

日英機械翻訳における共起表現の扱い

鈴木 克志* 太細 孝**

* 三菱電機(株) 情報電子研究所
** (財)コンピュータ教育開発センター

多様な構文形態を有する日本語共起表現について分類を行ない、慣用的表現など強い結付きを持つ共起表現と、格関係など弱い結付きを持つ共起表現を統一的に記述できる表層格記述形式を設計し、さらに日英機械翻訳への適用を試みた。得られた記述形式は、(1)副詞句と用言、用言と用言など、多様な共起表現に関する制約が記述可能、(2)慣用的表現と格関係の間の曖昧さの認識に利用しやすい、(3)格要素削除、生成、格繰上げなど、日英間の構造変換に関する処理を制約として同時に記述可能、などの特徴を有する。この結果、従来の翻訳システムで扱われていた述語慣用句の処理に加えて、より広い範囲の共起表現の処理が可能になった。

A PROCESSING OF CO-OCCURRENCE IN
JAPANESE ENGLISH MACHINE TRANSLATION

Katsushi Suzuki* Takashi Dasai**

* Information Systems & Electronics Development Laboratory,
MITSUBISHI ELECTRIC CORPORATION, 5-1-1 Ofuna, Kamakura, Kanagawa 247 Japan.
(e-mail: suzuki@mjbv870.islab.melco.co.jp)
** CENTER FOR EDUCATIONAL COMPUTING,
Terayama Pacific Building, 1-23-11 Toranomon, Minatoku, Tokyo 105 Japan.

Classification of Japanese co-occurrence expressions and its processing in Japanese machine translation system are described. Format for description of surface case structure is designed, so that it can describe both close relations such as idiomatic expressions and loose relations such as case relations in an integrated manner.

The characteristics of this format are as follows: (1) Various relationships such as that between verbs can be described, (2) It is applicable to resolution of ambiguity between close relations and loose relations, and (3) Structure transfer processing from Japanese into English can be described.

**この研究は著者の一人である太細が三菱電機所屬時に行われたものである。

1. はじめに

共起表現の処理は、日本語解析において重要な課題である。特に、日本語から英語への機械翻訳を行なう場合、共起関係の認識を行わずに共起表現を構成する各要素を通常の翻訳処理により独立に処理すると、訳文が意味を成さなくなり訳質低下の原因となるので、適切な共起表現の扱いが要求される。

共起表現には、他の語からの修飾に関する制約や、接続される付属語的表現に関する制約など、多様な制約が存在することが知られている。それらの制約は、最近の研究により機械処理の観点から整理が進んできているが、⁽¹⁾⁽²⁾ 従来の機械翻訳において必ずしも十分に扱われているとはいえない。具体的に言えば、間に他の要素が挿入可能な共起表現の場合、名詞と述語から成る格関係的な共起表現（いわゆる述語慣用句）や、動作名詞と機能動詞から成る機能動詞表現を扱う従来手法は存在するが、⁽⁴⁾⁽⁵⁾ これら以外の多様な構文形態を有する共起表現については、制約の記述手法や翻訳処理への利用手法の検討が必要である。

また、述語慣用句については、全体の意味を各要素の意味から合成することが可能な通常の格関係との識別が微妙である場合が多い。たとえ共起が成立するための制約を満たすときでも、述語慣用句ではなくて通常の格関係表現と見なすべき場合がある。従って、述語慣用句の認識と通常の格関係の認識は、別々の処理ではなくて同時に行なう必要があり、よりもっともらしいものと認識された方を採用できるような解析アルゴリズムが要求される。このアルゴリズムを実現するためには、述語慣用句の記述と通常の格関係の記述が、同一の記述形式で、かつ、保守性の観点からも同一の辞書に記述されていることが望ましい。すなわち、多様な共起表現に関するすべての制約を統一的に記述することが、曖昧さ処理と辞書記述の保守性の観点から理想的であるが、従来はこの点に関する考慮があまり成されていなかった。

さらに、日英翻訳においては、従来から指摘されているように、格構造の変換が必要な場合が多い。⁽¹⁰⁾ 特に共起表現に関しては、修飾要素まで含めた変換の仕方を考える必要がある。従って、共起表現の修飾要素に対する制約を記述する際に、修飾要素に影響を与えるような構造変換の指定についても同一枠組みで記述できれば、辞書記述が効率的に行なえるものと考えられる。

