

語彙の概念と知識について

牧野 武則
東邦大学理学部

一般性がある自然言語理解や問題解決のためには、現実の世界で成り立つ知識を構造化したオントロジカルな知識ベースが必要である。この論文では、単純命題としての概念の成り立ち、複合命題としての知識、及びオントロジュカルな知識について述べ、それらの知識を集積する大規模知識ベースについて考察する。自然言語の語彙が持っている概念の集合が人間が認知できる事象や事物の概念の集合であるとし、それらの概念で現実世界の知識が記述されると考える。単純命題を単一の事象概念を含むものとし、事象概念を概念素の分解し、概念素の物理的、論理的な成立を明らかにし、単純命題の構造について議論する。そして、自然言語で記述されたドキュメントから知識を抽出し、大規模なオントロジカルな知識ベースの構築について考察する。

On Knowledge and Concept of Lexicon

Takenori MAKINO
Faculty of Science, Toho University
2-2-1 Miyama, Fumanashi 274 Japan

For natural language understanding and problem solving with generality, it is essential to provide an ontological knowledge base which contains knowledge in practical world. As such knowledge is described by lexicons of natural language in practical documents, it is possible to represent all concepts of events and things in the practical world as the concepts of lexicons. Simple proposition include the single event concept, and the knowledge is represented as compound proposition combined the simple proposition by the relation between the event concepts. The event concept is decomposed into conceptual primitives which has the condition that the primitives hold logically and physically. The conceptual primitives have contextual information which reflects the compound propositions. Based on the concept structure, the ontological knowledge base architecture is discussed.

1.はじめに

言葉は人間が認知しうる事象や事物を説明する最も強力なメディアであり、人間の知的活動の源でもある。自然言語理解システムを構築する動機の一つは、言葉が持っている事象や事物に対する知識表現能力と、その知識表現能力をもとにした問題解決能力を計算機の上に実現することにある。

自然言語処理は既にワードプロセッサから機械翻訳など様々な応用システムを生んできたし、また知識処理もエキスパートシステムなどの実用的な応用を提供してきた。しかし言語の意味や知識の内容に係わる応用システムの開発や使用経験の蓄積と共に、それらのシステムの脆弱さが一般的な知識、常識に関する知識の欠如によるものであるということが認識されるようになってきた[1]。

言語や知識を扱うシステムの脆弱さを解決するために、大規模な知識ベースの開発が行なわれつつある。自然言語理解のための知識ベースの開発がEDRで約50万の概念について進められており[2]、一般の問題解決のための知識ベースの開発はMCCのCYCプロジェクトで進められている[3]。ところが、現在の段階ではEDRプロジェクトでは一般の常識の蓄積までは考えておらず、一方、CYCプロジェクトは言語の意味に関しては注意を払っていない。

こうした知識ベースで、世の中の全ての命題公理、理論を書き出すことができるだろうか。数学での数学公式や化学変化の公式では根拠が明確であるが、一般の知識については、公理や理論を導く根拠が明白ではない。しかしドキュメントの中では一般の知識について自然言語を用いて記述されており、自然言語の記述能力がその条件を備えていると考えるのは不自然ではない。

ここでは、人間が認知しうる現象は自然言語によって表現されることを前提とし、自然言語の意味と知識は連続していると考え、この仮説に従えば、知識の表現は、自然言語の理解と関係付けられ、より一般的な知識の表現へ道を開くことが可能になる。

この論文の目的は、自然言語の語彙が表している概念と、文(命題)として表される知識の間の関係を与えることにある。

意味や知識に関する研究は、形式論理学や意味論の中で精力的になされてきた。最近の努力は、自然言語や多様な知識を表現するために自然論理

の構築に向けられている。命題論理が文としての知識の真偽を議論するのに対して、述語論理では文を構成する語の間の関係について議論してきた。しかしそれらの論理では、例えば自然言語の意味内容を形式化するには制限が大きいことが知られており、様相論理[4]や時制論理[5]など自然論理[6]を目指した論理が提案されてきた。しかし可能世界や状況を前提にした論理は記述能力の点ではかなり進歩してとはいえ、一般の問題解決のための知識の表現から見ると知識表現の複雑さを増加させている。

