

知識情報処理の現状と動向

Overview of Knowledge Information Processing Systems

石原孝一郎* *Kôichirô Ishihara*

人工知能、知識工学などの言葉でも呼ばれる知識情報処理は急速に進展しつつある。本号はその情報処理分野への応用の特集号であるが、後続の論文の背景にある知識表現、推論、知識処理言語などの基本技術と、エキスパートシステムと自然言語処理の2大応用に関して研究開発の歴史と技術の動向について概観した。

日立製作所では早くからこの分野の研究開発に力を入れ、1985年12月号の本誌で、知識工学とその産業への応用を特集して成果を報告したが、それ以後本号でも後述するLISP、PROLOGやES/KERNELなどのユーザーのシステム構築を助ける言語やツールを続々と商品化し、それとともに応用システムの面でも実用システムの開発を進めてきた。しかし、システムのより大規模化、高機能化を目指して解決すべき研究課題も多い。

1 緒言

知識工学や人工知能及びこれらに基づいた応用システムの研究開発が活発になっている。両用語の一般に合意された定義は存在せず、区別もあいまいになっているが、ここでは人工知能は「基本原理を追求する科学」、知識工学は「応用面に重点を置いた工学」と区別しておく。知識情報処理という用語はなじみが薄いですが、「ソフトウェア工学や関連研究分野を自然な方法で統一的に吸収した拡張された知識工学」と定義されている¹⁾。知識工学がエキスパートシステムとほぼ同義に誤用されることもあり、これを避けるためにも本稿ではこの「知識情報処理」という用語を用いた。

なお、「知識処理」という用語も、「知識情報処理」とほぼ同義に用いることがある。特に、後に言語、システムなどの語が続き、形容詞的に用いる場合は、「知識情報処理」では長くなるので、「知識処理」を用いることにする。

人工知能は30年、知識工学は10年の歴史があるが、実用的な応用システムは少なかった。しかし、ここ1、2年、急速な応用範囲の広がりとともに、実用システムも出現するようになった。従来は、問題の定式化がはっきりしやすい産業分野への応用が多かったが、情報処理、ビジネス分野への適用も最近になって多く行われている。次論文以降で、日立製作所での基本ソフトウェアや応用の代表的なものを紹介するが、本稿ではそれらの応用システムを支えている技術の歴史、現状と動向について概説する。

2 研究開発の歴史

人工知能の研究の萌芽は1950年代に見られるが、1956年のMcCarthyらによるDartmouth会議で初めて人工知能という

言葉が用いられた。これ以後、研究交流が促進され、組織的な研究が行われるようになった。1960年ごろ開発された記号処理言語LISP(List Processor)は、その後改良されながら人工知能研究の主要な言語として発展した。

1960年代前半には、計算機による自然語処理の研究が盛んになった。特に、宇宙開発でソビエトに遅れをとった米国では、露英翻訳を中心とする機械翻訳の研究が加速されたが、1966年に出されたいわゆるALPAC(Automatic Language Processing Advisory Committee)報告が、機械翻訳の実用化に否定的な見解を示したため²⁾、研究は急に下火になった。

1965年には、Feigenbaumらによって最初のエキスパートシステムであるDENDRAL(Dendritic Algorithm)³⁾のプロジェクトが始まった。これは化学物質の構造を推定するもので、プロジェクトは10年以上にわたって推進され、後続の研究に多大な影響を与えた。

1969年には第1回の人工知能国際会議が開かれた。この会議はその後隔年に開催され、この分野で世界最大の研究交流の場となっている。

1971年ごろColmerauerらによって提案された論理形プログラミング言語PROLOG(Programming in Logic)⁴⁾は、1977年にエジンバラ大学で実用的な処理系が作られて以後に普及し、今日ではLISPとともに、この分野の2大言語として使用されている。

1976年にShortliffeらにより開発された血液病診断エキスパートシステムであるMYCIN⁵⁾(薬の名から命名された。)は説明機能や確信度などの機能を持ち、エキスパートシステムの一つの典型となった。また、MYCINの経験に基づき、MYCIN

* 日立製作所システム開発研究所

表1 日立製作所でのエキスパートシステム開発例 1980年以後、日立製作所では幅広い分野で多くの知識システムの開発を進めてきた。

種類	分野	機能
計画・設計	計算機	計算機室機器レイアウト
		システム構成設計
	プラント	配管ルーチング
診断	電子回路	回路系変換
	原子力	プラント故障診断
	半導体	プロセス診断
制御	プロジェクト管理	リスク管理
	鉄鋼・電力	炉の制御
	鉄道ヤード	列車運行制御 コンテナクレーン自動制御
情報処理	自然言語	機械翻訳
	ソフトウェア	部品再利用
	金融	融資エキスパート 資金運用エキスパート

としての研究を開始した。それ以後、応用面では計算機室内の機器の配置を行うレイアウトCAD (Computer Aided Design), ファジィ論理によるコンテナクレーンの自動運転システム, 日英機械翻訳システムなど多くの実用的システムを開発してきた(表1)。また, 基本システムの面でも, 高速かつ他言語とのインタフェースを持つなど実用性を重視したPROLOG[開発コード名LONLI(Logic Oriented Language Inferencer)]やLISPなどの知識処理言語及びユーザーのエキスパートシステム構築を助けるツールであるS-LONLI(Super LONLI)⁶⁾, ES/KERNEL (Expert System KERNEL), EUREKA⁷⁾(Electronic Understanding and Reasoning by Knowledge Activation)などを開発し, その多くを製品化してきた。基本システムの一例として図1にS-LONLIを示す。これは各種の知識表現, 推論機構が統合的に扱えるとともに, 知識ベースやユーザーインタフェースを具備した総合システムである。

