人工知能学会誌 Vol. 3 No. 3 (May 1988)〔別刷〕

特 集「次世代自然言語処理技術」

言語理解研究の諸相

Aspects of Language Understanding Study

田中 穂積*
Hozumi Tanaka

* 東京工業大学工学部情報工学科
Dept. of Computer Science, Faculty of Eng., Tokyo Institute of Technology.

1988年3月2日 受理

Keywords: natural language understanding, thesaurus, HPSG incremental disambiguation.

1. はじめに

コンピュータによる(自然)言語理解が真の意味で可能かどうかは議論の分かれるところだろう．しかし不可能であるとする立場に立つとしても，コンピュータによる自然言語理解の研究が無意味であると断ずる人は少ないだろう．それはこの種の研究により言語理解の本質に漸近的に接近し，部分的であるにせよ，以下の成果が期待されるからである．

(1) 自然言語を理解して応答するシステム(自然言語理解システム；Natural Language Understanding System)や，機械翻訳システムの作成．
(2) 人間の行う自然言語理解過程の解明．
(3) 大量の言語データの統計分析．
(4) コンピュータの将来像への知見．

(1)は人工知能の主要な研究課題である．(2)は認知科学の主要な研究課題である．(3)は言語学の研究を側面から支える重要な研究であるが，以下では(3)についてはふれない．

自然言語理解システムの概観をつかんでいただくために，図1に自然言語理解システムのアーキテクチャを示す．自然言語理解システムを構築するうえで最も重要な部分は，言語解析を行う部分である．図1から明らかなように，言語解析により文の言語的な処理を行い意味構造を作り出す．意味構造は，言語を理解した結果であると考えてよいだろう．このとき，推論マシンの助けを借りて，中央にある知識ベース中の知識を使うことになる．このように自然言語理解は，知識と推論の問題とをきりはなして論じるわけにはいかない．そこで2章では自然言語理解と知識に関する問題，特に言語解析におけるシソーラスの利用と効用について論じる．そして3章でシソーラス作成の問題点と解決策を探る．4章では最近の言語理論と言語解析の問題を述べる．5章では言語理解結果の内部表現はどのようなものでなければならないかを考察する．6章では言語理解研究の新動向を紹介する．

2. 言語理解と上位/下位関係，部分/全体関係シソーラス

言語解析は言語理解の中核に位置するが，言語解析には概念の上位/下位，部分/全体関係についての知識が必要になる．以下では具体例をあげて説明する．

2・1 格フレームと意味解析

自然言語解析では，文を統語解析したとき，文法規則に含まれる曖昧さのために多数の解析結果が得られてしまうことが問題である．長文の統語解析結果として，数百の結果が得られてしまうことも珍しくない．どのようにして解析結果の数を絞るかという問題に対する悪戦苦闘の歴史が，言語解析研究の歴史であったといってもよいだろう．言語解析の研究が意味論に傾斜していった理由もこの辺りにある．意味的な情報を用いて統語解析結果を少数に絞り込もうとしたからである．

意味は，格フレームに似た枠組みで解析し，このとき，意味マーカを用いることが多い[21]．意味マーカについては，コンセンサスが得られた体系が存在しないことも問題であるが，筆者は，意味マーカ以上に微

