自然言語処理 110－7
（1995．11．17）

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

白井 諭† 池原 悟† 横尾昭男† 井上浩子‡

†ＮＴＴコミュニケーション科学研究所
〒238-03 横須賀市武1-2356
{shirai, ikehara, ayokoo}@nttkb.ntt.jp

‡ＮＴＴアドバンステクノロジ（株）
〒244 横浜市戸塚区川上町90 6
inoue@totsuka.ntt-at.co.jp

あらまし 機械翻訳における意味解析には，動詞と名詞の意味的な共起関係に着目した結合価が有効であることが知られている。最近，用例から自動学習によって結合価パターンを自動生成する方法が研究されている。しかし，どれだけの数のパターン対を用意すればよいかが不明であること，また，それを作成するに必要な用例を現実の文書から網羅的に収集するのは困難であること等の問題があった。

これらの問題を解決するため，本論文では，人用の辞書の知識と人の知識を内省する方法で作成した用例を使用して，和語動詞の一部を対象にパターン対の作成実験を行った。具体的には，(1)和英辞書から収集する方法，(2)和英辞書に加え日本語動詞の語義対応の用例を使用する方法，(3)それらを参考に人の知識に基づいて用例を作成する方法，の３種の方法を取り上げ，30種の和語動詞に対するパターン対の数を比較した。その結果，(2)の方法は(1)の方法の約２倍，(3)の方法はさらにその２倍のパターン対が収集できることが分かった。また，日英機械翻訳において訳し分けの必要な約1,000種の和語動詞に対し約7,500件のパターン対が必要であり，それを作成するには15,000件の用例収集が必要であることなどが分かった。

これらの結果，漢語動詞や用言性慣用表現を含む日本語述語全体では約25,000件のパターン対が必要であること，また，人手でパターン対を作成する場合は，辞書情報などを参考に人の知識を内省して用例を作成する方法によって，パターン対作成に必要な用例をほぼ網羅的に収集できることが分かった。

キーワード 結合価パターン，意味解析，日英機械翻訳，対訳用例

The Quantity of Valency Pattern Pairs Required for Japanese to English MT and Their Compilation

Satoshi Shirai †, Satoru Ikehara †, Akio Yokoo † and Hiroko Inoue ‡

† NTT Communication Science Laboratories
Take 1-2356, Yokosuka, 238-03, Japan
{shirai, ikehara, ayokoo}@nttkb.ntt.jp

‡ NTT Advanced Technology Corporation
Kawakami-cho 90-6, Totsuka-ku, Yokohama, 244, Japan
inoue@totsuka.ntt-at.co.jp

Abstract In order to realize the valency pattern method, which is used in the semantic analysis of co-occurrence of verbs and nouns, this paper discusses how many pattern pairs should be prepared and the method of collectively gathering these patterns. A pattern pair preparation method is proposed that combines existing knowledge compiled in dictionaries for human use with examples prepared manually by relying on personal knowledge.

Specifically, three methods are examined. The results show that Japanese to English machine translation requires about 7,500 pattern pairs to cover the 1,000 Japanese origin verbs that are critical to differentiated translation. Preparing this number of pairs requires the collection of 15,000 examples. It is also predicted that about 25,000 pattern pairs would be required to cover all Japanese predicates including verbs of Chinese origin and idiomatic expressions of declinable word type. Furthermore, the method of preparing examples through human knowledge is shown to be entirely feasible.

Keywords valency pattern, semantic analysis, Japanese to English machine translation, bilingual corpus

1 はじめに

機械翻訳において意味解析技術の重要性が指摘されている。意味解析の方法としては，単語の共起関係に着目して単語相互の意味を決定する方法が研究されているが，中でも，動詞の意味解析においては，動詞と名詞の意味的な共起関係に着目した結合価パターンを使用する方法が有効であることが知られている。この方法を実現するには，パターンの記述精度の問題とパターン対収集方法の問題がある。

パターン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，慣用表現を除き，日本語の動詞を訳し分けられるようなパターン対が記述できることが知られている[池原93]。

これに対して，パターン対収集の問題についても，既に，種々のヒューリスティックスや学習技術を応用した方法が提案されている。しかし，どれだけのパターン対を作成すればよいか不明であること，網羅的にパターン対を作成するのに必要な用例を実際の文書から収集するのは困難なことなど，様々な問題があり，実用できるレベルにない。例えば，黒橋らは例文とシソーラスを用いて文型を同定する方法を提案した[黒橋92]。また，アルモアリムらは自動学習の手法を用いた翻訳ルールの自動抽出方法を提案し，6動詞に対し各27～80の対訳用例を用いた抽出実験に成功している[アルムアリム94a, 94b]。しかし，これらの方法を使用するには，学習に必要なだけの種類と量の例文を入手できることが前提となる*1。例えば，日英翻訳の場合，使用頻度の高い和語動詞のパターン対をほぼ網羅的に学習させるには1,000万ペアの日英対訳文が必要であると言われている[金田94]。しかも，その対訳文は動詞と名詞を組み合わせた単純な文形式で与えられなければならない。実際の文書から得られた例文は通常，複雑な構造を持つ場合が多いので，目的にあわせて単純化する作業が必要となる。このように，膨大な量の単純化された用例を実際の文書から機械的に収集することは，事実上，不可能である。

これに対して機械翻訳では，一度網羅的なパターン対が完成すれば*2，繰り返しパターン対を作成する必要はない*3。また，訓練されたアナリストによれば，適切な対訳用例があれば，類推能力によって，1用例から1パターン作成することができると推定される。これらの点を考えれば，現状では，パターン対作成作業はむしろ人手を中心に進め，計算機はあくまで作業支援に使用するのが現実的と考えられる*4。