上記の観点を踏まえて、本報告では、次の特徴を有する共起表現記述の枠組みと日英翻訳への利用手法を提案する。

- (1) 副詞句と川言、あるいは川言と川言など、多様な共起表現に関する制約が記述可能である。
- (2) 制約の記述形式は通常の格関係の記述と統一され、格関係と慣用的表現の間の曖昧さの認識に利用しやすい。
- (3) 格要素削除・生成や格繰り上げ等、日英間の構造変換に関する処理が制約として同時に記述可能である。

以下では、まず、2章において共起表現の分類について述べ、3章で共起表現を記述するための枠組みについて述べる。そして、4章で同様枠組みを日英翻訳へ適用した例について述べる。なお、本報告では、一般に慣用的表現と言われる「語の強い結び付き」による表現に加えて、通常の格関係のような「弱い結び付き」をも扱うという意味で、共起表現と呼ぶことにした。

2. 共起表現の分類

2. 1 分類上の問題点

従来の機械翻訳等の自然言語処理システムで主に扱われてきた共起表現は、次の(1)式の構造を持つような慣用的表現であった。⁽⁴⁾

(名詞+助詞)+(川言+付属語的表現) (1)

これは、国語学の分野で議論されてきた慣用表現のうち、文中において述語相当の働きをする慣用句を網羅する構文である。

(1)の構文により、慣用的表現のかなりのものを扱うことができる。しかし、最近、機械処理の観点からの共起表現の分類・整理が進んできており、共起表現にはより広範な構文形態が存在することが明らかになってきている。⁽¹⁾⁽²⁾ これらを参考にして、ここでは次のいくつかの表現が(1)の構文では扱えないことを重視する。

(a) 副詞的表現と川言の組合わせ

例えば、「強く+出る」で「強い態度を示す」の意味を表現するように、副詞的表現と川言との共起により特定の意味を示すようになる表現がある。

(b) 文と川言の組合わせ

埋込み文と川言もしくは従属文と川言の共起により意味が変化することがある。例えば、「・・・したいと思う」や「・・・してほしいと思う」では、主文の川言の意味を埋込み文の付属語的表現が補足する。

また、従来の翻訳システムでは、(1)の構文の中で、特に「影響を受ける」のように、動作名詞と機能動詞が結びついたいわゆる機能動詞表現の処理に関する議論が進んでいた。⁽⁴⁾ 機能動詞表現では、全体的意味を主に動作名詞の方が担い、機能動詞は主に文法的機能を果たしているため、表層構文上で動作名詞を川言化することにより解析が可能である（「影響を受ける」→「影響される」）。しかし、「自信+ある」や「自信+持っている」の表現のように、動作名詞以外の名詞と機能動詞的川言との共起表現が数多く存在する。この場合、名詞を川言化できないので、別の処理が必要となり、このことを考慮した分類を行わなければならない。

2. 2 共起表現の一分類

本報告では、図1に示すように共起表現を分類する。

A. 間に他の語が挿入可能な2語の組合わせ（挿入可能表現）

A-1. 自立語と形式川言の組合わせ

自立語+助詞+形式川言（例：自信がある、基盤とする、迷子になる）

A-2. 自立語と川言の組合わせ

A-2-1. 非慣用語（+助詞）+川言

(i) 格要素と川言（例：新年を迎える、条約を結ぶ、油を売る）

(ii) 副詞句と川言

- ・副詞と川言（例：じっと見る、しきりに勧める）
- ・副詞相当表現と川言（例：強く出る、案に暮らす）
- A-2-2. 川言（＋付属語相当表現）＋川言
 - (i) 埋込み文と川言（例：～したいと思う、～かどうか思案する）
 - (ii) 従属文と川言（例：～しないように注意する、～して喜ぶ）
- A-3. その他（休言＋休言、呼応表現、慣用パターン等）
- B. 間に他の語が挿入不可能な2語の組み合わせ（挿入不可能表現）
 - B-1. 川言＋付属語相当表現（例：使役、受身、～ている、等）
 - B-2. 接頭語・接尾語と他の表現の組み合わせ（例：逆比例する、不熱心である）
 - B-3. 数量と他の表現の組み合わせ
 - B-4. 複合語
 - B-5. その他（格言・ことわざ等）

（助詞には助詞相当表現を含む）

図1. 共起表現の分類

A. 挿入可能表現

まず、一般の手法を踏襲し、共起表現を2語間に他の要素が挿入可能(A)か否(B)かによって2つに分類する。これは、多語性表現と一語性表現⁽²⁾と呼ばれるものにそれぞれ対応する。