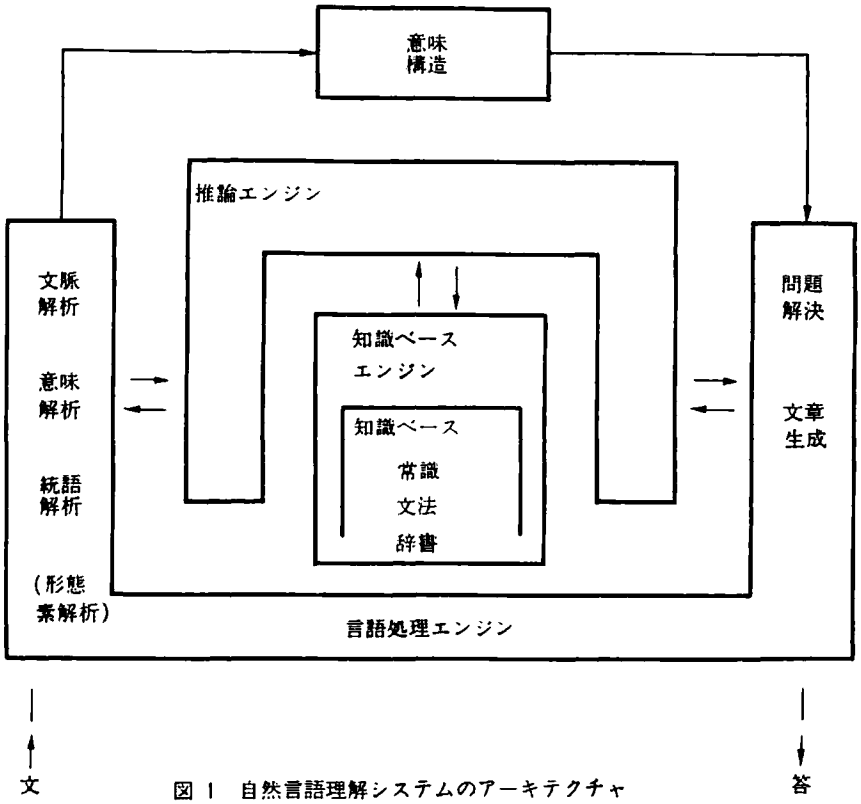

図 1　自然言語理解システムのアーキテクチャ

細な構造をもつ上位/下位関係シソーラスを開発し，利用すべきであると考えている．たとえば，

①　『彼は犬を飼っている．』

という文を意味解析して，「飼う」の対象格が「犬」で，「飼う」の動作主格が「彼」である，という深層格構造を抽出するためには，「飼う」の対象格や動作主格になるべきものが，それぞれ「人間以外の動物」や「人間」であるといった知識を使わなければならない．こうした知識は，動詞「飼う」の格フレーム中に記述される．この格フレームを用いて「彼は犬を飼っている」という文の「彼」が「人間」であり，「犬」が「人間以外の動物」であるという上位/下位関係の知識，言い換えると上位/下位関係シソーラスを使った推論を行って，先に述べた深層格構造を抽出することができる．

2・2　助詞「と」で結ばれた並立関係の抽出

並立の助詞「と」で結ばれた名詞句中の名詞のどれとどれとが並立するかを決めるためには，統語情報だけでは不十分で意味的な情報を用いなければならない．そのとき使う意味情報は，意味マーカの詳細さのレベルを越える．

②と③の例文を考えてみよう．

②　『日本の動物と植物の歴史を調べる．』

③　『東洋の仏教と西洋のキリスト教を調べる．』

文②では「動物」と「植物」とが並立する．一方，文③では「仏教」と「キリスト教」とが並立する．このような並立する名詞を決めるために，上位/下位関係のシソーラスを利用することができる．

たとえばシソーラス上で，「動物」と「植物」の親概念に「生物」が，「キリスト教」と「仏教」の親概念に「宗教」が，「東洋」と「西洋」の親概念に「自然界の領域」が位置しているものとしよう．このシソーラスを利用すると，次のようにして文②と③に含まれる名詞で，互いに並立するものを決めることができる．まず文③では，「動物」と並立する可能性がある名詞は「植物」と「歴史」である．これは統語解析により容易に知ることができる．しかし，われわれの上位/下位関係シソーラスによれば，「動物」と「植物」とは親概念「生物」を共有するが，「動物」と「歴史」とは如何なる上位概念をも共有しない．なぜなら，前者「動物」の最上位概念は「具体物」で，後者「歴史」の最上位概念は「活動」と「現象」で相互に排他的な関係にあるからである．このように上位/下位関係シソーラスを用いて，「動物」と意味的に近いものは「歴史」ではなく「植物」であると解釈することができる．文③でも同様に，「仏教」と意味的に近いものは「西洋」ではなく「キリスト教」であると解釈することができる．以上のようにして並立の助詞「と」で結ばれる名詞がどれであるかを決めることができる[34]．

2・3　助詞「の」の意味機能

部分/全体関係が意味解析で必要になることは，次の例を考えるとわかる．

④　『東京　の　目黒区』

⑤　『目黒区　の　東京』

名詞句④，⑤ともに「名詞$_1$＋の＋名詞$_2$」の形をしているが，④は日本語として意味的に問題のない名詞句であるのに対して，⑤は意味的に異常な名詞句である．日本語では，「名詞$_1$＋の＋名詞$_2$」という名詞句では，名詞$_1$と名詞$_2$の間に部分/全体関係が成り立つとき，先行する名詞$_1$が全体で，後続する名詞句$_2$が部分を表す場合には意味的に適格な名詞句である．⑤は，その順序が逆のために，意味的に不適格な名詞句になっている．意味的に適格な文とそうでない文とを見分けることは意味解析の基本であるから，部分/全体関係シソーラスは，意味解析を行うための有用な知識であると考えられる[30]．そして，それにより④に現れる助詞「の」の意味的な機能を決めることができる．

2・4　形容詞の修飾先

筆者らがかつて提案した「名詞関係先決の原則」は，「名詞$_1$＋の＋形容詞連体形＋名詞$_2$」という品詞連鎖パターンがあったとき，形容詞連体形が名詞$_1$と名詞$_2$のどちらを修飾するかを決める原則である[31]．この原則の適用にあたり，部分/全体関係シソーラスを使うことがある．次の⑥と⑦を考えてみよう．

