05-R 008

大規模知識ベースに関する調査研究

報　告　書

平 成 6 年 3 月

JIPDEC

財団法人　日本情報処理開発協会

この報告書は、日本自転車振興会から競輪収益の一部である機械工業振興資金の補助を受けて平成５年度に実施した「大規模知識ベースに関する調査研究」の成果の一部をとりまとめたものである。

KEIRIN
この報告書は、競輪の補助金を受けて作成したものです。

はじめに

近年、情報技術分野における研究開発の中で、知識（情報が体系化されたもの）を扱う知識情報処理技術が、従来のデータや情報の処理方式を追求するアプローチから知識の体系・構造など、知識そのもののあり方からのアプローチが新しい情報技術を創出するものと期待されております。そこで知識情報処理では、大量の知識を体系的に蓄積し、多くの情報システムから利用できる大規模知識ベース（Very Large-Scale Knowledge Base:VLKB）の構築と共有に関する技術の開発が当面の課題となっております。

大規模知識ベースは、自然言語や図形、画像などの多様なメディアで記述、表現された知識を対象に、これらを大量に、かつ意味内容までを扱うため、研究開発には、人工知能技術をはじめとする情報処理技術のほとんどの分野、また言語学、心理学、さらには、人文・社会の分野とも関連があり、各学問分野との連携が不可欠であるとともに、国際的な協力が重要となります。

このような状況のもと、当協会では、平成５年度「大規模知識ベースに関する調査研究」の一環として、大規模知識ベースに関し、多方面から議論し、取り組みへの緊急性と研究開発を進めるにあたっての国際協力を共通認識することを目的に、平成５年１２月１日から４日の４日間にわたり「大規模知識ベースの構築と共有に関する国際会議１９９３（ＫＢ＆ＫＳ’９３国際会議）」、「同ＫＢ＆ＫＳ’９３国際ワークショップ」を開催いたしました。

同国際会議ならびにワークショップの実施にあたっては、ＫＢ＆ＫＳ’９３組織委員会（委員長　渕　一博　東京大学工学部電子情報工学科教授）、ＫＢ＆ＫＳ’９３プログラム・実行委員会（委員長　横井俊夫　㈱日本電子化辞書研究所　研究所長）を設置し、国際会議ならびにワークショップの性格づけ、具体的実施内容等について審議して実施しました。

本報告書は、同国際会議ならびにワークショップの会議内容をまとめたものです。内容は、第１部「ＫＢ＆ＫＳ’９３国際会議会議録」、第２部「ＫＢ＆ＫＳ’９３国際ワークショップ概要」で構成されています。

最後に、本調査研究ならびにＫＢ＆ＫＳ’９３の実施にあたり、ご指導ご協力いただいた組織委員会ならびにプログラム・実行委員会の委員長をはじめ委員各位、ＫＢ＆ＫＳ’９３の講演者各位ならびに関係各位に深甚なる謝意を表する次第であります。

なお、本書が今後の知識情報処理技術ひいては高度情報化の発展に寄与することを念願する次第です。

平成６年３月

財団法人 日本情報処理開発協会

大規模知識ベースの構築と共有に関する国際会議１９９３（ＫＢ＆ＫＳ'９３）

組織委員会・名簿

（平成５年11月現在）

（敬称略）

委員長 ： 渕　一博　東京大学 工学部電子情報工学科 教授

国内委員 ： 飯沼　一元　日本電気㈱ 情報メディア研究所 所長
（50音順）
池田　吉紀　日本経済新聞社 データバンク局 次長
市川　隆　(財)日本情報処理開発協会 常務理事
伊藤　直　朝日新聞社 ニューメディア本部 本部長
大須賀節雄　東京大学 先端科学技術研究センター 教授
神田　利彦　日本科学技術情報センター 技術開発部 部長
佐藤　繁　㈱富士通研究所 常務取締役 情報社会科学研究所 所長
曽我　正和　三菱電機㈱ 情報システム研究所 所長
田中　英彦　東京大学 工学部電気工学科 教授
田中　穂積　東京工業大学 工学部情報工学科 教授
田村浩一郎　通商産業省 工業技術院 電子技術総合研究所 次長
都村　友紀　松下電器産業㈱ AV&CCシステム研究開発センター
東京情報システム研究所 所長
鶴保　征城　ＮＴＴデータ通信㈱ 取締役 開発本部長
堂免　信義　㈱日立製作所 電機システム事業本部 技師長
長尾　眞　京都大学 工学部電気第２教室 教授
中島　隆之　シャープ㈱ マルチメディア開発本部
応用システム研究所 所長
中村　直司　日本電信電話㈱ ＮＴＴ情報通信網研究所 所長
羽下雄之輔　沖電気工業㈱ 研究開発本部 総合システム研究所 所長
服部　茂久　(財)金融情報システムセンター 理事
溝口　文雄　東京理科大学 理工学部経営工学科 教授
南　正名　㈱東芝 研究開発センター 情報・通信システム研究所 所長
山田　尚勇　文部省 学術情報センター 教授 研究開発部長
横井　俊夫　㈱日本電子化辞書研究所 所長

海外委員 ： Edward A. Feigenbaum

(アルファベット順) Professor, Knowledge Systems Laboratory

Department of Computer Science, Stanford University

U.S.A.

Peter M. D. Gray

Professor, Department of Computing Science, University of Aberdeen

U.K.

Chen Liwei

Research Center for Computer and Microelectronics Industrial

Development (CCID), China

Jacques Mathieu

Direction Générale des Stratégies Industrielles

Sous-Direction Informatique et Bureautique

Ministère de l'Industrie et du Commerce Extérieur, France

Reind P. van de Riet

Professor, Department of Mathematics and Computer Science

Section of Information Systems, Free University Amsterdam

The Netherlands

Christian Rohrer

Professor Institute for Computational Linguistics

University of Stuttgart, Germany

Vadim L. Stefanuk

Professor, Vice-chairman of coordinating Council of Soviet

Association for AI, Institute for Information Transmission

Problems, Russian Academy of Science, Russia

Donald E. Walker

Bell Communications Research (Bellcore)

U.S.A.

オブザーバ：　古瀬　利博　通商産業省 機械情報産業局電子政策課 課長補佐

西川　勇　通商産業省 機械情報産業局電子政策課 技術係長

武涛雄一郎　通商産業省 大臣官房情報管理課　総括班長

垣本　和則　通商産業省 特許庁電子計算機業務課機械化企画室 企画班長

佐藤　真輔　科学技術庁 科学技術振興局科学技術情報課 課長補佐

事務局　：　片岡　幸一　(財)日本情報処理開発協会 AI・ﾌｧｼﾞｨ振興ｾﾝﾀｰ 調査課長

大規模知識ベースの構築と共有に関する国際会議１９９３（ＫＢ＆ＫＳ’９３）

プログラム・実行委員会・名簿（平成５年11月現在）

（敬称略）

委員長 ： 横井 俊夫 (株)日本電子化辞書研究所 研究所長

国内委員 ： 麻田 治男 (株)東芝 研究開発センター情報・通信システム研究所
（50音順） 第二研究所 所長

内田 裕士 (株)富士通研究所 情報処理研究部門知識処理研究部 部長

北野 宏明 カーネギー・メロン大学 機械翻訳研究所 研究員

土屋 俊 千葉大学 文学部行動科学科 助教授

徳永 健伸 東京工業大学 工学部情報工学科 助教授

西尾章治郎 大阪大学 工学部情報システム工学科 教授

西田 豊明 奈良先端科学技術大学院大学 情報科学研究科 教授

新田 克己 (財)新世代コンピュータ技術開発機構 研究所 第２研究部長

服部 文夫 日本電信電話(株) ＮＴＴ情報通信網研究所知識処理研究部
主幹研究員

日夏 健一 日本科学技術情報センター 技術開発部技術開発課
副主任情報員

堀 浩一 東京大学 先端科学技術研究センター 助教授

松本 裕治 奈良先端科学技術大学院大学 情報科学研究科 教授

溝口理一郎 大阪大学 産業科学研究所 教授

村木 一至 日本電気(株) 情報メディア研究所
メディアテクノロジィ研究部 部長

元吉 文男 工業技術院電子技術総合研究所 知能情報部自然言語研究室
室長

横山 晶一 山形大学 工学部電子情報工学科 教授

海外委員 ： Hyan Alshawi
（ｱﾙﾌｧﾍﾞｯﾄ順） Cambridge Computer Science Research Center
SRI International, U.K.

Christian Boitet
Professor, Grenoble University
France

Ronald J. Brachman
Department Head, Artificial Intelligence
Principles Research Department, AT&T Bell Laboratories, U.S.A.

Jaime G. Carbonell
Professor, Center for Machine Translation
Carnegie-Mellon University, U.S.A.

Richard Fikes
Professor, Department of Computer Science
Stanford University, U.S.A.

Kenneth Haase
Professor, Media Laboratory
Massachusetts Institute of Technology, U.S.A.

Jiawei Han
Associate Professor, School of Computing Science
Simon Fraser University, Canada

James A. Hendler
Associate Professor, Computer Science Department
University of Maryland, U.S.A.

Bob Jansen
Program Manager, Knowledge Based Systems
Division of Information Technology
CSIRO, Australia

Dimitris Karagiannis
Professor, Knowledge-Bases Systems Department
University of Vienna, Austria

Douglas B. Lenat
Director of the Cyc Project
Microelectronics and Computer Technology Corporation (MCC), U.S.A.

Mark Liberman
Professor, Department of Linguistics
University of Pennsylvania, U.S.A.

Nicolaas J. I. Mars
Professor, Department of Computer Science
University of Twente, The Netherlands

John McNaught
Professor, Center for Computational Linguistics
University of Manchester Institute of Science and Technology, U.K.

Desai Narasimhalu
Professor & Program Manager, Institute of Systems Science
National University of Singapore, Singapore

Robert Neches
Professor, Information Science Institute
University of Southern California, U.S.A.

Michael Sperberg-McQueen
Professor, Computer Center
University of Illinois at Chicago, U.S.A.

Luc Steels
Professor, AI Laboratory
Free University of Brussels, Belgium

Oliviero Stock
Istituto per la Ricerca Scientifica e Tecnologica (IRST)
ITALY

Jun-ichi Tsujii
Professor, Center for Computational Linguistics
University of Manchester Institute of Science and Technology, U.K.

Yorick Wilks
Professor, Department of Computer Science
University of Sheffield, U.K.

オブザーバ： 古瀬　利博　通商産業省 機械情報産業局電子政策課 課長補佐
西川　勇　通商産業省 機械情報産業局電子政策課 技術係長
佐藤　真輔　科学技術庁 科学技術振興局科学技術情報課 課長補佐
村松　弘和　科学技術庁 科学技術振興課科学技術情報課
田中　裕一　㈱富士通研究所 ソフトウェア研究部

事 務 局 ： 市川　隆　(財)日本情報処理開発協会 常務理事
片岡　幸一　(財)日本情報処理開発協会 AI・ﾌｧｼﾞｨ振興ｾﾝﾀｰ 調査課長

目　次

第２部　ＫＢ＆ＫＳ’９３国際ワークショップ　－　発表要旨・質疑応答　－

資 料

第　1　部

大規模知識ベースの構築と共有に関する国際会議１９９３

ＫＢ＆ＫＳ’　９３　国際会議

―　会　議　録　―

1　開会セッション

1. 開会セッション

1.1 主催者挨拶

財団法人　日本情報処理開発協会
会　長　　影　山　衛　司

皆様、おはようございます。本日は皆様方、大変お忙しい中を大勢の皆様にお越しを頂きまして誠に有り難うございました。大規模知識ベースの構築と共有に関する国際会議１９９３の開催にあたりまして、主催者といたしまして、一言ご挨拶を申し上げさせて頂きます。この国際会議は、本日１２月１日から４日まで、内外から著名な学者・研究者の方々をお招きいたしまして、この会議を２日間と、その後ワークショップを２日間開催するものでございます。

また本日の開会にあたりましては、公務ご多忙な中を、来賓として、通商産業省機械情報産業局の荒井次長、科学技術庁の加藤審議官、ならびにアメリカ合衆国国立科学財団プログラムディレクターのＭｒ．Ｓｕ－Ｓｈｉｎｇ　Ｃｈｅｎさんのご出席を賜っておりまして、後ほどご祝辞を賜ることになっております。厚く御礼を申し上げる次第でございます。さて、皆様ご承知の通り、コンピュータと通信におけるめざましい技術革新によりまして、世界の主要国における情報化は急速に進展をしておりますが、２１世紀に向けた高度情報化社会の発展のためには、人間の持つ知識を活用することによって様々な問題の解決をはかる知識情報処理の実現のための、新しい情報化基盤の確立が求められておるのでございます。特にコンピュータで直接利用できるように加工された辞書や百科事典、あるいは各種の学術研究情報・特許情報・経済情報などの経済文化の様々な分野における知識情報を、大規模知識ベースとして、世界の人類共通の財産として利用することは、社会・経済・文化の発展に大きく寄与するものでありまして、このような将来の知識処理インフラストラクチャーの中核となる大規模知識ベース技術のグローバルな研究開発が必要になってきておるわけでございます。本国際会議はこのような趣旨に基づきまして、フランス・ドイツ・イギリス・アメリカ・欧州共同体並びにわが国の政府、政府機関の御後援と、関係諸団体の御協賛のもとに、各国におけるこの分野の第一級の学者・研究者による講演と発表、ワークショップにおける若手研究者による研究交流を行いまして、今後における大規模知識ベース技術の発展に寄与してまいりたいとするものでございます。

本国際会議の開催にあたりましては、渕先生を委員長とする組織委員会、また横井先生を委員長とするプログラム実行委員の委員各位に多大な御尽力をいただきました。厚く御礼を申し上げます。本会議が先の成果を納めますように皆様の御声援と御協力をお願いを致しまして、私の開会の御挨拶とさせていただきます。どうも有り難うございました。

1．2　来賓挨拶

通商産業省 機械情報産業局
次　長　　荒　井　寿　光

皆様、おはようございます。世界各国からようこそご参加いただきまして本当に有り難うございます。ＫＢ＆ＫＳ’９３国際会議の開催にあたりまして、一言ご挨拶を申し上げます。

大規模知識ベースは、２１世紀の地球規模の高度情報化社会を築く情報インフラとして重要であります。また、人間の知的活動を飛躍的に高度化させるものとして、その発展が非常に期待されております。本国際会議は知識ベース分野での初めての本格的な国際会議であります。数多くの諸外国の参加や協力のもとに開催され国際的な議論を図りますことは、極めて画期的なことであります。今コンピュータ産業は大きな変革の時期を迎えております。新たなコンピュータ革新技術が求められております。コンピュータの利用技術に関する研究開発は活発に行われております。情報化の波は、産業や行政のみならず家庭にも浸透しつつあります。今後も情報化や情報環境整備を積極的に推進しなければなりません。知識ベースの分野について見ますと、大規模ということ自体が本質的な技術開発課題であります。これをブレイクスルーすることが重要な課題になっております。このような状況の中で、大規模な知識ベースの構築やこれを共有し利用する技術に関する国際会議が開催されますことは、誠に時宜を得たものであります。

情報化は、知的活動の質及び生産性を高めるものとして決定的な重要性を持つものであります。高度情報化社会へ向けた創造的基盤として積極的に推進する必要があります。このため通産省は次のような政策を進めております。

第一に、研究・教育・行政といった公共分野の情報化投資の促進や、ネットワーク化等の推進であります。

第二に、民間の情報化を促進する環境の整備であります。

第三に、情報化を支える情報産業の構造改革の推進であります。

第四に、リアル・ワールドコンピューティング、第五世代コンピュータといった、情報化のフロンティアを拡大する基礎的情報技術開発の推進であります。

本国際会議のテーマであります知識ベース分野におきましても、わが国では電子化辞書などの研究開発が多くの成果をあげてきております。国際的にも様々な研究が活発に行われております。より高度で、より大規模な知識ベースを構築し、これを利用するための技術的環境は整ってまいりました。研究者の世界では、研究用ネットワークを通じた国際的な情報交換が常識となっております。２１世紀には様々な分野におきまして地球規模の情報流通が発展すると考えられております。しかしながらこのためには、技術的・社会的課題を解決しなければなりません。今回国際的な枠組みの中で議論を行い、コンセンサ

スを得られようとすることは極めて有意義なことであります。最後になりましたが、本国際会議の開催にあたりご尽力されました影山会長、渕委員長をはじめ国内外の委員の方々をはじめ関係者の皆様に対し、敬意を表します。本国際会議が大きな成果を納め、この分野がますます発展することをお祈りいたします。有り難うございました。

科学技術庁
長官官房審議官　加藤康宏

本日は、このような盛大な会議、それも情報処理技術の将来に向けて国際的に議論を深めていく場にお招き頂き、光栄に存じます。

今や社会の隅々まで情報化の波が押し寄せる時代となっております。科学技術を活用し、より豊かな社会を育んで行くためにも、科学技術の情報化への対応を適切にしていく必要がございます。この為私どもは、まず研究機関相互の研究情報の交換や研究協力を促進していくために、研究領域、省庁、さらには国の枠を超えて研究機関を接続する省際研究情報ネットワークの整備を行うこととしております。わが国は、諸外国と比べてネットワークの整備が遅れているという状況にございますので、今後、関係省庁と連携を図りつつ、計画的にその整備を図っていきたいと考えております。また、併せて、国研等でつくられる基礎的・基盤的データ等のデータベース化を行うことにより、研究情報の整備も図っていくとともに、本ネットワークを介して流通させることを考えております。また、傘下の日本科学技術情報センターについては広くいろいろな方からご利用されているところでございますが、さらに先月より文部省学術情報センターのネットワークとゲートウェイで接続されることにより、利用できるデータの種類や量が大幅に増大しました。また、日本科学技術情報センターのネットワークは海外とも接続されており、本日お集まりの方々を始め、広く国内外での利用がいっそう促進されるよう我々も努めていきたいと考えております。

さて、本会議のテーマでございます、大規模知識ベースの構築とその共有ということにつきましては、ご案内の通り２１世紀における高度情報化社会の実現をするためにもその基盤となるものでございます。科学技術庁としましても、科学振興のために不可欠なものとして大いに関心を持っているところでございますし、本分野における研究開発の進展により、科学技術分野だけではなく社会や経済にとっても、大きなブレークスルーがもたらされることを期待しております。

最後になりましたが、本国際会議の成功と今後の一層の発展を祈念して挨拶とさせていただきます。どうも有り難うございました。

National Science Foundation
Program Director　Su-Shing Chen

おはようございます。先進工業国、また、実に全世界が今、第一世代情報時代から第二世代情報時代への転換期にさしかかっています。第一世代情報時代は、ポスト産業サービス社会、あるいはAlvin Toffler（１９８０）の言う第３の波の社会です。しかし、第二世代、すなわち来るべき情報時代は、高度の情報もしくは知識集約社会です。大規模知識ベースの構築と共有は、この来るべき時代に欠くべからざるものとなるでしょう。この国際会議はこうした研究にとって有意義かつ時宜を得たものです。

米国はいわゆるＮＩＩ（国家的情報基盤 National Information Infrastructure)を計画しています。これは通信ネットワーク、コンピュータ、データベース、および民生用電子機器の継ぎ目のない網です。大量の情報がユーザーの指先にもたらされます。普遍的サービスの概念を導入し、情報資源が誰にでも手頃な価格で提供されることを目指しています。

私は、共有資源の中の知識と言語処理の方法論は、ＮＩＩの上を将来流れるであろう情報製品を実現するための要となる科学であり、技術であると考えています。従って、大規模知識ベースの構築と共有は重要であり、また時宜を得ているわけです。

第一に、ＡＩおよび言語研究は、１０年前よりはるかに成熟しています。ＮＳＦでは、ＡＩや学習、また言語に関する非常に興味深いプロジェクトを援助しております。コネクショニストによるＡＩパラダイムの追究、種々の学習方式、高度の発話・言語方法論、および人-コンピュータ対話における認知科学など、大変活発で刺激的な研究分野をなしています。

第二に、初期の研究開発プロジェクトを通じて我々は貴重な経験をつんできました。第二世代の大規模知識ベースは新しいＡＩ学習および言語の成果を利用することができるかもしれません。これは大変やりがいのあるおもしろい仕事です。現在この方向での研究活動にＮＳＦが援助を行なっています。

最後に、この国際的な共同研究のために、ＮＳＦの援助によって１０人以上のアメリカの科学者の方々がこの会議とワークショップに参加しておられます。皆さんがこの重要な催しの成功のために活発な貢献をなさると信じております。有り難うございました。

1．3　基調講演

(1)「幼年期から青年期へ」

KB＆KS'９３ 組織委員会
委員長　渕　一博

皆さん、おはようございます。最初に組織委員長として、ある意味で非常に画期的な最初の国際会議に皆様多数お集まりいただいて、大変嬉しく思うと同時にお礼を申し上げたいと思います。この会議の準備については、いろいろな人々の作業とか、関係のいろいろな機関のご支援がありましたが、これにも深くお礼を申し上げたいと思います。

さて、私の基調講演という題名でのお話ですけれども、予稿には個人的な歴史を少し書いてみました。しかし、これは後でお読み頂くことにして、ここで私が言いたかったことというのは次のようなことであります。

一番言いたいことは、技術の発展というのは大変早いように見えますが、よくよく見てみると技術の発展は、必ず段階的な発展をするということであります。ということは、技術屋あるいは皆さんもいろいろな夢をお持ちだと思いますけれども、その夢というのは一足跳びに実現されるのではなくて、いろいろな技術開発のステップ、あるいは産業の展開のステップ、あるいは社会の発展のステップを踏まえながら次のステップにいく。そのステップをクリアしますと、また次の努力が始まって次のステップへ進んでいく、というようなパターンだろうと思うわけであります。これは、コンピュータとか情報処理の技術だけではなくて、他の技術一般にも当てはまるのではないかと思っておりますけれども、情報処理の技術の発展についても私どもはそう感じているわけです。

これは私の個人的なことでいいますと、学生時代から現在までの４０年近い体験で得られた感想であります。そのことを別な言葉でいいますと、技術の発展においても歴史の非常に大きな流れがある。非常に大きな流れが底流としてあるわけでありまして、その線に沿って一歩一歩技術が展開していくというふうに思うわけであります。この４０年程を考えますと、コンピュータとか情報処理の発展というのは大変目ざましいものがあって、こんなに急速に発展した技術があっただろうかというふうな驚きの念を禁じ得ないわけでありますが、かたやそのコンピュータとか情報処理が目指す夢の目標、その内の１つは人間の知能に限りなく近づこうというような目標から測りますと、技術の進歩というのはまだまだ最初の段階、非常に萌芽的な段階であると言っていいと思います。３０年４０年といいますと大変長いような気がいたしますけれども、しかし、人類の歴史の長さ、発展の未来ということを考えますとこの数十年というのは、一瞬の出来事かも知れないわけです。この一瞬の中にいろいろな発展がありましたし、また片方では、この技術の発展に対して、ある時は非常に楽天的に思う雰囲気が、ある時は非常に

悲観的に思う時期があったり、いろいろな波があったと思います。現在はどちらでしょうか。世の中全体の大変な大きな変動の中で、やや悲観的な観測が多いわけですけれども、しかしながら、私自身はこれもまた次の発展への過渡的な現象だと思っております。

ところで、私自身はこの十数年間、第五世代コンピュータ・プロジェクトというものに関わってきたわけであります。このプロジェクトを始める時にいろいろな議論がございました。日本の中でも多数の人が議論に参加しましたし、また国際会議の開催を通じて世界の研究者とか、あるいは研究に関わるような人々との議論を行ったわけであります。その時に議論してまとめた一つの方向性というものがございます。これは、情報処理の将来の方向というのは、大きくみれば人間の知能ということでしょうけれども、１０年、２０年、３０年というオーダーで考えれば、知識情報処理という方向に進むであろうということを議論して、そういう結論に達したわけであります。そういう大きな流れの中で、基礎的な研究としては何をやるべきか、これもいろいろ議論しました。直接的にこの知識あるいは人工知能の研究というところに狭く焦点を絞っていくということも議論されたわけであります。片方では、そういう方向を支えるコンピュータ技術自身を見直す必要があるのではないかという議論をしたわけであります。そういう議論の中から私たちは、コンピュータというのは一つには、並列コンピュータの方向に進むであろう。それと同時にそのコンピュータの構成、これはハードだけではなくソウトウェアを含めて考えますと、コンピュータの構成は知識処理の根幹であります推論という操作を、基礎として、基本として持つような体系にいくのではないか。そういう議論をしまして、「並列推論」という技術的なスローガンをつくったわけであります。実際に推進した第五世代コンピュータのプロジェクトというのは、その当時の多くの人が思ったような、この人工知能ないし知識処理技術を１０年間で仕上げるということではなくて、そういう時代をつくるための技術的な基礎として並列・推論型のコンピュータ、あるいはコンピュータ技術をつくろうということであったわけです。

このプロジェクトは幸いなことに通産省の努力があって、ナショナルプロジェクトとしてスタートしまして、その後多くの若い意欲的な研究者たちの努力によって、その方向での具体的な研究成果を積み上げてきたと思っております。ということで、この第五世代コンピュータのプロジェクトというものもそれだけで終わるものではなくて、これはその後の時代のための一つの礎石であろうと思って取り組んできたわけですし、そういう観点でいいますとそれなりの研究成果を上げてきたと思っております。しかしながら、世の中全体を考えますと、この第五世代コンピュータも必要なことの一つであったわけで、すべてであったわけではありません。その他に関連するいくつかのプロジェクトが進行しましたし、また片方では、産業のレベルで技術の発展があったのはご承知の通りであります。

１０数年経った現在考えますと、この１９８０年代というのは、私達の第五世代コンピュータの研究プロジェクトを含めて実にいろいろな努力がなされたと思います。ものによっては、非常に着実に成果を上げたものがありますし、テーマによっては期待した程進まないということもあったかもしれません。あるいは、予想以上に進展した技術もあった。そういうこと全体を考えますと、８０年代というのは大変大きな技術的な蓄積、あるいはそれを踏まえた社会的な活動の広がりがあったと思います。しかしな

がら、私が感じますのはこれはまさに次の１０年間、言い換えれば、２１世紀に向けてのいろいろな準備であったというふうに言ってもいいと思うわけです。私どもが１０数年前に情報処理の方向性として主張した知識情報処理への動きというのは過去の１０年間で終わったわけではなくて、それはほんの序の口であったと思うわけです。現在、やや悲観的な風潮もありますけれども、私の感じではこの知識情報処理の時代というのはこれから始まる。１９８０年代のいろいろな技術的な蓄積を踏まえてこれから始まっていくものだと思っております。これは１０数年前にもそういうふうに思っていたわけですが、それから１０年あまり経った現在でも、そういうふうに私は感じているわけであります。そういうことで、私どもが進めた第五世代コンピュータのプロジェクトもその準備の一つであったわけですし、このプロジェクトと非常に密接な関連を持って進められたもう一つの電子化辞書プロジェクトというものも、次の時代への準備であったと思います。電子化辞書プロジェクトについては、その推進者である横井さんが後ほどいろいろ述べられると思いますけれども、７、８年のプロジェクトで、今辞書ができつつあるわけですけれど、これができてそれで終わるということではなくて、実はこのデータをベースにして次のいろいろな展開がある。これは単にその電子化辞書が製品化されるというのではなくて、むしろ大きな意義というのは、これが次の、これからさらに必要である基礎的な研究の土台になるという意味で、私は大きな意義があると思っているわけです。

いろいろな研究をやりますと、研究者の中にはくたびれて、これが我々の力の限界だと思う人も出てくるわけですが、しかしながら、実はその後に若い人達が育ってきているわけです。そういう人達は、実は大変な苦労をしながら研究をしたり、作業をしてくたびれはてたその前の世代の人達の残した成果を踏み台にして次のステップに進むというのが、一般的なパターンだと思いますが、今日から議論する分野においても、次の世代の人達が次のステップに進むための土台をつくった。この電子化辞書というものもそういう土台をつくったものであると思っております。そういうふうに思いますといろいろな土台が出来ているのでありまして、ほんとうに難しくて面白い人工知能的な研究テーマというのは、実はこれから始まるといっていいと思います。かつていろいろな夢を見てくたびれた研究者もいるわけですけれども、これは正に先人の苦労であるわけですけれども、ある意味で時代が早すぎた、そのためのいろいろな苦労があって場合によっては挫折したと思うんですけれども、これからの時代というのは、そういう人達の時代とは違っていろいろな土台が増えてきている。そういうことを踏まえて次のステップに進めるという、非常にいい時代がこれから始まると思っております。そういう時代のためにこの８０年代を含めていろいろな努力が成されたわけです。

もう一つ私が思いますのは、単に先端的な研究だけを進めるというのではなくて、これから必要になる基礎研究のためには、それを支えるそのインフラストラクチャーの整備というものが大事だと思っています。そのインフラストラクチャーのための素材は、これまで私が述べましたように日本でもいくつかの努力が成されましたし、世界中を見渡しますといろいろな人の努力があって、素材としては大変豊富に育ってきているわけです。それらを有機的に構成して、先ほど言った次の時代の若者達が活躍してもらうために、これまでの蓄積されたそれらの素材を有効に有機的に構成して、一種の研究インフラス

トラクチャーを提供する必要があると思っております。これに関していいますと、実はこの２、３年大変大きな動きが始まりつつありまして、日本の中でも、例えば、コンピュータ・ネットワークの整備が必要だということが、研究者の社会だけではなくて、行政の人達、あるいは政治の人達でも議論されるようになっています。その点では大変な進歩ではありますけれども、しかしまた逆をいいますとコンピュータネットワークを含めたその研究環境、あるいは研究のためのインフラストラクチャーの整備というのは日本は実は大変遅れていたわけです。見かけは日本は経済的あるいは技術的に大進歩したように見えますけれども、それは、それ以前の蓄積を食いつぶして発展したものだったと思います。次の時代のために残す貯金というのは、実はだいぶ減ってしまっているわけです。勿論先ほど述べましたようにいくつかの新しい素材が生まれたわけですけれども、研究インフラという意味ではいろいろなところを実はなおざりにしていたと思います。これは一部のプロジェクトではなくて、社会全体を考えるとそう私は思うわけです。一つの典型的な例というのは日本における大学の人達の研究環境でありまして、これは遠くから見るより非常に劣悪なものがあるわけです。幸いなことにこの数年、これではいけない、大学の研究機関を含めて、日本の研究機関のあり方を基本から見直そうという動きが始っています。それは、どういう研究をするかという研究のテーマも片方で大事ですが、そういう研究を展開するための基本的な環境というものも大事だという事であります。

そういう認識がやっとこの数年始まったと思っております。そういうことで、実はこれからそういうことの本格的な展開が実施される事を期待しているわけでありますが、その時にまたいくつかの事を考える必要があると思うんです。例えばコンピュータ・ネットワークというのは一番ベースにあるインフラストラクチャーであります。しかしながら、情報処理の研究ということを考えますと、そういうハード的な環境だけではなくて、その上に蓄積されている、いわゆるソフトウェアという蓄積もまた次のステップのための研究インフラストラクチャーだと思うわけです。そういう観点でいいますと、また日本の例でいいますと、この７、８年進められた電子化辞書というようなものは、辞書という分野で一つの蓄積をしたわけですが、それだけで十分なわけではないわけであります。これから本格的に始まるこの知識情報処理の時代のためには、英和辞典のワンセットがあれば十分というわけではなくて、それは一つのサンプルです。そういう辞書を含めたこの膨大な知識、人間が過去何十年、何百年をかけてあるいは何千年をかけて蓄積してきた膨大な知識というものを、電子化といいますか、コンピュータ技術と組み合わせて流動化する。そういう時代がこれから来るわけです。そのためには単に図書館に書物の形として存在する知識の集合だけではなくて、それ自体を自由に電子的にアクセスして利用するという、そういう非常に大規模な知識の集積というものも次の時代のための研究のインフラストラクチャーとして必要だと思うわけであります。そのための先駆的なひな型的な努力というのが電子化辞書であったわけですし、コンピュータ技術としていえば、第五世代コンピュータの技術であったと思うわけです。そういうことで、今日から始まります４日の会議というのは非常に大きな歴史の変わり目でのイベントとして捉えていいのではないかと私は思っているわけであります。

さて、いろいろこれから皆さんに議論していただくわけですけれども、この大規模な知識の集積、こ

れをどういう形で行うかというのはいろいろな観点があっていいと思います。結果的にはいろいろなグループでのいろいろなこの努力の蓄積というものが集まってできるわけですが、その中で大いに議論して頂きたいと思う事は、今回の会議の題名にもあります「シェアリング」という観点であります。今までですとどちらかというと「ビルディング」、どうつくるか、どういう研究を展開するか、という議論に集中することは多かったと思うんですけれども、そういうものが積み上がっていく時にもう一つ大事な観点というのはこのシェアリングということだと思います。これは、またいろいろな意見を皆さんお持ちだと思います。場合によってはこの大事な知的な財産をシェアするというのは大変問題があるという人もいるかもしれません。あるいは私のように、知識とか技術というものは本来人類共通の資産であるから、本来はすべて自由にみんなが共有してその果実を利用できるべきだと思う人もいるかもしれません。その間にいろいろな意見が存在していいと私は思っておりますけれど、そういうことも大いに議論していただくという機会としてはこれは大変画期的だと思うわけであります。このシェアリングというテーマというのはほとんど真面目に議論されたことがないと私思っておりまして、これを機会に議論が進展することを期待するわけであります。

一つのイグザンプルを申し上げますと、第五世代コンピュータ・プロジェクト、これはいろいろな意見がありましたけれども、私どもスタートした時から、そこで得られた技術的な知見とか成果というのは、基本的には、すべての人に自由に利用して頂きたいということで研究を進めてきたわけです。しかしながら、１０数年前は、まだ研究成果がないわけで、ないものについてこれを自由に利用して頂くといっても机上の空論であったわけですが、プロジェクトの結果、大変多くの成果が生まれております。それは主としてソフトウェアという形で表現されているわけですけれど、これをフリーソフトといいますか、世界中からコンピュータネットワークを通してもいいし、直接郵便でもいいかもしれませんけれど、自由に皆さんにアクセスしていただいて、それを調べて、この技術はいいとか、あるいは場合によってはまだ不十分であるとか、その次のステップのための素材にして頂くということが実現しています。しかしながら、広く世の中全体を考えますとこれはまだほんの一例でありまして、そういう主旨で進めてきたプロジェクトというのはまだ少ないわけです。これからはいろいろなプロジェクト、特にナショナルプロジェクトあるいはインターナショナルプロジェクトでもいいんですが、ナショナルプロジェクトというものは、方向としては五世代プロジェクトで行なったように、その研究成果を一種のフリーアセットとして全世界に提供するということになっていくのではないか、あるいはそうあるべきではないかと思います。特に日本のナショナルプロジェクトはそういう方向に進むべきだろうと思っております。そう思う人は増えているわけで、それの最初の例が第五世代コンピュータプロジェクトの成果であるソフトウェアの、自由なアクセスということであります。ご興味のある方は、その資料もこの会議で入手していただけると思いますので、見ていただければと思います。

ということで、基調講演としては少しざっくばらんなお話をさせていただきましたけれども、今日から始まります２日間の本会議、それから、あと２日間のワークショップを通じてこの次の時代のための大規模な知識ベースの、このビルディングとシェアリングについて活発に意見を交わして頂いて、これ

からいろいろなところで始まる活動の基礎として頂ければ幸いだと思っております。それでは、これで私の話を終わらせて頂きたいと思います。どうもありがとうございました。

(2)「知の空間を構成する大規模知識ベース」

KB&KS'93 プログラム・実行委員会
委員長　横井俊夫

ただ今ご紹介に預かりました横井です。それでは、基調講演の後を受けまして、大規模な知識ベース、「知の空間を構成する大規模知識ベース」と題しましてお話をさせて頂きます。通訳の便宜のために原稿が出来ておりますので、これを見ながらお話させていただきます。

大規模知識、大規模知識ベース、大規模知識ベースシステムを人工知能を含む情報処理分野における最重要課題と位置づけたいと思います。「大規模知識」という言葉では、知識として利用できるように構造化された大量の情報そのものを指すことに致します。「大規模知識ベースシステム」という言葉では大規模な知識の構築・管理・利用を自動化ないしは高度に支援する機能を持つシステムを指すことに致します。そして、「大規模知識」と「大規模知識ベースシステム」を併せて代表させる時に、「大規模知識ベース」という言葉を用いることに致します。

◎何のために必要か

3つの観点から、大規模知識ベースの位置付けを試みます。本テーマの重要性と取り組みのための緊急性をご理解いただけるものと思います。

第一、知識インフラの構築と高度化の技術として：

グーデンベルグ以来の印刷文字文化から、電子文字文化への壮大な移行が始まったといわれております。電子化された知の空間の構築であります。巨大な知の空間が整備されますと、コンピュータの利便性は格段に向上致します。知の空間をコンピュータによって縦横に駆け巡ることによりまして、人々の知の生産性は飛躍的に向上することになります。

これは「コンピュータの人工知能化」というより「知の人工空間の建設」という方がふさわしいと思います。また、「コンピュータ中心の見方」から「情報中心の見方」へと大きく世界観を変えるものでもあります。

「知の空間」は、「物の空間」という物理的世界と等しい、あるいはそれ以上の広がりを持つ情報の世界を形作ります。その建設には21世紀の大半をかけることになりましょう。そして、「知の空間」の多くは社会や組織で共同利用される施設として建設されます。ここに、「知識インフラ」、「社会資本としての知識」という考え方が生まれることになります。

知識インフラは、「機械可読化」、「ハイパー化」、「インテリジェント化」のステップを大きく踏みながら高度な知の構造物へと成長して行くことになります。「機械可読化」とは、情報をとりあえずそのままの形態でコンピュータ処理可能な状態にするステップであります。「ハイパー化」とは、情報をそ

の本来の構造へと再構築するステップであります。これは「ハイパーメディア」の「ハイパー」とほぼ同じ趣旨であることから借用した言葉であります。「インテリジェント化」とは、情報を知識としての体系に整えるステップであります。

現在、知識インフラを本格的に議論できる諸条件がようやく整ってきました。その中でも特に重要なのは「知の空間」が「物の空間」の制限から解放され、「知の空間」そのものとして議論できるようになったことであります。例えば、コンピュータのパーソナル化は誰もが自由に「知の空間」を作り、利用できるようにしてくれております。コンピュータネットワークは物理的な距離にわずらわされることなく「知の空間」を構成できるようにしてくれております。マルチメディアは人間がもともと持っている「知の空間」とコンピュータ上の「知の空間」の間の隔たりを無くしてくれております。もちろん、物理的制限が全く無くなったわけではありません。「知識インフラ」が巨大になるにつれ、制限の解消に向けた新たな努力が必要になるでしょう。以上のような高い利用価値のある、巨大な「知識インフラ」を効率良く構築していくには、そのための新しい学問や技術が必要であります。知識はどのような構造となっているのか。日常的知識・専門分野に共通の知識・分野ごとに特有な知識、それぞれどのようになるのか。人間やデータベースやテキストデータから多様な知識を効率良く抽出するには、どのようにすれば良いのか。大量の蓄えられた知識を誰もが有効に再利用できるようにするには、どのようにすれば良いのか。このような新しい多くの課題に取り組まなければなりません。「大規模知識ベースシステム」はこれらの課題に総合的に取り組み、そして「知識インフラ」の構築と利用のためのプロトタイプシステムを実現するのがその役割であります。

第二、人工知能のブレークスルーを求めて：

今までのお話しでは、「コンピュータの人工知能化」というよりも「知の人工空間の建設」を、という観点から大規模知識ベースを位置付けました。しかし、人工知能の側からもブレークスルーを達成する大きな手がかりとして「大規模知識ベース」への取り組みが始まろうとしております。

従来の人工知能研究では小規模な問題、小規模な知識を対象にし、処理の仕組みを研究し、出来上がった仕組みを大規模な問題・大規模な知識へとスケールアップするというアプローチがとられてきました。しかし、得られた処理の仕組みが結局は小規模さの特徴に依存してしまうため、多くはスケールアップに失敗しました。いわゆるトイプロブレム問題、「トイプロブレム・プロブレム」であります。

これに対し、「大規模な知識」からアプローチすべきであるという指摘がなされるようになってきております。処理の仕組の精度は悪く、処理のレベルは低いのですが、大量の知識から出発し、少しづつ精度やレベルを上げていこうというアプローチであります。

「大量の知識」を対象にするのでありますから、厳密な論理的な処理ばかりをしていたのでは、たちまち処理時間の爆発を生ずることになります。確率・統計的な処理や解析的な処理も重要になります。ただし、基本となるのは論理的な処理であります。また、「メモリベース推論」のような超並列処理への期待も大きくなります。さらに、「柔らかい処理」とか「あいまいな処理」とかいわれている一連の

処理モデル、すなわち、「ファジイ論理（ファジイロジック）」、「ニューロコンピューティング」、「ＧＡ」、「人工生命」などが有効性を発揮することにもなります。

また、「大規模知識」も「大規模データ」という程度のものであるならば、さほどの特別の努力は必要とは致しませんが、本格的な知識となると、「大規模知識」を用意するまでに多大な努力が求められることになります。また、「大規模知識ベースシステム」そのものが、「大規模知識」を対象とする典型的な人工知能システムにもなります。

「大規模知識」を人工知能のブレークスルーのためのより本質的なテーマととらえる考え方があります。知能現象の解明のためには、対象となる現象を正確に把握する必要があります。従来はそのような現象は無限の多様性を持つものと考えられてきました。しかし、実際には非常に大きな有限と考えてよく、この十年程の研究開発の努力によって十分に到達できる量であるという考え方が有力になってきております。

人間の日常的常識も、知識として整理されれば有限の量でおさえられます。Ｍｉｎｓｋｙ教授はその量を約２，０００万程と推定しております。どれ位のものを単位として考えるのか、いかなる知識表現法を用いるのか、いろいろな試みが必要であります。日常一般に用いる自然言語文も、プロトタイプ文としては１００万から１，０００万位の量で押さえられるといわれております。これらを正確に集積すれば、自然言語現象全体が把握されることになります。さらに、常識を表現するには自然言語が最も好ましいという意見もあります。いずれにしましても、このような「大規模知識」を収集するためには「大規模知識ベース」としての本格的な取組みが必要であります。

「人工知能」を実証的な学問に育て上げるためには、実験学的な土台をしっかりとさせる必要があります。知能現象は本来多彩なものであり、理論やモデルでとらえられるのはごく限られた側面であります。実際の知能現象を客観的に観測できるものとしてとらえ、大規模な実験を行い現象を解明していく、長期にわたる努力が必要であります。この実験人工知能学の拠り所となりますのが、「大規模知識ベースシステム」であります。

第三番目、意味処理に踏み込む情報処理へ：

今まで情報処理の各分野は対象とするものには深入りせず、できるだけ共通となる仕組を考え出すことを本分として研究開発が進められてきました。そして、今までの成果が得られてきました。しかし、より発展させ、より高度な仕組を実現するためには対象とする世界を積極的に取り込むことが重要になってきております。すなわち、「対象を取込んだ意味処理」に踏込むことが必要になってきているわけです。

自然言語処理や文書処理においても、意味処理や文脈処理への試みが重要になってきております。形態素処理や構文処理が安定して利用できる技術になるにつれ、単語や文や文章の意味の記述法や、電子化辞書などの言語知識ベースの研究が大きなテーマとして取り上げられるようになってきております。知識処理や知識工学においても、知識表現言語や知識ベースの構築の議論から、対象となる知識そのも

のの議論が大きくなってきております。「オントロジー」に関する議論などが代表例であります。データベースにおいても、データベースの機構に関する議論と並行して、対象データそのものをとらえようという研究が行われるようになってきております。データからの知識獲得や、関係や実体の共通語彙設定に関する研究であります。マルチメディアにおいても、ハイパーメディアの研究開発が本格化するにつれ、対象世界をどうとらえるかの研究が重要になり、知識処理への接近が試みられるようになってきております。ソフトウェア工学においても、再利用技術の研究開発の努力はやはり知識処理への接近となってきております。

本来、知識処理といいますのは、「情報処理のある特定の分野である」というより、「情報処理において対象世界を取り込んでいく際の共通の手法である」とみるべきものであります。したがって、情報処理を意味処理に踏み込ませるということは、情報処理の各分野を「知識ベースド情報処理」にするということであると考えることができます。この「知識ベースド情報処理」によって、個々の応用システムは、インテリジェントなシステム、次世代のシステムへと高機能化していくことになります。

◎どのような試みがなされているのか

常識を目標として：これはＣＹＣプロジェクトが代表例であります。これに関しましては明日レナートさんがその成果、最新成果をご報告してくれると思いますので省略させて頂きます。

二番目、語彙知識を対象として

代表例としてＥＤＲ電子化辞書が挙げられます。自然言語の基本語を主とする語彙周辺の知識を記述対象に選んだものであります。言語、特に汎用性の高い言語に関する知識というのは、現時点で安定した一般性を得る重要な方法であります。但しこれは、言語処理のためにとりあえず有用である知識に限定したものであります。

次は、工学分野の共通知識ベースとして

まだ、大規模とは言えませんが、知的ＣＡＤなどを目的とした共通知識ベースの試みがいくつか始まっております。記述対象を限定することによって、問題解決方式や記述方式にある程度の汎用性を確保しようというものであります。機械設計を対象に、「フィジカルフィーチャー」と呼ぶ単位を知識の基本単位にして、表現・収集を行う試みや、物理デバイスをモデリングするために必要な工学的知識などを対象に、工学分野の共用知識ベースの構築を目標にしたものなどがあります。

次は、分散協調システムとして

今まで述べてきたものとは別の方式で大規模さを達成しようというアプローチであります。ネットワーク上で各所に分散する様々な知識ベースを、協調して働くようにしようというものであります。そのためにインターフェースや、プロトコルの標準化や、システムアーキテクチャのオープン化などを計

ろうというものであります。今までの議論に沿いますと、「知識ベース」のシステム仕様を記述対象にした共通知識ベース作りと見ることもできます。アメリカのKnowledge Sharing Effort内のグループによります、"Collaborative Testbed"という試みがあります。分散協調するものの単位をさらに小さくとろうという試みもあります。知識コミュニティーなどの提案であります。これは、「大規模知識ベース」を「大規模ソフトウェア」と見立てるアプローチであります。再利用性の高い柔軟な構造を持つ「大規模ソフトウェア」を実現する手立てとして、自立して動作し、互いに協調し合うエージェントの集合という構成法を採用します。エージェントの粒度をどの位にするかが大きな論点となりますが、究極はＭｉｎｓｋｙ教授の「心の社会」、"Society mind"でありましょう。このあたりになりますと賛否はいろいろさまざまでありましょう。

◎どのようなものでなければならないのか

「大規模知識ベース」はどのようなものでなければならないのか、「大規模知識ベース」の基本的な要件を列挙することにいたします。

第一番目に解放系であること：「大規模知識ベース」を「閉じた知識システム」としてはなりません。人間の頭脳の中にある知識システムと連続的につながっていることであります。このつながりが知識が利用される環境や状況へのつながりを保証してくれることになります。「大規模知識ベース」と人間の知識システムを連続的につなげるための土台となるのは、知識を表現するメディアに、ある種の共通性・連続性があるということです。

二番目、分野に対する一般性を持つこと：基本となる知識の構造・知識ベースの機構・システムの機能はいろいろな知識分野に広く適用できるものでなければなりません。したがって、あまりに複雑で特殊な仕組みではなく、単純で適応範囲の広い仕組みを土台にすることになります。

第三、多様な表現手法に対応できること：知識表現や知識処理のためのさまざまな言語やモデルや手法に、広く対応できることであります。これは、「すべての言語やモデルや手法を包括する一つのもの」を作ろうというのではありません。外部仕様の統一的な記述法や相互変換・相互翻訳の仕組みを用意し、対応するということであります。

第四、進化しうること：質の向上・量の増大とともに漸進的に進化していけることであります。さらに、学習機能・自己組織化機能によって進化の効率が加速的に良くなっていくということであります。

第五番目、大規模な集積が可能であること：現状においても、研究開発の努力を行えば、諸条件を満たす十分に大規模な知識の集積が達成できるということであります。その集積されたものが十分な有用性を持ち、さらなる研究開発の努力の基盤となるということであります。

◎どのようなものを知識とするのか

知識をどのような形式のものにするのかによりまして、「大規模知識ベース」の性質や構造が定まっ

てきます。対象にしうる知識は、客観的に観測しうる形式をとるものだけにします。言語、広くは各種メディアによって表現されることにより、知識は、分析したり処理したりできるものになります。知識を表現するメディアを何にするかによって、「大規模知識ベース」の性質や構造が定まることになります。したがいまして、大規模知識ベースのあるべき姿を見通すために、それぞれの知識表現メディアの性質・役割・互いの関係などを的確に見極めねばなりません。

知識表現メディアは大きく二つに大別されます。とりあえずそれぞれをヒューマンメディア、及びコンピュータメディアと呼ぶことに致します。

ヒューマンメディアは、人類が知識の表現・蓄積・利用・伝達のためにその長い歴史の中で育んできたもので、人間を処理系・推論系・理解系とする知識表現メディアであります。記号化能力と汎用性の両面から最も高い能力を持つのが自然言語。これを「汎用知識表現言語」といたしまして、形式言語・図形言語という「応用向知識表現言語」があります。そして、情報・知識の直感的な表現能力に富む画像、音響が加わります。それぞれのメディアは他で置き換えることのできない固有の役割を持ち、適切に組合わせることによって、高い表現能力を発揮します。

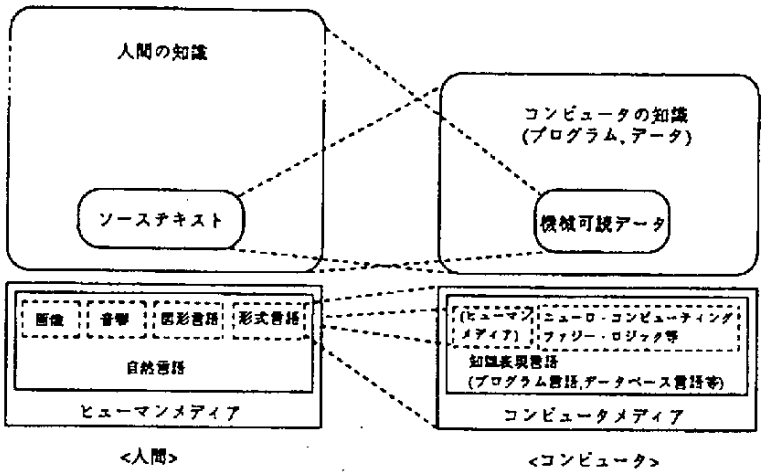
コンピュータメディアはコンピュータを処理系・推論系とする知識表現メディアであります。人工知能分野での知識表現言語・プログラム言語・データベース言語など多くのものが含まれますが、表現の対象や能力、そして処理効率などによって、それぞれの役割が割り振られます。また、ニューロコンピューティングやファジイロジックなど、まだ表現能力に十分な広がりを得るまでにはいたっていませんが、新しい能力を求めての工夫が続けられております。

重要なことは「大規模知識ベース」としてこの２つのメディアにいかなる役割を割り振るかということであります。知識工学からのアプローチでは、大規模な知識をコンピュータメディアによって表現しつくそうといたします。一方、ハイパーメディアのようなアプローチではヒューマンメディアに頼って進められます。しかしながら、研究開発の現状から、次のことは明らかになってきたといえるでありましょう。

まず、たとえ分野を限ったとしても、人間の持つ知識をコンピュータメディアで表現しつくすにはまだ程遠いところにあります。コンピュータメディアにおける最近の新しい工夫も、その効果のほどはごく限られたものであります。

コンピュータメディアでさえも「コンピュータが確実に処理できる」、あるいは、「確実に理解できる」のは、オブジェクトとして実行することだけであります。メタな処理、例えば「正しさを判定する」・「等価性を判定する」・「表現を作り出す」などに関しては、コンピュータが理解できるのはごくわずかであります。やはりほとんどは人間の役割であります。この事実を反映して、ＯＨＰのこの図では、コンピュータの知識がソーステキストとして人間の知識に含まれております。また、コンピュータメディア全体が、形式言語の一部としてヒューマンメディアに含まれております。

ただし、次の２点の例に見るように、メディアの技術にも新しい展開があります。



OHP－1

まず、マルチメディアの技術によって、ヒューマンメディアで表現された知識の相当広範囲のものを、とりあえず機械可読データとして扱うことが可能になりました。このレベルに限るならば、ヒューマンメディアもコンピュータメディアとなりつつあります。さらに、コンピュータ上で融合することにより、ヒューマンメディアは新しい機能を持つに至っております。この事実を反映して、この図では、人間の知識が機械可読データとしてコンピュータの知識に含まれております。また、コンピュータメディアがその表層部分もという意味で括弧を付け、「ヒューマンメディア」としてコンピュータメディアに含まれております。

次は、「ヒューマンメディアを、マルチメディア的なレベルを越えてコンピュータに理解可能なものにしよう」という努力も少しづつ実を結びはじめております。特に自然言語に関しては、確実な進歩が得られようとしております。コンピュータの処理能力は、マルチメディア的な文字列レベルから相当向上し、形態素や構文レベルに至っては、かなり広範囲に利用できる、ほぼ安定したものとなりつつあります。意味処理に関しては、ごく浅い理解に限れば、もう少しの努力の段階になってきております。文脈処理に関しても、限られた範囲内ではありますが、同様の試みが行われております。このようにして、図のヒューマンメディアは、表層から少しづつ深層へと機能を持つようになってきております。「機械可読データ」は、いずれは「機械認識可能データ」、「機械理解可能データ」へと高度化していくことになりましょう。

以上のことから、「大規模知識ベース」でのメディアの役割を考える上での指針としては、「知識は人間とコンピュータの複合系に対して表現される」ということなります。すなわち、理解の主役はあくまでも人間であり、人間の理解を適切に支援する機能を十分に発揮しうる程度にはコンピュータも理解できるようにする、というところが出発点としては妥当であります。

◎どのように実現していくのか

「大規模知識ベース」をどのように実現していくのか、実現へのステップをお話しいたします。まず、すべてに先立ちまして、「大規模知識ベース」が最終的にどのようなシステムに統合化されていくのか、そのイメージをお話することにいたします。そのイメージは「知識インフラのシステムイメージ」であります。

知識インフラのシステムイメージ：「知識の利用」「知識の管理」「知識の収集」、これらの機能をいろいろなレベルや比重で持つ、無数の「大規模知識ベースシステム」が、ネットワークを介し、分散協調しながら、全知識に対する知の世界を作り上げることになります。それぞれの「大規模知識ベースシステム」は、どのような知識を対象にするのかによってバラエティに富んだものになります。知識は、分野の違いや共有化の形態の違いによって多様なものとなります。共有化の形態としては、「国際的な共有化」・「国や社会による共有化」・「組織や機関による共有化」・「個人のプライベートな所有」などがあります。知識は人間用の知識表現メディアであります「ヒューマンメディア」と、コンピュータ用の知識表現メディアであります「コンピュータメディア」の両メディアが適切に組合わされ表現されることになります。人間の知的活動領域の拡大にしたがいまして、「ヒューマンメディア」の機能も向上していきます。技術の進歩にしたがいまして、「コンピュータメディア」の機能も向上していくことになります。

このイメージからもお分かり頂けますように、知識インフラは統合的な人工知能システムであります。この「知識インフラのシステムイメージ」に向けてこれからの研究開発が目標とすべき、「大規模知識ベースのプロトタイプシステム」をご説明いたします。

大規模知識ベースのプロトタイプ：

○まず最初に「システム機能」、システムの基本機能は次の３つであります。

第一が知識ベース機能：大量の知識を体系的に蓄積するための機能であります。適切なレベルの学習能力、自己組織化能力を持つことになります。また、この機能を規定する知識表現言語には、高い汎用性と効率の良い処理系の存在が求められます。これには第五世代コンピュータプロジェクトの成果が土台となります。

二番目は知識獲得支援機能：テキストデータやデータベースという知識素材、そして専門家からの知識獲得・収集を高度に支援する機能であります。知識素材からの知識獲得は、極力自動化するのが望ましいのでありますけれども、良質の知識を得るには、獲得過程への人間の介入や人手による事前編集が必要であります。専門家からの知識獲得では、獲得作業を行うインタビュアを介入させる時もありますが、専門家自身やグループによる作成を支援する環境の整備の方が重要であります。

第三番目、知識利用支援機能：色々な利用に特化された知識ベースの作成を高度に支援する機能であります。「大規模知識ベースシステム」は、プロトタイプとしては「マスター知識ベース」であります。ある部分・あるレベルのみを取り出したり、適切に組み替えたり、知識コンパイルしたり、バラエティ

に富んだ、広範囲の、応用向きの、効率の良い知識ベースが生成できなくてはなりません。また、知識獲得支援機能の実現には、この「システム機能」も利用されることになります。

○知識表現メディア

システムの中核となる「知識表現メディア」としては、次の２つを選ぶことになります。ヒューマンメディアの中核となる「自然言語」と、コンピュータメディアの中核となる「知識表現言語」であります。ヒューマンメディア全体については、中核となる「自然言語」から次のような段階を追って研究開発を進めることになります。

まず、第一は、「自然言語」のみを用います。他のメディアによる知識は「自然言語」による近似表現で置き換えられます。また、知識抽出が容易になるように、「制限を加えた『自然言語』の仕様」も考案することになります。

二番目、「自然言語」と「形式言語」を用います。「形式言語」としましては、知識表現言語・プログラム言語・代数式・論理式などを用います。「知識表現言語」としては、「大規模知識ベースシステム」用のものをはじめとして、代表的なものをすべて取り上げることになります。

三番目、「自然言語」として「母国語」と「諸外国語」、および「形式言語」「図形言語」「画像」「音響」などを総合的に用いる最後の段階です。ただし、「図形言語」「画像」「音響」に関するコンピュータの理解能力は、ごく限られたものに限定するのが無難であると思います。

○対象とする知識

ある分野の学術的・科学技術的知識、歴史的・科学技術論的知識、経済や法律に関する知識、産業活動や工業技術に関する知識、教育や資格試験にかかわる知識、製品・システム仕様・言語仕様・製造技術に関する知識、機関や人物にかかわる知識、社会問題や日常生活にかかわる知識などを対象にすることになります。

○研究開発課題のトピックス

多くの新しい課題がありますが、トピックス的には次の４点を挙げることに致します。

まず、第一、自然言語コンピューティング：自然言語をコンピュータシステムの情報表現・処理の中核言語に設定します。ソフトウェアからハードウェアまでを含む一般的なコンピュータシステムの中核に、自然言語を位置づけてみようという試みであります。自然言語が、情報側からシステムアーキテクチャを規定することになります。自然言語で表現された情報の作成・変換・蓄積・検索・伝達がシステムの基本機能となります。ただし、コンピュータ側からシステムアーキテクチャを規定するのはプログラム言語であります。プログラムはプログラム言語で書かれます。いわゆる「自然言語プログラミング」と混同しないでいただきたいというところです。

二番目、コーパスベース言語処理：大量の言語データを収集し、現実の言語現象に対応づけながら、

自然言語処理ソフトウェアを頑健な（ロバストな）ものに育て上げたり、新しい言語理論の研究を展開することになります。言語データに関しましては、ＥＤＲ電子化辞書によって単語レベルのものの整備は一応はなされました。次は、句・文・文章レベル、いうなれば「コーパスレベル」での電子化辞書の研究・開発が必要であります。これによって文・文章を単位とした意味処理、さらには文脈処理への本格的な取組みが始まることになります。

三番目、ケースベース知識処理：知識はほとんどが「ケース（事例）」の形で蓄積されます。「ケース」は知識現象を直載に捕捉し、再利用に適した形式であります。蓄積された大量の「ケース」を分析することによりまして、より複雑な知識、ルール化された知識などが取り出されることになります。言うなれば「大規模知識ベース」は「大規模ケースベース」であります。

四番目、大規模オントロジー：言語処理と知識処理を結びつけるのが「オントロジー」であります。通常、知識工学で言われます「オントロジー」は、どちらかといえば小語彙でありますが、ここでは大語彙のものを想定することにします。言語処理におけます「シソーラス」が知識処理の「オントロジー」と融合することになります。概念のゆれの少ない専門用語を対象に、融合化がはかられることになります。単語だけではなく、句や文も対象になります。「オントロジー」の規模の拡大や精度の向上を、学習能力によって手順を追って達成していくことになります。そして、「オントロジー」を刻々入力される知識にダイナミックに適応させたり、観点の違いに合わせダイナミックに骨組みを変化させたりする「ダイナミックオントロジー」ともいうべきものが目標であります。「大規模知識ベース」の知識構造としましても、研究テーマとして「大規模オントロジー」が焦点になります。

○研究開発環境

既存の技術を使いまして「『大規模知識ベースシテム』のプロトタイプのプロトタイプ」を作ることができます。この「プロトタイプのプロトタイプシステム」が、研究情報や研究知識の蓄積と共有化の場となり、「大規模知識ベース」の研究開発環境を形作ることになります。いろいろな研究グループのシステムがネットワークで結ばれます。さらに、各国のシステムも国際ネットワークで結ばれることになります。このように作られました「知識インフラのプロトタイプ」が、「研究開発インフラ」を形作ることになります。研究開発の成果は、この「研究開発インフラ」を徐々に革新していくのに使われます。したがって、「研究開発インフラ」は研究開発環境であると同時に、最新の成果物そのものともなります。

◎どのような協力体制が必要か

「大規模知識ベース」の研究開発を進めるにあたりまして、新しい形態の協力体制が必要であります。従来の枠組みを大きく越えるダイナミックな協力が重要であります。その要点となるところを以下に列挙いたしまして、私の話のまとめといたします。

まず、学際的な協力：人文科学・社会科学とコンピュータ科学の協力であります。人間の知識や、そ

の社会的な出現形態を研究の課題にしてきたのが人文科学・社会科学であります。これらとコンピュータ科学との協力は、知識研究の新しい学問領域を作り出すことになりましょう。

次は、業際的な、インターインダストリアルな協力であります。新聞・放送・出版・印刷などのメディア産業、さらに金融・教育などの産業は、情報や知識を扱う産業であります。そして、膨大な知識資源を保有し、知識に関する多くのノウハウを蓄積している産業であります。いうなれば、「情報のメーカー」「知識のメーカー」であります。これらの産業が「コンピュータユーザ」としてではなく、「情報・知識メーカー」としてコンピュータ産業と新しい協力関係を築くことによりまして新しい情報産業が生まれることになります。

次は、情報関連学問・技術の融合：今、情報科学技術は新しい枠組みを求めて大きく展開しようとしております。情報関連の学問や技術は、互いの壁を越え、融合し、新しい枠組みへと脱皮しなければなりません。例えば、人工知能と情報処理であります。人工知能は情報処理の特異な一分野というものではありません。情報処理全体があげて取組むべき分野であります。人工知能も軽薄なブームの時代は終わりました。コンピュータ科学にしっかりと根を降ろし、着実な努力をする時となっております。

次に情報学と情報処理学（コンピュータ科学）であります。情報学は古い歴史を持っています。そして、今、その役割がクローズアップされはじめております。しかしながら、その多くはまだ古典的な分類学の領域を出ておりません。コンピュータ科学の最新の理論や技術と融合させ、若い研究者にとって魅力のある分野へと脱皮させなくてはなりません。そして、本国際会議の二大テーマであります「知識処理」と「言語処理」であります。「大規模知識ベース」の研究の核となるのがこの２つの技術であります。両者の適切な融合が「大規模知識ベース」の鍵となります。その融合に向けての努力が本国際会議であります。

そして、最後に重要なのが国際協力であります。２１世紀の情報化社会、Chenさんがお話しされましたが、セカンドジェネレイションの情報社会に向けて、今、世界は大きく動き始めました。おそらく、情報や知識をめぐって、場合によりましては、国際的に非常に厳しい競争が始まると思います。逆にそれだからこそ、世界的な安定を求めるならば、知識の構築と共有に関する国際的な協力、この分野の先端技術の研究開発における、全面的な国際協力が非常に重要になってくると思います。少し早めですが、昼休みが短いものですから少し早めに切り上げます。以上で私の基調講演を終わらせていただきます。ありがとうございました。

2. セッションI

社会的・学際的要請

2. セッションI：社会的・学際的要請

2.1 座長挨拶

千葉大学 文学部行動科学科
助教授　土屋　俊

それでは、午後の最初のセッションを始めさせて頂きます。この種類の技術的な会議には珍しいセッションだと思いますが、大規模知識ベースという問題を考えるときには、単に技術的な面だけではなくて、その人間が行なっている社会的・経済的な活動、それからさらにその学術的・科学的な活動というものに対して、このような大規模知識ベースがどのような影響を与えるか、あるいは、そのような今までの経験の蓄積から、大規模知識ベースというものがどのようにして生まれてくるかということについて考えなければいけないだろうと思います。そういう意味で、このセッションでは、社会科学・人文科学という観点からいろいろな、様々な考察を伺うことにしたいと考えております。

2.2 講　演

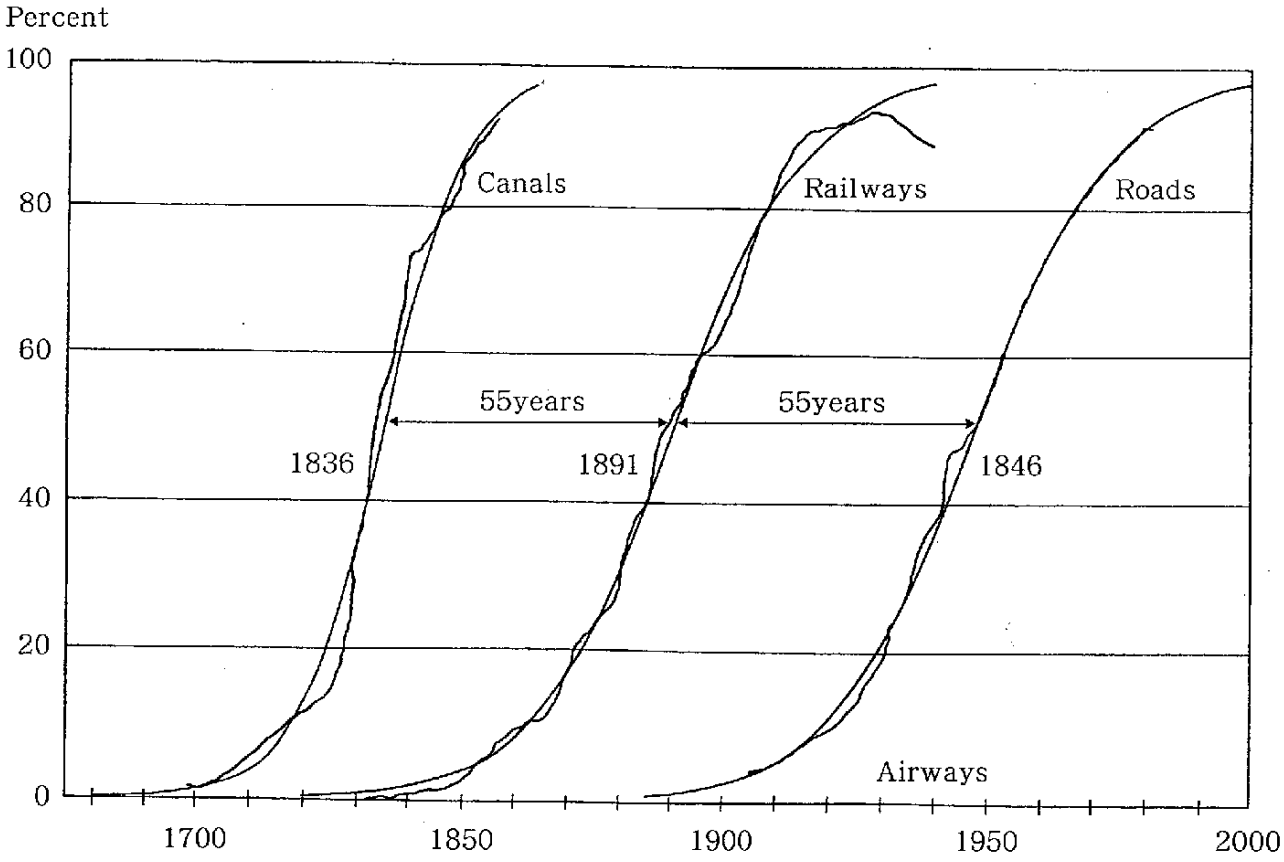
(1)「新しい経済的・社会的インフラストラクチャとしてのＫＢ＆ＫＳ」

座　長：

社会科学の分野から、今ご紹介がありましたようにスタンフォード大学の今井賢一先生にお話を伺うことに致したいと思います。テーマは大規模知識ベースが社会的・経済的インフラストラクチャとしてどのような役割を果たすかということになっております。それで、先生の簡単なご紹介をさせていただきます。一橋大学を１９６３年に卒業された後、一橋大学で経済学部の教授を長く務められて、現在、スタンフォード大学の教授及び、スタンフォード大学日本センターの研究所の所長として、研究それから教育に携われております。また、通産省の様々な審議会で様々な貢献をされていらっしゃいます。では、お願いします。

スタンフォード大学日本センター
研究所長　　今　井　賢　一

ご紹介をいただきました今井でございます。私は経済学者でありまして、私に与えられた課題は社会科学の観点から「このコンファレンスでいうＫＢ＆ＫＳ、大規模な知識ベースというのがどういう役割を持つのか」、あるいは、「なぜ今の経済社会システムでそういうものが新しいインフラストラクチャとして必要なのか」ということを申し上げてみたいと思います。ただ、「これからの経済社会システム」と申しましてもやや漠然としておりますので、最初に、私がこれからの経済社会はどういうふうに考えているかということを簡単に二点に要約して、その二つの軸から議論を申し上げてみたいと思うわけであります。私は、これからの経済社会は二つの軸を中心に形成されていくと考えます。第一はここでいう情報化社会、あるいはインフォメーション・オリエンテッド・ソサエティでありまして、「情報技術の潜在的な能力をどういうふうに生かしながら、我々の経済社会あるいは文化を作っていくか」ということが一つの軸であります。もう一つはご承知のように、「いわゆる『地球環境問題』というものにどういうふうに対処していくか」、単純にいいますと、グリーン・オリエンテッドなソサエティをどういうふうに作っていくか。それに対して、情報技術の潜在力というものが基本的に重要なわけでありまして、したがいまして、いわゆるサスティンナブルなグロース、すなわち持続可能な経済成長というものは、「情報化」および「地球環境問題の対応」と、その二つを軸に、これから形成されていくだろうと思うわけで、そういう意味から言いまして、ここでいう「大規模知識ベース」というものは、これから２１世紀たる基本的なインフラストラクチャとして本質的な役割を持つというふうに考えるわけであります。



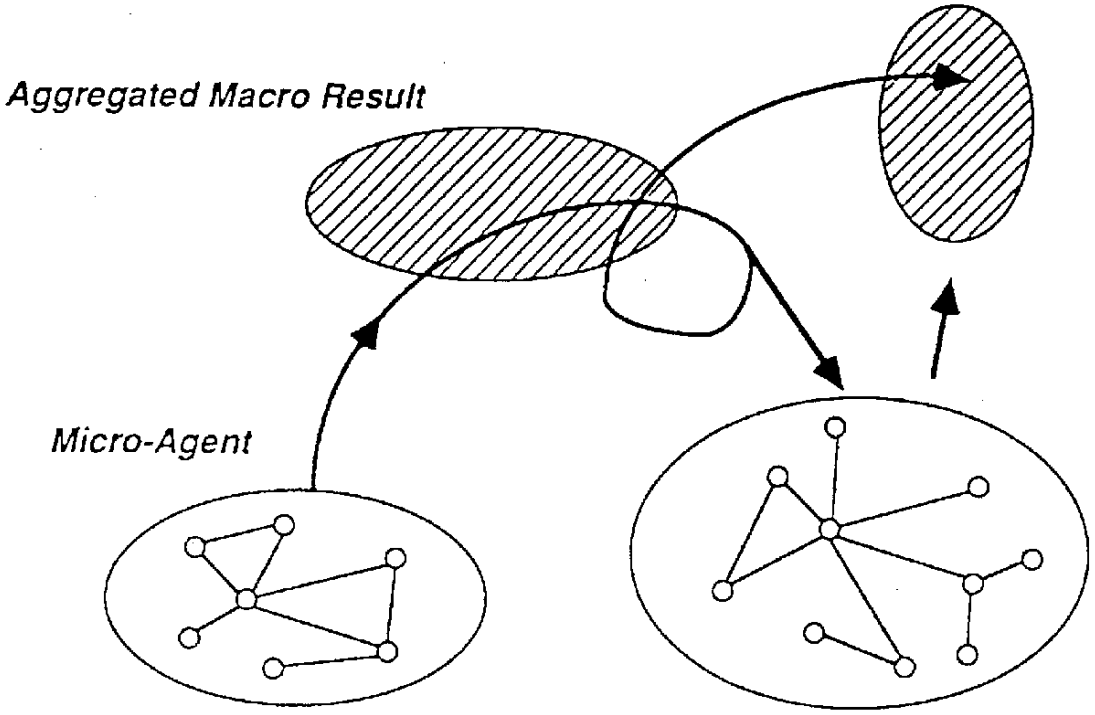
OHP－1

今までインフラストラクチャというものは、大体この図に書きましたように最初に、これは１８００年代から２０００年までのインフラストラクチャの転換でありますけど、最初に運河の時代、ヨーロッパは運河でインフラストラクチャができたわけですが、運河の時代から鉄道、道路というふうに、インフラストラクチャが、大体ピークでとると５５年位の周期、それから大体一つの－例えば鉄道のインフラストラクチャの役割が終了するのには１００年かかるわけでありまして、そういうふうに第一、第二、第三というふうにこう変化してきたわけでありまして、そして、これからその新しいインフラストラクチャが１９９０年代のはじめから形成されていく、その場合に大事なことは、それは多層なインフラストラクチャになるということでありまして、単に針金のネットワークが作られるだけではなくて、その上に産業的なインフラストラクチャもできますし、それからここで赤いので書きましたのは、その上にソフトウェアのインフラストラクチャが必要であり、ＫＢ＆ＫＳもその一つなわけであります。したがいまして、たまたま現在は、この新しいインフラストラクチャが転換する時期にあたっておりまして、今ちょうど、そういう情報通信系－ソフトウェアを含めた情報通信系－のインフラストラクチャを作ることによって、先ほど申しましたサスティンナブルな、持続可能な経済成長の基盤ができるわけでありますので、その中でＫＢ＆ＫＳ、ここで議論されるような問題というのは、本質的に重要な役割を持つだろうというふうに考えます。それで、今、二つの軸のことを申し上げたわけでありますが、その前提

として、当然のことながら「市場経済」というものを前提にしているわけです。ご承知のように「社会主義経済の計画経済」というものは行き詰まったわけでありまして、やはり「マーケットに基づいた経済」というものに、我々の社会は依存せざるを得ないわけであります。そこで、市場経済を前提としたときに、その情報という基礎にどういう問題があるか、ということから私なりに経済学者としての議論をしてみたいと思うわけであります。

といいますのは、なぜ市場経済が社会主義経済より優れたかといいますと、結局はここにいらっしゃる皆様方に、すぐに分かりやすい言葉でいえば、非常にうまい具合に情報効率の高いシステムを作った。あるいは、分散型の情報システムをうまく作ったということであります。これは私のペーパー、フルテキストはちょっと遅れて後ろの方に入っていて恐縮なんですが、イグザンプルで書いてありますが、例えば、色々なところに、世界に、マーケットのセンターがありまして、例えばシカゴでいえば、シカゴで毎日のように穀物の値段が決まっているわけでありますが、その背後には膨大な情報があるわけであります。それぞれの専門家がサテライトのセンサーを使ったり、あるいはそれぞれ専門家を使って、気象条件であるとかあるいは作付けの情報であるとか、そういうものを全部、そういう情報を集約して、シカゴでは毎日のように、「今日の小麦の値段はいくらである」ということ、それが先物を含めて決まるわけでありまして、これは非常に能率のいい、インフォメーショナルエフィシェントなシステムでありまして、いうならば、プライスメカニズムというのは、そういう非常に単純な形で分散型の情報を集約する、一つのプライスというインジケータに集約することに成功し、そして、そういう意味で「分散型の情報システム」をうまく作ったということ、これが基本的な成功の原因なわけであります。ところが、市場経済にはまだ、未解決の問題がたくさんあるわけでありますが、情報論的にいうとその一つは、そのペーパーには引用しておいたのですが、例えば、こういう市場経済の情報論的な基礎でノーベル賞をもらいました、フリードリッヒ・ハイエクという亡くなられた教授がいるわけでありますが、彼がこういうことを言っているわけです。「市場経済には、まだ情報的に未解決な問題がある」。それはそれぞれの意志決定をする主体が、マイクロな主体がいろいろあって、企業だとか消費者だとかいうのが、それぞれの情報に基づいて行動するわけでありますが、その人達が適切な行動をするためには、その結果として合成された、アグリゲイト（aggregate）された、マクロの結果がどうなるかということをその人に知らせてやらなければいけない、単純にいえば、「『全体はどうなっているんだ』ということを知らせてやらなければいけないということが残された課題なのだ」。したがいまして、それぞれが行動した結果、例えば、先ほどの例で言えば、シカゴのそれぞれのその穀物の業者がいて、それがシカゴで米の値段がいくらというふうに、マクロのインジケータとして集約されるわけでありまして、それは　単純なプライスメカニズムがありますが、今の問題はもう少し複雑でありまして、そういうこと全体の関係を、もう少し複雑な全体の状況を、それぞれの人に示してやる、例えば、「環境問題はどうなっているんだ」ということをそれぞれの企業がやった結果、「こういう状態になっている」ということを意志決定者に伝えて、そして、その人達が次の行動を起こして、さらにそれが次のマクロの結果を生み出す。その間に当然のことながら、個人は全体を見ながら行動するわけでありますから、そこにリフレクティブ

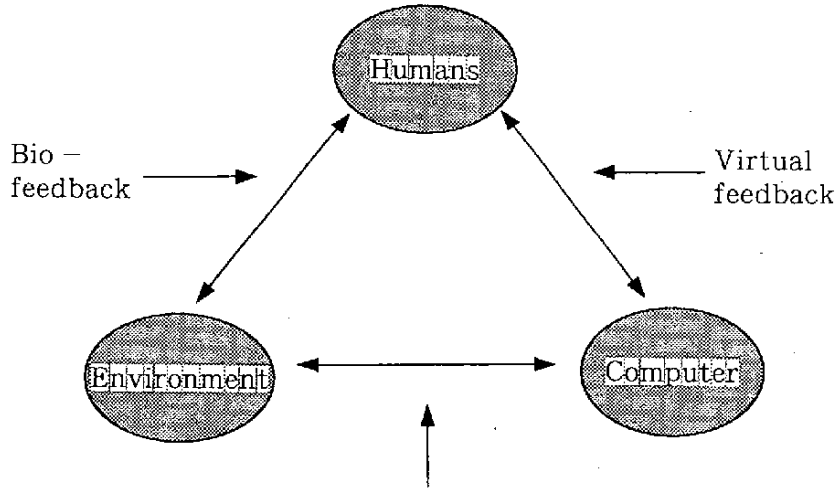
な、ある意味で反省的な行動といいますか－これは道徳的意味を含めた反省ではないのですが－そういう、全体をみながら行動をする、あるいは、情報に基づいてリアクションする、そういうことが起こるわけでありまして、地球環境問題なんかを解決するにも、結局はこの「情報ループ」をどう作るかということに、私は重要な焦点があると思っているわけであります。いずれにせよ、そういう、私の言葉でいうと「マイクロ－マクロの情報のループを作る」という仕事が残されているわけであります。これができませんと、市場経済というのはやはり行き詰まるんではないかというふうに思います。それが、な



OHP－2：MICRO－MACRO LOOP

ぜかということなんでありますが、単純な経済の場合は、「品質が良くて安いものが売れる」、そして、あるいは、企業でいえば、「品質が良くて安い材料がくればそれに取り替える」。それが、サブスティテューション、代替でありまして、市場経済というのはこれには非常にうまく効くんですね。安いパソコンでいい品質のものが出れば、必ずそっちに皆が買い換えるというわけでありまして。したがって、そのプライスというインジケータにあらゆる情報を集約するということは、非常にいい情報集約のやり方であったわけであります。しかしながら、申すまでもなく、現在の財サービスというものは、非常に複雑なシステムになっているわけでありますから、単なるサブスティテューションではなくてコンプリメンタリティ、経済の言葉でいうと、こなれない言葉かもしれませんけれど、コンプリメンタリティ、補完関係を作っていくということがシステムを作ることの基本でありますから、「こういうハードには

こういうソフトがいる。さらにそういうサービスがいる。」と、こういう補完関係を作って、補完関係に基づいてシステムを作るということのためには、単純なプライスメカニズム、価格に単に情報を集約するだけではなくて、その間の補完的な関係を示す情報のループが必要なわけでありまして、これが先程申しましたように「マイクロ－マクロ」情報の環、「全体をつなぐループ」というものがどうしても必要である。例えば、皆様方がいろいろな仕事をされる。例えば企業のグループで仕事をされる。すると全体はどうなっているんだろう。それは、いろいろな補完関係を含んでいるわけでありますから、そういう現代の経済システムのもとでは、やはり新しい情報ループが必要なわけであります。それをもう少し具体的に申しますと、これは、ちょっとわき道にそれるかもしれないんですが、私はその中の情報ループで重要なことは、まず、人間がいて、コンピュータがあって、環境があるという３つのエージェントを考えた時に、「人間とコンピュータの間にフィードバックのループを作る」。これは、やはり最近のバーチャルリアリティーとかというものはまさに「お互いの環境を伝えあえる」ということでありますから、そこで、この情報の理解が進む、先程、渕先生のお話で「シェアリングというのが重要だ」というお話があったんですが、私も、まったくその通りだと思います。そういうループをまず作る。それから、「環境とコンピュータの間には、エコロジカルなフィードバックを作る」ということで、わざわ

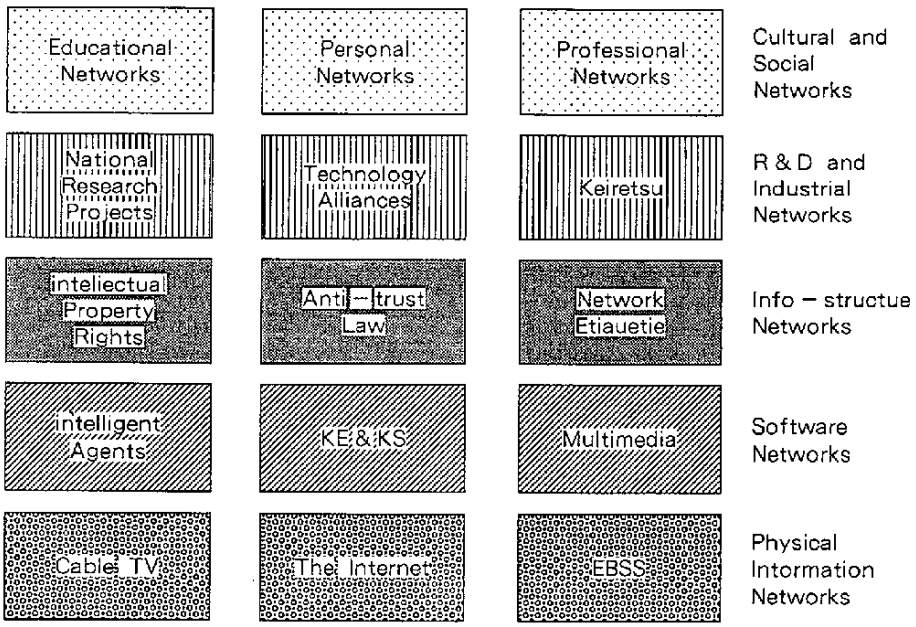


ＯＨＰ－３：Eco－Feedback

ざセンサーと書いておきましたけど、センサーをいろいろなところに置いてですね、そして「今の二酸化炭素がどういうふうに、各地点でどうなっているか」ということが即座に集計されて、全体の状況がわかるようになりつつあるわけでありますので、そういう情報ループを作る。ということが、私が申し上げている「マイクロ－マクロ情報ループ」ということの意味であります。次は、今のＫＢ＆ＫＳということですが、ここで申し上げたいことはこういうことなんです。「市場経済」というのはやはりマーケットのプロセスなわけです。つまり、プロセスとして理解するということが、市場経済の良い点の根

本だと私は思っているわけでありまして、それは一時点でプライスが決まって、そこで均衡ができるということも大事でありますけれど、マーケットというのはプロセスとして動いていくわけでありまして、最初に、例えば、ある環境があると、その環境についてー環境というのは、もう少し、地球環境だけではなくて、広い意味での我々の世界、我々のその世界に関する環境でありますけれどーそれについていろいろな情報があると。例えば、それが一期前に、例えばそのＫＢ＆ＫＳとして知識ベースができている。そうするとそれに基づいて、我々のこのニューロンは反応しだして、さらにその反応の仕方は、アプリオリーな我々のナレッジに基づいて行動していくわけでありまして、そしていろいろな判断をして、そしてそれがさらに事後的な知識として作られるわけであり、そしてそれが、そのＴ時点の（Ｔ－１）のＫＢ＆ＫＳに対応して行動した結果が、結果として今度は（Ｔ）期のＫＢ＆ＫＳになるわけでありまして、さらにそれが（Ｔ＋１）期のものに、「プロセス」として累積的に蓄積されていくと、こういう「プロセス」を考えることが非常に重要でありまして、そしてなぜ、「プロセス」ということを強調するかといいますと、現在、情報のインプットは、昔のように必要に応じて、例えば「データベースをつくるために何かインプットをしましょう」と、そういうことではなくて、それぞれのこのマーケットの「プロセス」、つまり、ビジネスが実際に行われる「プロセス」の中で、自動的にデータというものは蓄積されるようになってきているわけです。それから研究者の間でも、論文をどんどん書いて、それはもう完全に、わざわざいちいち入れなくても、今日のコンファレンスのペーパープロシーディングなんかも、もうそこへ蓄積されていくわけでありますから、「今のプロシーディングができれば、この次のプロシーディングはそれに新しいものが加わってできてくる」と、こういう「プロセス」になっていくわけでありまして、経済活動の「プロセス」の、こういう実行の「プロセス」と同時に、ナレッジベースができていくと、こういう時代になるのではないだろうかというふうに思うわけであります。しかし、それでは、すべて問題が解決するのかということではありませんで、おそらく今日のコンファレンスは、「そのためにどういう分析の道具があるか」ということを議論するということであると思います。有名な経済学者でジョセフ・シュンペーターという、日本では割合いろいろな翻訳が出たりして有名な経済学者で、ウィーン生まれの経済学者でありますが、彼はそのイノベーションということを機軸にした経済理論を作りまして、非常にうまいことを言っているわけであります。それはイノベーションというのは、「馬車を何台繋げても汽車にはならないんだ」と、だから、「こういう発明があった」「こういう発見があった」あるいは「いろいろ新しいものがでてきた」。だけど、それはいくつ繋げたって、それは汽車にはならないんだと。汽車にするためには、やはり新しいプリンシプルなり、新しいそのまとめ方が必要なんで、したがって、例えば汽車で言えば、昔の馬車と、客車の方は同じなんだけれど、最初にくっついているそのエンジンが違うわけであります。そういう意味で、既に個別に情報ループというものが存在しているわけでありますけれど、それをどういうふうに、今のように「汽車にする」にはどうしたらいいかということでありまして、これは私、専門ではありませんが、まず私ども経済学者からみて、こういう最近の動きの中で注目したい点の一つは、まあこれはもう今日ずっとお話があり、あるいは横井さんがおっしゃった事で、「自然言語からコンピュータのアーキテクチャの方に迫っていく」と、