挿入可能表現とは、文法的結合の緩やかなものを指し、代表的構文は格要素と川言の共起である。結合度の強弱は一般に判定が難しい。例えば、「油を＋売る」を挿入不可能表現に分類し、2語連続の「油を売る」の形で用いられた場合にだけ慣用表現であると見なす手法もあるが、その場合、2語連続に用いられたにもかかわらず、共起でなく通常の表現として解釈できる場合が扱いにくい。その逆に、やや例外的であるが、「彼は毎朝喫茶店で油を長時間売っている」のように、間に挿入があるが共起として解釈できるような場合も扱えない。そこで、構文的には間に他の要素が挿入可能なので、すべての格要素と川言の組み合わせをここに分類しておき、通常の意味であるか共起の意味であるかを、挿入の有無と他の修飾要素の制約とを合わせて判断できるようにしておく。

A-1. 自立語と形式川言の組み合わせ

2.1 で述べたように動作名詞以外の一般の名詞と機能動詞的川言との組み合わせを扱う必要がある。そこで、本分類では、機能動詞表現を、動作名詞以外の名詞と機能動詞的な川言に拡張し、ここに分類する。この機能動詞的な川言（以下では形式川言と呼ぶ）とは、「ある、ない、なる、よい、わるい、多い、少ない、する、である、だ」という付属語性の強い述語とする。このように、形容詞と形容動詞が含まれる点が形式川言と機能動詞との異なる点である。また、従来機能動詞表現として扱われてきたもので、「影響を受ける」のように、自立語性の強い「受ける」等を用いた表現は、通常の格要素と川言の中で特定2語の共起として扱う。

A-2. 自立語と川言の組み合わせ

2.1 で述べたように、副詞的表現と川言の組み合わせ、及び、文と川言との組み合わせを扱うことができるように、従来の述語慣用句の概念を拡張し、非活用語と川言の組み合わせ、及び、川言と川言の組み合わせ、の2つに分類する。

A-2-1. 非活用語（＋助詞）＋川言

(i) 格要素と川言

川言の表層格パターンに適合するような格要素と川言の組み合わせは、共起の一種としてここに分類する。

(ii) 副詞句と川言

副詞及び副詞に相当する表現と川言との共起は、次の働きを行なう意味で機械翻訳において重要である。

- ・特別の意味を表現する。

例：「強く＋出る」＝強い態度を示す、「案に＋暮らす」＝金銭的に余裕がある

- ・川言が単独で持つ複数の語義を特定する。

例：「訴える」は単独では「告訴する」、「不平不満を述べる」、「人の気持ちに働きかける」のような意味を持ち多義語であるが、「強く＋訴える」のように程度の副詞と共起すると「告訴する」の可能性が弱まり曖昧さが減少する。

- ・川言の訳語選択の曖昧さ解消の際、名詞との共起とともに影響を与える。

例：「TVを見る」ならwatch、「じっと見る」ならlook at、「ざっと見る」ならglance over。

- ・副詞の意味を特定する。

例：「しきりに＋勧める」の場合、「しきりに」は頻度ではなくて熱心さの程度を表わす。

A-2-2. 川言（＋付属語相当表現）＋川言

(i) 埋込み文と川言

埋込み文と川言との間の制約は、格制約として文を格要素にとることと実際の表層助詞の制約を記述するだけでは不十分であり、埋込み文の付属語相当表現と川言との共起関係として認識する必要がある。例えば、「～したいと思う」、「～してほしいと思う」、「～すれば良いと思う」等における埋込み文の付属語相当表現は「思う」の意味を特定化する働きを持ち、訳語選択に影響を与える。さらに、「～するかどうか疑問である」は埋込み文の時制を特定しないが、「～するかどうか思案する」は主に未来のことを言う埋込み文と共起する。

また、このような埋込み文のアスペクトやテンスに対する制約に加えて、埋込み文が支配する表層格の主文への繰り上げ制約も記述できる。これについては後述する。

(ii) 従属文と川言

埋込み文と同様に、従属文と用言との共起においても、付属語相当表現を考慮する必要がある。例えば、「～して喜ぶ」は「～する(した)ことを喜ぶ」と同義であるので、従属文と用言との共起を、埋込み文と用言との共起と同様に分類・記述することが望ましい。その他の例としては、「注意する」は注目、用心、忠告などの多義を持つが、「～しないように+注意する」の共起の場合は注目の意味が落ち、「～してはいけない+注意する」の場合は忠告の可能性が強くなる、という例がある。

A-3. その他

休言間の共起、用言付属語との共起表現(例:「たとえ+～しても」)、パタンの表現(例:「走っても走っても追いつかない」)については、重要であるが本報告では扱っていない。

