

構文意味辞書と構文体系

白井 諭*¹ 横尾 昭男*² 中岩 浩巳*¹ 渡邊いづみ*³
高橋 直美*³ 関 嘉代*³ 池原 悟*⁴ 宮崎 正弘*⁵

*¹NTT コミュニケーション科学研究所 *²ATR 音声翻訳通信研究所 *³NTT アドバンステクノロジー
*⁴鳥取大学 工学部 *⁵新潟大学 工学部

1 はじめに

日英翻訳の意味解析と日英変換の精度向上を目指す、体言と用言の意味的共起を結合価パターンとして記述し、日英で対にしたパターン対の収集を進めている[池原 93]。構文意味辞書は一般と慣用のパターン対を束ねたもので、体言への制約条件が一般名詞意味属性(2,700 分類)を用いて記され、用言意味属性(107 分類)が付与されている。先に上梓した構文体系は、この構文意味辞書の日本語パターン、英語パターン、体言への制約条件、用言意味属性を、人間が利用することを意識して編集し直したものである。本稿では、構文意味辞書から構文体系への編集の概要を報告する。次に、一連の編集作業を通してわかった構文意味辞書の問題点をまとめ、今後の方針を述べる。

2 ALT-J/E におけるシステム辞書の構成

ALT-J/E のシステム辞書は図 1 のように構成されている。

意味属性体系は、対象の見方や捉え方、すなわち、対象を概念化する際の視点を整理したものであり、話者から見れば単語の用法に相当する分類となっている。具体的には、分類観点として、上位-下位(is-a)関係のほかに、全体-部分(has-a)関係にも着目し、階層的な木構造にまとめたものである[池原 97, 宮崎 97, 中岩 97]。一般名詞意味属性体系は、固有名詞を除くすべての名詞の体系的な分類を行なうためのものである。固有名詞意味属性は、人名、地名等の詳細かつ体系的な分類を行なうためのものであり、その上位の属性は一般名詞意味属性と対応づけることが可能である。用言意味属性は、用言だけでなく格要素を含めた文型の体系的な分類を行なうためのものである。そして、日英翻訳における訳し分けや文型記述を通して妥当性が検証されてきた。

以下では、構文意味辞書について概要を述べる。構文意味辞書では、格要素の条件記述に一般名詞意味属性体系が使用され、文型パターンに対して用言意味属性体系が使用される。

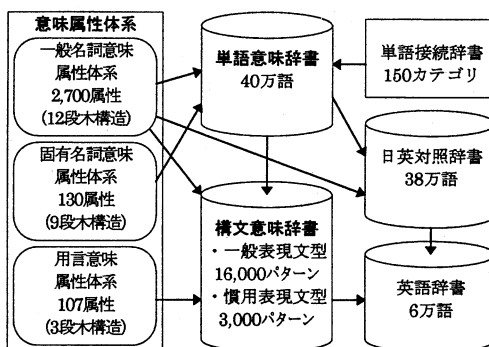


図 1 ALT-J/E におけるシステム辞書の構成

3 構文意味辞書の概要[池原 97, 白井 97]

3.1 結合価パターン対の収集経過

(1)和英辞書からの収集

初期は、和英辞書の用言見出しを対象に、語釈の対訳例文から日英の基本構造を結合価パターン対として記述することにより、構文意味辞書の構築を開始した[林 87, 奥 87]。具体的には、中辞典クラスの和英辞書数冊と、慣用表現辞書を参照し、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 3,000 件を収集した。これらを用いた翻訳実験では、語義数の多い用言のパターン不足が目立ち、構文辞書はこの約 2 倍の規模に充実させる必要があることがわかった[白井 95]。

(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の充実を目標として、和語動詞を詳細に記述した IPAL 動詞辞書[IPA87]の用例文 (861 語に対する 5,243 文) に着目し、この用例文に忠実かつ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。その結果、一般表現パターン 1,532 件と 500 件の修正情報が得られた[白井 96]。

(3)内省による用例文とその英訳文からの収集

多数の用例文を収集するため、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し、その対訳データからパターン対の収集を試みた。IPAL 動詞辞書と同じ 861 動詞に対し、10,500 文が収集された[池原 96]。このうち 30 動詞につい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。

3.2 結合価パターン対の記述内容

(1)一般表現パターン

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も英語として通用しない。一般表現パターンとしては、まず動詞や形容詞の単文表現を主な収集対象としてきた。「述べる」や「命じる」など、文相当の内容を格要素として要求する動詞構文は、複合文ではあるが一般パターンとして収集した。

名詞構文「XはYだ」は、英語でも「X be Y」の形でXとYを名詞のまま訳せばよい場合が多い。しかし、「今日は天気だ」→「It is fine today.」では英語表現には形容詞が使われる。これらもパターン対の収集対象に加えた。

(2)慣用表現パターン

構文意味辞書としては、「油を売る」のように、いわゆる慣用表現のほか、日英翻訳の観点から慣用表現に準じて扱う方が都合がよいもの（「背が高い」では全体が「be tall」に対応する）も収集した。