「自然言語の豊かさを利用する」、これはまあ、私ども画期的なことであり、またそれをロバストな形で利用されるということは、私ども経済学者から見ると、それは「機械と人間との分業を可能にする」わけでありますので、完全にオートメーションにする、完全に人工知能を作ってしまったら、今の失業は益々増えるわけでありますので、そういう意味で、横井さん達のグループの研究というのは非常に注目している、というのが”A”であります。それからもう一つ”B”は、「ケースベースでのアプローチを取る」ということで、これも私どもの世界では、アメリカでビジネススクールとかいうのは全部ケースメソッドでありまして、先程申しました様に、ビジネスの「プロセス」の中でいろいろな知識が蓄積されてきているわけであります。しかしながら、ケースベースメソッドの弱点は、それを一般的な推論にするのが、一般的なそこからなんらかの意味なりインプリケーションを導きだすところが、まったく個人のノウハウあるいは直感に基づいているわけでありまして、それをもう少し理論的にやるとすれば、おそらく「アブダクション」ということになる。帰納でも演繹でもなく、いろいろな実証を矛盾無く解釈していく方法論を捜す。このことについてはいろいろな議論があったわけでありますが、最近、私も素人でありますけれど、いろいろな論文等を読んでみますと、こういうことをコンピュータのプログラムで書いていこうということで、非常に勇気づけられる点があります。もう一つはセンサーのフュージョンの研究、これは日本でも非常にやられているわけでありますが、やはり人間の判断というのは、人間の行動というのは、いろいろな五感に基づいて、単にこのテキストを解釈するだけではなくて、いろいろな五感に基づいて判断をするわけでありますので、そういうことが本格的な研究として始まっているということは、誠に勇気づけられる点であります。そうしてもう一つ、これはまったくの試論であり、多少未熟なものなわけでありますけれど、先ほどの「アブダクション」と同じように、いろいろな情報なりがあった時に、「それをどういうふうに編集していくか」という技術、これも、文化系、社会科学系の連中は「雑誌を編集する」とか、あるいは「論文を書く」時に、常にやっていることなんですが、それをもう少し、いろいろな意見なり情報があった時に、「それをどう区別して、それをどう方向づける」とか、あるいは「少し論争をしかけてエッセンスをもうちょっと引っ張り出す」、あるいは「インターフェースを作る」とかあるいは「それをさらに方向付けて編集して、一つの結果を導き出していく」と、こういう技術が必要なわけでありまして、こういうことも、社会科学系あるいは文化系の中でいくつかの研究をやり始めているということでありますので、まさに情報のインプット・その編集ということに関しまして、インターディシプリンな研究が、徐々に、最近かなり急速に立ち上がっているのではないかというふうに思います。最後にもう一つ、経済学者として注目したい点は、「問題は価値ということに関わる」。価値論というのは根本問題でありますから、常に議論があるわけでありますが、単純な現代の経済理論というのは「値段が価値を決める」と、要するに高いものはいいものであり、企業で言えば「株が高い」ということは企業の価値を決めているということなわけで、そういう複雑なものを、例えば「『企業の価値』というようなものを、『株価』と言うようなプライスのインジケータにしていく」ということは非常に、まさにエフィシェントなことなんであります。しかしそこにやっぱり、今限界があるわけでありまして、やはり人間が行動する時、我々が行動する時に、バリューとい



OHP－4

うものがどういう影響を受けているかというと、やはり、まさにプライスも一つの要素でありますけれど、その他の情報、あるいは「将来がどうなる」というような問題、それから先程申しましたように「地球環境がどうなる」、「そういうマクロの合成された結果がどうなる」、こういうことがバリューに影響するわけであります。勿論プライスも重要であります。従いまして、今申しましたような情報のループができ、そしてそれはプロセスとして、さらに大規模な知識ベースとして集積されていって、そのことがやはり人々の行動を、次の行動を動かしていくと、こういう経済のビジョンを考えないと、これからの持続可能な経済発展というのは不可能なんではないかというふうに思いますので、そういう意味から考えますと、我々のこの社会というのは、私流に言いますと「マルチレイヤのネットワーク」として形成されてくるんではないだろうかと。一番下にフィジカルな情報ネットワークがございます。例えばインターネットであるとか、あるいはＣＡＴＶのネットワークであるとかというのができます。その上にソフトウェアのネットワークがあって、これは今のＫＢ＆ＫＳなんていうのがそれぞれの中核にあって、それぞれのところでいろいろなシェアリングが行われて、そのソフトウェアのネットワークができているわけです。その上に、これはある人の言葉で「インフラストラクチャ」に対して「インフォストラクチャ」というのは、これは情報化社会のルールを決める部分のインフラでありまして、例えば知的所有権であるとか、独禁法であるとか、あるいはネットワークを利用していく上でのルール、ここはエチケットとしていますが、ルールですね。その上に研究Ｒ＆Ｄ、及び産業のネットワークが今のビジネス活動として形成されていく。その上に、文化的・社会的なネットワーク、教育のネットワークであるとか、パーソナルネットワークであるとか、あるいはプロフェッショナルネットワーク、こういうふうなネットワークが形成されていくわけでありまして、２１世紀へのインフラストラクチャというの

は、このインターネットを中心に非常に世界中が結ばれつつあるわけありますが、やはり基本的にこの赤い部分ですね。繰り返しになりますが、やはりそのためのソフトウェア、それからそれを動かしていく法律とかルールですね。そういうものを形成していく時代に入ってきて、この１０年くらいの間に、まさにそういうことをやる転換点になる。そういうことを実際に進め得るまさにいいターニングポイント、ちょうどいい時期にあるわけでありますし、またそれを作らなければ、我々の将来の経済社会に対する展望は開けないんではないかというふうに考えまして、この大規模知識ベースに関する研究なりプロジェクトというのが進んでいくことを、一介の一経済学者として心から願って、私のお話を終わらせて頂きます。どうも有り難うございました。

座長：

どうも有り難うございました。特に質疑という時間は取ってありません。可能であればセッションの最後に、もしご意見ある方あれば伺いたいと思います。

(2)「人間の知識の本性について」

座　長：

引き続いて藤澤令夫先生にお話を頂きたいと思います。藤澤先生は京都大学を１９５１年に卒業された後でギリシァ哲学を中心として研究を続けられ、戦後の日本の哲学の振興に非常に力を尽くされました。退官された後は現在、京都国立博物館の館長をされていらっしゃいます。また、日本哲学会の会長、西洋古典学会の会長なども務められていらっしゃいます。ではよろしくお願いいたします。タイトルは”人間の知識の本性とはなにか”ということです。

京都国立博物館館長　京都大学名誉教授　　藤　澤　令　夫

「大規模知識ベース」に関するいろいろな報告書がございますが、それを拝見していますと、非常に強調されている点は、「大規模知識ベース」ということで考えておられることが、「ただの情報の集積物ではない。そうではなくて、知識そのものの技術と取り組む。そのために意味、ミーニングの領域に踏み込む。そういうことによって知識そのものの技術と取り組むものである。」ということが強調されておりまして、この点に私は感応いたしまして、大変新鮮な力強い意欲ではないか、野心ではないかと思われまして、そこで、以下私の立場から、それでは「知識そのものの技術」といわれる場合の、その「知識」というのが、本当のところ人間の求める「知識」というものは、どういう性格、どういう本性のものであるかということをここで申し上げまして、そのことによって一つの視点を提供することができれば、というくらいの気持ちでおります。徹頭徹尾なるべく基本的なことをお話したいと考えておりまして、しかもプラトンとかアリストテレスとかに言及いたしますが、これは私の専攻が、先程ご紹介のようにギリシァ哲学が中心ですので、そのためもございますけど、西洋の場合の一番淵源にあるところだから、こういう知識とか学問の基礎的な問題を扱う場合にはふさわしいんではないかという気持ちもございます。

それで、私自身の研究におきまして、実際に主としてやっていることは、どういうことをやっているかというと結局、本を読んでいるわけなんです。古代ギリシァの哲学者の原典は当然でございますけど、それから始まって以後、今日までに蓄積されたものすごい量の注釈書とか研究書がございます。ギリシァ語・ラテン語・英・独・仏・各国語で書かれたそういう文献の蓄積がございます。結局それを読んで学ぶという作業に、実際のところ一番時間を取られている。これは、別に私共だけではなくて、人文系の学問の場合には、多かれ少なかれそういう書物の伝統の存在が、かなり本質的な重要性を持つだろうと考えられます。

そこで、まずこのことを手がかりといたしまして、そういう書物とか文献とか、これは結局情報の集積物でございますね、そういう書物とか文献に書かれてある「情報」は、そのままとりもなおさず学問の求める「知識」であるかという問いを、作業仮説的な問いとして、まず問うてみたいと思います。こ

れは一つは、電子技術が発達してますます便利になる。情報化社会になる。そのことによって、本当にそれがそのまま本当の意味での知識の前進につながるのだろうかというような気持ちからでもございますけれども。それで、プラトンを引き合いに出すわけですが、プラトンは、古代の哲学者の中では、自分の書いた著作というものが今日まで完全な形で残っている、唯一例外の哲学者でございますけれども－アリストテレスの場合は、講義の草稿みたいなものですからちょっと違うんですね－そのプラトンが書物ということについて、「パイドロス」という著作の中でどういうふうに認定しているかということを見てみますと、第一点は、人が書物から得る情報というのは、決してそのままでは本当の知識ではない。その外見に過ぎない。知識ではなくて知識の外見に過ぎない。人は書物のおかげで、自ら自分自身で探求することなしに、いわゆる情報通となるために、プラトンの言葉の引用ですが、「多くの場合、本当は無知であるのに博識家であると思われるようになって、知者とはならずに知者であるという自惚れだけが発達して、つき合いにくい人間となるだろう」と。確かにあまりに情報通とかあるいは、文化系でも、あまりコンピュータにのめり込んでばかりいると、そういう人は大体付き合いにくい人間が多いんではないかと思いますけれども。それが第一点。要するに書物から得られる情報というのは、そのまま知識ではないということです。

プラトンの認定の第二点として、直接引用しますが「書物というものは、あたかも何事かを知って語っている様に見えるけれども、そのどれかについて、『本当に勉強したい、学びたい』と欲して質問すると、いつもただ一つの同じ合図を送ってよこすだけだ」と。そして、「言葉というものはいったん書き物にされると、それを理解できる人々のところであろうと、まったく不適当な人々のところであろうとお構いなしに、点々と巡り歩く」と。

第三点として、それでは書物の果たし得る積極的な役割は何かというと、それは結局、「そこで扱われている事柄について知っている人に対して、それを思い出させるということ以上ではない」。だけど、これは裏を返していいますと、それ故に、「書物を書くということは自分自身のために、また同じ足跡をたどって探求の道を進むすべての人々のために、覚え書きを蓄える」という、積極的な意味を持つということになります。

プラトンという人が「情報の入れ物としての書物」について認定した以上の点が、わりと人間にとっての知識の本性を考える上で手がかりになる、重要ではないかと思うわけですが、まず、その積極的な役割として認められている書物の機能は、「覚え書きを蓄える」ということ、つまり言い替えれば情報を伝え残していくということですね。実際これが、学問における、文献の伝統というものを形作ってきたわけなんですね。こういう覚え書きの機能というものが、文字による情報というものが、電子化されてコンピュータに載せられるということによって、飛躍的に増大あるいは強化されつつあります。そのことによって、我々が膨大な文献の蓄積と取り組むための労力は非常に軽減されて、今後もますますそうなっていくだろうと思います。

しかし、プラトンは今見ましたように、そういう意味での「情報」がそのままとりもなおさず知識であるということを、明確に否定しております。例えば百科辞典に載っている内容を全部暗記している人

がいたとしましても、その人は諸事万端についての知者であると本当は言えないんだと、本当の意味での知識を持っている者ではないと、そういうことなんですね。私たちの今日の生活を振り返って見ますと、日常的な場面で何らかの小さい目的、小目的－さし当たってどこへ旅行するのが一番いいかとか－、そういう小さい目的があって、そのために集める情報、その情報をその事柄に関する「知識」と呼んでいる傾向が、習慣がありまして、そういう習慣が拡大されて、もっと重要な文脈においても、「情報」という言葉と「知識」という言葉が、何となく相互に置換可能な、置き換えることができるような同義語のように扱われる風潮があるように思われますので、それだけに「なぜ、どの点で、書物に書かれた情報、あるいはそれが電子化された従来型の知識ベースシステムの情報は、そのまま知識で有り得ないのか」ということを、もうちょっと立ち入って考えてみたいと思うわけです。

まずプラトンの第一の認定というのは、「自分で探求しないでいろいろなことを書物から習う、けれどもそれは本当の知識ではない」と言っていたことですが、知識が成立するために何が不可欠の条件であるとプラトンやソクラテスは考えたかといいますと、「無知の知」ということなんですね。自分が知らないということを知っているということ。自分が関心を持っている事柄を知らないと自覚して始めて、それについての本当の知識へ向けて、苦労がかえってプラスに働くような、そういう欲求が発動する。それで、与えられた情報によって始めからわかったつもりになっていますと、そこで停滞して安定してしまう他はない。知識というものは「自分で第一歩から求めて成立するものである」のに対して、いわゆる情報というものは、そういう不可欠のプロセスを抜きにして「最初から他から与えられる」という基本的性格を持っているように思います。私の知っている、最近非常に活躍しているルポライターというかノン・フィクション作家がおりますけれども、その人がいつか書いていたのを見ると、ある人について書こうとする時に、（ルポライターですから）さぞその人に関する情報を集めて、そこから出発するんだろうと思われるけれど、「自分は一切そういうことをしないで、自分がその人についてまったく何も知らないというところから出発する」ということが書いてありましたけれども、そういうようなことなんでしょう。自分の関心にそってのみ、その人についての取材を深めていくということですね。それが知識一般についても、非常に重要なことである。

それから第二の点は、「書物というのは質問しても答えてくれない」ということですが、これは今の点を補強してくれるようにも思われます。つまり、「学びたいと欲しても、質問しても、いつも同じ合図を送ってよこすだけだ」ということは、こういう書物の形での「情報」というものは、それ自体としては、その「情報」の受け手に－「情報」の受け手というのは知識の求め手ですが－とっての関心、そしてその関心の中にある当人にとっての「意味」と「価値」からは、完全に独立別個の、閉ざされた固定的なシステムを成しているものである、ということに他ならないだろうと思うんですね。逆にいいますと「知識」というのは、求め手自身の関心に支えられて、当人にとっての意味と価値と一体的である、ということになるでしょう。一般にこの種の「情報」は、受け手の関心とは無関係に、「他によってすでに一定の意味付けを与えられた情報」として与えられるわけですから。インフォメーションという言葉の元にあるラテン語のインフォルマーレというのは、フォルマは「形」という意味ですから、「形を

与える」「秩序を与える」という意味で、「インストラクトする」という意味でございますけども、そういう意味を持っているインフォメーションという言葉にも関わらず、むやみに過剰な情報を与えられると、秩序ではなくて、逆にカオス、無秩序と混とんをもたらすことを我々は経験しておりますが、それは今言ったことのためであると思われます。つまり本当の自分自身の関心、あるいはその人にとっての「意味」と「価値」からは独立の、閉ざされた体系であるということ。それに対して「知識」というものには本性上、「過剰」だとか、いわゆる「情報過多」という場合の「過多」とかいうことはありえないと思うわけです。

それで、なぜそういうことが出てくるかということを、ここで「人間が環境としての世界の中で生きて行動している」という、これが一番ベーシックな基本的な事実だと思うんですけども、そこへ立ち返って考えてみたいと思う。つまり「人間が環境としての世界の中で生きて行動している」ということはどういうことかというと、環境世界、あるいはその中の当面の対象ですが、そういう環境世界のあり方とか、あるいは状況なり構造なりを認知して、まさにその認知・知見によって、「どういうふうにそれに対して反応したらいいか、働きかけたらいいか、対応したらいいか」ということの指示を得ながら生きている、行動しているということでございますね。つまり、もともと「環境あるいはその中の対象のあり方を知る」ということは、そのままとりもなおさず「その環境、対象に対していかに対処すべきかを知る」ということに他ならない、両方切り離せないということ。つまり、いかにあるか、英語の”ｉｓ”に関わる知と、それからいかに対応すべきかという”ｏｕｇｈｔ”に関わる知とは、本来一体的であるより他はない。そういう全一的な知が、ほんとうに人間が求める自然本来の知識であるというふうに思われます。だからさっき言いましたように、「知識がその求め手であるところの人間自身にとっての意味と価値と一体的である」という所以もそこにある。「”ｉｓ”から”ｏｕｇｈｔ”を引き出せない」とか、いろんなことがあたかも自明のように言われてきましたけれども、それは嘘だと思うんですね、基本的な状況に立ち返ってみると。

ですからアリストテレスという人が人間の「知」のあり方を非常に厳格に区別しまして、いわゆる観想的な、あるいは理論的な「知」と、それから人間の生き方とか行動や働きかけに関わる実践的な「知」とを、厳格に区別いたしますが、これは本来からいえば非常に疑問です。「知」を表す言葉も、それに対応して非常に厳格に区別してしまうわけなんですね。私たち自身も、何となく「知識」というものと「知恵」というものは区別する傾向がございますけども、しかしそういう区別は結局、人工的なというか、私利的なフィクションであって、自然本来には存在しないと言わなければならない。その点プラトンの方は、いわゆるソフィアとかエピステーメとかフロネシスとかテクネーとかいった言葉を、まったく端的に「知」を表わす同義語として一体的にとらえておりますが、この方が人間の先程申しましたような基本的あるいは原初的な事実に合っているように私には思われます。ギリシァ語としてもそれがホーマー以来の使い方だったわけです。

で、学問というのは結局そういうところに根がある。人文系に限らない、あらゆるサイエンスというものはそういうところに根を持っている。さっきの基本的な事実の中に根を持っている。人間の学問的

営為というものは、すべて先程述べましたような意味での「全一的な知識」を求める努力を、そのまま延ばして発展させたところに成立する営みに他ならないと思われます。勿論、環境としての世界・人間－自然だけではなくて人間社会も含めてですけども－、それには当面の関心と注意がそこへ向けられることになる、いろいろなアスペクトというのがございます。ですからそういうそれぞれのアスペクトに応じて、それぞれの個別の学問分野というのが、成立することになります。

そして、知識の有効な蓄積と推進のためには、どうしても、さし当たって自分が責任を持って詳しく調べることのできる小さい範囲に、できるだけ自分の分担領域を限定いたしまして、他には一切わき目を振らない、そこへ全注意を集中するのが得策であると思われるところから、学問の歴史の中では、この「領域の専門分化・細分化」ということが進行する傾向は避けられませんでした。そういうふうにして細分化された、個別の学問あるいは科学には、申しましたような「事実認識」と実践的な「価値判断」を総合するような全一的な知識の追求という、そういう元々の性格はかなり希薄にならざるを得ないだろうと思います。そこから、「科学というものは人間的な価値を一切排除して、冷静に知識のための知識を追求するべきものである」と、そういう常識みたいなものができちゃったという面があると思います。実際、西洋におきましては、１９世紀の半ば頃に、特にこれは顕著になった学問の姿でございました。

けれども、人間の求める知識ですから、それはどれほどいわゆる価値中立的な、バリューニュートラルな、知識自体のための知識の追求、純粋科学であることを標榜したといたしましても、しかし人間の知識としてのその素性というものを完全に振り切るということは有り得ない。本性上かならず「べき」、”ｏｕｇｈｔ”に対応する実践的価値的な連関を、濃い薄い、強い弱いの様々な差異はありましても、潜在的にせよ内包しておりまして、その発展は、いつかは何らかの倫理的価値問題に直面することが避けられないと思うんですね。それから、おおもとの根であった「全一的な知識」への志向というものを分け持っているということで、そこへいつか収斂されることを求めているということですね。現在では、いろいろな科学や学問の先端部分に、そのことを告げる徴候が出てきつつあるように思われる。だいたい計算機工学というものが「知識そのものの技術」なんてことを言い出して、「人間の認識機能とはいかなるものか」ということを真剣に問うてるわけですから。「人間の認識機能とはどういうものか」と問うことは、ほとんど「人間とは何か」を問うにもう一歩のところなんですね。計算機工学というような先端的なところがそういう問いを問うようになっているということは、そういうことの一つの徴候ではないかと思われます。つまり、元々の全一的な知識に収斂されたがっているということが可能的にあるということです。

そういうふうにして、さっきプラトンの発言に関連して考えられた、「書物に書かれた情報と本当の知識との相違点」とか、あるいは「知識であるための条件」とか「特色」というのは、そういう知識のそもそもの本性、素性を顧みるならば、いずれもそこから由来していることだとして理解できるだろうと思います。そうであるとすると、この大規模知識ベースとの関係がどういうことになるかということを基本的に考えてみたいと思いますけど、おそらく一番基本的な問題は、そういうふうに「知識という

ものが、求め手である人間自身にとっての意味や価値と切り離せない、本来、一体的である」ということ。意味とか価値とかいうのは、求め手の関心のあり方によってどういう方向にでも、どこまでも伸びていく。原理上、無限にといってもいいでしょうけども、相互含意、ミューチュアル・インプリケーションの網を通じて、どんどん広がっていく可能性を持っております。つまり、知識というのは本質的に「開かれた構造」を持っているということですね。オープン・テクスチャーである。そういうふうにオープン・テクスチャーを持っているとするならば、それを「閉じられて完結した有限のシステム」へと移すことは、これは勿論原理上不可能でございます。したがって、大規模知識ベースとか知識アーカイブとか、それが言われるように、単なる「大量情報の入れ物」ではなくて、「意味をとらえるものとしての知識」を扱う、あるいは「的確な意味処理に基づく情報処理を志向する」ということになりますと、そういう意味連関、あるいは意味の相互含意、ミューチュアル・インプリケーションということの本質的なオープン構造を、何らかの仕方で、少なくとも近似的に再現しなければならないだろうということになってまいります。

あるいは、もうちょっと慎ましやかというか控えめに考えますと、情報の集積を文字どおり、極めて大きな規模、ヴェリー・ラージ・スケールなものとすることによって、情報集積がそれ自体で持っている「覚え書き」の機能－さきに述べましたような覚え書きの機能－を、まさに飛躍的に強化拡大することによって、覚え書きであることから一歩進めてですね、もはや覚え書きに留まらず、その潜在的な意味連関あるいは価値連関というものが、顕在的・明示的イクスプリシットになることを積極的に促す、いわば「触媒」ですね。「触媒」たらしめること、それによって新たな知識の形成を支援する、ということは考えられるだろうと思います。このことは、一応まとまった体系的な知見というものを形作っているこれまでの学問分野の範囲内では、「当面必要な新しい知識の獲得を支援する」という仕方で十分実現可能だろうと思います。ですけれども、その場合も、「当面必要な知識」と簡単にいいますけども、「当面必要な知識」とはいってもその一番基層におきましては、知識本来のオープン構造を抜けきるということは絶対できないということを忘れてはならないだろうということです。いろいろな場面で「学問分野の再構成」「学問分野の再統合」ということが要請されている今日の状況の中では、特にそうだろうと思います。決してそのオープン構造を忘れてはならない。

最後に一つだけ、先程今井先生も触れられましたけども、この大規模知識ベースは、「自然言語」というものを、コンピュータシステムとしても、知識表現メディアとして採用するということを方針としていますので、それは戦略として非常に有効であるということは直感的に納得できますけれども、その戦略と直接どう関係するかは別といたしまして、そもそも「自然言語」というものの最も基本的なあり方はどういうものであるかということを、一言申しておきたい。何を念頭においてそう言っているかというと、いわゆる「主語－述語構造」のことなんです。自然言語がとる一番標準的あるいは普遍的なパターンは「主語－述語型」の知識記述方式だと思われている面が多い。個別的なあるものを、まず、主語として独立に主題的に立てておいて、しかる後にその主語として立てられたものがどういう普遍的な属性を持っているか、つまり、「何の類・何の種に分類されるか」とか「どういう性質なり様態なりを

持っているか」と、そういうことを述べる、そういうタイプの方式、この方式は今申しました手続きそのものが示しておりますように、おのずから、述語の方は主語あっての述語、属性は個物があってこその属性、個物の方はそれだけで独立に存在し得るけれども、その属性の方、これは黒いとかこれは甘いとかいう、そういう属性の方は、主語があって始めて、主語として立てられた実体に依存して始めて、存在し得るという見方と直通しているわけです。これがアリストテレスによって史上初めて明確な形で提示された「主語－述語」イコール「実体－属性」の記述方式。これでいきますと「意味」とか「価値」とか、先程重要な論点として申しましたものは、みんな述語の方に入ってしまうことによって、第二次的な存在ということになります。

ですけども、この「主語－述語」方式というのは、確かに日常言語の常識的なパターンではありますけども、「自然言語」といった場合の一番基本的なあり方であるかどうかは非常に疑わしい。あらゆる民族の言語がこんなパターンを基本としているわけでは決してない。サピア・フォアフの言語相対理論の中に詳しくデータが出ています。「主語－述語」と結びつく「実体と属性」という事物の把握方式、これが決して世界のありのままの見方であるとは思われません。「世界のあり方を知る」と言うことは、先程申しましたように「意味と価値を知る」ことに他ならなかったわけですから、決してその「意味と価値」というものは、「優先的に実体というものが存在して、それにディペンドしてはじめて現れる」というような、二次的なものではないはずなんです。世界そのものが始めから意味と価値の相互連関ネットワークとして存在しているといって過言ではない。だから「主語」に対応させた「実体」というのは、やはり一つの虚構であると私は断定しておりまして、これは消し去らなければならない。

ですから、自然言語の一番基本的なパターンも、そういう世界のあり方に相応したものでなければならないということです。一番基本的な場面は知覚の場面ですけども、例えば、「これ（主語）は薔薇の花（述語）である」とか「この薔薇の花（主語）が美しい（述語）」といった言い方が基本パターンではなくて、むしろその場合「これは」とか「この」というのは、本当はその知覚的な性状が現れる「場所」を指定する副詞的な言葉であると、そう理解すべきだろうと。で、むしろ「ここに薔薇の花がある」とか「ここに美しい薔薇の花が見える」とか、そういう言い方の方が、自然言語としてのより基本的な記述方式だといっていいですね。どっちにしましても事態そのものが決して、主語になる「これ」とか「この薔薇の花」とかそういったものがまずあって、それが「薔薇」とか「美しい」とかいう属性をその次にとるという、そんな二段階構造になっているのでは全然なくて、端的に「薔薇の花」という知覚的性状、「美しい」という知覚像がそこに現れて、我々に訴えかけてきているという、そういう事態だと思います。現代の哲学者の中ではP．E．F．ストローソンという人が「フィーチャー・プレイシング・ステイトメント」というようなことを言いまして、やはり「主語－述語」構造よりも「フィーチャー・プレイシング」、「特色を置く」というのか、「ここに水がある」とか「ここに雨が降っている」とかそういうタイプの方がむしろ基本的だといっていますが、しかし、ストローソンは気が付いているわけではないけど、そういう言い方を一番最初に開拓したのはプラトンだったわけです。主語の方は大体名詞で表されますね、それが物に対応し、他方述語は形容詞とか性質に対応するといった、そう

いう区別を最終的には抹消してしまう。同じステイタスのものとしてとらえます。詳しいことは省略致しますけれども、そういうところに人間の世界把握と世界記述の基本型を考えているということ。そう考えておかないと、知識の進歩ということーつまり水なら水について、「水とは何であるか」ということを追求していって、それで「水とは何であるか」ということに段々と目が開かれていくというような私たちの経験を十分説明することが、できないだろうと考えた。「人間とは」とか「美とは」とか「平和とは」とか、みんなそういう形で考えておいた方がいいのではないかと。

おそらくコンピュータに馴染み易いのは「主語－述語」構造の方で、主題的にあるものを立てておいて「それがこうこうだ」という、そっちの方が馴染み易いのではないかと想像するのですけれども、ただ大規模知識ベースがおっしゃるところによると、「意味を考慮した知識そのものの技術である」ということを言っておられるので、それをやるためには念頭の片隅でもいいから以上のことを置いておいた方が望ましいのではないか、まったく目を閉ざしてしまうのは得策ではないだろうかと考える次第でございます。私の話は以上で、５分超過いたしましたが申し訳ありませんでした。

座長：

どうも有り難うございました。大変残念なんですけども時間の都合があると思いますので、皆さんいろいろお考えあると思いますが、会場の方から意見を頂くということは省略させていただいて、このセッションをこれで終わりにしたいと思います。お話し頂いた先生方にもう一度感謝をしたいと思います。有り難うございました。

3. セッションII

言　語　処　理

3. セッションII：言語処理

3.1 座長挨拶

奈良先端科学技術大学院大学 情報科学研究科
教 授　松 本 裕 治

奈良先端科学技術大学院大学の松本と申します。よろしくお願いします。

このセッションは、「自然言語処理関係」のセッションということで、今朝、ＥＤＲの横井所長より紹介がありましたけれども、自然言語処理の技術というのは、こういう「大規模データベース」の「構築」というか、それを「ある程度自動的に作る」という意味で非常に重要な技術だと思います。今日は、この分野の５人の先生方にお願いしまして、「自然言語処理技術の現状がどういうふうにあるか」と、「大規模知識ベースの構築」についての関連について、議論して頂こうと思います。

ちょっとこの会議とはずれますが、今年の８月に京都で"COLING"という、COMPUTATIONAL LINGUISTICSの会議がありまして、自然言語の分野でも、「言語からの知識獲得」というか、「言語処理をして、そこからいかに言語的な知識ないし一般的な知識を抽出するか」という話題は、非常にホットな話題になっていまして、そのあたりの議論も聞けるのではないかと思います。今日最初にご講演頂く長尾先生とかYorick Wilks先生達が、その運営委員長であるとか、プログラム委員長をなさっています。ぜひ皆さんも参加して頂きたいと思います。

3.2 講　演

(1)「言語処理の現状と将来動向」

座　長：

最初のご講演ですけれども、京都大学の長尾先生にお話して頂きます。長尾先生は、紹介はもうそれほど必要ないかと思いますけれども、ずっと京都大学で、日本の自然言語処理の指導者として、この分野をずっと引っ張って下さる指導者的な立場をとって頂いている先生です。ご講演のタイトルは、"Current Status and Future Trends of Natural Language Processing"というタイトルでお願いします。よろしくお願いします。

京都大学 工学部電気第2教室
教 授　長 尾　眞

只今ご紹介頂きました長尾でございます。自然言語処理技術というものがＫＢ＆ＫＳにどういうふうな位置づけを持つか、ということについてはすでに横井さんのほうからいろいろ今日お話があったわけでございますので、それじゃ自然言語処理というものはどういうところまで現在進んできているか、今後どうなりそうか、そういったことに焦点をあててお話したいと、こういうふうに思います。

そこで、さっそくですけれども、自然言語処理には、典型的にはどういう種類のものがあるかということですけれども、まず、テキストの記憶と処理、そしてテキストのリトゥリーヴァル、こういった問題がありますが、そういったことについてはまた、後ほどのスピーカーがお話になりますので、ここでは、殆どお話致しません。その次の段階としましては、モルフォロジーの問題がある。形態素解析といわれている、モルフォロジカルアナリシスという問題がありまして、それから、シンタックスの問題がある。それからセマティックアナリシスといったものがある。それから、もっといきますと、ノウレッジの方にそこから繋がって行くと、こういうふうな形でございまして、それではそのモルフォロジックアナリシスとかシンタックティクアナリシスとか、そういったものは現在どういう程度の所まできているかということについて、簡潔にお話をしたいと思います。時間が２５分というふうに限られておりますので、細かいアルゴリズムとか、そういったことについては殆ど触れることはできませんので、悪しからずご了承頂きたいと思います。

形態素解析につきましては、英語なんかでも相当なプログラムがございますけれども、日本語におきましては、現時点ではだいたい単語単位で測って９７％から９８％位の精度で形態素解析ができるというところまできている、と言っていいと思います。たとえば、ＮＴＴの研究所がお造りになって新聞社等で使われております形態素解析のプログラムは特殊な分野のテキストかもしれませんが、９９．８％

程度まで行っている、あるいは私どもの所で造りました、ジュマーンというソフトウェア、これはその後随分改良したんですけれども、そういったものでは大体９７あるいは９８％位にはいくというようなところでございます。

文を解析する場合に形態素解析をやって構文解析をやるわけですが、このシンタックティクアナリシスでエラーが出てくる、セマティックアナリシスでもエラーが出てくる、とこういうふうになりますので、最終的に何が欲しいかというところのパーセントを、例えば９０％にしようとすると、形態素解析のプログラムは９８％位の精度では満足できないということになります。たぶん９９．９％位までは持って行かないといけない。これが、今後我々に課されたテーマでございます。これには、どういうことをすればいいか、というのはそう簡単ではございません。分野特有の辞書の用意、あるいは分野特有の言葉の言い回し、そういうものをきちっと捉えて辞書に入れて処理をする、といったことが必要になってくるわけですけれども、先程の藤澤先生のお話にありましたように、分野特有というのが本当にどこまで有効であるか、もっとオールラウンドなものでなければならないんじゃないか、というようなことを考えますと、なかなか難しい問題になります。

しかしながら、自然言語処理の、これまで３０年以上の研究の歴史を考えますと、現時点というのは出来ることはほぼやった。あとはパーセントをどこまで上げることができるかという時代に入ってきているのではないか、特に、形態素解析の分野におきましては、そういうふうに考えることができますので、精度を０．１％上げるためにどれだけの努力をしなければならないかといった時代であるというふうに認識されます。その次にシンタックティクアナリシスですけれども、シンタックティクアナリシスにはいろんな方法があります。すでに開発されているものとしましては、フレーズストラクチャーグラマーがある、あるいはディペンデンシーグラマーのやり方がある、あるいはもっとほかの方法があるということがあります。けれども、あんまり時間がありませんので、細かくはお話できませんけれども、日本語のように語順が任意であって、省略といったことが非常に多いような言語の場合にはフレーズストラクチャーグラマーというもので日本語文を解析するのは非常に難しい。それに対してディペンデンシーグラマーで解析するということの方が、まだすんなりいけるということです。ただ、ディペンデンシーグラマーで解析する場合には、かかりうけ関係の可能性が爆発的に増えるという問題が出てきます。そういったことを解決するための方法というのはあとでちょっとお話し致します。

その他にケースグラマーというのがありまして、これは意味を考えてやるというやり方でございますが、この場合は辞書をどれだけ精密に作ることができるかということが一番の問題になる。つまり、電子辞書というものの最も基本的な部分というのは、格文法に関する情報をどれだけ精密に用意することができるかという問題でありまして、これについては、ＥＤＲの辞書、あるいはＮＴＴがお作りになりました辞書というのが現在おそらく日本で利用できる最も精密なものであろうというふうに思われますが、これはやはり量及び質の両面におきまして、今後もっと充実させていく必要がある。これはしかし、非常にコストの高い作業を伴うものでありますので、これをいかに自動化できるかという研究は、今日司会しておられる松本先生なんかも随分研究しておられますけれどもどこまでうまく行くか、というこ

とについては今後の研究の発展に待たなければならないというふうに思われます。

今簡単にご紹介しました解析の各種のプロセスを日本語の解析にあてはめて考えるとどういうやり方が最も妥当であるかということになります。それは、まず形態素解析をやる。これは有限オートマトンモデルで実現できることになります。それから、構文解析は、ディペンデンシーアナリシスをやりまして、そこからケーススストラクチャーアナリシスをしてケースフレームに直す。そういうプロセスをとるのが、フレーズストラクチャーグラマーで解析をしていくよりも、日本語の場合には非常に良いというのが私どもの考え方でございます。ところが先程言いましたように、ディペンデンシーアナリシスをする場合には、かかりうけ関係の可能性が非常に多くなりますので、何とかしてそれを減少させる方法を考えなければならない。そのために考えましたのが、文の中に含まれる並列的構造をうまく発見する。そうすることによってかかりうけ関係の可能性を極端に減らして、できるだけ単一な解析結果を得るということができるようにする。そういうふうなシステムを作ることができるわけであります。

そういうふうにして作りましたシステムを、ＫＮパーサーと名付けまして、現在パブリックドメインでどなたも利用して頂けるように公開しております。そのシステムが現在どういった能力をもっているかということを少し簡単にお話をしてみたいと思います。まず、ディペンデンシーアナリシスの部分につきましては文節単位にして９７％の精度で大体解析ができる、というところまできております。これはまだまだ改良の余地がありますので、９８％位まではちょっとした努力で行くと思います。けれども、これを９９あるいは９９．５％まで持っていこうとすると、これはなかなか難しい問題を含んでいるわけでございます。

次にディペンデンシーアナリシスを文節単位でやる。それをうまく成功させるためには、長い日本語文の中に含まれる並列構造をうまく取り出すことをしなければいけないわけでして、それをうまくやる方法を考えました。実際ＫＮパーサーという我々のプログラムの中に入っているわけですけれども、それの詳しい内容は、お手元の資料の文献を見て頂きたいと思います。大体どの程度の成功率になっているかと言いますと、日本語の仮名漢字混じり文で３０～５０文字から成る文につきましては、大体７８％位の精度で解析ができる。５０文字から８０文字位から成る文については７０％位の解析の成功の程度になる。８０文字から１５０文字位の文の場合に４６％位の成功度に落ちる。平均しますと大体６５％、これは単純平均ですけれども、それを、どういう長さの文がどの程度出てくるかという、フリーケンシーウェイトをかけてやりますと大体７４％位になってくるということになります。

現在の広く使われておりますような一般的なコマーシャルシステムの場合にはどの程度になっているかというと、これはそれほど正確に調べたわけではありませんので、あるいはもっと精度がいいかも知れませんけれども、大体短い文においてはそう精度の差はないわけですけれども、長い文になりますと非常に精度の差がでてくるということです。それからＫＮパーサーの場合、ディペンデンシーアナリシスをしたあと、ケーススストラクチャーに変換していく部分、それがどれくらいの精度であるかというと大体９３％位になる。ですから、トータルでいいますと、９７％×６５％、あるいは７４％×９３％というような精度に落ちてくるわけですね。それから形態素解析が９７％とかそういうことになりますか

らさらに落ちるわけでして、それは止むを得ないわけであります。たとえば、こういうことになりますね。８０文字から１５０文字位からなる文というのは大体文節の数にしまして２０文節位になる。文節のかかりうけの正確度は大体９７％位になりますから、それが２０文節ありますと文単位の成功率は４０％に落ちるわけですね。このセンテンス単位での解析の精度を、８０％迄上げようと思うと、ディペンデンシーアナリシスの精度を９９％に持ってこないといけない、ということになります。つまり、現在の９７％というのを９９％に持ってこないといけない。まあ、９８％位迄は行くでしょうけど、９８から９９あるいは９９．５までもってくるにはどうしたらいいかというのは、これからの私どもの大きな課題であります。

５０文字から８０文字位の長さの文については、文節の数でいうと大体平均的に１０文節位でありまして、その解析精度は大体７０％位ということになります。文の解析の方式というのは、これより短い文についてはほぼできるわけですが、あとはいかにしてその解析の精度を１００％に近い所まで持ってくるか、そのためには何をすべきかということが問題になる、そういうテクノロジーの現状ではないかと、こういうふうに考えるわけでございます。

そういうふうにして文を解析しまして、かかり受け関係あるいは格構造がどういうふうになっているかというのがわかるわけですが、それで解析が大体済んだかと言いますと決してそうではありませんで、いろんな問題がまだまだ残っております。その１つは、構文の解釈、構文的解釈の多様性の問題というわけです。これにつきましては通常は意味を考えることによって関係のないものを排除していくわけですが、排除できないものがたくさんでてきます。例えば、「東京は物価が高いが田舎は安い。」というような場合に、ふつう「は」がついているものは文末の述語を修飾することが多いわけですが、この場合はそうではない。ところが「東京は物価が高いが人が集まる」というような例ですと両方にかかる、というふうに考えることができる。こういう問題を解決するためにはどうしたらいいかということになりますと、非常に微妙なヒューリスティックルールを書かないといけないということになってきます。

つまり、ある単語とある単語がかかりうけ関係にあるかどうかという時に、その２つの単語がどの程度の距離離れているか、その間にはどういう単語が存在するか、文の他の部分にはどういう単語とか構造が存在するかというようなことをよく考えてヒューリスティックルールを書くということが必要になります。そういうことをヒューリスティックルールの形でいろいろと書きますと、結構いい結果がでてくるということが充分期待できます。

そのほかにもいろんな問題がありますけれども、例えばエリプシスアナフォラレゾリューションという問題があります。エリプシスというのは語句が省略されている場合ですね。日本語の場合は非常に多く主語が省略されるとかあるいは目的語が省略されるとかいろいろあります。アナフォラは照応関係。指示詞が何を指すかという問題です。「これは本です」とかいう場合の「これ」とかですね、「これは何々です」というようなときの「これ」というのはいったい何を指すか、そういった問題を解決する必要があるわけです。我々のＫＮパーサーの場合には、格構造解析をした後、省略されている部分に対してある程度の推定結果を与えるということができるようになっておりますが、これがなかなか完璧には

行なえない。これはディスコースの問題、文脈の問題ですから、いくつかの文に跨がって情報を取り出して考えないと、解決が根本的にはできない問題なんです。しかしながら１つの文の中だけを調べるだけでもある程度の判断はできる。つまり、１つの文は、非常に微妙な、いろんな情報を含んでいるわけでして、現在のシンタックティクアナリシス、あるいはセマンティックアナリシスは、その文に含まれている情報のほんの僅かしか取り扱っていない。もっとよく考えると、微妙な情報がいろいろ含まれていて、そういうことをいかにしてうまく抽出することができるかというのは今後の大きな問題だろうと思います。

そういうことをやりますと、省略の問題とか指示詞の問題も、文に跨がってやらなくてもかなり解決できる可能性があるということが言えます。勿論いくつもの文に跨がった解釈をしなければいけないということは事実なんですけれども、それをやる前に、１つの文の中だけでどこまでのことがわかるかということを、やはり徹底的に調べる必要があるわけです。

そういったことで我々がやりました１つの例をお示しします。ある単語がどういう対象の指示性をもっているかという問題です。日本語の場合は、名詞は数が明示されませんので、その数を推定しなければいけません。これは単数の名詞であるか複数の名詞であるか、あるいは抽象名詞だから数を持たないとかですね。そういったことを推定するのが１つの文の中での情報だけでできるかどうかという問題がでてきます。例えば、「彼は学生です。」と言う場合は、「彼」は一人の人間ですから、「学生」というのは単数である。あるいは、「彼は昨日一等賞を貰った学生です。」というふうに、「昨日一等賞を貰った」というようなスペシフィケーションがついている場合は、この「学生」というのは特定な学生であるという意味でｔｈｅがつくというようなことになるとか、そういうことを判断することができるわけですね。そういうヒューリスティックルールをいろいろ書きますと、結構おもしろい結果が得られます。例えば、ある名詞がインディフィニットな内容を示しているのかディフィニットな内容を示しているのか、抽象的なセンスで使われているかどうかということですね。そういうことを調べると大体８５％位合う。１つの文の中から微妙な情報を取り出すだけで８５％まで合う。あるいは名詞の単数、複数の推定に関しては、８９％位まで合う。つまりシンギュラーであるかプルーラルであるかアンカウンタブルであるかというのを日本語の名詞について推定することができる、というような結果もでてきます。

さて、そこで辞書の問題をお話しないといけないんですけれども、もう殆ど時間がなくなりましたので、詳しくはお話できないんですけれども、辞書に関しては、現在我々が利用できる辞書としましては、格に関する情報であるとか、シソーラスの情報とか、そういうものが辞書に大体入っている。しかしながら、これから必要とされる辞書の情報というのはどういうものであるかというと、いろんなファンクショナルリレーション、あるいはコンポーネンシャルデコンポジションリレーションであるとかですね、あるいはアトリビュート・ヴァリューの問題、あるいはオブジェクト、あるいはコーズエフェクトのリレーションであるとか、そういったある単語が他の単語に対してどういう関係性を持って、どういうファンクショナルリレーションでもって結びついているかという問題を明らかにしなければならない。そういう膨大な、マルチアスペクトの単語関係のネットワーク、そういうものを作っていかなければな

らないということになるわけです。

特にこの知識ベースのプロジェクトにおいては、汎用の知識ベースを作っていくということが勿論大切な目的なんでしょうけれども、一足跳びにそこへ行くというのは非常に難しいわけでして、単語の辞書をどこまで言語内知識から言語外知識の世界に広げていくことができるかというアプローチを考える必要があるのではないかと思うわけであります。例えば、具体名詞なんかの場合には、その具体名詞の持ついろんな性質として、例えば形であるとかサイズであるとか色であるとかテクスチャーであるとか、こういうものをいろいろと集めていくとか、そういったことを考えなければいけない。そういうかたちでどんどん辞書の情報を豊富にしていく。そういうふうにしていくことが最も堅実で健全で失敗のない知識辞書作りの方法ではないか。そういうふうに、言語処理の立場から言うとなる。

自然言語処理につきましては、応用分野ですからいろいろあるわけでして、機械翻訳というのはその典型的なものであります。機械翻訳には、新しいパラダイムというのがいろいろ出されてきておりまして、これの実用化が急がれているわけでありますが、その他にナチュラルランゲージプロセッシングに関して、新しいやり方、あるいは新しいターゲット、あるいは自然言語処理の新しい応用分野、そういったものが沢山出てきつつあるわけでありまして、自然言語処理は、これから非常におもしろい、希望のある分野であると思います。お手元の資料にはもう少し詳しい内容を書いておきましたのですけれども、２５分ということですので、この辺で終わらせて頂きます。

座長：

どうも有り難うございました。それでは、あまり時間がありませんが、もし何か質問などありましら、会場から頂きたいと思いますが…どなたかよろしいでしょうか。

また、セッションの終わりに、時間がありましたら、そういう質問の時間もとりたいと思いますので…。

はい、それでは、長尾先生どうも有り難うございました。

(2)「解析・生成技術」

座　長：

それでは、次の御講演ですけれども、東京工業大学の田中穂積先生にお願いします。

簡単にご紹介しておきますと、田中穂積先生は、通産省の電子技術総合研究所から'８５年に東京工業大学に移られまして、現在教授でいらっしゃいます。今日のお話は、「自然言語処理の解析及び生成技術」についてお話して頂きます。それではよろしくお願いします。

東京工業大学 工学部情報工学科
教　授　田　中　穂　積

ただ今ご紹介にあずかりました東京工業大学の田中です。本日は、"Natural Language Analysis And Generation Technologies"ということでお話をさせて頂きたいと思います。プログラム委員の方から、自然言語の解析と生成の現状と、「将来こういったことを一所懸命やるべきだ」というようなことでまとめて欲しいということでしたので、その線に沿ってお話を進めたいと思います。

限られた時間ですので、ここでは一応「解析アルゴリズム」として、どんなものが今世界各国でよく使われているか、また、それはどういう優れた点があるかというようなことをお話したいと思います。これは主として構文論的な処理が中心となります。本来ならば、意味解析ということも非常に重要な問題となるわけですが、それについてはちょっと触れる時間がありませんので、お話することを省略させて頂きたいと思います。

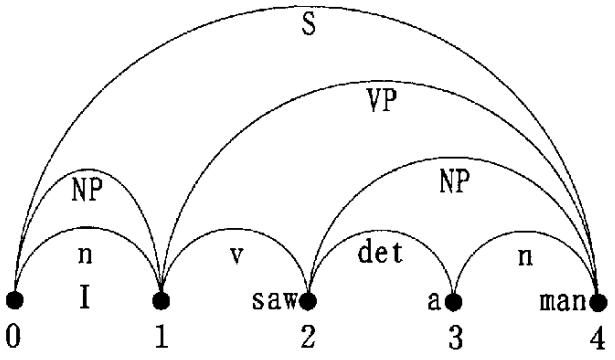
これまでのお話の中にも何回か出てきたかとは思うのですけれども、自然言語処理をする為のこういったいろんなアルゴリズムというものは、ある意味で言語的な知識を処理するエンジンの役割を果たすわけで、このエンジンがうまく働かないとそれから後のいろんな処理がうまくいきません。あるいは知識を蓄えるにしても、おそらくいちばん豊富にある知識は自然言語の形で書かれた知識でしょうから、それから機械可読な形の知識を獲得するとなった場合にはこういったエンジンが必要になってきます。そこで、まず初めに、「解析アルゴリズム」についてお話をします。

２番目に、最近非常に重要になってまいりました「対話」についてお話ししたいと思います。「対話」を調べてみますと、文法に叶った言い方をした「対話」というのは結構少なくて、私のこの話もそうだと思うんですけれども、途中で「えーと」だとか「あのー」だとかいろんなものが入ったり、言い間違いをして言い直したりということがあります。そのような場合も、頑健な解析ができるようにするため、「非文をどう扱うか」というようなことについてもお話したいと思います。

自然言語を解析すると、一般に非常に多数の構造的な曖昧性が出ますが、その曖昧性の解消をするために、「確率的な情報」を利用することによって、優先順位をつけるという技術についても触れたいと思います。最後に生成に関する最新の話題についてお話をしたいと思います。

自然言語の解析につきましては、これまでいろんなアルゴリズムがあったわけですけれども、その中の「チャート法」と、「Generalized LR法」について、その概略をお話ししたいと思います。そして、これらの方法が「非文をパーズするのにどういう使われ方をするか」、あるいは「確率的な解析をするのにどう使われるか」ということをお話したいと思います。

An Example:



[n -->"I"., 0,1]
[NP --> n., 0,1]
[v -->"saw"., 1,2]
[det -->"a"., 2,3]
[n -->"man"., 3,4]
[NP --> det, n., 2,4]
[VP --> v,NP., 1,4]
[S --> NP,VP., 0,4]

OHP－1

まず、「チャート法」というのは、文が与えられますと、文の単語と単語の間にこう、番号をふりまして、まず最初に辞書引きを行います。辞書引きにより「Iというのはnounだよ」ということが分かれば、ここにこう、アークを張ります。このアークを張ったものが文法規則によってどういうふうにまとまっていくか、というようなことをずっと追跡していくわけです。ルールに従いながら、例えば、"determiner"と"noun"の間に、こういうアークが張られていますと、これら２つが"NP"というアークを作りだします。一方「"noun"は"NP"だ」という規則があれば、別のアークを張る。さらに"v"と"NP"というアークが一緒になって"VP"というアークを張ります。"NP"と"VP"を組み合わせて、「これを"S"としてまとめて良い」という規則があれば"S"としてまとめます。今ボトムアップの方法をお話していますけれども、大体こういった形でパージングが進みます。実際には、問題はこんなに単純ではありませんで、文を左から右に見ていったときに、「これからどういった規則がここからあと適用可能か」というようなことを表すアークなども、いろいろ張り巡らします。そういったアークをうまくつなぎながら、全体として、文の先頭と末尾の間に"S"というアークができればこれで解析が終了となります。こういったやり方の特徴ですけれども、１つは、「辞書引きは同時にやってもよろしい」、あるいは『「これらを結合したアークを作る」ということとは同時にやってもよろしい』ということがありますので、非常に「並列性が見える」という利点があります。実際そのあたりのことで、Prologの上ですと、今司会をしております松本先生がＩＣＯＴで開発した「ＳＡＸ」という、非常に効率の良いアルゴリズムが実際に動いております。

これに対して最近、特にカーネギー・メロン大学を中心にして、"Generalized LR法"が脚光を浴びています。私はこのやり方に沿っていまのところ研究を進めておりますが、この"Generalized LR法"は、実際には富田さんが開発されたアルゴリズムがいちばん有名なわけですけれども、いろんな特徴を持っています。「解析結果にいろいろ曖昧性があったら、それを圧縮して、非常にコンパクトな形で保持する」というようなメカニズムや同じ計算を何回もやらないで済むようなメカニズムがあります。また、経験的には、チャート法よりも高速です。この方法は、クヌースが開発した"LR 法"をベースにしていますが、現実に自然言語の文をパーズする場合には、そのクヌースが開発したやり方では不具合が生じまして、それを一般化した、という意味で"Generalized LR法"という名前がつけられているわけです。このやりかたですと、先読み語の情報を利用しながら効率よく構文解析を行なうことができるという特徴がございます。

計算の複雑性の観点からすると、"チャート法"の方が、文の長さを"n"としますと、"n"の３乗のオーダーの時間がかかるということが示されているわけですけれども、実は"Generalized LR法"は「"n"の"m+1"乗」、"m"というのはルールの長さにディペンドした数字で、非常に長いルールがありますと、最悪の場合には３乗以上のオーダーになることもありますが、経験的には、このアルゴリズムは非常に高速であるということが知られています。それから、これはあとで述べますけれども、このやりかたは、音声認識にも使われており、わが国のＡＴＲ、ＣＭＵも使っています。LRは先読み語を使いながら構文解析をするのですが、その先読み語に関する情報のところを、スピーチの音素に置き換えることができます。そうすると、途中まで解析していったときに、「これから先はどういう音素が現れるか」ということを、ある意味で逆手に利用しまして、その予測の音素があるかどうかというふうに、現実の音声の方を調べるというやり方ができるわけで、音声認識にも応用されています。

それで、このやり方を簡単に説明いたします。これはルールの書き方で、「文というのは、名詞句（NP）と動詞句（VP）から成り立つ」ということです。午前中のお話にもありましたように、「NPが主部で、VPが述部である」というふうに見ても良いと思います。それから、"S→S　PP"は「文には、後ろに前置詞句が補語としてつくことがあり得る」という規則です。こういった形で規則を書きまして、一方で辞書を用意し、この規則から、LR表を作ります。このLR表を使いながら殆ど直線的に解析を進めることができるというやり方になっているわけです。お手元の資料には、そのあたりのことが少しステップを追って書いてありますので、わかるかと思いますが。ほんの頭のところだけちょっとやりますと、例えば"I saw a man in the park"とか何か、そういう文章を解析しようと思ったときに、最初先読み語は"I"です。これを１つ見ておくわけですね。で、いちばん最初、スタックに「状態０」をこうおいておきまして、この先読みの品詞を見ます。この品詞のことをプリターミナルといいます。ここに"det"、"n"、"v"、"p"とありますけれども、これですね。「状態０」で、LR表を見まして、先読み語の品詞が"n"ですと、「シフト４」と書いてあります。「シフト４」ということはどういうことかというと、「単語を１つ処理」し、その、「操作を１つ右に進めろ」ということで、こちらのスタックに、"I"という単語をプッシュするということになります。その時、プッシュすると同時に、こういった「０」という構

造を作って、それとペアで「状態４」に行くことになります。次は「状態４」に行きまして、次の語は"saw"の処理にいくという具合で、LR表に沿って解析を進めるというやり方です。これは４４ページから４５ページの間に例が書いてありますので、あとで追って見られると、どういうふうに動くかということがおわかり頂けるかと思います。

A sample English CFG and its LR Table

(1) S → NP VP
(2) S → S PP
(3) NP → n
(4) NP → det n
(5) NP → NP PP
(6) PP → p NP
(7) VP → v NP
(8) n → "I"
(9) n → "man"
(10) n → "park"
(11) v → "saw"
(12) det → "a"
(13) det → "the"
(14) p → "in"

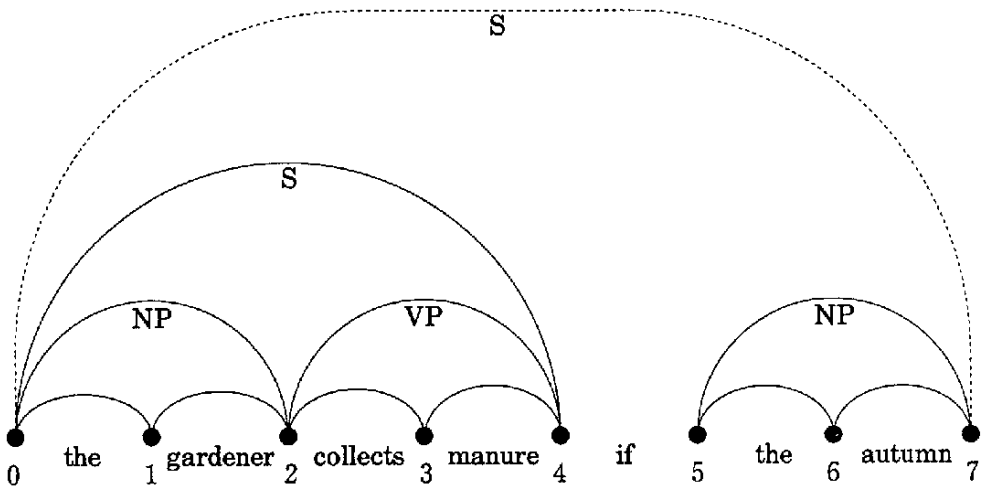
	Action field					Goto field			
State	det	n	v	p	$	NP	PP	VP	S
0	sh3	sh4				2			1
1				sh6	acc		5		
2			sh7	sh6			9	8	
3		sh10							
4			re3	re3	re3				
5				re2	re2				
6	sh3	sh4				11			
7	sh3	sh4				12			
8				re1	re1				
9			re5	re5	re5				
10			re4	re4	re4				
11			re6	re6/ sh6	re6		9		
12				re7/ sh6	re7		9		

ＯＨＰ－２

こういったやり方の特徴は、解析が進みますと、だんだんスタックの内容も複雑になったり、また簡単になったり、いろんなことを繰り返していくわけです。例えばある段階で、スタックの枝をどう延ばすかということですが、できた構造が全く同じだったら、両者をマージしてしまうんですね。マージしてしまうと、以後同じ計算をやることが避けられます。マージによる、再計算を避けるメカニズムがいろいろあって、非常に効率的なパージングができるようになっており、この"Generalized LR法"というのは非常に応用範囲が広い。それから、LR表の中身ですけれども、先程言いましたように、先読み語の品詞のかわりに、もう少し細いレベルを書くことができます。そうしますと、例えば語の一文字一文字をルックアヘッドしますと、LR表は何になるかというと、ＴＲＩＥ構造というデータ構造の辞書になります。それから、一文字の代りに音素にしますと、先程言いましたように音声認識に都合がいいという

ようなことになります。

だんだん時間がなくなってきましたが、次に、「非文をどう扱うか」ということで、Mellishという人のやり方をちょっと紹介致します。これは「チャート法」に基づいており、基本的には、「チャート法」のボトムアップ・ヴァージョンというものを使うわけですけれども、ボトムアップ・ヴァージョンでうまくいかなかったらば、トップダウンから攻めて、エラーの箇所を同定してやるというやり方です。それで、例えばチャート法で、"The gardener collects manure if the autumn"、ここは"in"のつもりが"if"になっています。こういう文章で、ifが間違っているわけですが、チャート法ですと、「ここからここまではNPができている」だとか、「ここはVPができていて、これがSにまとまっている」というものは財産として保存します。「if文」が今解析できない文法ですと、この"if"でエラーが起こるわけですが、ここまで、以前の解析結果が財産として残る。エラー以後は辞書引きだけはしておきます。



Before Error Focusing

Focusing on Error

TD Phase:

<Need S form 0 to 7>	(hypothesis)
<Need NP+VP from 0 to 7>	(by top−down rule)
<Need VP from 2 to 7>	(by fundamental rule with NP found bottom−up)
<Need VP+PP from 2 to 7>	(by top−down rule)
<Need PP from 4 to 7>	(by fundamental rule with VP found bottom−up)
<Need P+NP from 4 to 7>	(by top−down rule)
<Need P from 4 to 5>	(by fundamental rule with NP found bottom−up)

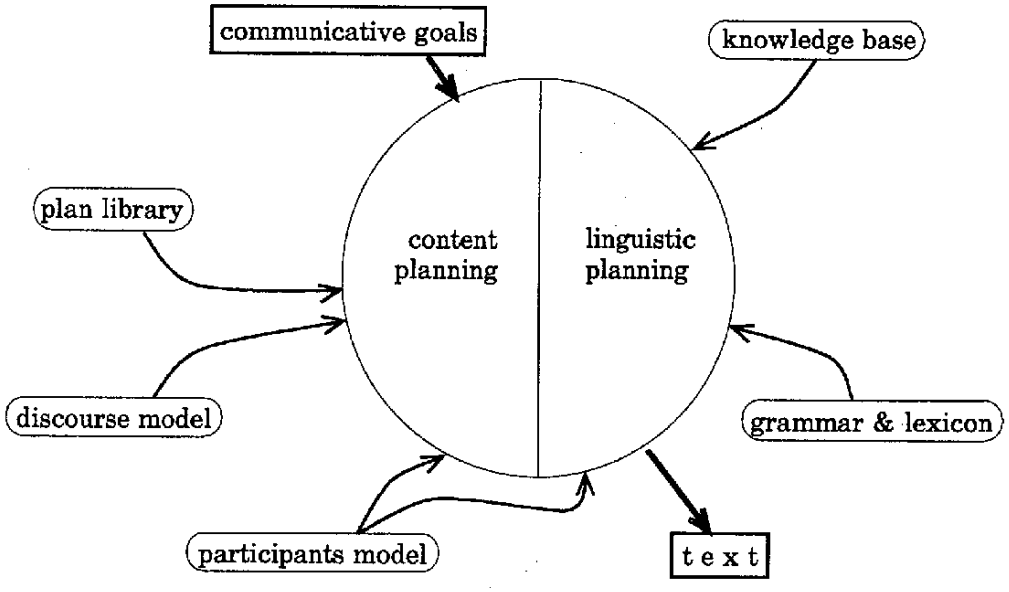
OHP－3

文法的なまとまりを、このアルゴリズムを使ってボトムアップに作っておきます。作っておいたら、トップダウンでここからここまで"S"だろうということで解析を、上から今度攻める。たとえば文法を使い「0から7の間が"S"であるはずだ」という仮定をたてるわけです。"S"というものを文法規則に

従って見ますと、"S"というのは実は、主部と述部から成るということになります。そこで、"S"をもう少し細かくNPとVPに展開しておきます。"NP"が０から２にすでにできてますから、「０から７のうちの０から２は"NP"ができていますから、２から７の間は"VP"があるだろう」という予測を次にたてるわけですね。"VP"はもうちょっと細かくみると、これトップダウンのプロセスなんですが、「"VP"に"PP"がついたものである」ということが分かる。というようなことをやっていくと、「ここからここまでは前置詞があるはずなのに、それがない」ということがわかる、こういったやり方でエラーの絞り込みをやることができます。しかし現実にはこれ非常に難しくて、時間がかかる操作になります。ですから、パラレルに処理をしたりというようなことが多分必要になってくるでしょう。

"LR"でも同じようなことができます。最初これをやった人は、ＣＭＵで、今慶應に行かれた斉藤さんだったと思いますけど、斉藤さんは、ちょっと効率が悪いやり方になっています。原理的にはかなりこのやり方と似ていますが、もう少し効率良くすることもできます。

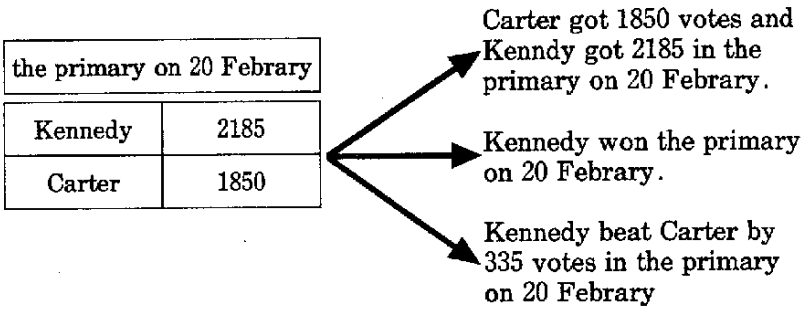
あと「確率構文解析」ですけれども、これはちょっと時間がなくなりましたので、省略させて頂きます。ここまでの話で、"Generalized LR法"、あるいは「チャート法」は、先ほど言いましたように非文の解析にも都合がいいし、効率のいいやり方であるということですが、実は確率を埋め込む、確率のスコアを計算する場合にも"Generalized LR法"は都合のいい機構を持っております。そのあたりが予稿集に書いてありますので、ご参照下さい。



Natural Language Generation

ＯＨＰ－４

次は、自然言語の生成です。これは最近の動きなんですが、以前は意味構造がありまして、意味表現なら意味表現、あるいは意味構造がありまして、それから１つの文章を生成するということが中心であったわけですが、１９８０年代の半ば以降からは、もうちょっとそこをフレキシブルに考えて、伝えたいことをゴールとしてたてまして、それからテキスト、文章の系列、そういうものを生成する。その間に何を言うかとか、文法的に合った文を言わなくちゃいけませんから、文法とレキシコンを参照しながらプランニングを行うモジュールを中心にすえて、「どういった文をどういうふうな順序で言うか」とかを決めるためのプランライブラリーを利用して文の生成を行います。さらにディスコースのモデル、それから相手のモデルも使いながら文章を生成する研究が盛んになってきています。そのように問題は非常に難しくなってきているわけで、例えばこういった技術が直ちに機械翻訳の技術に使えるかというとまだその段階にはないわけですけれども、対話システムを考えますと、こういった柔軟な枠組みで自然言語の生成の研究をするということが、非常に重要になってきます。特に、相手のモデルを考えながら、「どういう言い方をしたら、最も相手に自分の意図がよく伝わるか」というようなことも考慮に入れながら、文章を生成する研究が盛んになってきております。



ＯＨＰ－５

それで、幾つかトピックスがあるんですけれども、これはホビーという人の例なんですが、ケネディが、これだけの得票数があり、カーターがこれだけの得票数があるとします。これを相手に伝えるときにいろんな言い方が考えられます。ある場合にはケネディはただ単に数値を言うんではなくて、「ケネディは選挙に勝った」、「予備選挙に勝った」というようなことを文章にして出す。あるいは「ケネディはカーターをやぶった」というようにも言えるとか、いろんな選択枝があるわけですが、それが先ほどのゴールのパラメータによっていろいろな文章が出てきたりする、という技術が研究されております。ただ、この場合の問題は、かなりデリケートな単語の選択をやらなくてはいけないわけで、そのためには今やられている、大規模な知識ベースに含まれている情報程度ではちょっと不十分なので、どうしてもこういった研究はトイ問題のレベルにとどまっておりますけれども、非常に重要な研究の方向ではないかと考えられます。それから、最近'９０年代に入って、マルチモーダル・ジェネレーションという

ことで、コミュニケーションをするためには、自然言語の文章を作り出して相手に伝えるだけじゃなくて、それと一緒に図をつけるとか、音声をつけるとか、そういったことが必要になってきます。それと同時に、例えば音声で相手にあることを伝えようと思ったときに図を出す。図を出すときに、音声と図をどう同期させるかとかという問題も非常によく研究されてきております。

ということで、駆け足でアナリシスとジェネレーションの問題を話してまいりました。

それでは、先程スキップしてしまった確率の話をしたいと思います。確率は意味の問題にいく前に、もしかすると使えるかも知れないという話です。それで、確率というのはコーパスからいろんな確率的な情報を抽出して、それを使っていろんな構文解析の結果の曖昧性を解消してやろうということです。

そのとき、大きく分けて２つのアプローチがあります。１つは、文脈自由文法というのがあります。この文法規則にも確率をふってしまいます。よく使われる規則は確率を高くしておくというようなことです。そういうアプローチと、先程LR表というのをお見せ致しましたけど、この表の中のアクションに確率をふるという、この２つのやり方があると思います。

前者は、確率を直接規則にふるわけですね。構文解析結果の木が、ルールの組み合わせでできているわけで、そのルールに確率値がふってありますから、この確率値を全部あとで掛け算してやる。そうするとこの木の「もっともらしさ」がでるという、こういう考え方であります。

それに対して、LR表のアクションに確率値をふってやるという考え方があって、それはWrightという人が考えたわけですけれども、先程のLR表がありますね。LR表中に「アクションを何回通ったか」ということを解析中に積算しておけば、確率に相当するものが計算できるということで、それに基づきアクションに確率値をふっておきます。これは何が前者と違うかというと、前者はルールに確率値をふっていたわけですけれども、そのルールがアプライされる時期というのは、LR表でいえばレデュース・アクションというところです。このテーブルの上では、あるルールが適用する前に幾つかの、この「シフト」というアクションが起こって、レデュース・アクションが動作することになります。そうしますと、「そのルールが適用されるまで待つのではなくて、もうちょっと細かいレベルで確率値を段階的に計算できる」ということができますから、音声認識で使ってみると良い結果がでています。

さらに、Briscoeという人は、例えば、こういった解析木がでますと、この"I"というのが"noun"で、"noun"が"NP"になるという、規則は、目的語の辺りよりも、主部に現れることが多いということに着目しました。「同じ規則であっても、現れる場所によって確率の値を変えられるようにしよう」ということを実現しました。

お話ししたかったことは、自然言語処理については、きちっとした技術があるということです。その中で有望なものは、私が見る限りでは「チャート法」と「Generalized LR法」であると考えられます。私は個人的には「Generalized LR法」の方がちょっと好きなわけですが、それは好みの問題もあるかも知れません。これは、音声認識にも使えます。それから、「コンテクスト・センシティヴィティー」というようなことも、いろいろ考えなくてはいけない段階になってきているのかな、ということですね。それから文生成の問題について簡単に触れました。細かいことにつきましてはレファレンスを参照にな

されますよう、お願い申し上げます。以上です。

座長：

どうも有り難うございました。

この会場のほうは５時に終了になっていますけれども、会場自体はそんなに時間の制約はないそうですので、あまりストリクトにする必要はないと思います。それで、質問とかありましたら受けたいと思うんですけれども、特によろしいでしょうか…。よろしいですか？

それではどうも、田中先生有り難うございました。

(3)「知識獲得の自動化を目指して」

座　長：

それでは次の御講演に移りたいと思います。

We are not so restricted in time using this room. So, feel free to use time your speech, O.K.?

次の御講演はYorick Wilks先生に話して頂きたいと思います。Yorick Wilks先生は、"New Mexico State University"にいらっしゃったことはよく御存知だと思うんですけれども、そこで"Computing Research Laboratory"の所長をされまして、今年の春から、イギリスの"Sheffield University"の教授をなさっています。今日のタイトルは、「知識獲得の自動化を目指して」ということでお願いしてあります。それではよろしくおねがいします。

University of Scheffield, Department of Computer Science

Professor　Yorick Wilks

この論文に対して、専門的な同意はほとんど得られていませんので、皆さんにはお喜び頂けるかもしれません。ここでは「自然言語処理の中の２つの下位分野」について、ごく手短にまとめようと思っています。ごく手短といっても、あくまで「研究テーマの豊富さをお見せできる範囲で」ということです。本論文の主旨は、「１つの論点をごく手短に、時間の許す限りにおいて、この会議のテーマとの関連で提示すること」であり、私自身がかかわっている仕事については手短に触れるにとどめざるをえないでしょう。

本論文の論点とは、大規模知識ベース構築、多くのタスクの中でもとりわけ自然言語処理のためのそれについては、進む方向は一つであるということです。すなわち２つの分野における既存の技術を拡張することによって、知識獲得を自動化することです。その分野の１つは辞書の自動構築、これはこの会議ではお馴染みで、特に日本ではＥＤＲの仕事を通じて知られています。もう１つはコーパスからの情報抽出で、日本ではおそらくそれほど知られていないと思います。そしてそれをトップダウン方式とボトムアップ方式に統合することです。

第２の論点は、もっと言語政策的な問題で、すなわち、私は、知識獲得に関する大規模な国際協力のセンターでは、インターリンガ（知識共有のためのなんらかのインターフェース）についてもっと実効ある決定が必要であると考えています。「表現のための基準について」ならびに「世界中で研究されている多くの表現において英語中心主義が大勢となっていること」について再考の必要があります。本論文ではオントロジー獲得にかかわる提言を行なっていますが、表現が自然言語に似ている度合の問題は度外視し、表現というものを「自然言語処理のために開発された技術によって制御される一つの言語」として扱います。ポイント2Bは、横井博士のお話の論点にごく近いもので、表現の自然言語性ならびに

表現が自然言語であるまたはその類似である程度に関するものです。

まず最初に、このような知識ベース構築におけるクロス言語的およびクロス文化的な国際協力のあり方、modalityと私が呼ぶところのものについて、一言だけ述べさせていただきます。知識というものが、多少とも言語依存的および文化依存的であるならば－私は知識はおおいに言語・文化依存的であると考えておりますが、数学を扱う場合でない限り、数学はある意味で依存的でないと私は考えるのですが－、もしそうであるならば、我々はその事実から必然的に生じる事柄に、正面から取り組まなくてはなりません。つまり、例えば、２つのポイントがあります。すなわち、パートナー団体それぞれがベストを尽くすべきか、それとも１人１人がすべての試行錯誤を繰り返すのか。各人が母国語から得られた知識を研究するべきか、それとも知識を他国語で考察するべきなのか。私が申しあげたいのは、それぞれのグループや国は、得意とする技術分野がそれぞれ異なっており、パートナー団体それぞれが自前のインターリンガを使うか、あるいは自前の知識表現インターフェースもしくは表現言語を使うことによって、それは将来の協力の基盤となり得るのです。私が暗に申しあげたいのは、言うなれば、私の使っているものを皆さんもまた使うことができるなら、私はますますそれに自信を深めるだろうということです。私の考えでは、知識交換インターフェースは、もっと高い基準にしたがって構築される必要があります。国際的な協力が可能であるならば、です。この最後のポイントについては、ここにおいでの皆さんの中で、ロジックベースの知識表現派の方々には－多くはアメリカからの代表の方々ですが－、まったく自明のことと思われるかもしれません。私はここに大きな相違点の一つが姿を表わそうとしていると考えます。ワークショップで討論するされることになりますが、それは「大規模知識ベースは、自然言語処理にみられる条件を基礎として構築されるべきか否か」と言うことです。これから明らかになるように、本論文の立場は、大規模知識ベースは自然言語技術によるべきとするものです。会場の皆さんのうちかなりの方々が、まったく反対の意見をお持ちであり、大規模知識ベースは論理の条件に基づくべきであり、それが優先条件であると考えておられることは存じていますが、私はそれに賛成しかねます。

ここで自然言語関連の現在の研究分野をざっとお見せします。（国名を挙げたのは、今私が述べた個々の国やグループはそれぞれ異なる得意分野を持っているという私自身の論点に従ってのことです。）例えば、イギリスは、私の国ですが、コーパス機械可読辞書の研究にたいへん力をいれています。日本では、ＥＤＲプロジェクトにおいて機械可読辞書の研究におおいに力を注いでいます。ここでは仮に、なんらかの形でのタスク型（tasklike)機械翻訳を中心目的と仮定していますが、実際の中心目的は何であってもかまわないのです。これらの技術全ての応用目標が必要であり、機械翻訳なら向いているということです。このスピーチで私が明らかにしたいのは、情報抽出のような大きなプロジェクト、これは主にアメリカでTipsterやMUC Enterprisesによって行なわれていますが、それとこのワークショップで討議することになる新しい種類の企てであるコーパス、機械可読辞書などからの大規模データベース抽出との関係であり、私がこれから論じるのは、これらが底のところでどのようにつながっているのか、どのようにつながるべきなのかについてなのです。

この問題についての私自身が経験してきたことは－それこそ私が今ここでみなさんにお話している理

由なのですが－、私はニューメキシコ大学のコンピュータ調査研究所(Computer Research Laboratory)と仕事をしてきましたが、このグループは、このようなタスクのうち２つに取り組んできました。実際にコーパス内の機械可読辞書からデータベースを抽出すること、そして機械翻訳に放り込むことで、これはPanglossという、南カリフォルニア大学、カーネギー・メロン大学、ニューメキシコの私の共著者の大学の関わる大きな共同プロジェクトであり、Tipsterプロジェクトにニューメキシコが参加して行なった、コーパス機械可読辞書からの大規模情報データベース抽出はDeteroプロジェクトと呼んでいました。この２つでサイクルが成り立っており、まさにこの経験によって、私はここにやってくる資格ができたわけですが、この点にはあまりこだわらないことにしましょう。

この２つの成分技術について少しだけ述べておきます。これは本論文の概観部分として、２つの技術がいかに世界中に広がっているかをお分かりいただくためのものです。「コーパスとして扱われる辞書の研究」「辞書からの情報抽出」「大規模辞書データベースもしくは辞書知識ベースの構築」。世界には実に多くのプロジェクトがあり、ごく近くで行なわれているものもあれば、大変遠方で行なわれているものもあり、その他の場所にもたくさんあります。これは１０年にわたった、たいへん大規模な企てで、論文中にもうすこし詳しく述べてあります。問題は、もし機械可読辞書をコーパスとみなすなら、そこからどのような辞書的意味論的情報を得ることができるのかということでした。

これらのグループはそれぞれ大きな機械可読辞書を研究しており、テキストコーパスによって、様々な技術を利用してそのような構造を構築しようとしています。この企てには大きな問題があります。問題というのは、私の話の最後の部分で重要になるのですが、辞書の見出しそのものが自然言語によるものであり、曖昧性のある言葉からなっているということで、ロングマンやEldosのように厳密な辞書でも、定義はたった２０００セットの曖昧性のある言葉で書かれており、そのような「曖昧性のある言葉」を使って「曖昧性を除去した構造」といえるものを作り出すことが問題となってきたのです。これについてはかなりの研究努力が注がれてきており、特にガスリーたち、ガスリーとコーニーによって－ガスリーはここにご出席ですが－、「曖昧性を除去した構造」を自動構築して意味論的および辞書的情報を表現することができるということを証明しようとしています。これは大変重要なことだと思います。それは実用的な意味からだけでなく、この研究が、自動的に抽出された、それ自体曖昧性を除去した辞書的構造を得ることができるか、という技術的な鍵となるポイントに焦点を合わせているためです。「曖昧性を除去した知識表現言語を得る」とはどういうことを意味するのでしょうか。これは知識表現言語は自然言語かという疑問に極めて近いものです。もしも知識表現言語が自然言語であるなら－私の考えではその通りなのですが－、もしくは自然言語にきわめて近いものであるならば、自然言語同様に「曖昧性を除去される」必要があります。もしも皆さんが、「知識表現言語は、曖昧性を気にせず、またそのアトムが曖昧性のあるものであろうとなかろうと論理的に定義できるもの」とする考え方の伝統に属していらっしゃるなら、これが問題だとはお思いにならないでしょうが、私の考えでは、それは誤りであり、これは問題として取り組むべきなのです。

では次にさらに馴染みの薄い、しかし、アメリカでは最近５年間非常に重要な問題となっている技術

に話を進めましょう。テキスト抽出です。コーパスから、特定のドメインをカバーするスロットのついた、出来合いのテンプレートへの情報抽出のことです。この研究の膨大な部分はアメリカでTipsterプロジェクトのもとで行なわれました。ここでの課題はそのようなテンプレートを選ぶなり、うまくいけば発見して、それをコーパスに合うように調整することでした。いくつかのプロジェクトでは、例えば私の関わっていたプロジェクトでは、辞書の見出しをコーパスに合うように調整することを目指していました。私の関わっていたプロジェクトの目的は「自動的に取り出された辞書的構造を使うこと、そしてそれをコーパスに合うように調整すること」でした。重要な点は、生の情報を、できる限り自動的に、コーパスから取り出すことを試みるということです。

我々が辞書的システムを研究していくにつれ、最近まで行なわれていた情報抽出について多大な問題点が出てきています。なぜでしょう。これはまた本日の他の講演で既に挙げられた哲学的疑問にごく近いものです。大きな疑問点の一つは、「知識一般はテンプレート状のものなのか」というものです。例えば、国の歴史はテンプレート状でしょうか。昔々は－私が子供のころですが－、歴史はまるでテンプレート状であるかのように教えられ、我々は歴史とは年号と戦争と王様と皇帝から成り立っている、非常に単純なテンプレート状の形をしたものと考えていました。歴史がそのようなものではないことは誰でも知っているでしょう。歴史はテンプレート状のものではないのですが、あたかもそうであるかのように教えることもできます。知識一般は、テンプレート化できるドメインを統合したものなのか。読みだし率や精度を商業ベースに見合うまでに高めることはできるのか。現時点で情報抽出の到達度は。現時点で商業的可能性はあるのか。

最初の２つの問題には、テキスト抽出システムを大規模データベースシステムのための入力として使うことができるかどうかという問題が関わってきます。テキスト抽出システムを見ればわかるように、理論言語学と人工知能がほとんど何もなし得ていないこの近年、これはこの極めて単純明快なタスクに実際に使われているのです。非常に単純な技術によって驚くほど優れた成果が得られているのです。

アメリカでのテキスト抽出の動きは、自然言語処理を理論的な関心の段階を超えて言語工学の方向へ推し進める一助となってきました。皆さんはこれを良しとされるかもしれないし、されないかもしれません。私自身も確信はありません。問題の半分は、将来、すぐ目先の将来、この２つの技術を組み合わせて大規模知識データベース研究の一部を創出しようと考えることに意味はあるのかということです。問題は２つの組に分かれます。テキスト抽出にはドメインとテンプレートの限界の問題が有り、また辞書的知識ベースの循環性の可能性が有ります。これらは自然言語に存在するもので、曖昧性が解決されたとしても、次には制約された語彙と統語法が有ります。テキスト抽出と辞書的データベース構築とは「大規模データベース」というこの最終目標に向かって力を合わせることができます。ここで私は１つ質問をします。「この２つの技術に現存する固有の問題があってもなお、それが可能か」と。私は可能だと考えますし、やってみる価値はあると思いますし、また、大規模データベース構築についての私の見解に基づいていえば、この方向が最も進むに値するのではないかと考えます。しかし、我々は「知識表現は、我々の多くが主張するような自然言語から独立したものではない」という哲学的問題と正面か

ら向き合わなければなりません。私はもはや、「『述語や辞書的素性は単なるラベルにすぎない』と言ってしまえば知識表現の問題から抜け出して先に進める」とは考えていません。英語においてはこのように主張するのが一般的になっています。「表現言語における述語はラベルにすぎない、ほぼ全て大文字の英単語で表わすことのできるラベルにすぎない」と言うのが普通になっています。我々は、それはラベルにすぎないといって済ませるわけにはいかないし、言語に似た性質を無視するわけにはいかないのです。

しかし、私はこの苦境から脱するためには、「そうだ、知識とは自然言語によって表わせるし、表わすべきなのだ」といえば済むとは思いません。そんなことを言っているのではありません。結局のところ、もしそう主張するならば、我々は帰って図書館にこもっていればいいわけです。図書館には知識が詰まっており、自然言語によって保存されている。それが図書館なのです。もしも知識は全て自然言語によるものであるなら、我々は図書館に座ってそこにある本を読んでいるべきなのです。それでコンピュータ化の問題はおしまい、それだけと言うことになりますが、もちろん私はそんなことを言っているのではありません。私は、辞書抽出に使われる種類の技術が必要であると考えているのですが、しかし、表現言語の自然言語性に発する問題を認識し、それに対処しなければならない、それが鍵であると言いたいのです。ここで見てきた問題によって、我々は諸技術の規模・成長率・予想外の新規性に目を向けることになったわけですが、そうした問題がなくなるわけではなく、したがって、図書館に戻って自然言語で読むことだけを考えていれば問題が解決するというわけにはいかないのです。

私が主張したいのは、「表現言語は自然言語に似ているという事実を直視すること、そしてそれを制御するために全力を尽すことが解決の道だ」ということです。我々は辞書知識ベース研究および構築の成果に目を向けるべきで、ＥＤＲはこの分野における画期的な大成功例として、表現における曖昧性の問題をコントロールしているのだから、今度、我々は人手によってあるいは自動的に苦労して取り出されたこのような辞書的表現を、どうすればコーパスに合うよう調整することができるかを考えるべきなのです。その両方ともが必要だということを我々は分かっています。なぜなら我々の手元にある辞書的表現は静的なものであり、コーパスは、我々の知る通り、新しい情報、新しい言語学的用法を表わすからです。我々にはこの静的なものと動的なもの、この古い部分と新しい部分の両方が必要です。両者を統合することこそ、将来我々が行うことの中核となるに違いないと私は考えます。ここで皆さんにお知らせしたいのですが、使ってもよさそうな一つの技術が実用化されつつあります。「制限言語技術」です。その一例はPhil Hayesを長とする、アメリカのカーネギーグループの制限言語環境です。彼らは英語を特定の機械翻訳プロジェクトの入力としていかに制御するかを研究してきて、大変優れた方法を編み出しました。しかし、世界中の至る所で人々は技術的言語を一常にではないが英語であることが多いそれを一、技術目的のためにいかに制御するかを考えているのです。ここでまじめに考えてもよいかもしれないことが一つあります。それは、実際にスクリーン上でリアルタイムで英語を制御することに成功すれば、それも知識表現言語を制御するためにも使えるかもしれない、我々の仲介的インターリンガと表現言語の言語類似性を制御するために使えるかもしれないということです。探求の価値が十分にあ

る分野だと思います（どうすればいいか分かっているわけではありませんが、私が言いたいのは、「ここで、辞書的なものとコーパス的なものを統合すること」、これは多くの人が言ってきたことです、それと、「制限された英語を使うこと」、これは表現言語をコントロールするもう一つの方法ですが、この二つのことを我々は試みるべきだということなのです。どうすればよいという秘訣を私が知っているということでは毛頭ありません。）

第３に、研究すること、これは日本でのこの会議においてはとりわけ非常に重要だと考えますが、「表現言語をいかにして英語という一つの自然言語から独立させるか」を研究することが重要です。現時点では大変不幸なことに、奇妙な歴史的および学術的理由から、英語が表現言語となる傾向が支配的であるのは事実です。日本の機械翻訳システムの中にも英単語を用いた英語様のインターリンガを仲介言語として使用したものがあります。それがいけないと申すのではありませんが、そのことによって我々はそれを用いることがどういうことであるかを非常に深刻に考えさせられるのです。これは、インターリンガと辞書の述語を複数の自然言語から合成して作り出すというような実験になるのでしょう。このテーマはワークショップで、「表現言語をいかに構築し拡張するか」の問題に伴って出てくると思いますが、私としては、「自然言語からインターリンガを拡張・合成することにより表現言語を構築する方法を考えてはどうか」と提案したいのです。

Panegloss プロジェクトの中でのことですが、南カリフォルニア大学のISIで、Ed Hoveyは多数の語彙ソースからインターリンガの項目を合成することを提案しており、これは本日の他の研究でも取り上げられたか、またはこれから取り上げられると思いますが、Miller's word netから、またL-DOOS そのものからMSUの抽出作業を通じるなどによって、インターリンガを合成することです。さて、これはある意味で難しいものの、それでも比較的簡単な仕事です。なぜなら述語は全て同一の自然言語に由来する、英語の様々なソースに由来するものだからです。我々は、「知識交換言語における述語を異なる言語から合成することはできるのだろうか」と自問してみるべきかもしれません。それはどのようなものになるのでしょうか。最初は馬鹿げたものに聞こえるかもしれません、フランス語なまりのドイツ語は一つの言語といえるのだろうかというようなものです。フランス語とドイツ語の語彙を合わせたもの、これは言語でしょうか。私にはわかりませんが、いい質問ではありませんか。この言語を呼ぶ名前はありません、ついでに言えばこれは歴史からして英語ではありませんので念のため。しかし、もしもHoveyや同じ様な考えの人々が、このような統合によって作られた完全な知識言語であるインターリンガを凝縮させる効果的なやり方を定めることができるならば、重複性を取り去って凝縮する効果的な手順を明らかにするならば、現存する知識交換言語から、英語性あるいはいずれかの自然言語性を－それはなんであろうとかまいませんが－、それを一応全て取り去ったということで、大した業績になるだろうと思います。言い換えれば、我々は知識ベースのための表現言語構築の議論を始めるにあたって、このような技術に目を向ける必要があると言ってよいでしょう。

さて、今申し上げたばかりの事に対する別の形の答えとして、よく知られているのは、日本の機械翻訳研究ではしばしば言われているようですが、インターリンガとその他の言語とを取り持つ規則を作る

という言い方です。日本のグループのいくつかは、これを意味トランスファーとよび、日本語的インターリンガ構造から英語的インターリンガへ、またその逆の橋渡しのためのトランスファー規則となると言っています。このアイデアはなかなか興味深いものです。この考え方は、この複数のインターリンガは真のインターリンガではないと認めるものです。日本のインターリンガから英語のインターリンガに移動する必要があるならば、定義に照らして真の知識言語ではない、あり得ない、さもなければ２つもあるはずはないからです。これは問題にたいする一つの考え方です。これが特に魅力的だとは思いません、なぜならこれには経験的根拠があるとは思えないからです。意味トランスファーの概念は、例えば松本さんが統語トランスファーに対する経験的根拠をあたえたような方向にあるのでしょうか。松本さんは言語間、コーパス間の統語トランスファー規則を学習するメカニズムを示しています。それは完全に、経験的根拠のある概念です。これが根拠のある概念かどうか確信はありませんが、私は経験主義者なので、確かに興味深いアイデアではあると思います。

しかし、私が強く信じるところでは、インターリンガ構造、あるいは少なくともそのclosed-class itemsについては－ここではopen-class itemsについての話はしませんが－、私がお話してきたような技術をなんらかの形で組み合わせることによってチェックする必要があります。つまりＣＲＯで曖昧性を除去した辞書見出しを構築するため、または少なくともそれ自体の情報源である辞書に対して曖昧性を除去したものを構築するため、これもやはり堂々巡りとなるアイデアであることは先程も申しあげましたが、そこに用いられている曖昧性解決技術、それとカーネギーの制限言語環境のオーサリング法のような技術とです。もしも共同研究パートナーの２つ以上が同じインターリンガを使うのであれば、これは決定的な問題となります。'９２年のフィラデルフィアでのＥＤＲワークショップを振り返っていただけば、興味深い論文があったことを思い出していただけるでしょう。ＥＤＲインターリンガ表現を母国語以外の言語で理解する場合の問題についてのものでした。この問題はここにいる誰にとっても重大な問題であると私は考えます。私は、インターリンガや知識表現言語に反対しているのではありません。ただこれらを、もっと真剣に考えて、伸展性のあるものにするべきではないかと申しあげているのです。機械可読辞書を基礎とした研究によって、昔はまったく推測的なものだった辞書がより真摯で一貫性のあるものとなったのに倣いたいものです。

第４に、これで最後ですが、私の考えでは、また私の共著者のNirenbergも同様の意見なのですが、サブドメインの一つとしてのオントロジーを確立するためのできる限り文化に規定されない、独立した研究が必要ではないでしょうか。これはワークショップでの主なトピックの一つになるはずなので、ここでは余り多くは申しあげません。この種の研究の先鞭として、必要ならばいくつかの自然言語の範囲内にあるコーパスから作られた既存の機械可読辞書からでもよく、カーネギーメロン大学のMacrocosmosプロジェクトはそれを研究していますが、ＩＢＭのJellenyk-Mercer-Brownグループにより、皆さん聞いたことがおありかもしれませんが、コーパスから純粋に経験的にオントロジーを構築するための、高度に経験的・統計的な研究が行なわれてきました。このような意味論的クラスーこれらをオントロジーと呼ばれても結構ですがー、これらは純粋に経験的に構築されたもので、これらは旧来の知識

には基づかず、これらのクラスは、完全に経験的に約３億の英語の語彙のコーパスから構築されたのです。これらは極めて優れたものです。これらのクラスが、完全に経験的関連の手法empirical association techniquesだけによって巨大なコーパスから構築されたもので、このように一貫性のある緊密なクラスが構築できたという事実は非常に興味深いものであり、私はこれがある種のオントロジーの基礎となるのではないかと考えます。ＩＢＭは出来の悪いクラスは見ません。出来のよいクラスだけです。これらは特別うまくいったものなのです。モンスターが生まれているかもしれませんが見せてもらえません。これはたいへん感銘深いことです。このようなコンピュータ技術は１０年か２０年前には不可能だと考えられていたまさにそのものなのです。

結びとして、いかなる困難があろうとも、我々はコーパスから抽出された新しい情報を使って半永久的辞書知識ベースを合成する努力を続けなければなりません。新しいというのは事実に関しても、語の用法に関してもです。言葉は変わらずにはいません、事実も変わらずにはいません。ただ既存の自動あるいは手作りの辞書をもって、それが将来を捉えていると考えることは出来ません。そんなことができないのは分かっています。それらをリアルタイムで、自然言語から実際に生まれている知識によって増強していく方法が必要です。ＥＤＲに既に投じられたような大きな投資から引き続き利益を生み出していくにはそれが唯一の方法なのです。国際的な知識交換のための新しい基準を、言語と文化のインターフェースとしての知識表現言語の特性という中心問題と取り組むことによって作り出さなければなりません。標準化は可能です。この会議に参加している多くのグループは、ワークショップで十分な討議を持つことになりますが、そのやり方については私の考えとはかなり違う方法をとるべきだと考えておられるでしょう。ワークショップに出されている論文をご覧になれば、専門用語expertiseドメインから作り上げるという全く異なる方法で行うべきだという考えも有ります。私の考えとしては、コーパスから、そして辞書から作ることも考慮すべきであり、特定の専門的技術ドメインのための英語の論理構造のようなものから単純に作るべきではないということです。これについてワークショップでは活発な議論が期待されると思います。これらの問題に対する一つのアプローチが紀要に記されています。イギリスとアメリカの大学の研究グループは、合同プロジェクトとして新しい形式の並列実行のための研究を行なっています。有り難うございました。

座長：

Thank you very much.

