

56-R 011

# 文章情報データベース総合利用 調査研究報告書

昭和 57 年 3 月



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昭和56年度に実施した「文章情報データベースの総合利用に関する調査研究」の成果をとりまとめたものであります。





## は　じ　め　に

本報告書は、財団法人日本エネルギー経済研究所に委託して実施した「文章情報データベースの総合利用に関する調査」の成果をとりまとめたものである。

近年の貿易摩擦問題等にも見られるように、国際情勢の変化は、即時、わが国各界に大きく波及し、こうした事態への迅速・的確な対応はもとより、事前に予知できる体制づくりが急務となっている。

このためには、内外の情報資源を活用し、常時こうした動向を把握できる情報処理システム体制を確立する必要がある。

とくに、記事情報等の文章情報をデータベース化し、コンテンツ・アナリシス等の高度な手法を駆使して情報内容を分析し、客観的な判断材料を求めることは有効と思われる。

また、これらの処理過程においては、なるべく人手を介さない、自動的な機械処理が可能な形が望まれている。

こうしたことから、本事業においては、文章情報データベースを効果的に利用するために必要な総合解析システムを開発することを目的として、初年度にあたる本年度は、基礎的な調査研究を実施した。

今後は、実用レベルでのシステム開発に着手するとともに理想的なシステム構築に向けて研究開発を推進していくこととしたい。

最後に、本調査研究にあたって、ご指導、ご協力をいただいた委員並びに関係各位に感謝の意を表します。

昭和57年3月



# 「文章情報データベース総合利用調査委員会」委員名簿

(順不同、敬称略)

|     |     |   |     |                                                      |
|-----|-----|---|-----|------------------------------------------------------|
| 委員長 | 淵   | 一 | 博   | 電子技術総合研究所パターン情報部長                                    |
| 委員  | 田   | 中 | 穂積  | 電子技術総合研究所パターン情報部<br>推論機構研究室長                         |
| "   | 矢   | 田 | 光治  | 電子技術総合研究所情報計算センター長                                   |
| "   | 中   | 瀬 | 純夫  | リソースシェアリング(株)開発技術第二部部長                               |
| "   | 小   | 沢 | 大二  | 国際協力事業団派遣事業部管理課長                                     |
| "   | 下   | 岡 | 克幸  | 日本貿易振興会企画部電子計算室                                      |
| "   | 木   | 南 | 公統司 | (株)開発計算センターシステム管理部部長                                 |
| "   | 小   | 泉 | 秀雄  | 日本ハネウエル・インフォメーション・システムズ(株)<br>営業本部営業推進第二部部長代理        |
| "   | 長谷川 |   | 亨   | (株)ソフトウェア・リサーチ・アソシエイツ<br>ソフトウェア・マーケティング本部システム第2部部長代理 |
| "   | 山   | 本 | 恵美子 | ソフトウェア研究会副会長                                         |
| "   | 三   | 上 | 喜貴  | 通商産業省産業政策局産業構造課産業構造専門職                               |
| "   | 木   | 地 | 三千子 | 通商産業大臣官房情報管理課<br>政策情報システム室企画係長                       |
| "   | 小   | 谷 | 泰久  | 通商産業大臣官房情報管理課政策情報システム室                               |
| "   | 笹   | 川 | 丞也  | 通商産業大臣官房情報管理課政策情報システム室                               |
| "   | 栗   | 川 | 正仁  | 通商産業大臣官房情報管理課計画班第二係長                                 |
| "   | 所   | 沢 | 仁   | (株)日本エネルギー経済研究所第6研究室長                                |
| "   | 市   | 川 | 隆   | (株)日本情報処理開発協会技術調査部次長                                 |

# 「文章情報データベース定量化利用研究 ワーキング・グループ」委員名簿

(順不同, 敬称略)

|    |         |                                                        |
|----|---------|--------------------------------------------------------|
| 主査 | 山本 毅 雄  | 図書館情報大図書館情報学部教授                                        |
| 委員 | 石川 徹 也  | 図書館情報大学図書館情報学部助手                                       |
|    | 石 塚 英 弘 | 国立国文学研究資料館助教授                                          |
| "  | 石 井 義 章 | NCC. (日本コンピュータ・センター)<br>ソフトウェア・エンジニアリング(株) 代表取締役       |
| "  | 木 南 公統司 | (株)開発計画センター・システム管理部課長                                  |
| "  | 岡 克 幸   | 日本貿易振興会企画部電子計算室                                        |
| "  | 長谷川 亨   | (株) ソフトウェア・リサーチ・アソシエイツ ソフトウェア・<br>マーケティング本部システム第2部部長代理 |
| "  | 中 瀬 純 夫 | リソースシェアリング(株)開発技術第二部長                                  |
| "  | 木 地 三千子 | 通商産業大臣官房情報管理課<br>政策情報システム室企画係長                         |
| "  | 小 谷 泰 久 | 通商産業大臣官房情報管理課政策情報システム室                                 |
| "  | 笹 川 丞 也 | 通商産業大臣官房情報管理課政策情報システム室                                 |
| "  | 所 沢 仁   | (財)日本エネルギー経済研究所第6研究室長                                  |
| "  | 小 川 芳 樹 | (財)日本エネルギー経済研究所第6研究室                                   |
| "  | 加 藤 泰 宏 | (財)日本情報処理開発協会開発部                                       |



# 「自動翻訳システム研究ワーキング・グループ」委員名簿

(順不同, 敬称略)

|     |         |                                     |
|-----|---------|-------------------------------------|
| 主 査 | 矢 田 光 治 | 電子技術総合研究所情報計算センター長                  |
| 委 員 | 田 中 穂 積 | 電子技術総合研究所パターン情報部推論機構研究室長            |
| "   | 横 山 晶 一 | 電子技術総合研究所パターン情報部推論機構研究室             |
| "   | 田 中 隆   | 電子技術総合研究所情報計算センター                   |
| "   | 柿 元 俊 博 | 富士通(株)システム統轄部科学システム開発部<br>第一科学システム課 |
| "   | 中 瀬 純 夫 | リソースシェアリング(株)開発技術第二部長               |
| "   | 山 本 恵美子 | ソフトウェア研究会副会長                        |
| "   | 栗 川 正 仁 | 通商産業大臣官房情報管理課計画班第二係長                |
| "   | 木 地 三千子 | 通商産業大臣官房情報管理課政策情報システム室<br>企画係長      |
| "   | 小 谷 泰 久 | 通商産業大臣官房情報管理課政策情報システム室              |
| "   | 笹 川 丞 也 | 通商産業大臣官房情報管理課政策情報システム室              |



# 目 次

## I 総 論

|                                  |    |
|----------------------------------|----|
| 1. 事 業 概 要 .....                 | 1  |
| 1.1 目 的 と 背 景 .....              | 1  |
| 1.2 事 業 計 画 .....                | 2  |
| 1.3 実 施 経 過 .....                | 5  |
| 1.4 本報告書の構成 .....                | 6  |
| 2. 文章情報データベースの総合利用調査 .....       | 7  |
| 2.1 文章情報データベース利用の現状 .....        | 7  |
| 2.1.1 データベースの普及 .....            | 7  |
| 2.1.2 文章情報データベース利用の問題点 .....     | 12 |
| 2.2 総合解析システムの必要性 .....           | 14 |
| 2.2.1 情報需要の変化 .....              | 14 |
| 2.2.2 総合解析システム化の方向 .....         | 17 |
| 2.2.3 文章情報データベース高度利用による効果 .....  | 18 |
| 2.3 文章情報解析手法の研究 .....            | 19 |
| 2.3.1 定量化利用の発展過程 .....           | 19 |
| 2.3.2 内容分析の方法 .....              | 20 |
| 2.3.3 内容分析データの高度利用 .....         | 23 |
| 3. 文章情報データベース総合解析システムの基本構想 ..... | 25 |
| 3.1 総合解析システム像 .....              | 25 |
| 3.1.1 総合解析システムの概念像 .....         | 25 |
| 3.1.2 基 礎 解 析 .....              | 27 |
| 3.1.3 知識ベース作成 .....              | 27 |
| 3.1.4 内容検索と分析 .....              | 29 |

|         |                    |    |
|---------|--------------------|----|
| 3.1.5   | 翻訳について             | 30 |
| 3.2     | システム機能             | 31 |
| 3.2.1   | システム機能の概要          | 31 |
| 3.2.2   | 機能別サブシステム          | 32 |
| 3.2.2.1 | 基礎解析               | 32 |
| 3.2.2.2 | 知識ベース              | 35 |
| 3.2.2.3 | 内容検索と分析            | 35 |
| 3.2.2.4 | 機械翻訳               | 36 |
| 3.3     | システム開発への課題         | 37 |
| 3.3.1   | 文章情報の特性            | 37 |
| 3.3.2   | 文章中に出現する語の利用のための課題 | 37 |
| 3.3.3   | 文章の意味内容の解析         | 39 |
| 3.3.4   | 文章からの内容抽出のための課題    | 41 |
| 3.3.5   | 内容分析と自動翻訳技術との関連    | 43 |
| 3.3.6   | 課題解決のための展望         | 43 |

## II 各 論

|     |                       |     |
|-----|-----------------------|-----|
| 1.  | 文章情報定量化利用の研究          | 47  |
| 1.1 | 研究目的と活動概要             | 47  |
| 1.2 | 既存システムによる文章情報定量化利用事例  | 50  |
| 1.3 | 入力データ作成にあたっての問題点      | 64  |
| 1.4 | 単語およびキーワードの自動抽出       | 72  |
| 1.5 | シソーラスの作成              | 78  |
| 1.6 | 定量化利用の方法とユーザの使い易いシステム | 84  |
| 1.7 | 新しい記事情報総合利用システムの構築    | 93  |
| 2.  | 自動翻訳システムの研究           | 108 |

|                           |                    |     |
|---------------------------|--------------------|-----|
| 2.1                       | 研究目的と活動概要          | 108 |
| 2.2                       | 翻訳システム技術の現状と将来     | 110 |
| 2.2.1                     | 歴史的背景              | 110 |
| 2.2.2                     | 現 状                | 110 |
| 2.2.3                     | 機械翻訳システムの構成        | 113 |
| 2.2.4                     | RIPSの翻訳システム        | 117 |
| 2.2.5                     | 今後の課題と将来           | 127 |
| 2.3                       | 実験データによるトランスレータの分析 | 130 |
| 2.3.1                     | INSPECデータによる実験     | 130 |
| 2.4                       | 翻 訳 辞 書            | 137 |
| 2.5                       | イメージ文書処理と入出力装置     | 140 |
| 2.6                       | LC-MARCの実験準備       | 144 |
| 付属資料 - 文章情報データベース関連資料一覧 - |                    | 147 |
| あ と が き                   |                    | 193 |



# I 總 論





# 1. 事業概要

## 1.1 目的と背景

コンピュータ情報処理は、広く深く社会に浸透し、情報化社会は名実ともに確立されつつある。

近年のオフィス・オートメーションやパーソナル・コンピュータの急激な普及にもみられるように、職場や家庭における一般利用者の比較的容易な操作によるコンピュータ利活用が可能な時代となった。

こうした状況の背景には、ハードウェア、ソフトウェアの技術向上に伴う諸条件整備はもとより、情報資源に対する社会及び個人の関心、意識の向上が大きく作用していると思われる。

情報そのものの重要性が認識され、情報量が拡大されるに伴い、データベースや、情報検索システムといった情報流通機能の需要も拡大するのは必須であり、益々増大する情報利用者のニーズに対応すべく、情報の量的、質的整備と簡便な利用体制づくりを中心とした情報処理システムの確立が急務となってきた。

とくに文章情報データベースは日本語情報処理の普及や海外からの文献情報データベース等の導入により、その蓄積及び利用は急速に高まるものと思われる。

それらを効果的に利用するには通常の情報検索に加えて、データベースの持つ各種の情報をコンテナンツ・アナリシス等の高度な分析（キーワードの頻度の時系列分析、出現頻度の相関分析、意味論的分析）をし、利用することが極めて重要である。

このため、文章情報データベースを効果的に利用するために必要な総合解析システムを開発することを目的として、調査研究を実施する。

## 1.2 事業計画

本事業の計画概要は図1-1に示すとおりである。

第1年度は、文章情報データベースの総合解析システムに必要な各種分析手法並びに具備すべき諸機能の検討を中心とした基礎的な調査研究を実施する。また、第2年度以降のシステム開発に向けて基本構想を作成することとし、この際、現状で実現可能なシステムを検討すると同時に、理想的な総合解析システムのあり方も研究する。

第2年度は、第1年度の調査研究を継続するとともに海外先進例の実態調査を行い、システム開発の参考に資する。また、新データベースの構築並びに一部システム開発に着手する。

第3年度は、第1,2年度の研究成果に基づき、可能なところから実用化レベルでのシステム開発を実施し、併せて、総合解析システムとしてあるべき姿を追究する。

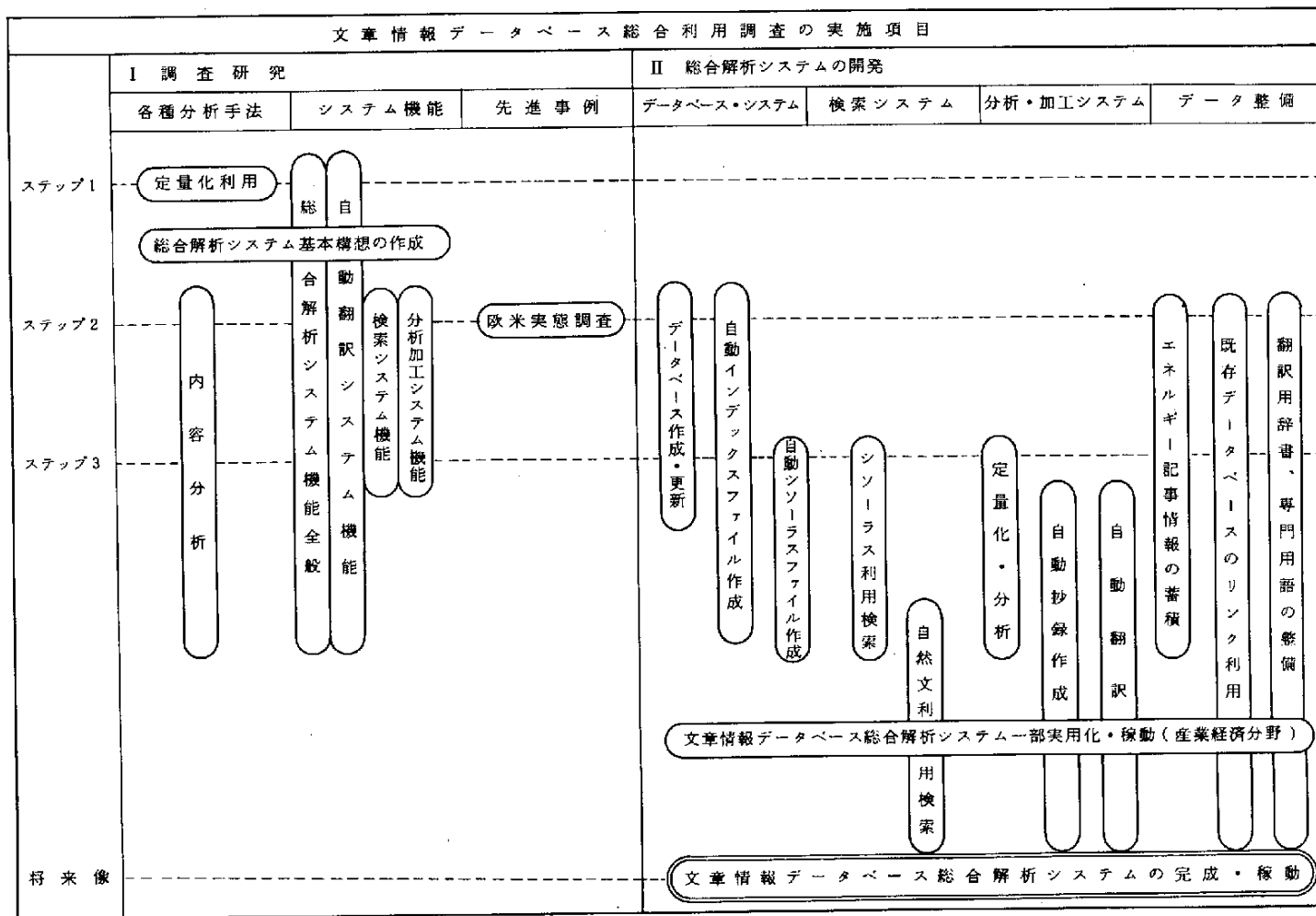


図 1 - 1 計 画 概 要

### 1.3 実 施 経 過

本調査を実施するにあたり、「文章情報データベース総合利用調査委員会」を設置し、4回の委員会を開催、本調査研究全体を推進・統括した。

また、具体的には、

- ① カントリーリスク・エネルギー動向把握のための文章情報コンテンツ  
・アナリシス手法の研究
- ② 海外情報の有効活用、問題別把握、分析を行うための翻訳システム実  
用化の基礎研究

のテーマに基づく調査研究を、委員会メンバーを中心とするワーキング・グループ形式で実施した。

経過概要は表1-1に示すとおりである。

表 1 - 1 昭和 5 6 年度調査実施経過表

|                                | 56/4 | 5 | 6                                                                                                                                                  | 7 | 8 | 9 | 10 | 11 | 12 | 57/1 | 2 | 3 |
|--------------------------------|------|---|----------------------------------------------------------------------------------------------------------------------------------------------------|---|---|---|----|----|----|------|---|---|
| I 実施計画の策定                      | ←→   |   |                                                                                                                                                    |   |   |   |    |    |    |      |   |   |
| II 委員会の開催                      |      |   |                                                                                                                                                    | ○ |   |   | ○  |    |    | ○    |   | ○ |
| III<br>文章情報コンテンツ・アナリシス手法研究     |      |   | 研究課題の検討<br>←→<br>キーワード自動抽出法の研究<br>←→<br>入力データ作成上の課題<br>←→<br>既存システム利用事例研究<br>←→<br>シソーラス辞書作成の研究<br>←→<br>定量化利用方法論の研究<br>←→<br>新記事情報利用システムの研究<br>←→ |   |   |   |    |    |    |      |   |   |
| IV<br>翻訳システム化<br>実用化基礎調査<br>研究 |      |   | 研究課題の検討<br>←→<br>実用可能性の研究<br>←→<br>機械翻訳技術の研究<br>←→<br>構文解析技術の研究<br>←→<br>入出力インターフェースの研究<br>←→<br>LC-MARCの実験準備<br>←→                                |   |   |   |    |    |    |      |   |   |
| V 報告書の作成                       |      |   |                                                                                                                                                    |   |   |   |    |    | ←→ |      |   |   |

## 1.4 本報告書の構成

本報告書は第一部総論と第2部各論からなっている。

総論編として、第2章に文章情報データベース総合利用のための基礎調査を、第3章に今後のシステム開発に向けて文章情報データベース総合解析システムの基本構想をまとめた。

第2章は、2.1に情報処理技術やデータベース流通の現状と問題点を、2.2に文章情報有効活用のための総合解析システムを開発する必要性を、2.3に文章情報の解析手法の研究として、定量化利用を中心とした内容分析の方法論、分析データの高度利用法を各々まとめた。

第3章は、3.1に総合解析システムの全体像と機能別のシステム概念を図示して、システム・イメージの統一を図った。この際、将来のあるべき姿と本事業計画における開発範囲の両面からの検討を行った。3.2は、3.1のシステム像に基づき、具備すべきシステム諸機能につき解説した。さらに、3.3にシステム開発への課題として、今後に残された問題点を中心に述べた。

各論編は、ワーキング・グループで実施した研究成果を、第1章文章情報定量化利用研究、第2章自動翻訳システム研究として詳述した。

また、付属資料として、文章情報の解析や自然言語処理に関連した文献の一覧を付し、今後の参考に資することとした。

## 2. 文章情報データベースの総合利用調査

### 2.1 文章情報データベース利用の現状

#### 2.1.1 データベースの普及

情報処理技術の進展に伴い、コンピュータを利用した情報の蓄積・提供・利用が急速に普及し、大きな関心を集めている。

データベースという用語は、広く使われ意味が定着しつつあるが、微妙に異なった使われかたをする場合がある。ここではデータベースを、「機械可読な形になっている情報で、何らかの検索システムを介してユーザーに提供されているもの」と定義する。

現在、利用されている主要なデータベースは2種に大別できる。1つは、数値データを主としたもので“数値情報データベース”。他方は、非数値すなわち言語で書かれたデータを主としたもので“文章情報データベース”と呼ばれる。

数値情報データベースは、さらにデータの内容から、統計データ、企業関係の財務データ、株式・商品等の市況データ、科学数値等に分けることができる。最も代表的な数値情報データベースは、統計のデータベースであり、中心は経済社会統計である。各国の中央統計局や国際機関等の統計作成機関自身のものから、統計作成機関からデータの提供を受けて構築されたものまでさまざまである。例えば前者ではカナダ中央統計局のCANSIM、国連統計局のUNSD、後者ではGEの米国産業統計データベースSIC72、日本経済新聞社のNEEDS-IFSが挙げられる。企業財務データのデータベースは、民間の調査機関、銀行などにより数多く作成され利用されている。Standard & Poor's社のCOMPUSTATや日本経済新聞社のNEEDS-TS/Iなどはその例である

表 2-1 数値情報データベース（経済、財務・市況）

出所：データベースの利用法 欄旭リサーチセンター

| データベース名                            | 内 容                                                                                          | プロデューサー                                      | ディストリビューター |
|------------------------------------|----------------------------------------------------------------------------------------------|----------------------------------------------|------------|
| PTS INTERNATIONAL TIME SERIES      | 世界の主要30ヶ国（鉄本国）に関する約500主要系列からなる25000の予測時系列（資源データと予測データ）及び1957年から現在までの年間データを世界中の約125000系列として収録 | Predicasts, Inc.                             | Lockheed   |
| PTS INTERNATIONAL FORECASTS        | 世界主要国（鉄本国）の出版された予測データと資源とともに収録。マクロ経済、産業統計、消費統計に及ぶ                                            | Predicasts, Inc.                             | Lockheed   |
| CANADIAN ECONOMIC DB               | カナダのマクロ経済、人口統計、米国の連邦統計局マクロモデル予測                                                              | DRI, Statistics Canada, CANSIM Division      | DRI        |
| NEEDS-IFS                          | IMF作成の国際財政、金融データ。200ヶ国、16900系列                                                               | 日本経済新聞社                                      | 日本経済新聞社    |
| CITIBASE                           | 米国のマクロ経済、産業および予測データ。約3800項目の時系列データ                                                           | Citibank                                     | GE         |
| PTS U.S. TIME SERIES               | 米国に関する1957年からの変遷データと1994年までの出版された予測データ約500の時系列。及び米国の生産、消費などの約30000系列の資源統計                    | Predicasts, Inc.                             | Lockheed   |
| PTS U.S. FORECASTS                 | 雑誌、新聞、報告書などに発表された米国のマクロ経済、産業などの予測データを収録。（短期予測、長期予測）                                          | Predicasts, Inc.                             | Lockheed   |
| SIC 72                             | 米国の568の産業分類（SIC）の19項目の時系列データ。各産業の出産額、付加価値額、原料消費、従業員数、賃金などを含む                                 | U. S. Department of Commerce                 | GE         |
| U. S. Central DB                   | 米国の経済発展に不可欠な経済統計データ約15000系列を収録。毎月更新の短期マクロ予測。                                                 | DRI                                          | DRI        |
| UCLA                               | 米国のマクロ経済・産業の460系列の資源データ及び短期、長期予測データ                                                          | University of California, Los Angeles (UCLA) | GE         |
| DRI CAP                            | 米国のマクロ経済指標（3700系列）                                                                           | DRI                                          | CDC        |
| EVANS ECONOMIC - CS - USA DATABASE | 米国のマクロ経済指標2万系列。予測も可能                                                                         | Evans Economics, Inc.                        | CDC        |
| BI/DATA                            | 世界70ヶ国の150項目にわたる国民経済、およびその予測の時系列データ                                                          | Business International Corporation           | GE         |
| EUROPEAN ECONOMIC DB               | 欧州各国の統計当局から直接入手した経済統計データベース（IMF, OECD, 世界などのデータも収録。欧州主要3ヶ国経済のモデル予測）                          | DRI, INSEE(仏), IFO(西独), CSO(意)               | DRI        |
| NRI/E                              | 日本のマクロ・ミクロ経済・産業・消費などの約2500項目の時系列データ                                                          | 野村総合研究所                                      | GE         |
| NEEDS-TS/E                         | 日経総合経済ファイル（国内約6800系列の時系列データ）。主要産業の生産、出荷、在庫統計、消費・家計統計など。                                      | 日本経済新聞社                                      | 日本経済新聞社    |
| NEEDS-TS/I, O                      | 産業調査（74部門）                                                                                   | 日本経済新聞社                                      | 日本経済新聞社    |
| ENERGY DB                          | 本国内外のエネルギーの消費、輸入、価格など22000系列をカバーするデータベース。経済・技術などを敏感に反映する價格面分析用のモデル予測。                        | DRI                                          | DRI        |
| NEEDS-IEE                          | 約23000系列のエネルギー関連データ。基盤データの予測及び中長期予測も可能                                                       | 日本経済新聞社<br>日本エネルギー経済研究所                      | 日本経済新聞社    |



| 財務・市場情報       | データベース名                                         | 内 容                                                            | プロデューサー                                                                                             | ディストリビューター    |
|---------------|-------------------------------------------------|----------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|---------------|
|               | COMPUSTAT                                       | 米国内外7,500社の財務諸表                                                | Standard & Poor's Compustat Services, Inc.                                                          | DRI CDC       |
|               | DISCLOSURE                                      | SECに提出された米企業1,000社の10-K, 10-Q, 8-K リポート、証券投資会社のNIR, NIQ リポート   | Disclosure Inc.                                                                                     | Lockheed      |
|               | EIS INDUSTRIAL PLANTS: EBI5*                    | 米国の年間売上高1,500万以上の約6,700社の内13万の事業所に關するリスト売上、従業員数、マーケットシェアなどから成る | Economic Information Systems, Inc.                                                                  | Lockheed CDC* |
|               | EXSTAT                                          | 米国外の2,000社(欧州、オーストラリア、日本)の財務諸表                                 | Extel Statistical Services, Ltd.                                                                    | CDC           |
|               | FTCDATA                                         | 米国の42産業分類における114項目の財務データを時系列データベースとする連続データベースのデータベース           | U. S. Federal Trade Commission                                                                      | GE            |
|               | STANDARD & POOR'S NEWS                          | 米国の9,000以上の企業の活動状況及びSEC提出用の財務データ                               | Standard and Poor's Corporation                                                                     | Lockheed      |
|               | VALUE LINE                                      | 米国の著名企業1,600社の250項目の財務データ及び売上、収益、株価の予測も可能                      | Arnold Bernard & Co., Ltd.                                                                          | GE DRI        |
|               | BANKANAL                                        | 米国の上位145の銀行の財務データ。収益の予測も可能                                     | Robinson-Humphrey Company                                                                           | CDC           |
|               | EURABANK                                        | 欧州、南米、北米、オーストラリアなど72ヶ国の1,000の銀行のマニュアルレポートより作成した財務データ           | European-American Bank & Trust Company                                                              | CDC           |
|               | FINANCIAL INSTITUTION DATA BASE (FDIC/FSL/NCUA) | 米国の金融機関約11,000社の財務データ                                          | Federal Deposit Insurance Corp., Federal Home Loan Bank Board, National Credit Union Administration | CDC           |
|               | INSURANCE INDUSTRY DATABASE                     | 米国の生命、財産保険会社500社の損益計算書                                         | National Association of Insurance Commissioners                                                     | CDC           |
|               | IBJ DATA                                        | 東京市場約100社の財務データ(1964年~)                                        | 日本興業銀行                                                                                              | GE            |
|               | TSR DATA                                        | 東京市場約1,700社と京上場約15,000社の財務データ及び企業業績                            | 東京証券リサーチ                                                                                            | GE            |
|               | NEEDS-TS/1 (日経財務データ)                            | 上場会社(1,700社)の財務データ及び中間財務データ、決算予想・通算データ                         | 日本経済新聞社                                                                                             | 日本経済新聞社       |
|               | 共同株価                                            | ニューヨーク証券取引所の上場の全銘柄の株価(日次)                                      | 共同通信社                                                                                               | GE            |
|               | 時事株価                                            | 東京1、2部上場1,700銘柄、全債券の日次価格及びニューヨーク証券取引所、アメリカン証券取引所上場の全銘柄の株価      | 時事通信社                                                                                               | GE            |
|               | CAPLOAN                                         | 欧州市場を中心としたシンジャー・ローニングデータ、(約3,000の銀行と2,500の債券)                  | Caploan International Finance Data Co.                                                              | GE            |
|               | NEEDS/TS-1 (日経株価ファイル)                           | 東京全銘柄、大証単独銘柄の株価                                                | 日本経済新聞社                                                                                             | 日本経済新聞社       |
|               | QUICK ビデオ-1、ビデオ-2                               | 国内の上場会社の株価、市況情報、経済ニュース                                         | 市況情報センター                                                                                            | 市況情報センター      |
| QUICK ビデオマスター | ニューヨーク、ロンドンなど海外の主要取引所の株価、商品情報11,000件            | ロイター通信社                                                        | ロイター通信社 市況情報センター                                                                                    |               |

(表2-1参照)。

一方、文章情報データベースは、科学技術・学術情報、特許情報、法令、図書目録、新聞・雑誌等のデータベースに分けることができる。

最も代表的で多種あるものは科学技術・学術情報のデータベースで、国際的に広く知られたものには、化学文献の抄録データベースの CAS SEARCH, 生物科学・医学の BIOSIS, PREVIEWS, 医学の MEDLARS, 物理の INSPEC 等がある。わが国では日本科学技術情報センター (JICST) が理工学文献のデータベースを作成し、提供している。特許情報のデータベースは世界45カ国の特許を集めた INPADOC や、わが国の特許情報を提供している (財) 日本特許情報センター (JAPATIC) の PATOLIS 等が知られている。

科学技術関係のデータベースは、米国において科学技術振興政策に沿った形で比較的早くから公的機関等を通じて開発され利用されてきた。わが国でも、CAS, MEDLARS 等の海外データベースを含めて昭和40年代始めから利用され最も普及した分野である。

これら科学技術文献データベースに加えて、最近では経済社会関係分野の情報を主とした文章情報のデータベースも増えつつある (表2-2参照)。米国内外の企業・産業情報を索引化した、Predicasts 社の PTS F&S INDEX, 経済・産業関係の情報を抄録化した同じく Predicasts 社の PTS PROMT などはその例である。また、ニューヨーク・タイムズ・インフォメーション・バンクは主要新聞雑誌の記事抄録データベースを提供している。ニューヨーク・タイムズ紙については、オンラインによる全文記事が検索できるデータベースが利用可能である。

わが国では、日本経済新聞社が自社発行紙誌の要旨・抄録をデータベース化し、NEEDS-IRとして提供している。

データベースの利用を促進している背景の1つに国際間でのオンライン検索ができる国際通信サービスの普及がある。国際的に広く知られた各

表 2 - 2 文章情報データベース (企業経営, 新聞情報)  
出所: データベースの利用法 (朝日サーチセンター)

|        | データベース名                                                              | 内 容                                                                                                                                                                                 | プロデューサー                                                                | ディストリビューター                |
|--------|----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|---------------------------|
| 企業経営情報 | ABI/INFORM <sup>®</sup> INFORM <sup>™</sup>                          | 企業経営, 管理に関する文献                                                                                                                                                                      | Data Courier, Inc.                                                     | Lockheed SDC <sup>™</sup> |
|        | CBPI                                                                 | カナダの経済, 産業に関する文献                                                                                                                                                                    | Information Access of Toronto                                          | SDC                       |
|        | ECONOMIC ABSTRACTS INTERNATIONAL                                     | 世界の経済, 企業経営に関する文献                                                                                                                                                                   | Learned Information Ltd.                                               | Lockheed                  |
|        | MANAGEMENT CONTENTS                                                  | 意思決定や予測の役に立つ, 企業, 産業に関する文献                                                                                                                                                          | Management Contents, Inc.                                              | Lockheed SDC              |
|        | PTS F&S INDEXES                                                      | 米国内外の会社, 製品, 産業情報, 企業買収, 合併や新製品開発, 社会政策的要因に関する情報を含む。                                                                                                                                | Predicasts, Inc.                                                       | Lockheed                  |
|        | PTS PROMT                                                            | 世界の企業活動, 企業環境, 市場, 新製品開発, 消費者等に関する, 数千の新聞, 雑誌, 報告書から選ばれた文献                                                                                                                          | Predicasts, Inc.                                                       | Lockheed SDC              |
|        | AMI                                                                  | 広告やマーケティングに有用な新製品, 消費性向などに関する情報                                                                                                                                                     | The New York Times Information Service Co., J. Walter Thompson Company | N.Y. Times                |
| 新聞情報   | NEEDS-IR                                                             | 日経新聞, 日経経済新聞, 日経産業新聞, 日経ビジネス, 日経エレクトロニクス, 日経コミュニケーションからの記事要約と抄録。                                                                                                                    | 日本経済新聞社                                                                | 電電公社 (1981/4 よりサービス開始)    |
|        | The New York Times Information Bank (The Information Bank)           | The New York Times などの米国の有力新聞, 雑誌などの 90 種からの抄録。                                                                                                                                     | N. Y. Times                                                            | N. Y. Times               |
|        | NATIONAL NEWS-PAPER INDEX                                            | The Christian Science Monitor, The New York Times, The Wall Street Journal からの記事抄録                                                                                                  | Information Access Corp.                                               | Lockheed                  |
|        | CNI                                                                  | カナダの主要新聞 5 誌 - Montreal Star, Toronto Globe & Mail, Toronto Star, Vancouver Sun, Winnipeg Free Press からの抄録                                                                          | Information Access of Toronto                                          | SDC                       |
|        | NEWSPAPER INDEX (INDEX)                                              | 米国主要新聞 7 誌 - The San Francisco Chronicle, The Detroit News, The Chicago Tribune, The Los Angeles Times, The Houston Post, The New Orleans Times Picayune, The Washington Post からの抄録 | Bell & Howell                                                          | SDC                       |
|        | QUEBEC-ACTUALITE                                                     | カナダ, ケベック州の新聞 3 誌 (Le Devoir, La Presse, Le Soleil) のフランス語による抄録。                                                                                                                    | Bell & Howell                                                          | SDC                       |
|        | RIT                                                                  | New York Times 紙, The Associated Press Washington Newswire service 産 50 以上の出版物からの政治, 経済, 社会の主要記事の抄録。                                                                                | N. Y. Times                                                            | N. Y. Times               |
|        | MONITOR                                                              | Christian Science Monitor 紙からの記事抄録。                                                                                                                                                 | Bell & Howell                                                          | SDC                       |
|        | LIBCON                                                               | 米国会図書館の図書目録情報                                                                                                                                                                       | 3-M Library Systems                                                    | SDC                       |
|        | CIS/INDEX                                                            | 米国会議記録, 各委員会報告, 国会文書                                                                                                                                                                | Congressional Information Service, Inc.                                | Lockheed                  |
|        | GPO MONTHLY CATALOG                                                  | 米国議会を含むすべての米国連邦政府機関が発行するレポート, 研究報告, 会議録等。                                                                                                                                           | U. S. Government Printing Office                                       | Lockheed                  |
|        | COMPREHENSIVE DISSERTATION ABSTRACTS <sup>®</sup> : CDI <sup>™</sup> | 1861 年以降のすべての米国立論文のリスト, 全分野にわたる                                                                                                                                                     | Xerox University Microfilms                                            | Lockheed SDC <sup>™</sup> |
|        | SOCIAL SCISEARCH                                                     | 1,000 以上の重要な高い社会科学領域の雑誌と 22,000 以上の自然科学からの重要な社会科学関連の文献の索引, 引用関係の検索を含む                                                                                                               | Institute for Scientific Information                                   | Lockheed SDC              |

種のデータベースが、簡単に利用できるメリットは大きい。国際電信電話(株)(KDD)のICAS(国際コンピュータ・アクセス・サービス)が開通して、米国の優れたデータベースが利用しやすくなってきた。三大文献情報データベース検索システムといわれるロッキード社のDIALOG SDC社のORBIT、そしてBRS社のBRS検索システム等を通じて、データベースがより身近に利用できるようになってきた。

## 2.1.2 文章情報データベース利用の問題点

前項に述べたように、データベースは多くの分野で作成されてきており、今後文章情報データベースに関する関心はますます高まり、利用は急速に広まるものとみられる。利用者にとって、内容豊富で良質なデータベースが低コストで利用できることが望ましい。このためには、データベースの作成・提供にあたって情報公開の促進、回線の共同利用など基本的な課題も少なくない。

文章情報データベース固有の(特に新聞雑誌等を情報源とする一般的な文章情報データベース)問題点のうち、利用者からみた主要なものをあげることとする。その際に、データベースの作成・提供、利用の各面から問題を考えてみる。もちろんこれらは相互に関連し合って利用者に影響を及ぼしており、それぞれの相互依存、相乗作用を考える必要がある。

### (1) 作成・提供上の問題

データベース作成・提供上の問題として、

- (i) 必要情報の全体量が不足している。
- (ii) オリジナル情報が不足している。
- (iii) 利用までのタイムラグが大きい。

等が挙げられるが、それらのほとんどは、入力時のコスト及び労力がかかりすぎることが原因しているといっても過言ではない。特にわが

国においては、欧米に比較すると文章情報のデータベース化は遅れている。その要因としては、日本語のわかちがき問題や文字種による問題が大きく作用していると思われる。また、データベース作成時におけるキーワード付け、抄録作成などはほとんど人手による作業のため、多大な時間と労力を要する。このため、タイムリーな情報提供ができにくい状況にある。これらのことを利用者から見れば、網羅性に欠ける情報源となる、全文による照合ができない、利用時におけるデータ内容の価値が低下する、等の不満を残す結果となる。

しかし、上記の問題はいずれも技術的進歩に伴って解消されつつある問題である。その例としては

- ① 朝日新聞社の NELSON、日本経済新聞社の ANNEX、サンケイ新聞社の SUCCESS 等の新聞編集システムにおける記事原稿のコンピュータ可読化
  - ② 印刷、出版分野での電算写植による全文のコンピュータ可読化
  - ③ オンライン利用による海外情報の容易な入手
  - ④ A P、共同通信社などの受信情報の活用
  - ⑤ O A 機器などの普及による抄録作成等文書編集の簡略化
  - ⑥ 特許情報センターの JAPATIC システムにみられるドキュメント類の直接入力によるデータベース作成の自動化
  - ⑦ JICST のキーワード自動抽出システム
- 等が挙げられる。

このような動向は、文章情報データベースの作成を有利にし、かつ蓄積情報を豊富にし、より多角的な使用を可能とするものである。

## (2) 利用上の問題

データベース利用上の問題としては、

- (イ) 情報の高度解析利用の問題
- (ロ) 自国語と他国語の問題

## い) 検索方法の簡易化と標準化

等が挙げられる。

利用者にとって従来の一般検索により得られる情報のみでは不足の場合が多い。にもかかわらず情報の時系列的な変化を分析したり、情報の傾向や特色を把握するといった高度な解析は、現状のシステムでは不十分であり、情報の高度利用は遅れている。また海外情報量の増大、海外情報に対するニーズの高まりにより、データベース利用の機会が増えている。しかし自国語に比べ他国語を利用する場合は、理解力が劣る、翻訳に時間を要する等の阻害要因がある。情報源となる言語も英語のみに限らず、各種言語に拡大する傾向にある。さらに検索方法についても、書誌情報などに使われる分類コードの不統一や、キーワード選定の難しさなど、検索時の煩雑さは初心者への検索能率の障害となっている。

しかし、これらの問題についても、いくつかの新しい試みがなされている。たとえば、

- ① JICSTにおけるシソーラス検索
- ② IBMにおいて実験中である自然語検索システム“やちまた”
- ③ 海外においては、医療用応答システム“MYSIN”の稼動
- ④ 米国議会図書館のLC-MARCシステムに見られるような検索コードの統一化の方向
- ⑤ 工業技術院情報計算センターにおける技術文献表題の英文和訳システムの実用化

等があげられる。

このような、文章情報解析技術の進展と実用システム開発の気運の向上は、今後の文章情報データベース利用をさらに促進するものである。

## 2.2 総合解析システムの必要性

### 2.2.1 情報需要の変化

情報化社会・知識集約型産業社会といわれてすでに久しい。複雑化した社会環境に応じて情報に対する需要も変化してきている。最近の情報需要として、次の3点に顕著な傾向を見いだすことができる。

第1の特色は情報需要の「国際化」傾向である。海外情報の需要は従来にも増して急激な高まりをみせている。例えばイラン革命、ポーランド危機等各地で発生する様々な事件は、国際的に大きな波及効果を及ぼしている。相互依存関係が深まるにつれて海外の政治・経済動向がわが国の政治・経済に及ぼす影響も大きくなっており、海外情報を的確かつ迅速に収集し、分析することが大きな課題になっている。

