コンピュータと教育　10－10
（1990. 4.25）

集合論の定理証明向けＩＣＡＩ

ＩＣＡＩ　for theorem proving in set theory

小林学、吉川厚　大場勇治郎
Manabu KOBAYASHI Atsushi YOSHIKAWA Yujiro OHBA
慶應義塾大学理工学部
ＫＥＩＯ　University　Faculty　of　Sci.　&　Tec.

あらまし　学生が主体的に学習を行える知的ＣＡＩシステムには、自然言語インターフェースは不可欠である。しかしながら、自然言語インターフェースを組み込むことはたいへん難しい。ここでは、具体的に集合論の定理証明を取り上げ、そこで生じる問題点とその解決案について述べている。特に、数式混交文の処理を問題記述言語とメタ言語の考え方から、数式を名詞と同等に扱うことによって行っている。これによりローマ字漢字変換で数式混交文の日本語を入力しても、正しく切り出すことができる。また、文節の係り受け処理も助詞の性質を使っている。文法を定理証明専用にしたことで、比較的容易に自然言語インターフェースができることを述べている。

キーワード　ＣＡＩ　教育情報処理　大学教育　知識工学　自然言語処理 インターフェース

１．はじめに

人工知能の技法を用いた知的ＣＡＩの研究が盛んに行われてきている［1-4］。知的ＣＡＩの構成に関する研究は、いままでかなりの成果をあげてきた［5,6］。現在、研究の対象は教材知識モジュール、問題解決モジュール、解答解析モジュール、問題生成モジュール、そしてインターフェースの各モジュールへと移っている。それらの中で、インターフェースの研究は重要視されている。知的ＣＡＩは従来のＣＡＩとは異なり、多肢選択の問題を出題するのではないので、学生の入力が番号や数字といったものではない。あらかじめ学生の答えは設定されておらず、解答にどの様なものがくるのか予想もつかない。また、学生は知りたい項目の質問がいつでもできるなど、学生に対しても主導権を与えているのも知的ＣＡＩの大きな特徴である。このため、単に学習環境の整備にとどまらず、学習者とシステムとの間を取り持つインターフェースが知的ＣＡＩの能力をも左右するほど、重要になってくるのである［5］。

しかし、現在のこの分野はかなりたちおくれているといってよい。それは、学生が主体的に主導権を発揮できるよう、ある程度の自然言語を理解できるインターフェースにしなければならないからである。たとえば、集合論の定理証明の問題に対する学生の答えとして、

「ｘ∈Ａ∩Ｂ」、　「∴　ｘ∈Ａ　（∵　共通要素の定義）」

の様な数式を用いたものや、

「ｘをＡ∩Ｂの任意の要素とする」、　「ゆえに共通要素の定義よりｘはＡに含まれる」

の様な自然言語を用いたものも考えられる。従来は、数式やなんらかの記号入力しか入力を認めていなかったが、これは設定した入力形式に合わせられる学生しか利用できない、もしくは、利用意欲を減退させるという欠点がある。両者とも入力形式として認めておかなくては、学生に”使ってもらえる”知的ＣＡＩにはほど遠いものとなってしまう。そこで、学生に使ってもらえる知的ＣＡＩにするための、インターフェースの拡充について報告する。具体的に、集合論の定理証明問題を取り上げる。

２．インターフェースの問題点

数学や物理学等、理系科目の学習を行う知的ＣＡＩのインターフェースは、文系科目用の知的ＣＡＩとは、異なった難しさを持っている。それは、数式を自然言語文に埋め込んでいる、数式混交文を扱うことである。例えば、どこからどこまでが式であるのかの判別を行ったり、記号の切り出しも行わなければならない。この問題は日本語を扱い、ローマ字漢字変換を行う場合は顕著である。さらに、式そのものをどういったカテゴリーで扱うのかという問題もある。自然言語と式が同じ内容を指している場合、これらの記述形式が異なるものをどう統一的に扱うかという問題も考えなくてはならない。また形式的にも、式が混在する文は句点が無い場合が多く、文としての切り出しも難しい点がある。それゆえ、これら数式と自然言語の混在する文を扱う際には非常に多くの問題を処理しなければならない。以下にそれらの問題点を具体的に示し、数式混交文の解析が抱えている問題を明確にする。

２．１　入力について

ここでは、学習者がキーボードを通じて入力することを想定する。その際生じる問題点には以下のようなものがある。

〔１〕キーボードに不慣れな学習者はキー入力に煩わしさを感じる。
〔２〕日本人が使うため、日本語で入力できるようにしなければならない。
〔３〕数学で使う独特な記号（∈や∩等）がキーボード上にない。
〔４〕入力は、たいていのシステムではコマンドラインを用いて一行ずつ行われる。このため、一度に多くの入力を受け付けることができないので、学習者は一度に多くの事柄や式変形を思い付いたとき、入力にもどかしさを感じる。また、以前の入力を見直したり書き直したりすることができないか、もしくはたいへん面倒である。
〔５〕ユーザーがインターフェースの能力を正確に把握していないと、能力以上の要求をしたり、また逆にインターフェースの能力を十分に生かすことができない。しかし、使用のために多くの時間を割くことは、学生に使う気をなくさせる。

