日英機械翻訳における意味解析のための構文辞書

白井 諭[*1] 横尾昭男[*1] 井上浩子[*2] 中岩浩巳[*1] 池原 悟[*3] 八木晶子[*4]

[*1]NTTコミュニケーション科学研究所 [*2]NTTアドバンステクノロジ [*3]鳥取大学 工学部 [*4]フリー

1 はじめに

機械翻訳における意味解析では，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，人用の辞書や対訳用例文からパターン対の収集を進めてきた。本稿では，収集の経過と現在の到達点を示すとともに，今後の展望について述べる。

2 構文の収集の背景

機械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。

そこで，石綿らの方法を継承し，単語辞書[横尾97]に収録された用言性の1万数千語の規模で文型を体系化することを目指した。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。以上から，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し，体言の意味特徴に一般名詞の意味属性2,800分類を適用した日本語パターンとそれに対応する英語パターンを対にして持つ。

また，作業に当たっては，言語知識の範囲をファクトベースで記述することにした。具体的には，単文の肯定形を記述し，体言は一般名詞の意味属性により英語文型の選択が可能な段階まで抽象化した。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，体言と用言の結びつきが強く個々の単語から全体の意味が導き出せないものは，慣用表現パターンとして体言を個別的に指定する。これに対して，体言を個別指定しないものは一般表現パターンと呼ぶ。

3 パターン対の収集方法

パターン対の収集の経過について概要を述べる。

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や用例文などが記載されている。

日英翻訳の観点で和英辞書と英和辞書を比べると，英和辞書では英語の表現の意味を日本語で示すため，現実の日本語の文書では使用されない説明的な表現がしばしば使用されているのに対して，和英辞書では日本語の表現に対応する英語の表現が記述されている。

第1ステップとしては，和英辞書の見出し語を基準に，主として対訳例文から日英の基本構造をパターン対として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」は5つの語義に分類され，第2語義に次の例文が示されている。

彼の学校の成績が*上がった*。
His school record *has improved*.

この対訳文からは，例えば次の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，汎用化などの見直しにより，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。また，これらを用いた翻訳実験により，語義数の多い用言のパターン不足が目立ち，構文辞書はこの倍程度の規模に充実させる必要があることがわかった[白井95a]。

3.2 日本語辞書の用例文とその英訳文からの収集

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例文を収集分析する研究が進められてきた。こうしてまとめられたIPAL動詞辞書[IPA87]の用例文に対し，日本語原文に忠実で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パターン対を収集した。この用例文は和語動詞861語（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）に対する5,243文（日本語7.5万字，英語4万語）で，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。語義とパターン対が1:1に対応するケースは4割程度にとどまっていることから，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。

3.3 内省による用例文とその英訳文からの収集

用例文を収集するため，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を試みた。このようにして表現のバリエーションを網羅的に収集し，その対訳データからパターン対を収集することにした。

和語動詞に対してこの対訳用例文の作成作業を試みたところ，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]。このうちいくつかの動詞に対してパターン対の収集を試行したところ，和語動詞については2倍程度の規模に拡張できそうな感触を得た。この作業は現在継続中である。

4 パターン対の記述内容

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も機械用の辞書には丹念に記述しておく必要がある。本節では日本語および英語の結合価パターンをどう記述するかについて，いくつかの観点から整理する[林87,奥87]。

4.1 一般表現パターン

述語を中心とする表現には，動詞による動作性の表現（何が何をどうする，何がどうなる）と形容詞（形容動詞を含む；区別が必要ならイ型，ナ型と呼ぶ）による状態性の表現（何がどんなだ）のほか，述語に名詞を用いた断定文（何が何だ）がある。

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，一般表現パターンとしては，まず動詞構文や形容詞構文のように訳し分けの必要があるものがパターン対の収集の中心として考えられる。

これに対して，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。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は収集対象に加えた。

なお，動詞構文には，「述べる」や「命じる」のように，文相当の内容を格要素として必要とする表現がある。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，これらは一般パターンとして収集対象に加えた。

一方，形容詞構文には，「象は鼻が長い」のような二重主格構文や「AがBよりCだ」などの比較構文がある。これらは一定の変形処理により，一般表現パターンに還元することが可能であるので，原則として収集しない。

「研究開発する」といった複合動詞表現は2つの一般表現パターンに分解できるので，別々に登録した。ただし，別登録すると合成が困難になる場合はこの限りではない。

4.2 慣用表現パターン

一般表現パターン以外の構文としては慣用表現パターンがある。慣用表現を厳密に定義するのは難しく，「単語の二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。

