

Muプロジェクトにおける意味マーカ概念と体系

石川徹也
(情報大)

坂本義行
(電総研)

佐藤雅之
(JICST)

科技庁の機械翻訳プロジェクト (Muプロジェクト) の日英翻訳において、日本語の構文・意味解析で用いた名詞に関する意味マーカ概念と体系について述べる。この意味マーカ体系は、13のファセットと49の意味スロットから成る。日本語構文解析において、動詞と名詞との間の格支配関係の詳細な解析を行うために意味マーカを設定した。名詞の表す概念の分類に関して、その分類の定義、体系、方法について述べ、我々の意味マーカ体系とその意味マーカの付与の方法について述べる。最後に、日本語解析における格助詞が示す深層格の決定ならびに日英変換における訳語選択において、例を用いて、意味マーカが果している役割について述べる。

Concept and Structure for Semantic Information in Mu-Project

Tetsuya Ishikawa,¹⁾ Yoshiyuki Sakamoto,²⁾ Masayuki Sato³⁾

1) University of Library and Information Science (ULIS)

2) Electrotechnical Laboratory (ETL)

3) The Japan Information Center for Science and Technology (JICST)

This paper discusses the semantic features of nouns classified into categories in Japanese-to-English translation, and proposes a system for semantic markers. In our system syntactic analysis is carried out by checking the semantic compatibility between verbs and nouns. The semantic structure of a sentence can be extracted at the same time as its syntactic analysis.

We also use semantic markers to select words in the transfer phase for translation into English.

The system of the Semantic Markers for Nouns consists of 13 conceptual facets, and is made up of 49 filial slots (semantic markers) as terminals.

1. はしがき

Muシステムは*、解析・変換・生成の3段階から成るトランスファ方式が用いられている。この各段階に、意味素性を導入することにより、より正確な翻訳が可能となる。日英翻訳におけるこの意味素性をを用いた解析・変換では、どのように正確で、高品質の結果がもたらされるかを示す。日本語解析では、この意味素性が用言の格フレーム内に埋め込まれた構造で解析される。一方、日本語から英語への変換では、訳語選択において大きな役割を果たしている。この意味素性は、より言語に固有のものであり、格フレーム中の深層格と名詞の対として、辞書の中に記述して置かねばならない。しかし、その名詞の個数は非常に多く、これをすべて格フレーム(格支配情報)として、辞書に記述しておくことは困難である。従って、われわれはその代わりとして、その意味素性を分析し、意味マーカを設定した。これを用言の辞書中に記述することにより、効率的な処理を可能とした。

2. 格文法と意味マーカ

Muシステムの日英翻訳において、日本語解析は格文法に基づく依存構造解析を行っている。英語のような語順を重視する言語では、句構造文法が適している。動詞が後置するといった例外を別にすれば、語順が自由で、その格支配情報は格助詞が担っている日本語では、格文法が適していると言われている。この格文法では、用言が中心となり、これに支配される形で、その他の品詞がこれに従属している。即ち、構文解析において動詞と名詞の意味的な解析が必要となり、構文解析と意味解析が同時並行して行われなければならない。用言によって支配される格支配情報は、格助詞(表層格)、格ラベル(深層格)及び名詞意味マーカによって決定される。用言は、動詞、形容詞、形容動詞から成る。格ラベルには、必須格を表すものと任意格を表すものがある。また、一つの格助詞が複数の深層格の役割を果たしているものがある。例えば、格助詞「に」は、場所、場所・終点、時、役割、方式、終状態、相手、構成要素、条件、範囲規定等10以上の格ラベルの役割を果たしている。Muシステムでは、34個の格ラベルを設定している。

用言辞書の格支配情報欄に、各格パターン別に格ラベルとそれと共に起する名詞意味マーカを記述するとともに、名詞辞書の各単語に対し該当する意味マーカを記述している。現在、約7万語が収録されている。但し、用言辞書では、原則として、翻訳対象文中に出現した用例のみについて格パターンが記述されている。

