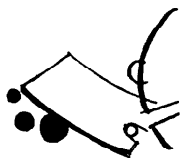


報告

パネル討論

自然言語の知的処理†



パネリスト

荻野 綱男¹⁾, 鶴丸 弘昭²⁾, 辻井 潤一³⁾, 島津 明⁴⁾
 石崎 俊⁵⁾, 司会 野村 浩郷⁶⁾

司会 「自然言語の知的処理」というタイトルはあまり聞き慣れないかとも思われますが、これは将来の高度な言語処理を考える場合に、人間の知能の働きを勘案しなければならないということでもあります。この観点から、本パネルでは将来を見越した上で、それに至るステップとして地に足を付けて各段階をどのように進めていかなければならないか、という問題を考えてみたいと思います。

「言語の知的処理の研究というのは、人間の社会生活における言語活動を種々の側面から研究して、それによって高度な言語処理技術の基盤を確立するとともに、それを応用して人間の社会生活を豊かにすることを目的とする」ということができます。それを具体的な技術として実現するために追究していかなければならない大きな問題として、1) 意味を理解する技術、2) 問題を解決する技術（たとえば、算数の問題を解くこと、質問応答を行うこと、翻訳すること、など）、3) 知識を学習する技術、があげられます。これらの3つの問題は、言語処理の技術の基本的な問題であるとともに、人間が社会生活を行う上でコミュニケーションを支える基本的な問題でもと思われる。

その中で、特に、意味理解技術に焦点を当てて考えると、「意味モデル」が基本となります。この意味モデルはさらに3つに分けて考えることができます。

第1は、句構造変形生成文法だとか、格文法だとかの従来のものを超越した高度な言語モデルです。第2は、コミュニケーションにおいて当事者同士の間仮定されている背景知識の状況モデルです。第3は、種々の情報を表現・蓄積するための枠組みの記憶モデルです。

この意味モデルの応用の一つに知識ベースがありま

す。知識ベースには、個々の言語に内在する知識（言語内知識）、異なる言語の間を結ぶ知識（言語間知識）、いわゆる知識といわれている言語外の知識（言語外知識）があります。意味モデルのもう一つの応用は処理機構であり、これには解析機構（文の解析）、変換機構（内部表現の変換）、生成機構（文の生成）の3つに分けられます。この中には当然のことながら知識を使った推論も含まれます。知識ベースと処理機構を統合しますと「意味理解システム」というものが考えられるようになり、その応用として質問応答システムや、文書作成システムや、機械翻訳システムが考えられます。

本日のパネル討論ではこの言語の知的処理のすべての側面をカバーすることはできませんので、その中から、特に、辞書に関する問題と、翻訳に使われる概念あるいは知識と、いわゆる知識を使った処理を取り上げて議論してみたいと思います。

荻野 シソーラスに関係してお話したいと思います。シソーラスとは「階層構造をなす概念体系の各部分に単語を位置づけたものである。」と定義します。従来情報検索用のシソーラスが、各専門分野ごとにたくさんあります。しかし、本パネルで議論する自然言語処理あるいは自然言語理解に役立つような一般語のシソーラスは、ないと思います。私は、シソーラスは、言葉の意味的情報すべてを扱う、というように広く考えています。

従来の言語学研究の観点は、「言語は体系をなす」であり、これは構造言語学とか、生成文法とかの20世紀の言語学の大きな潮流になっています。この中で特にシソーラスに最も関係するのは意味論ですが、そこでの普通の研究法は、単語とか、句とかの相互の意味的な関連を研究することです。

しかし、これでは、シソーラスおよび、その裏にある言語の知的処理にとって不十分です。というのは

†日時 昭和59年11月6日、7日の「自然言語」シンポジウム
 場所 機械振興会館大ホール（地下2階）

1) 埼大, 2) 長崎大, 3) 京大, 4) 武通研, 5) 電総研, 6) 武通研

第1に、言葉の表層形だけを追いかけ、その裏にあるたとえば人間の行動のもとになる心理(意図、概念、心情などを含めて考える)が捉えられなくなります。

言語学の立場では、言語を記号の一部とみなして、その体系を追求しますが、人工知能の立場ではそれではまったくだめだと思っております。対話の理解は、発話の意図、言葉の運用、言葉の機能に基づいて解析しなければ役立たないと思います。したがって、言語学者が語彙を分類するという立場でいままでに作成した一般語のソーラスでは不十分で、もっと多くの情報が必要となります。そして、ソーラスの使い方についての検討もほとんど無視されています。

にもかかわらず、現実にはこのソーラスは、自然言語理解の研究、人間の社会生活、言語生活でわりと広く使われております。人間は膨大な知識が頭の中にあるので使えますが、計算機にはそのような常識を期待できないので、いままでの言語学の立場からつくられてきたソーラスでは不十分で、これから言語の知能的処理に適したソーラスを新しくつくるべきだと考えています。

いま日本語では分類語彙表と、角川の類語新辞典という2種類のソーラスがありますが、それらの「は」から「ほ」までで始まる全単語の意味区分の数を数えてみると、平均1.054ないし1.087という個数になり、ほとんど1見出しが1つの意味に分類されていることがわかります。比較のために三省堂国語辞典・第3版の「は」から「ひ」までの意味区分の数を数えてみると、平均で1.398と意味分類が非常に細かくなっています。三省堂国語辞典の場合は、品詞が違えば別の意味区分で数えますので、多義語が多くなる傾向があります。

さらに、用法が違うと独立の意味として多義を設定するので、意味が細かく区分されます。また、語釈には「金貸し」のような簡略化した表記がありました。その意味記述は「金を貸す人」、「金を貸すこと」の2つですから、これは概念の立場からみると2つに分けたほうがいい。そうすると1.398という平均の意味区分数は決して多くはなく、むしろもう少し多くなる可能性があると思います。小型の国語辞典でもこのぐらい意味区分を細かく分けているわけですから、ソーラスとしての分類語彙表とか、類語新辞典の意味区分の数ではまだ不足ではないかと考えます。

さて、それではソーラスには一体どういうふうな情報を盛り込むといいいのでしょうか。私は、文法的情

報と重ならない意味的情報は、すべてソーラスのなかに入りうると考えます。

まず、概念あるいは概念体系上の位置がソーラスの中心的情報になると考えます。これは、従来、言語学の意味論では単語間の関係、たとえば同義関係あるいは類義関係、上下関係、反対関係、反義関係、非同立関係などとして捉えられてきたものであります。そのほかに指示物(referent)に関する情報も必要であると思います。その一つに全体部分関係というのがありますが、これは単語間の関係と捉えられる場合もありますが、正しくは指示物に関する情報だろうと私は思っております。ただ、この指示物に関する情報をどのぐらいまで記述できるかというのはなかなかむずかしくて、丁寧に書くと膨大になってしまいます。

それから、語義に関する情報。これは、いわゆる単語の意味記述であります。どういうふうに着意すればいいのかいまのところ代案は持ってありません。

それから、文体に関する情報。特に類義語間の文体差であります。私は、同義関係というものをやや広く考え、概念上の同義であれば類義語も同義語と考えますが、その場合でも文体に関する情報は別に書いておいたほうがよいと思っております。文語・口語、俗語・雅語、古語・新語、方言的、年齢差、いろいろなものがことばのニュアンスとしてまつわりついております。「めし：ごはん：食事」の例では、これらと同義語と考えることもできますが、めしは俗的な表現、食事はやや固い言い方になるので、単なる同義とするのではなく、文体差を伴った同義というぐらいに考えておきたいわけです。

そのほかに、喚情に関する情報(ことばを聞いたときの響き)、連想に関する情報(連想語彙表のようなもので十分)があります。

このようないろいろな意味的情報を全部盛り込むのが望ましいと思いますが、これをいっぺんに全部取り上げるのは大変むずかしいですから、概念体系上の位置に関する情報あたりから少しずつつくっていくことになるだろうと思っております。

このような情報は、従来言語学のなかで、特に意味論という部分で研究されてきました。その意味論の目標は、それぞれの単語なり句なりの意味を明らかにすることで、類義を整理した上でそれらの間の差異、違いを追求していく。これをつきつめると、完全な同義語は存在しないこととなります。たとえば、文体差が伴うということは、単語の意味のなかの文体の特徴が

違うということになる。しかし、機械のための意味論ではそのような細かい差異は少し目をつぶるぐらいにして分類しないと、語彙体系の全体が理解できないと考えております。

シソーラスのつくり方については、とりあえず一般の国語辞典が利用できると思います。それには、いくつかの理由がありますが、最初は、一般語の見出しの選定に有用な資料である、ということです。シソーラスに含めるべき見出しの範囲を選定するのは大変むずかしいので、とりあえずは国語辞典にある見出しを利用すると、第1次近似は十分できるだろうと思います。ただ、複合語などがかなり落ちているので、これは別途考えなければいけません。

2番目に、国語辞典には意味記述がのっているので、初めて接する単語でも意味がわかる。

3番目に、その意味記述が非常に改良されてきており、単なる同義語のいいかえでなく他の言葉でいいかえようとする努力がみられます。したがって、この意味記述には上位概念なども含まれており、このままシソーラスに翻訳できるような形が非常に多い。

4番目は、付随的な情報ですが、文体と喚情に対しても一応まとまっていることです。

鶴丸 国語辞典の計算機での利用という立場から「意味辞典作成の機械化について」少し話してみたいと思います。

実用規模の意味辞書（広い意味でのシソーラス）の機械化を考ますときの問題として、(1)膨大な数の単語の意味に関する情報（意味情報）を、何から、どのような方法で獲得するか、(2)獲得された意味情報をどのような枠組みで構造化するか、などがあります。