また, 1982年には, 国家的プロジェクトとしての第5世代コンピュータの研究が10年計画で開始された。これに刺激されて, その後欧米でも幾つかの大型プロジェクトが始まっている。

1986年には日本人工知能学会も創設され, 理論的な研究と実用システムの開発が車の両輪となって, この分野の研究が加速されることが期待されている。

3 技術の現状

3.1 知識表現と推論

人間の持つ様々な種類の知識をシステムへ入力し, システムが処理できる一定の形式で記述するのが知識表現であり, 多くの表現法が研究されている⁵⁾が, 以下の二つが代表的で

から専門知識を除いたEMYCIN (Empty MYCIN)は, その後のエキスパートシステム構築ツールの開発により刺激を与えた。

1977年の第5回人工知能国際会議でFeigenbaumは知識工学を提唱し, 以後エキスパートシステムなど応用研究が盛んになった。

我が国に目を転ずると, 日立製作所でもロボット, パターン認識, 日本語処理などの研究を経て, 1980年から知識工学

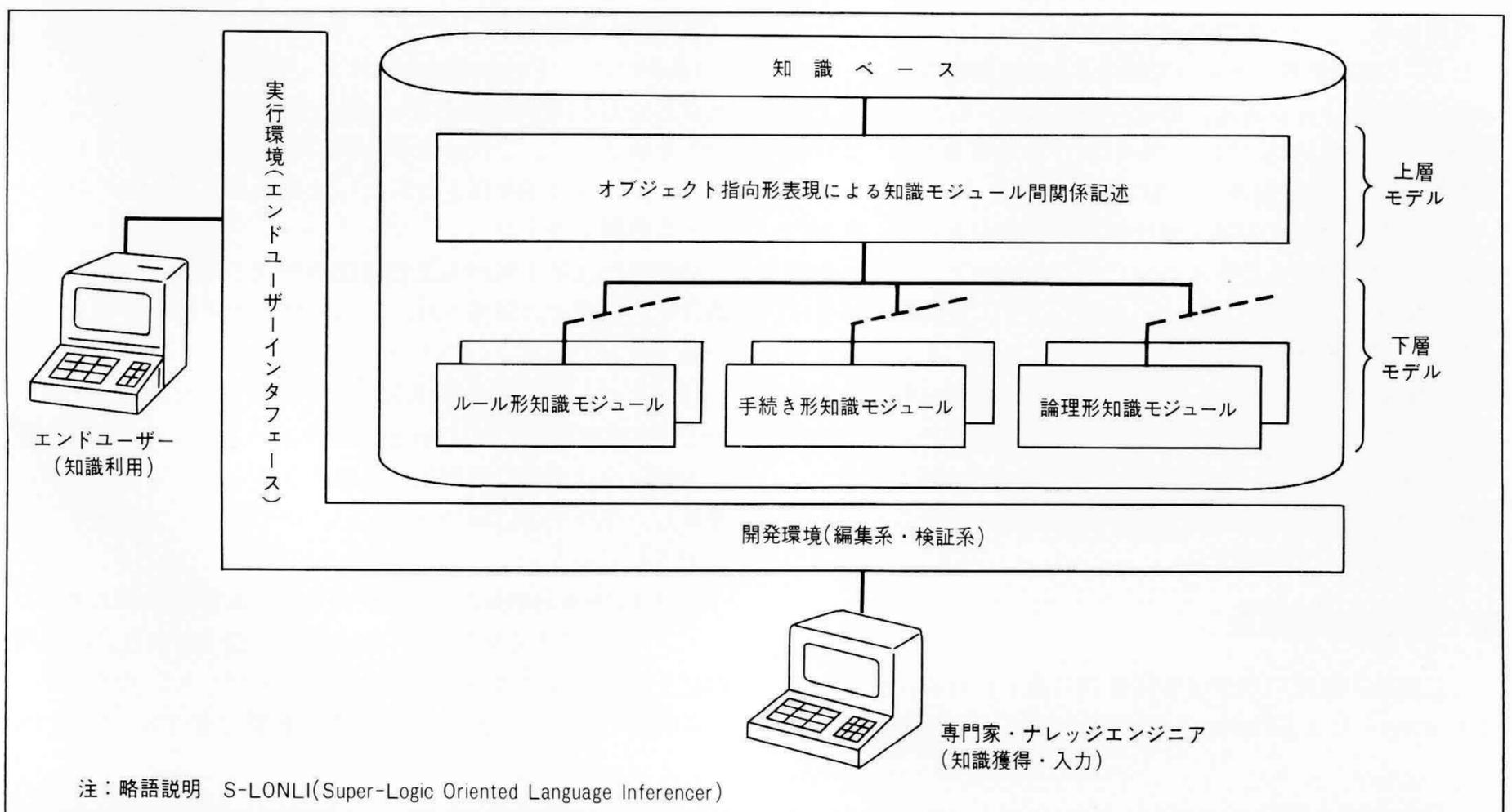


図1 知識処理システムS-LONLIの構成 異なる知識表現法の混在を許し, 2階層モデルにより, これらを統一的に扱う。

ある。

(1) プロダクションシステム

事実と規則及び推論方法で構成される。規則は通常「IF 条件 THEN 実行部」の形で表現されIF-THENルールと呼ばれることもある。事実と規則が知識ベースを構成する。推論方法は大きく分けて二つあり、IF部が満たされればTHEN部を実行する前向き推論と、THEN部を実行するために、IFが成立するかを調べる後ろ向き推論に分類できる。プロダクションシステムは人間の思考になじみやすく、広く使用されている。