3. 意味マーカと意味マーカ体系

1) 分類の定義

分類(classification)とは、「分類する目的に対して対象となる物事を、一定の規準に従い、一定の体系に仕分ける行為」と考えることができる。物事とは人間が認識し得る“物体”と“事象(現象)”をいう。一定の規準とは仕分ける規則(分類規準)をいい、体系とは仕分ける規則によって生成される配列構造(分類体系)をいう。故に、分類する為には、目的、分類対象、分類規準及び分類体系が必要になる。分類は、目的に即し、対象の外延又は内包によってなされる。分類に際し、どちらの系で分類するかは、対象の特性に依存するものではなく、あくまでも分類の目的に対し選択されるものである。故に、分類規準及び分類体系は対象と独立に設定することができる。分類規準及び分類体系の深度(詳細性)及び形式化は、目的によって決定づけられる。

2) 分類体系

分類の目的は、対象を分類すること自体よりもその結果の利用(検索)の為にある。故に、分類結果と検索との一致性の為に、分類体系(classification system)がツールとして必要になる。

* 本研究は、国の科学技術振興調整費による『日英科学技術文献の速報システムに関する研究』の一部として行ったものである。その遂行のために作業分科会を組織し、その審議のもとに研究を進めている。

分類体系は、人間が認識しうる概念 (concepts) の認識単位を分類標識 (classification item) として表現し、分類標識の配列によって体系化したものである。

分類標識は、番号、単語、特殊記号等によって表現される。番号で表現した分類標識を分類番号 (classification number) といい、単語で表現した分類標識を分類名辞、一般的には索引語 (index term) といい、特殊記号で表現した分類標識を分類記号 (classification symbol 又は marker) という。

各分類標識が意味するところの定義 (definition)、及び時に範疇注記 (scope note) が明示される。体系化は、分類標識の配列、及び配列に伴う分類標識間の関係を明示することによって表現される。

分類標識の配列には、順次配列体系 (sequential order) と、関係配列体系 (relational order) とがある。この配列の体系化は、データ構造と同質とみなせる。順次配列体系は分類番号の場合は番号順に、索引語及び分類記号の場合はその読みの五十音順又はアルファベット順になる。関係配列は、順次配列体系のように形式的に配列するのではなく、分類標識が表す概念の類別化により配列する。

3) 概念の類別化と体系化

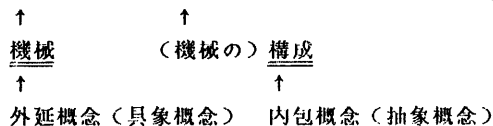
概念とは、「物事に対する人間の認識単位 (category)」と考えることができる。概念の体系化とは、「人間が認識し得る概念の認識単位を組織化すること」をいう。認識単位の表現は、一般的に言葉によって示される。

① 概念の類別化 (categorization)

・外延概念と内包概念

外延概念とは、独立し存在及び機能する物事の一般的総称名をいい、内包概念とは、物事の内的特性の一般的総称名をいう。故に、外延概念と内包概念は、相対的に認識されるものである。

(ex) 電子計算機 : 入出力部、記憶部、演算部、制御部

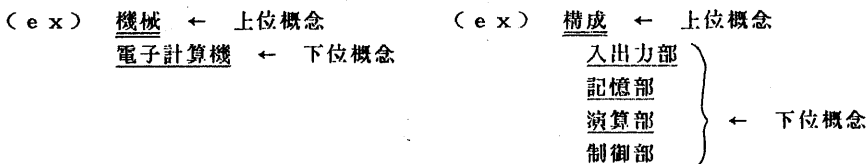


具象概念は、具体的に存在する具象物に対する概念であり、抽象概念は、様相に対する概念とみることができる。

② 概念の体系

・種概念と類概念

概念の認識単位には、範疇が存在する。一つの範疇を種概念といい、その範疇内の細分化概念を類概念という。種概念は類概念を包摂する。一般に、種概念を上位概念、類概念を下位概念という。



③ 概念の体系化

上位概念と下位概念は、相対的なものである。上位概念の最上位概念は“物事”になり、最下位概念は“個々の物事の名称 (単語)”になる。それぞれの認識単位の深度は、分類する目的によって決定づけられる。抽出された上位概念は、体系化の目的による配列、及び下位概念はその下位概念の上位概念への関係づけによって体系化できる。

4) 概念の分類体系の例

概念のカテゴリ化は、古くから哲学上の見解に基づき数多くの提示がある。以下に、時代を追って代表例を掲げる。

① アリストテレス (Aristoteles: B.C. 384~332) のカテゴリ

i) 物質 (Substance)、ii) 量 (Quantity)、iii) 質 (Quality)、iv) 関連 (Relation)、v) 場所 (Place)、vi) 時間 (Time)、vii) 状態又は位置 (Situation or Position)、viii) 所有