それでは、質問があれば受けたいと思いますけれども、よろしいでしょうか。

それでは、時間もつまっています。質問の時間は後でまとめてとりたいと思いますので、その時にまた、どなたにでも質問、よろしくお願い致します。

それでは、どうも有り難うございました。Thank you very much.

(4)「言語処理のためのテキスト資源の収集と利用」

座　長：

それでは次の講演ですけれども、Susan Armstrongさんにお願いします。Susan Armstrongさんは、現在ジュネーブ大学及びジュネーブのISSCOの所属でいらっしゃいます。今日のタイトルは、現在、彼女が積極的になさっている仕事に関してですけれども、「テキスト資源の収集」ということで、言語処理の為のテキスト資源の収集と利用ということについてお願いします。

University of Geneva

Professor　Susan Armstrong

私はComputational Linguistics Communityを代表してお話させていただきます。この分野ではコーパスの使用が新しい方向として現れてきています。コーパスから情報を取り出す新しい技術、これについて皆さんにお話したいと思います。初めにこの資源を利用するための技術そのものについてお話し、次に私達が共有でき、誰もが使うことのできるコーパスを獲得する上での、非常に実際的な問題についてお話します。私の話は基本的に２つの部分からなっています。最初は問題解決のための新しい方法、コーパスによる私共の研究の内容について少し触れた後、最近の技術、およびその応用可能性について簡単にさらい、その後で、私共がこのような資源を得るためにどのような研究を行なってきたか、そしてそれをどのように使おうとしているかをお話しようと思います。

新しい技術とは何のことでしょうか。これはデータ指向と呼ぶべき分野での比較的新しい方向で、テキスト型資源による研究のことです。以前には、ルールベース・システムを記述文法writing grammerとみなしていました。新しい方法ではコーパス、つまりこのデータがある場合、「データ指向」という言い方をしており、私共がやろうとしているのは、まさにそこから情報を抽出することなのです。私は言語をいかに構造化すべきと考えているかという内向的な話をするのではなく、言語を表層指向で考え、何が観察され何の情報が得られるかを考えていきたいと思います。その方法は、したがって、非常に蓋然的なモデルを使い、そして情報抽出は統計的方法によって行ないます。大量のデータを扱い、現れる規則性を抽出するという方法です。

動機は何か。機械翻訳を例に挙げてみると、writing grammer応用のための他の研究を挙げてもよいのですが、そうすると、私達はこれまで長い間一つの方向に進んできたこと、しかしあまり大きな前進はできなかったことに気がつきます。では、新しいコーパスベースの研究をルールベースのやり方と比較してみましょう。ルールベースのＮＬＰシステムを、私達は長期にわたって開発研究してきましたが、このシステムでは、どの言語であれ、ごくごく小さな断片しか説明できません。あなたが文法を記述したとします。そして新聞を渡されました。自分の文法でどのくらいカバーできるか試してみようと思います。多分、１つの文すらカバーできないでしょう。それで私達は、実際の言語の用法は私達の使って

きた方法で説明できるよりもはるかに複雑であることに気付いたのです。またこれらの方法は拡大性がないように思われます。新しい規則を次々に付け加えていくことはできますが、新しく追加しようとする規則ほど長くなり、そして言語の中でそれがカバーできる部分は小さくなっていくのです。この新しい方向に人気が集まるようになったもう１つの理由は－実はこの方向は１９５０年代に追究されていたのにも関わらずですが－、今になってこのような資源が得られる少なくとも可能性が出てきたこと、そして今ではそれを扱うことのできる機械があることです。このように、今や私達は一定の基礎的情報を与えてくれる信頼性のある方法を開発するにいたっており、そしておそらく、これが私達全員にとって非常に重要なことだと思いますが、今私達は、おもちゃのようなシステムを作るのではなく、機械翻訳のためであれ、他の何かのタスクを実行するためであれ、コンピュータをよりよく使うために私達の技術がどのように役立つかを実際に示すことを求められているのです。

別の観点から見れば、これらの新しい方法、これらの新しい技術は、少なくともこれまでどうしても満足すべき解決法のなかった問題に解答をもたらしてくれることが期待できます。その第１は獲得の問題です。私達は辞書やその類を使って来ましたが、それではコンピュータ上で言語を有効に使用する目的でのテキスト研究に必要な情報全てを得ることはできませんでした。そのため、獲得の問題は、いまだに重要な研究課題です。従来の方法は非常に労働集約的だった上、なおかつあらゆる情報を与えてくれるものではありませんでした。大量のデータの中からの情報抽出によって、大きな前進が期待されます。

coverageの問題ですが、以前ルールベース・システムにおいて、coverageはごくごく限られたものにすぎないことがわかり、従って、私達は非常に小さいドメイン、ごくわずかなテキストを対象とするしかありませんでした。大量のテキスト、大量で多様性のあるコーパスにアクセスできればcoverageは向上します。３番目に問題になるのは耐性robustnessで、これはルールベース・システムにおいては非常に現実的な問題です。というのは、このシステムでは私達がこうあるべきだと考えるのと異なる事柄は、何も説明できないからです。例えば、新聞を見れば分かりますが、テキストの中で毎日使われており、整っていない、または少なくとも、私達の説明できない通常の範囲を超えた変則であるとみなすものは、しょっちゅう見られます。コーパスを扱う場合、私達は「言語はただこのように使われてきたのだ、これを説明できるようにすべきだ、そうすればより耐性のあるシステムを作れるようになるだろうし、そしてあわよくば拡張性も得られるかもしれない」というふうに言えるのです。もしもある一つのコーパス上の言語をどう説明すればよいかを知る方法が開発できれば、その同じ方法を、別のドメインの他のテキスト群にも適用することができるでしょう。このようにいろいろな可能性が考えられるのです。

手短に、私共がコーパスによってどのような種類の研究を行なっているかを一通りご紹介したいと思います。これを言語学的応用と呼ぶことにします。私共の行なっている研究は、少しもおおげさでないものなのですが、しかしこれによって、かつてルールベース・システムによる研究で、深いレベルを目指していたときよりも、遙かに先へと進むことができました。品詞付けは、おそらく、このような相対的にごく単純な方法によっていかに大きく前進できるか示す、第１番目の、そして最良の例ではないで

しょうか。私共は、コーパスから文法を帰納的に推論するため、文法の信憑性を高めるための新しい方法を開発中です。辞書獲得、すなわちコーパスから辞書に必要な情報を獲得すること、そして最後に、翻訳研究です。また、翻訳された並列コーパス・テキストに、より多くアクセスできるほど、多くの可能性が見出されます。これによって、もしかしたら私達は従来の機械翻訳を研究していた時よりも少し先に進むことができるかもしれません。

品詞付けについて。ひと続きの言葉があって、その一語一語に、曖昧さ無く品詞を割り当てたいとします。これは皆さんがどのような研究の基盤とするにしても、非常に重要なことです。このタスクを分析してみると、何よりまずトークン化（tokenize）すること、すなわち語、数、等々、また句読法が何かを同定して、文なり何なりに区分します。考えられる品詞を結び付けるには、なんらかの辞書あるいは形態素解析が必要です。なぜなら、当然のことながら、言葉は曖昧なものだからで、そのために様々な方法が開発されており、そこからモデルを探すことにより、それらしい品詞を決めることができるのです。現時点で最も典型的なものは、an-em gramで、これはいわばtri-gramモデルであって、あらかじめ２つの品詞が分かっていれば、３番目にどんな品詞がくるかが分かるというものです。この研究は非常にうまくいっています。非常に単純な方法によりコーパスを対象として９５パーセントもしくはさらに高率の正解を得ています。これによって何ができるでしょうか。自動品詞付けはどのようなタスクにとっても、その後のどのような自然言語処理タスクにとっても基本となるものなのです。何か複雑な構造を分析しようとする場合、なんらかの文法を用いるなら、まず最初にするのは品詞を決定することです。私はここで皆さんのご参考になりそうなことをいくつかお話ししていますが、私が支持している方法は、お聞きのとおり非常に単純で、おおげさでなく、実際に使える方法であり、これにより知識ソースを構築することができます。これは文解析用の前処理系として使え、文解析の効率を向上させることができます。これは、例えば、句認識の基礎となります。これによってごく単純なパターンを書いたり認識したりでき、英語の場合、一列の名詞があれば、それはおそらく名詞句またはなんらかの固有名詞なのだろうし、英語で名詞が１つあり、次にofのような前置詞があれば、次は名詞、等々。これが基本で、文法機能を割り当てる研究もいくつか行なわれており、もし並んで動詞が１つ、次に名詞が１つあれば、それは目的語である可能性がある、ということをやっています。語の意味の曖昧性の除去、インテリジェント・インデクシングと検索、また例えば、精密コンコーダンス作成と、それを利用した半自動あるいは人間によるテキスト精査、また例えば辞書編集法について求める情報への絞込みなどです。

コーパス・ベースの研究におけるもう１つの重要な発展は、辞書獲得の自動化です。最近コーパスからこの種の情報を得ようとする試みが盛んになっています。８０年代には機械可読辞書からこの種の情報を抽出しようとする研究が盛んだったのですが、限界が知られるようになり、現在では大量のテキストからこの種の情報を得ようとする試みが勢いを増しているのです。ここではこれまで行なわれた研究の幾つかをご紹介するにとどめます。例えば、テキスト内に見出される手掛かりcueによってサブカテゴリー分類枠（subcategorization frame）を自動識別すること、固有名詞を同定分類すること、また未知の語について、テキスト内にあるがままに未知の語の属性を推測しようとする研究もあり、句認識

については、語の意味の曖昧性の除去と辞書的分類の発見、さらにシソーラス上の関係の自動構築などの研究が行なわれています。このような研究は大量のテキストデータにアクセスすることによってうまくいくはずのものであり、実際うまくいっています。さてそこで最後の１つは、並列コーパスを利用した自動翻訳というタスクです。ここでむしろ強調しておきたいのですが、私達にとって機械翻訳はいまだに少し難しすぎるのです。ですから、まず小さなステップからはじめましょう。さて何ができるか、どの部分が自動化できるでしょうか。これまでに分かったのは、セグメントとその翻訳を対応させる自動整列（automatic alignment）ができることです（ここにご紹介している研究では、文の長さや文中の語数、また、２つの文中に共通してあらわれる語の組み合わせの規則性などに基づいています)。この研究でもまた９５パーセントから１００パーセントの正解率が得られており、これもまたもう少し翻訳的な問題といえるものの、研究をはじめる土台となりうるものです。なぜなら翻訳上の問題としてみなすべき範囲を限定することができ、またすでにテキストの断片の整列が終わっているので、つまり対応づけが済んでいるので、次にはある１つの単語が現れる場所を全て探し、それから翻訳中で現れる場所を全て探せばよいのです。その外にも単語対応を自動的に見出そうという研究も将来性のあるものです。可能性のありそうな最後の分野は翻訳者のためのツールで、この場合でも、私達はもはや直接に自動翻訳を行なうことを考えているのではなく、確かにコーパスを使って機械翻訳を行なう研究はありますが、明らかに将来性があるのは翻訳者のための優れたツールを作ることに絞った方向なのです。

ごく手短に、現在コンピュータ言語学の世界でコーパスを使って行なわれている研究のいくつかを見ているのですが、ここにきわめて新しい方向が見えています。それは何かというと、テキスト資源の必要性が非常に大きくなっており、さらに高まろうとしているということです。これらの方法は、全て大量に集めたテキストにアクセスするということに根本的に依存しており、テキストが多く集まるほど、ただ手当り次第集めただけでなく、よく整理されたテキストの集積ができるのです。それからまた、このような資源を共有できるよう新たなアクセス法を開発することも必要です。そして最後のポイントは、私がここでご報告した研究、また文献に記載されている研究、これらは全て英語を使用したものであり、そろそろ他の言語のテキストを集めることにももっと力を注ぐ時だということです。特に、翻訳研究をさらに盛りたて、フランス語と英語以外に広げるためにも、もっと並列データが必要なのです。その資源はどこにあるのか、テキストはあらゆるところにあります。本を買うこともできるし、図書館に行くこともできるし、世の中にはテキストは山ほどあるのですが、そのデータをどうやって手に入れればよいのでしょうか。まず最初によくぶつかる問題は、物理的な取得の問題です。新しいテキスト、古いテキストというのはどう意味かというと、新しいテキストというのは現代の電子機器よって今作り出されているものであり、そのデータにアクセスするのは簡単です。古いテキストではこう簡単には行きません。例えば、翻訳されるテキストの種類は限られています、限られた種類のテキストだけが翻訳されます。さらに電子的方法では、簡単に手に入れられるのは文書になった材料だけであり、発話された材料についてそうはいきません。こうした問題はあと少なくとも１０年は続くと思われますし、その後には少しはよくなるでしょう。しかし、過去に印刷されるためだけに作られたテキストにアクセスし、その

データを操作するための情報を得るというのは、必ずしもごく簡単にできることではありません。これらのテキストに付けられた印刷用コードは種々異なっており、たとえば印刷についてのコードと、これから抽出しようとしているテキスト内容そのものについてのコードは分離できません。マーク法が標準化されていないということは、新しいデータを手に入れるたびに違うマークが付いていて、そのたびにそのデータにアクセスする問題が新たに生じるということなのです。もう１つこの会議にとって重要だと思われるのは、購入が可能かどうかの問題です。問題はこうです。電子テキストのライブラリーについてお話しますが、それはまさに、それこそ私達があって欲しいと思っているものだからです。先程、「いつでも図書館にいけばいい、本を買えばいい、さもなければどこかから借りてくればよい」と申しましたが、電子的データについてはそのようなアクセス法がないのです。電子的データは出かけて行って手に入れるというわけにはいきません。研究に必要な電子的データの１つ１つを、今まで本を買ってきたのと同じ様に、行って買ってくればよいと考えるわけにはいきません。私はテキスト入手について少々研究してきたので、いくつかの方法をお話しましょう。この種の資源を開発する時に問題となるのは何でしょうか。私が最も重要だと思うのは、最初のものと最後のものの２つです。少なくともこの種のデータを入手しようとして大きな組織と交渉した時に、私が経験したのがそれです。最初の問題は、一番重要な問題でもあると思うのですが、価値の考え方で、内容ではなくキロ単位で計ると言うべきものです。他の用途のためのデータが生み出されてきたのは情報内容のためです。私達が新聞を買うのはその情報内容のためですし、小説を買うのはその芸術内容のためです。データというものには必ず内容的価値があったわけですが、さて私達が雑多なテキストを使って研究したいことは、そのような内容的価値についてではなく、ただ言語がどう使われているかをそれが反映しているからなのです。従って、誰かのところに出かけて行って、従来通りの規則にしたがって「大量のデータをください」ということはできません。また購入可能性の問題に戻ってきます。小説を買いに行くときなら、何のためにお金を払うのか分かっていますが、自然言語処理のためのコンピュータ言語学の研究の対象にしたい小説全部に、同じ様にお金を払うことはとてもできません。そしてこれが最後の問題につながっていきます。それは、単に著作権法だけの問題ではないのですが、著作権法が取得や利益を生む可能性に関する事柄を保護しようとしていることです。引き出されたデータの地位の問題もあります。このようなテキストを用いて研究を行なう場合、テキストから新しい単語のリストや新し情報を引き出す時、これをどうやって保護すればよいのでしょうか。それから機密性のあるデータが私達に、私達の自然言語処理研究にとって重要である場合、先にも申しあげたように私達はデータがどのように用いられているかを、あらゆるコンテクストについて知りたいのですが、なんらかの理由で機密であるため、アクセスしてはならないものがあるのです。例えば医学報告書だけについて考えてみると、もしも医学報告書を自動筆記するためのインテリジェントシステムを作りたい場合、そのなかに書かれている人を保護しなければなりません。なぜなら私達にはそれに対する権利が無いからです。テキスト資源を手に入れようとするとこのような問題に直面しなければならないのです。

それではテキスト入手の手続きはどんなものかをごく簡単にお話します。まず最初に、そのデータが

どこにあるかを知る必要があります。物理的にどこにあるかが分かったら、入手できるかどうかを確かめます。それを手に入れることができるか、フォーマットが合うか、使用可能な形式になっているか。私はいろいろな国際組織と交渉して研究目的のために資料を手に入れようとした経験から、国際的組織ではそのようなデータを外部に渡す権利のある人がいないことを知りました。そのための仕組みができていないのです。そのための基本的制度がないのです。本を売るための仕組みならありますし、報告書を配布するための仕組みもあるのですが、このようなデータを外部に渡す職務を持った人はいないのです。

この問題に対処するため、いくつかのテキスト収集計画が始まっています。どの研究所もどのグループも、単独では必要なデータ全てを手に入れることはできないため、協力することが必要です。ここで私が直接関わってきた計画のいくつかと、まさにこれから始まろうとしているプロジェクトのいくつかをご紹介します。このようなプロジェクトは、明日からみなさんがデータを収集しようとする時に、お耳に入る機会が多くなるに違いありません。最初の１つはＡＣＬデータ収集計画という、１９８９年に始まったもので、私達にはデータが必要であることを認めたものでした。これは１９８９年に設立され、大規模なテキスト・コーパスの獲得を支援することを目的としていました。かなりボランティア的な努力によるもので、ここでもまた、「はいこれが仕事です。給料を払うからやりなさい。」という人も制度もなかったので、各研究団体は非公式に協力して、この問題をどうにかしようということになりました。私達はテキストを収集し、それをすべてCD-ROM化して、いろいろなところに配布しました。まもなくアップデート・バージョンが出るはずです。アメリカのACLDCIによって、私達が必要とするものにずっと近いものが設立されました。The Linguistic Data Consortiumです（私がこれをお話するのは、単にＬＤＣの代表の方が誰もここにきておられないので話す人がいないと思ったからです）。この必要性を認めたのがアメリカ政府であり、かなり多額の基金を出して、テキストを収集し配布するための組織を作りました。設立したその年の内にCD-ROMを１００種類作って配布し、これからもそのペースで、もしかしたらペースはさらに早まるかもしれませんが、必要なデータの収集が続けられます。来年は並列データ収集が主なテーマになるでしょう。ACLDCI、また始めのうちはLinguistic Data Consortiumも、どちらもアメリカ中心なので、英語のデータだけを集めました。それで私達は同じやり方でヨーロッパ語のコーパス計画を作ることを決め、これによって英語以外の言語のデータを集めることにしました。英語だけの研究はしたくなかったからです（私達のやり方もまた偏っているかもしれませんが、いずれにせよそれぞれの国で自国語で研究ができることは重要ですし、また他国の人もいろいろな言語のテキストにアクセスできることは重要です）。私達はヨーロッパ・コーパス計画を開始しました。これもまたボランティア的努力による作業を続け、今年の終わりまでにはCD-ROMを完成する予定で、それはごく手頃な価格になるはずです。

これによって著作権問題と権利保護の問題を多少とも克服できたのですが、私共はあることを行ないました。それは協定書を作成し、ACLDCIとよく似たものですが、そのポイントは、私達は「保護する」ことを－データ提供者に対してできる限り権利を保護し、データを悪用しないことを－約束したことで

す。また１つ重要なことは、データをもらう時に、資料に追加の制限をつけたいかどうか提供者に尋ねて、あれば協定書に記入することでしょう。例えば、私はデータをＩＴＵ、つまり国際電気通信連合から収集しましたが、ＩＴＵは確かにもう一つ制限を追加したい、そのデータを使用して何人も多言語グロサリーを作れないようにしたいとのことでした。私達は現在多言語グロサリーを作る技術を持っているのですが、ＩＴＵが多言語グロサリーを販売しているため、その利益を損なうおそれがあったのです。このようにデータ提供者の権利を保護する手段を持つことはたいへん重要です。同じ理由で、私達は資料を受け取るユーザー各人に向けた協定書簡を作り、研究目的のみに使い、部外者に再配布はしない旨を署名誓約してもらっています。これは誰でも資料にアクセスできるが、私達がそれをコントロールしてとんでもないことにならないようにしようということです。ＥＣＩの資料は、喜ばしいことに様々な言語のデータが集まっており、少なくともヨーロッパの主要言語だけで少なくとも５００万語を集めたのですが、それでもまだ本気で研究するにはとても足りないのです。しかしこれをよい契機に、例えば英語以外の言語における品詞付けなどが進んでいくでしょう。並列データのほとんどはＩＴＵのような国際組織から得られ、そのための日本語データも国際労働機関から得られることが確認され、私共はできる限りあらゆる資料を収集しました。英語、フランス語、スペイン語だけでなく、アラビア語、ロシア語もです。中国語については収集できるかどうか確実でなく、全て国連の会議報告なのですが、私達はジュネーブから得られる資料は全て収集し、また一部はニューヨークから集めました。これがテキスト収集計画で私達が行なっている仕事です。

さてこれでもう１つのポイントをお話しできることになりました。次には基本的アクセスツールが必要であり、そしてそれは全て共有性および汎用性のあるものでなければならないことです。この種の研究があまりたくさん重複するのを防ぐ必要があります。より高度な、データを利用するためのより高度な方法を開発するための基礎として、誰でもやりたいのは明らかだからです。私達は誰にでも手の届くツールができればよいと思っています。この事に関して申し上げておきたいのは、ヨーロッパで１つの大プロジェクトがまさに始まりつつあり、これはMultextと呼ばれていますが、そのプロジェクトでの私共の目標はまさにこのようなツールを作ることなのです。これは２年計画で、１月に始まることになっています。また、ヨーロッパの少なくとも６つの言語について大量のデータ収集を計画しています。また、あらゆる基本的なアクセス・ツールを作ろうとしています。重要なのは、私達はこのような成果をあげるために全身全霊を捧げていることです。つまりコーパスとアクセス・ツール、仕様、その他の付属文献を無料で一般に使用できるようにすることを目指しているのです。

まとめとしまして、コーパス・ベースの研究からは、自然言語処理における新しい可能性が開けてきます。将来性のある研究が進んでおり、お手元の紀要に掲載されている私の論文にもその幾つかを紹介してあります。データの新しい利用法もそうですが、しかし何よりも今この時点で心に刻んでいただきたいのは、収集と加工は共同の仕事として、国際的に共同して、学術および産業の各分野を超えて行なう必要があること、そしてそうすることによって、未だに解決されたというには程遠いたくさんの実際的な問題に答えていかなければならないということです。

座長：

どうも有り難うございました。We'll take the questions later after this session, O.K.?

(5)「機械翻訳における知識処理」

座　長：

それでは、次の講演に移らせて頂きたいと思います。

このセッション最後の講演になりますけれども、辻井先生に講演して頂きます。

辻井先生は、京都大学で、長尾先生の研究室で研究を続けられたあと、約５年前にマンチェスターに移られて、現在、マンチェスターのUMISTの言語関係の研究所の所長をなさっています。今日のご講演のタイトルは、「機械翻訳における知識処理」という御講演です。よろしくお願いします。

The University of Manchester Institute Science and Technology
Center for Computational Linguistics
教　授　辻　井　潤　一

それでは機械翻訳における知識処理ということでお話したいと思います。現在機械翻訳の方は言語の構造に注目して翻訳を進めていこうという動きがあります。私の予稿ではリングィスティックスベースドＭＴという言葉で呼んだのですが、言語学を基本的な理論的枠組みとして、翻訳システムを作っていこう、そういう動きと、それから、言語の構造というよりも、むしろ言語が表現しているもの、あるいはそれを理解することによって翻訳を進めていこうと。私の予稿ではそれをナレッジベースドＭＴというふうに呼んでいるのですが、そういう２つの動きと、それとは全く別に、先程の講演にもありましたように、人間の翻訳家が作り出した翻訳のコーパス、テキスト。それを基本的な翻訳をするときの枠組みとして使おうと。その中で、例えばそれを例用として使うという意味でイグザンプルベースドＭＴですとか、あるいは実際のテキストに見られる統計的性質を抜きだして、それによって翻訳しようというマスターテスティックベースドＭＴ、そういう枠組みが現在提案されているわけです。私の予稿のほうでは、そういう４つの枠組みが「機械翻訳と知識という観点からどういうふうに整理できるか」ということを整理して書いております。今日の話は、それを少し見方を変えて、この会議の目的である「大規模知識ベース」を作るときに、我々が機械翻訳でやっていたようなことが、どういうふうに役に立つか、あるいは、自然言語処理あるいは機械翻訳、特に機械翻訳の方から今回の会議のテーマに対してどのようなコントリビューションができるかという観点から少し整理し直してお話したいと思います。

午前中の横井さんのお話ですとか、Yorick先生、Susan先生の話にありましたように、ナレッジベースと、テキストベースの２つをどういうふうに噛み合わせるかというのが現在の１つの研究の焦点になっているかと思うんですが、それを、私なりに一旦整理してみますと、一般的に使えるような、いろんな目的によってシェアできるようなナレッジベースを作りましょうと言ったときに、ナレッジというのは非常に抽象的な物なわけです。「ここに知識がありますよ」と言っても、「それを実際に書きなさい」というふうに言われると人によって違うふうに捉えているし、違う知識の記述ができてしまう。あるいは、

目的によっても、同じ分野の知識を書いていても、違った知識の記述ができてしまう。そういう意味では、知識というのは現在のところ人間の直観だとか、あるいは人間のその分野に関する洞察みたいなものを中心にして作っていかれるわけです。それだけで「大規模な知識ベースを作りましょう」ということをやってしまうと、実際には人による個人差だとか分野による個人差というのがものすごくでてきてしまって、系統的に大きな知識ベースというものを作り上げていくことができない、そういう実際の困難が出てくるというのが１つの大きい問題だろうというふうに考えるわけです。

そこで今日の午前中のお話ですとか午後のお話にあったように、知識に迫るには、結局もう一回言葉の方に戻って、言葉の方から知識というのをアプローチせざるを得ない。言葉というのは我々人間の持っている一番ユニバーサルな、知識を表現するときの手段ですから、言葉を調べることによって一応、抽象的な存在である知識というものがどういう構造を持っているのかとか、どういう特質を持っているのかというのが洗い出されるであろう。だから言葉から迫りましょう、ということになるわけです。ところが言葉の方も実際には、自然言語処理の研究をやっている人だと同意されると思うんですが、シンタックスの所までは一応安定した理論ができているんですが、シンタックスから今度セマンティックス、つまり、「言葉が何を表現しているか」というところになると、また個人個人の主観性がものすごく強く出てきてしまって、安定した、意味に関する理論ができない。そういう意味では、知識の問題を言葉の問題に置き換えても、結局言葉のところで同じような抽象的な問題に出くわしてしまうわけです。

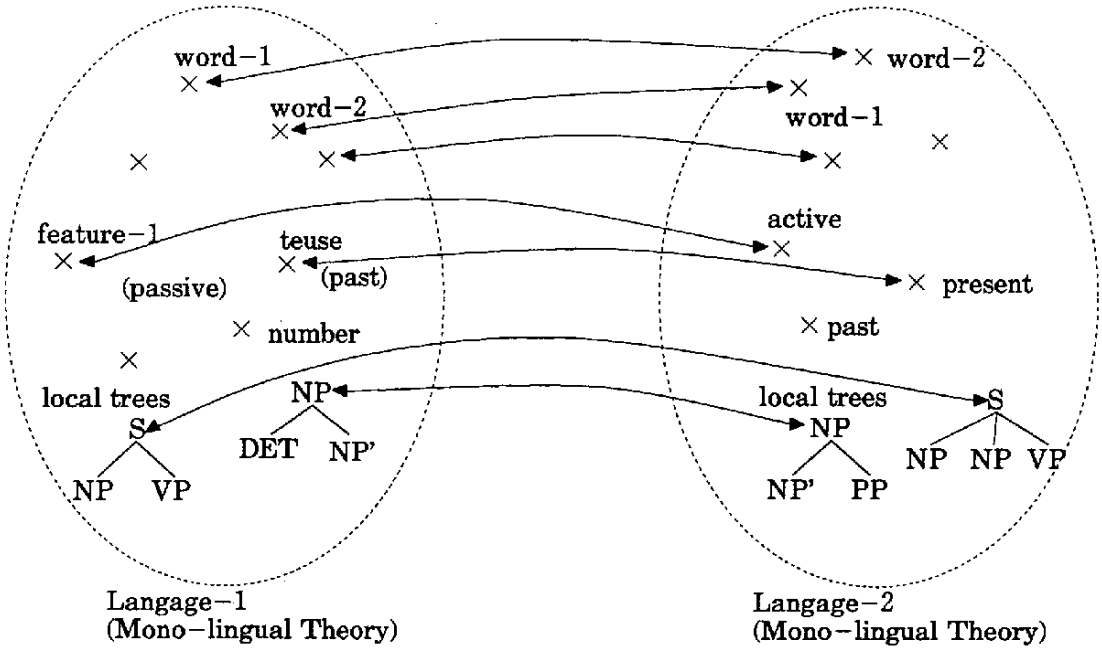
そこで、今日のSusan先生の話ですとか、Yorick先生の話にありましたように、言葉の方から知識に迫っていくときでも、結局具体的なデータから出発して、言葉から知識に迫るときに、「言葉が実際にどういうふうに使われているのか」あるいは「ある特定の表現がどういう場面で使われて、だから実際にはそれが１つの意味の単位を成している」とか、ということを、抽象的に議論していくのではなくて、現在計算機の中で使える言語データを中心にして、人間が言葉をどう使っているかということを調べることによって知識というものに迫っていく、そしてそれが多分大きな知識ベースというような、つまり抽象的な対象である知識というものを大規模に作っていこうというときの、一番安定した方法論ではないか、つまり、実際に使われている言葉を、「どういうふうに使われているか」ということを調べていくことによって、知識というものに迫って行きましょうと。で、そういう意味では、このテキスト、あるいは事象、百科事典というものが、そういう人間のインティヴィションとか人間の洞察によってやっていたようなことに対して、ある種のシステマティックな、あるアプローチの方法論を与えてくれるのではないかという、それが多分、言葉と知識の結びつきが、この会議の１つの大きなテーマになっている理由であろうかと思います。

そこで機械翻訳ですが、機械翻訳の方はさらにこれがつけ加わるわけですね。つまり、僕等は言葉と知識というのが一応対応しているように思うんだけども、どうも英語と日本語とフランス語と中国語というのを調べてみると、必ずしもそれ程綺麗な対応がない。つまり、幾つかの言葉を調べてみることによって、言語が、実際の知識だとか僕等が意味だとかと思っているものとどういうふうに結びついているかというのがもっと相対的にみられると、もっと客観的に、１つの言葉に固着することなく見られる。

そういう意味では、Yorick先生が言ってましたように、今までのナレッジベースの方の概念というのがみんな英語の言葉になっていたんですけれども、それを幾つかの言葉を比べることによって、僕等の概念の姿というのがもっと立体的に把握できるんではないか、というのが多分機械翻訳の方から大規模知識ベースというものの構築にコントリビュートできる一番大きなポイントであろうかと思います。実際機械翻訳の方で知識が要るというのは、私の予稿の方でどういう形で要るかということが整理してありますので、今回はそういう観点から少しお話してみたいと思います。

そうしますと、結局言葉と知識がどういうふうに結びついているかということを具体的に考えてみないと駄目だということになると思うんですが、そこを、我々が機械翻訳の研究を通じてどういうふうに見ているかということを少しお話したいと思います。

まず、その知識と言葉の結びつきに行く前に、先程言いました、機械翻訳の中の１つの立場、つまり「言語の構造を中心にして翻訳して行きましょう」、私の予稿の方ではリングィスティックスベースドＭＴという立場で呼んでいる機械翻訳の考え方が、どういう立場で翻訳を見ているのかというのを少し見ておきたいと思うんですが、基本的な立場というのは非常に単純でして、ここに日本語に関する言語学の理論がある。こちらの方に英語に関する言語学の理論がある。だから、これが日本語の理論、こちらが英語の理論、というふうにしますと、各々の理論は、「その言葉がどういう部分的な構造からでき



−The dominant Layer : Linguistic Level
−Units of Translation : Determined by Mono−Lingual Theories

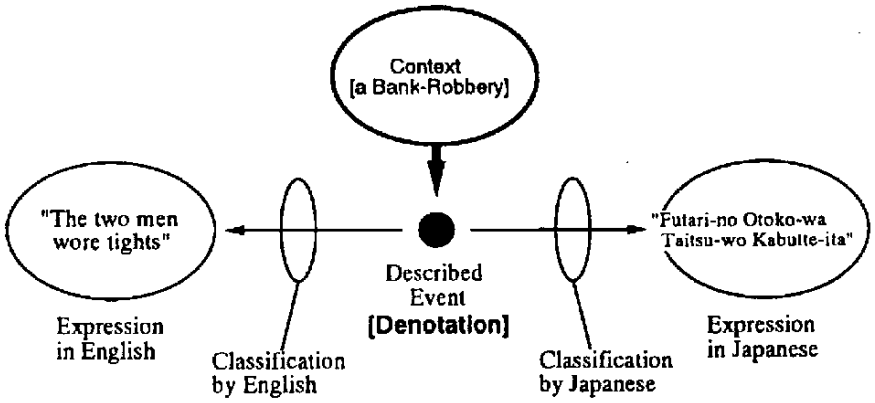
ＯＨＰ－１

ているか」ということを実際に調べ上げるわけです。それが組み合わさって我々の言葉というのを作っている。だから、日本語だったら日本語というものが持っている基本的な規則性を要素に分解して、それでもって理論を作る。英語だったら英語という言語が持っている規則性を要素に分解して、その要素の組み合わせでもって英語の世界というものを説明する。そういう２つの理論がありますと、今度は理論の間で結びつきをやればいい。例えば、日本語のこういう構造、要素的な構造に対しては英語のこういう要素的な構造が対応している。日本語のこういう単語に対しては、英語のこういう単語が対応している。日本語を作っている要素的なものと英語を作っている要素的なものの個別的な対応をつけてやると、２つの間の翻訳ができる。そういう発想で言語学の方からの翻訳システムというのが構成されてきたわけです。

実際そういうやり方でやっていきますと、結局意味の部分がすっぽり抜けちゃっているわけです。つまり、言葉の構造でもって、要素の対応をつけることをしますから、そういうやり方の研究の基本的な欠陥というのがいくつかでてきてまして、その１つは、今までは英語だったら英語、日本語だったら日本語のときには、英語の基本的な単位は何かというと英語の単語がある。日本語の基本的な単位は何かというと日本語の単語がある。その２つ同志が対応している。というふうな感じで考えていたんですが、どうも翻訳の単位とは、それほど小さくはなく、よく「イディオムだとか、フレーズで対応する」というのを言いますけれども、翻訳の単位というのがもっと大きいんじゃないかという話があるわけです。そういう問題を実際に系統的に考えて行こうと思うと、どうも１つの言語の構造だけを議論していてもはじまらなくて、言語が表現しているものに立ち入って、そこで翻訳というものを考えざるを得ないということになるかと思います。

実際にどういう問題があらわれてくるかと言いますと、２つありまして、ここで、少しあとで説明しますが、１つは翻訳というのが非常にコンテクストにディペンドしてしまう。だから、ある１つの表現を取り出してくると、その表現が相手の言葉に行くときに幾つもに翻訳し分けないと駄目で、その翻訳をし分けるのが、実はその部分だけを見ていても翻訳できなくて、結局周りの状況を見ないと翻訳できない。それがコンテクストディペンデンシーという問題です。それからもう１つは、先程言いかけました、「翻訳の単位」という問題なんですが、フレーズだとか少し大きな単位でどうも翻訳されていっているようで、それが、ここの「慣習化された対応」という言葉で呼んでいる問題です。それが実際にどういう問題であるかというのを、少し具体的な例で考えてみたいと思います。

先程も言いましたように、リングィスティックベースのＭＴからものを理解して、つまり言葉が表現しているもの、知識との対応でもって翻訳を考えて行こうというときの基本的なモデルというのは、ここで書いたような感じになるわけですね。つまり、真ん中に表現されているものがある。表現されているものを、一方の言語では何かある表現ですると、もう一方の言語では、また別の表現をするということになっているんですが、この真ん中の部分が、いまここで話題になっている「知識の領域」と言われているものです。あるいはこの会議でしばしば出ている言葉ですと、「オントロジー」と言われている部分です。つまり言葉とは無関係に、それによって表現されているもの。それからその両端に各々の言



OHP－2

葉というのがあるわけです。２つの言葉があります。その言葉とオントロジーとを結び付けるものとしてリンクがあるんですが、そのリンクというのが言葉の持っている言語的意味と言われている部分です。

辞書というのは、この言語的意味というものを実際にデータベースとして持つもので、ですから、ＥＤＲが実際に中間言語による辞書を作るというときには、こういうフレームワークをとったわけです。両端に言語があって、真ん中で、辞書というのがそれを結び付けているということになっているわけです。実際に今日のお話というのは、この結びつきというのが、実は非常に難しいというか、それほど単純ではないというのが、機械翻訳の方からの１つの結論と言いますか、あるいは逆に言いますと、今やられているＥＤＲの辞書をさらに次に一歩進めるための方法論を考えるときの問題を提供してくれるということになると思います。

そこでまず、このフレームワークが具体的にどういう事を指しているかということを考えるために、少し具体的な例を考えますと、例えばよく挙げられる例なんですが、「タイツを履く」という例です。それを英語の方では"wear tights"というふうに言っているわけです。これは機械翻訳の分野ではよく使われる例でして、実際には"wear"というのは、英語から日本語に翻訳するときによく問題を引き起こすものでして、例えば"wear shoes"と言われると、「靴を履く」。それから"wear hat"というと「帽子を被る」。また、"wear specs"というと「眼鏡をかける」というふうに、英語の方では同じ"wear"なんですけれども、日本語の方では、「履く」だとか「被る」だとか「身につける」だとか「かける」だとか、というふうに幾つにも別れちゃうわけです。どうして別れているかといいますと、"wear"の方が、単に人間があるものを身につけているということだけであるのに対して、日本語の動詞というのは全て「どこに身につけているのか」によって使い分けないと駄目なんですね。だからたとえば、頭に「身につけている」場合ですと「被る」と言わないと駄目ですし、ここに「身につけている」場合ですと、「履く」という言葉を使いますし、ここにつけている場合ですと「かける」という言葉を使う。だから身体部位によってみんな、日本語の方は言葉が変わってしまうわけですね。

それは、もう少し図式的に言いますと、英語の方が単純に人間とストッキングが、ある種の関係に

ひっついているということだけを言っているのに対して、日本語の場合ですと、それが、実際にこの人がここにつけている場合ですと「ストッキングを履く」という話になりますし、頭の方につけている場合ですと、実は同じストッキングでも「ストッキングを被る」というふうに言い分けないと駄目なわけですね。ですから、"wear"とある特定の名詞の組み合わせで翻訳するというと大体うまく行くんだけども、実際にはストッキングの場合でも、銀行強盗なんかの場合ですと「２人の銀行強盗が入ってきて、黒いタイツを履いていた」と言うと何か気持ちの悪い状況ですけれど、実際にはタイツを「被っていた」という話になるわけですね。ですから、細かな、この"wear"とある名詞の関係だけ見ていても実はうまく訳し分けられなくて、全体の状況の中でどういうことが起こっているのかということを見極めないと実際の翻訳はできないという話がでてくるわけです。ここでは何が起こっているかというと、"wear"というのが実は、"wear"とか日本語の「被る」とか「かける」とか何とかというのが、言葉の意味としては非常に柔らかい意味を持っているわけですね。どこにつけるかということだけを見て、それによって「被る」が使えるとか「かける」が使えるとか何とかが使えるという、そういう状況になっていて、実際に現場でどういうことが起こっているかということがわかると、どの言葉を使っていいかがわかる。

ですから、言葉と、それが表現しているものというのは非常に柔軟な関係である。そういう条件が満たされていれば、その単語が使えるという状況にあるということです。ですから、言葉と、それが表現している対象との間に、少しそういう柔らかいマッピングの構造があるというのが１つの難しさになっているのではないかと思います。

それからもう１つは、これも機械翻訳の分野ではよく使われる例なんですが、例えばバスとか電車なんかに乗りまして、切符を持っているわけですね。これはMartin Kayeという人が使った例なんですが、バスとか電車に乗って切符を持っていく、その切符を機械にカシャンと入れて印字をしてもらう、と、そういう動作があるわけですね。その動作を、例えば日本語だと多分「パンチを入れる」だとか「切符を切る」だとかという言葉で言い表すと思うんですが、ドイツ語のほうだと、「その切符を有効にする」、あるいはフランス語のほうだと、「その切符を無効にする」とか、そう言葉で言うわけですね。つまり、ある切符に、「パンチを入れる」ことによって「有効になる」というふうに言う場合と、それからもう一方の言語だと、その切符は１回パンチを入れられると２度と使えないから「切符を無効にする」という言葉を使っちゃうわけですね。日本語の方は、切符が「有効になる」とか「無効になる」とかということとは全く無関係に「切符にパンチを入れる」という言葉を使う。

ここでは何が起こっているかというと、表現されていることは１つなんですね。つまり、なんかある動作があると、その動作を１つの言語は「切符が有効になるか無効になるか」という観点で言葉に換える。もう一方の言語は、日本語のように「どういう動作をしているか」ということで、言語に換える。例えば「パンチを入れる」とか「機械に差し込む」とかという感じの言語化をするわけですね。そうすると言葉の意味のレベルでは、例えばドイツ語の方だと「切符を有効にする」、フランス語の方だと「切符を無効にする」、全く翻訳にはなっていないわけですね。全く正反対のことを言っているという

状況になる。あるいは、「パンチを入れる」みたいに、「パンチを入れる」と「有効にする」というのは全然違う概念ですから、それが、言葉の意味だけを見ていると全く繋がらないのに、実はそれが翻訳の関係で繋がっている。それがなぜ翻訳で繋がるかというと、その言葉は、その言葉を使うときにはその行為はどういうふうに表現しないと駄目かというのが決まっている、という話になるわけです。

実はどういうことかと言うと、こういう、知識と結びつく、例えばある動作、バスだったらバスの中で「どういう動作をする」というのが１つの知識の単位を構成しているとしたら、知識の中でも１つの動作を表現しているとしたら、単語というのが直接そういう知識の１つの単位に対応しているのではなくて、こういうフレーズが１つの知識の単位に対応している、ということになるわけです。それが翻訳のレベルでは、そういう知識のレイヤーを通って繋がっていますから、表面上全く言葉の意味のレベルでは翻訳にならないようなことが、翻訳の関係で繋がってしまう、という話になります。こういう話というのは、結局翻訳のユニットというのが、実は知識のユニットと非常に関係していて、しかもその知識のユニットと、あるいは翻訳のユニットというのは単語のユニットではなくて、知識の中で１つの存在を占めているような単位が１つの翻訳の単位になっているということになります。

これからちょっと少し技術的な話になりますので飛ばしますが、翻訳の単位は、２つの種類の翻訳の単位というのがあって、それがイグザンプルベースドとかナレッジベースドとかリングィスティックベースドという、違う機械翻訳の枠組みで、それぞれ違ったふうに捉えられていて、実際の将来の機械翻訳のシステムとしては、この違った単位の翻訳の単位というのをもう少し真剣に考えて行く必要があるだろうというふうに考えています。

それから、機械翻訳とは離れて、知識と機械翻訳という観点からすると、言語の単位というものと、知識の単位の相互関係みたいなものをもっと真剣に考えて行く必要があるだろうというふうに思っています。それから、もう１つの観点は、もしそういうふうに、ある１つの言葉がある特定の見方で現象を見ているということであるとしたら、そういう話と言うのは機械が実際にダイナミックに処理をしても出てこないことなんですね。つまり日本語は、日本語の世界で物事をそういうふうに見ているということがもう決まっているわけですから、それは、機械がそこで理解をして、あるいは推論をして「実は日本語の方ではこういうふうに見方を換えるんだよ」ということを出してくるというのは、実は非常に難しい作業をすることになるんじゃないか。だから、機械翻訳の中に知識を入れてくるとしたら、これまで考えていたような、理解に基づく、理解したあとでのダイナミックなプロセスではなくて、今日のSusanの話だとか、Yorickの話にあったように、むしろどういうふうな言語がどういう風に知識の単位を捉えていて、それが２つの言語の間でどういう差があるのかというのを事前に整理するような段階、つまり、テキストから知識に結び付けるのを、整理する段階ですね、知識を入れていく段階のほうにむしろ大きな問題があって、この会議のテーマである、ナレッジアクイズィションとか、あるいはテキストからナレッジに結び付けるときのいろんなツールだとか、そういうものがむしろ機械翻訳においては大きな役割を果たすようになるんではないか、というふうに考えています。

そこで現在の機械翻訳の枠組みの話を少ししたかったんですが、そこはちょっと時間がないようなの

で省略しますけれども、１つは、言語と知識の結びつきというのを「言語とワールド」というふうに書いていますが、ここの話のコンテクストの中では「言語と知識」というふうに読み換えてもらってもいいですが、ここがどういうふうに柔軟に結びついているかというのを考えるときに、１つの考え方は、我々の認知の枠組みを形作っているような基本的なもの、それを私は「空間」というものと「時間」だと思うんですけれども、「空間」と「時間」というものの捉え方が言葉によってどういうふうにかわっているかというのを少し考えてみて、そのなかで、翻訳と知識というものがどういうふうに関係しているのかということを少し研究を進めているところです。

具体的な例を少し挙げて、現在の機械翻訳で知識を結び付けるというのがどういうことに対応しているかという例を少し挙げたいんですが、ちょっと小さな字で少し読みにくいかも知れませんが、ここで書いているのは、英語のセンテンスでして、"The students will be examined in mathematics next month"という、非常に単純な例文なわけです。この例というのは「学生さんが来月数学の試験をされますよ」という感じの文章なんですが、実はこういう単純な文章でも時間の関係が複雑に絡んでいて、それによって、ある言語によっては翻訳の仕方が変わってしまうんですね。

例えば、これ日本語で翻訳するときには全然変わらないんですが、ギリシァ語に翻訳、ギリシァ語というのは僕もよく知らないんですが、ギリシァ語に翻訳しようと思うと、それが２つに分かれる。それがなぜ分かれるかというと、先程の文章、つまり「来月学生さんが数学の試験をされる」という、そういう文章なんですが、その理解の仕方で２つ違った状況が考えられるんですね。１つの状況は、「学生さんが集められて、１回のセッションで学生全員が試験される」、そういう状況が１つです。あと、そういう試験の仕方ではなくて、例えば１０人位の学生が月曜日に試験をされて、その次の木曜日には３人位の学生がまた試験されて、次のときにはまた７人位の学生が試験をされる。つまり、「試験される」という事態が１回しか起こらないか、あるいは不定回起こるか、つまり、１回きりの現象なのか、何回か起こる現象なのかというのを、ギリシァ語のほうでは区別するわけです。１回しか起こらないと、「完了時制」というのを使わないといけないのですが、複数回起こる場合だと、「不完了」の時制を使う、というふうに分かれてしまいます。言葉によって少し時間の捉え方、時間とイベントの考え方も少し違うわけです。

そこで、そういう問題を本当にやろうと思うと、実はいくつかの、ここにありますように非常によく似た例なんですが、例えば、「ジョンは来年卒業しますよ」とか、「ジョンは来週卒業しますよ」とか、「次の月曜日ジョンは早く起きる」とか、「ジョンは来月早く起きる」とかいうのが、全て違って翻訳されないと駄目なんです。例えば、「月曜日に早く起きる」と言うと、１回きりの現象ですが、「次の月曜日に早く起きる」と言うと、１日しかないわけですから、１回しか起こらない。これが「来月早く起きる」というと、３０日ほど日がありますから、毎回早く起きるということになりますから、不定回起こっているという話になりますね。それから、例えばこういうふうに「来週ジョンは月曜日にどこどこに行く」というのと、「来月ジョンは月曜日にどこどこに行く」というのでも、例えば月曜日が、週の間には１回しかないんだけども、月の間には何回もあるから、また不定回起こる。そういうふうに、言

葉と時間をどういうふうに捉えるかというのが分らないとうまく翻訳できない。時間をうまく捉えるためには、今度は例えば「月」だとか「年」だとかというのが「実際に時間としてどういう構造をしているか」とか、あるいは、「ある特定のイベントがどういう周期で起こるか」ということを知ってないとこういう場合はうまく訳せないという話になります。

ですから、我々の翻訳のモデルでは、この翻訳するのに一応３つ程違ったタイプに分けてまして、違ったタイプの知識というふうに分けてまして、１つのタイプの知識は言葉と時間のエンティティーを結び付けているような知識、例えば、「月曜日」というのは１つの時間の概念をリンクしてます。だから、言葉と知識の世界でもエンティティーを結び付ける単位というのが１つの知識の単位ですね。それからもう１つはそういう結び付けられた知識のレベルで、今度は出来事に関する知識、例えば「卒業する」というのは一応一生のうちに１回しか起こらないとか、あるいは「起きる」というのは１日のうちに１回しか起こらないとか、あるいは「教会に行く」というのは毎日起こるようなことだとか、今度は出来事に関する知識ですね。それは、言葉に関する知識ではなくて、言葉がある出来事と結びついて、今度はその出来事に関する知識というのがまた別にある。それからその次に、一番微妙なのは、今度はそういう出来事の知識から推論されることをまた言葉の世界に引き戻してくるような知識、つまり、複数回それが起こっているんだったら相手の言語では何に翻訳しないと駄目ですよというようなことを結び付けている知識ですね。だから結局機械翻訳というのを、知識と、言葉と、それから知識から導かれる情報によって言葉がどう使われるかというのを結び付けるような、そういう３つの構成要素によって機械翻訳をやっていく、ということを現在考えています。

そこで、結論なんですが、この会議で議論されていますように、言葉を手掛かりにしてそこから知識に入っていくというのは、多分大規模な知識にアプローチするほとんど唯一の安定した手法だということは多分合意できると思うんですね。ただ、知識と言語というのは、それほどストレートフォワードには繋がっていない。だから、そういう意味で研究のチャレンジとしては、言葉の使用のあり方を観察することによって、そこから知識という、言葉とは一応無関係の世界の構造を見つけだしていく。そのときに、コーパスを中心にしたマニピュレーションですね。そこでシステマティックな方法論が作れるのかどうか、というのは研究として非常にチャレンジングな話だと思います。しかも、こういう大規模な知識ベースを成功させるというときのキーになるようなテクノロジーじゃないかというふうに考えています。

それからもう１つ、そういうことを考えるときの、私にとって一番重要な問題だと思われるのは、知識の単位というのと、言葉の単位というのがまたずれているわけですね。先程言いましたように、例えば"Slot a ticket"というのが１つの行為を表している表現になっているというふうに、必ずしも単語が１つの単位ではありませんから、そういう言葉の単位というものと知識の単位というものが、どういうふうに相互に関係しているのかというのを、やはりコーパスの中からうまく見つけていく、というのも１つの大きなテクニックになるでしょうし、それからもう１つは、今回の議論が少し分りにくくなっているというのは、言葉に関する知識なのか、あるいは言葉が表現しているものに関する知識なのか、

例えば、辞書という話を強調しますと、それは言葉に関する知識というふうに見えるのですが、言葉とは無関係に、言葉が表現しているものに関する知識になっていくわけですね。そういう、言葉の意味ということと、それから言葉が表現しているもの、というものをまたうまく分けないと、"wear"の例にありますように、言葉の意味としてはもっと柔らかなもので、それが実際に表現しているものは実際の客観世界の中での１つのエンティティーを成している。そこらをうまく捉えられないんじゃないかというふうに思います。

ですから、結論としてもう１回言いますと、多分言葉を中心にして、知識に迫っていくというのは、非常にいい方法論だろうと、その時に注意しなければならないのは、言葉と知識の関係というのがかなり複雑な関係をしていますから、そこのところに１つ大きなテクノロジーが要るのではないかということです。以上で終わります。

座長：

それでは、会場からもし質問とかありましたら、受けたいと思います…。辻井先生の講演だけでなく、このセッション全体に関するご質問でも結構です…。よろしいでしょうか…。

それでは、予定の時間も少し過ぎましたので、このセッションを終わらせて頂きます。もう一度、御講演の方々に拍手をお願いします。どうも有り難うございました。

4. セッションIII

知 識 処 理

4. セッションIII：知識処理

4.1 座長挨拶

東京理科大学 理工学部経営工学科
教 授　溝 口 文 雄

おはようございます。日本語でやらせて頂きますので、少しゆっくり話すかもしれません。このセッションは知識処理ということで、講師の先生方はいずれも１５年から２０年の、いろいろと知識表現を経験された先生方で、そういういろいろの経験を通じて現在いろいろと何を考えておられるか、というようなことが、ディスカッションされるんじゃないかと思います。

この中で、おそらくキーワードは、「知識の再利用」とか、それから「オントロジー」という言葉が出てくると思うんですけれども、ちょっと「オントロジー」という言葉が、馴染みが、もしかするとないかもしれませんが、その辺の話もあると思います。知識を表現して獲得して再利用するというのは長年の課題でして、今日１日では解決できないんですが、まあ、今後の動向を、こういう講演を通じて皆様がシェアし合えればいいんじゃないかと、まあそんな感じがするわけです。

4.2 講　演

(1)「大規模知識ベースの共有法」

座　長：

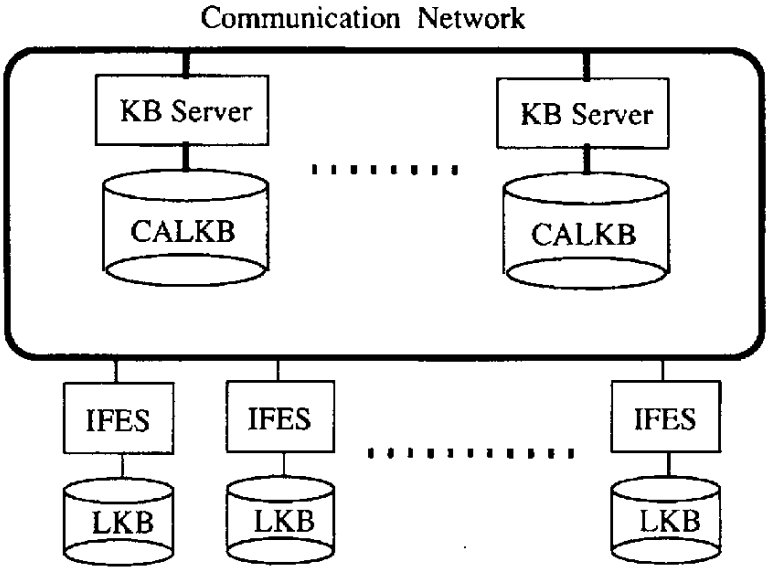
最初の講演は、日本語のタイトルでは「大規模知識ベースの共有法」ということですが、英語のタイトルでは"How can people share large knowledge base"というタイトルで、東京大学の大須賀先生でございます。大須賀先生は、'５７年に東京大学工学部を卒業されまして、富士通精密機械というところに勤務されてから、再び東京大学に戻られて、１９８１年に東京大学工学部の教授で、'８７年から東京大学先端科学技術研究センターの教授でございます。先生はデータベース、あるいはそれの拡張での知的データベース、とくにCAD関係についての表現と、そういう関係の仕事がたくさんおありでございます。多分今日の話も、そういうことに関連した話が発表されるんじゃないかと思いますので、よろしくお願い致します。

東京大学 先端科学技術研究センター
教　授　　大須賀　節　雄

おはようございます。ご紹介頂きました東京大学の大須賀でございます。時間がありませんので早速始めさせて頂きます。今日の私の講演は、How can people share large Knowledge baseというタイトルでやらせて頂きます。私の今日の話の趣旨は、大規模知識ベースを構築するということが、１つの社会的なイベントである以上、そこに蓄えられる知識は非常に有用度の高いものでなくてはいけない、ということでプラクティカルな立場から大規模知識ベースシステムがいかにあるべきかということについて、私の考え方をお話させて頂きたいと思います。

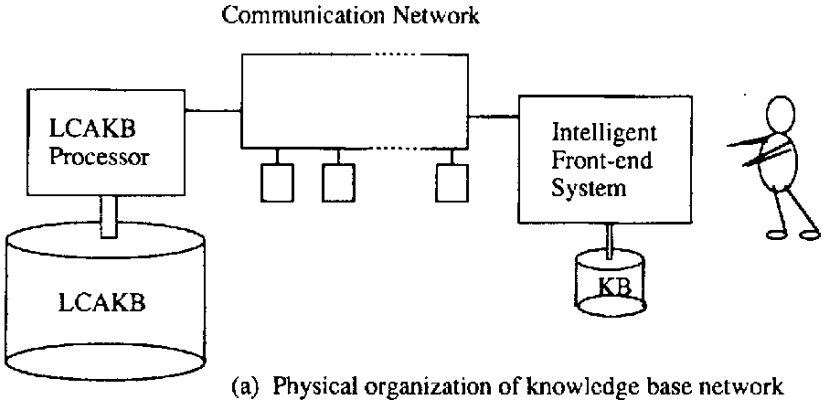
この議論の基本になりますシステムは、図１のようなものとして考えております。図の一番上に共有される大規模知識ベースがありまして、多数のインテリジェント端末によってシェアされるという状況です。システム全体の動作のバランスを保つためには、大規模知識ベース自身が複雑な問題解決を行なうといったようなことはできるだけ避けたい。それは多数のインテリジェント端末からの要求を処理するためには、時間のかかる問題解決をするのではなくて、インテリジェント端末側で行なわれる、非常に高度の情報処理を行なう過程に於いて必要となる知識をできるだけ大規模知識ベース側からサプライするという立場でものを考えたいと思います。

したがって、図２のような見方をします。これは図１を書き直したものですけれども、上の図が物理的な構成を表しています。このインテリジェント端末を以下では、インテリジェント・システムと呼ばせて頂きますが、ユーザーはそれを使って、高度の問題解決を行います。高度の問題解決というときに

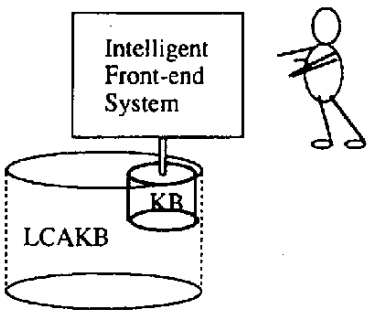


CALKB: Commonly Accessible Large Knowledge Base
LKB : Local Knowledge Base
IFES : Intelligent Front-End System

図1 : Configuration of large scale knowledge based systems



(a) Physical organization of knowledge base network



(b) Logical organization of knowledge base network

図2 : Physical organization and logical organization

は、私どもはコンピュータ自身がかなり自律的な問題解決機能を持っているということを前提として話を進めたいと思います。

このような状況のもとで大規模知識ベースは、実はロジカルには、個々のインテリジェント・システムがもつ局所的知識ベースと一体化したもの、少なくともユーザーからみた限りではこの大きな知識ベースがインテリジェント・システム内にある、図２の(b)の形で動いてくれるものでないと、実際には役に立たない、というふうに考えております。これを満たすように大規模知識ベースの中身を決めるためには、まずこのインテリジェント・システムがどういう構成をとり、どういう形で自律的な問題解決を可能にするか、ということについて議論しなくてはなりません。このことは勿論、従来のＡＩのシステムの、１つの目標でもありました。しかし現状においても、ＡＩシステムが、非常に高度の機能を発揮するまでには到っていません。我々はまず最初に実際的な場で役に立つような、自律的な問題解決能力をもつＡＩのシステムを早急に開発していかなくてはいけないと考えております。

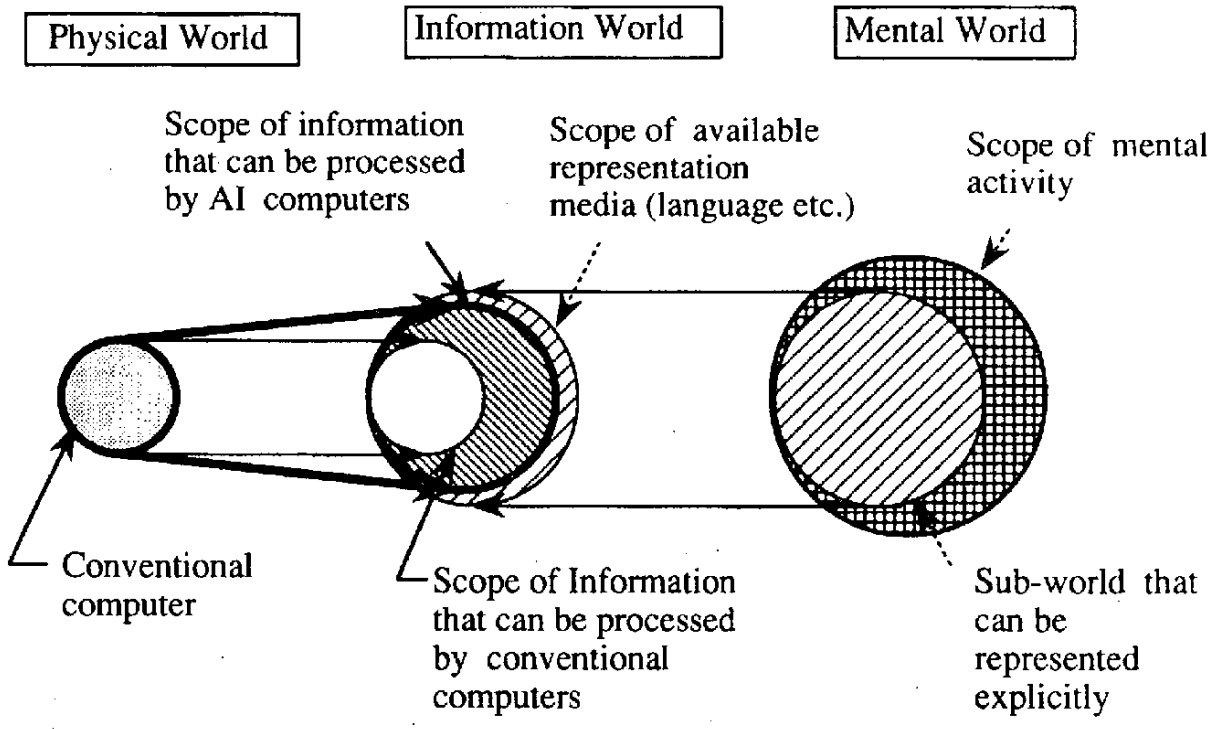


図３　：Physical world,information world and mental world

そのために何が必要かというと、情報とは一体何なのかということを我々がよく理解することです。そのためには、人間が行っている様々なアクティヴィティーを分析し、よりよく理解することが必要に思われます。ここで図３をお見せします。メンタル・ワールドと示してあるのは、人間の知的な世界、人間のアイディアとか思考とかというものを表すとしますと、情報とはこれを表現する何らかの媒体で

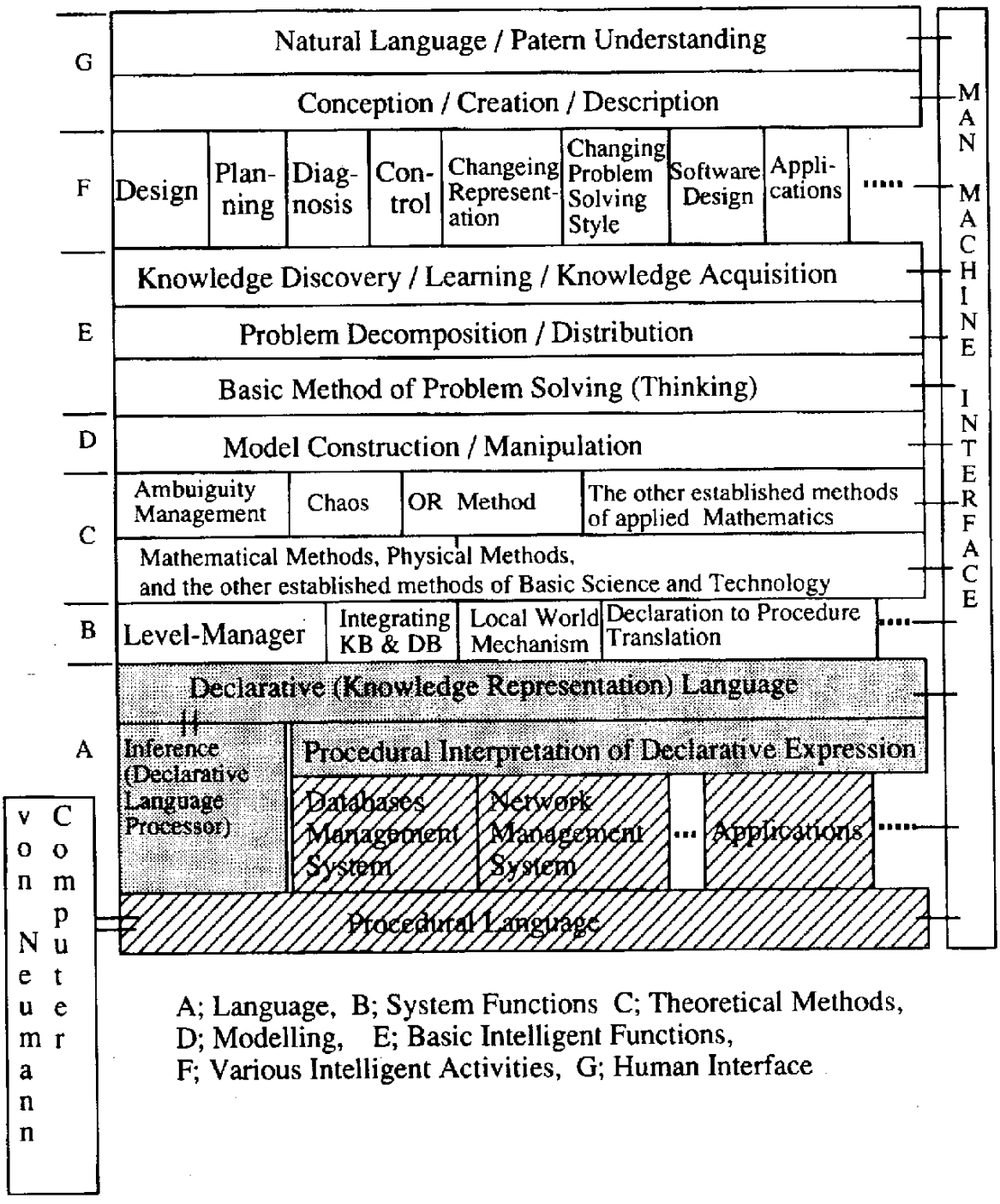
す。それは言葉であり、その言葉によって表される様々な方法論があります。この全体をインフォメーション・ワールドと示しています。人間は、このインフォメーション・ワールドを、長い時間かけて、長い歴史をかけて築き上げてきました。コンピュータというのは、ある時期に、物理的なものとして作られた特殊な機械です。その機械がこの情報の世界の一部分を扱うものとして今まで使われてまいりました。

しかしながら、今のコンピュータで扱われる情報の世界が、人間の持つ情報世界に比べて小さい為に、様々な問題が起こってくるわけです。したがってＡＩシステムを開発する際の目標は、機械が扱える情報の範囲を、図３にありますように、より大きくすることです。人間の情報の世界のかなりの部分をカヴァーできるような、そういう機械は作れないだろうか、ということです。そのためには、まず人間の情報世界、人間の知的アクティヴィティーを表す情報の世界を分析をして、それをフォーマルに表して、それを機械化する、機械的な実現を図る、というアプローチが必要なのではないだろうか、こういう考え方から、私どもは、人間の機能というものをいろいろな形で分析をしてきました。

図４はこの情報世界の一つ、新しいＡＩのシステムが備えるべき情報世界の、１つの提案です。この図で、Ｆと記した層には現実の世界における人間の様々なアクティヴィティーに対応する知的機能があります。デザインであるとかプランニングであるとかというものです。一番下は、言語層です。知的機能は言語で表されない限りは、実現されませんから、それを表す為の言語です。この層の一番下に従来のプログラミング言語に相当する言語を置いておきます。この言語は、コンピュータのハードウェアによって処理されます。しかしながら、この言語よって表される世界は非常に小さい世界です。言語を定義するということは、その言語によって表現される世界を定義するということと同じです。このプログラミング言語は、人間の解釈を通してはじめて意味を持ちますから、その制約のために、現在のプログラミング言語の表す世界は、情報世界全体に比べますと非常に小さい範囲になっています。

人間の行なっているアクティヴィティーを、分析していくと、より基本的な、いろいろな知的な機能に分かれてきます。このような分析自身が、１つの非常に重要な研究の目的でもありますが、それを表すことのできる新しい言語が必要です。これは、いわゆる知識表現言語と言われているものです。したがって知識表現言語というものは、その言語によって表現される世界が人間の行っている様々なアクティヴィティーを表現するだけの記述力をもったもの、そうでなければならないわけです。それをはっきりさせるために、我々はまずこの実際の機能レベルからトップダウンに、その仕事を分析していったわけです。ここにもありますように、人間のアクティヴィティーは多様です。この全てをもちろん分析し尽くすことはできませんけれども、現在私どもは、できるだけ多くの分野を分析しています。今日はその中で、特に代表的なものとして、デザインの問題を挙げたいと思います。

我々は、この人間のアクティヴィティーの中で、デザインという機能が特別な意味を持つものと考えています。それは情報世界の構造を表す為に大きな情報を与えてくれるからです。この議論に入る前に、言語について触れておきます。ここには少なくとも２種類の言語が必要です。手続き型言語はハードウェアで処理されますけれども、宣言型言語は手続き型言語によるプログラムで表された処理系に



A; Language, B; System Functions C; Theoretical Methods,
D; Modelling, E; Basic Intelligent Functions,
F; Various Intelligent Activities, G; Human Interface

図4：A proposed structure of information world

よって処理されます。この処理系が、いわゆる推論と言われているものです。もちろん、このレベルの言語の処理系をハードウェアで作ることも可能です。しかしながらこの新らしいシステムは、当然従来のソフトウェア資産というものを全部利用できるということが大前提になりますので、いずれにしても手続き型言語の処理メカニズムは持たなくてはいけないという意味で、ここでは、この推論機構は全部

ソフトウェアで表すという形で考えておきます。

さて知的と言われている、人間の様々なアクティヴィティーがレベルFの層にありますけれども、これを順次分析していこうというのが、我々の目的です。ここではその中で、デザインをとりあげていこうというわけです。デザインという行為は、より基本的な様々な機能に分割されていきます。デザインという仕事は、人間がやっている多くの仕事の中でも、非常に高度な仕事ですけれども、インテリジェント・システムが実際の場で役に立つようなデザインの支援をしようと思いますと、そのシステムには、少なくともこの、4種類の機能を持つことが要求されてくると思います。1つは、デザインは非常に創造的なプロセスですから、その創造的な過程を表現できなくてはいけないこと。それから第2には、コーペラティヴなワーク、共同作業ができるような環境を作っていくことです。共同作業をするということは、大きな仕事を小さな仕事に分割していって、最後にその分割した仕事をもう1回再統合する、といったプロセスを支援するということです。第3にデザインの定義そのものですが、「要求を満たすような構造を見つける」という問題を解決することです。これは非決定的な問題解決です。第4は生産に関わるものですが、今回は時間の都合で、3番目だけに話を限定させて頂きます。

この部分を形式化して表現致しますと、図5のような構造で表されます。処理の構造です。非決定的な問題解決を行うためには、まず対象のモデルを作ってみて、そのモデルを解析して、評価をしてみます。そして、それが要求に合うかどうかをチェックします。もしこれが目的に合わなければそのモデルを修正するという、このプロセスを繰り返すということが、デザイン、あるいは新しいものを発見するプロセス、あるいは非決定型の問題解決の基本的プロセスです。これがコンピュータの中できちんと表現されれば、そのコンピュータはこのスキームに従ってあるものを発見してくれる、構造を見つけ出してくれる、そういうことが期待できるわけです。これを実現するためには、少なくとも言語にいくつか

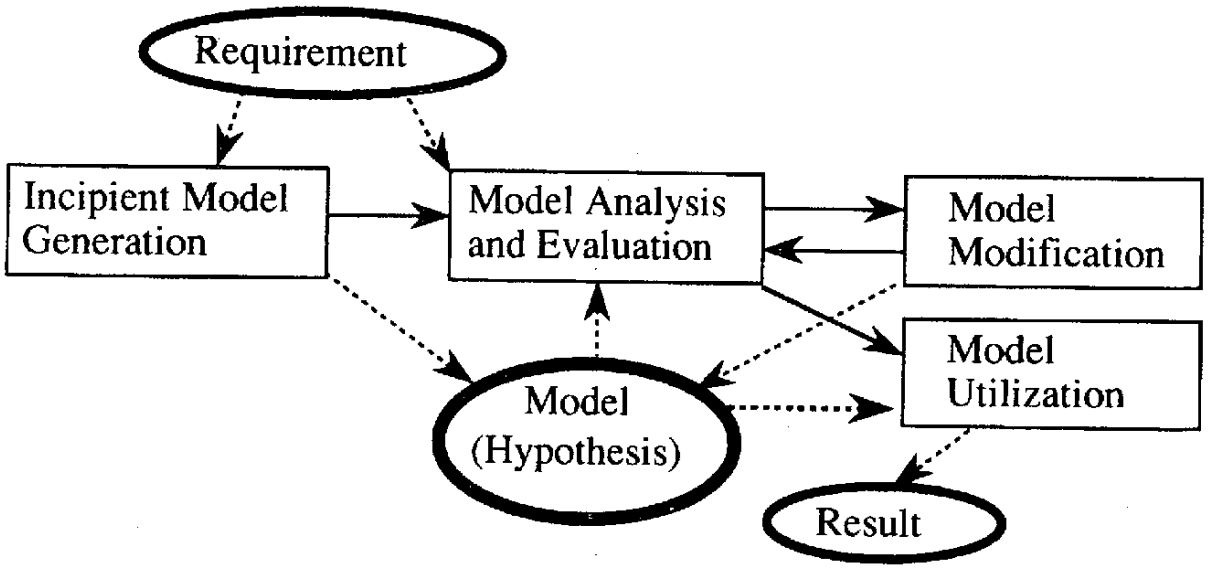


図5 ：Standard design process

の要求が課せられてきます。まず第一は少なくともモデルを表現できなくてはいけません。一般的にモデルというのは、非常に複雑な構造を持ちます。設計対象は、例えばビルディングであったり、飛行機であったりしますけれども、それは非常に大きくて複雑な構造を持ちます。それと同時に多様な機能を持つわけです。それは構造に伴って生じてくる機能です。したがって、対象モデルを表す為には、構造－機能関係をきちんと表せるような言語でなくてはいけない。また、設計というのはそれを変換する仕事ですから、そのモデルを変換するということ、これをルールの形で表す必要があります。これはモデルを含んだ形のルールです。さらに、こういうプロセスの構造を表し、且つ制御の部分を表すための言語機能が必要です。これにはメタのレベルの規則というものを必要としてきます。このメタのレベルの機能を表すためにはどういうふうにするかというのを簡単にお話致します。

図６は図５と全く同じ図ですけれども、モデルとその解析・評価、モデル修正など、オペレーションは全て、それぞれの知識ベースによって実現されるものとして示されています。同時に、枠の外にプロセスの表現と制御の部分があって、ここに戦略ルールベースがあります。この戦略ルールベースは、枠内の知識ベースを定義し、それらがいつ、どういう条件のもとでモデルに適用されるか、ということを決めるルールベースです。これを表すには、図７のような機構が実現できればよいわけです。上の図は、領域知識が単純に、大きな知識ベースに集められている状況を示しています。これをメタのレベルで下

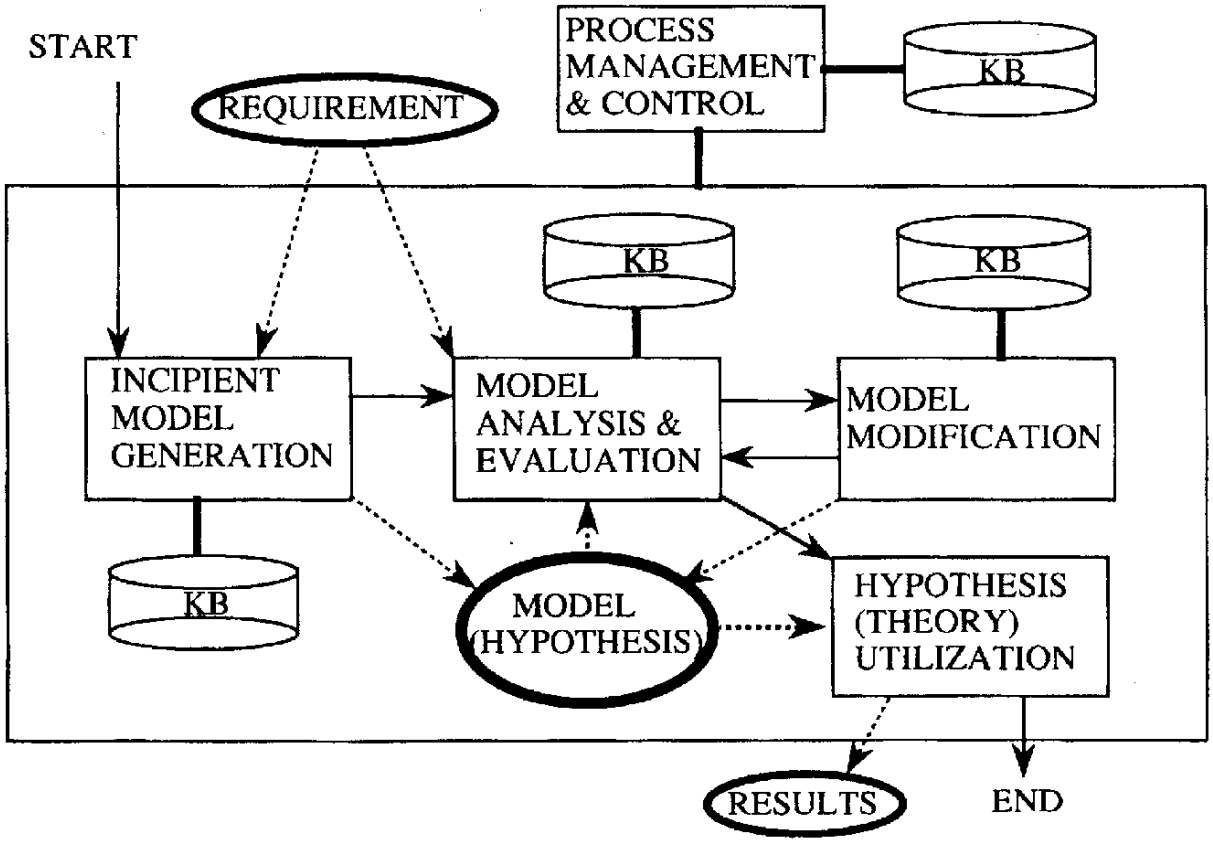


図６：A knowledge based design process

図のような構造として表してやることができます。この構造をさらに精密な構造に表し、この各部分はそれぞれ別個の領域知識源を表現するようにし、且つ、その各知識源がいつ活性化されるか、ということを表現することによって、制御が表現できます。