ここでは、自然言語の一部の現象について、知識を自然言語の語彙の概念によって記述される命題としてとらえ、オントロジーとよばれる一般的な知識の表現を考察する。

まず語彙の概念を分析することによって、概念の成り立ちについて述べる。とくに概念の構造を状態と状態変化に分けてその間の関連を述べ、概念が成り立つ条件について調べ、単純命題としての概念の構造について述べる。それをもとに論理的に文の概念として表現される知識について定義する。

そして最後に、現実世界での知識からなるオントロジカルな知識ベースの構築について考察する。

2.概念と知識

知識を記述する際、その知識が表している事象や事物を指し示すために自然言語の語彙を用いるか、自然言語で定義された記号が用いられる。自然言語のみならず、知識を表現するプログラムや数式などの形式言語においても、知識表現言語においてもそうである。つまりあらゆるドキュメントにおける知識表現において、自然言語の語彙がそれ以上説明を必要としない言明あるいは終端記号となっている。このことは知識ベースの構築において語彙、さらに言うところ語彙の概念が定義されていなければならないことを示している。

概念とは、人間が認知しうる事象や事物に対して、それらの特徴に対して与えられたシンボルであり、すべての事象や事物はそれらに対応する概念を持つ。そしてそれらの概念に対して語彙が与えられている。ここでは、自然言語として表現できる事象や事物を対象とし、概念の集合は語彙が表わす概念の集合とする。

知識は、適切な概念の組み合わせで表現される

命題である。ある一つの事象を表わす概念の集まりがもっとも基本的な単純命題に対応する。それらの単純概念が事象間の関係によって結合され複合命題を形成する。知識ベースはそれらの命題としての知識の真偽を記述したものである。ここで知識の真偽について次のように分けて考える。

- 1) 語彙の概念の知識：語彙によって表わされる概念が成り立つ条件が記述される。
- 2) 概念を組み合わせた知識（単純命題）：知識の記述は概念が論理的に成り立つための条件で束縛されており、その条件のもとで知識が記述される。
- 3) オントロジカルな知識：宇宙空間やおとぎ話の世界での真偽を扱うわけではなく、ごく普通の実世界における知識の真偽を記述する。

Fig. 1 に示すように、知識の構造は、語彙と概念の関連に関する知識、概念に関する知識、減事実世界での文（命題）に関する知識からなると考える。例えば「住む」という概念を例に上げると、「町に住む」という現実世界で真である知識がオントロジカルな知識であり、それに対して、現実世界ではありえない「本に住む」という自然言語で表現できなくはない比喩に当たる知識や「ペンギンが飛ぶ（飛ばうとした）」といった願望を含む知識はオントロジカルな知識ではなく、区別される。また、「町から住む」という文にもならず概念として成り立たない知識も区別することにする。

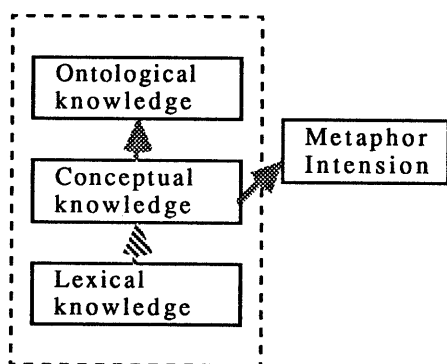


Fig. 1 Knowledge structure

3. 概念の成り立ち

概念の意味とその構造は、単純命題の意味を与える。

概念を、事象概念と事物概念、それらの属性概念に分ける。特に断らないかぎり、議論を簡単にするために事象概念と事物概念を対象とする。

「行く」という語彙の場所移動としての概念は、移動するもの、移動元、移動先が存在して始めて成り立つ概念である。このうちどれが欠けても論理的に場所移動の概念は成り立たない。格文法[7]で与えられているように、ある事象概念 C_e について、相手概念への概念関係子（格）、 r_1, \dots, r_n 、を与え、次のように記述する。