このような、入力における問題点は、あらゆる知的ＣＡＩのインターフェースが抱えている問題である。このように入力形式自体にも様々な問題があり、これらはユーザーとシステムのスムースなコミュニケーションを妨げ、ユーザーの学習意欲を低減させる要因となる。

次にこれらの問題点について、その解決の手がかりを考え、我々がとった方法について述べる。
〔１〕この問題は入力にキーボードを使う形態に固執するならば、ユーザーにキー操作になれてもらう方法しかない。入力形態を、解答を解答用紙に書いてもらいイメージスキャナで読み取る方法にすれば、この問題は解決できる。さらに、手書き文字入力は、〔２〕、〔３〕、〔４〕、〔５〕の問題に対してもかなり改善することができる。しかし、手書き文字読みとりは、現在の技術レベルでは実現が難しい。特に、$\int_0^a xdx$のように字の大きさが異なっていたり、上付け、下付け文字が他の記号と重なるような場合は特に難しい。我々は、”慣れてもらう”という消極的な方法をとった。
〔２〕これは、現行のキーボードを使うのであれば、日本語ワープロが行なっているローマ字漢字変換、あるいは、かな漢字変換を行なうことが考えられる。この場合でも、数式混交文の場合には多くの難しさが伴う。かな漢字変換の場合には”ｘ＝ｙヨリ”の形で入力が可能であるので数式の同定は、アスキーコードによって判別できるのでさほど困難ではない。しかし、学習者はアルファベット配列のキーにはたやすく慣れるが、カタカナ配列はなかなか覚えない。そのためローマ字漢字変換を行うことにした。この場合は、”ｘ＝ｈｈａ”は”ｘ＝ｈは”なのか”ｘ＝っは”なのか、表面上からは判断ができないという問題がある。この点は、２．２で詳しく述べる。
〔３〕この問題は、次の２つの方法が考えられる。第１の方法は、”＄”や”＆”などのすでにキーボード上にある記号を用いて代用することである。あるいは、ファンクションキーにそれらの記号を登録しておくなどの処理が考えられる。この方法の欠点は、登録できる記号が高々２０程度であることである。この数は、集合論の演習にとって、最小限にすぎない。第２の方法は、たとえば、”ａｎｄ”と入力すると、自動的に表示が”＆”となるようなプリプロセッサを知的ＣＡＩに組み込むことが考えられる。我々は前者の方法を採用した。
〔４〕スクリーンエディタをシステムに組み込んで、エディタで解答の編集が終わってからシステムに渡す方法をとれば、学習者は通常の試験と同じ感覚でできるので、かなり改善できると考えられる。我々もこの方法で行った。
〔５〕については、システムの機能を学習者が知りたい場合即座に説明できるオンラインマニュアルが備わっていれば、かなり改善される。ただし、このオンラインマニュアルは、できればコマンド入力ではなく、自然言語入力を許し、学習者の意図を把握して、それに対応することができるシステムが望ましい。しかし、学生の解答の他に、学生の要求をも自然言語の使用を許すと、自然言語処理機能がかなりたいへんになるので、我々はコマンド形式のオンラインマニュアルにした。

ここに挙げた解決策の多くは、自然言語処理の単独の問題ではなく、文字認識などマンーマシンインターフェース全体に関わる研究課題である。これらが理想的に実現できれば、それだけでかなり使いやすいシステムとなる。次に、学習者とシステムとの間の問題から、学習者の入力をシステムがどう解析していくのかという処理の問題、つまり、入力解析について同様に問題点を挙げ、その解決策を検討してみる。

２．２　入力解析について

自然言語の「意味」とは言語による表現が表わしている実体のことであり、意味をコンピュータが理解するということは、コンピュータの内部表現に実体を表わすことである。このように、ユーザーの入力からコンピュータの内部表現に変換する作業を入力解析という。

学習者の入力が終了した後、入力解析にはいるが、この入力解析過程での問題点を列挙しておく。ただし、入力の問題はすべて解決しているものとして考える。入力解析の際に生じる問題を、具体的にＦｉｇ．１およびＦｉｇ．２に示す入力を例にとって述べる。図中の右肩についているｘは補集合を示す。また下線は学習者の入力ではなく、後の説明のためにつけたものである。まず、問題点を列挙する。

【１】一文への分解

学習者の意図している意味を知るために、通常、構文解析、意味解析というステップで入力解析を行なう。構文解析は、その解析の対象が一文であるため、入力を一文ずつに分解しなければならない。しかし、Ｆｉｇ．１を見てもわかるように、文末に必ずしも読点がつけられているとは限らない。その際、どこからどこまでを一つの文として判断するのかが難しい問題である。