(2)の枠組みに関しては、たとえば、単語間にどのような意味的關係を用意すればよいか、また、単語間の階層關係を基本に考えた場合、上位語の情報が下位語に適切に伝わるように構造化するにはどうすればよいか、などの問題があります。これらは、知識の表現、単語の意味の定義の問題と関連があり、重要な問題ですが、話が広がりすぎますので、(1)の問題に絞りたいと思います。また、機械処理の対象として、一般用語から専門用語まで考える必要があるわけですが、このような膨大な数の単語の意味を同一レベルで定義するのは、単語間の關係が非常に錯綜すると考えられます。そこで、ここでは、一般用語を対象に、その意味情報を、国語辞典から計算機を利用して獲得するときの問題について考えてみたいと思います。

現在私が使っている機械可読な形式の辞書は市販の国語辞典がそのままの記述形式で磁気テープ化されたものです。そのような磁気テープデータから情報を抽出する場合には、国語辞典に含まれている情報を外部情報と内部情報に分けて考えたらどうだろうかと思えます。外部情報とは国語辞典に含まれています特殊記号とか、その辞典の編集方針、などを手がかりに求められる情報、たとえば見出し語、品詞、活用、語義文、用例、反対語などです。これらをまず構造化、データベース化しておくで非常に有用です。

内部情報とは、語義文（見出し語の意味記述文）や補足的説明などに含まれている情報のことで、意味辞書に必要と思われる意味情報の多くがここにあると考えます。ですから、階層關係など単語間の意味情報を抽出するには、語義文などの内容、構造などを調査整理しておく必要があります。たとえば、見出し語「折り尺」の語義文「…しまっておけるものさし」では文末の単語「ものさし」が、「折り尺」の上位語になっています。ここで、見出し語に階層關係のある語を定義語と呼ぶことにしますと、語義文の特徴の一つとして、文末に定義語が現れる場合が多いことがあげられます。しかし、なかには文末の語が見出し語と階層關係にない場合もあります。たとえば、見出し語「青蛙」の語義文「…大型のカエルの一種」では文末表現が「～の一種」となっており、定義語と見出し語との上位一下位關係を積極的に規定する表現になっています。また、見出し語「青柳」の語義文「…を茂ったヤナギ」の意の雅語的表現」では、定義語としては「ヤナギ」が含まれていますが、文末表現「～の雅語的表現」は先ほど荻野先生のお話の文体に関する情報を表すものと考えられます。

また、見出し語「秋虫」の語義文「スズムシ・マツムシなど」では、文末表現が「など」となっています。この場合、二つの定義語「スズムシ」と「マツムシ」があり、これらは見出し語「秋虫」の下位語になっています。このように文末に「など」という表現があったり複数個の定義語があったりすると、定義語と見出し語との階層關係が前に述べた例と逆になるということがあります。

これらの例より、語義文の第二の特徴として、定義語と見出し語との階層關係を規定するような表現が文末に表れる場合があることがあげられます。

また、語義文が見出し語の、単語によるいいかえになっている場合もあります。この場合は、一種の同義

関係と考えてよいと思われます。

そこで、このような内部情報、すなわち定義語および見出し語との階層関係を抽出するためには、語義文の文末表現を整理しておく必要があります。

さて、語義文の第三の特徴として、通常、語義文では定義語がいくつかの側面から規定されているような記述になっていることがあげられます。たとえば、見出し語「青桐」の語義文「庭木・街路樹に用いる落葉高木」では「落葉高木」が定義語、すなわちこの場合は上位語になっているわけですが、これを修飾するといえますか、規定する表現として「庭木・街路樹に用いる」があります。「…に用いる」という表現は、定義語を使用に関する側面から規定していると考えられます。国語辞典の中では、使用に関する側面は「…に用いる」のほかに、「…に使う」とか、「…のために用いる」とか、「…用」とかで表現されているようです。意味辞書では、このような側面も重要な意味情報と考えられますので、それらに対応した表現を明らかにしておく必要があります。

市販の国語辞典は、本来人の使用が前提になっていきますので、使用者が当然知っているであろうことは、省略されている場合があります。このため、国語辞典からだけでは、機械処理用の意味辞書の情報として当然不足します。たとえば、単語に多義があったり、語義文に適切な上位語が用いられてなかったりする場合がありますので、単語間の階層構造を求める場合に、途中で変な意味の単語が現われたり、適当な上位語が現われないということが起こります。欠落している情報を補うために、人間の援助が必要となります。

意味辞書の機械化には、多くの問題が山積みされています。単語の意味の定義の問題から辞書の応用までも含めた総合的な立場からの研究が必要です。

辻井 機械翻訳と知能処理の関係について、何か話題提供をするというのが私の役目です。機械翻訳と知識処理というのは、どちらも大きな分野ですが、その中で特に「語」と「概念」の関係について考えてみたいと思います。

まず、簡単な文として、「夜が遅くなったので、花子を家に送った」という文と「協議の必要が出てきたので、太郎を東京に送った」という2つの文を考えます。1番目の文では実際にこの話者が「花子について家まで行った」ということを想い浮かべますし、逆に2番目の文の場合には必ずしも話者は、太郎と一緒に東京に行ったわけではなくて、「派遣した」の意味で

あるのがわかります。すなわち、「送る」という簡単な単語をとりあげても、少し内省的に考えると、全く異なった状況を描写するのに使われることがわかります。この「送る」を英語に翻訳する場合には、それぞれ take, send に訳し分けられます。

こういう例題を考えて見ますと、このパネルの主題である知能処理あるいは知識処理、それにもとづく深い理解が質の良い機械翻訳に必須であることが、よくわかると思います。しかし、それでは一体深い理解の結果は、具体的に計算機の中でどう表現されるのか、この「送る」の2つの「意味」を区別するための記号系はどうあるべきか、ということは、現時点の自然言語処理ではよくわかっていないと思います。したがって「深い理解が必要だ」という主張はあるのですが、実際のシステムで意味を深く取り扱えば扱うほどあいまいの数が増えるという議論になってしまい、必ずしも深い理解と機械翻訳とがうまく結びつかないという状況があります。

上の例からもわかりますように、同じ単語が使われたとしてもその単語で記述されている状況として、人間ならばいくつかの違った場面が想定できます。すなわち、同じ単語が使われていても、適当な文脈が与えられると、意味の区別ができます。このように「語」と、それが記述している外界の状況は、必ずしも一対一の対応ではなくて、同じ「語」が記述する状況の側に広がりがあることとなります。

人間の場合には同じ単語で記述されている少しずつ違う状況が必要に応じていくらかでも区別できる。この区別できるという能力が翻訳の場合決定的で、単語の訳し分けの問題として出てきます。

「送る」を普通の国語辞典で調べてみると、だいたい7つほどの「意味」が列挙されています。その中には、1)「物」を送る、2)「人」を案内して送りどける、といった意味があって、前述の2の例では、「夜が遅く…」の場合の「送る」が2の用法、「協議の必要が生じたので…」の場合が1の用法となります。

しかし、この2つの意味は目的語が「物」か「人」で分かれているのではなくて、この状況において、「話者が一緒にいて行ったかどうか」によって分かれています。したがって、計算機はこの状況を適切に把握していないとこれらの意味の違いが分からないこととなります。

この辞書では、「送る」の用法として、「人を派遣する」の用法を1の特殊な場合として取り扱っているわ

けですが、たとえば、「手紙を送る」とは言えても、「手紙を派遣する」とはいえないわけで、厳密には、この1の用法は、概念・意味の立場からすると、もっと細分化されることになります。また、同じようなことですが、「合図を送る」といった用法も、この辞書では、1の用法の比喩的用法に含めていますが、「手紙を送る」とは随分と違った状況を表していることは明らかです。その状況の差に応じて、「手紙を送る」と「合図を送る」に違った語を対応させる言語があっても、ちっとも不思議ではない、ということになります。

このように考えてゆきますと、普通、われわれは「シソーラスは、概念をもとに構成すべきだ」とよく主張しますが、「語」と「概念」とはどのように関係しているのか、また、そもそも「概念」とはなんなのかが、あまり明確ではないことがよく判ると思います。「語」の表す概念や意味を、その「語」の指示する外界世界との対応で考え、外界世界でなんらかの差があればすべて異なった意味・概念であるという立場に立つと、「語」の意味は無限に細分化されてゆくことになります。

このような問題を考えていくと、機械翻訳以外の自然言語処理での知識処理と機械翻訳のそれとでは、少し様相が違うことがわかります。

よくいわれていますように、データベースのフロントエンドとしての自然語処理の場合ですと、対象分野の限定がかなり大きいわけですが、機械翻訳の場合には、そういう限定が比較的少なくなっています。この対象の分野の広さの差は直接的には量的側面、たとえば、単語の数が増えるとか、表現の種類が増えてくるということになるわけですが、このような量的な差だけでなく、質的な側面で差が出てきます。

それは、「送る」を例にとって話をしてきましたように、ある言語表現、あるいは単語といったものとそれが表す外界世界との関係が、かなりばやけてくるということです。

機械翻訳の場合には、この外界での状況の差が直接訳語の差になってあらわれることがあります。状況の差を別の単語で表現するかどうかは、言語に依存するわけで、一方の言語では同一の単語で表現できることが、他方の言語では2つの単語で区分して表現しなければならぬ、といったことがよく起こります。

