

文脈理解のための概念記述法

石崎 俊 井佐原 均 橋田 浩一 内田 ユリ子 横山 晶一
電子技術総合研究所

本報告では、文脈理解のための概念記述法を提案し、その有効性を示す。概念記述は概念の階層構造を「制約」と呼ばれる概念定義によって詳細に記述し、多言語に基づく概念記述を統一的に取扱うと共に、文脈処理の全過程を「概念の照合」と「概念の展開」の二つのプロセスで取扱おうとするものである。この概念記述法では、概念は、インスタンスの集合であり、各概念は上位概念・下位概念によって階層化されている。また、概念のインスタンスの持つべき属性も指定されている。ある対象について言語表現から得られる情報を視点として関連付け、複数の表現に対応する複数の視点を1つのindexにまとめて記述する方式を採用した。

関係概念の一部である関係子の階層化を試み、関係子の特定化プロセス、多言語間の関係子の統一などについて検討する。

A Method of Concept Description for Contextual Analysis

Shun ISHIZAKI, Hitoshi ISAHARA, Kouichi HASHIDA, Yuriko UCHIDA and
Syouichi YOKOYAMA

Machine Inference Section, Information Sciences Division,
Electrotechnical Laboratory
1-1-4, Umezono, Sakura-mura, Niihari-gun, Ibaraki, 305 JAPAN

A method of concept definition and concept hierarchy construction is proposed. Concepts are defined with "CONSTRAINTS", which describes relations among the concept and other concepts. This method is aiming at building multilingual concept dictionaries using a unified description format.

The context analysis method adopted here uses "index" and "view" to represent an instance obtained from input sentences. A few views are connected to an index when an entity is referred by a few expressions.

Finally, relations such as case, causal, and modifying ones are structured into a hierarchy to extend the system to a multilingual one.

1. はじめに

我々は新聞記事を翻訳する概念レベルにおける文脈情報変換型機械翻訳システムCONTRAST (CONtext TRAnslation)を開発してきた。その開発過程において行った人間の言語理解過程の検討に基づいて、各種概念の表現法を研究している。本稿では、そこで得られた知見を基にした概念表現について述べる。まず、概念として取扱うものを定義し、概念の表現法及びそれらの階層関係について述べる。また、関係子を概念として取扱う手法及びそれによる利点について説明する。さらに、概念のインスタンスの表現法にも触れる。

2. 概念記述

2.1 概念とは

CONTRASTは、CRS (Contextual Representation Structure) という文脈情報を考慮した概念表現構造を用いている。CRSは処理対象言語に応じて導入される様々な概念を互に関連付けることにより、多言語の処理を統一的に取扱おうとするものである。この構造を構成する要素としては、大きく分けて「概念」と各概念に結び付いた「インスタンス」の二つがある。各概念はインスタンスの集合であるとも言える。たとえば、「*弟」という概念は「弟のインスタンス (@弟)」の集合である。本稿では、説明上の混乱を避けるために、概念は「*」で始まる表記（「*弟」）とし、インスタンスは「@」で始まる表記（「@弟」）として、言語の表層の単語（「弟」）と区別する。

概念は予め作られた概念辞書上で上位概念・下位概念による階層構造をなしており、各インスタンスの持つべきスロットや部分構造、先で述べる「制約」が規定されている。インスタンスは一般に、入力文の解析によって概念から導出されるが、固有名詞が指示するような一部のインスタンス（「@マイアミ」、「@USA」等）は予め概念辞書中に登録されている。概念とインスタンスの構成の模式図を図1に示す。

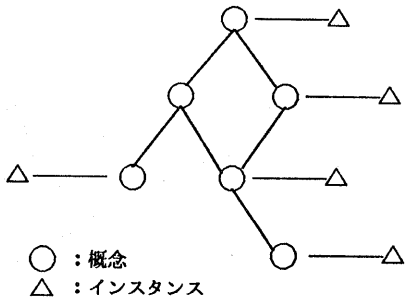


図1 概念とインスタンスの構成

2.2 概念の表現形式

概念辞書の書式を図2に、また、概念辞書の記述例を図3に示す。概念辞書の記述の中で、UPPER-CONCEPTS (上位概念)とLOWER-CONCEPTS (下位概念)とは概念の階層構造を示すために用いられる。LOWER-CONCEPTSには下位概念が並べられていて、互いに背反な概念同士は括弧でくくられている。たとえば、「*人間」において「*男」と「*

