

オブジェクト指向分析支援システムCAMEO —日本語文章記述からの設計要素の自動抽出—

大野 雅志*1
原田 実

青山学院大学大学院
青山学院大学理工学部助教授

オブジェクト指向分析を自動化するために、自然言語による要求文からオブジェクト図やイベントトレース図を構成する要素を抽出する過程を自動化するシステムCAMEOを開発した。

CAMEOは一連の日本語解析処理を経て生成される動詞・名詞・形容詞意味フレーム、並びにそれらを補助する2つの意味フレームを基に、以下の(1)分析ルール、(2)調整ルールを順次適用し、要素を抽出するシステムである。

分析ルール: 名詞意味フレームからクラス・属性/集約要素・属性値・ロール・名詞イベント・個数/件数・数を、動詞意味フレームから関連(集約)・イベント・動作・冗長関係を、形容詞意味フレームから属性値となる要素を抽出し、これらのどれであるかを分類し所定の形式で出力する。この際、フレーム間の修飾関係も分析に用いる。全部で19種類ある。

調整ルール: まず名詞意味フレームから抽出された要素で曖昧さが存在するものに対し、競合解消優先順位を基に、競合を解消する。そしてクラス候補に対してアクティブかパッシブの決定を行い、最後に冗長関係にあるクラスや属性間で代表を検索し、冗長なクラスや属性は代表で置き換えるルール。

今後は、学習による常識の獲得や文脈解析や対話構築による曖昧性除去、抽出された設計要素を設計図に自動レイアウトすること、イベントフローや状態遷移の条件などを厳密に獲得することなどが必要である。

Computer Automated Modering Engine for Objects

- automatic extraction and classification of design elements from texts -

Masashi Ohno, Minoru Harada

Faculty of Science and Engineering, Aoyama Gakuin University

The purpose of our study is to develop a system CAMEO-Computer Automated Modeling Enginefor Objects-, which automates object oriented analysis and design activity. CAMEO inputs requirement statements written in Japanese. First, CAMEO parses and analyses each input sentence into the semantic frames (of verb, noun, adjective, time-noun, and predicate parallel). Next, CAMEO extracts and classifies design elements from those sentences according to the following rules.

[Extraction rule]:Classes, attributes or aggregation elements, attribute values, roles, and events are extracted from noun frames mainly depending on its semantic marker and by what words it is modified. Associations, aggregations, inheritances, redundancies, events, and actions are extracted from verb frames depending on its semantic marker and what cases it has. Similarly, attribute values are from adjective frames.

[Adjustment rule]:When different classifications of the same word appear in the former steps, it is solved basing on the criteria such that the classification supported by many sentences and the more specific classification shuld to be preferred.

CAMEO was applied to a sample problem consisnting of 25 sentences, and actually extracted 79 design elements. These elements included all the elements and classified properly done by well experienced analysts.

* 1 平成6年4月より北電情報システムサービス(株)

概要

具体的には、自然言語による要求文からオブジェクト図やイベントトレース図を構成する要素を抽出する過程を自動化するシステムCAMEOを開発した。CAMEOは一連の日本語解析処理(JUMAN**, SAX**, JANUS)を経て生成される動詞・名詞・形容詞意味フレーム、並びにそれらを補助する2つの意味フレームを基に、以下の(1)分析ルール、(2)調整ルールを順次適用し、要素を抽出するシステムである。

(1)分析ルール:名詞意味フレームからクラス・属性/集約要素・属性値・ロール・名詞イベント・個数/件数・数、動詞意味フレームから関連(集約)・イベント・動作・冗長関係、形容詞フレームから属性値となる要素を抽出し、これらのどれであるかを分類し所定の形式で出力する。この際、フレーム間の修飾関係も分析に用いる。全部で19種類ある。

(2)調整ルール:まず名詞意味フレームから抽出された要素で曖昧さが存在するものに対し、競合解消優先順位を基に、競合を解消する。そしてクラス候補に対してアクティブかパッシブの決定を行い、最後に冗長関係にあるクラスや属性間で代表を検索し、冗長なクラスや属性は代表で置き換えるルール。