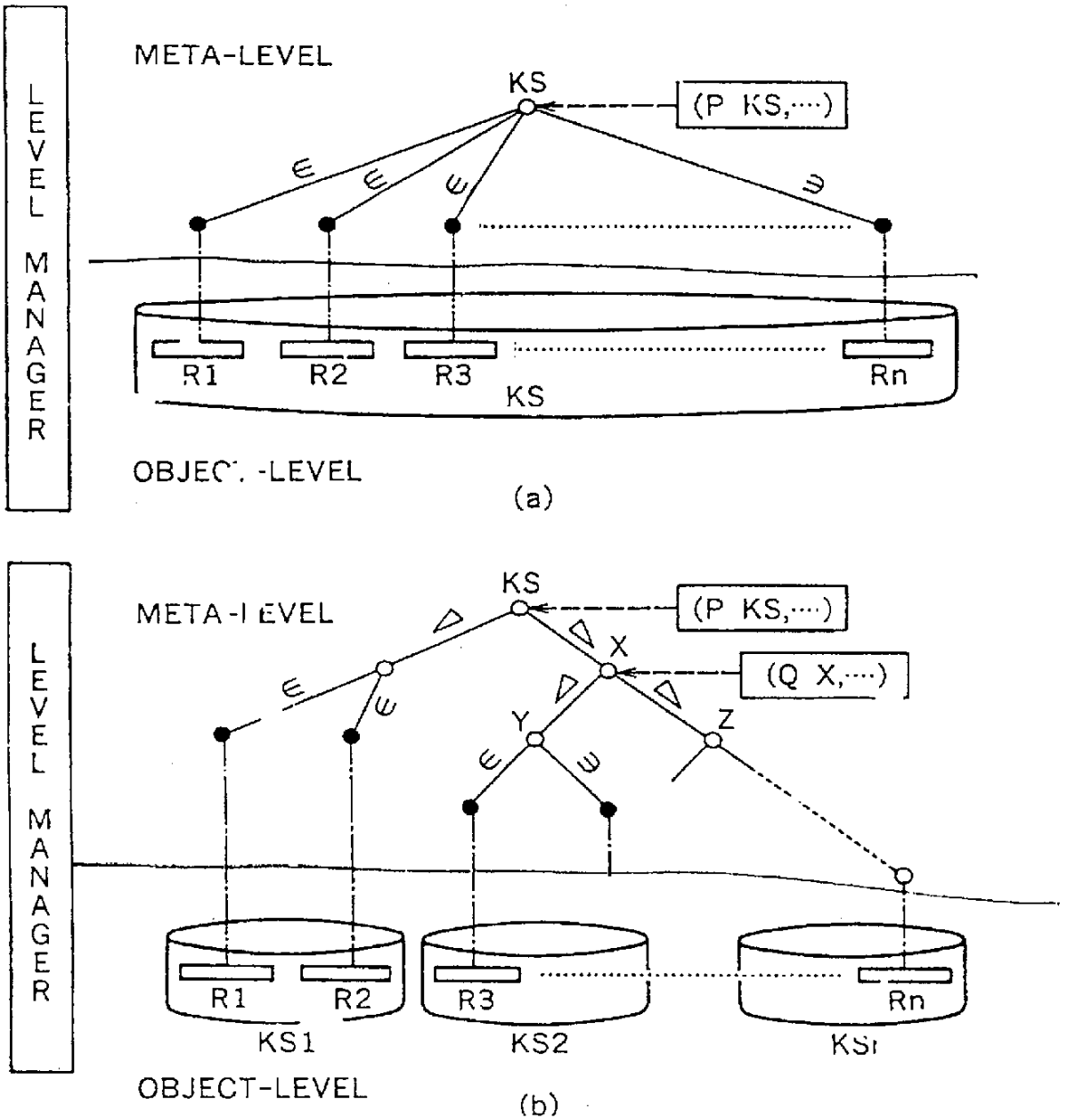


図7：Classification and struetuing of object rules

そういう表現するための言語として、私どもは図8のような言語開発をしてまいりました。これは、述語論理を基本とした言語です。まず一番基本形として通常の述語から出発します。この場合には項は、

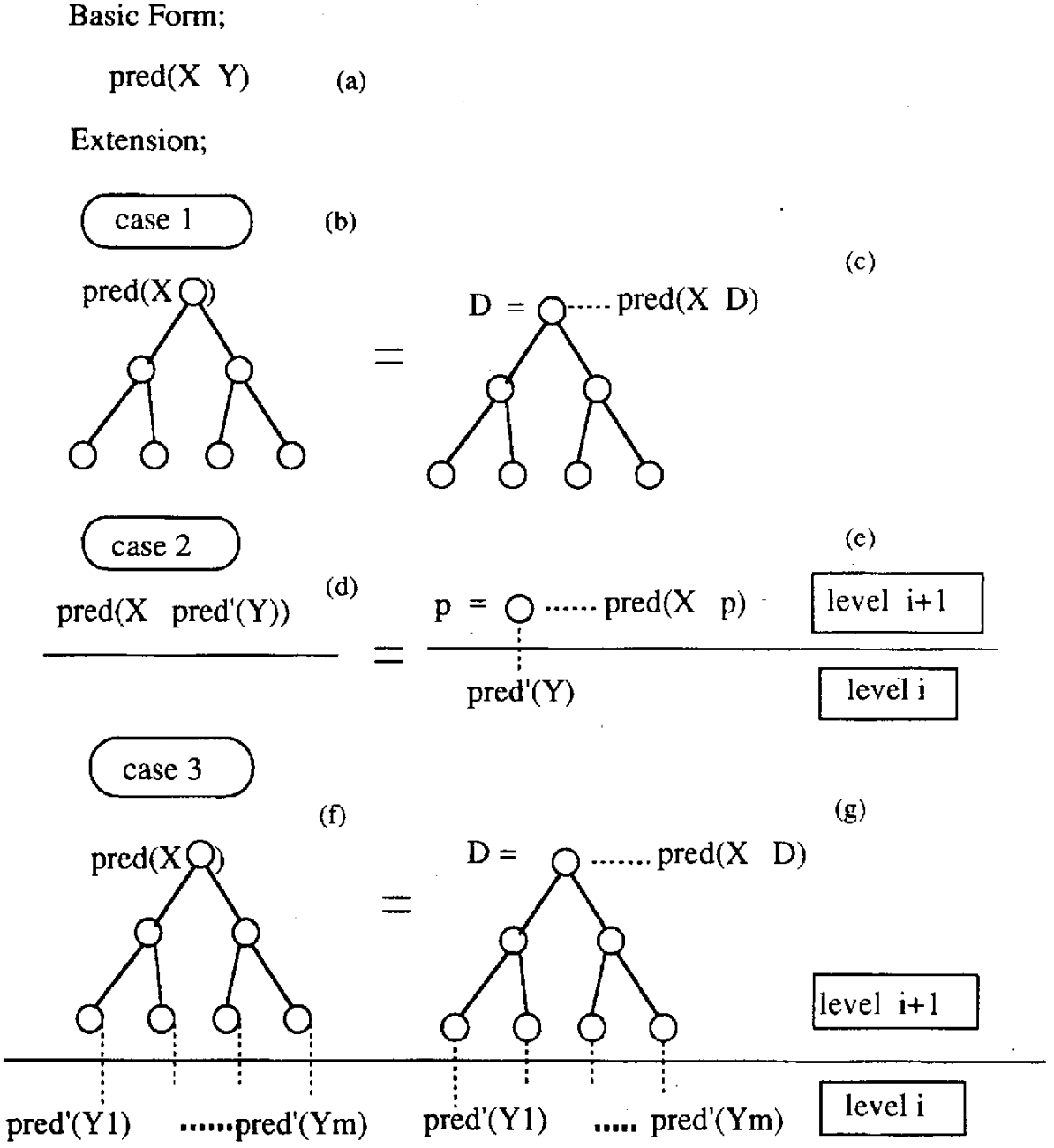


図８：Structure of MLL

それ以上分割のできない単体です。これを、ここに記述される対象を表わす項として、データ構造を含むような拡張をまずしてみます。デザインにおいて「モデルを変える」部分を含むと申しましたが、それは、「モデルの構造と機能を同時に変える」ということよりも、通常人間は構造の部分を変えて、それに伴って生ずる機能の変化をまた再チェックするという措置をとりますので、この構造の記述と、その述語によるその構造の機能の記述、あるいは性質の記述は、ルーズ・カップリングであったほうがよろしい、そういうことで構造表現を分離した形式をとるようにしておきます。２番目の拡張は、「述語

の中に述語を含む」という形を許すことで、これは見かけ上高階論理になりますけれども、実際にはある制約をつけて、実現をしやすいような形をとります。この制約によって、中の述語表現はレベルｉにおき、外側の述語表現をレベルｉプラス１、すなわち一段高いレベルに置くという、こういうオブジェクト・レベルとメタ・レベルの関係を実現することができます。さらに、これを単一の述語にとどめて置かないで、多数のオブジェクト・レベルの述語を用意し、これに対してそれぞれのIDを与えておき、そのIDに基づいた構造を作り、それについての記述を表す述語を与えるというような階層的な表現ができるような拡張をしています。こうしますと、見かけ上、どのレベルにおいても、表現形式は元の基本の述語形式と同じですから、同じ推論メカニズムが使われるということになります。

こういう言語を使って、現実に非決定型の問題を処理するシステムを表したのが図９です。これは先程と同じで、まず対象モデルを作ります。対象モデルを解析する部分と、それからモデルを変換する部分があります。これはこの間を繰り返えし処理するわけですけれども、モデルの解析並びに評価は、このモデル解析・評価の知識ベースを使って行われます。また、モデルの変換は、多数のモデル変換の規則を集めて、これを適用することによって行ないます。これは基本としては図５と同じです。しかし、現実においてはこのモデル解析部分は、このモデルが、現在要求された各項目に対してどういう値を持っているかという解析をしていって、１つのこういうテーブルを作ります。これをモデル・ステイタスとしてやりますと、このモデル・ステイタスをみて、この戦略知識ベースが働きだします。これはメタ・レベルに置かれています。戦略知識は、このステイタスから、現在のモデルの状態が、ゴールとするモデルの状態までどの位の距離がまだ残っているかということを見て、一番その距離を縮めるように働くモデル変更知識を知識ベースの中からセレクトします。セレクトされたモデル変更知識がモデルに適用されて新しいモデルを作り上げていく。そのモデルに対してまた再びモデル解析が働いて、モデルスティタスを作り直す、という操作を繰り返すわけです。

このようにして作られたシステムの、１つの応用例を示します。図１０は、今のシステムを分子構造設計に適用した例です。この場合、モデル変更知識ベースには、分子構造の可能な部分構造の変換例が多数集められて蓄えられています。例えばこれは古い例で、ヒスタミンの系列ですけれども、古い形の分子構造を持ったものから、新らしい分子構造を持ったものが作られてきました。この各対ごとに、実際に変換された部分があります。ある部分構造が、他の部分構造に変わっています。こういう物理化学的にみて、変換の可能な部分、そういう部分構造が抜き出されて、それと同時にこの変換にともなって生じているこの物質の性質変化を併せて、ひとつひとつのモデル変換知識を築いていきます。その結果がモデル変換知識ベースに入れられてくるわけです。そのようにして作られたモデル変換知識を繰り返し適用することによって、既知の物質から未知の物質を作っていくというのが基本の考え方です。しかし実際には、その変換の規則というのは非常にたくさんあります。先程古い例を用いましたけれども、現在でも化学の分野では毎月のように新しい反応が発見されて、それが登録されてくるわけですから、それがどんどん知識ベースに入ってきます。非常に多数の変換則が作られるわけです。

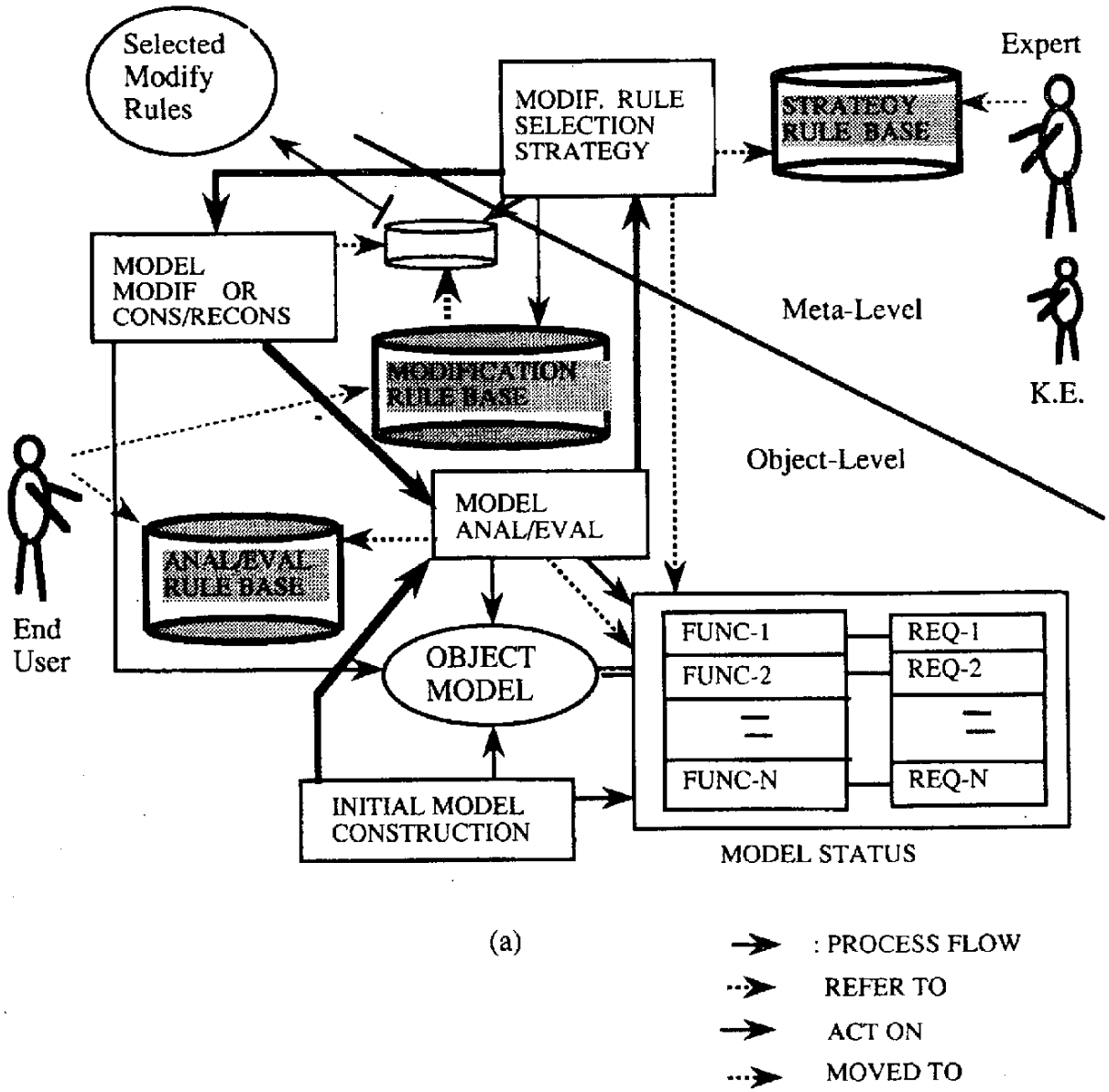


図9 : Object model design process

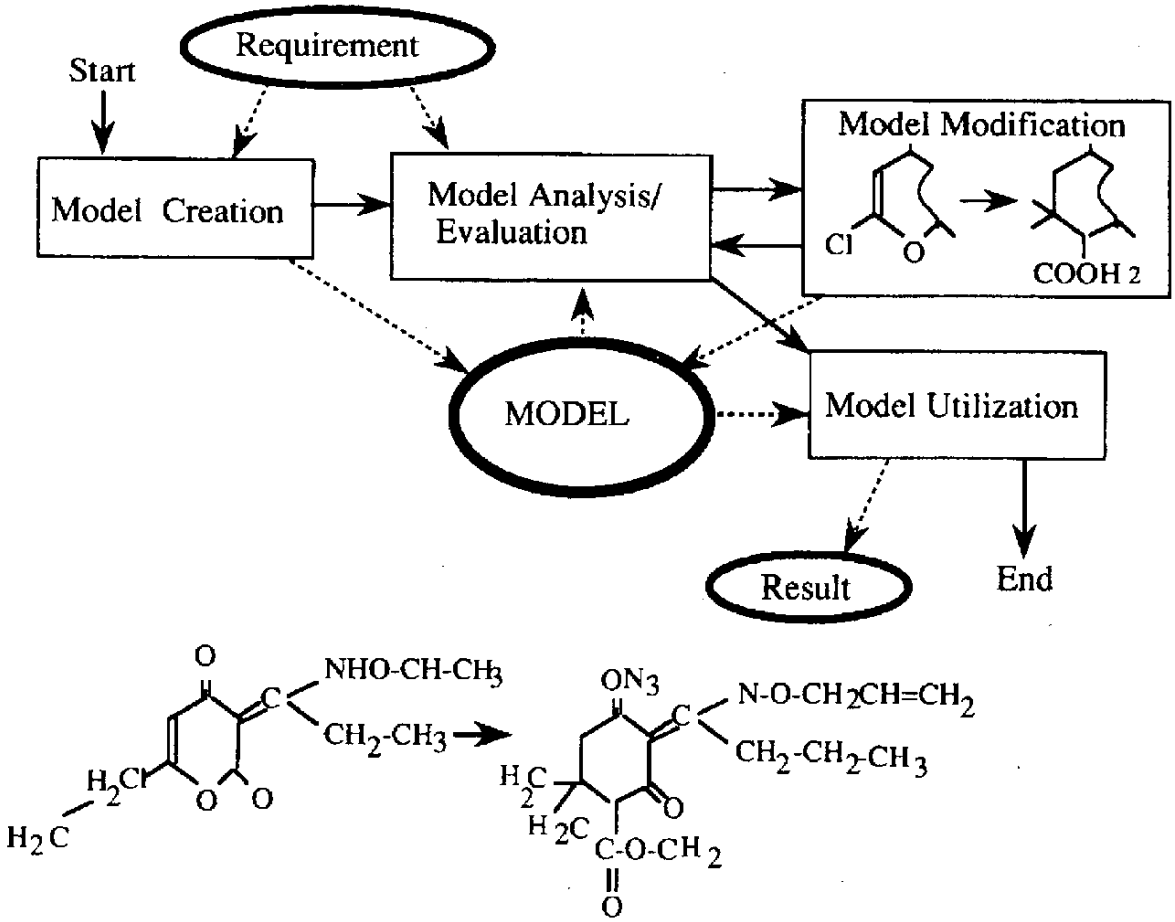


図１０：Design process applied to chemical compound design

それを、ランダムに適応するというわけにはとてもいきません。セレクティブに使わなくてはならないのです。そこで、そのための戦略が働きだします。この戦略は、変換則の集まりを性質によって分類する知識ベースです。この例はドラッグ・デザイン、薬の設計をするというケースですが、ある構造の既知の物質の性質、例えば毒性が非常に強すぎる、したがって、この次の変換はその毒性を減らすような変換をしたい、そうデザイナーが考えたとします。この場合には、この多数ある変換則の中からそういう目的にあったものだけが選ばれるのが、非常に望ましい訳です。そのためにここでは、変換則が及ぼす効果等に基づいてメタ・レベルでこれが構造化されています。例えば、この３つが、毒性を減らすように作用する変換知識が３種あるとします。そのことは戦略メタ知識で表現されています。これがディスプレイに表現される。デザイナーは、これを例えばクリックしますと、そのグループだけが選ばれて、モデルを変換するように適用されるわけです。そのようにして作られた結果が図１１のです。この場合にモデル変換則自身は数千、将来はこれは数万にもなると考えられますけれども、その中からごく小数のものだけが選ばれます。図１１では、最初出発した物質から２段目に示したものが作り出されました。この場合には複数の候補が作られますので、デザイナーはこれを見て、その次のプロセスとし

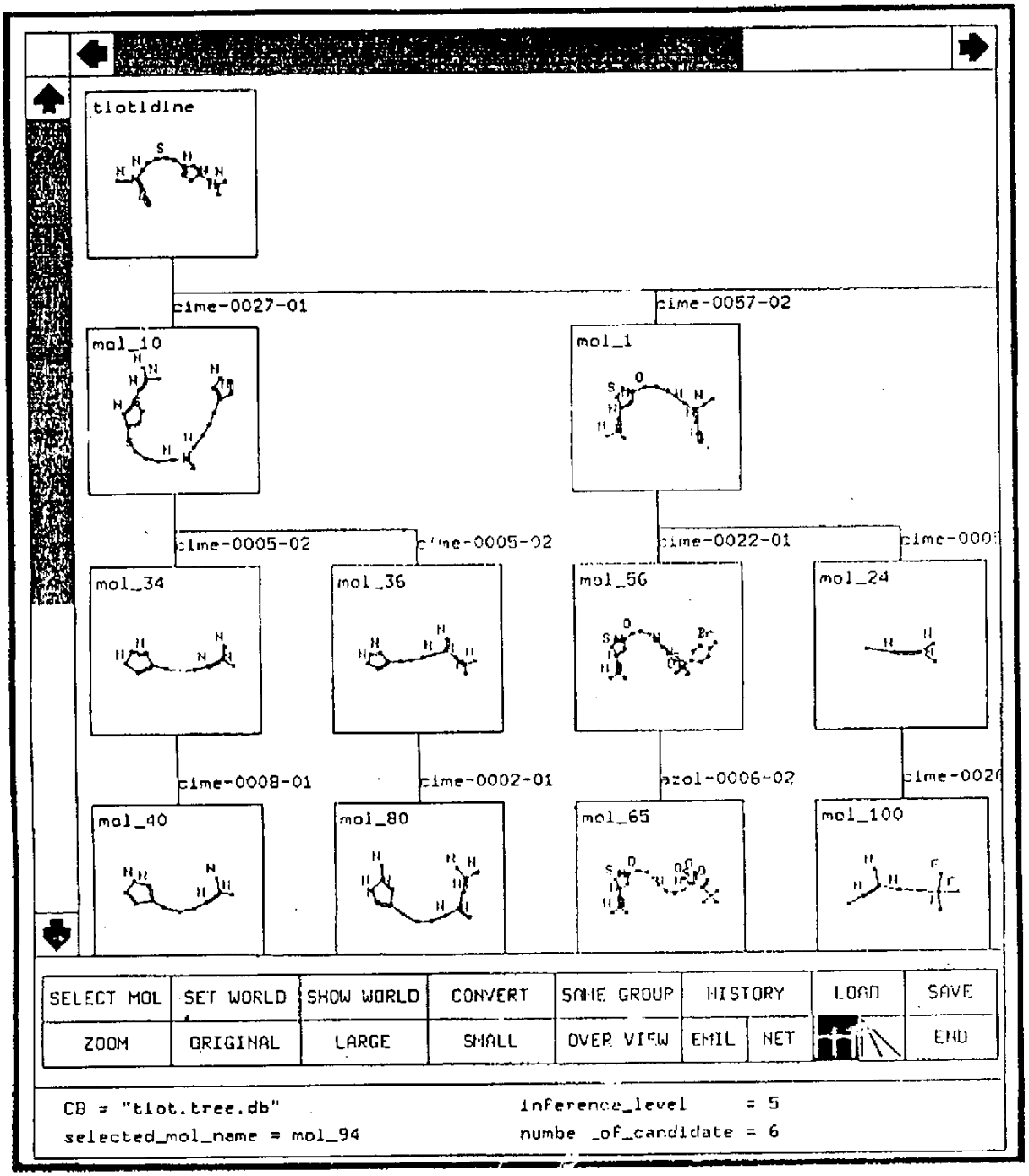


図１１：An example of chemical compound design

て、用いるものを選ぶことができます。もし、この戦略知識を、別な形に作って、この変換則の中から一番いいと思われるものを１つだけ選ぶようにしますと、これは全体が自動設計システムになります。そのような例として、私どもは既に自動制御系の設計システムといったようなものを例として、これで自動設計ができるということを実際に示してまいりました。

さて、以上は設計ということを中心にして、インテリジェント・システムがどういうように働かなく

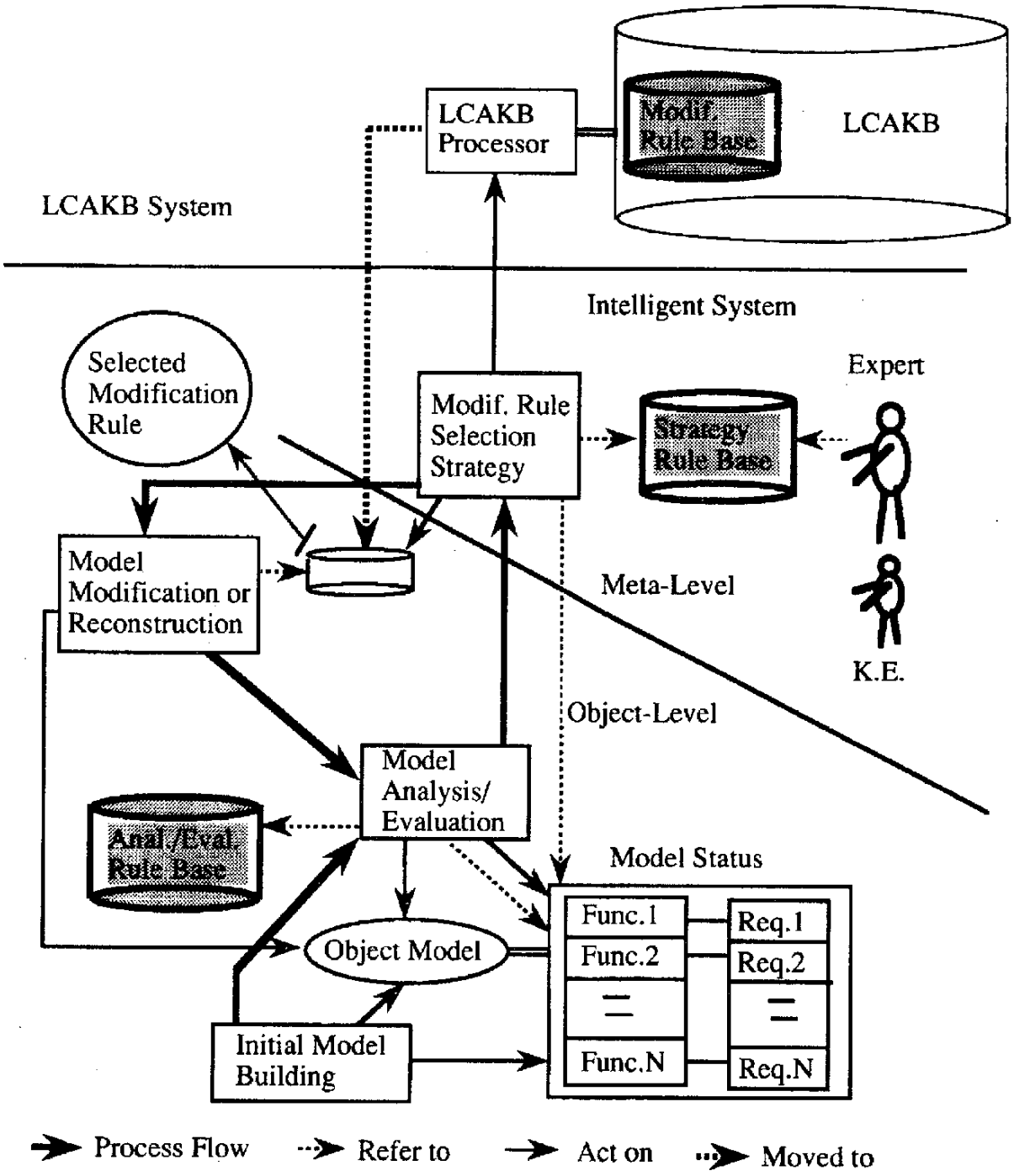


図１２：Moving object−level knowledge to LCAKB

てはいけないか、というための議論でした。これを大規模知識ベースの共有という視点から考えてみます。図９を少し変えてみます。図９でモデル修正知識ベースは、このシステムの中では一番たくさんの知識を持っている部分です。しかもそれは、日常的に増えて行きます。ところがこの構成では、モデル修正知識全体が使われるのではなくて、戦略知識が、どういうモデル修正知識を選ぶかということをま

ず決めて、その中からセレクティブにモデル修正知識が選ばれているわけです。この戦略知識が、各インテリジェント・システムの中でそういうセレクションを行なう代わりに、このモデル修正知識ベースを外に出してしまう。そして、それを大規模知識ベースとして登録しておく。そうすることによって、今度は戦略知識が働いたその結果を大規模知識ベースに送ってやる。大規模知識ベースはそれに基づいて、このモデル修正知識の中から、最も有用と思われるものだけを選んでこちらに返してやる。そういった程度の知識検索の機能をここにもたせる。そうすることによって、そのインテリジェントシステムと、バック・エンドである、大規模知識ベースが非常にうまく結びついて使われることになります（図１２）。

同じようなことはこの解析知識ベースに関しても同じです。様々な解析がありますけれども、今必要なのは、この要求された項目に対する解析ですから、やはりこの全体の中から、今の問題解決に必要な部分だけが選ばれるようにしておけば良い。そして、この本体をこちらに移すことができます。そのような形で作られた、大規模知識ベースは、インテリジェント・システムの場合と同じような形で、多数の知識ベースから成っていますけれども、やはりメタのレベルを持っていて、まず問題領域で分類し、その問題領域のなかで、その戦略に対応して、知識が持つ性質によって区分けされます。そしてインテリジェント・システムから送られてくる要求、メタ・レベルにおける要求とマッチングをとって、マッチした部分を送り返してやる、そういった形でこれが働くことがメモリーのバランスからいっても、それから、それを使う、処理の速度のバランスからいっても、非常に合理的なものなのではないだろうか、これが私どもの１つの提案であります。

以上の議論をもとに、ここで結論を申し上げます。大規模知識ベースというのは、バック・エンド知識源です。これは、インテリジェント・システムに対して、必要な知識をサプライするようなものとして私どもは捉えています。そして、その知識ベースは、全てのインテリジェント・システムとロジカルには一体化して働くようなものとすべきです。２番目に、その大規模知識ベースは、全てのインテリジェント・システムに対して、非常に効率よくサービスを提供しなくてはいけません。そのためには、この大規模知識ベース自身が特定の問題解決をするというのではなくて、これはあくまでもそのバック・エンドとして、必要な知識をサプライする、そういった役割に徹するべきと思います。そういう役割をきちんと果たすためには、知識に関するインデックシングが必要ですけれども、これはメタのレベルの知識になります。したがって当然メタという概念を実現できるような言語をインテリジェント・システムに関しても、大規模知識ベースに関しても、定義する必要があるでしょう。以上が私の提案でもありますし、この、今日のプレゼンテーションの結論でもあります。どうも有り難うございました。

座長

大変興味深いお話ありがとうございました。それでは質問の時間が少しありますので、何か質問、或いはコメントがありましたら…。最初に質問する人はかなり勇気が要ると思うんですけれども、なかなかしにくいかなという気がするかもしれませんが、誰かし始めると、こういう所はバッと広がるという

のが…。大須賀先生の長年のいろんなお仕事の、いろいろの集大成された部分もありますし、それからさらに、新たに提案された部分もあります。それで、表現というのは、普通だとすると、表現を計算する部分とか、その辺のアーキテクチャーというのが、いままであんまりはっきりしてきていなかったのですが、大須賀先生の研究では、それをボトムからアップまでそれをすっきりした形でまとめて、それをメタ・レベルの処理の中で戦略決定をすればどうかということと、それをドラッグ・デザインとか、分子の設計に応用されて、具体的な例を示されたという話だと思うのですけども…はい。

質問者：

簡単な質問をさせて頂いてよろしいでしょうか。あなたのマルチレイヤーロジック(MLL)はまだ一階述語論理であり、見かけ上高階論理でもあります。例えば非述語的なシンボルと述語的なシンボルを含めますと、それが高階論理であった場合、なにか良い影響はありますでしょうか。

大須賀：

もし私どもが非常に一般的な意味で高階述語論理を使おうとすると、それを実現するのは非常に困難です。それでそれを使うには利用上の制限をしなければなりません。この考え方にたってシステムを実現するのに私たちが行なったのは、使用上の制限をしていることです。つまりこの述語内述語は閉じられたフォームの中で文を作っている、つまりこの内部述語の中では自由な変数の使用を認めないわけです。もう少し正確にいいますと、外部述語内変数と内部述語内変数への代入の手順に一定の制約をつけていることです。それが外部述語と内部述語の評価を分けることができる理由になっています。したがって内部述語と外部述語を別のレベルに置き、低位のものを識別子で代表します。上位述語がその識別子を含むわけです。とても簡単で、基本的には、一階述語の枠組みでありながら高階論理のもつ機能のみならず、一階述語のみですと非述語記号を導入しないと表現できないものも一部は表現できます。

座長:

それでは、大体時間になりましたので、もう１度拍手をお願い致します。

(2)「知識共有：予測と課題」

座　長：

それでは次の講演に移りたいと思います。日本語のタイトルでは「知識の共有：予測と課題」ということですが、"Knowledge Sharing Prospects and Challenges"ということで、論文の方は３人のco-autherになってますけども、ご発表はBill Swartoutさんです。この３人の先生方も長い間知識表現の研究で、Nechesさんは、ワークショップの方でもお話されます。簡単にSwartoutさんの紹介をさせて頂きますと、Swartoutさんは、University of Southern Californiaという所のInformation Science Institute、これはあの、Marina del Rayという非常に風光明媚な所にある研究所で、さぞかし研究が進んでいるんじゃないかと思うんですけど、そこのインテリジェント・システムのディレクターでございます。それで、Bill Swartoutさんは、ＭＩＴのご出身で、特にMedical Diagnosisですね、Peter Solvisさんという方がThesis Advisorで、論文をまとめた方でございます。特に、エキスパート・システムでは、エクスプラネーションということが重要なポイントだと思うんですけども、その辺のご研究を長くやっていらっしゃる方です。

University of Southern California, Infometion Science Institute
Professor　William R. Swartout

おはようございます。私がこれからお話する研究では、知識共有の分野で、将来もっと効率的に知識を共有できるようにするための技術を開発するためにはどうしたらいいかを考えています。最初にあたって申し上げておきたいのは、私がお話する研究には、実に何百人もの人が参加しているということです。私の話の最後に、この研究についてもっと知りたい皆様のために、連絡先のリストのスライドをお見せいたします。

初めに現在の状況をお話ししましょう。知識に関する現在の状況はどのようなものか。次に、将来どのような成果が望まれるか、そしてなぜ今現在そこに至っていないか、なぜそれが今できないのかついてお話することにします。そのあとに知識共有計画についてお話し、その計画において現在知識共有の妨げとなっているいろいろな問題を私共がどのように克服しようとしているか、そして知識共有計画に参加している４つの研究グループがどのような成果を示しているか、何をやっているのか、何が達成されたか、何がまだ問題として残っているのか、というお話をしようと思います。

まず現在の状況について少し考えてみましょう。現時点では、もし知識ベースシステムを作っても、１つのシステムに入れた知識を他のシステムで共有したり再利用することは大変難しいのです。たとえその２つのシステムがよく似たものであっても、よく似たドメインに関わるものであっても同じです。

問題は、知識ベースが今現在のところ、特定のシステムや特定のタスクのためにカスタムメードされていることです。そのために、ひとつのシステムを作るたびにゼロから出発しなければならず、土台に

できるものがほとんどないのです。知識ベース開発のコストは、システム毎に個々に負担しなければなりません。たくさんのシステム間でコストを分担することは不可能なのです。またそれぞれのシステムは独力でバグ処理をしなければならず、そのため同じ誤りを何度も何度も繰り返す可能性がたくさんあるわけです。

もう１つ、現時点での問題は、知識エンジニアには非常に高度な能力が求められることです。いくつものことができなければなりません。システムを構築するためには、ドメインエキスパートから知識を抽出すること、その知識を表現するための適切なフォーマリズムを開発すること、そしてドメインエキスパートからそのフォーマリズムに捕捉した知識を表現する能力が必要なのです。

こうしたことすべてが原因となって、現時点では、知識ベースシステム開発は非常に費用と時間のかかるものとなっています。このような問題もあって、ＡＩは、'８０年代に私達が期待したようなインパクトをいまだ持ち得ていないのではないかと思われます。当時、私達はＡＩに目を向けはじめ、ＡＩの商業的利用の可能性について真剣に考えはじめていたのです。

私達が目指すのは何なのでしょうか。将来私達がやりたいことは、知識ベースを、ゼロから築きあげるのでなく、ライブラリーにある既存の断片や部品を再利用して作れるようにすること、ライブラリーから情報を取り出しそれを利用することによって大きなシステムを構築できるようにすることなのです。このような環境になったとき、知識エンジニアに求められる能力は今とは違ったものになるでしょう。

知識エンジニアにとって重要な能力は、再利用できる知識を見つけだして、その既存の知識を、新しく作るシステムで利用できるよう改造して特殊化することになるでしょう。その結果システム構築は迅速になるはずです。なぜなら今の時点で始める場合よりもはるかに上の段階から出発できることになるからです。システム毎のコストは低くなります。なぜなら多くのシステムの間でコストを分担することができるからです。そして信頼性もはるかに高くなることが期待できます。なぜならこのような既存の知識ベースから得られる知識は、既にバグ処理がほとんど終わっているはずだからです。

なぜ今この段階に到達できないのでしょう。私達の前進を妨げる障害は何なのでしょうか。障害の１つは知識表現言語が数多くあることです。表現方法が多種多様なのです。そのため、異なる表現言語間で知識を行き来させることが困難です。この種の多様性は実際のところ避けがたいものであり、また実は大変望ましいものなのです。知識表現言語を１つしか作り出せないし、それが何にでも最適だということはここではあてはまりません。様々なことをあらわすのに様々な言語があったほうがよいのですが、しかし複数の言語が有るという事実のために知識の共有は難しくなります。

もう１つの問題は、同じファミリー、つまり同じ知識表現法の内部でさえ、例えばKL/ONEファミリーと呼ばれる知識表現言語の中にも－正式にはターミノロジカル・ロジックスとか記述ロジックスと呼ばれますが－、その中にも多くの方言が有り、多くのインプリメンテーションがあり、そのインプリメンテーション中にもまた小さな統語法の相違があって、ひとつのファミリー内での、それどころかひとつのグループ内でさえ、知識の共有を妨げているのです。

最後に、知識表現に対するもうひとつのアプローチがあります。１つのシステムから知識ベースを取

り出してそれをそのままの形で別のシステムに移すのではなく、知識共有のために利用できるもうひとつの方法は、基本的には知識サーバを持つことで、その知識サーバに問い合わせ、答えを得て、処理を進め、また問い合わせ、また答えを得る、ということを繰り返すやり方です。そうすれば知識共有のやり方は、知識ベースそのものの移動ではなく、質問しながら進める（QUERY-BY-QUERY）ものになります。このやり方をとる場合の問題は、今のところ知識サーバへの質問を共有するためのよい通信プロトコルが無いことで、それを開発することが必要だということです。

この最後の問題は、既に話題に登ったオントロジー上の問題です。つまり共有された共通のターミノロジーが無いことです。たとえ同じ表現言語で、あるいは同じ方言で知識を表現できたとしてもなお、一方の知識ベースで使われている用語がもう一方で使われている用語と完全に違ったものであるなら、システム間で知識を共有することは非常に難しいものになるでしょう。従ってはじめの３つの問題を解決できたとしてもなお、ターミノロジー、事物の表現の仕方を調整するという問題が残ります。つまり、土台となる共有のオントロジーをもつこと、出発点となる共通の基礎を作ることが必要なのです。

知識共有計画とは何の仕事をしているのでしょうか。それは、国際的な努力により、知識共有と再利用のための技術を開発することです。この計画をコーディネートしているのはUSC/ISIのRober NechesとRamesh Patilです。私共は高等研究計画局（The Advanced Research Agency）すなわちARPAから一定の支援を受けています。これは標準のためではなく、私共は標準作りを目指すものではなく、私共の仕事とＩＳＯなどの国際標準化団体との間の整合を図っているのです。全体は４つの作業部会に別れ、それぞれがさきほど知識再利用のところでお話した問題に取り組んでいます。

このように、４つのグループが有ります。第１のグループはインターリンガグループで、複数の表現法の間での知識の移動の問題に取り組んでいます。基本的には、仲介的表現法を使って各表現法間の翻訳を行なおうとしています。

複数の方言の問題に取り組んでいるのは、知識表現システム仕様グループ、略してＫＲＳＳで、知識表現システムの言語ファミリー間の共通仕様を作り出そうとしています。

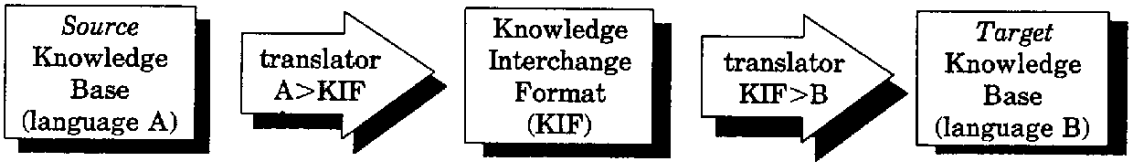
外部インターフェースグループはプロトコルの欠如の問題を扱っており、質問ベースの共有を促進するための対話プロトコルを開発中です。これについては明日のワークショップでTim Fininが詳しくお話することになっています。

最後に、共通ターミノロジーの欠如の問題に取り組んでいるのは、共有再利用知識ベースグループです。このグループは共有ターミノロジーあるいはオントロジーを開発しようとしています。

では、各グループについて順番にみていきましょう。今どのような地点にいるのか、どんな問題と直面しているのでしょうか。

インターリンガグループは、仲介言語すなわちインターリンガを用いて各表現言語間の翻訳を試みる方法をとっています。従って、基本的な考え方はこうなります。まず言語Ａによる知識ベースがあります。言語Ａから言語交換フォーマットであるＫＩＦへのトランスレータがあります。次にこのＫＩＦと目的の知識ベースである言語Ｂを行き来できるトランスレータがあります。ここで、インターリンガが

この仲介の役目を果たすことにより、必要なトランスレータの数を基本的に２ｎに減らそうという考え方です。ここでｎは、やり取りを行ないたい言語の数を表わし、これに対しN（大文字）は、やり取りを行ないたい全言語をペアにしてトランスレータを作った場合を表わします。



◆Interlingua (KIF) serves as intermediate representation
◆reduces number of translators required (2N vs. N^2)
◆KIF could serve as language for library repository
◆concerns:
　－expressiveness
　－translation support
　－extension

ＯＨＰ－１

それに加えて、ＫＩＦは再利用したい知識のライブラリーのための言語の役目も果たします。その考え方は、ライブラリーからＫＩＦに知識を引き出してからそれを実際に使いたい言語に翻訳するというものです。ＫＩＦ開発において重要なのは、十分な表現力を確保すること、つまりソース言語と目標言語における情報内容を十分に捉え、翻訳であまり多くが失われないようにすることです。また、翻訳を支援すること、そして拡張性があって、世界をどう表現したらよいかがもっとよく分かってきた時に、そのような情報をもＫＩＦ言語で表現できることです。

インターリンガグループの成果はＫＩＦのための仕様を開発したことです。これは一階述語論理を発展させたもので、形式的モデル理論による意味論を持ち、オブジェクト定義のための、すなわちターミノロジー定義のためのサブ言語を内部に持っています。知識についての知識、例えば信念を表現するためのメタ知識のためのサポートもあります。また拡張機能があり、それにより言語を拡張することができます。これまでに、主としてLOOMやCLASSICなどのKL/ONE式言語内外について、またフレーム言語間についての翻訳実験を行なってきました。

もう１つ、どちらかといえば意外な結果だったのですが、ＫＩＦは教育学的用途にも役立つことが分かりました。表現言語の意味論を明示し、異なる表現言語を比較するための一種のツールとしても役に立つものだったのです。

このグループの直面した問題は、まず初めに分かったのは、ＫＩＦへのソース言語からの翻訳は容易なのですが、ＫＩＦからの翻訳が難しいことです。とにかく翻訳の形をしたものさえ得られない場合があります。問題は、ＫＩＦは所与の命題、与えられた文をいろいろな統語法的に異なるやり方で表現で

きることです。トランスレータが認識できるのはその一部だけなので、必ずしも常にＫＩＦの中身を翻訳して出すことはできないのです。

第２の問題は（これも同様に重大なのですが)、翻訳から得られた結果が目標言語の慣習に合わない場合があることです。表現言語で物事を表現する場合、ある種の慣習、ある種のスタイルによって表現の仕方を決めることがたいへん多いのです。問題なのはそのような約束事は論理的意味論に属するのではないため、翻訳の際に迷子になってしまうことです。

例えば、いくつかの言語は型(type)を支持しており、「犬は動物の一型である」という言い方ができ、全ての型を階層化しています。別の言語は型を全く持たず、そのかわりにすべての事物を単項述語として表現します。ＫＩＦに翻訳する場合、ＫＩＦは全てを単項述語として表現しますが、ＫＩＦから翻訳する場合、型を支持する言語に翻訳する場合には、ＫＩＦにおける単項述語を型として表現したいか、それとも、もとのＫＩＦでそうであるとおりに単項述語として表現したいかを決めなくてはなりません。問題は、その選択はＫＩＦ表現のどこにも指示されていないことで、そのため、知識ベースをどのように構築することを好むかを定める慣習があっても、そのような慣習は出来上がった翻訳には反映されないことになります。

私共がこの問題について考えている解決策は２つあります。１つは、基本的に、自動、完全自動翻訳を行なおうとすることは恐らく実現不可能であるから、人間の関与が必要である、翻訳プロセス全体を導く人間が必要であると考えることです。そのため人間主導の翻訳ができるような枠組みを構築しているところです。

もう１つの考え方はＫＩＦを拡張することによって翻訳で失われるこの種の情報をもっと捕捉できるようにするというものです。

第２のグループは知識表現システム仕様グループです。このグループは特定の表現システムのための共通仕様を作ることによってひとつのファミリー内での方言的相違を低減しようとしています。単にひとつの表現言語を作るのではなく、完全な仕様を作成しようとしています。その意味するところは、つまりその言語の統語法と意味論を持つのに加えて、知識表現システムに典型的に備わっているアップデート質問および推論機構を備えるということなのです。

このようなシステム、このような仕様があれば、既存のシステムの最良の特徴を組み合わせることができます。これがシステム製作者にとって理解できるものになれば、構築したいシステムを簡単に実現できるでしょう。これが拡張性を持つようになれば、知識表現の方法についての理解が進んだ時にこれを拡張することができます。最終的には、私共の作り出す仕様のどれも多くの実際のシステムによって使われ支持されることが望まれます。

これはCommon Lispでの経験にいくらか類似しています。Common Lispが現れる前は多数の異なるLisp方言があり、異なる方言間でプログラムを移動することは非常に困難でした。Common Lispが現れた時，このプロセス全体ははるかに容易になったのです。第一のターゲットは記述論理LOGICSすなわち表現言語のKL/ONEファミリーです。

基本的にここで私達がやってきたことは、仕様の草案を起草することでした。それを各研究団体に回覧してコメントを求めました。何度か回覧修正を重ねました。最終草案はAAAI-93で起草されました。この草案の最終承認は、草案は今回覧中ですが、ドイツのボンでのKR'94でなされるはずです。もし承認されなかったら私は非常に失望することになります。

ＫＲＳＳ仕様を作る際にぶつかった問題はどちらかというと意外なものでした。表現のコンストラクトの統語法と意味論に関して同意に達することは比較的簡単でした。難しかったのは行うべき推論の種類、特にKL/ONEclassifierによって行なうべき推論の種類を定めることでした。

ここでの問題は（これは知識表現システムがLISPのようなプログラミング言語と異なる場合に実際に起ることなのですが)、LISPには基本的に一連のコンストラクトがあり、それが何をなすべきかをきちんと定義した意味論があります。表現システムにおいてはそれ以上が期待されるのです。そのようなコンストラクトを持つだけではなく、表現システムにはある程度の推論を行ってくれることが期待されます。そしてここで問題となるのは正確にどこまでシステムが推論する必要があるかということです。どの程度の推論をさせればよいのでしょうか。

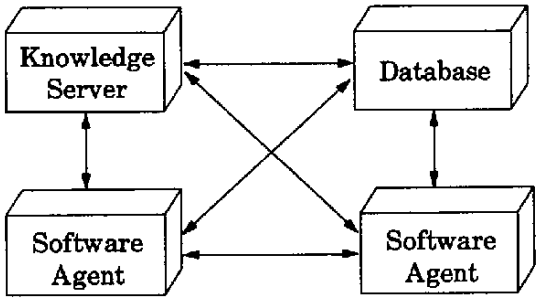
問題は言語の表現力であり、その言語でどこまで表現できるかがclassifierもしくは推論の完全さと扱いやすさを左右します。表現力の高い言語を使うほど多くを表現でき、推論機構の行うべき仕事が多くなります。推論機構で処理できる仕事が増えるのです。現在KL/ONEファミリーで実現されている各システムでは表現力と推論の完全さのレベルの異なるものが選ばれています。その理由は、各システムが別々の種類のアプリケーションを想定して構築されたためで、どのシステムにも正しいという選択はないからです。

この問題の解決法は、基本的には多層的仕様を作ることで、その内側に小さなサブセットを持ちそれによって完全な推論を多項式時間で行なう事ができるようにすることです。その外側にはより表現力を高く定義された拡張のセットがおかれます。そしてこの内層コアによって推論するためには、仕様に合致した、つまり仕様に合ったシステムであることが必要です。システムは外層コアを解析することが必要で、任意に外層コアによって推論することもできます。このようにこのアプローチは基本的にシステム製作者にかなりの自由を与えるものになっています。システム製作者はどの程度推論を支援したいかを選ぶことができ、なおかつユーザーは知識ベースを共有することができるからです。外層レベルに書かれている知識ベースも共有できます。そしてシステム間で最低限の推論が保証されています。

外部インターフェースグループは、ソフトウェアエージェント間の相互運用のための、支援プロトコルの問題を考えています。例えば知識サーバが１つあったとすると、それは一種のソフトウェアエージェントである、データベースが１つあると、それはまた別のソフトウェアエージェントです。基本的にはエージェント同士がやりとりし、質問を送り合って知識共有することが望ましいわけです。

このグループが開発してきたのは、メッセージフォーマットを備えたプロトコル、メッセージ処理プロトコルと実行時間共有支援のための戦略です。そのためにこのグループが提示した言語はＫＱＭＬ、つまり知識質問操作言語（Knowledge Query Manipulation Language)と呼ぶものです。これは基本的に

◆Creating common high−level language & protocol
for inter−operation among software agents



◆provides message format, message handling
protocol & policies to support runtime sharing

ＯＨＰ－２：External Interfaces (KQML)

一連の遂行的発言を中心に構築されており、これは一方のエージェントが他方のエージェントに頼む行動や依頼を定義したものと考えてください。これがおよそ２０個あります。

個々の遂行的発言には、そのメッセージをどのように解釈すべきかを定義したパラメータ・フィールドが一組づつあります。従って、例えば一方のシステムがもう一方に問い合わせて「ロサンゼルス空港（LAXと呼ばれます）はどこですか？」と聞いたとします。するとこれは以下のような遂行的発言によって表現できます。これは質問的（ask-about)遂行的発言です。送信側はあるプランニング・システムであるかもしれず、受信側は知識サーバであるかもしれません。問い合わせはこのようなものになります。「LAX、すなわちロサンゼルス空港の地理的位置はどこか、経度と緯度によって知らせよ、また応答はＫＩＦで受けたい。DRPI-GEOの定める用語を使用して欲しい。」このメッセージフォーマットが伝えている内容は、基本的に、誰が、誰に対して質問しているのか、質問は何か、どのように表現されているか、質問はどの統語法によって表現されているか、返答にはどのオントロジー、すなわち用語のセットを使うべきかです。

もう１つお話しておくべきことは、このグループはfacilitatorsという特別なソフトウェアエージェントについても研究しており、これは異なるエージェント間の対話を実現するために作られ基本的に必要な時に情報を探す助けをします。

このグループもまたいくつかの成果をあげ、またいくつかの課題も残しています。彼らは仕様の草案を作成しました。ＫＱＭＬエージェントを作るためのソフトウェアツールのキットを考案しました。このアプローチは何回かのデモンストレーションプロジェクトでテストされています。

このグループに残る問題は、第１にどの程度の情報を返すべきかを決定することです。知識ベースからひとつの項目を引き出す時に気がつくのはそれが他のたくさんのもの全てと結び付いていることです。まるで皿いっぱいのヌードルの中から１本のヌードルを引っ張り出そうとするようなものです。他のヌードルもいっしょにくっついてきます。ここで考えているのは、問い合わせに必要なディテールのレ

ベルを示すことによって、あるシステムが情報を依頼する時に、どの程度のディテールまで戻して欲しいかをそのシステムが指定するようにすることです。

もう１つは、情報を探す時のエージェント間の共同作業を促進することです。そして最後に、異なる表現間およびオントロジー間の翻訳の問題です。質問がある表現法またはあるオントロジーで表現されて入ってくると、答えは別のオントロジーで返ることになり、その２つの用語の間を行き来するための支援が必要になります。

最後のグループは共有再利用知識ベースグループで、知識ベースを構築する際に使うターミノロジーの調整を扱っています。問題は、上部構造の異なる知識ベースは共有が難しいことです。そのため、例えば２つの知識ベースがあり、両方とも最下位のレベルに至るまで全く同じ事物を表現しているとします。飛行機の部品の１つ、ランディング・ギア・ストラットという、着陸装置についている支柱です。ところが上部構造では、こちらではこれは最上位レベルではそのものであり、次に可算名詞で次に支柱、次に、ランディング・ギア・ストラットは一種の支柱であるとなります。もう一方では実在物であり、物質的事物であり、飛行機の部品であり、そしてランディング・ギア・ストラットなのです。この２つの知識ベースを共有しようとするのは非常に困難です。なぜなら上位レベルの用語が調整されていないからです。このグループがやっているのはハイレベル用語、あるいはオントロジーと呼ばれるセットになったターミノロジーを開発することで、これは大規模知識ベースを構築している人が土台として使うことができます。これがあれば複数のシステム開発者が共通の土台、共通の基礎から出発することができ、その結果構築された知識ベースはより共有しやすいものになるはずです。

このグループの成果はいくつかのオントロジーを開発したことです。その中のいくつかは、エンジニアリング・モデルや環境問題などのいわば問題解決領域にかかわるものです。そのほかのものは自然言語などにかかわるもので、自然言語の単語のための大規模なオントロジーもあります。

ここでの課題はこのようなオントロジーを利用できるソフトウェアを開発することです。特に問題なのは、１つのオントロジー、つまり１組の用語と、それが意図されている用途との間の関係を捉えることです。これは問題解決にどう使われることになっているのか。なぜなら、この場合、あるひとつのオントロジーと共に使おうとするような種類の問題解決法は、そのオントロジーの本質に非常に強い影響を与えるという問題があるからです。この関係を認識して、それをオントロジーそのものの中に表現することにより、オントロジーの使用者に、そのオントロジーがどんな種類の問題解決法のためのものなのか分かるようにする必要があります。このようなものを開発するために、またこのようなオントロジーにどうアクセスし、それを維持し、アップデートするかの問題のためのツールが必要です。

最後の締めくくりとして、この計画に大きな寄与をなさっている方々の一部のリストをお見せしたいと思います。またここでお話してきたことは４つの別々の研究であることを申しあげておきたいと思います。これらの研究が全部成功しないと、私達が知識を共有する能力が向上しないとはお考えにならないでください。それどころか、このひとつでも成功すれば知識共有への道をおおいに前進したことになるのです。これを終わりの言葉といたします。有り難うございました。

座長：

Thank you very much for your clear presention.それでは、残された時間がありますので、質問、あるいはコメントがありましたら…。はい。所属と名前を。Name and affliation.

質問者１：

ＫＩＦの表現力について２つ質問があります。１つは、ＫＩＦは一階述語論理の拡張であると申されましたが、これは一種の高階論理なのでしょうか。もう１つは、表現言語についてですが、これを用語を競合的データ構造によってＫＩＦに表現できるような機械語に直接翻訳することは可能でしょうか。

Swartout：

ＫＩＦの現在の構築法の基本には一階述語論理があります。二階バンプがいくつかその上に乗っており、あるメカニズム、一種のエスケープ機構によってその言語で述べられた表現を得ることができ、その表現をどう解釈すべきかを指定することができるようになっています。それによって、二階論理に進むことができるのですが、解決すべき問題がいくつもあります。翻訳の問題は特に大きいと私は考えています。

Randy Gable：

東京大学とアルバータ大学のRandy Gableです。特に差し迫った問題の中でも一番論議を呼ぶのではないかと思われる問題について、ここで少し時間の許す限りお答えいただきたいのですが、長年の間、コンピュータ科学者はより大規模な、よりすぐれた言語を構築しようとしている人々だと考えられてきました。新たなインターリンガを作るというのは、多種多様な表現言語が必要だとおっしゃることと矛盾しているように思われます。この研究には何か根本的に異なるものがあるはずです。あなたは翻訳過程の中に人間を関与させておられましたが、これは他の一連の問題を素通りしています。この一見しての矛盾のために生じる反論や疑問には主にどのようなものがあるのでしょうか。

Swartout：

私は必ずしも矛盾とは考えていません。しかし、ＫＩＦ設計の際に生じた主要な問題はどのようなものであったかというと、これは実際には推論を支援する言語を意図したものではないということなのです。別の言い方をすれば、これは我々が通常考えるような意味での知識表現言語、それを中心に築かれた完全なシステムとしての推論を支援する知識表現言語ではないのです。むしろ意図しているのは、単に知識を補足し、ひとつの表現から別の表現へいく際の軸の役目をする言語とすることです。従って、問題が異なるわけです。

つまり、ＫＩＦは容易に翻訳を支援できる言語でなければならないのです。人間が容易に調べられるものであることが必要なので、他の表現言語に求められるものとは異なる条件をたくさん満たす必要が

あります。

プロセスをどこまで自動化できるかについての問題があります。私は完全に自動化できるということには懐疑的です。なぜなら基本的にそれは、他の考えられるあらゆる知識言語と、少なくとも同等の表現力を持つ言語を作ることを意味するからで、それは考えられないからです。それでも、もしも知識ベースをこちらの表現からこちらへ、そうですね、９０パーセント移せるようにしてくれるツールがあったとしたら、何もしてくれないものを持っているよりはましでしょう。ですから私は、ループ内の人間および翻訳過程への人間による支援を、完全に追い出してしまうことは、たいへん重大な問題だと考えます。

座長：

話は、AUPAの、４つのグループの話で、数年前まではインターリンガが中心だったのが、何年かすると４つのグループに分かれていくという、まあ、アメリカというのは、そういう点では組織力があるなという感じがします。

またいろんなディスカッション、標準化は、"KILL"、"KILL of KISS"とか何か、いろいろと論議を呼んでおりますので、その辺のおもしろい話題がたくさんあると思います。標準化というのは逆に言うと、いろんなアクティビティ、創造性なんかを殺してしまうという部分もあるし、たぶんそれから、「品質の問題」というのもあると思うんですね。"reuse"というときには必ず「品質保証」とか、「クオリティー」というような問題があると思うんで、「それをどう使うか」というのも大きな問題だと思います。多分いろんなディスカッションを経て、そういうグループが動いていっていると思うんです。

それではもう一度拍手をお願い致します。

(3)「共有知識ベース：ヨーロッパの観点から」

座　長：

３番目の講演に移ります。日本語のタイトルは、「共有知識ベース：ヨーロッパの観点から」ということで、"Reusable and Shareable Knowledge Bases: A European Perspective"ということで、アムステルダム大学の、Bob Wielingaさんです。

簡単にご紹介しますと、Wielingaさんは、核物理学で学位をとられております。ですから、最初のお仕事はコンピュータ・コントロール・システム、特にリニア・エレクトロン・アクセレーター。アムステルダムのニュークリア・フィズィックス研究所の、そういうエレクトロンのアクセレーターの制御をやっていたという方で、それはだいぶん古い仕事だと思うんですけれども。

１９７０年代の後半になってからは、特にコンピュータ・ビジョンの知識表現の分野で研究されて、最近では特に、Acquisition、それからLearning Processesとか、Intelligent Coach for Teaching Physicsとか、認知科学的な立場とか、そういう非常に多様な研究をされている先生でございます。

University of Amsterdam, Department of Social Science Informatics
Professor　Bob J. Wielinga

私がお話申し上げることも皆さんのお話と大体同じような問題になりますが、私としては「知識ベースの再利用性と共有性の問題」についてヨーロッパの視点から考えてみたいと思います。

最初の２枚のシートは前の話でBillが見せてくれた最初の２枚とよく似ています。基本的に問題は、「大量の知識を手に入れること」は可能なのですが、「必要な知識にアクセスすること」、そして「知識ベースの宇宙を巡航していくこと」は難しいということです。第二に、異質の知識を一つのベースとして、異なる種類の知識を一つの知識ベースで表現するのは難しいことです。知識を広い用途に使えるように改変するのは難しいことです。Billの話にもあったように、ほとんどの知識ベースは一つの特定のタスクもしくは一つの特定の目的のために特別に開発されたものだからです。知識獲得をしようとする時に分かるように、異なるソースからの知識を統合するのは非常に難しいことです。最後に、なんらかの知識を見つけだしたとして、「それが実際に何を意味するのか」を評価すること、「それを推論にどう使えるか、その知識の質と範囲はどういうものか」を評価するのは難しいことです。このような問題をお話しようと思います。

私共のビジョンは知識共有計画の背景となっているビジョンとよく似ています。私共が考えていることは、ポイントは結局２つで、アクセスが容易でライブラリーの大規模な「標準知識貯蔵庫」をつくること、そして特定の目的に合った知識を選びだす方法を持ち、選びだした知識を他のソースからの知識と統合する方法を持つことです。また出来上がった知識ベースを特定の応用タスクに、あるいは特定の質問の組み合わせに対して使えるようにすることを考えています。

このビジョンは、私の理解が正しければ、ＣＹＣの背景にあるビジョンとは多少異なっており、私共は汎用知識ベースを目指しているのではなく、いくつかの別々の知識ベースを作って、そこから必要な部分を取り出し、特定のアプリケーションのために結合したり統合したりできるようなものを目指しています。従って、どのようにしてそのビジョンに至ったかをお話したいとおもいます。

私のいう「ヨーロッパの視点」とは、どういうものでしょうか。私はヨーロッパESPRIT基金による知識ベースシステムの分野での、いろいろなプロジェクトのアイディアを取り上げてみたいとおもっています。ごく手短に一覧することにします。

The European Perspective:

Sources of ideas are a number of European ESPRIT-funded projects

- KADS and CommonKADS
- VITAL
- ACKnowledge
- MLT
- GAMES-II
- KACTUS
- EUROKNOWLEDGE

Total budget of these projects exceeds 35 MEcus.

A number of principles are emerging from these projects, that can serve as a basis for designing large, reusable and shareable knowledge bases.

Principle 1: Structuring Knowledge according to function and scope

Rationale:

- Structured Knowledge Modelling
- Modularity
- Maintainability
- Reusability
- Knowledge Acquisition

A consensus exists: functional separation between static knowledge about the aplication domain and knowledge concerning the reasoning process.

ＯＨＰ－１

これまで１０年以上にわたって、私共はKADS-ＩおよびKADS-Ⅱプロジェクトに携わり、知識ベースシステム構築のための方法論の開発に取り組んできました。その結果得られた方法論は"CommonKADS"と呼ばれています。２つめの大きなプロジェクトは、現在なお進行中の"VITAL"で、"KADS"方法論プロジェクトと目的は同様ですが、その方法論を支援するためのツール構築に重点を置いています。"ACKnowledge"というプロジェクトでは、知識獲得と機械学習のためのツール開発を行っており、これは実際はワークベンチなのですが、そこでは知識獲得を支援できる幾つかのツールの統合を試みました。"MLT"も最終目標は"ACKnowledge"とほぼ同じだったのですが、やはり他の知識獲得ツールよりも機械学習ツールを使うことに焦点を絞っていました。

最初の４つのプロジェクトは、なんらかの形で方法論を目指すものでした。いずれも知識ベースシス

テム開発を支援する方向を目指していたのですが、ただし再利用性と共有性に焦点を合わせていました。次の３つのプロジェクト"GAMES","KACTUS","EUROKNOWLEDGE"は知識再利用と共有を目標としています。

"GAMES"というのは医学知識の再利用に目標を特定したプロジェクトです。"KACTUS"というのはごく新しいプロジェクトで、実は昨日発足したばかりなのですが、再利用可能なオントロジーの構築を目指しています。"KACTUS"については、あまり申し上げることはないのですが、ただこれは本質的に先程お話したのと同じビジョンに基づいています。そして"EUROKNOWLEDGE"は、知識ベースの分野でのなんらかの標準化を達成することを目標とした共同研究です。

そこで私がこれからお話しようとしているのは、これら様々なプロジェクトからどのようなアイディアが得られ、しかしそれがいかに再利用可能、共有可能な知識ベースの土台となりうるかということです。個々の成果についてではなく、知識共有という文脈において有用とおもわれる、一般的なアイディアや原則のいくつかについてお話するつもりです。

このように多様なプロジェクトから得られた第１の原則は、「私達の知識ベースを構造化する必要がある」ということです。一つの考え方は知識ベースをその知識の機能と範囲に従って構造化することです。その論理的根拠は、一旦知識を構造化する方法を得られれば、構造化した知識をモデル化する方法を開発できるということです。２番目に、この点は重要なので、のちほどまたお話しますが、再利用可能・共有可能な知識ベースを構築するためにはなんらかの形のモジュール性を持つことが必要です。モジュール性は、種類の異なる知識をその機能にしたがって構造化し、区分することによって得られます。モジュール性が得られれば保守性も向上し、またさきほど申しあげたように、再利用性もある程度モジュール性から得られるのです。

この知識構造化の論理的根拠のもうひとつの一面は「知識獲得の問題」です。どのような方法で構造化を行うのか。最初に分解を行うのは静的な事実に基づくドメインの知識と推論的知識の間であり、これは「制御知識」と「静的な獲得ドメインの知識」の二種類の知識の間の、ごく単純な分解です。たいしておもしろいものではありませんが、このような知識の分解を行なったプロジェクトのいくつかでは、知識ベースシステムのもっと複雑な構造を見出しています。

最上層にある知識を私達は「戦略的知識」と呼ぶのですが、これは特定の問題にどうアプローチするかに関するものです。この分野の知識は「問題解決法」、特定の問題にどのように対処するかに関するものです。その次に２番目のクラス、もしくはタイプの知識として、「応用文脈内での実際の推論にかかわる知識」があり、私達は２種類の推論知識を区別しています。「推論過程をどう制御するかに関する知識」と、「ある特定の知識ベースについてどのような推論方法をとることができるかについての推論知識」の２つです。そして最下層には、「そのドメインの実際の応用知識」があります。ここでは多様に区別されたドメインモデルに更に区分づけを行ない、特定のデバイスの構造モデルのような様々なドメイン知識を表現します。例えばあるドメインの因果的モデル、行動モデルなどなどです。

多様な種類の知識を区別することにより、再利用可能性への手掛かりが得られます。まず何よりも、診断や実際のドメイン知識のような応用タスクを分離し、order outしてあるので、タスク知識の再利

用が、"KADS"適応モデルライブラリーで行なわれたのと同様にできるようになります。このライブラリーは、どのようにして特定の種類のタスクを、内容を指定することなく、または指定する必要なく実現できるかを記述したタスクモデルのライブラリーでした。

このモデルの２番目の意味は、いろいろな標準推論を定義できることです。標準推論とは、静的ドメイン知識ベースにアクセスし質問するための標準の方法、すなわち標準化された方法です。

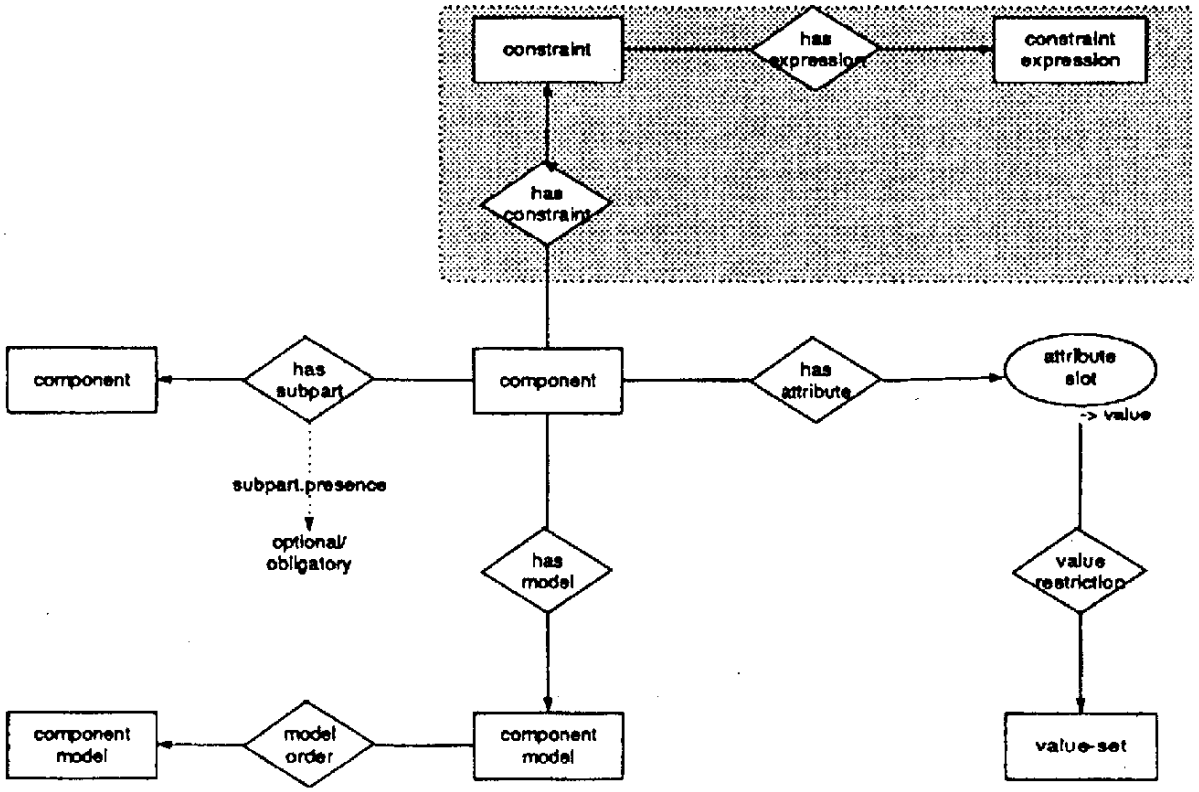
この原則の第三の意味は、"VITAL"プロジェクトで行なわれたように、問題解決法のライブラリーを開発することができることで、それにより特定の種類の問題に対する戦略的アプローチが表現されます。"VITAL"プロジェクトでは文法表現問題解決法を開発しており、それは問題の特定のアスペクトに従って特定の応用タスク・モデルを生成するものです。さきほど「このドメイン知識は様々な種類のドメイン・モデルに従って区別できる」と申しあげたばかりですが、必要なのはこれらドメイン・モデルのタイプを分類するなんらかの方法なのです。この応用ドメイン知識の特定の部分に何が含まれているのかを記述する方法が必要なのです。

このことにより第２の原則が導かれます。それは「知識ベースの構造および内容はどのようなものかを記述する方法が必要だ」ということです。ご紹介した様々なプロジェクトで、メタモデルを知識ベースの抽象的記述として用いることを試みてきました。メタモデルについて確実にいえることは、これは知識ベース内の知識項目の構造をデータベース・スキーマによく似た方法によって記述するということです。

この図式により、機能タイプの異なる知識を連結する方法が得られます。これにより、知識ベース内の特定の構造に対する手掛かりを得ることができ、それを例えばひとつの推論に結び付けることができます。またこれにより、トップダウン方式によって、知識獲得のためのガイドラインの手掛かりを得ることができます。いったん知識ベースの図式を設計してしまえば、様々な知識獲得手法を使って、「どの知識項目が必要か」「どうやってそれを構造化するか」に焦点をあてることができます。しかし、のちほどお話するように、これによって知識ベースの知識抽象化と構成要素の集約もまた可能になるのです。さらに、これは知識巡航と知識アクセスのための索引ともなります。

このようなポイントのいくつかをごく簡単な例でご紹介しましょう。ここにひとつのデバイスに関する知識を含む知識ベースのための第１のメタモデルがあります。さて概念に表現された成分があり、データベースで使うＥＲ図によく似た関係があります。これが第１の非常に単純なメタモデルです。しかしこれでは複雑な知識ベースの構造を記述するには十分ではありません。複雑な表現を扱えなくてはなりません。ここにＣＭＬ、すなわち概念モデル化言語Conceptual Modeling Languageと呼ばれる言語の一例をお見せします。これはKADS-Ⅱで私共が開発したもので、私共の知識ベースのもっと複雑なコンストラクトを扱うことができます。

私共は「状態の観念」を導入しました。これは省略によって示すことが望ましいもので、つまりこれは概念成分すなわちCAR成分の状態変数による、複雑な表現なのです。「状態」という考え方は、「成分概念のあるクラスに属する一定のCAR変数による表現がある」ということを示す抽象です。さてこれで



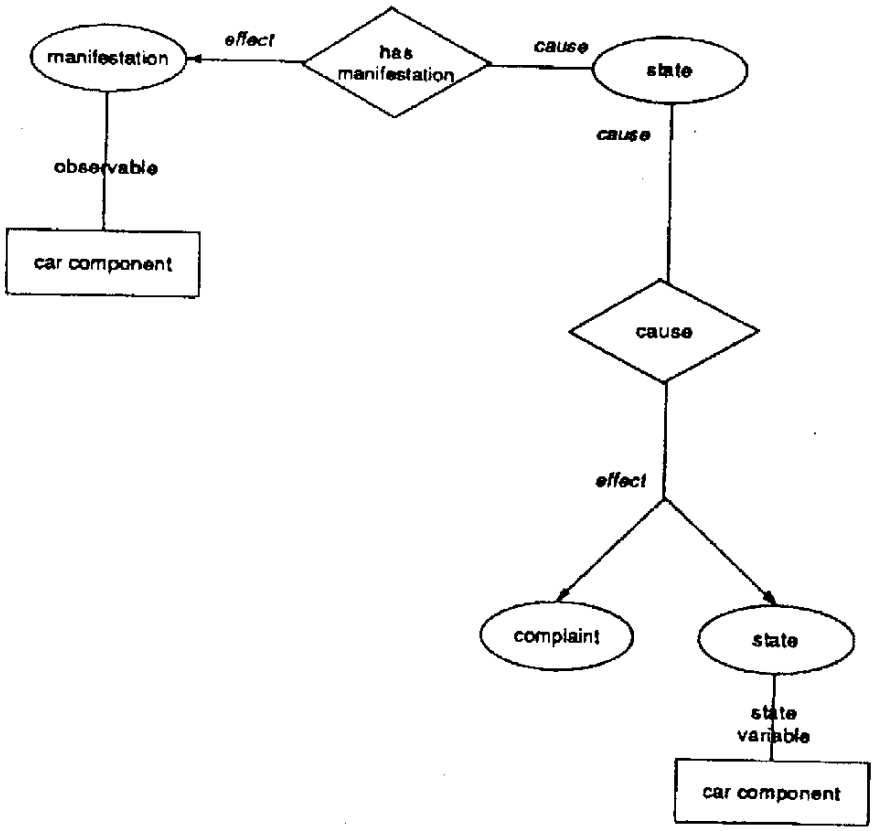
OHP－2

この「状態の概念」と、もう一つの「状態の概念」との関係を定義しました。

さらに、manifestationの観念を定義して、ひとつの成分の可観測性による表現としたので、状態とmanifestation(これをHAS manifestationと呼びます)との関係が得られました。こうして因果的関係やHAS manifestationsや状態の観念を定義するより複雑なスキーマが得られました。

さてこれで一歩先に進み、一定のドメイン・モデルを定義することができます。例えば因果モデルと行動モデルで、この前者は今定義したばかりの因果関係を含む、むしろ因果関係のインスタンスを含むもので、本質的にこれはこの因果関係の各組が１セットになったものなのです。同様に行動モデルも「組」、すなわちHAS表出関係のインスタンスを含みます。

本質的に私達が行なってきたことは、私が「知識抽象」と呼ぶものを通して、知識ベースの各部分の抽象的記述を作り出すことでした。私達が行なったことは、本質的には、ドメイン項をとって、一定の型へのメカニズムとしての型づけを通して抽象化し、スキーマ式に表現されていることを確認し、これらの式に状態とか表出とかの名前を与え、これらの抽象概念間の関係のためのスキーマを構築し、そしてドメイン・モデルについてのスキーマとは、この関係スキーマのインスタンスであると定義しました。そして最後に、今度はドメイン・モデル・スキーマが、どんな種類の組を知識ベースに保存できるかを定義します。このように、知識抽出というのは、知識ベースの特定の部分に入っている項目の内容、ま



OHP－3：CML meta−model of relations

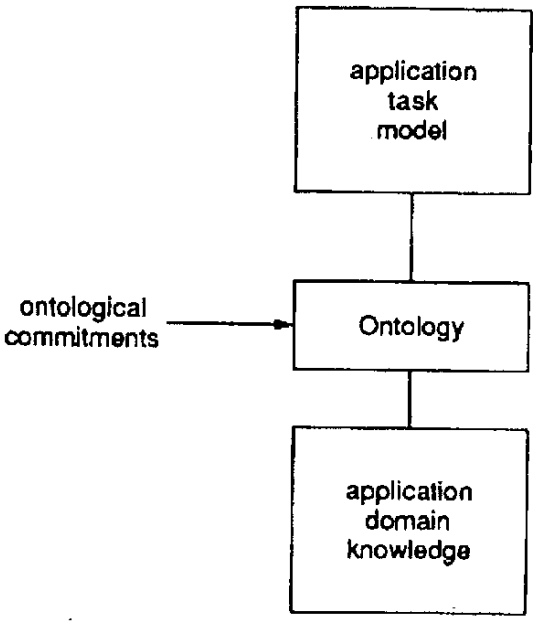
たは少なくとも構造が何であるかを記述する方法なのです。

もちろんこれは多分お馴染みのものでしょう。なぜならこれは現在私達が「オントロジーの考え方」と呼ぶものに非常に関連が深いからです。オントロジーという言葉は非常にいろいろの意味で使われており、また哲学で使われている意味とは違っているため、私はオントロジーとは「コンピュータ・プログラム内に存在することができ、そのコンピュータ・プログラムの意味を解読する表現理論である」と定義することにします。さて私達の例では不完全CARのstatusは状態stateとして表現され、次にこの状態stateは成分の状態変数の式として表現されます。お分かりのようにこの文は、知識ベースの構成要素の構造について何かを示しているだけではなく、「このような文に実際に表現されているのは世界のどの部分であるのか」についても示しているのです。これを私達は、オントロジーを分析するオントロジカル・コミットメント(ontological commitment)と呼びます。ただしオントロジカル・コミットメントとは、そのオントロジーの構成要素によって、世界の特定の特徴をどのように表現するかについての決定なのです。このように私達はオントロジーとオントロジカル・コミットメントを定義します。

さて、メタモデルとみなされるこのオントロジーを、オントロジカル・コミットメントと共に、ドメイン知識ベースと応用タスクモデルとの間の仲介的コンストラクトとして使うことができます。

これによって、私が検討したい第３の問題にたどり着きます。それはこういう疑問です。「１つのオ

ントロジーは、そのタスクからどの程度独立であると定義できるのか」「オントロジーをタスクから分離し、そのタスクを無視してどのドメイン知識ベースにも当てはまるオントロジーを構築することは実際にできるのか」。これは実際によく知られた問題で、もともとはChadrasekaranが対話の問題として述べたもので、実は、その答は相対的です。



Ontology as intermediary between task model and domain knowledge

OHP－4

私共のドメイン・オントロジーのある部分は、成分に、部分や成分等々があるように、「それがどう使われるか」からはほとんど独立なのです。デバイスにおいて、その各部分に関して推論を行ういかなるシステムも、この種の関係を使わなくてはならなくなるでしょう。しかし、デバイスの環境のためのオントロジーを構築したければ、更に別のコンストラクトを加える必要があります。例えば、コンストラクト制約が必要です。しかし、制約はデバイスに本来備わっているものではありません。むしろ私達が環境、タスクのことを語っているという、その事実から生じるものです。したがって、オントロジーのこの部分は、灰色になっている部分ですが、オントロジーのこの下のほうに示してある部分よりもある面でタスク依存的になります。

私共が第３の原則として示したいのは、「あるオントロジーのさまざまな部分のさまざまなスコープとタスク依存性の背後にあるこれらのオントロジカル・コミットメントは、そのオントロジカル・コミットメントを説明(explicate)しなければならない」というものです。基本的な考え方は、「知識ベー

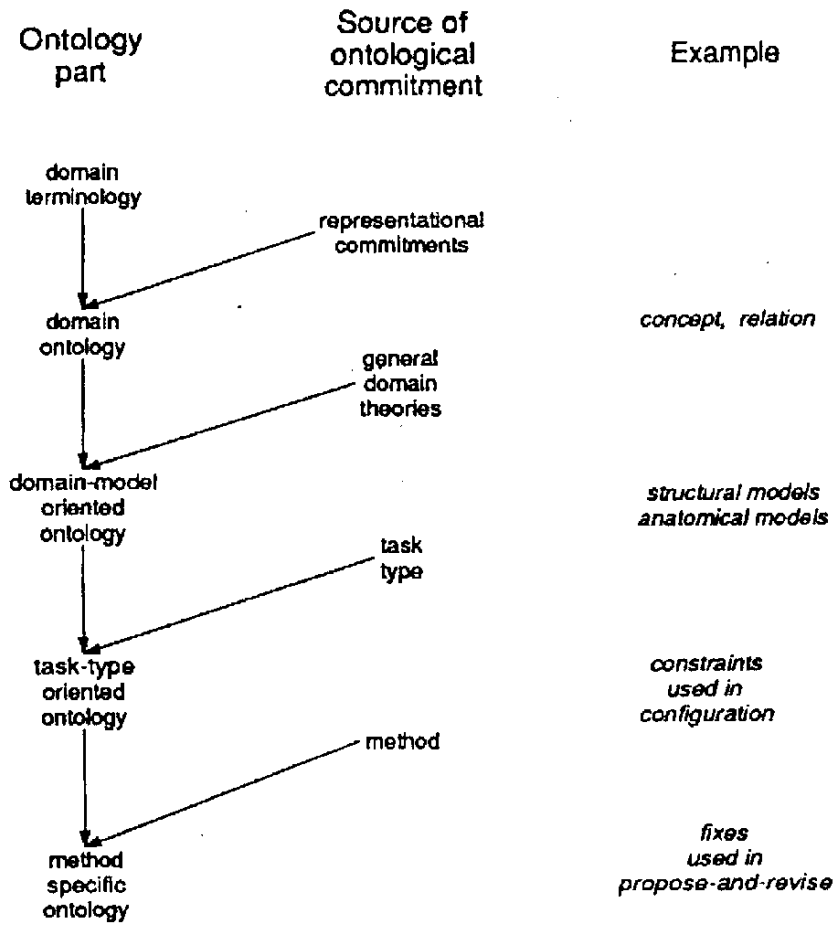
スの中のある知識項目は他の項目より再利用性が高い」というものです。本質的に、１つの知識ベース内の各々の部分は、幾つものオントロジカル・コミットメントに従属し、そのコミットメントには、非常に一般的なものもあれば非常に個別的なものもあり、タスク依存的なものもあればタスク独立的なものもあります。このようなコミットメントは重要です。なぜならこれらは、知識ベースが、特定のアプリケーションにおける推論プロセスにおいてばかりでなく、異なるソースからの知識の統合や、知識変形、知識獲得のようなその他の知識処理タスクにおいて果たすことのできる役割を、ある意味で定義するものだからです。従って、私の考えでは、そのオントロジカルコミットメントを説明(explicate)することが知識共有と知識再利用のための必要条件として決定的に重要なのです。

このようなオントロジカル・コミットメントはどこから来るのでしょうか。私達が見たある例では、タスクと応用タスクの必要条件からきたものでした。オントロジカル・コミットメントの出所として考慮に入れるべきものは実にさまざまです。

私達はそのドメインの用語から取り掛かります。当面これを、「特定のドメインを記述するために用いられる自然言語以外には、特定のコミットメントを持たないもの」とみなします。しかしこの後オントロジー構築を始めると、次々にオントロジカル・コミットメントを導き入れることになります。まず第１に、何かを表現するのに、「概念によるか」「関係によるか」「属性によるか」を指定する、つまり決定する時に、私達は表現的コミットメントを行うのです。これにより、土台となる第１のオントロジーとしての、ドメインの基本オントロジーが得られます。次にドメイン理論を導入します。私達が導入する理論は、例えば「成分が互いにどのように適合するか」などです。これにより私達の呼ぶドメイン・モデル指向のオントロジーが得られ、これはある特定のドメインについての理論に由来する特定のコミットメントを組み合わせたものになります。

第３のタイプのオントロジカル・コミットメントはタスクの種類によって決まります。環境を扱う場合なら、制約などの観念を新たに導入します。これにより知識ベースを適用したいタスクの種類に向いたオントロジーが得られます。しかしこれでは十分でない場合が多いのです。もしもタスクを達成するために、ある特定の方法を、例えば環境のためのproposeとreviseなどを選ぶ場合、オントロジーに更に新たな概念を導入したことになります。proposeやreviseのような問題解決法は、オントロジーに新たな概念とコンストラクトを導入するのです。例えば、もしproposeとreviseを用いるとするとーMarcusがPTシステムで行なったようにですがー、破られた制約をfixする場合、fixの観念を導入する必要があります。fixは知識ベースにおいて表現しなければならないものなのです。しかし方法についての非常に個別的な概念なのです。fixは方法依存的です。構成依存的でさえなく、方法依存的なのです。こうして最後には方法特異的なオントロジーに行き着きます。

基本的な考え方は、「異なるタイプのオントロジーにおいて行なわれる、これら異なるオントロジカル・コミットメントを区別することによって、知識ベース内の異なる部分を区別することができる」というものです。これらの部分は多少とも再利用可能性があります。このようなオントロジカルコミットメントを明らかにすることにより、どの部分をどのような環境下で再利用できるかを実際に知ることが



Different types of ontological commitments and their corresponding ontologies

OHP－5

できます。

これがまさに"GAMES"プロジェクトで私共が行なおうとしていることなのです。先に申しあげたように、このプロジェクトは医学知識ベースシステムの構築を目指すものです。基本的に私共が構築中のシステムには３つの異なる種類の事物のライブラリーが存在します。第１に、医学ドメインからの包括的タスクのライブラリーがあります。この中には「診断モデル」「治療計画モデル」「監視モデル」が入っています。ライブラリーの第２の部分は、オントロジー・ライブラリーで、ここには「知見」「疾患」「治療」などの物事についての、定義を含む多数の理論が入っています。つまりオントロジー・ライブラリーには複数の異なるオントロジーが入っているのです。そして最後に医学用語の辞書があり、これは本質的に標準的医学語彙のデータベースです。

特定のアプリケーションのための知識ベースを構築したい場合、知識エンジニアはまずはじめに、ここに示したタスクモデル・ライブラリーの中の様々な構成要素から、そのアプリケーションのためのタスクモデルを構築します。それによってタスクモデルが得られ、またそのアプリケーション・ドメイン

知識が満たすべき、知識役割が得られます。

この知識エンジニアが次に行うのは、オントロジー・ライブラリーから特定のオントロジーを選びだし、それを「アプリケーション・オントロジー」と呼ぶものに構成することです。この中に「疾患」「知見」「可観測」などの観念があります。知識エンジニアはアプリケーションのためのオントロジーを構築します。そして知識役割をアプリケーション・オントロジーに写像し、そして最後にこれらのオントロジーを知識ベースによって開始する必要があります。この開始プロセスは、既存の知識ベースのライブラリを使って行なうこともできるし、医学辞書を使ったエキスパートシステムによって行なう事もできます。現在の私共の持っているシステムは、エキスパートに対して、知識項目についてのごく個別的な質問をすることができ、そしてオントロジーを定義した場合、「Intoness知識ベースのような利用可能な知識ベースをどうしたらアプリケーション知識に移入することができるか」を考えています。この一例は、このような高度に抽象的な原則を基礎とすることによって、知識ベースシステムを柔軟かつ効率的に構築するための具体的なシステムをどのようにして実際に作り出すことができるかを示しています。

終わりに、私のお話の主なポイントをまとめてみたいと思います。まず第１に、知識構造化によるモジュール性と再利用性によって、複雑性に対処し、再利用性を促進できることをお話しました。第２のポイントは、メタモデルは、知識構造を記述し、知識減算(subtraction)を通じてモデルの複雑さを減じるための重要なツールだということでした。３番目は、オントロジーとは、オントロジカル・コミットメントによって注釈されるメタモデルとみなされ、そのようなオントロジカル・コミットメントを説明することにより、さまざまな程度の共有性が得られるということです。私がご説明しようとした基本的なポイントは、「知識をその機能やそのオントロジカル・コミットメントの性質にしたがって区別すること」、また「メタ記述は大規模知識ベースの再利用性と共有性の鍵である」ということです。有り難うございました。

座長：

Thank you very much for interesting talk.