(2) 論理

論理学を基礎にした知識表現も多く用いられている。命題論理は命題(真偽を判定できる文)を演算子(論理積, 論理和, 否定など)で結合して新しい命題を作るものである。1階の述語論理はこの命題論理に変数と数量子を導入したもので、例えば「すべての人は死ぬ」という知識は次のように表現される(∇はすべてを意味する数量子)。

$$\forall x/\text{人}(x) \rightarrow \text{死ぬ}(x)$$

現在の論理による知識表現の多くは、この1階述語論理の上に組み立てられているが、他の形式の論理も研究されている⁵⁾。

(3) その他の知識表現法

Minskyらの提案したフレーム理論⁵⁾は、階層的知識表現が可能で静的な事実を記述するのに適しており、これに基づいた言語も多く開発されている。

Quillianらの提案した意味ネットワーク⁵⁾による知識表現は、自然言語理解に適用されることが多いが、他の知識システムでも用いられている。

また、最近では、どれか特定の表現法だけでなく目的に応じて適切な表現法を組み合わせる言語も研究されている⁶⁾。

(4) 推論

前述した前向き、後ろ向きの推論や、あるいは推論の手順を探索木で表わしたとき、深さ方向に解を追求する縦形探索と可能性のある解をすべて探っていく横形探索などの基本技術はほぼ確立された。現在は、確信度などを用いたあいまい推論⁵⁾や、高速化のための並列推論などの研究が盛んになりつつある。

(5) 知識ベース

エキスパートシステムを知識ベースシステムと呼ぶこともあるが、ここでは推論機構と対比して知識の集合を知識ベースと呼ぶことにする。従来は個々の知識システムに内蔵された形のものが多いが、知識の共有・管理の面からは汎用データベースと同じように個別システムから独立した知識ベースが望ましく、その研究が行われている。日立製作所でのKBAF⁷⁾(Knowledge Base Administration Feature)もその一例であり、図2にその構成を示す。同図に見るように、各種の構造の知識のほか、従来のデータベースも一種の知識として扱うことができる。知識ベースの大きな課題は、人間の持つ知識をどのようにして機械に教えるかという知識獲得技術と、獲得した知識の正しさを検証する技術であり、幾つかの試みはあるがまだ十分ではない。

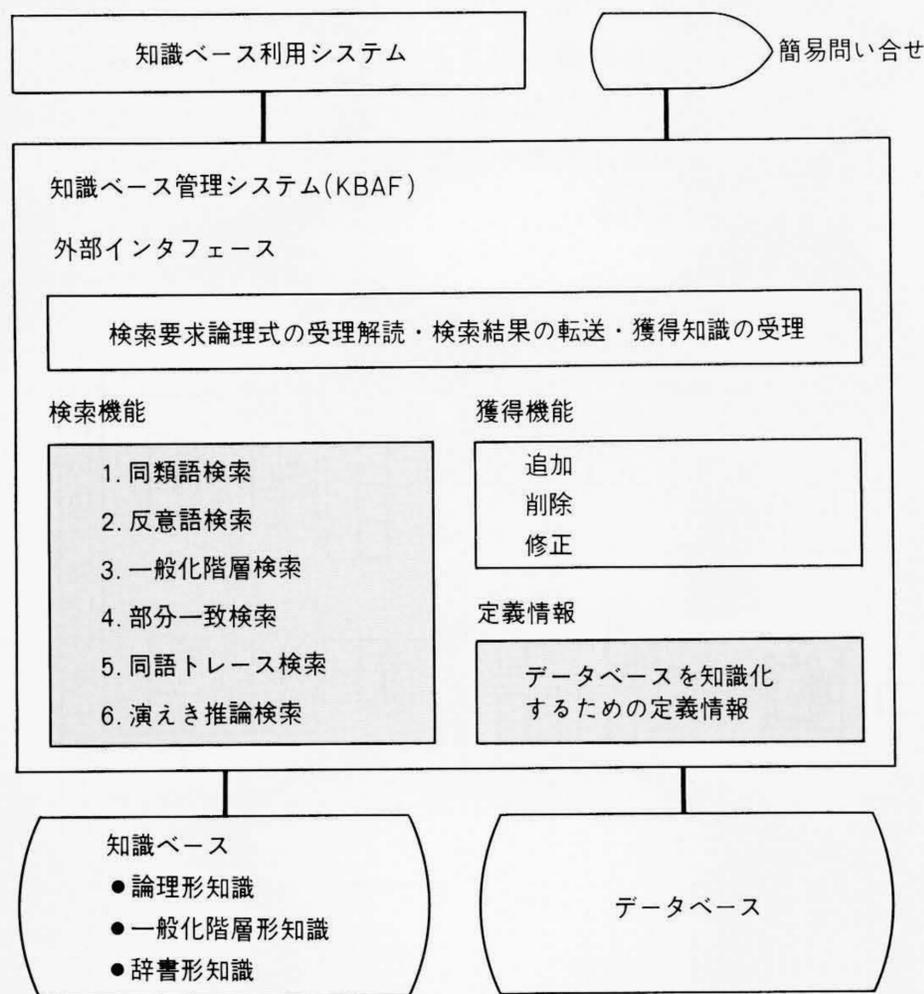
3.2 知識処理言語とツール

知識工学は、それ自身大きな研究分野であるが、応用システムを構築する一手法とみなすこともできる。手法には適切な手段が必要であり、その手段としての知識処理言語とツールについて簡単に述べる。

