

# 入力文と結合価パターン対辞書の照合に関する一手法

C1-2

白井 諭<sup>\*2\*1</sup> Francis Bond<sup>\*1</sup> 野沢 弥生<sup>\*3</sup> 佐々木 富子<sup>\*3</sup> 上田洋美<sup>\*3</sup>

<sup>\*1</sup>NTTコミュニケーション科学基礎研究所 <sup>\*2</sup>ATR音声翻訳通信研究所 <sup>\*3</sup>NTTアドバンステクノロジー(株)

## 1 はじめに

機械翻訳における意味解析精度の向上には、体言と用言の意味的共起関係に着目した結合価パターンの利用が有効である。筆者らは、入力文と結合価パターンの照合による意味解析を提案し、日英翻訳を対象にパターン対の網羅的収集を進めている[白井99]。

入力文と結合価パターンの照合による意味解析を行なうALT-J/E[池原91]では、入力文から様相機能語や副助詞等を取り除いたパターンの基本形を推定し、それと辞書登録パターンを照合している[河合87]。多数の特殊構文への対応を進めるうち、基本形推定処理が複雑化し、照合精度の低下が問題となってきた。

本稿では、入力文を辞書に合わせるのではなく、逆に、辞書を入力文に合わせることを考える。具体的には、入力文の様相機能語や助詞に着目して特殊構文の可能性を検出し、辞書登録パターンから可能な展開形を複数派生させ、展開形と入力文を照合する方法を提案する。

## 2 従来の照合方法と問題点

### 2.1 従来の照合方法

入力文に対し、形態素解析、構文解析を行ない、述語を中心としたまとまり(単位文)に分割する。述語の中心用言に対する日本語パターンを結合価パターン対辞書からすべて取り出し、副助詞を格助詞に変更しながら、格要素の名詞の条件を最も満たす日本語パターンを照合結果とし、対応する英語パターンを用いて英語の基本構造を生成する。

述語に様相機能語が付加されている場合、単位文の構成単語の状況に応じて様相属性の同定を行なう。同定された様相属性に応じて、格助詞を変更することにより、日本語パターンの基本形を1つ推定する。このとき、副助詞を格助詞に変更する処理を行なう。推定した基本形が結合価パターン対辞書に登録されていれば、それを照合結果とする。なければ、別の基本形を

推定し、前記の処理を繰り返す。

以下に、従来の照合方法による基本形推定の処理例を示す。そして、次章ではこれらに対する新手法の適用を検討する。

#### (a)「れる/られる」

以下の順で「れる/られる」の様相が“自発/尊敬/受身/可能”のいずれであるか判定し表層格を変更する。

1. 「XがVられる」のVが自発動詞(思い出す、偲ぶ、等)なら自発と判定し「XをV+自発」に変形する。
2. 「Xが…Vられる」のXが尊敬対象語(王、先生、等)なら尊敬と判定し「Xが…V+尊敬」に変形する。
3. 「…Vられる」のVがサ変動詞または五段動詞なら受身と判定し、それ以外は受身・可能の両方の可能性を考える。受身に対する表層格変換として、日本語パターン「ガ/ヲ/ニ/カラ」の名詞条件に応じて、「に」によって「/から/に」を「が」に、「は/が」を「を/に/から」に変更する。可能に対する表層格変換として、日本語パターン「ヲ」の名詞条件を満たすなら、「は/が」を「を」に変更する。

#### (b)機能動詞変形

機能動詞とは、動詞自身は実質的な意味を持たず文法的な役割のみを担う語で[村木80]、「(を)行なう」「(を)手掛ける」「(を)集める」など、約30語が該当する。

例えば、「彼は機械翻訳の研究を行なう」は「彼は機械翻訳を研究する」と意味的には等価である。英訳も“He performs a research of machine translation.”より“He researches machine translation.”の方が望ましいと考えられる。

一般には、「XのY[サ変名詞]を行なう」を「XをYする」と変形し、「Yする」のパターン対(上例では「研究する」)を用いて翻訳する。

ただし、「注目を集める」は「注目する+受身(注目される)」に変形されるのに対し、「署名を集める」を同様に変形すると誤りとなる。機能動詞ごとに変形禁止語を

登録した一覧表を作成し、副作用を防止している。

### (c)二重主格文

「象は鼻が長い」のように、1単位文中にハ格とガ格が同時に存在する文をいう。ハをガ以外の格助詞(ノを除く)に変更できない場合、どのように翻訳するかが問題となる。

例えば、前記の文は「象の鼻が長い」とほぼ等価である[三上60]とみれば、“Trunks of elephants are long.”のように英訳できる。しかし、「鼻」が「象」の体の一部であることに着目すれば、“Elephants have long trunks.”とする方が適訳と考えられる。