海外投資の増加、取引規模の大型化によって、海外の情報の需要者が大企業から従来関心が乏しかった中小企業にまで広がっている。海外との取引は「もの」中心の取引から技術・ノウハウなど「情報」の取引へと拡大し、情報の相互利用を積極的に促進することが要請されるようになった。

- ① 海外情報を得るため海外の新聞、雑誌等の輸入量が年々増加している。
- ② DIALOG, MARK IIIなど国際間のデータ通信システムを利用した情報サービスが普及して、国内で入手できる海外情報量は極めて大きくなっている。
- ③ 日経NEEDS-IRなどわが国のオンライン情報サービスシステムによるデータベースも、海外情報を得る一手段として利用されている。

こうした傾向はさらに増大すると予測され、海外情報に対する分類・索引の作成、抄録・翻訳等の処理が容易に行える必要がある。

しかし、前節で述べたように、膨大な情報量に対して人間が手作業で処理を行うには自ずと限界があり、コンピュータを利用した自動的な分類、インデックス作成、翻訳等の処理システムの開発が望まれている。

第2の特色は情報の「総合化」需要の傾向である。従来の文章情報の検索はデータベースの中から個別の情報をとり出すことが中心であった。今後は従来の検索につけ加えて、データベースに蓄積された文章情報の内容

を把握し要点を分析することが極めて重要視されている。

たとえばカントリー・リスク，経済安全保障，通商摩擦等のテーマにとり組むことを考えてみよう。これらのテーマに対処するには個々の断片的な，情報を抽出するだけでは問題点を理解することが容易でなく不十分である。個々の情報では把握しにくいこれらのテーマに関係する情報を分析して内容の傾向や特徴を簡単に取り出せるようにすることが必要である。また，利用可能な情報量が膨大になるにつれて，利用者はすべての情報にあたってみることは不可能である。大きな政治・経済・社会問題をとくに表出する前にあらかじめ予知できるようにするには，データベースの中の情報から何らかの方法によって，見落としがちな変化やわずかな兆候を発見することが重要となっている。

このような需要に対し，最近の情報解析技術の進歩と，分析に適したデータベースの構築がすすめられていることから，情報の総合的利用が行えるシステムの開発への期待が高まっている。

第3の特色は情報需要の「多様化」傾向である。この傾向は従来からみられるが，近年著しく強まっている。

その背景としては情報を必要とする需要者の拡大が挙げられる。大企業に偏りがちであった情報需要者は中小企業，家庭・個人へ広がりをみせている。企業内でみると，特定の部門（たとえば調査部門）に集中しがちであったものがあらゆる部門に拡大している。とかく専門家に限られるきらいのあった文章情報データベースの利用者が，いわゆる情報の“素人”にいたるまで情報を求める層が広がっている。

このような，需要者層の広がりに伴って，情報需要者の利用目的・利用方法は急速に拡大してその要求内容は極めて多様化している。個々の需要者の要求を満たすためには，従来の利用システムとは異なって，よりキメの細かなサービスが必要である。特に需要に応じて処理方法・提供方法・提供媒体・提供形態等を，利用者が自由に選べるような需要者オリエンテ



ッドなサービスができるシステムの開発が求められている。

## 2.2.2 総合解析システム化の方向

前項で述べたように、最近の情報需要は新しい傾向を示している。この傾向は今後さらに強まるものと予測できる。

このため、需要者が求めている情報の質・量ともに拡大し、需要に比例してデータベースの作成がすすみ利用できる情報量は飛躍的な増加が見込まれている。また、需要者が求めているサービスはデータベースから検索して提供をうけるだけではない。さまざまな目的に応じた需要者本位のサービスを要求している。

しかし、現状の問題点（2.1.2項）で指摘したように現在のデータベース利用システムで対応するには限界がある。迅速かつ的確に処理してニーズに合った個々の情報を提供するだけにとどまらない。個々の情報を利用者が望むかたちで提供できるよう加工分析し、そのうえ、それらの情報を有機的に組み合わせて解析し総合化できる処理システム — 総合解析システム — の開発が不可欠となっている。

総合解析システムが持つべき主要な機能は以下のとおりである。

- (1) 文章情報を処理するための言語処理能力
  - ・ 利用者が希望する言語で情報を入手できるための翻訳機能。
  - ・ 利用者の求めに応じて、長文の文章情報を要約あるいは抄録にして提供できる機能。
  - ・ 利用者が検索しやすいように、蓄積対象の文章情報のキーワードを設定したり、索引を作成する機能。
- (2) 文章情報を分析して内容を把握する解析能力
  - ・ 情報内容を分析して特徴をとらえる機能。
  - ・ 情報内容の時系列的分析ができる機能。
  - ・ 蓄積された情報内容の傾向をとらえる機能。

- (3) 文章情報のなかから数値データ部分を抽出する能力
  - ・統計資料のようにまとまっていない断片的な統計数値、予測値等を文章情報から抽出して利用できる機能。
- (4) 文章情報を分析して内容を数値化して処理する機能。
- (5) 文章情報から数値化したデータと、数値データベースのデータを組み合わせて分析する機能。

## 2.23 文章情報データベース高度利用による効果

文章情報データベースを総合解析システムや、内容分析手法等を駆使して高度に多角的に利用することによって、以下のような効果が期待できるであろう。

- ① 国際政治情勢の分析
  - ・カントリー・リスク分析
  - ・エネルギー問題分析
  - ・通商問題分析
  - ・経済安全保障分析       等
- ② 企業経営・産業政策等意思決定への支援
  - ・マーケティングへの応用
  - ・新製品開発計画への応用
  - ・広告・PRへの応用
  - ・国際会議の運営補助       等
- ③ 公共サービス
  - ・海外情報の日本語翻訳サービス
  - ・図書館のインデックス発行サービス       等

この他、利用していく過程において更に有効な使い道が数多く見い出されると予測され、その種類や内容は無限に含まれているといっても過言ではないと思われる。

## 2.3 文章情報解析手法の研究

### 2.3.1 定量化利用の発展過程

報告書、議事録、新聞記事などの大量の文章情報を有効に活用するためには、文章情報を数量化して扱う技術が不可欠である。そのため、いわゆる内容分析の方法が注目されてきている。内容分析は、文章化された情報を個人の主観的判断に頼らずに、体系的かつ数量的に分析し、そこに必ずしも明示的に表現されていない筆者の意図を推論することを目的としている。例えば、一連の論説が、特定の方針に対し、賛成の立場をとるのか、反対の立場をとるのか、それとも中立の立場をとるのかを判断したい場合に、内容分析はその客観的な判断基準を提供してくれる。

近年、内容分析は国際政治の研究分野で盛んに利用されてきたが、その理由の1つには、内容分析のもつ、厳密で数量的な史料操作の有効性がある。国際政治の分野で過去の政策決定のプロセスを数量的に分析する場合既存の史料（公文書、政治の発言記録、新聞記事）の数量的分析が必要だからである。さらに内容分析が関心をひいたのは、国際政治の中でのコミュニケーションの役割が認識され、メッセージが当事国あるいは相手国との関係において、さまざまな形で分析されてきたからである。

内容分析は国際政治の分野ばかりでなく、心理学、精神分析学、民俗学、文学、歴史学、言語学など多くの分野で研究、活用されてきた手法である。

内容分析と呼べる手法が最初に使われたのは、1900年代のアメリカの新聞調査である。当時の調査は、スポーツやファッションといったトピックスに払われる関心を測るのに、その話題にさかれているスペースの大きさ（コラムのインチ数など）、紙面の位置、見出しの大きさなどの形で数量的に分析するのが普通であった。

政治学の分野で内容分析が用いられはじめたのは、1920年代から30年代にかけてである。ラスウェル、ライツなどが、主にプロパガンダの研

究で、文章内の言葉や主題の頻度を数えるという形で数量的な分析を試みた。

第二次世界大戦中は、米国の国家的要請もあり、内容分析は敵のメッセージを分析し、敵国の将来の行動を予測することに重点が置かれた。この時期は宣伝の裏にひそむ意図の発見という点で長足の進歩をしたが、理論的な進歩はあまりなかった。

第二次世界大戦後の1950年代に入ると、単に言葉の数を数えるだけでなく、その言葉の持つ意味の内容まで考慮した内容分析の手法が、オズグッドにより開発された。これが2.3.2の(3)で説明する評価主張分析(Evaluation Assertion Analysis)である。

1960年代には、ノースはこの分析法をベースにした、コンピュータを使用した自動内容分析法を開発し、国際政治に応用し多くの成果をあげた。ノースの開発したこの分析法は内容分析の発展の可能性を示すものと言える。

## 2.3.2 内容分析の方法

### (1) 頻度分析(Frequency Analysis)

頻度分析は内容分析の手法の中で、最も古くから用いられてきた基本的な手法である。

この方法は、文章の中に繰返し出現する言葉または主題の頻度を分析することにより、筆者の意図とか、興味の対象を分析しようとするものである。ある主題が記されている記事の件数を時系列的に調べたり、またその記事の書誌データとのクロスセクションを調べたりすることにより、有益な結果が得られる。

頻度のカウント方法を改善することにより、内容分析の精度を高め、より効果的に使用することができる場合が多い。その1つは、カテゴリーを構成する言葉に、その重要性に応じたウエイトを与え、ウエイト化

された頻度をカウントする方法である。言葉の重要性を測定する方法については、オズグッドなどにより、各種の方法が研究されている。

つぎは言葉の分布に関する問題である。出現頻度が高いカテゴリであっても、それが1つのセクションに集中している場合は、そのカテゴリが文書を代表しているとは言い難い。理想的な指標は、出現の頻度と分布の両方を考慮したものでなければならない。

## (2) 関連性分析 (Contingency Analysis)

関連性分析は、頻度をカウントするということでは、(1)で述べた単純な頻度分析と同じであるが、その目的および形態は異なっている。頻度分析が、ある言葉が何回出現するかを問うのに対し、関連性分析は、ある言葉が他の言葉との関連において何回出現するかを問う。前者が単純な確率を問題にするのに対し、後者は条件付の確率を問題にする。

つまり、どの言葉と言葉、主題と主題が同じ文章中に出現するかを調べることにより、例えば情報の送り手の連想語をつきとめ、それから送り手の連想のパターンを推測することが可能になる。

2つの言葉または主題の同時出現を記録する文脈の単位としては、“XとYが5語以内にある”というように単に語数で区切る方法、“Xの前5語以内にYがある”というように方向をも指定する方法、文または文章の段落、文章全体という自然な単位を指定する方法、さらには“XがYの修飾語であれば”と云ったシンタックスで指定する方法も考えられる。

同時出現の測定の結果は図2-1のように関連性マトリックスとして表現することができる。このマトリックスの対角線の左下半分には、実際に測定された同時出現確率が、また対角線の右上半分には、例えばカテゴリAとカテゴリBが同時に出現する期待値が記入される。この期待値は $P_A \cdot P_B$ で計算される。

もし実際の同時出現確率が、その期待値より大であれば、こうした事

象は偶然以上によく出現する、言いかえれば、たとえばAとBは関連性があると解釈される。偶然との有意性の検定はカイ自乗検定などの統計的な方法により検定することができる。

この関連性マトリックスを用いた各種の分析手法が開発されている。

|   |       | カ テ ゴ リ ー |       |       |        |
|---|-------|-----------|-------|-------|--------|
|   |       | A         | B     | C     | .....N |
| カ | A     | —         | •08   | •24   | •32    |
| テ | B     | •06       | —     |       | ← 期待値  |
| ゴ | C     | •38       | •02   | —     |        |
| リ | ..... | .....     | ← 測定値 | ..... |        |
| 1 | N     | •18       | ..... | ..... |        |

図 2-1 関連性マトリックス

### (3) 評価主張分析 (Evaluation Assertion Analysis)

ある主題の出現頻度を調べるだけでは、その主題が筆者によって好意的に扱われているのか、中立的なのか、あるいは敵対的なのかを知ることとはできない。このような筆者の対象に対する態度の強度を測定するために考案されたのが、オズグッドの評価主張分析である。

オズグッドは、言葉の意味をその使われ方に着目し分析した場合、多くの言葉に共通する基本的な意味が見出されることを発見した。オズグッドはそれを、

「よい—わるい」で代表される Evaluation (E)

「つよい—よわい」で代表される Potency (P)

「Active—Passive」で代表される Activity (A)

と名づけ、各言葉の方向及び強度を、各E、P、Aについて、+3から-3の7段階の評価値で表わした。

評価主張分析では、まず文章を、

- ① 誰によれば
- ② 誰を
- ③ 何に関して
- ④ どう評価している

の4つの要素を含む命題に分割する。つぎに各命題単位に、命題に含まれている評価語のE, P, Aの評価値を求める。命題に2つ以上の評価語があれば、平均をとるなどの処理がなされる。つぎにこれら命題を、主体または客体の区分ごとにまとめ、その平均値を求める。こうすることにより、AのBに対する態度の評価値が得られたことになる。

### 2.3.3 内容分析データの高度利用

ある文章、メッセージについて内容分析を試みる場合、その分析目的はそれらの情報に表現されたメッセージの送り手の意図、方向性、態度の強さなどをより科学的に、数量的に把握しようとするものである。

ところで、ある文章、メッセージを分析する場合、我々はその意図の経済的、社会的、政治的状況に関連づけて分析しなければならないケースが多い。特に複雑な社会現象を分析する場合には、多くの変数間の関係を同時に分析しなければならない。こうしたケースに、内容分析と合せ良く用いられる手法に 多変量解析、数量化理論などの分析手法がある。

また複雑な国際関係を分析する場合には、メッセージの内容分析と共に、シミュレーション・モデルなども有効に用いられてきている。国際社会環境におけるほとんどのメッセージ、すなわち内容分析の対象となる情報は、その受け手の反応を考慮して、送り手の意図や方針が表現されている。換言すれば、コミュニケーションを行っている二者又は二国間が、互いに相手の次の行動を自己の価値体系の中で予測しながら、メッセージを送っていると言えよう。

そこで、自己の意図を相手に正確に伝達し、かつその目的、例えば輸出

の増大、緊張緩和を達成しようとするとき、何らかの形で相手の対応行動を予測しなければならない。このような状況下ではシミュレーションが有効な手段となる。

各種の文書、メッセージを内容分析の手法を用いてその意図をさぐり、統計的手法により、相手の置かれている環境を把握し、自国のもつ選択肢のうち、最良のものを見出すために、その選択肢を変数としたシミュレーションを実施するといった、総合的なシステム体系の構築が必要とされてこよう。



### 3. 文章情報データベース総合解析システムの基本構想

#### 3.1 総合解析システム像

##### 3.1.1 総合解析システムの概念像

前章でも述べた様に、文章情報のコンピュータ可読媒体化、処理技術の向上、及びハードウェアの機能向上、低コスト化は、今後ますます進展するものと思われる。

本事業では、それらの条件をふまえ、コンピュータを有機的に利用することにより、大量の文章情報を分析し、事態の客観的把握、さらに政策立案、決定を行う際の有力なサポートシステムを作成することを目指すものである。但し、初めから高度なレベルのシステムを目指すものではなく、既存技術をもれなく応用し、技術の進歩を適宜とり入れられる拡充性を有したシステムとする。

まず、内外の新聞、本、文献、報告書などの文章情報を解析し、その後の対応が効果的に行えるような文章情報データベース総合解析システムとはどのようなものであるかということについて考えてみた。

概念図を示すと図3-1のとおりである。

ここでは基礎解析、内容検索・分析の部分がメインとなり、知識ベース作成及び翻訳部分はそのサポート機能となる。今後は日本から海外へ情報を出す必要性が生じたり、これまで以上に海外の情報分析の必要性が高くなることが予想され、翻訳システムの重要性が、今後さらに増加することが考えられる。

しかしながら、どの部分をとり出しても、いくつかのレベルがあり、解決すべき問題も含まれている。

それらについては3.3に詳述する。

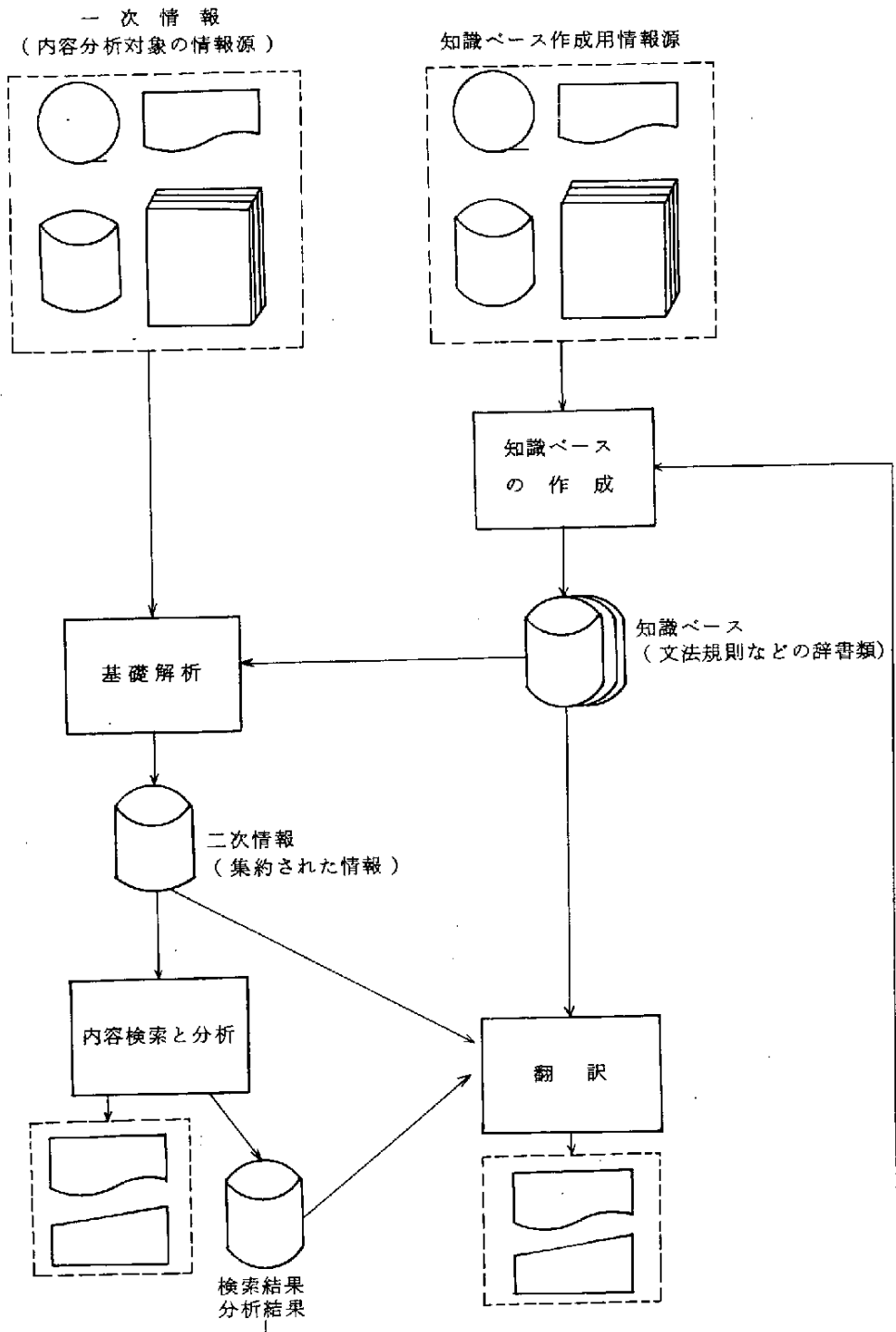


図3-1 文章情報総合解析システムの概念図

### 3.1.2 基礎解析

基礎解析の目的は、一次情報（広義の内容分析対象とするもの）に情報の抄録、インデックス、キーワード、構文解析や意味解析結果、文章中に含まれる数値情報の抽出、単語頻度数などの一次情報の集約された情報（二次情報）を付加し、データ構造の一元化や文章情報の定量化を行うことによって、次のステップである内容分析が効果的に行えるように一次情報を解析し蓄積しておくことである。

本事業で目指すような総合的な解析システムが存在しない現状では、前述した二次情報の中で、どういう解析が効果的であるかという結論を出すのは難しい。従って、当面は翻訳技術のノウハウを利用した一次情報の一元化と簡単な定量化にとどめ、効果的な内容分析の手法が明確化して来た段階でよりよい二次情報を作成する方向で順次発展的に検討するのがよいと考える。

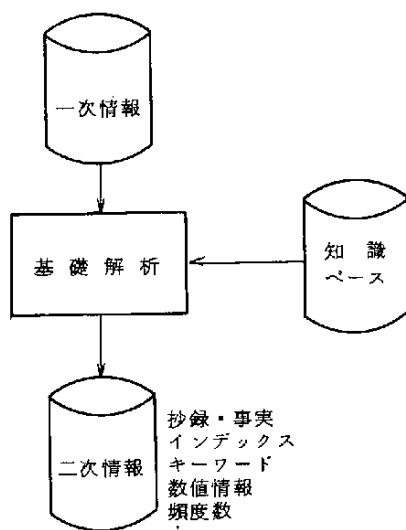


図 3 - 2 基礎解析の概念図

### 3.1.3 知識ベース作成

知識ベースの作成は、基礎解析、内容分析および翻訳に必要な辞書類や

シソーラスを作成するものであるが、今までの辞書やシソーラスにはない意味、属性を含むことで、より事実に近い抽出を可能とするものである。

従来、辞書やシソーラス等の作成は、手作業によるものが多く、期間や費用の面で膨大になったり、担当者や作成時期によって、同一データから作成したシソーラスが異なることが生じたり、時間を経るに従って辞書やシソーラスが陳腐化する傾向にある。このような状況を改善するため、今までの翻訳システムの技術、経験を生かし、コンピュータを利用して、辞書やシソーラスなどの知識ベースを作成することによって、期間および費用の削減やそれら自身の増殖の可能性を容易にして、陳腐化を妨げることが当システムの目標とするところである。また、最近のハードウェアも含めたテキストプロセッサ技術などの向上を併せて考えると、当システムの開発には好条件であると言える。

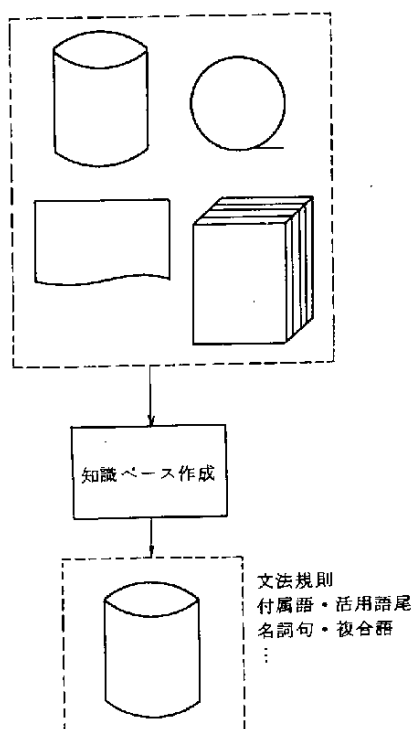


図 3-3 知識ベースの概念図

#### 3.1.4 内容検索と分析

内容分析は、基礎解析の次のステップとして行われるものであって、次の(1)、(2)のようなことがらが内容分析の効果としてあげられる。また、これらの効果がより一層に有効となり、利用者にとって利用しやすい解析システムに機能アップされたり、改善されて行くためには、以下の(3)～(5)に示したことが利用者からの要求に基づいて随時行われる必要がある。

- (1) 二次情報から利用者が内容分析したい情報源を抽出する。(母集団の抽出または内容検索)
- (2) 抽出した母集団に対して、利用者独自の分析方法をマンマシン・インターフェースで実行したり用意された手法(例えば、統計的手法)を利用して分析する。
- (3) 母集団の再抽出(変更)を行う。
- (4) 利用者が上記の様に再抽出したり、データを整理するために辞書、ソースラスなどの知識ベースの増殖機能や追加または変更できるマン・マシン・インターフェースを充実させる。
- (5) 抽出された結果を分析する手法の充実。

内容分析の現状は、抽出したデータをグラフ化したり、統計的手法で頻度数などを整理し、その結果を利用者独自の考えに基づいて分析し、結果を導き出している場合が多い。このことを考慮して、当面のシステム化の課題としては、情報の一般的検索と統計的分析手法のみとするが、一般的検索結果から行った分析手法のなかで効果的な方法であるという評価ができるようになったものについては、随時、機械化し、分析方法を充実させていくことを考える。

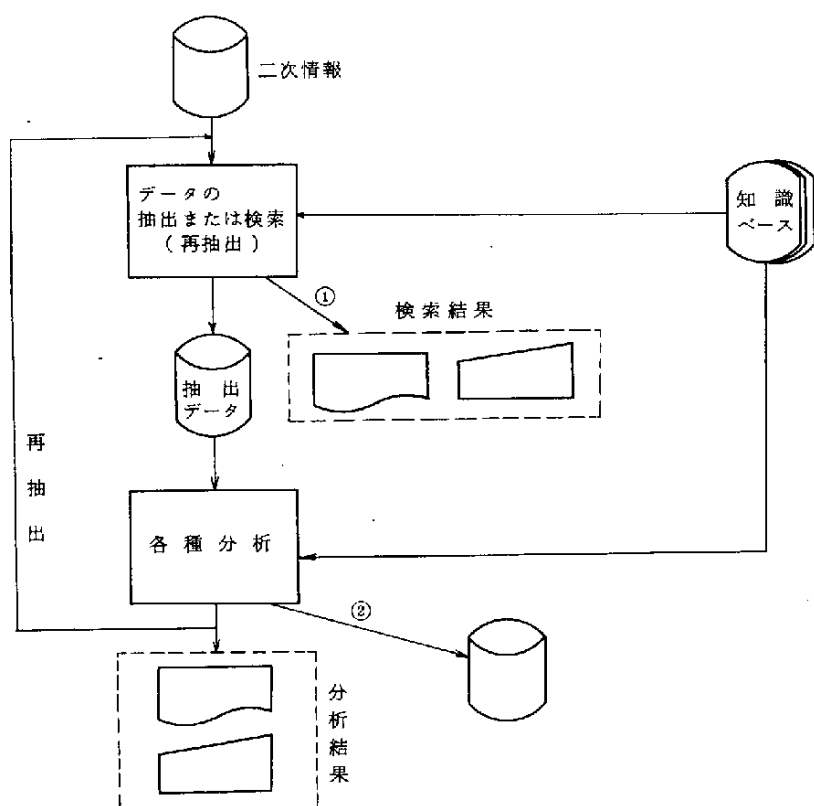


図 3 - 4 内容検索と分析の概念図

- ① 二次情報からデータを抽出する検索のルート。必要ならば知識ベースも利用する。
- ② 分析結果を他システムで利用する場合（例えば、翻訳システム）の携帯性を考慮したときのルート。なお、抽出データについても他システムで利用することが考えられる。

### 3.1.5 翻訳について

情報には、日本語の文章ばかりでなく海外のものも存在する。内外における日本の状況を考慮するならば、日本語→外国語、外国語→日本語の両方向への翻訳が重要なことはいうまでもない。日本語↔欧米語以外の翻訳は、翻訳者が少ないという点をカバーしたり、巾広い情報を収集するとい

う意味で機械翻訳の必要がある。また、日本語⇄欧米語の翻訳は、多量のデータから早く、適切な情報を得るというためには機械化の必要がある。

このシステムにおける翻訳機能では、海外の情報を日本語で検索し、分析するためには日本語への翻訳が必要であり、結果を海外へ知らせるというためには、日本語→他国語への翻訳が必要となる。

## 3.2 システム機能

### 3.2.1 システム機能の概要

内外の新聞・雑誌等、文章情報有効利用の要求が高まってきている。このためにはコンピュータ処理の過程で個別データを取り出す一般的な情報検索機能に加えて、文章情報データベースから何らかの潜在的な事象・傾向をとらえられる統計的・意味論的手法による総合解析機能が重要となる。このために具備すべきシステム機能は以下のとおりである。

これらの文章情報データベースを検索・解析利用するためには、あらかじめ二次情報として圧縮加工して蓄積しておく必要がある。一般的にはこの作業は大部分を人手にたよっているが、しかし、人手では多大な経費・期間を要し、かつ個人差などによる精粗のばらつきに問題がある。これを避けるためには、可能な限りコンピュータにより自動化すべきである。この場合、現状で不可能な部分は人間が介在して担当し、将来技術の進展をみながら段階的に機械側に移行していく方法により実用的なシステムを目指すことが重要である。

二次情報への加工機能として次のものがあげられる。

- ・ 単語の切り出しと品詞認定
- ・ キーワードの自動抽出
- ・ 事実の抽出及び自動インデックス作成
- ・ シソーラスの自動生成
- ・ 数値情報の取り出し

これらの二次情報は、一次情報と共に蓄積され、情報の検索・分析に利用される。

検索機能は通常のキーワード、シソーラスを用いた検索に加えて、キーワードの文中での役割りを加味した検索機能によって検索効率、精度を上げることができる。検索機能により文章情報中の書誌的情報、記事抄録・全文あるいは分析を行う主題に関する母集団データが出力される。

分析機能はキーワード(単語)の出現頻度をベースとした出現頻度分析・時系列分析・相関分析あるいは文や文章を対象とした意味論的分析により、大量の文章情報の中から、諸現象間の関係を明らかにしたり、共通の意味内容をひき出し、総合判断の素材を提供する機能である。各種の分析手法をパッケージ化しておき、容易に利用可能にしておく必要がある。

近年、機械翻訳も実用に近づいてきているので、検索結果の出力にあたって欧米語の書誌的情報、抄録、全文などは機械翻訳システムにより日本語に翻訳し、対訳的に出力表示することにより利用の便宜をはかることができる。また、検索質問中のキーワードを日本語で指定して、欧米語の文章情報データベースを検索できる機能もあれば便利である。

文章情報の中には、数値情報が含まれている場合が多い。これらの数値情報は速報性があり、また文章情報から得られる情報を補足する役割もある。これらの数値情報を抽出して、適当なキーワードを付して蓄積し、有効利用する機能も考えられる。

### 3.2.2 機能別サブシステム

#### 3.2.2.1 基礎解析

基礎解析は、検索・分析機能のために文章情報の内容を解析して、二次情報として圧縮加工・蓄積する機能である。

##### (1) 一次情報の入力

すでに機械可読となっている一次情報は、本システムでの蓄積・利



用のために標準化しておく必要がある。すなわち各情報源の情報構造は固有の形となっているので、以後の利用のために、これを統一した情報構造へ変換する機能であり、一次情報源として全文を蓄積したり、基礎解析の入力とするためのサブ・システムである。

## (2) 単語の切り出しと品詞認定

文章中から単語を切り出す機能は、以後の単語をベースとした処理の前提となるものである。欧米語文章の場合には、単語の間はスペースで区切られているので、単語切り出しそのものは単純な機能である。一方、日本語文章の場合には、わかち書きなどのない自然文形式を対象としているので、単語の切り出しは高度な処理となるが、人手では画一的な繰返し作業となるので、当然機械による自動化を考える必要がある。

単語認定の方法は、1つの方法のみで完全にはできない。文字種の変り目、辞書マッチング及び文法規則により付属語・活用語尾などを検出して、他の語を文型パターンから認定する方法などを折り混ぜて、精度を上げることができる。

また、この過程で辞書マッチングのとれなかった語を未登録語としてファイルしておき、単語管理に利用することも考えられる。

## (3) キーワードの自動抽出

文章情報データベースの内容を検索あるいは分類する際に使用するキーワードは、通常、人手により適切なキーワードを選定して、これを索引に使用している。しかし、膨大な量の情報源が対象の場合、人手では費用、期間がかかり過ぎ、また個人差などによる違いが生じるという問題がある。したがってキーワード付けにおいても機械により自動化する必要がある。

一般に、文章中でその主題との関連の深い語は出現頻度が高いので、文章中でのその単語の出現頻度を調べ、統計的処理により、ある一定

の基準値の範囲内にある単語をキーワードとして自動的に抽出することができる。この場合、付属語（助詞・助動詞 接続詞、英語の冠詞・前置詞など）は出現頻度は高いが、明らかに主題とはあまり関係がないので不要語として捨てることになる。

また自動抽出の効率を上げるために、活用語尾変化や同義語を代表語に置き替えてキーワードとすることも必要である。

更に、自動抽出したキーワードの妥当性の確認のために人手によるチェック機能を組み込んでおくことも必要である。

#### (4) 事実の抽出及び自動インデックス作成

抽出したキーワードに関して、該当文での述語処理を施し、主体・客体・場所・時間・手段などの意味的役割り属性を付与して、レコードとして収納する。これらの文レベルでのレコードを文章レベルで組み合わせ、1つの構造へまとめ上げ、インデックスを作成する。この処理は、人間の介在により曖昧性の解決や編集を手助けしてやる必要がある。

#### (5) シソーラスの自動生成

一般に、検索効率を上げるためにシソーラス辞書が使用されるが、このシソーラスの作成にあたっては、人手による個人差を避けるために、機械による自動生成機能を用意することができる。

これは、ある語とある語の同じ文献中での出現度合を調べ関連係数を求めて、確立モデルなどの統計的手法により階層関係や関連概念として、シソーラスを自動生成する機能である。この場合の自動生成の精度を高め、より実用をめざすために、機械による曖昧さを人間が補助するヒューマン・エイドの操作を前提とすべきである。

#### (6) 数値情報の取り出し

これは、文章情報の中に埋もれている数値データを構文解析により抜き出して、適当なキーワードを付与して蓄積し、有効利用するため

の機能である。これらの数値データは、定期的に発表される統計・指標データとは異なる観点での利用が期待できる。

### 3.2.2.2. 知識ベース

これは基礎解析・内容検索・内容分析あるいは機械翻訳の際に使用される辞書類や文法規則などの集合である。また、より高度な利用を実現するためには、一般知識をも含まれよう。そして、これらの知識ベースの維持管理を行い、常に質的拡充をはかるためのサポート・システムの完備が重要である。

知識ベースに含まれる辞書類には、概略次のものが考えられるが、必ずしも明確に区分できるものではない。

- (1) 単語切り出し用辞書 — 現在、国語辞書が機械可読となっているので、これを一般語として、あらかじめ登録しておくことができる。また専門用語、活用語尾、文法規則なども含まれる。専門用語などは、一次情報により増殖が可能である。
- (2) シソーラス辞書 — 主題別のシソーラス辞書であり、一次情報により増殖・改訂することにより陳腐化を防ぐことができる。
- (3) 機械翻訳用辞書 — 機械翻訳時に使用される辞書であり、一般の辞書をベースにファイル化されたものに加えて、特定分野辞書も必要となる。

### 3.2.2.3. 内容検索と分析

#### (1) 内容検索

キーワード、シソーラスを用いて該当する文献の個数、書誌情報・記事抄録及び記事全文などを出力する通常の情報検索機能である。

また、上記の機能に加えて、キーワードの文中での役割属性を加味した検索も可能である。更に、この機能によって後述の統計的分析用の入力情報の母集団を得る機能も重要である。

運用面では、一次情報が入力される都度あらかじめ用意された特

定テーマに関しての質問に答える S D I ( Selective Dissemination of Information ) 的な検索機能も必要となろう。

## (2) 統計的分析

前述の二次情報の中から、必要に応じて、検索機能で特定主題に関するデータを抽出して、キーワード(単語)の出現頻度・出現属性をベースにして、各種統計分析手法によりキーワードの示す諸特性同志の種々の関係を明らかにすることにより、情報の集約・整理を行う機能である。これによって、膨大な文章情報の中から、現実の新しい徴候を見い出したり、仮説の立証のための素材を得ること等ができる。

分析の特性データとしては、単語レベルのみならず、文・文章・テキストレベルでの概念を対象とすることも考えられる。

分析手法としては、下記のものが考えられる。

- ・ 各種の多変量解析(主成分分析など)
- ・ 数量化理論
- ・ グラフ理論など

運用面では、これらの各種の分析手法やデータを自由に変えて試行錯誤が可能のように、手法のパッケージ化やオンラインでの利用を考慮しておく必要がある。

## 3.2.2.4. 機械翻訳

検索結果の出力にあたって、欧米語の書誌的情報、抄録、全文などは指定により日本語での表記もとれるように、機械翻訳システムにより日本語に翻訳出力する機能である。この場合に、機械翻訳の結果を手直しする手間を省くために欧米文も対訳的に並記した形式が望ましい。

また、欧米語の文章情報データの検索質問中のキーワードを日本語で指定できる機能も利用者にとっては便利である。

### 3.3 システム開発への課題

#### 3.3.1 文章情報の特性

前節までで提示された総合解析システムの構想は、多様化する情報資源の効果的活用のために、大量かつ多種多様な情報の中から意図するものを把握することを可能とするものである。その為には、文章情報の様な豊富な内容を有しつつも、不均一であり、曖昧性をもつ情報を計算機上で、明確に論理的な情報構造として表現し操作する技術が必要不可欠となる。そのような技術の上になつて初めて、①利用者による文章情報などに対する内容分析の機能、②計算機が一次情報からその内容を抽出し、加工・蓄積する基礎解析の機能、③一次情報や加工・分析結果の諸情報を各国語に翻訳する為の自動翻訳機能、④これらを可能にする為の知識ベース等の諸機能が実現されると言える。

#### 3.3.2 文章中に出現する語の利用の為の課題

システムが文章情報をどのような情報構造として把えるかということは、そのシステムが文章情報からどのような情報を抽出し蓄積するのかということと一体である。自動インデックス作成、ソース検索、定量化分析を行うためには、文章あるいは文中での語の出現をもととして文章情報を把えるという手法が有効であろう。入力情報中で出現した語を収集し、編集することによって、語の出現頻度の蓄積や性格付けなどを実現することができる。このとき、出現した（あるいは抽出された）語の各々に属性を付与することが必要である。そのような属性情報としては、入力された各種書誌的情報と入力に対してあらかじめ先験的なものとして与えられる知識ベース中の情報との二つのレベルが考えられる。ただし、これらは独立なものではない。出現した語に対して辞書から属性を抽出するにあたって

は、照合された語そのものとともに、その語が含まれた文章情報の属性も考慮されなければならない。また、出現した語の集合に対しては、文章情報や文に対する語の同時出現と言う関係も、付与される。

これを計算機上で表現する為には、各種頻度や包括概念を用いたサマリー、属性などから生まれる自然なインバーテッドストラクチャ、同時出現や同一属性、それに概念関係などによるレコード間のリンクと云った事柄を反映したダイナミックな情報構造を実現しなければならない。その為にはデータ構造としてのリレーショナル モデルの概念、数値情報などで現在インプリメントされているディシジョン・サポート・システムにおける経験なども援用することができる。

このようなシステムを実現する上では、十分な規模をもつシソーラス辞書、文法規則等の知識ベースを用意する必要がある。

システム開発は当初から大規模な辞書を用いることを前提にするのではなく、単語抽出等を行うに足る比較的作成が容易にできる辞書から整備し、次のステップとして、語形変化熟語や複合語の問題を解決する形態素レベルの処理を行うための大規模な辞書の作成とその辞書を利用するシステムを開発する必要がある。

抽出、インデクシング、解析といった必要なプロセスに辞書を利用する場合、形態素レベルでの認定の曖昧度の問題など解決すべき技術的課題は残るにしても、実用上、種々のソフトウェア技術や運用体制で適切に対応することにより、アルゴリズム上の問題は殆ど解決することができる。むしろ、辞書、シソーラスのメンテナンスの問題に留意しなければならない。日々に増大し、その意味、関係も分野と時間経過により変化する単語や、処理中で発生する未登録語の処理を自動的に行うシステムの構築が必要である。

そこで、辞書、シソーラスの作成に当っては、まず堅固な予備分析からはじめる必要がある。すなわち、そのフレームの設定、概念のカテゴリー化

(属性付与規則の決定)および辞書中のリレーションの策定を行うとともに、語が多義性をもち、適用分野によってもその優先度が異なるという点を反映させる手段を考えなければならない。このような辞書作成の為の分析作業を、その辞書ファイルの使われる用途に応じて行うことにより、はじめて実用に耐えうる辞書を作成することができる。

### 3.3.3 文章の意味内容の解析

前項で述べた語の出現に基づくシステムは、元の文章情報に表現された豊富な情報の中で、その極く一部をなす情報をとりあげてきたにすぎない。

内容分析、内容検索で語を対象として扱う場合においても、文中でのその語の使われ方や役割が重要になることも多いであろう。例えば、同じ「レーガン大統領」であっても、「レーガン大統領が批判した」と「レーガン大統領を批判した」では、同じ扱いはできない。

また、文章情報から「賛成」、「反対」という2語が抽出されても、「賛成」と「反対」があったのか、「賛成」から「反対」に変わったのかなどは判らない。

更には、文章情報の中である人物や団体が肯定的に扱われているのか、否定的に扱われているのかという点は、語の出現はもとより、その語の出現する文単体で明示的に示されるとは限らない。

また、文章情報の主題、主張をとり出す為には、語をとり出しただけでは殆ど無力であるだけでなく、主題の語自体が単語として表われているか否かも疑問である。

このように、語の出現現象をもととした処理では、その不十分さが統計的に丸められるといった類のアプリケーションの場合に、その妥当性のある程度検定しながら用いるといった考慮が必要である。