B. 挿入不可能表現

間に他の語が挿入不可能な2語の組合わせを挿入不可能表現と呼ぶ。その多くは全体を1語として見なすことが可能な表現であるが、共起として扱うべき表現もある。ここでは共起表現として4種類(B1～B4)に分類し、それ以外をその他(B5)として1語とみなすことにした。

B-1. 用言+付属語相当表現

付属語の連接からなる付属語的表現は従来から詳細な分析と整理が進んでおり、⁽³⁾ 本報告の付属語的表現もそれらの分類に基づいている。付属語的表現は全体を1語として扱うことにしている。ここでは、用言と、付属語相当表現のうちのいわゆる助述表現に相当する助動詞的表現との間の共起を扱う。例えば、「覚える」は暗記するの意味が強いが、「ている」が承接した「覚えている」は「記憶して忘れない」の意味になり、訳語を選択する場合、memorizeとrememberのように異なる訳語となる。また、使役や受身等の付属語的表現の承接が表層格構造に影響を与えることはよく知られている。⁽⁹⁾

B-2. 接頭語・接尾語と他の表現の組合わせ

例えば、「逆比例する」は「逆に+比例する」の共起表現に同義と見なせるし、「不熱心である」は「熱心で(ある)+ない」と同義である。このように、接辞と自立語の組合わせのうち、これまでに分類した共起表現に属する表現で書き換えられるものがある。また、「線分+上+にある」や「悪+習慣を+断つ」のように、接辞を1要素と数えると3語共起と見なせるような表現も多い。しかし、現状では、接辞と自立語とを合わせて1語として見なすことにしており、共起表現としての扱いは今後の課題になっている。

B-3. 数量と他の表現の組合わせ

数量表現の処理は、実用的な機械翻訳において非常に重要であるが、まだ十分に整理されていないのが実情である。数量表現に関しては、名詞とそれが取り得る助数詞の間の共起関係(例:「鉛筆+3本」)、数値とそれが直接連結し得る名詞との共起関係(例:「3+文字」とは言うが「3+学生」とはあまり言わない。⁽¹¹⁾)などがある。これらの共起表現は、現在、数量表現に特化したアルゴリズム的な処理で扱っており、本報告の範囲外とする。

B-4. 複合語

複合語はまとめて1語として複合語辞書中に登録している。語彙分解により共起として扱う手法は今後の課題である。

B-5. その他

いわゆる格言や諺の類は、ここに分類し、まとめて1語として扱う。

3. 共起表現記述の枠組み

3. 1 記述方式

共起表現を辞書中に記述して文解析に利用する場合、共起表現の中心となる単語に関する辞書記述中に共起情報を合わせて記述する手法⁽¹⁰⁾と、共起表現辞書を一般の単語辞書と分離して持つ手法⁽⁴⁾がある。後者の利点は、通常の用法の記述を乱すことなく共起表現に関する情報を記述できるため辞書の保守性に優れている点にある、と言われている。確かに、共起表現についての情報の記述は多くの制約を扱うため複雑なものとなり、熟知者以外の者には編集が困難である。従って、辞書を分離することにより、例えば翻訳システムのユーザには通常の辞書のみ編集を許可する、という現実的に妥当な使用上の制約が実現できる。しかし、通常用法と共起表現の区別は微妙である。強い結び付きの2語からなる共起を慣用的表現として慣用辞書に持ち、格制約などの弱い結び付きを通常の辞書に持った上で、慣用表現の解析の後に格解析を行なうような方式の場合、慣用表現としての認識に格制約が参照できない。すなわち、慣用表現が通常用法かの区別が微妙な入力文に対してもっともらしい方を比較選択する、という処理が行えないという欠点がある。そこで、語の強い結び付きと弱い結び付きをともに共起表現という統一された論理的枠組みで扱うという方針とし、物理的にも一つの辞書にまとめて記述することにした。保守性を確保するための方策としては、厳密な記述形式を用意することにした。

3. 2 記述の概要

前節で述べた理由から、2.2で述べた共起表現の記述を、共起表現を構成する2語のうち、主要となる語の表層格を記述する格記述辞書中にまとめて記述することを考える。ただし、現在A-3、B-2、B-3、B-4、及びB-5については別の枠組みで扱っており、本報告の対象外としている。主要となる語とは、A-1では自立語(例:自信がある)、A-2では後の方の用言(例:強く+出る)、B-1では用言(例:覚えて+いる)、である。ただし、A-2-1の非活用語と用言の組み合わせにおいては、場合により非活用語の方に記述できる。