そこで、ALT-J/Eでは、一般名詞意味属性体系[池原97]により、ハ格の名詞とガ格の名詞がhas-a関係にあるかどうかを調べ、have構文に訳すか、ofを使うかを決定している[奥98]。ofを使うのが適当でないと判定されれば、“As for elephants, trunks are long.”のように英訳されることもある。

なお、ALT-J/Eでは、形容詞に対する二重主格文のみを対象としており、上記の例ではガ格の名詞と形容詞に対する結合価パターン対が使用される。

## 2.2 従来法の問題点

様相機能語の解釈に多義がある場合、標準形を推定する順序に依存して、判定結果が変動する。また、格要素の省略に弱い。このため、「れる/られる」では、“尊敬/受身/可能”の判定をしばしば誤るほか、“被害の受身”が正しく判定される割合がかなり低い。

機能動詞変形は正しく処理される割合は高いが、「XがY[サ変名詞]を行なう」を「XがYする」と変形した時、「Yする」が他動詞である場合は目的語がないという問題が生じている。

二重主格文では、副助詞の揺らぎに弱い。例えば、「鼻は象が長い」は“As for noses, elephants are long.”となり、「象は鼻が長い」が“Elephants have long trunks.”と英訳されるに比べ相当見劣りがする。

このように、様相や特殊構文への対応が画一的である点が問題である。これは、あらゆる場合を想定しながら、様相の語義を判定したり、特殊構文から単位文を検出したりするための処理アルゴリズムを記述するのが容易ではないことに起因すると考えられる。

## 3 辞書登録パターンを展開する方法

### 3.1 基本的な考え方

単位文と日本語パターンの照合において、様相の付加や特殊構文のない単位文であれば、単位文の述語に対する結合価パターン対辞書の全エントリのうち最も適したものが選ばれるので、辞書への項目追加や記述改良が翻訳精度の向上に直結しやすい。

そこで、ある様相や特殊構文がどのような場合に使われるかという可能性に応じて結合価パターン対辞書に登録されている日本語パターン(標準形)から展開形を生成し、その展開形と入力文を照合する方法を提案する。具体的には、入力文の解析結果から様相機能語や助詞情報等を抽出し、様相や特殊構文の特徴の諸元を記述した一覧表により変形の可能性を網羅的に列挙し、それに応じて辞書登録パターンから展開形を生成した後、入力文に近いものを選択する。

この方法では、どのような意図で様相機能語や副助詞が使われているかをトップダウン的に設定できるため、入力文の意図に対する仮説が立てやすい。また、類似表現の系列を意識して表現を整理するのも容易になると考えられる。例えば、表1のような表現である。

表1 日本語表現と使役動詞を用いた英語表現

様相	日本語表現	英語表現
(基本形)	Xが 走る	X run
ヲ使役	Zが Xを 走らせる	Z make X run
ニ使役	Zが Xに 走らせる	Z let X run
受益	Zが Xに 走ってもらう	Z get X to run
被害	Zが Xに 走られる	Z have X run

### 3.2 文型変化を伴う様相への適用性

本節では、様相属性の判定が必要となる「れる/られる」について検討する。「れる/られる」の処理には多数の要因を考慮する必要があるため、他の様相機能語へ適用する上でモデルケースになると考えられる。

「れる/られる」には、“受身/可能/尊敬/自発”の属性のほか、“被害の受身”がある。自発は、特定の動詞にしか成立しない。また、英訳も基本形を変形しても適切なものが得られにくい。そこで、パターン対辞書に記述することにする。

従って、「れる/られる」に対しては、辞書に登録された基本形に対し、“受身/可能/尊敬/被害”に対する展開形を生成する。格助詞の変化パターンはテーブルで

与えることが可能であると考えている。表2に例を示す。ただし、受身変形では、ガ→カラが可能な動詞(先生が生徒を叱る→生徒が先生から叱られる)、二格に視点を置ける動詞(AがBに影響する→BがAに影響される)、ヲ格・二格の両方に視点を置ける動詞(AがBをCに紹介する→BがCにAによって/から紹介される、CがAに/によって/からBを紹介される)などがあり、さらなる検討が必要である。