又は維持形質 (Possession or Acquired Character)、ix) 活動 (Activity)、x) 受動 (Passivity or Passion)

②カント (Kant, I.:1724~1804) のカテゴリー

i) 量 (Quantity) : 単一体、多数体、全体、ii) 質 (Quality) : 真実、否定、限定、iii) 関連 (Relation) : 実在性、因果関係、相互関係、iv) 様相 (Modality) : 可能性、現実性、必然性

③ランガナタン (Ranganathan, S.R.:1892~1972) のカテゴリー

i) 時間 (Time)、ii) 空間 (Space)、iii) エネルギー又は作用 (Energy or Action)、iv) 物 (Matter)、v) 個性 (Personality)

④シェラー (Shera, J.H.:1982~1903) のカテゴリー

i) 行為者 (Agent)、ii) 行為 (Act)、iii) 手段 (Instrumentarity)、iv) 目的 (Object)、v) 時間 (Time)、vi) 空間 (Space)、vii) 結果 (Product)

ここに掲げた例は、“知識の整理”の為に、その知識の表現された“物”としての図書の分類に活用されてきた。以後、図書館学 (library science) の中で、図書の形態分類 (主題分類: subject classification) 及び書誌分類 (内容分類: contents classification) の体系化に発展し、今日、前者の例として、例えば日本十進分類表 (NDC: Nippon Decimal Classification)、後者の例として、例えば国際十進分類表 (UDC: Universal Decimal Classification) に至っている。これらの分類表は、知識の体系化の為に、学問分野を10単位ずつ細分化し、体系化したものである。知識の体系化に対して、単語の意味の類別化を図り、体系化したのがシソーラス (thesaurus) である。シソーラスは、結果として概念のカテゴリー化を提示している。シソーラスの原型を提示したのは、Rojet, P.M. (1779~1869) であり、以後、論文等の索引付与 (indexing) のツールとして使用されるようになり、今日、主題分野別に多くのシソーラスが作成・利用されるに至っている。

5) 分類体系の作成

分類体系の作成には、次の通りの方式がある。

①帰納的方法 (ボトム・アップ方式)

全ての対象の外延又は内包項目をそれぞれ分析・抽出し、類別化し、分類標識を設定し体系化する方法である。分類する対象が事前に明確に確定している場合には、この帰納的方法により体系化できる。

②演繹的方法 (トップ・ダウン方式)

対象となる全てを想像し、考える対象の特性を設定し、類別化し、体系化する方法である。分類の目的が明確になっていても、対象が確定し得ない場合には、この演繹的方法により体系化する。帰納的方法による分類体系は、対象が更新される場合には、新規の対象にどれ程対応できるかという問題が発生する。演繹的方法による分類体系は、帰納的方法による分類体系の問題と合わせ、体系化自体の策定、例えば分類標識の確定等に、困難性が伴う。

6) 分類作業の方法

分類作業は、目的に即し、対象の特性を外延又は内包に類別することによってなされる。一般に、外延による分類を形態分類 (morphological classification) といい、内包による分類をファセット分類 (facets classification) という。対象それ自体の分類は、整理・保管を意図し形態分類がなされる。対象の内容を多項目の特性によって取出せるようにファセット分類がなされる。分類は一般的に図1の過程によって行われる。

①内容分析 (contents analysis) …… 対象の特性分析を行う作業である。

②対象の特性の抽出 (extracting for classification features) …… 分類規準に従い、必要とする対象の特性を抽出する作業である。

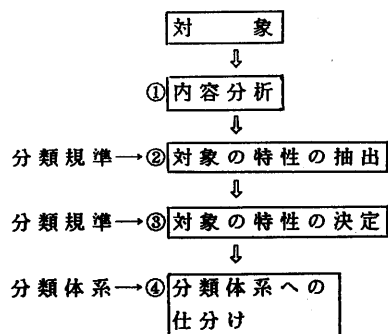


図1 分類作業工程

③対象の特性の決定 (decision for classification features) …… ②によって複数抽出された対象の特性の中から分類規準に従い最適な対象の特性を決定する作業である。