発展的な文章情報データベース総合解析システムの実現過程は、文章情報中から抽出された語の集合にかかわる諸技術をもととしたシステムから出発し、文章情報の内容の自動把握技術を順次取り込んでゆくものとなる。

それでは、文章情報からの意味内容の抽出をどのようにして、実用というフェーズで実現していくかを見ていこう。

文学作品などは別にして、各種業務で必要な事実情報を扱う場合、文中の述語とそれが支配する格という構造で文を把えるのが、最も一般的に考えられている手法である。このような文の構造の自動的な把握は、現状の構文解析技法を用いることにより、十分実用に耐えうるであろうことが予想される。勿論、ここで構文解析は「文法論」の範囲内に拘泥する必要はなく、意味論的、語用論的なメカニズムをも取り込んだものになる。

このような格支配の構造を考えると、述語（『行為』など）に対し、それが受ける格が分類される。例えばある動詞に対しては、それが支配する『目的語』や『（前置詞）＋（意味カテゴリ）』（日本語ならば、『（意味カテゴリ）＋（助詞）』など）の集合を定めるということである。

これは裏返すと、名詞などの意味カテゴリを決定しておくという事でもある。そのような意味カテゴリは、個々の名詞（個体、概念など）の上で、構造をもったレコードとして表現されよう。（これが、シソーラスの構造と分類・内容を定めることにほぼ対応しているということにも注意されたい）ここで、名詞あるいは複合語の問題にも着目しなければならない。種々の属性をもった名詞がどのような規則でまとめ上げられ、総体としてどのような意味カテゴリになってゆくかという点はシステム開発に際して検討する必要がある。また、述語として行為が表現されるだけでなく、名詞句や名詞自体がその文での叙述の中心となるべき述語の役割を担う場合が多い。例えば、「会議が行われた」は、「行う」を行為として取り上げても、何ら情報としては有効とされない。この場合鍵は名詞の中に総て含まれている。この意味でも単なる構造の解析だけでは、実用上意味がないことになる。

これらの事は、文を処理単位とした技術であるが、実際の処理対象となるのは文章レベルである。内容分析などのために文章情報の表現している



ところのものを得るには、文の並び、叙述の並びを一つの構造としてまとめ上げる必要がある。この過程も純理論的にはいくつかの研究が為されているが、本システムの開発においては、ある意味で割り切った分析により、実用化を図らなければならない。

個々の文には、上に述べたように、解決すべき問題は残っているものの、セマンティック・ネットワークやフレーム構造などで表現することが出来る。システム・インプリメントに際しての分析により指示代名詞や語の省略（不足している「必須格」、他の情報により埋め込み可能になっている「任意格」など）に対する文章中の他の場所からの引用や暗黙値の採用をすることで対応出来るであろう。それに加えて、文の並び（できれば、意味構造化されているとして）を、時間的経過関係、因果関係、並列関係といった関係によりつなぎ合せ、1～2の構造へとまとめ上げる必要がある。例えば、一人の人物をテキスト中の文の並びについて追跡し、そのテキストで表出されている各種情報をその人物に関する情報として抽出し位置付けることを、どのようにして実現できるかという事である。

### 3.3.4 文章からの内容抽出のための課題

文章情報の意味内容を抽出し解析する為には、その文章情報をどの様に把握するかという立場を明確にしておくことが必要である。これは文章情報中の何に着目し、何を取り出したいのかという事と同じである。文章情報自身にはそこで先に述べている1つ又は複数の事実あるいは主張や、同時に文章情報から派生して理解されるいくつかの事実あるいは主張が含まれる。それらのどれも、機械的に取り出すのは必ずしも不可能ではない。しかし、それを一般的に見ている限り、内容分析、内容検索の対象としての情報構造として具体化して下りてくることはない。そこで、文章情報からの内容抽出に関する実用の場合からの分析作業が必要となる。それは机上での、あるいは各種サポート・ツールを用いた分析と、基礎資料の作成で

あると共に、具体的にシステムに組み込んで動かすことによる経験の蓄積でもなければならない。これは、辞書（シソーラス）の作成のための作業とも密接な関連を持っていることも考慮しなければならない。

ここで、次の様な項目が実用の中で具体的に明らかにする必要がある。

- ① 明示的な述語とその支配する出現した格という把握でどこまで文の内容を抽出することが可能であるのかを検定する。
- ② 設定すべき意味カテゴリーはどのようなもので実用上十分であるのかを確かめる。また意味カテゴリーは各々の語に対してどのような構造をもって付与されるかを決定する。
- ③ 「行為」はどのような格をとるのかを整理する。これは②と分離不能なものである。
- ④ 名詞句の構成を分析し、その扱いを明らかにする必要がある複合語の問題や日本語におけるサ変名詞（「する」が付いて動詞化する名詞）などの扱いの問題も同様に解明しなければならない。
- ⑤ 実際の文章について、文脈からの判定方法を検定する。同じく暗黙値の設定は各アプリケーションにおいてどのように想定され、それが満足のゆく結果を得るのかを確かめる。
- ⑥ 文の並びを、所定の内容について追跡することによって、文章情報からの内容抽出を実行し、その経験によりその精度をあげていく。
- ⑦ 文章情報の解析の上で必要な「知識」はどのようなもので、どのようにして文章情報の自動解析に援用するかを決定する。
- ⑧ テキストとその集合としての文章情報データベースの中で、諸事実、傾向などを機械的に理解する為の構造を設定する方法。このためにはデータベースのリレーショナルモデルとして想定することである。例えば、個体や概念に対して、意味カテゴリーや属性を付与するとともに、そこにダイナミックに属性値を代入していくメカニズムと、事象（叙述）を行為とそれが受ける格（辞書における振興された属性に対応する）という

構造で把握し、実現することである。これらの構造はアルゴリズムとの接点においては文法規則などの形で反映される。

### 3.3.5 内容分析と自動翻訳技術との関連

解析すべき一次情報は、必ずしも一カ国語には限らない。その解析を均質的に行うためには、過程のどこかに自動翻訳のメカニズムを組み込む必要がある。また、システム内部から離れても、文章情報データベースを保持するとき、そのなかの検索、分析された情報を自国語に、あるいは他国語に翻訳して流布するという要求は、システムとして答えなければならないものである。

前項までに述べた文章の解析機能の大部分は実は自動翻訳に必要な各種機能と本質において同一のものである。文章情報からの内容抽出のために、文章情報の内容が計算機の情報空間上に展開されうるとすれば、そこに翻訳用ツールを組み合わせることにより自動翻訳を実現することができる。内容分析のためのシステム機能が、文章情報のより深い構造にまではいることは、自動翻訳のメカニズムが深い解析まではいることと軌を一にする。

従って、翻訳システムと文章解析技術の開発は、相互に共通的な技術要素を用いて行うことになる。

### 3.3.6 課題解決のための展望

以上に述べた種々の課題は、現実的には前節までで提示されたシステムの実現の過程のなかで解いていくことになる。それを総合解析システムの開発のなかに位置づけておく。

#### (1) 語の抽出、蓄積とその結果を用いた内容分析

文章情報から語を（辞書を援用して）抽出し、種々の属性を付与したデータベースを作成・更新する。それに対するアクセスを行うことにより、内容分析の一部機能が開発される。このとき、各種分析技法の

開発はもとより、辞書の構成と内容、データベースの構造を確かめることができる。

## (2) 文レベルの内容分析

文レベルの分析により、述語とそれを修飾する格の内容が取りだされ、文レベルの叙述レコードが作成されることによるデータベースが構築される。ここでデータベースに収納された語は、(1)における属性のほかに、文のなかでの役割りと他の語との関係をもったものである。そして、(1)におけるよりも高度な検索、分析機能を開発し実現することが可能となる。このなかで、文構造や文中の名詞句などの問題が分析されるとともに、辞書、データベースもより深い考察がなされ、「第2段階」のデータベース・システムとなる。

## (3) 文章レベルの内容分析

(2)で作成されるデータベース中で表現される各文の内容表現を組み合わせ、文章情報の内容分析を行う。検索、分析機能、データベース、辞書機能などは、情報のより深い構造と内容とに対応した「第3段階」のものとなる。

## (4) 開発運用サポート

辞書の構成や内容抽出のための情報構造を決定するには、テキスト調査・実験サポート・ツールがあらゆる段階で必要である。それはまた、辞書の作成、知識の増殖などのための運用サポート・ツールとも共通性をもつ。合わせて文章情報データベース・システムの重要な一要素となる。

本節で述べてきた課題は、研究室段階でも非常に難度の高いものと言われる問題も少なからず含んでいる。しかし、アプリケーションの視点から、それらを把え直し、実用化を図ったとき、予想外に簡単に解消されることも十分に考えられる。言語処理の困難さは、言語が、それを用いる人間とその世界全体に結び合わされながら常に発せられているという点にある。しかし、

実用化に際しては、その範囲を限定し、把え直す必要がある。そこに実用に向けた言語処理が、純理論的には最も困難な地点から、遂に解かれていくという可能性が大きく開けてくると思われる。現在、ここに欠けているのは、何よりも実用の上での経験と、基礎研究である。本システムがそれ自身の機能と共に、それらを豊富に蓄積していくことが次段階への第一歩である。



## II 各 論





# 1. 文章情報の定量化利用の研究

## 1.1 研究目的と活動概要

### (1) 研究目的

情報処理分野における技術の進歩は長足のものがあり、これまでの大型計算センターのコンピュータ能力がスーツケース程度のパーソナル・コンピュータにおさまリ、磁気記憶装置の記憶能力が名刺大の大きさに集約化されるのは時間の問題である。これまで、数値情報処理に対するコンピュータ利用は、データベースも含めて相当幅広く進められてきているが、文章情報処理に対するコンピュータ利用はまだまだの段階である。しかしながら、最近の企業内のOA化、ワードプロセッサの導入などに伴い、コンピュータによる文章情報処理に対する関心は急速に高まりつつある。

これまでの文章情報データベースは、内容の抄録にキーワードを付加してデータベースに登録し、キーワード検索によって抄録をリスト・アップして検討するという形式のものが中心である。このような文章情報処理システムでは、人間が最終的に文章情報全文までさかのぼり、内容を検討して整理してこななければならない。この意味では、文章情報は非常に豊富な情報を内容としているが、情報密度がきわめて薄められたソースということができよう。文章情報の有効利用を考えた場合、上述のような一般検索による利用もさることながら、文章情報が薄く幅広く豊富に持っている情報を特定のテーマに関して集約化して利用する方向性を導き出していくことが重要である。

海外でも、ニューヨーク・タイムズ・インフォメーション・データベースなどは、10年以上前から、世界有数紙誌の記事情報抄録に関するデータベース化をはかってきており、最近ではニューヨーク・タイムズの記事

全文データバンクの開発と文章情報内の単語頻度数カウントのソフトウェア開発に力を注いでいる。また、米国国防省などが共産圏との関係を考えていく場合の有効な道具の1つとして利用しているシステムも、記事情報をソース・データに置いたものである。上述の一般検索による文章情報の利用という観点からならば、Chemical Abstracts社の各種抄録データベース、Environment Information CenterのEnergyline, Environmentline など数限りがない。このような海外における文章情報データベースの利用はきわめて活発である。

文章情報を集約化して利用していく方法の1つとして、内容分析の手法により文章情報から様々な情報を数値情報に変換して取り出し、これに多変量解析など種々の数量化理論を適用していく方向性が考えられる。このような文章情報の定量化利用の方法論が確立できれば、統計数値情報の分析による結果とは全く異なった視点に立つ分析が可能になると考えられ、例えば、エネルギー問題とか、カントリー・リスク問題の分野などで大いに役立つであろうと考えられる。本研究の目的は、このような文章情報の定量化利用に関する基礎的な調査を行うことと、文章情報の定量化利用機能をも含む文章情報総合利用システムのシステム像に関する基礎的な検討を行うことである。過去に、文章情報の定量化利用を念頭においたきわめてプリミティブなパイロット・システムの開発を行っており、本研究ではこの成果も踏まえて検討を行っている。

## (2) 活動の概要

(1)で述べた研究の主旨に従って、学識経験者も含むワーキング・グループを形成し、文章情報総合解析システムを構築していくのに必要な種々の項目について検討した。ワーキング・グループによる研究会は、各回ごとに講師をお願いし、特定のテーマで講師にお話をいただいた後、ワーキング・グループのメンバーによって討論を加えるという形式で行った。研究会での検討内容は以下の通りである。

- 第1回 ① メンバー紹介  
② これまでの経緯と計画説明  
③ キーワードあるいは単語の自動抽出について
- 第2回 入力データ作成にあたっての問題点
- 第3回 ① Chemical Abstracts 被抄録文献数の国際比較  
② 2部グラフ理論を用いた新聞記事キーワード間の概念構造分析  
③ キーワードの出現頻度数分析
- 第4回 MITI-IRの経験について
- 第5回 新しいデータベースの構築について——記事情報総合利用システム——
- 第6回 講演「ユーザから見て使い易い記事情報システムについて」

東京大学東洋文化研究所

助教授 猪 口 孝 氏

以下の各節で、ワーキング・グループによる研究会での検討内容を詳しく述べる。

## 1.2 既存システムによる文章情報定量化利用事例

本節では、第3回ワーキング・グループの資料に基づいて、現在すでにある文章情報検索システムを利用した文章情報定量化の事例を紹介する。併せて、既存の文章情報検索システムによる定量化利用の問題点を考えてみる。

### (1) Chemical Abstracts 被抄録文献数の国際比較

東京大学大型計算センターでは、オンライン情報検索システム TOOR-IR を開発し、全国の学術研究者を対象として日本最初のオンライン文献情報サービスを1975年7月に開始した。現在、TOOR-IRはCA Search, Crystallographic Data Base, INSPEC-C, COMPENDEX, CA Ecology and Environment等をデータベースとしている。1980年9月には、東京大学大型計算センターの機種更新に伴って、これまでのTOOR-IRの利用者インタフェースを保存しつつ、別の情報検索システム ORION (日立製作所)の上にTOOR-IRを構築するという新しい試みも行われた。ORIONとの結合により、データベース内容についての統計情報が得やすくなったことが、大きな特徴である。

CA (Chemical Abstracts) は、化学及び化学工学についての研究・技術情報を抄録したもので、ハードコピー及び磁気テープ・データベースとして販売されている。ソース源は学術誌、単行書、会議録、特許から各国の抄録誌やDissertation Abstractsにも及んでいる。抄録件数も最近では年間45万件程度となり、対象分野も化学、化学工学はもとより応用数学、物理学、生物学、医学、農学、金属工学、材料力学等と多岐にわたっている。従って、化学を中心とする関連諸分野の研究活動の指標の一つとして、国別のCA被抄録件数を取り、国際比較を行うことは妥当と考えられる。

本事例では、TOOR-IR上のCA Search データベース (CA91 巻1-9 2巻, 1979年下期-1980年上期) から、ORION の EXPANDサブ

コマンドを用いて、分野別・国別の被抄録件数を求め、国際比較を行っている。単純に被抄録件数の多い国をリストアップし、構成比を求めるだけでなく、国の大きさ（人口を指標とする）に対する比や国の経済力（GNP又はGDPを指標とする）に対する比に関する国際比較も行っている。

表1-1は、主要各国の化学における研究活動をCA被抄録文献数からみたものである。左欄は被抄録件数の多い順に、中欄は人口1万人当たりの被抄録件数の多い順に、右欄はGNP又はGDP100万ドル当たりの被抄録件数の多い順に上位15カ国をリストアップしてある。被抄録件数を見ると米国が圧倒的に多く、これをソ連、日本が追い、さらに西独、英国、フランスといった国々が続いている。これは比較的順当な結果であるが、第7位に発展途上国のインドが入っているのは大変興味深い。人口1万人当たりの被抄録件数あるいはGNP又はGDP100万ドル当たりの被抄録件数に関して、イスラエル、スイスといった国々が上位に登場してく

表1-1 CA被抄録文献数からみた主要各国の化学における研究活動

| 順位 | 被抄録件数   |      | 人口1万人当たり<br>被抄録件数 |     | GNP又はGDP 100万\$<br>当たり被抄録件数 |      |
|----|---------|------|-------------------|-----|-----------------------------|------|
| 1  | 米 国     | 11.6 | スイス               | 8.0 | イスラエル                       | 19.5 |
| 2  | ソ 連     | 7.6  | イスラエル             | 7.2 | ブルガリア                       | 12.3 |
| 3  | 日 本     | 6.2  | スウェーデン            | 5.6 | ハンガリー                       | 11.8 |
| 4  | 西 独     | 2.7  | 日 本               | 5.4 | ポーランド                       | 11.0 |
| 5  | 英 国     | 2.6  | 米 国               | 5.3 | 英 国                         | 10.4 |
| 6  | フランス    | 1.8  | カナダ               | 4.8 | 日 本                         | 9.8  |
| 7  | インド     | 1.4  | 英 国               | 4.7 | チェコ                         | 9.8  |
| 8  | カナダ     | 1.1  | 西 独               | 4.4 | インド                         | 9.8  |
| 9  | ポーランド   | 1.0  | デンマーク             | 4.2 | スイス                         | 9.1  |
| 10 | イタリア    | 1.0  | オランダ              | 4.0 | ソ 連                         | 8.5  |
| 11 | チェコ     | 0.6  | チェコ               | 4.0 | 東 独                         | 8.3  |
| 12 | 東 独     | 0.6  | フィンランド            | 3.9 | ニュージーランド                    | 7.6  |
| 13 | オランダ    | 0.6  | 東 独               | 3.4 | 米 国                         | 6.4  |
| 14 | スイス     | 0.5  | ベルギー              | 3.4 | スウェーデン                      | 6.4  |
| 15 | オーストラリア | 0.5  | ニュージーランド          | 3.4 | フィンランド                      | 6.2  |

ることは、これらの国々で化学における研究活動が活発であることを示す結果と考えられる。GNP又はGDP 100万ドル当たりの被抄録件数についてブルガリア、ハンガリーといった共産圏の国々が上位に現われるのは、これらの国々のGNPが相対的に他よりも低すぎるためであろう。

表1-2は、計算機、汚染、農業の3分野で各国の被抄録件数と構成比を上位10カ国について示したものである。図1-1は、この構成比の対数を米国、日本、ソ連、西独、カナダ、英国、イタリア、フランス、インド、ポーランド、チェコの11カ国について計算機、汚染、農業の3分野でプロットしたものである。西欧先進国の米国、カナダ、イタリアは、汚染の分野における被抄録件数の構成比が3分野の中で最も高いが、共産圏のソ連とチェコは構成比が最も低い。計算機の分野の構成比が3分野の中で最も高い値を持っているのは、米国、西独、英国、フランスといった国々である。農業の分野の構成比が最も高い値を持っている国は、ソ連、日本、インド、ポーランドである。

表1-2 3つの分野における各国の被抄録件数と構成比

| 順位 | 計 算 機 |      |       | 汚 染   |       |       | 農 業   |       |       |
|----|-------|------|-------|-------|-------|-------|-------|-------|-------|
| 1  | 米     | 1474 | 40.5  | 米     | 2860  | 40.5  | 米     | 2500  | 21.3  |
| 2  | ソ     | 362  | 13.7  | 日     | 714   | 10.1  | ソ     | 2277  | 19.4  |
| 3  | 西独    | 288  | 7.9   | ソ     | 595   | 8.4   | 日     | 1226  | 10.5  |
| 4  | 日     | 286  | 7.9   | 西独    | 474   | 6.7   | 印     | 745   | 6.4   |
| 5  | 英     | 214  | 5.9   | 加     | 322   | 4.6   | 英     | 541   | 4.6   |
| 6  | 仏     | 123  | 3.4   | 英     | 273   | 3.9   | 西独    | 501   | 4.3   |
| 7  | 加     | 81   | 2.2   | 伊     | 238   | 3.4   | ポーランド | 340   | 2.9   |
| 8  | 伊     | 78   | 2.1   | 仏     | 193   | 2.7   | 加     | 316   | 2.7   |
| 9  | 印     | 72   | 2.0   | ポーランド | 172   | 2.4   | 仏     | 285   | 2.4   |
| 10 | チェコ   | 59   | 1.6   | 印     | 157   | 2.2   | 伊     | 198   | 1.7   |
|    | 件 %   |      |       | 件 %   |       |       | 件 %   |       |       |
|    | Total | 3643 | 100.0 |       | 7,067 | 100.0 |       | 11731 | 100.0 |

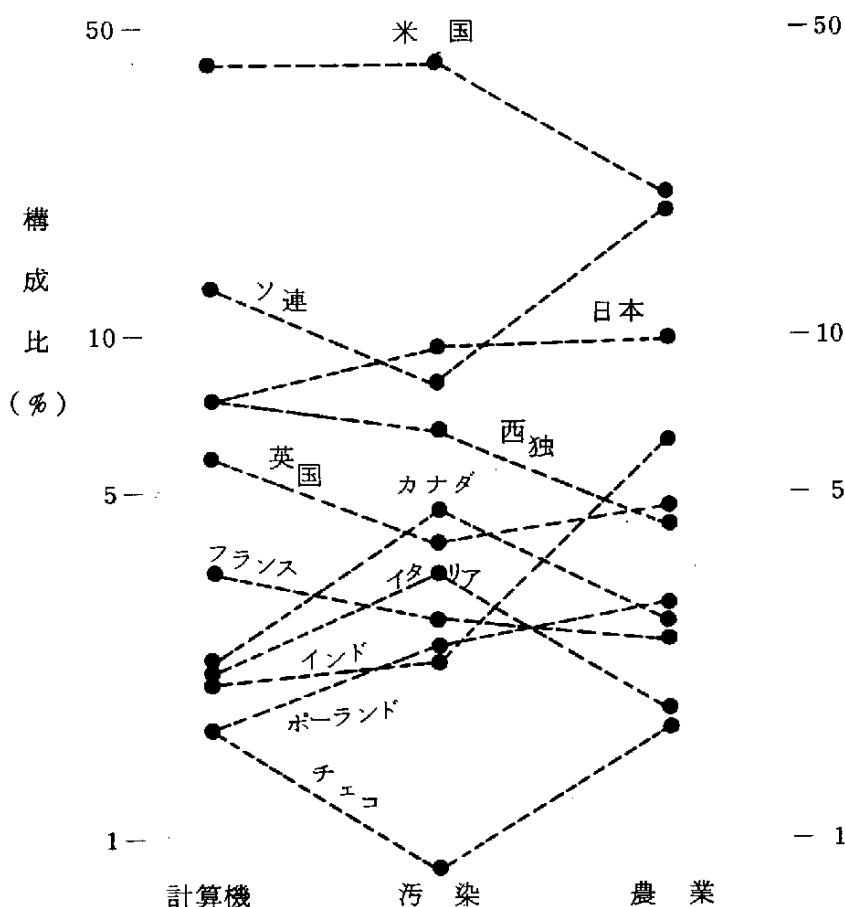


図 1 - 1 計算機・汚染・農業各分野の各国別被抄録文献数構成比

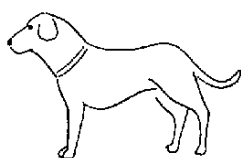
ここに示した結果は、全体の極く一部である。結果の持つ深い意味を解釈するにはさらに多くの情報に基づいて解析を加えることが必要であろう。このようにデータベースの登録件数を種々に分類して分析することは、キーワード検索によりリストアップした情報の内容を単純に見る方法以外のデータベース利用法を1つ提供するものである。時系列に各国の研究活動の変化を観測したり、対象分野を3分野からさらに広げることにより、一層興味深い結果が得られるものと期待される。

## (2) 2部グラフ理論を用いた新聞記事キーワード間の概念構造分析

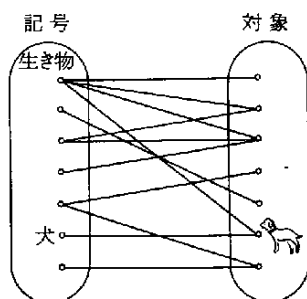
通商産業省情報管理課では、日本経済新聞社から記事情報インデックスの磁気テープを購入し、独自に開発した検索システムを通じて省内での利用を進めている。記事情報インデックスは、日本経済新聞社が発行している新聞と雑誌について、記事内容を40字以内のインデックスにまとめたものである。各インデックスには、掲載紙誌名、掲載年月日、掲載紙面あるいは掲載ページの他、日本経済新聞社で体系付けたキーワード群からインデックスの内容に適切ないくつかのキーワードが付けられている。また、重要記事については、40字のインデックスの他に、200字の要旨も付加されている。本事例では、この記事情報インデックスに対して、2部グラフの理論を適用したキーワード間の概念構造分析を紹介する。

2部グラフの理論を用いると、図1-2に示すように、ある対象をある記号で名指すという形の言語行動データの構造を分析することが可能である。図1-2の1)に示したものを、人は「犬」、「ペット」、「生き物」、「ワンちゃん」、「漫画」など種々の言葉を使って名指す。ここでは1)のようなものを対象、「犬」、「生き物」などの言葉を記号と定義する。このような記号と対象の集合を考えた場合、2)に示すように、記号と対象の間の対応は一般に多対多となり、様々な結び付きを生ずる。3)に示すように、記号と対象の集合の間に新しい概念の集合を入れ、対応関係を整理して構造化できないであろうか。4)に示す記号と対象を実例として、対応関係の構造化を考えてみよう。4)に示した2部グラフのデータを、記号を列とし対象を行としたマトリックスに置き換えたものが5)である。×印は記号と対象の間に対応関係のあることを示しており、このマトリックスでは、×印が対角線の右側に集まるように、行と列の並べ換えを行っている。このマトリックスを作成することが、実は3)に示した新しい概念の集合を求める操作となっている。マトリックスから記号が持っている概念構造を整理すると6)となる。「人工物」はすべての対象と結び付

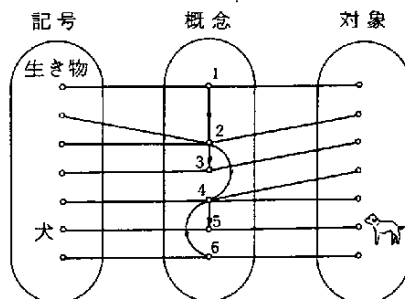




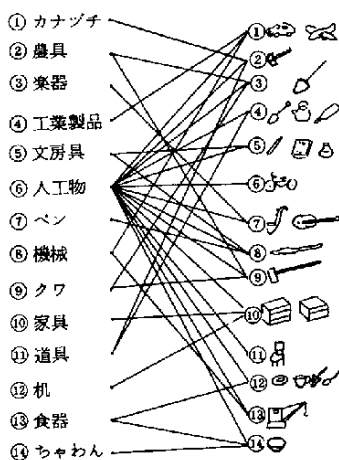
1) 対象の例



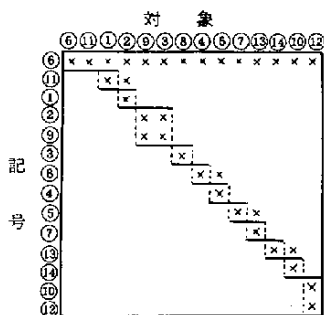
2) 記号と対象の2部グラフ



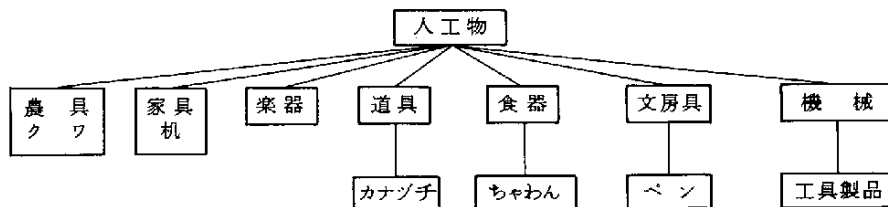
3) 記号と対象の間に介在する概念



4) 記号と対象の2部グラフ  
データ例



5) 概念構造を示すマトリックス



6) 4) のデータから得られる概念構造

図1-2 2部グラフによる言語行動データの概念構造分析

いているので最上位である。「道具」は④，②の対象と結び付き，「カナヅチ」は②の対象とだけ結び付いているので，「道具」は「人工物」の下に位置し，「カナヅチ」は道具の下に位置する。「農具」と「クワ」は共に③，⑨の対象と結び付いているので，「人工物」の下に同等の関係で位置する。6)の概念構造は，このようなマトリックスに対する判別を通して作成されたものである。

さて，記事情報インデックスを対象とし，インデックスに付されているキーワードを記号として考えると，図1-3に示すように，キーワードと記事インデックスの間に2部グラフを考えることが可能である。通商産業省情報管理課では，図1-2の実例で示したような2部グラフの解析手法に基づいて，キーワード間の概念構造分析を行うシステムを開発した。図1-4にシステム・フローを示す。

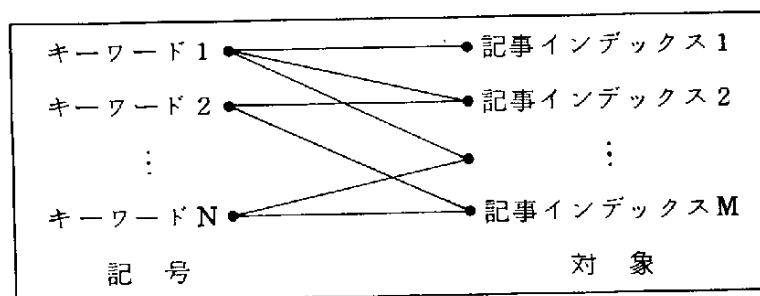


図1-3 記事情報インデックスとキーワードの  
2部グラフ

この解析システムを用いて，1978年7月と10月の記事情報インデックスの中から，イランの国名キーワードが付いた記事インデックスを抽出し，それぞれの記事インデックスの集合に対してキーワードの概念構造分析を行ったところ，大変興味深い結果が得られた。図1-5に1978年の10月の解析結果を示す。イランに関する種々の事件を示すキーワードがうまくグルーピングされている様子がよくわかる。イランで革命が起

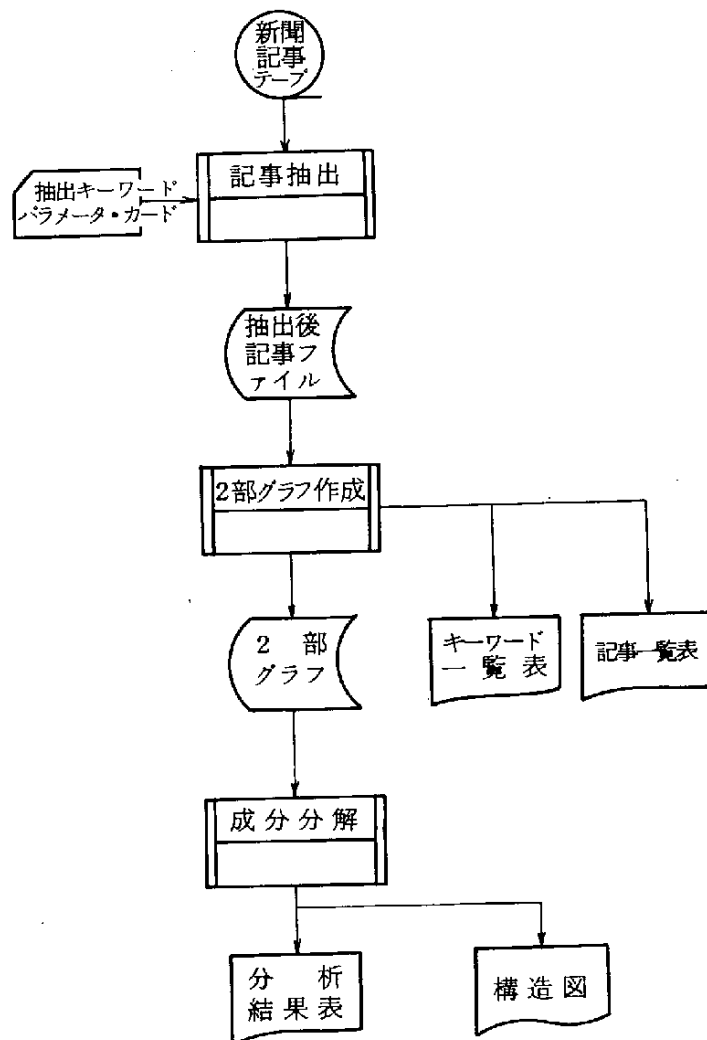


図1-4 キーワードの概念構造解析システムのフロー

ったのは1978年の12月であるが、2カ月前の10月には日本経済新聞でも、反政府運動、暴動、治安、労働争議といった内容がずいぶん話題に上っていたということである。5カ月前の7月に対しても同様のキーワード概念構造分析を行ったが、このような話題を示すキーワードは全く登

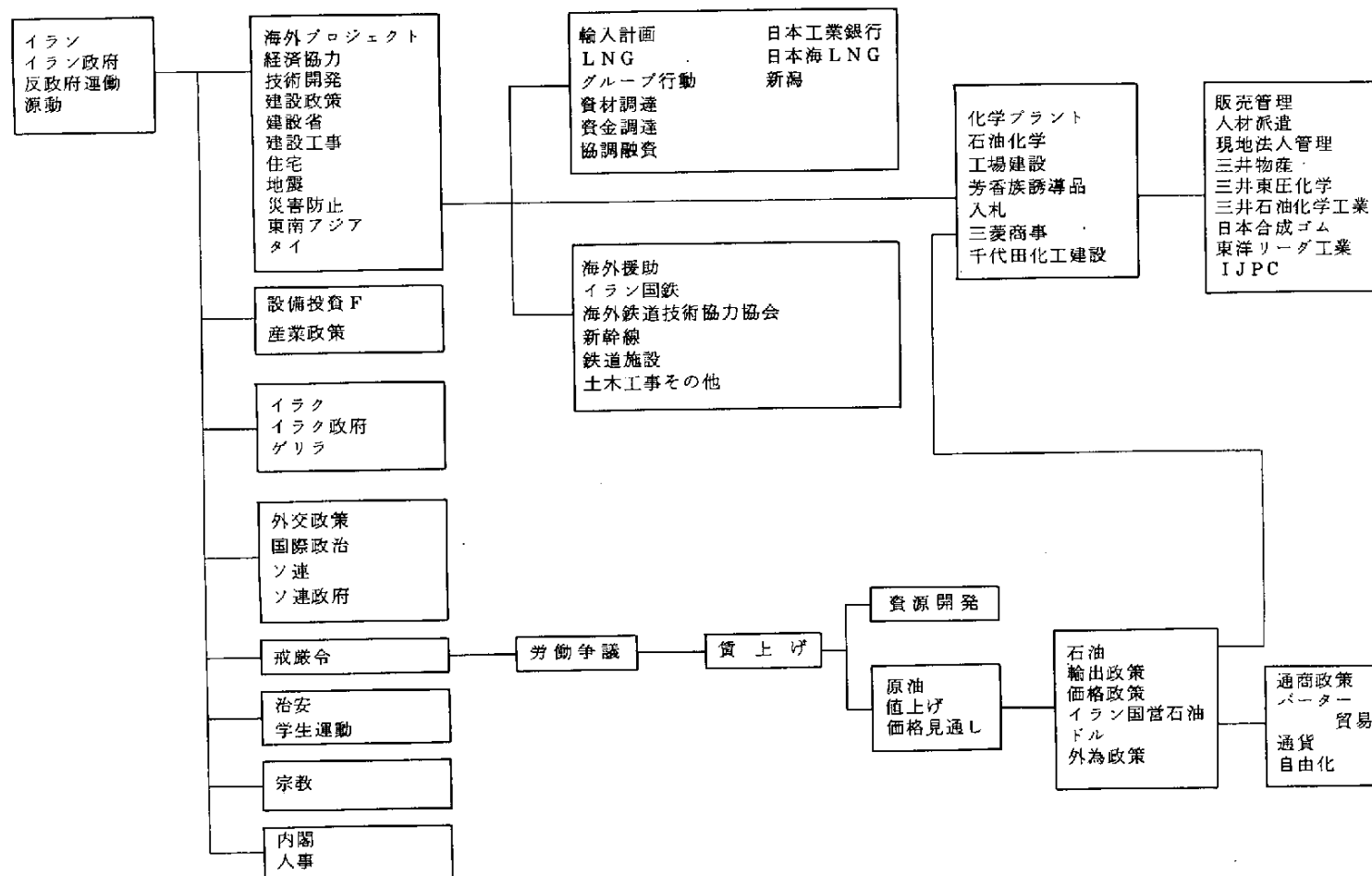


図1-5 イランに関する記事インデックスのキーワード概念構造分析の  
解析結果(1978年10月)

場していない。

2部グラフによるキーワードの概念構造分析という手法は、このように解釈しがいのある重要なデータを我々に提供してくれる。記事内容をキーワードに集約させ、この手法を用いてキーワードをグルーピングすることにより、各国のイメージとイメージの時系列変化を観測することも可能であろう。しかしながら、この手法の大きな欠点は常にキーワードのグルーピングが図1-5に示したようにきれいな形にまとまるとは限らないことである。なんらかの束縛条件を見出して、常に有効な結果を提供する分析パターンを確立することが今後の大きな課題であろう。

### (3) 世界エネルギー情報によるキーワード出現頻度分析

(財)日本エネルギー経済研究所では、日常業務として、70余の海外紙誌からエネルギーに関連した記事を切り抜き、種々の海外動向分析を加えている。切り抜き記事を過去に遡及して使用できるようにするため、1979年1月から日本経済新聞社の記事情報インデックスの書式に従って世界エネルギー情報データベースを作成してきた。世界エネルギー情報データベースの内容は、40字以内の記事インデックス、掲載年月日、掲載紙誌略号、掲載紙面あるいは掲載ページ、記事行数、200字以内の記事要旨(部分的)、各種キーワードである。最近、きわめて小型の漢字出力端末装置が電々公社から発売され、世界エネルギー情報データベースも電々公社のDEMOSEセンターを公衆電話回線網を通じてオンライン利用することが可能となった。図1-6に世界エネルギー情報データベースのオンライン検索例を示す。西独の原子力開発について検索した結果の一部である。

本事例では、1979年3月末米国で起ったスリーマイル島原発事故以後の原子力の動きを、米国、西独、フランス、英国の4カ国について原子力開発と原子力問題の2つのキーワードの頻度数の変化を通して眺めてみる。電電公社の情報検索システムDRISによるオンライン検索は、図1-6に示したようなリストで情報を提供できるが、データベースの統計情報を

|                                      |           |    |     |
|--------------------------------------|-----------|----|-----|
| ・西独の原発建設、再び暗礁にーハンフルク社長宛、反対投票。        | 810204 WJ | 27 | 73+ |
| ・西独ハンフルク社長宛、ブロクトルフ原発建設に反対の投票。        | 810204 HB | 4  | 80+ |
| ・西独ブロクトルフ原発建設再開へーハンフルク市長宛3月2日反対決議。   | 810205 FT | 2  | 81  |
| ・西独の北部2州、政府の原子力政策の混乱を反映し原発建設で意見対立。   | 810213 FT | 2  | 62  |
| ・ハンフルク社長宛、ブロクトルフ原発建設を阻止ーシェミット政権に大打撃。 | 810223 BW | 43 | 63  |
| ・フランス原子力公社、西独で原発建設資金として最高2億ドル調達へ。    | 810225 FT | 16 | 84  |
| ・論説ー西独政府、説得力をもって原子力エネルギー拡大さるべき。      | 810204 HB | 2  | 85  |
| ・西独のエネルギー開発にカルカー計画は必須ー高速増殖原発社長長談。    | 810227 HB | 12 | 36  |
| ・もしも原子力技術会議で西独の原発建設遅延について報告。         | 810326 HB | 16 | 37  |

著作権者 日本エネルギー経済研究所

図1-6 世界エネルギー情報データベースの  
オンライン検索例

提供することには必ずしも長けていない。従って、このようなリストを出した後の解析はすべて手作業で行った。

図1-7は、各国別に原子力問題と原子力開発というキーワードの月別頻度数変化を示したものである。1979年3月-4月は、スリーマイル島原発事故の直後で、事故のあった米国はもちろん西独、フランスの両国でも原子力問題、原子力開発のキーワードの出現頻度数が大幅に増加している。スリーマイル島原発事故を契機として、新聞・雑誌の上に国内の原子力開発、原子力問題が大いに取り上げられたということである。ところが、大変興味深いのは英国の頻度数変化で、1979年3月-4月の時期には、他の3国と同じようなピークは全く現われていない。このことは、Financial Timesなどで米国のスリーマイル島原発事故は取り上げられたが、事故直後に英国内の原子力発電所の問題はあまり論じられなかったことを意味している。英国で原子力問題が大きく取り上げられているのは、むしろ1980年1月-4月の時期で、この時期には核燃料廃棄物プラント