女」とは相互に背反であるが、「*男」も「*女」も「*学生」とは背反な概念ではない。SLOTS-OF-INSTANCEには、この概念から導出されるインスタンスが持つ可能性のあるスロット名 (AGENT, OBJECT, LOCATION, TIME など)とスロットを埋める値についての制約条件が記述される。概念からインスタンスが導出されるときに、そのインスタンスはSLOTS-OF-INSTANCEに指定されたスロットを持つ。スロットの情報は概念の階層構造に沿って継承される。したがって上位概念で定義されたスロットは、より詳細な定義をする必要がない限り、下位概念では定義する必要がない。

SLOTS-OF-INSTANCEの中で、キーワード“SCENES”に記述された情報は、この概念 (のインスタンス) が時間関係あるいは因果関係によってどのような概念 (のインスタンス) に分割されるかが示されている。たとえば、「*誘拐事件」の“SCENES”としては、「*誘拐する」「*交渉する」「*解決する」という3つの概念の並びが定義されている。また、キーワード“PART-OF”を用いれば、この概念の部分構造を示すことが出来る。たとえば、「*自動車」の“PART-OF”としては、「*タイヤ」「*ハンドル」等が定義される。なお、SCENESは時間軸による概念の部分構造化であるから、本来、PART-OFに含まれるものである。しかし、文脈処理においては、このようなシーン関係は重要な役割を果たし、頻繁に参照されるために別扱いをしている。

```
(概念 (UPPER-CONCEPT 概念1 概念2 ...)
(LOWER-CONCEPT 概念3 概念4 ...)
(SLOTS-OF-INSTANCE
 (スロット1 条件1)
 (スロット2 条件2)
 ...)
(SCENES シーン1 シーン2 ...)
(PART-OF パート1 パート2 ...))
(CONSTRAINT ...))
```

図2 概念辞書の書式

```
(*人間 (UPPER-CONCEPT *哺乳類)
(LOWER-CONCEPT (*男 *女) (*学生) (*犯人))
(SLOTS-OF-INSTANCE
 (年齢 時間) (名前 文字列) (国籍 国名))
(*犯人 (UPPER-CONCEPT *人間)
(CONSTRAINT ((@@ (CONCEPT = *人間)
 (@1 (CONCEPT = *罪を犯す)
 (@2 (CONCEPT = *行動する)
 (動作主 = @@))
 (@3 (VIEW = (@1 @2)))))))
(*誘拐する (UPPER-CONCEPT *罪を犯す)
(SLOTS-OF-INSTANCE
 (対象 *人間)
(SCENES (*捕まえる)
(*運ぶ)
(*監禁する))))
```

図3 概念辞書の記述例

2.3 インスタンスの表現

以上述べてきたような概念体系から、実際の文章を解析する過程で順次インスタンスが導出されてゆき、全体の文脈表現構造が抽出される。この過程で、事象や物に関する1つの対象を幾つかの概念のインスタンスの組合せで表現する必要が生じる場合がある。ここで「対象」とは文脈表現構造が表現しようとするものであり、物概念においては、その外延となる。ここでは、複数のインスタンスのそれぞれがその対象を別の視点 (view) から説明しているものと考え、一つの対象に対する視点 (言い替えると、一つの対象を説明している情報) をまとめて指す節点として index という節点を導入した。ここで、視点 (view) は複数のインスタンスの組合せで表現されるような対象の持つ情報を、各インスタンスごとに切り出した切り出し方を意味している。また、index は、視点の一部あるいは全部を指す節点であり、これにより対象の一部を特定して言及することが可能になる。外部からの対象の参照は index あるいは view に対して行われる。「夫であり、かつ大使である外延」を示す表現を図4に示す。ここで、「@夫」と「@大使」とは、ある外延をそれぞれの視点から説明していることになり、それら全てを指す index 節点が、全ての視点からの情報を含んだ外延全体を示している。