そこで，本論文では，人手によるいくつかのパターン対作成の方法について部分的な作業実験をし，その結果から，日英機械翻訳ではどれだけの数のパターン対が必要か，また，それは実際にはどのような方法によれば作成できるかを明らかにする。

具体的には，単語当たりの語義数が多いためパターン対の相互関係が問題となる和語動詞約1,000語の中の代表的な動詞を対象に，(1)人間用の和英辞書に記載された語義に着目する方法，(2)日本語の語義に着目する方法，(3)人間の知識を内省して用例を作成し，その用例からパターン対を作成する方法の3種類の方法を示し，それらの方法でどれだけのパターン対が収集できるかを検討する。また，得られたパターン対の数から，和語動詞全体では最終的にどれだけの数のパターン対を作成すればよいかを推定し，その作成方法について議論する。

最後に，漢語動詞，形容詞系述語，名詞述語のほか，用言性の慣用表現を含むパターン対全体で必要なパターン対の数についても考察する。

2 前提条件

2.1 パターン対記述の枠組み

機械翻訳において用言と名詞の共起関係の知識を結合価パターンにまとめるには，対象となる用言の種類，名詞の意味分類の方法等が問題となる。特に，名詞の意味分類では，翻訳する言語ペアによって必要とされる分解能に差が生じる。日英機械翻訳の場合は，日本語の用言と英語の用言の意味的な対応関係が記述できる程度の分解能を得るため，日本語の名詞の意味を2,000種程度以上に分解整理することが必要とされている[池原93]。本論文では，この条件を満たしていると見られる日英機械翻訳システムALT-J/E[池原89]の枠組みを用いてパターン対の作成方法を検討する。以下では，ALTのパターン記述の枠組みを示す。

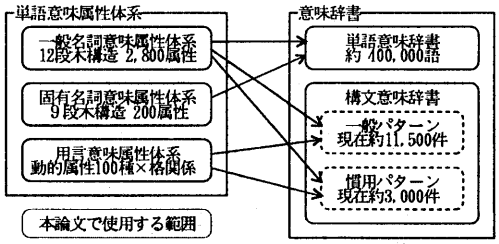

図1 ALT-J/Eにおけるパターン対記述の枠組み

ALTの結合価パターン記述の枠組みは，図1に示すように，日本語名詞に対する単語意味属性体系と2つの意味辞書（単語意味辞書，構文意味辞書）から構成される。単語意味属性体系は，3種類の意味属性体系から構成されるが，結合価パターンの記述には，そのうちの一般名詞意味属性体系が使用される。これは，日本語名詞の意味的な用法を約2,800種の属性名で表現し，それらの相互の意味的関係を12段の木構造に整

*1 自動学習の方法では，モデルを単純化せざるを得ないなどの理由で精度上もまだ問題があり，実用に展開するのは困難とみられる。
*2 様々な言語現象で見られるような使用頻度の低いパターン対でも，それを合計した出現頻度は無視できない頻度となることが予想されるため，機械翻訳システムにおいては（専門分野依存は別にして）一般的なパターン対はあらかじめ網羅的に装備する必要がある。
*3 ここでは専門分野固有のパターン対を除く。専門パターン対については後で触れる。
*4 計算機による支援としては，不足しているパターン対の作成支援のほか，人手で作成されたパターン対の相互無矛盾性の検証支援も重要である。自動学習技術の応用研究には，パターン対と単語意味属性体系の間の相互矛盾を論理的に検出する仕組みの開発が期待される。

理したものである。単語意味辞書では，単語約40万語の持つ意味（１単語１つ以上）が単語意味属性を用いて記述されている。また，構文意味辞書は日本語の結合価パターンとそれに対応する英語の構文パターンをペアとして持つ。これらの辞書は，構文解析結果の絞り込み，動詞の訳語の選択，名詞訳語の選択等の意味解析に使用される。

ALTの結合価パターンは，用言（動詞，形容詞），格要素（主名詞＋助詞），副詞要素，様相情報から構成される。主名詞は，通常，日英の動詞が訳し分けられる最小限の深さの意味属性を用いて記述される[池原93]。意味属性で代表できないような名詞の場合は，名詞そのものが使用される。格要素の主名詞が意味属性によって指定されたパターンを一般パターン，１つ以上の格要素の主名詞が名詞そのものによって特定されたパターンを慣用パターンと呼ぶ*1。慣用パターンは，慣用表現や固定化した比喩的な表現に対する日英間の対応付けのために使用される。本論文では，一般パターン対の収集を対象とする。

結合価パターンは，述語となる用言（動詞，形容詞）毎に作成される。日本語では名詞が述語になる場合があり，この「名詞＋だ(です)」型の述語は一般に英語では名詞補語として訳出されるが，名詞補語には訳出できないものに対し名詞を述語とするパターンが作成される。例えば，「今日は晴れだ。➡ It is fine today.」や「あなたに質問です。➡ I ask you a question.」などである。また，述語が複合語の場合，例えば，「成功は努力次第だ。➡ Success depends on one's efforts.」に対しても同様にパターン対が作成される。

2．2 パターン対作成の方法

精度の良いパターン対を効率的に作成するには，対訳用例からパターン化すべきものを発見しパターン対の作成を支援する仕組みと，作成したパターンと既存のパターンとの間の無矛盾性を検証する仕組みが大切である。ALTでは，パターン対作成の過程を支援するために図２に示すような仕組みを実現した。