たとえば、ごく卑近な例ですが、「テニスをする」とか、「ピアノをひく」とか、「ドラムをたたく」を取り上げますと、こちらの英語の世界では、基本的には

play という1語で表されていることが、日本語の世界では、「弾く」とか、「たたく」とか、「する」ということばに使い分けられる。この差をどこで認識するかというと、結局この「プレイ」という単語が、スポーツという性質をもった単語とつながっていると、楽器という性質をもった単語とつながっていると、というような差でもって、相手の言語表現へどう移すかという訳語の選択をせざるをえないわけですが。

機械翻訳と意味処理、あるいは自然言語処理と意味処理という関係を考えますと、基本的に、こういう意味の世界を表現するのに、どういう記号系を使うかということが問題になるわけですが、それには二つの立場があります。

1つは、言語表現をより抽象的な縮退したレベルで捉える立場です。たとえば Schank の概念素のように、表面上ものすごくたくさんある単語の意味をごく少数のプリミティブで記述する、そういった縮退したレベルで捉える立場。もう1つは、意味というときに、言語表現をさらに細分化する立場です。「送る」という単語を「送る1」「送る2」「送る3」「送る4」「送る5」はすべて違う意味である、というふうに、より細かな概念に細分化してやるという立場です。

現在の自然言語処理、意味処理を論じるときに、この2つの立場が混同されているように思います。機械翻訳の場合でも、中間表現としてのピポット表現を考えるときに、縮退したレベルで意味を捉えようという立場の人と、単語よりも細かい概念語のレベルで捉えようという立場の人がいます。

われわれがトランスファ・アプローチでの機械翻訳を主張するのは、このどちらの立場も機械翻訳システムの間言語表現とするには、不十分だと考えているわけです。ただ、現在、私が考えておりますのは、どうも縮退したレベルというのはかなり大事ではないか、ということです。縮退したレベルで捉える第一歩は、意味素性の考え方ですが、これは単語の意味全体を表現するのではなくて、単語の意味の「ある側面」を抽象化して、それを素性で表現するという立場です。意味とか、知識の規則を素性のレベルで書くというのは、ちょうど文法的な規則性が単語を直接参照することで書かれるわけではなくて、たとえば、名詞とか動詞とかの統語的な素性を使って書いているのと同じことです。

こういうことが、意味のレベルでもやはり必要で、ある単語の意味を完全に表現するための意味表現では

なくて、単語のもっている意味のある側面を抽象した表現というのを考えないと、言語表現をどんどん細分化していったら、概念だといわれている部分で規則を立てようと思うと、とんでもないことになるのではないかと考えています。

問題点をまとめると、第1番目に、シソーラスをつくるときに概念の単位をどういう基準で設定するかという問題点。たとえば、「引く」という動詞をとってみますと、「電話を引く」「電気を引く」「水を引く」「気を引く」「例を引く」みたいな用法があるが、たとえば、この中で「電話を引く」「電気を引く」「水を引く」というのは、どうもよく似た用例であるという感じがします。「なにが有益なものを自分のところまで持ってくる」という点でこの3つの用法は共通しています。したがって、同じ概念と見たらいいのか、それとも実際には違った行為を行うわけだから、違った概念だというふうに見るべきなのか、そういう区別をどう考えるかということをもまず考えないと、自然言語処理のシソーラスの基本となる概念の単位そのものが設定できないこととなります。

2番目は、意味素性というもので、語の意味を完全に記述しないとしても、先ほどいいましたように、われわれが外界の世界で区別できるようなことであれば、なんらかの形で記号表現しておかないと翻訳には役にたたないわけですから、たとえば、楽器という程度の意味素性では不十分で、たとえば打楽器だとか、ものすごく細かいレベルの意味素性というのが必要になってきます。意味素性というふうに少し退いて考えたとしても、かなりの量の素性表現があるだろうという問題点です。

3番目は、「送る」の例のようなものをほんとうに翻訳しようと思うと、意味素性程度ではなかなかすまなくて、その状況で、この場合には一緒についていったんだ、この場合には単なる旅費を出して派遣したんだ、というふうに分離しないとだめですが、これには非常に深い処理が必要です。そういう推論だとか、状況の把握をどうするかという問題点です。

島津 「機械翻訳におけるエキストラリンギスティック・ノーリッジ (Extra-Linguistic Knowledge = ELK)」をしゃべれといわれております。けれど現在の自然言語処理の研究は、こういう ELK を使う以前のレベルでかなり苦労していて、ELK を使うレベルまで到達していないのが実状です。

こういう ELK が、実際われわれが日常、言語を使

って生活しているとき、かなり重要な役割をはたしています。自然言語の世界というのは、非常にオープンエンディッドだといわれていますが、自然言語処理のどの辺ができるのか、どの辺がむずかしいのかという点をまず明らかにすることが重要と思います。

ELK は、いわゆる文法的な情報じゃなくて、外界の情報とか、文化的な情報とかに関するような知識です。そのような知識が絶対に必要だという例として “I saw the man on the hill with a telescope.” という文があります。この文の構造をちゃんと決定するためには、“I”、“man”、“hill”、“telescope” が外界でどういう位置にあるかをちゃんとみないと、できません。“on the hill” が “man” にかかっている場合、“saw” にかかっている場合があります。それから、“on the hill” が “man” も “saw” も形容しているような状況もありうると思います。

こういう文を処理するための知識を扱うためには、システムはいわゆる世界モデルを持っていない限りなりません。システムは、この世界モデルの上でテキストを意味解釈していくこととなります。そのとき、視点とか、話題とか、焦点とかの情報も必要になります。さらに、もっと細かくいうと、送り手のもっている知識に関するモデルも必要になります。

翻訳の場合、そして通訳の場合はもっと、これらが複雑になると思います。すなわち、送り手がいて、その人がなんらかのテキストを發して、それを送り手のモデルに則してシステムが解釈して、今度は受け手の知識をおもんばかって、訳文に写像する、というように複雑になります。

ELK の例と考えられるものを以下に列挙してみます。

まず「事物の存在や事物間の関係に関する知識」です。先ほど出た「丘の上にいる人を望遠鏡で見た」という文の解釈に必要なものです。また、グルノーブル大学のボアテ先生の例ですけど、“carbon and nitrogen tetraoxide” という並列句では、2つの解釈がありますが、「carbon tetraoxide」が存在しないので “carbon” と “nitrogen tetraoxide” という構造だということがわかります。こういうのは、「エキスパート・システム」の関連で、かなり意味のあるような応用だと思います。

次は、「話題に関する知識」です。“What is the registration fee?” といったときに、“What” は「いくら」とか「なに」とかですが、もし会議が話題の文

豚だと、「いくらですか」となります。あまり英語も知らない人とか、そういうことを話題にしてない場合には、“registration fee”とは「なんのことですか」ということになります。

次は、「文化・制度の違いに関する知識」です。これは、Schank のグループの論文にある例ですけど、“The police department did not file charges”という文で、交通事故があったときに、西側の国だと普通警察は“file charges”というようなことをするわけですけど、ソ連だと、そもそもそういうことがない。それで、こういうのは普通は訳さなくてもいいというようなことがいわれています。

次は、「視点に関する知識」です。これは絶対に翻訳に必要というよりも、文の質がよくなるというたぐいだと思います。“My mother send me a book”. といったとき、視点により「くれた」とか「もらう」とか、いわゆる受給表現といわれている表現をとって日本語らしい表現にします。

次は、「特定の事象に関する知識」です。いわゆる埋め込み文のやっかいな例としてよく出てくる文で、「魚を焼くにおい」とか、「コーヒーを入れる香り」とかの文です。英語だと関係詞というより“the aroma of coffee brewing”というふうになります。これは日本語特有の知識らしくて、個別的においなどに関して覚えておかないといけない。そういう意味では、これはあんまり言語外知識といえないかもしれません。

それから、「前提 (presupposition) と断定 (assertion)」に関連した知識です。「どうして神田へ行きましたか」という文は「きのう神田へ行きました」といわれたときには、理由を聞く意味は普通はとれます。電車のストなどがあったときには手段を聞いたりするという意味になります。文章をいわれたとき、どの文が前提で、どの部分が新しくいわれているかをちゃんと解析することが必要という例です。

いまましいいましたけど、次は、「状況に関する知識」です。日本で外国人に日本語を教えている学校で、先生が外人の生徒さんに、その横にすわっている人の姿が見えないので、「彼はどうしましたか」というと、「彼は来ています」と答えるんですけど、「そこにはいないじゃないか」といっても、「彼は来ています」といったそうです。多分その外人さんは“He is coming”というのをそのまま訳したんで、その場合だと、「こちらに向かってるところです。」が正しい。ing がついているからといって「ている」というふう

につけるわけにはいかないという例です。

次は、「連想の知識」です。“Move our tail.”という、あるアメリカの航空会社の宣伝文句があったそうです。“tail”というのは、飛行機だと尾翼とか、しっぽのほうを意味して、「飛行機を飛ばす」という意味で使ったらいいんですけど、そのうち会社内で、スチュワーデスの tail の意味にも取れるというんで、物議をかましたという話があるそうです。これは文化的知識が必要という例です。

次に、「意見とか意図に関する知識」です。「野球の試合を見た」というときに、通りがかりで見たのか、どこかの野球場に行って見たのかで、訳し方が違うと思います。

それから、「社会言語学的知識」です。たとえば、“I don't have to tell you anything, Noman Page?”といった場合、こういうふうに名字をちゃんといったときには、怒っていることになります。「なにもいう必要はない」と憤然としていったなどと、訳されます。これは質に関する話だと思います。

次は、状況と連想知識と両方からかかっているような話です。外国人に日本語を教えている学校で、外人さんが職員室に来て、「先生、外で女が待ってるよ」といわれて、ギクッとしたという話があるそうです。その人は「woman=女」としか覚えてないものだからこうなったという例で、「女性が待たれてますよ」というべきところなのでしょうけど、こういうふうには、そのまま運用するとおかしくなるという例です。