【２】数式混交文

数式混交文としては、”$x \in A \cap B$よりｘはＡに含まれる”のように文中に式が埋め込まれている場合や、”$x \in (A \cap B) \cup C \Rightarrow x \in A \cap B$または$x \in C$”のように論理演算子として自然言語が使われている場合がある。さらに、自然言語が入っていなくても”$(A \cup B)^{\times} = ((A^{\times})^{\times} \cup (B^{\times})^{\times})^{\times} = ((A^{\times} \cap B^{\times})^{\times})^{\times} = A^{\times} \cap B^{\times}$”のように連続的な式変形による数式の列挙の場合もある。これらの場合、数式をどう扱うかたいへん難しい。

【３】記号

この問題は、記号の表記上の問題と、意味的な扱いの問題の２つがある。前者の問題としては、次のようなものがある。括弧については”$A \cap (B \cup C)$”のように式中に使われている場合と”Ａを（一つの）集合とする”のように自然言語の文中に使用されている場合とがある。さらに数式の部分の解析では二重括弧の付け方がユーザーによっては｛（…）｝であったり（（…））であることがある。後者の問題については、次のようなものがある。”⇒の証：”と”左辺ならば右辺が成り立つことの証明”が同じものであることを認識したり、”（１）⇒（３）の証明は（１）⇒（２）の証明においてＡとＢを交換すればよい”において（１）、（２）、（３）が式番号であることを認識しなければならない。また、”$x = \phi$すなわちｘは空集合である”というような数式と自然言語の意味的な同一をはかる必要性がでてくる。

課題：ド・モルガンの定理（１）、（２）を証明せよ

（１）　$(A \cap B)^{\times} = A^{\times} \cup B^{\times}$

（２）　$(A \cup B)^{\times} = A^{\times} \cap B^{\times}$

入力

上式（１）を示そう。1

$x \in (A \cap B)^{\times}$　⇔　$x \in X$かつ$x \notin A \cap B$

ここで、$x \notin A \cap B$⇔　”$x \in A$かつ$x \in B$”でない

⇔　$x \notin A$であるか$x \notin B$

に注意することによって次のように⇔が続けられる：

⇔　”$x \in X$かつ$x \notin A$”あるいは”$x \in X$かつ$x \notin B$”

⇔　$x \in A^{\times}$あるいは$x \in B^{\times}$

⇔　$x \in A^{\times} \cup B^{\times}$.

こうして、2 上式が示せた。3 4 下式（２）は上式との相補律から次のように導ける。

$(A \cup B)^{\times} = ((A^{\times})^{\times} \cup (B^{\times})^{\times})^{\times} = ((A^{\times} \cap B^{\times})^{\times})^{\times}$

$= A^{\times} \cap B^{\times}$.

Ｆｉｇ．１　入力例（その１）

【４】単語への分解と単語の意味の決定

構文解析をするまでに、一文を単語に、さらに単語を形態素に分解しなければならない。この時、ユーザーの意図に合うように正確に分解しなければならない。例えば、”または”という語には論理式の”∨”を表す場合や集合論で用いる”∪”を表す場合、あるいは単に接続語として使っている場合がある。このように一つの単語が複数の意味をもつような場合、その意味の決定も十分検討されなければならない。

【５】文構造の決定

文を構成する単語の意味が分かっただけではその文の意味を理解したことにはならない。各単語や、文節の意味的なつながりを把握する必要がある。たとえば、”Ａの任意の元はＢに含まれる”という文では、係り受けの関係を矢印で示すと”Ａの→元”、”任意の→元”、”元は→含まれる”、”Ｂに→含まれる”のような構造を把握しなければならない。また、接続詞は、例えば、”ゆえに”が何を受けているのかや”いずれにせよ”が何と何を指しているのか、あるいは、助詞では”ＡとＢ”、”ｘ∈Ａとする”の”と”の意味決定（ここでは並列か仮定順接か）など難しい問題である。

さらに、”$A \cap B$に属する任意の元をｘとすると、$x \in A$かつ$x \in B$”という文において、文末に”が成り立つ”あるいは”である”という述語が省略されていることや、Ｆｉｇ．２の下線部８において”これは矛盾である”のように文節（仮

課題：$A \subset B$と$A^{\times} \supset B^{\times}$は同値であることを証明せよ

入力

この 5 定理は、”$A \subset B$ならば$A^{\times} \supset B^{\times}$、又その逆、$A^{\times} \supset B^{\times}$ならば$A \subset B$も正しい”ことを意味する。まず、6 $A \subset B$ならば$A^{\times} \supset B^{\times}$であることを証明する。$x \in B^{\times}$と仮定して$x \in A^{\times}$を示せばよい。7 背理法を使う。もし$x \in A$ならば、条件$A \subset B$によって$x \in B$であるがこれは矛盾する。8 ゆえに 9 $x \in A^{\times}$。つぎに、10 $A^{\times} \supset B^{\times}$ならば$A \subset B$であることを証明する。$A^{\times} \supset B^{\times}$とすると、はじめに証明したことから、11 $(A^{\times})^{\times} \subset (B^{\times})^{\times}$である。ゆえに、補集合の定理より$A \subset B$。