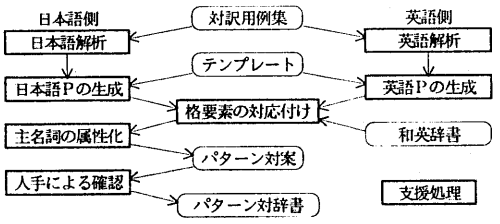

図２ ALT-J/Eにおけるパターン対作成支援の仕組み

(1) パターン対作成支援の方法

日英機械翻訳用のパターン対の構造はその大半が約10種類のテンプレートで記述できることが知られている[横尾94]。従って，これを使用して与えられた日英対訳用例の中から日本語側，英語側のパターン要素を指定すれば，最低限のパターンは容易に作成できる。しかし，質が良く汎用性の高いパターン対を作成するには，パターンの適用範囲を決定する名詞要素の記述が大きな問題となる。この作業を支援するため，ALTでは以下のコンピュータ支援処理を実現した。

例えば，用例「彼は電話を引いた。➡He installed a telephone.」に対して，まず「X〔主体〕が/“電話”を/引く ➡ X install a telephone」というパターンが作成される。支援処理は単語意味辞書を見て名詞“電話”の意味属性とその上位の意味属性を表示するので，アナリストはこれを見て“電話”の部分を汎用的な意味属性に置き換えてパターンを作成するか，そのまま辞書に登録するかする。そのまま登録した場合，その後，日本語動詞“引く”，英語動詞“ install ”である用例が追加されたとき，支援処理が再度，ヲ格の名詞（複数）に共通する意味属性を表示するから，それを見てアナリストはパターンを汎用化できる。用例が増加すれば意味属性候補の判断はより正確になる。

(2) パターン対相互チェック支援の方法

結合価パターンは述語を見出し語として登録されるから，見出し語が異なるパターンの間で相互に干渉することはない。従って，パターン相互の無矛盾性をチェックするには，同一の見出し語を持つ用例を対象に翻訳実験を行えばよい。そこでALTでは，パターン相互の無矛盾性チェックを支援するため，以下の手順を半自動的に実施する仕組みを実現した。まず，前述の(1)の処理のあと，パターン作成に使用した用例とそれに対する機械翻訳の結果を保存する。再び(1)の手順で新パターンを作成したときは，新パターンを暫定的に登録した後，同一の見出し語を持つ既存の用例を対象に翻訳実験を行う。その結果を過去の翻訳結果と比較して，差分の生じた用例とその翻訳に使用されたパターン対を出力する。アナリストはそれを見て，新パターンの最終的な登録の可否を判断する。

無矛盾性チェックの結果によっては，新パターンの作成だけでなく，既存パターンの修正が必要な場合もある。パターンの修正はまた(1)に戻って実行される。

3 パターン対収集の方法

上記の支援システムはあくまで人手作業を支援するものであり，すべての知的判断は人手で行われる。そして，判断に使用される基本情報は日本語用言の語義もしくはその用例である。そこで，用言の語義および用例の入手方法に着目して，パターン対収集の手順を３段階に分けて考える。すなわち，(1)和英辞書の語義を参照する方法，(2)日本語の語義に基づく方法，(3)人の知識を内省する方法の３種類の方法を順に適用してパターン対を作成するものとする。

3．1 和英辞書の語義分類に基づく方法

(1) パターン対収集の方法

パターン対を収集する第１の方法として和英辞書の情報を参照する方法を考える。人間用の和英辞書には，

*1 一般パターン，慣用パターンのほかに，特定の専門分野を対象とする専門パターンがあるが，本論文では専門パターンは扱わない。

日本語の用言に対して，語義とそれに対応する英語の動詞や語法，例文などが記載されている。従って，これらの辞書に記載された語法や例文を分析し，格要素，副詞要素などの日本語側の制約条件を整理すれば，日本語動詞と英語動詞のペアに対してパターン対を作成することができる。例えば，ライトハウス和英辞典[研究社84]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文がある。

彼の学校の成績が上がった。

His school record ***has improved.***

この例文の文要素を分析し，若干の情報追加を行えば，図３のようなパターン対が得られる。

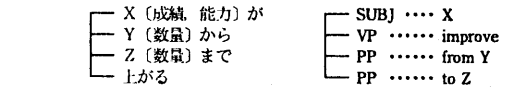

図３ 人間用の辞書を使用したパターン対作成の例

本論文では，何冊かの和英辞書*1を使用してパターン対を作成した。

(2)収集されたパターン対の数

和英辞書に含まれる主な用言5,600語に対して上記の方法でパターン対を作成した。得られたパターン対は，当初，一般パターン10,000件，慣用パターン5,000件であった。その後の見直しにより，一般パターンの中に統合できるものが含まれていること，また，慣用パターンの中にも汎用化できるものがあることなどが分かり，辞書から収集したパターン対は一般表現10,000パターンと慣用表現3,000パターンとなった。

(3)翻訳実験での充足性

上記で得られたパターン対を使用して，情報処理装置関連の仕様書（1,361文）の翻訳実験を行った。その結果によれば，試験文中に現れた用言の種類は142件，翻訳に必要なパターンは201件であるのに対して，本節の方法であらかじめ準備できていたパターン対は120用言に対する154件であった。試験文中の22の用言（22パターン）はパターン対が登録されていないこと，また，23の用言に対しては合計25のパターンが不足していることが分かった。

この例から見れば，用言数で15%(22/142)，パターン数で23%(22+25/201)が不足していることになる。中でも，パターン対が不足している用言は，単語当たりの語義の多い和語動詞が多い。

３．２ 日本語辞書の語義分類に基づく方法

(1)パターン対用例収集の方法

前節で見たように，和語動詞は語義が多いため，通常の和英辞書の語義分類だけでは翻訳パターンを網羅的に収集することは困難である。これに対して，和語動詞については，かねてより日本の言語学者（20名あまり）を中心にその語義と対応する用例を収集分析する研究が進められており，既に861動詞に対して語義と語義毎の用例（ただし，日本語用例のみ）がＩＰＡＬ動詞辞書[IPA 87]*2としてまとめられている。そこで，本節では，第２の方法として，日本語の語義をより詳細に分類する立場から，この辞書の用例を使用したパターン対の収集を考える。