⑥　『髪　の　長い　花子』

⑦　『花子　の　長い　髪』

⑥と⑦は，名詞$_1$と名詞$_2$とを入れ換えたものである．形容詞連体形「長い」は，⑥では先行する名詞$_1$の「髪」を，また⑦では後続する名詞$_2$の「髪」を修飾している．修飾の向きが逆転するのである．しかし，前後の名詞を入れ換えた⑥，⑦とも，2つの名詞間に「花子の髪」という意味的依存関係が保存されていることに注意したい．以上から次の「名詞関係先決の原則」が成り立つ．

「名詞$_1$の名詞$_2$」という意味的依存関係をはじめに調べて，それが成り立てば，形容詞連体形は名詞$_2$を修飾する．逆に「名詞$_2$の名詞$_1$」という意味的依存関係が成り立てば，形容詞連体形は名詞$_1$を修飾する．また「名詞$_1$の名詞$_2$」と「名詞$_2$の名詞$_1$」の双方が成り立てば，形容詞連体形は名詞$_1$，名詞$_2$の何れをも修飾可能性なことがあり，曖昧になることがある．

このように「名詞$_1$の名詞$_2$」や「名詞$_2$の名詞$_1$」といった意味的依存関係が名詞$_1$と名詞$_2$の間に成り立つかどうかをはじめに調べるために，名詞$_1$と名詞$_2$の間に，部分/全体関係が成り立つかどうかを見ることが（完全ではないにしても）有用であることは⑥，⑦の例から明らかだろう．

2・5　照応関係の解析

解析すべき文に表れる名詞句が，前後の文の名詞句などとどのような照応関係をもつかを決めることは，文脈処理の困難な問題の1つである．さまざまな言語的な知識や常識などを用いなければならないからである．照応関係の解析に上位/下位関係，部分/全体関係シソーラスが役立つことがある．例をあげる[31]．

⑧　『すずめがいる．その鳥は足に怪我をしている．』

⑨　『花子は自転車を買ったが，サドルが高くて乗りにくかった．』

⑧の「その鳥」は前文の「すずめ」を指す．「鳥」は「すずめ」の上位概念であると考えられる．このように前文にあるものを後方から参照する場合，繰り返しを避けるために上位概念を用いることがしばしばある．英語やドイツ語でも同様な現象があることが知られている．次に⑨の「サドル」は，前文の「自転車」の部分の「サドル」を指している．自然言語理解を行うためには，このような照応関係を解析する必要があるが，そのとき上位/下位関係，部分/全体関係シソーラスが役立つことがある．

2・6　機械翻訳における訳語選択

ソース言語のレベルでみると，語源が同じであるという理由で1つの辞書項目に納められているものが，ターゲット言語では意義が異なるとして，複数個の訳語が対応している場合の問題を，動詞 take の訳し分けを例にして考えてみよう[32]．

⑩　『I take a plane.』

文⑩の場合には，take の目的語が「乗り物」であれば take を「乗る」と訳すという知識を用いることになるだろう．そのためには plane が「乗り物」であるという知識の利用，言い換えると上位/下位関係シソーラスを用いた推論が必要になる．

われわれは，訳語選択には意味マーカは十分ではないと考えている．訳し分けに当たって，意味マーカ以上に微細な概念が必要になることが多いからである．このことを見るために，次の⑪と⑫に含まれる take の訳し分けを考えてみよう．

⑪　『I take a cup of water.』

⑫　『I take medicine.』

文⑪の場合には，意味マーカを使って，目的語が+liquid という意味マーカをもてば take を「飲む」と訳す，という知識を記述することは可能だろう．しかし文⑫については，目的語に貼るべき意味マーカが用意されていないとみるのが妥当であろう．そこで目的語を直接指定して，take の目的語が medicine なら take を「飲む」と訳すという翻訳用知識を組み込むことも考えられる（語直接指定方式）．ところが，目的語には pill（−liquid という意味マーカをもつことに注意）とか薬品名がくるなど際限がない．以上の理由から語直接指定方式と意味マーカを用いた訳語選択方式は，いずれも好ましい方法であるとはいえない．

この問題を解決するためには，約 50 の意味マーカ[21]による粗い意味分類ではなく，もう少し微細で大量の概念の分類体系が必要になる．特に，上位/下位関係シソーラスを用意しておくことは，訳語選択に極めて有効である．文⑪と⑫の例では，目的語が「薬」の場合に take を「飲む」と訳す，という訳語選択用の一般的な知識と，medicine や pill の上位に「薬」が位置する上位/下位関係シソーラスを利用した推論を行い，適切な訳語選択を行うことができるからである．これまで，大量の概念の分類は困難であるという理由から，意味マーカによる粗い分類と語直接指定方式とがとられてきた．しかし，より高度な機械翻訳システムを開発するためには，大量の概念を分類したシソーラスの作成が不可欠であろう．そのようなシソーラスの階層の上位には，意味マーカの体系が位置することになるだろう．