Ｆｉｇ．２　入力例（その２）

定x∈Bに対して）が省略されているような場合もある。この場合、省略された部分を補う必要があるが、省略された部分を推定することは難しい。また、倒置文のような文では、係受け関係にある語が隣接しておらず極端にはなれていたり、また、受ける立場にある語が文中では前（左）にあるなど、その関係が複雑になっているため処理が難しい。

【6】指示語、指示的表現

”これ”や”前者”、”上述”のような指示語の指示する内容を決定する問題である。例えば、”ゆえに”という語が指しているのは前の一文のみか、あるいは、前記の文章全体を指しているのかを決めなければならない。また、語ではなく、Ｆｉｇ．２の下線部１１”はじめに証明したこと”や”…についても同様である”という指示的記述に対しても同様の問題がある。

【7】ＭＴＵ（Meta Techinical Utterance）

定理証明では具体的な式変形のほかにＦｉｇ．１の下線部１”示そう”や，Ｆｉｇ．２の下線部７”x∈Aを示せばよい”や”背理法により証明する”、”この問題は…を示せばよい”というような記述が含まれている。これらの記述は証明の方法を示すものであり、また、”以上で…が示せた”という文は証明全体を振り返る発言である。これらは証明の直接的なプロセスではないため、それらとは区別されたほうが学生の解答評価はより正確に行うことができる。このような直接の発話内容には関係しないが、発話内容を理解させやすくする発話をＭＴＵという。このＭＴＵと非ＭＴＵの分離も難しい問題である。

このように、入力解析には様々な問題があり、これらの問題は自然言語を扱う以上、避けることのできないものである。これらの問題はすぐに解決できるものばかりではない。しかしこれらが解決されなければ、ユーザーの意志を忠実に取り込むことのできる、ユーザーにとって使いやすいシステムをつくることはできない。以下では具体的に本システムで行ってきた解決案を考える。

【1】一文への分解

一文の切り出しは日本語の文法どおり、まず句点でわける。また画面表示幅の途中で改行がしてあれば、次が式でないかぎり、そこまでを一文とみなす。次が式であるときは【2】に従うとする。

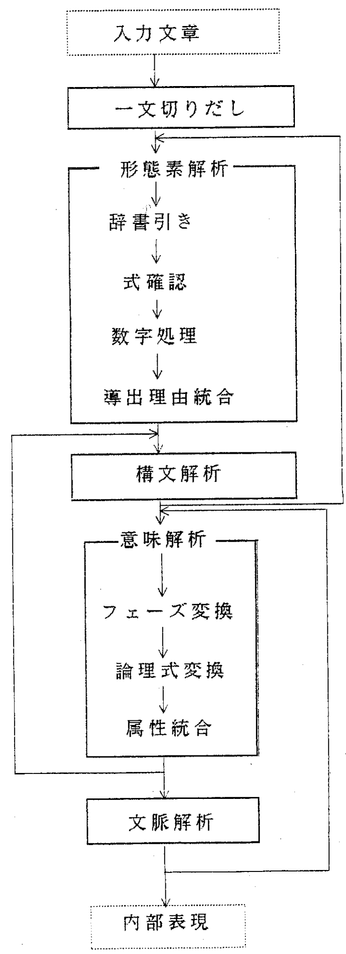
【2】数式混交文

”$(A\cup B)^{\times}=((A^{\times})^{\times}\cup(B^{\times})^{\times})^{\times}=((A^{\times}\cap B^{\times})^{\times})^{\times}=A^{\times}\cap B^{\times}$”の様な連続的な式変形は、”$(A\cup B)^{\times}=((A^{\times})^{\times}\cup(B^{\times})^{\times})^{\times}$、$((A^{\times})^{\times}\cup(B^{\times})^{\times})^{\times}=((A^{\times}\cap B^{\times})^{\times})^{\times}$、$((A^{\times}\cap B^{\times})^{\times})^{\times}=A^{\times}\cap B^{\times}$”のように、”＝”や”⇒”、”⇔”を目印に一ステップずつ分解して一文とする。

また、文中に組み込まれた式か、そうでないかの区別は式の次にくる語に着目する。そのためには、文を区切る語とそうでない語とに、予め語をクラスわけしておくことにする。たとえば、”なので”のような接続詞や、”から”のような助詞は、文を区切らない語である。”ここで”のような語は文を切る語は文を切る語である。従って、”ここで”が次ぎに語として出できたときには、”ここで”の前を文末とし、文を切り出すことが考えられる。たとえば、”x∈A回　ここで…”とあれば、”x∈A”と”ここで…”に区切る。しかし、”x∈Aなので”の場合は、”なので”は文を区切らない語なので、文中に式が組み込まれているものと考えて処理を行う。このように、語の役割わけをしておけば、文中に組み込まれた式か、そうでないかの区別ができる。この役割わけを語クラスと呼ぶ。