④分類体系への仕分け (classification) …… ③によって決定された対象の特性に従い、対象を分類標識のもとに仕分ける作業である。

7) Muプロジェクトの意味マーカ体系

Muプロジェクトにおいて、名詞及び代名詞を対象に、単語の概念分類を目的とし、分類方法を演繹的に構築し、単語74,444語に対し概念分析を行い、意味マーカを付与し、分類した。

①概念カテゴリーの設定

名詞は、人間が認識し得る物事の概念記号であると考え、概念カテゴリーの設定を行った。最上位の概念カテゴリーに、次の4つがあると考えた。

i) 具象物 (Concrete objects) ……世の中に存在する具体物をいう。

ii) 抽象的概念 (Abstract concepts) ……人間の知的活動の結果生成された規範、規準等をいう。

iii) 現象 (Phenomena) ……社会及び自然界に起こる様をいう。

iv) 人間の行動 (Human actions) ……人間の行動そのものをいう。

この4つの基本概念カテゴリーをもとに、以下のような細分化を図った。

i) の具象物を、生物 (Animate objects) と無生物 (Inanimate objects) に分け、ii) の抽象的概念を知的抽象物 (Intellectual objects) と名付けた。上記の3つの生物、無生物、知的抽象物には、それぞれに全体と部分が存在する。そこで、上記3つの概念カテゴリーは全体を表すものとし、あらたに上記3つの概念カテゴリーの部分を表す概念として部分 (Parts) を表す概念カテゴリーを設けた。iv) の人間の行動を、内的状態と外的状態に分け、内的状態を心情 (Sense & Feeling)、外的状態を行動 (Actions) に分けた。全体と部分、あるいは部分と部分のように物事の存在の関係を明示する概念カテゴリーとして、属性 (Attributes) を設けた。国・機関・組織 (Nations & Organizations)、測度 (Measurements)、場所・空間 (Spaces & Topography)、及び時間 (Time)。以上、12の概念カテゴリー (conceptual category) を設け、これ以外の概念を有する名詞の為にその他 (Others) というカテゴリーを設けた。

②意味マーカの設定

上記12の各概念カテゴリーの範囲内において、明示できる下位概念をそれぞれに設定した。その数は、全体で36個である。各上位概念に含まれない概念を、その他 (Others) の概念とした。

③体系化

1つの概念カテゴリー内の下位概念を意味スロット (semantic slot) と名付け、その記号を意味マーカ (semantic marker) と名付けた。意味スロットの集合の代表をファセット (facet) と名付け、ファセット名 (facet name) にファセット記号を名付けた。

体系化は、1つのファセットを独立に位置づけ、1つのファセット内の意味スロットを一階の階層構造で表した (図3)。

以上のようにして作成した「名詞意味マーカ体系」(System of Semantic Marker for Nouns) は、およそ外延系の体系となっている。ここで、特に“事象”を意味する単語について、特別の意味マーカを別に用意し、図3に示すように、その他を含め13個設定した。概念カテゴリー、ファセット、意味スロットの関係を図2に示す。