3．シソーラスの作成

自然言語理解過程でシソーラスが重要な役割を果たすことを前章で説明した．それではシソーラスを作成する場合にどのような問題があるかを考えてみよう．過去においても Roget のシソーラス[26]，分類語彙表[8]，角川類語新辞典[22]などのシソーラスが作られてきた．しかし，残念ながらそれらは意味解析の観点を考慮に入れたシソーラスではない．連想されるものをただまとめただけで，まとめたもの相互間が決められた関係（たとえば上位/下位関係や部分/全体関係）をもつとして作成されていない．前章で説明したように，意味解析を行うためには上位/下位関係，部分/全体関係などといった，概念と概念との間の関係を明確にしたシソーラスが必要である[28]．

このような階層関係のシソーラスを作ることは容易なことではない[10][24][33]．たとえば，上位/下位関係のシソーラスを，語彙項目（単語名）をそのまま用いて作成することを考えてみよう．一般に語彙項目はさまざまな語義をもつ．「桜」という語彙項目は，「花としての桜」「樹木としての桜」「だまし役としてのさくら」といった語義をもっている．したがって，語彙項目をそのまま用いてシソーラスを作成すると，語義に関する曖昧性のために，シソーラスに矛盾が含まれ収拾がつかなくなることが予想される．そこで語義を対象にしたシソーラスの作成が考えられる．これについて問題がないかどうかみてみよう．

直ちに考えられることは，語彙項目が数万にも及ぶ場合には，語義の数はその数倍にもなるため，シソーラスの作成が一段と困難になることが予想される．シソーラスの作成者は量の問題を克服しなければならない．これは困難であるからといって避けてはならないとする考えもあろう．そこで，この立場に立ってシソーラスを作成する場合に直面する問題にどのようなものがあるかをさらに検討してみよう．まず語義を対象にしたシソーラスとして矛盾を含まぬ体系の構築は不可能のように思える．なぜなら，(1)われわれがもつ概念の体系は完全ではなく常に例外的な事実が含まれている．しかも，(2)シソーラス作成時にわれわれが誤りを犯すことは避けられない．

(1)については，例外に対処する仕組みを考えればよい．例外リンクを導入するなどして解決するのである（具体例については文献 (34), (35) を参照されたい）．(2)に対処するためにわれわれは，シソーラス作成作業を支援するツールの作成が必要であると考えている．シソーラス作成支援ツールの中には，矛盾を発見しやすい手だてを組み込んでおくことのほかに，人間の読む市販の国語辞典が即座に，しかもあらゆる角度から検索可能にしておくことが望まれる[9]．国語辞典には，シソーラス作成に必要なノウハウが多数詰め込まれているので，それを利用しない手はないからである．

必要なら現在作成中のシソーラスの階層構造を木状に表示したりする機能も必要になるだろう．このようなシソーラス作成支援ツールを十分に生かしながら，慎重に作業を進める必要がある．シソーラスの作成には，作成者の言語的な知識と，とぎすまされた言語感覚に依存しなければならない．また作成者の作業を補助するものとして，数百万からなる例文集を用意し，それらがさまざまな角度から直ちに検索できるシステムを構築することも長期的には重要である．

それでは，労力をかけさえすれば質がよくしかも大規模なシソーラスの作成が可能と言えるだろうか．残念ながら答えは否である．シソーラス作成の基本にか

かわる問題が2つ残されている.

第1は,語義をシソーラス作成の基本要素として導入することの是非である.これは概念体系を構成する基本要素は何かという問題を提起する.たとえば,「桜」は「花としての桜」と「樹木としての桜」という少なくとも2つの語義をもつとしても,これらは果たして基本要素と言えるだろうか.これらはもっと基本的な要素を設定し,その組合せで表現できる可能性はないだろうか.基本要素とそうでない要素とはどのようにして区別されるか.両者に明確な境界はあるのか,などといった問題がある.

第2は,語義が異なるといっても,語義相互間に共通点が見られることが多い.語義相互を横断的に見て,それらの間の類似性を見るという視点が,多数の語義に細分化することで失われてしまう.Lakoff は比喩理解を例にして,語彙項目を必要以上に細かい語義に分割して,多数の同音異義語を設定する立場を厳しく批判している[40].

以上の問題はあるにしても,その解決策がいつ得られるか定かではない.当面,辞書に現れる語義をベースにしたシソーラスを作ることは有意義であろう.語義をベースにしたシソーラスを作成することにより,上の2つの問題がより鮮明になり,それらの解決に寄与するということも十分ありうる.

4. 言語理論と言語解析

言語解析は言語理論にも影響される.最近の言語理論では,文法規則のほかに,語彙のもつ知識を重要視する.語彙とは単語の集まりのことであるが,これまで語彙の研究は,その記述量が膨大であること,地道な用例分析の積み上げを長年にわたって多数行う必要があることから,言語学者の多くはこれを泥沼とみなし敬遠してきたように思われる.ところが,言語学の新しい潮流として非変形文法の枠組みが注目されるようになってきて,事情が少しずつ変わりつつある[5][27].その代表例として LFG (Lexical Functional Grammar) がある.また GPSG (Generalized Phrase Structure Grammar), HPSG (Head Driven Phrase Structure Grammar) がある[7].