どこまでが言語外知識なのかというのは、なかなか区別もむずかしい。次のは、こういう言語外知識もあるのではないかと例です。

「1フィート=2.5字~3字」

少し情報を追加すると「1フィート=16コマ」。なんのことかという映画の字幕のことなんです。フィルムの464フィート8インチのところから“Raise your head a litte and let me see your face from the side.”というセリフが始まって、468フィート9インチのところまで終わって、長さが4フィート1インチだとすると、さっきの式をあてはめて、2.5字ないし3字が1フィートだから、これは大体12文字以内に訳せばいいということになります。この場合だと、「横顔を見せてくれないうか」と11文字で訳されているそうです。

ちょっと話は脱線しましたが、こういった翻訳との関連でみたとき、いくつも問題が山積みしていま

す。そのなかのいくつかをピックアップすると、たとえば、知識の表現が重要だと思います。先ほど辻井先生の話で細かくいわれたんで、ここではあまりいいませんけども、知識をちゃんと扱うようになると、「字面」とか「語」とか、「概念」といったものをちゃんと区別していく必要があると思います。

普通のシステムだと字面レベルで単語とか辞書のエントリとかを identify していると思いますけども、本当はある性質を持ったものとか、あるいは、ある性質の条件が充たされているものとかといったような、そういう属性的なものから、ある知識の chunk にアクセスしていく必要があると思います。

もう1つあげておくと、「情報の価値」という問題があると思います。人間の言語理解には、価値判断に基づいて、かなり決めつけ的にやっているような感じがします。翻訳の場合、たとえば“Would you like some nice warm Bovril?”という文があったとき、これを LONGMAN の英語辞書でみると、濃縮スープの登録商標と書いてあります。ある映画にあったんですけど、「熱いスープはいかが?」としか訳してないそうです。ヨーロッパやアメリカでは有名なやつらしいんですけど、日本ではなじみがないので、こうしているそうです。

ちょっと話が雑駁になったかもしれませんが、最後に、まとめ的なことをしてみますと、ELK というのは、自然言語処理、特に機械翻訳の技術のキーコンセプトではないかと思います。自然言語処理のジャンルの中で、できることとできないことを整理していくキーだと思います。それらの関連をはっきりさせて、可能な応用はなにか、当面やるべき応用はなにかを、明確にして、進めるべきではないかと思います。

石崎「AIのアプローチによる言語処理」という題をいただきましたが、人工知能的といっても非常に広い分野を指してしまして、それと言語処理とのかかわりといいますとさらに発散する危険性がありますので、ここでは研究のシーズという立場からお話いたします。

先ほどから、自然言語処理での理解というのが深ければ深いほどいいというわけではない、という話がありました。たとえば、機械翻訳システムをつくる立場からいえば、現在の構文解析技術と意味解析技術のレベルを考え、理解の深さについて、いい折衷点を選んで、そこで大きなシステムをつくるという立場があると思います。

ただ、人工知能的な知識、たとえば常識などを使って自然言語を理解する立場からいうと、いままでは構文解析が中心であったのに対し、意味という問題を扱ってみると、いろいろな意味のレベルがあることが分かります。

では、その意味のレベルをいくつかに分類して、そのなかでもう少し詳しい議論をするところが現在の出発点だと思います。

そういう意味で、ここで自然言語処理における5つの理解レベルを設定してみました。

第一は、いわゆる構文解析によって文の構造を理解することです。これは従来から盛んに研究されてきました。

2番目は、単語レベルの意味解析です。これは、いろいろな単語の共起性ですとか、かかりうけですとか、そういったものが入っています。

3番目は、1つの文を処理して、ひとまとまりの意味表現に変換すると、これは、1つの文章じゃなくて、文ということを意識して書いてあるんですが、たとえば、動詞を中心にして矛盾なく1つの文を意味的にまとめるレベルです。

それから、さらに4番目としては、いくつかの文を処理して、文章全体といいますか、1つの段落の理解があります。「1つの段落は1つのトピック、テーマを持つべきだ」という話がありますが、そういう段落におけるトピックを把握することです。

5番目としては、これは先の話になるかと思うんですが、スピーチアクト (Speech Act) ですとか、比喩のように、文章そのままを理解したのでは分からず言語外の意味を理解する必要のあるものです。

こういう5つのレベルの中では、1番目とか2番目のレベルはかなりいろいろなところで研究されていると思います。

3番目の文の理解についても研究されていますが、4番目に書いてあるような文脈をどうやって把握するかという研究は、いままではわりと欠けていると思います。

ただ、先ほどから話が出てますが、アメリカの Schank のところでは、かなり前からこういう研究がされています。ただそこでは、スクリプトですとか、MOP とかいういわゆる文脈処理用の研究はしているんですが、逆にいいますと、構文解析というレベルはかなり抜かしています。

それから、単語レベルの意味解析でも、興味のある

ところだけは拾うけれども、興味のないところは捨ててしまうという、ある意味では非常に大ざっぱなことしかやっていない。

それが、人間が自然言語をどうやって処理しているかという、いわゆる認知科学的なアプローチからいえば、Schank のいっていることは、いわば、斜め読みをするといえますか、あるテーマについて中心部分だけを把握するような読み方なんだということがいえると思います。一方人間はそれだけじゃなくて、新聞を読んでいて斜め読みもできれば、逆に興味のあるところは詳しく読むという両方ができるわけです。そういう二つの機能を持たせることは、どうすればいいかという話があります。

以上で、自然言語理解のレベルの分類という話をしました。次に、これらのレベルで、いわゆる AI 的な推論が、どういう役割をはたすかという話をしたいと思います。たとえば、常識を使って文章を理解するために、いろいろな推論をするというプロセスがあると思います。

ここで、具体的な話に入る前に、推論の深さの分類を行ってみたので、それをご紹介します。

推論をここでは照応の問題と関連させた4つのタイプに分けてあります。

最も代表的なのが、直接照応です。先行詞が文脈中に存在するような照応です。直接照応の例としては、たとえば、「太郎はきのう大阪へ行った。彼はきょう阪大で開かれる学会へ出席する。」ここで、「彼」は一体どこを指すのかという指示代名詞の照応の解析です。この場合には、前の文の「太郎」を指すというふうに判断します。実際に先行詞がある場合には、ある程度探しやすいということがいえるかもしれません。

2番目の深さの推論というのは、これは仮りに「1次推論型」と名づけてあるんですが、先行詞が文脈中に明示的に表現されていない。ところが、文脈中の名詞句とか、事象句に含まれる標準的なスロットを指示する場合があります。

ここでいう標準的なスロットというのは、デフォルトみたいに用意されている、いわゆる常識的に考えられるようなプロパティを指します。たとえば、「殺人事件が発生したが、まだ犯人はつかまっていない。」という例では、この「犯人」というのが前の節の「殺人事件が発生した」とどう関係しているのかが問題です。これは直接照応じゃないわけで、むしろ殺人事件における事件という単語に付随するスロットと

して、犯人や、犯罪があります。その犯人というスロットが「殺人事件が発生した」というところまで読んでいただけでは、まだデフォルトであるわけです。

その次に、「まだ犯人はつかまっていない」というところで、そのデフォルトへの参照が起こったというように解釈します。このような推論のことを1次推論型と名づけています。

それからさらに、深い推論というのが考えられまして、2次推論型とも名付けてあります。文脈中の事象と因果関係にあるような事象や状態を指示するような照応です。例としては、「太郎は塩を水に溶かし、できた溶液を凍らした」があります。できた溶液というのが、前の文のどこに対応するのかが問題です。

これは1次推論型でも、直接照応でもありません。つまり、溶かすという事象の結果できたもの、ということで、一応2次推論型に分類できます。

このような1次推論型、2次推論型は、具体的に普段身の周りにいくらかもある例だと思います。

それから、さらにもっと難しい推論をひっくろめて、多次推論型といっています。2次推論以上の高度な推論を要する照応の場合です。

たとえば、「太郎は山登りが大好きだ。それが自分に挑む姿に引きつけられるからである。」が一例です。「それが自分に挑む姿」というような形で表現された場合に、普通に解析しますと、「それ」は「山」であるとか、「山登りである」に照応していくと思われま。ところが「挑む」という動詞がありますので、「山が自分に挑む」という文ができます。そこでたとえば、共起性みたいなものを使ってチェックしてみますと、無生物は「挑む」という動詞の主語になれないのではねられてしまいます。はねられたから、そこで初めてそれが比喩であるということがわかるという見方もありますが、そういう例でわかるように多次推論は、非常に難しいと思います。

いまの話の続きとしては、たとえば「殺人事件が発生したが、まだ犯人はつかまっていない」という1次推論型の話ですが、ではこれをきちんと把握するには、どういう知識があればいいのかというのが次の話題です。

この例では、一応3つの知識を用意すべきだと思います。「事件には、犯人と犯罪が付随する」というのが知識の1番目です。

2番目は、「警察による犯人の捜査が行われる」で、3番目は「警察による犯人の逮捕が通常行われる」です。

これらの知識を使えば、先ほどの「まだ犯人はつかまっていない、では一体だれがつかまえようとしているのか」という質問の答えは警察ということが分かります。「警察はいまなにをやっているのか」という質問に「警察は犯人の捜査をしている」と答えるわけです。そういう一種の質問応答のための予備知識にもなるということが考えられます。

こういうような常識的な知識にも、考えてみればいろいろなレベルがあるということがいえます。

その1つの例として、これは、Schank が MOP というものを考えるときに基礎にした考えですが、情報の記憶の形態に大きく分けて2種類あるというものです。事象記憶と状況記憶の2種類です。