記述の例として、「自信がある」に対する格記述を図2に示す。

彼は、来年は試験に合格する自信がある。

日本語の表層格構造記述

【構文解析】

<日本語の依存構造>

彼 自信 合格
ある
来年 試験

【格処理】

<英語の依存構造>

he confident pass
next year examination

【生成】

He is confident of passing an examination next year.

```
jcase([自信, [cf(1, jishin, [case(A, e, [ga], [ani], nil),
                             case(B, o, [ni], nil, nil)],
                             [formal(aru, a)], eng(confident1, 1)),
        cf(2, jishin, [case(C, e, [ga], [ani], nil),
                       case(D, o, [no], nil, [form(sen)]))],
        [formal(aru, a)], eng(confident2, 2))];
```

英語の表層格構造記述

```
ecase(confident2, spec(confident, adj),
       [cf(2, [case(A, e, subj, nil, nil),
                case(B, o, obj1, prep(of), [cltype(prcl)]))]);
```

図2. 格記述とそれを用いた依存構造変換の例

日本語の表層格記述においては、主要語の「自信」を見出しとする2つの格フレームを与える。

両方のフレームとも形式用言指定 formal(aru, a)を持つ。この指定は、自立語と形式用言との共起表現全体に対して格構造を記述していることを示し、英語依存構造への変換時に1つのノードに縮退される。formalの第一引数のaru が「ある」との共起を示し、第二引数のa が「が」格で共起することを示す。

例文に対応するのは2番目のフレーム（以下ではJf2と呼ぶ）であり、この格記述形式の特徴的な面を示している。Jf2について詳細を説明する。Jf2は、2つの格、e(経験者格)とo(対象格)から成る。対象格に対しては格標識noが指定されており、これは「の」格または「・・・する」という連体修飾が可能であることを指定する。さらに、格オプションとして、格に文要素を取ることを表わす文指定form(sen)の指定がある。これらから、「埋込み文(サ変) + の + 自信がある」という表現と、「埋込み文(・・・する) + 自信がある」という表現との対応を指定していることになる。Jf2は、英語の格構造記述へのポイントにより英語のconfident2フレームに対応づけられており、「ある」のノードと「自信」のノードがconfident2フレームに対応するひとつのノードに縮退する。その結果、図に示すような構造変換が起きる。

英語の表層格記述においてはJf2と対応づけられたconfident2のフレーム（以下では、Ef2と呼ぶ）はJf2と同じ2つの格を持つ。通常は、日本語と対応する英語の格フレームは同じ個数の格を持つ。日英間の対応する格フレームにおいて、同じ概念を表わすにもかかわらず表層格の個数が異なる場合もある。この場合は格の削除や生成を行なう機能が用意されている。

対象格に対しては、英語格標識として第一目的語のobj1、英語格形式として前置詞のprep、さらに、英語格オプションとして現在分詞節のprcl、がそれぞれ指定されている。これらから、「・・・of～ing」という表現を指定していることになる。その結果、図2に示すような英文が訳出される。

3.3 日本語の表層格構造の記述

日本語の表層格記述の1単位は、複数の格フレームから成る。格フレームは複数の格から成る。格には動作主(agent, aで示す)などの名前が与えられる。格の名前は英語側の格構造へ変換する際に、英語側の格への対応をとるための識別子としての役割を果たす。格には、格標識、意味マーカ、及び格オプションの3種類の制約が記述できる。格標識は、構文上の制約であり、入力文の助詞に相当する表現と照合される。意味マーカは意味上の制約条件である。格オプションは本記述形式の特徴の一つであり、各種の共起表現に関する制約が集約的に記述される場所である。また、格フレームは格フレームオプションと英語格記述へのポイントを持つ。格フレームオプションにも共起表現に関する制約が記述できる。日本語の表層格記述形式を図3に示す（英語格記述については省略）。

<格構造記述> ::= jcase(<見出し>, <格記述>)

<見出し> ::= 動詞・形容詞・形容動詞の語幹; サ変名詞; 名詞

<格記述> ::= <格フレーム>のリスト

<格フレーム> ::= cf(<語義番号>, <読み>, <格リスト>, <格フレームオプション>のリスト, <英語格記述へのポイント>)

<語義番号> ::= 1, 2, 3, ...