この研究において開発したCAMEOやJANUSによって、従来の分析者が手作業で抽出してきた要素を、すべて自動的に抽出可能となった。この結果、従来要求文にラインマーカなどを付けて行われていた設計要素の抽出・分類が自動的に行え、オブジェクト指向分析作業の効率と信頼性が大きく向上した。

しかし、CAMEOはまだ人間としての常識に欠けるため、同意語などを判断できないので、クラス候補が手作業による場合よりかなり多くなる。また、文間の修飾を理解できない。従って今後は学習による常識の獲得や文脈解析や対話構築による曖昧性除去などを行う必要がある。また、CASEツールとしては、抽出された設計要素を設計図に自動レイアウトすることが考えられる。さらに、自動プログラミングに繋げるためには、イベントフローや状態遷移の条件などを意味的に厳密に獲得する必要がある。

1. 序論

1. 1 研究の背景研究テーマ

近年、オブジェクト指向分析(OOA)が注目され始めてきている。その理由として、今日開発されるシステムは巨大化、複雑化し、たえず変更を余儀なくされると同時に、従来のシステムに比べよりデータ指向的かつ、データのモデル化についても優先されるようになったためである。OOAに従った分析技法は、以上の問題点を解決するものとして期待されている。

2 オブジェクト指向をもとに解析する分析者は、システム内容が記述された日本語要求文を受け取ると、日本語解析技術やオブジェクト指向分析技法に基づいてクラス、関連、属性、ロールを含むオブジェクト図や、イベント、動作の流れを示すイベントト

レース図への展開を、すべて手作業で行っているのが現状である。

しかし、数多くの日本語要求文を分析することで、日本語文章の中には潜在的なルールが数多く存在していることが分かり始めた。にもかかわらず、現状のシステムにはまだこの一連のルールを知識としてデータベース化し、自然言語による要求文から自動的に分析抽出するシステムはまだ皆無である。

1. 2 研究テーマ

オブジェクト指向分析における分析図の設計要素を日本語文章から自動抽出、分類する自動システムを作成することを研究テーマとする。

具体的な処理方法は、次の2段階にわけて行う。

1つは事務処理に関する日本語要求文に対し、形態素解析、統語解析、意味解析を行った後、意味表現として格概念を取り込んだ意味フレームに変換すること、もう1つは、抽出された意味フレームからオブジェクト指向分析に基づいてクラス、関連、動作等の候補を抽出、分類することである。

1. 3 自然言語解析における現在の技術水準

コンピュータの能力向上、言語学の成果蓄積、ソフトウェアの成果蓄積により、人間とコンピュータとの間に新しいインタフェースを確立しようと自然言語理解システムの研究が1970年代によく本格化され、現在に至っている。

しかし、計算機による日本語解析は膠着言語であるため、欧米の言語解析と比べて形態素解析が問題となり、これには入力文内の個々の形態素を認識できるだけの実用に耐えられる大きな辞書が必要となる。同時に、如何に辞書を整備するかという問題も存在していることから、共通のツールとして世の中に流布された形態素解析システムはまだ存在しない。

また、統語解析を行う際の解析には広く認められ同意を得た日本語文法、文法用語がない現実があり、意味解析にあたって「意味」を表す枠組みは「意味ネットワーク:Quillian,1968」、「格文法:Fillmore,1968」、「論理式:Montague,1970」、「概念依存図式:Shank,1972」等のうちどのアプローチが最善であるか分からず、模索しているのが現状である。

以上のような状況から、既存のシステムは適応領域をかなり限定し、ある決まった極少数の語彙、典型的構文しか扱えないシステムが多い。よって精細な形態素解析、統語解析、意味解析は行われていないことになる。

英語解析も含む今日まで開発された主なシステムとしては

- ・統語解析中心の「ELIZA」(Weizenbaum:1968)
- ・意味解析に格文法を用いた「SHRDLU」(Winograd:1972)
- ・意味解析に論理式を用いた「TELL」(東工大の佐伯:1984)
- ・意味解析に意味ネットワークを用いた「KIPS」(富士通の杉山:1984)
- ・意味解析に意味ネットワークを用いた「ARIES」