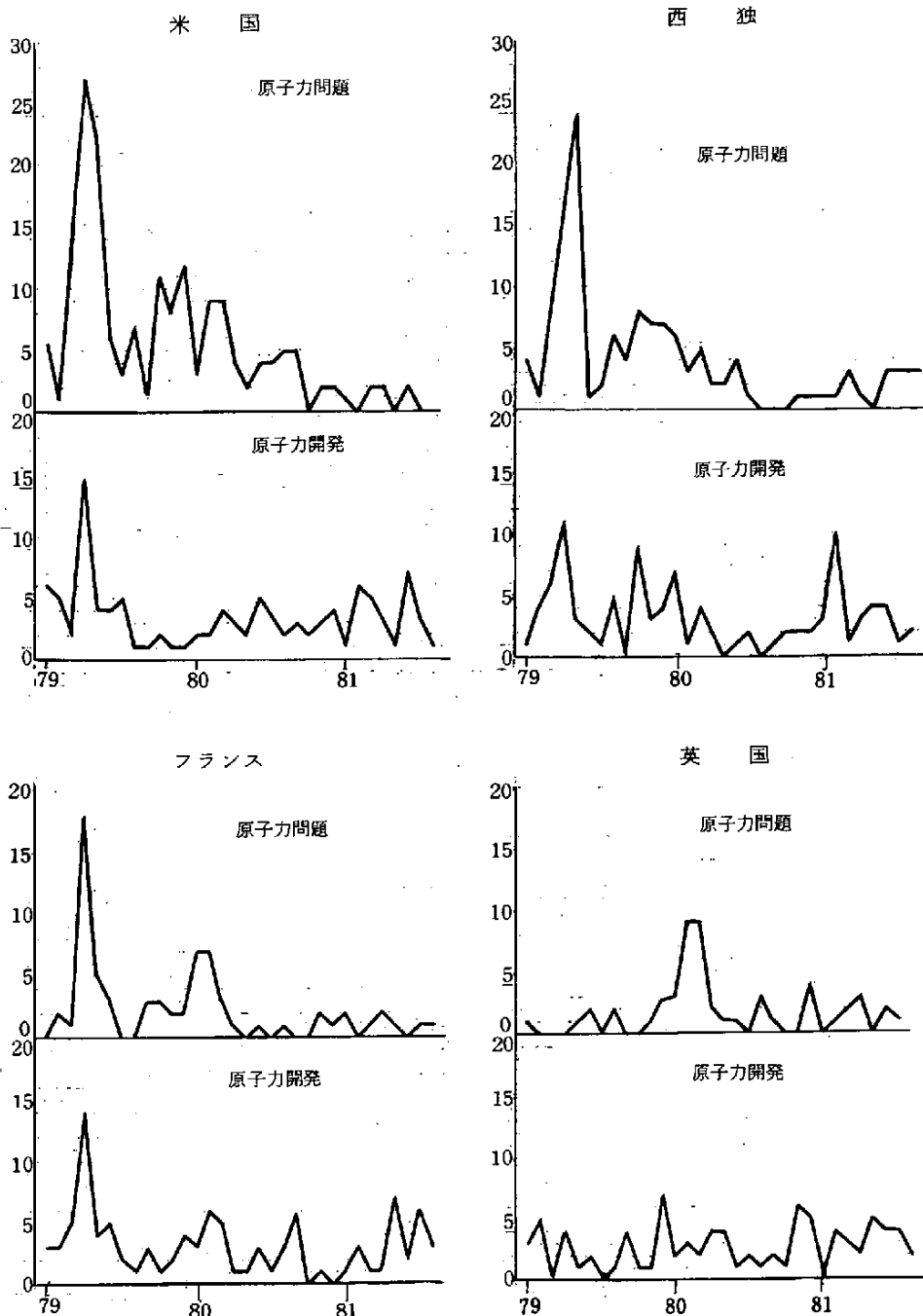


図 1 - 7 キーワード ( 原子力問題, 原子力開発 ) の  
月別頻度数変化

の問題が紙誌上をにぎわせている。

全体としては、各国とも原子力問題というキーワードの出現頻度数が事故後しだいに減少し、原子力開発のキーワードの出現頻度数がしだいに増加しているように見受けられる。ここでは図示しないが、原子力開発の出現頻度数を横軸とし、原子力問題の出現頻度数を縦軸として年別に各月の出現頻度数の2次元プロットを作成するとキーワードの増減の傾向は一層はっきりとらえることができる。

このように頻度数の単純なプロットを取るだけでも、いろいろな情報を整理することができるが、日本経済新聞社によって整備されたキーワード体系には、文章情報を定量化して取り扱うことを考えた場合、大きな欠陥があることを指摘しておかねばならない。本事例のキーワードを例に取るならば、原子力開発というキーワードを持つ記事には、原子力開発反対も原子力開発賛成も双方とも含まれてしまうという問題である。文章情報の定量化を行うには、キーワードで表現される行為に対して少なくともネガティブな方向とポジティブな方向との2種類の分類を持つことが必要であろう。

図1-7に示した月別頻度数の変化は、キーワードの出現頻度数を用いたのでは母集団が小さすぎるため、有意な変化を十分に反映しきれない可能性を示唆している。表1-3は、月別頻度数を四半期、半期、年に集計したものであるが、キーワードの出現頻度数も半期ぐらいまで集計されると有意な変化がはっきり見えてくるようである。この点では、記事内に出現する単語頻度を情報源とするほうが、母集団ははるかに大きくなると考えられる。

世界エネルギー情報データベースだけでなく、1.2(2)の分析に用いられた日本経済新聞社の記事インデックス・テープに関しても共通に言えることであるが、これらのインデックスのキーワード付けは人為的に行われている。従って、キーワード付け作業者の主観による種々のバイアス



がかかっていることは否めない。文章情報の定量化利用を行うためには、できるだけ客観的立場からのデータベース作成作業を行えるシステム体制を構築する必要がある。

表 1 - 3 キーワード出現頻度数の集計

|         | 米 国   |       | 西 独   |       | フランス  |       | 英 国   |       |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|
|         | 原子力問題 | 原子力開発 | 原子力問題 | 原子力開発 | 原子力問題 | 原子力開発 | 原子力問題 | 原子力開発 |
| 1979 1Q | 19    | 13    | 14    | 11    | 3     | 11    | 1     | 8     |
| 2Q      | 55    | 23    | 41    | 16    | 26    | 23    | 3     | 7     |
| 3Q      | 11    | 7     | 12    | 6     | 3     | 6     | 2     | 5     |
| 4Q      | 31    | 4     | 22    | 16    | 7     | 7     | 4     | 9     |
| 1980 1Q | 21    | 8     | 14    | 12    | 18    | 14    | 21    | 7     |
| 2Q      | 10    | 10    | 8     | 3     | 2     | 5     | 4     | 9     |
| 3Q      | 14    | 6     | 1     | 3     | 1     | 10    | 4     | 5     |
| 4Q      | 4     | 9     | 2     | 6     | 3     | 1     | 4     | 12    |
| 1981 1Q | 3     | 12    | 5     | 14    | 3     | 5     | 3     | 7     |
| 2Q      | 4     | 11    | 4     | 11    | 3     | 10    | 5     | 11    |
| 1979 上期 | 74    | 36    | 55    | 27    | 29    | 34    | 4     | 15    |
| 下期      | 42    | 11    | 34    | 22    | 10    | 13    | 6     | 14    |
| 1980 上期 | 31    | 18    | 22    | 15    | 20    | 19    | 25    | 16    |
| 下期      | 18    | 15    | 3     | 9     | 4     | 11    | 8     | 17    |
| 1981 上期 | 7     | 23    | 9     | 25    | 6     | 15    | 8     | 18    |
| 1979    | 116   | 47    | 89    | 49    | 39    | 47    | 10    | 29    |
| 1980    | 49    | 33    | 25    | 24    | 24    | 30    | 33    | 33    |

### 1.3 入力データ作成にあたっての問題点

文章情報データベースに限らず、一般のデータベース・システムを考える場合、入力データの処理の問題が占めるウェイトはきわめて大きい。特に、システム完成後のメンテナンス体制下では80%以上をこの問題が占めると言っても過言ではなからう。本節では、第2回ワーキング・グループの資料に基づいて、入力処理の種々の問題点と日本語漢字情報の入力について考えてみる。

#### (1) データベース・システムにおける入力処理系の位置付けと各種作業

入力処理系はデータベース構築の入口であり、人間による作業の比率の最も高い部門といえることができる。文章情報データベースの一般像(図1-8)を考えてみると、このようなデータベースは、情報の網羅性、最新性、正確性、利用者の要求に合った情報の収集性・加工性を要求される。また利用面から見た場合、IRにおいては情報の集合性と同時に分散性が要求され、CTS出力・KLP出力では内容の均一性と統一性が要求される。データベースにこのような諸条件を保証するためには、情報の収集、保管、管理等に複雑な作業を必要とするわけであるが、人力の最も関与する入力処理系に作業が集中することは否めない。

入力処理系の作業は、①データベースが単一目的利用か多目的利用か、②入力マスター・テープ作成のための初期入力データがフリー入力かコントロール入力か、③ファイル構造・データ要素の決定・各種表記規則・コード体系の決定、④担当者間での作業統一、⑤検索ソフトウェアの種類といった諸点に依存しているといえることができる。現在あるデータベースの大半は、一定のルールでデータ・コーディングを行い、初期入力データを作成するコントロール入力の形式を取っている。入力処理系の作業を分類すると図1-9となる。

入力処理の過程で問題が生じた場合に常に原票に戻れるようにするため、

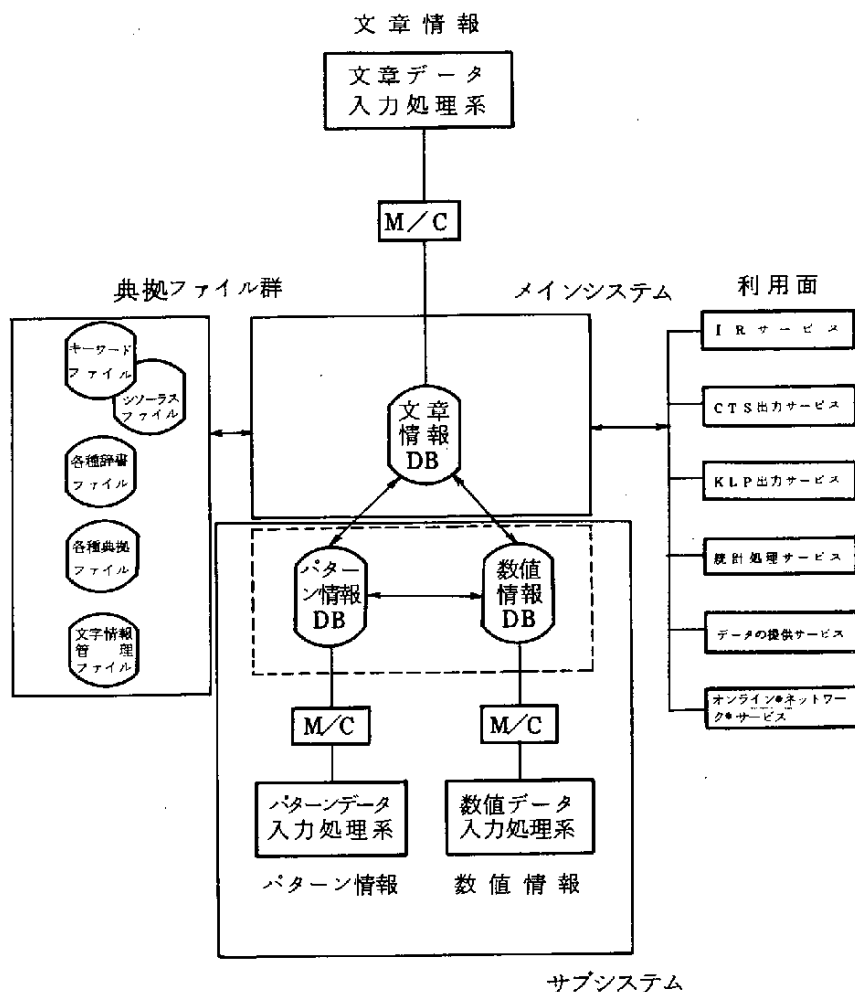


図 1 - 8 一般的な文章情報データベース

原票管理台帳の作成，原票棚の整理といった入力資料の管理保管作業が大切である。この作業における問題点は，大量データの場合に，他の原票との混在が避けられないこととスペースの確保が難しくなることである。入力原稿の作成作業あるいは入力前作業では，特にコントロール入力の場合，送り仮名規定，翻字法，使用文字種，使用単位といった諸点に関して厳密な規定を設けたコーディング・マニュアルを作成することが必要であ

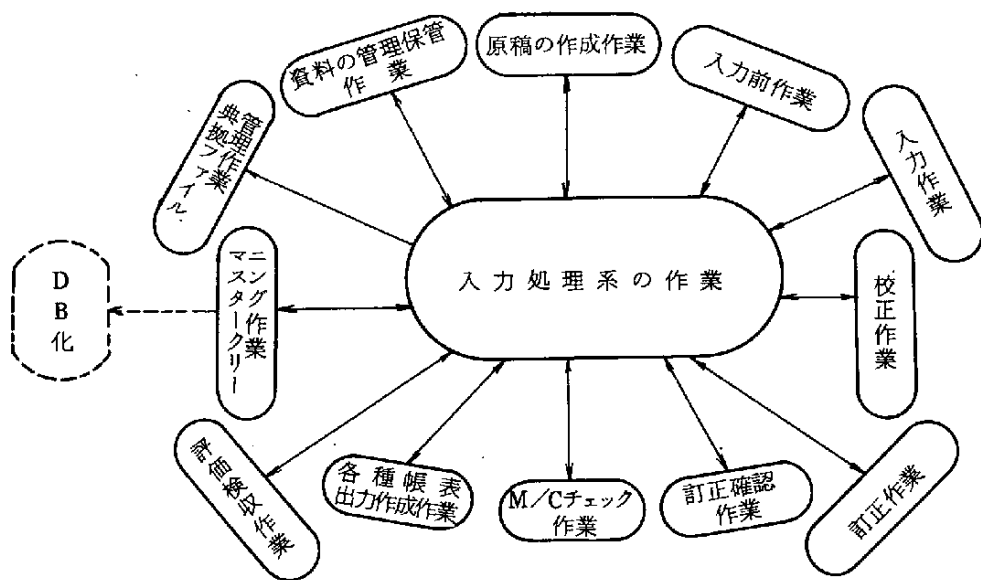


図 1 - 9 入力処理系の各種作業

る。これらの作業中には、マニュアルで処理できないケースが絶えず生ずるので、逐次マニュアルを更新する体制を取っておかねばならない。入力作業においても、各種入力規定を取り決めた入力マニュアルが必要である。入力タッチ数の記入、頻度調査、入力ミスの原因追求などを行えるようにし、作業の質を保てるようにしなければならない。この作業における問題点は、大量入力だと各種漢字コードの混在に伴って作字・外字処理が困難になることと、入力機器のハードによるトラブルを発見するのが困難なことである。

作成した入力データに対しては、M/Cチェック作業が行われる。アルゴリズムを立てることができる論理的な整合性のチェックは、チェック・プログラムを作成し、各種典拠ファイルに基づいて行う。チェック・プログラムが複雑になると完全にプログラム・ミスを除けない点が問題である。典拠ファイルは、メンテナンス・マニュアルを作成し、最新の状態

に保つための管理が必要である。典拠ファイルに対する大量処理や更新の時系列管理はきわめて難しい。M/Cチェック作業の結果として、校正作業者に負担のかからない各種帳票作成作業を行い、校正しやすいリストを出力する必要がある。

校正作業では一般には文字さえ合っていれば良いという段階で作業を終えてしまう傾向にあるが、校正マニュアルに基づいて各種辞書との整合性チェック、関連項目間での整合性チェックを行い、統一した表示でエラーや校正結果を示すことが必要である。新造語の判断、固有名詞の読み、分かち書きの不統一といった諸点は、校正作業において十分処理しきれない問題である。校正の結果に基づいて、訂正作業と訂正確認作業が行われる。バッチ訂正の場合はM/C処理後でないと訂正確認ができない点が、ディスプレイ訂正の場合は長時間の作業に耐えられない点がそれぞれの欠点である。校正、訂正、訂正確認の3作業とも、全入力処理業務に精通したベテランが担当することが望ましい。

上述の作業を総合してマスター・クリーニング作業が行われ、データベース化直前の入力マスター・テープが作成される。入力処理系全体の質を保つため、エラー文字数、エラー件数などの統計を取り、評価検収作業を行うことが必要である。以上が入力処理系の各種作業とその問題点であるが、新しい文章情報データベースを構築するにあたっては、これら諸点を十分考慮した入力処理系のシステムをデザインしていくことが重要である。

## (2) 日本語漢字データの入力について

漢字データ入力の難しさは、①字種が多い（漢字48,902字、ひらがな87字、かたかな86字、アルファベット52字、数字10字、特殊記号70～500字）、②字の形が複雑である、③類似した字が多い、④異体字（体→體）、絵文字が存在する、⑤適用分野で必要字数に差異がある、といった諸点に帰因している。現在の漢字入力方式をまとめると図1-10

となる。パターン認識による漢字入力には、まだ実用の段階までいたっていない。生産性、精度、教育、外字入力、特殊入力、入力可能文字数、ハード的互換性などといった観点からみると、各方式とも種々の長短を有している。

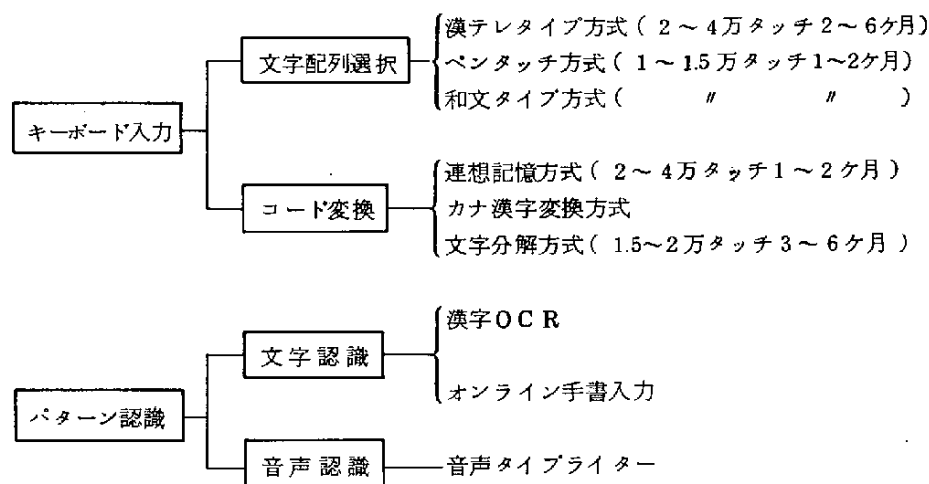


図 1 - 1 0 現在の漢字入力方式

漢字データ入力を左右する要因としては、原票の状態、使用文字量、盤外文字の使用、タッチ度と目の疲労度、記憶媒体の種類、訂正機能、ハードミス表示機能など様々である。漢テレタイプ方式で今迄の入力処理の問題点を考えてみよう。漢テレタイプ方式の入力処理系の場合、入力後のM/Cチェック、校正、訂正作業はHOST M/Cを使っての処理が中心となっており、訂正の確認はM/C処理後に行われる。従って、入力処理系のシステム設計を行う際、初期入力しやすい方法、効率の良いマスター・テープ設計、校正しやすい帳表の作成、訂正しやすい方法、訂正の確認、人間の作業が繁雑にならない管理可能な全体運用、といった諸点を十分考慮しなければならない。こうした事情は、漢テレタイプ方式の入力処理にいくつかの問題点を生じさせている。第1は、図1-11に示すように

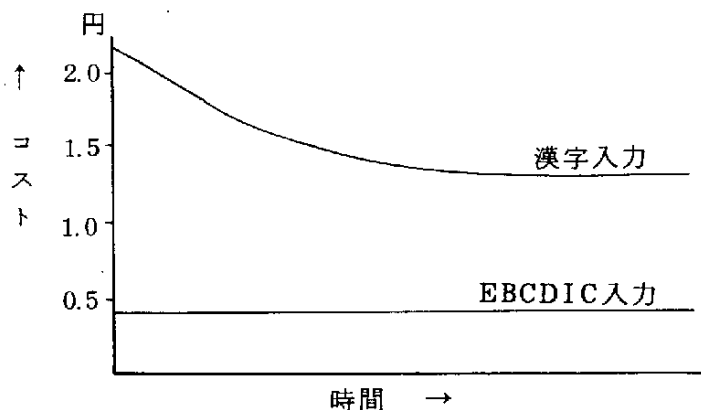


図 1-11 漢字入力のコスト

漢字入力がEBCDIC入力に比べて2～3倍コスト高になっていることである。第2は、原票作成から入力マスター・テープ完成までの入力処理系の作業に多大な時間がかかることである。第3は、入力精度をある一定の所までは改善できるが、それ以上は一定の割合でエラーが残ることである。第4は、記憶媒体が比較的不安定だということである。

漢字入力機器の歴史をふりかえってみると、漢テレタイプ方式に始まり、ペンタッチ方式、和文タイプ方式が導入され、漢字キーto ディスク、漢字キーtoフロッピーから現在は日本語ワードプロセッサの登場にいたっている。表1-4に示すように、日本語ワードプロセッサは各社から様々な種類のものが発売され、現在全盛時代を迎えようとしている。日本語ワードプロセッサの特徴としては、2～3日の教育で可能になる使い易さ、豊富な校正編集機能、文章の蓄積機能、文章の印刷機能、文字種の豊富さなどが掲げられる。これらの点から考えて、日本語ワードプロセッサは文書処理の合理化のため、今後ますます各方面に導入されていくと考えられる。ただし、日本語ワードプロセッサの特徴から考えて、一旦蓄積した情報をディスプレイ編集することには大きな威力を発揮するが、日本語情報の大量入力には必ずしも十分な能力を持っていないと考え

表 1-4 日 本 語 ワードプロセッサ

| 方 式                       | メーカ             | モ デ ル                          | 価 格                     |           |
|---------------------------|-----------------|--------------------------------|-------------------------|-----------|
| かな漢字変換<br>(かなキーボ<br>ード使用) | 自 動<br>(単語辞書有り) | 東 芝 JW-10モデル2                  | 340万                    |           |
|                           |                 | JW-5                           | 260万                    |           |
|                           |                 | キ ャ ノ ン キヤノワード55(ローマ字可)        | 260万                    |           |
|                           |                 | 三 菱 M8510                      | 293万                    |           |
|                           | 文法処理なし          | 富 士 通 オアシス100(独自キー)            | 270と320万                |           |
|                           | 逐 次<br>(単語辞書なし) | 表 示 選 択 日 立<br>(沖, JDL)<br>カシオ | BW-20B(ローマ字可)           | 260万      |
|                           |                 | 連 想 コー ド 日 立<br>リ コ ー          | BW-20B                  | 260万      |
|                           |                 |                                | R I P O R T 410         | 270万      |
|                           |                 |                                | " 600モデル140<br>(FAX機能付) | 448万      |
|                           | 三 角 信 号 伊藤忠データ  | WANG・OIS                       |                         |           |
| フ ル キ ー<br>(本文字配列)        | タ ブ レ ッ ト       | シ ャ ー プ 沖                      | レターメイト                  | 295万      |
|                           |                 | 日 電                            | 文豪(NWP-20N)シリーズ         | 185万      |
|                           |                 |                                | NWP-21N(入力専用)           | 140万      |
|                           |                 |                                | 22N(ラインプリンター)           | 298万      |
|                           |                 |                                | 23N(ページプリンター)           | 458万      |
|                           |                 |                                | クラスター                   | —         |
|                           |                 | 松 下 通 信                        | パーソワード1000              | 約 350万    |
|                           |                 | リ コ ー                          | R I P O R T 410         | 未 定       |
|                           |                 |                                | " 600モデル240             | 458万      |
|                           |                 |                                | (FAX機能付き)               | 約 250万    |
|                           |                 | 日 立                            | BW-20A                  | 250万      |
|                           |                 | ベ ン て る                        | レタコンPW-802              | 295と約350万 |
|                           |                 | J D L                          | 文 作                     | 198万      |
|                           |                 | 伊藤忠データ                         | OIS                     | —         |
|                           |                 | ベ ー ジ                          | ワープロワン                  | 290万      |
|                           | 多 段 シ フ ト       | 漢字ストローク                        | WORDIX                  | 235万      |
|                           |                 | 横河・<br>ブラザー                    | 入力専用機                   | 120万      |



られる。

新しい文章情報データベースを構築する場合、日本語情報については、これまでに述べた漢字入力のようなハンディを乗り越えて、大量入力を行うための入力処理系を設計していかなければならないし、大量入力可能なソース源を求めていかなければならない。現在、朝日新聞社や日本経済新聞は、新聞のコンピュータによる自動編集を開始している。これらの作業の副産物として、今後新聞記事情報は低コストで磁気テープ化されていくであろう。これが大量入力可能な日本語情報ソース源の第1である。日本語ワードプロセッサの利用が盛んになると、各方面でデータベースを作成する以外の目的で、様々な文書がフロッピー・ディスクに蓄積されていくであろう。各フロッピーから蓄積された文章情報を吸い上げるしくみを作ることができれば、大量ソース源の第2の可能性が生まれてくる。第3の可能性は、いまだ実用段階に至っていないが、漢字OCRの発達が期待される。パターン認識による漢字入力方式の中では、漢字OCRが最も大量入力実現の可能性を秘めているといえることができる。

## 1.4 単語およびキーワードの自動抽出

現在のデータベースでは、検索のためのキーワード付けは、整備されたキーワード体系に基づいて人力で行っているのが一般的である。人間による作業にはムラがあり、様々な主観的バイアスが上乘せされてしまう。検索効率、適合率を均質化し、データベース作成作業を合理化するためにも、キーワード付けの作業が一定のルールで自動化されることは望ましい。また、文章情報の定量化利用のためには、できるだけ客観的な方法によって分析の対象となる二次情報が一次情報から集約されてくる必要がある。本節では、第1回ワーキング・グループの資料に基づいて、単語およびキーワードの自動抽出の問題を考えてみる。

### (1) 単語の自動抽出について

単語の自動切り出しの方法としては、漢字・仮名という文字種の違いを利用する方法と辞書とのマッチングによって切り出す方法の2種類がある。ここでは、工業技術院で開発した単語頻度数集計のプログラムを例として両者の特徴を考えてみる。システム・フローの概略を図1-12に示す。

このシステムにおける辞書を用いないで単語を抽出するためのルールは、2文字以上連続した漢字列を拾うということである。図1-13に辞書を用いないで単語抽出を行った例を示す。このように、2文字以上連続した漢字列を拾うというきわめて単純なルールによっても相当数の単語を抽出することが可能である。この方法で問題になる点は、図1-13の中には適切な例が出ていないが、平仮名を含む名詞が脱落してしまうということである。また、この例の中にあるように「実際の」とか、ここには出ていないが「徹底的に」、「記述的に」、「理論上の」といった副詞あるいは形容詞として働いている漢字列も抽出してしまうことは拾いすぎかもしれない。この例の中で抽出された「使用」は、使用するという動詞であるが、中には使用という名詞で出現するケースも当然考えられる。従っ

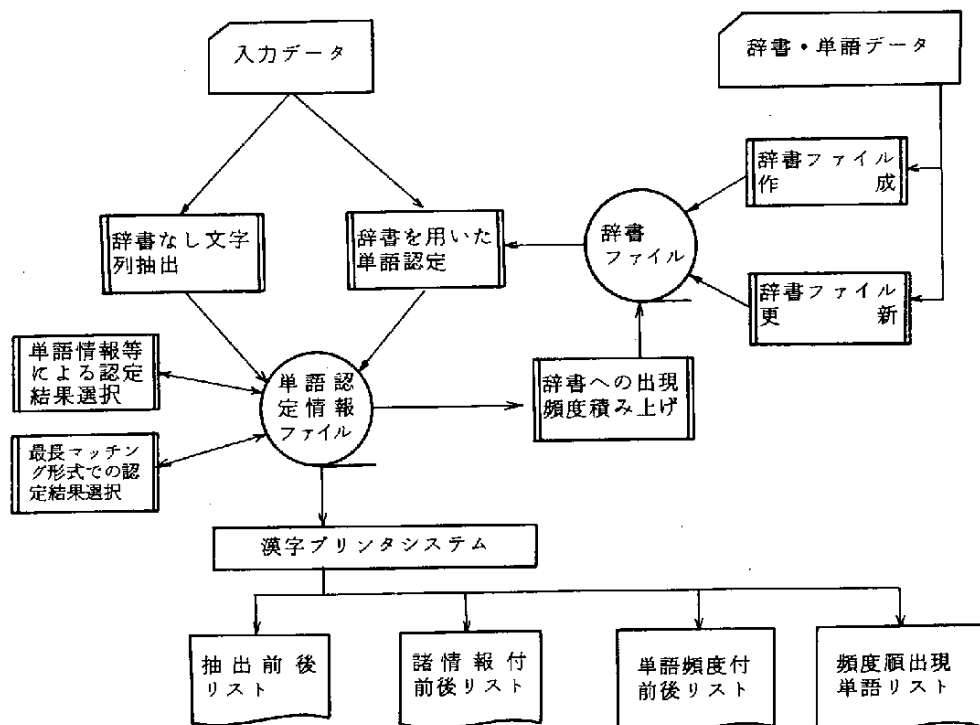


図 1 - 1 2 単語頻度数集計のためのシステム・フロー

| (入力文レコードID部) | (文字数) | (直前文字列)     | (抽出文字列) | (直後文字列)                          |
|--------------|-------|-------------|---------|----------------------------------|
| 00000005     | 102   | 会に知っており、その  | 言語      | に関する自分の知識を、実際の言語運用 (performance) |
| 000'00005    | 102   | り、その言語に関する  | 自分      | の知識を、実際の言語運用 (performance)       |
| 00000005     | 102   | の言語に関する自分の  | 知識      | を、実際の言語運用 (performance) にお       |
| 00000005     | 102   | 関する自分の知識を、  | 実際      | の言語運用 (performance) において使用       |
| 00000005     | 104   | 自分の知識を、実際の  | 言語運用    | (performance) において使用する際          |
| 00000005     | 102   | manca) において | 使用      | する際、文法とは特に関係をもたない条件、たとえば         |

図 1 - 1 3 辞書を用いない方法による単語抽出例

て、漢字列で単語を抽出した場合には品詞の情報を取捨してしまう結果となる。文章から動詞を抽出することに力点を置く場合には、2文字以上の漢字列に限ってしまい、1文字の漢字を切り捨てることはいささか問題がある。

次に、図1-13と同じ文章を用いて辞書マッチングの方法で単語を抽出した例を図1-14に示す。この例を見るとわかるように、辞書を用いない単語抽出よりも一層きめの細かい単語抽出が辞書マッチングの方法によって行えると言えるであろう。ただし、「関する」という単語を関すると関するで抽出したり、「実際」という単語を実際と実際で抽出したりしているのは、単純な辞書マッチングの欠点である。また「言語運用」という複合名詞は、漢字列抽出では言語運用で拾い出されたのに対し、辞書マッチングではこの複合名詞が辞書に登録されていないため、辞書中にある言語と運用に分解して抽出されている。このような複合名詞がそのまま抽出されるべきか、分解して抽出されるべきかは大いに議論の別れる所であるが、単純な辞書マッチングだけでなく抽出した単語を認定するためにもう一工夫必要である。ここでは、最長マッチングの方法を採用して、「関する」や「実際」は関すると実際の形で一意的に認定されるようにしている。「言語運用」という言葉が辞書に登録されていれば言語運用という分解されない形で抽出されてくるであろう。いづれにしても、辞書マッチングの方式を使用して単語を抽出する場合には、各種の辞書を単語抽出の目的に併せて、必要なものを完全に登録した形で準備しなければならない。

より高度な単語抽出の方法としては、抽出した単語の品詞の識別を行ったり、語尾変化のある品詞に関しては語尾変化をおさえて基本形に一元化したりするような文法論を反映した方法が考えられる。また、抽出した単語の品詞の種類あるいは単語に付属している助詞の種類などから単語の文中での役割を判断していく語用論を反映した方法を応用していくことも

| (認定シーケンス) | (単語ID) | (品詞)  | (文字位置) | (文字数) | (直前文字列)     | (認定<br>単語) | (直後文字列)                       |
|-----------|--------|-------|--------|-------|-------------|------------|-------------------------------|
| 453       | 192    | JOSHI | [119]  | 1     | 完全に知っており、その | の          | 言語に関する自分の知識を、実際の言語運用 (perform |
| 454       | 392    | KANJI | [120]  | 2     | 全に知っており、その  | 言語         | に関する自分の知識を、実際の言語運用 (perform   |
| 455       | 191    | JOSHI | [122]  | 1     | 知っており、その言語  | に          | 関する自分の知識を、実際の言語運用 (performan  |
| 456       | 333    | VERB  | [123]  | 3     | っており、その言語に  | 関する        | 自分の知識を、実際の言語運用 (performanc    |
| 457       | 144    | VERB  | [124]  | 2     | ており、その言語に関  | する         | 自分の知識を、実際の言語運用 (performance   |
| 458       | 457    | NOUN  | [126]  | 2     | り、その言語に関する  | 自分         | の知識を、実際の言語運用 (performance) に  |
| 459       | 192    | JOSHI | [128]  | 1     | その言語に関する自分  | の          | 知識を、実際の言語運用 (performance) におい |
| 460       | 578    | NOUN  | [129]  | 2     | の言語に関する自分の  | 知識         | を、実際の言語運用 (performance) において  |
| 461       | 229    | JOSHI | [131]  | 1     | 語に関する自分の知識  | を、         | 実際の言語運用 (performance) において使用  |
| 462       | 460    | NOUN  | [133]  | 2     | 関する自分の知識を、  | 実際         | の言語運用 (performance) において使用する  |
| 463       | 437    | NOUN  | [134]  | 1     | する自分の知識を、実  | 際          | の言語運用 (performance) において使用する際 |
| 464       | 192    | JOSHI | [135]  | 1     | る自分の知識を、実際  | の          | 言語運用 (performance) において使用する際、 |
| 465       | 392    | KANJI | [136]  | 2     | 自分の知識を、実際の  | 言語         | 運用 (performance) において使用する際、文  |
| 466       | 278    | NOUN  | [138]  | 2     | の知識を、実際の言語  | 運用         | (performance) において使用する際、文法と   |

図1-14 辞書マッチングによる単語抽出例

考えられる。

## (2) キーワードとしての意味付け

このようにして抽出した単語に対して、いかんしてキーワードの意味を与えていくかということが、次の段階での問題である。現在までにキーワードの意味付けとして使われている方法は、重要語抽出である。すなわち単語出現頻度数など統計的手法を応用したもので、SaltonのSMARTシステムとか言ったものを例としてあげることができる。抽出した単語から出現頻度数のきわめて高い不要語を除いてみた場合、確かに出現頻度数の高い単語は重要語であるという判断が一般には成り立つと考えられる。

しかし、こうした方法で単語をキーワード化した場合、単語が文中で持つ役割の情報は完全に脱落してしまう。キーワードとしての意味を単語に持たせるためには、文あるいは文章中での単語の役割を抽出し、単語に付加することが必要ではなかろうか。このためには、文中における格あるいは5W1Hを判断して、これらに該当する単語を重要語としてキーワードの意味を持たせていかなければならない。例えば、同じ「米国」でも、主体として働いているか、対象として働いているか、場所として働いているかでは、文中の役割は全く異なっている。従って、キーワードとしての意味も異なってくると考えられる。現在までのところ、単語の文中での役割まで追究してキーワード付けの自動化を行ったシステムは十分な実用の段階にいたっていない。

キーワードとしての意味付けを行う場合に考えなければならない重要な問題の一つは、図1-13あるいは図1-14に現われている「言語運用」のような複合語の問題である。日本語は、英語などに比べてきわめて造語能力の大きい言語である。年々新たに造語される複合語をキーワードとして登録していくならば、辞書は無限の大きさを必要とすることになる。また複合語が増加すればするほど、キーワード認定のステップにおける不確度も増加していくことになるであろう。「言語運用」を「言語」

と「運用」に分けて登録すべきかそのまま登録すべきかは、種々の観点から大きく議論の分れるところであるが、語管理という観点から見た場合、できるだけ最小の単位で管理できることが望ましい。このためには、最小単位である「語基」の研究と「複合語合成の規則」に対する研究に力を注ぐことが必要であろう。

新しい文章情報データベースを構想していくにあたって、入力文からの単語の自動抽出化の可能性を考えてみると、漢字列抽出、辞書マッチング、最長マッチングなどの手法を組み合わせて単語認定を行う方法で相当の部分まで日本語情報について自動化できるのではなからうか。これに加えて語尾変化の判定による基本形への還元、品詞の判定、助詞などの判定による単語の役割付けといった各種情報を獲得するアルゴリズムが確立できれば、単語抽出自動化機能としては申し分ないと考えられる。

これに対して、キーワード付けの自動化は、まだまだ難しいと考えられる。人間がある整備されたキーワード体系からキーワード付けを行う場合には、単にその単語あるいは同義語が見付かったから行うのではなく、文あるいは文章全体の意味を判断してそれを集約させた形でキーワード付けを行っている。このことを考えただけでも、機械的に現在できる方法との間には大きなギャップがあると言えよう。キーワード付けの自動化のアルゴリズムは、文を単語にまでバラバラに分解してから意味付けしていく方法ではなく、文の状態から含まれる単語も含めて一定ルールで情報を抽出していくような方法の中に見出される可能性があるのではないだろうか。

## 1.5 シソーラスの作成

一般にデータベース・システムでは、データベースの作成サイドにおいても、データベースの利用サイドにおいても、各種辞書類が陰に陽に重要な役割を果たしている。辞書類の中でもキーワードのシソーラスは、検索効率を上げ、呼び出し率と適合率を改善してユーザの望む検索結果を提供するために、きわめて重要な辞書である。しかしながら、シソーラスの作成は、膨大な作業量を伴い広範な知識による体系的な分析を必要とする困難な作業である。本節では、通商産業省情報管理課におけるMITI-IRの経験を通して、シソーラス作成における種々の問題点を考えてみる。

### (1) 手作業によるシソーラス作成

通商産業省におけるコンピュータ利用は、電気試験所のETL MARK-4の完成を契機として本格的に開始されている。数値情報と文章情報の双方が、このコンピュータによる情報処理の可能性を追究する対象として選ばれている。数値情報に関しては、比較的早期に利用の目度が立ったが、文章情報に関してはコンピュータによる情報処理の種々の体験を行い、問題点の洗い出しを行うということで昭和37年にスタートした。この研究は、昭和45年に成果をまとめて一区切りをつけているが、その中でシソーラス作成の基礎となる各種の実験も行っている。

万能シソーラスは考えられるかという問題が出発点となって、手作業によるシソーラス作成の実験は開始された。第1歩は、通商産業省および外務省における漢字の頻度調査である。この結果、約1000字種のところを境目として100%のカバレッジに近づき、飽和状態にいたることが判明した。また通商産業省あるいは外務省全体でのカバレッジ曲線が、原局別のカバレッジ曲線よりもやや上方にくることが判明した。

次に、国際関係、貿易、個別産業・企業経営、学術用語、政策、生産、工程、消費・流通の8つの分野から無作為抽出で26種の資料を選択し、



用語抽出を行った。表 1-5 に示すように抽出された用語数は約 32 万語で、ダブリを除去して約 15,000 語の用語にまとめられている。これらの用語の中から、きわめて頻度の高い言葉と 4 回以下のきわめて頻度の低い言葉とを除去して絞りを加え、10 種類の項目に各用語を分類し

表 1-5 通商産業省におけるシソーラス作成例

| 内 容            | 件 数          |
|----------------|--------------|
| 資 料            | 8 グループ, 26 種 |
| 抽 出 用 語        | 約 32 万語      |
| 用 語            | 約 15,000 語   |
| 項 目            | 10           |
| 頻度 5 ~ 400 の用語 | 4,000 語      |
| 国際貿易関係でのテスト    | 132 語        |
| 全 シ ソ ー ラ ス    | 約 10,000 語   |

た。10 種類の項目すべてに登場する言葉はコモン・ワードとしてやはり除去されている。分類したなかの 1 項目ごとに、カルタ取りの要領で各用語の同義語、上位概念、下位概念、関連語、反意語を選び出し、完全なシソーラスを作成していった。この手作業には約 3 年の日時と当時の金額で 2,000 万円の経費を必要とした。

さて、このような手作業によるシソーラス作成の経験を通して、いくつかの点が明らかとなった。第 1 に大きな点は、すべての作業に人手を使ってシソーラスを作ってはいけないということである。その理由は、複数の人間が作業に関与した場合、10 項目の分類あるいは上位概念、下位概念の分類に関して、異なった人間の間で全く統一性と一貫性が取れないということである。1 人の人間がある言葉間に上位、下位の順序を付けたとすると、別の人間はこれとは全く逆の上位、下位の順序を付けてしまう問題が頻繁に起こってしまうのである。そこで、同一の人間が作業を行った

らどうであろうかという点をチェックしてみた。資料の一部を抽出して用語の分類を行い、それから3カ月後に全く同一の人間が同一の資料に対して同一の作業を繰り返してみたわけである。結果は、2つの用語の分類が全く異なったものとなってしまったということである。1人の人間の中でも、判断の基準がいかに揺れ動くかということを示す典型的な例となってしまった。また、1回目の作業の学習効果が2回目の作業に影響を与えている可能性もある。従って、シソーラス作成作業のすべてを人間の手作業を通して行うのではなく、分類の判断を行う過程は一定のルールを確立して客観的な機械操作を取り込むことが必要である。