具体的には，ＩＰＡＬ動詞辞書の各語義に示されている用例に対して，日本語原文に忠実で，かつ，英語としても十分通用する英訳文を翻訳家に作成してもらい，その対訳データからアナリストがパターン対を作成する方法でパターン対の収集を試みる。

(2)収集されたパターン対の数

上記の方法では，861の和語動詞に対して，5,243文（和文7.5万字，英文４万語）の対訳例文が得られた。これらの対訳用例を使用したパターン対作成作業では1,399パターン対が新規に作成され，既存のパターン対のうち414件が修正された（一部見直し中）。

(3)追加拡充の程度

ＩＰＡＬ動詞辞書は，日本語動詞の語義分類に基づいて用例が作成されている。従って，日英翻訳用のパターン対から見ると，日本語動詞の語義とパターン対との対応関係（１語義が１パターンに対応するか）が問題となる。そこで，日本語の語義の多い代表的な４動詞について，語義とパターン対の対応関係を調査した。その結果を表１に示す。この表から，日本語用言の語義とパターン対が１対１に対応するものは４割にとどまり，両者は必ずしも対応しないことが分かった。このことは，日英機械翻訳から見れば，ＩＰＡＬ辞書の語義分類は，英語に訳出するうえで，必ずしも適切ではないことを意味している。すなわち，日英翻訳では，日本語と英語の意味的対応関係に即して，日本語の語義分類をする必要のあることが分かる。

表１ ＩＰＡＬ語義と文型の対応

動詞＼分類	「語義」対「パターン」の関係					合計
	１対１	１対ｎ	ｍ対１	ｍ対ｎ	保留	
あがる	8	5	1	3	1	18
あげる	14	2	1	1	3	21
だす	8	9	5	4	1	27
でる	13	3	10	4	2	32
合計	43 43.9%	19 19.4%	17 17.3%	12 12.2%	7 7.1%	98 100%

３．３ 人の知識を内省する方法

(1)パターン用例収集の方法

前節までの結果から，人間用の和英辞書，日本語辞書の双方から用例を収集しても十分なパターン対が作成できないことが分かった。日本語例文とパターン対の関係を観察すれば，同じ動詞を使用していても，動詞の使われ方のニュアンスが異なるときに新たな英語パターンが必要となる場合が多いことに気がつく。そこで，第３の方法として，英語の理解できる日本人が

---

*1 慣用パターン作成では，一般辞書のほか慣用表現辞書も使用した。

*2 日本語の動詞において語彙体系上ならびに使用頻度上重要であると考えられる基本的な和語動詞861語（ひらがな表記した場合で，漢字表記では1,301語に相当する）について，意味および統語的特徴に基づいて下位区分し，それを１つの単位として，意味，形態，統語，文法的カテゴリ，慣用表現などに関わる情報が詳細に記述されている。また，各下位区分ごとに１～３文の用例が付されている。

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法を可能な限り列挙するという方法で用例の収集を試みる。

列挙する用例は，作業に掛ける時間にもよるが，ある程度の時間以上考えても用例が思い浮かばなくなるまで抽出することとした*1。用例数としては，いくつかの動詞について思考実験した結果に従い，ＩＰＡＬ動詞辞書の語義数の２～３倍を目標とした。また，これらの日本語用例に対する英訳は翻訳専門家に依頼し，対訳用例集を作成することとした。

(2)収集されたパターン対の数

上記の方法による作業結果では，約1.5人年の作業により，861動詞*2に対し用例10,500文（和文13万字，英文6.8万語）が収集された*3。

また，収集した用例から，語義数の多い動詞と少ない動詞が混合するように36動詞（1,100用例文）を選び，パターン対の抽出を試行したところ，新たに300パターンが抽出された。第１，２の方法で得られなかったパターン対が１動詞当たり平均10パターン見つかったことになる。

4 収集された用例とパターン対の数の比較推定

4．1 収集された用例の数とパターン対の数の比較

前述の36の和語動詞について，３章で述べた３つの方法によって得られた用例数とパターン対の数を比較して表２に示す。参考のため，この表のほぼ中間位置にある動詞「上がる」について，３種の方法で得られた動詞用例とそれらの用例から得られたパターン対の関係を付表１～３に示す。