(青山学院の原田:1986)
 ・形態素解析の「JUMAN」(奈良先端科学技術大学院大学の松本:1993)
 ・統語解析中心の「SAX」(奈良先端科学技術大学院大学の松本:1993)
 などが挙げられる。

2. 本研究の基本的考え方
 2. 1 システム全体構成
 解析過程(システム構成)図を以下に示す。解析は4ステップからなる。

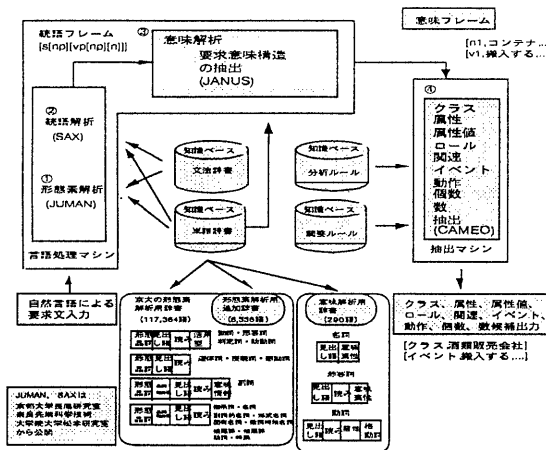


図1 システム構成図

CAMEOは、入力された日本語要求文を①形態素解析(JUMAN)②統語解析(SAX)③意味解析(JANUS)の順に処理した結果の意味フレームを基に、④オブジェクト指向分析の自動化を行う。①と②については奈良先端科学技術大学院大学の松本裕治先生のJUMAN⁽¹⁾、SAX⁽²⁾を用いた。以降では③と④について詳しく論じる。

2. 2 本研究の特長と工夫した点
 2. 2. 1 日本語処理における意味解析(JANUS)

単語レベルの意味解析と文レベルの意味解析を同時に行い、解析結果として意味フレームを作成し、単語の意味情報や、それぞれのフレームのリンク情報を表現する。また、オブジェクト指向分析にも活用できるように十分留意した内容となっている。

[1] 動詞の場合

1) 動詞意味フレーム

動詞意味フレームの中で中心的存在となるのは格(CASE)であり、述語に対してどんな役割であるかを示すものである。JANUSでは21パターンの格ラベル(Subject, Object, Receiver, Sender, Secondary, At, From, To, Start, Goal, Cause, Tool, Component, Manner, Condition, Range, Topic, Viewpoint, Degree,

Predicative, Assert)を用意し、動詞によって取り得る格ラベルを分類している。

[Vn, 基本表記, 読み, 活用型, [活用形, 文中表記], 自|他|判, [[Case分類, 接辞品詞, 名詞・動詞・形容詞のFrameID]], [被修飾名詞のFrameID], [[副詞, 動詞意味分類]], [態], [アスペクト], [ムード], [その他の格尾, 文番号, [類義語], [上位語], [反義語], [その他]]]

図2 動詞意味フレームの一般形

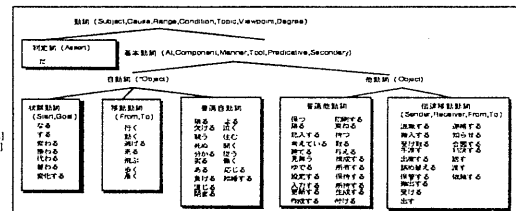


図3 動詞意味分類

2) 動詞意味辞書

前節で述べた動詞意味分類の情報等を格納する辞書である。この動詞意味辞書はJANUSの動詞意味フレーム作成時に参照される。

フォーマット:

[見出し語, 読み, 自|他|判, 動詞意味分類]

例: [着く, つく, 自動詞, 移動動詞]

[持つ, もつ, 他動詞, 普通他動詞]

[2] 名詞の場合

1) 名詞意味フレーム

普通名詞、サ変名詞、固有名詞、それらの複合名詞がフレームとなる。「名詞間の係り受け情報」は、名詞間の修飾・被修飾関係、AND関係、OR関係を示す。「数修飾、数、名詞助数辞」はその名詞が持つ数情報を示す。名詞意味分類(有形{人間、動物、植物、固体、液体、気体}、役割、出来事、相互作用、詳述、記述、地名)は、オブジェクト指向分析のクラス抽出に適した分類法としている。