(3)記述する修飾要素の範囲

「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。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が主な記述対象となる。副詞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)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。

名詞句に対する制約条件は、名詞句の中心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記述した。条件指定を

柔軟にするため、中心名詞の字面指定や名詞句の構成要素の個別指定なども許容した。

助詞表現に対する制約条件は、標準的に使用される格助詞 1 語により指定した。助詞表現のバリエーションは日本語解析で標準化するが、それが困難な場合には複数個を指定した。

(5)パターン対への付加要素

述語に対する使役、受身、可能などの語尾表現は原則として処理系で対処する。ただし、「花を持させる」は「花を／持つ＋使役（せる）」と解析されるが、「せる」を伴って初めて慣用表現であるため記述対象とした。「回転させる」→「rotate」は「回転する＋使役（せる）」というよりも自動詞「回転する」の他動詞化であり、「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られる。これらは必要に応じて記述した。

(6)用言意味属性

文と文の関係は、文の骨組みである日本語パターンの関係で近似される。日本語パターンの中心要素は用言であり、用言の意味的用法を分類体系化した用言意味属性を用いることにより文間の意味の追跡が可能になる。用言意味属性は、まず状態と行動を分け、それぞれを数種類に細分類する。最終分類として、用言に強く結びつく格を表示する。

(7)英語パターンの記述

英語パターンには、主語や目的語、動詞などの文法的な機能を併せて記述した。英語パターンには多用される形式があり、英語パターンの 70%までが骨格構造 9 個、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[横尾 94]。

(8)現状の到達点

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 件であった。実験中のものを除いた件数は表 1 のようになっており、これらが最終的

表 1 結合価パターン対の収集情况

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	513	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

に構文体系の収録対象として選定された。

4 構文体系の編集

4.1 編集方針

構文意味辞書には、日本語と英語の結合価パターンが対にして記述されている。このため、日英機械翻訳の意味解析において日本語パターンが選択された時点で英語のパターンも決定されることになり、構文に関しては改めて変換の必要がないという特徴がある。構文意味辞書では、論理的には、例えば次のようなパターン対の構造が記述されている。

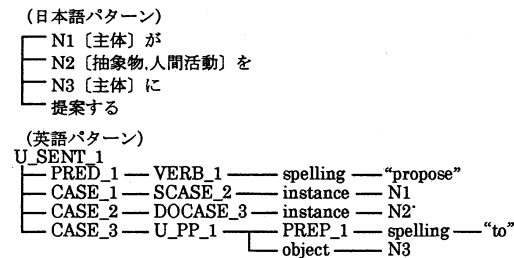


図2 パターン対の記述例

これに準じた構造を表示すれば結合価構造が直感的に理解できるが、1つのパターン対の記述に広いスペースを必要とするため一覧性が低下する。このタイプの辞書には他に類似するものがないため、どのような形式で編集すれば人間用の辞書として適切であるか、様々な編集形式を評価した。

試行錯誤の結果、1行に表示しても構造の再現は可能で、一覧性も良いことから、ベタで表示することとした。最終的に、構文体系は、用言を国語辞書配列し、パターン対ごとに優先順位に基づいた通番を付し、用言意味属性と英語動詞の条件(動作/状態, 受身の可否)を添えた。パターン対をベタで表示し、格要素の条件などは最後に表示した。例えば、上記の例は次のようになる。

提案する (ていあんする)

(1) 29 精神活動 (動作)

N1 が N2 を N3 に提案する N1 propose N2 to N3

[N1 (3 主体) N2 (1001 抽象物 1236 人間活動) N3 (3 主体)]

図3 構文体系におけるパターン対の記述例

また、用言以外の語から検索できるよう、慣用表現に使われる名詞の索引と、英語パターンに現れる英単語の索引を作成した。単語体系に倣って、用言意味属性別の単語一覧表を作成した。

なお、構文体系には上述のパターン対情報をすべ

て記載するが、名詞索引、英単語索引、用言意味属性別の単語一覧表についてはパターン対のベタ表示を記述するにとどめる。用言とパターン対の通番により、詳細情報は構文体系とリンクする。

4.2 記述内容の見直し

(1) パターン対の検証

日本語パターンと英語パターンの対応やそれぞれの記述内容の見直しを行なった。日本語パターンの格要素は一般名詞意味属性を用いて条件指定されているため、格要素に具体的名詞をイメージしながら妥当性をチェックした。この過程で、どのような表現を意図したパターン対か直感的に不明なものが散見され、作業進行の都合上、構文体系の収録対象からは除外した。構文意味辞書の構築初期のパターン対は格要素の条件指定を抽象化しすぎたものが多く、それが主な原因と思われる。

(2) 一般表現と慣用表現の区分の見直し

一般表現パターン対はすべての格要素が変数として扱われるのに対して、慣用表現パターン対では1つの格要素が特定の名詞の指定を伴って固定的に扱われ、それ以外の格要素は変数として扱われるという違いがある。しかし、最近是一般表現においても必要に応じて名詞の字面指定が可能であり、両者の区分は揺らいでいる。構文体系の編集に当たり、日本語パターンに字面指定が行なわれ、英語パターンと単純に対応しないものは慣用表現とした。逆に、慣用表現に区分されていても変数化が可能な場合は一般表現に移行した。