それでは、質問とコメントを…。

お話は、知識共有のために、３つのプリンシプルですね、１つは知識をどうストラクチュアリングするか、２番目は、「メタ・モデル」をどういうふうに考えるか、３番目は、「オントロジカル・コミットメント」という言葉があったと思うんですけれども、それを使ってどういうふうに知識を扱うか、という話であったと思います。

Swartout:

ＵＳＣのBill Swartoutです。あなたはオントロジカル・コミットメントをどう表現なさるのか、知識ベースをどのように注釈してそのオントロジカル・コミットメントが何であるかを示すのかについて

少しご説明ください。

Wielinga：

それはまさに私共が今研究中の問題の一つです。それらをどうやって表現するか、オントロジカル・コミットメントとは何なのか、またオントロジカル・コミットメントの背景となる理論は何かなど、多くが未解決です。

"GAMES"システムで、私共はオントロジーライブラリーの様々な部分について、そのオントロジーの範囲は何かを表現しました。もしタスク依存的であるならどのようなタスクに適用されるのか。その背景にどのようなドメイン・コミットメントがあるのか、それは例えば基本的前提は何か、単一疾患か多疾患かというような典型的な事柄か。これは医学領域で用いるオントロジーになんらかの影響を与えています。

従って、私達はこのようなオントロジカル・コミットメントにアドホックな表現を与えていますが、今のところ完全な理論にはなっていません。それを研究中なのです。

座長：

まだ質問・ご意見あると思いますけれども、時間が迫ってきましたので。

ではもう一度拍手をお願い致します。

(4)「知識表現とデータ」

座　長：

それでは４番目の発表に移りたいと思います。日本語のタイトルで「知識表現とデータ」、で、英語のタイトルでは"Viewing Data Through a Knowledge Representation Lens"ということで、AT&TのRon Brachmanさんです。

Ron Brachmanさんは、１９７７年にハーバード大学のApplied MathematicsでPh.D、で、"What is link?"で有名な、あのBill WoodsさんがThesis Advisorです。それで、ハーバードを卒業してから、BBN、Bolt, Benarek and Newmanというところに勤められて、それからFairchild/Schlumberger Palo Alto Research、そのときに、Kryptonという言語ですね。最初はKL1だったんですが、その頃知識表現のシステムというのは、名前をいろいろと凝っておりまして、KryptonとかKandorとか、いろんな名前が出てきたんですけれども、そういう研究をされていて、現在は、Bell Lab.のDepartmaent Headをされており、Artificial Intelligence Principles Research Departmentというところに所属しております。

AT&T Bell Laboratories, AI Principles Research Department
Department Head　Ronald J. Brachman

ミゾグチサン、ドウモアリガトウゴザイマシタ。ミナサン、オハヨウゴザイマス。スミマセンガ、ワタシハニホンゴガジョウズニハナセナイノデ、エイゴデハナシマス。スミマセン。

私に割り当てられましたこの短い時間の中で、知識交換に関する貴重な講演から、私たちが有している「非常に大きな知識ベース(VLKB)」への情報入手に最も関係の深い問題に至るまで、お話をして行きたいと考えます。特に、今までの講演のほとんどで話されていない話題の中で、私たちが将来成功するためにはきわめて重要であり、また、不可欠と考える話題を紹介したいと思います。

特に、学術的で人工的な知識社会の外側では非常に一般的な、現実の世界のデータを処理する必要性についてお話しようと思います。VLKBに関する私たちの研究は、現実にはまだ着手したばかりの状態です。それは非常に新しいタイプの分野です。何年間にもわたり、非常に大量のデータが、これまた非常に多数のシステムにより集められてきました。この分野では、私たちがどのような努力をしようとも、それが次の１０年間にかかるものであるとしても、今までのシステムで今までのニーズに対して集められてきたデータは、やはり同じ理由で、同じ方法で継続して集められることになるでしょう。そのため、データ処理の世界や科学、ビジネスの社会により集められたデータを処理することができるように準備をしておくことが、非常に重要であると思います。そうしないと、キーとなる科学やビジネスの応用面から隔離される危険性を有しているからです。もちろん、ビジネスの話といえば、私たちの将来の健康がおそらく最大の中心となるでしょう。積極的な側に立てば、VLKBシステムを開始するためのジャンプ

をするチャンスにつながるものと見ることができます。

データについて大きく焦点を当ててみたいと思います。私のコメントは、この会議で既にお聞きになった他の講演を補足するものであるということを明確にしてみたいと思います。これは「自然言語の情報や、VLKB企業のテキスト・コーポラや、他の面がそれほど重要ではない」ということを言っているのでは決してありません。しかし、世界中に存在しているデータについて少し時間を割いてみることが非常に重要であると考えます。

その結果、話の行き着くところは非常に簡単なものであります。私は基本的にはある一つの主要なポイントを示したいのです。すなわち、データがそこにあるのであれば、その収集作業は継続され、それを非常に真剣に行う必要があるということです。

私の話すことは、この２日間にわたって行われた講演の補足となると考えられる研究問題を僅かではありますが特定することです。これは私たちのＫＢ＆ＫＳ研究で考えることが重要なものなのです。また、最後には私たちが構築した実験的なプロトタイプを非常に簡単ではありますがお見せすることで、ＡＴ＆Ｔで行った特定の研究を例示することとします。

第１のポイントは非常に簡単なものであり、皆さんが納得して下さるのにさして時間はかからないと考えます。事実、この会議用の予稿集の表紙をご覧になれば、本会議の組織委員会の皆さんが賢くも示して下さっています。私たちが直面している問題の１つは、この世界がデータという海でおおわれているということです。その大きさたるや、まったく面食らうほどです。

この「データの海」がどのくらい広く、また、どのくらい深いかを理解するために、私たちが毎日目にする各種データ源と収集のタイプについて一度考えてみましょう。このデータは八百屋やデパートにあるキャッシュ・レジスターや販売時点端末から自動金銭支払機や非常に広範囲の国際金融取引に至るまで、範囲は広がっています。私にとってより親密なものとしては、数多くの国際電話ネットワークから瞬時に発生する大量の情報もあります。

患者の保険に関するデータから体温や放射線治療データに至るまで、病院の中で収集されるデータもあります。そのほかにも生物学的なデータ、化学データ、天文学、人工衛星、スペースシャトルなどに関する科学データなどもあります。この２５年間で収集された大量の軍事関係のデータは、おそらく圧倒されるほどのものでしょう。

私たちは映像やビデオの形で収集されたデータも持っています。これらについては、昨今の議論では比較的無視されることが多いのが事実ですが、人間の知識共有の大部分は、グラフィックス、映像、及びアニメーションに基づいています。そのため、もし私たちがテレビ・ラジオ局やビデオ・クリップのライブラリーを構築しているテレビ局を取り上げるならば、これらのものを私たちのVLKBシステムに組み込むデータとして扱うことは非常に重要なことになります。一般的なＰＣスライドによるプレゼンテーション・プログラムにおいて、これと同じ様なスライドを構築してみようとするように、映像データは私たちが関心を持たなくてはならないビットの量において単純なテキスト・データをはるかに圧倒するものです。

「私たちが世界中で有しているデータの量はただ増え続けるだけである」という事実は議論の余地のないところです。毎日毎日、会社・政府がオンラインで新しいデータベースやデータ収集システムを持ち込み、私たちが知識表現システムやVLKBシステムに向かって動きだそうとして努力しようとしても、このデータはコンピュータでデータベースや通常のフラット・ファイルに収集され続けるのです。そのデータが商業の世界に、科学や連邦の世界にとって重要であるとし、またそれが私たちを圧倒し続けるとすれば、この事業で研究する必要がある重要なキーとなる研究テーマはいくつかあると思います。これは非常に短いリストではあり、決して総括的なものであるとするものではありません。

私たちが心配するのは、知識ベースの中で非常に多くの個人的なものを対象としているということです。ＣＹＣプロジェクトを例にとれば、今、耳にする議論の殆んどと、今日後から聞く議論には、非常に多くのオントロジーや概念、等級などの取扱いが含まれています。これは非常に重要な作業であることは明らかで、過去において、私たちがデータベースによって見ていたものとはたいへん異なっています。一方、私たちはこれらの知識ベースを、膨大な数の個人対象のデータの中に植え込む必要があります。たとえば、各個人が今年かけた電話回数はVLKBシステムには非常に有用なデータとして入力されます。

このことは効率に関して良い機会を与えることになります。特に個人の対象構造が非常に規則的になってきており、一方、私たちが自分のオントロジーに見いだす概念構造は、概念ごとに非常に変わってきているからです。しかし、やはり、その量の多さに圧倒される危険性があるという点については同じです。

次にもう１つの重要な問題は、もし私たちがこのデータの重要性を認めたとして、どのようにそれをこのVLKB管理システムに入れていくかということです。これは２つの部分に分かれます。１つはVLKBをある非常に大きな対象セットにより初期化することであり、もう１つは私たちの人工衛星が新しい気象データを収集したり、軍事システムが新しい映像データを収集したり、電話やＡＴＭシステムが通話や取引に関するデータを収集したりするように常に更新を行うというものです。

アーキテクチャは非常に重要な問題です。単純に考えて、ある種の知識ベースをもつシステムが曖昧な方法で一つまたは多くのデータベースを持つシステムに繋がれていたとします。これらの２つのシステムはどのように互換するべきでしょうか。

システムは２ついるのでしょうか。それとも１つだけのものなのでしょうか。

このデータを処理する責任の配分の問題を話す必要が有ります。どちらのシステムがそれを分類するのでしょうか。どちらのシステムが照会をするのでしょうか。どちらのシステムがこの質問に答えてくれるのでしょうか。２つのシステムを持たなければならないのでしょうか。既存のデータベースを利用しようとすることは、少なくともそれが関係のあるものである場合には私たちにとって非常に重要です。このことは、早々に述べましたように、知識ベースが本質的に理屈を有している為、万能薬であることはおそらく無いでしょう。しかし、私たちにとって関係がある場合には、非常に大きなデータベース技術を考慮することは重要です。もう１つの重要な問題はこのデータの多くが非常に広範囲に及ぶセット

に記憶されているということです。このことは場所が地理的に散らばっているということばかりでなく、ハイテクの新オブジェクトに基づくデータベースからUNIXやPCsに至るまでの非常に異なったデータ・フォーマットまでを含んでいるということです。

その他の関心のある問題はと言えば、例えば、知識ベースの更新が有ります。入力されるデータばかりでなく、もし私たちがデータを開発し、規則的なパターンが見つかったと決定したならば、知識ベース、オントロジーを変更したくなるでしょう。それはデータ自身にどのような影響を与えるでしょう？個人に関する生データについてはどうでしょう。

これは大まかに言えば、データベースにおける考え方のアップデートに関する問題です。そしてこれは極端に難しい問題で、VLKBを利用しようとする場合に、より難しくなると思います。

もう一つの重要なことは、データの中の変更をモニターすることです。ドメイン（活動範囲）中の変更箇所を探そうとする場合に、これら両方を利用したくなります。これは知識ベース用語にはいくらかの高レベルの定常オーダーを与えたくなるものです。しかし、非常に低いレベルの変更箇所を察知する、積極的なデータベース技術も利用したくなります。例えば、地震の時などに、あまりに多くの通話が行われて電話線の幹線がパンクしそうになると、私たちは、VLKBシステムが起動して自動的に地震を察知し、私達が何秒も何分もかけて手でデータを処理する前に、そのトラブルを私たちに警告してほしくなります。

私たちが、知識ベースに持ち込まれた、通常のデータ収集方法から入ってきたデータを見ると、これには翻訳の問題も生じます。例えば、あるリレーショナルデータから、世界のよりオブジェクト指向な考え方に翻訳するというケースです。これは私たちのプロトタイプの中で関係しているため、これについてもう少しだけ述べてみたいと思います。ここで生じるもう１つの問題は、非常に低レベルの単純なものですが、とても重要な問題で、データ収集の世界と知識ベースの世界との間の名前の互換性に関してです。

もう１つの非常に重要な問題は、テキスチャル・コーポラに関連して、Susan Armstrong氏が昨日の講演で述べられましたように、私たちが所有している最高のデータベース技術にもかかわらず、私たちが見つけたほとんどのデータは、VLKBシステムに取り入れたいとおもっている大規模データベースにおいて、いくぶん損なわれてしまうということです。ピュアでクリーンなデータを得ることは非常にむずかしいことです。最小限でも、私たちはナル値、もしくは雑音によりシステム的に、もしくは、ランダムにこわれているデータ構造、ファイル、またはデータの中の単純な穴を処理する必要があります。私たちがリレーショナルデータベース中のクリーンで、クリアーであると考えているデータは何らかの点で多くの場合粗悪なものです。知識の表現の形式化、例えば、とくにYorick Wilks氏の言う「論理的見解」に影響されているものは、データの広範囲にわたる破壊や質の欠損を扱うにははあまりにも壊れ易いものです。しかし、もしこのデータを真剣に受け取るつもりならば絶対にこのことを考慮しなければなりません。

最後にパフォーマンスの問題を検討することが非常に重要です。誰であろうとも、もし私たちのユー

ザーにシステムを処理する我慢強さがなければ、何日も、何週間もかかる何十億という数のオブジェクトを見ているだけのことになるでしょう。そうなれば、私たちの事業が学術的なものになってしまいます。そのため、VLKBシステムにおける性能の問題は現実的に非常に重要です。

私の残りの話は、データから入ってくる知識を得るまでの簡単な概要を見ながら、ＡＴ＆Ｔにおいて私たちが行ってきた実験を紹介してみたいと思います。

インターラクティブな、人間の調査に関するある作業を取り上げてみます。データアナリストが腰をかけ、パターンやトレンド、例外的な状態のための大量のデータを調査します。私が思うに、これは産業及び科学の両方において非常に一般的な問題であります。私たちはそれをデータ考古学と呼んで来ました。より自発的な活動の代わりに、データ発掘システムと呼ばれるものが見えてきます。そこではパターンを見ながら誘導的なアルゴリズムを操作します。データ考古学は本質的に、一般的な考古学者が破片や残骸、もしくは少量のデータからどのようにパターンが生じるかを発見するのとほとんど同じような、人間の作業を対象としています。

このタスクの有用性は詐欺の探知、クレジットカードの利用、電話の利用から市場データ分析などに至るまでのビジネスにおいて明らかです。また科学分野の研究においてもM規則性、たとえば、金融データ、軍事データ、気象データなどを常に探求しています。インターラクティブなデータ分析の現在の環境は、それらが現代のデータベース技術に依存しているものの、データ考古学者にとってはあまり援助にはなりません。最善の環境においてさえ、人間がキーパターンを見つけるのを困難にする使いにくい、統合性のないツールであることが分かっています。

このプロジェクトにおける私たちの仮説は、今日あなたがたが聞いたこの種の知識表現システムは、これらの種類の分析アプリケーションに対する統合力として働くというものです。もし知識表現システムをこの統合モデルとして利用できるなら、どのような利益が得られるのでしょうか。それらのいくつかは明らかですので、それをすばやくここに挙げて見ましょう。

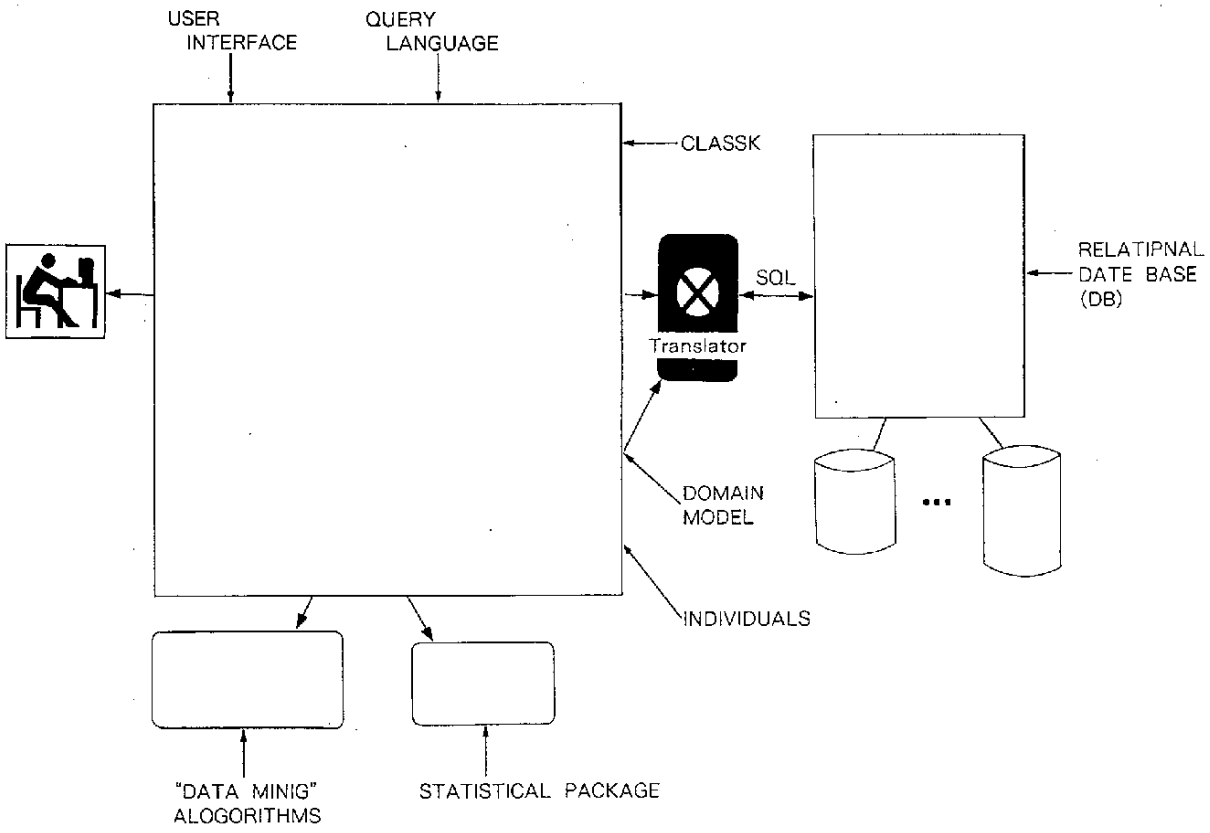
その内の１つはデータ・アナリスト用のドメイン（活動領域）のより自然なモデルです。個々の人々、場所や物を確認することを困難にする単純なリレーショナルモデルの代わりに、オブジェクト指向のモデルはより自然で、洗練されたドメイン（活動領域）の概念を与えることができます。そのため、より自然な複雑な概念を非常に簡単に表すことができます。たとえば、顧客の購入量が１月当たり１０％増加するようなデパートのドメインや、休日の売上が２０万ドルを超えるデパートにおける共通概念がそうです。これらは知識表現言語で表すことがきわめて簡単で、かつ、直接的であります。

もうひとつ重要な利点は問い合わせの結果といっしょに問い合わせを統合できることです。これはデータベースのセッティングでは、はるかに困難なものです。私たちは問い合わせに名前をつけ、再利用し、それらの抜粋を作り、それらをパラメータ化し、知識表現セッティングにおける問い合わせの結果が最良のオブジェクトになります。

さらに、考えられる知識表現システムの一般的な長所は、データの考古学者にとっても非常に有用なものとなるでしょう。そのため、ドメイン知識に対する一般的な推論能力はここでも利用できます。た

とえば、以前にBill Swartoutが言っていたこの種の自動分類や一般的な分類表記からの遺産を利用できるかも知れません。知識表現システムを利用する上で私たちが見いだすその他の重要な長所のすべてはＡＩに入っています。

私たちのプロトタイプの、非常に高レベルのスナップショットは少しこれに似ています。私たちはこれをIMACSと呼び、それはインタラクティブ市場分析及び分類システム(Interactive Marketing Analysis and Classification System)の略です。IMACSの核はこの四角の中にあります。これには私た



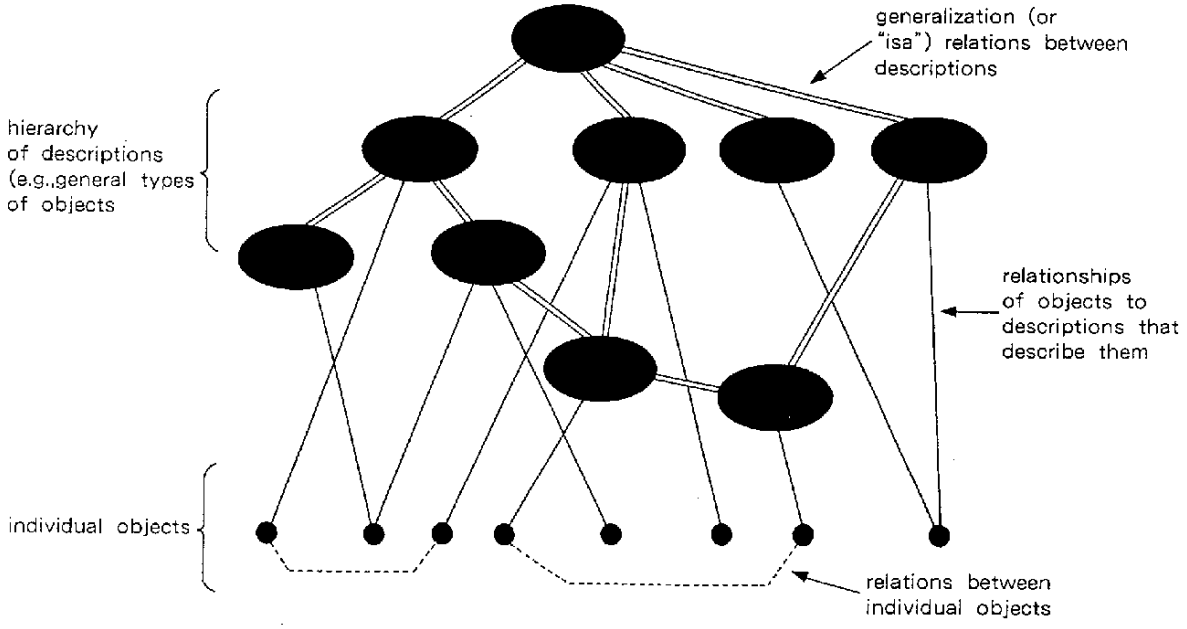
ＯＨＰ－１

ちがCLASSICと呼ぶ知識表現システムを含んでいます。Billが少し前に述べたようにこれは論理システムを表したものです。もう少しだけこれについて話してみます。私たちはドメインのエキスパートと協力してCLASSICにドメインモデルを構築します。次に、トランスレータを通じて生データから入力される、膨大な数の個々のデータと共存させます。

私たちはこれからお見せする複雑なユーザー・インターフェースとリレーショナルデータベースやSQLの言語ではなく、CLASSIC自身で話すことのできる問い合わせ言語を持っています。IMACSシステムはデータ発掘や誘導的な学習アルゴリズムへアクセスできます。そのため、自動ルーティンはユーザー

によるインタラクティブ分析に対して利用できます。また、S&SAS、GPSSのような、従来の統計的分析パッケージへアクセスすることも可能にしています。これは市場調査を行う場合に必要なデータ分析には絶対的に重要です。

さて、今申し上げましたように、CLASSICはこのシステムの心臓です。時間が余りないので多くは話せませんが、実際のところ、これについては東京での昨年の第五世代の会議におきまして、すでに十分に話しております。キーポイントは、CLASSICは単純なものであるということです。これはオブジェクト指向であり、あなた方の家族のように、クラスを受け継いだ正常な形のヒエラルキーを有しています。公式的に特化した表現言語であり、公式の意味論を持った推論規則であり、また、Billが述べましたように、知識ベースを管理するためのＫＲＳＳバナー・サポーティングソフトウエアの下にあります。ですからこれは統合システムなのです。



For IMACS, CLASSIC provides
- an object-oriented, natural domain model ; expresses general (class) as well as specific (individual) facts
- inferences, such as inheritance and "classification" (determines automatically when an individual satisfies the criteria of a class)

ＯＨＰ－２

しかしこれには限度があります。これはＡＩや知識表現システムの万能薬ではありません。そのために私たちが利用可能にしようとしているのであり、それを高速で操作しようとしているのであり、内部利用の範囲内でそれをポータブルにしようとしているのです。

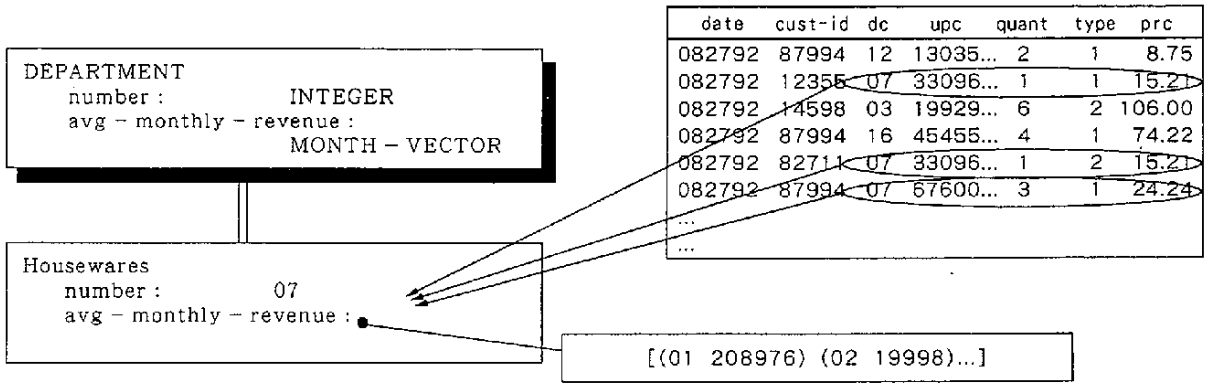
CLASSICは普通の文章表現に加え、ディスクリプション（表現）を管理しています。つまり概念やクラス、もしくはフレームであります。Billも言っているようにディスクリプションロジックです。またここで申し上げておかねばならないのはベル研究所のＣ言語とCOMONN LISPの両方で、CLASSICが実行されているということです。図式的にみれば、CLASSICの知識ベースはこのようなものです。OVALは通常のISAヒエラルキーにおける高レベルの概念構造です。私たちには数多くの個々のオブジェクトがあり、それぞれが多くの概念で表すことができます。また個々のオブジェクト間の関係を議論したり、表現したりもします。IMACSシステムの設定に当たっては，CLASSICはオブジェクト指向で自然ドメインモデルを与えています。これについてはもうすでに何回も述べました。これにより私たちは一般的情報を表すことができたり、特定の事実を表すことができたり、また重要、かつ、有用な推論も提供できます。

CLASSICはこの種のアプリケーションでは様々な点で有用な推論の形式を大体１０～２０提供しているということです。その中心は分類であり、与えられたオブジェクトに応用する知識ベースにおける表現をすべて示しています。データ考古学作業のセッティングの際、アナリストが関心を持つドメインのモデルをCLASSICで作らなければなりません。デパートのモデルをパラダイムとして例にとってみましょう。私たちはデータを分析したいエキスパートと座って作業します。そして彼らの世界を、彼らが受け継いでいるリレーショナルスキーマや、専門的な言葉ではなく、自然なオブジェクト指向な方法で話してもらいます。われわれは、その気になれは、オブジェクト指向知識エンジニアリングをすることが出来ます。そこでわれわれはこの種の継承ヒエラルキーをつくり、それぞれの概念に対し、その属性と、バリュー、数値に関するいわゆる役割、制限について語っているディスクリプションを所有します。

ここで重要なことは、既存のデータベースにあるものは何も参照しないで、世界中のエキスパートの意見に基づいてドメインモデルを構築するということです。次に、もちろんのことながらこのオブジェクト指向クラシック構造の中に表の形で入ってくるデータベースから、現実のデータをマッピングする方法を見つけなければなりません。これを行う方法は、CLASSICでいわゆるプリミティブクラスの情報を取ってくるSQLを手動で提供します。これらは基本的な概念であり、完全な、必要性のある、十分な定義は持っていません。これの良い点は、一度行なえば、プリミティブクラスに基づいたクラスがすべて新たに生成されるという点です。新しいSQLを自動生成してみましょう。

ユーザーまたは知識エンジニアは、ある限定された有限のクラスについて、一度だけSQLコードを書くだけです。ここで重要でおもしろいことは、知識ベースではデータベースには対応するものがないリアルクラスを生成できるということです。もし私たちのデータベースが商店の購入表で、「いつ」「どの品物を」「いくらで」「誰が払ったか」について簡単に書かれているとすると、情報を、家庭用品の売り場のようなオブジェクト表現のように見えるものに、実際に変換することができるのです。これは売り場の数や毎月の売上が入っています。そのほかの特徴はオブジェクト指向に見えることです。

さて、SQLは非常にややこしいもののようです。事実、これがそのポイントの一部なのです。今、データアナリストが毎月の売上のベクトルを含むデータ分析をしようとする場合、デバッグ支援がほとんど望めない面倒なSQLコードを書かなければなりません。一旦プリミティブなSQLを書いて、生データ



- Need to provide SQL manually for "primitive" classes
  - e.g., need to define how to build DEPARTMENT class from actual relations in DB
  - done once
  - newly created classes can be populated automatically
- Can create real classes that don't exist in DB

OHP－3 ：Mapping from the Data to CLASSIC

から知識ベースを構築したならば、ユーザーは再びSQLを書く必要はありません。ユーザーは私たちの知識表現及び問い合わせ言語で書くだけで良いのです。

データベースから翻訳する場合に生じるのはネーミングの問題です。データベースにからむデータは、記号や数字の使用について特定の規約を持っています。また知識ベースもまったく異なった規約を持っています。たとえば、データベースの中の文字列を取り込み、知識ベースの中で単一の識別子を持とうとするかも知れません、またデータベースの中では１つのオブジェクトに対してただ１つのキーを持ちますが、知識ベースの中では終わりにキーを付けているクラス名付きの識別子を作ろうとするかも知れません。もしくはキー自身が、このようにまったく対応付けなしに、直接的に私たちを単一のユニークな名前に誘導しようとすることもあります。例えば、データベースの例Ａ７は知識ベースにおける家庭用品を意味しているかも知れません。

● Key↔name	"Joe Smith" ↔Joe_Smith
● class + key	ACCOUNT##1124ABD112
● KEY ⇒ name	A7 ⇒ Housewares
	promo_id=01164 ⇒ SpringClearance
● many−to−one	M ⇒ {Male,Married,Manager,...}
● more complex	(FILES has−features? T)⇒
	*exists (select * from advanced where...)*

OHP－４：Naming

私たちには、多数対１のマッピングに関して非常に厳しい問題があります。データベースの中の記号Mは、ある関係においては男性のMを表しているかも知れませんが、他の関係では既婚者のMを、また、別のケースではマネージャのMであったりします。同じ問題が知識ベースの中で別の形で現れます。ネームの間に非常に複雑なマッピングがあるということになります。このような、あるものが真のものである特徴をもっているCLASSICの構造は、複雑なSQL問い合わせ法にマッピング処理を行わなければなりません。そのため、ネーミングはこのようなセッティングでは明確なものでないが、とても重要な関心事となります。

稼働中のIMACSの絵を１～２枚お見せて、終わりにさせて下さい。皆さんはマルチメディアの作品を楽しんでくれると思います。後ろの方にお座りの皆さんには、細かい部分が見えなくて申し訳ありませんが、細かい部分は重要ではありません。重要なのはインターフェースのスタイルなのです。データアナリストがまずするのは、データの回りを調べ、それを見つめ、その感覚をつかむことです。それには非常に広範囲の方法でリレーショナルデータを表示することも含んでいます。典型的な事なのですが、SQLを書く場合、ビューイングにデータを正しく反映させるため、問い合わせでとても混乱しなくてはなりません。IMACSでは非常に単純なテンプレート機構を開発しました。これを使用すれば問い合わせ言語で表現ができ、表の中に新しいコラムを作成することができます。例えば、安売りにきた人が買い物をするパーセンテージは、データベースの中の単一の項目には関連していませんので、それは表の中の通常の項目であるかのように表示して終了します。例えば、消費額と購入品の数は直接的にデータベースの中に表すことができますが、販売購入件数のパーセンテージは、販売時点でなされた購入件数をカウントする式で計算しなければなりません。そのため非常に柔軟なテンプレート・エディターとプレゼンテーション機構があるのです。データアナリストが行うもう一つの重要な作業は、セグメンテーションと呼ばれるもので、これはドメイン内にあるオブジェクトを興味のあるサブクラスに分解し、一定の挙動を示します。これはマーケティングでは確かに非常に重要なものです。その結果、最低から最高まで購入件数を見つめて、顧客全員のグラフを作ることもできます。また、データの中に自然のブレークもあることもあります。この世界を知れば、これらは自然のブレークのように思われます。私たちのインターフェースでは実際にライン・セグメントを引くことができます、それは問い合わせ言語で自動的にセグメンテーション・コマンドに翻訳され、私たちの概念ヒエラルキーに新概念を作ることができ、それにネームを与えられます。そのため、例えば、このグループを「非常客」と呼び、以前には私たちの知識ベースに存在しなかった新しい概念を作成できます。

中間のグループは「半常客」と呼ばれ、それより上のグループを「常客」と呼びます。これは彼らが、本当に関心のあるのは大量購入客であるからです。これらの概念をいったん構築したならば、それらを来月にも再利用でき、その働きをチェックすることができます。事実、IMACSはまたデータベースにおける変更のインタラクティブニモニターをサポートします。私たちはFORMSと呼ぶ一連の項目も所有しており、これを使用すれば何度でも分析を繰り返し行うことができます。これは問い合わせ言語による、単純に抽象化されたステートメントで、変数、個々の役割チェーン属性などを用いています。そこで

FORMSに埋めることができ、以前には見たことがないような興味のある表を入手することができます。例えば、各ユーザーの購入品を購入が行われた部門毎にセグメンテーションを行うことができます。通常のアナリストには書き込むことがむずかしいかも知れませんが、私たちが一度このFORMSを書いて、それをすべてのデータアナリストと共有することが可能となります。
IMACSでは他にもいろいろなことができます。今はそれらの詳細をお話する訳にはまいりませんが、これだけ言わせて下さい。インターフェースは知識表現システムCLASSICを使用してグラフィック情報、概念情報、モニタリングの変更、ヒストリー、再利用可能なフォーム、テンプレートとセグメンテーションを統合して、非常に複雑なデータ考古学的な作業の支援をしています。

最後に私がこのプレゼンテーションの主要点であると感じていることを繰り返して申し上げて、結論といたしたいと思います。単純にして最も重要なことは、世界はてんでばらばらな大量のデータで満ちており、しかもそれは常に増加する方向にあります。事実、この話をしている間ですら、世界は人工知能知識ベースの歴史全体に相当するよりも多くのデータを収集していることでしょう。それはかなり大げさな予測であると思いますが、私たちの非常に大きな知識管理システムを構築する上で考慮しなければならないもので、さもなければ商業の世界や科学の世界では私たちの研究を真剣に取り上げてはくれないでしょう。非常に多くの研究テーマが生じておりますが、そのいくつかについて手短にお話をさせて頂きました。これらは私たちの研究において重要かつ中心的な問題であると信じています。IMACSについてVLKB管理システムを要求する多くの問題の内のひとつに対するアプローチとして少しお話をさせていただきました。

座長：

Thank you very much for your clear presentation and inpressive slides.

それでは、かなり時間が押してきたんですけれども、質問・コメントがありましたら…はい。

質問者１：

まず最初に、私はデータベース・システムにおける知識発見というテーマで研究しておりますのであなたの話にはとても興味を持ちました。一つだけ非常に短い質問があります。あなたの言われるIMACSシステムでは、データベース・パートとしてリレーショナルデータベース・システムを考えておられます。オブジェクト指向データベース・システムをあなたのIMACSシステムに採用した場合に、IMACSに含まれるデータ発掘ワークベンチの技術を、オブジェクト・データベース・システムまで拡張することが可能でしょうか。

Brachman：

非常によい質問です。それはオブジェクト指向データベースがますます一般的になってきているからです。商業の世界ではまだ時間がかかるかも知れません。データベースを利用しているIMACSのバック

エンドは、まったく正直に言いまして、むしろ一般的な方法で構築されますから、ひとつの問題から別の問題へと移動できるのです。

マーケティングから科学的なデータ分析に至るまでかなり簡単に移ることができます。しかし、私たちが書いた基本的なメカニズムはSQLを書くことも入っています。それをすることについては十分な理由がありまして、それは普遍的な規格があるからです。リレーショナルからオブジェクト指向問い合わせ言語までの翻訳をリライトする方法はきわめて直接的であるとは思いますが、あるものから別のものへ単純に変更する切り替えは簡単なものではありません。ですから作業の進め方は分かっているとは思いますが、異なった技術に移行するにはかなり多くのことをしなければなりません。

質問者１：

どうもありがとうございました。

質問者２：

あなたが研究を続けてこられたデータベースの大きさとそのスピードについて一言お願いします。

Brachman：

すばらしい質問です。私たちはこれをプロトタイプとして研究を行ってきました。実際の生きたエキスパートと対話をし、そのニーズをより直接的に満足させようと努力してきました。しかし、Lispで書かれたCLASSICのバージョンを使用し、SPARKステーションで走らせています。

さて私たちはCLASSICの継続的なバージョンを書いており、これは原則的にはオープンエンド規模の知識ベースを認めています。しかし、まったく正直に言うと、このプロセスには分解しなければならない部分が多くあります。ひとつは最初にリレーショナルデータベースからCLASSICへデータを翻訳する場合であり、512個のメモリーの位置にある知識ベースの中のデータに穴があることを見つけました。それはネットワーク伝送の過程においてただ単純に落ちたものです。

そのため、数十万の個々のオブジェクトから成る知識ベースを構築することができましたが、それだけです。性能も適当なものです。合理的な問い合わせに回答するのにかかるのは、おそらく３０秒です。生のSQLにおいてすら問い合わせを書くのはそんなに難しいことではありません、しかし２週間はかかるわけですから、その基本的な問題を克服することはできていません。しかし、制限付き問い合わせでSPARKステーション上で許容性能は得ることができました。しかし、大きな問題が残っており、それが解決できたとは言えません。すなわち、現在のアーキテクチャでは直ちに何百万、いや、それ以上に重要な何十億というオブジェクトまでスケールアップできるとは思えないということです。アメリカ国内の通話回数を見ても何百万ではなく何十億バイトという情報を取り上げているのです。そのためにはやらなければならない仕事が山のようにあります。おそらく、パラレル・アーキテクチャが今後進む道になるでしょう。私たちが所有することになる、テラ単位のデータによって作られた大型分散データベー

スがその重要な要素となるでしょう。ご静聴ありがとうございました。

座長：

それでは、前のFGCSのときには、KRがリアルジーにぶつかるとどうなるかという話で、今日はそれの１つの回答だと思うんですけれども、そういうインテグレーションとか、そういう話で、現実をどういうふうに解決しているかと。たぶんデータというのが非常に、知識表現の枠組みと現実のデータが違う。それをどういうふうに解決するかという問題だったと思うんですけれども。

それでは、だいぶ押してきましたので、まだ質問があると思いますが、また昼休みの時間にでもお願い致します。再度拍手をお願いします。

(5)「知識獲得とオントロジー」

座　長：

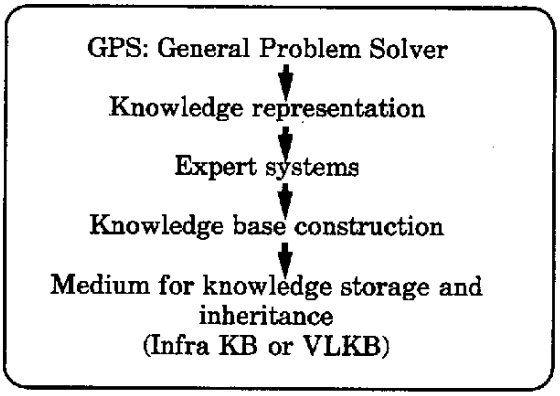
それでは、私と同じ名前なんですけれども溝口理一郎さんで、R. Mizoguchiです。溝口理一郎先生は、'７２年に大阪大学基礎工学部電気工学科を卒業されまして、引き続き大阪大学に残られて、現在大阪大学産業科学研究所の教授でございます。パターン認識、それから音声の研究、最近では知的エキスパート・システムです。溝口という名前ですから、どこか遠い祖先では繋がっているんじゃないかと思うんですけども、ちょっと感じが違うんですが、髭を最近剃られちゃって、ちょっと特徴がなくなっちゃったんですけども、じゃ、お願いします。

大阪大学 産業科学研究所
教　授　　溝　口　理一郎

どうも有り難うございます。皆さんも昼ごはんで少しおなかが空いているかもしれないけれども、あと少し辛抱して下さい。私の話は、「知識獲得とオントロジー」ということで、大体こういうふうな内容でお話しさせて頂きたいと思います。

まず最初に、ほんの簡単な、今までの知識処理の研究をサーベイしまして、それから知識の共有と再利用に関して、何が難しいかということをちょっとだけ議論してから、それに対してどういう方法論があるかということを整理して、その方法論に対して、オントロジーが非常に重要なんですが、そのオントロジーの分類というお話をして行きながら、時間があれば、日本におけるオントロジーの研究をお話したいと思います。

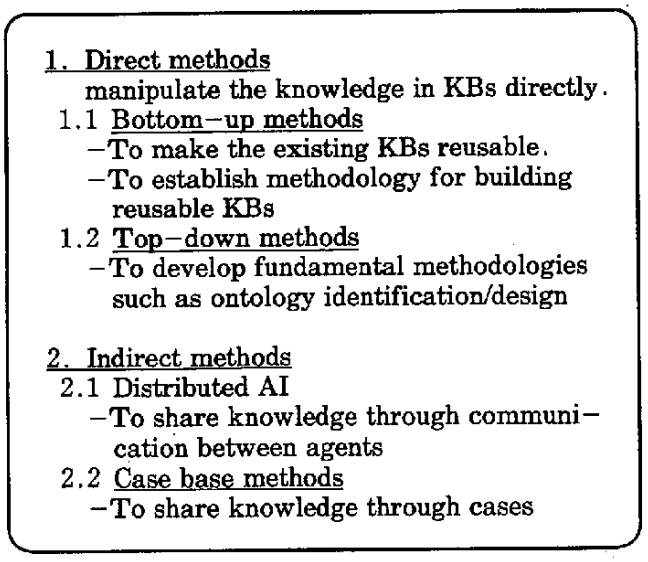
まず、最初にお話したいことは、今まで知識処理がどういうふうに変遷してきたか、ということなんですけども、これはみなさんよく御存知と思いますが、まず最初、まあ非常にこれ、大雑把な流れですけれども、一般問題解決器というのがあって、ここではあまり知識というのは意識されていなかったわけですね。で、知識表現という、非常に大きな研究の流れが出てきて、しかしここの、知識表現では所謂そのリアル・プロブレムという、現実に存在する難しい問題解決をする時の知識というのはあまりここでは意識されてなくて、知識表現の方法論自体が注目されているわけです。ところが、エキスパート・システムが出てきてからは、これは所謂そのもっと困難な問題というのをどうやって解くかという、あるいはどうやって表現するかという話がでてきて、所謂その現実の人間が抱えている難しい問題を解くということに適用できる知識、といったことが注目されてきたわけです。で、知識ベースを作る必要があるんですが、実はこの知識ベースの構築が難しくて、専門家から知識をどうやってとってこようかということからずっと、知識ベースの構築という問題は認識が非常に高まってきて、現在は所謂この、非常に大規模な知識ベース、そして、基盤となるような知識ベースというものを皆で共有して再利用していこうというふうに、まあ動きが出てきたわけです。したがって、大きく言うと、「処理や表現」と



ＯＨＰ－１：Brief Overview of Knowledge processing Research

いうところから、「知識そのもの」、「知識の内容」というものに注目が移ってきたと言っていいかと思います。じゃあ、そういうことで、いままでのお話、たくさんありましたので、「知識の共有」がなにゆえ大事かというお話はとばしまして、「なにゆえ共有することが難しいか」ということをちょっとだけお話しますと、基本的にはこれも、今までの先生方のご講演とだいぶん重複があるんですが、基本的には表現、ようするに「知識の表現」というところに関係する問題と、「知識の内容」に関する問題と、両方の原因で、共有が難しい、或いは再利用が難しいというわけですね。で、私の話は、「知識の内容」に関するお話ということに限定して議論したいと思います。

ちょっと考えますと、先程のご講演の中心は、「自然言語処理」だったわけですが、こんどは、所謂「問題解決」という観点から知識ベースを見ようというわけです。したがって、現実世界に存在する困難な問題、これを解くための知識ベースということを考えた場合には、すこしは自然言語処理の方々が必要とする知識ベースとは少し違ったニュアンスがあるわけですね。そう考えてみたいんですが、やはり結局、いまエキスパート・システム等で使われている知識ベースというのは、「問題解決の為の知識」なわけです。そうすると、実際に今、エキスパート・システムの中にある知識ベースにどういう知識があるかと言うと、基本的には専門家という方の経験的知識が入っているわけですが、これは所謂ヒューリスティックス、ある特定の問題に対してだけしか使えないルール、経験則がいろいろな知識源をコンパイルした形で入っているわけです。したがって、これはいろんな内容の知識がコンパイルされていますから、これを解きほぐす必要があるというわけですね。で、そうでないと、なかなか今現実の知識ベースの中に入っている知識を再利用、或いは共有することは難しい、と言っていいかと思います。で、基本的内容の問題のことを言いたかったわけですが、そこで今少し、こういった知識の共有と再利用に関して、どういうふうなアプローチがあるかということを、一応自分なりに整理してみたのがこのスライドです。



ＯＨＰ－２：Classification of Approaches

これは、まず大きく分けますと、「直接法」と「間接法」というふうに私は呼んだわけですが、直接法というのは、実際の知識ベースの中に入っている知識そのものを操作する、何かそれに手を加えるという方法ですけれども、間接法というのは、基本的には直接は使わなくて、間接的に共有していこうという方法論です。

まず直接法ですけれども、さらに２つに分かれまして、所謂「ボトム・アップ法」と「トップ・ダウン法」というのがあると、私は思うんですが、「ボトム・アップ法」というのは、所謂本当に今、知識工学者の方々が実際お作りになった知識ベース、それを解析して、それをもう一度再利用可能なように再編成しよう、ということをやっている動きがあるわけですが、それは非常に重要な動きだと私は思っています。要するに今ある知識ベースをもう一度考え直して、再利用できるような知識ベースに変換していくには、何をする必要があるかということ、そういう経験を通じまして、所謂その再利用可能な知識ベースというものを作っていく方法論を経験から確立していこう、という動きが、ある意味でボトムアップな方法。もう一つは、今日Ｗｉｅｌｉｎｇａ先生がお話になった、所謂トップダウンの方法ですが、基本的に、こういう方法論があるべきであるという、わりと理論的な立場から、基本的な方法論、中心になるのは要するにオントロジーの同定、あるいは設計という問題、この辺から議論をしていって、それを現実の知識ベースにアプライしようという方法があると思います。

一方、間接法と私が呼んだものは、所謂その分散ＡＩといったような、マルチ・エージェントのようなものを考えて、具体的にマルチ・エージェントの中の構成、各エージェントが持っている知識ベースの内容を、直接には手を加えないで、お互いがお互いに通信しあい、コミュニケーションしあいながら協調して問題を解くという方法で、結果的に知識を共有しているというアーキテクチュアがあります。もちろんこれも非常に有効な方法です。もう一つは、所謂ケース・ベース・メソッドですね。これは、

先程の横井さんの講演にありましたように、基本的にケースという事例を通して間接的に知識を共有しようという方法論です。みんな、いろんな、もちろん全部有効な方法であるけれど、立場が違うので興味深いんですが、私の今日のお話は直接法ということ。具体的にエキスパート・システムを作っていって、知識ベースの中身をどう触っていくか、という話をしたいと思うわけです。

この範囲で、オントロジーというのを僕は喋っているつもりですが、実はいろいろ、この立場の方が使うオントロジーと、この立場の人が使うオントロジーと、多少ニュアンスが違うわけですね。で、私自身も含めて多少混乱があるような気がしております。

基本的にまずオントロジーというのは何だろうということは、まあいろいろ定義があるわけですが、自分なりには基本的にはオントロジーというのは、ある人工物を作る場合、ソフトウェアで作る場合の基本コンセプトになるようなヴォキャブラリーの体系であるというふうに定義してみました。したがって、これはきわめて内容に関連する問題である、と僕自身は認識しております。で、今の分類でしますと、ある意味で直接法というのは、私は「知識の再利用」ということに重点をおいたような考え方であって、所謂分散協調、マルチ・エージェント方式のようなアプローチというのは、「知識共有」の方に重点をおいているわけです。したがって、「再利用」と「共有」というのはあまり区別されないことが多いんですが、僕自身は少し区別してみたいと思っていて、基本的に中身をもう一回使うということに意識した場合と、中身はあまり気にしないで間接的に共有しようという方法というふうに、少しこれはニュアンスが違ってくると思います。

こういうふうな、要するに分散ＡＩでやるようなのに必要なオントロジーというのは、私は「コミュニケーション・オントロジー」と呼んでもいいような気がするんですが。で、もう一つは、これは「ケース・ベース」の方には、当然ケースにはインデックスをつける必要がありますから、その意味では、インデックス自身が勿論重要な知識表現ですから、この知識表現を支えているオントロジーが当然必要になってきます。したがって、「インデックスィング・オントロジー」と呼んでみようと思います。まあ、私の理解では３つ位の処理がまずあると思うんですが、まずその、これを、何を喋っているかということを言った方が、ある意味で議論の時の混乱を防げるような気が致します。で、その一つの例としまして、「オントロジーは関係データ・ベースの概念スキーマ」というのをよくアナロジーに使われて説明されますが、「関係データ・ベースにおける概念スキーマ」といった物と、オントロジーは非常によく似ていると。確かに似ているんですが、それは部分的でしかないという気がします。と言うのは、基本的にはオントロジーと言う部分的にはフォーマル・スペシフィケーションというか、要するに「概念」というもの或いは「概念化操作」というもののスペシフィケーションであると、形式的なスペシフィケーションであると言えますが、これはある意味で、一部でしかない。と言いますのは、関係データ・ベースにおいては、これは、概念スキーマというのは基本的にはもう共有されて、皆が理解しています。したがってもう、関係データ・ベースに関しては、皆で内容に関する議論をする必要がないわけです。そういった意味でこの辺のところはもう「共通の理解がある」ということが前提になっています。ところが、私の言う「内容のオントロジー」というのは、ちょっとそういうところは表現できていない

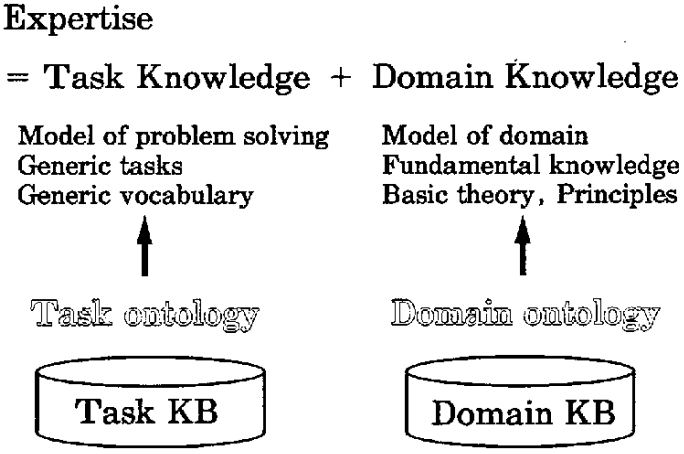
という気がするわけです。

では、「オントロジーはポータブルであるか」という話ですが、「ポータブル」と言うものの定義というのは基本的に「ポート」されたあとの内容は、その「ポート」された先のコンピュータが実行できるということを前提にしてますから、基本的にもしオントロジーを「ポート」した場合に、コンピュータがどうやってそれを使うのかというと、実は既にある、対象となったコンピュータ、「ポート」された後のコンピュータ側にあるプログラムは、当然そのオントロジーを使うことができません。といいますのは、当然そのオントロジーを人間がインタープリットして、人間が理解して、それ用のソフトウェアを使わなければいけないからです。ということは即ち、そのオントロジーに関するコミットメントがもう終了している場合には、勿論その終了したコミットメントの範囲で、いろんな人がソフトウェアを作っていますから、これはある意味で形式的なことを変換すれば使えるんですが、まあなかなかそうはいかない。したがって、コミュニケーション・オントロジーに関しては私はポータブルだと思うわけですが、まあそこら辺が多少こう、ややこしいような気がします。

結局何が違うのかと言いますと、私としては、基本的に既にオントロジカル・コミットメントに関して合意がある場合と、まだ合意がなくて、いまそれを一生懸命作っていると言う場合とは少し違うような気がします。したがって、この内容に関するオントロジーというのを議論したい時には、今まだどういうものがオントロジーであって、これをどう使うということの表現に関する合意がまだ得られていない時の議論と、得られてからの議論、あるいはこれ言い換えますと、要するに、コミュニケーション・オントロジーというのは基本的にエージェントの中身を気にしないで、間接的に使う。こっちの方は知識の再利用を頑張っていて、いまどういう内容の知識が入っているかと言うことを議論したいという時には少し違うような・・・。したがって私自身はまあ、これを区別して議論したいと思っていて、私は内容に関することを議論したいので、コンテント・オントロジーに関してお話をしたいと思うわけです。

こういう意味でＷｉｅｌｉｎｇａ先生のお話というのは、コンテント・オントロジーだと私は思っています。で、その次にこれを前提にしまして、基本的には、じゃあその次に何をするかというと、まず大事なことは、「問題解決のために知識を再利用したい」ということを中心にお話したいと思うわけですが、基本的には、いちばん大事なことは、今、問題解決におけるコンテクストから、その知識を切り放すことが必要です。そうでなかったら再利用は非常に難しいわけですね。したがって、どのようにして今あるその問題解決のコンテクストから知識を切り放すかと言うことが、非常に大事な課題になります。

そのうちの、そのコンテクストから切り放すことの一つの方法として、いわゆる「タスク・ノレッジとドメイン・ノレッジへの分解」ということが挙げられると思います。これはＷｉｅｌｉｎｇａ先生の話とだいぶ重複する部分があるんですが、まあ私自身は知識の分解というのは非常に大事と思ってまして、要するに「タスクに依存するノレッジとドメイン・ノレッジへの分解」、実は、ここをしっかりして、ここのリンクが非常に大事でして、基本的にドメイン・ノレッジというのは、ある意味で客観性があって、これ単独で記述できるような動きがあるわけですが、実際はこれは、ある問題解決に使おうと



OHP－3：Knowledge decompilation

思えば、問題解決に利用する際のコンテクストを決定するタスク・ノレッジとここのリンクをしっかり張っておかなければ実際は使えないわけですね。ところがこのリンクを張るためには、こっち側の問題解決の側のオントロジーをはっきりしておいて、なおかつこちらのリンクを、まあ両方で持っているということが大事なわけです。それを意識してタスク・オントロジーというものを考えていかなきゃいけない。これはある意味でドメイン独立ですから、ここでドメインが変わった場合でも再利用できるし、こっちの方はタスク独立ですから、ある意味で同一ドメインであれば、タスクが違っても使えると。しかし、このリンクは非常に大事というわけであります。

そこで、問題解決でのコンテクストから切り放すという問題というのは非常に大事なことで、そのちょっと一例をお話したいと思いますが、設計に関するお話を例にとろうと思います。まず、私は「タスク解析」という言葉をよく使うんですが、これは、要するに人間の問題解決のモデルを作りたいんですが、人間の問題解決のモデルを表現するためのヴォキャブラリーとしてタスク・オントロジーというものを設定しようというわけであります。勿論これは、あらゆる問題解決は難しいですから、基本的には今、エンジニアリング、工学の分野での問題解決ということに限定して議論をしたいと思っています。

基本的にこういうことが、モデルをしっかりと記述するための語彙としてのタスク・オントロジーが定義できれば、いろいろないいことがあると思っています。非常に大雑把な方法論ですけれども、ここにまあ４つのセットを書きまして、基本的には、ある適切な抽象レベルで、あるタスクを記述する、ここは難しいですけども、その記述を人間が解釈をして、ドメイン依存の知識とタスク依存知識のところを切り分けて、その切り分けたところから、タスクに依存する語彙をしっかりと記述していこうというお話なわけです。ほんの少ししか時間がないんですが、まあ、設計というお話を例にとりますと、設計というのは非常に大雑把に言うと生成・検査と修正の過程の繰り返しですから、これを、まず生成・検査・修正というモジュールを抽象化して、その次は生成に関するモジュールにおける人間のアクティヴィティーというものを少し検査していくと、例えばケース・データ・ベースからの検索とか、あるい

は数式の計算による値の決定とか、いろんな部分解の組み合わせによる解の合成とか、あるいは、勿論基本的には、ある探索空間における探索が中心ですが、探索の種類も分解していくぞというふうにして、例えば機能の部分機能への分解とか、あるいは部分機能の属性への写像とかいったことがあるわけですね。ま、これはまあわりと、いろんなシステムをチェックすればすぐこの位のものは出てきます。今のはジェネレータの部分、もっと詳しくあるんですが、ほんの簡単なものとして例をお見せしたわけです。あとはああいった作業を続けていって、例えばこういうふうな、センテンスとして作っちゃいます。要するに、要求仕様としての機能を部分機能へ分解するとか、適切な部分品を選択するというふうにして、まあ、これは抽象レベルはだいぶ高いですが、こういった意味で、いろんな設計における抽象的な記述を、動詞と名詞で区別して基本センテンスを作っていくというふうにして、アクティヴィティーを記述するわけです。

実際そういうことをやっていくと、これは全部を書いたものじゃなくて、或る一部ですが、基本的には一般的に使えるような名詞というものと、設計に固有の制約というものに関する語彙と、ゴールに関する語彙と、それと動詞、これは基本的には設計における基本的なアクティヴィティーを表すような動詞です。ここで大事なことは、動詞に関してはメカニズム、要するにメソッドとしての計算機で実行可能な様なメカニズムを必ず考えます。したがってこの動詞には必ず計算機で実行可能な構造がくっついております。で、実際にはその設計のための問題解決用の基本的なエンジンというものを、この動詞を組み合わせて作っていけるように処理をする事が大事なわけです。

そういったようにして、基本的には問題解決のアクティヴィティーを分析して、問題解決におけるコンテクストとドメイン・ノレッジの切り分けというのをやって、なおかつ、ドメイン・ノレッジとのリンクを考えていくということが大事なわけです。今、実はこれほんの一例でお話ししたわけですが、基本的にそういった動きで、今、日本におけるオントロジー研究というのは、わりと今盛んに行われつつあります。日本における研究は、大雑把に３つあると思うんですが、３つに分類できまして、所謂私が今お話したタスク・オントロジーということに関する話としては、ま、私自身の研究と、それと日本ＩＢＭにおけるーこれはコンピュータ・アシステッド・ノレッジ・エンジニアリングの省略で”ＣＡＫＥ”というらしいですがーの話とか、それから、所謂このドメイン・オントロジーに関する動きとしては、東京大学の動きで、”Ｐｈｙｓｉｃａｌ　ｆｅａｔｕｒｅ”、３次元ＣＡＤにおける物理的なある対象とその対象のビヘイビアを分類していって、ドメイン・オントロジーをしっかり整備していこうというお話と、それと、医療におけるオントロジーとして内科に関する語彙を整理していこうという動きが、もう始まっています。そして、私の研究で恐縮ですが、所謂定性的モデルを作るためのオントロジーというものを今研究しています。・・・それから、３つめの分散ＡＩでは、奈良先端科学技術大学院大学でやってらっしゃる「知識コミュニティー」という動きがあります。

数年前から動きがありまして、結果もだいぶん得られたと思うんですが、ちょっと時間の関係であまり詳しくはお話できないんですが、「MULTIS」というのは、タスク解析用の知識獲得システムで、基本的にはタスク・オントロジーというのを整備してあって、この、今動いている例題はスケジューリング

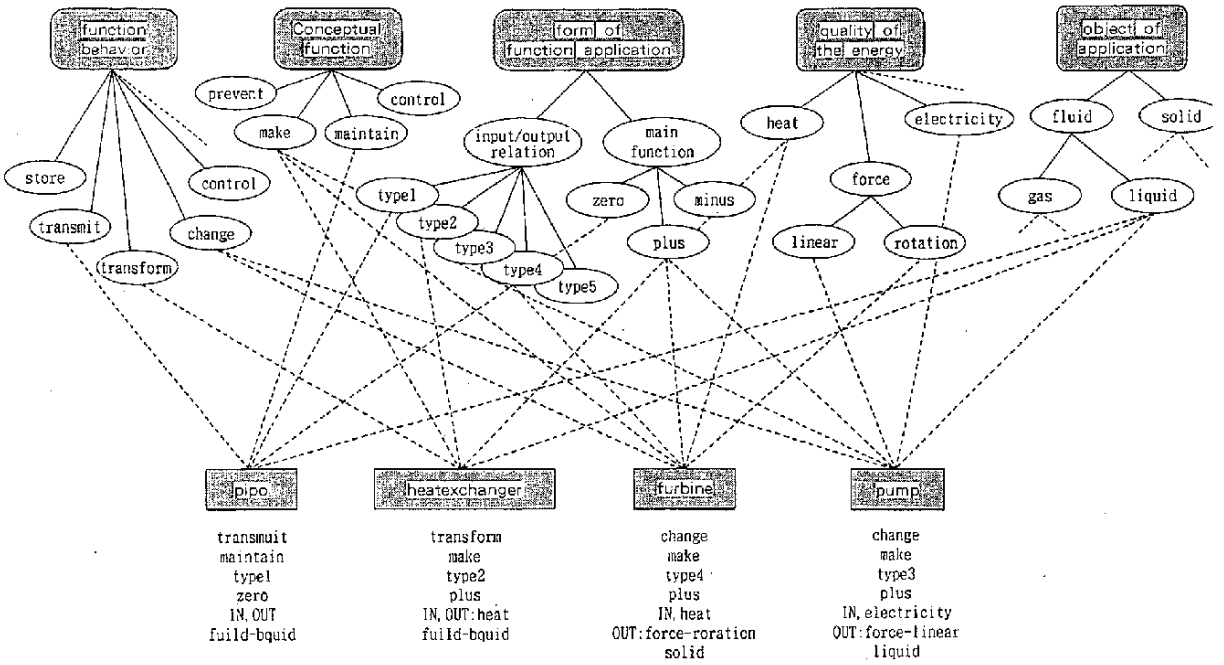
を例にとっていますが、一応スケジューリングの、全てのスケジューリングのシステムを解析して、それに基本的なタスク・オントロジーを同定して、それで、ある未知の問題がきた場合に、それをどういうふうな問題解決構造にしていくかというようなタスク解析を計算機がやってくれて、最終的には実際に計算機で実行可能なコードを出力するというところまで動いてます。

同じように”ＣＡＫＥ”でも、ま、ほぼこれと同じ機能が作られています。少し違うところはタスク・オントロジーとドメイン・オントロジーの間に、「問題解決するためのオントロジー」というのを、３つ目のオントロジーとして用意して、まあ、３段階でオントロジーをやっているということと、所謂ジョブ・ショップ・スケジューリングに特化して、そのかわり非常に精密なオントロジーを作っているというところが特徴です。

”Ｐｈｙｓｉｃａｌ　ｆｅａｔｕｒｅ”は、所謂３次元ＣＡＤの定性的プロセス・セオリーを使って、いろんなビヘイビアとか、３次元ＣＡＤで中心となるようなオブジェクトを記述していこうという動きで、いろんなローテーションとか往復運動とか振動といったようなことをしっかりと定義していって、それを部品として組み合わせていって、３次元ＣＡＤで作った機械部品の振る舞いを、シミュレーションできるようにしていこうという話であります。基本的にはこの定性的モデリングのためのオントロジーともほぼ同じ話となっております。少し立場が違いますが、基本的には同じような方法であります。医学の方でもオントロジーの研究が行われているという話です。少し例をお話致しますと、例えば、所謂ドメイン・オントロジーといった場合も非常に範囲が広いですので、かなり困るわけですが、基本的には所謂モデリングのためのオントロジーというものは、基本的にはその対象としているモデルをいくつかの部品を組み合わせていって、作られた全体の結果のもののビヘイビアとファンクション、というような性質を議論できるようなものというのがわりとベイシックな理解だと思うんです。これは簡単な例なんですが、基本的に抽象的な機能と言うのはもっとありますが、こういうのがあって、例えばこう、伝達とか変換というのがあって、これは全くドメインが違います。これは電気回路、これはパワー・プ

Abstract function	Electric circuit	Power plant	Circulatory organ
transmission	wire	pipe	blood vessel
transformation1	amplifier	heater	stomach
	voltage source	pump	heart
transformation2	electric range	turbine	muscle
change	transformer	heat exchanger	lunge
separation	filter	demineralizer	kidney
storage	condenser	tank	liver
control	regulation	controler	liver

ＯＨＰ－４：Possibility of knowledge sharing



OHP－5：Device modeling using abstract primitives.

ラント、これは所謂人間の血流とか言う話ですね。こういった、違うドメインですが同じように伝達という機能を持ったものが勿論あるわけで、あるいは変換という機能を持ったものもあるわけです。こういった意味で、抽象レベルはビヘイビアとかファンクションと言うのは相当共有して書けるわけですが、こういったものをいろんな観点から整理していこうというのがある意味でビヘイビアとファンクションに於ける、オントロジーだと言っていいと思うんですが、これは、いろんな観点があります。その観点を少し挙げますと、要するにファンクションと振る舞い、振る舞いと機能に関する合意を成立していって、その合意をしっかりと定義をしていく。勿論、所謂その、説明をするときに必要なような語彙も大事です。まあ、ちょっと時間がないので、これ簡単にしますけど、基本的にはそういったような抽象クラスを作って、抽象クラスのマルティプル・インヘリタンスで、パイプとか熱交換系とかタービンといったようなものを、いろんな抽象部品からの再利用で定義しようというわけで、基本的にこの部分のとこが、所謂ビヘイビアと機能に関するオントロジーとして、皆さんで合意があるようなものを作っていくことが非常に大事かと思っています。これは、問題はこういうふうな抽象レベルで書いたものの合意が－勿論さっき言ったオントロジカル・コミットメントと言ってもいいんですが－それがなかなか得られないんですが、それを得るためには皆さんで一緒に努力していく必要があるような気がします。

結論と言っているものは、基本的にはオントロジーというものがあれば、まずオントロジーを抽出して、みんなでそれを議論すれば、どういったものを基本的にアプリートメントして、タスクにしろドメインにしろ、モデルを作っていくことができるかということは、非常に知識の共有と再利用に関しては重要な役割を果たすわけです。私の理解では、数種類のオントロジーがあって、基本的に内容に関する

研究というのは非常に難しいわけですけれども、僕たちとしては、知識の問題により深く入っていこうと思えば、知識の内容というものをこのコンピュータ科学で扱う必要があるわけです。その辺のところが、基本的には知識の共有とか再利用というものをキーにして、一体内容に関して僕たちは、「何をコンピュータで触れるか、どういうふうにすれば知識の共有ができるか、再利用が可能か」ということで、踏み込んで行くという研究の方向だと思うんですが、そういった意味で、知識の内容を意識した、処理ではなくて、処理も大事だけれども、処理以外に知識の内容に関することで共通な議論ができるようになればいいと思っております。では、以上で終わらせて頂きます。

座長：

どうも有り難うございました。

だいぶ時間がオーバーしてしまいましたので、一つだけ短いコメントなりご意見があれば…。

溝口理一郎先生の話は、オントロジーという言葉が何気なく使われているんですが、それをもうすこし深く考察して、それに基づいてどういうふうに研究を進めていけばいいかという、そういうアプローチを示されたと思うんですけれども、まあ、今日の午前中のセッション「知識処理」で、非常に話は多様に渡ってますので、まとめるのは非常に難しいんですけれども、聴衆の方々、それなりのヒントを得てですね、「大規模とはこんなイメージである。」或いは「表現はこんなことである。」というようなことを掴まれたんじゃないかと思いますので、今回のこのセッションの目的は充分果たせたんじゃないかと思います。

最後に、討論に参加された方と、それからもう一度スピーカーに拍手を以て終わりたいと思います。

それではセッションを終わります。

5. セッションIV

利用可能な大規模知識ソース

5. セッションIV：利用可能な大規模知識ソース

5.1 座長挨拶

大阪大学 工学部情報システム工学科
教 授　西 尾 章治郎

こんにちは、大阪大学の西尾です。今日、4時からパネル・セッションがあるわけですけれども、その前の、プレ・ファイナルなセッションとしまして、これから「利用可能な大規模知識ソース」というタイトルのセッションを始めさせて頂きます。昨日、今日と「大規模知識ベースの構築と共有」ということで、それに関するいろんな技術、あるいは理論的な側面、あるいは現実のシステムの開発動向の発表があったわけですけども、もう一つ大事な討論をしておかなければならないこととして、「その時、どのような知識データを共有するのか」という問題がクローズ・アップされてきます。そこで、このセッションでは、ここに書いてありますような4つの興味深いテーマに関しまして、日本、それからヨーロッパ、アメリカからの代表者のかたがたに、ご講演、それから現在の動向等をお話して頂きます。

最初の発表は学術データに関するもの、それから2つめ・3つめは、言語データ、テキスト・データに関する発表、それから最後には、エキスパート知識ベースに関する共有データに関しての発表をして頂くことになります。

5.2 講　演

(1)「学術情報サービスの将来像」

座　長：

第１番目の発表者であります、山田先生のご紹介をさせて頂きたいと思います。山田先生は、日本、アメリカ、それから世界を股にかけて、いろいろ活躍してこられた先生であるということは、皆さんご存知だと思いますが、東京大学をご卒業になりまして、それからペンシルベニア大学の大学院で博士号を取られまして、それからゼネラル・ダイナミックス社、それからＩＢＭのトーマス・Ｊ・ワトソン・リサーチ・センター、それからペンシルベニア大学に、今度は先生として、准教授としてお戻りになられまして、それからさらに日本にお戻りになられまして、東京大学の教授をなさっておられました。１９６３年の４月から、文部省の学術情報センター研究開発部長をなさっておられまして、東京大学の名誉教授でもございます。

先生は日本のコンピュータの草分け的な研究をいろいろなさってこられましたことはご存知だと思いますが、最近ではさらに「心理言語学」とか「知能の科学」等に興味を色々持っておられます。学会にもたくさん所属しておられますけれども、最近ちょっとお知らせすべきこととしましては、米国のコンピュータ関係の、一番よく世界的にも知られております学会のＡＣＭの、日本セッションが今年できました。山田先生は、そのセッション・プレジデントとして現在ご活躍中でございます。

それでは先生お願いします。

学術情報センター
教授・研究開発部長　　山　田　尚　勇

議長、ありがとうございます。皆さんおはようございます。最初にお断りしておきますが、本日は、主として外国人の皆さんを対象として、英語でお話したいと思います。その理由は日本人の皆さんは別の機会に私の話に触れることができると考えるからです。

私の話は研究報告ではありません。国立の学術情報センターが今までに行なってきた活動状況の報告や、私達が将来何をなすべきかについて個人的な見解をお話するつもりです。

皆さんもご存知の通り、第二次世界大戦後に始まった日本の再工業化において、私たちが産業において西洋諸国と肩を並べるに至るまでには２５年間の努力が必要であったわけです。このプロセスの最中で、私たちは西洋諸国の知識と、また、もちろんのことながら技術を学び、同時にこの知識と技術により大きな利益を得たわけであります。しかし、私たちが産業活動の最高のレベルに近づいていた２５年

ほど前から、「日本は西洋の知識を非常に広範に利用しているが、自分達の行なった発見については西洋諸国には隠している」とときどき非難されるようになってきたのです。

これは誤解でありまして、私たちの情報サービスが非常に粗末なものであったために、私たち自身でさえ自分達が作った情報を利用することができなかったのです。そのため、１９７３年初期に、すなわちちょうど２０年前に、学術審議会が日本における情報流通サービスを改善するよう勧告を行い、日本ではふつうのことですが、十分に時間をかけてこれを検討し、まず特定研究として実際の検討を開始し、ちょうど１０年前に東京大学の中に学内共同利用施設として、文献情報センターが作られました。

それがうまくいったので、１年後には全国の大学間の共同利用施設に改組されました。このセンターは東京大学に属してはいましたが、東京大学だけに情報提供サービスを行うのではなく、センターのサービスは日本国内のすべての大学に利用できるようになりました。１９８６年になって、この組織は全国の大学共同利用機関のひとつに改組され、いまでは文部省の直轄となっています。この活動が今日の午後、私が報告を行う内容です。

もっと別の形でやればまた別の形で多くのことができるであろうと思われますが、それにしても現在進行中の活動は広範囲にわたっています。その組織は各大学と大学共同利用機関であるセンターとがゆるやかに結合されて、協力し合うもので、現時点では約２８５の大学が直接私たちとオンラインサービスをし合っています。

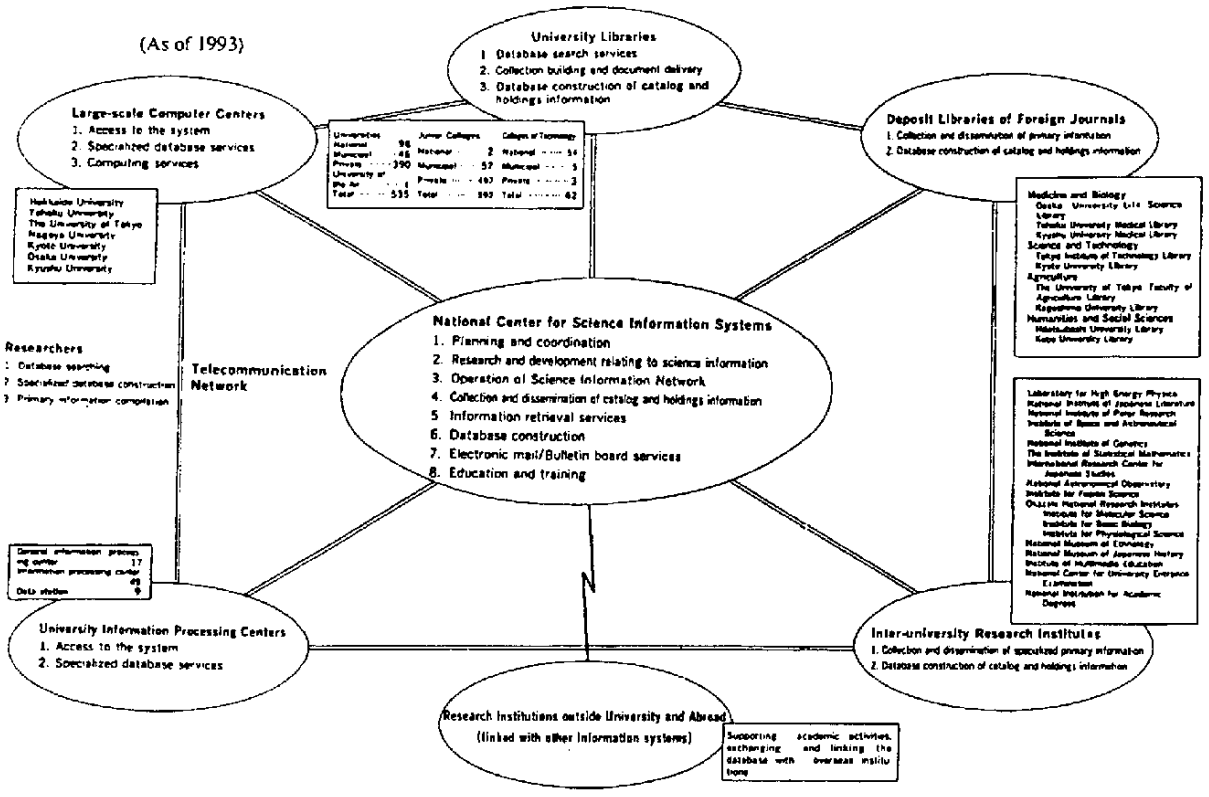


図１　Scope of Science Information System

それ以外にもまだ２５０の総合大学や単科大学などがサービスを希望しており、またさらには約５００の短期大学やそのほかの研究機関などが私たちのサービスを享受することを希望できますが、予算も限られており、そのため、すぐさまその希望のすべてをかなえるというわけにはなりません。したがって、しばらくの時間はかかりますが、私たちとしてはその役割を、遅かれ早かれ－もちろん、早い方が良いに決まっていますが－果たすつもりで努力しております。

全体的な計画は図１に示した方法で行われています。世界の重要な一流の雑誌類をすべて組織的に収集している大学図書館が計画的に設置されているため、研究分野に盲点はありません。もちろんこのことは、他の大学では自分達が必要としている雑誌も入手できないという制限をしているのではありません。また、私たちは２０余りの大学共同利用機関を文部省の管轄下に有しており、文部省下では当然のことながら図書館のあり方についても研究活動を行っています。またこれらの共同利用機関はパケット・スイッチング・ネットワークで相互に結ばれております。

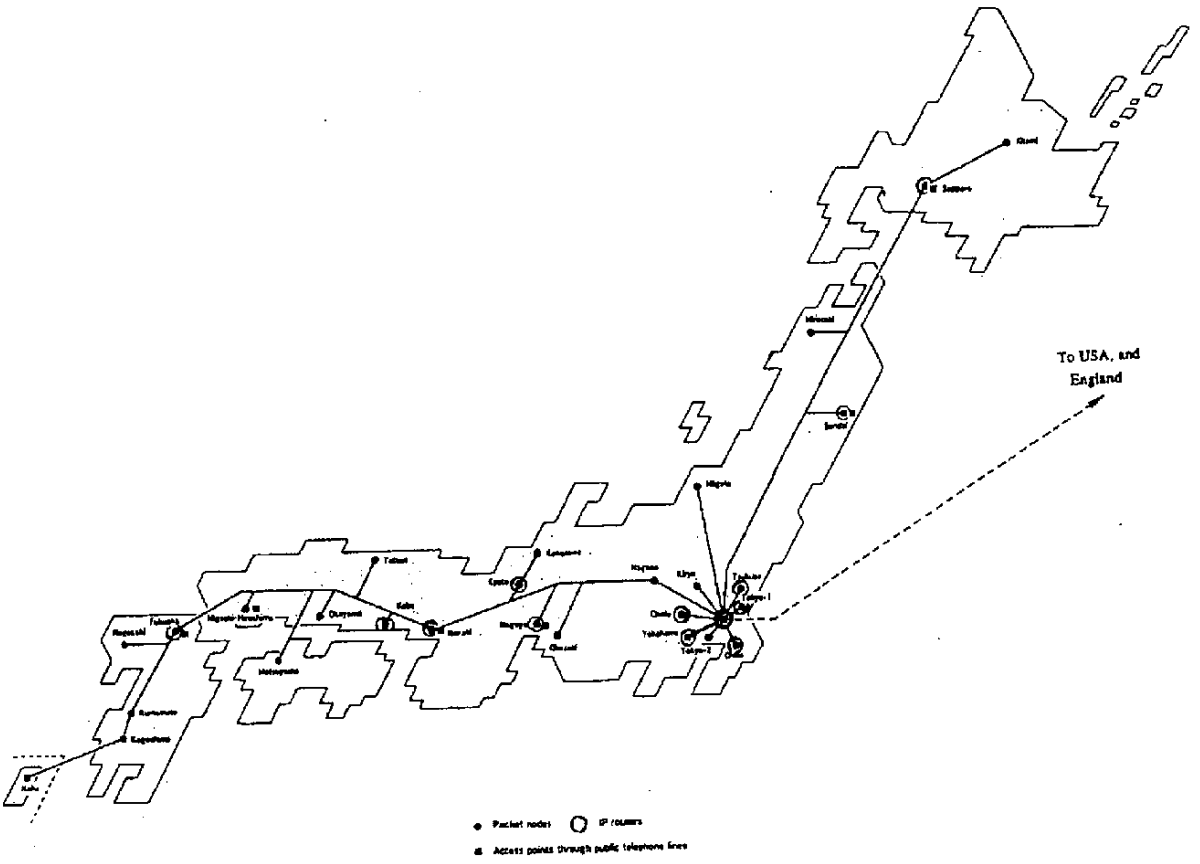


図２　Science Information Network

このネットワークを考えるにあたってはTCP/IPについて聞いたことがある人がほとんどいなかった

２０年前にすでに計画されてたものであったという事実を忘れてはなりません。そのため、現用のいわゆるＮ１プロトコルは文部省下で利用するために設計され、図２の通信ネットワークはこのＮ１プロトコルの下に設立されました。リンク・ラインを結ぶノード・ポイントは現在２９ヶ所あり、２９ヶ所のノードから支線などを介して、いま約３００近の機関が接続されており、約５００の研究機関が今後更にだんだんと接続される予定です。

またしばらく前に、アメリカやイギリス向けの専用通信線を設置しました。現在ではこれはインターネット・バックボーンにとり換えつつあります。この外国向けの通信網の拡張により、いまパイロット・サービスを行っております。現行の法令の下では、私たちは国外へは直接サービスを提供できません。しかし、研究活動は別のものとされており、そのカテゴリーのものについては海外の研究機関にもサービスを提供できます。

今まで述べたものは単に、物理的なネットワークのことです。しかし、この物理的な単一のネットワークの上、各種の仮想的な独立した論理ネットワークを設置しています。これらは表面的には約１０件ほどあります。これについてはもう少し詳細にご説明しましょう。

まず最初にあるひとつのネットワークについて話をします。たとえば、医療歴に関するサービスがあります。このネットワークで通信される情報には、プライバシーに関する情報が多く含まれています。すなわち、患者の病歴が含まれているため、資格のある医師だけがこのネットワークに入って見ることができるという意味で、厳しくガードされたネットワークです。当然のことながら、私たちがそのサービスを行なっているからといって勝手にその内容を見るわけにはいきません。