数式が自然言語文の中に組み込まれている場合には、自然言語を数式を説明しているメタ言語（meta-language）と考え、数式を一単語（名詞）扱いにする。これはたとえば、”A⊂Bならば$A^{\times}\supset B^{\times}$である”の場合を考えても、”人間ならば哺乳類である”と変わりはない。自然言語の側面から、数式を一単語扱いしても良い。そして、数式については、数式自体で処理を行う。数式自体の意味解析は、本システムでは前置記法としている。たとえば、”A⊂B”であれば、”（⊂　A　B）”である。

また、ローマ字漢字変換をとったとき、数式混交文の処理は、特に難しい。たとえば、”ｓｙｕｕｇｏｕｈｈａ”は、”集合っは”あるいは、”集合ｈは”の両方に解釈できる。この場合は、日本語文としての妥当性も検討しなければならないので、ローマ字漢字変換と構文解析を並列的に行うことが望ましい。本システムでは、この問題を、各処理ブロックごとにフィードバックループをもうけることにより、並列処理を行わないで、同等のことを行えるようにした



Ｆｉｇ．３　入力解析部の構成

（Ｆｉｇ．３）。つまり、形態素解析は、最長一致法を使って単語を引き、この時にローマ字漢字変換を行う。次に構文解析を行い、構文上あわない単語があればもう一度形態素解析を行うことにしている。そのため、”ｓｙｕｕｇｏｕｈｈａ”は、まず、”集合っは”に解析されるが、構文解析の結果、再度、形態素解析が行われ、”集合ｈは”となる。より詳しい具体例は３で述べる。

論理演算子の代用として自然言語を扱うときの処理は意味解析がきわめてたいへんになるので、本システムでは行っていない。

【３】記号

数式中の括弧と自然言語文の括弧の区別は、左括弧と右括弧の対応をとり、その中に自然言語が入っていたら、自然言語文の括弧として判定し、そうでなければ数式中の括弧として判別できる。その後、”（　）”と”｛　｝”の役割が同様であったなら、処理をしやすくするために内部表現では、”（　）”のみの一種類の表現に統一する。

また、”⇒の証：”や、”（１）⇒（３）の証明は…”の様な表現は、次の２つが考えられる。一つは証明の言いまわしに独特なものであるから、イディオムとして扱うことである。この場合、イディオム辞書が必要になるが、比較的、簡単に処理できる。この方法の欠点は、言いまわしが少しでも異なれば、解析不可能となることである。もう一つの方法は”⇒”の前後の語句を調べて、そこから推論することである。まず、”⇒”の記号として考えられるのは論理記号であるから、”⇒”の前後が数式か否かの判断を行えば良い。数式でなければ証明であることがわかるので、証明の記号として扱う。具体的には、その文より前で（１）、（２）、（３）の記述を探して、それらを当てはめる。本システムでは、この両者の考えを使用している。中心的にはイディオム辞書を使用し、前後からの意味解析による方法は説明した例題が行える程度の、一部の機能についてだけ実現した。

また、”ｘ＝φすなわちｘは空集合である”というような数式と自然言語による重複的な記述は、接続語の語クラスを調べ、単に換言的な記述なら内部表現では”ｘ＝φ”として統一している。

【４】単語への分解と単語の意味の決定

一つの単語が複数の意味をもつ多義語の意味決定は、その語と共起し得る語との係り受け関係をチェックする。たとえば、”へいほう”という語は”平方”と”閉包”がある。これは格文法の考え方を採用することによって解決できる。たとえば、平方は数字に対して行うものであり、閉包は集合に対して行うものであるから、語クラスを設定しておけば、係り受け関係を調べた際に判別できる。たとえば”へいほう”という語に係る方の語が、”集合”という語クラスであれば、この語は”閉包”となる。しかし、このような方法でも決定できない場合は、文脈処理を行う必要があるが、これはかなり難しいので本システムでは行っていない。

【５】文構造の決定

本システムでは、倒置文は扱わず、少なくとも係り受け関係が非交差であることを前提としている。この制限のもとで、例えば係り受け関係は次の２つの方法で解決している。第一は、例えば助詞”の”は”の”が含まれている文節が”の”の次のくる体言に係ることから決定するという、文法上の制限から係り受けの可能性を制限している。次に語クラスからの制限を行う。これは、集合論定理証明用という前提から、例えば、”集合の任意の要素”などは”集合→要素”と”任意→要素”に係る関係をあらかじめ係り受け辞書に登録しておくことで処理している。”集合→任意”という係り受け関係はこの辞書に登録されていないので、この係り受け関係は文法上認められても、成立しないことがわかる。

また、語や文節が省略されている場合、これをシステムが推定して補わなければならないが、通常、このような略語の推定は非常に多くの知識を必要とする。本システムでは、証明に独特な省略についてのみ、この推定を実現した。それはＦｉｇ．２の最後の文のように”成り立つ”が省略されているか、あるいは、”である”という述語の省略を補うことを行う。これに、数式混交文において、述語があった場合、文の最後が式であれば”成り立つ”を自動的につけることにしている。