既に述べたように、どのようなスロットを持ちうるかは概念ごとに特定されており、それらは概念の階層構造に沿って継承される。index によって指される複数の view ができた場合には、各概念の特定するスロットはそのインスタンス中に記述される。たとえば、「学生の田中さん」という表現からは「@学生」が導出されるが、同時に「*学生」の上位概念である「*人間」のSLOTS-OF-INSTANCESで指定される「名前スロット」に値を入れることが要求される。このような場合には、それを指定する概念のインスタンスを (もし未だインスタンス化されていないならば) 新たに作り出し、そこに記述する。ここで、既にあるインスタンスと新しく作ったインスタンスとが同一の対象を説明していることを示すために、index 節点がいられる (図5)。

このような場合に、このスロットを既に導出されているインスタンスに書き込む方法と、そのスロットを指定している概念のインスタンスを新たに作り出して、そのインスタンスの中にスロットを記述する方法とが考えられる。しかし、既に導出されているインスタンスに書き込む方法では、既に複数のインスタンスが存在している場合に、それらの内のどのインスタンスにスロットを書き込むかという曖昧性が起こる。

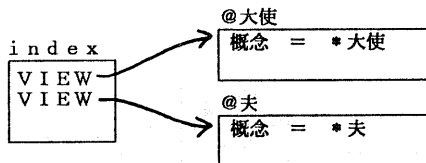


図4 「夫でありかつ大使である外延」

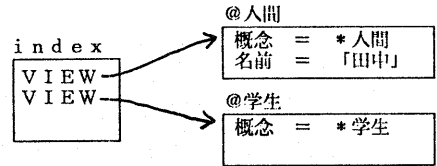


図5 スロットが記述されるインスタンス

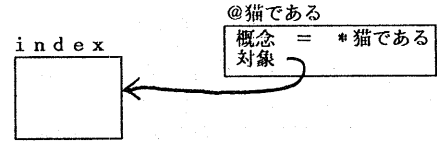


図6 「猫である」index

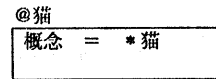


図7 「猫」の概念表現

3. 概念とインスタンス

3.1 概念をどう表すか

概念としてどのようなものを採用するかには任意性がある。概念は物概念と事象概念に、事象概念はさらに動作概念と状態概念とに分けられるが、物概念を「物」として概念化する方法と「物である」という状態概念を用いて概念化する方法とが考えられる。CONTRASTでは、物概念を用いて表現している。たとえば、「猫」という単語から導かれる情報を表現するには (a) 「*猫である」

(b) 「*猫 (という対象)」の2つの概念をそれぞれを用いて表現することが可能である。図6に (a) の場合の表現を示す。ここでは、全ての具象物は「*~である」という状態概念の下位概念として表されている。左側の箱は「猫」の外延を示す index 節点である。「である」型表現では全ての具象物は (それが単独の概念で説明されようとして複数の概念で説明されようとして) 外延を示す節点とそれにリンクを張る「*~である」との組合せで表現される。

図7に (b) の場合の表現を示す。ここでは、「*猫」は「猫という対象 (具象物)」を示すのであるから、「*猫」という概念のインスタンスだけで、「猫」の持つ情報を全て示していることになる。したがって、ある具象物が猫であることだけを表現すれば良い場合には、「対象」型の表現が効率の良い表現となる。一般に、ある対象を単一の視点からのみ表現している場合には、「*~である」+「index 節点」よりも効率がよい。また、複数の視点から表現されている場合には、(a)、(b) の二つの表現はリンクの方向が逆転しているだけで、等価なものとなる。図8に示す「夫である大使」の「である」型表現を図4と比較されたい。