(3) 格要素の条件の見直し

格要素の一般名詞意味属性指定の見直しを行なった。特に、構文意味辞書の構築初期のパターン対は格要素の条件指定を抽象化しすぎたものが多く、逆に最近作成したばかりのパターン対は実験的検証が不足している場合には抽象化が不十分であるものがある。用例を思い浮かべた際、直感的に違和感を感じたものについては見直しを行なった。

(4) 格ラベルの単純化

格要素は、格助詞と連動して、N1(ガ), N2(ヲ), N3(間接目的のニ), N4(カラ), N5(N4と対応するニへ)などと変数化されている。従って、パターン対をベタで表示すると欠番が生じる。また、構文意味辞書の構築初期は深層格的な分類を試みたことがあり、一部については現在の基準での見直しが完了していない。次の(5)の問題と併せ、日本語パター

ンの出現順に N1, N2, ... と単純化した。

(5) 用言意味属性の一般化

用言意味属性体系において、最終段階の分類は用言と強く結びつく格要素を整理しているが、(4)で述べた格ラベルの不統一に伴う不整合がある。構文体系として、その影響を受けないレベルを検討した結果、上位の 36 分類のみを編集対象とした。

(6) 読みの付与

構文意味辞書は、意味解析において日本語の文型を決定し、対応する英語の構文を与えるのが目的であるため、読みの情報は必要ない。構文体系において、見出しとなる用言と、慣用表現に使われる名詞は、いずれも国語辞書に準じた配列とするため、これらの語に読みの付与を行なった。

4.3 編集手順

構文意味辞書には様々な情報が記されているのに対し、構文体系の編集に使用される情報は用言、読み、ベタの日本語パターン、格要素の意味条件、ベタの英語パターン、英語動詞の屈折変化情報である。また、前節で見直した情報は、必ずしも構文意味辞書の記述方針に合致していないため、単純に構文意味辞書に反映すると悪い副作用を生じる恐れが高い。そこで、構文意味辞書から構文体系の編集に使用する情報を取り出し、それに見直した情報の反映を行なった“中間辞書”を作成し、構文体系、慣用表現名詞索引、英語単語索引、用言意味属性別の単語表を並列的に作成した。

5 今後の課題

前節で述べたように、初期に作成したパターン対の徹底的な見直しが必要である。主な項目は、一般表現と慣用表現の区分、格ラベルの基準の統一と用言意味属性の適正化、格要素の一般名詞意味属性指定の適正化が挙げられる。中期以降はパターン対作成に使用した用例のメモを保存しており、これが見直しの際、有効に働いた。今後は、構文意味辞書に用例を収録するとともに、コーパスと連携を検討したい。辞書としては読みの情報も有効であるので、これも収録したい。

このほかには、格要素に指定する助詞を、代表的なものだけでなく、助詞相当語などを積極的に指定する必要がありそうである。また、必須的な格であるかどうかは日本語解析に大きく影響するので、この点からも見直しを進めたい。

英語のパターンでは、前置詞の選択条件を記述する必要があることが分かった。また、日本語と英語のパターンの時制が異なる場合がある。例えば、

N1 が身の破滅にと成る N1 be the cause of ruin
では、過去時制ならよいが、日本語が現在の場合、英語には would be (または will be) を使う必要がある。このような観点からの見直しも必要である。

パターン対の収集可能な量は、一般表現パターン対 20,000 件、慣用表現パターン対 5,000 件と予想されている[白井 95]。表 1 を見ると、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、今後はこれらの補充を進めたい。

6 おわりに

本稿では、日英機械翻訳用の構文意味辞書の概要と、それを編集して作成した構文体系について報告した。今後は、この編集を通して明らかになった構文意味辞書の問題点の解決や、辞書とコーパスとの連携について検討する予定である。

謝辞 日本語語彙大系の編集にご指導、ご討論頂いた岩波書店の宮内久男氏、上野真志氏、岡本潤氏に感謝する。

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, p.39-44 (1987)
- [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, 情報処理学会論文誌, Vol.34, No.8, pp.1692-1704 (1993)
- [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)
- [池原97] 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 (1997)
- [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)
- [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)
- [宮崎97] 宮崎, 池原, 横尾, 白井: 日英翻訳のための意味属性体系, 電子情報通信学会, 技術研究報告, NLC97-12, pp.29-36 (1997)
- [中岩97] 中岩, 池原: 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, 情報処理学会論文誌, Vol.38, No.2, pp.215-225 (1997)
- [白井 95] 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, NLP95, pp. 443-448 (1995)
- [白井 97] 白井, 横尾, 中岩, 池原, 宮崎: 日英機械翻訳のための構文辞書, 電子情報通信学会技術研究報告, NLC97-14, pp.45-52 (1997)
- [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, 情報処理学会第 48 回全国大会, 6Q-8, pp.3-139-140 (1994)