[Nn, 表記, 読み, 品詞種分類, [[名詞間の係り受け情報]], [[名詞修飾動詞のFrameID], [[名詞修飾形容詞のFrameID], [[名詞修飾連体・指示詞], [[数修飾, 数, 名詞助数辞], 名詞意味分類, 文番号, [類義語], [上位語], [反義語], [その他]]]

図4 名詞意味フレームの一般形

3) 名詞意味辞書

前節で述べた名詞意味分類情報を格納する辞書である。この名詞意味辞書はJANUSの名詞意味フレーム作成時に参照される。

フォーマット: [見出し語, 名詞意味分類]

例: [倉庫係, 役割]

[コンテナ, 固体]

[3] 形容詞の場合

1) 形容詞意味フレーム

動詞同様、格が存在するが、取り得る格ラベルは「Subject, Cause, Range, Condition, Topic, Viewpoint, Degree」の7パターンに限っている。

[An, 基本表記, 読み, 活用型, [活用形, 文中裏記], 形容詞意味分類, [[Case分類, 接統品詞, 名詞・動詞・形容詞のFrameID], [被修飾名詞のFrameID], [副詞, 副詞意味分類], [格属], 文番号, [類属語], [上位語], [反義語], [その他]]]

図5 形容詞意味フレームの一般形

2) 形容詞意味辞書

前節で述べた形容詞意味分類情報を格納する辞書である。この形容詞意味辞書はJANUSシステムの形容詞意味フレーム作成時に参照される。

フォーマット: [見出し語, 読み, 形容詞意味分類]
例: [赤い, あかい, 物理的/色]

[嬉しい, うれしい, 心理的/感情]

[4] 時相名詞の場合

時相名詞名詞(毎日、その都度etc)は、重文の中で存在する場合、一度に多数の述語フレームに係るのが一般的である。そのたびに述語フレームにその情報を格納するのは得策ではない。よって特別な意味フレームとして時相名詞意味フレームを設定した。時相名詞意味フレームの一般形: [Zn, 表記, 読み, 品詞, [関連フレームID], 文番号]

[5] 述語並列の場合

文が重文である場合、各単文ごとの述語フレーム(動詞、形容詞)を述語フレームとしてリンクする。文はSubject格を持つのが一般的であり、もしSubject格を持たない述語フレームが存在し、述語並列フレームが存在すればリンクしている述語フレームのSubject格を参照し補完できる利点がある。よって特別な意味フレームとして述語並列意味フレームを設定した。

述語並列意味フレームの一般形:

[Pn, 述語並列, [関連フレームID], 文番号]

2. 2. 2 オブジェクト指向分析支援システム (CAMEO)

CAMEOは、名詞意味フレーム、動詞意味フレーム、形容詞意味フレームから、分析ルールを用いてクラス、属性、属性値、ロール、関連(集約)、イベント、動作、個数、数の候補を抽出するシステムである。この際、分類上曖昧性が存在する要素については調整ルールを適用し競合を解消する。

[1] 分析ルール

名詞意味フレーム、動詞意味フレーム、形容詞意味フレームに関する分析ルールを述べる。

1) 名詞に関して

JANUSが出力した名詞意味フレームからはクラス、属性/集約要素、属性値、ロール、イベント、個数、数を抽出する。この際、名詞がどのような意味素性を持つかや名詞がどのような修飾語がついているかによって、種々の分析ルールが起動される。後者については特定の言い回しの修飾語があるかを調べていく。以下、これらを抽出する分析ルールを説明する。

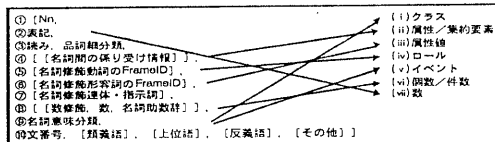


図6 名詞分析ルール

(i) クラスルール

2.2.1[2]の2で説明した名詞意味分類の中で、有形(人間、動物、植物、固体、液体、気体)、役割、出来事、相互作用、詳述を表している名詞は、アクティブクラス候補として抽出する。同様に、記述を表すものはアクティブクラスかパッシブクラスとする。パッシブクラスとは、「~票」、「~書」、「~リスト」、「~データ」などデータを表すもので、アクティブクラス間を受け渡されるものである。

例えば、以下のようなものがある。
コンテナ(固体)、倉庫係(役割)、飛行(出来事)、依頼(相互作用)、モデル(詳述)、積荷票(記述)

クラスフォーマット: [クラス, クラス候補名詞, アクティブ/パッシブ, [文番号....]]