したがって、(b) 型の、すなわち対象そのものを概念とする方法を採用する方が全体としての効率は高くなるので、CONTRASTでは「対象」型表現を採用している。

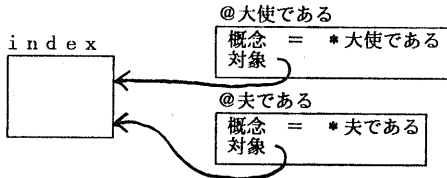


図8 「夫でありかつ大使である外延」の「である」型表記

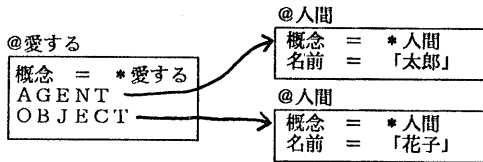


図9 スロットによる表現

3.2 関係概念について

概念には既に述べたように物概念と事象概念とがある。事象概念は動作概念と状態概念に分かれる。状態概念はさらに関係概念と呼ばれるものを含み、このインスタンスは概念のインスタンス同士の関係を定義する。「*動作主 (agent)」や「*対象 (object)」といったスロットも概念の一つとして定義されている。関係概念の内、このように二項関係として定義されるものは、そのインスタンスが結び付ける2つの概念のうちのいずれかにスロットとして埋め込むことが可能である(図9、10)。このように二項関係を示す概念は、別のインスタンスから指されていない場合には、外部にインスタンスとして表現することも、内部にスロットとして表現することも可能である。CONTRASTでは、スロット表現はインスタンス表現の簡略表記と考え、スロット表現による表現も認めておき、必要に応じてスロットを外部に取り出すこととしている。

このようにスロットで示される2項関係をも概念化することにより、先で述べる視点や関係子の表現が統一的に取扱えるようになる。

3.3 視点の役割

視点を含むindex表現は、ある対象を指定するために幾つかの情報を用いていることを示すものである。しかし、一連の文章のなかで、このような情報の全てが常に用いられているわけではない。一つの対象の持つ情報のうちの一部分だけに対する言明を行うことがある。(「金星」「明けの明星」「宵の明星」などがその例である。)このような場合には、必要な情報だけをリンクでつないだindex節点を作りだして、その節点を指すことにより、ある対象の一部に対する言明の指示先とすることができる(図11)。

なお、「対象」型の表現を用いた場合に「～である」という事象を示したい場合には、index節点がある特定の視点を含んでいることを示すことになる。実際には、index節点中のVIEWスロットを「*VIEW」概念のインスタンス「@VIEW」として外部に取り出し、それを指すことによりこれが示される(図12)。筆者ら

の用いる概念体系においては、既に述べたようにスロットによる関係記述も概念として取扱うためこのような表現が可能となっている。

4. 制約

4.1 概念の細分類

以上述べたように「対象」型の概念を採用することと決定しても、それらの階層構造をどのように定めるかという問題が残っている。ある概念をより詳細な幾つかの分類に分けようとする場合、下位概念として分類する方法と、その概念のインスタンスの持つ属性値によって分類する方法とが考えられる。図13に下位概念による分類を図14に属性値による分類を示す。これらは表現能力は等価であり、どのレベルまでを下位概念とし、どのレベルより詳細な分類は属性値で分けるかということ、明確に理由付ける根拠はない。しかしながら、意味的な制約条件を各スロットに手続的に記述するよりも、概念間の制約として宣言的に記述しておく方が、処理過程に一貫性を保てる。そこで、筆者らの概念体系では、スロットの値が(固有名詞で示されるような)特定の個体を示すようなもの(たとえば、事象概念における「動作主」や「場所」等)以外は、下位概念化して細分類している。