$$C_e(r_1, \dots, r_n) \quad (1)$$

ここで、 C_e は事象概念であり、事象概念が与えられれば関連する概念関係子のセットは規定される。

それぞれの概念関係子に対して適切な他の概念が結び付いて文としての知識を形成する。次のように記述する。

$$C_e(r_1/C_1, r_2/C_2, \dots, r_n/C_n) \quad (2)$$

ここで、 C_1, C_2, \dots, C_n は事物概念あるいは事象概念である。

格文法では、関係子のセットをすべての概念について一様に定義しようとしてきた。しかしその試みは、多くの改良の積み重ねにも関わらず、いまだに決定版は存在していない。この原因は、後で詳しく議論するように、概念には多様な観点があり、それらの観点到に依存して関係子の意味が異なってくることによる[8]。

ふたたび「行く」という概念を考える。この概念に関係する関係子の中で、明らかにこの概念に固有の関係子と、そうではなく他の概念にも関係する関係子のセットがある。「行く」という概念の移動するもの(obj)、移動もと(source)、移動先(goal)は「行く」という概念に固有である。他の概念についても、例えば、所有移動の概念である「与える」の概念関係子、与えるもの(obj)、与える人(source)、もらう人(goal)と概念関係子の意味は概念に固有である。ここで「行く」の概念関係子と似ていることに注意する。それらに対して事象が生起する時間(time)や場所(location)などの概念関係

子は概念によらずどの概念でもその意味は同じである。これらの概念関係子のほかに事象概念と事象概念の間の関係を示す因果関係子がある。

ある事象概念の関係する概念関係子を次のように分類する。

- 1) 事象概念に依存する関係子 : { p_i }
- 2) 事象概念にわたって共通の関係子 : { r_j }
- 3) 事象概念間の関係を示す関係子 : { s_k }

そして、(1)に上げた事象概念と概念関係子の表現を次のように書く。

$$Pe : Ce(\{p_i\}), \{r_j\} \quad (3)$$

事象概念の関係は、事象概念を含む単純命題を Pe_1, Pe_2 とすると、

$$s_k(Pe_1, Pe_2) \quad (4)$$

ここで(3)が単純命題の対応している。複合命題は(4)で示すように(3)の形式で記述された命題間の結合、つまり事象間の関係子 s_k で結び、与えられる。また、(2)に対しては、記述の便宜のために次のように記述する。

$$Ce(\{C_i\}), \{r_j(C_j)\} \quad (5)$$

ここで、 p_i は概念により固有であるために省略している。

次に事象概念に依存する関係子の成り立ちを調べるために、概念の構造について述べる。

変化の概念

すでに Anderson[9] が語彙分解で導入した静と動の考え方をここでも用いる。事象概念は状態を表わす状態概念とその状態の変化を表わす状態変化概念に分ける。つまり状態変化概念は状態概念に変化の概念が付加されて形成されると考える。

例えば、「在る」や「いる」に代表される場所存在の概念 EXT(obj, location) に対して、location に変化の概念 TRANS(source, goal) が付与されると、場所移動 EXT-TRANS(obj, source-loc, goal-loc) が形成される[10]。

この変化の概念の導入は、他の状態変化の概念の成り立ちを説明するうえで役立つ。所有の概念

の POS(obj, possessor) に変化の概念を possessor について付与すると、「与える」などの概念である所有移動 POS_TRANS(obj, source-pos, goal-pos) が得られ、「知っている」などの知識所有の概念である INF(content, possessor) から「教える」「知らせる」の概念である情報移動の概念 INF_TRANS(content, source-pos, goal-pos) の概念が生まれる。