具体例:

[クラス, 倉庫係, アクティブ, [1, 2]]

(ii) 属性/集約要素ルール

以下の3つのルールに従うものは、属性かあるいは集約要素とする。属性か集約要素であるかの曖昧性の解決は、調整ルールによって後で決定する。
属性ルール1: 特定名詞「名前」、「年齢」、「住所」、「重さ」、「長さ」、「速さ」、「色」、「種類」、「日時」、「データ」、「番号/NO/NO./No./No./no/no.」は属性とする。

属性フォーマット: [属性, 属性候補名詞, class (属性を持つクラス), [文番号....]]

具体例:

[属性, 名前, class(倉庫係), [1, 2]]

属性/集約要素ルール1:

クラス名詞の所有句に続く名詞意味フレームは属性か集約要素のどちらかと考えられる。例えば「酒類販売会社の倉庫」において酒類販売会社がクラス候補である場合、倉庫は属性か集約要素である。

このような曖昧性がある場合、CAMEOは、この事実に対して次の2つの出力を行う。

属性フォーマット: [属性, 属性候補名詞, class (属性を持つクラス), [文番号....]]

具体例:

[属性, 倉庫, class(酒類販売会社), [1]]

集約要素フォーマット:

[集約, 集約候補名詞接続助詞, 所有, class(親クラス候補名詞), hasA(子クラス候補名詞), [文番号....]]

具体例：[集約, の, 所有, class(酒類販売会社), hasA(倉庫), [1]]

属性／集約要素ルール2：

特定動詞「持つ」、「とる」の目的語、「ある」、「与えられている」の主語、「なる」の構成語は属性か集約要素のどちらかと考えられる。例えば「人は行き先階を持つ」において人がクラス候補である場合、行き先階は人の属性か集約要素である。

このような曖昧性がある場合、CAMEOは、この事実に対して次の2つの出力を行う。

属性フォーマット：[属性, 属性候補名詞, class(属性を持つクラス), [文番号....]]

具体例：[属性, 行き先階, class(人), [1]]

55

集約要素フォーマット：[集約, 集約候補動詞, 所有, class(親クラス候補名詞), hasA(子クラス候補名詞), [文番号....]]

具体例：[関連, 持つ, 所有, class(人), hasA(行き先階), [1]]

(iii) 属性値ルール

属性値ルールは以下の2つのルールから分析している。

属性値ルール1：

クラスを表す名詞意味フレームに係っている形容詞は属性値と考えられる。例えば「赤い花」や「赤い色の花」において「花」がクラス候補である場合、「赤い」はこの花の色の属性値となる。

以下にCAMEOが出力する具体例を出力フォーマットと共に示す。

属性置フォーマット：[属性値, 属性値候補形容詞, attribute(属性値を持つ属性), class(属性を持つクラス), [文番号....]]

具体例：

[属性値, 赤い, attribute(—), class(花), [1]]

[属性値, 赤い, attribute(色), class(花), [1]]

属性値ルール2：

「～には、...がある」があるという文型において、「～」が属性なら、「...」は属性値の列挙である。例えば、「テトリスの色には、赤、黄、緑がある」において「色」が属性なら、「赤、黄、緑」は属性値となる。

属性値フォーマット：

[属性値, 属性値候補名詞, attribute(属性値を持つ属性), class(属性を持つクラス), [文番号....]]