表２から，各手法で得られたパターン対の数の関係を示すと図４の通りとなる。これらの図表より，以下のことが分かる。

1)第１の方法に加えて第２の方法を実施すれば，第１で得られるパターン対の数の約２倍のパターン対が収集できる。

2)第１，第２の方法に加えて第３の方法を実施すれば，第１，第２で得られるパターン対の数のさらに２倍以上のパターン対が収集できる。

これらの結果から，和語動詞について見れば，和英辞書がら収集されるパターン対の約４倍が人間の知識の内省によって得られることになる。

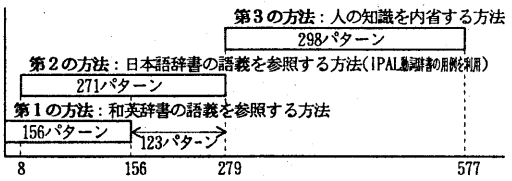

図４ ３種の方法で得られたパターン対の種類の関係

表２ 収集されたパターン対の数の比較（和語動詞の例）

収集方法		第１	第２の方法			第３の方法		パターン数	[続]
IPAL表記	漢字表記	P数	語義	例文	追加P	用例	試用P	合計	信頼P
でる	出る	22	32	49	5	145	38	65	18
だす	出す	16	27	53	15	95	22	53	21
あける	空ける	4	11	17	1	14	5	10	0
	明ける	4			0	9	2	6	1
	開ける	3			1	9	2	6	1
たつ	立つ	5	13	24	4	75	30	39	11
	発つ	2			0	6	1	3	0
	建つ	1			0	5	0	1	0
	経つ	2			0	3	0	2	0
あく	空く	4	10	12	4	13	4	12	1
	開く	5			2	12	4	11	0
たてる	立てる	8	9	17	7	69	29	44	7
	建てる	1	1		0	5	0	1	0
あげる	上げる	8	21	31	13	98	16	37	14
おちる	落ちる	8	11	21	7	53	23	38	1
たつ	断つ	4	1		0	6	1	5	0
	絶つ	4	3		0	6	2	6	0
あがる	上がる	7	18	31	16	90	16	39	12
はいる	入る	7	23	34	11	105	31	49	5
おとす	落とす	6	14	19	5	53	15	26	3
くずす	崩す	6	4	4	2	8	2	10	0
いれる	入れる	5	19	30	12	113	28	45	10
くずれる	崩れる	5	4	6	2	13	4	11	0
きめる	決める	3	14	20	4	28	5	12	0
さける	避ける	3	6	11	0	9	2	5	0
きまる	決まる	3	8	17	2	32	10	15	2
うめる	埋める	3			1	9	0	4	1
	（うめる）		4	5					
	（うずめる）		4	4					
さく	割く	2	5	7	0	4	2	4	0
	裂く	1			5	3	0	6	0
うまる	埋まる	2			2	5	1	5	0
	（うまる）		5	6					
	（うずまる）		3	4					
さける	裂ける	1	1	3	2	4	1	4	0
さく	咲く	1	1	1	0	3	2	3	0
計		156	271	426	123	1102	298	577	108

4．2 必要なパターン数と用例数の見積もり

(1)和語動詞の場合

第３の方法で得られるパターン対の網羅性を調べるために，アナリストを代えて用例作成を行い，その用例から得られたパターン対を比較した。その結果，各アナリストの作成した用例から得られたパターン対はほぼ一致することが分かった*4。従って，前章で取り上げた個々の動詞のパターン対の数は，ほぼ，それぞれの動詞に必要なパターン対の数とみなせる。この結果に基づき，和語動詞に対して日英機械翻訳でどれだけの数のパターン対が必要とされるかを予測する。

まず，第１の方法で得られたパターン対の数を図５の実線で示す。次に，前節で取り上げた動詞に対して第２，第３の方法で得られたパターン対の数をプロットし，それらの点をなめらかに結べば，それぞれ波線，一点鎖線の結果が得られる。

この図から，和語動詞に対して必要となるパターン

---

*1 後に述べるように，経験によれば，作業開始当初は「出る」「上がる」「掛ける」など，語彙数の多い動詞の場合，１動詞の用例を書き出すのに１人日程度かかった。しかし，慣れてくるにつれて速くなったこと，通常の和語系動詞はそんなに語義がないことにより，１日平均で３動詞前後の用例が抽出でるようになった。

*2 見出し語は仮名表記のめ，漢字仮名混じりの表記では語数は増大する。実際に収集した漢字仮名混じりでの語数は約1,100語である。

*3 人件費で見ると，和文用例作成のコストとその英語への翻訳コストはほぼ同じである。

*4 意味属性の指定等では揺らぎがあり必ずしも一致しないが，必要なパターンの種類ではほぼ一致する。例えば，20～30パターンを持つ動詞の場合，２人のアナリストが独立に作成した用例から得られたパターン対のうち，一方が作成したが他方が作成しなかったパターン対は１～２件程度（再現率90%以上）である。パターン対数の少ない用言の場合の再現率はさらに高い。

対の数はおおよそ9,000件と推定される。

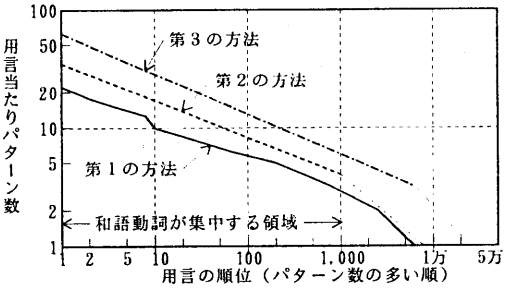

図5 日本語用言に対するパターン対の数の分布

(2)最終的な規模予測

日英機械翻訳システムにおいて，パターン対に整理することが適切と見られる述語としては，和語動詞のほかに漢語動詞，形容詞系の述語等がある。また，前章までは一般パターンについて述べたが，用言性の慣用表現もパターン対とすることが適切と考えられる。

これらのうち，形容詞系の述語は和語動詞と同様の性質を持つため，本論文と同様の方法が適切と考えられる。慣用パターンも同様である。漢語動詞は，通常，1単語当たりのパターン対数はほぼ1～2件であるため，用例が得られれば比較的容易に収集可能である*1。

そこで，これらの語を含む用言全体として必要と見られるパターン対の数を推定すると表3を得る。表3では，推定されたパターン対の数に対して，本論文の方法でどれだけ収集できる見込みかについても示す。

この表から，日英機械翻訳では，一般パターン，慣用パターンを含め，約25,000のパターン対が必要と推定される。

4.3 パターン対収集方法に関する考察

(1)対訳用例収集の問題

出現頻度の高いパターンの場合は比較的少量のコーパスからその用例が得られるが，出現頻度の小さいパターンの場合は用例はなかなか得られない。しかし，使用頻度は小さくてもそのようなパターンの種類は多いため，合計した出現頻度は無視できない。

本論文では，このような現状を打開し，網羅的なパターン対が作成できるようにするため，和英辞書を使用するパターン作成の方法など，人手処理を中心とする3つの方法を提案し，その可能性を試行実験によって調べた。その実験結果をもとに，用例収集の問題を整理すると以下のようになる。