たとえば、事象記憶というのは、われわれがなんらかの事柄を経験したときに、それをそのまままっておくような記憶場所です。これはたとえば、歯医者に行って歯を抜いてもらったとき、非常に痛かったとか、料金が高かったとか、そういうことが記憶されます。それがイベントメモリ、事象記憶と呼ぶものです。

ところが、状況記憶というのがもう1つあります。これはそういう事象などの場面一般についての情報からなります。その場面に関連した文脈ですとか、規則、標準的な経験、たとえば病気になれば医者に行くという知識を表しておくような記憶の種類であります。これが一応 Shank のいう MOP に対応するような記憶の形態であるといわれております。こういうものを利用すれば、ある程度文脈というものがあるんじゃないかと思われま。

シャंकが以前から提案していたスクリプトとこの MOP の違いについて簡単に言えばスクリプトというのは、1つ1つのイベントを細かく指定表現したものといたえますが、それを更に一般的に柔軟に扱えるような構造にしたものが MOP であるということが出来ます。

それで、私どもは電総研で MOP の考え方に基づいた文脈理解の研究を進めていますが、研究を行う上での課題がいくつかありまして、その第一は話題(トピック)の決定です。入力文章で現在話題になっているのが一体なにかをまず最初に決定するのが非常にむずかしい問題です。

それから、次の問題はそれに関連しますが、話題が転換されたときに、どの段階でどうやって判断するのかということです。段落の切れ目などで、話題が転換

されているかどうかの判断がむずかしい問題です。

それから、実際にこういうシステムを少しずつつくっていき、先ほどから話があるようなソーラスをどうやって構成するかが非常に大きな問題です。

事象概念とか名詞概念を別々につくっていきます。もちろん多義性というのが出てくるんですが、その多義性をどう処理するかというと、やはりある程度文脈に応じた処理の仕方が必要だろうと思います。われわれの場合には、領域を限定してやっていますので重要な単語の多義性は少なく、単語の各意味について階層構造を構成していきます。そういった形で領域ごとに組み上げていき、いろいろな分野について比較統合するというような、ボトムアップ的なつくり方もあると考えております。

司会 では、先ほどの5人のパネラの方々のお話に対して、フロアの方からの質問などがありましたらお伺いしたいと思います。

(質問) ソーラスとして、名詞と名詞の関係の主にお考えのようですが、自然言語処理技術者がソーラスに熱い眼を注ぐのは、1つには、動詞と名詞の共起関係についての情報がソーラスから得られるのではないか、という点があるかと思えます。

荻野 名詞と動詞の共起関係も大切な意味の情報の一部です。これをどう扱うかは大変に難しいですが、共起は、連語に関する情報に相当すると思えます。ソーラスはそこまで扱う必要はないと考えます。たとえば動詞と名詞の共起関係は、間に格助詞をはさんで、どういう動詞とどういう名詞が共起できるかでもって、名詞のグループ分けとか、動詞のグループ分けができます。それをもとに、ある種の階層構造がつくり出せるようにも思えます。

ところが、そうやって作り出した階層関係と、それからここで私が述べたようないわゆる概念体系としての上位・下位で導かれた階層構造は、部分的には重なっても厳密な1対1の対応はなしません。したがって両方を1つのソーラスという形で統合していくのは大変難しく、概念体系なり概念構造を基準として用いて、共起関係についてはこれとは別の扱い方をすべきだというふうな考え方をしております。

辻井 普通のソーラスでいう上位・下位は、単語が表す概念を基本にして組んでいきます。ところが、動詞と名詞との共起関係を考えますと、「電話を引く」、「道路を引く」、「水道を引く」という一連の用例があります。そのときに、水道だとか、電気だとか、鉄道

だとかいうのは、「ある種のネットワーク的な人工システム」という点で共通性を持っていて、「引く」というのは「そのネットワークを自分のところに引っ張ってくる」という意味用法です。したがって「鉄道」・「道路」・「水道」という名詞は、そういう意味の素性のレベルで共通していると考えた方がよい。

これは、個々の語をほんとうにプリミティブにしてその上位・下位関係で動詞と名詞の共起を説明するということは、違うという気はします。結局、荻野さんのいわれたように、知識体系としての概念の上位一下位関係と、動詞との共起性を意識するために、むしろ素性で表現した方がいいものも上位一下位で表現することにすると、かなり性質の異なる情報が同じ枠組みで表現されることになって、かえって混乱することになると思われます。

〔質問〕 荻野先生のお話は随分プラクティカルな話だと思います。情報をいくつかに分けていますが、たとえば同義関係というのは、完全に日本語に依存した問題であり、また、上下関係とか、反対関係は日本語や英語という言語には依存しないと思います。指示物に関する情報の知識も、日本語とは関係なく、ほかの英語とかフランス語のなかでも使える知識だと思います。ですから、言語に独立したものと、依存したものとを分けたほうがいいのではないかと思いますか。

荻野 言語に依存する部分と、独立な部分と、両方あるように思いますが、基本的には依存する部分の方が大きいだろうと思います。したがって、シソーラスはマルチリンガルなものをつくるのは大変むずかしい。専門用語(たとえば情報検索用)ならば、個々の世界における概念共通性みたいなものが保証されますので、まだマルチリンガルなものが可能かもしれません。

しかし、一般語というのは、その国なり民族なりの文化と非常に密接な関係があって、概念の切り取り方そのものがとても違う。たとえば日本だったら季節、場所、自然に対して非常に細かく区分されているが、他の国ではそうではない、というようなこともあり、文化に依存する部分が非常に多い。したがって、一般語のシソーラスは、言語ごとに別個にした方がすっきりすると考えます。

確かに、同義関係以外の上下とか、反対とか、反義とかというものについては、ほかの言語でも成り立つように見えますけれども、必ず成り立つかどうかは、自信の限りではありません。たとえば、「イヌとテリア」という上下関係に対して、テリアという犬を知ら

ない文化があっても全然おかしくないわけです。結局その地域の自然や文化によって規定されるものですから、しかたがありません。

〔質問〕 シソーラスのつくり方に関して、荻野先生は計算機を使ってもあまりうまくいかないだろうというお話があり、鶴丸先生のほうは、うまくいくというお話でした。私自身も昔ちょっと簡単な実験的なことで英語の辞書からある種の特定の語彙を引き出そうとやってみたことがありますが、うまくいくのか、いかないのかははっきりしませんでした。この件についても少しコメントをいただきたい。

鶴丸 大規模な意味辞書をつくる場合には、作業手順が確立していることが一番大切です。その部分に計算機の利用が可能で、できる部分は積極的に利用した方がよいと考えています。私が話したのは、国語辞典からの情報収集で、構造的に捉えられる部分に計算機を利用していることです。

しかし、国語辞典は人間が利用するようにできてますので、結果をみたあとやはり人間がチェックしなければならぬということも当然あり得るわけです。結局、支援という形で計算機を利用するということが最も重要だと思います。たとえば、たくさんのデータの整合性のチェックなどは、計算機を使わないと、手作業ではむずかしからうという気はします。

司会 計算機はツールとして使われるわけですが、その能力がどれだけあるかということが、非常に大きな問題になってくると思います。

特に言語学の観点からみて、いかに使えるか、あるいは役に立つかについて荻野さんからご意見を伺いたいと思います。

荻野 計算機を使う必要はない、というのではなく、基本的には使ったほうがいいでしょう。ただ、きちんとしたものをつくるためには、計算機を使おうと使うまいと、同じぐらいの手間がかかるだろうということです。私もいまシソーラスづくりを手がけておりますが、ある段階からあとは計算機にのせてやらなくとも今度は質が落ちるだろうと思っております。

たとえば、見出しの選定が終わって、ある程度の上位・下位の関係などが関係づけられたあと、実際にその概念構造を組み立てる段階においては、計算機でチェックしないと、統一した整合性のとれたシソーラスはできにくいだろうと思っております。要はどこから使うかという問題、および使うことによって、どれぐらい速くなるのか、あるいは質が高くなるのか、その

辺の見通しの違いが、鶴丸さんと私の違いなんだろうと思っております。

〔質問〕 荻野先生にお伺いいたします。上位・下位の概念から申しますと、シソーラスはセマンティックネットワークの一種だと思います。セマンティックネットワークの場合は、一番上の親が複数あり、しかも上位・下位概念はおろか、同義語とか、それから全体部分まで含めた膨大なものになっておりますが、それはシソーラスで代用できるのか、をお伺いします。

荻野 シソーラスはその辺までカバーするべきだと考えます。とりあえずは上下関係などからスタートしても、順次情報をふくらませて、シソーラスの見出し語を強化し、関係づけのために多くの情報をつぎ込んでいきたい。いままでに現れたシソーラスとはだいぶ形は違うかもしれませんが、私はそれは広義のシソーラスであると考えます。

〔質問〕 荻野先生のおっしゃっているシソーラスというのは、自然言語の処理の立場でなにに使うのか、鶴丸先生のおっしゃっているシソーラスはなにに使うのか、それをお聞かせ願いたい。

鶴丸 ここでお話しましたのは、一般的な自然言語の意味処理を目指した意味辞書の機械化のために、単語の意味に関する情報をどのようにして獲得したらよいかという問題についてだったわけです。意味辞書では、上位・下位関係が基本ですので、これは、推論などに利用できるでしょうし、上位語がうまく得られれば動詞との共起関係も整理できるでしょうし、因果関係とか動詞の定義ができれば、MOP などでも利用できるのではと思っています。しかし、ここでは、単語の意味の定義をどうするかが中心ですので、そこで導入されます意味は、人間が知識を整理する場合に用いるものになっているわけです。ですから、これらの意味的關係が、実際の機械翻訳や言語理解などで、どのような機能を持たされるか、または持つことができるかは、もう1つ別のレベルの問題になるのではと思っています。