さてこれはどちらかと言えば例外でありましたが、その他のネットワークはきわめて開放的です。センターの施設を詳細に説明することは省略します。しかし１つだけ申し上げておきたいのは、それはおそらく、日本国内で一般国民に知られており、開放されているものの中では最大のコンピュータ組織です。このコンピュータのディスク・メモリーは現在のところ、１テラバイトを超える容量を有しており、もちろんこれ以外に多くの磁気カートリッジ・メモリーを利用しています。現時点ではこれで十分ですが、ほどなく需要にはとても追いつかなくなるでしょう。事実、私たちのサービスは、全文テキストのデータベースに向かって動いているところであり、それには現在の利用対象者の範囲でもメモリー量が今後少なくとも1000倍は必要となるでしょう。その単位はペタバイトです。この単位は余りお聞きになったことがないかも知れませんがペタバイトとは1000テラバイトのことです。現在、テラバイトを使用しているのと同じ金額でペタバイトにしようとすれば、そのコストダウンには約１０年から１１年はかかります。しかし、これは近い将来に実現すると考えていいと思います。ですから日本国内でのネットワーク接続がかなり短い間に実現できることになります。NACSISはこうしたインターネットと呼ばれるネットワークの一部を担当しております。このネットワークはすでにWide、Joinなどの他の国際的な各種ネットワークとも密接に接続されています。

それでは私たちが維持している、単一の物理的ネットワークの上でサービスしている、独立した論理的ネットワークのサービスをもう少し詳細に見てみましょう。まず最初はユニオン・カタログ・データ

ベース・サービスです。すでに述べたようにいま２８５ほどの図書館及び研究機関が実際にこのサービスを利用しており、直接・間接に私たちのところに接続しています。これは現在日本の総合大学、単科大学、研究機関が保有していながら、いままでは簡単には利用できなかった、すべての本の総合カタログを設立することを目的としています。法的な規制のために、いまのところサービスは依然として主として文部省の管轄下にある機関の図書館などに限られてはいますが、現在、これらの機関の保有図書数は１億6000万冊を超え、雑誌数は200万種となっています。今までにカタログされたものの詳細につきましてはお手元の資料の表１に示してあります。過去における情報の蓄積量の増加のカーブから見ると蓄積はますますスピードアップしているわけですが、詳細は省略します。

Table 1. NATIONAL ACADEMIC UNION CATALOG DATABASE
(as of April 1993)

		Bibliography	Holding
Books	Japanese	720,000	5,820,000
	Foreign	1,710,000	3,930,000
Serials	Japanese	70,000	1,520,000
	Foreign	110,000	940,000
Author Authority			536,000
Title Change Map	Japanese		8,457
	Foreign		12,843

私たちのカタログは参加館のすべての書籍と雑誌に関するユニオン・カタログをなしており、そのため、このサービスに加入している各図書館はこのデータベースを検索し、新規に購入した書籍や雑誌の情報がすでにデータベースに入っているならば、繰り返して詳細な情報を入力する必要はありません。すでに私たちのデータベースに入っているからです。この情報を自分の図書館にダウンロードして、それを自分のカタログに取り入れるだけでよいのです。しかし、同時に各図書館はそれを新しく保有したという情報を私たちに送り、私たちはこのユニオンカタログのデータベースにその保有情報を追加します。もしその新しい書籍などが私たちのデータベースに入っていなければ、それを入手した図書館は、私たちのデータベースにその文献及び所在（保有）情報を入力します。このユニオンデータベースを一歩一歩構築するためには１５年以上もかかりましたが、お陰さまで、今では広範囲に利用されています。

今述べたのが第１の論理ネットワーク上のサービスです。第２の独立した論理ネットワークは情報検索ネットワーク・データベース・サービス用です。私たちのところでは現在、約４０種類のデータベースのサービスを行なっています。当然のことながらこれはすべての参加図書館にオンラインでつながっており、また、文部省の管轄下にあるなど、資格をもった検索利用者の数は、現時点で約7000人であり、これだけの人びとがデータベース・サービスを利用しているのです。しかし、利用資格のある、文部省の関連の研究者が総数で１５万人以上いるという事実を考えれば、この数字はわずか５％にしかならず、

表2　Databases at NACSIS (as of April 1993)

	No. of Records	File Starting
I. IMPORTED DATABASES (Type A)		
Life Sciences Collection	1,140,000	1982~
MathSci	1,480,000	1940~
Compendex Plus	2,450,000	1976~
Harvard Business Review	2,600	1927~
ISTP&B (Index to Sci. & Tech. Proceedings)	1,950,000	1982~
EMBASE [medical]	2,670,000	1984~
SciSearch	4,250,000	1987~
Social SciSearch	740,000	1987~
A & H Search [art & humanities]	680,000	1987~
II. DATABASES CREATED BY NACSIS (Type B)		
*Grant-in-Aid Research Reports	93,000	1985~
*Dissertation Index	57,000	1984~
*Database Directory	1,300	1992~
*Directory of Researchers	130,000	1988~
Laws in Force	3,600	Latest
III. DATABASES CREATED IN COLLABORATION WITH ACADEMIC SOCIETIES (Type C)		
***Conference Papers of Academic Societies and Associations**		
Series 1 (Elect. Eng., Info. Processing and Control)		
Series 2 (Chemistry)		
Series 3 (Architecuture and Civil Engineering)		
Series 4 (Biology and Agricultural Sciences)	125,000	1984~
Series 5 (Physical Sciences)		
Series 6 (Engineering and Technology)		
Series 7 (Medical Sciences)		
Series 8 (Humanities and Social Sciences)		
Scientific Papers (full text, *: abstracts)		
Series 1 (Electronics)	930 (2,500 *)	1989~
Series 2 (Chemistry)	8,900	1983~
Series 5 (Physical Sciences)	1,500	1992~
EXIRPTS [research projects of 8 countries]	81,000	1985~
*Private Grants-in-Aids Research	960	1964~
*Economic Titles Japan	91,000	1969~
Electronic File of Academic Conference Papers	43,000	1969~
Clinical Case Reports	2,000	1988~
IV. REPOSITORY DATABASES (Type D)		
*Summary of Materials of Ishin History (1846~1871)	20,000	-
*Unearthed Wooden Tablets of Japan	15,000	-
*Index for General Information of Home Economics Research	20,000	1979~
Japanese Periodicals Index (National Diet Library, NDL)	910,000	1984~
NDL Catalog of Science & Technology Proceedings	33,000	1985~
*RAMBIOS (molecular biosciences)	5,400	1983~
*Chemical Sensor DB (preparations & propeties)	9,000	1975~
Electric Chemistry DB (numeric)		Latest
Japanese Slavic and East European Studies DB	7,000	1988~
Academic Conferences (all fields; Science Council)	3,500	Latest
Academic Meetings (engineering; Japan Fed. of Engrg. Socs.)	800	Latest
V. CATALOGING DATABASES (Type E)		
*National Academic Union Catalogs	see Table 1	
Japan MARC (Monographs)	1,170,000	1956~
LC MARC (Books)	3,390,000	1968~
LC MARC (Serials)	570,000	1973~
American Center Union Catalog	6,500	Latest
GPO MARC (Government Printing Office)	270,000	
UK MARC (Books)	1,140,000	1950~
Bibliographia Germanistica Japonica (humanities)	8,000	1988~

そのためにはこの利用を促進するために何らかの手だてを行う必要があります。なぜなら、研究者活動の各段階において利用可能なデータを参照したり、研究結果や、新規の発見と思われるものが、はたして他の機関で行なった仕事の焼き直しでないかどうかを自問してみる必要があるからです。そのためには、サービスを利用する上でデータベースをもっと使いやすい形にしたり、端末と人間のインターフェースを改善するなどいろいろなことをしなければなりません。

これら４０種ほどのデータベースは便宜上、５種類のカテゴリーに分割されています。表２をご覧下さい。最初のクラスはライセンス付きのもの、そのほとんどは外国製です。第２番目のクラスは私達が作成したもので、そのいくつかは表２に掲載してあります。即ち、グラント・イン・エイド研究報告がこの例です。さきにお話しましたようにわずか１５年前には、情報サービスがまだ整備されていなかったため、私にとっては、コンピュータの分野において、北海道大学で書かれた学位論文の表題を知るよりは、アメリカのＭＩＴやハーバード、スタンフォードなどの大学の学位論文の表題や、全文さえも、それらを入手することの方がずっと簡単でした。

その結果、日本人の研究者の多くが国内の論文はあまり参照できず、外国の文献を参照していたのです。これが２０年前に学術審議会が苦慮した状態だったのでした。幸いなことに現在では、研究報告を書くために役立つデータベースがいろいろとあります。たとえば学位論文の索引です。これは日本で書かれたすべての学位論文に関するデータベースです。最初に述べたグラント・イン・エイド研究報告のものは、文部省が科学研究費補助金によって支援している主要研究活動に関する研究成果概要のデータベースです。

その次のものはデータベースのディレクトリーです。もちろん、私たちはすべてのデータベースのサービスを提供することはできません。そのため、データベースのためのデータベースを所有しており、もしあなたが適当なデータベースを私どものサービスで見つけられない場合には、少なくともこのデータベースを使用してどこか他の所を探してみることを可能にします。

データベースの第３番目のカテゴリーは、学会の協力を得て私どもが作成したデータベースです。日本学術会議には1100団体を超える数の学会が登録されており、最大の学会には１０万人程度の会員が登録されており、最小の学会には100人から200人程度の会員しかいません。しかし、規模は重要ではありません。これらの小さな学会は特殊分野の学術団体であるからです。例えば、ある特定の血球の性質に関する重要な学会があり、その分野ではきわめて重要なものなのです。これらのそれぞれの学会が国内で年次会議、シンポジウム、ワークショップなどを催しており、毎年、毎月と会合が行われていて、そこで発表されたすべての論文に関するデータベースを、これらの論文が実際に発表されてから３ヶ月以内に、できれば２ヶ月以内に作成しようと努力しています。具体的には表題、発表者、所属、抄録などが、多くの場合英和両方で記載されています。

このシステムは日本の研究機関にとって重要なデータベースであると思います。それはこのデータベースを使用すれば、論文が実際に発表されてから２～３ヶ月以内で、日本国内で現在行われているすべての研究を網羅的に見ることができるからです。このシステムによって日本国内の研究活動が相互に

刺激し合い、外国の情報源に余り頼らなくてもよくなるものと確信しています。しかし、この種のデータベースは、たとえ研究報告が優れていたとしても速く陳腐になってしまいやすいものです。例えば、よい論文ならフル・ペーパーが学会雑誌にせいぜい２年以内に掲載されるでしょうから、それ以降はほとんどの人がデータベース化されたショート・ペーパーではなくフル・ペーパーの方を読むのが確実であるため、このデータベースはほとんど無価値なものになってしまうからです。しかしそれでも私たちはこれが重要なものであると考え、そのデータベースの構築にお金をかけているのです。お分かりのようにこれは商業的には有効に元のとれるものではありません。

今述べましたこれらの抄録のほかに、ある種の論文については、私たちはASCII系の文字コードを利用して、全文テキストのデータベースのサービスの提供を始めました。細かくは、文書はコード化された文字列であり、もちろんオンライン化されています。しかし、この方法ではグラフや図は転送できません。そのため、私たちはそれらを非テキスト型データベースで構築しました。即ち、この非テキスト型情報はすべて光ディスク・メモリーに記憶され、しかもシステムは自動化されており、いまある特定の論文を検索する場合には、グラフなどのイメージの部分はユーザーが前もって指定したファックスに送られます。もちろん、あなたが別のファックスを使用したい場合にはそれを優先できます。

また、私たちは電子ライブラリーについても開発を行っており、これを使用すれば、全ページのイメージをビット・マップのイメージとしてカラーで受け取ることができます。これは現在開発中であり、予算がつけばこのサービスをほどなく試験的に提供できることになっています。ご存知のように、すぐさまの全面実施はかなりむずかしいことでしょう。

時間が制約されているため、第４番目のカテゴリーにいきますと、これは委託的な性質を持ったタイプのデータベースです。日本国内で活躍している研究グループの多くは、その研究活動においてまず何らかの学術的なデータベースを作成するのが普通です。研究が終了したあと、このデータベースを維持していくことに必ずしも関心がない場合でも、特定な項目については価値のあるデータベースであることもあります。しかし、研究が終了したあとは、このようなデータベースを維持していくことができないことも多いのです。そのため、特定の部門にとって重要であると考えられ、しかも、特定の基準を満足するものについては、私たちが代わってこれらのデータベースを提供することを引受けています。その例のリストは表２のⅣです。

表２の第５番目のデータベースは、目録情報に関するデータベースです。これは主として図書館で作業を行なうライブラリアンが使用するものですが、もちろん、研究者にも利用できますし、私を含めてこのデータベースを利用する研究者がいま日本には7000人もいます。

外国のユーザーに対してはかつて全米科学財団向けの専用の通信線などを持っていましたが現在ではインターネットによっています。実際にはこれらの設備を通して、アメリカやイギリスのいくつかの大学に実験的にデータベースを提供しております。私が強調したいのは、今のところ実験でなければ日本の政府機関の会計関係法令により、実施がむずかしくなっていることです。しかし、これらはほとんどが研究項目に関するデータベースです。いま、イギリスの図書館については、オックスフォード、ケン

ブリッジ、シェフィールド、スターリング、さらにはロンドン大学などの５ヶ所ほどに対して、私たちはデータベース・サービスを提供しています。これらの大学は優れた日本研究や日本関係の資料を保有している大学です。そこではライブラリアンは主として私たちのデータベースをカタログ作成の目的に利用しています。

表２のリストをご覧になれば、特定のデータベースにアスタリスクがついています。これらは特定の外国ユーザーに利用可能なものです。

一般に利用料金はかなり名目的のもので、それは費用のほとんどが文部省の助成金でまかなわれているからです。そのため、このサービス料金を類似の商業ベースのサービス料金と比較すると通常は６分の１から１０分の１となっています。

私たちは図書館間自動貸出サービスも行なっています。私たちは貸出と呼んでいますが、このサービスを通じて論文のハードコピーも入手可能です。この図書館間貸出サービスのおかげで、研究者は検索の結果、興味のある論文を見つけたならば、その所在情報を自動的に検索してコピーを請求することもできます。いまのところ自分の図書館を通じて自動的にコピーを請求しますし、また、図書館としても他館からの情報提供のサービスを行うことも可能です。このシステムは特定のアルゴリズムを有しており、利用負担を平均化したり、また、ある館の所蔵がすぐに、もしくは、利用できないときには、ほかの適切な図書館を自動探索することもします。しかし、日本においては、貸出規則がまだ厳しくないために、ときにはハードコピーを請求してから到着するまでに何日もかかるという弱点があります。日常のペースですが、平均値として、コピーを入手するまでに３日以上かかります。しかし１週間後には到着率は９０％まで上昇しています。

これらはサービスの現状でありますが、将来的には私たちはもっと多くのことをしなければなりません。というのは現在のサービスはすでに２０年も前に計画されたものであり、そのため、現在のサービスはまだオンラインですが、それ以降、約１０年前にCD-ROMが利用できるようになり、現在、世界中ではCD-ROMによるサービスがかなり盛んになっており、これからもますます盛んになるでしょうから、この形態でのサービスを無視することはできません。CD-ROMに関して現時点で私たちのしていることは、各館が自館についての目録情報を確保するためのカタログをCD-ROMで供給することであり、それは、ある図書館から請求があった場合にはその図書館が保有している目録情報を私たちのユニオンカタログから抜き出して、CD-ROMを焼きこみ、これらのCD-ROMを発注元の図書館へ送付し、その大学内に布設されているＬＡＮ（地域ネットワーク）のカタログサービスの情報とすることです。これが新しいサービスの一つですが、そのほか、既に試行サービスを開始していることですが、全文テキストのデータ・サービスが現在重要な問題になってきているのです。昨日、Susan Armstrong教授がデータの利用性と非利用性について述べておられましたが、ここで確約はできないものの、これからも、NACSISは利用者が希望する情報を提供していくものと確信しております。

こうしたサービス一般の将来は明るいものと思われますが、皆さんもお分かりのとおり、昨日と本日、数人の講演者が述べられたように、現在この種の学術情報のデータベースは急速に蓄積中であり、その

結果、研究者自身がいったい何を探したら良いかが分からない学際的な分野で初めて研究を始める場合には、膨大なデータベースをどう利用するかがますます困難になってきています。将来的にはサービスの核となる人々は、情報サービスの機構や世界中のデータベースの地域分布に詳しい知識を有することが必要になるでしょう。その上、これらの人々は特定の分野の知識に加え、いくつかの関連した分野の知識が必要となり、要求に応じてデータベースの海を案内してくれる、ナビゲータとしての役割を果たすことができなければなりません。

この種の大型のデータベースが効果的に利用できるためにはこうした専門的な技能が必要です。しかし、２１世紀にかけては、データとはある場所に受動的に座っているべきものではなく、私としてはＵＰ、ＡＰ、共同、ロイターなどの報道通信サービスなどのように、ある意味では世界中に向けて研究結果がいつも配信されているような世界を予測しています。そして情報は受取る側の末端で選択をすることになるわけです。もちろん、現在でもそうした選択は手作業でできるわけですが、将来的にはオフィスや家庭に強力なワークステーションを所有して、これが希望に応じて選択する対象のプロフィールを持ち、それに適した配信情報のみが、全文テキスト、抄録、表題のいずれかの、これも指定した形で表示されることになろうかと思います。そしてこれは、取りこんだすべての情報がかなりうまく整理された書式で出力されるように、いつも自動的に編集されることになります。この種の活動ではシソーラスやオントロジー（本質論）が重要であり、取りこんだ情報は、それらを活用して読みたいときにきちっとした割付でプリントアウトしたり、スクリーンに表示したりできる、うまい構造にまとめて整理されています。これが２１世紀になっての効果的で、競争力のある研究活動を行う上で不可欠なものとなるのです。

この辺でやめることにしますが、私たちが今、何をしており、そして何処まできているかということについて少なくとも何らかのアイディアをお伝えできたことと期待しています。私たちの活動の詳細に興味がお有りの方は、センターの要覧があります。この要求があれば喜んでパンフレットをご提供いたします。ご静聴有り難うございました。

座長：

山田教授、本日はNACSISの開発の歴史と最新の状況、学術関係の情報サービスの将来の見通しについて非常に有効なお話をありがとうございました。

何か質問やコメントはございませんか。時間に制約がございます関係から…何かございませんでしょうか。

貴重なお話を大変有り難うございました。

(2)「研究開発領域における言語リソース：その課題と展望」

座　長：

次はAntonio Zampolli教授の講演に移りましょう。教授の講演のタイトルは研究開発コミュニティー問題とその見通しに関する言語的資源というものです。ザンポリ教授は現在、ピサ大学のコンピュータ言語研究所の教授であられます。また、同時にイタリアの国立研究理事会のコンピュータ言語研究所の理事も兼務しておられます。また識字及び言語コンピュータ協会の会長もしておられます。

University of Pisa, Department of Linguistics
Professor　Antonio Zampolli

残念なことに、私は前の講演者のように講演をする事はできません。私は日本語ができませんし、私の英語もひどいものです。しかし、通訳の方が私の下手な英語をうまく訳してくれるでしょうから、通訳の話の方に耳を傾けていて下さい。

私たちは言語資源を必要としていることには同意しています。そして自然言語プロセスによるコンピュータ言語を４０年以上もやっていて、ニーズに合った適切な言語資源を手に入れることができないのかと自問することもあります。私の講演では、「私たちのニーズはその資源を共有することができる再利用可能な資源を入手することである」というポイントを示してみようと思います。それではこの目標に向かっているヨーロッパのいくつかの例と、将来の展望のいくつかをお話することにします。

まず最初に、最初の疑問に対して私の個人的な答をさせて下さい。コンピュータ言語を４０年間もやっていて、なぜ適切な言語資源が得られないのでしょうか。もちろん、お金の問題、組織の問題はありますが、２０年前は技術が適切なものではありませんでした。しかし、コミュニティーとして私たちも責任があると思います。私は非常に年を取っていますが、世界中のほとんどすべてのＭＴプロジェクトを停止した、'６６年のＡＬＰＡＣ報告において初めてコンピュータ言語という言葉を導入しました。その時のパネルの勧告は大型の言語表現、モノリンガル及び対比コーポラ、語彙、文法などの促進をすぐに始めよというものだったのです。しかし、コンピュータ言語という私たちのコミュニティーはこれらの勧告から離反したのです。責任はそれぞれ異なりますが、特に、現代の語学学校の影響であると思います。言語学校はそのためモデルの研究に集中しており、モデルを研究するために言語現象の研究に集中しております。モデルを研究することはよりおもしろいことですが、現実には次に示すテキストにあるようなものではありませんでした。

状況が変わったのは'８６年になってからのことに過ぎません。私の意見では、この変化は言語産業が出現したことによるものです。産業界、国家機関、国際機関などが、私たちのノウハウを実用化するために用いるという可能性を信じ始めました。この事実が、私たちのコミュニティーに大型の多言語の語彙目録にアクセスを有し、再利用可能な技術上の文法を含んだ現実的な言語利用の根拠に基づく、強

力な要素の必要性を検討させることになりました。

私は、キーポイントは'８６年にグロセトで組織されたワークショップであると思います。このワークショップの動機付けは、語彙の研究をしていた人々が、当時のおもちゃのような語彙ではこれ以上研究を続けることは不可能だということを認識したという事実によるものであると思います。'８６年にはマンチェスターで行われた調査では、コンピュータ・システムにおける語彙の平均的な大きさは１２単語であり、もちろんのことながら、この１２単語では重要な研究は何もできません。

また開発者も、新しい応用法を開発するたびごとに語彙を前と同じように開発することは不可能だということが分かったのです。私たちも同じ会社の中でも、また同じプロジェクトにおいてさえも、プロジェクトを更新するためにはスクラッチから語彙を作成することから始めなければならないことを見てきました。金がかかりすぎるし、効率は悪いばかりでなく、分野ごとに努力を集中させることに障害でもありました。

'８６年にはヨーロッパのＥＣ評議会を開催し、初めて各コミュニティーの代表を参集し、そこでまとまった勧告は多機能型の再利用可能な資源を開発する概念と緊急度でした。

私たちは次の方法で、再利用可能な資源を定義しています。すでに確立された２つの見解から、再利用可能な言語資源を定義します。過去の見解は、すでに存在している資源を再利用することができるというものです。Susan Armstrong氏が昨日の講演で、'８０年代には機械で読むことができるディクショナリーから知識を構築しようとする努力が広く行われた事実に言及したと思います。

将来については、そのキー・アイディアは新しい資源が再利用可能であり、その応用方法に異なった言語サービスが含まれていたとしても異なった応用法について、利用可能な方法で構築されているという目標を特に念頭において新しい資源を開発することにあります。たとえば、Ｘ７と呼ばれるヨーロッパのプロジェクトでは語彙に対するいわゆるコンパイル系列が出現しています。また、この語彙データベースから、適当なインターフェースによりアプリケーション語彙を構築できる方法で大型の語彙データベースを構築する可能性を科学的に証明したのがこのプロジェクトです。

現在私の頭に入っているある考えをお話しましょう。これは公式なものではありませんが、現在のヨーロッパでの状況といえましょう。ヨーロッパでは言語資源はいくつかの理由から、インフラの一部を形成しなければならないと信じている人々が多くいます。私たちが考えるには、ヨーロッパでは言語資源の構築に多くの資金が使用されたものの、研究、教育、補助開発については私たちが必要とするものは何も手にいれていないからでしょう。これは努力をするということと、努力のデュプリケーションという異なった要因のためです。特にその再利用可能性を保証できる規格がないことが上げられ、事実、イタリアではご存知のように数種類の言語を有しているからです。

私たちはすでにヨーロッパで数多くのインタビューを受けました。私自身としては、特にイタリアでのインタビューが多かったのですが、産業界が私たちに尋ねることは、私たちが語彙とコーポラを提供しているということです。このことは彼らが最初に尋ねる質問であり、少なくとも、いつそれが利用できるのか、また何が利用できるのかということを知りたいからです。情報と拡張の見地から、適切な言

語資源を持たずに製品におけるプロトタイプを作ることは、ほとんど不可能に近いと言っています。

たとえば、特定の利用目的と無関係に、開発をすることは不可能だと信じている人もいます。資源は特定の利用目的の範囲内で開発しなければならないと考えている人もいます。私はそれは危険な幻想であると考えます。

これがヨーロッパで今起こっていることです。私がヨーロッパで起こるであろうと考えることは、ＥＣができるかぎり早く別のヨーロッパの戦略を開発するということです。現実に、各種の委員会がこれについて研究を行っており、私もその目標は、インフラ資源としての言語資源を検討することであろうと考えております。公共領域で開発しなければならないものは、共有でき、利用性を保証できる公共資源であり、その標準化は公共領域で開発しなければなりません。これらがただで良いというつもりはありません。これらが公共領域で利用できなければならないと言っているのです。ヨーロッパのコミュニティーはこの分野では大きな責任を負うものと見られています。

現在いくつかのプロジェクトが進行しています。これらのプロジェクトの目標は、ヨーロッパ向けの共有可能な言語資源を作成することです。プロジェクトは技術的なものと組織的なものの２つのカテゴリーに分けることができます。

技術プロジェクトは語彙、コーポラ、文法などに対する一般的な仕様を特定したいと考えています。また組織プロジェクトは一般的な組織的な枠組みを特定したいと考えています。彼らは多くの問題を抱えています。たとえば、これを支援することについて誰が責任を取るかということです。現在進行中のプロジェクトのいくつかの例を紹介します。

一つのプロジェクトはNERCであり、これについてはすでに昨日講演が行われています。NERCはNetwork European Network for Corporaの略で、１０ヶ国のコンソーシアムで実行されており、事実、代表権のないのはルクセンブルグのみであり、その理由は非常に小さいことであります。知っている限りではこのプロジェクトで研究を行っている研究者は各研究所を代表しており、ヨーロッパ全体では300名程度です。NERCは委員会に対し、今年末までに、現在必要なソフトウェアのコーポラの仕様に基づいた、コーポラ設計関係の技術仕様に基づく語彙、及びコーポラの内容に関する勧告を報告する予定です。将来各種の要因が持たなければならない特徴や、ニーズの知識やコストの評価に基づいたすべてのものを特定する予定です。

これについてはまったく同感で、私たちはＥＣＩやその他の研究を支持してきましたが、ＡＤＣのモデルは、重要であるとは思いますが十分ではありません。それぞれの言語に対してその国のコーポラを開発する必要があります。その構造について勧告をしてきました。

特に私たちは、それぞれの言語が基準コーポラを有しなければならないと考えます。ヨーロッパ言語に対して、同じ特徴を持って構築されているのがコーポラであり、このコーポラというものは静止したままのものではありえないとも考えます。Sinclair教授により導入されたモニター・コーポラというアイディアを紹介してきました。このモニター・コーポラの考え方は書かれたり、話されたり、注釈をつけたりする、非常に大きな、バランスの取れたコーポラのことであります。しかしこのコーポラは、た

くさんのコーポラや、異なったタスクに対するサブ・コーポラをその辺りに構築するための中心となる核のことであり、これらは時間と共に変化するものです。将来的には大量のデータが入ってくるものと考えます。このデータは特定の瞬間のコーポラと比較され、適切なフィルターを通した結果、別のコーポラになります。これが言語の変化を観察する時間の窓としてのモニター・コーポラであると考えています。

もう一つのプロジェクトはEAGLEです。EAGLEの目標は以下に示す通りで、言語資源を構築するために基準を作成することです。EAGLEの構造は次の通りで、ピサ大学の私たちがチェックをして支持してきました。これは管理委員会を有していて、５つの作業グループに分かれています。５つの作業グループのリストはここに示します。

コーポラ、語彙、形式主義、評価、及び話言葉に対する仕様と標準化を取り扱うことにします。話言葉にもグループがあるとは不思議に思われるかも知れません。評価や、形式主義のコーポラ、語彙は話言葉やスピーチやNLPを対象としたものであることは勿論なのですが、しばらくの間はこれらをひとまとめにすることにしました。その理由は、スピーチとNLPの統合はまだ先のことであるからです。２つのコミュニティーがまとまろうとしているとは思いますが、そのプロセスの初期段階にあります。そしてそのことがここに反映されているのです。

管理委員会は、資源のことを取り扱う全てのヨーロッパプロジェクトの代表で構成されています。そして全体としておよそ５００人の研究者を代表しています。半数は産業界を、そして、残りの半数は大学を代表しています。ここの議長はRohrer教授で、私の後から講演をなさる予定です。

各グループはサブ・グループに分割され、ユーザー・グループと連けいしています。その理由は、「基準を見つけられない基本的な考えは、基本的な考えとは定義できない」ということであるからです。しかしそれは、分野における主要な関係者の同意が得られた場合に限って定義することができます。ユーザーは非常に重要なものです。私たちは関係したプロジェクト・システムを有しており、ＴＥＩやテストから関連プロジェクトは非常に重要なものであることを学んでおります。そのため、仕様のテストも行う予定です。'９５年にはガイドラインと共通仕様の第一版が発行される予定です。

以上は、技術レベルでのプロジェクトの例です。組織レベルの方はプロジェクトは委員会で決定された２つの目標を持っています。最初の目標は、ヨーロッパにおける言語資源の開発の枠組みを定義することであります。それは何を意味するのでしょうか。それは分野をまとめることのできる可能な方法を委員会に提案することです。資源を提供する関係者によりまとめることができるでしょうか。支持することでまとまるでしょうか。それは責任の問題です。どうしたらＥＣ、各種の国内機関、私的企業部門、公共部門などの間でこの責任を分けることができるでしょうか。どれが言語資源の状況なのでしょうか。関連してくる法的な問題には何があるでしょうか。組織とは、私たちがこのプロジェクトを作ったため、ピサに産業政策決定委員会を有している構造体を意味しています。私たちは高いレベルのコンサルタントを有しており、彼らは国内及び国際機関と共同で、アイディアを促進する上での援助をしてくれますし、また、私たちはNLPやスピーチ・コミュニティーから参加している研究所も持っています。

一方、私たちはＡＤＣに匹敵するものを構築しようとしています。私たちはヨーロッパという性格を考えると、作るものはまったく同じものという分けには行きません。私たちは資源の実験保管場所から始めることにしました。できる限り多くの資源を集め、たとえば１年間で資源を配布できるインフラを構築する予定です。このインフラが日本とアメリカという他の２つの主要経済ブロックとの協力ポイントになるものと期待しています。

このプロジェクトの目標の１つは、非ヨーロッパの関係者との結びつきを求める政策を提案することです。国際協力とはどういうものでしょうか。私たちは国際協力を徐々に促進を開始するモデル化した許可をEAGLE内部で得ることに成功しています。現在までに決定した活動としては、米国立科学財団及びALPACとの協定という成果が上げられます。これらは非常に小さな活動ですが、この方法で始めるのが重要であると信じています。以前のような会議での話題に上ることは希望していません。

最初の目標はNSFと共に最高のものをまとめることで、科学の本を作ることではなく、分野の自己評価となるような本を作ることです。私たちが必要とするのは「成果の評価」の他に「技術の評価」です。ニーズのある状態が主たる問題です。私たちの意見では、この本を見る人は決まっていますが、その分野で専門知識を持っている人も関連した分野で情報を探しています。コンピュータ言語は非常に分化していますから、期待をかけられても困ります。この本はおそらく来年の後半には準備が整うことと思います。またたとえNSFやECが購入することを促進したとしても、著者は日本の方も数人招待することでしょう。

これらの行動においては、運動に沿ったテストを支持しております。Susan Hockey氏がその講演をしてくれるでしょうから、私は何も話しません。私たちがEAGLEとTIEの間に公式の連帯を確立したという事実だけを強調したかっただけです。この協力が今後どのように働くかについてお話する時間はありません。すでに昨日、Susan Armstrong氏がこのプロジェクトについてお話しているからです。このプロジェクトはEAGLEのモデルで、EAGLE内部でも支持されています。私がこれについて国際協力の枠組みの中で述べた理由は、目標の１つが、アメリカがMUC-TIPSTERにおいて開発した材料に匹敵するものをヨーロッパが手に入れることであるからです。

もちろん、この努力の結果は仕様規格の開発に取り入れています。申し上げたように、規格はコンセンサスが得られると考えるものを基本にしています。もちろん、この規格がヨーロッパのコンセンサスを得るだけでなく、アメリカや日本の協力が得られることを希望しています。アメリカについては、すでにEAGLEとALPACの間で公式の連帯を確立しており、５つのグループを有しており、文書も交換しております。日本にも同じ様なものがあるかどうかは知りませんが、もし類似のものがあれば、ぜひそれを知りたいと思っています。EAGLEの協力者として連帯を確立することには非常に興味があります。

私たちは科学的な協力も推進したいと思っています。２月にヨーロッパ会議を開催する計画があり、産業界、学会、ユーザーなどの間で、国際協力に関するヨーロッパの共通の立場を確立する予定です。その後５月か６月に、他の国からも関係者を招待する組織づくりを計画しています。皆さんもすでにご存知のように、COCOSDA研究やLERIC計画は学術的なものであり、時としてアカデミック過ぎることもあ

ります。ヨーロッパ、アメリカ、及び日本は私たちが必要としている言語資源について論じようとしており、NLPに関するスピーチと同じ種類のものを開発しようとしています。国際協力に関するこれらのプロジェクトの中で私たちが考えている目標の１つは、国際レベルで協力を刺激することばかりでなく、共同の国際協力構造を設立することです。会議からは何かが生まれてくると期待しておりますが、NLPとスピーチ・コミュニティーの間の協力も支援するつもりです。

将来の展望については話さないつもりです。それはこの会議の話題であるということを私は十分に理解しているからです。少しだけ、議論を提供しましょう。言語資源を共有することは各方面から考えてむずかしい問題であると思います。既存の資源の共有に限定するべきではないと考えていることが議論の種になるでしょう。より関心のあるものは、新しい資源を設立するために協力をする事です。もし現在ある言語資源が適切でないことが分かれば、新しい適切な資源の開発のための協力に集中する必要があります。

私たちは各種のレベルで協力をする事ができます。設計面でも協力はできます。言語資源の内容は何でしょうか。例を上げてみましょう。NERCはあるモデルを開発しました。コーポラに関する日本のプロジェクトがこのモデルを議論することを始めてくれれば非常にうれしく思います。

技術的な仕様のレベルでも協力はできます。共有性を確実なものにするためには、最初の段階から適合性、交換性を確実なものにしておくことが好ましいと思います。現在、EAGLEではあらゆる入力、評価、フィードバックを受け入れる態勢を取っています。より困難なのは、将来の資源の開発を計画する上で協力を行うことだと思います。

計画については、私たちは同じ優先度を持っているでしょうか。どちらの方向で計画を始めたら良いのでしょうか。１つ例を上げてみます。ヨーロッパにおけるある例は、語彙目録用にある種のパイロット・プログラムを開発することです。それぞれの言語についてたとえば、25000個の単語で構成された仕様、ツール、及び核を指定したとします。このアイディアが優れているかどうかは分かりません。これが計画という言葉や構築における協力という言葉で意味していることです。

これはそれほど単純ではありません。将来の資源は単一言語であることはなく、たとえば２ヶ国語辞書や対比文法などのように、多言語方式となるでしょう。イタリアでは、日本語－イタリア語の２ヶ国語の辞書を構築する力があるかどうかは分かりません。通常は、もしこの種の辞書を構築したい場合には、長尾眞教授がそのお話をされております。その仕様については合意する必要がありますから、できれば協力することの重要性の理由はすぐにお分かりになると思います。もちろん、協力できるかどうかは言語資源の状況に依存します。もし言語資源が公共領域で利用可能であるならば協力はずっと簡単になるでしょう。

お話申し上げたように、あるヨーロッパのプロジェクトではヨーロッパにおける協力を見つけようとしております。国際レベルでの協力の必要性を考慮した、この会議の組織関係者の皆様にお礼を申し上げます。有り難うございました。私の英語がひどかったことをおわび申し上げます。

座長：

Zampolli教授、言語資源を目的とした国際活動に関するご講演をありがとうございました。

時間が迫っていますけども、１件あるいは２件位、短い質問あるいはコメントがありましたら、お願いします。Professor Zampolli, I'm asking question or comment O.K.?

ありますか。Is there question or comments?

それでは、Zampolli教授、どうも有り難うございました。

(3)「ドキュメンテーションつき多目的電子化リソースの作成、保守、利用におけるＴＥＩの役割」

座　長：

それでは、今日のこのセッションの３番目の発表に移りたいと思います。Susan Hockey先生によります、「ドキュメンテーションつき多目的電子化リソースの作成、保守、利用に関するＴＥＩの役割」ということで、お話頂きます。

Susan Hockey先生の紹介に関しましては、皆さんアクティビィティーを御存知だと思いますが、現在、ラトガーズ大学とプリンストン大学の共同によって１９９１年に設立されました、Center for Electronic Texts in the Humanitiesのディレクターをなさっておられます。御存知のように、Hockey先生は長い間オックスフォード大学のComputing Serviceのセンターにおられまして、それから、１９９１年にアメリカに移られまして、このセンターのディレクターをなさっておられます。皆さん御存知のように、"A Guide to Computer Applications in the Humanities and SNOBOL Programming for the Humanities"の、世界的によく知られています本とか、いろいろ出版されておられますが、現在Association for Literary and Linguistic Computingのチェアーをなさっておられます。Please.

Center for Electronic Texts in the Humanities
Professor and Director　Susan Hockey

有り難うございます。それでは私は昔はヒューマニティーにその源を持っていましたが、現在では多くのいろいろな分野に応用されているプロジェクトについてお話致したいと思います。これは電子テキスト資源に関連しており、言いかえれば、コーポラや、その関連資料に関係しています。ヒューマニティーに源を持っていたというのは、ヒューマニティーの研究者たちが１９４０年代の末ごろからヒューマニティーの研究の応用として、電子テキストを利用していたからです。そのプロジェクトはＴＥＩ(Text Encoding Initiative)であり、これらはコンピュータ言語、コンピュータ計算、それにヒューマニティーにおける３つの主要な学術研究協会が協賛しています。

講演の中で、私は１９８７年末以来のＴＥＩの概要についてお話したいと思います。次にＴＥＩのガイドラインについて考えてみたいと思います。ＴＥＩがどのようなものか、そして次にこの会議で討論された分野のいくつかにおけるＴＥＩの役割についても考えてみたいと思います。

ＴＥＩは各種の資金を受けており、始めにこれをお知らせするのは適当でしょう。主たる資金はアメリカ国立ヒューマニティー基金とヨーロッパコミュニティー委員会から得られたものです。メロン基金及びカナダの社会科学研究理事会からの補助金もあります。しかし最も重要なものはその他の研究所のホストからの直接、間接の支援です。ボランティアによる非常に多くの時間と努力が寄せられています。

このプロジェクトはどうして始まったのでしょうか。以前、電子テキストをコード化したり作成して

いたＴＥＩの人々は、非常に多くのいろいろなエンコーディング・スキームやメイクアップ・スキームを有していました。既存のスキームに関係する問題のいくつかをお見せしましょう。これらは原作者の研究の関心を反映するように設計されたもので、そのため、特定の対象とする分野に対してのみ利用可能でした。あるアプリケーション・プログラムに対し特定であるため、拡張規定がなければ完全には柔軟性があるとは言えません。これらは非常に粗末なもので、文書もありません。そのため、電子テキストの変換には非常に多くの時間がかかりました。そしてこの作業はまだ続いています。ＴＥＩはこの問題をすべて取り除くものと考えられます。

ＴＥＩは１９８７年１１月にバサール大学での計画会議を開催することから始めました。これはNCIとコンピュータ及びヒューマニティー協会が召集したものです。その会議には３１人の参加者が集まり、テキスト及びコーディング・スキームのニーズについて討論が２日間集中して行われました。この会議はポウキプシー原理と呼ばれる結論を出して閉幕しました。ポウキプシーにあるバサール大学で確立された一連の原理は、ＴＥＩがそれに基づいて構築されたものであります。

これはＴＥＩ作成を前進させるる最初の試みではありませんでした。なぜこの会議がその終わりにおいて多くの合意が得られ、それ以来なぜ多くのことが起こったのかを非常に簡単ですが振り返ってみるのも意味があると思います。

明らかにある時点までは、私たちは電子テキストについて約４０年間も研究を続けてきました。そしてコード化に関連する問題点についてより多くのことを知りました。またその会議で、参加者は、その分野における主要組織のほとんどを代表していたと思います。また、ＴＥＩやその他の分野にとって非常に重要となった第３のポイントは、その時までにSGMLが国際規格となっていたということです。これはStandard Generalized Mark-up Language（標準一般化マークアップ言語）の略です。

それではＴＥＩは何をしようとしているのでしょうか。その目的は「ヒューマニティー及び言語産業においてもっと一般的に応用できる電子テキスト言語資源のための共通コード化フォーマットを定義すること」であります。それは新しいテキストをコード化したり既存のテキストを交換したりするにもどちらにも利用できるからです。

バサール大学での会議後、１つのプロジェクトの枠組みができあがりました。それについてごく手短にその概要をお話します。ＴＥＩの方針決定委員会は３つの協賛団体のそれぞれから、２名づつの代表を参加させております。そしてこの委員会は、現実的には専門委員会として非常にうまく機能しています。プロジェクトには２名のエディターが作業をしており、１名は本日もアメリカから来ておられるマイケル・スパーバーグ・マッキーン氏であり、もう１名はヨーロッパの方でオックスフォード大学のルー・バーナード氏であります。

このプロジェクトは諮問委員会ももっており、その代表は１５の主要な学術協会から来ておられ、そのほとんどはアメリカをベースとしていますが、アメリカ現代言語協会、アメリカ情報科学協会、ACL、アメリカ歴史協会などのような外部グループも入っています。それに国際図書館協会連合からの代表者も参加しています。

諮問委員会は、ＴＥＩが応用できる主要な学術機関をカバーする意向であります。Antonioが前の講演で述べたように、これには一連の関連プロジェクトがあり、それについては少し後で述べることとします。

バサール会議後、方針決定委員会が発足し、プロジェクトの資金をを見つけ、技術的な作業をするためのプロジェクトの枠組みを確立しました。この枠組みは計画会議でなされた提案をベースとしています。コード化の問題を見ると、現実には４つの異なった観点からそれぞれの見方で見ることができます。１９８８年の中ごろ、少なくとも１９８８年の終わりまでには４つの作業委員会を設立しました。それについてもう少し詳しくその分野のことを説明しましょう。ここにその関係のスライドがあります。

私の知る限りでは、ＴＥＩが行った作業の１つは、現実には他ではまったく行われたことのないものでありまして、電子テキストの文書化を行うメカニズムを提供することであります。そしてそれは、文書化委員会の焦点でもありました。そして図書館や文書管理の世界から多くの専門知識が寄せられ、情報管理を行う経験を持った人々や、情報管理のための規格の重要性を認識している人々が集まりました。

第２のグループは、物理的な表現法とテキストの論理構造を取り扱いました。それはテキスト表現委員会と呼ばれています。また彼らは文字セットの表現の問題も検討していました。第３のグループは、分析と解釈の問題を取り扱っていました。オリジナルテキストではそれほど明確ではない情報についてです。その委員会は言語解析と解釈に焦点を当てました。

第４のグループは、シンタックス及びメタ言語委員会と呼ばれており、間もなくSGMLが、ＴＥＩが前進できるベースを提供するものであるという決定を下しました。そしてそれ以来、その提供に関して研究を続けています。これについては私はＴＥＩの中でのSGMLの利用に関するある種のハウス・スタイルと呼んでみたいと思います。

実際には１９８８年の末から１９９０年の始めまでにこれらの作業委員会がいくつか開催された後、最初のＴＥＩ勧告が出され、これがＰ１と呼ばれる文書です。草稿の最初の文書は１９９０年７月に印刷されて発行されました。それはおよそ３００ページのものです。これにはいろいろなものが欠けていました。最新の版では、現在非常に大きなものである付録がまるでついていませんでした。ある部分は詳細に研究が行われていましたが、ある部分は詳しくありませんでした。また全然記載されていない部分もありました。しかしながら、その考え方は皆さんがコメントできるように何かを作り出すというものでありました。それでこの文書が日本でもかなり広く出回ったということを私も知っております。

Ｐ１の出版後、プロジェクトは第２開発期間の段階に移行します。その目的は実際には２つありました。バージョンＰ１の中で詳細に渡って述べられているガイドラインをテストとすることであり、もう１つは適用範囲を拡張し、前に行われていなかった部分を検討することでした。さらに２つのグループが、あるいは作業グループのようなものが、第２開発期間中に出来上がりました。

第２開発期間中、文書委員会、SGML、シンタックス・メタ言語の作業はある程度続けられましたが、その時までに、その作業の大部分は終了していました。技術的な作業の殆どは、これらの厳密に焦点を絞っていた作業グループにより終了していました。そしてこれらが作業グループの中で対象となった分

野のいくつかです。

作業グループはテキスト表現委員会、分析もしくは解釈委員会のいずれかに割り当てられました。テキスト表現グループは文字セットを研究しているサブグループをもち、また、別のグループはテキスト評価、テキストの異なったバージョンの問題点、及びテキストの多様性に対応する方法を対象としていました。別のグループはハイパーメディアなどのメカニズムを取り扱っていました。別のグループは式とテーブルを対象としていました。別のグループは言語コーポラを研究していました。また他のグループは、ソース・テキスト、主として手書きのもの、初期の印刷本などの物理的特性の表記方法を研究していました。主にドラマや文学的な詩を含む、散文のテキスト、演劇のテキストも研究されていました。

分析及び解釈作業グループは、当初これらの分野ー言語分析のメカニズム、コード化された話言葉、歴史的研究、歴史分析、及び解釈、辞書、コンピュータに関する語彙及び用語ーを対象とする計画でありました。用語作業グループについては、本日、先ほどお話があったと思います。これは実際にはいくつかの他の用語（グループ）の努力の結果の一部であり、ＴＥＩから全体の用語の分野に渡る１つのリンクを提供してきました。そして、これはＴＥＩがすでに存在していて、作業中のグループに参加できた分野の１つでありました。

作業グループが設立されたちょうどそのころ、ＴＥＩは「関連プロジェクト」と呼ばれるものを多数設立しました。その発想はガイドラインをテストするため、電子テキストの作成や、保持に取り組んでいたいくつかのグループ・プロジェクトを獲得することでした。そして彼らは異なる分野の代表者として選ばれたのです。彼らは改訂を含むＰ１勧告に従ってテキストのサンプルをコード化し、その結果について報告するよう依頼されました。

ＴＥＩは各プロジェクトにコンサルタントを割り当てました。これらの人はＴＥＩのガイドラインに非常に詳しく、プロジェクトを支援しました。プロジェクトやコンサルタントのために、ワークショップもいくつか開催しました。関連プロジェクトの範囲に関するいくつかの見解をご紹介しましょう。英国国立コーパスもそのひとつですが、ストックホルムにももう１つのコーパス・プロジェクトがあり、それはライデンにある、アメリカのテキスト研究をしているグループです。もう１つのグループは、初期の女性の筆記を多くコード化しており、それは地球規模のユダヤ人のデータベースとなっています。この考えは異なった分野、異なった自然言語を対象としようとするものであり、これらの成果から共通なものをまとめようとしています。

第２開発期間中、更にいくつかの技術検討会議が開かれました。１９９１年にはノルウェーのミルダルにおいて開催されました。１９９２年の５月にはシカゴで開催され、第３回目の会議はオックスフォードで開催されました。ここではこれまでの完全な草案をもう一度検討しようというものでした。これらの会議は実際には約２０人の参加者があり、２～３日の非常にインテンシブな作業を行いました。

これがプロジェクトの概要です。それではSGMLについて、ほんの少しだけお話することにしましょう。なぜ私たちはこれを利用することを決定したのでしょうか。もう皆さんの中にはSGMLについて十分ご存知の方がおられるかと思いますが、それの内容について概観をお話します。

まず最初に、それはメタ言語であります。それはコード化スキームの一種のフレームワークです。それは「規範的というよりは記述的である」という原則に基づいて構築されています。これは情報を記述し、プログラムとは別にデータを保存しておく方法です。SGMLはデータを記述します。それはプログラムであり、もしくは特定のアプリケーションであり、それを使用して何を行うかを決定するものです。そのため、SGMLのエンコーダー・テキストはいろいろの異なった目的に利用することができます。それは非常に柔軟なものであり、拡張も可能で、このことはＴＥＩのようなアプリケーションにとっては非常に重要です。

これはマシンやアプリケーションからまったく独立しており、形式的構造であるため、コンピュテーショナルな意味で処理可能であり、他のものよりはるかに優れています。これは多目的のテキスト、多目的資源用のためのフレームワーク、土台を提供するもので、次世紀を超えて十分に利用できる寿命を持っています。このことは、私たちが今までにあった非常に多くの電子資源について持っていなかったものです。これは非テキスト型資料を記述することにも利用でき、私たちがマルチメディアの世界に入って行くため、非常に重要となります。

ＴＥＩガイドラインの原則とは何でしょうか。そのガイドラインが何を目的としているかということについては、ごくごく単純にお話をしてきました。それは電子テキストをコード化する際、何を、またどうやってコード化するかについてガイダンスを与えることです。実際、ガイドラインは非常に包括的なもので、人が電子テキストの言語資源にコード化したがるような、あらゆる種類の特徴を見渡したものとなっています。それは非常に多くのSGMLエレメント、SGMLテキストを提案してきました。絶対的な規則はほとんどありません。それはこの特徴をコード化したい場合、この方法でやるという一種の哲学みたいなもので、多くの場合それをコード化する必要はありません。

これについてもう少し掘り下げてみましょう。私が前に言ったように、考え方は「様々な異なった目的に利用できるテキストや資源を持つこと」であります。そのため、ガイドラインはすべてのテキスト、すべての資源はいくつかの共通な、核となる特徴を共有するという考え方に基づいて作られています。そしてこれらの特徴に加え、ある規則、応用、もしくは理論的方向付けに特化した特徴を追加することができるのです。これはあなたが望めば自分自身のSGMLテキストを定義するメカニズムとして柔軟で、拡張性があるものです。非常に重要なことはこれらはテキストのマルチ・ビュー（多角的視点）を認めるということです。これはどんな特別な意見にも特権を与えるものです。それは同じ特徴となるもののマルチプル・エンコーディングを認めます。そのため、同じテキストにコード化されている同じ特徴について、非常に多くの異なった見方をもつことが可能になります。

しかし、もう一度繰り返しますが、強制的なマークアップSGMLタグやエレメントなどはほとんどありません。また、これらは適切な文書が電子テキストに提供されなければならないという考えに基づいて作られています。もちろん、SGMLはASCIIファイルであり、これは簡単にすべての既存ネットワークを使用して転送することができます。またＴＥＩでは、全ての既存のネットワークを通じて転送される、最大公約数のような文字列の勧告を出しています。

それではＴＥＩ適合テキストとはどんなものでしょうか。それは２つの主要な要素からなっています。テキストを文書化する部分のヘッダと、テキスト自身です。これらについてもう少し詳しく見てみましょう。最初のものについて少し詳しくみてみます。SGMLに詳しい人は、文書タイプ定義の原則がお分かりになるかと思います。これはもし希望すれば、テキスト中の、互いに関係しながら、DTSとも関係しているエレメント、またはマークアップ・タイプを定義するものです。

```
<TEI.2>
  <teiHeader>
    <fileDesc>
      <titlStmt>
        <title>Title of Work
      <publicationStmt>
        <p>publication/distribution information
      <sourceDesc>
        <p>bibliographic description of source
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
    All the text here
  </text>
</TEI.2>
```

ＯＨＰ－１：Minimum TEI Document

ＴＥＩはアメリカではピザを作るのになぞらえられる方法で、そのDTDを構築しています。ベースを選択し、次にそれの上に乗せるトッピングを選択します。いつでも文書化部分であるヘッダの中のタグ（コア・タグと呼んでいる）を取り出し、すべてのタイプのテキストに共通している一連のタグを取り出します。次に特定のタグ・セットのベースを選択する必要があります。そこに私たちが持っている主なものを略述しておきました。これらを組み合わせることも可能です。もし何らかの理由で辞書と散文をいっしょに欲しいという場合には、これらを１つのベースに組み合わせることが可能です。

ピザのアナロジーを続けますと、ベースを選択したら、次はトッピングの選択になります。追加のタグ・セットについては、そこに変数となるものを示しておきました。ペーパーには、ＴＥＩのガイドラインの内容表をのせてあります。これらとガイドラインの中の特定の章との間には大量のマッピングがあることがおわかりでしょう。これについてはあまり詳細にはお話しませんが、私たちが研究している分野の幾つかを説明しております。

ＴＥＩヘッダー、電子テキストの文書化についてもう少し詳しい情報をお話しします。そこにＴＥＩ

ヘッダーの４つの主要なセクションについての概要を述べておきました。これについて手短に話してみようと思います。その理由はこれがＴＥＩの非常に重要な側面となるからです。特に、共有された再利用可能資源の分野に行くと何を得るかということがとても重要になってきます。

Documentation of electronic text

<TeiHeader>

<fileDesc> ... </fileDesc>

<encodingDesc> ... </encodingDesc>

<profileDesc> ... </profileDesc>

<revisionDesc> ... </revisionDesc>

</TeiHeader>

ＯＨＰ－２：TEI Header

ファイルの記述はテキストの出所を記述します。実際、ファイルの記述は、マークレコードをそのフィールドにほぼダイレクトに対応させるエレメントをその中に持っています。そのため、ライブラリアンはカタログレコードを作成するためのワークシートとしてそれを利用します。これにはタイトルの記述、出版の記述、テキストがどこから来たのかというソースの記述などが含まれます。

コード化の記述には、テキストにどのような特徴がコード化されているかという記述が含まれています。そして、もし分析や解釈がテキストの中に入っている場合、なぜそれをそこに入れたのか、もしくはどの特定の言語スクールに追随したのかということを聞くことができます。繰り返しますが、テキストユーザーにとって、なぜ特定のものがその中にコード化されたかを知ることは非常に重要です。

プロファイルの記述はある追加情報を含んでいます。それはテキストの中で生じる自然言語リストのように、他のものには論理的にフィットしません。もしテキストが再び、ライブラリーの世界から何かを持ってきて、よく知られた分類システムを利用する場合にはそれを文書に書いておきます。それがスポークンテキストであるならば、会話中の参加者についての情報も記述されるべきです。改訂の記述はコンピュータコードのどの部分でも、発見したその時にテキストになされた変更のヒストリーを示しています。叙事詩のテキストはこれに似ています。テキストはオプションの前方事項、本体、及びオプションの後方事項から構成されています。

ミニマムＴＥＩ文書は必要であるものの概念を与えてくれます。事実、そこにはいった時、強制的と

なる唯一のタイプは、人々が文書化情報やテキストを取り入れることに関してのみです。

今、私たちがどちらの方向へ向かっているかということについて、ＴＥＩとの関連性を検討してこの講演を終わることと致したいと思います。SGMLは他のどのマークアップ・スキームやコード化スキームよりはるかに豊かなもので、確実にテキストコード化に含まれる知的問題を処理することができます。私が知っている、あるいは今まで読んだどのマークアップよりも優れています。更に、非テキスト系の資料にもリンクしていますし、音声や映像もSGMLで描写することができます。

ＴＥＩにいる最高のものとして努力してみても１つ問題があります。SGMLはシングル・ツリーである文書の概念に基づいて構築されたものです。そして、事実、私たちが取り扱おうとしているテキストはほとんどありません。また私たちが検討しているアプリケーションはシングル・ヒエラルキーでありますが、それについては何らかの解決策を提案しています。

ＴＥＩが使用されるようになると、ディジタル・ライブラリーや知識ベースとはより深い関係になるものと思います。ディジタル・ライブラリーについては、アメリカでは現在かなり関心がもたれています。来世紀に入ろうとしている時に知識ベースとしてここに記載したものからあまり離れているとは思いません。また、単純に、ＴＥＩはプロジェクトの範囲、取り扱ってきた資料の種類などこの分野で多くの利点を有しています。事実、マルチ・ビューと解釈は同じ資料の中に入るでしょう。

ヘッダは人々が「自分は何を持っているか」を知っているという点で、図書館のカタログやテキストに対する認証の方法にリンクを与えます。また、標準的な図書記述法も利用します。図書館のカタログ収集者が資料を集めることができるように、ＴＥＩのヘッダーだけを交換する方法があるでしょう。

非テキスト系資料とのリンクはますます重要になると思います。昨日横井博士はマルチメディア、ハイパーメディア及び知識ベースにおけるそれらの役割について多くのことを話されました。ＴＥＩはマルチメディアやハイパーメディアの情報も記述することができます。

最後に一言、最も重要であると考えられることは昨日、横井博士が述べられたことです。ＴＥＩはある広範なポインター・メカニズムに対する提案を有しており、それらは実際にはグローバルなハイパーテキストの基礎を提供するものです。そのため、グローバルなネットワークで結ばれたライブラリーや知識ベースなどを有している状況を想像することができます。

ＴＥＩの拡張されたポインター・メカニズムはテキスト資源とその中のエレメントの間にリンクを提供し、異なった場所で異なった人々が資料に注釈をつける通信コミュニティーを持っているという状況を想像することができます。それは注釈を文書化するメカニズムを提供し、これはもし同じ資源で研究をしている多くの異なった人々がいる場合には、極端に重要なこととなります。また、同じエレメントについて多数の注釈をつけるメカニズムを提供するため、同じ資料について異なった見解を与えることになります。ご静聴有り難うございました。

座長：

共有可能な知識資源に対するＴＥＩの持つ重要な役割についてすばらしいプレゼンテーションを有り

難うございました。このお話に対して何かご質問、またはコメントはございませんか。

それでは、どうも有り難うございました。

(4)「Ｃｙｃ：知識共有の先駆け」

座　長：

それではこのセッションの最後の講演に移りたいとおもいます。講演者はＭＣＣのDr. Douglas Lenatさんです。タイトルとしては「Ｃｙｃ：知識共有の先駆け」というタイトルでお話になります。Lenatさんらしく、非常にアトラクティヴなタイトルをつけておられますが。もうLenatさんに関しましては説明の必要もないかとも思いますけれども、現在ＭＣＣのプリンシパル・サイエンティストでご活躍でございまして、Cycプロジェクトを引っ張っておられる方です。それと同時に、現在スタンフォード大学でコンサルティング・プロフェッサーもなさっておられます。

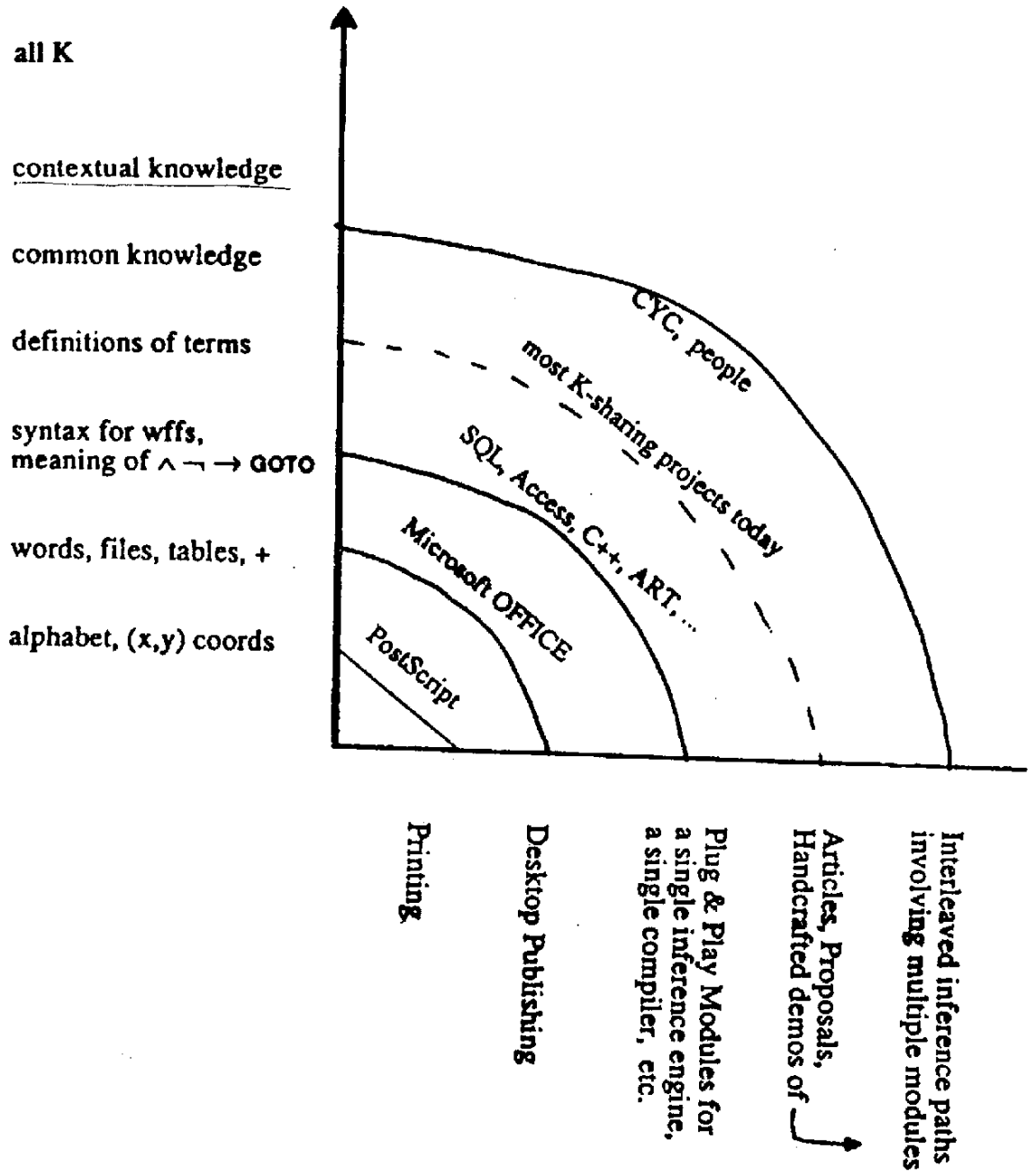
今日のタイトルは、非常に興味深いタイトルでございますが、"Priming"という言葉は、「前もって通報しておく」とか「予備知識を与えておく」という意味もありますけれども、一方ではちょっと変わった意味で、「火薬を詰める」という意味もございまして、そういう意味で、これからのこの分野の研究をさらに押し進めるという意味も込められているのではないかと思います。

Cycプロジェクトが始まって、もうだいたい１０年になりつつありますけれども、現在までの開発の動向を振り返って、それから最近の、新たに加わっている機能、それから１９９５年以降のいろんなプロジェクトの事に関しましても、お話になられるかと思います。So please, Dr. Lenat.

Microelectronics and Computer Technology Corporation
Director of the Cyc Project　Douglas B. Lenat

本日は２つの事についてお話したいと思います。まず、私たちが共用していなければならないものは何か、ということであり、２番目は、皆さんと共用するものについて、今までのところ何をやってきたか、ということです。

もし、われわれが共用するあるものについて考察を行えば、それは殆どゼロから無限に近いところまでのスペクトル（領域）になるでしょう。皆さんもお考えのように、共用するものが多くなればなるほど、共用が可能にする機能が増大するのです。ですから、例えばここに同じようなスペクトルがあるとします。そして皆さんが共用したいと思うものが、アルファベットであったり、X-Y座標の概念であったり、無彩色スケールやカラースケールのようなものであれば、本を印刷する際に、後書きなどの決りきったものを書くよりは、もう少しましなものができるでしょう。もし皆さんが次のステップに進みたいと考え、例えば次にSusan Hockeyさんが話されるような情報を共用したいと考えられるならば、皆さんはＴＥＩのような仕事ができるわけです。マイクロソフト・ワードからある事項を取り出し、エクセルかマイクロソフト・ナウに入力するだけで、デスクトップ・パブリッシングができるのです。



OHP－1

もし皆さんがプログラムの式の文法や式がうまくできていることを示す構文に専心したいと思っているなら、また条件のあるIF文やGO TO文やAND,OR,APPLYなどの意味する構文に専心したいなら、ある形式でプログラミング言語を持つことができるのです。またSQL(structured query language)や特定分野の問題解決プログラム・ソフトのArt and Keyなどのような、プログラミング言語とは考えられないものや、C言語などのようにプログラミング言語として考えられるものもあります。一般的に、ひとたびそれらのプログラミングを行えば、皆さんは本質的にプラグ・イン・タイプのモジュールを持つことが

でき、ひとつのシングル・エンジンや翻訳プログラムで異なったモジュールを実行することができ、最低限異なったモジュールの実行から得られるアウトプットを共用するというレベルにおいて、さらにもう少しの多くの共用が得られるというわけです。

もし、皆さんが用語の意味や、それらの用語などについての知識を決定したいと思っているなら、皆さんはそれと同時に、まるで密閉されている状態のモジュールではなくて、むしろ我々がみんな欲しがっているであろう、フル・インターリーブさせておくモジュールを含んだインターリーブ推論経路にアプローチを始めることができるのです。そして、ある意味では、私は愚直にRon Brachman氏の発言に全面的に賛成であり、溝口理一郎氏にも全面的に賛成ですし、Bob Wielinga氏の発言にも大々的に賛成であると申し上げれることによって、話を簡単にすることができると思います。小さなことですが<私がある意味でBill Swartout氏と意見が一致しないことに関係するのですが、私はただ用語を共用することでなく、用語に関する知識、それも結局文脈的につながっている知識や文脈的に異なる知識を共用することが重要だと思うのです。ですからある意味で、このスピーチではその点に焦点をあて、あるいは課題とし、あるいは論点としたいと思います。

これらの異なった、私たちが共用するもの全てについてお話する時間もありませんので、できるだけ省略をして、興味深く議論の対象になるようなものをお話したいと思います。例えば、皆さんが全く議論の余地がないとお考えの、辞書に載っている定義の共用というような低次元の話にしても、複雑な問題点があるのです。異なった語義・慣用句をどう扱うか、その言葉の持つ隠喩表現をどう扱うか、俗語を含んでよいかなどです。ですから辞書の定義ということに加えて、あるいはそれ以上に、私たちはこうした用語についての一般的な知識を共用したいということなのです。そして、この一般的な知識によって、私たちは一度も会ったことがなくても、あるものについてお互いがそのことを知っていると仮定できるということなのです。例えば人は「夜眠る」ということを知っている、「ものを食べれば空腹でなくなる」「死ねば動かない」などなどです。もし、これらのことに同意が得られないとしたら、意思の疎通が困難になるばかりでなく、協力はほとんど不可能となるでしょう。

さらに、疑わしい領域がいくつかあります。たとえば、隠喩の共用も含むのか、そしてそれは「一部」を「全体」といい「精神」を「肉体」という、普通の隠喩なのか。あるいは「永久に前後関係のある知識」とも言うべき、歴史・文学・神話なども共用するのか。皆さんは次第にお分かりだと思いますが、大変異なった文化を持つ人々同士では共用するものがどんどん少なくなっていくのです。

今、前後関係のある知識についてお話しましたが、それはしばしば異文化間の知識でもあるのです。例えば誰かが私に「どうしてMCCに働きに行ったのだろう」と言ったとして、私が「ああ、満月だったからねえ」と答えたとします。日本人でしたら、私がまもなく人生を円満かつ完璧に終わろうとしているから、そこへ行ったのだと思うでしょう。ところがアメリカ人だったら、私が狂気に導かれてそこへ行ったと思うことでしょう。このように文化による脈絡の差というものがしばしば起こり得ますが、日常生活においても同様に脈絡の差があるものなのです。例えばこの部屋で電話が鳴ったとすると、会議をしているのに邪魔だな、と感じるものですが、これがロビーにいる時とか自分のオフィスに座って

いる時には違って感じられるものなのです。

知識の断片によってなされる仮定を用いてお話しましょう。もし雨が降っている時に「傘を持っていったほうがいいですよ」というのは自然に聞こえますし、もし皆さんが注意深くなければ、「もっと身を入れなさい」とか「システムを起動しなくては」というのは前後関係においては正しいものと言えます。また例えば皆さんがもし室内にいて傘を持って行こうとなさるのでしたら、外出されるんだなと仮定することができます。助言というものがあてはまるのは、子供でないこと、麻痺していないこと、正気なこと、そして「人」であることだと思います。私は75,000年もの昔に正しかった助言についてお話しているのではありません。また、貧しくて傘が買えないような人も含めた、地球上の全ての「人」にあてはまる助言についてお話しているわけでもありません。

重要なことは、もし皆さんがある主張や基準を一つの文脈から別の文脈に移す時には、最初の文脈の仮定－必ずしも別の文脈の仮定にはならない－を持って来るでしょうし、必要ならばそれを明確にするだろうということです。ここで大事なのは「エトセトラ」という言葉なのです。なぜなら、一般的に言って、皆さんはこうした仮定の全リストを完璧に作成することは絶対にできないからです。必ず漏れが出ることになります。ある意味ではこれは私たちのプロジェクトを可能にしてきた鍵となる洞察力なのです。皆さんが規則の前後関係を完全になくすことは決してできないということ、そしてそれゆえに、明白な文脈を用いて包括的な一致を得ることなしに特定の一致が得られるのです。そしてそのおかげで、皆さんは「多量ではあるが莫大ではない」明確に表現されたWielingaの用語にある「分離され、特定の分野で一致された知識のベースであるプレート」を持つことができるのです。私たちに意見の不一致はないと思いますが、彼が「分離された知識」と呼ぶものは、私が「分離された文脈」と読んでいるものなのです。

「知識の共用の必要性に拍車をかけている問題は何か」という話に移りましょう。これはフォーチュン・マガジンの最新号に全面記事で掲載されたものですが、おわかりのように本質的には隠喩で、「情報は蛇口から滴りおちているものであるから、情報の一滴までも手に入れるためには蛇口に口をつけて必死に吸い込まなければならない」と言うものです。「これが私たちが直面している問題か」という質問に対しては、答は「ノー」ということだと思います。これは基本的には私たちが直面している問題とは関係がないのです。私たちの問題はほとんど反対のものです。現代は情報が過多になっています。私たちの直面しているのが雫の滴っている蛇口でなくて、捻るばかりになっている消火栓のホースの先だという現象が多くなるでしょう。そして問題なのは、Ron Brachmanが「情報の大洋」と呼んだ情報の洪水の中から適当な情報を見つけ出すということなのです。

ですから私たちが本当に望み、また、知識の共用を望む理由は、私たちを導くのに役立つような強力な道具、私たちが適切な情報を見つけ、多量の不適切な情報は見つけないように役立つ強力な道具を必要としているからです。もしそれに適した道具が手に入るなら、情報が暗号化されていてもかまいません。集計用紙でもデータベースでも、表題のついたビデオのクリッピングでも構造は関係ありません。情報を見つけることができるはずだからです。またその情報がどんなに異質なものであってもかまいま

せん。私たちはユーザーとして異なったスキーマや異なった問い合わせ言語、異なった識別プロセス規約などを習いたくないからです。

それらの情報源に蓄えられている情報はできるかぎりチェックされなければなりません。論理的であるかどうか、一貫性があるかどうか、誤りがないかどうか、またその誤りがRon Brachmanが指摘していた種の誤りなのか、単純なミスによる誤りなのか、ほとんど共用はしているものの、完全には知識を十分に共用していないスキーマ同士の衝突によっておこる矛盾なのかをチェックするのです。問題の源が何であれ、私たちはこうした矛盾を見つける強力な道具が必要なのです。

私たちがシステムに質問する問い合わせは、論理的であるかどうかチェックされなければなりません。もし皆さんが回答に長時間を要するような質問をしたならば、それはそのシステムが論理的かどうかをチェックするよい機会なのです。もし皆さんの質問の中に冗長な節が含まれていたら、皆さんが単に問い合わせの誤記をしたという表示がでるかどうかなどをチェックするのによい機会です。

あるものに対して質問をすれば、構文上の一致だけよりも、単なるキーワードや同義語よりも多くのものが得られるはずです。理解がほんの少しだと、不幸なことにほんの少ししか理解を必要としない情報の洪水におそわれることになるのです。

それでは、なぜ「単なる用語より以上のものを共用する必要がある」と申し上げたか、「こうした用語を含む、一般的な知識の幾分かを共用する必要があるのか」を示すいくつかの例をあげてお話してみたいと思います。

例えば、皆さんがある会社の従業員に関する、リレーショナル・データベースか集計用紙のチェックをしているとします。名前・生年月日・入社年月日・配偶者名・緊急連絡先などが各欄に記入されています。表の２列目のMarry Willsの項を見ていて、いくつかの問題点があるのがわかります。生年月日のほうが入社年月日よりあとだったり、配偶者が誰か別の人と同じだったり、緊急連絡先が本人だったりなどです。人間である皆さんはもう何が間違っているかお分かりだと思います。すなわち、生まれる前に何かするなどということはありえませんし、意識不明で病院へ運ばれた時に、自分の名前が緊急連絡先というのもあまり賢いとは言えないでしょう。

皆さんはこうした誤りを見つけられる日常社会に関する知識はお持ちでしょう。私たちはいろいろな種類の専門家の推論についてお話しているわけではありません。私たちは正気で大人で…。どんな人でもこうしたものを見ていればできる推論についてお話しているのです。しかし、Ron Brachmanが言っているように、常に非常に多くの情報が増えており、人間が常に全ての情報を調べることは可能でもないし、コストもかかり過ぎるのです。したがって通常は、誰も全ての情報を調べませんし、ですからこうした情報の誤りが繁殖していくというわけです。

さて、人が生まれる前には何もしないなどということに対して、私たちはエキスパート・システムにルールを書き入れる事ができないのでしょうか。またアプリケーションにそうしたことを注意するようなＣコードを２行書くことはできないのでしょうか。答えは「イエス」です。もちろんできるのですが、もし皆さんが注意深くないと、たえずこうした断定を付け加えていることになるのです。それも何百万

という断定をです。ですから、集計用紙を作成する仕事は2，３日で終わる仕事ではなく、人が数世紀にもわたってする仕事となり、もちろんばかばかしいものですから、かつてなされたことがなかったのです。

嬉しいことに、こうした試みが一度だけされたことがあるのです。そして、それが私たちのとった考え方だったのです。Marvin Minsky、Alan Key、それに私が集まり、「我々は地獄のような狂気である」と自分たちを断定し、もはやそれ以後はこの作業をする予定はありません。私たちは１回だけ集まって、この知識を全て暗号化しようとしました。ですから、これが１９８４年以来私たちのグループとＭＣＣが取り組んでいることなのです。私たちは誰でもが知っていると思われる何百万のことを分類し、平均的な人が、集計用紙やデータベースや、説明つきのイメージ・データベースの中で誤りを発見できることを可能にする、何百万ものことを分類し続けています。もう１つうれしいのは、こうした試みは１度なされなくてはならなかっただけでなく、同じ知識が集計用紙ばかりかこうした仕事全てに用いられることができるようになったことです。

例えばMaryが土曜日にお話する用紙、Cycの用紙ですが、この中には説明つきのビデオ・クリップやオリンピック競技の静止画像が入っています。そしてここにいくつかの仮説のキャプションがあり、「幸福な人々の例をあげよ」のような質問があります。そうすると、それはＡとＢを適切な例として見つけ出します。キャプションの中には幸福に関して述べているものは何もないにもかかわらず、です。もっと面白いことに、シャツを着ていない男を呼び出すとすると、ＡとＣを適切な例として示します、というのは彼らは水着を着る水泳と飛び込みの競技に参加しているからです。マルチタレントの選手を呼び出せばＡとＢが返ってきます。これらのキャプションのどれ１つとして、それだけでは皆さんがその人をマルチタレントだとは結論づけられないことに気づいて下さい。オリンピックのレベルでは大変希な、水泳とトラック競技とフィールド競技のどれもに参加していることから結びつけられたものなのです。

失望している人々は飛び込みで腹を打った人に加えられます。ＡとＢがまた適例として返ってきます。なぜならビデオ・クリップの中のどこかで、競争に敗れた人々も勝者と同様に写されているからです。ですからＥ、Ｄ以外の全てが大変失望しているということになるわけです。

このように少量の理解でも、構文的にすぐれたシソーラスがないにもかかわらず、またイメージ・ベースではほとんど全てのものが返らずにこうした適例が返ってくるという期待はなかったにもかかわらず、説明のあるイメージを引き出すことが可能であることをお感じになったと思います。

異なる成分、私が前に述べた異なる種類の異なる成分に対処するかぎりでは、私たち独自の異なる成分を作りだすことができる、４つの攻略法があります。自然言語のインターフェース、ユーザー・モデル、自動的に推論するリレーショナル・データベース、リレーション、集計用紙の欄などです。そして、異なる成分をいっしょに写すために、科学者間で用いられる国際語に明確な公理を使用することです。これらの全てが私たちが過去１０年間にわたってCycに集めてきた知識のようなものを必要とします。これについては詳しくはお話しませんが、わかりやすい例をあげてみましょう。その中には例えば自然

言語の理解が含まれています。もし私が英語で「ペンが箱の中にある」と言ったとすると、皆さんはこれは何か小さな箱に入った筆記具だなと仮定するでしょう。もし「箱がペンの中にある」と言えば、ペンはある種の檻になり、箱はもはや小さな箱である必要がなく、大きな篭かカートンボックスであることになります。皆さんがそれを理解できるのは、英語や英語の単語の知識とは全く関係がないのです。それはそれらのものがどれくらい大きいか、通常どこにあるか、なぜいろいろな場所にあるのかなどという皆さんの知識に関係があるのです。

いつでもある種の省略とか「エトセトラ」を含む文がありました。皆さんは「エトセトラ」の意味するところをどのように描きますか。またどこでも見られる「彼ら」という代名詞ですが、「警察当局はデモ参加者を逮捕した。なぜなら彼らは暴力を恐れたから」という場合、「彼ら」というのはおそらく警察当局のことでしょうし、「警察当局はデモ参加者を逮捕した。なぜなら彼らは暴力を主張したから」という場合でしたら、「彼ら」はデモ参加者のことでしょう。それは英語には関係がありませんし、用語の意味にも関係がないのです。皆さんがお持ちのデモ参加者、暴力、警察当局などの一般的な知識に関係があるのです。

もし、例えばその文章を日本語に翻訳したいのでしたら、ここにあるコールド・ウォーターに対する日本語の訳語、ホット・ウォーターに対する日本語の訳語をどう使用したらいいか、皆さんはどうやって知りますか。それはまたもや英語には関係がないのです。ウォーターとかホット・ウォーターとかコールド・ウォーターという言葉の意味にも関係がないのです。皆さんはお茶とやかんについてご存じだと思います。「やかんが鳴ったらもう熱いのでそれ以上沸かす必要がない」ということもおわかりだと思います。

最後に、と言っても短いわけではないのですが、類推と隠喩があります。「昨日の株式市場はまるでジェット・コースターみたいだった」というようなものです。私たちが用いるほとんどすべての文章が、ある範囲までは、類推と隠喩をその中に作り上げているのです、そしてもし皆さんが「自分はかなり意味の限定された文章で書かれた資料しか扱わない」とお考えでも、その資料を取り扱うのに含まなければならない、表面的に広範囲な知識の量に驚くでしょう。これが人々が類推と隠喩を除去したり、トーン・ダウンしようとしてコントロールされた言語を使用しようとしている理由の一つなのです。

不均質な情報が与えられれば、自動的に推論されます。集計用紙の欄の意味やデータベースの範囲などです。ここに人々の氏名と電話番号が載っている集計用紙があるとします。この欄は事務所の内線番号でしょう。この欄は多分自宅の電話番号でしょう。この欄はおそらく入社した年度でしょう。この欄は最初の２人については時給を示し、３番目の人については年俸を示しているのでしょう。ここは緊急連絡先の氏名と電話番号でしょう。この欄のうち、３つまでが電話番号であるにもかかわらず、私たちはそれが何を表しているかを示すのに、なにも問題がないのです。一般的に、例えこれらの言葉の意味が非常にあいまいであっても、できるかぎりの定義集があっても役には立ちません。ほんのちょっと常識を働かせれば、その言葉の意味するところ、例えば時給なのか年俸なのかがわかるのです。見れば総額がわかり、職業もわかるというわけです。

この点についてはすでに他の方が述べておられますので、くどくど述べるつもりはありません。私たちが１９８０年におこなった実験は関係のあるものをとりあげ、物を優先し、テキスト・データベースを使用していました。さらに国際語としてCycを使用し、新車の特集記事とか顧客の満足度調査などのデータベースを含んでいました。Cycはある人がが新車を購入するのに役立つような、異なるデータベースや集計用紙からのデータを相互に関係付ける共通言語として用いられました。この実験はコンスタントに日本車ばかりを勧めるのですぐにやめてしまいましたが、考えはおわかりいただけたと思います。

では第１部から第２部に移りましょう。第１部では主に「私たちが共用すべきものには何があるか」「要求される機能性を入手するには何を共用しなければならないか」ということについてお話しました。第２部でも「何を共用しなければならないか」という同じタイトルになりますが、ここでの意味は「私たちが持っているのは何か、私たちが作りだしたもので共用する価値のあるものは何か」ということです。つぎに、これらについて少々お話したいと思います。ご存じのようにこれらについて詳細にお話する時間がないのですが、興味をお持ちでしたら、ご自由に質問なさるなり、休憩時に質問なさるなり、ワークショップでお話下さるようお願いします。

私たちが現在の状態になるまでに追求してきたいくつかの原理を調べることがさらに重要だと思います。どんな情報が含まれるべきかに関する限り、テキストの断片を見て、「誰がこの小さなテキストの断片を見ただけで何らかの回答が得られるのだろう」と言うようないくつかの例によって動かされるアイデアが含まれています。もし、仕事からの帰りに列車が転覆して誰かが怪我をしたとします。私が皆さんに「もし列車の乗客がその朝もっと早く仕事に来ていたら、どうだったと思いますか」と質問したとすると、皆さんは多分「ああ、多分その列車に乗ったでしょうね」と答えることでしょう。これはなぜでしょうか。それは皆さんが通常ある場所から戻るときには行ったときと同じ手段で戻るものだということをご存じだからなのです。もしCycがこの質問に対して誤った回答を出したら、私たちはシステムに正しい回答が出せるような規則を付加することでしょう。私たちは百科事典や小説や広告などから、何百何千というこうしたテキストの断片を見ることにより、知識ベースの向上を図り、現在の知識ベースを作り上げたのです。

知識ベースの中に何を含ませたいか、何を含ませたくないかはデータベースの中により良く、十分に表現されています。このデータベースには合衆国気象庁の予報情報や、Ronが先に触れていたような種類のデータベースが含まれています。ですから、もしそれが概要と中心的知識ベースの間に位置するこれらの明瞭な公理を、書くことによって、事実上含んでいるなら、それを知識ベースに「知識」として含む理由はありません。それゆえに、データベースの誤りのようなものがおきるわけで、それも私たちが仮想メモリを使用している際におきてているページの誤りと同じようなほとんど知識ベースの誤りです。

とにかく、原理は知識ベースが表現言語になんの関心も持たないような、ビルディングに焦点をあてたようなものを含んでいました。ですから、最初は平凡なフレーム・ベースのシステムから始めて、あ

とから皆さんの必要に応じて追加し、発展させていけばいいのです。重要なのは表現言語ではなく、構造でもなく、知識ベースなのです。目標は調査をする事でなく、ものを構築することなのです。そうした結果、私たちは過去１０年間、業績も少ししかあがらず、会議にもほとんど出席しなかったため、低姿勢をとってきたのです。

私たちはそれを働かせるのを目標にして、働くために出現させるのを目標にしてきたわけではないので、鴨猟だの徴兵猶予だの難問を遅らせるだの、時間や宇宙の因果関係や意図や信念や本質の心配をすることだのには、まるで使い道がなかったのです。まず、それらを処理することに私たちは'８０年代の半ばから終わりまでを費やしましたが、それが重要だったのです。

私たちがした唯一のイデオロギー的公約は、ものを機能させることで、それは同時に他の人の解決法を借用し、一体化してきたことを意味するのです。また、それは私たちが１つのすばらしい時間のモデルや、宇宙の完全なモデルでなく、全ての時をカバーするだけでなく、もっとも普通の一時的な表現と一時的な推論の使用法をカバーする、セットになったモデルや異なった仮定を作りだすこれらのモデル、一時的なモデルの集合で終わった事を示しているのです。

前にもお話しましたが、私たちは知識ベースを積み上げによって構築してきました。最近ではその知識ベースは十分完璧なものになり、私たちのシステムはトップダウンで動くようになっています。

私たちは自然言語の前面を研究してきており、特に人が知識を取り入れるように知識のエントリーを行うことを狙いとしております。英語の文をCycの表現言語に置き換えて関数計算のような形にすることをしています。

私たちは２年以内に、知識のエントリー全部が、Cycプロジェクトの終点近くまで進んでくれれば良いと思っています。私たちがCycから学んだ重要な事柄を提示することができます。最初のものは「フレームアンドスロットの言語では非常に不十分である」ということです。「否定」「分裂」「確信」「欲望」などの助動詞、高レベルの知識などを能率良く簡単に表現するには不十分なのです。ですから、これからは関数計算に非常に良く似たものを発展させたので、どうやったら効率的にできるかお話できると思います。これについては質問があると思います。

２番目の理解はMcArthyとHayesが長年にわたり述べてきた、認識論的な問題と発見的手法の問題との分離を重大なことと考えることです。どうやったら、システムが、知ったことを効果的に考えることができるでしょう。そのためには私たちはこの２番目の認識論的で発見的レベルの言語をただ認識論的な言語を理解する、知識のエンタラーにします。認識論的な言語は関数計算に非常に良く似ているのです。

次の１０年間のプロジェクトの基礎を作るため、層として私たちはCycを利用して行こうとしています。すなわち、次に真剣に行わなければならないことは機械学習であり、これは小さなものから学ぶと言うよりは大きな知識ベースから始める方が良いでしょう。自然言語の理解についてはこの知識をすべて利用して始めます。異なった国々、異なった年齢のグループ、異なった社会－経済グループなどに関するマルチ文化の知識の導入が必要となるでしょう。最終的には百科辞典、年鑑、子供の教科書などの内容のように共通の知識を導入することになるでしょう。

この辺でやめることにいたしましょう。おそらく、私たちがしなければならないことは休憩をする事であり、もし、ご質問がある方はここで私の辺りに集まって頂ければ幸いです。どうか、ここまで上がってきてご質問をなさって下さい。有り難うございました。

座長：

Lenat博士、Cycプロジェクトに関しまして情報をありがとうございました。ご質問、コメントはございませんか。Lenat博士が言われましたように、どうぞそばまで行って、ご質問をなさって下さい。

すばらしいプレゼンテーションを有り難うございました。

このセッションを締めくくるに当たり、このセッションは非常に情報に富んでおり、ヨーロッパ、アメリカ、日本における活動の状態をより良く理解する上で非常に有効であったと思います。最近の共有可能な知識資源の現状に関するすばらしいプレゼンテーションを頂きましたホスト・スピーカーの皆さんに感謝をいたしたいと存じます。講演者の皆さんにもう一度大きな拍手をお願いいたします。
有り難うございました。

6. パネル・ディスカッション

情報インフラストラクチャの構築と国際協力

6. パネル・ディスカッション：
情報インフラストラクチャの構築と国際協力

6.1 パネリスト

渕　一博	東京大学 工学部電子情報工学科 教授
Jacques Mathieu	Ministère de l'Industrie et du Commerce Extérieur
Cristian Rohrer	Professor, Institute for Computational Linguistcs Stutgart University
Peter M. D. Gray	Professor, Department of Computer Science University of Aberdeen
Su-Shing Chen	Program Director National Science Foundation
Brian Oakley	Director, Logica plc.