(i) 1語当たりのパターン数が多い用言（和語系の動詞約1,000語および「い」型形容詞約200語）の場合は，和英辞書に基づく方法では十分なパターンが収集できない。また，日本語動詞の語義分析の結果に基づく方法でも，日本語動詞の語義分類が必ずしも英語動詞と対応関係にないため，十分とは言えない。むしろ，これらの情報を参照しつつ，両言語に素養のある人によって，自分の知識を内省し，日英でニュアンスの異なるとみられる用法を用例として書き出していく方法が適当と言える*2。

(ii) これに対して，1語当たりのパターン数が少ない用言（漢語系の動詞約7,000語および「な」型形容詞約2,000語）の場合は，和英辞書にエントリがあれば，必要なパターン対はほぼ収集されていると思われる。和英辞書にエントリがなければ用例を集める必要が生じるが，このタイプの用言は訳し分けの必要性があまりないから，パターン対辞書化しないで用例翻訳に期待することができる。

(2)日本語側のパターン記述範囲を決定する問題

用言毎に，必要なパターン対の種類を網羅するような対訳用例が収集されたとき，次に問題となるのは，(i)どの用例からパターン対を作成するか（複数の用例が1つのパターンに対応する場合の判断），(ii)どの格要素を日本語側のパターンに規定するか，それに伴って英語側で規定される要素は何か，また，(iii)日本語側の格要素の意味属性をどの範囲まで汎用化できるか，の問題である。

これらの問題について，本実験を通して得られた知見をまとめると以下の通りである。

(i)用例とパターン対の対応関係の問題

複数用例が1つのパターン対に対応するような場

表3 日英機械翻訳で必要なパタン対の数とその収集に必要な用例の数の見積もり（＊ 漢字表記による異なり数）

パターン対の種類 ＼ 比較項目		必要量の見込み		第1の方法（結果）		第2の方法（追加見込み一部済）			第3の方法（追加見込み一部済）		
		用言数	パターン対の数	用言数	パターン対の数	*対象用言数	用例数	パターン対見込	対象用言数	見込み用例数	パターン対見込
一般パターン対	和語動詞	1,500	9,000	1,500	4,000	1,000	5,200	1,500	1,200	15,000	3,500
	サ変動詞	6,500	8,000	3,000	4,000	(50)	(141)	(9)	4,000	8,000	4,000
	形容詞系	2,000	3,000	1,100	2,000	200	2,400	500	500	2,000	500
	小計	10,000	20,000	5,600	10,000	1,200	7,600	2,000	5,700	25,000	8,000
慣用パターン対		---	5,000	---	3,000	---	---	---	---	不明	不明
合計		10,000	25,000	5,600	13,000	1,200	7,600	2,000	5,700	25,000	8,000

*1 また，パターン対間の相互作用の心配は少ないため，パターン対当たりの作成効率は和語系の用言の場合に比べてはるかに良い。
*2 実験によれば，1日で3～4種類の用言に対して用例が抽出できるから，用例抽出の工数は和語系の約1,100種類の用言全体で1人年あまりと見積もられ，組織的な取り組みを行えば十分実行可能な工数と考えられる。

合，通常，英語訳の動詞と前置詞が同一であるような対訳用例（複数）が1つのパターン対に対応すると考えて良いが，日本文に対する正しい英語訳が常に1つとは限らないため，用例とパターン対の関係を人手によって見分けるのは困難な場合がある。その場合はとりあえず用例を含む狭い範囲のパターンを作成しておき，着目する用言のパターン対が一通り作成された段階で，共通的なパターンを縮退させることにより，作業効率を向上させることができる。

(ii)規定する格要素の選択の問題

用例を見れば，日本語側のパターンとして，最低限規定の必要な格要素の種類はおおよそ判断できる。また，英語生成に立場から見れば，前置詞の選択などの情報を付与するため日本語側での規定が望まれる要素のある場合があるが，これも人手で比較的容易に判定できる。

(iii)格要素の意味属性選択の問題

1つの用例から1つのパターンを作成するとき，格要素の名詞の意味属性を決定するには，名詞の意味辞書と意味属性体系リストの利用が大変効率的である。まず，名詞の意味辞書から用例に使用された名詞の意味属性を調べ，その意味属性の上位の意味属性を意味属性体系リストから調べて，得られた意味属性名を順に例文の名詞要素と置き換えてみる。このようにすれば，指定できる意味属性の上限が比較的容易に判断できる。

(3)今後のパターン対作成で期待されること

人手でパターン対を作成する場合，その能率は支援環境によって大きく影響される。汎用のエディタを使用してパターン対を作成する場合は，アナリストが知的判断に使用する時間よりも，むしろ，判断に必要な情報を単語意味辞書や意味属性体系リストから検索する作業や日本語・英語のパターンを記述する作業，さらには翻訳実験による作成パターン対の適切性のチェックに要する時間が圧倒的に多くなる。しかし，このように問題となる作業の多くは機械化が可能である。

本研究では2．2節で示したような支援環境を使用したが，この支援環境を使用するようになって以降，与えられた用例からパターン対を作成する速度は支援環境のない場合に比べ5～6倍に向上した[白井95]*1。今後，格要素となる名詞の意味属性指定法の改良，共通化可能な複数のパターン対の発見とそれを統合したパターン対候補の半自動作成，翻訳実験結果から不適切なパターン対を半自動的に指摘できる仕組みなどの実現により，パターン対の作成効率は支援環境のない場合に比べ10倍以上の効率化を目指している。

すでに述べたように，自動学習の方法では1パターン当たり多数の単純化された対訳用例を必要とすることなどのため，網羅的なパターン収集の作業に応用するのは困難である。そのため，本論文では，結合価パターン対の作成において当面する問題の解決を急ぐ立場から，人手をベースとする方法について議論した。しかし，一部の出現頻度の高いパターン対や分野に依存した専門パターン対の作成では自動学習技術の適用が期待される。特に，専門パターン対の場合は，あらかじめ汎用システムで準備しておくことは困難なこと，分野を決めれば比較的多くの用例が収集できることなどにより，学習技術による自動化が期待される。