LFG は,各語彙項目にかなりの情報をもたせておくことにより,これまで説明が困難であった幾つかの言語現象をきれいに説明し直すことが可能になることを明らかにしている.この新しい言語学では,語彙に書かれる知識が,言語学的な現象を説明する鍵になる,という考え方に基づいている.従来の言語学では,語彙に書かれる知識の量が貧困であったために,理論的な困難さが生まれてきたとするのである.こうした考え方の1つの極が HPSG と呼ばれる言語理論である.これは,語彙項目の内容を豊富にすることにより,親 (M) と子 (H, C) の関係からなるただ2つの文法規則 (M→H C, M→C H) で済ませることができると主張する.子は頭部 (Head) と補部 (Compliment) の対からなる.このようにすることにより,語順が自由な言語が扱いやすくなる.さまざまな語順に対応した文法規則をいちいち用意しなくてすむからである.日本語の語順が比較的自由であることを考慮して,日本語解析に適した HPSG ベースの,JPSG と呼ばれる文法が,第5世代コンピュータ計画で開発されている[7].

一般に,文法規則の数を増やせば各語彙項目に記載する内容を減らすことができ,それを扱うプログラムが単純になる.一方,文法規則の数を減らせば各語彙項目の記載内容を増やさなければならず,それを扱うプログラムは複雑になる.文法規則の数と語彙項目の記載内容量との間にはトレードオフがある.おそらく最適な解は両者の中間にある.筆者は何千という数の文法規則を作ることには賛成しない.300以下の文法規則数であれば,解析のスピードに問題はないし,われわれの手に余ることはないだろう.われわれの経験では,最初370ほどであった英語の文法規則を170に減らし,しかもカバーできる文の種類を増やすことができた.このことは,200ほどの文法規則数で相当広範な言語現象がカバーできることを示唆している.

とはいえ,各語彙項目にどのような知識をどのような形式で記述すべきかについての研究が言語学者によって十分になされているとはまだいえない.しかも言語学者の興味は現在のところ統語的な知識に限られている.ところが,語彙項目に書かれている知識で最も重要なものは意味に関する知識だろう.意味解析に関連させてこの問題はすでに2,3章で論じた.

以上に述べた非変形文法の枠組みのもう1つの特徴は,ユニフィケーションをベースにしている点である.ユニフィケーションはパターン照合の1つであるが,Prolog に代表されるロジックプログラム言語の基本計算機構がユニフィケーションであることを考慮すると,これらの文法をプログラムとして組み込むことが容易になることが予想される.ロジックプログラム言語と言語解析の整合性の良さについては文献 (4), (14), (15), (25), (35) を参照してほしい.

ここで自然言語解析の立場からの HPSG 理論の問題点を指摘しておこう.HPSG では頭部と補部がど

れであるかを見分けなければならないが，それが容易でないことがある．たとえば2・2節で述べたように，「AのBとCのD」という名詞句では，頭部に相当する部分がBとDなのか，それともDだけなのか明らかではない．どの名詞句とどの名詞句とが並立するかが決まらない限り，どれが頭部でどれが補部であるかを決めることができない．それには意味解析が必要になるということを既に述べた．言語学者はこの問題を深く追求しない．しかし，言語理解システムの作成者は，この問題の解決なしにはシステムが作成できないことを知っている[36]．

人間は言語解析を決定的に行っているとする考え方がある．そこで数語を先読みするというトリックを用いて，統語解析を決定的に行うことが考えられた[13]．このシステムはPARSIFALと呼ばれており，チョムスキーの痕跡理論をベースにしている．筆者は先読みという考え方それ自体に反対するものではない．しかし，人間が言語解析をほぼ決定的に行える主な理由は，統語解析と意味解析とが融合しているからだと考える．意味解析の視点を無視した決定的な言語解析アルゴリズムには自ずと限界がある．なおMarcusのアルゴリズムは，LR(k)パージングの再発見であるとする指摘もある．

5．増進的曖昧さ解消問題

言語理解の研究は，人間の行う自然言語理解過程がどのようなものかという研究を促すことになった．言語学がともすればスタティックな言語現象を扱うことが多かったのに対し，人間の言語理解過程という，ダイナミズムを含む言語現象の解明の重要性が認識されるようになってきた．その延長上に，既存の正統的な心理学にあきたらない心理学者との交流があった．そして言語理解の研究者と心理学者との交流が，認知科学の誕生に大きな役割を果たしたことはよく知られている．人間とコンピュータとの間のセマンティックギャップを解消するための工学的ともいえる研究が，人間それ自身の研究に回帰してきていることは興味深い．