【６】指示語、指示的表現

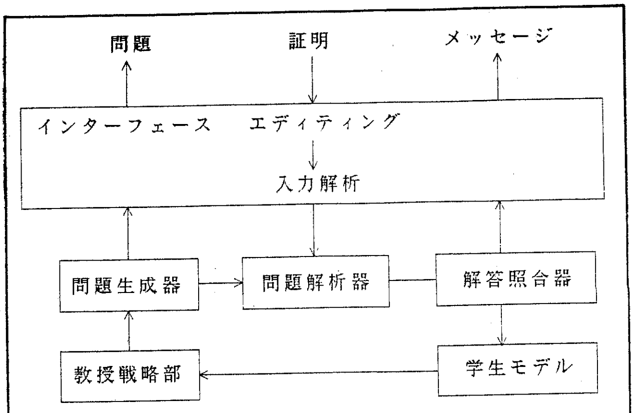
本システムでは、意味解析過程で、”これ”と”このこと”なこの簡単なものだけ次のように決めて、指示語処理を行っている。たとえば”ｘ∈Ａ∩Ｂよりｘ∈Ａである。これはｘ∉Ａに矛盾する。”という文では”これ”は”ｘ∈Ａ”を意味している。このように、”これ”という語は前の文の主語を指すとことに決めて処理できる。また、”ＡはＢに含まれ、ＢはＡに含まれる。ゆえに、このことからＡ＝Ｂが成り立つ。”のように”このこと”という表現は、その語の直前の文を指している。

”前者”、”後者”という表現はその前に並列的あるいは対比的な二つの事柄がなければならず、”かつ”、”と”、”そして”、”あるいは”、”または”、”か”という語を目印として検索することを考えているが、これは、この方法だけではうまくいかない。また、”ゆえに”という語がさしているのは前の一文のみか、あるいは前記の文章全体をさしているのか、Ｆｉｇ．２の下線部１１”はじめに証明したこと”や”…についても同様である”という指示的記述の内容の決定には、より広い範囲にわたる文脈解析を行なわなければならないので、この処理についてはほとんどできていない。

【７】ＭＴＵ

ある文がＭＴＵであるかを判断するには、たとえば、”背理法により証明する”のような定理証明に独特な言いまわしは、イディオム辞書に登録しておき、一文ずつ引き出して解析している。また、”この問題は…を示せばよい”というような記述も、定理証明では”よい”で終わる文はやはりＭＴＵと考えられる。このように、ある決まった形の文をＭＴＵとしてあ

らかじめ設定しておき、それに基づく解析を行っている。しかし、これに対する明確な方法は見つかっていない。　入力解析は、形態素解析、構文解析、意味解析、文脈解析といろいろなレベルで処理を行っていくが、一つのレベルのみの情報では問題を解決することはできない。異なるレベルからの情報を用いなければ一つのレベルの問題を処理することも難しい。本質的には入力解析は分散協調システムであるので、各レベルの処理を独立して行なうことは難しいであろう。単語の意味決定や、係受けなどの文構造の決定、文脈解析は、入力の始めから並列的に行なわれなければならない。これを実現するためにも、多くの知識やヒューリスティクスを必要とするためシステムは非常に大きくならざるを得ず、また、その処理もたいへん複雑になる。我々はこれをフィードバックループを導入したシステムで、並列処理を行わずに処理をしている。このため、各解析モジュールごとにプログラムが組め、比較的容易に入力文解析ができる。以下では、実現したシステムについて説明する。



Ｆｉｇ．４　システム全体の構成

３．定理証明向けＩＣＡＩ

本システムは集合論の定理証明向けＩＣＡＩであり、特にその入出力として数式と自然言語の混在する文章を受け付けることができるようにインターフェース部を工夫してみた。システム全体の構成図をＦｉｇ．４に示す。

我々の作成したシステムの入力の仕様は以下のとおりである。入力はキーボードから一文ずつ入力し、キーボード上にない記号はＴａｂｌｅ．１に示す記号で代用するものとする。入力はスクリーンエディタ上で行い、学生は入力終了後にファンクションキーを押す。

入力解析は、一文の切り出し、形態素解析、構文解析、意味解析、文脈解析の順でシーケンシャルに行っている。ただし、一文の切り出し作業以外は、一つ前のモジュールへのフィードバック機能を持っている（Ｆｉｇ．３）。以下、Ｆｉｇ．５を例にとって入力解析を簡単に説明する。