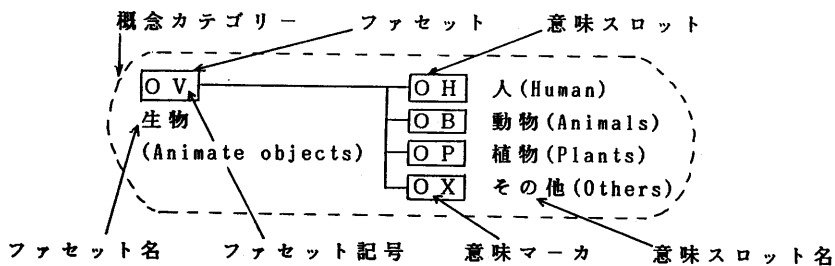


図2 意味マーカの体系構造

④意味マーカ付与規準

各意味スロットの決定及び意味マーカ付与規準を表1（一部分）に示す。

表1 意味マーカの意味定義表（一部分）

ファセット	意味スロット	定義及び語の範囲	語例
〔OF〕 国・機関・組織		人間の機能集団（国家、党派、会社、統治主体、機関、団体、行政区画名等）。規約を持つ集団。 意志動詞の主格になれる語。	アメリカ、ブラジル、学会、アリゾナ州、東京都、国鉄、JICST、科技厅、IBM、会社、国際標準化機構、日本、フランス、英国
〔OV〕 生物			生物
	〔OH〕 人	人名・職名・地位・構成員・親族の名称等。 ☆「立場」、「役目」、「役割」等人間の社会的役割を表す語は対象とせず、これらの語は、〔AU〕（機能）に入れる。 ☆人の器官の名称は〔EL〕（生物の器官及び構成要素）に入れる。 ☆人の病気の名称は〔PB〕（生理的現象）に入れる。	兄弟、構成員、専門家、オペレータ、デカルト、教授、冤柱者、教授陣、報道陣、医師団
	〔OB〕 動物	☆動物の種の名称、慣用名、総称名など。 ☆動物の器官の名称は〔EL〕（生物の器官及び構成要素）に入れる。 ☆動物の病気の名称は〔PB〕（生理的現象）に入れる。	動物、ねこ、プランクトン、ペット、家畜
	〔OP〕 植物	☆植物の種の名称、慣用名、総称名など。 ☆植物の器官の名称は〔EL〕（生物の器官及び構成要素）に入れる。 ☆植物の病気の名称は〔PB〕（生理的現象）に入れる。	穀物、アオイマメ、植物、エンドウマメ、高山植物、プランクトン
	〔OX〕 その他		ウイルス、媒体、目標物、実体、生命体

8) 意味マーカの付与

対象の単語に対し、以下の手順に従って作業を行う。意味マーカ付与は、事前の名詞品詞細分類の後に行う。

①単語の意味理解

単語の意味を、辞書等を用いて内容を理解する。

②単語の意味素性の抽出

単語の意味素性をすべて分析し抽出する。

③意味マーカの付与

個々の意味素性に意味マーカを全て付与する。意味マーカの範疇理解は、表1の定義及び語例に従う。

・注記

i) 意味マーカ付与対象名詞

普通名詞、固有名詞、動作名詞（サ変、その他）、疑問代名詞、人称代名詞、指示代名詞（場所を示す名詞のみ）副詞的名詞（時と場所を示す名詞のみ）

ii) 次の単語には付与しない。

化合物名、数式、製品等のモデル名

iii) ファセット名と同一の単語は、ファセット記号を付与する。

iv) 対応するファセットはあるが、その中で対応する意味マーカがない場合は、そのファセット内のその他のマーカ（～x）を付与する。

v) どのファセットにも対応しない単語は、その他のファセット名のマーカ（xx）を付与する。

vi) 複数の語基からなる複合語は、最後の語基のみで意味素性を抽出することなく、複合語全体の意味をとらえ付与する。

vii) “事象”を意味する単語について、別に定めている特別の意味マーカを定義表の語例に準拠し本来の意味マーカとは別に付与する。

viii) 属性は、物事のある関係を示す概念であるが、存在のあり方の総称（属性名）と、その程度・状態等の名称（属性値）に分けられる。そこで、属性の概念カテゴリー内に、属性名と属性