第2に大きな点は、この作業に要した3年という歳月は作業時間としては長すぎるということである。言葉の体系は、時間が経過するにつれて変動しているとみることができる。従って、このような長期の時間を要してシソーラスが完成した時期には、もはやそのシソーラスの体系は現実の言葉の体系からははずれてしまい、陳腐化が始まっているといえることができる。これに対処するためには、いずれにしても人間による作業でなくもっとハイ・スピードでシソーラス作成の作業ができるような仕組みが必要である。

## (2) コンピュータの部分作業によるシソーラス作成

上述のような手作業によるシソーラス作成の経験に基づいて、コンピュータの部分作業を加えたシソーラス作成作業を行ってみた。すなわち用語の分類を行う操作の一部分に、コンピュータ作業を挿入し、分類作業の客観性の保持と作業時間の短縮化をはかったわけである。当時、語と語の間の関連度をコンピュータを用いて求め、シソーラス作成の際の補助手段とする実験的試みが、日本科学技術情報センター（JICST）、日本貿易振興会（JETRO）などで行われていた。関連度を求めるベースは「情報あるいは記事のなかで同時に含まれる確率の高い用語はお互いの間になんらかの関係がある」という考え方である。語の関連度を表わす係数

は関連係数と呼ばれるが、日本科学技術情報センターの実験報告によると、図 1-15 に示すような包含関係を考えた場合、関係係数を求める式は表 1-16 に示すような①～⑨のものが考えられる。これらの式は、いず

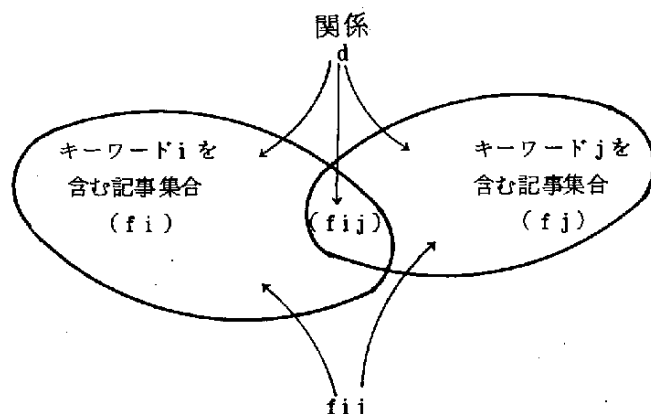


図 1-15 キーワードが関連する記事の包含

表 1-6 関連係数を求める式

- ①  $A_{ij} = f_{ij} - (f_i \times f_j) / d$
- ②  $A_{ij} = f_{ij} + \bar{f}_{ij}$
- ③  $A_{ij} = f_{ij} / (f_i + f_j - f_{ij})$
- ④  $A_{ij} = \{f_{ij} \times \bar{f}_{ij} - (f_i - f_{ij})(f_j - f_{ij})\} / \{f_{ij} \times \bar{f}_{ij} + (f_i - f_{ij})(f_j - f_{ij})\}$
- ⑤  $A_{ij} = \log_{10} \left[ \left( \left| f_{ij} \times d - f_i \times f_j \right| - \frac{d}{2} \right)^2 \times d / \{f_i \times f_j \times (d - f_i)(d - f_j)\} \right]$
- ⑥  $A_{ij} = \log_{10} [ \{ (f_{ij} \times d - f_i \times f_j)^2 \times d \} / \{ f_i \times f_j \times (d - f_i)(d - f_j) \} ]$
- ⑦  $A_{ij} = (f_{ij} - (f_i \times f_j) / d) / \sqrt{f_i \times f_j}$
- ⑧  $A_{ij} = \{f_{ij} - (f_i \times f_j) / d\} / \{(f_i + f_j) / 2\}$
- ⑨  $A_{ij} = (f_{ij} \times d - f_i \times f_j) / \sqrt{f_i \times f_j \times (d - f_i)(d - f_j)}$

$f_i$ : キーワード  $i$  が含まれる記事数

$f_{ij}$ : キーワード  $i$  と  $j$  が同時に含まれる記事数

$\bar{f}_{ij}$ : キーワード  $i$  と  $j$  が同時に含まれない記事数

$d$ : 全記事数

$A_{ij}$ : キーワード  $i$  と  $j$  の関連係数

れを用いても得られる結果に大きな差異はない。そこで①の式を用いて、シソーラスの作成を行ってみた。この結果を手作業で行ったシソーラスと比較してみると、分類の整合性と客観性、作業時間など種々の面で、コンピュータを用いた作業によるシソーラスの方がすぐれていることが判明した。従って、シソーラス作成作業の中には、できうる限りアルゴリズムを確立して、コンピュータによる作業を取り込んだほうが良いことが結論付けられる。

以上、手作業とコンピュータの部分作業によるシソーラス作成の2つの経験を通して、シソーラス作成はできるだけ人間の作業を避け、コンピュータを利用した機械作業で行うべきであるとまとめられる。ましてや作成したシソーラスのメンテナンスのことまで考慮すると、相当部分が機械作業で処理できるようになっていなければ対応しきれないであろう。文章情報のコンピュータ処理に関して種々の体験を重ね、問題点の洗い出しを行おうとする通商産業省の試みは、昭和45年度をもって中断した。中断の理由は、ハードのコストダウン、コンピュータの処理能力のレベルアップに10年程度の歳月が必要で、入出力関係のコストが安くなるまで待とうということである。また、人間とコンピュータとの関係すなわちマン・マシン・インタラクションの有るべき姿ができ上がるのにもかなりの時間が必要であろうということも理由の一つとなっている。

この20年余の進歩でコンピュータの価格は1万分の1となり、TSSシステム導入といった形でマン・マシン・インタラクションのシステム確立も可能な状況となってきた。このような観点から考えると、中断した文章情報処理に対する新しい挑戦を開始すべき時期に十分至っていると判断できる。新しい文章情報総合解析システムを構築するにあたっては、この通商産業省における貴重な実験結果を十分生かして、辞書類の整備にあたる必要があるであろう。特にシソーラスについては、できるだけ多くの作業をアルゴリズム化してコンピュータ作業とし、短時間の編成作業に

よる分類基準のしっかりした陣腐化しないシソーラスを作成することが重要である。このためには、すべての領域をカバーした万能シソーラスではなく、分野別にコンパクトな固有のシソーラスを作成し、分野別シソーラスの集合体として万能シソーラスを実現するような方向性を取ることも必要であるかもしれない。シソーラスの作成については、完全自動化、半自動化による方法など海外でも種々の検討がなされているので、これらの結果も取り込んで新しい文章情報総合解析システムの構築に取り組んでいかなければならない。

## 1.6 定量化利用の方法とユーザの使い易いシステム

新しい文章情報総合解析システムの中では、これまでの一般検索によるデータベース内容のリストアップだけでなく、多角的な文章情報の利用方法を考えていきたい。その一つの方向性として、文章情報の定量化利用が考えられる。これは文章情報の内容を内容分析などの手法により定量化して数値情報に変換し、この数値情報に対して多変量解析、数量化理論などを適用して有効な分析結果を導き出すという方法である。本節では、第6回ワーキング・グループの資料に基づいて、政治学の分野における文章情報の定量化利用事例を検討しながら定量化の方法と問題点を考える。また政治学の分野におけるユーザの観点からみて使い易いシステムのあり方についても言及する。

### (1) 政治学の分野における利用事例と方法

政治は、無言の威圧による政治といったことが行われるケースもあるが、一般には種々の政治コミュニケーションを通して行われる。従って、政治を取り扱う政治学の分野では、早くから言語情報を分析する手法が取り入れられてきた。特に、国際政治学の分野では、記事情報として掲載される政治コミュニケーションを分析する手法が重要である。政治学の分野における文章情報の定量化利用事例をいくつか紹介する。

第1の例は、新聞に掲載された政治コミュニケーションから内容分析の手法を用いて種々のキーの出現頻度数などを抽出し、これを主成分分析、因子分析、数量化理論などを使って解析したもので、ソ連、中国、北朝鮮の友好敵対関係、ソ連、中国、ベトナムの友好敵対関係に関して大変興味深い結果を提供している。政治の世界では、本当に重要な言葉は限られていて、政治コミュニケーションの文章もあまり複雑な論理構成を取らず、主張、宣伝、訴えを行っているものが多い。従って、

ある種のキーワードを着目して追究することにより、かなりの分析を加えることが可能である。

共産圏の場合には、建国記念日とか革命記念日とかいった重要な日、党機関誌などの重要メディアに、党書記長、党主席、首相といった重要人物の発信した祝電が掲載される。祝電の内容は、祖国統一、人民団結など毎年毎年同じ主張を繰り返しているかのように見えて、例えば「一緒にやろう」と「めいめい頑張れ」では大きな違いがあるように、種々の変動が起っている。従って、このような政治的コンテキストを踏まえて、重要語の出現頻度数、文章の長さ、掲載面、掲載日といった情報を抽出して分析するとかなり有意な結果を引き出すことが可能である。

表1-7と図1-16に解析結果の一例を示す。これは、ソ連、中国、北朝鮮3カ国の政策決定者間で交した交友的な内容で政治的に重要な文言の数から3×3のソシオメトリー・マトリックスを1961～1966年の6表を作成し、林知己夫の数量化理論第IV類を適用して1次元値を計算したものである。この1次元値は、国家間の距離すなわち国家間の友好敵対の度合を示す一つの指標となっている。これらの指標の変化は、この期間に3カ国の間で起った重要事件と密接に関連するものである。

表1-7 共産3カ国に与えられた1次元値

|       | 1961年  | 1962年  | 1963年  | 1964年  | 1965年  | 1966年  | 1961～<br>1966年 |
|-------|--------|--------|--------|--------|--------|--------|----------------|
| 中 国   | 0.743  | 0.799  | 0.462  | 0.751  | 0.707  | 0.769  | 0.755          |
| 北 朝 鮮 | -0.665 | -0.256 | 0.352  | -0.099 | 0.000  | -0.147 | -0.108         |
| ソ 連   | -0.078 | -0.543 | -0.814 | -0.652 | -0.707 | -0.622 | -0.647         |

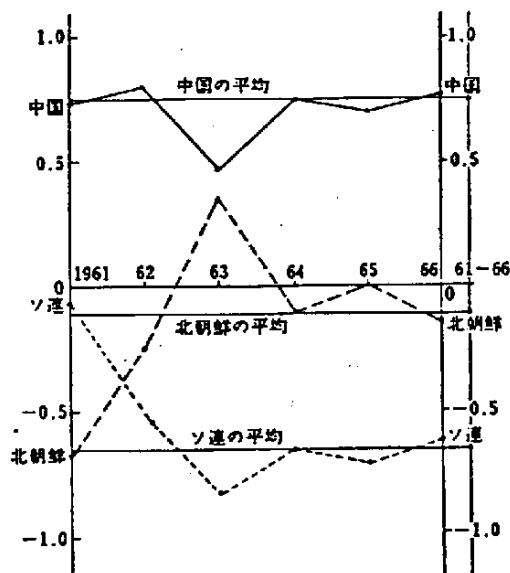


図 1 - 1 6 国 家 間 の 距 離

国際政治学の分野，特に共産圏の場合，このように比較的単純な重要語に対するサーベイで，有意な結果を導き出すことができる。しかしながら，エネルギー問題，カントリー・リスク問題といった分野で，有効な解析ができるかどうかは必ずしも明らかではない。いずれにしても，それぞれ独自の重要語の分析と解析パターンの検討が必要となるであろう。

第2の例は，R. Wich による「Sino-Soviet Crisis Politics」である。これは，1968年2月のベトナムでのテト攻勢，ジョンソン大統領の再選出馬断念，8月のソ連によるチェコ介入，1969年3月の中・ソ国境衝突，9月のベトナム・ホーチミン大統領の死，1979年2月の中越紛争といった事件を新聞，雑誌，ラジオ情報などが取り上げ，それぞれの事件の因果関係を論理的に解明したものである。このように事件のエッセンスを直感で取り出し，因果関係を論理立てていく方法は，



一種の神託的な結論を与えるものであるが、大がかりなシステムを必要としない一つの方法論としてカウントできるものである。エネルギー問題の分野でも、このような方法論で各種の分析が行われているはずである。

第3の例は、山本吉宣による認知構造図を使った政策シミュレーションである。この分析では、日中平和友好条約締結当時、外務省アジア局長であった中江要介が、日中平和友好条約に関連してインタビューに答えた資料をもとにして認知構造図を作成している。認知構造図は、資料の中から因果関係についての主張（コンセプト）を抽出し、有向グラフを作成したものである。抽出したコンセプトとしては、000 日本の効用，001 日中条約の締結，002 将来のアジアの安定，アジアの調和，……101 日本の基地使用に対する中国の圧力，……301 日中条約早期締結，……401 日本が中国に及ぼした大きな害を反省，……525 北朝鮮の効用などが抽出されている。これらのコンセプトから有向グラフを作成したものが図1-17である。

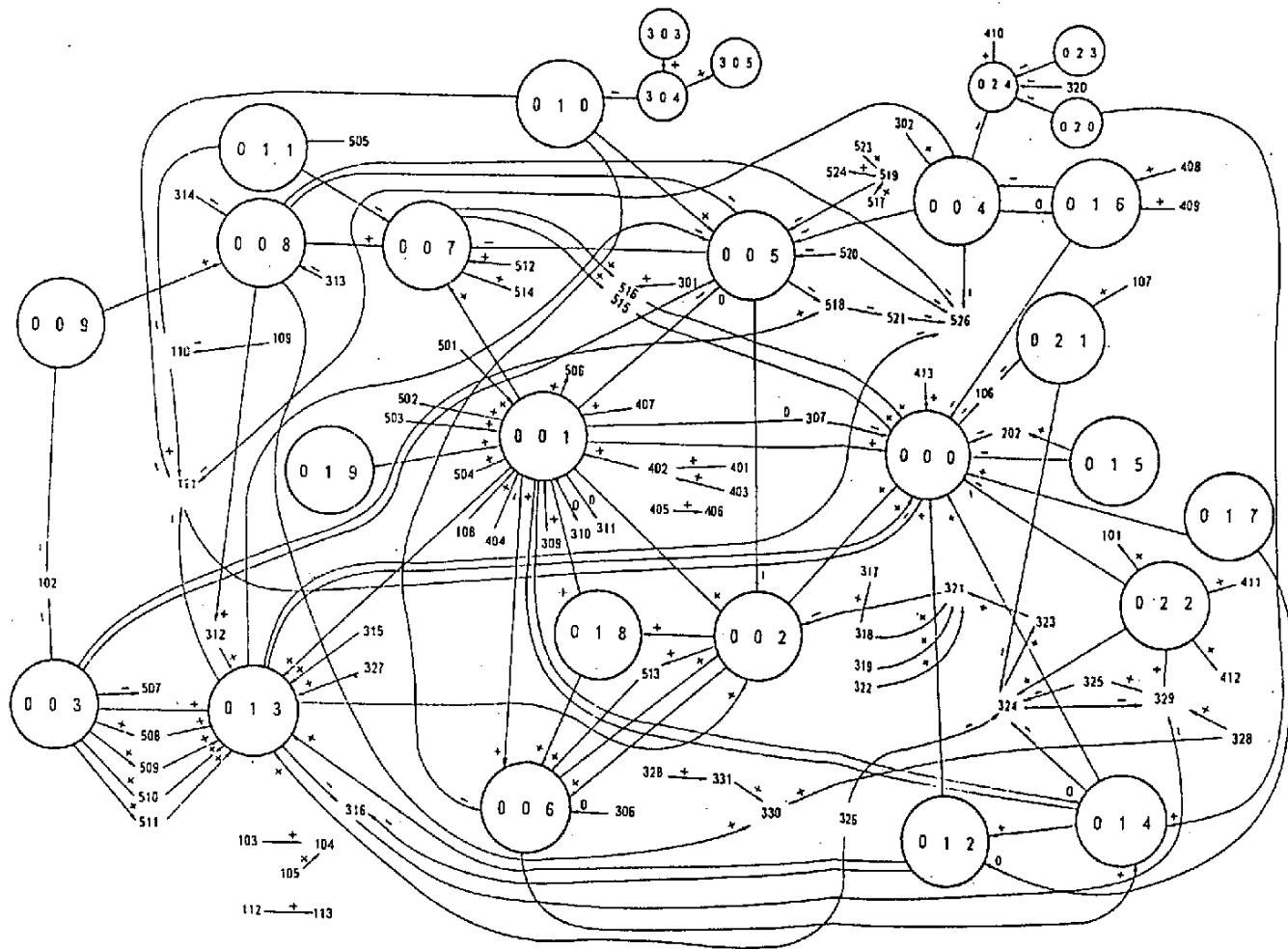


図1-17 中江要介氏の認知構造図

コンセプトの中には、政策(P)コンセプト、効用(U)コンセプト、価値(V)コンセプト、認知(C)コンセプト4種がある。日中平和条約に関連する資料から作成した認知構造図を基にシミュレーション用の認知構造図を作成し、中越紛争に関する政策シミュレーションを行ったものが、表1-8である。これは、政策コンセプトと価値コンセプトの関係を示したものである。この中でP<sub>2</sub>、P<sub>4</sub>という政策が日本の価値にプラスの因子となっている。P<sub>3</sub>は逆にマイナスの因子となっている。日本政府はP<sub>2</sub>、P<sub>4</sub>の政策を取る可能性が最も高く、P<sub>3</sub>を採らないであろうということがこのシミュレーションの予測結果である。「中国とできるだけ一線を画す」という日本の原則が貫かれている、というのがこの分析の主張の一つである。

表1-8 シミュレーションの結果

| 仮<br>りの<br>順<br>位<br>(もとの番号) | 選 択 肢<br>日本<br>の<br>価<br>値 | P <sub>1</sub><br>静 観 | P <sub>2</sub><br>中国を<br>単に批<br>判 | P <sub>3</sub><br>ベトナムを<br>単に批判 | P <sub>4</sub><br>国連での<br>対中決議<br>を勧める | P <sub>5</sub><br>中国に<br>対し<br>て<br>中<br>味<br>援<br>助<br>す | P <sub>6</sub><br>ベトナム<br>に<br>対<br>し<br>て<br>味<br>援<br>助<br>す |
|------------------------------|----------------------------|-----------------------|-----------------------------------|---------------------------------|----------------------------------------|-----------------------------------------------------------|----------------------------------------------------------------|
| 1<br>(63)                    | 日本の独立と安全                   | +                     | +                                 | -                               | +                                      | +                                                         | -                                                              |
| 2<br>(62)                    | 日 本 の 貿 易                  | +                     | +                                 | -                               | +                                      | +                                                         | -                                                              |
| 3<br>(58)                    | 中ソ対立に日本が<br>まきこまれない        | 0                     | +                                 | -                               | +                                      | -                                                         | +                                                              |
| 4<br>(60)                    | 東南アジア諸国の<br>効用             | 0                     | +                                 | -                               | +                                      | +                                                         | -                                                              |
| 選 択 肢 の 順 位                  |                            | 3                     | 1                                 | 6                               | 1                                      | 4                                                         | 5                                                              |

〔注〕 +(-)：政策P<sub>i</sub>が、日本の価値変数にプラス(マイナス)に効く。

第4の例は、日本の選挙公約に対する分析である。これは、日本の政党が右か左かという位置関係を分類するため行ったもので、言語情報の数量化の1例となっている。この分析では投票前に新聞に発表される各党の選挙公約から、法と秩序、経済の運営、社会のしくみ、外交といった部類で種々の問題点について各党の態度を抽出し、これらの情報をもとに各政党の位置付けを行っている。約20年分の選挙公約が取り扱われており、使っている分析手法は因子群分析である。因子群分析の結果は、政党の位置付けに関して案外と良好な解を与えている。ただし、この場合に日本の政党を右と左に分けている因子は何かということについて、ある程度しっかりした認識を持っていないといけない。この認識に基づいて因子群分析を行えばそれなりの結果が得られるが、情報全部を因子にして分析を行えばわけのわからない結果が出てくるのは当然である。

第5の例は、田中昭彦の中国の対外構造に関するシミュレーション分析である。世界の出来事を非常に丹念にまとめたFacts on Fileという刊行物があるが、これから中国のいろいろな国々とか、いろいろなものに対してどのような態度を取ったかという情報を数量化して抽出している。抽出した数量情報を基にして、中国国内の動きと対外的アグレッシブネス、パッシブネスとを関係付けたモデル化を行い、シミュレーション分析を行っているわけである。

この分析で取り上げている手法は、アメリカ国防省の外郭団体であるCACI社が開発した方法が一つのベースとなっている。CACI社は、危機管理システムを組織しようということでこれまで研究を進めてきている。このシステムでは、アメリカの対外政策の全領域を対象として、政策決定の方向性を与える結論を与えようとしている。すなわち、アメリカの政策のゴールは何か、そのゴールを実現するための行動としてはどのような選択があるか、ウェイト配分はどの程度かといった情報がシ

システム内に組み込まれていて、ある事件を入力するとそれに対してアメリカの取るべき行動がウェイト付けられて出力してくるという仕組みである。前述したFacts on Fileを内容分析して得た数値情報の結果が、このシステムにおけるベース情報の一つとなっているのである。

日本の政治学者というのは、500人前後しかいないが、以上5種の例に示したように、結構各種の文章情報を利用したシミュレーションを行っている。これらの方法が、そのままエネルギー問題やカントリー・リスク問題の分析に適用できるかどうかは不明であるが、それぞれの問題に適した方法論を立てることは可能であろう。

## (2) 政治学のユーザからみて使い易いシステム

(1)で述べたように、政治学の分野では言語情報を利用したかなり多くの分析が行われており、この中の相当部分は手作業による研究である。コンピュータ作業による研究ができるようになることは望ましいことであるが、政治学者としてのユーザからみた場合、使い易い解析システムの在り方についていくつか言及する。

第1は、入出力がユーザにとって親しみ易くわかり易いものであってほしいということである。例えば、入力に関しては、入れるコンセプト、キーワードあるいは記号が、種々の政策と論理的に因果的に深く結びついているものを使用するのがよい。こうすることによって、政策思考の流れに沿ってシステム操作ができるようになるはずである。出力に関しても、できるだけ単純なものであることが望ましい。しかも、ドライな情報の羅列ではなく、ポジティブかネガティブか、どのような判断をくださか、どのような行動を取るか、といった政策決定者に本質的な内容が出力結果に一目瞭然となって表われるシステムがよい。システム操作に関しては、インタラクティブな形で、ユーザが出力結果を見ては、システムと応答を繰り返していけるものが便利であろう。

第2は、前述の要望とも深く関連することであるが、システム全体が

多くのオブションの中からユーザに取って価値のある判断を下した結論をその場で出力するような仕組みになってほしいということである。膨大で無味乾燥な出力を与えられて、その中から必要な価値判断を下して、ほしい結果を得て下さいというのでは、ユーザはとてもついていけないであろう。このようなシステムを作り上げるためには、あらかじめ種々のオブションと価値判断の基準を体系付けた因果関係図を作成しておかねばならない。十分吟味した因果関係図をシステム内部にセットし、入力として何か事件を与えると、因果関係図に沿った判断の結果として、ある行動なり政策が出力されてくるシステムが必要である。アメリカのCACI社のシステムは、こうした要請をかなり満したシステムとなっている。現実には海外ではこのような種類のシステムができていのであるから、日本で作れないということはないと考えられる。新しい文章情報総合解析システムを作成する場合には、このようなユーザ・サイドの観点からの要請も十分考慮していくことが必要であろう。

最後に、これは政治の分野では、ある程度宿命的と言えるが、いくら優秀なシステムでいくら確度の高い結果が出たとしても、高度に政治的な判断として結果が無視されることはいくらでも覚悟しなければならない。もちろん、そのような判断が下されるようなシステムになれば、それだけでも十分すぐれたシステムといえるであろう。話は情報システムとは異なるが、エジプトとシリアがイスラエルに侵攻した時、イスラエルの情報網は動きを十分キャッチしていたにもかかわらず、イスラエル上層部の攻撃開始時間の読みが4時間ずれたために被害が莫大なものとなったという事例がある。このように、システムの出力結果を政策決定レベルが見向きもしないということは往々起りうることであるから、政治の分野ではシステムの宿命としてある程度肝に命じておかねばなるまい。

## 1.7 新しい記事情報総合利用システムの構築

1.2～1.6まで新しい文章情報総合解析システムを構築するにあたって考慮しなければならない種々の問題を検討してきた。本節では、全体のまとめとして第5回ワーキング・グループ資料と1.2～1.6の各節の内容に基づいて新しい記事情報総合利用システムのシステム像と開発のタイム・スケジュールを検討する。

### (1) 記事情報の利用の方向

記事情報は、文章情報の1つとして大きな一群を形成している。新聞・雑誌に現われては消えていく記事情報をソース・データとして、記事情報総合利用システムを考える時、利用の方向性として大きく次の3つの方向が考えられる。

第1の利用の方向は、これまでに実在する様々な文章情報検索システムですでに実現している方向であるが、特定のテーマに関する過去の記事の全文あるいは抄録の把握を行うものである。この場合、一般には自分の把握したいテーマに関するキーワードをシステムに入力して検索と絞り込みを行い、システムから抄録あるいは全文のリストを引き出すことになる。システムそのものでは、記事内容の解析は行わず、リストを出力後ユーザが内容を解説して種々の判断を下すことになる。

第2の利用の方向は、記事情報データベース全体を集合体として捉え、社会現象を観測するための有力な材料として取り扱おうという方向である。データベースを構成している個々の記事情報は、主観的で偏りがあり分散が激しいと考えられるが、記事全体を集合体として捉えることにより平均化された性質を取り出すことができ、これをベースとして社会現象の様々な法則性を導き出せる可能性がある。特に、記事情報が持っている内容を数量化して抽出した場合には、数の言語を用いて科学と

して現象を観測，分析していくことが可能である。この場合も，システムは人間が判断しやすい形に情報を変換するのが主たる役割で，最終的判断は人間が行う。

第3の利用の方向は，記事情報の中に豊富に散在する数値情報を抽出して利用しようという方向である。現在，各種統計数値情報は，いろいろな機関からの月報，年報といった形で公表されているが，これらから得られる数値情報の内容は限られているといえることができる。これに対して記事情報の中では，各種数値情報がうたかたのごとく現われては埋もれていく。統計数値情報とは異なった性質のものも多い。これらの数値情報を記事情報から抽出し，データベースに蓄積して利用することはできないであろうか。

この他にも，記事情報の利用法は種々考えられるが，このような3つの利用方向を目指して，記事情報総合利用システムが構築されれば，それは大いに役立つものとなるであろう。

さて，このような記事情報総合利用システムを考える時，ソース・データである新聞・雑誌の記事情報を大量入力する可能性がどこにあるかということが最も大きな問題である。国内記事情報に関しては，1.3で述べたように，現在朝日新聞，日本経済新聞といった新聞が編集の電算写植化を開始しており，副産物として記事全文の磁気テープなどを廉価に作成できる体制に入ったといえることができる。また，オンライン手書き入力，音声タイプライター，漢字光学読み取り装置といった開発段階の入力機器の中では，パターン認識による漢字光学読み取り装置が大量入力のための本命機器として近い将来実用化が期待される。海外記事情報については，例えばニューヨーク・タイムズの記事全文データベースがオープンされ，衛星中継でコンピュータ・リーダブルな情報を受信することが可能である。また記事情報に関するいくつかの磁気テープも入手可能であろう。従って，ある程度大量入力可能な文章情報とし



て、記事情報をソース・データとしたデータバンク・システムを考えていくことは、無理がないであろう。

## (2) 記事情報総合利用システムの理想像

ここで、システム内の各機能が実用レベルにあるか、開発レベルにあるかといった点は無視して、(1)で述べた記事情報の利用方向を考慮した記事情報総合利用システム像を述べてみよう。図1-18に具体的なシステム像を図示する。

必要なデータベースは、第1の利用方向を考えた場合、記事情報全文データベースと記事情報抄録データベースである。全文データベースは、入力情報の全文を登録したものである。抄録データベース作成のためには、できれば自動抄録が望ましい。自動抄録の一つの可能性については後述する。この2つのデータベースの利用形態は、まず記事抄録データベースを検索して抄録をリストアップし、抄録の内容を判断して全文を要求するという形を取るようになるであろう。

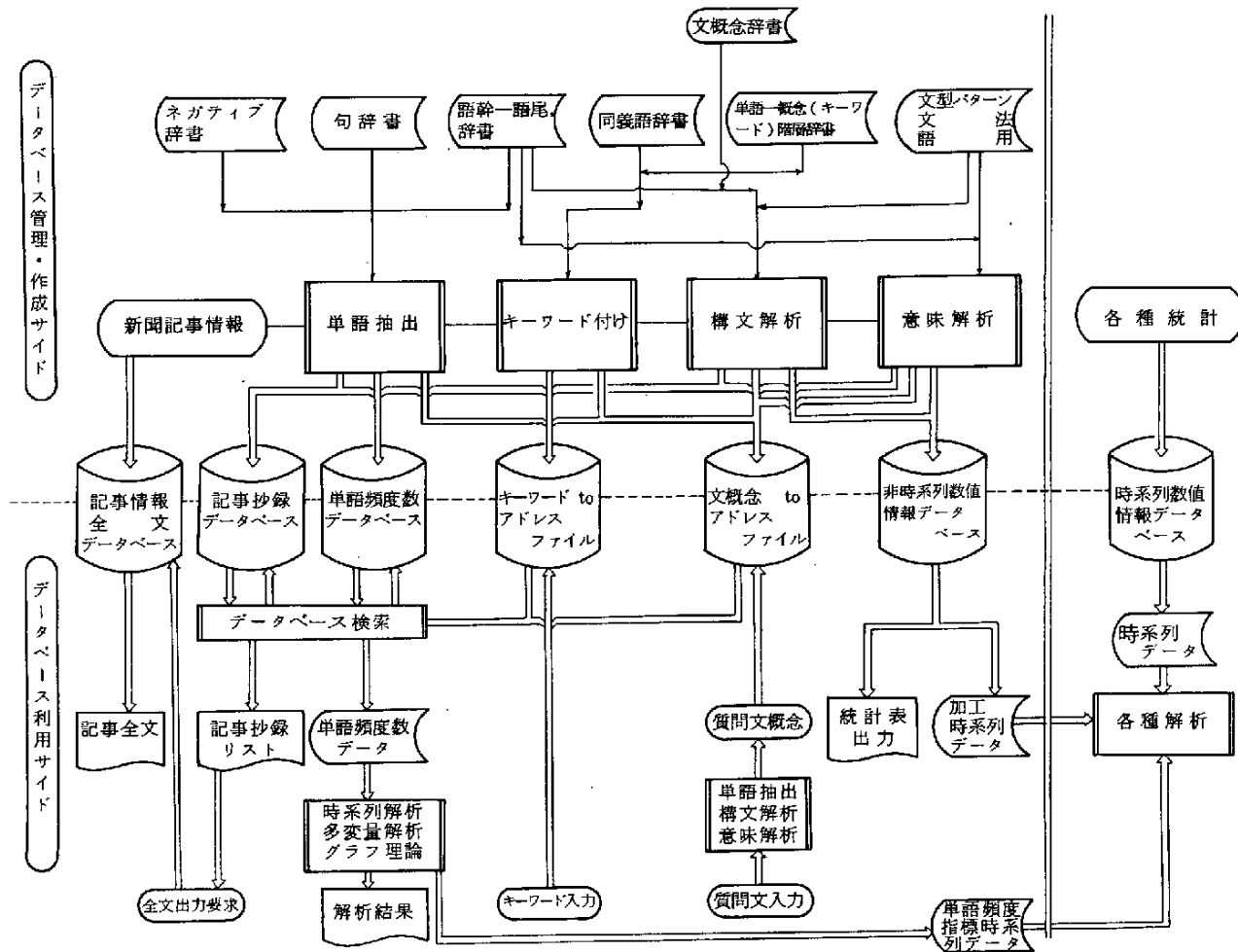


図 1-18 記事情報総合利用システム像

第2の利用方向を考えた場合、単語頻度数データベースが有力な数量化情報源の一つとなるであろう。この場合には、単語の出現頻度数が社会現象の性格を表わしていると仮定することになる。データベース作成サイドでは、記事情報の自然文からシステムで単語抽出を行い、同じ単語の頻度をシステムで積算して単語出現頻度数のデータを作り出す。単語抽出にあたっては、主観的なバイアスがかからないようにしたいので、完全自動単語抽出が望ましい。データベース利用サイドでは、単語頻度数データベースに対する検索を行って単語頻度をファイルへ出力し、各種解析プログラムを通して分析の結果を得ることになる。

第3の利用方向を考えた場合、非時系列数値情報データベースが必要である。データベース作成サイドでは、非時系列数値情報データベースを作成する場合、記事情報の自然文からシステムで数値の発見を行い、発見した数値の前後の文に対する単語抽出、構文解析、意味解析といった作業でキーワードの抽出を行い、キーワードと数値の双方をデータベースに登録する。データベース利用者サイドでは、このようなキーワードをいくつか指定して非時系列数値情報データベースから必要な数値を検索し、行と列にこのキーワードを設定することで二次元のマトリックスの統計表を出力することが可能となる。

次に、各データベースの検索適合率あるいは検索再現率をあげるためには、システムはできるだけ良い検索の仕組みを持たなければならない。記事情報総合利用システムにおいて考慮している2つの検索方法について言及してみよう。

第1の検索方法は、きわめてオーソドックスなキーワード付けによる方法である。キーワード付けのソース・データとしては、記事情報の自然文から自動抽出した単語を利用することができる。ただし、抽出した単語をそのままキーワードにしたのでは、同義語等の問題があり検索の密度が淡くなると考えられる。従って、単語とキーワードの関係を整理

して体系付けることが必要である。

このような単語とキーワードの整理を考えていく場合、いくつかの問題点を上げることができる。一つは、上に述べたように同義語の問題で、同義語辞書を作成することが必要である。最下層のキーワードをどの次元に求めるかの判断は大きな問題であるが、一応同義語の代表名をキーワードの最下層の一つとして選ぶことができるであろう。

次の問題は概念化したキーワードの問題である。世界エネルギー情報のキーワード付けに用いている日本経済新聞社のキーワード分類(表1-9)を例として話を進めよう。

表1-9 日本経済新聞社によるキーワードの分類

|           |         |
|-----------|---------|
| 会社 キーワード  | 固 有 名 詞 |
| 団 体 キーワード | "       |
| 人 名 キーワード | "       |
| 地 域 キーワード | "       |
| 品 目 キーワード | 普 通 名 詞 |
| 業 界 キーワード | "       |
| 項 目 キーワード | "       |

会社、団体、人名、地域といったキーワードは、固有名詞であるから文の中から単語自動抽出ができて、同義語辞書が整えられていれば、ほぼ問題はないと考えられる。品目キーワードは、ガソリン、石油製品といった物を表わすキーワードであるから、例外が生じる可能性はあるが、同義語辞書でほぼ対応できるのではないかと考えられる。これに対して、業界キーワードと社会のアクティビティを示す項目キーワードは単純な同意語の対応関係では処理できない。例えば、「エネルギー政策」(項目キーワード)とか「石油化学業界」(業界キーワード)を見た場

合、単語抽出で単純に抜き出せるケースと文あるいは文章全体でこの概念を現わしているケースとがありうる。後者の場合、抽出された単語のインフォメーションだけで、キーワードに結び付けるのはきわめて難しい。

次の問題は、検索者が指定したキーワードを包含する概念を要求する問題である。ガソリン、灯油、軽油、重油を総称する言葉として石油製品があるが、ユーザが「石油製品」で検索する場合、石油製品という単語の出現している記事を要求するケースと石油製品という単語は出現していないがガソリン、灯油、軽油、重油といった単語の出現している記事も含めて要求するケースの両方が考えられる。

第4番目の問題は、名詞をキーワードとして登録していくことは比較的容易であるが、単語抽出の操作で取り出される動詞、形容詞をキーワードにいかに関結び付けるかという問題である。以上、単語とキーワードの整理を行うにあたって障害となる4種の問題から、少なくとも同義語辞書とキーワードの階層辞書(シソーラス)は必要であると結論付けられよう。検索の場合にも、検索者が自由に自分の思いつく単語(名詞)を入れ、必要な同義語辞書、シソーラスを参照して自分の検索したいキーワードに到達する機能が必要である。

第2の検索方法は、文あるいは文章全体から概念を示す数量を作り出し、それをキーワードに対応するものとして利用する方法である。この実例としては、G. Saltonによる「SMART情報検索システム」がある。このシステムにおける基本的な考え方の一つは、情報検索システムにおける検索方法は、いずれはキーワード入力でなく、自然文による質問となるべきであるという考え方である。これは、基本的な考え方として妥当である。

「SMARTシステム」では、このような考え方を実現するため、入力データ・テキスト中から単語抽出を行い、あらかじめ登録してある単

語辞書から単語番号とウェイトの情報を引き出し、これらの数値情報に基づいて概念ベクトルを作成している。この概念ベクトルが一種のキーワードの働きをしているわけである。検索の時は、自然文による質問文が入力され、入力テキスト処理ルーチンと全く同じステップを踏み、質問文の概念ベクトルが作成される。この質問文の概念ベクトルと入力テキストの概念ベクトルの適合性が判別され、必要な入力テキストが検索されてくるという仕組みである。このシステムの場合、欠陥となったのは入力テキスト全体の概念ベクトルと短い質問文の概念ベクトルでは本質的に違うということであった。この批判に対して Salton が与えた解答は、検索したい内容に合致する文献を1つか2つ入力すれば、このような方法を検索に使えるというものである。

我々がキーワード付けを行う場合、文あるいは文章全体の意味を考慮して与えていくことを考えると、文全体あるいは文章全体から数量情報の形でキーワードに該当する文概念を与えようという方法は理にかなっているといえる。その場合、単純な頻度数によるウェイトだけでなく、構文解析や意味解析から得られる情報も含めることが必要であろう。このような文概念の情報をシステム化することは、きわめて難しいと考えられるが、記事情報総合利用システムでは、これら2つの検索方法の観点から、キーワード to アドレス・ファイル、文概念 to アドレス・ファイルをシステムの中に取り入れている。

次に、データベースを効率的に作成し、有効に利用するためにシステム内部に設置しなければならない各種辞書について述べる。単語抽出の機能には、抽出する単語をリストアップした単語辞書の他に、抽出不要語を規定したネガティブ辞書、複合名詞、慣用句を規定した句辞書、動詞、形容詞を基本形に導くための語幹一語尾辞書などが必要である。キーワード付けの機能には、キーワード付けの検索方法の所でふれたように、少なくとも同義語辞書、キーワードの階層辞書が必要である。より高

度な機能として構文解析，意味解析といった機能をデータベースの作成あるいはデータベースの利用に取り入れていくとすれば，文形パターン，文法，語用に関する情報をまとめた辞書が必要になると考えられる。

ここで，記事情報抄録の自動作成に関する一つの可能性について述べておく。もし単語抽出，構文解析，意味解析といった機能を用いて，すでに述べたような文概念を作り出すことができるとすれば，自動抄録にこの文概念を適用することが考えられる。というのは，記事情報は大体において見出しを持っているので，見出しの文概念を作成し，これと適合性の高い文をいくつか本文から抜き出してくるということで抄録を作成できる可能性があるからである。

さて，これまでは日本文記事情報と外国文記事情報を特に区別しないで述べてきたが，記事情報総合利用システムで考えている両者の取り扱いについて触れておこう。単語抽出，構文解析，意味解析などのシステム機能に関しては，日本文と外国文で別々のルーチンが必要になると考えられる。記事情報全文データベースには，外国文の記事情報はそのまま登録するということがよい。単語頻度数データベースに関しては，日本文記事情報と外国文記事情報とを併せて全体で頻度解析をしたいケースと別々の解析しか意味のないケースとで事情が異なってくる。前者の場合には，外国文の単語と日本文の単語を対応付けた辞書が必要である。この場合，単語が日本文と外国文で1：1に対応するとは限らないので，文あるいは文章全体を判断して複数の対応の中から1つを選ぶ機構が必要である。後者の場合は，外国文の単語の頻度数をそのまま登録して分析するということが対処できる。検索のためのキーワード付けはすべて日本語で行うことが望ましい。外国文の単語と日本文の単語の対応辞書ができていれば，日本文の単語からキーワードに結び付けることができるであろう。そうでない場合は，同義語辞書の中に外国文の単語を登

録すればよい。非時系列数値情報データベースに関しても、やはり日本語のキーを付して数値を登録することが望ましい。

次に、記事情報抄録データベースであるが、これもできれば日本語の抄録を付けることが望ましい。最近、機械翻訳の研究活動が相当活発になっているが、記事情報総合利用システムで考えている各ルーチンが自動的に機能するということは、機械翻訳に必要なツールがかなりの部分そろそろことを意味している。従って、外国文記事情報の抄録には機械翻訳の機能を利用することが考えられる。外国文記事情報の全文に対しても、記事全文出力の際のファクションの1つとして機械翻訳を適用するのが妥当と考えられる。