表2 「られる」を伴う日本語表現とその英語表現

様相	日本語表現 / 英語表現 / 備考	
(基本形)	Xが Yを 食べる	
	X eat Y	(辞書登録形)
受身	Yが Xに/よって 食べられる	
	Y be eaten by X	基本形+受身
可能	Xに Yが 食べられる	
	X can eat Y	基本形+可能
尊敬	Xが Yを 食べられる	
	X think Y	基本形+尊敬
被害	Zが Xに Yを 食べられる	
	Z have Y eaten by X	基本形を変形する

被害の受身は「自動詞の受身」といわれるが、それは正確ではない。他動詞「食う」の例文、「猫が鼠を食う」に対し「私は猫に鼠を食われる」とすれば被害の意味となる。即ち、事象を構成する要素以外に視点を置くことにより成立する。英語には被害の受身の概念がないため、適切に訳すには文脈を考慮する必要がある。しかし、表3の表現は文脈にあまり影響されずに使用できる見込みである。表3の上から、それぞれ、人以外を動作主とする自動詞、人に関する非動作の自動詞、自分が被害者となる意思動詞、他動詞の表現である。使い分けとしてこの条件でよいかは今後の検討課題である。

表3 被害の受身の日本語表現とその英語表現

基本形	被害の受身の表現
雨が 降る	私は 雨に 降られる
It rains	It rains on me
父が 死ぬ	太郎は 父に 死なれる
(My) father dies	(His) father dies on Taro
大男が 前の席に 座る	私は 大男に 前の席に 座られる
A big man sits down on the seat in front of me	A big man goes and sits down on the seat in front of me
車が 足を 轢く	私は 車に 足を 轢かれる
A car run over my leg	I have my leg run over by a car

3.3 特殊構文への適用性

本節では、機能動詞表現と二重主格表現について検討する。

機能動詞表現に対する現処理の問題点は、目的語が欠落する場合があることである。従って、その恐れがある場合は機能動詞変形を行なわないようにすればよいのであるが、例外処理の導入は得策ではないと考えられる。そこで、本稿では、機能動詞処理を行なった場合の日本語パターンと、通常の動詞とみた場合の日本語パターンのうち、適切な方が選ばれるようにする枠組みを考える。

例えば、「選挙を行なう」に対し、「行なう」は機能動詞になりうるので、「選挙する」に対する結合価パターン対を検索し、日本語パターンを「選挙を行なう」に変形する。このとき、英語パターンは変形しないが、機能動詞変形適用を注記する。また、「行なう」に対する結合価パターン対を検索し、この両者のうち適切な表現を選択する。即ち、表4において、機能動詞ではXとYの2要素が欠落するのにに対し、一般動詞ではXの欠落だけですむので、一般動詞のパターンが選択される。

表4 「選挙を行なう」に対するパターン対検索結果

種別	日本語パターン	英語パターン
機能動詞 (元の形)	Xが Yの 選挙を 行なう (Xが Yを 選挙する)	X elect Y
一般動詞	Xが Yを 行なう	X conduct Y
	Xが Yを 行なう	X carry out Y

(注) 一般動詞のパターン対は上記以外に数件検索される。

「署名を集める」の場合も同様に表5となる。この場合も一般動詞の方が欠落が少ない。その前に、「署名する」の二格は受身の視点になれないため、本来はその点から機能動詞変形を施してはいけないことがわかる。辞書に記述されている情報のうち、どの情報をどの順序でチェックするのがよいかは今後の課題である。

表5 「署名を集める」に対するパターン対検索結果

種別	日本語パターン	英語パターン
機能動詞 (元の形)	Xが Yに 署名を 集める (Xが Yに 署名する)	X sign Y(+受身)
一般動詞	Xが Yを Zに/へ 集める	X gather Y in Z
	Xが Yを 集める	X collect Y

(注) 一般動詞のパターン対は上記以外に数件検索される。

次に、二重主格文に対する現処理の問題点は、副助詞の揺らぎに弱いことである。例えば、ハ格とガ格が has-a 関係(全体-部分の関係)にある表現「象は鼻が長い」についても、表6のような類似表現が考えられる。それぞれに対する英語表現として、どのようなものが適切であるかについては検討中であるが、定/不定や総称/指示の情報を考慮するのが有効であると考えられる[Bond99]。当面は現処理で適訳が生成されるハガの場合と同様の英訳を目指すことにする。