<読み> ::= ローマ字綴り

<格リスト> ::= <格>のリスト

<格> ::= case(<格変数>, <格の名前>, <格標識>のリスト, <意味マーカ>のリスト, <格オプション>のリスト)

<格変数> ::= Prologの論理変数

<格の名前> ::= 動作主; 経験者; 対象; 場所; 起点; 目標; 授与; 相手; 手段・道具; 材料; 原因・理由; 資格・役割; ... など

<格標識> ::= ga, ni, wo, ... など。

<意味マーカ> ::= abs, ani, ins, loc, ... など。

<格オプション> ::= <共起単語指定>; <分類番号指定>; <分類番号範囲指定>; <文指定>;
<dscm指定>; <raise指定>; <down指定>; <delete指定>; <adv指定>

<直接単語指定> ::= dir(日本語見出し)

<分類番号指定> ::= scm(分類番号)

<分類番号範囲指定> ::= scm-range(分類番号from, 分類番号to)

(4) 形式用言指定(formal 指定)

先に図2の例で見たように、formal指定により自立語と形式用言との共起を扱うことができる。ここでの自立語は動作名詞に限定されず、一般の名詞をも扱う。

(5) 共起単語指定(co-oc指定)と共起側活用指定

自分が他の単語の格要素となるような共起関係を、ここにco-oc 指定としてその単語の字面を記述することにより、扱うことができる。この機能により、2語間の共起関係を扱う場合、どちらの語の格構造に対しても共起関係の指定が行なえる。例えば、「油を売る」においては、「売る」の格フレームの対象格に格オプションとして「油」との共起指定を行なうことと、「油」という名詞に格構造を与えて格フレームオプションとして「売る」との共起指定を行なうことの両方が選択可能である。

3. 6 共起表現に対する制約の扱い

共起表現には、他の語からの修飾に関する制約や、接続される付属語の表現に関する制約など、多様な制約が存在する。例えば、すでに図2の例で見たように、「～の自信が+ある」では「合格の自信がある」のように連体修飾する要素にはサ変名詞を指定する制約があり、「～に自信が+ある」では「数字に自信がある」のように特定の問題領域を示す名詞に修飾される。また、「間が+持てない」では「持つ」に対して否定助動詞が接続するという制約が存在する。これらの制約は、最近になって機械処理の観点からの整理が進んできた。(1)(2) 前節まで(3.4～3.5)に述べた各種のオプションは、共起表現に対する制約を扱うことが可能である。これらのオプションは、格処理における格フレーム適用時に入力文が条件を満たすかどうかチェックされ、満たされた場合、その格フレーム適用の可能性を高める働きをする。しかし、制約には、このような適用可能性を高める制約だけではなく、満たされた場合に適用可能性を減少させるような制約も存在する。そこで、制約を次のように分類する。

制約	必須制約	必ず満たさなければならない制約。	メタ指定=must.
	任意制約	満たされた場合、妥当性を高める制約。	デフォルト.
	非推奨制約	満たされた場合、妥当性を低くする制約。	メタ指定=unlike.
	禁止制約	満たしてはならない制約。	メタ指定=not .

各種のオプションは、任意制約である。ただし、オプションにメタ指定を用いることにより他の3種類の制約として働くようになる。また、任意制約の中にも妥当性に重み付けがしてあり、直接単語指定、形式用言指定、及び共起単語指定は強い結び付きに対応するので特に重くなっており、必須制約に近いものになっている。

各種の制約は文献(1) によく分類されている。以下の例はこれから引用した。

(i) 「頭が切れつつある」とは言わない。・・・not(dsem(X))を格フレームオプションに与える。X は変数を示す。

(ii) 「喉から手さえてる」とは言わない。・・・must(ga)を格標識に指定する。

(iii) 「攻守(が)所を変える」の「が」は省略可能。・・・vmodif(助詞省略可能)とgaを格標識に指定する。

(iv) 「へそに茶をわさせる」とは言わない。・・・not(voice(cause))を指定する。ただし、(i)と異なり、使役のみ禁止する。

以上の制約は、語の修飾や共起に関する制約であり、依存構造の範囲内での制約とも言える。この他に、構造形に関する制約が重要である。次の例(1)を特別な格オプションで扱うことを検討中である。

(v) 「手が喉から出る」とは言わない。(語順変更)・・・「喉から」にbefore(a)オプションを与える。a はagent を示す。

(vi) 「棒になった足」とは言わない。(関係節化)・・・「足が」にnot-relativeオプションを与える。

4. 日英機械翻訳への適用

共起表現の記述の枠組みを評価する目的で、日英機械翻訳への適用実験システムを試作した。翻訳処理の概要を図4に示す。システムはHELCON PSI II上に実装され、約10万語の辞書を有する実用的規模のシステムである。プログラムは専用の文法記述言語とPSIのシステム記述言語SP(Prologベースのオブジェクト指向言語)で記述され、辞書もPrologの項形式で構造的に記述されている。トランスフォーマ方式であり、拡張文脈自由文法や手続きの段階的適用による文解析が終了した後、構造変換を経て英文生成が行われる。