したがって、事象概念について状態を示す概念を整理し、それらの状態概念について変化の概念を付与することで状態変化の概念が形成される。ここで重要なことは、状態概念から派生した状態変化概念の関係子はシステマテックに状態概念から得られることである。ここに概念素(プリミティブ)を導入する可能性と利点が見られる。

概念素(プリミティブ)

概念をそれ以上分解できない概念素の組み合わせによって表わすことには批判があるが[11]、概念の成り立ちを考えるうえではそれらの批判を考慮しても有用である。

ここでは、次のような概念素を用意し、それらの組み合わせについて議論をする。これらの概念素がすべての概念を記述するわけではないが、状態概念を設定することで同様に他の概念も説明できる。

状態概念

- 場所存在 : EXT(obj, location)
- 所有 : POS(obj, possessor)
- 情報所有 : INF(content, possessor)
- 感覚 : MEN(content, possessor)
- 属性 : ATT(obj, value)

変化概念

- TRANS(source, goal)
- アスペクト概念 ASP(aspect)
- 使役概念 CAU(causer)

状態概念と変化概念の組み合わせで次の概念が形成される。

- EXT-TRANS (場所移動)
- POS-TRANS (所有移動)
- INF-TRANS (情報移動)
- MEN-TRANS (感覚変化)

ATT-TRANS (属性変化)

 注意: 「歩く」という語彙の概念では、「公園を歩く」という意味の場所存在と示す状態概念と、「東京から大阪まで歩く」という場所移動を示す状態変化概念が区別される。

変化とアスペクト

状態を変えるのは、変化概念だけではない。アスペクトの概念によっても変化する。

アスペクトによる状態変化は、Fig.2 に示すように状態の時間的変化を伴う。変化の開始、継続、終止を示すのがアスペクトである。

一般に「開始する」、「続ける」、「終わる」の語彙の概念がアスペクトであり、「講義を始める」という文のように、状態概念に付与されてアスペクトが表わされる。そのほかにも「林檎が赤い」は状態、「林檎が赤くなる」は状態変化、「林檎が赤くなりだす」はアスペクトが開始を表わす。

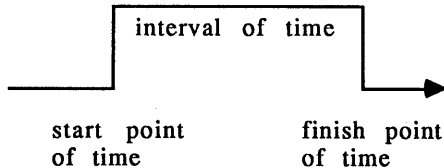


Fig.2 State changing of time

場所存在の概念「いる」は状態、その状態変化の概念は「行く」、「行く」のアスペクトが開始のとき「出発」、終止のとき「到着」の概念が対応する。一方場所存在の概念に開始のアスペクトが付与されると「現われる」といった概念が形成される。

格文法で様態格、mannar、とされている関係子のうち時間に関係する関係子は、アスペクトと密接な関係を持っている。このことは、関係子はそれぞれ独立ではなく、依存関係を持っていることを示唆している。

変化概念とアスペクトの概念を説明するために時相を考える。ある時間 t について、 $+t$ を t よりも前の時間、 $+t$ を t よりも後の時間、そして $t1$ と