言語としてはLISPとPROLOGが代表的である。LISPはList Processorの名が示すとおり、木構造のリストをデータとして扱い、記号処理に優れている。1960年ごろの開発以降、多くの派生LISPが開発されたが、最近ではCommon LISPに統一しようという動きがある。後続の論文で述べるHi LISPも、Common LISPに準拠し、大形OSであるVOS 3の下で動く実用システムである。

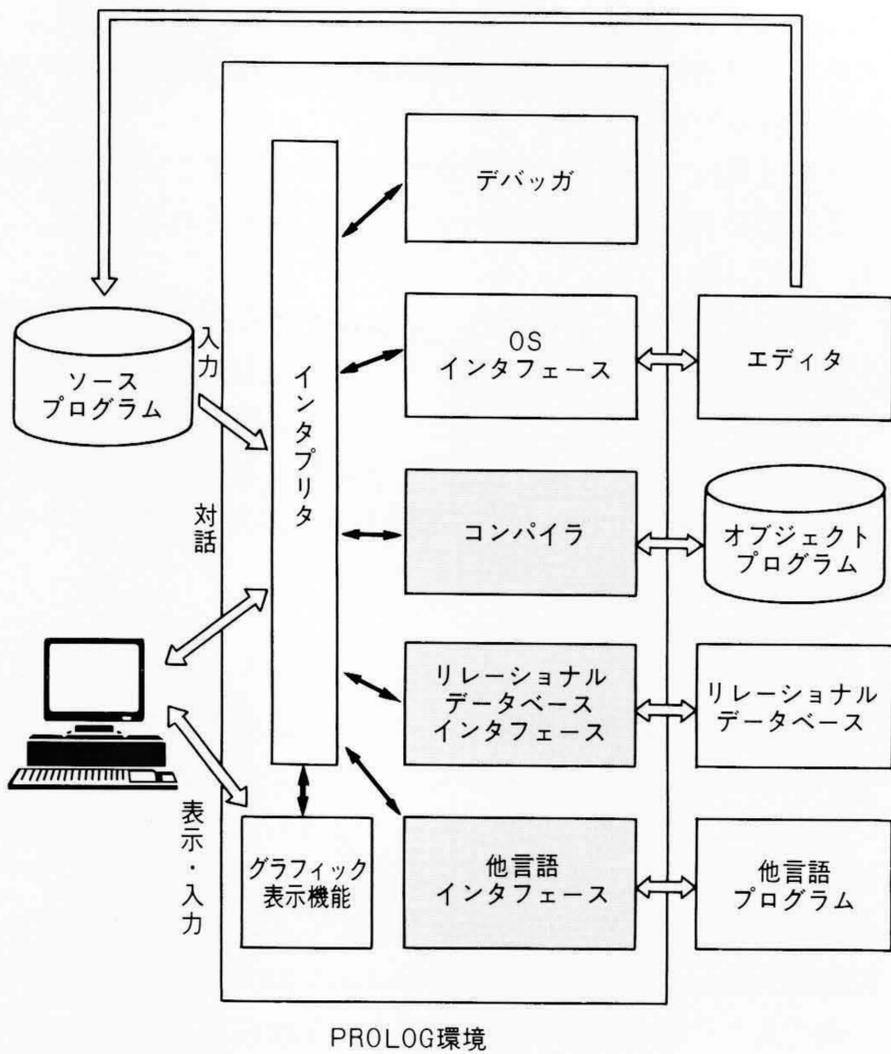
PROLOGは前述の述語論理に基づいたプログラミング言語であり、その特徴は知識を宣言的に書くことができ、推論手続きは処理系が実行してくれる点にあり、歴史は浅いが急速に普及しつつある。次論文のLONLIは、HITAC Mシリーズのホストコンピュータ及び2050ワークステーション上で動くPROLOGの開発名称である。図3にその構成を示すが、PL/I, FORTRAN, C, リレーショナルデータベースなどのインタフェースを持ち、かつ700k LIPS (Logical Inference per Second)と高速な実用言語である。

LISPやPROLOGを用いてエキスパートシステムを作ることでもできるが、推論機能や知識記述機能などの標準的なものを備えたツールがあれば、システム構築者は知識記述に専念することができる。効率は大いに向上する。これがエキスパート



注：略語説明 KBAF(Knowledge Base Administration Feature)

図2 知識ベース管理システム(KBAF)の構成 論理形, 辞書形など各種の構造の知識のほか、従来のデータベース中のデータも知識とみなして扱うことができる。



注：略語説明 OS(Operating System)

図3 LONLIの構成 従来のPROLOG機能のほかに、FORTRAN, PL/I, Cなどの他言語やリレーショナルデータベースのインタフェースを加え、高速(700k LIPS)化を図った知識処理言語である。

システム構築ツールであり、エキスパートシェルと呼ぶこともある。その始まりはMYCINから専門知識を除いたEMYCINであるが、最近では多くの商用ツールが開発されている。後続の論文で述べるES/KERNELはその一例である。