辻井 鶴丸先生のお仕事は、たとえば構文解析に直接使おうというのではなくて、たとえば、国語辞典の中には、これだけの意味的關係があったとか、かなり多くの語の上位概念になっているようなカテゴリとしては、どのような語があるか、あるいは共通の上位概念のもとにある下位概念の平均の個数はどうかとか、意味の世界を探索するときの基礎的な資料ができたと考えたほうがいいのではないのでしょうか。

鶴丸 現在の作業状況は、辻井先生のおっしゃる通

りです。

司会 ほかのパネラの方にも、シソーラスが直接言語処理のなかで使い得るかどうかということに関してのご意見をお伺いしたいと思います。

荻野 私も、自然言語処理での応用を考えないでもないですが、具体的な目的は、はっきりしない面があります。

それはまず第1に、現在のシソーラスがあまりにも不備であるということです。これではどうやっても使えないものにならないから、もっといいものをつくりたいという基本的な欲求があります。いいものができれば、それをどう使うかはそこから先の問題です。はっきりとした概念体系を持つシソーラスというものができれば、計算機だけじゃなくて、人間が使うにも非常に有用な資料になると思います。特に、日本語を知らない外国人や語集の習得中の子供たちにとって、大変メリットがあるでしょう。ですから、自然言語処理のなにに役立つ、なにに使うかということは、私自身はあまり強く考えたことはありません。むしろ、つくり方をはっきりさせて、いいシソーラスをつくっておけば、なににでも使えるんじゃないかというぐらいのあいまいな考え方をしております。

島津 私も先ほど辻井先生がいったように、基礎データとしてみるという考え方に興味があります。特に鶴丸先生がいておられるみたいに、実際の国語辞典で、メタ的な語の種類がどれくらいあるかというのは、よくわかっていないような気がします。

たとえば上位・下位関係としてみたときに、必ず上段にある語とか、末端でしか出てこない語とか、いわゆるジェネリックな語のインスタンスとか、いろいろな種類があると思いますが、これらが実際われわれ人間の頭のなかでどれぐらい区別して持っているのかということの手がかりをうる上でも、興味があります。

それからひとつ質問をさせて下さい。荻野先生の話で、「シソーラスとは、階層構造をなす概念体系の各部分に単語をあてはめたもの」とありますけど、概念と単語というのは同じものとして考えておられるのか、違うものとして考えておられるのか、違うとしたら概念的なものはどこにどのように位置づけられるのでしょうか。

荻野 単語というのは音の形と意味とが結びあったものであり、その意味の一番中心にあるのが概念である、と考えています。意味には、コンテキストであるとか、文法的なものの一部などの概念以外のさま

ざまなものも入ります。そこで、シソーラスは単なる概念を中心に分類したにもかかわらず、その階層構造をつくると、たとえば、ノードの部分に概念はあるがびったりした単語はないということがいくらでもあり得ます。

そういう場合に、現在のシソーラスですと、たとえばいろいろなものを列挙して、「米・糠・小麦粉など」というふうなグループ名が使われたりします。下位概念にしても、いくらでも細かく分けていくことができます。それは、人間の認識能力の普遍的な特徴だと思いますが、ただ、細かく分けていけば分けていくほど、それにあてはまる単語というものもなくなってきて、たとえば、頭に修飾節がついたような、文に近いようなもので定義されるようなそういうものになりかねない。そうすると、概念という形で整理しているといっても、ある程度言語化された単語という外形をもたなければいけないので、すでにある単語を概念体系にしたがって配置したというふうに考えるのが、一番妥当だろうと考えております。

石崎 言語学・国語学の先生方が辞書をつくるときは、相当なタイムスケールでお考えになっていると思いますが、われわれ計算機屋が使えるようなシソーラスは、いつでもできるのでしょいか。

荻野 これはむずかしい。どの程度のものを作るかによって大分時間が変わってきます。ただ、シソーラスという以上は、いわゆる基本的な単語の数千語くらいで作ったのではその効果はあまりない。複合語の一部まで含めると、少なくとも数万語以上、もしかすると10万語、20万語までいきます。たとえば10万語のシソーラスに、入れるべき情報としてあれもこれもというふうに欲ばりますと、おそらく数十年ぐらいかかってしまいそうな気がします。しかし、たとえば、同義関係とか、上下関係とかのあたりからやっつけば、ある程度のタイムスパンでできるのではないかとこの期待があります。

実際に、情報処理振興事業協会で、三省堂国語辞典からカードを作り、名詞を中心に同義関係、上下関係についてやりましたが、見出しの選定として7万語を選ぶのに10人で100時間、見出しのダブリを落とし50音順に並べるのに10人で合計120時間かかりました。そして64,000枚のカードで同義語判定を行うのに400時間(5日間合宿、1日10時間)以上かけてその約半分までしかいきませんでした。ですから、同義関係、上下関係だけで数年のオーダーがかかるだろうとい

うのが私の実感です。

〔質問〕 シソーラスは、エキストラリンギスティックな知識の経済的な整理の仕方としてある程度使えるのではないのでしょうか。

荻野 いまの意見にまったく同感です。将来的には、それぞれのレフェレントの性質までも、シソーラスのなかに書き込みたいと思っております。

司会 シソーラスのなかの、たとえば上位概念などは、多分、いわゆる知識として扱われるようなもののほうが多いのではないかという気がいたします。

少し話題を拡大いたしまして、先ほどの辻井さんからの「語か、概念か」と、島津さんのエキストラリンギスティックな知識まで含めて討論をしていきたいと思ひます。

〔質問〕 意味素性とシソーラスとの関係についてお聞きしたい。たとえば、「太郎」の素性として「+human」とか「+male」とかがありますが、それらはシソーラス的のいうと、「太郎」の上位概念となるのでしょうか。

辻井 多分かなりの重なりがあることは事実だと思います。ただ、先ほどの荻野さんのご意見にもありましたように、シソーラスというのは外界世界 (extra-linguistic) の知識を表現する、すなわち、知識のある種の体系的整理だと思います。

そういう知識が言語表現上にどう反映されるか、また、それをどのように規則化するかという部分は、少し別ではないかということです。名詞と動詞の共起は、シソーラスの体系とはちょっと別のレベルで捉える必要があるといったのはそういうことでして、たとえば「移動可能」であるとか、あるいは、「形が丸い」とか、そういう共通性をシソーラスの上位一下位で捉えるのは、外界知識の静的体系としてのシソーラスとはかなり異質なものになる、と考えているわけです。

島津 私の考えは、いま辻井先生がいわれたことと、ほとんど似ています。私どものほうでは、知識や、そのなかの意味辞書を、いわゆるフレームというシステムで捉えています。荻野先生が説明されたシソーラスの例でいうと、「犬」とか「テリア」とか「買う」とかは、1つのフレームのノードに対応して、そのなかに言語的な意味情報として、意味素性がいくつか入っており、そしてその意味素性自体もフレームのノードとなりえるという構造を持っています。

その結果、全体としてセマンティックネットワーク的な構成になっているという感じですが。

司会 従来から、素性は、定義が十分であるかどうかは別にして、言語処理では非常によく使われてきましたが、シソーラスそのものを直接言語処理に使うことは、あまり行われていなかった。その辺の概念的な整理が非常に重要であるということが、今日の議論で強調されているわけでありませう。そこで具体的にどうつくっていったら、どう利用していけばよいかという点をもう少しつめてみたらおもしろいのではないかと思っています。

〔質問〕 従来のシソーラスはなにゆえに使いものにならないのでしょうか。たとえば、従来よく使われているロジェのシソーラスは、荻野先生の考えているシソーラスの範疇に入らないものなのか、それから、シソーラスを階層構造に組み上げていく際のセオリーがあるのかお伺いしたい。

荻野 「なぜ使いものにならないか」については、分類がはっきりしていないからです。たとえば、概念の階層構造、個々の単語の配置について客観的な規準がない。これは、分類語彙表、類語新辞典だけでなく、ロジェの場合もまったく同じです。ロジェの分類そのものが間違っているというつもりはありませんが、一番下のレベルでは、同義語、類義語がただずらずらと列挙してあるだけです。分類語彙表の場合でも、単語が20ぐらい入っているパラグラフという切れ目があるが、その20の単語が一体どういう順序に並んでいるのか、どこにもなにも書いてない。角川の類語新辞典も同じです。一番下の3桁の数字の1,000分類の下に、さらにもっと細かい区分があり、その1つ1つの区分のなかには単語が少なくとも数個から、多い場合には40個ぐらいまで入っている。しかし、それが相互にどういう関係にあるかということが全然明示されていない。

すなわち、いままでのシソーラスが使いものにならない理由は、そのように区分の仕方がきわめておざっぱであるということにつきまします。それでもいままで使われてきたのは使うのが人間だったからです。しかし、計算機というのは常識がありませんから、常識がない人が使うためには、あまりにも粗すぎて使いものにならないというのが私の結論であります。

第3点ですが、同義語を整理した上で階層関係を構築していく場合、国語辞典の意味記述のなかにある上位・下位関係を想定させるいい回しがたくさんあります。ですから、辞典の意味記述を利用してシソーラスをつくっていく場合には、とりあえずはこの辞典に書