$t2$ の時間間隔を $t1: t2$ とする。

アスペクトは、時間上の状態の変化であり、場所存在「いる」の開始を表わす「地点 A に現われる」は、

$EXT(obj, A). TIME(+t)$

そして、「去る」という終止の概念は、

$EXT(obj, A). TIME(+t)$

また、時間 $t1$ から $t2$ までいるという継続の概念は、

$EXT(obj, A). TIME(t1: t2)$

と表わすことにする。

場所移動の概念「行く」は、ある時間 $t1$ より前は、地点 A にいて、その後地点 B に存在場所を移すことを意味している。つまり、

$EXT-TRANS(obj, A, B). TIME(t1)$

は、分解すると、

$EXT(obj, A). TIME(+t1) \&$

$EXT(obj, B). TIME(t1+)$

である。記号 $\&$ は事象間の関係を示し、後で議論をする。

ここでさらに地点 C へ時間 $t2$ に移動したとすると、

$EXT-TRANS(obj, B, C). TIME(t2)$

つまり、

$EXT(obj, B). TIME(+t2) \&$

$EXT(obj, C). TIME(t2+)$

と表わされる。

この2つの事象をマージすると、

$EXT(obj, A). TIME(+t1) \&$

$EXT(obj, B). TIME(t1+) \&$

$EXT(obj, B). TIME(+t2) \&$

$EXT(obj, C). TIME(t2+)$

ここで同件事象 $EXT(obj, B)$ について $TIME$ の演算を定義し、 $TIME(t1+) \& TIME(+t2)$ を $TIME(t1: t2)$ と書き換えると、

$EXT(obj, B). TIME(t1: t2)$

無論ここで $t1 < t2$ でなければならない。Fig.2 に状態変化の様子を示す。

このように状態変化の概念には、 $TIME(+t)$ や $TIME(+t)$ 、 $TIME(t1: t2)$ のアスペクトに関する概

念が含まれている。

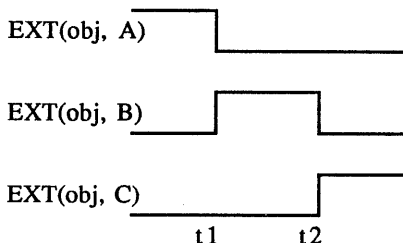


Fig.3 State changing in EXT-TRANS

属性、量、質の変化

事物は属性を持ち、属性の変化を伴う。これらの属性や量、質の変化も変化概念と同じように表現される。例を用いて説明する。

例えば「林檎が10個ある」は量を示す概念 QUAN を用いて次のように表わされ、「林檎が10個から20個に増加する」は、状態変化概念である QUAN-TRANS で表わされる。

量 : $QUAN(obj, value)$

増加 : $QUAN-TRANS(obj, val1, val2). TIME(t)$

この概念を分解すると、

$QUAN(obj, val1). TIME(+t) \&$

$QUAN(obj, val2). TIME(+t)$

一般の属性について同様に、

$ATT(obj, value)$

の時刻 t における状態変化は、

$ATT-TRANS(obj, val1, val2). TIME(t)$

と表現され、この概念を分解すると、

$ATT(obj, val1). TIME(+t) \&$

$ATT(obj, val2). TIME(+t)$

である。

前提を必要とする概念

「帰る」とか「戻す」の語彙に代表されるように、この概念が成り立つためには前提を必要とする概念がある。これらの概念に対しては、概念の分解の際に、その前提が記述される。例えば「帰

る」の概念については、場所移動 EXT-TRANS に対して、

$EXT-TRANS(obj, A, B)$

は、

$EXT(obj, B). TIME(++t) \&$

$EXT(obj, A). TIME(+t) \&$

$EXT(obj, B). TIME(+t)$

このように概念自体が内在している文脈上の情報が記述される。

概念の構造

単純命題の意味を与える事象概念を概念素に分解し、その構造について述べた。

- 1) 事象概念は、状態概念とその状態概念に変化概念が結び付き形成される状態変化概念、そしてそれらに付与されるアスペクト概念、使役概念が結び付き、形成される。そして事象概念に関わる概念関係子はそれらの概念に依存している。
- 2) 状態概念、状態変化概念は、それらの概念が成り立つために必須の概念関係子のセットを持っており、それらの概念関係子の意味はそれぞれの概念によって規定される。
- 3) 概念が生起する時間、場所に関わる概念関係子は事象概念にわたって定義される。
- 4) 変化概念、アスペクト概念を持つ概念は、時間上の文脈を指示しており、状態概念に分解することで、その文脈の内容を表わされる。

このように語彙の概念には、それ自体に文脈の情報、つまり複合命題としての知識が含まれており、複合命題の形成の基盤を与えている。

4. 事象間の関係と知識

事象間の概念関係は、自然言語では様々な意味をもって表現される。ここでは自然言語が持つ微妙なニュアンスを形式化することが目的ではない。論理的な事象間の関係を表現することが目的である。