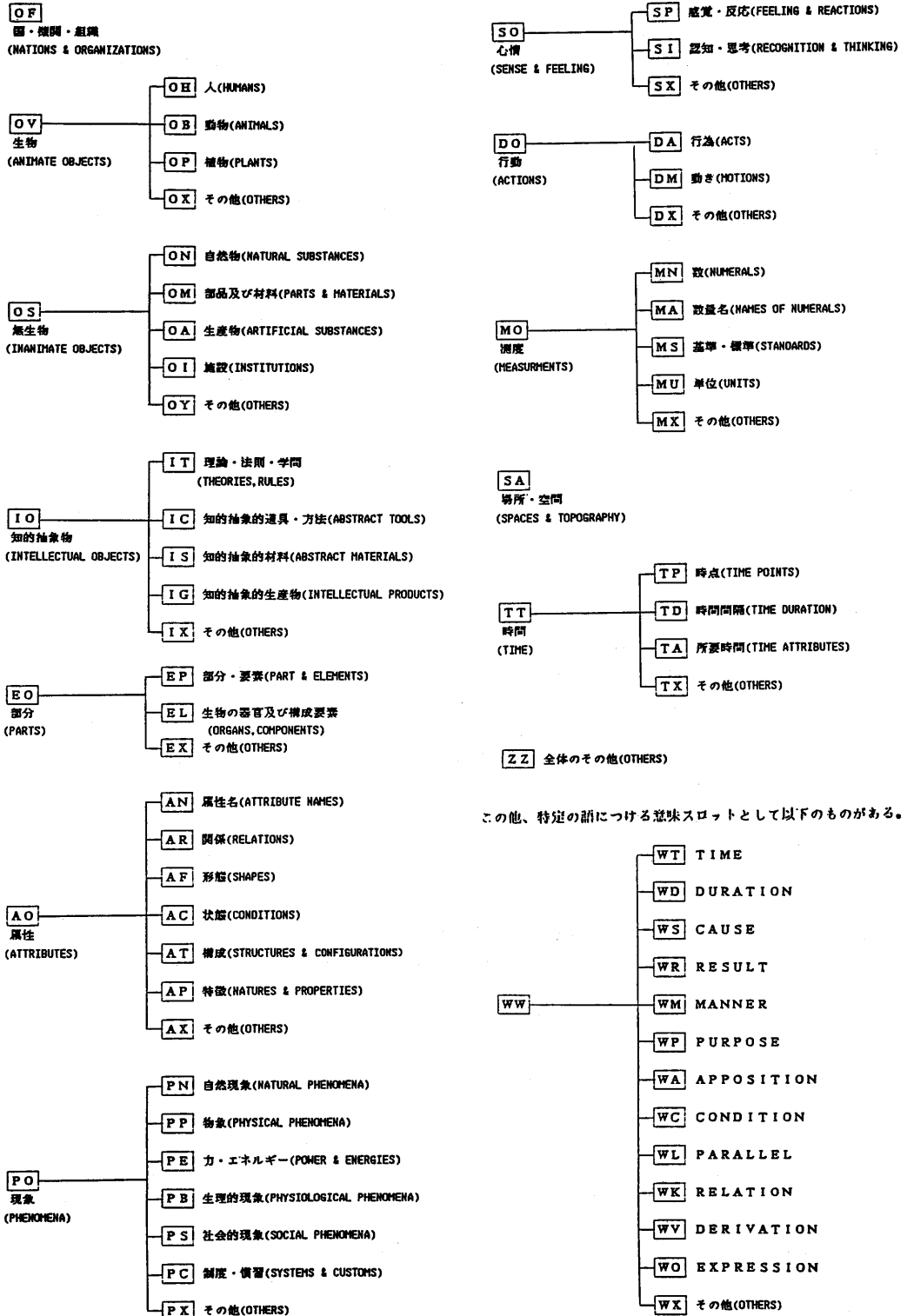


図3 名詞意味マーカ体系表

値を示すレベルの異なる下位概念が意味スロットとして並置されているので、使い分けに注意する。
ix) 現象は、事象が顕在化した様子を示す概念である。それには内在(潜在)する原因(因果)がある。その原因が、人間にとって好ましい現象であれば“+”、好ましくない現象であるならば“-”という因果記号(causal property)を、特に顕著な単語のみにつける。

4. 副詞・連体詞・接統詞の構文・意味情報

副詞・連体詞・接統詞に、それぞれ下記の構文・意味情報を辞書記入シートに記入する

①副詞の構文情報(品詞細分類)

- i) 状況副詞・・・文中の述語にかかる副詞であり、他の細分類に属さない一般の副詞をいう。
- ii) 程度副詞・・・形容詞・形容動詞・状況副詞にかかって、その程度を強調する副詞をいう。例えば、「非常に大量の」「さらに高度な」等の用法であり、但し、「常に大量の」のように程度に関与しない副詞や、「さらに研究する」のように直接動詞にかかる副詞は除く。
- iii) 陳述副詞・・・仮定・打消・疑問等の表現形式を誘導する副詞をいう。例えば、「もし」: 仮定、「必ずしも」「決して」: 打消、「何故」: 疑問、のように、常に呼応を引き起こす副詞を収める。
- iv) 数量副詞・・・動詞と共に起する名詞に対し数量的限定をもたらす副詞をいう。例えば、「Aをすべて処理する」を「すべてAを処理する」と言い換えられるように、「すべて」は動詞よりむしろ名詞と強く結びつく。逆に「Aを毎日処理する」を「毎日のAを処理する」とは言えないから「すべて」のような副詞と区別する。