以上が、各機能の段階が実用レベルにあるか、開発レベルにあるかを無視して描いた記事情報総合利用システムの全体像である。

## (2) 開発のタイム・スケジュールに関する見通し

上述のような記事情報総合利用システムをどのように開発していくかということであるが、以下に示すような3ステップぐらいの段階を取ることが必要と考えられる。

第1ステップでは、記事情報を定量化して高度利用をはかろうとする当初の目的に従って、単語抽出システム及び頻度数解析システムに力点を置くのが妥当と考えられる。単語抽出に関しては、1.4で述べているように、漢字列抽出と辞書マッチングの2つの方法を用いることによりかなりの線まで実現することが可能である。この他に語尾、品詞の判定及び語用の判定などの機能を加えると、さほど遠くない時期に完全自動化ができる段階に達しているとみることができよう。

記事情報の定量化に関しては、単語抽出の他に内容分析の手法を利用した種々の方法が考えられる。日本でも、政治学あるいは言語学の分野で手作業ではあるが内容分析の手法を利用した各種の研究が行われている。単語抽出の他に、これらの研究で行われている各種手法



をアルゴリズム化してシステムに取り入れていくことが必要である。

頻度数解析システムに関しては，表 1-10 に示す各種多変量解析理論の他，多次元尺度法，パターン分類法，ペーズの決定理論 2 部グラフ理論 (1.2 (2))，有向グラフ理論，行動科学で用いられている諸方法を適用することが考えられる。これらの方法は，いずれもすでにアルゴリズムが確立しており，システム化自体には大きな問題はないと考えられる。例えば，いくつかの単語の出現頻度を年別

表 1-10 多種多変量解析

| パターン | 目 的                                         | 使 用 す る 分 析     |                      |                   |
|------|---------------------------------------------|-----------------|----------------------|-------------------|
|      |                                             | 量のデータのみの場合      | 質に関するデータまたは量のデータも含む時 | 質に関するデータのみの場合     |
| 1 型  | 予測式 (関係式) の発見<br>量の推定                       | 重回帰分析<br>正準相関分析 | 数量化分析Ⅰ類              |                   |
| 2 型  | 標本の分類<br>質の推定                               | 判別分析            | 数量化分析Ⅱ類              | 分割表の分析<br>クラスター分析 |
| 3 型  | 多変量の統合整理<br>(減らす)<br>変量の分類，代表変量の<br>発見 (選定) | 主成分分析<br>因子分析   | 数量化分析Ⅲ類，<br>Ⅳ類       | 潜在構造分析            |
| 4 型  | 分析の検定<br>(test による justify)                 | 分散分析            |                      |                   |

に 1 月から 12 月まで並べ，この数値情報から相関マトリックスを作り，主成分分析を行うことによって，1 つの国のある年の時系列変化をみたり，いくつかの国について国別に特徴を比較したりといった方法が考えられる。

全単語を並べて解析を行うことはもちろん可能であるが，1.6 でも再々触れているように，出力結果が膨大となりユーザは解説の意欲を失ってしまうであろう。この問題を避けるためには，エネルギー問題，カンントリー・リスク問題とそれぞれのテーマ別に有効な解析方法をパターン化し，これをシステム内部に組み込んでしまうこと

が必要であると考えられる。最近では、パーソナル・コンピュータの発達  
が著しく、現在の普通の大型コンピュータが有している能力は近い将来  
パーソナル・コンピュータの上にのってしまふであろう。こうした点を  
考えると、ドライで膨大な一次情報のデータベースを大型コンピュータ  
で管理し、大型コンピュータから検索により必要な情報をパーソナル・  
コンピュータに移動し、パーソナル・コンピュータ上のユーザ・オリ  
エンテッドなソフトウェアで解析を行うようなシステムの開発がこれ  
からは最も望まれる方向であるかもしれない。

単語抽出の完全自動化と個別の問題に有効な解析パターンの開発にあ  
る程度時間がかかると考えられるので、第1ステップの実行に少なくと  
も2～3年の時間が必要であろう。

第2ステップとしては、記事情報総合利用システムで用いようと思  
っているシステム機能の大半は現在開発段階にあり、完全自動化に持ち込  
むことはすぐには困難なので、重要な判断は人間が下し、大量データ処  
理や論理的なチェックはコンピュータで行うようなシステムをまず開  
発することである。この場合、開発中のシステム機能を十分考慮して、  
開発が終われば人間の操作からその部分を開発したコンピュータ操作に  
容易に直換できるようにシステム構成を組立てておかねばならない。

1.3でも述べているように、現在のデータ入力の大半は、コー  
ディング・シートに書かれたデータに一定のコントロール情報を付加し  
てパンチし、種々のチェックを加えて作成したマスター・テープからデ  
ータベースに展開するコントロール入力である。また現状では、キーワ  
ード付け、構文解析、意味解析、抄録作成、非時系列数値情報の抽出と  
いった機能は、人間が重要なポイントである程度判断を下してやらない  
と無理である。そこで図1-19に示すように、自然文をソース  
としたフリー入力を、辞書によるコンピュータ・チェックや人力による  
各種の判断を通して画面編集でデータベースに展開していくシステムの

開発が必要である。

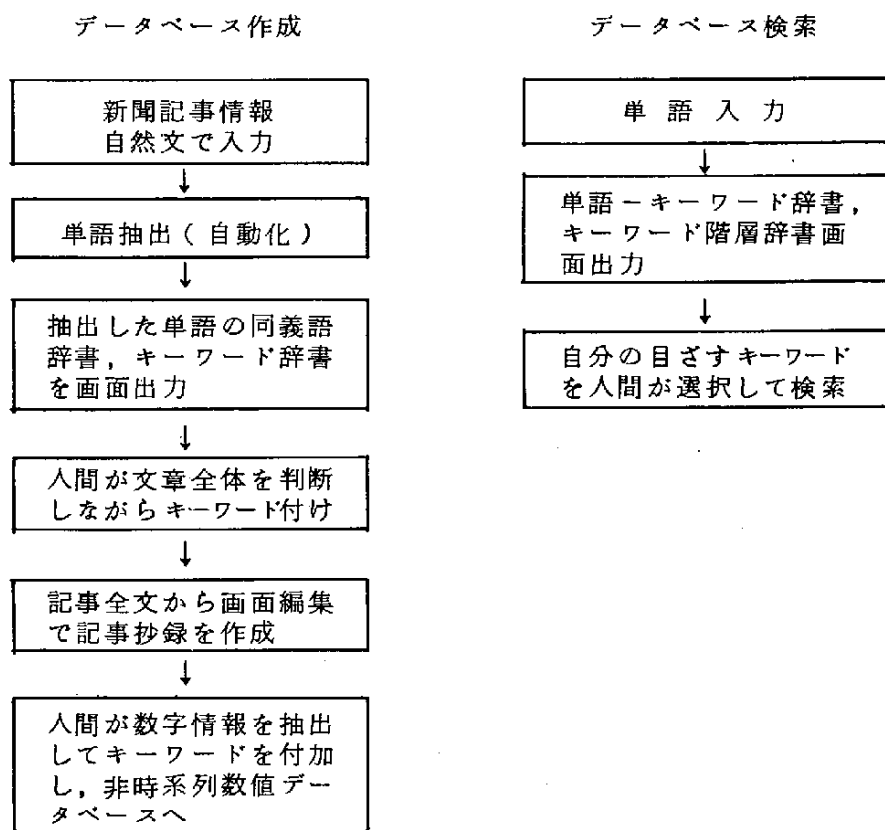


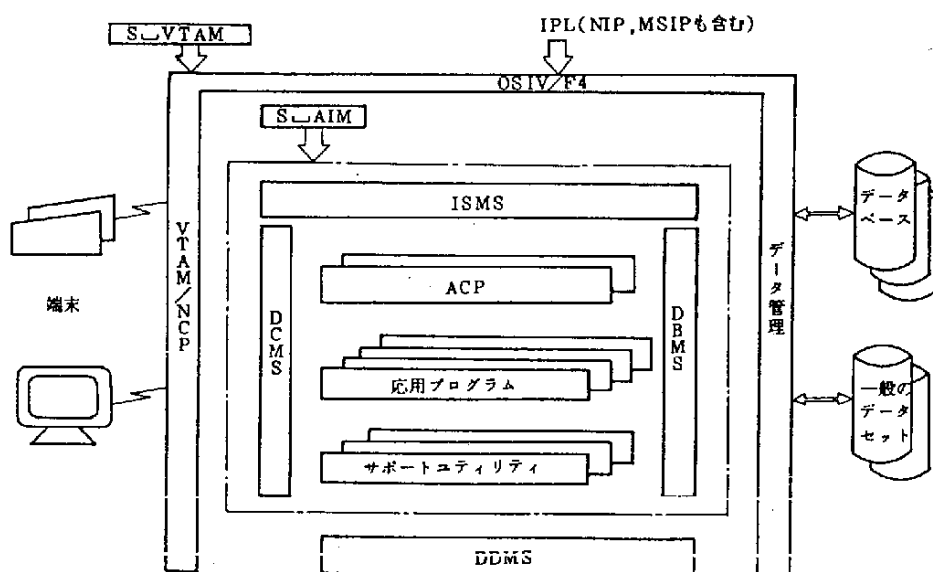
図1-19 人間の判断にウェイトを置いたシステム

1.5でも述べたように、シソーラス辞書の作成1つを取っても辞書作成の作業はきわめて大変なものである。従って、第2ステップとして、人間の判断にウェイトを置いたシステム開発を行うには、3～5年の年月が必要となるであろう。

第3ステップは、第2ステップで人力が介入している部分をできるだけ自動ルーチンに置き換え、人力の介入を最小限にとどめるシステム開発を行う段階である。この最終段階では自然文の質問による検索利用

の実現を目指すことになると考えられるが、これには10年以上の歳月を必要とするであろう。

最後に、富士通のAIMデータベース・マネージメント・システムについて触れておく。AIMシステムは、図1-20に示すようにDDMS（辞書、ディレクトリー管理プログラム）、ISMS（機能、選択、環境設定のプログラム）、DBMS（データベース管理プログラム）、DSMS（データの連結関係マネージメント）、サポート・ユーティリティ、ACP（応用プログラムのコントロール）、応用プログラムの7システムから構成されている。



(注) 一内は、オンラインデータベースシステムを運転する場合に、有機的に連結されて動作するAIMのコンポーネントを示す。

DDMS: Dictionary and Directory Management Subsystem

ACP : Application Control Program

ISMS: Integrity and Schedule Management Subsystem

IPL : Initial Program Loader

DBMS: Data Base Management Subsystem

NIP : Nucleus Initialization Program

DCMS: Data Communication Management Subsystem

MSIP: Master Scheduler Initialization Program

図1-20 AIMシステムの構成

AIMシステムを記事情報総合利用システムに適用することを考えた場合、各種辞書あるいはデータベースの管理にはAIMシステムの機能を使用できるが、単語抽出、キーワード付け、構文解析、意味解析といった機能に関しては独自の応用プログラムを開発しなければならない。従って、応用プログラムでAIMの受信可能なデータ形式に変換し、AIM管理下のデータベースに持ち込むようなシステム体制を作ることが必要である。

## 2. 自動翻訳システムの研究

### 2.1 研究目的と活動概要

近年、情報の量、質、速さには注目せざるを得ない状況にある。特に新聞のように、製作過程の変化や流通手段の変化は、国内はもとより海外の情報も翌朝にはオンライン端末を介して入手できるようになってきている。さらに、国際社会への対応をせまられる現在、各国、各産業の状況と迅速、的確に対応することが急務となっている。一方、それらの処理には、コンピュータによる処理が拡大されつつあり、その技術も高度化している。特に自動翻訳の要望は強く、その研究、実験には長い歴史があり、ここ数年間の研究においては実用化の兆しも見えている状況にある。

本研究会では、それらをふまえ、ハードウェア、ソフトウェア技術の動向を調査するとともに、処理対象の分野、言語を限定した実用システムの開発を目的として、翻訳トランスレータの分析、評価、実験等を行った。なお、翻訳トランスレータの分析、評価、実験は、工業技術院情報計算センター（RIPS）で稼働中のシステムを使用した。研究会での検討内容は以下のとおりである。

- 第1回 ① 研究会発足の経緯及び今年度計画の検討  
② RIPSの翻訳システムの見学
- 第2回 ① 自動翻訳システムの実用可能性について調査、検討  
② RIPS翻訳システムのトランスレータの分析、評価  
③ 翻訳システム用辞書についての検討
- 第3回 ① 自動翻訳技術の現状と将来について

講師 京都大学工学部教授

長 尾 真 氏

第4回 ① 構文解析技術について

② 翻訳システム開発時の留意事項について

第5回 ① 入出力インタフェースについて

② LC-MARC (アメリカ議会図書館入庫書誌データ)  
の実験について

## 2.2 翻訳システム技術の現状と将来

### 2.2.1 歴史的背景

翻訳の機械化の提案は、既に1930年代前半に行われている。1940年後半からは、コンピュータによる機械翻訳が「人類の夢の実現」という期待をもって始められた。

1950年代後半から1960年代前半にかけて、アメリカでは、スプートニクス・ショックによりソ連の科学技術文献の大量の翻訳に対する要求がおこり、機械翻訳の研究が盛んに援助された。しかし、言語理論、コンピュータ技術が、まだ機械翻訳を実現するために十分でなく、良い結果が得られなかった。そのため、1966年のALPACレポートにより「現時点では、役に立つ機械翻訳はできそうもないし、また、ニーズもない。」と評価され、アメリカのみならず、世界的にも機械翻訳の研究が衰退した。

しかし、ヨーロッパやカナダでは、現実問題として翻訳の要求が強く、機械翻訳の研究が継続されていた。また、アメリカでも空軍や原子力委員会で実際に機械翻訳システムが使用されていた。そして、ヨーロッパやカナダでは、単なる機械翻訳のみならず、翻訳者をコンピュータによって支援するための用語データバンクなども開発されていた。

このように、実際のニーズのあるところでは実用化の努力が重ねられてきた。近年、各国間の情報流通量の増大と共に翻訳のニーズが増加してきている。また、言語理論の発展やコンピュータ技術の向上によって、機械翻訳に必要な技術も整いつつあり、再び、機械翻訳の研究が盛んになってきた。(図2-1, 図2-2参照)

### 2.2.2 現 状

現時点では、高品質の全自動機械翻訳システムを作成するのは不可能に



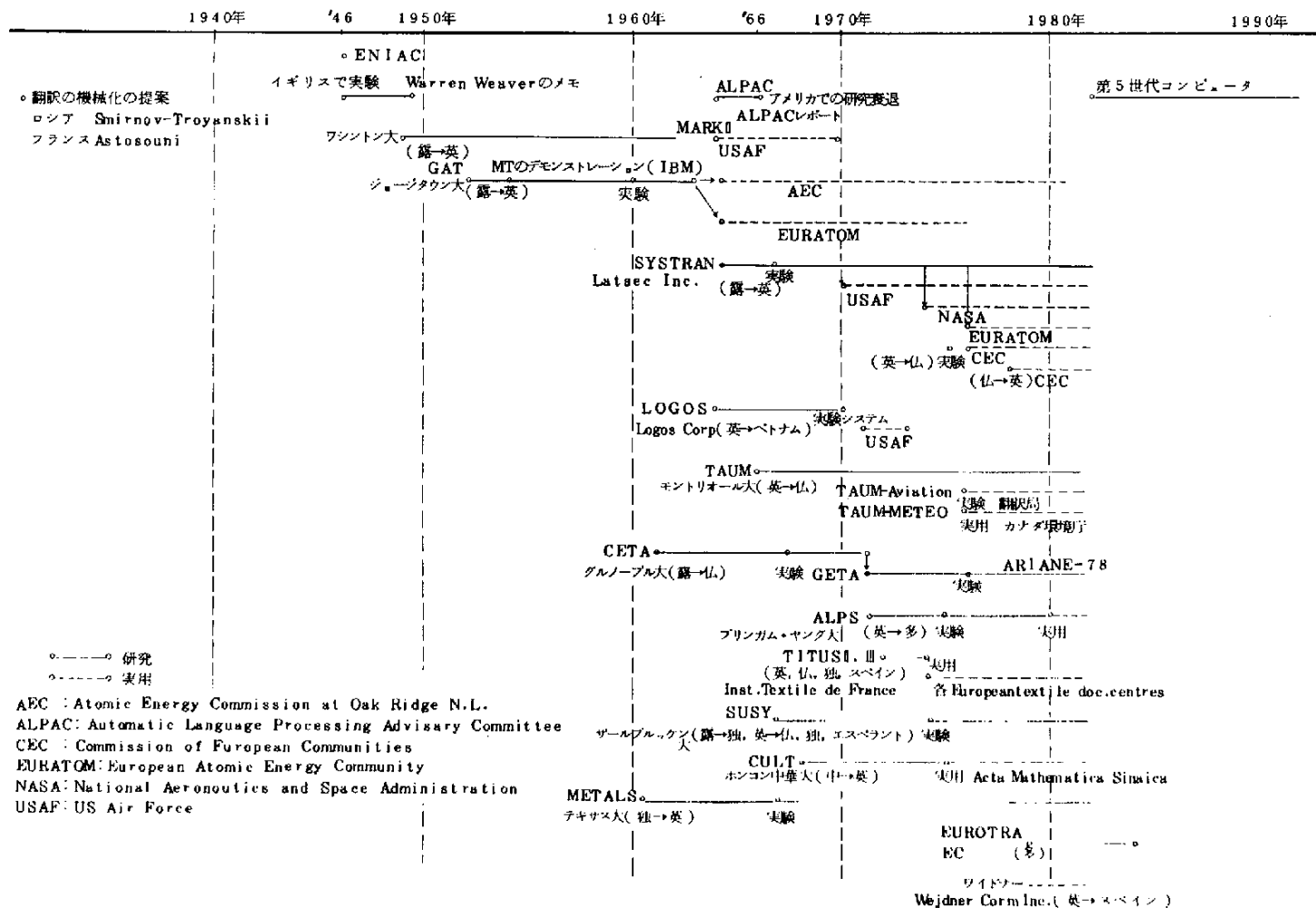


図 2-1 機械翻訳システムの発展 (ソビエト, 東欧, 日本除く)

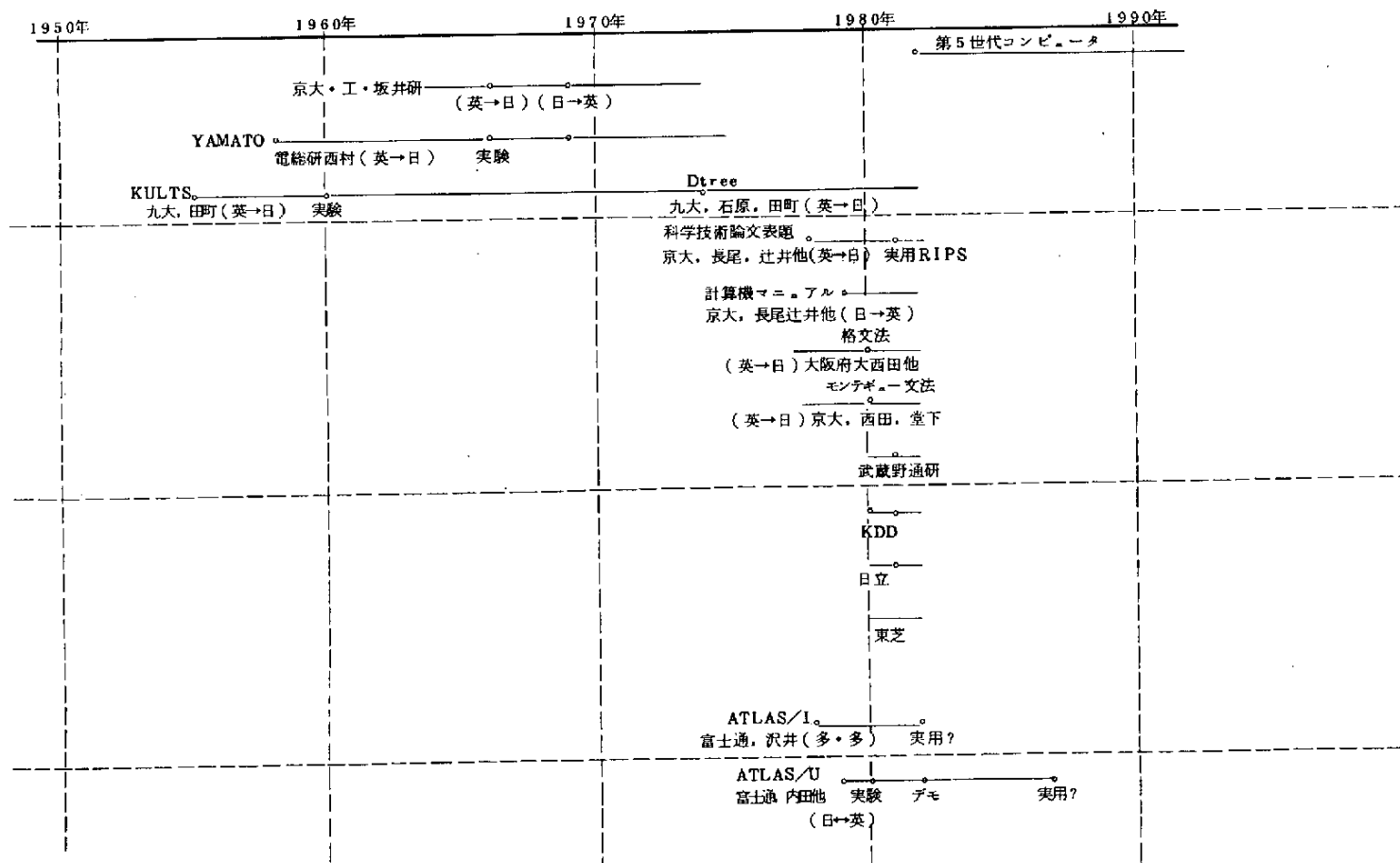


図 2 - 2 日本での機械翻訳システムの発展

近い。そのため、人間とコンピュータとの共同作業により翻訳の効率を上げるシステムが考えられており、次の2つの概念が提案されている。

一つは、人間（翻訳者）を中心にして、コンピュータは人間の翻訳作業を支援するために使うというものである。これは、Machine aids for translator、又は、Machine Aided Human Translation (MAHT) と呼ばれ、ヨーロッパやカナダでは、10年以上前から用語データバンクとそれをサポートするシステムとして開発が行われているものである。

もう一つは、機械翻訳システムは不完全なものという認識から、機械翻訳のできないことを人間が行うというものである。これは、Human Aided Machine Translation (HAMT) と呼ばれていて、前編集、機械翻訳、後編集と機械翻訳の前後に人間が介在するものと、機械翻訳の過程で人間と会話しながら翻訳を行うシステムが考えられる。

これら2つの考えは、共に人間の作業を少なくする方向で、より良いマンマシン・インターフェースの開発が望まれる。

ECの機械翻訳のコスト評価では、人手で行った場合を100%として、実用化されているSYSTRANでは70%であり、現在、開発中のEUROTRAでは40%を目指している。また、「実用的機械翻訳システムは、制限された適用分野では、妥当な受け入れ可能な翻訳を供給できるようになった。」という見方も現われている。

このように、現在機械翻訳システムの実用化に関して、ある程度の期待をもっても良いように思われる。

## 223 機械翻訳システムの構成

### (1) 機械翻訳の過程

機械翻訳システムを機械翻訳の過程からみると図2-3のようになる。言語データの入力、前編集を経て、形態素解析から形態素生成までの機械翻訳の中心過程を通して、後編集、出力までであるが、意味解

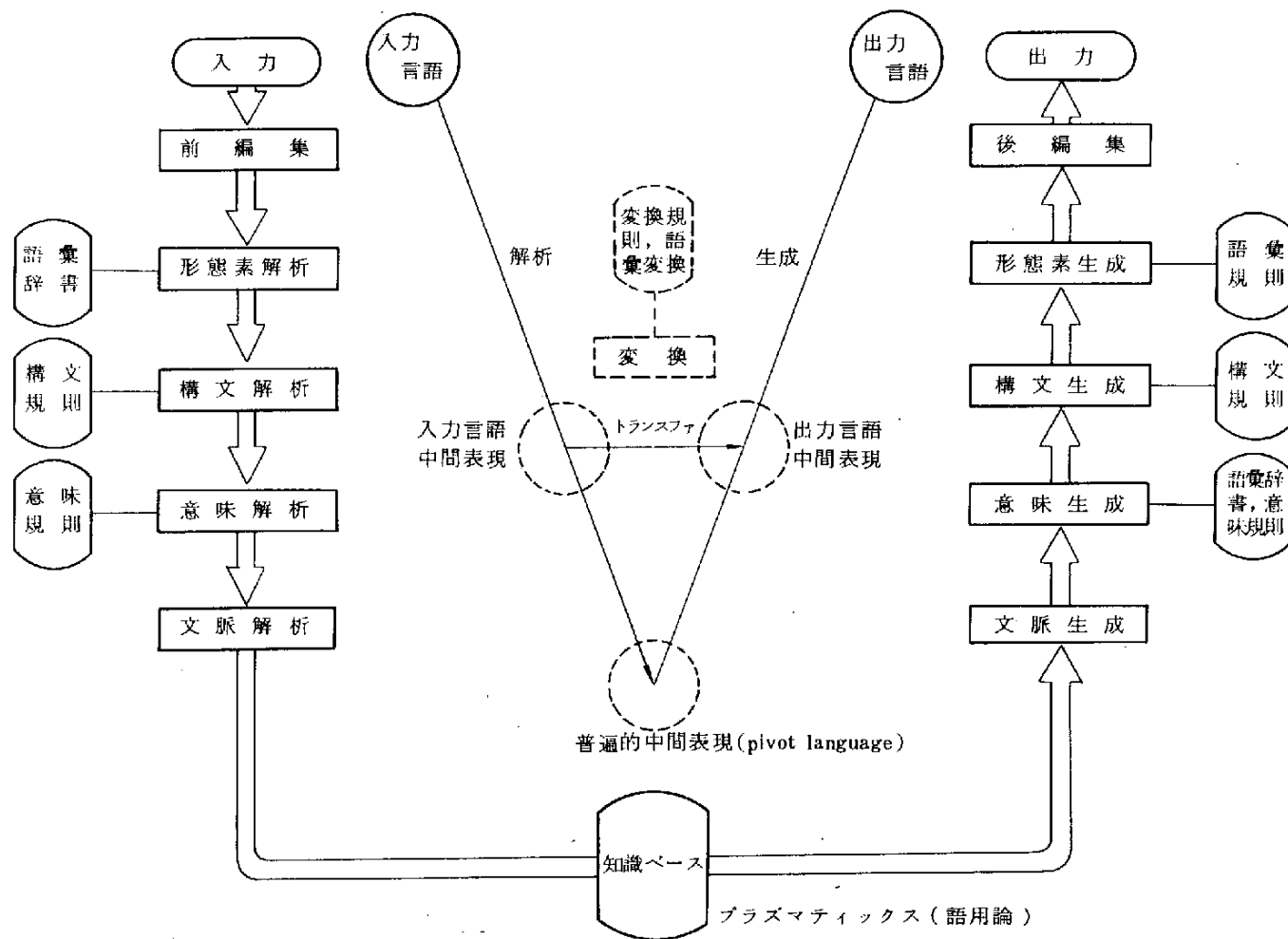


図 2-3 機械翻訳過程

析から意味生成までの処理には、まだ未解決な問題が多い。

図に従って各処理内容を以下に簡単に述べる。

入力と前編集では、言語データの入力や文の境界、大文字の指示などの文の形式的処理を行う。

形態素解析では、辞書と関連して、語形変化、熟語、複合語などの処理や、日本語では、語分割処理を行う。

構文解析では、品詞を元にして文の構造の解析を、句構造文法、変形文法、依存文法などに従って行う。

意味解析、文脈解析、プラグマティックス(語用論)解析では、構文解析までに解決されなかった「曖昧さ」の処理を行う。「曖昧さ」には、多義語、代名詞参照、省略、句の係り受けなどがあり、文の要素の意味チェックをすれば良いものや、文脈を見なければいけないものや、対象としている世界の知識を利用しなければならないものがある。

トランスファでは、入力言語を解析した結果の構造(入力言語中間表現)から出力言語を生成するために必要な構造(出力言語中間表現)への変換と語彙の変換を行う。これは、各中間表現の設定の仕方(解析、生成のレベル)により処理の複雑さが変化し、入力言語、出力言語に共通の中間表現が設定されれば、語彙変換のみで良くなる。

文脈、意味生成では、出力言語の文脈、意味表現を考慮し、文の順序、語彙の選択、代名詞の使用などを行う。

構文生成では、出力言語の語順に文を生成する。

形態素生成では、語形変化処理などを行う。

後編集、出力では、翻訳結果の修正、文の形式の変換や、ラインブリクタ出力、写植、マイクロフィッシュ出力を行う。

## (2) 機械翻訳システムの分類

機械翻訳システムを分類する場合には、いろいろな立場からその方法が考えられている。代表的な分類方法を挙げると以下のようなものがあ

る。

#### A 機械翻訳の過程による分類

機械翻訳の過程を解析と生成を対象させて考え、形態素（語彙）レベルの翻訳から、構文レベル、意味レベル、文脈レベル、プラグマティックスレベルと、浅い解析から深い解析までその解析レベルによって機械翻訳システムが分類される。これらは、現時点では、コンピュータ技術や利用できるコンピュータ資源の規模、機械翻訳そのものの技術レベル、目的によって選択される。

#### B システム構成からの分類

システム構成の面から、翻訳する対象言語の数、効率、汎用性、辞書の構成を考えて、以下のような分類が行われている。

- ① ダイレクト（direct）：入力言語の解析を行いながら出力言語の生成も行う。辞書は1つにすることができ、入力言語と出力言語情報に文法、意味情報が格納されている。これは、効率を重視した翻訳が可能だが、多言語間の翻訳システムを作成する場合にはシステム作成に時間がかかる。例としては、SYSTRAN、LOGOSがある。
- ② トランスファ（transfer）：入力言語の解析、中間表現とトランスファ、出力言語の生成と明確に分離されており、それぞれに辞書が使われる。これは、多言語間翻訳システムに使われる方法があるが、中間表現のレベルをどの程度にするかによってシステム作成の労力が異なる。例としては、ARIANE-78、TAUM-AVIATIONがある。
- ③ 中間言語（interlingual）：入力言語と出力言語の間に中間言語を設定して、解析、生成を独立に行う。それぞれに1つ以上必要である。この中間言語を言語に依存しない普遍的中間表現にすることが理想である。例としては、METALSやCETAのシステムがある。

現在、実用化されているシステムは、構文レベル＋意味程度の処理を行い、ダイレクト方式が多い。そして、実用化のために研究、開発中のシステムは、多言語間翻訳を目的として、処理のレベルは同程度であるが、トランスファ方式のものが多い。

表 2-1, 2-2 に主な機械翻訳システムを、表 2-3 に用語データベースシステムをまとめる。

## 2.2.4 RIPS の翻訳システム

### (1) 英和翻訳プログラム

現在、RIPS で使用している英和翻訳プログラムは、京都大学工学部の長尾研究室で開発された「科学技術論文表題の英和翻訳プログラム」である。これは、英語のサブセット（表題程度）を日本語のサブセットに翻訳するもので、構文レベル＋意味の制限された解析・生成を行うダイレクト方式の翻訳プログラムである。

図 2-4 に従って簡単に処理を説明する。

#### ① 前処理

- ・文の区切り記号に従って文の抽出を行う。
- ・単語の抽出と数値の切り出しを行う。
- ・「' S 」の処理を行う。

#### ② 辞書引きと熟語処理

- ・単語単位の辞書引きを行い、日本語訳、品詞、熟語、意味カテゴリを得る。
- ・冠詞＋動詞を冠詞＋名詞に変える。
- ・単語ごとに登録されている熟語によるマッチング

#### ③ AND の処理

- ・AND の前後が同一品詞（名詞除く）の場合と、AND の前が形容詞の場合、AND の処理を行う。
- ・AND の前後を前置詞、接続詞、動詞を区切りとしてまとめ、AND

表2-1 主な機械翻訳システム（ソビエト，東欧，日本除く）

| システム名         | 研究開発機関                     | 機械翻訳レベル                   | 対象言語対                                 | 対象分野                                             | ハードウェア                         | ソフトウェア              | 研究開発期間          | 実用化時期            | 使用機関                                                                        | 評価，その他                                                                                                                                 |
|---------------|----------------------------|---------------------------|---------------------------------------|--------------------------------------------------|--------------------------------|---------------------|-----------------|------------------|-----------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| (MARKI)       | ワシントン大                     | 語彙レベル+語順変更<br>direct      | 露→英                                   |                                                  |                                |                     | 1949年～<br>1962年 | 1964年<br>MARK II | USAF<br>(1964～70)                                                           | 200頁標準→7,573のエラー<br>35 語句の削除，26 語句の追加<br>12 語句の誤った選択，13 語句のエラー                                                                         |
| GAT(-SLC)     | ジョージタウン大                   | 構文レベル(初歩的)<br>direct      | 露→英                                   |                                                  | IBM7090<br>IBM370              | アセンブラ               | 1952年<br>～1963年 | 1964年            | AEC('64～)<br>EURATOM('64～<br>'76)                                           | 使用者(58回答)の評価<br>良い又は受容: 9.2%, 文の8.7%<br>正しい，7.6%技術用語正しい，<br>9.6%の正確                                                                    |
| SYSTRAN       | Latsee Inc.                | 構文レベル+意味<br>direct        | 露→英<br>英→仏<br>仏→英                     |                                                  | IBM<br>シーメンス                   | アセンブラ<br>マクロ言語      | 1964年～          | 1970年            | USAF(1970～)<br>NASA(1974～)<br>EURATOM(1976<br>～)<br>CEC(1976～)<br>* (1978～) | CECの各種文献での評価(英→仏)<br>了解度: 4.4～4.7%, 6.6% (評<br>者修正)<br>忠実度: 8.6～9.5%, 語，名詞句<br>6.1～8.5%文法<br>速 度: 300,000～350,000語/<br>H(CPU) 1370/155 |
| LOGOS         | Logos Corp.                | 構文レベル<br>direct           | 英→ベトナム<br>(LOGOS)<br>英→露，仏<br>(LOGOS) |                                                  | IBM370/145<br>(CMS)            | FORTRAN<br>アセンブラ    | 1964年～?         | 1971年            | USAF(1971～73)                                                               | 速 度: 60900～80,000語/H<br>(CPU)<br>250,000語/H(CPU)<br>(1370～168)                                                                         |
| METALS        | テキサス大                      | 構文レベル<br>interlingual     | 独→英                                   | 物理学，電気<br>技術の資料                                  | DEC-2060                       | LISP<br>(Interlisp) | 1961年～          |                  |                                                                             |                                                                                                                                        |
| CETA システム     | ダルトン大<br>CETA              | 構文レベル<br>interlingual     | 露→仏<br>(ロ→仏)                          |                                                  | IBM7044→<br>IBM360/67<br>(CMS) | アセンブラ<br>メタ語        | 1961年<br>～1971年 |                  |                                                                             | 評価(1971年): 42%正しい，61<br>%理解できる                                                                                                         |
| ARIANE-78     | GETA                       | 構文レベル+意味<br>transfer      | 露→仏，英→日，<br>英→仏，英→中，<br>英→マレーシア       |                                                  | IBM4331                        | アセンブラ<br>特殊仕様言語     | 1971年～          | (1986年頃)         |                                                                             |                                                                                                                                        |
| TAUM          | セントリオール大                   | 構文レベル+意味<br>transfer      | 英→仏                                   | 一般                                               | CYBER-74                       | PASCAL              | 1966年～          |                  |                                                                             |                                                                                                                                        |
| TAUM-METEO    |                            | 構文レベル(制限)<br>direct       | 英→仏                                   | 天気予報                                             | *                              | FORTRAN             |                 | 1976年            | カナダ環境庁                                                                      | 80 語句は後編集不要<br>interactive 後編集機能                                                                                                       |
| TAUM-AVIATION |                            | 構文レベル+意味<br>transfer      | 英→仏                                   | 航空機マニアル                                          | CYBER-173                      | PASCAL<br>特殊仕様言語    | 1976年～          |                  | 翻訳局                                                                         |                                                                                                                                        |
| SUSY          | ザールブルゲン大                   | 構文レベル<br>transfer         | 英→独，露→独，独<br>→日，独→仏，独→<br>エスベラント      | 科学技術文献の<br>文(独→英，独<br>→日)<br>メンテナンスマニ<br>アル(英→独) | シーメンス<br>7760                  | FORTRAN<br>アセンブラ    | 1967年～          |                  |                                                                             |                                                                                                                                        |
| TITUS I       | Inst. Textile<br>de France | 構文レベル(制限)<br>interlingual | 英，仏，独，スベ<br>イン語間                      | 繊維関係の文献                                          | IBM360/30                      | アセンブラ               | 1972年～          | 1974年            | 各 European textile<br>doc. centers                                          | 速 度: 12 文(20 語/文) 秒                                                                                                                    |
| CULT          | ボンコン中蔵大                    | 構文レベル<br>direct           | 中→英                                   |                                                  | ICL1904A                       | FORTRAN             | 1968年～          | 1975年            | Acta Mathematica<br>Siniicaの翻訳                                              | interactive 編集機能                                                                                                                       |
| ALPS          | ブリュッセル大                    | 構文レベル<br>interlingual     | 英→多(仏，スペイン<br>独，ポルトガル)                |                                                  | IBM360/65                      | PL/I<br>アセンブラ       | 1971年～          | 1980年            |                                                                             | interactive 編集機能                                                                                                                       |
| CAT System    | Weidner Comm.<br>Inc.      | 構文レベル<br>direct           | 英→スペイン                                |                                                  | PDP-11                         |                     |                 | (1980年)          | BRAVICE Int.<br>Inc.                                                        | interactive 編集機能                                                                                                                       |
| EUROTRA       | EC                         | (構文レベル+α)<br>transfer     | 多                                     | 国際通信(編集あり)<br>し)出版(編集あり)                         | 任意                             |                     | 1979年～          |                  |                                                                             | 目標，了解度99%，忠実度90%                                                                                                                       |



表 2-2 日本の主な機械翻訳システム

| システム名                     | 研究開発機関           | 機械翻訳レベル                             | 対象言語対           | 対象分野     | ハードウェア                         | ソフトウェア           | 研究開発期間                      | 実用化時期  | 使用機関              | 評価, その他                                       |
|---------------------------|------------------|-------------------------------------|-----------------|----------|--------------------------------|------------------|-----------------------------|--------|-------------------|-----------------------------------------------|
|                           | 京大, 工, 坂井研       | 構文レベル<br>direct                     | 英→日             |          | N2200/200                      | アセンブラ            | ?~(1966<br>年)~(1969<br>年)~? |        |                   | 科学・技術文の70%正しい<br>3,600語/H                     |
| YAMATO                    | 電総研              | 構文レベル<br>transfer                   |                 |          | Honeywell 1200,<br>600 F230/50 | COBOL            | 1958年~?                     |        |                   | 20,000語/H                                     |
| KULTS                     | 九大, 田町           | 構文レベル(Direct)<br>interlingual       | 英→日             |          | F230-60, 75<br>(M200)          | FORTRAN<br>アセンブラ | 1955年~                      |        |                   |                                               |
| (科学技術論文表題の<br>英和機械翻訳システム) | 京大, 工, 長尾研       | 構文レベル(制限)<br>+意味<br>direct          | 英→日             | 科学技術論文表題 | M200                           | PL/L             | 1978年~<br>1979年             | 1981年  | 工業技術院<br>情報計算センター | 評価: INSPEC表題93% (詳細<br>修正)<br>速度: 10表題/秒(CPU) |
| (計算機マニアルの<br>日英機械翻訳システム)  |                  | 構文レベル+意味<br>transfer                | 日→英             | 計算機マニアル  | M200                           | LISP             | 1979年~                      |        |                   |                                               |
|                           | 大阪府立大<br>西田      | 意味レベル(格文法)                          | 英→日             |          |                                |                  | ?~(1980<br>年)~              |        |                   |                                               |
|                           | 京大, 工,<br>西田, 室下 | 意味レベル(モンテギ<br>ュー文法)<br>interlingual | 英→日             |          |                                | LISP             | ?~(1980<br>年)~              |        |                   |                                               |
| ATLAS/I                   | 富士通              | 構文レベル+意味<br>(direct)                | 多(日→英, 英→<br>日) |          | Mシリーズ<br>(OS/IV/F4)            | PL/1             | 1978年~                      |        |                   |                                               |
| ATLAS/U                   | 富士通研             | 意味レベル(概念依存<br>構造) transfer          | 日→英             |          | Mシリーズ<br>OSIV/F4)              | アセンブラ            | 1979年~                      | 1987年頃 | 富士通社内             |                                               |