6.2 パネル・ディスカッション

渕：

それでは、２日間の会議の最後のセッションを始めたいと思います。これまでのセッションでいろいろな研究の進行の状況であるとか、あるいは、これまで進行したプロジェクトの内容であるとか、あるいはこれからの将来ビジョンに関していろいろなスピーカーに話して頂いて、このパネルのための素材もたくさん貯まっていると思います。パネルとしましては、いろいろな国の方にお願いしました。政府関係で先進的な研究をプロモートしている方々もいらっしゃいますし、大学で基礎を含めた研究を進めておられる方々もいらっしゃる。そういう方にパネルに上がって頂いて、それぞれの立場から、最初に１０分間を見当にしてプレゼンテーションをお願いしてあります。その後、皆様にお願いですけれどもフロアの方からどんどんいろいろな意見とかご質問を寄せて、賑やかなパネルになることを期待してお

りますので、よろしくお願いいたします。

テーマはまとめみたいなことでありまして、「情報インフラストラクチャの構築」、その中でもキーポイントであると思われる「大規模知識ベースの構築とシェアリング」、これをどういうビジョンを持って進めるか。あるいは、どういう国際的な協力のもとに進めるのがいいか。各国でいろいろそういうことを目指した試みというのは、進行していたり、あるいは、準備中であったりすると思いますので、そういうことも差し支えない範囲でご説明頂いて、皆様のこれからの研究の参考になればと思っております。それでは、順番としては、こちらに並んで頂いている順番でいきたいと思いますが、最初のスピーカーはフランスの、先程ご紹介頂きましたけれども、大体、日本で言いますと「通産省の産業政策局の次長さん」という肩書きであります、Mathieuさんにお願いしたいと思います。

Mathieu：

議長、有り難うございます。まず、いろいろトラブルがございまして、大急ぎで新しく作らなければならなかったのでスライドの出来があまりよくないことをお詫びします。

時間が限られているため、あまり長いお話はできませんので、まず申しあげておきますが、前の講演者のお話で、技術的見地から見て主たるポイントは何か、巨大知識ベースに関る主要なポイントは何か、次世代の情報化はどうなるかということは明らかになりました。ですから、このような点については繰り返しませんが、１点だけ申し上げておきたいのは、このような大規模な技術上の問題点を解決するためには、国家間および企業間の大規模な共同研究が必要だということです。なぜなら、巨大知識ベース構築は非常に困難であり、また多言語が関わってくるからです。それから２つの点が議論の的になっています。１つはどうやって情報にアクセスするかであり、もう１点はどうやって情報を利用するかです。まず初めに申しあげたいのはこういうことです。もし国際的な共同研究をしたいのであれば、１つの重要なポイントは、知識に関する著作権はどういうものになるかをはっきりさせることだということです。その知識は無料なのかそれとも誰かが決めるのか、どうすればその知識にアクセスできるのか、知識の価格はどのくらいなのか。著作権というものがあるように、知識の著作権はどのようなものになるのか。Susan Armstrong氏もお話になったように、これは非常に重要な問題であり、私たちは多大な困難に直面することになります。

そこで、まず第１にお話するのは、フランスやその他の国でこの分野の研究機関はどうなるのか、現状はどうなのかということです。フランスでは研究資金の出所は２種類あり、これが非常に重要な意味を持っています。一つはフランス政府であり、もうひとつは欧州連合政府です。ですから研究目標に関して、異なる省庁、異なる政府の資金が関わってきます。

基礎研究についてお話すると、基礎研究といえば研究所の話になります。例をあげますと、フランスには現在この分野で活動している主な研究機関が３つあります。パリには７００から１０００人の研究者がいます。グルノーブル大学のGETA-IMAGは、ここにおいでのBoitet博士のもとに５００人の研究者

を抱えたこれも資金力のある研究所です。マルセイユ大学、ここからも本日この会議に出席している方々がおられます。これらに共通する一番のポイントはそれが国家資金によるものであり、研究結果は全く自由に利用でき、誰でも自由にこの研究にアクセスできるということです。公的な研究ですから、この点についてはなんら問題がありません。そして資金の１０パーセントはヨーロッパ連合、ＥＣから出ており、これは様々なプログラムに関する契約に対するものです。

このように、基礎研究は主としてプログラムに関するものであり、研究目標を定めるのは研究者の側であり、政府ではありません。ですからさまざまな分野の研究が可能であり、さまざまな種類のプログラムを研究することができ、研究者は将来を考えて最善と思われる方向に研究改善していくことができます。この分野では、各種のプログラムをご存じの研究者が沢山いらっしゃるので、本当にあまり長くお話はいたしませんが、ひとつだけ申し上げておきます。また、グルノーブル大学やマルセイユ大学からもそれぞれプレゼンテーションがなされると思います。大事なことは、すべてが無料であり国際協力が非常に容易であることです。これが基礎研究にとって大切なことだからです。

しかし、別の観点に立ってみますと、これは競争以前の研究です。競争以前の研究というのは、１０年後に最初の成果を得ることを目標としているような研究のことです。たとえばＥＣのESPRITプロジェクトの目標ですが、これについては明朝お話することになっていますが、これは知識ソフトウェアとソフトウェア・ファクトリーに関するものであり、数多くのプロジェクトがあります。ここでの考え方は、このような研究をする研究所が必要だということです。例えばフランスにはINRIAがあり、本日もINRIA代表の教授がお一人出席されていますが、企業も参加しており、資金は主としてヨーロッパ政府もしくはESPRITから出ています。この契約は各国政府が定める契約です。その考え方は、各国政府がＥＣに資金を出し、ＥＣがこの研究の次なる目標は何か、ＥＣにとって興味ある主題は何かを決めるというものです。そして、このプロジェクト終了の時点では、５年ほどかかり、５ないし１０の契約者が関わることになるでしょうが、結果を商品化することが問題となるでしょう。このように国内的にも国際的にも、資金は主としてヨーロッパ連合から出るので、もし共同研究による知識に関する規定を知りたければ、ヨーロッパ連合政府に聞かなければならないでしょう。

第３の種類のプロジェクトはイニシャル・プロジェクトです。このほうが容易なもので、目標は必ず３年から５年後には企業が販売できる製品を作り上げることです。これは先程のように１０年もかからない短期のもので、フランスは、ヨーロッパには２種類の資金源があります。１つは国家資金で、私共は特別のプログラムを作り毎年テーマを定め、次のプロジェクトの目標を定めます。これが関係省庁によって指定されます。政府が企業のアドバイスによって指定します。第２のポイントは、私共が呼ぶところのEURIKAプログラムによるものです。EURIKAは国際的プログラムであり、このプログラムにはヨーロッパその他の国々、たとえばロシア、カナダ、スカンジナビア諸国、スイスなどが参加しています。ルールはＥＣとは違います。ここでは、いくつかの企業が企業共同体を作り、その企業共同体が各国政府に提案を行い、各国政府が自国の会社もしくは研究所に出資します。これでフランスのあるプロジェクトでは資金の１０％、スペインでは６０％、ベルギーでは２０％というふうになります。このように

それぞれの国がプロジェクトに対する資金を定め、製品化を目指します。そしてもう１つのポイントは、その企業が独自にまた企業間で提携協定を結び、プロジェクト終了段階および新製品についてロイヤルティーをどうするかを決めるというものです。ですから非常に分かりやすく簡単です。今日関心のもたれる分野で、３ないし４の主要なプロジェクトがあります。例をあげれば、文法、辞書、機械翻訳、メモリーと呼ばれる、コーパス・メモリーに関するプロジェクトなどがあります。このようにここではポイントを定めることがルールになります。各国がそれぞれにどのプロジェクトに資金を与えるかを決定します。

ですから結論として申し上げると、新しい情報時代には、先程申しあげたように、多言語に関る新しい問題があります。また著作権と知識が第２の問題となるでしょう。なぜならどのプロジェクトにおいても、どうしたらアクセスできるか、どうやって購入するか、どうやって知識を売るかについての情報が必要とされるからです。

結びの言葉になりますが、この４日間の会議で、この問題について皆さんと討論する機会を得たことをたいへん喜ばしく存じております。私たちの考えているものは最良の問題であり、またどうしたら共同研究できるかを解決することは重大な問題です。なぜなら、それがこの種のプロジェクトにとって非常に重要なことだからです。有り難うございました。

渕：

Merci.フランスの状況についてお話頂いたわけですが、いくつかのポイントで皆様の方もさらにお聞きしたいこともあるかと思いますが、後ほどまとめてお聞きしたいと思います。

それでは次のスピーカーですが、今度はドイツのStuttgart大学のComputational Linguistics研究所の教授をなさっておられます、Christian Rohrerさんにお願いしたいと思います。

Rohrer：

皆様、まず第１に、本会議に参加できますことを大変光栄に存じていることを申し上げたいと思います。これまで日本の皆さんは非常に重要なイニシアチブを取ってこられました。またこのたび本会議を開催されたことも、まことに時宜を得たことであります。本会議は将来に向けて正しい決定をするために役立ってくれると考えます。私の専門は言語学ですので、言語学的な見地からお話するつもりですが、その前に、この主題の重要性について少しコメントさせていただきたいと思います。私の考えでは、将来、情報工学の進歩はハードウェアではなくソフトウェアに依存するようになります。そして、政治的には、私たちが１つの研究分野として将来の研究に必要な資金を求める際には、このことを根拠とすることができると考えます。今日、テキスト処理のために使われるコンピュータの数は、データ処理目的のものを既に上回っています。ヨーロッパの立場から見ると、言語は特別な機会を与えてくれます。なぜならヨーロッパには多言語環境があり、地域市場も大きく、またすぐれた研究界もあるからです。

将来の展望と見通しについて述べれば、今日、研究者の間では非常に大規模な知識ベースが重要だと

いう認識が広がっています。大規模知識ベースは理論および応用研究に必須の前提条件だと考えられています。もし皆さんがコーディングやACLのような会議の紀要をご覧になれば、今日、コーパス・ベースの研究がほんの５年前に比べてさえはるかに目立っていることに気がつかれるでしょう。また今や数多くのプロジェクトが始まろうとしています。第２に、産業界も認識し始めています。自然言語製品や自然言語に関わるソフトウェアの潜在的な市場は巨大なものであるということです。言語処理、すなわち文書作成、保存、検索、翻訳などは、ＥＣにおいて農業と同じくらい大規模なビジネスとして期待できます。しかし、本日の午後の講演でZampolli教授も指摘されたように、知識ベースの欠如が障害となって自然言語応用の開発が妨げられているのです。これはヨーロッパにおいては重要な問題です。ヨーロッパでは、自然言語製品はごく小規模もしくは中規模の会社で作られているのが一般的です。このような会社は資源、すなわち、必要投資のための財源を持っていません。第３に、出資機関や政府機関のレベルでも認識し始めています。ほんの幾つか例をあげれば、ＥＣの計画には、EAGLESやRELATORがあり、そのほか、本日Zampolli教授が、また昨日Susan Armstrong氏が挙げられた計画も全てここに含めていただくことができます。

この分野の潮流はどうなっているのでしょうか。私は言語学者であると申し上げたので、言語学的な見地から見ますと、主な流れはコーパスから知識を自動的もしくは少なくとも半自動に生成することに向かっています。主な研究テーマとしては、ざっとあげると、辞書見出しの半自動構築、文例による単一言語もしくは多言語機械翻訳、これも私はここに含まれると思います。意味判別、文法生成、および統計的情報による文法の多様化もあります。これらのテーマはすべて考えられるというだけであり、もし本当にこのような分野を研究したいのであれば、統計的に意味のある成果を得るためには非常に大きな知識ベースが必要となります。

さてここお見せするのは、このスライドは少々違ったものになっていますが、これも私が現時点で非常に重要であると考えるものです。これはオフィスオートメーションに関係する自然言語処理なのです。私は年度休暇をスタンフォードのゼロックス・パークで過ごしましたが、このスライドに現れているのはもちろんその影響です。私は、自然言語処理を、文書の生成と処理の観点から考えることができると思います。文書はオフィスワークの要であり、そしてオフィスは経済成長の要です。私が見てきたデータによると、オフィス部門の上昇に伴って労働者総数が６０％増加しています。そしてこの数十年間、オフィス労働者の生産性は全く向上していないのです。これをどうにかする必要があります。そして、すでに申し上げたようにオフィス作業の要は、書く、読む、保存するなどといったことなのです。従って、自然言語は文書処理における重要な要素となるという結論が得られます。唯一のものというわけではありませんが、実現化技術の１つとなるでしょう。

さて、今、私たちは何が言えるでしょうか、この会議の重要テーマの１つであるということで、この分野における国際協力について私達は何をいうことができるでしょうか。私の個人的意見では、自然言語資源は国家によって構築され維持されるべきです。ただし、データのコード化やデータの交換に関する共通標準についてはもちろん国際的合意が必要です。そして最も重要なことは、研究や開発のために、

資源は世界的に利用できなければならないということです。Zampolli教授がEAGLEプロジェクトについてお話になりましたが、私がEAGLEプロジェクト運営委員会の長を務めておりますことから、EAGLEについて少しお話させていただきます。少なくともヨーロッパのレベルでEAGLEがどんな役に立つのか、国際協力の例としてお話したいと思います。このプロジェクトの目的は大規模言語資源の開発、利用、評価のための標準の作成を促進することです。

EAGLEは次のような機構になっています。これは主な産業および学術センターの勧告、そしてヨーロッパ委員会の言語技術戦略委員会の勧告によってできたものです。全部で３つのセクターからなっています。すなわち産業界、学術界、政府です。３０を超す研究所、企業組織、専門団体、ならびにＥＣ内のさまざまなネットワークがこの研究に協力しています。そしてEAGLEの目的は、現在のヨーロッパにおける研究結果を蓄積し、専門家のネットワークを利用することによって、言語工学の個別の分野について幅広い同意を得た公的仕様およびガイドラインを作成することにあります。そして私共はヨーロッパにおける各研究開発プロジェクトを補完し、これらの標準の採用を促進して、その成果を国内および国際的な標準化計画に活かしたいと考えています。

さて、このスライドを見て思ったのですが、何と言うか、あまりにも楽観的だといわれるかもしれません。まだほかにもいくつか問題があります。何が問題なのでしょうか。これもまた個人的意見になりますが、少なくともヨーロッパのレベルでは、私共は標準とかフォーマットばかりを問題にしていて、資源を構築していないのではないかと私には時々感じられます。例えば、Zampolli教授のお話を聞かれたと思いますが、教授はNercやEAGLESのようなプロジェクトについて、ここに３００人の研究者がいる、そこに５００人の研究者がいるということを言っておられましたが、その全員が委員会から給与を受けているわけではないし、資金は非常に少なく、それを分散してしまうため、非常に乏しいものになってしまうのです。したがって、私共は相当額の資金が得られない限り、資源を手に入れることはできない、少なくとも、辞書や文法を手に入れることはできないのです。コーパスなら得られるかもしれませんが、それはより安価だからです。

もう１つの問題ですが、組織が一定の資源を持つと、それを共有するよりも金もうけに利用しようとする傾向があります。そうすると法的な問題になりますが、これについてはすでにMathieu氏が指摘されました。特に辞書の発行者は、著作権問題が未解決のため、資源を提供することを恐れています。日本の電子関係の研究組織でさえ、著作権問題が未解決のために、コーパス提供には問題があるとうかがっています。

また心配なのは、ガットの交渉を見ていてのことですが、法律家にまかせておくと、私たちの法的問題を解決するのに最終ラウンドまでかかるのではないかということです。これもどうにかしなければなりません。有り難うございました。

渕：

Danke Shoen. Rohrer先生を中心にいろいろ進んでいる研究の話と、最後に問題点のご指摘まで頂き

まして、どうも有り難うございました。

それでは３番目ですが、英国のKing's College、Abereen大学のComputer Scienceの教授をしておられます、Peter Grayさんにお願いしたいと思います。

Gray：

渕博士、有り難うございます。本会議に参加できますことは、長年に亘ってデータベースと知識ベースの歴史に関わってきた私にとりまして、誠に喜ばしいことであります。日本に来られたことを大変うれしく存じております。皆さんは既にフランスとヨーロッパの調査研究についてお聞きになっておられますから、私はイギリスにおける研究について簡単にお話したいと思います。その次に、昨日と本日の講演に耳を傾けている間に私が特に強い印象を受けたいくつかの点につきましてお話することにします。特に再利用の点と、ネットワークの点に重点をおいてみたいと思います。

研究資金について背景を手短かに述べますと、イギリスではちょうど新しいシステムに移りつつあるところです。現在は、内閣の大臣の１人が科学技術局の運営を担当しているのですが、これを再編成し、もっと技術的展望をもって運営できるようにしようとしています。私としてはこの展望の中に大規模知識ベースが含まれてくれればいいと期待していますが、結果を待ちましょう。特に、富の創生のために産業界に技術を移行しようとする動きがありますが、これはイギリス人が余り得意ではないと思われている問題の最たるものと言っていいかとおもいます。しかし、私共にはたくさんの強力な研究グループがありますし、データベースの分野ではことさらです。私共には非常に優れた通信およびインターネットのインフラストラクチャーがあり、微生物学ネットワークの１つが現在ケンブリッジに移されつつあるのも、１つはこの理由によっています。また最近、先端データベースと大規模知識ベース研究のためのコーディネータをおくことが合意されました。しかし、この再編成が進んでいる最中も、私共はまだプログラムの資金を待っているのです。

知識の再利用について一言いわせて頂きますと、私は昨日横井博士の言われた、「大規模知識ベースを柔軟な構造を持った非常に大型のソフトウェアとして扱いたい」というお話、これは刊行物から引用したのですが、最大限の再利用を行なうために、というお話に関心を持っております。私自身の経験から言っても、たしかにこのようなものが必要なのです。アバーディーンで私共が開発してきたアーキテクチャーでは、データベースの世界で知られている、私共が意味データモデルと呼ぶもののもとに知識を蓄積します。これは、科学的な、高度に構造化されたデータを対象としたものであり、それを大規模なオブジェクト指向データベースとして処理します。また私は、オブジェクト指向のプログラミング関係の人が「動作知識」と呼ぶものは、論理型データベースで行なわれる古典的な方法のように手続きとして保持するのではなく、特にPrologをソースとして使用することにより、宣言的に保持するべきであると考えます。その理由は、再利用するためには知識を再編成し変換する必要があると考えるからです。私共のプロジェクトではこのような経験をし、このようにして拘束サティスファイヤと共にPrologを利用することに大きな成功を納めています。しかし、これは大規模知識ベースにおいてははるかに汎用性

があると思われます。もし知識を再利用したいのであれば、変換したり再編成したりできるように知識を構造化する必要があります。パイプに注ぎこんでそれを集めるというようなものではありません。どうやって利用するかは極めて扱いにくい、複雑な問題です。

第２番目のポイントとして、渕博士が昨日の基調講演でお話になったことを補足させていただきたいと思います。実際には言語コーパスに関するお話であったと思いますが、それはもっと広く一般に当てはまることだと申しあげたいのです。大量に蓄積された言語情報を変換するには、研究だけでなく、非常に基礎的な骨の折れる作業が必要となります。そしてオブジェクト指向のプログラミングの世界には、これは私にとって親しい分野ですが、ここには知識再利用のおもしろい例があります。スモールトークＶクラスのライブラリーです。これは産業的には大きな成功を納めました。これは進化するのにつれて、数回は書き直されています。進化にかかった時間はおよそ１０年程だったかとおもいますが、今では採算に乗り始めています。しかし、再利用が達成できたのはこの非常に面倒な作業を行ったからこそなのです。この種の努力が必要であることはオブジェクト指向のプログラミングの世界では周知のことです。

したがって、私は渕博士や、これまでの講演者の方々がおっしゃったことを支持いたします。言語コーパスの分野だけではなく、その他の分野でも、再利用可能な知識ソースを本当に開発しようとするならば、何らかの国家資金が必要であるということです。これは非常に困難な仕事なのです。

それではデータベース技術の話に移りたいと思います。これは私自身の専門分野なのですが、ここでは大規模知識ベースを本当に成功させようとするなら、共有・分散データベースから学ぶべきことがあるという事実に驚かされてきました。理由はいくつもあります。大規模知識ベースは複数のソースから得られるであろうこと、それらが非常に大きなものになるであろうことは明らかです。これは規模の問題といえます。本日の午後早くに行われたいくつかの講演の中で数字がでてきたのを聞いた覚えがありますが、データベースは伝統的に非常に大規模ものを取り扱ってきたということです。従って、そこには技術が関わってきます。ユーザーの数も膨大になるでしょう。そのため、ユーザーはただデータベースのコピーを得ればよいというものではなく、それに選択的にアクセスしなければならないでしょう。ソースは知識ベースの更新を続けます。これらはあたりまえのことに思われますが、これに気づいたとき大規模データベースというのはCD-ROMで買って家に持ち帰るというようなものではないことが分かるのです。

またこれは、データベース関係者がデータベーススキーマと設計、スキーマと統合の分野で問題とする点がすべて関わってくることを意味します。これに関する論文発表も既にいくつか聞きました。Brachman氏の研究は、このような分野で何をする必要があるかを示した一例です。明日のワークショップでは、John Mylopoulos氏のグループによる論文が発表されることになっています。彼らもまたデータベース実現化技術を利用しています。この理由から私は大規模データベース(VLKB)で成功するには、データベース技術、知識ベースの理解、並びにネットワークの利用を統合することが必要だと私が考えるのです。この３つを組み合わせることが必要なのです。

分子生物学データの研究のためにごく最近できた専用のネットワークがあります。これもまた私がE

Cのあるグループと共同研究をしている専門分野です。タンパク質の分子構造に関するデータは事実、非常に複雑なものです。これはマークアップできるテキストのようなものではありません。計算に使うデータ値なのです。またこれは3次元構造であり、人間の遺伝子やDNA塩基配列とは同列のものではありません。このこともまた重要です。しかしこの今も結晶学者やNMRの研究者により発見されるタンパク質分子構造の数は急速に増大しています。またこのデータを適切に利用するためにはきわめて複雑な計算が必要です。これは簡単なことではありません。

ソースとしては、ブルックヘブンやEMBLのものが有名ですが、ケンブリッジに新しい機関として，EBIができようとしており、これもまた歴史のおもしろい一断片なのですが、元はテープにコピーを取って送っていたものが、今ではネットワークにより選択的に入手できるものと期待されているのです。すでに様々な人が指摘したデータ検証(validation)の研究は、根本的に重要です。データベースの仕事をすると、データがいかにひどいものであるかが分かります。このように科学者の提供するデータであってもそれを検証するにはかなりの努力が必要です。ですからそれが人々がお金を払って最新バージョンを見たがる理由なのです。

さて、一番おもしろいと思われる方向は、生データにアクセスするだけではなく、ネットワークを通じて興味深いグラフィック・フロントエンドを提供される機会が増えていることです。ファイルトランスファーを利用することにより、ネットワークを通じてこのグラフィック・フロントエンド・プログラムを取り出し、ＰＣなり何なりにインストールして、求める知識ベースにアクセスを開始します。私が引用する例はNCBIではアントレと呼ばれており、実際にはややテキスト・ベースのものでありますが、非常に広範な仕事をすることができます。また、これは大規模知識ベースアクセスの発展の方向を示すパラダイムではないかと私は考えます。いろいろな企業がこのような装置を用意するようになるでしょう。

CMPネットがこれほどの進歩を遂げることができた理由は、研究者がデータ理解と共有機械可読スキーマ開発の長い伝統を持っていたことであると私は考えます。しかしそれでさえ、結晶学者の間では標準に関する論争の原因となっています。従って、これはSusan Hockey氏が述べられたテキスト・コーパスとはまったく異なる分野です。むしろ、より複雑なものであると言えるでしょう。しかし、科学者が相当期間研究してきたので、かなり進歩してきました。

面白いことに、私が考えついたのは、このスライドを作ったのは他の講演者の方々が何をお話になるか知る前だったのですが、ネットワークの経済性の問題があるという同じ様な考え方です。また私共は国際協力に関する問題について話をするように依頼されました。私の考えでは、ある種の料金制度が必要です。これは、今は私の研究仲間の皆さんには余り人気が無いかもしれません。しかし、低額なものになることを期待します。しかし、私たちは電話料金の請求書には慣れていますし、気に入らないまでも、受け入れています。ケーブルテレビは別個の料金がかかります。知識ベースの提供者たちにも、即ち、ネットワークに出てきて興味深い知識ベースを提供してくれる人々にも資金が必要なのだと私は考えます。

私が付加価値提供者と呼ぶものを鼓舞する必要もあります。興味深いグラフィック・インターフェース、専門的なエキストラ・データ、ネットワークを通じて入ってくる色々なものを提供してくれる人々のことです。この人々が何らかの見返りを得られるようにしなければなりません。その答は、大きなコミュニティーが少額の料金を支払うようになることだと思います。現時点での問題は、このような会社のほとんどが、複雑なシステムを非常に大きな企業に高額で販売することによってしか採算をとることができないことです。

もし私たちが本当にこの分野全体でのブレークスルーを目指すなら、非常に大きなマーケットを存在させ、そのマーケットで多くの人々が取り引きできるようなシステムを構築することが必要となります。ごく未来指向的になるなら、それが新しい情報社会の形となるだろうと申し上げるでしょう。ただし、それが２０１０年あたりに実現するかどうかは私には分かりません。

最後に、警告を一言、Rohrer教授がおっしゃったことを繰り返すことになりますが、大きな問題は弁護士です。既にアメリカにおいては憶測的な、非常に高額の訴訟がおきています。これについては新聞でお読みなっていることと思います。実際のヒトDNAが特許の対象になるかどうかについては議論のあるところです。新薬の発明でなく、自然界に実在するものなのです。調査研究を行うビジネスや調査データの所有権のことが引き合いに出されています。よく注意していないと、弁護士が情報全体を拘束してしまい、本当に前に進むことができなくなってしまいます。この非常に大きな問題を乗り越えて次の段階に進むことができれば、人々はマーケットが存在すること、そこで実際に取り引きができると思えるようになり、そしてこのような知識を増大させることができるようになるでしょう。有り難うございました。

渕：

Thank you very much.今日のお昼の、食事をしながらの打ち合わせの時に「今回の企画は大変よかったけれども、全体的に見ると自然言語との関連の話の比率が大きすぎる様な気がした」というのは、ある意味で的確な指摘をされたんですが、データベース・ネットワークとのつながりについて、実例を挙げながら説明して頂きました。それから問題点の指摘も頂き、ありがとうございました。

それでは次に、アメリカ合衆国、ナショナル・サイエンス・ファウンデーション(NSF)の、知識モデル及び認知システムプログラムのプログラムディレクターをされていますSu-Shing Chenさんにお話を伺いたいと思います。

Chen：

有り難うございます、座長。私は米国のHPCC計画を簡単にご説明し、私共がHPCCを国際情報基盤にどのように適用しようとしているかをお話しようと思います。HPCCは約２年前に発足したきわめて大きなプロジェクトです。初年度の予算は６億５千万ドルで、現在３年目に入っています。様々な機関が参加しており、事実上国家的研究活動といってよく、ほとんどの政府機関が関わっています。元々は４つの

部門がありました。第１は高機能コンピュータシステム、第２は先端ソフトウェア技術およびアルゴリズム、第３は国家研究教育ネットワーク、第４は基礎研究および人的資源です。知識ベースと自然言語処理については、そのプロジェクトの多くが私共とNSLの支援を受けていますが、ソフトウェアの部門に入っています。もちろんソフトウェア再利用など様々な事柄も検討しています。

最後にもう１つ、今年追加されるもので、今年が３年目であることは申しあげましたが、新しい部門を追加しました。情報基盤技術および応用、ITAです。そこでITAについてご説明させていただきます。ITAはHPCCの新しい部門で、いくつかのテーマがあります。その第１は、情報基盤サービス、第２はシステムおよび支援環境、第３はインテリジェント・ヒューマン・インターフェース、第４は国家的問題に対する解決です。このようにHPCCには重大な課題がありますが、民生用基盤、ヘルスケアおよびデジタルライブラリーなどの国家的課題にも注目していきたいと考えています。そのため今年度はデジタルライブラリー計画を開始します。従って基本的な考え方は、ITAによりHPCCの焦点をテレフロップからメガユーザーに移すということです。私共はHPCCコンピュータ通信の全てが国内の全ユーザーの手に届くようにしたいと考えています。

さて、国家情報基盤は皆さんのお話のとおり、通信ネットワークに基礎を置いています。私共がやろうとしているのは、大学・企業・病院をひとつのマルチメディア環境の中に結び付けることです。さて、現在、いわゆるNREN、国家研究教育ネットワークと国家情報基盤との関係は次のようになっています。現在NRENの中継線スピードはT-3です。国内で6,000以上また全世界で12,000以上のネットワークとつながっています。 従って、普及率は指数級数的に上昇しています。私共は、次世代ネットワークの研究プロジェクトも行なっています。つまりギガビット・ネットワークです。時間が無いのでこれには立ち入りません。しかし将来、今後５年間の内には、ギガバイトの中継線スピードが得られると期待しています。私の個人的見解ではこれには政府資金は必要ないと考えています。各企業、すなわちAT&T、MCI、スプリングなどがこのギガビット中継線スピードを達成しようとするはずだからです。また全国的、世界的な連結性があります。この事については後ほどお話します。ということでNRENは５年のうちには研究、教育、社会的応用のための新しい情報サービスを支援できるようになるでしょう。ということで国際間協力のお話をさせていただきたいと思います。さきほども申しあげたとおり、国際的連結性、共同研究と政府の役割についてひとつずつお話します。

国際的連結性に関して、全世界の通信の相互依存性を強調しなくてはなりません。例えば、アメリカと日本は互いに連結されています。米国内にはAT&Tの地上線があります。AT&Tの海底光ファイバーケーブルもあります。これは中間で日本からの海底光ファイバーケーブルと繋がっています。これによってNTTの地上線と繋がっていることになります。国際協力は国際的連結性のために不可欠であると考えます。さてこれは一例にすぎません。他の例もお見せしましょう。いかに多くの企業が関わっているかお分かりになると思います。焦点が合っているかどうか分かりませんので、私がお話しましょう。まず第１はこれ、いわゆるWersoceと呼ばれるものです。参加企業はAT&T、シンガポール・テレコム、KDDジャパン、それから、たしかテストラ・オーストラリア、カナダのユニテル、韓国テレコムです。これは環

太平洋諸国をつなぐ共同事業です。２つ目の共同事業はNewcomです。参加企業はMCIとブリティッシュ・テレコムで、米国とイギリスをつなぐものです。３番目はGBPNパートナーシップで、スプリング、カナダのユニテル、もう１つのカナダ企業、オランダPTTテレコム、日本ではIDC、よく読めませんので間違いがありましたら御訂正ください、オーストラリアもあります、香港テレコム、マーキュリー・コミュニケーション、これはイギリスだと思います、それとスウェーデン・テレコムです。このように国際的連結性がいかに国際協力に依存しているかお分かりいただきたいと思います。

さて共同研究についてもう少しお話させていただきます。ちょっとばかなことを言ってみましょう。共同研究なら、例えば国際会議など、居ながらにテレビ会議をすればよいのであって、時差ボケになる必要も、旅費を払ってここにやってくる必要もない、と冗談はさておき、現在ネットワークに関して様々な共同研究が考えられ、それができればおおいに旅費も節約でき、リアルタイムで、オンラインで、研究の成果をあげることができるのです。また、ネットワークなら費用もそれほどかかりません。皆様ご存知のとおり、インターネットはほとんど無料です。

最後に政府の役割ですが、私がお話できることは多くありませんが、NSFには様々なプロジェクトの国際部門があります。予算は年間およそ2500万ドルで、世界各国との共同研究を支援しています。以上、有り難うございました。

渕：

Thank you very much. HPCCという、日本でも大変興味を持たれているというか、いろいろ知りたいテーマについてもお話し頂いて、その中でのごく最近の動き、ご紹介頂きました。いろいろな点があるかと思いますが、Chenさんは明日、さっきちょっと言われたデジタルライブラリー、電子図書館の計画のために、飛行機でお帰りにならなければいけないということで、先ほどのテレビ会議の冗談もあったんではないかと思いますが。どうもありがとうございました。

最後にBrian Oakleyさんにお願いしたいんですが、現在の肩書きは英国のロジカというソフトウェア会社、非常に大きなソフトウェア会社のディレクターなんですけども、実は十数年前に最初にイギリスでアルベー委員会の計画、アルベー計画というものができた時に、英国の貿易産業省でアルベー・ディレクトリーとその研究計画の総指揮官をなさっておられました。その後、企業の方でも活躍されているということであります。では、Oakleyさん、お願いします。

Oakley：

どうも有り難うございます、渕先生。またここでお話できること、そして渕先生と同じ壇上でまたお話できることをうれしく思っております。年寄りの特権のひとつは、そうしたい時に人よりゆっくりそこらを見て回れることでありまして、今回はこの偉大なお国の歴史と文化を少々見せていただくことができました。実に素晴らしいと申しあげるしかありません。「我が国が長い歴史を持った古い国だ」と思っている私にとって、私共がまだ暗黒時代と呼ばれる時代を生きていた頃、皆さんのお国が高度の文

化を持った文明国だったということは新鮮な発見です。

ヨーロッパ委員会(European Commission)の好意により、本会議に参加させていただきましたことをたいへん光栄に存じます。私のバックグラウンドをご説明申しあげましょう。私がコンピュータの分野で研究を始めたのは４０年以上も昔のことだとおもいます。政府研究機関でリアルタイム・コンピュータ技術に関する仕事をしてまいりました。Alan Clearingと同じ機関で働いていたと申しあげることができるのは喜ばしいことです。短い期間ではありましたが、しかし彼の講演を聞いたことがあります。彼の友人だったと申しあげられればもっとよいのにと思うのですが、それでは誤解を招くかもしれません。どちらにせよ真実とはいえないでしょう。

さて、２０年の研究生活の後、私は貿易産業省の研究計画の長官になりました。これは渕先生のおっしゃったとおり、イギリスでの通産省にあたる機関です。しばらく科学技術研究協議会の会長を勤めました。これが文部省や科学技術庁にあたると言えるのかどうか、私には分かりません。その後、私は光栄にもALVE計画の長官となりました。この計画は日本の第五世代計画に直接負うところがあると申しあげなければなりません。ついでながら、渕先生はさきほど随分とこの（第五世代）計画の成功を過小評価なさっていたように思います。私はこの計画で論理方針を慎重に追究したことは、たいへん果敢であり正しい決断であったと思っています。効率の問題は、私が長いコンピュータ生活で発見したところでは重要なものではありません。ハードウェアが追いついて来るので大丈夫です。たとえばこれが産業界の話なら、商売が成り立ってきたのは、数年前に顧客にシステムを売って、それが完成し、そしてもっと高い計算力が必要だということが分かった時には、新しいコンピュータは同じコストでさらに優れていますと顧客を説得することができたからです。しかしながら、この計画から訓練された人的資源が得られたという点からも、日本の技術研究全体にとってこの計画がいかに重要なものであったかはこの先明らかになってくるものと思います。さらにこの計画によって日本の研究の名声が世界的に認められたことには全く疑いの余地がありません。

並列処理の研究に疑問を抱く人もいるかもしれませんが、そうだとしても、それがどのようなものか私達には分かりません。イギリスでは、ICLが、もちろん現在富士通が所有していますが、大規模並列処理システムを発表したばかりで、皆さんにとって喜ばしいことだと思いますが、これは大規模データベースに目標を絞ったものです。機械は嬉しいことに、ALVE計画のもとで始まった研究からできあがったもので、この研究には特にImperial Collegeとマンチェスター大学が参加していました。その後ヨーロッパ委員会のESPRIT計画から資金を得て研究を続け、この計画にはフランスのINRIAの研究などが参加してきました。ミュンヘンにある、ヨーロッパコンピュータ産業研究センターも関わっていました。そして第五世代の研究も関係していたことも確かです。 なぜなら研究の途上でその関係の大学やセンターの専門家がこのICOTにやってきてはしばらくの間働いていたからです。

さて、渕先生のご紹介の通り私はLogicaに入りました。これは世界的な活動をしているソフトウェア会社で、この日本も例外ではなく、私共は今まさに東京オフィスを作ろうとしているところです。しかし実のところ政府での勤めを退いた後、私は主としてヨーロッパ委員会の研究計画に関わっており、計

画の評価を行なったり、運営を行なったりしてきておりまして、そのような理由で私はこの会議に参加させていただいているわけです。

私はヨーロッパ共同体、というより今やヨーロッパ連合というべきですが、それを代表してまいっているわけですが、その立場でのお話はできないことを申しあげておきます。私がお話することは私自身の考え方です。私はヨーロッパからこの会議に参加している方々と協力して、もちろん委員会を通してですが、ヨーロッパ連合に対して、日本との協力によりヨーロッパで行なわれている同様の計画についての助言を行なうようにしたいと考えています。ところで、ヨーロッパ連合という耳慣れない名称について少しご説明申しあげたほうがよいでしょう。私共はしょっちゅう名前を変えています。初めはヨーロピアン・コミュニティーズいう複数形だったのです。ローマ条約後ヨーロピアン・コミュニティーになり、ほんの数週間前マーストリヒト条約が発効して、私共はヨーロッパ連合となったのです。これもヨーロッパの結束へ向かう荘厳なる歩みの一歩なのです。長い時間がかかります。

ある言葉を思い出します。確か中国の傑出した指導者周恩来の言葉だったと思いますが、ミッテラン大統領との会談でのことだったと思います。ミッテラン大統領が２００年前のフランス革命の影響、インパクトについてどう思うかとたずねたのです。すると周恩来はしばらく考えてこう言いました。「そうですね、まだそれを言うには早すぎますね」。日本の方々は中国人に比べて短気なのではないかとは思いますが、ヨーロッパで私共が互いの違いを乗り越えるまで待っていていただきたいのです。

さて、私はこの大規模データベース計画の提案を２つの理由からおおいに歓迎いたします。まず第１に、私にとって自然言語処理は最後のフロンティアであるからです。コンピュータ技術にとって最後の知的チャレンジなのです。これを克服したと本当に言えた時、はじめてバイオテクノロジーだのなんだのといった暇つぶしに取り掛かることができるのです。

幸いなことに、少なくとも私にとってはですが、そこに至るまでまだ少し時間があると考えます。これはあまり知られていないことだと思うのですが、１９５５年にAlan Touringが「自分の試みは２０００年までに実現されるだろう」と予言しました。私は、自然言語処理の側面から見て、これはとてもあり得ないことといわざるをえないと思います。

昨日の講演の中で私がとりわけ同感したのは、おもちゃのようなシステムから抜け出るというお話です。委員会のいろいろな計画を見て歩き、またKBSとその自然言語計画を見るたびに痛感させられるのは、私共はこのような計画のために規模から言えばおもちゃのような辞書やコーパスを作るという馬鹿げたことをしているということです。このようなものは応用の役には立ちません。産業界に入ってみて、私は産業用システムでは何が痛切に求められているかが理解できます。委員会では、散々の悪評を浴びたSYSTRANシステムを使っていましが、これは正直に言ってたいへん古い機械翻訳システムです。最初からよくはありませんでしたが、今となっては全くお粗末なものです。しかしそれでもこのシステムは政府の役人にはたいへん役に立ちました。本当に翻訳が必要かどうかを決めるための下訳として目を通すのに使えたからです。このような用途に役立った理由は、広範な辞書を持っていたからであるということに尽きます。残念なことに、これらの辞書は編制がまずいため、取り出すことが全く不可能です。

このことは今度の計画に取り掛かる際には大いに教訓となると思います。

第２の理由、「なぜ私が大規模データベースに挑戦したいと考えるか」は、今日の午後最後のお話を聞けばよく分かっていただけると思います。ところで私は夢想家のDoug Lenatにたいへん共感しています。彼のやり方が正しいかどうかは問題ではなく、彼がやろうとしていることが、目的として絶対に正しいと私には思われるからです。話によると彼の言うには、２０１５年までには、我々が今ワードプロセッサーを持っているようにバックグラウンドとなる常識を持ったデータベースを持つのがあたりまえになるということです。

私は、データについての細かい問題はさておいて、これは完全に正しいと思っています。現在の機械がいかに賢いかにはいつも驚かされます。これはどうしても必要なのです。処理パワーの保存コストは今では全く瑣末な問題になっています。問題となるのは完全にソフトウェアに関すること、特にデータベースの構造です。あなたがたがこのように２つの研究の流れを合流させようとお考えになるのは大変勇気あることだと考えます。それをやってみようというのは全く正しいことだと思います。イギリスにこういう諺があります。「never the twain shall meet」多分本当はスコットランドの諺だと思いますが、これは２つの流れをいっしょにするのは大変な困難だという意味です。正直に言って、私は、ヨーロッパの視点から見て、知識ベースと自然言語研究の世界を真に融合させようとするなら、あなた方は多大な困難に直面することになるでしょうと言わざるを得ません。ご健闘をお祈りします。

必要なことは、分野の壁を超えて挑戦すること、これは決定的に重要なことです。私個人としては心理学者にも参加して欲しいと考えますし、哲学者も不可欠になるでしょう。そのほかこれには多くの分野からの協力が必要になるのではないかと思います。科学だけでなく人文科学の分野からの協力はどうしても必要です。研究に飛び込む前に十分な検討時間をもたれることを切に望みます。私には、計画の正しい構造を得ること、特に辞書とコーパスの正しい構造を得ることが絶対に必要だと考えます。そしてこのような大規模な構築に取り掛かる前に、初期の段階で、慎重を期すことが本当に必要であると思います。

国際協力について言えば、私の考えではこれは絶対必要条件です。世界はあまりにも小さくなってしまいました。これについては多言を要さないでしょう。我々は互いに依存しあっており、もし協力が不可能だなどと考えるなら我々は自らを知識人と称することはできなくなります。

言語の問題そのものが非常に大きな問題であり、あらゆる助けが必要だと思います。ヨーロッパがこの問題の発祥の地であることを思い起こして頂きたいと思います。公式には、ヨーロッパ連合には９つの公用語があります。ということは委員会の公式文書の言語の組み合わせは７２種類あるということを意味します。ヨーロッパ連合に1500人以上もの翻訳者が雇われていることも不思議ではありません。

しかし、もちろん実際には９つの公用語以外にもたくさんの言語があります。例えば、アイルランド人はどうでしょう。驚くべきことに自国語を公用語の１つとはしないことにしています。自ら選んでそうしたのです。しかし、ゲール語を使う人はたくさんいます。スペインと南西フランスにはカタロニア語を使う人々が１００万人以上いるでしょう。それからウェールズ人などヨーロッパの中で自分達の言

語を維持していきたいと考える小数派が他にも多くいることも申しあげておくべきでしょう。その理由はつまるところ、言語とは文化の核心であり、文化は我々の文明の基礎だということです。従って、もしも自らの言語を失えば、文化をも失うことになるのです。

私は並行した計画が必要であると思います。それは梯子である。礎を日本とヨーロッパと米国に置いた三角形の梯子であると考えます。しかしそれは互いに連結され、特に標準に関して並列に連結されたものです。ただしこれは正式の標準とするのではなく、研究を共にするための非公式の標準とするべきだと考えます。多言語コーパスと辞書を構築するための協力は切に必要とされています。

私は委員会に対して同様の計画を設置することを個人的に助言したつもりですが、しかしこれは完全に新規の計画とする必要はないと思います。むしろ既に進行している研究の上にマトリックスを乗せて、調和と統合を図ることが必要なのです。日本の皆さんにとっては込み入った問題になるでしょう。正直に言って、ヨーロッパにいる我々にとっては手に負えないことです。しかし実際にいくつかの計画が平行して行なわれているのです。ESPRITは今ではもはやDG-13のもとではなく、DG-3のもとで行なわれています。これはindustry director generalshipのことです。ESPRITそのものは２つの部分からなっています。基礎研究計画、日本との共同研究はほとんどこの部門に入ると思います。そしてより大きいESPRIT計画の本体です。この中の様々なプロジェクトについては今日既に皆さんがお聞きになった通りです。明日話題に登るはずのIdomeneusは、「これがIdomeneoであれば魔笛を連想させてくれてよいのに」と思います。私はいつも魔笛を思い出すのが好きなのです。これは高級ネットワークで、つまり全研究センターを結ぶものという意味です。この場合の目標は、データベース、マルチメディア、情報検索などの技術に関する研究のために、ライブラリーやアーカイブなどをオンラインで、開放型メディア分散処理環境方式で結ぶことです。しかし、ESPRIT以外にも、DG-13の言語工学計画があり、それはブリュッセルよりもルクセンブルグに拠点を置いています。地理をご存知の方のため申し上げておきます。それからもちろんEAGLES標準計画などの名前もお聞きおよびとおもいます。Zampolli教授はREALTORについてお話くださいましたが、この計画はヨーロッパでの辞書およびコーパス構築に強力な貢献をしてくれるものと信じます。またここから日本への協力が広がっていくと思います。私はこのプロジェクトに参加することができて喜ばしく思っております。それからまたElsnet高級言語ネットワークがあり、スピーチおよび言語の研究をつないでいます。

他の方々と同様、私が考えるところでは、渕博士の理想主義は、先生が「自由なfree」情報交換ということをいわれると行きすぎになってしまうのではないでしょうか。この「自由なfree」という言葉そのものが英語において極端に曖昧な言葉で、世界中が市場原理主義に突っ走っている昨今、この「自由なfree」という言葉は政治方面では死語になってしまっているのではないかと思います。私の考えでは、我々は特許の本来の原理に戻るべきなのです。特許は国王と個々の発明家の間の協定だったのですが、一方でこう言います。「よろしい、あなたに一定の独占権を与える。あなたは、それを作るという意図でそれを利用する限りにおいてあなたの発明によって報酬を得ることができる。そして私は自由に（freely）という言葉を、自由に手に入る（freely available）という言葉を使用する。」広く手に入

る（widely available）という言葉を使ったほうが安全かもしれません。我々の研究における原則はこのようなものになるべきだと思います。法人または個人がコーパスまたは辞書に自分の仕事をつぎ込んだ場合、当然その報酬を求めるはずだということです。そして我々が前進を望むならそのことを直視する必要があります。

しかしここでもう一度パーソナルコンピュータ革命の柱のひとつを思い出していただきたいのです。これはあまりにも忘れ去られていることが多いのですが、今日我々全員の机の上にパソコンがあるのは、Gary Kildollの勇断に負うところが大であるということです。彼はオペレーティングシステムCMPの作者であり、彼を説得したのはあるエレクトロニクス雑誌の編集者で、そのオペレーティングシステムを大企業向けの２０万ドルで売るのをやめて、１回につき６０ドルで売るようにさせたのです。この勇断のお陰でマイクロソフトなどが莫大な富を手にしたのはもちろんで、これが良いことなのか悪いことなのか分かりませんが、しかし、パソコン革命におけるその重要性は誰も否定できるものではありません。

コーパスの分野でもこれが必要なのです。我々は、言うところの「高く積んで安く売る」ことが必要です。実際にはそうしていません、この言い方をするのはアメリカ人で、全くものの言い方が違うのです。別の言語です。さて、最後に申しあげておきたいことは、この計画には４つ、できれば５つの異なる部門が必要だということです。研究部門が必要だと思います。率直に言って、なすべきことはまだたくさん残っています。今後何年も残るでしょう。昨日の藤澤令夫教授の、知識の本質についてのお話は非常に素晴らしいお話だったと思います。我々皆にとって、とりわけDoug Lenatにとって注目すべき内容だったおもいます。その方面でなすべきことは多くあります。特に知識表現と知識索引の問題を強調しておきたいと思います。

コーパス構築について。我々は柔軟性と適応性が求められます。machine-transferringとself-referentialの能力が求められます。品質基準は不可欠であり、それについておよび標準規制手続きについての国際的合意を目指すことになります。各国の政府がヨーロッパにおいて、もちろん協議会の形になりますが、このような研究に資金を提供することによってそれぞれの役割を果たすことはどうしても必要であると考えます。さもなければ計画は甚だしく遅れることになるでしょう。計画の一部はシステム・アプリケーション構築に向けるべきです。これは非常に重要なことです、一部には業界を、そして何よりユーザーを巻き込むこと、それによってある種の検定、我々が前進しているあいだに現実の世界で何が起っているかを教えてくれるユーザーからのある種のフィードバックが得られるからです。これはしばしばないがしろにされます。これが協力に様々な困難をもたらす等の理由からです。しかし私の考えではこれは絶対に不可欠なことです。もしユーザーを巻き込むことができれば応用の道のり全体が加速されるでしょう。

私は標準化の研究が分かれ目であると考えています。私が言っているのは事実上の標準のことで、もしお望みならIntercept標準といっても結構です。Zampolli教授は私達が進もうとしている道がどのようなものかを示してくださいました。ヨーロッパのEAGLESについてもお話もありましたが、私はこれを世界的な仕事に拡大しなければならないと思います。

もう１つ付け加えさせていただきたいのですが、このような計画では人々の移動の計画が必要だと考えます。これは技術移動のため、車輪に油をさすための最善の方法で、若い研究者が簡単によその国に行って仕事ができるようにすることが非常に大切だと思います。振返ってみるとこの協力計画の波は日本の第五世代宣言から発しており、人的訓練はその中でも最も重要なものであったと思います。そして国際協力を成し遂げるのに、人々の移動以上に優れた方法はないのです。

さて、話が長くなりすぎたようです。私は車椅子に座って、我々の間の協力が発展していくのを見せていただきたいと思います。ヨーロッパ連合、協議会、そして特にここに参加しているヨーロッパ人を代表して本計画へのごあいさつとはげましをお伝えいたします。有り難うございました。

渕：

どうもありがとうございます。Thank you very much. 全ヨーロッパといいますか、ヨーロピアン・ユニオン的な立場、あるいはいろいろな観点からお話を頂いて、どうもありがとうございました。

それでは、これからディスカッションを始めたいと思いますけれども、パネルの方々もあるいはお互いにコメントがあったり、あるいは「もうちょっと言いたい」という事もおありかもしれませんが、ございましたら若干の補足をして頂いても結構です。

それともフロアの方のご質問等を先にいたしましょうか。

それでは、いろいろご意見は、パネルの人達もたくさんおありだと思いますが、最初の予定の通りに、皆さんの方からご意見とか質問を寄せて頂いて、それをきっかけに、またパネルの方々それぞれの立場の補足をして頂こうと思います。キーワードとしては「大規模知識ベース」という１つでも済むかもしれませんけれども、関連していろいろな非常に多岐にわたる論点があるわけですが、どのあたりからでも結構ですから活発な議論ができるかと思います。どなたかまず、口火を切って頂けるかた、おりますか…。プリーズ。

Bob Jansen：

オーストラリアのCSIROのBob Jansenです。私はこの計画全体について１つ心配なことがあります。それは基本的に我々はスタートレック型とも言うべき活動を始めようとしているのではないかということです。本当に正直に言ってしまえば、もし私が政府の誰かのところに行って、我々はスタートレック・タイプの活動を開始しようとしており、それは２０年くらいは商業的利益を生まないでしょうと話したとしたら、返ってくる答はこうなるでしょう、よし思い切ってやれ、タオルを持っていけ、ヒッチハイカー用ガイドブックも忘れるな、そして１９年と半分たったら戻って来い、その時には何かできるかもしれない。現実問題として、予測可能な将来のことです。私の心配は現実の応用分野がまだ無いことです。それがあれば、その時間枠内に十分な利益をあげることによって、政府を研究資金を出そうという気にさせることができるのですが。このパネルが、幅広い専門知識によって、このような計画のために何か新しい応用分野を示すものになればと考えています。

渕：

最後のポイントでは「この分野でアプリケーションがほんとうにあるんだろうか」ということですけれども、いかがでしょうか、パネルの方、あるいはフロアの方からのお答えでもいいんですけれど。では、You're going to? Please.

Oakley：

議長、直ちにワードプロセッシングの分野をあげたいと思います。あらゆるワードプロセッサーには自然言語機能にかかわる部分があります。ワードプロセッサーは、多言語であろうとするなら、コーパスに依存しており、そのコーパスは概して適切な形式になっていません。この分野では自社でこのようなコーパスを作り上げることができる企業は数少ないのです。この分野では特にオフィス環境に関して直接の応用ができると考えます。作り上げたコーパスを利用できるようなものがです。

Gray：

バイオテクノロジーの分野については既にお話しましたが、皆さんはいずれにせよ我々はそれにごく近づいているのだとおっしゃるかもしれません。私は今後５年から１０年のうちに工学デザインがキラー・アプリケーションとなるのではないだろうかと考えています。これはBobが考えておられる方向だと思います。このキラー・アプリケーションというのは随分おかしな言葉ですが、競争相手を全滅させてしまうという意味のようです。しかし本当にこれができるようになる必要があるのです。そしてそれが工学デザインの分野になるのではないかと私は考えるのです。人々がさらに野心的になるほど、デザイナーはさらに野心的なデザインをすることを迫られます、デザイン技術を推し進め、先端制約推論や、ある種の限界まで充足した技術を使うことが求められるようになります。この種のことが推進力になると考えます。それが私の考えです。

渕：

すみません、しばらくお待ちください。彼が答えます。

回答者：

それではアプリケーションについて一言お話します。つまり「大規模知識ベース」。今大量さが、ある種のブレークスルーがすべてに必要でして、それは個々の企業ではなかなか達成できない。従ってある程度プリコンペティティブなプロジェクトとして、ナショナルあるいはインターナショナルプロジェクトとして実現すると。その「量のバリア」が解決しますと、今度は一気にその上に個別の企業が－小さなグループでもいいです－小さなグループがいろいろなアプリケーションを開発できる、共通の土台が出来上がるのです。自然言語も、機械翻訳の例を考えて頂ければ分かりますように、非常に大規模な辞書、あるいは大規模なコーパス、こういうものを用意しないと有効な機械翻訳システムを開発できま

せん。しかし、これはすべての企業が個別に用意していたのでは、これはものすごい浪費になります。ですからこういうものは共通に作る。それをシェアして、今度は「その上で、それを使ってアプリケーションを開発する」というような、自由に競争して、いろいろな有効なお金儲けの手段を考える。今それをやる必要がある、という段階なんだと私は思っております。

渕：

まあ確かにいろいろな意見があってもいい時期でありまして、まだ存在しないものについてはどちらの議論も成り立つわけで、私としてもいろいろな皆さんの意見をお聞きしたいわけですが、次の論点に移ってもいいかもしれないと思いますので…プリーズ。

質問者：

もう１つの質問をしようとしていたのですが、しかしアメリカの知識共有研究を代表して、私共が短期的収益可能性があると考える応用分野についてお話させていただきます。その応用分野の一部は設計製造であり、一部は我々が電子商業と呼び始めている分野、もうひとつはオフィス・オートメーションの分野です。基本的に、これらの分野は、企業内の様々なグループでの小規模な個別の機会を通じて、情報を共有するものになると私共は考えています。それによって製品設計が迅速になり、そのような製品の生産ネットワーク内での問題に関する情報伝達が迅速になり、それにより各企業が協同して新たに一時的提携を組むことができるようになります。即席の提携といってもよいでしょう。またそれによって各企業は新たな製品をもっと早く市場に出すことができます。ここに多くの収益可能性があると考えます。もう１つは輸送計画の分野です。私共の研究の多くの部分が商品とサービスの動きのプランニング支援に向けられており、私共はこの分野に短期的収益可能性があるとみています。

話題を変えてよろしければ、Gray博士の話された問題をフォローしたいと思います。パネリストの皆さんのご意見をうかがいたいのですが、公益のため公的資金によって標準化すべきと考えられるのは大規模知識ベースのどの分野だとお考えになりますか。またソフトウェア製品を作っている企業が商機を求めるトピックとしてどの分野が残るべきとお考えになりますか。

渕：

All right? それでは、お２人に質問がいったようですがChenさん、Could you comment? Mr.Chen? In sequence.

Chen：

それは大変難しい質問だと思います。そうですね、NSFの立場は、知的所有権、著作権の問題には関わらないというものです。アメリカでは大学や企業に、また大学教授のスピンオフにも同様に、そのような権利を享受させています。政府はただ研究を促進することによって、国家経済を支えることができ

るのが望ましいと思います。これが私の考え方です。NSFの立場から見るとこれは答えにくい問題です。もう少し検討が必要だと思います。

渕：

結構です。Gray教授、コメントあるいはお答えがいただけますか。

Gray：

私は分子生物学の例をあげます。政府が分子生物学データベースを準備できる研究所に対して資金を出すことに同意しなかったなら、今のような早さで物事が進んだとは考えられません。もちろん多数の結晶構造が公的資金によって大学の研究所で得られました。従って、私はこれを一種の事実ベースとして、これを組織化し統合することによって、真に完成させ、多くの人々がこの共有情報を利用することができるようにしなければならないと考えます。またこのようなコーパスに関する議論は優れたものだったと思います。自然言語処理研究がお話のように小規模な辞書その他のものを用いて長く行われてきたことをうかがって驚いております。私にはこうした研究には大規模なものが絶対に重要だと思われます。私共イギリスでは市場そのものが原因でその物が実現しない時、それを「市場の失敗」と呼ぶと思います。コーパスの分野には明らかに市場の失敗があります。

Chen：

それは非常に重要なことだと思います。米国においては、ご存知のように、LDCデータが世界中の全ての人に平等に開かれています。これは本当に模範として示すべきものだと考えます。NSFの立場からいえば、全てが公的領域にあることが望ましいことは前にも申しあげました。そしてLDCモデルが優れたモデルとなりうる理由ですが...コストはどのくらいでしょうか。たしかSusan Armstrongさんがおよそ2,000ドルと言っておられたと思います。会費が2,000ドルですね。それですべてのデータにアクセスできるわけです。既に１００個のCD-ROMがあるのだったと思います。１００個から１５０個のCD-ROMがたったの2,000ドルです。これはモデルとなります。

渕：

ありがとうございました。ではZampolli先生、コメントをお願いします。

Zampolli：

幾つかコメントしたいと思います。ひとつは応用と言語資源との関係についての答です。通常、確かに自然言語処理とスピーチに対する市場があります。自然言語とスピーチの市場を不死薬の市場と比較してみましょう。もしも不死をもたらしてくれる薬があったとすると、その市場は全人類になります。唯一の問題はどうやってそれを作り出すかです。それを作る技術が必要です。さて自然言語処理とス

ピーチ技術が現在できることは限られており、必要なことすべてができるわけではありません。そしてその技術の状態をもとにして何が可能かを確認することが重要なのです。しかし、確かなことは応用に進みたければ原型から出発して問題に進まなければならないということです。１年から３年の間に市場化が可能な問題にです。またもし我々がしたように産業界のことを話題にするならば、産業界が求めているのは、少なくともヨーロッパでは直ちに言語資源を得ることです。言語資源が直ちに得られることが製品化の可能性の前提条件です。私がお答えできるのはそれだけです。

ヨーロッパでは資源の定義のためのプロジェクトはたくさんありますが、資源構築のためのプロジェクトはまだないという現実の問題についていえば、私はRohrer教授に全面的に賛成します。今朝もお話したように'８６年にDont WorkerのFidelが、ご存知のように数日前に亡くなった方ですが、我々には言語資源が必要であるというアイディアを打ち出しました。それから現在までに我々が得たものといえば定義段階への資金だけなのです。我々はもちろん資金獲得のために戦っています。この点に付いては私はRohrer教授と少々意見が異なります。彼は言語資源の構築は完全に国家機関の仕事であるべきだとおっしゃっています。私は少なくとも言語資源の房室に当たる部分の構築は、例えばヨーロッパでならＥＣが行うべきであると考えます。ここではお話していませんが、理由はいろいろあります。もちろん、これは非常に重要なトピックの１つです。Oakley教授に申しあげたようにRelatorやEagleのようなプロジェクトが日本の方々の自信を深めるものとなれば喜ばしいと思います。またOakley教授が、ＥＣに対して持っておられる影響力によって、それが非常に必要だということを説得してくださることを望んでいます。

LDCが非常に重要な計画であることについては同意見ですし、私共も協力しています。確かに研究目的のために2,000ドル払わなければならないところを、もしこれが企業だったら20,000ドル払うことになるでしょうが、これはたいしたことではありません。私が確信を持っているのは、LDCが手に入るものを収集していることは正しいという点についてだけです。我々がこのモデルに付け加えるべきだと思われるのは、必要なもの全部は得られないかもしれないという事実であって、何か欠けているものがあれば、それを自分で作らなければなりません。例えばLDCの中のARPAは市販の辞書が無いために、辞書を作るためのプロジェクトを開始しました。コーパスについても、事情は同じだと思います。現在ARPA内にある大規模なコーパスだけでは、研究者の要求を満足させることができません。なぜならそのコーパスは厳密な基準にしたがって構築されたものではないからです。バランスの取れたコーパスではないのです。単にあるものの集めにすぎません。ということで私もRohrer教授やOakley教授に声を合わせて、我々はただあるものを寄せ集めるのではなくモデルに従ってオブジェクトコーパスを構築する必要があると申し上げましょう。

渕：

有り難うございました。どうぞ。

Oakley：

資金がどこから来るのかということと、誰が研究を行なうのかということの間の区別だけはっきりさせたいと思います。ヨーロッパでは、研究の段階では、資金提供または研究の資金繰りの援助を行なう責任が国または協議会にあることは、当然のこととして同意されていると思います。国が適切な規模の責任を果たしているかどうかについては当然意見が分かれます。メンバー国の間で見解の相違が出てくるのは、完全なコーパスを構築するというビジネスの段階に入った時です。私としては協議会や国は品質管理を続け、基準の監督権を維持して、良い研究だけが受け入れられるようにするなどの必要があると考えています。

しかし国が研究をすべきかどうかというのは全く別の問題です。実現可能性がはっきりした時には直ちに産業界の参加を求め、その仕事を手伝ってもらうことが重要だと思います。もちろんリードできる学術団体と協力する事によって、どうしたらいいかを示してもらうことによってです。しかしもしこの分野を強力な産業にしたいのなら、これは自然言語の分野では誰にとっても最優先のことだと思いますが、研究初期の段階から産業界の参加を求めることが重要です。

渕：

どうぞ。

質問者：

ほんの少しだけオーストラリアの方のご質問に話を戻したいと思います。様々な応用分野をお考えのようですので、少し違った意見を申し上げたいと思います。私は、政府機関やまた幾つかの企業を説得して自然言語処理にもっと資金を出してもらうための、論拠を探し求めてきました。今、企業が利益を得ることができるような応用分野を示すことは容易ではないのです。また今私は１年の半分を米国ですごしており、いろいろな企業の人々に会っては、「お宅では自然言語処理からどのくらいの利益を得ていますか」とたずねています。１つおもしろい答が返ってきました。ある人がこう言いました。「いいですか、フォルクスワーゲンに行って、お宅の車のリア・アクセルからどのくらいの利益を得ているか聞いたとしても、答えようがないでしょう。」自然言語についても同じことです。要素はより大きなソフトウェアパッケージの一部となることができるのです。だから、私達はこれまで間違っていたのではないかと思います。私達が「機械翻訳で利益をあげることができますよ」と言って、むやみに難しいことをやろうとしていたことです。「私達は、今、ソフトウェアパッケージの一部を提供することができます」と言うなら、チャンスはもっと増えると私は思います。

Brachman：

Ronald J. Brachmanです。今朝の短かった持ち時間での、この知識共有に関する素晴らしい議論を離れ、もっと大規模知識ベースへの知識獲得に関る問題に議論を移したいと思います。特に議事ではほと

んど触れられていないものの、私は非常に重要であり将来の成功にとって決定的であると考える問題です。特に、学術界および人工知能の世界の外にある、一般的な現実世界のデータの必要性についてお話したいのです。大規模知識ベースについての我々の研究はようやくその緒についたばかりです。これはごく新しい種類の分野なのです。長年にわたって大量のデータが多数の従来のシステムによって収集されてきました。そして、この分野で我々がいかに努力しようと、まだあと１０年間は、従来のシステムによる従来の需要のためのデータ収集が同じ理由によって同じ方法で続けられていくことも、また明らかだと思います。したがって、データ処理の世界および科学界、産業界によって、収集されたデータを扱えるような体制を整えることが重要だと考えられます。さもなければ主要な科学および産業の応用分野から切り離されてしまう危険を冒すことになります。またビジネスの問題になれば、当然それは我々の将来の富の、最大の源になるでしょう。肯定的に見れば、これは大規模知識ベースへの飛躍的出発と考えることができます。今朝はデータの問題に、特に焦点をあてたかったのです。私のコメントは、会議で皆さんがお話になったことを補足することを意図するものです。これは、決して自然言語情報やテキストコーパスなど、大規模知識ベースの事業に関わる事柄に重要性がないというつもりではないのです。しかし私は実際に存在するデータについて、もう少し時間をかけて話し合って見ることが大切だと考えます。結果として、その話しで私が目指す目的はごく単純です。私は基本的に一つの点を明らかにしたいのです。データはそこにあり、これからも収集され続けるのだから、それをまじめに考える必要があるということです。従って私の目的はただ皆さんにこのトピックの重要性を知っていただくことなのです。私のコメントの中でもう１つやりたいことは、ほんの幾つかの研究課題を明らかにすることです。これもまた、２日間で話し合ってきたことの補足になると思いますが、ＫＢ＆ＫＳの研究において考慮すべき、重要な問題です。最後に、私共が作った実験的プロトタイプをお見せすることにより、ＡＴ＆Ｔで私共がとっている一つのアプローチをご紹介したいと思います。最初のポイントは大変単純なので、ご理解いただくのに時間はかからないと思います。本会議の紀要の表紙をご覧になると、オーガナイザーは賢明にも私達の直面している問題のひとつを示してくれています。それは世界は海とデータの海で覆われているということです。その量は実に圧倒的なものですが、日本での本会議では、故意によってか失念かは分かりませんが、非常に根本的な問題が語られないままになっています。すなわちコードセットとその標準化の方法の問題です。昨日、Yorick Wilks教授が知識の文化間移動のことをお話になりました。もしかしたら呼び方は違うかもしれません。またSusan Hockey教授がＴＥＩと標準化などのことをお話になりました。しかしこれは皆アルファベットコードセットについて、というより文字コードセットについてのことであって、非アルファベット文字についてのお話はありませんでした。実際の応用ではこれが大きな問題になるのです。

中国と日本の優れたプログラマーの約６０パーセントは、主として欧米で開発されたソフトウェアを現地化するためだけに忙殺されているという根拠のある見積もりがあります。もし我々が大規模データベース、大規模知識ベースを作ろうとするならば、コード、すなわち非アルファベット文字のためのストリップコードの問題という、恐らくは基本的な問題を考えてみる必要があると私は考えます。

もしかしたら私はここでの真の問題を理解していないのかもしれませんが、これは当面現実の問題です。過去５年間文部省の研究資金を受けて私共はこの、もしかしたらあなた方からみればごく基礎的なものかもしれませんが、中国・日本・韓国の間の書誌情報データベースをいかにコーディネートするかを研究してきており、今も継続中です。この３ヵ国で、通常の方法で情報交換ができるかどうかです。これについて、このごく基礎的な問題に関するご意見をうかがいたいと存じます。

渕：

Ask me why? or us, O.K.

Please, Oakleyさん…。

じゃあ、Professor Yamada, Please.

山田：

どうも有り難うございました。私はパネルメンバーの方々に向けて質問したつもりだったのです。というのはパネリストの方々は、もちろん司会者を除いて、欧米の方ばかりだからです。しかし、私の質問は本当にこの会場の日本人に向けたものではなかったのです。どうもありがとうございました。

渕：

O.K.…You're not….

（？）：

山田先生のような流暢な英語が喋れませんので、日本語でやらさせて頂きます。Hockeyさんがお話されたＴＥＩとの、アクティビティーの関係を、コメント的にちょっとご説明しておきます。やはりＴＥＩの活動が、アメリカとヨーロッパの方々がイニシアティブをとられてスタートされて、いろいろの仕様が決まった段階で、やはり幹事権の問題があまりとらわれていません。長尾先生とお話をしまして、Hockeyさんを始めとしてＴＥＩの方々に提案をして、これに関しましては非常に快く、やはり日本及びアジアの人達の協力がいるということで、実は一番働いておられるのは千葉大学の土屋先生なんですが、それで現在、そのＴＥＩのワーキング・ループでもできるだけ日本から土屋さんが出ていかれて、そして幹事権の問題からいろいろ意見をお話しております。これに関しましては日本側にももう１つ問題がありまして、やはりヨーロッパとアメリカのファンディングで動いているわけで、やはり日本もちゃんとしたファンドがいると思います。お金のサポートをしないと、向こうで決めてくれたものが気に入らないといって文句を言っているというのでは話にならないんでして、それで長尾先生ともお話をして、何とか日本もコントリビューションができるようにという議論をしております。しかしながらまだ解決策が出てないので、山田先生にもご協力を頂ければと思います。

このようにしてＴＥＩの活動を、日本及びアジアにも定着させていきたいと思っております。さらに、

Zampolli先生が言われたコーポラに関しても、より密な協力関係ができればと思っております。そのためには、日本にも問題があります。アメリカがLDCを作ったり、そういうセンター的なものをきちんと作っておられるのですが、それに見合う日本の仕組みがまだありません。日本はこのVLKBというコンセプトを提案はしておりますが、足元が固まっておりません。至急これは我々日本側の努力すべき問題で、是非とも学術情報センターの強力なるご協力がいるのではないか思います。

渕：

O.K.いろいろな論点が含まれてまして、「アプリケーションがあるかどうか」という議論だけでも１時間や２時間できる問題ですし、その先の「技術的にこういう知識ベースをどうシェアするか」という他に、「著作権問題を含めたシェアリングの問題」というのもあるわけです。アプリケーションがなければ、権利の問題もなくて済んで楽なんですけど。ということでたくさん議論があると思いますが、やはり時間というものが限られておりますので、私の最後のコメントといいますか、若干の意見を述べて２日間の会議を終わりたいと思います。

具体的なことは別にしまして、やはりこの２日間を通じて感じたことは、現在はこれまでの数十年間の情報処理の歴史、これはエクスペリエンシャルにも見えたんですが、さらにその先につながるためのトランジェントな時代だということをこの２日でも感じました。物事と言うのは連続に伸びていくものと、あるいは飛躍的に大きくなると見える部分との組み合わせになるわけですが、過去を単純に効奏しただけではそう大きな展望は開けないと思います。しかし、この「知識ベース」という世界だけではなくて世の中全体を見ると、これも大きな変革期、大きな時代への進展だと私は感じています。コンピュータの周りだけでも「知識ベース」といわず、そのベースになるコンピュータネットワークーこれは非常に広がっていこうとしていますし、日本もそういうことが大事だということになってきているわけですがーこういうものの発展とまた知識ベースの発展というのは当然深い相関があるというのは、既に指摘されているわけです。それだけではなくて、やはり私は経済学は素人ですけれども、やはり人々が働く環境というものが大きく変わろうとしているという気がしています。

今までの観念でいうと、大きな企業がたくさんの人を抱えてものを生産して、それを「顔のない無数の大衆」に売って儲けると。例えばそれを「アプリケーション」というようなことが多かったと思うんですけれど。そのパターンが崩れ始めているのではないかという気がしています。大きな企業では何かを作る。これはこれからも必要な場面は残ると思いますけれども、それよりはもっと小さな世界、グループで、そこで創造性をつぎ込んだ色々なものができてくる。ソフトウェアの世界なんか特にそうだと思いますが、そういう時代。それから使う方のユーザーの方も、単にカタログに書いてあることを指向するのではなく、むしろアクティブに反応するようなユーザーが増えていくというようなことになっていくような気がしています。ですから、過去の延長で「知識ベースが儲かるか、自然言語処理が儲かるか」というだけではない要素を入れて考えていい時期が来ているんではないかという気がします。そういう観点でいうと、例えば、米国で「メガ・ユーザー」というような意識があるわけですが、これは

十年後か二十年後か数十年後を考えれば、「メガ」ではなくて、「ギガユーザー」であるかー「テラ」まで人口が増えると地球がパンクしたりして、別問題が増えますけどもーそういう風に情報処理の世界が広がっていく。その中身は何かという議論をしだすと話が元に戻ってしまいますけれども、歴史の流れとしてはそうだと思います。その中でこの「知識ベース」というものがいらないのかいるのか、いるとすれば「どういう技術的な問題を解決していくべきか」と、「そのための体制はどうか」と、１つの企業でできなければ、人が集まってナショナルプロジェクトもしなければいけないでしょうし、１国では足りないような分野であるわけですから、これは当然ながらインターナショナル・コーポレーションということで、「知識」というものの本性自身が、基本的にはやはりみんなのためのものを作ってるというのがこの世界の人達の「創造者の気持ち」だと思うんですね。儲けるためというのは生きて行かなければいけないから儲けるということがあるんでしょうけど、本当に優れたものを作った人というのは「みんなに使ってもらって有効に活用してほしい」というのが基本的にある。それだけの理想論ではいけませんけれども、そういうものがベースにあるという風に私は思っておりまして、まあ、そういうことをベースにすれば、国際協力あるいはそれぞれの国でのナショナルプロジェクトなんかをも、これまでよりずっとうまく成果を上げるんじゃないかという気がしております。この会議を通じて何か１つの非常にまとまった具体的なメッセージを出すということは目的にしていなくて、むしろ色々な問題点を議論したり発表して頂こうという主旨でしたから、そういう意味では色々なプレゼンテーション、あるいはディスカッション頂いて言うべきだったと思います。

私はやや、楽天的な理想家のところがあるーあんまり楽天的でもなくて慎重なところもあるんですがー楽天的に考えますと、今まで述べましたように、この「大規模知識ベース」へのステップというのは、世界中で大きく展開するに足るテーマであり、そのための準備を始める必要があると思っています。これは、この半年、１年は急がなければいけないというものではなくて、場合によっては３年でも５年でも議論を詰めなければいけないかもしれませんし、必要なら来年からスタートする必要もあるかもしれないと、その辺の緊急性の議論も含めて、これから私としては世界中で議論をすると同時にそれぞれのところでもスタートできればと思っています。

また、私自身は国籍が日本ですから、日本もやはりその分野で貢献できて、それは「世界中の人達のためのそういうもの」を国のプロジェクトとしてやって欲しいと思っていますし、私も力が弱いかもしれませんが、私にもし少しでも力があれば、政府とかあるいは日本のいろいろな関係の人達に働きかけていきたいと思っております。

それでは、この２日間大変皆さん貴重な時間をさいて出席して頂き、また、それぞれにお話頂いたり、いろんな立場でお話頂いたパネルのメンバーの方々に感謝したいと思います。皆さんの拍手を得て感謝の意を表したいと思います。Thank you very much.。

7. 閉 会 挨 拶

7. 閉会挨拶

財団法人　日本情報処理開発協会
専務理事　　照　山　正　夫

閉会にあたりまして、一言御礼のご挨拶を申し上げます。

ご出席の皆様におかれましては、年末を控え何かと御多用のところ、長時間にわたり熱心にご参加頂きまして、誠にありがとうございました。また何よりも、内外からご参加頂きました２３人の講師の先生方、特に遠路海を越えてはるばるご参加賜りました外国からの先生方には、この国際会議の為に快くご講演・ご発表をお引受け頂きまして、誠にありがとうございました。あらためて心から感謝申し上げる次第でございます。

２日間にわたります、皆様の貴重な示唆に富んだご講演、また活発なご意見の交換を通じまして、この会議が２１世紀の知識処理インフラストラクチャーとして期待されます大規模知識ベースの構築と共有、またそのための国際協力の重要性についての共通理解の増進に、大いに寄与したものと確信するものでございます。

本会議を滞りなく、かつ盛会りに終えることができましたことに、重ねて御礼申し上げますとともに、最後になりましたが、会議の開催にご支援をいただきました通商産業省をはじめ、後援・協賛をいただきました諸外国ならびにわが国の関係省庁・関係諸団体各位に、あらためて深く感謝申し上げまして、閉会のご挨拶とさせていただきます。

どうもありがとうございました。Ｔｈａｎｋ　ｙｏｕ　ｖｅｒｙ　ｍｕｃｈ.