事象間を結び付ける関係子は、命題論理の結合子と同じであり、さらに自然言語での事象間の関係子が追加される。

記述形式は、

sk(C1, C2)

ここで、sk は事象間概念関係子である。事象概念 P1 と P2 間の概念関係子には、

- 1) 連言 (and) : and(P1, P2)
- 2) 選言 (or) : or(P1, P2)
- 3) 否定 (not) : not(P1)
- 4) 連続 (&) : &(P1, P2) : 時間上の前後関係
- 5) 条件 (con) : con(P1, P2) : if P1, then P2.
- 6) 理由 (rea) : rea(P1, P2) : P1はP2の理由

などが用意され、命題としての知識が記述される。ここで、連続は時間関係、条件、理由は因果関係を示しているがそれぞれ依存関係を持っている。

与えられた論理は真理値分析によって分析され、真理値が得られる。

5. オントロジカルな知識

自然言語理解や一般的な問題解決に有用な知識は、比喩や願望での知識ではない。現実世界の中の知識の集積が必要である。

ここでは、認知でき表現される知識は、自然言語を中心とする記号によって表わされる。そのような知識は、自然言語の語彙が代表している事象や事物の概念の集合で記述できるとしてきた。

そして、概念は自由に組み合わせることが出来るわけではなく、事象概念には、事象が持っている意味内容に応じて、論理的に概念(意味)が成り立つ条件があることを説明した。

意味的に成り立つ概念の組み合わせで記述される知識の真偽は、個別の概念では議論が出来ない。文(命題)としての知識に対してなされるが、真理値分析を行なうには、その命題に含まれるすべての命題について、その真偽が与えられていなければならない。

オントロジカルな知識とは、願望や比喩のように真偽が議論できないような命題を排除し、おとぎ話や宇宙空間など現実の世界では成り立たない命題も排除した命題である。いわばしっかりとした情報を持った知識をいう。

6. 概念と知識の体系

概念は、事象概念と事物概念、そしてそれらの属性概念に分けられ、事象概念と属性概念は状態

概念を概念素とし、それに変化概念、時相概念、使役概念、そのほかの概念素を作用させて状態変化概念などの概念を派生させる。状態概念はそれ自体が概念素であり、その特徴に応じて階層化される。一方、事物概念は、その概念が内在している属性のセットについて各属性に対応して体系化がなされなければならない[8]が、ある重要な属性に着目するば、体系はインヘリタンスを保証した階層をなす。

概念の体系は、知識の抽象化に有用であるだけでなく、知識ベースでの推論、さらには学習の仮説生成の道具として用いられる。知識の体系化は概念の体系化を密接の関連する。

Fig.4 に概念の体系と概念の記述からなる概念知識ベースの構成を示す。ここで概念の記述に含まれる概念に依存する知識の記述は省略している。

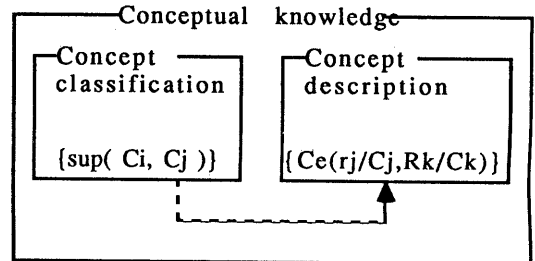


Fig. 4 Conceptual Knowledge Base

7. 大規模知識ベース

自然言語理解や一般性のある問題解決を行なうためには、広範な自然言語現象や知識の現象をカバーするオントロジカルな知識の集積が必要である。

語彙の概念を整備することで、知識として表現される事象や事物の概念が表現される。特殊な語彙を除くと一般に用いられている語彙の数のオーダーは10万語程度で、それらの語彙の数が、認知し区別できる概念の数のオーダーと考えられる。