言語解析には，統語解析（形態素解析），意味解析，文脈解析が含まれるということを1章で説明した．これらは独立した解析であろうか．人間はこれらを融合した言語解析を行っている．そして，文の部分を読み込んだ段階で部分的な意味構造を抽出し，以後の文の解析に利用し，意味ある言語解析結果を少数に絞り込み，文を読み終わった段階で（多くの場合）ただ1つの意味構造を得ているものと推察される．それをどのような計算モデルとして実現するかは，現在でも興味ある研究課題である．

この問題に関連して，われわれが名詞句を読み込んだとき，どのような実体を頭脳の中に作りながら言語理解を行っているかという問題がある．名詞句を読み込んだ段階では，その名詞句に何が後続するかはまだわからない．したがって，われわれの頭脳の中に作り出される実体は，その名詞句に先行する文脈から得られる言語解析結果を利用して作られることになる．ところが現実の自然言語を対象としたとき，名詞句に先行する文脈から得られる情報だけでは，作るべき実体を確定することができないことがある．ある名詞句に対応する実体が，その名詞句の後方の部分の意味が明らかになって初めて確定することがあるからである．文献(16)による例をあげる．

『The particles have mass b and c.』

という文で，the particlesという先頭の名詞句を読んだ段階でどのような実体を作ったらよいだろうか．実際には，目的語であるmass b and cまで読み進むと，the particlesに対して2つの実体を作るべきことがわかるのであるが，the particlesという語を読んだ段階では，それを知ることはできない．われわれはおそらく，the particlesに対して，このparticleの個数は複数であるが，数が不定の実体（言い換えると曖昧さを包含した実体）を頭脳の中に（とにかく）作っておき，それ以後の文脈に現れるmass b and cという部分を読み込んだときに不定の数を2と定める，という処理を行っていると思われる．このことは，文を読み進むにつれて次第に意味が明確化する実体についての研究の重要性を示唆している．

不定の情報を含む実体に対する形式をどのようなものにすべきかについてはまだ十分に研究がなされていない．そのためこれまでの多くのシステムでは，文全体を読み終わり，それから統語的な全体構造を抽出し，全体を見渡すことができる時点で意味解析を始める方法をとっていた．それによればthe particlesを処理する段階で，目的語がmass b and cであるということを知ることができるから，先の問題は起こらない．しかし，これは人間の行う言語解析法とは異なっているように思える．既に述べたように，人間はおそらく，統語解析と意味解析とを融合した言語解析を行っている．文全体を読み終わり統語解析が済むまで，文の解析をまったく行わずに待つなどということはない．文を読み進む過程で，部分的な意味解析を次々に行い，文全体の解析結果を得ていると考えられる．コンピュータ内部につくられた部分的な解析結果（実体）

は，その後の解析が進むにつれて次第に明確化し精密なものになっていく．その意味で，人間は「フレーゲの原則」に近い言語解析を行っているが，「フレーゲの原則」の限界をも克服する柔軟な解析機構をもっている．

以上の問題に関して，Mellish のフィルタと呼ばれる考え方と，ICOT の向井の部分項と呼ばれる考え方が注目される[17]．状況意味論[2]にも同様な考え方(不確定項)があり，向井はその実現を目指して部分項の考え方を導入している．そして部分項の操作を中核にした CIL と呼ばれる知識表現言語を開発している．フィルタと部分項とも，解析が進むにつれて解析結果が次第に精密化する機構を取り入れているが，これは知識表現の立場からも極めて重要な考え方である．

これまで述べてきた問題を「増進的曖昧さ解消問題」と呼ぶことにしよう．「増進的曖昧さ解消問題」をまとめると以下のようになる．

(1) 曖昧さを含む実体の表現形式をどのようにするか．

(2) どのような知識を用いて曖昧さを解消するか．

(3) 文を読み進んだとき，どこまで曖昧さを解消しておくか．

(4) いつ曖昧さを完全に解消するか．

(1)については，本章で述べた方法のほかに，曖昧さ解消に役立つ制約(Constraint)を単に貯めておき，それを実体の代わりとして使うことも考えられる．実体が何かを決める必要が生じたらそれらの制約を評価して適当な実体を取り出す．これは関数型言語で使われる遅延評価の考え方に近いといえる．(2)，(3)，(4)についてはさらに研究が必要である．

6．自然言語理解を巡る新しい研究動向

自然言語理解には5章までに説明した問題のほかにもさまざまな問題がある．たとえば，否定や名詞句についた限量詞のスコープを決める問題を解決しなければならない．限量詞のスコープは，名詞句が定冠詞付きかどうか，単数か複数か，each が付いているかどうかが関係する．これに関連して Mellish[16] の研究が注目される．言語行為の問題も重要である．意味ある文のまとまりである談話理解の問題も重要である．そこでは特に照応指示表現をどのように扱うか，また省略語の補強の問題が重要になるだろう．また文の理解と発話に際し，発話者の Intension と Attension の抽出やその利用が重要とされている．しかし，これらは残念ながら言語学的にも計算モデルとしてもまだ十分研究がなされているとは言えない．