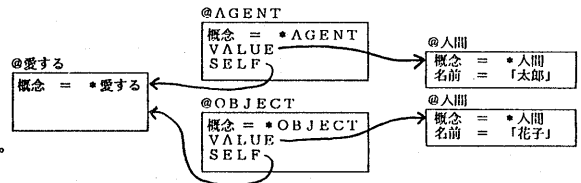


図10 概念のインスタンスによる表現

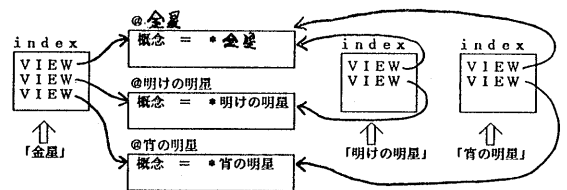


図11 indexによる部分的な言明の処理

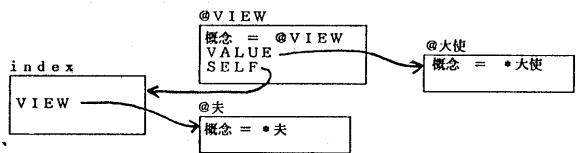


図12 「大使である」ことの表現法

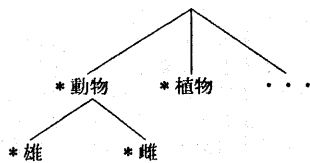


図13 下位概念による分類

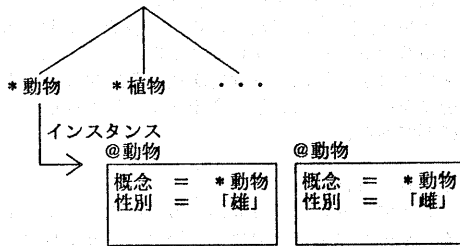


図14 属性値による分類

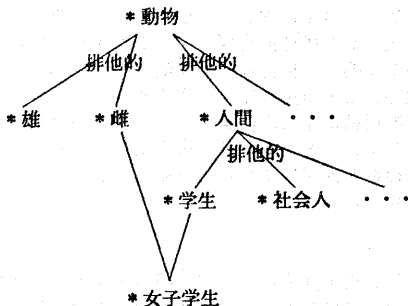


図15 複数の上位概念による指定

各下位概念はその上位概念を何らかの理由で分類したものであるが、このような分類の理由を表現する方法としては、(1) 複数の上位概念によるものと(2) 概念間の制約記述によるものがある。(1)は、図15に示すように、複数の概念のANDによって下位概念を指定するものである。ここでは、「*女子学生」が「*女性」と「*学生」のANDで表現されている。(2)は図16に示すように、下位概念が上位概念の内のどのような制約を満たすものであるかを記述するものである。ここでは、「*弟」が「*兄弟」のうちで「*年少」のものであるということが示されている。ここで、「*兄弟」は兄弟関係ではなく、兄弟かかのいずれかである実体を示す概念である。なお、「*年少」という概念は物概念ではなく関係概念であり、そのインスタンスは他のインスタンスと結び付いて初めて意味を持つ。したがって、このような制約を複数の上位概念を指定するだけで表すことは困難である。

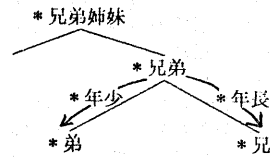


図16 制約記述

4.2 制約記述

概念の上下関係が示す情報を記述するには、上下関係の構造を示すだけでは十分でなく、その上下関係がどのような理由によって存在しているのかを記述しておくなくてはならない。既に述べたように、「*弟」は「*兄弟」の下位概念であるが、「*弟」は「*兄弟」の内で「年齢がより低い者」であるというような情報は、「制約」と呼ばれる概念間規則で概念辞書中に記述する。制約には上位概念と下位概念との関係やスロットとそのスロットを占めるものとの関係(たとえば、LOCATIONスロットを埋めるものは「場所」の下位概念でなくてはならないといった制約)が記述される。「制約」はその概念のインスタンスを上位概念のインスタンスと他のインスタンスを用いて表現する形に変換した場合に作られる表現全体を、中間表現と同一の書式で記述してある。図3中に「*犯人」に関する記述の中に「制約」が示されている。

ここで、「(@@ (概念 = *人間))」という記述があるが、「@@」は特別な予約語であり、この概念(「*犯人」)のインスタンス(「@犯人」)を別の表現で表現しなおした場合の人間という概念(*人間)のインスタンスを示している。「@1」「@2」に続く記述は、必要に応じて、より詳細な情報記述を得るために新しく作られるインスタンスを示しており、それら相互の接続の仕方が予め、スロットの形で記述されている(例:「(動作主 = @@)」。「@3」に続く部分は、indexを示しており、「@1」と「@2」が別の視点から同じ対象を修飾するものあることを示すものである。