まず、一文切りだし作業は２．２の【１】で示したように句点を見印に切り出す。例では入力文は２文切り出されている。次に、形態素解析が行われる。ここでは、区切り文字（スペース、読点、タブ等）を目印に連文節をとらえ、その連文節内で、最長一致法を用い、単語を決定していく。例えば”ｘｗｏ”の場合は、最初の文字”ｘ”に至っても辞書に該当する単語がないので、数学上の記号であると認定する。次の”ｗｏ”は”を”とする。このようにして、単語を切り出す。これがＦｉｇ．５の処理１である。各単語の下の記号は、品詞を示す。ｎは名詞、ｖは動詞、ｃｏｎｊは接続詞である。ｋｓ，ｋｌは数学上の記号であることを示し、前者は小文字、後者は大文字である。助詞に関しては、各助詞ごとに文法上の扱いを行う方が日本語の解析がしやすいので、そのままとしている。そしてこの時点で、ユーザーにその分解結果を示し、入力ミスや、分解ミスがないかを確認する方式にしている。また、式を構成し得る記号や、”かつ”、”または”という語には、特別な語クラスを設け、一つの記号に分解した後これらをもう一度一つの式につなげる処理を設けた（処理２）。これによって、式を単語扱いにすることができる。また同時に、複合名詞の処理も行っている。

構文解析（処理３）はトップダウン形の文脈自由文法に基づいている。文法規則は、一般の日本語用ではなく、定理証明専用にしている。このようにすると、例えば”成り立つ”等の省略も文法規則に埋め込んで処理できる。このため、意味解析もたいへん楽になる。なお、処理３のｊは助詞、ｓは文の意味である。

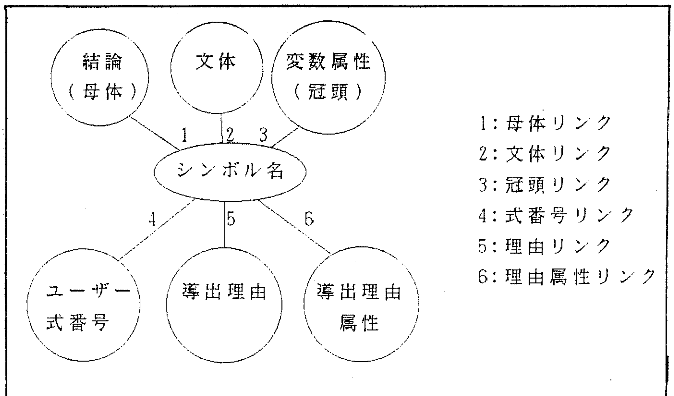
構文解析が終了したあと、意味解析過程にはいる。ここでは、”名詞句”、”動詞句”、”修飾部”とそれらにつづく”より”、”から”、”は”の様な助詞、そして、各語の意味を目印とした。例えば、”ｘをＡ∩Ｂの任意の要素とする”の係り受け関係は次のようにして調べる。まず、構文解析の結果、”ｘを”、”Ａ∩Ｂの””任意の”、”要素”、”と”、”する”に分解されているので、この単位（ここでは文節と呼ぶ）ごとに係り受け関係を調べる。まず、”ｘを”が、係る先を次ぎに続く文節の中から探す。”ｘを”の助詞”を”は、”の”がつく文節には係らないので、次の”任意の”で調べる。これも同様であるから次の”要素”を調べ、これが係る先の文節と判断される。同様に”Ａ∩Ｂ”も調べる。助詞”の”の場合は、”の”がつく文節に係れるが、”Ａ∩Ｂ”は”任意”に係れないという意味的規則から、次の”要素”に係る。このようにして係り受け関係は判定される（処理４）。それをもとに、例の文が意味的に”ｘ∈Ａ∩Ｂ　(all　ｘ)”であることがわかる。また、次の文は、導出理由を示す

もとの文字	代用文字
∈	／
∉	／／
⇒	－＞
⊂	＜
⊃	＞
∩	＊
∪	＋

Ｔａｂｌｅ．１　代用文字一覧

助詞”より”より、その前までが導出理由であったことを判定している。これらのことから各文ごとにＦｉｇ．６に示す内部表現に変換している。システムが作成した解答と学習者の解答との照合のために、入力文の結論の部分は一階述語表現に変換される。

他のモジュールの説明は割愛する。



Ｆｉｇ．６　内部表現

４．結論

システムはまだ、解答照合器が未完成なので知的ＣＡＩシステムとしては完成していない。問題生成機能や問題解決機能そしてインターフェースの各部につて試行されているだけである。したがって、学習者に試していない。ここでは、本報告で示してきたインターフェース部の評価だけを述べる。文献［７－１０］の集合論の５６個の問題の解答を試してみた結果、このうち３８問について解析は成功した。失敗の原因は主に２つある。一つは、ＭＴＵ、非ＭＴＵが分離できなかったことである。また、もう一つは、文脈解析が不十分なので、”ゆえに”などの指示子や、”はじめに示したこと”のような指示的表現に対しては、どこに係るかを十分決定できなかったことである。証明では文脈がかなり重要であるので文脈処理は、今後力をいれていく予定である。しかし、簡単な例題については、本報告の範囲で十分に解析できる。また、定理証明など、その使用する範囲を限定して文法や辞書をくめば、意味解析のときに便利なようにする事ができる。このことは、トータルな自然言語処理を行うシステムを簡便に作る１つの方法であると考えている。