例えば，「油を売る」では，「木陰で油を売る」は慣用表現といえるが，「安い油を売る」は慣用表現ではない（一般表現である）と判定しなければならない。しかし，「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。当面は慣用表現判定の手がかりとなりうる要素（「油を売る」では「どこで」）をパターンに記述することにより対処することにした。

日英翻訳の観点では，英単語が日本語の連語表現に対応するものは，慣用表現に準じて扱う方が良い。例えば，「背が高い」を文字通り，「背」→ “back”，「高い」→ “be 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。

このほか，慣用的といえる表現としては機能動詞表現[村木85]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）」といった一般表現パターンに還元することにした。

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。

パターン対では，このうち主として「格要素＋述語」に対し，英語との対比により記述するか否かを決定した。すなわち，英語の主語，目的語，補語や，英語の表現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対象とした。また，副詞要素や副詞的に働く時間表現，数量表現も，英語の表現で特徴的であれば，必要に応じてパターン対に記述した。

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，原則として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，意味体系で述べた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。しかし，多用される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あるが，特殊なパターンなどではそれが難しい。そこで，名詞句に対する制約条件として，中心名詞の字面そのものの指定や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。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。

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ことにより対処する。ただし，パターンに特徴的な助詞表現であるか，格助詞への還元が困難な場合にだけ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。

4.5 パターン対への付加要素

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，結合価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，原則として処理系で対処する。

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であるため，慣用表現パターンについては，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。このような付加要素は必要に応じて収集した。

4.6 英語パターンの記述

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。日本語パターンでは，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，英語パターンには，どのような文法機能を持つかを併せて記述した。

例えば，「XがYをZに提案する」→ “X propose Y to Z” は，具体的には次のように記述されている。

（日本語パターン）
┌X〔主体〕が
├Y〔人間活動〕を
├Z〔主体〕に
└提案する

（英語パターン）

┌SUBJ	NP	X
├VP	VT	propose
├OBJ	NP	Y
└PP	PREP	to
	NP	Z

最初は，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。この形式では，複雑なパターンでは全体の見通しが悪くなる。そこで，現在は，英語パターンの骨格部分と肉付部分を分けて記述する形式に改めている[横尾94]。

（骨格部分）

U_SENT_1	PRED_1 (CASE_1 CASE_2 CASE_3)	
PRED_1	VERB_1	
CASE_1	S	
CASE_2	DO	
CASE_3	PP	U_PP_1
U_PP_1	PREP_1	

（肉付部分）

VERB_1	spelling	“propose”
CASE_1	instance	X
CASE_2	instance	Y
U_PP_1	object	Z
PREP_1	spelling	“to”

この骨格構造と英語パターンの関係を見ると，英語パターンの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パターンが表現できる。

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。制約条件の汎用化などの見直し済みのものを集計すると表1のようになる。

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。

表 1 構文体系に収集したパターン対

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	1,308	4,112	493	2,510
複合和語動詞	506	874	30	51
サ変動詞	3,042	4,307	18	21
形容詞(イ型)	260	565	50	559
形容詞(ナ型)	874	1,313	1	3
名　詞	24	23	0	0
合　計	6,014	11,194	592	3,144

パターン対の収集可能な量を見積もったところと，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である。また，英和辞書は表現が説明的であるなどの理由で，パターン対の収集対象とはしてこなかったが，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97]ので，今後はその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，現状の到達点をまとめた。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予定である。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，構文辞書の完成を目指したい。

謝辞　本稿をまとめるにあたり，ご協力くださったNTT情報通信研究所の林 良彦氏に感謝する。林氏は構文辞書の構築開始時期の担当者である。

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, p.39-44 (1987)
[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)
[宮地 82] 宮地 裕 編: 慣用句の意味と用法，明治書院 (1982)
[水谷 83] 水谷,石綿,荻野,賀来,草薙,青山: 文法と意味I，朝倉日本語新講座 3，朝倉書店 (1983)
[村木 85] 村木 新次郎: 日本語の機能動詞表現をめぐって，国立国研究所報告 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，ジャパンタイムズ (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 95a] 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，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, *NLPRS 95*, pp. 443-448 (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)
[白井 97] 白井,大山,我妻,石崎: 英単語に対する連語的日本語表現の分析，言語処理学会第 3 回年次大会, B1-2 (1997)
[横尾 94] 横尾,中岩,白井,池原: 日英機械翻訳用スケルトンー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, 6Q-8, pp.3－139-140 (1994)
[横尾 97] 横尾,宮崎,阿部,池原,白井,細井: 日英機械翻訳における意味解析のための単語辞書，言語処理学会第 3 回年次大会, A2-2 (1997)