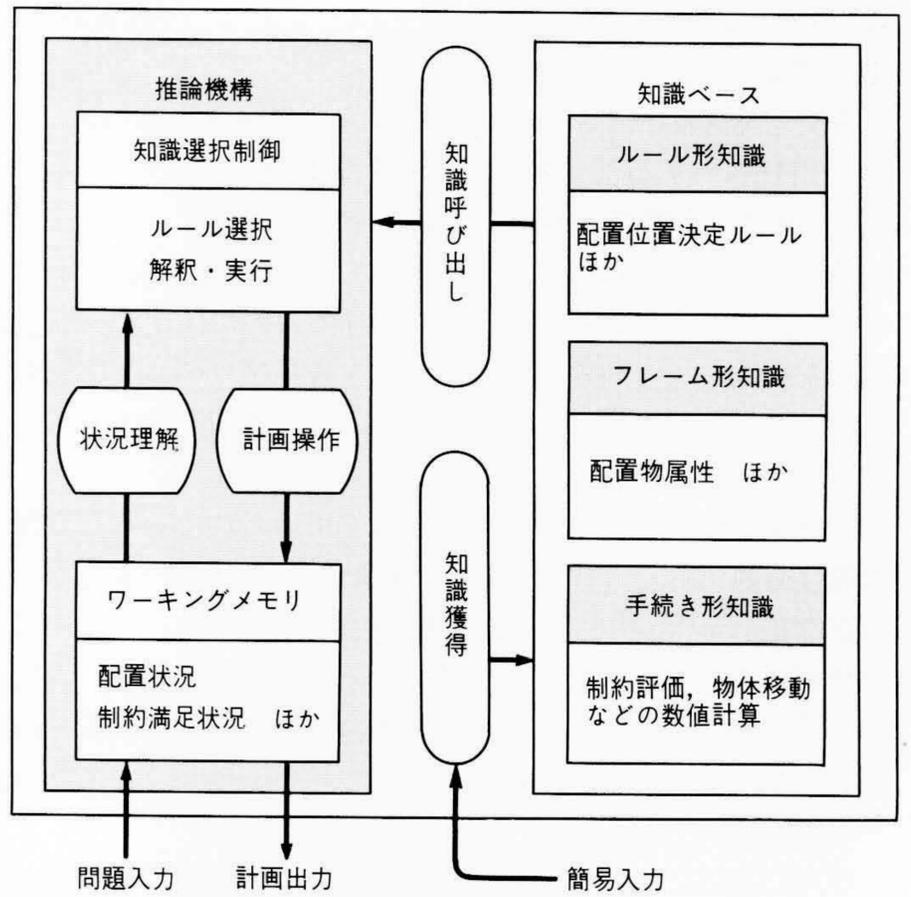


図4 計算機室レイアウトCADシステムの構成 知識ベースはルール形、フレーム形、手続き形の3種類の知識から成る。

3.3 エキスパートシステム

エキスパートシステムは、専門家の持つ知識を組み入れ、専門家に代わり、あるいはその助けをして専門業務を行うシステムであり、対象とする問題の性質に対応して、解釈、制御、設計、診断、予測、計画などの種類に分類される。