いてある意味記述を信用するならば、ある程度のことはいえます。

〔質問〕 今日のバネラの皆さん共通して、必要な知識はとにかくつくろうという基本方針をお持ちのように感じましたが、辻井先生からは「へたにつくればあまいさが爆発する」というご指摘もあったわけで、地道な努力を続けていけば将来展望が開けるとお思いになっているのか、それともたとえば学習というような逃げの道も考えるべきなのか。

辻井 必ずしも地道な努力だけを続けていくというふうには思っているわけじゃなくて、たとえば意味素性というのは、意味のレベルでのルール化をしようということですよ。

だから「引く」に電話がついたときには、特定の意味をもってという形で「引く」と共起し得る名詞をすべて列挙するというのがいやだから少し共通の概念として、たとえば、それらに共通する意味素性というのを考えて、ルール化してやろうということですよ。

その次は、そういう語に共通に見られる素性を自動的に学習していったら、規則性を抽出していくというのが当然であると思うんですね。

このような試みは、多分石崎先生が最後のほうでおっしゃられたみたいに、言語の比喩的使用といった問題と密接に関係していると思っています。

多分意味を細かく分けていったら爆発するというのは、言語の比喩的使用というか、人間が無意識に使っている比喩的表現がうまく処理できないということだと思います。

結局、単語の共起のような形で、表面だけで処理してゆく限りは、少しのゆらぎも許されないので、どっかでルール化が必要となり、それに対応するものとして素性を考えるという話をしたわけです。同じようなことが「現在のシソーラスがなぜ使いにくいのか」ということにも通じていると思います。荻野先生から「中核の意味を中心にしてシソーラスというのはつくっていく」という話がありましたが、実際の言語現象をみると、その中核の意味から派生してくる意味を考えないと処理できないことが非常に多くある。たとえば、「ハゲ」というのが、「ハゲが来た」のように特定の人間を指すのに使われる。中核的な意味をもとにしてつくられたシソーラスから、実際の語用法として、たとえば「ハゲ」は「人」のことも表現し得るみたいなことを導き出してくる部分が、いまの計算機処理の技術では欠けています。シソーラスが悪いのか、ある

いはそういう比喩的な用法を処理する機構は別に取るべきで、それが現在の自然言語処理では欠けているためにいまのソーラスを使いにくくしているのか、その辺は議論の余地があると思います。そういったことをもう少し明確にしないと、ソーラス自体にどれだけの情報を記述しておくべきかも明確にならないのではないかと考えています。

司会 地道な努力をするにはどうしたらいいかということで議論しているわけですが、質問された方はどうお考えでしょうか。

〔会場〕 私自身は、地道に全部つくらないで途中までつくっていくと、なにか意味空間ができてくるのではないかという気がします。そしてその中であいているすきまをうめることを学習とのアナログで考えると、ある程度つくってあとは自動化へのステップができるのではないかと考えています。

石崎 いまの方のご意見に賛成なんですけど、学習とひとこといっても、アナログですとか、アシミレーションですとか、いろいろな側面があって、実際には自然言語の解析にいままでその学習のファンクションはほとんど応用されていない。

そういう意味で、今後の可能性がかなりあります。学習的観点からの努力がされれば、展望が開かれる可能性があります。ただ、もちろん学習そのもののむずかしさは十分に自覚する必要があります。

「類推」ということで、私を感じたのですが、辻井先生もおっしゃったように、比喩的表現が非常に多いです。辞書の語釈文を見ても、ほとんど比喩的な表現によって読者をわからせようとしている面が多い。

比喩といっても、言語学の方に聞くと非常に細かく分類されてまして、たくさんあるんですが、なかでも死喩というんですか、デッドメタファというのがありまして、これは初めは陰喩だったものが、だんだん時間が経つにつれて市民権を獲得して1つの意味を持つようになる。そういうものが非常に多く、多義の問題と密接に関係していると思います。しかもそれが辻井先生がおっしゃったように、水道を引くとか、電気を引くとか、線を引きとか、その共通性は多分その辺からきているのではないかと思います。その辺をヒントにして、なにか1つのとっかかりがありそうな気もするし、それと学習というのやはり密接な関係があるように思います。

司会 本日は学習そのものはテーマとして取り上げませんでしたけれども、学習を機械の上で考える場合

には、それを可能とさせる枠組みが必要だと思えます。枠組みが定まってくると、たとえば、あいている所を埋めるとか、不十分なところを修正していくということが可能となってきます。そして、そのためのメカニズムが用意できるということになるかとも思えます。その辺は非常に重要な問題で、そしていまの石崎さんのご指摘がありましたように、まだほとんど進んでいないというのが現状だと思います。その意味でも地道な努力を進めていく必要があると思えます。

それでは、次に、ソーラスや辞書などの表現には、表現のための基準が必要となります。たとえば、概念と概念の間の関係の表し方、事象と事象の間の関係の表し方などです。このような問題についてご意見を伺ってみたいと思います。

鶴丸 国語辞典の計算機利用という立場から考えますと、現状の国語辞典は人間の利用が前提になってますので、いろいろ問題点があるわけです。たとえば、特殊記号にあいまいさがあったり、同じ内容を表現するのに異なった表現があったり、上位語が陽に表れない語義文などあります。特に、語義文などに使用されている単語の表す意味がなんらかの方法で区別されていたら、計算機処理上、好都合と考えられます。

司会 たとえば、ロングマンの辞書の語釈文を見ますと説明のし方がある程度形にはまっています。これは、利用者にとっては非常に理解しやすいと思えます。しかし、さらに、もっといろんな条件が持ち込めるのではないかと思います。どなたかご意見ございせんか。

〔質問〕 いったい最初萩野先生がおっしゃったソーラスの定義のなかで階層構造を前提にしていたが、もう少し別の角度からみますと、意味的特徴をひとまとめにして列挙しておくこともソーラスのなかでは大切ではないかと思えます。そういう意味で見ますと、意味情報のなかには構造的な情報と内容的な情報がございます。同義とか、類義とかの情報を主とすると内容的なものに重点をおいたソーラスになります。それだけでなく、構造的な面の特徴にも注目して、分類して列挙しておくことも、とても大切だと思います。いま司会の野村さんが指摘された、たとえば深層構造や文脈を分析するときに必要な事象と事象の間の因果関係のような構造的な特徴についても、同じ類を集めて並べておくことも大切な気がいたします。

辻井 単語の意味や単語について記述をするときに、構造的な情報と内容的な情報との2つがあるというこ

指摘は、多分正しいと思います。それがほんとうに相互の連絡がなくていいのかという話と、それからたとえば、内容的情報から構造的情報が自動的に導き出されるのか、ということもあるわけです。Schank なんかの立場もそうだったわけです。たとえば P-TRANS という語の深いレベルの意味の構造を規定しておけば、表層上の格は、そこから導出できるという立場も、基本的にはあるんじゃないかという気もするわけです。語の内容的情報を語彙分解して捉えて、それから構造的情報である格構造を予測するみたいな話です。

この2つの情報を独立して捉えておくか、あるいは、その相互の間に規則性を認めるか、その辺は議論しておく必要があるだろうという気がします。

司会 語彙分解をやっていきますと、意味素のところにも狭く閉じこもってしまうといったような感じもします。すなわち、構造的な情報が少し失われてきてしまうかも知れないということです。万一それが導出できるとしても、なんらかの形で情報を置いておくほうがより便利かも知れないという議論があるような気がします。

では、これから処理についての議論の方へ話を進めていきたいと思います。辞書の実際的な表現も含めてご意見など伺いたいと思います。ソーラスを計算機のなかで具体的に表現するときの表現の枠組みというのが常に気になります。すなわち、それを推論などにも使いたいわけですが、そのとき推論に便利な形で表現されている必要があります。言語学の立場からはこれを考えているのでしょうか。

荻野 いまの言語学の分野では、ソーラスは興味の対象になるところまでできていないのではないかと考えてます。ですから、言語学会会員の中からソーラスに対するなんらかの貢献ができるかということについては、疑問の面があります。言語学者が新しいソーラスを作るよりは、自然言語処理の分野に携わっている人が、自分でソーラスをつくる方が実現性が高いのではないかと思います。

鶴丸 いまのは結局、ソーラスを応用するのだから、そういう立場の自然言語処理をやっている人がつくれということです。

荻野 まあ、そうですね。つまり、言語学の人ですと、ソーラスをつくっていったにすぎないんだということがありますが、自然言語理解をやっている人は、ソーラスがないことによって非常に不便を感じており、欲しい欲しいという切実な要求がありますから、

自分でつくるのではないかと考えています。

司会 処理のほうに少し入っていきたいと思います。

石崎 電総研で最近文脈理解を扱ってやっています。わりとまじめにコツコツとやっていますが、必ずしも、先がはっきり見えない。コツコツやればやるほど先が見えない。そういう点があります。

ただ、われわれの場合は1つ別の要因がありまして第5世代プロジェクトでは非常にいい推論マシンをつくり、知識情報処理をやるということで、かなり盛んに研究を進めています。私もそのプロジェクトに少し関係がありますが、そのプロジェクトのなかで、自然言語処理システムをつくる時に、MOP 的考え方かなり役立つだろうと思います。他と比べて予算がありますから、そういう一面の期待は期待もっています。

それで、原理として MOP がほんとうに有効かというのはまだわからないです。いまのところ使えるのはこんなものしかないので、一応 MOP をたたき台にして、われわれなりに改良しながらやってみているというところでは。

ICOT では、またシチュエーションセマンティクスとかいろいろやっています。それが今後どうなるか非常に興味深いですけれども、現在ではそういうようなレベルだと思います。