具体例：

[属性値, 赤, attribute(色), class(テトリス), [1]]

[属性値, 黄, attribute(色), class(テトリス), [1]]

[属性値, 緑, attribute(色), class(テトリス), [1]]

(iv) ロールルール

クラスを表す名詞フレームに動詞が修飾している場合、それらの合成形をロールとする。例えば「待つ人」において人がクラス候補である場合、それに係っている動詞とその名詞の合成語である「待つ人」をクラス人のロールとする。

ロールフォーマット：[ロール, ロール候補動詞+名詞, class(ロールを持つクラス), [文番号....]]

具体例：[ロール, 待つ人, class(人), [1]]

(v) イベントからのルール

「AによるBへのCがあると」、「AはBへのCによって」という文型で「C」部分が「出来事」、「相互作用」を意味素性に持つ名詞である場合、その名詞をイベントとする。例えば「受付係による依頼があると、コンテナを出庫する」において、「依頼」はイベントになる。

イベントフォーマット：[イベント, イベント候補名詞C, 名詞意味素性(出来事／相互作用), sender(クラスA), receiver(クラスB), —, [文番号....]]

具体例：[イベント, 依頼, 相互作用, sender(A), receiver(B), —, [1]]

(vi) 個数／件数ルール

数で修飾された名詞は、その名詞が表すクラスの多重度や属性の限界値や特性値を与えられていると考えられる。例えば、「10銘柄」において「10」は「銘柄」の多重度を表している。

個数／件数フォーマット：[個数, 数値, modified(修飾されたクラス／属性), where(修飾された環境を表すクラス／属性), [文番号....]]

具体例：[個数, 10, modified(銘柄), where(コンテナ), [1]]

(vii) 数ルール

「～数」は「～」がクラスの時、そのクラスの多重度が問われている。「数」そのものはクラスでも属性でもないが、モデル化に重要な要件になっている。

数フォーマット：[数, numOf(クラス候補名詞), [文番号....]]

具体例：[数, numOf(コンテナ), [1]]

2) 動詞に関して

JANUSが出力した動詞意味フレームからは関連やイベント／動作を抽出する。以下、これらを抽出する分析ルールを説明する。分析ルールはそれぞれ図2.9に示すように文法ルール(1)と特定動詞ルール(2)に分けられる。また、各ルールにおいて格の後にくる名詞は主にクラス候補と考えられるが、ロールのことも有り得る。

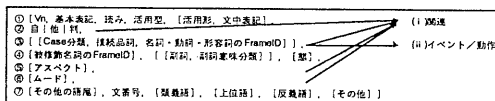


図7 動詞分析ルール

(i) 関連 (集約/冗長関係) ルール

「クラス間に静的関係がある」、「クラス間に継続的關係がある」、「クラス間に参照関係がある」ことを関連という。ルールを状態に関する文法ルール(1-a) (1-b)と所有権に関する特定動詞ルール(2-a)に分け、それぞれ抽出する。

(1-a)状態に関する関連抽出の文法ルール:

動詞のアスペクト語尾が「~ている」(過去形を含む)、ムード語尾が「~できる」の場合、関連があるとす。例えば「AはBとCで話している」、「AはCに混載できる」である。前者はSubject格がAで、Secondary格がBで、At格がCとなり、後者はSubject格がAで、At格がCになる。この場合、関連の一方はSubject格を持つクラスと特定できるが、他のどの格がもう一方のクラスかはこれだけでは不定である。

状態に関する関連抽出フォーマット:

[関連, 関連候補動詞, 状態, subject(クラス候補名詞), -, [その他の格(クラス候補名詞)....], [文番号....]]

具体例: [関連, 話している, 状態, subject(A), -, [secondary(B), at(C)], [1]]

(1-b)状態 (判定) に関する文法ルール:

判定詞を含む文型「AはBだ」、「AはBである」の場合、AやBが共にクラスや属性であるならば、「A isA B」なる冗長関係がある。これを同一視して冗長と考えるか、異なるクラス間の継承と考えるかはこの段階では不定である。

判定に関する抽出フォーマット:

[継承/冗長, 継承/冗長可能性名詞, isA(クラス候補名詞), [文番号....]]