EDRプロジェクトでなされている概念辞書は、語彙と概念の対応を数10万語にわたって記述し、それらの概念と他の概念との間の関係を記述することで概念に関する知識を記述している。

このような概念辞書の開発は、常識の知識を記述する大規模な知識ベースの構築の際の概念の関

する知識の基盤となる。自然言語現象をカバーする十分な量の概念の知識は、知識ベースの一般性の確保に欠くことが出来ない。

一方、オントロジカルな知識の宝庫は、ドキュメントであり、ドキュメントから何らかの手段で知識を形式化し、集積することで大規模知識ベースの構築が計られる。

Fig.5 にドキュメントベースのなかのテキストから概念知識、語彙知識を用いて自然言語理解、人手による検証のよって知識が獲得されるオントロジカル知識ベースの構成を示す。

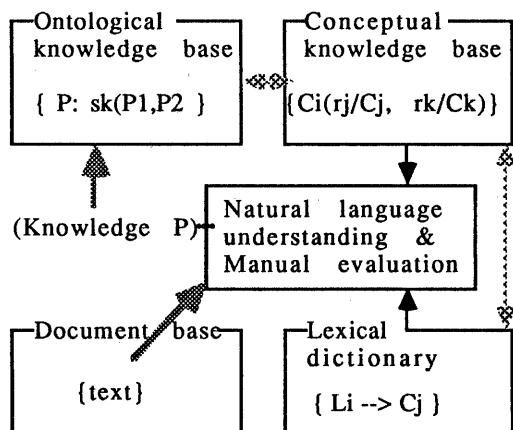


Fig. 5 Knowledge Base Structure

8. おわりに

自然言語理解や問題解決の一般性の向上のため、オントロジカルな知識を集積した大規模な知識ベースを構築することを目的として、語彙の概念と知識の関係について次のように議論した。

- 1) 命題としての知識の意味内容は、単純命題の意味を記述する概念によって与えられる。
- 2) 事象概念が論理的に成り立つためにその概念に固有の必須の概念関係子のセットがある。
- 3) 事象が生起する時間や場所に関わる概念関係子はすべての事象概念に関係する。
- 4) 事象概念は、状態概念とその状態概念に変化概念が結び付いて形成される状態変化概念からなり、さらに、アスペクト、使役概念など

が結び付きさらに多様な概念が形成される。

- 5) 状態変化の時間上の情報は、変化概念やアスペクト概念により与えられる。
- 6) 概念の記述により与えられる単純命題のうち、またそれらの組み合わせで形成される命題のうち、現実世界で成り立つ知識をオントロジカルな知識と呼ぶ。
- 7) 大規模知識ベースは、概念の知識とオントロジカルな知識を集積し体系付けたものである。

自然言語理解や問題解決に用いられる知識ベースについて、自然言語が知識表現のもっとも豊かなメディアであるという立場から、その構造について議論した。ここでの議論は、まだ初期のものであり十分とはいえないが、自然言語の語彙の概念が、知識の基盤であり、そのうえで知識が成り立っていると考えるのは自然であり、今後の知識ベース開発の方向を示している。

最後に、議論を頂いたEDRの横井所長、東芝の武井氏、JIPDECの平井氏に謝意を表わします。

参考文献

1. Lenat, D. "CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottleneck," The AI Magazine pp.65- (1989)
2. 「EDR概念辞書」EDR TR-020
3. Lenat, D. "The World According to CYC," MCC TR No.ACT-AI-455-89
4. 白井「形式意味論入門」産業図書
5. 松本ほか「時相論理とその応用」情報処理、Vol. 30, No.6, pp.651- (1989)
6. 西山「文の論理構造、英語学体系5 意味論」大修館書店
7. 田中ほか編「自然言語理解」オーム社
8. 牧野「概念分解について」人工知能学会大会 (1990)
9. 池上「するとなるの言語学」大修館書店
10. 牧野「ローカリスト理論と概念分解」人工知能学会大会(1991)
11. 安井ほか著「英語学体系5 意味論」大修館書店