また，学習技術の応用としては，人手で作成したパターン対の無矛盾性の検証の支援が考えられる。1動詞当たりのパターン数が増加するにつれて，その相互矛盾を人手で発見するのは困難となる。また，用言全体に対するパターン対記述の整合性の観点から，パターン対の記述に使用した意味属性体系の適切性を検証するのも容易でない。このような問題に対して，学習技術を応用した整合性チェックと問題指摘の自動化技術の実現が望まれる。

5 おわりに

日英機械翻訳において，用言（動詞，形容詞）の意味を訳し分けるのに必要な結合価パターン対の数とそれを収集する手段について検討した。

具体的には，単語当たりの語義が多いためパターン対作成が最も困難な和語動詞の場合を取り上げ，(1)和英辞書から収集する方法，(2)日本語動詞の語義対応の用例を使用する方法，(3)それらを参考に，人の知識に基づいて用例を作成して使用する方法の3種のパターン対の収集方法を比較した。その結果，主要な約1,000の和語動詞を意味によって訳し分けるには7,500件の結合価パターンが必要であることが分かった。これに対して，従来の和英辞書から収集できるパターン対の数は約1/4，和英辞書と日本語辞書の語義分類知識を使用する場合は約1/2であること，必要なパターン対を網羅的に収集するには，作業工数の面でも，和英辞書と日本語辞書の語義を参考に人の知識を内省して用例を作成する方法が適していることなどが分かった。

また，上記の結果から推定すると，漢語動詞，形容詞系の述語，用言性慣用表現などを含むパターン対全体では約25,000パターンが必要なこと，それらのパターンも辞書等を参考に人の知識を内省する方法で抽出された用例から比較的容易に収集できる見込みであることが分かった。

なお，現在，第1の方法で得られたパターン対を拡充するため，第2，第3の方法を並行して実施中であり，和語動詞，漢語動詞，形容詞系述語に対してそれぞれ，5,500件，4,000件，2,000件（合計11,500件）のパターン対を収集済みである。また，慣用表現では約3,000のパターンが収集されている。今後は，残されたパターン対（一般パターン約8,500件，慣用パタ

*1 パターン対の作成を開始した当初は紙ベースでパターン対を作成し一括入力していたが，そのときの最大の問題は，対訳用例の収集よりも，用例に関連する情報を調べること，それらからパターン対を決めて書き出すことの生産性の悪さにあった。しかし，支援環境の構築の後は，問題はむしろパターン対作成に使用する対訳用例をいかに収集するかの点に移った。そのため，いくつかの試行錯誤を行ったが，最終的には本論文の第3の方法によって用例を収集し，現在，パターン対作成作業を進行させている。

ーン約2,000件）を整備していく予定である。

謝辞　本論文をまとめるに当たり，パターン対の作成を担当してくださった井田倉紀子氏，小出ひとみ氏を始めとするＮＴＴアドバンステクノロジ(株)の関係各位，ならびに，用例収集にご協力くださった相澤弘氏を始めとするスバルインターナショナル(株)の関係各位に感謝する。

参考文献

[アルムアリA94a] Almuallim,H., Akiba,Y., Yamazaki,T., Yokoo,A. & Kaneda,S.: A tool for the acquisition of Japanese to English machine translation rules using inductive learning techniques, CAIA94, pp.194-201, San Antonio, Texas.

[アルムアリA94b] Almuallim,H., Akiba,Y., Yamazaki,T. & Kaneda,S.: Two methods for learning ALT-J/E translation rules from examples and a semantic hierarchy, COLING'94, pp.57-63, Kyoto.

[池原89] Ikehara,S.: Multi-level machine translation system, Future Computer System, Vol.2 No.3 pp.261-274.

[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, 情処学論 Vol. 34 No. 8 pp. 1692-1704.

[池原95] 池原, 白井, 横尾, ボンド, 小見: 日英機械翻訳における利用者辞書の意味属性の自動推定, 自然言語処理論文誌, Vol. 2 No. 1 pp. 3-17.

[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.

[金田94] 金田, 秋葉, 石井, アルムアリム: 事例に基づく英語動詞選択ルールの修正型学習方式,「自然言語処理における学習」シンポジウム論文集, pp. 158-165.

[研究社84] 小島, 竹林: ライトハウス和英辞典, 第1版, 研究社.

[黒橋92] 黒橋, 長尾: 格フレーム選択における意味マーカと例文の有効性について, 情処研報 NL-91-11.

[白井94] 白井, 横尾, 池原, 井上: 日英翻訳用構文意味辞書の記述精度の向上と作成支援, 第48情処全大 6Q-9.

[白井95] 白井, 兵藤, 上田, 横尾, 池原: 日英機械翻訳用構文意味辞書の作成支援, H7年電気関係学会関西支部連合大会 G14-3.

[横尾94] 横尾, 中岩, 白井, 池原: 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, 第48回情処全大 6Q-8.