第 2 部

ＫＢ＆ＫＳ’９３　国際ワークショップ

―　発表要旨・質疑応答　―

1. セッションI

知　識　共　有

１．セッションＩ：知識共有

(1)「知識コミュニティを目指して」

奈良先端科学技術大学院大学 情報科学研究科
教授　西田豊明

【発表要旨】

大規模知識ベース（ＶＬＫＢ）のアプローチでは２つのアプローチが考えられる。１つはアキュムレーション指向のものである。

ここでは知識コミュニティという、人工コミュニティを提案する。これは第１に組織構造、第２に人間とコンピュータ間の知識メディアの設計とから成る。枠組みとしては、通常のエージェント、ファシリテータのエージェント、メディエータの３つのカテゴリがある。

テストベットとして、分散情報システムＫＣ－Ｋａｎｓａｉを作っている。これは、関西圏の情報をイエローページの形で提供しようとするものである。

ファシリテータは下位レベルをモニタし、メディエータは上位レベルを扱うエージェントである。入出力はオントロジィ的に特定される。オントロジィサーバーは、入力された情報の意味を理解し、概念アソシエータに指示を与える。ここではアクティビティ空間と地理空間との対応をつける。ミスマッチした場合は人間の助け、協力を得ることになる。

【質疑応答】

質問１：どのような表現言語を用いているのか。

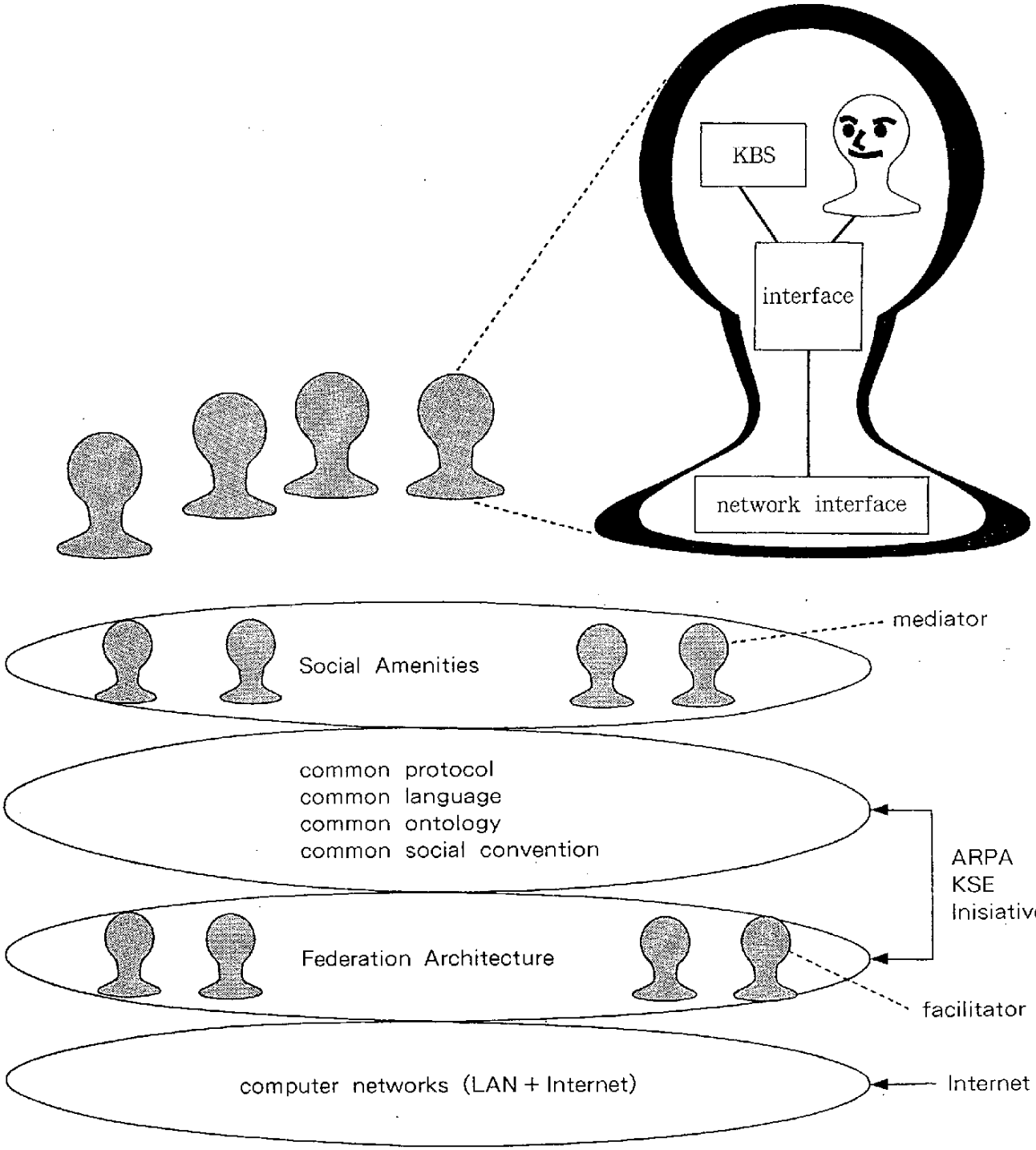
回答１：外部表現はＫＱＭＬを、内部表現はアドッホクである。

質問２：オントロジィはどのようなものか。

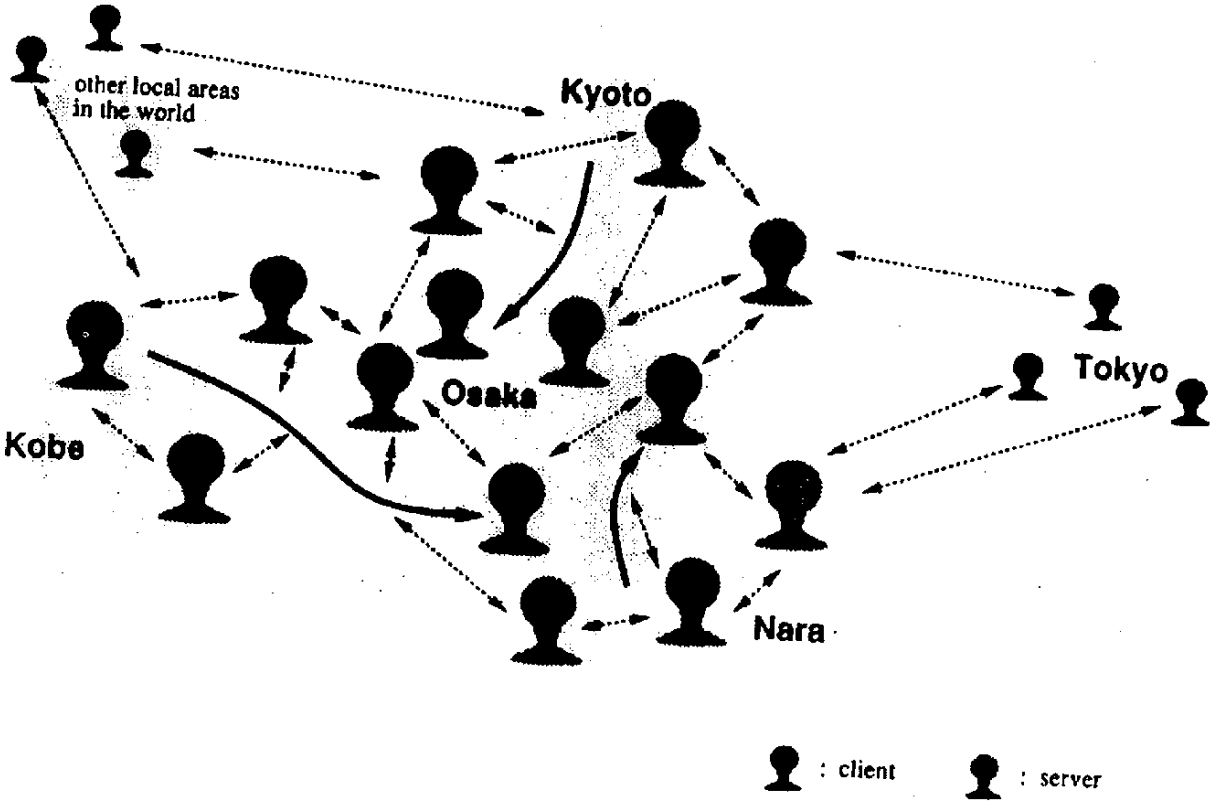
回答２：構築に時間がかかるので、現在は１つしか作成していない。

質問３：応用分野として何が考えられるか。

回答３：現在の案内のほか、会議参加受付などを考えている。将来は工学への応用を考えている。



OHP－1 : **Framework of the Knowledgeable Community**



OHP－2 ：KC−Kansai as a testbed

(2)「統合的ユーザ支援環境における知識共有：応用、フレームワーク、インフラストラクチャ」

University of Southern California, Information Science Institute（米国）

Project Leader & Associate Professor　Robert Neches

【発表要旨】

３つのトッピクについて話す。第１は知識共有の努力について、第２は応用と枠組みについて、第３はインフラストラクチャについてである。

具体的には、IN-USE という名の研究グループを組織している。この枠組みは、HUMUMANOID、BACKBORD、TINT、Scenarios / Agendasという４つのファシリティから成る。基本的にはモデルベーストメソドロジィを取り、ゴールは迅速なアセンブリである。

応用としては、たとえばジョージア工科大学との共同研究である MASTER MIND がある。これはモデル上にソフトウェアを作るための設計ツールを構築しようとするものである。その他にも FAST、DRAMA などがある。

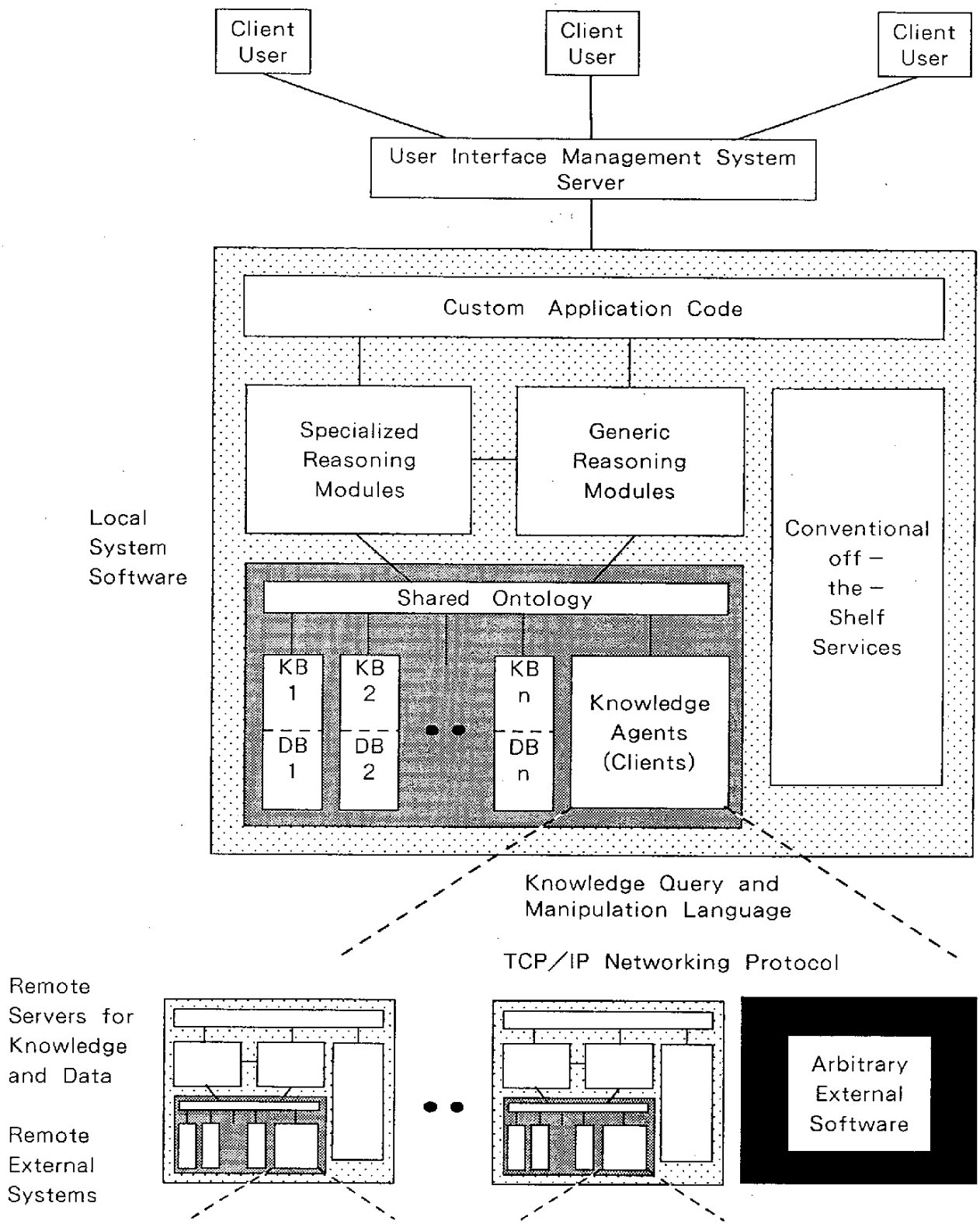
【質疑応答】

質問１：オントロジィの概念が、通常より大きいものようだが。

回答１：大きいとは考えていない。人によって異なるのは仕方がない。

質問２：設計に誤りはないか。

回答２：答えるのは難しい。モジュールごとに検討する必要がある。



OHP－1 : Accomplishments : Architectural model

(3)「コンテクスト：大規模共有知識ベースの実際問題」

CSIRO, Division of Information Technology（オーストラリア）
Program Manager　Bob Jansen

【発表要旨】

知識ベースシステム（KBS）には２つの問題が考えられる。１つは知識の詳細化、もう１つは知識の拡張である。

前者は、種々の例に見られるように、知識片がどんどんと細かくなっていく。後者は、ネットワークがどんどん複雑になっていき、文脈依存的な傾向が強まる。

文脈と共有 KBS との関係を考えると、マルチユーザ、マルチドメインになり、問題が複雑になってきて文脈が必ず必要になる。

これは知識獲得問題で、専門知識をエキスパートシステムに移植する場合にも起こる問題である。

また、KBS に対する質問に答えられないという場合にも問題になってくる。

これらの課題については、もっと国際協力して研究すべきである。

【質疑応答】

質問１：データモデルとの関係はどうか。

回答１：状況理論、コネクショニストがある。

(4)「大規模知識ベースの共有：ルール選択のアプローチ」

DFKI, General Research for AI（独国）
Research Scientist　Knuit Hinkermann

【発表要旨】

問題解決においては、さまざまな文脈の要因によってさまざまな知識が用いられることになる。その際の知識の選択の問題についてを論じる。

知識の選択は、質問や応用分野や領域知識やユーザの選好など、いろいろなパラメータに依存するが、こうしたパラメータを選択に利用できるかどうかは選択の方式による。本発表では、動的な処理の制御ではなく、前処理の段階において関係のある知識を選択する方法を扱ったものである。

ここでは、Horn節形式の知識ベースを用いる。選択は、同定（idendification）、適応（adaptation）、および予測（prediction）という３つのフェーズからなる。そこで、(無)関係性（(ir) relevance）という概念を提唱する。ある節が現在の脈において無関係であるとは、大雑把にいえば、その節が質問に対する証明に現われないことである。

この(無)関係性を求める上で、記号的な方法とニューラルネットワークを用いた方法とを提案し、両者を比較した。記号的な方法は、プログラムの構造をグラフで表現したものを用い、グラフの結合関係

Knowledge Base : Horn clauses

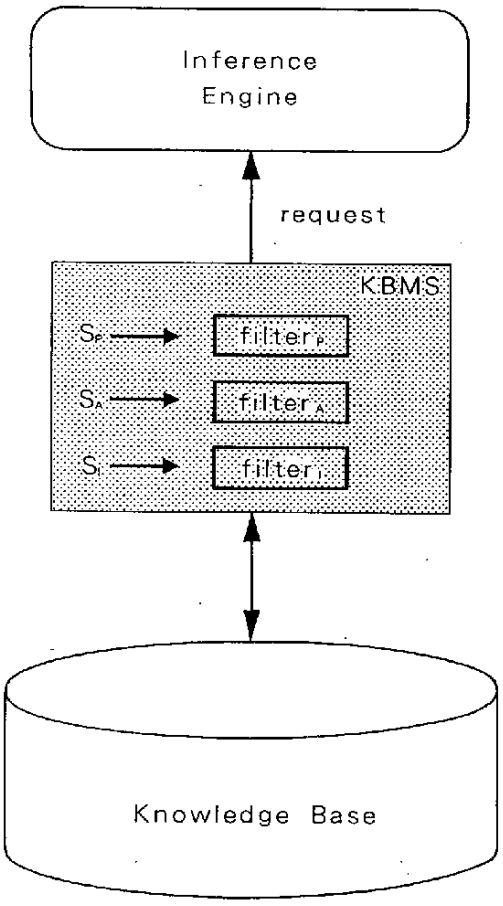
$\beta \leftarrow \alpha_1, \alpha_2 \cdots\cdots \alpha_n$

Selection Phases :

Identification : Internal Information
- Query
- Structural Representation

Adaptation : External Information
- Domain Knowledge
- Application Environment
- User's Preferences

Prediction : Context of Process State
- Short Term (actual consultation)
- Long Term (previous consultations)



OHP－1：Extending KBMS

から「強い意味での」無関係性を求めるものであり、ボトムアップとトップダウンの方法を検討した。

一方、ニューラルネットワークに基づく方法は、同様の木からニューラルネットワークを構成し、このネットワーク上の活性拡散によって関係のある節を求めるという方法である。ニューラルネットワークを用いた場合は、質問に関する情報の他、領域知識などの細かい情報も選択に用いることができる。また、大域的な情報を処理するために、束縛に関する情報を伝搬させるというテクニックなども用いることになる。

将来の課題としては、これら２つの方法の統合や、より大規模な知識ベースへの適用などが挙げられる。

【質疑応答】

質問１：プリコンパイルすることによって、この種の選択の効率が上がるという報告もある。プリコンパイルによって、選択がし易くなるのではないか。

回答１：現在は、選択の後でコンパイルしている。

質問２：記号的な方法では、関係のある節の集合を含む集合を選ぶが、ニューラルネットワークでは、その部分集合しか選択していないのではないか。

回答２：記号的な方法、ニューラルネットワークによる方法ともに、関係のある全ての節を選択してい

る。

(5)「大規模知識ベースの新しいフレームワーク：データベースと制約論理プログラミングの観点から」

(財)新世代コンピュータ技術開発機構 研究所
主席研究員　横田一正

【発表要旨】

新しく計画している知識ベースシステムの基礎的な概念に関する発表である。これは、多様（heterogeneous）で、分散的（distributive）で、協調的（cooperative）な問題解決システムである。このようなシステムでは、多くのエージェントが多くの言語や多様な問題解決戦略を用いることが前提となり、エージェントの間の協調に基づいた「知識エージェントの社会」を目指すものである。また、既存のデータベースなどの再利用も狙っている。

エージェントには単純なものと複合的ものがある。単純エージェントは、遂へいされた問題解決器であり、複合エージェントはエージェントたちが共有空間を介して連合したものである。個々のエージェントは、自分が単独で解決できない問題に遭遇すると、その問題を環境に投げ出し、その環境を共有空間とする複合エージェントにその解決を委ねる。最も外側の環境はユーザである。エージェント間の情報処理に関わる大域的な情報としてはメッセージ通信の他、環境とエージェント間の関係がある。環境は、その環境にどういうエージェントが属しているかに関する情報と、環境において共通の型システムを含む。また、エージェント間の関係は、それらのエージェントが共有するオントロジィと交渉の戦略からなる。このシステムのためのプログラミング言語として、CAPL（capsule language）とENVL（environment language）を考えている。CAPLは単純エージェント用の言語であり、ENVL は環境を記述するための言語である。

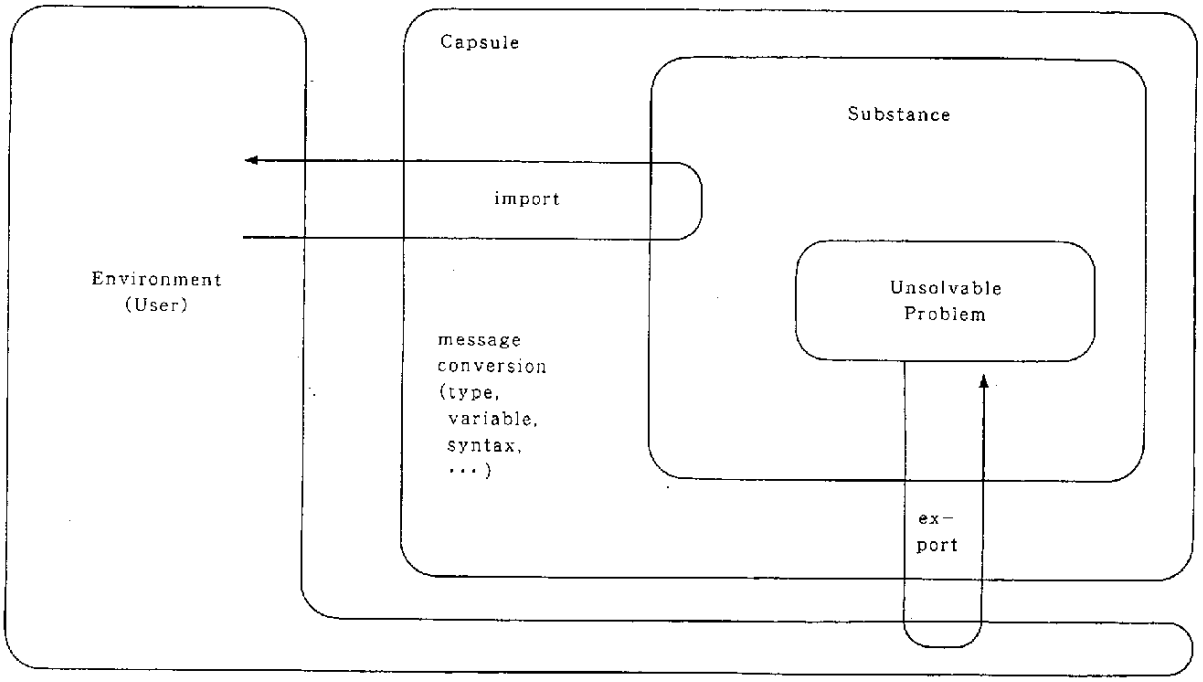
応用としては、現在、n乗問題、相撲、野球のフリーエージェント問題、heterogeneous ファイルシステムなどを扱っているが、将来は法的推論、自然言語処理、遺伝情報処理などの応用も考えたい。

このシステムは Knowledge Information Processing、Distributed Problem Solving、Multi Agent、Multiple Database などの既存の分野と関連する。たとえば Multiple Database においてはデータベース間の交渉がない、Multi Agent は一様なシステムしか扱っていないなど、これらのアプローチは一般性に欠けるという性質がある。

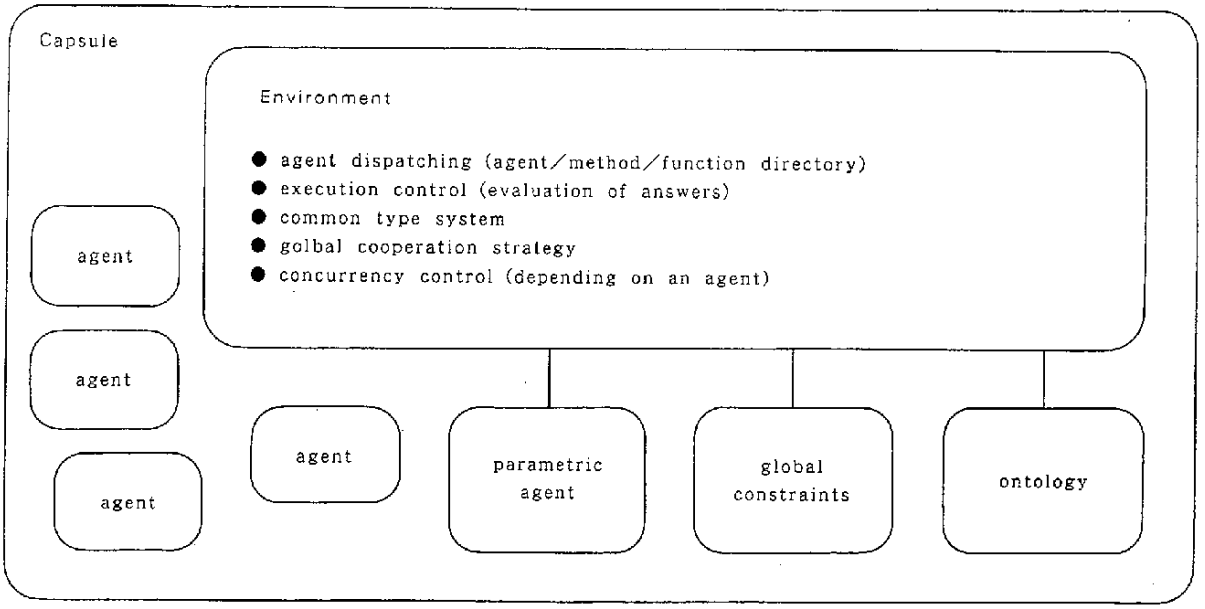
【質疑応答】

質問１：エージェントの知識の粒度に関する問題は発生しないか。

回答１：常識の扱いについては、必要に応じて他のエージェントを導入することで対処している。



OHP－1 : Simple Agent



OHP－2 : Complex Agent

質問２：知識ベースとのコミュニケーションの問題、たとえば、知識ベースの更新はどうか。
回答２：同期はとっていない。

質問３：Ｋａｐａ、ＱｕｉｘｏｔｅとＣＡＰＬ、ＥＮＶＬとの関係は。
回答３：ＣＡＰＬとＥＮＶＬは、基本的に新しい言語である。

質問４：共有型の利点は何か。
回答４：ＭＬの型システムのようなものに基づいてメッセージ通信ができる。

(6)「知識工学における言語的ツール」

Free University Amsterdam, Dept. of Mathematics and Computer Science（オランダ）
Professor　Reinr van de Riet

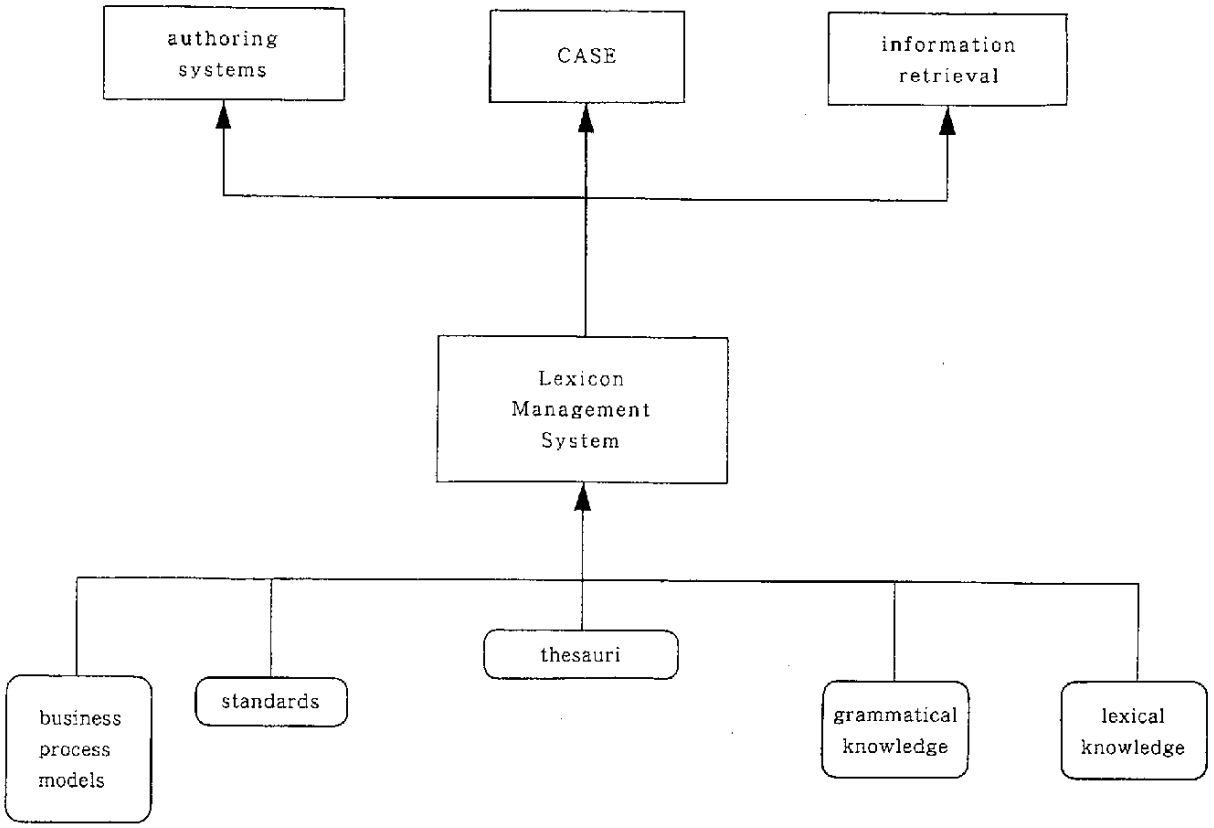
【発表要旨】

LIKE (Linguistics Instruments for Knowledge Engineering) は言語学と経営学と計算機科学を共通の辞書 (lexicon) のもとに統合しようというプロジェクトである。

また、MOKUN (Manipulating Objects with Knowledge and Understanding) は、Amsterdamの別名でもある。

LIKE では、言語学と経営学は言語行為論を介して、経営学と計算機科学はグラフィカルインタフェースを介して、計算機科学と言語学は機能文法を介して結び付いている。また、言語学は辞書 (dictionary) に関して、経営学はオフィスオートメーション／コミュニケーションに関して、計算機科学は MOKUN を介してこの LIKE に関わっている。

このプロジェクトにおいて解決したい基本的な問題は、言語の辞書と知識ベースシステムにおけるさまざまなツールとをいかにして結び付けるか、という問題である。ここで重要なのは、各々のツールがユーザによって定義されたオブジェクトを扱い、そこに現われる概念がそれぞれ一定の意味を担っているということである。たとえば、CASE ツールに関して考えみると、デザイナは何らかの問題を解く上で助けになるような設計をするわけだが、そこでデザイナはさまざまな図を描くツールを用いるとする。そこで、箱とか矢印などを識別するために言語を用いるだろう。その際、たとえば単語がその使用の状況に合致するようにその言語の使用を支援するような知的ツールがあれば便利である。単語の意味が辞書に書かれており、その意味は箱とか矢印などのオブジェクトを表するものである。また、それらの単語は「学生が本を借りる」などの知識を介して他の単語と関係付けられている。CASE ツールがこうした言語学的な意味と関連付けられるときに限り、辞書は有用なものとなる。同様のことが情報検索に関しても言える。実際この問題は非常に一般的な問題であり、多くの技術的課題を含んでいる。



OHP－1：An integrated Research Proposal

The framework for the LIKE project

【質疑応答】

質問１：ＣＹＣのような知識ベースの利用を考えているのか。

回答１：使いたいと考えている。しかし、この方法は、通常の辞書から出発し、次第に拡張しようというものである。

質問２：たとえば、辞書の意味が循環しているとか、矛盾しているという問題に対して、言語学の知見はどう役立つか。

回答２：そのような問題に対しては、計算機科学の方法が有効だと考える。

2. セッションII

データベースから知識ベースへ

2. セッションII：データベースから知識ベースへ

(1)「大規模知識ベースを管理するためのデータベース実装の適用について」

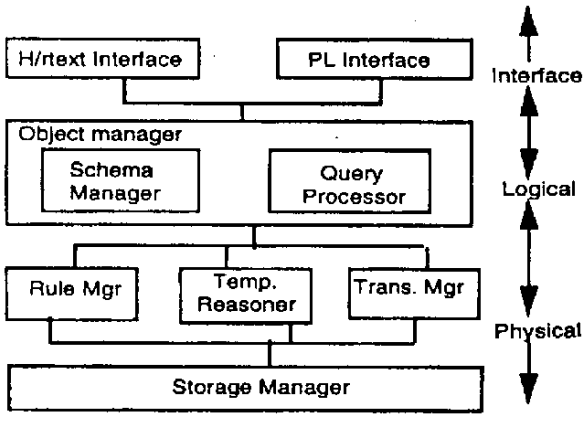
University of Toronto, Department of Computer Science（カナダ）

Professor　John Mylopoulos

【発表要旨】

超大規模知識ベースを管理するために従来から開発されてきたデータベースシステムの実装技術をいかに有効利用するかについて論じる。カナダのトロント大学の The Telos Knowledge Base Management プロジェクトで過去数年間に渡って推進されたきた研究報告である。特に、論理的／物理的な記憶構造や、与えられた知識ベースシステムに対する効率的な問合せ処理アルゴリズムを、従来のデータベースシステム技術をベースとして開発する方法論を提案する。

これは、超大規模な知識ベースシステムを開発するための方法として、システムを階層的に構築する方法である。下図がそのシステムアーキテクチャである。



ＯＨＰ－１：知識ベース管理システム・アーキテクチャ

【質疑応答】

質問１：システムを多層の分けて設計すると、最下位と最上位の間で、お互いの内容が把握できなくなり問題がないか。

回答１：指摘された問題は確かにある。しかし、各々の層が共通の土台（たとえば、オペレーティング・システム）のもとで設計されるのであれば、１つの包括的なシステム設計が可能であり、システムの最適化も可能と考える。

質問２：実際に構築したシステムではどうか。

回答２：ここで紹介したアルゴリズムなどは提案の段階であり、今後、実システムのもとで評価していく必要がある。

(2)「オブジェクト指向および能動データベースからの知識獲得」

Simon Fraser University, School of Computing Science（カナダ）

Associate Professor　Jiawei Han

【発表要旨】

大規模データベースから興味深い知識を抽出することは、データベースおよび知識ベースの開発にとって今後益々重要になるテーマである。現在までに、関係データベースを対象とした知識獲得の研究については、ある程度の成果が得られている。これらの従来の結果をベースとして、最近、データベースシステムの応用分野の拡大にともなって特に注目を集めているオブジェクト指向データベースおよび能動データベースを対象とした知識獲得の研究を行なうことは有意義なことである。本発表はこの分野における最近の成果をまとめたものである。

まず、データベースからの知識獲得の技法としては、従来関係データベースを対象として、属性指向の一般化アルゴリズムをオブジェクト指向データベースに拡張する方法がある。つまり、オブジェクト指向データベースシステムの特性である複合オブジェクト、メソッド、クラス階層に関する一般化アルゴリズムの提案である。

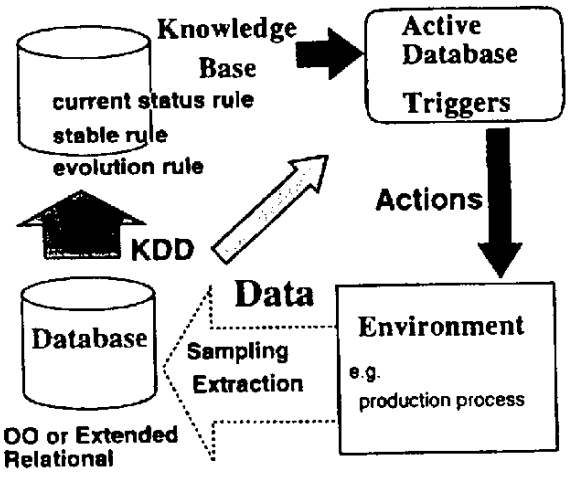
データベースの内容が動的に更新されていくシステム環境を反映した知識獲得技法を、特に能動データベースシステムと関連づけて提案する。獲得される知識を、現時点のデータから得られる一般化規則、定常的であまり変動しないデータから得られる一般化規則、いくつかのデータサンプル時間に渡って変動していくデータから得られる一般化規則に分類し、能動データベースに格納する方法である。さらに、動的にシステム状態が変動する環境下では、能動データベースのトリガー機能と知識獲得アルゴリズムの実行をいかに有効に融合するかが課題である。現在、生産プロセス制御システムを例にとりその方法を研究中である。

【質疑応答】

質問１：ここでいう能動データベースとはどのようなものなのか。

回答１：ＥＣＡルールに基づいたものであり、トリガー機能を備えたデータベースである。

質問２：データベースに非常に低レベルのデータが入っていて、能動データベースに記述されていないルールが概念的に高いレベルであるとき、迅速な応答ができないという問題はないか。



OHP－1

回答２：高レベルのルール記述に迅速に対応できるためにも、低レベルのデータからの知識獲得をおこない、高レベルのデータにすることが重要である。

質問３：対象としているオブジェクト指向データベースは標準的なものなのか。

回答３：現在、オブジェクト指向データベースの数学的な定義はないが、一般的な要件を挙げることで大体の合意はできている。本研究では、その要件に基づいたデータベースを仮定している。

質問４：一般的に、知識獲得アルゴリズムの計算の複雑度はどんなものか。また、関係データベースを対象とした知識獲得システムＤＢＬＥＡＲＮを実行したときのデータサイズはどの程度のものなのか。

回答４：計算のオーダーは、Ｏ（ｎ・ｌｏｇ ｎ）である、ただし、ｎはデータの個数でソーティングの計算オーダーと同じである。たとえば、テラバイトのデータがあるとすると膨大な計算量になる。しかし、通常は、データベースのある対象領域を限定して知識獲得アリゴリズムを実行することになり、単にデータベースサイズのみで計算量を計ることは難しい。
これまでにおこなった実験では、２．５メガバイト（カナダの科学研究費のデータベース）位のサイズで、Sun SPARC Station 2 で応答時間は２０秒であった。

質問５：採用しているサンプリング技法は何か。

回答５：Progressive Sampling 法である。つまり、サンプルして得られたデータに対して一般化を行った結果、あらかじめ決められたルール記述、制約などを満たさないような可能性が高い場合、さらに多くのデータサンプルを取って精度を上げる方法である。

3. セッションIII

知　識　表　現

3. セッションIII：知識表現

(1)「柔軟な知識表現のためのコンテキストリフレクション」

電子技術総合研究所 協調アーキテクチャ計画室
室長　中島秀之

【発表要旨】

大量の知識を扱う問題の中で重要となる柔軟性と効率の問題について論じる。すなわち、コンテキストを利用し、以下の課題解決策を提案するものである。

①コンテキストリフレクションを導入することによって関連する知識が存在する環境を特定して推論の効率を向上させる。

②環境の変化に伴ってコンテキストを切り替えることによって変化に柔軟に対応することができるメカニズム。

【質疑応答】

質問１：＜＜time, 4, JST＞＞　(抽象化)→　＜＜time, 4＞＞とするとき、どの変数（項目）を選んで抽象化するかをどのように決定するのか。

回答１：変数の役割はあらかじめ決められているのでコンテキストとの対応づけによって決定する。

質問２：IJCAI'93 の J. Mearthy の結果と比較してほしい。

回答２：J. Mearthy の結果は古典的な論理の中での定式化だが、私のは新しい（古典的論理の枠の外）での違う。

質問３：効率の向上の実現は関連するルールの数を減らすことができたからだということに気付いているか。

回答３：コンテキストをしぼると知識の数が減ることは分かっている。

```
ex1:    japan : : time (4, pm)
                world : : time (4, pm, JST)

ex2:    part time 4WD
        PT4WD / 4WD : : lever-change → out
        PT4WD / FF    : : lever-change → out
```

(2)「大規模知識ベースの構造化におけるオントロジィの役割」

University of Twente, Department of Computer Science（オランダ）

Professor　Nicolaas J.I. Mars

【発表要旨】

大規模知識ベースの開発におけるオントロジィの役割に関して３つの事例研究を紹介する。

事例１：工学分野における設計のためのオントロジィ：YMIR

事例２：セラミック科学のオントロジィ

事例３：測定単位のオントロジィ

本研究から以下のことが分かった。

①Simmons が指摘した認知的要素が関連するということは全くなかった、このことは、Simmons は常識を対象としていたからで、我々は工学における理論と実践を対象としたので、それは適用できないのであると考えている。

②オントロジィは何を別のものとして区別すればよいかを教えてくれるので有用である。

③既存の概念の再利用が可能であることが分かった。

④測定単位のオントロジィは予想に反して困難であった。

⑤システム理論は設計オントロジィの設計に極めて有用であった。

⑥ドメインのオントロジィの設計は、共有・再利用可能な大規模知識ベースの構築の最初のステップとして重要である。

【質疑応答】

質問１：オントロジィは知識獲得に利用できるか？

回答１：この事例においても利用している。

質問２：オントロジィは知識表現に依存するのか。

回答２：依存しない。知識レベルで設計するので依存しない。

質問３：測定オントロジィの知識サーバでＫＱＭＬを使っているとのことだが、どのようにして利用しているのか。

回答３：evrail を通じて（ＫＱＭＬプロトコルで）アクセスしている。

質問４：「標準」自体がアドホックなのに驚いたとのこただが、どうしてそうでないと思っていたのか。

回答４：特別な「標準」と考えていた。

質問５：アドホックでない「標準」はないと思うが。
回答５：そのとおりだ。

(3)「ＫＱＭＬ：インテリジェントなエージェントの相互運用性のための知識問い合せと操作言語」

University of Maryland Baltimore County, Dept.of Computer Science (米国)
Professor & Chair　Tim Finin

【発表要旨】

Knowbot のような知的で自立して活動するディジタル図書館などに代表される分散環境でのエージェント間の通信プロトコルを提案する。

ＫＱＭＬ (Knowledge Query and Manipulation Language) は、構文、意味、プラグマティックスの３つの層からなる。

通信の内容は、現在 S- 表現であるが、近い将来は KIF とか OMG が用いられる。内容の外側にはメッセージ performative ができる。

たとえば以下のとおりとなる。

```
( ask ( geoloc  lax (?long ?lat ))
        : number-answers 1
        : ontorogy geo-model 34)
```

ask が performative で、それ以外にreply、tell、～等がある。

(geoloc lax (?long ?lat)) が内容で、: ontorogy 等はオプションの引数である。

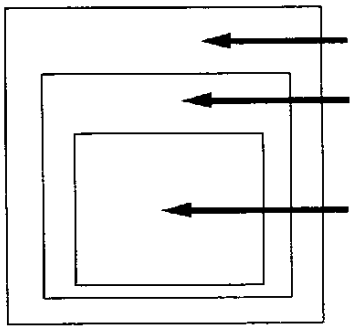
この例は答えは１つで、それをgeo-model 34 のオントロジィを使って答えてほしいということを意味している。performative は拡張が容易にできる。

通信の相手のエージェントを探す単なるフロントエンドではないファシリテータと表現言語やオントロジィの相違が原因で通信できない。そのため、エージェントを助けるメディエータ等が用意される。

【質疑応答】

質問１：ファシリテータは単なるフロントエンドなのか。
回答１：単なるフロントエンドではなく知的なもである。

Logically, a KQML message is a *content expression* enclosed in a *speech act* inside a *communication packet*.



OHP－1 : KQML is a layered language

4. セッションIV

自然言語処理と辞書知識

4. セッションIV：自然言語処理と辞書知識

(1)「計算機科学、認知科学と概念科学：多言語知識ベースのための制約条件の利用」

MIT, Artificial Intelligence Laboratory（米国）
Professor　Robert C. Berwick

【発表要旨】

現在までの計算言語学的なアプローチでは言語毎に固有の文法と固有の語彙とを開発しなければならなかった。これに対して、言語を X-bar理論+Head移動+... といった種々のパラメータ付き制約によって理論化すれば、各単語についての意味記述を与える語彙を開発することにより解析や生成が基本的には可能となる。

ここで発表する方式では、種々の言語毎に固有の言語現象とされたものに統一的な説明を与える。例えば、教会に寄付することを、"He churched the money"と言えないのは、"church"という語のそのような移動が ECP (Empty Category Principle) に違反する移動であるので生起しないからであるというように説明する。そして、この ECP は全ての言語において成り立つ原則であり、日本語において考えてみると、

（正）　かべにペンキをまく
（誤）　かべをペンキでまく

の説明は、基本とする意味構造から上記の表層構造への変形過程における原則違反が下の文に起ることで説明が可能である。そして、まさに、同じ意味構造から英語の表層構造への変形過程での原則違反として、

（誤）　paint smeared on(at) the wall

も説明できる。

このように、日本語と英語に固有の非文性の判断を一つの語彙からのパラメータの異なる原則の言語への変形として捉えることが可能であり、この方式を採用すれば、辞書の開発効率は、２倍となる。

なお、現在、ＥＤＲの辞書に基づいてこのような辞書開発が可能であるかの実験を実施している。

【質疑応答】

質問１：英語の"wear"は、日本語では、「はく」「きる」「かける」などに対応する。こういう、一対多対応となる単語はどのように扱うのか。

回答１：まだ、システム的に対応できるといいきる程の実験を行なっていないので、基本的な提案として紹介した。指摘されたような問題を考えると、２倍というのは言い過ぎかもしれない。

質問２：厳密に対応する単語がない場合などはどうするのか。（たとえば、日本語「なつかしい」←→英語「?????」）

回答２：基本的に問題となるのは、verb と verbal nouns だと考えている。従って、完全に一致する率は心配されるより多いと考える。

質問３：文レベルでの対応はどういうものを仮定しているのか。実際の翻訳では、原文の１文が翻訳結果では２文以上になったりする例がある。

回答３：今回の発表では簡単のため、１対１で文が対応すると仮定して理論化した。

(2)「テキストコンパイラと概念タグのついたコーパス」

㈱日本電子化辞書研究所 第６研究室

室長　安原　宏

【発表要旨】

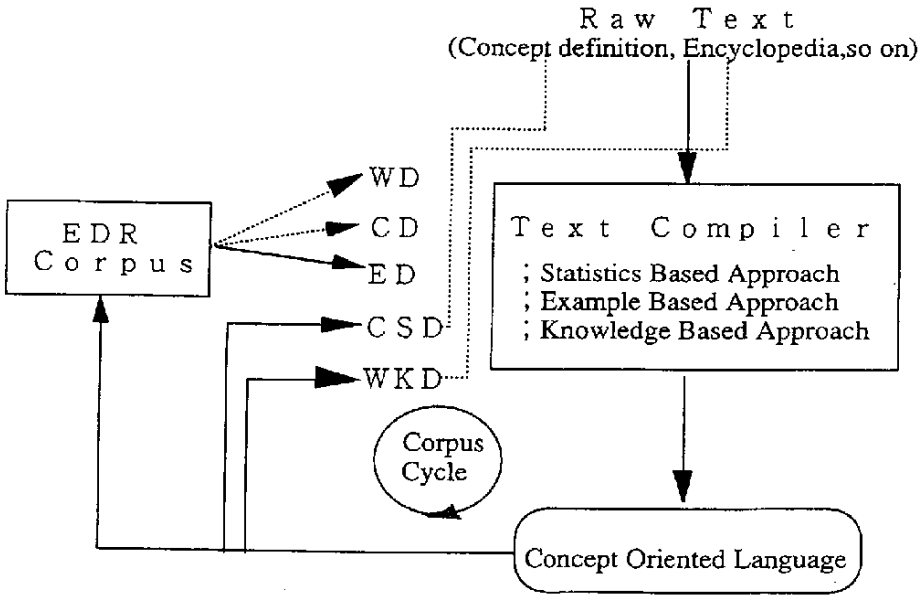
自然言語で書かれた種々の文書は人類の叡智の結晶である。それを機械処理可能な形に変形し表現することが今後の重要な研究テーマである。この発表は、概念指向言語というアイデアによりそれを実現する試みである。

ＥＤＲでは、語義からの抽象という形で概念というものを持ち込んだ。この概念を利用すると、もともとは曖昧性を持つ「文」から内容的曖昧性を排除した表現を付随させることができる。この表現は、次のような優れた特徴を有する。

①InterLingua(中間言語)となりうる
②概念体系の張る概念空間が利用可能となる
③概念体系と記述により分析や統合が可能である
④形式的な操作(operation)が可能である

以上のような特徴は、機械翻訳などの自然言語処理や知識処理において有用なものである。

生テキストに上のような概念指向表現を付随させる仕組みとしてテキストコンパイラというものを設計した。基本的には次のように処理を進める。まず、入力文は、統計的手法を併用した格パターン解析を実施する。これには、コーパスから抽出した「例文辞書」と概念体系とを利用するのが特徴である。続いて、照応や文間関係の解析をコーパスから抽出する知識を応用して実施する。さらに、ＥＤＲ辞書の概念説明を基に作成する「常識辞書」とでもいうべきものによる選択・補充を行なう。これに、必要ならば「世界知識辞書」による言語外知識に基づく処理を行なう。そして、この結果自身は、例文辞書・常識辞書・世界知識辞書の新たな項目となるのである。



OHP－1

以上の構想を進めることは、ＥＤＲ辞書を基盤に共有可能な知識ベースを構築するのに有効でかつ実現容易な方式である。

【質疑応答】

質問１：ＣＤ風のものを InterLingua にしているが、それを何に応用するつもりなのか。

回答１：ふたつの応用を考えている。ひとつは、Anaphor の解消などの処理への応用、 もう一つは、機械翻訳への応用である。

質問２：語彙（Vocaburary） の InterLingua はどのようなものを使うのか。たとえば、"born" の decompose はどのように行なうのか。

回答２：対話システムであれば、そういった細かな記述も必要となるであろうが、現在、応用テーマとして考えているのは機械翻訳なので、そこまで細かな分析は考えていない。

質問３：ＥＤＲ辞書を利用しているが、規模や粒度は十分か。

回答３：ＥＤＲ辞書は概ね必要十分な規模になっているように思う。また、ユーザの側で拡張可能な仕様になっているので調整がきく。

(3)「機械可読辞書からの知識ベース抽出」

Vassar College, Department of Computer Science（米国）
Associate Professor　Nancy Ide

【発表要旨】

この発表は、この１０－１５年に実施されて来た辞書からの知識抽出の研究が無駄であったかどうかを検討しようというものである。

最近、コーパスから知識を抽出しようとする研究の方に人々の興味が移りつつあるように見える。また、機械可読辞書から実際に知識ベースを構築したという例もまだない。では、どこかが間違っていたのであろうか。

辞書からの知識抽出研究は、次の２つのことを前提としている。(1) 機械可読辞書には自然言語処理に有用な情報が含まれている。(2) 機械可読辞書から簡単に情報を抽出することができる。

まず、辞書から概念体系を構築するという例題で、(1)の前提を検証実験を実施した。結果は、(a)体系が元となる辞書によって異なってしまう、(b)ループができてしまう、(c)上位語の記述がないことがある、(d)上位語が適切と思われるものより上すぎるというような欠陥が全データの半分以上を占めた。

次に前提 (2)については、(A)辞書の物理フォーマットが曖昧である、(B)上記が必要以上に難しい、(C)Meta Text が整合的でない、(D)語釈文が曖昧な場合があるなどの欠陥があった。

しかしながら、(1)複数の辞書からの抽出した情報をうまく組み合せれば、精度もエラー率５％程度まで向上可能となる、(2)今までの研究で実際の情報の枠組やモデルが定まってきたなどの貢献があったということを忘れてしまってはならない。

【質疑応答】

質問１：発表の意図が分からない。

回答１：主張したかったのは、ＭＲＤ（機械可読辞書）からの知識抽出研究は有意義であって、改良することで実用になるということである。われわれは、今のコーパスから知識を抽出しようという研究者もいずれまたＭＲＤに戻ってくるとさえ考えている。

質問２：Lexicographer に tool を与えるという方がいいのではないか。

回答２：わたしたちもそれはいい方法であると考えている。

質問３：Consistency は難しい問題になるだろう。辞書の問題というよりも語義（word sense）というアイデアそのものに問題があるように思うがどうか。

回答３：Sense distinction に必要なレベル程度ということで考えているのでそこまで難しい問題とはならない。

(4)「頻度情報つき機械翻訳用辞書の開発」

日本科学技術情報センター 技術開発部
副主任情報員　芦崎達雄

【発表要旨】

JICST では、1986年に機械翻訳システムを開発した。それにより1990年から科学技術文献の要旨の翻訳とそのデータベース化を実施してきた。その成果として、英日対応コーパスと頻度情報付き辞書が作成された。

JICST の翻訳システムにおいては、日本語単語と読みに英単語と分野コードおよび頻度情報を付けた用語対応辞書（correspondence dictionary）を使っている。この辞書と名詞翻訳辞書・動詞翻訳辞書とを翻訳の基礎データに利用している。名詞翻訳辞書は、日本語単語に発音・詳細品詞・語義分類・英単語・分野コードを付けたものである。動詞翻訳辞書は、名詞辞書に記載の情報の他に格パターン・態・相・自発・表層トランスファーパターンも記載される。これらの辞書は実行時には属性辞書・規則辞書・形態解析辞書・形態生成辞書にコンパイルして利用する。

以上の翻訳結果を基に、英日対応コーパスを作成している。ほぼ、一対一で文同志が対応する形式になっている。そのため、例文ベースの翻訳に利用可能である。

また、上記のコーパスから単語切り出しを実施して、対応辞書を拡充するとともに、実際の頻度を付したものを作成している。

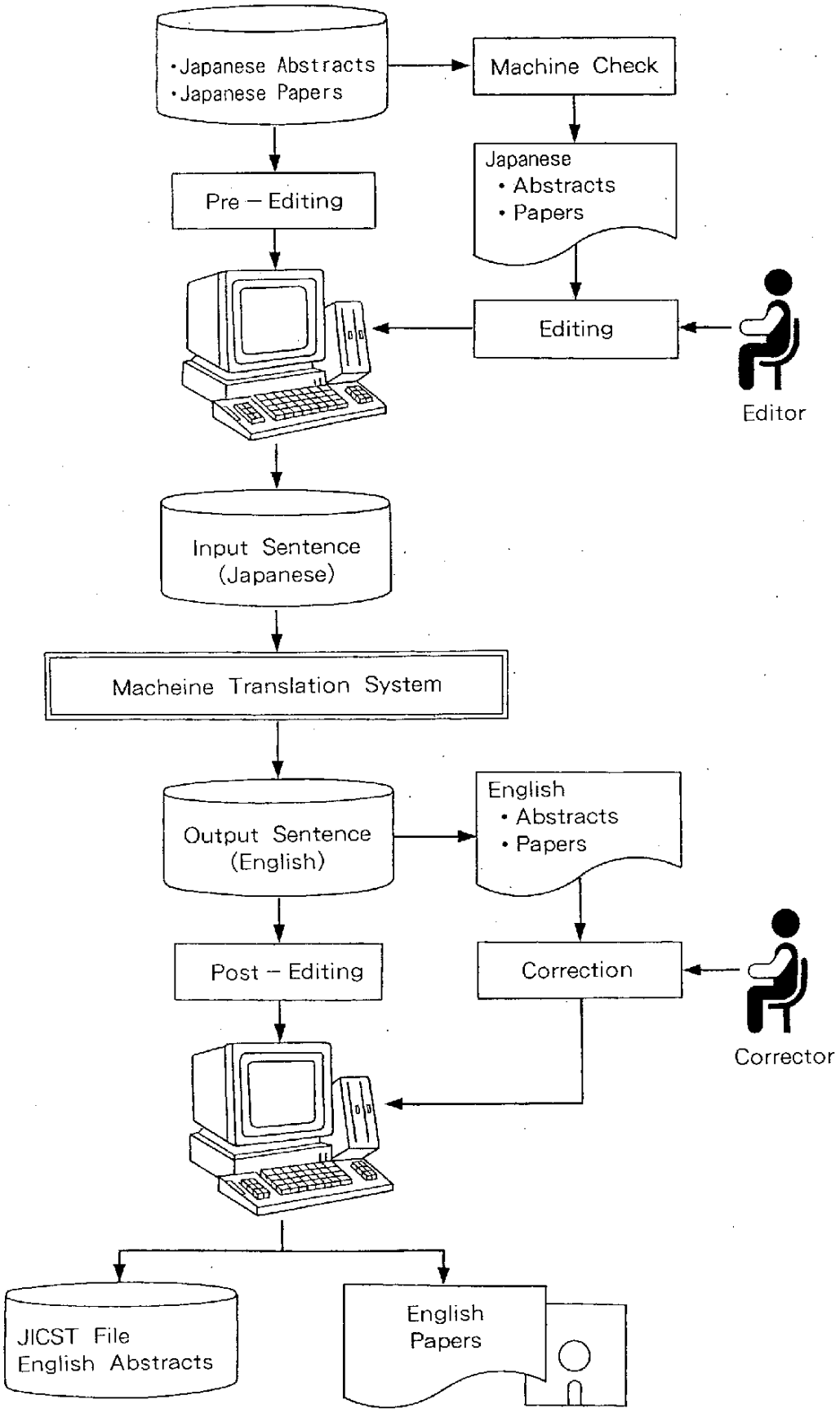
【質疑応答】

質問１：JICST の Bilingual Corpus と Frequency Dictionary の availability は？

回答１：A1 Bilingual Corpusは、1994年中にリリースしたい。Frequency Dictionary は、検討中である。

質問２：中頻度の語が人間にとって有用なキーワードとなりそうに思うが、そういう経験則は成り立つだろうか。

回答２：科学技術庁の２０万のテクニカルワードをキーワードに利用しているので統計をとっていないが、そういう傾向はあるかもしれない。



OHP－1 : JICST Machine Translation System (J/E)

FREQUENCY	'90	'91 (RATIO TO '90)	'91+'92 (RATIO TO '90)
10,000 ～	92	169 (1.84)	442 (4.80)
5,000 ～ 9,999	165	272 (1.65)	560 (3.39)
1,000 ～ 4,999	1,346	1,952 (1.45)	3,404 (2.53)
500 ～ 999	1,371	1,956 (1.43)	3,294 (2.40)
200 ～ 499	3,346	4,705 (1.41)	7,893 (2.36)
100 ～ 199	4,425	6,394 (1.44)	9,528 (2.15)
50 ～ 99	6,923	9,628 (1.39)	14,262 (2.06)
20 ～ 49	14,283	20,273 (1.42)	26,482 (1.85)
10 ～ 19	16,148	21,612 (1.34)	25,174 (1.56)
5 ～ 9	20,212	25,996 (1.29)	28,362 (1.40)
1 ～ 4	44,458	50,402 (1.13)	45,240 (1.02)
TOTAL	112,769	143,359 (1.27)	164,641 (1.46)
NUMBER OF WORDS IN DICTIONARY	345,692	364,660 (1.05)	364,660 (1.05)
NUMBER OF ABSTRACTS (JAPANESE)	213,737	318,258	640,684

Table.1.Frequency Count

用語Ｎｏ	M010013166
更新年月日	19910528
頻度	4
日本語見出し語	アイウエオ順
読み	アイウエオジュン
異形語	あいうえお順
(1)分野コード	AA00
英語見出し語 (1)	ａｌｐｈａｂｅｔｉｃａｌ　ｏｒｄｅｒ
用語Ｎｏ	M010106450
更新年月日	19901128
頻度	2
日本語見出し語	愛煙家
読み	アイエンカ
(1)分野コード	AA00
英語見出し語 (1)	ｈａｂｉｔｕａｌ　ｓｍｏｋｅｒ

Fig.2. Frequency Dictionary

OHP－2

5. セッションV

大規模知識ベースの作成支援と応用

5. セッションV：大規模知識ベースの作成支援と応用

(1)「知識構造の超並列マッチング」

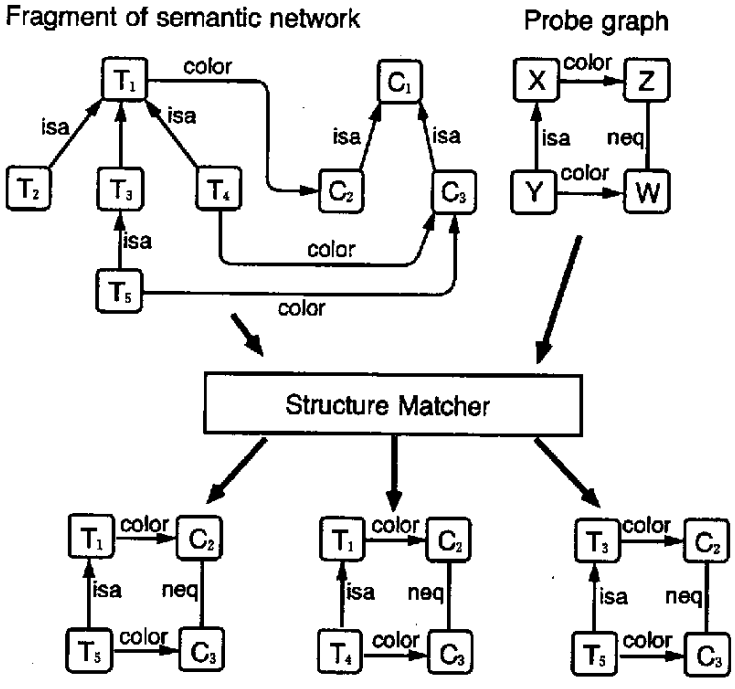
University of Maryland, Department of Computer Science（米国）

Associate Professor　James A. Hender

【発表要旨】

知識ベースが大規模化すると必要なデータを速く引き出すことが困難となる。知識ベースにおいては、データベースに比べて、(1) inheritance、(2) inference、(3) classification 、(4) recognition 、(5) structure matching といったことが必要となる。これらのうちで、一番問題となるのは(5)である。知識ベースにおいては、structure matching の structure 中に既知でない要素を含みうるので、データベースにおける indexing といった方式では解決できない。

そこで、超並列をベースとした知識ベースの高速マッチングの方法を提案する。今回の提案では、PARKAを知識表現方式として仮定した。PARKA は、一項と二項の関係を表現の基本とする汎用のフレー



OHP－1

ム型知識表現であり、高速な継承機能が備わっている。PARKA の semantics は基本的には KL-ONE と

同じものである。

これに今回提案したアルゴリズムを CM-2 という SIMD 型の Massively Parallel Computer で*Lisp によりインプリメントした。

その結果は、ワーストケースで、

(a) 単項 P(A)　k=0.1sec で o(1)

(b) $\wedge_{i=1}^{m} Pi(A)$　o(m)　　m=22で約1秒

という結果を得た。

上の結果は、本方式が知識ベースの良い retirieving 方式となることを示している。

【質疑応答】

質問１：ベンチマークには何を使用するのか。

回答１：知識ベースは法的問題がクリアーでないので困っている。ベンチマークに使える知識ベースを持っている人は教えて欲しい。

質問２：他の知識ベースを応用に取り込むことは考えないのか。

回答２：今はまだ、generic tool にこだわっている。

質問３：データベースでの経験から言うと特殊な計算機を利用していることが気になる。

回答３：並列のアルゴリズムの研究と見て欲しい。また、知識ベースでは入力の方が出力より難しいので、データベースとは逆になっている。 従って、特殊な計算機でもサーバとして利用できるのではないか。

(2)「知識指向工学を目指して」

東京大学 工学部総合試験所

助教授　冨山 哲男

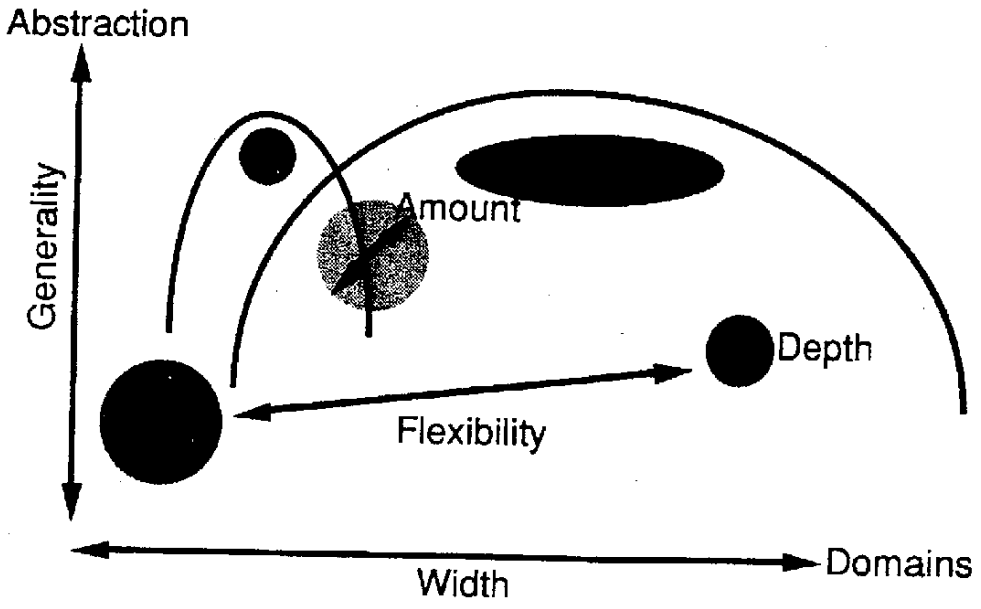
【発表要旨】

Intelligent Integrated Interactive Computer AidedDesign (IIICAD) システムを作成して、実際に試用してもらっているので報告する。

IIICAD は、Knowledge Intensive Engineering のひとつの例として作成したものである。この Knowledge Intensive ということばは、ただ知識を増やすのではなく、もっと柔軟な知識を増やすことこそが有用であるということを表現するために用いた。単なる知識の量だけでなくその深みも重要であ

り、それら２軸の和で intensiveness は測られる。

ＣＡＤにおいては、メタモデル 、質的モデル、 量的モデルという３つのモデルが利用されている。これらを 10,000 程度集めてある。それらは、3000 の chunk に独立のマイクロセオリとしてシステム化された知識の形で収集されている。それらは１つのオントロジィを共有している。



ＯＨＰ－１：Knowledge Intensiveness

現在では、それらのマイクロセオリ間のコミュニケーションはないが、統一したフレームワークの下での知識共有アーキテクチャにまとめていく予定である。

【質疑応答】

質問１：デザインは、どの段階から始めるのか。

回答１：機能デザインから始めるものと考えている。その後で、システムの助けによりデザインを進めていく。

質問２：学生を使って知識を増やすことができないというが、なぜか。

回答２：デザインを実際に行うことで知識獲得する仕組みになっているからである。

質問３：どんなデバイスが用意されているのか。

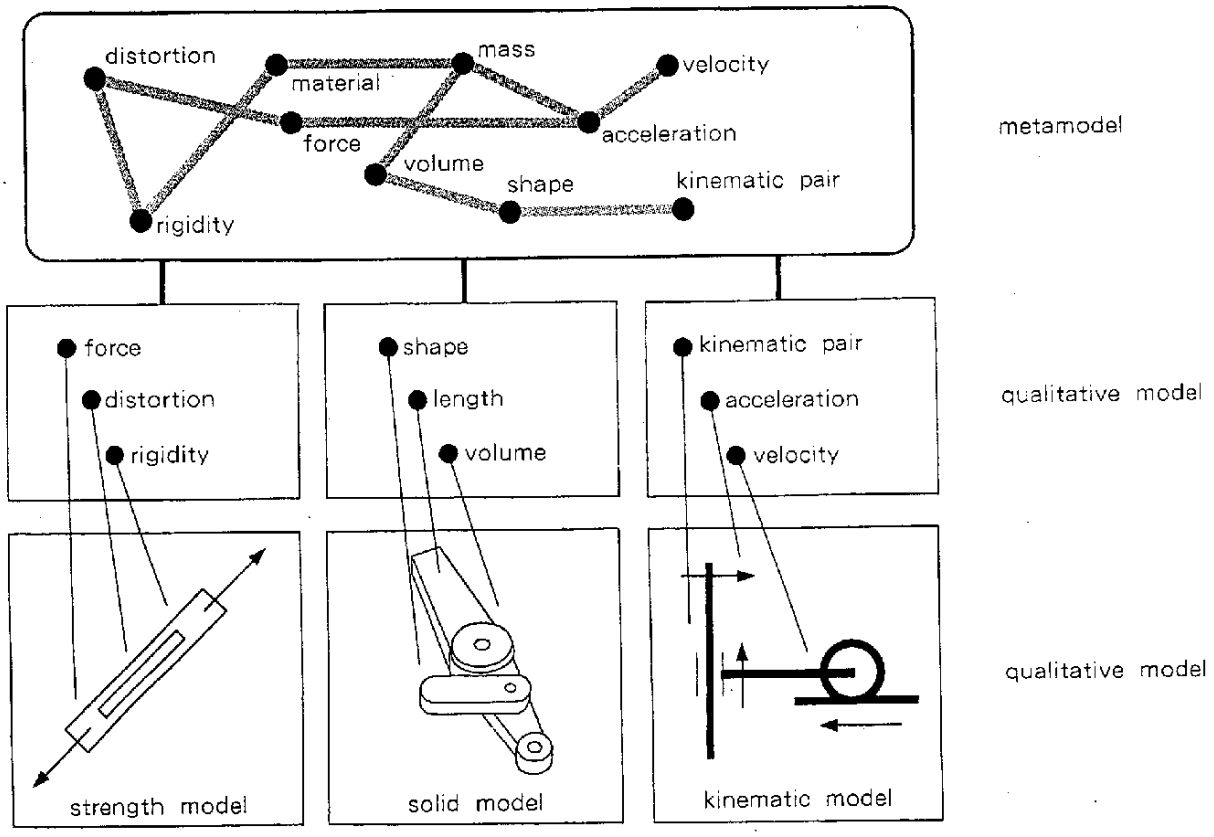
回答３：現在は、メカトロやキネメカが中心である。

質問４：どのようなシステムの上で動作するのか。

回答４：スモールトーク系の言語が動けばなんでもいい。

質問５：各国の標準の差異などは考慮するのか。

回答５：まだ、そういったレベルのことまではサポートしていない。



OHP－２：**Metamadel Mechanism**

(3)「分子生物学における大規模知識ベースの構築と共有」

INRIA（仏国）

Director of Research　Francois Rechenmann

【発表要旨】

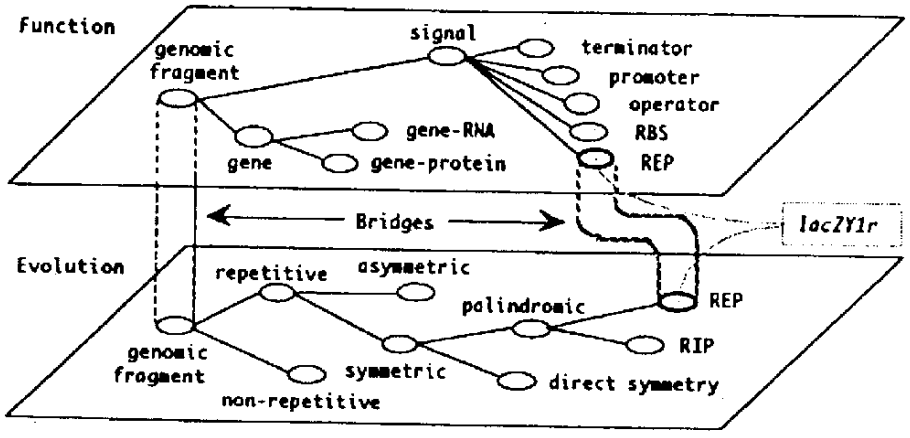
生体組織の遺伝子構造の分子レベルでの分析に利用される知識ベースを作成した。そこには、数値データとシンボルデータが混在し、種々の実体や構造が関係しており、複雑な構成となっている。規則数は、100,000 弱（94,000）である。

上記の知識ベースにおいては、下記のような要求項目があった。

1) Multiple View on entity and structure
2) Dynamic Relation
3) Task Decomposition and Specialization
4) Annotation

これらの要求項目を満たす形で、incremantal に知識ベースを構築した。

Multiple View on entity and structure を実現するために、Function と Evolution といった層をいくつか用意し、それらが、同一のオブジェクトを共有するという表現形式を用いた。そして、各層毎に、single-inheritance の hierarchy を定義することで、View の異なりによる Multiple inheritance の問題を解決した。



OHP－1

また、method と呼ぶソフトウェア的なモジュールを導入することで、数値データとシンボルデータの混在下での処理を実現するとともに、Task decomposition hierarchy を導入し、それにより入出力

属性を分類した。また、オブジェクトの属性間のリンクによりDynamic Modelの変更を表現した。

Annotation は Hyper-Text 方式で利用できるようにした。Incremental な構築を並列に実行するために、個人毎の personal Knowledgebase をうまく貼りあわせることができるような仕組みを導入した。これにより、うまく大規模化することができた。

【質疑応答】

質問１：再利用（Reuse）は可能か。

回答１：最初は、再利用可能ではない。利用者が注意深く検討した上でなら可能である。しかし、実際には、作成過程で作られる理論の方が重要であるため、直接、再利用可能とはならないのが現実である。

6　パネル・ディスカッション

ＫＢ＆ＫＳの応用とブレークスルー

6. パネル・ディスカッション：
ＫＢ＆ＫＳの応用とブレークスルー

6.1 パネリスト

座　長 ：	北 野 宏 明	ソニー コンピュータサイエンス研究所 リサーチャ
パネリスト：	James A. Hendler	Associate Professor, Dept. of Computer Science
		University of Maryland (米国)
	Mary Shephard	Menmber of CYC Project, MCC (米国)
	西 田 豊 名	奈良先端科学技術大学院大学 教授

6.2 パネル・ディスカッション　－ 要　旨 －

パネルは座長である北野氏の方針で、２つの質問にパネリストが答えを提示し、その後に、フロアとの議論を行なうという形式で実施された。

【質問１】ＫＢ＆ＫＳの応用展開はどういう方面になると考えるか？

[Mary Shephard]

(1) advice service （相談業）

(2) Online brokering on goods and services
（商品やサービスをネットワークで提供する）

(3) User Modellingに基づくサービス

・電子メールのフィルタリング

・Direct Marketing

・Corporate and asset management

・Enriched Aritificial Reality

・Video clipなどの情報へのアクセス

[James A. Hendler]

DB中からデータを引きだすのにKBの手法を利用するという分野

例えば、

・DRUG Company（製薬会社）のシステム

・Geographic Information Systems

などであろう。

[西　田]

日常生活で役にたつ知恵(Practical wisdom)を売る自動販売機のようなものができるであろう。

例えば、

・医療情報

・公式の場でのスピーチの仕方

などを売るものである。

[北　野]

ゲームでは「将棋や碁のソフト」、科学では「人ゲノムや脳地図の研究」、ロボット関連では「日常的な業務を代行するロボットや秘書ロボット」、社会基盤としては「音声翻訳など」に応用されると考える。

【質問２】ブレークスルーはどこでおきるか？

[西　田]

(1) Actual VLKB（実際にVLKBが存在すること）

(2) Common language, common ontology

（そしてそれが、共通の言語、オントロジィを持っていること）

という条件があれば、knowledge levelとSymbol level のどこにinterfaceを置くかによってブレークスルーが来る。

[James A. Hendler]

短期的には、二次記憶媒体の進歩による。長期的には、巨大な非構造的な知識ベースを探索することが必要である。

もっとも、すでに現在の internet は巨大規模のコーパスとなっている。

[Mary Shephard]

技術的な問題ではなく非技術的な問題の解決がブレークスルーの基となる。

例えば、

・政治的課題：共通のツールやフレームワーク

・財政的課題：長期の安定した Funding が必要である

・臨界量の突破

といったものである。

[北　野]

・不完全で不整合でノイズの多いデータの扱い

・知識ベースの創造・配布・利用の枠組み

・社会基盤の整備

・実時間処理

・分散協調処理

以下これを受けてフロアーと活発な討論があった。そのうち議論が多く行われたテーマは以下の２つであった。

(a) 既存のＤＢ研究の歩みは参考になるのか

研究の方向という意味ではなく、研究のための財政基盤を確立する手法は参考になる。データベースの研究者は、うまく他の分野の研究を取りこんで、それをＤＢユーザに自分達の成果としてうり込んでいるように見える。

このような方法をとれば Fund raising もできるのではないか。

(b) Application の広がり

(b-1) 現在の電話の利用法は、初期には考えられなかったものが多い。このように、ＫＢ＆ＫＳの応用システムも組織内での共有資産から始めて、それを互いに共有して広げていく方向が良いと思われる。当然、国際協力が必要となる。

(b-2) 子供の世代にはあたり前に利用されるものとなるように、Edutainment 的なものから始める手法が考えられる。

資　料

資料A.　　KB&KS'93 国際会議　－ 概 要 －

1. 目 的

本国際会議は、大規模知識ベースの整備の重要性とその構築・共有化技術の研究開発が人工知能を含む情報処理技術研究の最重要課題であるとの認識から、言語処理、知識処理などの情報科学の分野のみならず、社会・経済、人文科学の多方面から議論し、取組への緊急性と研究開発にあたっての広範な国際協力の重要性を世界の共通認識とすることを目的に開催された。

2. 実施要領

(1)主 催　財団法人日本情報処理開発協会

(2)後 援　通商産業省、文部省、科学技術庁
Ministère de l'Industrie et du Commerce Extérieur (仏)
Bundesministerium fur Forschung und Technologie (独)
Department of Trade and Industry (英)
National Science Foundation (米)
Commission of the European Communities

(3)協 賛　(社)電子情報通信学会、(社)情報処理学会、(社)人工知能学会
日本認知科学会、日本ソフトウェア科学会
The American Association for Artificial Intelligence
The Association for Computational Linguistics
The Association for Computing Machinery (Japan Chapter)
International Association for Machine Translation
The Europian Coodinating Committee on Artificial Intelligence

(4)会 期　平成5年12月1日（水）～2日（木）

(5)会 場　京王プラザホテル　南館5階　エミネンスホール
（東京都新宿区西新宿2-2-1）

(6)参加者数

① 国内／海外別参加者人数

国 内	海 外	合 計
383	49	432

② 国別参加者人数（参加国数：18カ国）

オーストラリア：1、オーストリア：1、カナダ：2、中国：1、
フランス：5、ドイツ：6、インドネシア：1、イタリア：2、

日本：３８３、韓国：１、マレーシア：１、ロシア：１、シンガポール：１、
スイス：１、タイ：２、オランダ：３、英国：４、米国：１６

(7)プログラム（講演テーマおよび講演者）

【１２月１日（水）】

開　会

10:00-10:30 主催者挨拶　影山衛司（(財)日本情報処理開発協会会長）
来賓挨拶　荒井寿光（通商産業省機械情報産業局次長）
来賓挨拶　加藤康宏（科学技術庁長官官房審議官）
来賓挨拶　Su-Shing Chen （Program Director, NSF）

10:30-11:15 基調講演１「幼年期から青年期へ」　渕　一博（組織委員長）

11:15-12:00 基調講演２「知の空間を構成する大規模知識ベース」
横井俊夫（プログラム・実行委員長）

12:00-13:00 （昼　食）

セッションⅠ：「社会的・学際的要請」座長　土屋　俊（千葉大学助教授）

13:00-13:30 講演「新しい経済的・社会的インフラストラクチャとしてのKB&KS」
今井賢一（スタンフォード日本センター研究所長）

13:30-14:00 講演「人間の知識の本性について」　藤澤令夫（京都国立博物館館長）

14:00-14:30 （休　憩）

セッションⅡ：「言語処理」座長　松本裕治（奈良先端科学技術大学院大学教授）

14:30-15:00 講演「言語処理の現状と将来動向」　長尾　眞（京都大学教授）

15:00-15:30 講演「解析・生成技術」　田中穂積（東京工業大学工学部教授）

15:30-16:00 講演「知識獲得の自動化を目指して」
Yorick Wilks (Professor, University of Sheffield, 英)

16:00-16:30 講演「言語処理のためのテキスト資源の収集と利用」
Susan Armstrong (Professor, University of Geneva, スイス)

16:30-17:00 講演「機械翻訳における知識処理」　辻井潤一（Professor, UMIST, 英）

18:00-20:00　レセプション

【１２月２日（木）】

セッション Ⅲ：「知識処理」座長　溝口文雄（東京理科大学教授）

9:00-10:00 講演「大規模知識ベースの共有法」　大須賀節雄（東京大学教授）

10:00-10:30 講演「知識共有：予測と課題」　William R. Swartout
(Professor, University of Southern California, 米)

10:30-11:00 講演「共有知識ベース：ヨーロッパの観点から」 Bob J. Wielinga
(Professor, University of Amsterdam, オランダ)

11:00-11:30 講演「知識表現とデータ」 Ronald J. Brachman
(Department Head, AT&T Bell Laboratories, 米)

11:30-12:00 講演「知識獲得とオントロジー」 溝口理一郎（大阪大学教授）

12:00-13:30 （昼 食）

セッション Ⅳ：「利用可能な大規模知識ソース」座長 西尾章治郎（大阪大学教授）

13:30-14:00 講演「学術情報サービスの将来像」
山田尚勇（学術情報センター教授・研究開発部長）

14:00-14:30 講演「研究開発領域における言語リソース：その課題と展望」 Antonio Zampolli
(Professor, University of Pisa, イタリア)

14:30-15:00 講演「ﾄﾞｷｭﾒﾝﾃｰｼｮﾝつき多目的電子化ﾘｿｰｽの
作成、保守、利用におけるTEIの役割」
Susan Hockey (Director and Professor, CETH, 米)

15:00-15:30 講演「Cyc ：知識共有の先駆け」 Douglas B. Lenat
(Director, MCC, 米)

15:30-16:00 （休 憩）

セッションV

16:00-18:00 パネルディスカッション「情報インストラクチャの構築と国際協力」

ﾊﾟﾈﾘｽﾄ： 渕 一博（東京大学教授）
Jacques Mathieu (Ministère de l'Industrie du Commerce Extérieur, 仏)
Christian Rohrer (Professor, Stuttgart University, 独)
Peter M. D. Gray (Professor, University of Aberdeen, 英)
Su-Shing Chen (Program Director, NSF, 米)
Brian Oakley (Director, Logica, plc, 英 (EC))

18:00 閉会挨拶 照山正夫 ((財)日本情報処理開発協会専務理事)

資料B.　　KB&KS'９３ 国際ワークショップ　－ 概 要 －

1. 目 的

本ワークショップは、KB&KS'９３ 国際会議に引き続き（９３年１２月３日～４日）に開催され、国際会議のオーバーオールの議論を踏まえ、データベース、言語処理、知識処理などの観点から、より技術的なテーマを議論し、大規模な知識ベースの実現のための意見交換を主な目的として開催したものである。

2. 実施要領

(1)主 催　（KB&KS'９３国際会議と同じ）

(2)後 援　（KB&KS'９３国際会議と同じ）

(3)協 賛　（KB&KS'９３国際会議と同じ）

(4)会 期

平成５年１２月３日（金）～４日（土）

(5)会 場　工学院大学 高層棟１１階 第５会議室、第６会議室

（東京都新宿区西新宿1-２４-２）

(6)参加者数

① 国内／海外別参加者人数

国 内	海 外	合 計
４９	４７	９６

② 国別参加者人数（参加国数：１７ヵ国）

オーストラリア：1、カナダ：２、中国：1、フランス：５、ドイツ：６、
インドネシア：1、イタリア：２、日本：４９、韓国：1、
マレーシア：1、ロシア：1、シンガポール：1、スイス：1、タイ：２、
オランダ：３、英国：４、米国：１５

(7)プログラム（発表テーマおよび発表者）

【１２月３日（金）】

開会挨拶　溝口理一郎（大阪大学教授）

セッションＩ：「知識共有 」

座長 Bob J. Wielinga (Professor, University of Amsterdam, オランダ)

服部文夫（NTT 主幹研究員）

9:00- 9:30 発表「知識コミュニティを目指して」 西田豊明

（奈良先端科学技術大学院大学教授）

9:30-10:00　発表「統合的ユーザ支援環境における知識共有：
応用、フレームワーク、インフラストラクチャ」
Robert Neches (Professor, University of Southern California, 米)

10:00-10:30　発表「コンテクスト：大規模共有知識ベースの実際問題」
Bob Jansen (Program Manager, CSIRO, オーストラリア)

10:30-11:00　(休　憩)

11:00-11:30　発表「大規模知識ベースの共有：ルール選択のアプローチ」
Knut Hinkelmann (DFKI, 独)

11:30-12:00　発表「大規模知識ベースの新しいフレームワーク：
データベースと制約論理プログラミングの観点から」
横田一正 (ICOT 主席研究員)

12:00-12:30　発表「知識工学における言語的ツール」　Reind van de Riet
(Professor, Vrije University, オランダ)

12:30-14:00　(昼　食)

セッションII：「データベースから知識ベースへ」
座長　Vadim L. Stefanuk
(Vice-chairman, Russian Academy of Science, ロシア)

14:00-14:30　発表「大規模知識ベースを管理するためのデータベース実装の適用について」
John Mylopoulos
(Professor, University of Toronto, カナダ)

14:30-15:00　発表「オブジェクト指向でかつアクティブなデータベースにおける知識発見」
Jiawei Han
(Professor, Simon Fraser University, カナダ)

15:00-15:30　(休　憩)

セッションIII：「知識表現」
座長　Ronald J. Brachman
(Department Head, AT&T Bell Laboratories, 米)

15:30-16:00　発表「柔軟な知識表現のためのコンテキストリフレクション」
中島秀之 (電子技術総合研究所室長)

16:00-16:30　発表「大規模知識ベースの構造化におけるオントロジの役割」
Nicolaas J.I. Mars
(Professor, University of Twente, オランダ)

16:30-17:00　発表「KQML：インテリジェントなエージェントの相互運用性のための知識問い合わせと操作言語」　Tim Finin

(Professor, University of Malyland Baltimore County, 米)

17:15-19:15　レセプション

【１２月４日（土)】

セッション Ⅳ：「自然言語処理と辞書知識」

座長　新田義彦（日立製作所主任研究員）

9:00-9:30　発表「計算科学、認知科学と概念科学：
多言語知識ベースのための制約条件の利用」
Robert C. Berwick (Professor, MIT, 米)

9:30-10:00　発表「テキストコンパイラと概念タグのついたコーパス」
安原　宏（EDR)

10:00-10:30　発表「機械可読辞書からの知識ベース抽出」
Nancy Ide (Professor, Vassar College, 米)

10:30-11:00　発表「頻度情報つきの機械翻訳用辞書の開発」
芦崎達雄（日本科学技術情報センター）

11:00-12:30　（昼　食）

セッション Ⅴ：「ＶＬＫＢの作成支援と応用」

座長　Desai Narasimhalu (Program Manager, ISS, シンガポール)

12:30-13:00　発表「知識構造の超並列マッチング」
James A. Hendler (Professor, University of Maryland, 米)

13:00-13:30　発表「知識指向工学を目指して」　冨山哲男（東京大学助教授）

13:30-14:00　発表「分子生物学における大規模知識ベースの構築と共有」
Francois Rechenmann (INRIA, 仏)

14:00-14:30　（休　憩）

14:30-16:00　パネルディスカッション「ＫＢ＆ＫＳのための応用とブレークスルー」

パネリスト：
北野宏明（ソニーコンピュータサイエンス研究所）
James A. Hendler (Professor, University of Maryland, 米)
Mary Shephard (CYC Project, MCC, 米)
西田豊明（奈良先端科学技術大学院大学教授）

16:00　　まとめ　James A. Hendler (Professor, University of Maryland, 米)

資料C. KB&KS'93国際ワークショップ参加者一覧
(平成5年12月3日現在)

【国内参加者】

相場　亮	(財)新世代コンピュータ技術開発機構 研究所第2研究部 部長代理
麻田　治男	㈱東芝 研究開発センター情報・通信システム研究所 第二研究所長
芦崎　達雄	日本科学技術情報センター 技術開発部技術開発課 副主任情報員
井佐原 均	電子技術総合研究所 知能情報部自然言語研究室 主任研究官
内田　裕士	㈱富士通研究所 情報処理研究部門情報処理研究部 部長
梅田　靖	東京大学 工学部総合試験所 助手
奥村　学	北陸先端科学技術大学院大学 情報科学研究科 助教授
河野　浩之	京都大学工学部数理工学科 助教授
木本　晴夫	日本電信電話㈱ NTT情報通信網研究所メッセージシステム研究部主幹研究員
北野　宏明	カーネギーメロン大学 機械翻訳研究所 研究員
桐山　孝司	東京大学 人工物工学センター知能科学部門 講師
黒橋　禎夫	京都大学 工学部
芥子　育雄	シャープ㈱ 技術本部情報技術研究所 主任
小泉　敦子	㈱日本電子化辞書研究所 第三研究室 主任研究員
佐藤　理史	北陸先端科学技術大学院大学 情報科学研究科 助教授
鈴木　浩之	松下電器産業㈱ 東京情報システム研究所計画部基盤開発室 主任技師
高橋　千恵	(財)日本情報処理開発協会 開発研究室開発研究第二課 係長
田中　秀俊	(財)新世代コンピュータ技術開発機構 研究所第2研究部 研究員
田中　穂積	東京工業大学 工学部情報工学科 教授
田中　裕一	㈱富士通研究所 ソフトウェア研究部
津田　宏	(財)新世代コンピュータ技術開発機構 研究所第二研究部 研究員
土屋　俊	千葉大学 文学部行動科学科 助教授
寺野　隆雄	筑波大学大学院 経営システム科学 助教授
徳永　健伸	東京工業大学 工学部情報工学科 助教授
富山　哲男	東京大学 工学部総合試験所 助教授
長尾　眞	京都大学 工学部 教授
中島　秀之	電子技術総合研究所 協調アーキテクチャ計画室 室長
長瀬　眞理	城西国際大学 助教授
西尾章治郎	大阪大学 工学部情報システム工学科 教授

西田　豊明	奈良先端科学技術大学院大学 情報科学研究科 教授
新田　克己	(財)新世代コンピュータ技術開発機構 研究所第２研究部 部長
新田　義彦	㈱日立製作所 基礎研究所 主任研究員
橋田　浩一	電子技術総合研究所 知能情報部自然言語研究室
服部　文夫	日本電信電話㈱ ＮＴＴ情報通信網研究所知識処理研究部 主幹研究員
日夏　健一	日本科学技術情報センター 技術開発部技術開発課 副主任情報員
渕　一博	東京大学 工学部電子・情報工学科 教授
堀　浩一	東京大学 先端科学技術研究センター 助教授
堀　雅洋	日本アイ・ビー・エム㈱ 東京研究所 副主任研究員
松本　裕治	奈良先端科学技術大学院大学 情報科学研究科 教授
水谷　博之	㈱東芝 研究開発センター情報・通信システム研究所第二研究所 主任研究員
溝口　文雄	東京理科大学 理工学部経営工学科 教授
溝口理一郎	大阪大学 産業科学研究所 教授
村木　至一	日本電気㈱ Ｃ＆Ｃ研究所メディアテクノロジィ研究部 部長
元吉　文男	電子技術総合研究所 知能情報部自然言語研究室 室長
安原　宏	㈱日本電子化辞書研究所 第六研究室 室長
横井　俊夫	㈱日本電子化辞書研究所 研究所長
横田　一正	(財)新世代コンピュータ技術開発機構 研究所 主席研究員
横山　晶一	山形大学 工学部電子情報工学科 教授
Pierre Morizet-Mahoudeaux	東京大学 先端科学技術研究センター

【海外参加者】

Bob Jansen	Project Manager, Division of Information Technology CSIRO, Australia
Jiawwei Han	Associate Professor, School of Computing Science Simon Fraser University, Canada
John Mylopiulos	Professor, Department of Compuer Science Univesity of Toronto, Canada
Changning Huang	Department of Computer Science Tsinghua University, Chaina

Francois Arlabosse Scientific and Technical Director
FRAMENTEC-COGNITECH, France

Christian Boitet Study Group for Machine Translation
GETA, IMAG-campus, France

Jacoues Mathieu Sous-Direction, Informatique Bureautique
Ministère de l'Industrie et du Commerce Extérieur
France

Francois Peccoud Study Group for Machine Translation
GETA, IMAG-campus, France

Francois Rechenmann Director of Research
INRIA, Rhone-Alpes, France

Stephan Bodenkamp Director, Foreign Affairs Office
Test Center for AI and MT, Germany

Knut Hinkelmann Research Scientist, General Research Center for AI
DFKI, Germany

Christian Rohrer Professor, Institute for Computational Linguistics
University of Stuttgart, Germany

Joachim Schmidt Professor, Computer Science
Hamburg University, Germany

Beate Springmann Director, Foreign Affairs Office
Test Center for AI and MT, Germany

Hans Voss Artificial Intelligence Research Division
GMD, Germany

Agung Santosa — Agency for the Assessment and Apllication of Technology
BPP Teknologi, Indonesia

Piero Torrigiani — FINSEL, Italy

Antonio Zampolli — Professor, Department of Lingistics
University of Pisa, Italy

key-Sun Choi — Associate Professor, Computer Science Department
Korea Advanced Institute of Science and Technology, Korea

Izhar Che-Mee — Manager, National Project Unit
Research & Development Division
Malaysian National Institute of Translation, Malaysia

Vadim L. Stefanuk — Professor, Institute for Information Transmission Problems
Russian Academy of Science, Russia

Desai Narasimhalu — Program Manager, Institute of Systems Science
National University of Singapore, Singapore

Susan Armstrong — Professor, University of Geneva, Swizerland

Wantanee Phantachat — Linguistics and Knowledge Science Laboratry
National Electronics and Compuer Technology Center
National Science and Technology Development Agency
Ministy of Science, Technology and Environment, Thailand

Vilas Wuwongse — Associate Professor, Computer Science Program
School of Advanced Technologies, Thailand

Nicolaas J.I.Mars — Professor, Department of Computer Science
University of Twente, The Netherlands

Reinder van de Riet Professor, Department of Mathematics and Computer Science
Section of Information Systems
Free University Amstrdam, The Netherlands

Bob J. Wielinga Professor, Department of Social Science Informatics
University of Amstrdam, The Netherlands

Peter M.D. Gray Professor, Department of Computer Science
University of Aberdeen, U.K.

Brian Oakley Logica pcl., U.K.

Jun-Ichi Tsujii Professor, Center for Computational Linguistics
University of Manchester Institute of Science
and Technology, U.K.

Yorick Wilks Professor, Department of Computer Science
University of Sheffield, U.K.

Robert C. Werwick Professor, Artificial Intelligence Laboratry
Massachusetts Institute of Technology, U.S.A.

Ronald J. Brachman Department Head, Artificial Intelligence
Principles Research Department
AT&T Laboratories, U.S.A.

Timothy W. Finn Professor & Chair, Department of Computer Science
University of Maryland Boltimore County, U.S.A.

Louise Guthrie Acting Director, Computing Research Laboratory
New Mexico State University, U.S.A.

Kenneth B. Haase Professor, Media Laboratory
Massachusetts Institute of Technology, U.S.A.

James A. Hendler Associate Professor, Department of Computer Science
University of Maryland, U.S.A.

Susan Hockey Director, Center for Electronic Texts in the Humanities
Rutgers and Princeton Universities, U.S.A.

Nancy Ide Associate Professor, Department of Computer Science
Vassar College, U.S.A.

Douglas B. Lenat Director of Cyc Project
Microelectronics and Compurt Technology and Corpration
U.S.A.

Robert Neches Project Leader, Integrated User-Support Environments Group
Information Sciences Institute
University of Southern California, U.S.A.

Hanan Samet Professor, Department of Computer Science
Center for Automation Research
University of Maryland, U.S.A.

Mary Shepherd Member of Cyc Project, Technical Staff
Microelectronics and Compurt Technology and Corpration
U.S.A.

C.M. Sperberg-Mcqueen
Editor, Text Encoding Initiative
The University of Illinois at Chicago, U.S.A.

Wiliam R. Swartout Director, Intelligent Systems Division
Information Sciences Institute
University of Southern California, U.S.A.

Jean Veronis Associate Professor
Department of Computer Science
Vassar College, U.S.A.

－ 禁無断転載 －

大規模知識ベースに関する調査研究報告書

発　行　平成６年３月
発行所　財団法人 日本情報処理開発協会
東京都港区芝公園３丁目５番８号
機械振興会館内
電話（０３）３４３２－９３９０

０５－Ｒ００８