鶴丸 私たちは、単語の意味の定義（記述）というところでもたたしている状況です。MOP というのは、事象概念とか名詞概念とか分けてつくる、これは私たちもそういうふうには思っておりますが、MOP では、階層関係のほかに、動詞の定義、特に因果関係などをきちんと整理しておく必要があるように思われますが。

石崎 それに関連しますと、スクリプトとか MOP というのは、いくつかの事象をくくったような枠組みですね。パラグラフを把握するような枠組みというふうに考えています。そのなかにいくつかのシーンがあって、それにいま表層的な文章のなかの動詞をあてはめるときには、決していま見ている文の動詞がそのまま MOP とかスクリプトに書いてあるわけではなくて、その必ず上位概念が書いてあって、たぐさんの下位概念が使えるようになっていくわけですね。

そういう意味で、上位・下位概念がしっかりつくられていると、その機能が非常によくなるということが1つですね。

それから、1つ枠組みが決まると、もう1つ別の段

落について、またいろんな事象が出てくると、その事象の持つ情報をどうやってすでに得ている知識に同化していくか、すなわちアシミレーションしていくかが重要です。前に出てきた文脈上のある概念との関連とか、動詞の上位・下位の関連とかをチェックしながらアシミレーションしていくわけです。

そういう意味で、もし処理対象の文章のなかに現われた単語すべてについてのシソーラスがあればかなり効率的にできるし、なければなくて少しずつ加えていくというレベルになっています。必要性というのはそういう意味で非常に大きいと思います。

島津 処理と辞書の観点からいいますと、一般に人工知能的な処理で気になるのは、たとえば、「爆発する」とか「死亡する」とかの意味表現が出てきますが、それは計算機内部ではローマ字で書かれています。すなわち概念的なレベルで書かれていると思うんですけども、そういうときに、辞書に最初にアクセスしていくときは、ローマ字の字面でもいいと思うんですけども、その後はこういう字面をコンカチネートしただけでは、一般には、概念的な表現にはアクセスできないと思います。多分こういうのはだれもまだうまく解決していないと思いますけど、そのようなものが辞書と処理の観点からいうと難しく、研究しなければいけないことだと思います。

〔質問〕 シソーラスというのは、パターンマッチのときに概念的に許せるかどうかをかなりきめの荒い程度でするときにしか使えないのではないかと思います。だからいかに細かくしても限界が見えていると思います。実世界では、時間とか場所とかに関する数値的なデータがものすごく入っています。だから実世界のいわゆる値の計算なくして意味処理は不可能ではないかという気がします。

司会 いまのご意見は、実際の文などを処理するとき現実の世界の知識およびその文が実際に発話された時間的・空間的状况を無視できないというご意見だと思います。世界モデル、状況知識を含めて処理するためには、どういった内容の知識をどのようにしなければならぬか、そしてどういった処理が必要か、という観点からのご意見をお伺いしたいと思います。

辻井 知能処理とか、知識処理といったときに、2つの違ったイメージがあるという気がします。

1つは、インスタンスに関する知識。たとえば「この人はどういう人か」とか、「計算機はこういう部品からなって、こういう形をして…」という話です。

もう1つは、石崎さんのお話されているような、たとえば、比喩的な表現をどう扱うかとか、あるいは一般の因果関係のようなもの、どういうできごとが起こっていくかみたいな事象知識と状況知識があると思います。この2つを分けるというのはなかなか難しいかもしれないですが、やはりきっちり考えておかないとだめだという気はしています。

司会 シソーラスあるいは長期記憶として蓄えておく知識というのは、どちらかというとゼネラルなものです。もちろんそのなかにはインスタンスとしての例示も含んでいます。それを実際に使うときには、その文などが発話された現実世界の知識、状況が当然必要になってくると思います。

〔質問〕 自然言語の意味論と人工知能における知識論との関係について、知識だとか、意味だとか、概念だとかはかなりオーバーラップしている部分があるように思われます。意味論と知識論は違うのでしょうか、同じなんでしょうか、分けて考えるべきなのか、べきでないのでしょうか。

萩野 知識と意味とは違うと思っています。ただ、その境界というのは実際にはややあいまいにならざるを得ない部分があります。意味は、発信者から受信者へいく言語記号が担うものだと思います。知識は、その言語記号を解釈する際に、聞き手が自分の頭の中にもっているものだと思います。ですから、厳密に知識と意味を区切ることは難しいだろうと思います。ただ、理想的には区切りたいというのが言語学をやっている者からの見方です。

石崎 私は工学的にみる立場ですが、ひとことであろうと、知識というのは要するに表現できたもの、コンピュータでもなんでもいいんですが、表現できたものを知識というと思います。ですから、メタ知識とか、メタ推論とか、メタがあるわけです。意味ということちょっと次元が違って、ワードならワード自身が持ってて、内容の底が知れないものですね。言語学的にいまおっしゃったようにぼんやりとしたものはあるんだけど、自分の頭のなかでは見えるものというような感じがあります。

司会 私は常に「知識を使って意味を理解し、理解した意味が知識になる」という説明をしています。意味処理をするためには知識は不可欠です。その知識は、計算機の場合には、なんらかの形で記述されます。また、意味理解された結果も計算機のなかで記述されます。一旦記述してしまえば、それは知識として

取り込むことができるようになります。

〔質問〕 翻訳のことはあまり聞かれてませんが、島津さんが「機械翻訳における言語外知識」のお話をされましたが、技術としてどの程度まで言語外知識を取り込んだ処理が行われているか、AI 的アプローチの言語処理の今後の発展のタイムスパン、どんな形でいつごろ実現されるかと思っているか聞きたい。

辻井 一般的に言って、島津さんがおっしゃられたました外界知識の関与というのは、ある言語での表現だけをとってみると implicit な情報が、相手の言語での訳を考えるとときには explicit にしないと訳が選べないというタイプの話です。その文だけみても、そのままでは情報はないんだけど、たとえば文化的な背景であるとか、百科事典的な知識を合わせると、それによって記述されている状況が再現できて相手の言語への適切な訳出形が選べる、といった話ですが、これは一般的にいうものすごく難しい。ですから、島津さんがあげられた例のかなりの部分というのは、少なくとも現在の機械翻訳では、困難な問題というのが多分正しいでしょう。ただ、こういう特定の事象に関する知識を入れておけばできますよ、という形で、システムをオープンにしておくことはできる。しかし、そういう情報をどう集めるかという問題がある。

それ以外、たとえば、視点だとか、状況に関する知識だとかいうレベルでは、それをどう表現していったらいいかとか、どう処理に反映したらいいのかということすらわかっていないような問題も島津さんの話の中にはあって、まだ解決しなければならない問題は山積しているという気がします。

石崎 非常にむずかしいご質問ですが、たとえば、電総研の研究ですと、いま文脈のほうを若干やっているところなんです。ですけれども、プロトタイプとか、商品化という話になると、もうちょっと長期的な話になりまして、第5世代のプロジェクトの計画書が一番参考になると思います。あれを見ますと、中期には意味処理を中心にやるとか、後期には文脈解析をやるとか書いてあります。かなりロマンチックで、ああいう方向でがんばれば非常に楽しいし、終わったときにはかなり成果はあるだろうと思います。もともと90年代に使われるようなコンピュータをつくるという趣旨でやっていますから、プロトタイプは多分90年代の初めごろにできるようなタイムスパンだろうし、商品という、さらにその技術がトランスファして、安くなっていく感じですから、さらに数年ぐらいかかるか

もしれないというのが、常識的な答えでしょう。

司会 時間も大分過ぎておりますので、パネルの方にもう一度一言ずつご意見をお伺いしてしめたいと思います。

島津 シソーラスをつくる時計算機は役に立つか、立たないかという話がありました。シソーラスをつくるには上下関係とか類義関係とか、要するに総当たりにチェックしていく必要があるから、組み合わせ的にもものすごく労力がかかるようになると思います。そういう点で、非常にいいマンマシンインタフェースやグラフィックディスプレイなどをうまく具合に使ってやると能率が上がるのではないかと思います。あいかわらずカードベースでやってたのでは能率が上がらないんじゃないかと思うんですが。

荻野 いろいろ問題がありますが、一番の問題は、私の手元にまともに使える計算機がないということです。グラフィックなんかとんでもない。いまでもカードでバッチ処理やっています。計算機があっても文科系の常で、予算がないですから、ICOT のようにお金が出るならいざ知らず、お金がなくて計算機がないところで、どうやって計算機を使ってシソーラスをつくるか、という問題もあります。もちろんいいツールがあれば使うのにやぶさかではないのは当然ですが、ただカードもあながち捨てたもんでありません。

鶴丸 大規模なシソーラスを考えますと、単語間の階層関係をもとにして統一的にすっきりした階層構造が得られるとは限らないと思います。ですから、実用規模の意味辞書をつくる場合には、大事な情報として、単語間の階層関係の規定のし方についての情報も与えておく必要があると思っています。

司会 本日は「自然言語の知能的処理」という題をかかげて議論させていただきました。当然のことながらそのほんの一部の問題しか追求することができませんでしたが、最後に言語科学としての言語の知能的処理の研究と応用には、言語学、心理学、計算機科学などの境界領域の探索とそれらの統合が必要である。これが、次世代のステップへのトリガであり、それへのアプローチは、学際的研究と段階的応用の二面性が不可欠である。換言すれば、基盤を与える理論、データの着実な収集と体系化、応用による評価・確認が、互いの主体性を保証しながら密に関連し、協力して進むことになる。理論の強靱化、データの充実、ツールの高度化が互いの進展を促す。これらは、長期的視野、超組織的協力のなかで進められなければならない。という言葉をしめして、本日のパネル討論を終わりたいと思います。