具体例: [継承/冗長, A, isA(B), [1]]

(2-a)所有権に関する集約要素抽出の特定動詞ルール:

Subject格がA、Object格がBの「AはBを持つ」、「AはBをとる」、「AはBを所有する」、「AはBを保持する」、「AはBを所持する」や、Subject格がA、Component格がBの「AはBからなる」、「AはBから構成される」や、Subject格がA、Bの「AにはBがある」、「AにはBが与えられている」などという文型が表れたとき、BはAの集約要素とする。なお、この集約要素は属性になる場合もあり、曖昧性を残したまま、解決を調整ルールにまわす。

このような曖昧性がある場合、CAMEOは、この事実に対して次の2つの出力を行う。

属性フォーマット: [属性, 属性候補名詞, class(属性を持つクラス), [文番号....]]

具体例: [属性, B, class(A), [文番号....]]

所有権に関する集約抽出フォーマット:

[集約, 集約候補動詞, 所有, class(親クラス候補名詞), hasA(子クラス候補名詞), [文番号....]]

具体例:

[集約, 持つ, 所有, class(A), hasA(B), [1]]

(ii) イベント/動作ルール

イベントと動作は表裏一体のものと考えられ、クラスPを中心に考えるとクラスPへ入ってくるpが「イベント」、出ていくqが「動作」となる。逆にクラスQを中心に考えるとqが「イベント」、pが「動作」となる。例えば「rするごとに、sする」、「rすると、sする」、「rすることで、sする」という日本語文型において、rに相当する動詞がイベント、sに相当する動詞が動作となる。

従って、能動形「AがVする」の場合は、VはAの動作であり、受身形「AがVされる」の場合は、VはAのイベントとなる。以降では、例として能動形を用いる。

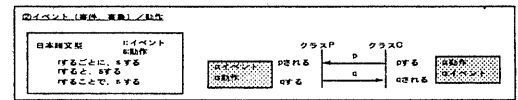


図8 イベント/動作

大別してイベント/動作は「メッセージ」、「信号」、「情報」、「物」を送ることと、状態の変化を表すものに分けられ、以下別々にルールを述べる。

◎「メッセージ」、「信号」、「情報」、「物」を送るイベント/動作は、対象の物理的な位置の異動に関する文法ルール(1-c)、対象の存在する動作に関する文法ルール(1-d)、コミュニケーション(交信)に関する特定動詞ルール(2-b)によって抽出される。

(1-c)対象の物理的な位置の移動に関する文法ルール:

Subject格がAで、Object格がBで、From格がCで、To格がDでTool格がFである他動詞「V」に対し、クラスAによるクラスBを対象とした「V」というイベント/動作がある。例えば、「AはBをCからDへEでFを用いてVする」などのように、対象物が移動する場合である。

対象の物理的な位置の移動に関するイベント/動作抽出フォーマット: [イベント/動作, 候補動詞, 対象移動, subject(クラスA), object(クラスB), [その他の格(クラス)....], [文番号....]]

具体例: [動作, Vする, 対象移動, subject(A), object(B), [from(C), to(D), at(E), tool(F)], [1]]

(1-d)対象の存在する動作に関する文法ルール:

Subject格がAで、Object格がBである動詞「V」に対し、クラスAからクラスBに「V」というイベント/動作がある。例えば、「AはBをCでDを使って保管する」などがある。

対象の存在する動作に関するイベント/動作抽出フォーマット: [イベント/動作, 候補動詞, 対象動作, subject(A), object(B), [その他の格(クラス)....], [文番号....]]

具体例：[動作, 保管する, 対象動作, subject
(A), object(B), [at(C), tool(D)], [1]]

(2-b) コミュニケーション (交信) に関する特定動詞 ルール:

Subject格がAで、Secondary格がBである動詞
「V」に対し、クラスAからクラスBに「V」という
イベント/動作がある。例えば、「AがBに応じ
る」、「AがBにCについてDで話す」、「AがBにC
のことをDを用いて知らせる」、「AがBに連絡する」
などがある。

交信に関するイベント/動作抽出フォーマット:

[イベント/動作, 候補動詞, 交信, subject(クラ
スA), -, [その他の格(クラス)....], [文番
号....]]