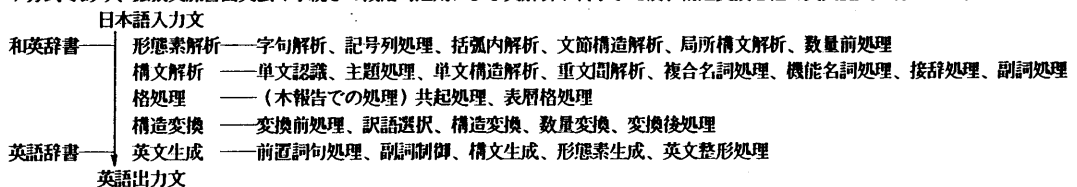


図4. 日英機械翻訳システムの翻訳処理

構文解析終了後に、本報告で述べた表層格記述を用いて格処理を行なう。格処理の内容を図5に示す。

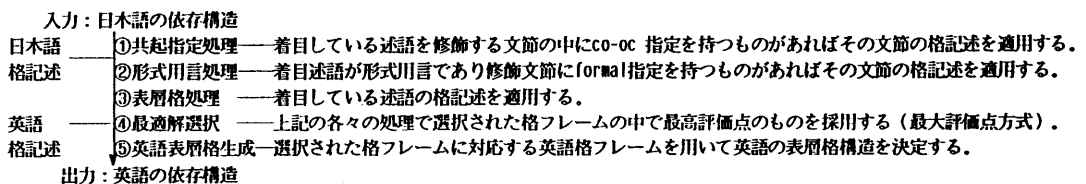


図5. 格処理アルゴリズム

格処理アルゴリズムは、日本語の依存構造の中の述語(川言)に対して順番に着目して図5の処理を繰り返す。①～③の格記述適用処理では、日本語の表層格記述の中から入力文に対して最も良く適合する格フレームを選択する。適合度は最大評価点方式により計算される。この方式は、オプションのそれぞれに重み付けを与え、満たされたオプションの重みの合計を評価点とする方式である。重みは負の制約ではマイナスとなる。必須制約と禁止制約は、他の制約を無視できるほど絶対値の大きな重みを与える。

一部のオプションは、解の選択ではなく、解が決定した後、英語の表層格生成処理において評価される。そのようなオプションには、構造変換を行なうdelete指定、raise指定、及びdown指定がある。これらの指定により、英語表層格生成において、日英間の表層格構造の相違が調整される。英語の表層格構造記述で指定された格の個数が日本語側の格よりも多い場合は、自動的に英語側で表層格が生成され、英語側の格記述で指定された訳語が選択される。

本アルゴリズムの特徴は、着目している述語の格記述の適用と、その述語を修飾する名詞の格記述の適用とを同時に行なうことである。従来の方式にあるような、慣用表現の認識と格関係の処理が順番かつ排他的に行われるアルゴリズムでは、強い結び付きの共起表現と弱い結び付きである格関係との間の曖昧さが扱えない。例えば、次の例文を考えてみる。

(i)彼は油を毎朝30分は売っている。・・・「油」と「売る」の間に挿入語句があるが慣用表現として解釈可能。

(ii)角のガソリンスタンドで油を売っています。・・・挿入語句が無いが通常の用法。

本方式では、情報を記述する場所として、慣用表現は「油」の格記述、通常用法は「売る」の格記述、にそれぞれ自然に記述できる。そして、「油」の格記述として場所格に禁止制約を与えて「ani が、loc で(禁止)、売る(co-oc指定)」、「売る」の格記述として「ani が、phy を、loc で」のフレームを与えておけば、(i)では「油」の格記述、(ii)では「売る」の格記述が最大評価点を取って選択される。しかし、この記述だけだと、次の(iii)が扱えない。

(iii)彼は油を毎朝喫茶店で売っている。・・・「で」格が存在するため禁止制約が適用されて通常用法が選択される。

(iii)を扱うためには、「油」の格記述を「ani が、loc で(油の販売店の分類番号を禁止)、売る(co-oc指定)」と拡張すれば良い。このように、本枠組みにおいて、語の強い結び付きと弱い結び付きを統一的な形式で記述して、さらに格処理時に比較考慮することにより、慣用の表現と通常用法の間の曖昧さが扱える枠組みとなった。