表 2 - 3 用語データバンクシステム

| システム名        | 研究開発機関              | 対象言語                                       | 用語データベース                                                     | 機能                    | 研究開発期間      | 備考                          |
|--------------|---------------------|--------------------------------------------|--------------------------------------------------------------|-----------------------|-------------|-----------------------------|
| EURODICAUTOM | ECの翻訳局              |                                            | 定義とテキスト例<br>約130,000項目                                       | 会話モードでアク<br>セス形態素分析   | 1969年～      | DICAUTOMが前身<br>1963～ 石炭鉄鋼同盟 |
| TERMIUM      | モントリオール大            | 英↔仏                                        | 定義とテキスト例                                                     |                       | 1970年～      |                             |
| LEXIS        | ドイツ連邦言語局            | 英↔独( $\frac{2}{3}$ )                       | 科学・技術分野<br>(航空, 通信, 電<br>子 etc)<br>3,000,000 エント<br>リ(1981年) | 多様な出力: LP,<br>COM, 写植 | 1965年～      | 更新量, 2,000～3,000<br>語/2週    |
| TEAM         | シーメンス               | 英, 独, 仏, 伊,<br>スペイン, ポルト<br>ガル, オランダ,<br>露 | 文法, 分野, 定義<br>情報の信頼性                                         | オンライン・アク<br>セス        | 1970年頃<br>～ |                             |
| NORMATERM    | AFNOR(フラン<br>ス規格協会) | 仏, 英                                       | 規格用語, シソー<br>ラス(ISO)                                         |                       | 1973年～      |                             |
|              |                     |                                            |                                                              |                       |             |                             |

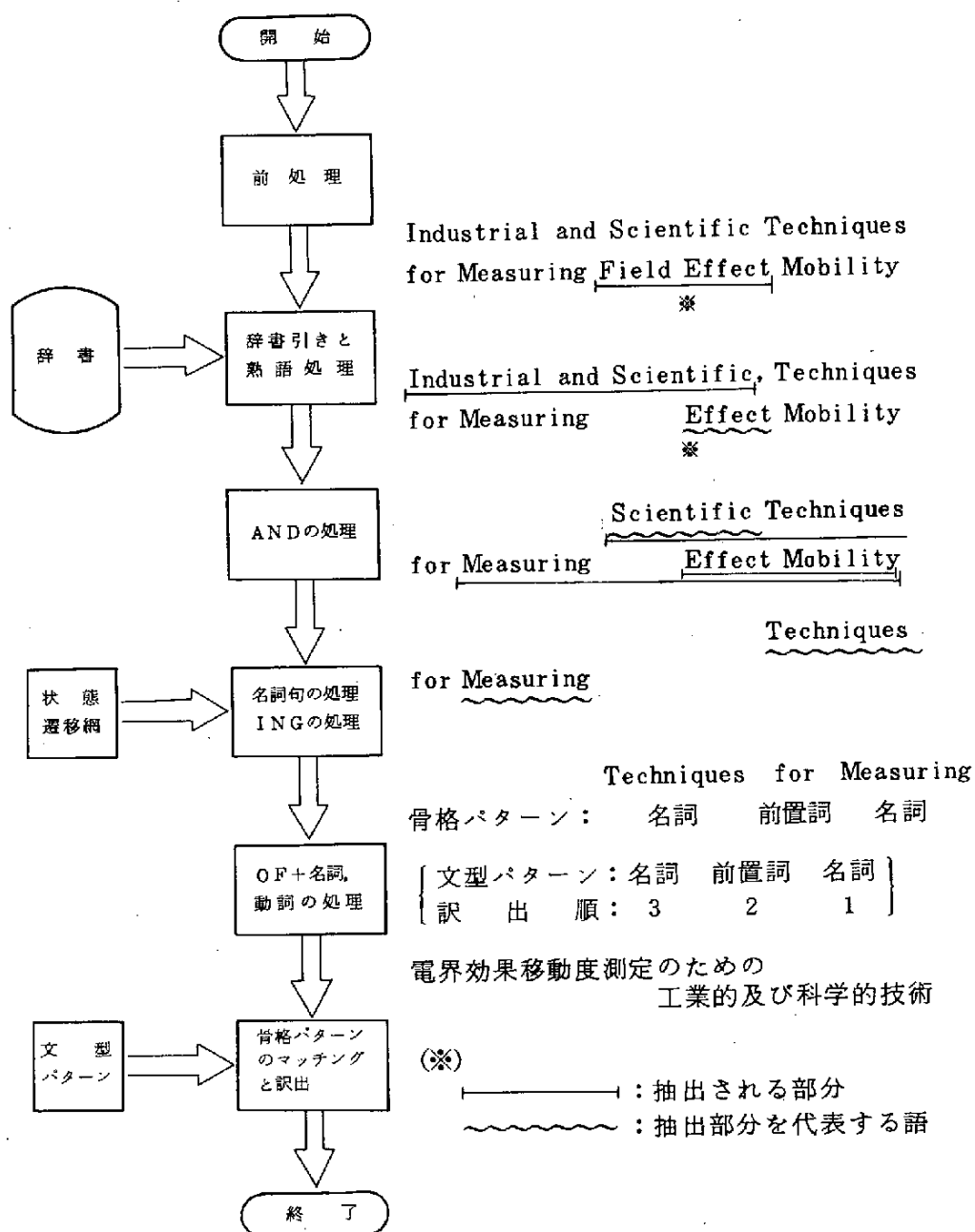


図 2-4 翻訳プログラムの流れと実行例

の前の品詞と同等の品詞が後部に含まれている場合にANDの処理を行う。

④ 名詞句の処理，INGの処理

- 状態遷移網（図2-5）による名詞句の抽出を行う。
- INGの後に前置詞が来る場合はINGが名詞として扱う。また，INGの後に名詞が来る場合は，その名詞がINGの主語になるか目的語になるか意味チェックが行われ，訳出順が決定される。

⑤ OF+名詞，動詞の処理

- 名詞句または名詞がOFの前後に来る場合まとめる。
- TO+原形，過去分詞+前置詞，助動詞+動詞の処理を行う。

⑥ 骨格パターンのマッチング

- 抽出された骨格パターンと文型パターンをマッチングする。
- マッチングした骨格パターンから，文型パターンに登録されている訳出順により日本語文を生成する。

(2) 文献速報システム

RIPSでは，上記英和翻訳プログラムを実際に利用するために，日本語エディタ，日本語情報検索システムと結合し，図2-6のような文献速報システムを作成した。

このシステムは，研究者が個別に収集した論文や，RIPSの共用データベースであるINSPEC，BDB，CASなどの文献データベースの英語の表題を日本語で速報するものである。さらに，将来の翻訳システムのために，用語を収集することも重要な目的となっている。このために，翻訳辞書修正用のサポートプログラムを特に重点的に開発した。図2-7に会話型翻訳辞書修正プログラムの使用例を示す。また，このシステムを構成する各プログラムの機能を表2-4に示す。

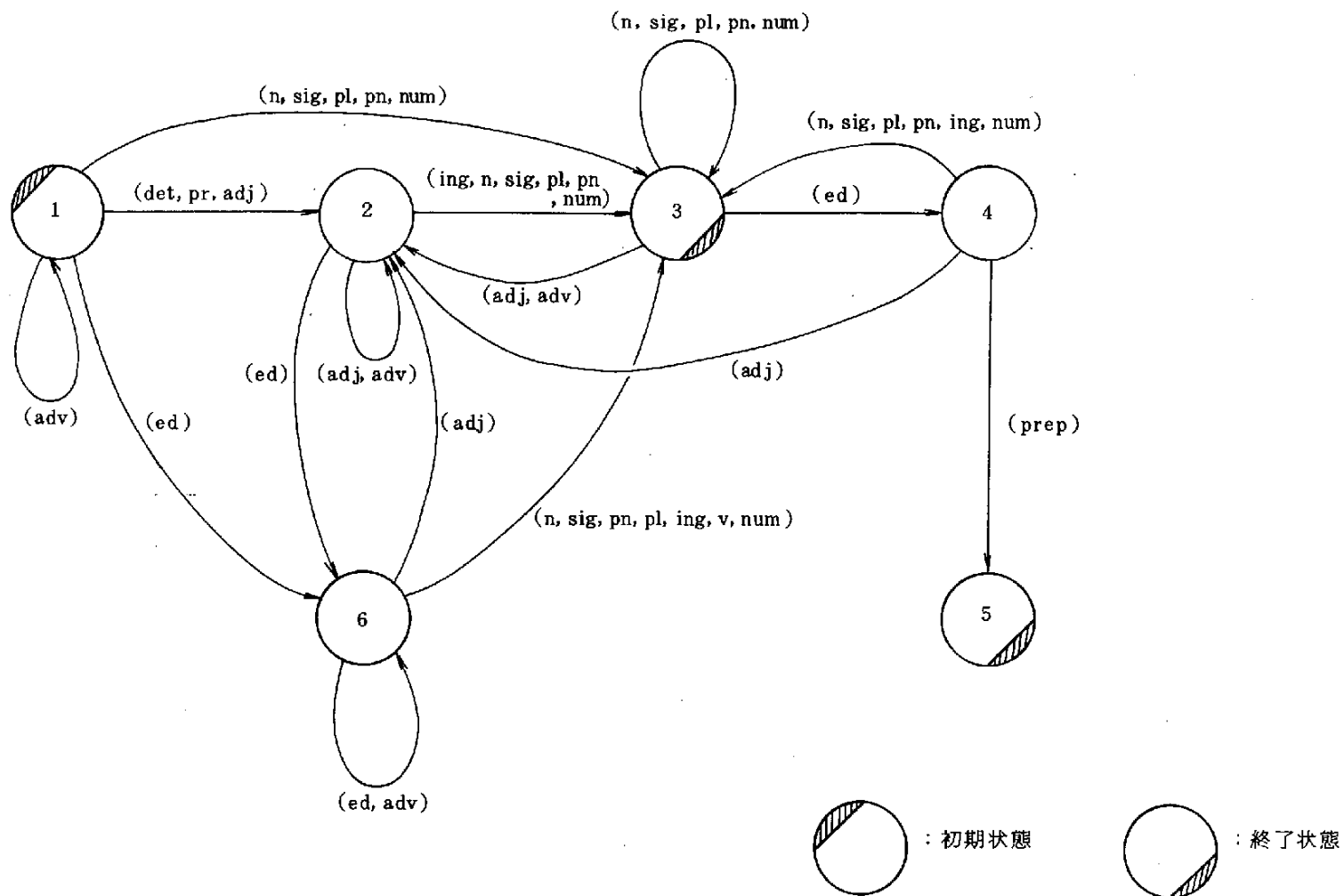


図 2-5 Transition Network (状態遷移網)

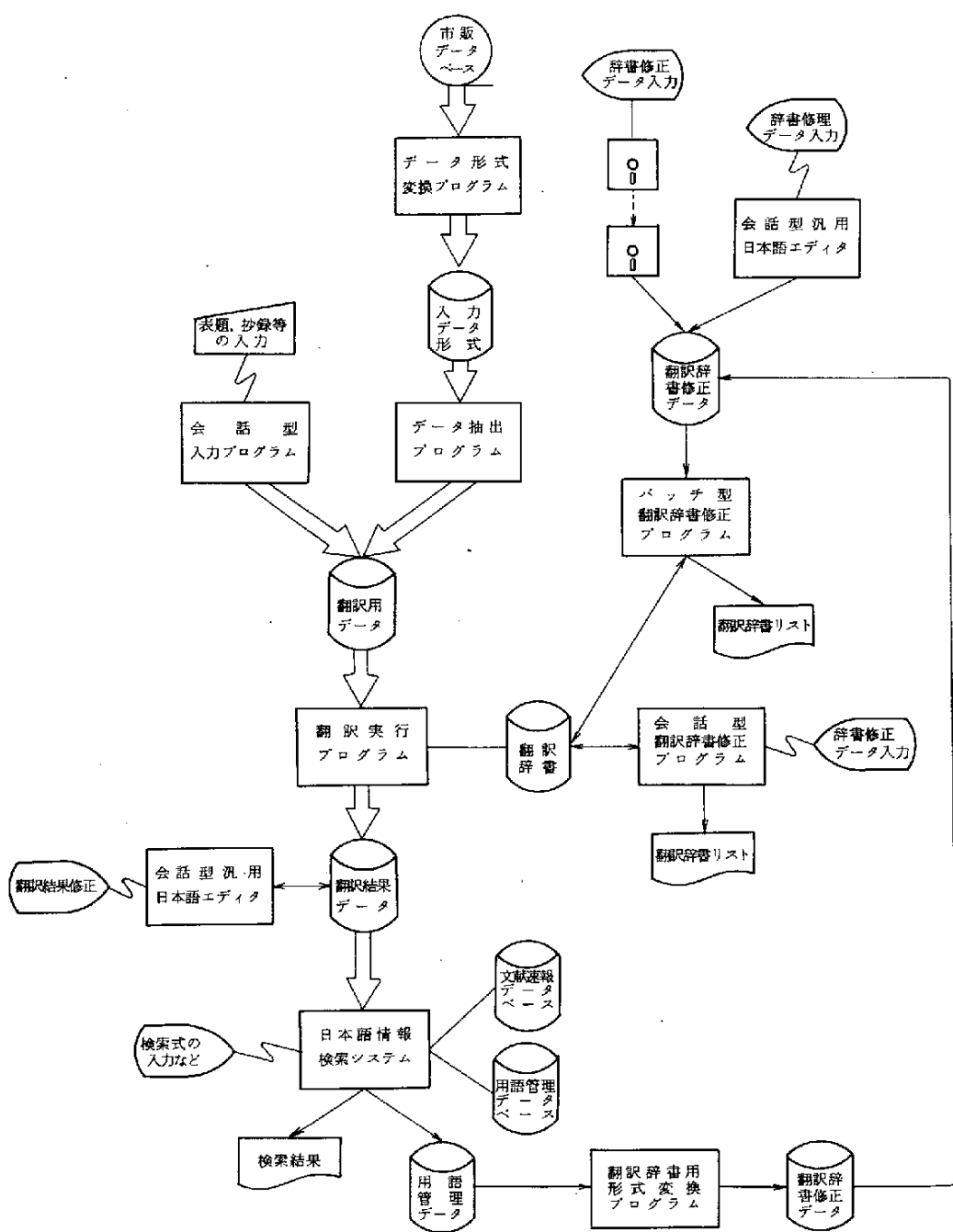


図 2-6 文献速報システムの構成

翻訳辞書編集 (V01/L00)----- 1981.05.08. 【表示】

コマンド入力 ==> DISP  
メッセージ ==>

変換作業行 ==>

| 英語インデックス |  | 翻訳辞書データ |        |        |        |          |                   |        |  |  |  |
|----------|--|---------|--------|--------|--------|----------|-------------------|--------|--|--|--|
| 1        |  | BLACK   |        |        |        |          |                   |        |  |  |  |
| 2        |  |         |        |        |        |          |                   |        |  |  |  |
| 3        |  |         |        |        |        |          |                   |        |  |  |  |
| 品詞 (番号)  |  | 1       | 2      | 3      | R (品詞) | R (表示番号) | R (数)             | 0      |  |  |  |
| R (日本語訳) |  |         |        |        |        |          |                   |        |  |  |  |
| R I      |  | 先行単語    |        |        |        |          | 後続単語              |        |  |  |  |
| 意味カテゴリ   |  | 動詞の主語   |        |        |        |          | 次の中から番号を選んで下さい。   |        |  |  |  |
|          |  | 動詞の目的語  |        |        |        |          | 1) 観点 2) 道具 3) 理論 |        |  |  |  |
|          |  | 名詞      |        |        |        |          | 4) 単位 5) 物体 6) 抽象 |        |  |  |  |
| 同義語      |  |         |        |        |        |          |                   |        |  |  |  |
| 品詞表      |  | 1 形容詞   | 2 副詞   | 3 助動詞  | 4 比較級  | 5 接尾詞    | 6 数詞              | 7 ED   |  |  |  |
|          |  | 8 助動詞   | 9 ING  | 10 同義語 | 11 名詞  | 12 接尾詞   | 13 固有             | 14 代名詞 |  |  |  |
|          |  | 15 前置詞  | 16 疑問詞 | 17 三単数 | 18 記号  | 19 最上級   | 20 動詞             | 21 数詞  |  |  |  |

RETURNキーを押す。



翻訳辞書編集 (V01/L00)----- 1981.05.08. 【表示】

コマンド入力 ==>  
メッセージ ==>

変換作業行 ==>

| 英語インデックス |  | 翻訳辞書データ |        |        |        |          |                   |        |  |  |  |
|----------|--|---------|--------|--------|--------|----------|-------------------|--------|--|--|--|
| 1        |  | 黒い      |        |        |        |          |                   |        |  |  |  |
| 2        |  |         |        |        |        |          |                   |        |  |  |  |
| 3        |  |         |        |        |        |          |                   |        |  |  |  |
| 品詞 (番号)  |  | 1       | 2      | 3      | R (品詞) | R (表示番号) | R (数)             | 2      |  |  |  |
| R (日本語訳) |  | 白馬      |        |        |        |          |                   |        |  |  |  |
| R I      |  | 先行単語    |        |        |        |          | 後続単語              |        |  |  |  |
|          |  |         |        |        |        |          | AND<br>WHITE      |        |  |  |  |
| 意味カテゴリ   |  | 動詞の主語   |        |        |        |          | 次の中から番号を選んで下さい。   |        |  |  |  |
|          |  | 動詞の目的語  |        |        |        |          | 1) 観点 2) 道具 3) 理論 |        |  |  |  |
|          |  | 名詞      |        |        |        |          | 4) 単位 5) 物体 6) 抽象 |        |  |  |  |
| 同義語      |  |         |        |        |        |          |                   |        |  |  |  |
| 品詞表      |  | 1 形容詞   | 2 副詞   | 3 助動詞  | 4 比較級  | 5 接尾詞    | 6 数詞              | 7 ED   |  |  |  |
|          |  | 8 助動詞   | 9 ING  | 10 同義語 | 11 名詞  | 12 接尾詞   | 13 固有             | 14 代名詞 |  |  |  |
|          |  | 15 前置詞  | 16 疑問詞 | 17 三単数 | 18 記号  | 19 最上級   | 20 動詞             | 21 数詞  |  |  |  |

図 2-7 会話型翻訳辞書修正プログラムの使用例

表 2 - 4 プログラムの機能概要

| プログラム   |                                | 機能概要                                                   |
|---------|--------------------------------|--------------------------------------------------------|
| 翻訳データ作成 | ※データ形式変換プログラム                  | 市販文献データ (INSPEC, USGRA等) を FAIRS-I のデータ入力形式に変換する。      |
|         | データ抽出プログラム                     | FAIRS-I データ入力形式に変換された市販文献データから文献速報データベース用のデータを抽出する。    |
|         | 会話型入力プログラム                     | 翻訳用のデータを含む文献速報データベース用のデータの、TSS 配下でのプロンプティング入力を実現する。    |
| 翻訳実行    | 翻訳実行プログラム                      | 英和翻訳プログラムにより、英語から日本語へ表題、キーワード等を翻訳する。                   |
|         | 日本語エディタ ※ (右筆)                 | 翻訳結果の日本語を含むデータを確認修正する。                                 |
| 情報検索    | 日本語情報検索システム<br>※ (FAIRS-I/JEF) | 文献速報データベース、用語管理データベースの蓄積・更新、検索・出力を行う。                  |
| 翻訳辞書修正  | 翻訳辞書修正プログラム                    | バッチで、翻訳辞書の修正を行うもので、データの追加、更新、削除、出力を行うことができる。大量処理に向く。   |
|         | 会話型<br>翻訳辞書修正プログラム             | 日本語ディスプレイを使つて、TSSで、翻訳辞書の修正を行う。日本語データ入力には、カナ漢字変換が利用できる。 |
|         | 翻訳辞書用<br>形式変換プログラム             | 用語管理データベースのデータを翻訳辞書修正データに形式変換する。                       |

※既存プログラム



## 225 今後の課題と将来

### (1) 今後の課題

これからの機械翻訳システムは、人間とコンピュータとの共同作業を元にして、徐々に人間の作業を少なくする方向で進展するものと思われる。そこで、ここでは、機械翻訳プログラムとマンマシン・インターフェースの2つの面から今後の課題を挙げる。

#### A. 機械翻訳プログラムの課題

##### ① 形態素処理，語彙辞書

- 各言語用の形態素処理パッケージ・プログラムの開発

例：英語語形変化処理，日本語分かち書き

- 各言語用分野別用語データベースの開発
- 語彙辞書編集プログラムの開発

##### ② 構文処理

- 機械翻訳用パーザの開発

変形文法，格文法などが適用可能なもの

パーシング・プログラム  
構文規則記述言語  
構文規則編集プログラム

- 各言語用の文法の開発

##### ③ 意味処理，文脈処理，プラグマティックス処理

- 意味，文脈，プラグマティックスの処理方式の研究開発
- 推論機構と知識ベースの研究開発
- 推論を直接サポートする推論マシンの研究開発
- 言語学習機能の研究開発

## B. マンマシン・インタフェースの課題

### ① 入 力

- 大量文書の入力用のOCR装置の開発  
印刷文字，手書き文字
- 音声入力装置の開発  
音声認識技術の研究

### ② 編 集

- 対話型テキスト編集装置または多言語ワードプロセッサの開発

|   |                 |
|---|-----------------|
| { | 画面エディタ：多言語サポート  |
|   | ウィンドウ機能：テキスト参照  |
|   | 辞書参照など          |
|   | 対話型翻訳機能         |
|   | 辞書，文法，規則の画面エディタ |
| { | イメージ・図形処理機能     |

### ③ 出 力

- ドキュメント・プリンタの開発
  - { テキスト，イメージ，図形の出力
  - { カラー出力
- 音声合成装置の開発  
各言語の任意の文を滑らかに抑揚をつけて発声

## (2) 将 来

機械翻訳システムは，世界の各地域間の交流が深まれば深まるほど必要とされるようになるだろう。そして，将来は，誰もが手軽に使用できる機械翻訳システムが期待される。現在の技術から出発すると，人間とコンピュータの共同作業を前提にしたシステムから，徐々にコンピュータ中心のシステムになっていくものと予想される。また，適用分野や言語の範囲も限定されたものから，より広いものになっていくだろう。そ

のためには、意味、文脈、プラグマティックスの処理技術を推論や知識ベースなどの知識情報処理技術と共に研究開発する必要がある。

これから機械翻訳システムを開発するには、機械翻訳そのものの基礎技術の研究開発の他に、マンマシン・インターフェース技術を始めとする実用化のための技術、例えば、多言語翻訳用のテキストエディタ、辞書、文法規則のエディタ、OCR、音声入出力装置などを開発しなければならないだろう。

## 2.3 実験データによるトランスレータの分析

### 2.3.1 INSPECデータによる実験

#### (1) 実験データ

INSPEC文献データベース(1981年)の物理, 化学, 数学関係の論文 3,000 文献

#### (2) 実験方法

1,000文献単位に以下の方法で行った。

- ① 翻訳を行い, 未登録語, 翻訳不可能文の抽出を行う。
- ② 翻訳不可能文を調べ, 熟語の登録を行う。
- ③ 未登録語の登録を行う。
- ④ 再度翻訳を行い, 漏れなどを調べる。
- ⑤ 次の1,000文献を選び, ①～③を繰り返す。

最初の1,000文献のみ全文を検査し, 意味, 翻訳の程度を調べた。

#### (3) 実験結果

表 2-5, 2-6, 図 2-8, 2-9 に翻訳実験結果をまとめる。これらの結果から以下のことが言える。

- ・ 翻訳速度は, 約 0.1 秒/文であった。
- ・ 文型パターンは, 次の使用頻度上位 4 パターンで 99%以上の文の処理が可能であった。

|                                |       |
|--------------------------------|-------|
| ① N, PREP, N                   | 45%   |
| ② N                            | 39%   |
| ③ N, PREP, N, PREP, N          | 12.3% |
| ④ N, PREP, N, PREP, N, PREP, N | 2.7%  |

( N : 名詞句, PREP : 前置詞句 )

表 2 - 5 翻訳実験結果

| 文献番号             | 計算機処理<br>時間 (CPU<br>タイム) | 翻訳不<br>可能文 | 単語・熟語<br>登録後の不<br>可能文 | 未登録語数 | 備 考                           |
|------------------|--------------------------|------------|-----------------------|-------|-------------------------------|
| 1 ~<br>1,000     | 100.16sec                | 49 文       | 11 文                  | 816 語 | 熟語登録数 79 プ<br>ログラムに機能追<br>加あり |
| 1,001 ~<br>2,000 | 105.74sec                | 39 文       | 16 文                  | 567 語 |                               |
| 2,001 ~<br>3,000 | 115.40sec                | 29 文       | 15 文                  | 479 語 |                               |

※ CPUタイムには、未登録語のファイル出力や、翻訳不可能文の骨格パ  
ターンの編集出力などの時間を含む。

表 2 - 6 実験前後の比較

| 比較項目  | 文型パターン数 | 辞書・単語数   | 辞書・熟語数 | 備 考 |
|-------|---------|----------|--------|-----|
| 実 験 前 | 13 パターン | 9,410 語  | —      |     |
| 実 験 後 | 18 パターン | 11,272 語 | 467    |     |

|        |                             |          |                                       |
|--------|-----------------------------|----------|---------------------------------------|
| 日本語訳出順 | ○ 正の数字は、対応するパターンの品詞の先頭からの番号 | !JPPTN01 | -101,5,4,3,2,1\$                      |
|        |                             | !JPPTN02 | -101,1,2\$                            |
|        |                             | !JPPTN03 | -101,3,4,2,1\$                        |
|        |                             | !JPPTN04 | -101,7,6,5,4,3,2,1\$                  |
|        |                             | !JPPTN05 | -101,6,5,3,4,2,1\$                    |
|        |                             | !JPPTN06 | -101,1,-23345,3,-23310,2\$            |
|        |                             | !JPPTN07 | -101,1,-23345,5,4,3,2\$               |
|        |                             | !JPPTN08 | -101,1,-23345,4,3,2\$                 |
|        |                             | !JPPTN09 | -101,3,2,1,-23345,5,-23310,4\$        |
|        |                             | !JPPTN10 | -101,4,3,2,-23310,1\$                 |
|        |                             | !JPPTN11 | -101,1,-23345,3,2\$                   |
|        |                             | !JPPTN12 | -101,1\$                              |
|        | ○ -101以後の数字を利用する。           | !JPPTN13 | -101,3,2,1\$                          |
|        |                             | !JPPTN14 | -101,9,8,7,6,5,4,3,2,1\$              |
|        | ○ -23345: JEF               | !JPPTN15 | -101,2,1\$                            |
|        | の「は」                        | !JPPTN16 | -1,1,2,1,-101,1,2\$                   |
|        | -23310: "「を」                | !JPPTN17 | -101,2,-23345,3,1,-23381\$            |
|        | -23381: "「か」                | !JPPTN18 | -101,11,10,9,8,7,6,5,4,3,2,1\$        |
| 文型パターン |                             | !SKPTN01 | -101,1,-23345,3\$                     |
|        |                             | !SKPTN02 | 9,11:16\$                             |
|        |                             | !SKPTN03 | 11:12\$                               |
|        |                             | !SKPTN04 | 11,9:2\$                              |
|        |                             | !SKPTN05 | 11,15,11:13\$                         |
|        |                             | !SKPTN06 | 11,15,11,9:3\$                        |
|        | ○ 「:」の前の数字は品詞番号             | !SKPTN07 | 11,15,11,9,15,11:5\$                  |
|        |                             | !SKPTN08 | 11,15,11,15,11:1\$                    |
|        |                             | !SKPTN09 | 11,15,11,15,11,15,11:4\$              |
|        |                             | !SKPTN10 | 11,15,11,15,11,15,11,15,11:14\$       |
|        | ○ 「:」の後の数字は日本語訳出順番号         | !SKPTN11 | 11,15,11,15,11,15,11,15,11,15,11:18\$ |
|        |                             | !SKPTN12 | 11,15,11,20,11:9\$                    |
|        |                             | !SKPTN13 | 11,20,1:11\$                          |
|        |                             | !SKPTN14 | 11,20,11:6\$                          |
|        |                             | !SKPTN15 | 11,20,11,15,11:7\$                    |
|        |                             | !SKPTN16 | 11,20,15,11:8\$                       |
|        |                             | !SKPTN17 | 20,11:15\$                            |
|        |                             | !SKPTN18 | 20,11,11:17\$                         |
|        |                             |          | 20,11,15,11:10\$                      |

図 2 - 8 日本語訳出順と文型パターン

\*\*\* FREQ. OF SKPTN USAGE \*\*\*

|                 |            |
|-----------------|------------|
| ◦ SKPTN NO.= 1  | FREQ.= 0   |
| ◦ SKPTN NO.= 2  | FREQ.= 466 |
| ◦ SKPTN NO.= 3  | FREQ.= 0   |
| ◦ SKPTN NO.= 4  | FREQ.= 536 |
| ◦ SKPTN NO.= 5  | FREQ.= 0   |
| ◦ SKPTN NO.= 6  | FREQ.= 0   |
| ◦ SKPTN NO.= 7  | FREQ.= 147 |
| ◦ SKPTN NO.= 8  | FREQ.= 32  |
| ◦ SKPTN NO.= 9  | FREQ.= 2   |
| ◦ SKPTN NO.= 10 | FREQ.= 0   |
| ◦ SKPTN NO.= 11 | FREQ.= 1   |
| ◦ SKPTN NO.= 12 | FREQ.= 0   |
| ◦ SKPTN NO.= 13 | FREQ.= 1   |
| ◦ SKPTN NO.= 14 | FREQ.= 1   |
| ◦ SKPTN NO.= 15 | FREQ.= 0   |
| ◦ SKPTN NO.= 16 | FREQ.= 2   |
| ◦ SKPTN NO.= 17 | FREQ.= 1   |
| ◦ SKPTN NO.= 18 | FREQ.= 1   |

INSPEC 1~1,000文献

「◦」は、最初のパターン

計 1,190

↑ 1文献の表題中に、2つ以上の解析対象文が含まれている場合があるため。

図 2-9 文型パターンの使用頻度 (INSPEC)

表 2-7 翻訳辞書の品詞別件数

| 品 詞 (英字表現)  | 件 数  | 品 詞 (英字表現) | 件 数   |
|-------------|------|------------|-------|
| 形容詞 (ADJ)   | 1330 | 複数名詞 (PL)  | 1291  |
| 副 詞 (ADV)   | 194  | 人 名 (PN)   | 516   |
| 助動詞 (AUX)   | 8    | 代名詞 (PR)   | 25    |
| 比較級 (COM)   | 18   | 前置詞 (PREP) | 59    |
| 接続詞 (CONJ)  | 9    | 疑問詞 (Q)    | 1     |
| 冠 詞 (DET)   | 5    | 三人称单称 (S)  | 45    |
| 過去分詞 (ED)   | 565  | 記 号 (SIG)  | 933   |
| 異綴字 (EQ)    | 0    | 最上級 (SUP)  | 12    |
| 現在進行形 (ING) | 643  | 動 詞 (V)    | 247   |
| 同義語 (MS)    | 0    | 数 詞 (NUM)  | 37    |
| 名 詞 (N)     | 161  |            | 26034 |

#### (4) 評 価

最初の 1,000 文献の詳細な分析で、以下のような機能を追加した。

##### ① 特殊記号の処理 (図 2-10, ①, ②)

- ・ 「,」, 「-」の取り扱い  
「-」+冠詞の場合、文の区切りとみなす。
- ・ (～), 「～」の中は埋め込み文として処理し、結果を名詞句とみなす。

##### ② 規則的な語形変化処理

複数形, 現在進行形, 過去形, 過去分詞形

##### ③ 多品詞語による再処理 (図 2-10, ③)

入力文の骨格パターンが文型パターンにマッチングしない場合

|            |   |                                 |
|------------|---|---------------------------------|
| (名詞句, 名詞句) | } | 下線の部分に多品詞語があれば次の品詞に置き換えて再処理を行う。 |
| (形容詞)      |   |                                 |
| (前置詞, 前置詞) |   |                                 |

##### ④ R・I・(Rigid Idiom) の拡張 (図 2-10, ④)

単語の指定の替りに、品詞の指定を可能にした。

上記の機能追加と未登録語の登録, 熟語の登録を行った後, 翻訳を行い以下のような結果を得た。

- ・ 処理不可能な文または処理の失敗 ..... 14 文
  - ANDの範囲の処理の失敗 ..... 4 文
  - 疑問文の未処理 ..... 4 文
  - 関係代名詞の未処理 ..... 1 文
  - 接続詞の処理の失敗 ..... 1 文
  - 等
- ・ 修正の必要な文 ..... 34 文

訳語の適切さ, 文の滑らかさ, 言い回しの点から判定

このことから, INSPEC のデータに関しては, 約 95% 以上の文が,



- ① ENGLISH : NOTES ON RADIOMETRY AND PHOTOMETRY, A USE IN TRANSFORM  
放射測定法及び測光法についての注意, 変換での使用
- ② ENGLISH : COMMENTS ON 'ABOUT THE EQUIVALENCE OF ABRAHAM'S AND MINKOWSKI'S ELECTRODYNAMICS'  
'アブラハムの及びミンコフスキーの電気力学の等価性について'についてのコメント
- ③ ENGLISH : A NOVEL LOW-TEMPERATURE X-RAY GONIOMETER WITH CLOSED-CYCLE COOLING TO ABOUT 18K  
約 18 K への閉じたサイクル冷却による新しい低温度 X 線角度計
- ④ ENGLISH : TRANSPORT OF THERMAL NEUTRONS IN D/SUB 2/O IN THE TEMPERATURE RANGE 5 TO 60 DEGREESC BASED ON A NEW SCATTERING KERNEL  
新しい散乱核に基づく温度範囲 5 から 60 度 C での D / サブ 2 / O での熱中性子の輸送

ある程度使用できる日本語文に翻訳できたと言える。また、翻訳の質の向上や翻訳不可能数の減少に最も効果があったのは、熟語の登録であることが分かった。図 2-11 に熟語の登録例を示す。

- ① 形容詞 + 前置詞 → 前置詞  
例 consistent with  
sensitive to

② 現在分詞 + 副詞 → 現在分詞  
過去分詞 + 副詞 → 過去分詞  
例 narrowing down  
based merely

③ from + 数詞 + to + 数詞  
例 from 170 to 138  
from 15.0 to 18.25

④ 名詞 + disulfide → 名詞  
名詞 + dioxide → 名詞  
例 carbon disulfide  
carbon dioxide

⑤ the + 数詞 + th  
例 the 29th  
the 19th

図 2-11 熟語の登録例

## 2.4 翻 訳 辞 書

自動翻訳システムを構成する場合には、単なる単語の置き換えという最も原始的なレベルから、各言語に共通なレベルを介しての翻訳というレベルに至るまで、いかなるシステムでも翻訳辞書の存在が不可欠である。

翻訳辞書は上記のレベルの種々の段階で異なった形態をとるものと考えられるが、そのいずれにも共通するのは次の2点である。

- (1) 翻訳に要する十分な語数を含んでいること。これは、どのくらいの語数といった明確な数字は示せないし、語数を増加すれば必ずしも翻訳結果が良くなるとは限らない(たとえば同音異義語の問題が生ずる)が、専門分野を限定した場合に、その分野で用いられる術語はほぼ網羅できるような辞書が必要であろう。
- (2) 拡張性に富むこと。たとえ分野を限定したとしても、新しい術語の発生、元からある語の意味の変化などが起りうる。その際に余り固定化した辞書が作成されていると、その改造に多大の労力を費すことになる。

たとえ専門分野を限定するにしても、上記の点を満たす辞書を全くの白紙から作るのは多大の労力と時間を要するし、言語の構造が未だに完全に解明されていない現状では、良い翻訳辞書を大規模に作成するのは困難である。

そこで既存の辞書をコンピュータのファイル上にのせてデータベース化し、翻訳システムに合うような形に変換することが当面では最も可能性のある方法といえる。既存の辞書(日英翻訳を行う場合には英和、和英辞典)は、語数の上からは十分であるし、これを分析することによって逆に翻訳システムの研究に寄与する面も大きいと思われる。以下では既存の辞書をコンピュータ上にのせた例として、三省堂の「新コンサイス英和辞典」について簡単に述べる。

この辞典は、文部省特定研究によってファイル化されたもので、この辞書の内容がすべてコンピュータのファイルに格納されている。また、見出し語、発音、語義などが段階的に構造化されたファイルも作成されており、データベース用の言語を使うと種々のレベルでの検索が可能になっている。

この辞書データベースは現在のところ、我々が英和辞典を普通のやり方で用いるのとはほとんど同じ使い方しかできないが、これを翻訳辞書を作る場合のベースとして用いることは、上に述べた点で意味のあることである。

しかしながら、これを翻訳辞書にするためには次のような問題点がある。

- (1) 複数の語義をどう扱うか。我々が使う身近な語であるほど一対一の訳というのはしにくい。曖昧さが生じたときにこれをどのように解消するかということは、辞書そのものの問題であるとともに、システムの構成にもかかわってくる。
- (2) 説明的な語義をどう扱うか。訳語が対応するものであれば問題ないが、この辞書はもともと人間が使いやすいように作られているので、語に対する説明となっている部分が多い。
- (3) 品詞の違い。英語では動詞でも日本語では形容詞、またその逆などといった例も多い。

実際の翻訳にあたってはたとえば of のかかり受けをどう扱うかなど、文法事項に直接かかわってくる問題も多いが、これを翻訳辞書レベルで扱うか、別のレベルにするかなども含めて、今後解決しなければならない問題が数多くある。従って、翻訳辞書の作成の問題は、言語の研究の進展、自然言語処理技術の研究などとともに、種々の実験、改良を行う必要があると思われる。

翻訳辞書の利用は、上記のような機械翻訳のほかにも、翻訳者の補助手段などが考えられる。そのためには、たとえば上記のデータベースをそのまま利用することも考えられるが、コンピュータの専門家でない者がデータベース言語のコマンドを打ち込むことは負担が大きいのので望

ましくない。そこで素人が使えるような知的なデータベースの開発が望まれる。これは、コンピュータとの対話が自然言語で行えることのみでなく、たとえばページをめくるようなイメージで使えるグラフィック端末など、システム構成の面からの検討も必要であろう。

## 2.5 イメージ文書処理と入出力装置

あらゆる翻訳システムにおいて、原文の入力、翻訳結果の表示や出力、誤まりの表示などが問題となる。翻訳前の原文は、その中に、テキストの他に、図や表、写真などのイメージが存在し、それらが重要な場合がしばしば見られる。例えば、社名入りのレターヘッドや、サイン、グラフ、カタログなどである。

このようなテキストとイメージの混合した情報を扱えるようにするにはシステムが数種類の情報表現を処理できなければならない。しかし、現在では、それぞれの表現に規格はあるものの、元が異なるので互換性をもつことはない。

イメージ文書処理ではパフォーマンスは非常に高いため、従来の紙、タイプライター、鉛筆、複写、はさみ、のりなどで行えたことが容易にできることが要求される。つまりイメージ文書では、画像処理システムとは異なった機能が要求されることになる。

そのためテキストの中に、画像や手書きのものを挿入、削除、複写、拡大、縮小などの機能が要求される。このようなイメージ文書を扱えるワークステーションとして、ゼロックス社のSTAR、スリーリバーズコンピュータ社のPERQ、シーメンス社のテキスト/ファックスなどがある。

図2-12はテキスト/ファックスの実験的ワークステーションの構成例を示したものである。このシステムの構成は、マルチ・マイクロプロセッサになっており、DMAバスを通して接続されている。入力にはカメラ、出力スキャナを用い、入出力用コンピュータによって接続されている。編集用にはフルページスクリーンの高解像ディスプレイが制御用コンピュータでコントロールされている。さらにオペレーティング・コンピュータにはキーボードとカーソル、操作ディスプレイが接続されている。またファイリング・コンピュータには、

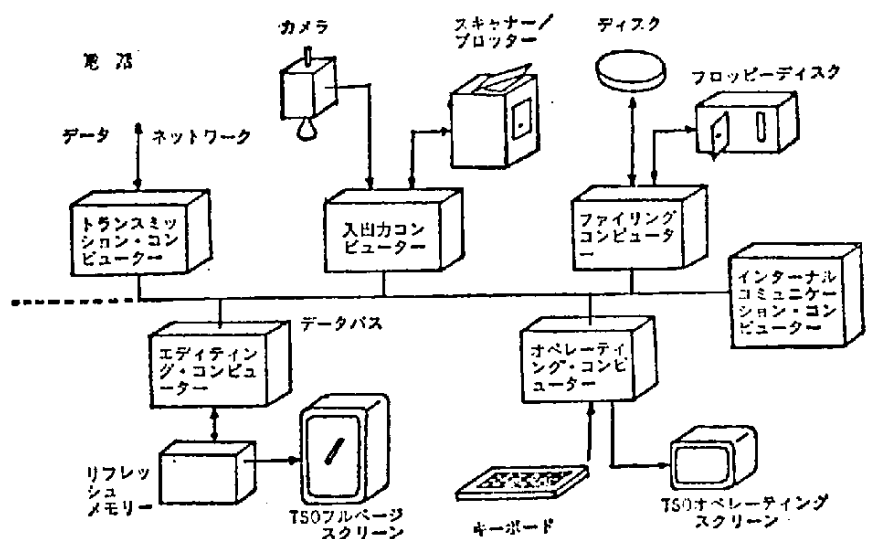


図 2-12 テキスト/ファックス実験的ワークステーション

ディスクとフロッピーディスクが接続されよう。外部との通信は伝送コンピュータ経由で接続される。このようにオフィスのワークステーションとしてはかなり充実したものが、すでに開発され実験されている。