②副詞の意味情報

品詞細分類に従って、対応する以下の意味情報を付与する。

- i) 法・・・陳述副詞が誘導する意味情報。
仮定、疑問、打消、願望、比況、談歩
- ii) 相・・・状況副詞が決定する述部の意味情報。
完了、進行、反復、習慣
- iii) 時制・・・状況副詞が決定する文の時制。
過去、現在、未来
- iv) 程度性・・・副詞が他の程度副詞により修飾されうるか否かを基準に、副詞自体の過程を示す意味情報。
尺度、極

③連体詞の構文情報(品詞細分類)

体言の修飾の仕方により、以下の4つに分類する。

- i) 指示連体詞・・・特定のもの(概念)を指すが、対象を複数指示する連体詞をいう。
(ex) この、その
- ii) 限定連体詞・・・特定のもの(概念)を指すが、対象が限定される連体詞をいう。
(ex) ある、さる
- iii) 疑問連体詞・・・対象を特定化できず、疑問文の前置詞的役割を果たす連体詞をいう。
(ex) どの、どのような
- iv) 形容詞的連体詞・・・対象を形容し、結果として対象(概念)に変化をもたらす連体詞をいう。
(ex) 小さな、少しの

④接統詞の構文情報(品詞細分類)

- i) 文接統詞・・・文と文の接統のみに用いる接統詞をいう。
- ii) 句接統詞・・・句と句、文と文のどちらにも用いられる接統詞をいう。

5. 日本語解析・日英変換での意味マーカの活用例

5.1 格パターンを用いた動詞の意味解析

(1) 一つの動詞がいろいろの意味を持つ場合に、各用例について意味を定めるのに意味マーカを用いる。これは特に和語動詞において有効である。例えば、「当たる」は以下に示すような意味を持つ。

a. ぶつかる、接触する

格パターン：A [物体・物象] が B [物体・物象] に 当たる。

① 石がガラスに当たる ② 光が斜面に当たる ③ (電波) 信号が山に当たって反射する

b. 従事する

格パターン：A [有意志体] が B [行為] に 当たる。

① 巡視艇が救助に当たる ② A氏が指揮に当たる ③ A社が点検修理に当たる

c. 該当する、相当する

格パターン：A [人] が B [人] に 当たる。A [場所] が B [場所] に 当たる。

A [時] が B [時] に 当たる。A [計量単位] が B [計量単位] に 当たる。

注) A, Bともにいろいろな値をとりうるが、同一用例中におけるAとBは、あまり異質のものであってはならない。

① サイゴンの銀座並木通りに当たるチュードー通り ② 今日がちょうど一年目に当たる

③ 1インチが2.5cmに当たる ④ 女性のブラウスに当たる開襟シャツの通動着

(2) 一つの文中に2個以上の動詞を持つ場合には、意味判別の必要があり、名詞と動詞の係り受け関係を判定するのにこの意味マーカが利用できる。

『AがBにXしたCをYする』という文には、名詞A, B, Cと動詞X, Yが含まれるが、このうち名詞A, BがXに係るのかYに係るのかは、単語の具体的な内容が理解できなければ判定できない。例えば、

① 私がポリバケツに腐ったリンゴを捨てる

ならば、 A が B に Xした C を Yする となり、

② 私が黄色に変色したシャツを捨てる

ならば、 A が B に Xした C を Yする となり、

③ 計画が失敗に帰した原因を究明する

ならば A が B に Xした C を Yする となる。

5.2 任意格の意味解釈

2章で前述したように、一つの格助詞が複数の深層格の役割を果たしている。

格助詞「で」を例として、名詞の意味マーカと動詞の表層格と共起する意味マーカから自由格を決定し、英語の格構造と前置詞“in”を決定する過程を示す。

例(1) 主として現在市場で活躍している数値制御超高速穴あけ機について説明。

→ The numerically-controlled superhigh speed drilling machines are explained which are active mainly now in markets.

例(1)では、名詞及び動詞の日本語解析辞書(図4)の内容を調べて名詞「市場」の意味マーカが(SAとPS)であるのに対し、動詞「活躍する」の格パターンの2番目の格スロットに記述されている名詞意味マーカは(SA又はOF)となっているので、この場合“SA”で一致するので、表層格「で」の深層格での解釈は、「場所」とであると決定できる。

例(2) 逆制動放射によるレーザ吸収項について二つの取り扱い方で問題を解き、数値解を求めた。

→ problems are solved by two handling methods about laser absorption terms by the inverse damping radiation, and numerical solutions are obtained.