知識を蓄積する知識ベースと、これを利用する推論部とを明確に分離しているため、次の二つの利点を生む。

- (1) この種の知識は変更や追加が頻繁に必要なになるが、知識の保守が容易である。

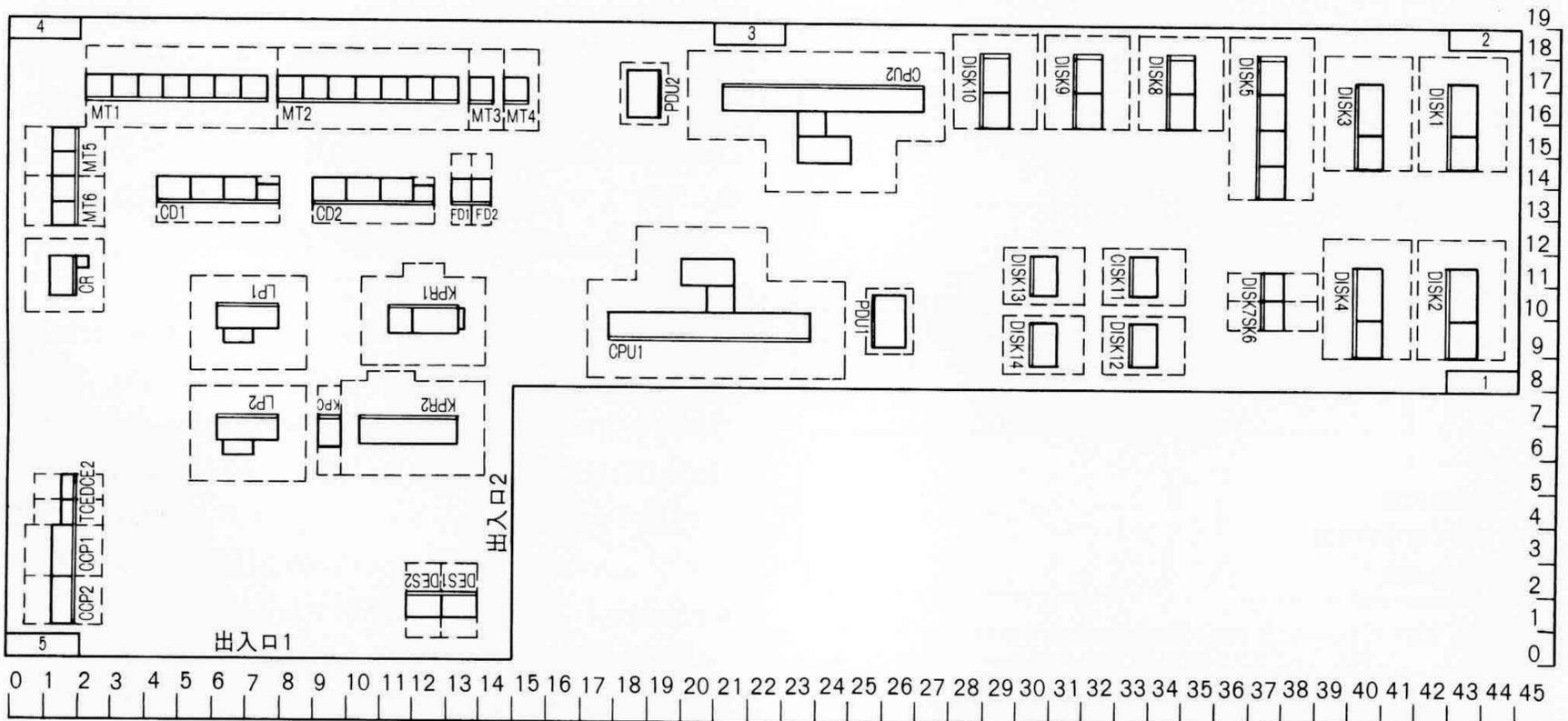


図5 計算機室のレイアウト例 実線で示すピストル状の部屋に、2 CPU(中央付近)を含むシステムを配置した例を示す。実線は機器形状を示し、破線は保守エリアを含めて必要なスペースを示す。

(2) 知識の記述には、計算機システムに関する深い経験や知識は必要としない。

これらの利点のため、エキスパートシステムの応用範囲は急速に拡大しつつある。日立製作所での開発事例は前掲の表1に示したが、具体例として計算機室のレイアウトCADの構成を図4に、また実際のレイアウト結果の例を図5に示す。図4に見るように、知識ベースはルール形、フレーム形、手続き形の三つのタイプの知識で構成され、これらを推論機構が呼び出しながらレイアウトをしてゆく。このシステムにより、レイアウト時間の短縮と熟練したシステムエンジニアと同等のレイアウトが可能となった。従来は知識の定式化が比較的容易なサイエンスや産業分野への適用が多かった⁷⁾が、しだいに意思決定などのビジネス分野への応用が増加しており、本号でもその幾つかについて報告する。

組織でのノウハウの伝承や熟練者の養成がますます困難になりつつある現状から、エキスパートシステムに対する期待は大きい。しかし、知識の獲得技術がまだ十分確立されていないこととシステムの処理能力の制限から、あまり大規模なシステムは実現されておらず、既存のほとんどのエキスパートシステムは数百の知識で構成されている。また、前述した(2)の利点はあるものの、現在ではシステムの構築に際して、知識工学技術と計算機システムの両面の専門知識を備えたナレッジエンジニアが必要であり、その養成が急務とされている。

3.4 自然言語処理

自然言語は人間の持つ最も強力な対話手段であるが、従来は人間同士の対話にしか利用できなかった。しかし、最近では徐々にではあるが、人間と機械の対話が自然言語を介して可能となりつつある。図6に示すように、自然言語処理を解析、変換、生成のプロセスに分解して考えると、代表的な応用システムは次のように整理できる。

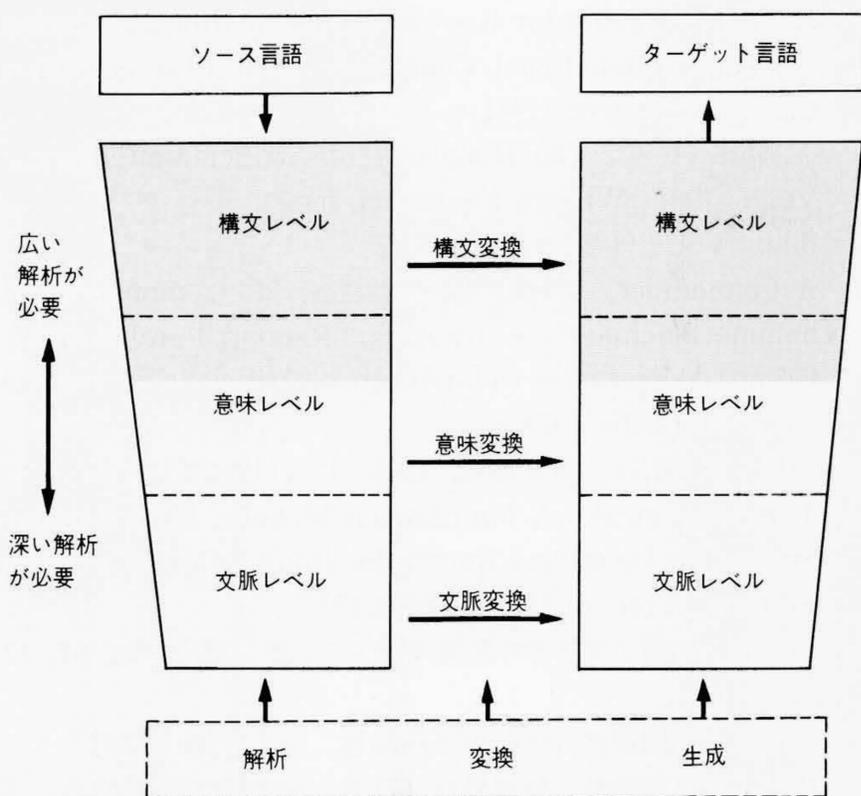


図6 自然言語処理システム 自然言語処理は、ソース言語を解析・変換・生成の3フェーズの処理を通してターゲット言語に変換する。

(1) 自然語インタフェース

ソースが自然言語で、ターゲットが計算機言語のシステム。

(2) 質問応答システム

ソースとターゲットが同一の自然言語であり、かつ内容の変換を伴うシステム。

(3) 機械翻訳

ソースとターゲットが異なる自然言語であるシステム。

計算機システムに自然言語を理解させるという点で、まず解析フェーズが重要な技術となる。解析の深さを示すのに、構文規則をもとにする構文レベル、意味内容の解析まで行う意味レベル、複数の文にわたって解析する文脈レベルの区別がされるが、前二者の差は明確ではない。現状の多くのシステムは、構文解析を中心に若干の意味解析を加えたものがほとんどで、文脈レベルの解析は一部で代名詞などの解析がなされている程度である。