1) Discourse, Intention, and Action
2) Rational Agency
3) Situation Theory and Situation Semantics
4) Visual Communication
5) Embedded Computation
6) Representation and Reasoning
7) Situated Automata
8) Analysis of Graphical Representation
9) Foundation of Grammar
10) Linguistic Approach to Computer Language
11) System Design Language
12) Computational Models of Spoken Language
13) Finite State Morphology
14) Grammatical Theory and Discourse Structure
15) Head-driven Phrase Structure Grammar
16) Lexical Initiative
17) Phonetics and Phonology
18) Theory of Aitiation Frames

図 2 CSLI で推進されているサブプロジェクト

最近，アメリカのスタンフォード大学の CSLI 研究所で，Situated Language Research Program (SLRP) が始まり，談話理解に関する基礎的な研究が行われている[3]．これは数学者の Barwise らが主導する状況意味論を基礎におくプロジェクトである．5章でふれた HPSG, LFG などの新しい言語理論も CSLI 中のサブプロジェクトとして研究されている．そのほか，言語と情報を巡って，数学の基礎の再構築を目指す極めて野心的な研究から，新しい推論エンジンを創り出す研究に至るまで，幅広く興味あるサブプロジェクトが推進されている(図2)．人工知能の主要な研究テーマも含まれている．状況意味論では，常識に関する推論を制約としてとらえる．その意味で Schank らのスクリプトによる談話理解研究も，制約という枠組みでとらえなおしてみるとおもしろい[28]．状況意味論の研究が進展するにつれて，人工知能でのこれまでの研究成果が新たな理論的枠組みの中で再解釈され，装いを変えて再登場することは十分ありうることである．わが国の ICOT で行っている談話理解の研究は状況意味論をベースにしている．状況意味論の詳細は，制約の説明も含めて本特集号の向井氏の稿を参照していただきたい[19]．

最後にシソーラスの作成に関連して，アメリカの MCC 研究所で，百科辞典の知識を計算機に可読な知識表現形式にかえて蓄積し，従来のエキスパートシステムのもつ弱点，すなわち脆弱性を克服しようとするプロジェクトが開始した[12]．これは，百科辞典的な知識の欠如が従来のエキスパートシステムの限界になっているという認識に基づいている．言語理解には

百科辞典的な知識も必要になるから，これは言語理解の研究にとっても重要なプロジェクトである．わが国の電子化辞書プロジェクト[11][39]と同様，今後の研究成果に注目したいと思う．

7. おわりに

これまで筆者にとって興味ある自然言語理解の研究課題と動向をさまざまな側面から述べてきた．最後に自然言語処理の高速化を図るために，並列処理の導入についてふれておきたい[6]．将来 VLSI 技術が進歩するにつれて大量の計算パワーが安価に得られる時代が到来する．ところで統語解析アルゴリズムのうち横型探索を行うものは並列計算に適している[1]から，統語解析の並列処理は今後の研究成果が期待できる課題であるといえよう．並列処理としてコネクショニズムの考えが最近注目されている．これらに，期待を寄せる研究者の数が増えていると聞く．言語理解に関しては，CSLI での研究と比べると，まだ理論的基盤が弱いように思う．形式的な考え方に基づく言語理解の現在の研究レベルに，コネクショニズムをベースにした研究が果たして到達できるかどうか，これからその真価が問われることになるだろう．

