナリストの方々により構築されてきた。また、英語パターン対の記述に関しては、小倉健太郎氏・Francis Bond 氏（NTT コミュニケーション科学研究所）の協力を得た。以上の方々に感謝する。

参考文献

- [林87] 林良彦：結合価構造に基づく日本文解析，情処研報，Vol.87, No.53, pp.39-44 (1987)
- [池原96] 池原悟，白井論，相沢弘：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)
- [IPA 87] 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)
- [宮地82] 宮地裕 編：慣用句の意味と用法，明治書院 (1982)
- [宮崎97] 宮崎正弘，池原悟，横尾昭男，白井論：日英機械翻訳のための意味属性体系，信学会NLC研究会&情処学会NL研究会 (1997)
- [中岩94] Nakaiwa, H., Yokoo, A., and Ikehara, S.: A system of verbal semantic attributes focused on the syntactic correspondence between Japanese and English, In Proc. of COLING94, pp.672-678 (1994)
- [中岩97] 中岩浩巳，池原悟：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，処理学会論文誌，Vol.38, No.2, pp.215-225 (1997)
- [水谷83] 水谷静夫，石綿敏雄，荻野孝野，賀来直子，草薙裕，青山文啓：「文法と意味I」朝倉日本語新講座3，朝倉書店 (1983)
- [村木80] 村木新次郎：日本語の機能動詞表現をめぐって，国立国語研究所報告65 研究報告集(2)，秀英出版 (1980)
- [中村83] 中村保男：翻訳はどこまで可能か，ジャパントイムズ (1983)
- [奥87] 奥雅博：日本語慣用表現の分析と日英翻訳への適用，情処研報，Vol.87, No.53, pp.9-14 (1987)
- [白井94] 白井論，横尾昭男，池原悟，井上浩子：日英機械翻訳用構文意味辞書の向上と作成支援，情報処理学会第48回全国大会，Vol.3, 6Q-9, pp.141-142 (1994)
- [白井95a] 白井論，池原悟，横尾昭男，井上浩子：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，Vol.95, No.110, pp.43-50 (1995)
- [白井95b] Shirai, S., Ikehara, S., Yokoo, A. and Inoue, H.: The quantity of valency pattern pairs required for Japanese to English machine translation and their compilation, In Proc. of NLP95, pp.443-448 (1995)
- [白井96] 白井論，井上浩子，小出ひとみ，井田倉紀子，横尾昭男：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，Vol.2, 4L-4, pp.59-60 (1996)
- [白井97a] 白井論，横尾昭男，池原悟，武智しのぶ，分部恵子：英単語に対する連語的日本語訳語の分析，言語処理学会第3回年次大会，D6-4, pp.469-472 (1997)
- [白井97b] 白井論，大山芳史，我妻知恵，石崎俊：英単語に対する連語的日本語訳語の分析，言語処理学会第3回年次大会，B1-2, pp.55-58 (1997)
- [横尾94] 横尾昭男，中岩浩巳，白井論，池原悟：日英機械翻訳用スケルトン-フレッシュ型構文意味辞書の構成，情報処理学会第48回全国大会，Vol.3, 6Q-8, pp.139-140 (1994)
- [横尾97] 横尾昭男，宮崎正弘，池原悟，白井論，阿部さつき：日英機械翻訳のための単語辞書，信学会NLC研究会&情処学会NL研究会 (1997)