必要とされる解析の深さも応用によって異なり、例えば自然語インタフェースではキーワードを抽出すればよいが、機械翻訳では深い解析を必要とする。

我が国では、社会的ニーズから機械翻訳の研究開発が盛んで⁸⁾、世界的に見ても技術レベルは高い。しかし、ユーザーの期待からするとまだ不十分であるが、辞書の充実や使用経験の蓄積によって実用性がしだいに増しつつある。

日立製作所での自然言語処理システムのうち、機械翻訳システム、自然語インタフェース及び知的検索システムについては後の別論文で報告する。

4 動向と課題

知識工学は実用期を迎えて急速に普及しつつあるが、更に大きな発展をするために解決しなければならない課題も多い。研究の方向としてシステムの大規模化と高機能化の二つの流れがあり、その動向と課題について考察する。

現在のエキスパートシステムの多くは数百の知識で構成されていることは前述したが、人間は通常もっと多くの専門知識や時には膨大な常識に基づいて推論や判断を行っている。システムの大規模化に際して、まず知識の体系化が必要となる。体系化しないと、多数の知識の維持・更新やそれに基づく検索・推論も極めて困難となる。フレーム理論もそのような体系化の一手法であるが、前述したS-LONLIはオブジェクト指向による2階層モデルの表現により大規模な知識記述への道を開いた⁹⁾。

大規模な知識ベースをすべて人間が記述するには限界があり、機械による学習の研究も推進されている。学習の方法としては、例えば実際の施行例から逆にルールを抽出するとか、マニュアルの記述をルール化するなどの研究がある。

推論の高速化もシステムの大規模化に必要な技術である。高速化技術の一例として、図7に示すようにES/KERNELは値の変化するデータだけを追跡する手法により、総ルール数にあまり依存しない高速推論方式を実現した¹⁰⁾。前述の知識の体系化も推論の高速化に有効であるが、これに加え、HITAC Mシリーズに付加する形で、あるいはマイクロプロセッサの形で知識処理アーキテクチャの研究も進めている。

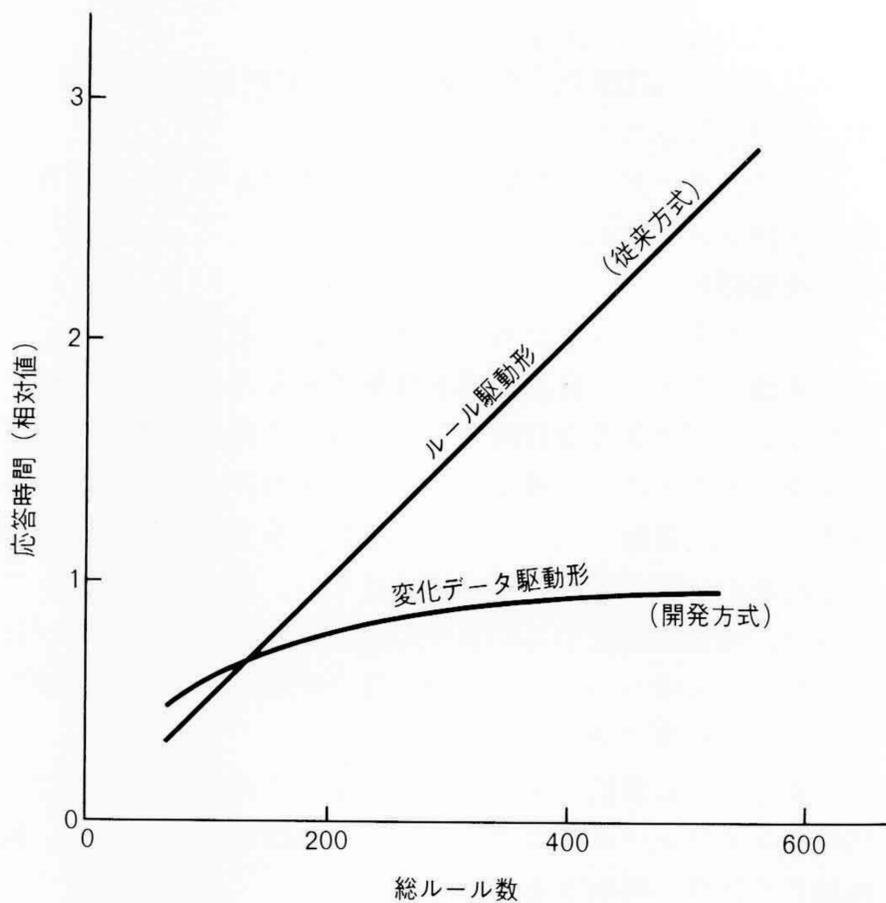
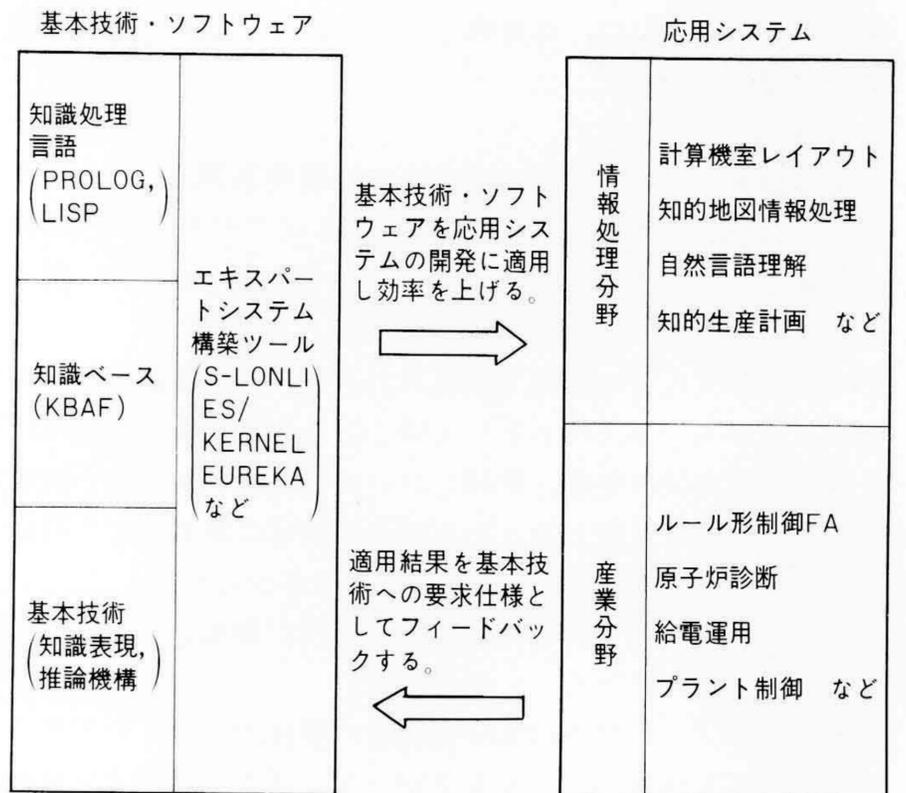