これまでに約7500語の見出しに対して表層格記述の付与作業を行なった。うち1500語は名詞を見出しとしており、co-oc指定やformal指定による強い結び付きの共起表現約2500個の記述を含んでいる。現在は任意制約のみ付与している。この作業の結果、2.2で述べたA-3、B-2～B-5以外の分類項目に属する共起表現について、本報告で提案した記述形式でほぼ記述できる見通しを得た。

5. おわりに

多様な構文形態を有する共起表現について分類を行ない、慣用の表現などの強い結び付きを持つ共起表現と、格関係などの弱い結び付きを持つ共起表現を統一的に記述するための表層格記述形式を提案し、さらに、日英機械翻訳への適用を試みた。その結果、

- (1) 副詞句と川言、川言と川言など、多様な共起表現が扱える
- (2) 格関係と慣用の表現との間の曖昧さの認識に利用しやすい
- (3) 格要素削除、生成や格繰上げ等、日英間の構造変換に関する情報も同時に記述できる

ような特徴を持つ枠組みが得られた。

本報告では詳細を議論できなかったが、本記述形式により、扱いが難しいとされている3語間の共起関係など、複雑な共起表現をある程度扱うことが可能となっている。例えば、「後ろに+手が+回る」の例では、「手」の格記述中に「回る」へのco-oc指定と「後ろ」へのdir指定及びdelete指定を行なうことにより、格処理時に「手が+回る」の2文節が1つのノードに縮退し、それを修飾する「後ろ」のノードが削除されて、全体を「be arrested」に変換するような処理が記述できる。このように、述語を2つの文節が修飾しているような3語共起は、本枠組みで記述可能である。さらに言えば、この時の修飾として川言間の修飾も扱えるので、「(乗物)に+乗って+(場所)へ+行く」のような表現も、記述の仕方により「(乗物)で+(場所)へ+行く」と同等に処理することが可能である。

しかし、「線分+上+にある」のように接辞がらみのもの、「特別+扱い+する」のように複合語がらみのものは、それぞれ接辞を含めた全体や複合語全体を1語として扱わなければならない。従って、3語共起に関しては、2語の共起で現在扱えないA-3、B-2～B-5を検討することと合わせて、さらに検討する必要がある。特に、述語への修飾を含まない名詞間のみから成る共起に関しては、現在別形式の辞書情報で扱っているため辞書の保守性が悪い。これらの拡張を今後検討し、研究の方向としては「各種制約を統一的に記述でき、制約間に生じる曖昧さを現実的処理時間と資源量で形式的に扱っていくようなシステム」の実現を目標としていく所存である。

謝辞 格記述辞書の作成にご尽力いただき多くの貴重な意見をいただきました山田道子氏に深い感謝の意を表します。また普段の熱心な議論と貴重な示唆をいただいた高山泰博氏、宮原浩二氏をはじめとする自然言語処理グループの皆様にも感謝いたします。

参考文献

- (1) 首藤、吉村：日本語における語の固定的共起、文法的知識と意味的知識の蓄積管理シンポジウム、pp. 1-6(1989)。
- (2) 首藤、吉村、武内、津田：日本語の慣用的表現について、情報処理学会自然言語処理研究会66-1(1988)。
- (3) 首藤、橋原、吉田：日本語の機械処理のための文節構造モデル、電子通信学会論文誌、Vol. J62-D、No. 12, pp. 872-879(1979)。
- (4) 奥：日本語慣用表現の分析と日英翻訳への適用：情報処理学会自然言語処理研究会62-2(1987)。
- (5) 奥：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)。
- (6) 林、奥、石崎：日英翻訳システムA L T-J/Eにおける日英変換技術、情報処理学会第33回全国大会、pp. 1753-1754(1986)。
- (7) 池原、宮崎、白井、林：言語における話者の認識と多段翻訳方式、情報処理学会論文誌、vol. 28, no. 12, pp. 1269-1279(1987)。
- (8) 小泉、梶、鶴：日英機械翻訳のための日本文の格構造モデル、情報処理学会第34回全国大会、pp. 1273-1274(1987)。
- (9) 平井、北橋：格の強度と述語の構文および意味属性を用いた格構造の変換生成、情報処理学会論文誌、vol. 28, no. 3, pp. 240-249(1987)。
- (10) 西山、高松：機械翻訳における中間表現の変換と文の合成手法、情報処理学会誌、Vol. 26、No. 10, pp. 1165-1173(1985)。
- (11) 吉武、高山、鈴木、太細：日本語の数量表現とその英語への機械翻訳に関する一考察、情報処理学会自然言語研究会74-7(1989)。
- (12) 首藤、藤田、鶴丸、吉田：叙述情報とその表現形式、電子通信学会オートマトンと言語研究会A176-42(1976)。