動詞「含む」の日英変換辞書(図5)の内容を調べると、訳し分け条件部の2番目の格スロット(すなわち対象格)をうめる語の意味マーカがOS又はOM又はPN又はPB又はPP又はPEの場合、“contain”を選択し、それ以外は“include”を選択するようになっている。例(1)では「ナトリウム」も「沃土スカンジウム」も意味マーカOMをもつので“contain”が選択されるが、例(2)では、「正視双直交系」の意味マーカはICなので“include”が選択される。

動詞の変換辞書には、図5で示すように ①深層格を埋めた名詞の意味マーカをチェックする訳し分け条件 ②条件を満たした場合の英語での訳語 ③条件を満たさない場合の英語の訳語 ④それぞれを選択した場合の日本語の格構造と英語の格構造の写像関係が記述されている。

6. 考察

単語74,444語に付与した作業経験から、意味マーカ体系及び意味マーカ付与の問題点を述べる。

- ①哲学の発展にみられるように、概念の体系化は、際限のない課題である。演繹的方法による分類体系では、無限に近い単語の意味分類をするには無理がある。構文解析上、意味解析のシミュレーションを行い、意味マーカの働きを分析した上で、必要とする意味マーカを設定し、それに基づき、体系化を図る必要がある。
- ②対象語数を、経験的には3,000語位に切って、意味マーカ付与作業を行い、その結果をフィードバックすることにより分類体系の修正及び付与規準を詳細に検討する必要がある。特に、リダクンダンシー・ルール(不可ルール)の設定は、この種の作業マニュアルには重要な役割を果たす。
- ③意味素性の分析また意味スロットへの仕分けに、作業員個人の知識・能力が反映されるだけに、この種の作業には人材の確保、教育訓練が重要な役割を果たす。
- ④意味素性の分析は人間が行うとしても、意味マーカ付与の“ゆれ”等の回避、作業の効率化等を考えると、支援システムが必要となる。例えば、同一意味マーカを有する単語の参考表示システムとか、データ修正システムとか、この種の事前の道具立てを充分図ることが必要である。

7. 謝辞

当意味マーカ体系は、Muプロジェクト推進の4年間の中で、試行錯誤の結果完成を見たものである。この間多くの議論が何回も繰り返された。初期の段階で中井 浩氏(現 常盤大学教授)が体系の原案を提示され、高井貞治氏(JCS)の協力を得て付与作業を行い、修正を行った。途中、収録語の拡張に伴い、辻井潤一助教授(京大)、鳥海 剛氏(JICST)、萬濃立夫氏(JCS)の協力を得て修正を行った。拡張語への意味マーカ付与作業をテクニカルサービス社が誠心誠意行って下さった。また、構文パターンへの意味マーカの埋め込みは、木村睦子氏(IBS)が担当された。この間、高所大所から、長尾 真教授(京大)から御指導を受けた。ここに、これら多くの方々へ謝意を表すしだいである。

8. 参考文献

- [1] 岩瀬泰郎(北島武彦 監修)、資料組織論(講座 図書館学 第三巻)、成文堂、314p.、昭和50年
- [2] 哲学事典、平凡社
- [3] 坂本義行、「格構造を中心とした用言と付属語辞書」情報処理学会 自然言語処理研究会、38-8、1983
- [4] 中井 浩、佐藤雅之、「語の収集と体系を中心とする辞書について」情報処理学会 自然言語処理研究会、38-7、1983

- [5] 石川徹也、佐藤雅之、高井貞治、「自然言語処理における名詞・代名詞の意味機能について」
情報処理学会 第 2 8 回全国大会、1 9 8 4
- [6] Sakamoto,Y., Satoh,M. and Ishikawa,T. : Lexicon features for Japanese Syntactic
Analysis in Mu-Project-JE, COLING 84, Stanford, 1984.
- [7] 長尾 真、辻井潤一、中村順一、「日英科学技術文献の速報システムに関する研究の概要」
情報管理、 Vol.28 No.1, 1985
- [8] Sakamoto,Y., Ishikawa,T. and Satoh,M. : Concept and Structure of Semantic Markers for
Machine Translation in Mu-Project, COLING 86, bonn, 1986