◇ 参 考 文 献 ◇

(1) Aho, A. V. and Ullman, J. D. : The Theory of Parsing, Translation and Compiling, Prentice-Hall (1972).
(2) Barwise, J. and Perry, J. : Situations and Attitudes, The MIT Press (1983).
(3) Center for the Study of Language and Information : Fourth Year Report of the Situated Language Research Program, CSLI-87-11, CSLI, Stanford University (1987).
(4) Colmeraure, A. : Metamorphosis Grammar. in Bolc (ed.), Natural Language Communication with Computers, Springer-Verlag, pp. 133-190 (1978).
(5) 淵　一博 (監修) : 自然言語の基礎理論，共立出版 (1986).
(6) Fordor, J. A. : Modulality of Mind, The MIT Press (1983).
(7) 郡司隆男 : 自然言語の文法理論，産業図書 (1987).
(8) 林　大 : 分類語彙表，国立国語研究所，秀英出版 (1966).
(9) 今津英世，田中穂積 : 電子国語辞典の構成と実現，日本ソフトウェア科学会「論理と自然言語」研究会ワークショップ資料 (1988).
(10) 石崎　俊 : 言語理解の問題—文脈理解を目指した概念の体系と記述—，松山公一 (編) : 高度自然言語処理，情報処理九州シンポジウム，情報処理学会九州支部 (1987).
(11) 柿崎尚弘 : 日本電子化辞書研究所における電子化辞書の研究開発計画，Vol. 19, No. 6, pp. 34-39 (1986).
(12) Lenat, D., Prakash, M. and Shepherd, M. : CYC : Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks, *The AI magazine*, Vol. 6, No. 4, pp, 65-85 (1986).
(13) Marcus, M. : A Theory of Syntactic Recognition for Natural Language, The MIT Press (1980).
(14) Matsumoto, Y., *et al.* : BUP : A Bottom-up Parser Embedded in Prolog, *New Generation Computing*, Vol. 1, No. 2, pp. 145-158 (1983).
(15) 松本裕治，杉村領一 : 論理型言語に基づく構文解析システム SAX，コンピュータソフトウェア，Vol. 3, No. 4, pp. 4-11, 日本ソフトウェア科学会編 (1986).
(16) Mellish, C. S. : Computer Interpretation of Natural Language Descriptions, Ellis Horwood Limited (1985), 田中穂積 (訳) : 自然言語意味理解の基礎，サイエンス社 (1987).
(17) Mukai, K. : Unification over Complex Indeterminates in Prolog, ICOT TR 101 (1985).
(18) 向井国昭 : 談話理解への応用，自然言語の基礎理論，pp. 171-197 (1986).
(19) 向井国昭 : 談話理解とロジック，人工知能学会誌，Vol. 3, No. 3, pp. 289-300 (1988).
(20) 長尾　真，ほか : 科学技術庁機械翻訳プロジェクトの概要，情報処理，Vol. 26, No. 10, pp. 1203-1213 (1985).
(21) 長尾　真，淵　一博 : 論理と意味，岩波講座情報科学-7，岩波書店 (1985).
(22) 大野　晋，浜西正人 : 角川類語新辞典，角川書店 (1981).
(23) 荻野綱男 : シソーラスについて，ソフトウェア文書のための日本語処理の研究-5, pp. 1-61, 情報処理振興事業協会 (1983).
(24) 荻野孝野 : 日本語の意味分類試案，計量国語学会第 31 回大会発表資料 (1987).
(25) Pereira, F., *et al.* : Definite Clause Grammar for Language Analysis—A Survey of the Formalism and a Comparison with Augmented Transition Networks, *Artif. Intell.*, Vol. 13, pp. 231-278 (1980).
(26) Roget, P. M. : Roget's Thesaurus (New Edition), (First edition by Peter Mark Roget 1852), Longman (1982).
(27) Sells, P. : Lectures on Contemporary Syntactic Theories, CSLI, Stanford University (1986).
(28) Schank, R. C. and Abelson, R. P. : Scripts, Plans, Goals and Understanding, Hillsdale, N. J., Erlbaum (1977).
(29) 島津　明，内藤昭三，野村浩郷 : 日本語文意味構造の分類—名詞句構造を中心にして—，情報処理学会自然言語処理研究会，47-4 (1985).
(30) 島津　明，内藤昭三，野村浩郷 : 助詞「の」が結ぶ名詞の意味関係の subcategorization，情報処理学会自然言語処理研究会，3-1 (1986).
(31) 田中穂積 : 知識情報処理の展望—自然言語処理の立場から—，電子通信学会誌，Vol. 69, No. 11, pp. 1080-1086 (1986).
(32) 田中穂積，仁科喜久子 : 上位/下位関係シソーラス ISAMAP の作成〔I〕，〔II〕，情報処理学会自然言語処理研究会資料 (1987).
(33) 田中穂積，辻井潤一 (編著) : 自然言語理解，オーム社 (1988).
(34) 田村直良，田中穂積 : 意味解析に基づく並列名詞句の構造解析，情報処理学会自然言語処理研究会，59-32 (1987).
(35) 徳永健伸，岩山　真，田中穂積，上脇　正 : 自然言語解析

システム Lang LAB (投稿中).
(36) 辻井潤一：機械翻訳のための文法とその問題点，日本語の特性と機械翻訳，第1回「大学と科学」公開シンポジウム (1987).
(37) 辻井潤一：文解析方式，情報処理，Vol. 27, p. 8 (1986).
(38) 鶴丸弘昭，藤田 毅，首藤公昭，吉田 将：日本語の機械処理，電子通信学会オートマトンと言語研究会，AL76-43, pp. 41-50 (1976).
(39) 内田裕士，ほか：自然言語処理のための電子化辞書の構成法，情報処理学会第35回大会，pp. 1213-1214 (1987).
(40) Lakoff, G. and Johnson, M. : Metaphors We Live by, The Univ. of Chicago Press (1980).

著 者 紹 介

田中 穂積（正会員）

昭和39年東京工業大学理工学部制御工学科卒業．昭和41年同大学院修士課程修了．同年電気試験所（現，電子技術総合研究所）入所．昭和58年東京工業大学工学部情報工学科助教授，昭和61年同大学教授となり現在に至る．工学博士．人工知能，自然言語処理の研究に従事．情報処理学会，電子情報通信学会，認知科学会，計量国語学会各会員．