表6 「象は鼻が長い」の類似表現

種別	日本語表現	英語表現
ハガ	象は 鼻が 長い	Elephants have long trunks
ガガ	象が 鼻が 長い	(The elephant has a long trunk)
ハハ	象は 鼻は 長い	(Elephants' trunks are long)
ガハ	象が 鼻は 長い	(The trunk of the elephant is long)

辞書情報の展開処理としては、例えば、入力文「鼻は象が長い」に対して、まず「長い」の結合価パターン対を検索して“Yが 長い → Y be long”を得る。意味属性体系上、Yに対しhas-a関係で上位となるノードがあればhave構文を、そうでなければof表現となることを考慮して、表7のように、それぞれを展開形として作成する。これらと入力文を照合すると、ガハ-haveが選択されると予想される。

ただし、has-a関係の上位のノードは一般に複数個あるので、条件をどのように書けばいいかには若干問題がある。また、英語表現としてas forを使う方が適切である場合も考えられる。これらについては引き続き検討を進める。

表7 「長い」のパターンに対する展開形

種別	日本語表現	条件	英語表現
(基本形)	Yが 長い	—	Y be long
ハガ-have	Xは Yが 長い	X has-a Y	X have long Y
ガガ-have	Xが Yが 長い		X have long Y
ハハ-have	Xは Yは 長い		X have long Y
ガハ-have	Xが Yは 長い		X have long Y
ハガ-of	Xは Yが 長い	上記以外	Y of X be long
ガガ-of	Xが Yが 長い		Y of X be long
ハハ-of	Xは Yは 長い		Y of X be long
ガハ-of	Xが Yは 長い		Y of X be long

#### 4 おわりに

本稿では、入力文の様相機能語や助詞に着目して

特殊構文の可能性を検出し、辞書登録パターンから可能な展開形を複数派生させ、展開形と入力文を照合する方法を提案した。

本稿で説明した例のうち、「れる/られる」と機能動詞については展開処理のプロトタイプを作成し、ほぼ正しい通りの結果が得られることを確認した。「れる/られる」の処理では、被害の受身も高い精度で判定されている。判定誤りとなったものを眺めると、日本語パターンとの適合度の評価に問題があることがわかった。

今後は、本文中で述べた個別の課題を考慮しながら、様相表現全般をカバーするように展開処理を拡張するとともに、特殊構文への対応を進めていく予定である。特に、機能動詞表現には、「システムを研究開発する」のように、動作名詞が複合語である場合があり、これについても検討する必要がある。また、プロトタイプにより明らかになった日本語パターンとの適合度の評価の見直しも行なう必要がある。

このほか、展開形の生成には副助詞も考慮することにした。現処理では副助詞は適当な格助詞に変換して日本語パターンと照合しているが、副助詞には表現の視点が反映されているため、適切な英訳を得るには無視できないと考えられる。

#### 参考文献

- [Bond99] Francis Bond & Kentaro Ogura. Reference in Japanese-to-English machine translation. *Machine Translation* (1999) (in press).
- [池原91] Satoru Ikehara, Satoshi Shirai, Akio Yokoo & Hiromi Nakaiwa. Toward an MT system without pre-editing -effects of new methods in ALT-J/E-. *In Proc. of MT Summit-III*, pp.101-106 (1991), URL=<http://xxx.lanl.gov/abs/cmp-lg/9510008>.
- [池原97] 池原 悟, 宮崎 正弘, 白井 諭, 横尾 昭男, 中岩 浩巳, 小倉 健太郎, 大山 芳史, 林 良彦. 日本語語彙大系. 岩波書店 (1997).
- [河合87] 河合 敦夫. 日英翻訳システムALT-J/Eにおける様相・時制の処理. 情報処理学会第34回全国大会, Vol.2, pp.1727-1734, 5W-2 (1987).
- [三上60] 三上 章. 象は鼻が長い. くろしお出版 (1960).
- [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), pp.17-75 (1980).
- [奥90] 奥 雅博. 日本文解析における述語相当の慣用的表現の扱い. 情報処理学会論文誌, Vol.31, No.12, pp.1727-1734 (1990).
- [奥98] 奥 雅博. 結合価構造に基づく日本語二重主格形容詞構文の解析. 情報処理学会論文誌, Vol.39, No.11, pp.2968-2977 (1998).
- [白井99] 白井 諭. 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(1999), URL=<http://www.etl.go.jp/etl/nl/sympo99/shirai.html>.