5. 関係子の階層化の試み

関係概念は状態概念の下位にあり、物、状態、動作に関する概念およびインスタンスの間に成り立つ関係の集合である。ここで扱う関係は表層関係ではなく、文脈処理に有効なレベルの深層関係である。

主な関係概念には、格関係、変化関係、事象関係、限定関係がある。それぞれの関係概念は図17に示すように多くの下位関係概念を包んでいる。

関係子として特に定義する関係概念は、自然言語処理関連システム(特に機械翻訳システムなど)で一般的に多く使われるものを、日本語、英語、中国語を考慮して選択整理したセットである。

従来の関係子セットは、一般に階層構造を持たず、フラットなものであったが、本論文では次のような理由により関係子の階層化を試みた。

- (1) 文脈解析で関係が確定化されるプロセスにおいて、関係子の階層関係を用いることができる。

- (2) 多言語間で関係子の統一を図る時、
 (a) 言語毎の関係子の詳細性の違いの明確化
 (b) 関係子の追加、変更
 が容易になる。
- (3) 言語理解技術のレベルに合わせて、関係子の詳細性(階層構造の深さのレベル)を設定できる。

ここで、(1)に関しては、任意の関係を示す“一般関係”を考慮することが便利であろう。これは二つの概念の間に、事象が介在するような関係も許し、特定の関係子に帰着できないものを含むようにする。

(3)に関連するが、この階層化した関係子で、最も詳細なレベルを使う必要は必ずしもない。翻訳に利用する時は対象言語における必要性にあわせて詳細化のレベルを変えることを考えている。

従ってこの階層化の試みでは、なるべく多くの関係子を列挙する方針で行っている。

5.1 格関係

一般には次に述べる変化関係も含める場合が多いが、ここでは動作主、対象、利用具、時、場所として階層化を試みた。経験者は対象(object1)に含めても良い。変形材料と非変形材料の細分類は英語のmake-ofとmake-fromを考慮している。

5.2 変化関係

ここでは階層化というよりもテーブル形式に分類している。縦軸は変化の始点、終点、経点であり、横軸は変化の対象となる時刻、場所、属性・状態、権利、情報伝達、制約・制御である。時刻、場所などの横軸をすべてまとめて、関係子として変化始点、変化終点、変化経点だけにするこゝとや、時刻と場所以外の変化の対象を一つにまとめるこゝとも考えられる。変化に関する量は4.4の程度に含まれる。

5.3 事象関係

主として事象の間の関係を扱うが、論理関係では物も対象とする場合がある。結果は原因・理由と対になるので省略可能である。排他選言は複数個の中から1つだけ選ぶこゝとしてしている。

5.4 限定関係

一般の修飾、非修飾関係の中でも、対象の状態、動作、変化等について限定的に修飾するものを階層化している。

5.5 文脈処理における関係子の確定化プロセス

階層化された関係子を用いると、曖昧な関係を先で確定化するという処理を、物や事象の確定化と同様な文脈解析プロセスとして処理できる。

このような処理は、事象概念に対する係り受け関係の決定においても同様に行いうる。埋込文や副助詞「は」の使用によって、係り受け関係を決定する場合に助詞の情報が利用出来ない場合がある。たとえば、「誘拐した人」という表現からは、この「@人」が「@誘拐する」の行為者であるのか対象であるのかが決定出来ない。このような場合に、まず「*行為者」と「*対象」との共通の上位概念で、「@人」と「@誘拐する」とを結び付けておき、先の処理で確定した段階で、「*行為者」あるいは「*対象」に置き換えれば良い。

しかしながら、このような処理が常に関係概念の範囲で閉じているわけではない。「ライト氏のホテル」という表現から、「*人間」と「*ホテル」とを関連付ける曖昧性のある関係概念を導いておき、解釈が進むに連れて、より詳細な関係概念を決定する場合に、詳細な関係概念としては、「*所有者」「*宿泊客」「*設計者」などがあげられるが、これらの内の「*宿泊客」「*設計者」等は、関係概念とみなすよりは、「*宿泊する」「*設計する」といった事象概念とみなす方が自然である。このような場合には、概念の階層構造と、概念のインスタンスのスロットに埋め込まれた情報とから、文脈解析における単純な「情報の詳細化」のプロセスで取扱うことは出来ない。関係概念と事象概念とを制約記述で関連付けておき、必要な変換を行うことにより、詳細な意味表現構造を得る処理が必要である。