さらに、これらの機能を組み入れた知的ＣＡＩシステムを早期に開発し、学生に使用させて、知的ＣＡＩシステムとしての自然言語処理機能には何が必要であり、何が重要なのかを見きわめたいと思っている。

［参考文献］

[1]宇都宮敏男、坂元昂監修：教育情報科学、１、教育とシステム、第１法規、(1988)

[2]大槻説乎、山本米男：知的ＣＡＩのパラダイムと実現環境、情報処理,29,11,pp,1255-1265,(1988)

[3]山本米男、岡本敏雄監訳：人工知能と知的ＣＡＩシステム、講談社サイエンティフィック、(1987)

[4]Burton,R.R Burton,J.S : "An investigation of computer coaching for informal learning activities", Int.J. Man-Machine Stud.,11,pp.5-24,(1979)

[5]Brown,J.S:"Uses of artificial intelligence and advanced computer technology in education", In R.J.Steidel and M.Rubin(Eds.), Computers and communications:Implications for education, New York Academic Press,pp253-270,(1977)

[6]Goldstein,I:"The genetic epistemology of rule systems" Int.J.of Man-Machine Stud.11,pp51-77.(1979)

[7]内田：数学シリーズ　集合と位相，PP.7-13,187-188，裳華房　(1986)

[8]竹之内、船越：入門　集合と位相演習，PP.25-39，実教出版　(1980)

[9]Ｓ．Ｌipschuts：マグロウヒル大学演習シリーズ　集合論，PP.24-31，マグロウヒル好学社　(1982)

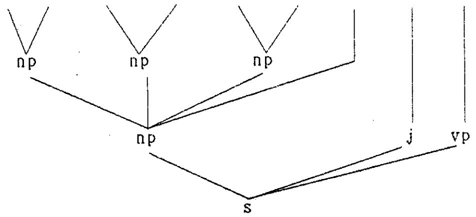
[10]小林：現代数学レクチャーシリーズＡ－１　集合と位相，PP.9-19，培風館　(1977)


```
入力：ｘｗｏＡ＊Ｂｎｏｎｉｎｉｎｏｙｏｕｓｏｔｏｓｕｒｕ．
      ｙｕｅｎｉｋｙｏｕｔｕｕｙｏｕｓｏｎｏｔｅｉｇｉｙｏｒｉｘｈａＡｎｉｆｕｋｕａｍｒｅｒｕ．

        （ｘをＡ∩Ｂの任意の要素とする。             ）
        （ゆえに共通要素の定義よりｘはＡに含まれる。）

  処理１  ｘ－を－Ａ－∩－Ｂ－の－任意－の－要素－と－する
          (ks  wo  kl  &  kl  no  n    no  n    to  v  )

          ゆえに－共通－要素－の－定義－より－ｘ－は－Ａ－に－含まれる
          (conj   n     n    no  n    yori  ks  ha  kl  ni  v    )
```

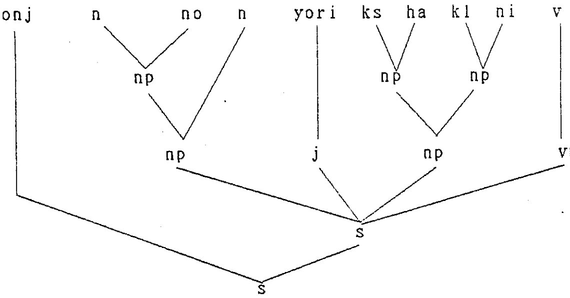
処理２　ｘ－を－Ａ∩Ｂ－の－任意－の－要素－と－する
(ks wo siki no n no n to v)

ゆえに－共通要素－の－定義－より－ｘ－は－Ａ－に－含まれる
(conj n no n yori ks ha kl ni v)

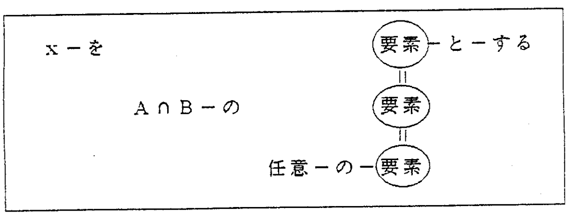
処理３　ｘ－を－Ａ∩Ｂ－の－任意－の－要素－と－する
(ks wo siki no n no n to v)



ゆえに－共通要素－の－定義－より－ｘ－は－Ａ－に－含まれる
(conj n no n yori ks ha kl ni v)



処理４　ｘ－を－Ａ∩Ｂ－の－任意－の－要素－と－する



↓

ｘ　Ａ∩Ｂ　(all ｘ)

ゆえに－共通要素－の－定義－より－ｘ－は－Ａ－に－含まれる

共通要素－の－定義－より　ｘ－は－Ａ－に－含まれる

導出理由　　　　ｘ　Ａ　(free ｘ)

Ｆｉｇ．５　入力解析例