図7 ES/KERNELの高速推論方式 ルールの適用順序を整理し、値の変化したデータだけをルールの条件部と照合する技術により、応答時間が総ルール数に依存しない高速推論方式を開発した。



注：略語説明
ES(Expert System)
EUREKA(Electronic Understanding and Reasoning by Knowledge Activation)

図8 基本技術と応用システムの両輪 課題解決のためには、基本技術と応用システムの両面からのアプローチと互いのフィードバックが不可欠である。

システムの高機能化の面では、あいまいな知識の表現とこれに基づくあいまい推論⁵⁾が必要となる。既存のシステムのほとんどは、真か偽かの2値による知識と推論によっている。確信度やファジィ論理などの研究もなされているが、これらの統一的な扱いが今後必要である。

高機能化の別の面として、知識の記述やシステムの検証をもっと容易かつ自然に行わせるユーザーインタフェースも重要であり、知識に関する知識とでも言うべきメタ知識に基づく知識入力誘導や、推論過程のトレースや説明機能などが多くのシステムに組み込まれつつある。

自然言語処理の分野では、意味解析や文脈解析の研究が盛んで⁹⁾、ごく限定された話題での実験例の報告などもあるが、はん用的な解析手法はまだ今後につとめるべきところが多い。機械翻訳については、任意の入力文を翻訳することは不可能であり、入力文の形式に一定の制限を課す方法や、入出力文の前編集や後編集を容易にする援助システムの研究も活発になりつつある。

これらの課題を解決するために、日立製作所では社内の豊富な事例に基づく応用システムの開発経験の上に、基本技術・ソフトウェアへの要求仕様をまとめ、これに基づいて基本システムを開発し、この基本システムを次の応用システムに使用するというサイクルを通して、基本技術と応用システムの両面の開発を推進している。この関係を図8に示す。

5 結 言

知識情報処理の研究の歴史・現状・動向について概観した。従来の計算機技術はソフトウェア生産性など幾つかの点で行き詰まりを見せており、それを打破するための知識情報処理

に対する期待は大きい。現在までのところ、その期待に技術のほうに十分にはこたえてこなかったが、他言語とのインタフェースを持ち高速で実用的なLISPやPROLOGあるいはエキスパートシステム構築用ツールなどの出現により応用システムも急速に拡大しており、特にエキスパートシステムや機械翻訳などの応用は実用期にさしかかっている。今後、基本技術と応用が車の両輪のようになって互いに刺激とフィードバックを与えながら発達するであろう。

参考文献

- 1) K.Fuchi: Aiming for Knowledge Information Processing Systems: International Conference on Fifth Generation Computer Systems(1981)
- 2) A.Bars et al.: The Handbook of Artificial Intelligence Vol.1 (1981, William Kaufmann, Inc.)
- 3) ibid. Vol.2 (1982)
- 4) A.Colmeraner, et al.: Un système de communication homme-machine en français: Rapport préliminaire, Groupe de Recherche en Intelligence Artificielle, Université d'Aix-Marseille(1972)
- 5) 特集: 知識工学: 情報処理, 26巻, 12号(1985年12月)
- 6) T.Chusho, et al.: A Multilingual Modular Programming System for Describing Knowledge Information Processing Systems: Proc. of IFIP 86(Sept.1986)
- 7) 小特集: 知識工学とその産業分野への応用: 日立評論, 67, 12(昭60-12)
- 8) 大特集 機械翻訳: 情報処理, 26巻, 10号(1985年10月)
- 9) 特集 計算言語学: 情報処理, 27巻, 8号(1986年8月)
- 10) 船橋, ほか: ES/KERNELとEUREKA-IIの推論高速化技術: 日経コンピュータ, 130号(1986年9月15日号)