5.6 まとめ

ここで提案している関係子の階層化は0次案である。ある一定の関係をどの関係子に帰着させるかについては今後の多言語を考慮した幅広い観点からの議論が必要である。

6. おわりに

以上述べてきた概念表現法は、機械翻訳システムCONTRASTで用いるためにインプリメント中である。CONTRASTの文脈解析部において、概念階層のチェックを主とした概念照合プロセスが稼働している。index表現に関する部分は一部稼働しているが、日本語一概念変換辞書との対応に関する変更を加えている。今後、関係子の階層構造を実際のシステム上で有効に用いる手法のインプリメントと、概念辞書の各項目の詳細化、制約記述に一般の命題が記述できるようにすることなどが、今後の予定である。

参考文献

- [1] 石崎・井佐原・内田・橋田・元吉:「機械翻訳システムCONTRASTにおける文脈情報の利用」情報処理学会第33回全国大会 1986
- [2] 井佐原・橋田・内田・石崎:「機械翻訳システムCONTRASTにおける概念変換について」情報処理学会第34回全国大会 1987
- [3] 石崎・井佐原:「文脈と言語理解」電子通信学会技術研究報告NLC86-4 1986
- [4] 井佐原・橋田・内田・石崎:「文脈解析における概念照合のための推論と制約記述について」情報処理学会第35回全国大会 1987
- [5] フィルモア、田中春美訳「格文法の原理」三省堂
- [6] CICC中間言語ワーキンググループ関係子資料 1987
- [7] 日-英科学技術用語研究データベースの開発に関する報告書 1986.3
- [8] 樋口、戸田「会話文処理用情報格構造」in「認知理論に基づく社会的相互作用過程の解明」1987.9
- [9] D.Dong, Z.Deling, "KY-1 MT System and Some Linguistic Aspect", Proc. of MT Meeting in Peking 1987

格関係 (case-relation)	動作主 (agent0)	動作主 (agent1) 随伴主 (coagent) 経験者 (experiencer)	限定関係 (modification)	役割・視点 内容 (content)	役割(role) 視点(viewpoint)		
	対象 (object0)	対象 (object1) 随伴物 (coobject)		比較 (comparison0)	比較(comparison1) 等価・同格(equibalance)		
	利用具 (implement)	道具 (tool) 手段・方法 (method) 材料 (material)		程度 (degree0)	程度(degree1) 頻度 特定頻度 (frequency) 一般的頻度		
	時 (time0)	時刻 (time1) 時間 (duration)		数量 (amount)	数値(number) 単位(unit)		
	場所 (place0)	場所 (place1) 対象場所 (object-place)		様態 (manner)			
				範囲 (range)			
				包含 (inclusion)	部分(part-of) 要素(element) 所有(owner)		
変化関係 (transition)	変化始点 (source0)	時刻始点, (time-from)	場所始点, (space-from)	属性・状態始点, (source1)	権利始点, (source2)	情報伝達始点, (source3)	制約・制御始点 (source4)
	変化終点 (goal0)	時刻終点, (time-to)	場所終点, (space-to)	属性・状態終点, (goal1)	権利終点, (recipient1)	情報伝達終点, (recipient2)	制約・制御終点 (recipient3)
	変化経点 (through0)	時刻経点, (time-th)	場所経点, (place-th)	属性・状態経点 (through1)			
事象関係 (event-relation)		前後関係 (sequence)	因果関係 (causal)	条件 (condition0) 結果 (result) 目的 (purpose)	条件(condition1) 原因・理由(cause)		
			同時 (concurrency)				
		論理関係 (logic)	連言 (and) 選言 (or)	順接(successive) 逆接(consessive) 選言(or) 排他選言(exclusive-or)			

図 1 7 関係子の階層化の一例