具体例: [動作, 話す, 交信, subject(A), -,
[secondary(B), topic(C), tool(D)], [1]]

◎状態の変化を表すイベント/動作には、対象の存在
しない物理的な位置の移動に関する文法ルール(1-
e)、条件に関する文法ルール(1-f)が適用できる。

(1-e) 対象の存在しない物理的な位置の移動に関する 文法ルール:

Subject格がAで、From格がBで、To格がCである
自動詞「V」に対し、自己状態遷移としてクラスA
に「V」というイベント/動作がある。例えば「A
はBからCへ着く」などがある。

対象の存在しない物理的な位置の移動に関するイベ
ント/動作抽出フォーマット: [イベント/動作,
候補動詞, 自己移動, subject(クラスA), -, [そ
の他の格(クラス)....], [文番号....]]

具体例: [動作, 着く, 自己移動, subject
(A), -, [from(B), to(C)], [1]]

(1-f) 状態遷移に関する文法ルール:

Subject格がAで、Object格を持たない動詞「V」に
対し、自己状態遷移としてクラスAに「V」という
イベント/動作がある。例えば、「AはBで開
く」、「AはBで閉まる」などがある。

条件に関するイベント/動作抽出フォーマット:

[イベント/動作, 候補動詞, 状態遷移, subject
(クラスA), -, [その他の格(クラス)....], [文番
号....]]

具体例: [動作, 開く, 状態遷移, subject
(A), -, [at(B)], [1]]

3) 形容詞に関して

JANUSが出力した形容詞意味フレームからは属性
値を抽出する。以下、抽出する分析ルールを説明す
る。

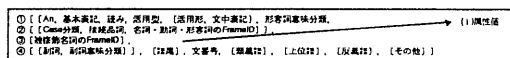


図9 形容詞分析ルール

(i) 属性値ルール

クラスを表す名詞意味フレームに係っている形容
詞は属性値になる。例えば、「赤い花」の場合には、
「花」がクラスであるならば「赤い」はクラス
「花」のなにか特定の属性 (例えば色) の属性値と
なる。また、赤い属性は、他の名詞から抽出され
ることになる。

属性値フォーマット: [属性値, 属性値候補形容詞,
attribute(属性値を持つ属性), class(属性を持つクラ
ス), [文番号....]]

具体例:

[属性値, 赤い, attribute(-), class(花), [1]]

[属性値, 赤い, attribute(色), class(花), [1]]

[2] 名詞調整ルール

ここでは、先の名詞の分析結果において、曖昧性
が存在する名詞に対して、以下の調整ルールを適用
し曖昧性を解消する。

1) 曖昧性解消優先順位

同一名詞に対して、異なる文で違う分析結果があ
る場合、その名詞の分析が曖昧であるとして調整ル
ルを以下の優先順位で適用する。

① イベント/動作のSubject格、Object格、Secondary
格、To格、From格、Sender格、Receiver格になっ
ているものはセンダやレシーバになる可能性が強い
のでクラスである。

② 関連の始点、終点になっているものはクラスであ
る。

③ 属性/集約要素の親になるものはクラスである。

④ 継承の親、子はクラスである。

⑤ イベントになっているものはイベントである。

⑥ 属性のクラスが存在するものは属性である。

⑦ 以上でまとまらない場合、分析ルールで抽出され
たもので多いものを優先させる。

⑧ それでもまとまらない場合、特殊なものを優先さ
せるために、属性値>イベント>ロール>属性>
クラスの優先順位に従うものとする。

2) アクティブ、パッシブの決定

クラス候補に対してアクティブ (センダ、レシー
バ) かパッシブの決定を行う。一般にイベント/動
作の送受信をしているもの (Subject格、Object格、
Secondary格、To格、From格、Sender格、Receiver格)
は、アクティブクラスとする。そうでないクラスは
パッシブクラスとする。

3) 継承/冗長の整理

冗長となったクラスや属性間に対し、それらが
1) の結果、属性と分かったものにおいては、その
代表を見つけ、冗長な属性名は代表で置き換える。
但し、この際冗長な名詞を結合したグラフが木であ
ることが必要である。この場合は根を表す名詞が代
表になる。そうでない場合は利用者に問い掛ける。