付表１　最終的に得られたパターン対(「上がる」の場合)

パターン番号	方法	パターン対（意味属性などの条件やパターンの構造は省略）	
		日本語パターン	英語パターン
P01	①	AがBからCに上がる	A rise from B to C
P02	①	Aが上がる	A go up
P03	③	AがBをCに上がる	A go up B to C
P04	①	AはBが上がる	A produce good B
P05	①	AがBで上がる	A get nervous at B
P06	①	AがB[数量]CからDに上がる	A be raised by B from C to D
P07	①	AがBに上がる	A appear as B
P08	①	Aが上がる	A be dead
P09	②	AがBに上がる	A splash over B
P10	②	AがBに上がる	A appear on B
P11	③	Aが上がる	A be raised
P12	②	Aが上がる	A stop
P13	②	AがBからCに上がる	A improve from B to C
P14	②	AがBを上がる	A would like some B
P15	②	AがBに上がる	A enter B
P16	②	Aが上がる	A arise
P17	②	AがBで上がる	A be completed in B
P18	②	AはBからCにD[地位]が上がる	A be promoted from B to C
P19	②	AがBで上がる	B be enough for A
P20	②	AがBを上がる	A fly into B
P21	②	AがBに上がる	A be landed on B
P22	②	Aが上がる	A be produced
P23	②	Aが上がる	A be arrested
P24	③	Aが上がる	A be sluggish
P25	②	AがBで上がる	A die as a result of B
P26	③	Aが上がる	A increase
P27	③	AはB[見かけ]が上がる	A improve in A's look
P28	③	AはB[気持]が上がる	A be in high spirits
P29	③	AがBから上がる	A be collected from B
P30	③	AがBに上がる	A go to B
P31	③	AがB[時代]をCに上がる	A go back to C
P32	③	Aが上がる	A end
P33	③	Aが上がる	A rise
P34	③	AはBが上がる	A go out of A's B
P35	②	AがBから上がる	A come out of B
P36	③	AがBに上がる	A be on B
P37	③	Aが上がる	A be found

(注) パターン番号はパターン選択をする際の優先順位を示す。

付表２　第２の方法によるパターン対作成(IPAL用例「あがる」)

No	IPAL語義	日本文用例	英語訳	既存	新規
1	生物が上方へ動く。	一行は階段を一階から五階に上がった。	The party went up the stairs from the 1st floor to the 5th floor.	P02	
		彼は坂道を一気に上がる。	He climbs slopes without stopping.	P02	
2	物が高く移動する。	水銀柱が三十度に上がった。	The mercury in the thermometer rose to 30 degrees.	P01	
		花火が夜空を空中高く上がっていく。	Fireworks are flying high into the sky.		P20
3	ある物の程度が高くなる。	会社は生産が上がった。	The company increased production.		P13
		勉強能率が上がった。	Study efficiency has improved.		P13
4	今までよりも上の階級になる。	国鉄は初乗り運賃を120円から140円に上がった。	JNR raised its basic fares from 120 yen to 140 yen.	P06	
		アパートの家賃が1万円上がった。	The apartment rent increased by 10,000 yen a month.	P06	
5	今までよりも上の階級になる。	彼は係長から課長へ地位が上がった。	He has been promoted to section chief from chief clerk.		P18
		娘の算数の成績が4から5に上がった。	My daughter's mathematics mark improved from four to five.		P13
6	(上の)学校に通うようになる。	娘は今年小学校に上がる。	My daughter will enter elementary school this year.		P15
7	ある目的のために決まった場所に入る。	落語家が高座に上がる。	The storyteller appears on the stage.		P10
		友人は歌手として舞台に上がった。	My friend appeared on the stage as a singer.		P10
8	水中から出る。	子供が風呂から上がった。	The child stepped out of the tub.		P35
		海亀が海から陸に上がる。	Sea turtles land on shore.		P21
9	ある場所で発生する。	会場に歓声が上がった。	A shout of joy arose in the hall.		P16
		辺りに水しぶきが上がった。	A spray of water splashed around.		P09
10	好ましい結果が得られる。	こうすれば利益が上がる。	If you do so, a profit will be obtained.	P04	
11	候補として名前が出る。	彼は次期社長の候補に名前が上がる。	He is running as a candidate for president.	P07	
		Aチームの名前が代表候補に上がった。	Team A was nominated as a representative team.	P07	
12	今までわからなかったものが明らかになる。	証拠が挙がった。(注1)	Evidence was deduced.		P22
		犯人が挙がる。(注1)	The suspect was arrested.		P23
13	物が完成・完了する。	夕立が上がった。	The shower has stopped.		P12
		原稿が上がった。	The manuscript has been prepared.		P17
14	ある額である。	会費が4000円で上がった。	Four thousand yen was enough for the membership fee.		P19
		設置に二時間上がる。	It takes two hours to install.		P17
15	活動が停止状態になる。	赤潮で魚が上がった。	Fish died as a result of red tide.		P25
		バッテリーが上がる。	The battery is dead.		P08
16	興奮する。	私は入試で上がってしまった。	I got nervous at the entrance examination.	P06	
17	「食べる，飲む，買う」の婉曲表現。	ビールを上がりませんか。	Would you like some beer?		P14
18	「行く，訪ねる」の謙譲表現。	私がお届けに上がります。	I will deliver it.	保留(注2)	

(注1)「挙がる」は「上がる」と表記されることが多いため，表記の揺れとして扱う。
(注2)日本語パターンの二格の条件記述が困難であるため，登録を保留しているもの。

付表３　第３の方法によるパターン対作成(「あがる」の全90用例の一部)

No	日本文用例	英語訳	既存	新規
1	時代を上がる。	Move back in time.		P31
2	梅雨が上がった。	The rainy season ended.		P32
3	物価が上がった。	Prices went up.	P02	
4	歓声が上がった。	(The crowd) cheered.		慣習P
5	遺体が上がった。	The body was found.		P37
6	悲鳴が上がった。	(The girl) screamed.		慣習P
7	犯人が上がった。	The criminal was found.		P37
8	娘が屋敷に上がる。	The girl goes up to the mansion.	P02	
9	7時に幕が上がる。	The curtain rises at 7:00.		P33
10	年貢は領地から上がる。	Land taxes are procured from the territories.	P22	
11	ダムの水位が上がった。	The water level of the dam rose.		P33
12	列車のスピードが上がった。	The train's speed increased.		P26