このテキストファックスを使用した例を図2-13に示すように、図や表、写真などのイメージやテキストなどが混在した資料の入出力や各種の処理や編集が自由に行える。

このようなシステムによれば、論文や出版などの分野にあたえるインパクトは大きく、著しい進歩をとげることになる。また、各種の翻訳システムのワークステーションやデザインオートメーション用ワークステーションとしても期待されるものである。

さらに、資料中の文字や図表、写真などはイメージとして認識し、文字認識やイメージ認識を行うことでさらに発展的処理が可能となるよう研究が行われている。

また、イメージ処理やイメージ編集、さらに、多重画面による各種処理の混合を可能にするワークステーションの例として、スリーリーバスコピュー

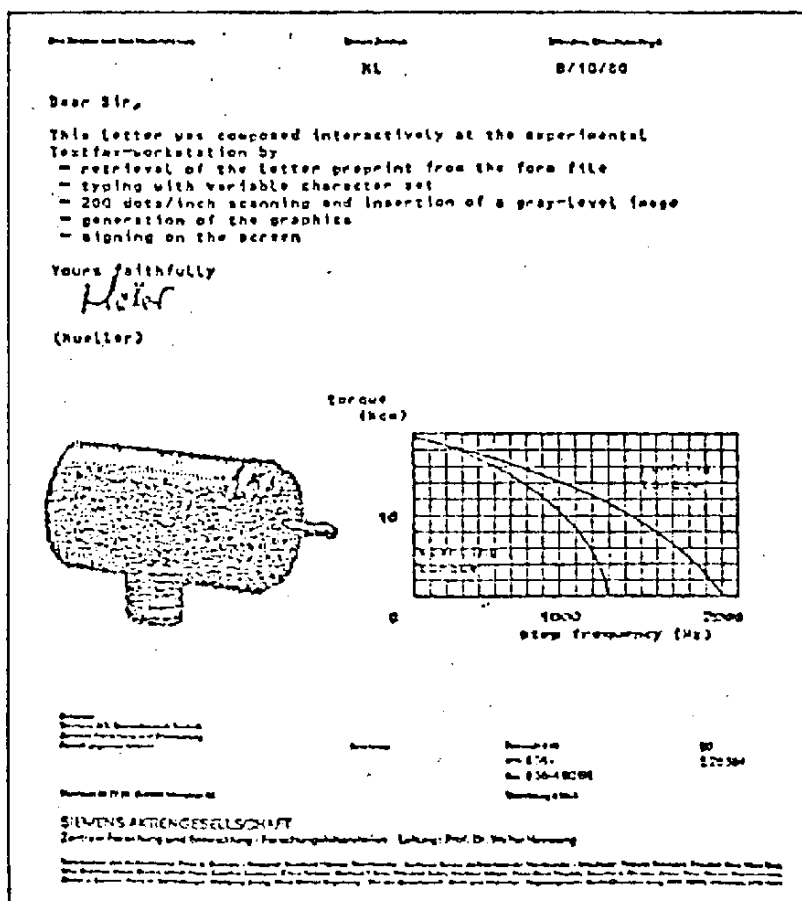


図 2 - 1 3 イメージ文書処理の例

タ社の PERQ がある。PERQ は、16 ビット、マイクロプログラム方式のプロセッサで、 $768 \times 1024$  ドットの A4 サイズの画面、60 回/sec のリフレッシュ機能、画素表示情報のメモリマッピング機能、多重仮想画面による任意画像の多重表示機能などを有している。そのために、各種のイメージ処理を可能にするとともに、翻訳システムに必要と思われる多重画面による原文の表示、翻訳結果の表示、誤まりの指示、各種コマンドなどの処理、翻訳結果の各種編集処理などを容易に行えるワークステーションとして期待されている（図 2 - 1 4 参照）。





图 2-14 PERQ 全景

## 2.6 LC-MARCの実験準備

LC-MARCとは、アメリカ議会図書館が1969年3月から頒布を開始した総合図書目録テープである。

今回実験に用いたテープは、800 bpi, 2048 バイト / 1レコードであった。一例として、次のようなデータが記録されている。それらはインデキータによってデータの有無が示されている。

- ・ 図書、雑誌等の名称
- ・ 著者名、称号
- ・ 刊年、年代推定刊年
- ・ 出版国、コード
- ・ さし絵コード
- ・ 知的水準
- ・ 参考資料コード
- ・ 言語コード
- ・ 目録ソースコード
- ・ その他

RIPSの英和翻訳システムによる実験では、翻訳がどの程度まで可能かを知ることが目的としたので、図書・雑誌等のタイトルのみを目録テープから抽出して、FAIRSデータベースとして、英和翻訳を会話型情報検索システムの中で行った。結果のサンプルを図2-15に示す。これは、COMPUTERという用語とMANAGEMENTという用語のANDで論理式を組んで検索し、結果について英和翻訳したものである。なお、FAIRSデータベースに登録された目録は約32,000件であった。

# 1  
GENERALIZED DATA MANAGEMENT SYSTEMS AND SCIENTIFIC INFORMATION :  
REPORT OF THE SPECIALIST STUDY ON COMPUTER SOFTWARE : THE USE  
OF GENERALIZED DATA MANAGEMENT SYSTEMS FOR HANDLING SCIENTIFIC I  
NFORMATION = SYST MES DE GESTION DE BASES DE DONN ES ET INFORM  
ATION SCIENTIFIQUE : RAPPORT D' TUDE DE SP  
一般化データ管理システム及び科学的情報：コンピュータソフトウェアについてのSPECIALIST研究の報告：  
科学的情報=SYST MES DE GESTION DE BASES DE DONN ES ET情報SCIENT  
IFIQUE取り扱いのための一般化データ管理システムの使用：RAPPORT D' TUDE DE SP

# 2  
MANAGEMENT GUIDE TO COMPUTER SECURITY AND CONTROL / WILLIAM E.  
PERRY.  
コンピュータ保安及び制御/WILLIAME. PERRYへの管理ガイド。

# 3  
TUTORIAL, SOFTWARE CONFIGURATION MANAGEMENT : INITIALLY PRESENT  
ED AT COMPSAC 80, OCTOBER 27-31, 1980, THE IEEE COMPUTER SOCIETY  
'S FOURTH INTERNATIONAL COMPUTER SOFTWARE & APPLICATIONS CONFERE  
NCE / COMPILED BY WILLIAM BRYAN, CHRISTOPHER CHADBOURNE, STAN  
SIEGEL.  
指導, ソフトウェア構成管理: INITIALLYは COMPSAC 80, OCTOBER 27-31, 1980を  
で紹介する. WILLIAM BRYAN, CHRISTOPHER CHADBOURNE, STAN SIEG  
ELによるIEEEコンピュータ協会の四次国際コンピュータソフトウェア&応用会議/ COMPILED.

# 4  
STRATEGIES FOR COMPUTER-ORIENTED PRODUCTIVITY AND EFFECTIVENESS  
IN THE 80'S : A LONG-RANGE PLAN FOR THE EDP COMMUNITY OF ILLINO  
IS STATE GOVERNMENT / PREPARED BY THE SCOPE 80'S TASK FORCE, DE  
PARTMENT OF ADMINISTRATIVE SERVICES, OFFICE OF MANAGEMENT INFORM  
ATION AND COMMUNICATIONS.  
80のでのコンピュータ配向生産性及び効率のための戦略: 管理情報及び通信のADMINISTRATIVEサービ  
ス, 事務所のスコープ80のタスク力, 部によって製作されるILLINOIS状態政府/のEDP共同体のための長  
範囲計画。

# 5  
COMPUTER SECURITY MANAGEMENT / DONN B. PARKER.  
コンピュータ保安管理/DONNB. PARKER.

# 6  
MANAGEMENT GUIDE TO COMPUTER SECURITY / J.R. TALBOT ; EDITED BY  
D.M. POWELL.  
コンピュータ保安/J. R. TALBOTへの管理ガイド; D. M. POWELLによるEDITED.

# 7  
MANAGEMENT GUIDE TO COMPUTER SECURITY AND CONTROL /  
コンピュータ保安及び制御/への管理ガイド

# 8  
COMPUTER SECURITY MANAGEMENT /  
コンピュータ保安管理/

# 9  
MANAGEMENT GUIDE TO COMPUTER SECURITY /  
コンピュータ保安/への管理ガイド

図 2 - 1 5 翻 訳 実 験 の 出 力 例



## 文章情報データベース関連資料一覧

1. 文章情報関連資料 ..... 147
2. 各論第 1 章参考文献 ..... 155
3. 「自然言語処理技術と言語理論」  
（電総研調査報告第 205 号）  
参考文献 ..... 158



## 1. 文章情報関連資料

- |                                                            |                            |                             |
|------------------------------------------------------------|----------------------------|-----------------------------|
| 1) 「用語検討上における一考察」                                          | 1965年11月                   | I R研究会・シソーラス部(財)機械振興協会経済研究所 |
| 2) 「I R研究会シソーラス部会議事録」<br>(42年第1回～第9回)                      | 1967年4月～<br>43年2月          | 同上                          |
| 3) 「シソーラス部用語委員会議事録」<br>(第3回～第5回)                           | 1967年4月,<br>9月, 11月        | シソーラス部会日本貿易振興会貿易資料センター      |
| 4) 「I R研究会議事録」(第1回, 第2回,<br>第6回)                           | 1967年5月,<br>43年2月          | (財) 機械振興協会経済研究所             |
| 5) 「ORA研究会報告書(インフォメーション・リトリバーブルに関する理論的研究)」                 | 1963年8月                    | ORA研究会, 日本機械工業連合会           |
| 6) 「同上 (アバーチャー・カード・システム)」                                  | 1964年3月                    | 同上                          |
| 7) 「インフォメーション・リトリバーブルに関する研究(2)」                            | 1965年7月                    | I R研究会・(財)機械振興協会経済研究所       |
| 8) 「Information Retrieval Systemの現状と動向」                    | 1965年8月                    | 日本電子工業振興協会                  |
| 9) 「通商産業省において使用される漢字の頻度調査」                                 | 1965年7月                    | 通商産業省データ・センター               |
| 10) 「情報検索問題の定式化」                                           | 1965年3月                    | I R研究会・機械振興協会経済研究所          |
| 11) 「経営科学に関する調査, インフォメーション・リトリバーブルに関する研究(5)」               | 1967年4月                    | 機械振興協会経済研究所                 |
| 12) 「インフォメーション・リトリバーブルに関する研究(6)」ーその後のハードウェアの進歩とシソーラス部会のまとめ | 1968年6月                    | 同上                          |
| 13) 「情報検索用シソーラス」                                           | 1965年3月                    | I R研究会・機械振興協会経済研究所          |
| 14) 「シソーラス部会第1次作業資料」(1), (3), (4), (5)                     | 1965年4月,<br>6月, 9月,<br>11月 | I R研究会                      |
| 15) 「通産省資料の用語調査(シソーラス作成作業報告書1)」                            | 1967年5月                    | 通産省官房調査課データ・センター            |

|                                              |          |                                                              |
|----------------------------------------------|----------|--------------------------------------------------------------|
| 16) 「省内用語調査結果」表5                             | 1967年5月  | 通産省官房調査課データ・センター                                             |
| 17) 「シソーラス修正資料」                              | 同上       | 同上                                                           |
| 18) 「情報検索機について」                              | 1967年4月  | 同上                                                           |
| 19) 「通産省における機械検索について」                        | 同上       | 同上                                                           |
| 20) 「SYSTEM MITI-IR第1部, 第2部」                 | 1967年8月  | 同上                                                           |
| 21) 「技術情報の機械検索についての現状」<br>(特許情報検索システムを中心として) | 1970年3月  | 日本情報処理開発センター                                                 |
| 22) 「特許情報機械検索システムの開発研究」                      | 同上       | 同上                                                           |
| 23) 「科学技術情報処理とサービス専門誌」<br>(情報管理)             | 1977年    | 日本科学技術情報センター                                                 |
| 24) 「情報検索用プログラム(KWOCシステム)」                   | 1966年6月  | (社)日本電子工業振興協会                                                |
| 25) 「数理科学11」(特集:情報検索)」                       | 1966年    | ダイヤモンド社                                                      |
| 26) 「情報検索機JETACオペレーション<br>ハンドブック」            | 1968年12月 | NEC日本電気株式会社                                                  |
| 27) 「FUJITSU(漢字処理特集号)」                       | 1970年    | 富士通株式会社 株式会社富士通研究所                                           |
| 28) 「記事索引カード記入の手引—JETAC<br>-IR SYSTEM—」      | 1970年3月  | 日本貿易振興会貿易資料センター                                              |
| 29) 「MITI情報検索システム第3部英文情<br>報DIAシステム」         | 1974年1月  | 通商産業大臣官房情報管理課                                                |
| 30) 「文字情報処理システム, 文書検索サブ<br>システムの設計」          | 1976年3月  | 電子計算機利用に関する技術研究会<br>開発研究分科会文字情報システム開<br>発研究班                 |
| 31) 「日本語情報処理の調査分析」                           | 1977年1月  | 同上                                                           |
| 32) 「特定用語自動抽出システム(システム<br>設計書)」              | 1979年2月  | 同上                                                           |
| 33) 「公衆漢字端末検討会報告書<br>—その1—」                  | 1979年12月 | 国立国会図書館, 日本科学技術情報<br>センター, 日本経済新聞社, 日本電<br>信電話公社, 日本特許情報センター |



- |                                                                        |                    |                                                  |
|------------------------------------------------------------------------|--------------------|--------------------------------------------------|
| 34) 「On-line 文献検索システムJOLDOR」                                           | 1973 年             | 財) 日本情報処理開発センター                                  |
| 35) 「第 1 回ドキュメンテーション研究集会予稿集」                                           | 1964 年 11 月        | 日本科学技術情報センター                                     |
| 36) 「検索技術委員会議事録」(第 1 回～第 5 回)                                          | 1967 年 1 月～<br>4 月 | シソーラス部会 日本貿易振興会<br>貿易資料センター                      |
| 37) 「JETRO 貿易資料センター マイクロシステムについて」                                      |                    | シソーラス部会検索技術委員会<br>日本貿易振興会貿易資料センター                |
| 38) 「特許情報管理に関する調査」                                                     | 1970 年 3 月         | 日本情報処理開発センター                                     |
| 39) 「記事情報検索のためのデータ・マネジメント」                                             | 1972 年 3 月         | 同上                                               |
| 40) 「外務省における情報検索」                                                      | 1966 年 2 月         | 外務省大臣官房電子計算機室                                    |
| 41) 「原型シソーラス作成の研究」                                                     | 1969 年 2 月         | 日本科学技術情報センター<br>日本語による標題索引技術委員会                  |
| 42) 「ドキュメンテーション・情報管理用語集 — 英和編 — 」                                      | 1968 年 3 月         | 同上                                               |
| 43) 「技術情報サービスの機械化」                                                     | 1967 年             | 日本ドキュメンテーション協会                                   |
| 44) 「シソーラスの作成統計確率モデルを採用して」                                             | 1970 年 5 月         | 日本経済新聞社                                          |
| 45) 「JETRO シソーラスに関する報告書」                                               | 1969 年 3 月         | 日本貿易振興会貿易資料センター                                  |
| 46) 「Computer Analysis of Natural Languages」                           | 1967 年 6 月<br>1 日  | Susumu Kuno                                      |
| 47) 「Research on Machine Aids to an Editor of Scientific Translations」 | 1968 年             | American Mathematical Society                    |
| 48) 「Information Storage and Retrieval」                                | 1964 年             | The computation Laboratory of Harvard University |
| 49) 「Syntax-Directed Recognition of Hand-Printed」                      | 1968 年             | Robert H. Anderson                               |
| 50) 「TWO-Dimensional Mathematics」                                      | 同上                 | 同上                                               |
| 51) 「情報検索システム入門」(基礎からオンラインまで)                                          | 1971 年 6 月         | オーム社 長谷川寿彦・梨山修共著                                 |

- |                                                                      |                    |                                                  |
|----------------------------------------------------------------------|--------------------|--------------------------------------------------|
| 52) 「情報検索システム《特性・試験・評価》」                                             | 1972年5月            | 紀伊国屋書店 R.ウィルフリッド・ランカスター, 中村幸雄監訳                  |
| 53) 「情報検索」                                                           | 1974年8月            | (社)電子通信学会                                        |
| 54) 「情報検索」(検索言語, 情報構成・ファイル処理)                                        | 1970年10月           | 日本経営出版会 チャールズ・T・メドウ著, 渡辺茂監訳                      |
| 55) 「海外研究開発レポート」(情報検索技術), (データ・ベースと情報検索), (会話型データ・ベース作成と質問のためのシステム)  |                    | Computer Data Center                             |
| 56) 「Mathematical Linguistics and Automatic Translation」             | 1966年8月~<br>70年12月 | The Computation Laboratory of Harvard University |
| 57) 「漢字情報のKWICシステム(詳細設計書)」                                           | 1973年3月            | 機 特許デイタセンター                                      |
| 58) 「BOOK THESAURUS」工業(I), (II), 政治(I, III), 経済・外報(I, III), 証券, 商品   |                    | 日本経済新聞社                                          |
| 59) 「金属工学機械検索用分類逆引表」                                                 | 1965年8月            | 日本科学技術情報センター                                     |
| 60) 「現代雑誌九十種の用語用字」(第1分冊 総記および語彙表)                                    | 1962年9月            | 秀英出版                                             |
| 61) 「専門分野による漢字および漢字単語の調査分析」(1), (2)                                  | 1974年12月           | (財)日本情報処理開発センター                                  |
| 62) 「行政情報処理用標準漢字選定のための漢字の使用頻度および対応分析結果」                              | 1974年3月            | 行政管理庁 行政管理局                                      |
| 63) 「行政情報処理用標準漢字の選定に関する調査研究報告書」「同資料編」                                | 同上                 | 同上                                               |
| 64) 「官報における漢字の使用頻度および分析結果(中間報告)」                                     | 1974年1月            | 行政管理庁 行政管理局                                      |
| 65) 「経済時系列作図プログラム パッケージ」                                             |                    | 東洋電機製造株式会社計算センター                                 |
| 66) 「IR研究会報告」<br>(インフォメーション・リトリバルに関する研究(4) 第1次シソーラス(I), (II), (III)) | 1966年              | (財)機械振興協会経済研究所                                   |
| 67) 卒業論文「情報検索語彙の変換」                                                  | 1967年              | 中央大学理工学部理工学科 竹中範隆 稲葉茂 高井二三勝                      |

- |                                                                |             |                                      |
|----------------------------------------------------------------|-------------|--------------------------------------|
| 68) 卒業論文「グラフ理論と電子計算機」<br>ー 東京都道路網における最短経路問題 ー                  | 1967 年 1 月  | 中央大学理工学部理工学科 田中博<br>宜 飯田次男 高橋誠 江口    |
| 69) 「技術情報サービスの機械化」ー訪米視<br>察団調査報告ー                              | 1967 年      | (財)日本ドキュメンテーション協会                    |
| 70) 「文献速報自動作成システム」                                             | 1970 年 7 月  | 特殊法人 日本科学技術情報センタ<br>ー電子計算機室          |
| 71) 「漢字電算植字システム適用例報告書<br>(工業統計表)」                              | 1974 年 7 月  | 通商産業省調査統計部統計企画室                      |
| 72) 「ドキュメンテーション・情報処理用語<br>集 ー和英編ー」                             | 1968 年 11 月 | 特殊法人 日本科学技術情報センタ<br>ー日本語による標題索引技術委員会 |
| 73) 「JICST情報検索システム ーBCD<br>モードー」                               | 1970 年 9 月  | 特殊法人 日本科学技術情報センタ<br>ー電子計算機室          |
| 74) 「漢字インプット・データのディスプレ<br>イ装置を用いた校正実験総合報告書」                    | 1970 年 11 月 | 同上                                   |
| 75) 「マイクロフォームによる情報管理セミ<br>ナー」                                  |             | 経済工学協会                               |
| 76) 「Teaching Machine, Training<br>Machine等の現状動向調査報告書」第<br>II部 |             | (財)日本電子工業振興協会                        |
| 77) 「Information Retrieval<br>Systemの現状と動向」                    | 1965 年 8 月  | (財)機械振興協会経済研究所                       |
| 78) 「日本語の作文技術」                                                 | 1976 年 6 月  | 朝日新聞社 本多勝一著                          |
| 79) 「日本語文献テキストの自動処理につ<br>いて」                                   |             |                                      |
| 80) 「JICSTシステム」                                                |             |                                      |
| 81) 「電気通信研究所における電子検索<br>(REWDAC)について」                          |             |                                      |
| 82) 「REWDAC」                                                   |             |                                      |
| 83) 「検索語辞典」                                                    |             |                                      |
| 84) 「自動抄録法および自動索引法の研究」                                         | 1966 年 3 月  | 外務省大臣官房電子計算機室                        |

- |                                                 |             |                                               |
|-------------------------------------------------|-------------|-----------------------------------------------|
| 85) 「JEIPACによる金属工学文献機械検索説明書」                    |             | 日本科学技術情報センター                                  |
| 86) 「THE CONCISE OXFORD DICTIONARY」             |             | OXFORD                                        |
| 87) 「情報処理 I 言語と概念」                              | 1968 年 10 月 | ㈱ 共立出版 中村幸雄著                                  |
| 88) 「工業技術院 試験研究所計画(昭和56年度版)」                    | 1981 年 10 月 | (財) 日本産業技術振興協会                                |
| 89) 「文字情報処理システム(カタカナ入力日本語情報処理システム)」             | 1975 年 3 月  | 電子計算機利用に関する技術研究会<br>開発研究分科会 文字情報システム<br>開発研究班 |
| 90) 「文字情報システム開発研究班(日本語情報処理の調査分析 その2)」           | 1978 年 5 月  | 同上                                            |
| 91) 「文字情報システム開発研究班報告書」<br>—自然言語情報処理の動向—         | 1979 年 6 月  | 同上                                            |
| 92) 「 同上 」—官庁における日本語情報処理—                       | 1980 年 5 月  | 同上 日本語情報処理システム研究班                             |
| 93) 「日本語情報処理システム研究報告書」<br>—自然語による情報検索システムの調査研究— | 1981 年 5 月  | 同上                                            |
| 94) 「自然言語処理技術及びシステムに関する動向調査報告書」                 | 1981 年 3 月  | (財) 日本情報処理開発協会                                |
| 95) 「第5世代のコンピュータ」—コンピュータ関連技術動向調査—               | 1981 年 2 月  | 同上                                            |
| 96) 「情報科学」(日本語の情報処理—VIS Index Systemの思想と構造)     | 1980 年 3 月  | 情報科学研究所                                       |
| 97) 「データ・バンク入門」                                 | 1971 年 2 月  | 日刊工業新聞社 渡辺龍雄編                                 |
| 98) 「貿易分類コード変換表と品目ソーラスの作成に関する調査研究報告書」           | 1971 年 3 月  | 行政管理庁, 通商産業省                                  |
| 99) 「情報処理と I R (情報検索) システム設計」                   | 1971 年 10 月 | (社) 日本能率協会                                    |
| 100) 「日本オペレーションズ・リサーチ学会論文誌」Volume 24, Number 2  | 1981 年 6 月  | (社) 日本オペレーションズ・リサーチ学会                         |

- 101) 「日本の言語学」第三巻 文法Ⅰ 1978年6月 (株)大修館書店 服部四郎 大野晋  
阪倉篤義 松村明
- 102) 「 同上 」第四巻 文法Ⅱ 1979年7月 同上
- 103) 「 同上 」第五巻 意味・語彙 1979年1月 同上 川本茂雄 国廣哲弥  
林大
- 104) 「日英語比較講座」第2巻 文法 同上 国廣哲弥編
- 105) 「電子計算機による新聞の語彙調査(Ⅲ)」 1972年5月 (株)秀英出版 国立国語研究所
- 106) 「行政機関における漢字情報処理に関する調査研究報告書」 1976年3月 (株)芙蓉情報センター  
Ⅰ. 法令検索システムの基礎調査  
Ⅱ. 法令検索システムの概要設計
- 107) 「法令検索システム」実験報告書 1976年6月 通商産業大臣官房 情報管理課
- 108) 「いんぷット」(データエントリー関連 1981年2月 インプット研究所  
記事索引集の概要と利用の方法)
- 109) 「電子計算機による新聞の語彙調査」(Ⅱ, 1971年3月, (株)秀英出版 国立国語研究所  
(Ⅳ) 1973年6月
- 110) 「漢文入門」 1962年12月 学燈社 藤堂明保著
- 111) 「数理言語学入門」 1972年10月 東京図書株式会社 グロス, ランタ  
ン共著
- 112) 「言語と論理」 1978年5月 紀伊国屋書店 M・J・クレスウェル
- 113) 「事務管理」(文書処理システム) 1981年5月 日刊工業新聞社  
Vol 20, No 5
- 114) 「機械翻訳システムに関する海外調査」 1981年 (社)日本電子工業振興協会・機械翻  
訳システム調査委員会
- 115) 「自然言語処理技術と言語理論」電子技 1981年4月 工業技術院 電子技術総合研究所  
術総合研究所調査報告 第205号
- 116) 「計算機による自然言語の意味処理に関 1981年7月 同上  
する研究」 同上 第797号
- 117) 「Proceedings of the COLING 1980年4月  
80」
- 118) 「情報処理学会 第23回(昭和56年後期) 1981年10月 (社)情報処理学会  
全国大会」講演論文集

- 119) 「文法の構造」 1963 年 11 月 研究社出版 チョムスキー著
- 120) 「情報処理」(大特集/プログラミング 1981 年 6 月 オーム社  
言語の最近の動向)
- 121) 「国際緊張緩和の政治過程」(年報 政 1969 年 岩波書店 日本政治学会編  
治学 1969)
- 122) 「The General Inquirer」" A 1966 年 MIT, P.J. Stone, et al  
Computer Approach to Content  
Analysis "
- 123) 「Trent in Content Analysis」 1959 年 I.D.S. pool (Ed)

## 2. 各論第1章参考文献

- 1) 「産業別海外情報に関する基礎調査報告書」, 1981年3月, (財)日本情報処理開発協会。
- 2) 「The Application of New Methodologies to analyze the Soviet perceptions of U.S. Policies Volume I」, 1975年, CACI Inc.  
「The Application of New Methodologies to analyze the Soviet perceptions of U.S. Policies Volume II」, 1975年, CACI Inc.  
「Analysis of the Soviet, Chinese, and U.S. Crisis Management Experiences: Interim Technical Report」, 1979年, CACI Inc.
- 3) C.E. Osgood, G.J. Suci, P.H. Tannenbaum, 「The Measurement of Meaning」, 1957年, University of Illinois Press.
- 4) R.C. North, O.R. Holsti, M.G. Zaninovich, D.A. Zinnes 「Content Analysis: a Handbook with Applications for the Study of International Crisis」, 1963年, Northwestern University Press.
- 5) O.R. Holsti, 「Content Analysis for the Social Sciences and Humanities」, 1969年, Addison-Wesley Publishing Company Inc.
- 6) 花井等「現代国際関係論」, P. 277, 1974年, ミネルヴァ書房
- 7) 「産業別情報拠点パイロット・システム研究開発報告書」, 1981年3月, (財)日本情報処理開発協会
- 8) T. Yamamoto, M. Negishi, M. Ushimaru, Y. Tozawa, K. Okabe, S. Fujiwara, 「TOOL-IR: An On-Line Information Retrieval System at an Inter-University Computer Center」, Proc. 2nd USA-Japan Computer Conf., P. 159, 1975年
- 9) 根岸正光, 山本毅雄, 「オンライン文献情報検索システム・TOOL-IRにおけるマン・マシン・インターフェース」, 情報処理, Vol 17, No.5, 1976年.
- 10) 山本毅雄, 「オンライン情報サービスの歴史」, 情報処理, Vol 21, P.99, 1980.
- 11) 山本毅雄, 「オンライン情報検索の技術とその現状」, 電気通信学会誌, Vol 63, P.1255, 1981年
- 12) 杉原厚吉, 伊理正夫, 「2部グラフの分割理論を利用した概念構造決定法」, オペレーションズ・リサーチ, 1981年8月号, P.504, OR学会

- 13) 石井成樹, 「立ち上がる日本語ワード・プロセッサ」, コンピュートピア, 1981年9月号,  
P.26
- 14) 長尾真「計算機による日本語文章の解析に関する研究」, 1979年, 京都大学。
- 15) G.Salton, 神保健二(訳), 「SMART情報検索システム - データベースへのアプローチ」,  
1974年, 企画センター
- 16) 渡辺龍雄, 林秀行, 山本恵美子, 地福光子, 「データバンク入門」, 1972年, 日刊工業新聞社
- 17) F.W. Lancaster, 松村多美子(訳), 「情報検索の言語」, 1976年, (社)日本ドクメンテ  
ーション協会
- 18) 猪口孝, 「国際関係の数量分析」, 1970年, 巖南堂
- 19) 猪口孝, 「Measuring Friendship and Hostility among Communist Powers :  
Some Unobstrusive Measures of Esoteric Communication」, Social  
Science Research, Vol.1, No.1, P.79, 1972, Seminar Press Inc.
- 20) 猪口孝, 「越中ノ関係, 1964年-1980年 - 探索的分析 - 」, 東京大学東洋文化研究  
所紀要, 第87冊, P.197.
- 21) 林知己夫, 樋口伊佐夫, 駒沢勉, 「情報処理と統計数理」, 1970年, 産業図書
- 22) R.Wich, 「Sino-Soviet Crisis Politics」, Harvard University Press.
- 23) 国際体制研究会「認知構造図による対外政策決定分析」, 1980年, (財)世界経済情報サービ  
ス
- 24) 猪口孝, 「Explaining and Predicting Japanese General Elections, 1960 -  
1980」, J.Jap. Studies, Vol.7, No.2, 1981年
- 25) A. Tanaka, 「Chinese International Conflict Behavior 1949-1978」,  
1981年, Ph.D Thesis of MIT
- 26) 緒方良彦, 竹谷一郎「新聞情報活用法」, 1980年, 産業能率大学出版部
- 27) 小川芳樹, 「エネルギー関連の非時系列数値情報のデータバンク化の研究」, エネルギー経済,  
第6巻, 第3号, P.23, 1980年, (財)日本エネルギー経済研究所
- 28) 西田富士夫, 「言語情報処理」, 1981年, コロナ社
- 29) Terry Winograd, 淵一博(訳), 田村浩一郎(訳), 白井良明(訳), 「言語理解の構造」,  
1976年, 産業図書
- 30) 伊理正夫, 「数と式と文の処理」, 1981年, 岩波書店
- 31) 衛藤瀋吉, 「中国をめぐる国際政治」, P.183, 1968年, 東京大学出版会



- 32) 渡辺昭夫,「近代日本の対外態度」, P.225, 1974年, 東京大学出版会
- 33) 武者小路公秀,「行動科学と国際政治」, 1970年, 東京大学出版会
- 34) 白井久和,「対外政策決定過程の日米比較」, P.347, 1977年, 東京大学出版会
- 35) 岡部達味,「中国の対日政策」, 1976年, 東京大学出版会
- 36) D.A. Graber,「Verbal Behavior and Politics」, 1976年, University of Illinois Press.
- 37) 水谷静夫,「数理言語学」, 1982年, 培風館
- 38) 安田三郎, 海野道郎,「社会統計学」, 1977年, 丸善
- 39) 河口至商,「多変量解析入門」, 1973年, 森北出版
- 40) 芝祐順,「因子分析法」, 1979年, 東京大学出版会
- 41) 林知己夫, 鮑戸弘,「多次元尺度解析法」, 1976年, サイエンス社
- 42) 小山昭雄, 森田道也,「オペレーションズ・リサーチ」, P.205, 1980年, 培風館
- 43) 安田三郎,「社会調査の計画と解析」, 1970年, 東京大学出版会
- 44) 池田央,「行動科学の方法」, 1971年, 東京大学出版会
- 45) 高木貞二,「現代心理学と数量化」, 1972年, 東京大学出版会
- 46) 安田三郎,「数理社会学」, 1973年, 東京大学出版会
- 47)「AIM操作手引書」, 1979年, 富士通
- 48)「AIM ADL文法手引書」, 1978年, 富士通

3 「自然言語処理技術と言語理論」参考文献（電子技術総合研究所調査報告  
第205号より）

（略号 AJCL≡American Journal of Computational Linguistics  
CACM≡Comm. of American Computing Machinery  
IJCAI≡International Joint Conference of Artificial Intelligence）

- [Allwood 77] Allwood, J., Andersson, L. and Dahl, O.: "Logic in Linguistics", Cambridge Univ. Press, 1977.
- [Austine 62] Austine, J. L.: "How to Do Things with Words", Oxford Univ. Press, 1962.
- [Bach 68] Bach, E. and Harms R. T. (Eds.): "Universals in Linguistic Theory", Holt, Rinehart and Winston, 1968.
- [Barr 80] Barr, A. and Davidson, J.: "Representation of Knowledge", Stanford Memo HPP-80-3, Computer Science Dept., Stanford Univ., 1980.
- [BBN 78] BBN Report 3878. "Research in Natural Language Understanding", Quaterly Progress Report No. 3, 1978.
- [Bever 71] Bever, T. G. and Rosenbaum P. S.: "Some Lexical Structures and Their Empirical Validity", in Steinberg and Jakobovits (Eds.): "Semantics", Cambridge Univ. Press, 1971.
- [Bobrow 75] Bobrow, D. G. and Collins, A. M. (Eds.): "Representation and Understanding", Academic Press, 1975 (淵 一博監訳:「人工知能の基礎」, 近代科学社, 1978).
- [Bobrow 77a] Bobrow, D. G. and Winograd, T.: "An Overview of KRL, a Knowledge Representation Language", Studies in Cognitive Science, Vol. 1, No. 1, 1977, 3-46.
- [Bobrow 77b] Bobrow, D. G., Winograd, T. and the KRL Research Group: "Experience with KRL -0, One Cycle of a Knowledge Representation Language", IJCAI5, 1977, 213-222.
- [Bobrow 77c] Bobrow, D. G., Kaplan, R. M., Kay, M., Norman, D. A. and Winograd, T.: "GUS, A Fame-Driven Dialog System", Artificial Intelligence, Spring, 1977, 155-173.
- [Bolc 78] Bolc L. (Ed.), "Natural Language Communication with Computers", Springer-Verlag, 1978.
- [Bonnet 78] Bonnet, A.: "BAOBAB, A Parser for a Rule-Based System Using a Semantic Grammar", STAN-CS-78-668, Stanford Univ., 1978.
- [Brown 75] Brown, J. S. and Burton, R. R.: "Multiple Representations of Knowledge", in Bobrow D. G. and Collins A. (Eds.): "Representation and Understanding", Academic Press, 1975, 331-349.
- [Bruce 75] Bruce, B.: "Case Systems for Natural Language", Artificial Intelligence, 6, 1975, 327-360.
- [Carbonell 70] Carbonell, J. R.: "AI in CAI: An Artificial Intelligence Approach to Computer-Assisted Instruction", IEEE, MMS-11, 4, Dec. 1970, 190-202.
- [Carbonell 73] Carbonell, J. R. and Collins, A. M.: "Natural Semantics in Artificial Intelligence", IJCAI3, 1973.
- [Celce-Murcia 76] Celce-Murcia, M.: "Verb Paradigms for Sentence Recognition", AJCL, Microfiche 38, 1976.

- [Chafe 70] Chafe, W. L. : "Meaning and Structure of Language", The Univ. of Chicago Press, 1970.
- [Chafe 72] Chafe, W. L. : "Discourse Structure and Human Knowledge", in Carrol and Freedle (Eds.) : "Language Comprehension and the Acquisition of Knowledge", John Wiley and Sons, 1972.
- [Chafe 73] Chafe, W. L. : "Language and Memory", *Language*, 49, 2, 1973, 261-281.
- [Chafe 74] Chafe, W. L. : "Language and Consciousness", *Language*, 50, 1, 1974, 111-133.
- [Chang 78] Chang, C. L. : "DEDUCE 2 Further Investigations of Deduction in Relational Data Bases", in Gallaire, H. and Minker, J. (Eds.) : "Logic and Data Bases", Plenum, 1978.
- [Codd 78] Codd, E. F. : "Rendezvous Version 1 : An Experimental English-Language Query Formulations System for Casual Users of Relational Data Bases", RJ2144 (29407) Computer Science Research Report, IBM Research Lab. San Jose, Calif., 1978.
- [Charniak 72] Charniak, E. : "Toward a Model of Children's Story Comprehension", AI TR-226, MIT, December 1972.
- [Charniak 76] Charniak, E. and Wilks, Y. : "Computational Semantics", North-Holland, 1976.
- [Chomsky 65] Chomsky, N. : "Aspects of the Theory of Syntax", The MIT Press, 1965.
- [Chomsky 71] Chomsky, N. : "Deep Structure, Surface Structure, and Semantic Interpretation", in Steinberg and Jakobovits (Eds.) : "Semantics", Cambridge University Press, 1971.
- [Chomsky 75] Chomsky, N. : "Reflections on Language", Pantheon, 1975.
- [Collins 75] Collins, A., Warnock, E. H., Niello, N. and Miller, M. L. : "Reasoning From Incomplete Knowledge", in Bobrow, D. and Collins, A. (Eds.) : "Representation and Understanding", Academic Press, 1975, 383-415.
- [Collins 75] Collins, A. M. and Loftud, E. F. : "A Spreading-Activation Theory of Semantic Processing", *Psychological Review*, Vol. 82, No. 6, 1975, 407-428.
- [Davis 75] Davis, R. and King, J. : "An Overview of Production Systems", Stanford AIM-271, Oct. 1975.
- [Dillon 77] Dillon, G. L. : "Introduction to Contemporary Linguistic Semantics", Prentice Hall, 1977.
- [Dowty 81] Dowty, D. R. : "Introduction to Montague Semantics", Reidel, SLL, 11, 1981.
- [Dresher 76] Dresher, B. E. and Hornstein, N. : "On Some Supposed Contributions of Artificial Intelligence to the Scientific Study of Language", *Cognition*, 4, 1976, 321-398.
- [Finin 79] Finin, T., Goodman, B. and Tennant, H. : "JETS : Achieving Completeness through Coverage and Closure", *IJCAI*, 1979, 275-281.
- [Fillmore 68] Fillmore, C. J. : "The Case for Case", in Bach and Harms (Eds.) : "Universals in Linguistic Theory", Holt, Rinehart, and Winston, 1968.
- [Fillmore 69] Fillmore, C. J. : "Types of Lexical Information", in Kieffer (Ed.) : "Studies in Syntax and Semantics", D. Reidel Publishing Co., 1969.
- [Fillmore 77] Fillmore, C. J. : "Scenes-and-Frames Semantics", in Zampolli (Ed.) : "Linguistic Structures Processing", North-Holland, 1977.
- [Fordor 77] Fordor, J. D. : "Semantics : Theories of Meaning in Generative Grammar", Thomas Y. Crowell Co., 1977.

- [Frederiksen 75] Frederiksen, C. H. : "Representing Logical and Semantic Structure of Knowledge Acquired from Discourse", *Cognitive Psychology*, 7, 1975, 371-458.
- [Gardner 79] Gardner, A., Davidson, J. and Winograd, T. : "Natural Language Understanding", Stanford Computer Science Dept. Rep. No. STAN-CS-79-754, 1979.
- [Goldman 75] Goldman, N. M. : "Sentence Paraphrasing from a Conceptual Base", *CACM*, 2, 18, 1975, 96-106.
- [Gregg 74] Gregg, L. W. (Ed.) : "Knowledge and Cognition", Laurence Erlbaum Associates Publishers, Potomac, Maryland, 1974.
- [Grice 75] Grice, H. P. : "Logic and Conversation", in Cole, P. and Morgan, J. L. (Eds.) : "Syntax and Semantics", Vol. 3, Academic Press, 1975, 41-58.
- [Grishman 76] Grishman, R. : "A Survey of Syntactic Analysis Procedures for Natural Language", *AJCL*, Microfiche 47, 1976.
- [Grosz 77] Grosz, B. J. : "The Representation and Use of Focus in a System for Understanding Dialogs", *IJCAI5*, 1977, 67-76.
- [Halle 78] Halle, M., Bresnan, J. and Miller, G. A. : "Linguistic Theory and Psychological Reality", The MIT Press, 1978.
- [Halliday 61] Halliday, M. A. K. : "Categories of the Theory of Grammar", *Word*, 17, 1961, 241-292.
- [Harada 77] Harada, S. I. : "Topics in Relational Grammar", *Ann. Bull. of RILP*, No. 11, University of Tokyo, 1977, 131-153.
- [Harris 77] Harris, L. R. : "User-Oriented Data Base Query with ROBOT Natural Language Query System", Third International Conf. on Very Large Data Bases, 1977, 303-311.
- [Harris 79] Harris, L. R. : "Experience with ROBOT in 12 Commercial Natural Language Data Base Query Application", *IJCAI6*, 1979, 365-368.
- [Hawkinson 75] Hawkinson, L. : "The Representation of Concepts in OWL", *IJCAI4*, 1975, 107-114.
- [Hays 76] Hays, D. G. and Mathias, J. (Eds.) : "FBIS Seminar on Machine Translation", *AJCL*, Microfiche 46, 1976.
- [Hayes 79] Hayes, P. and Reddy, R. : "Graceful Interaction in Man-Machine Communication", *IJCAI6*, 1979, 372-374.
- [Heidorn 75] Heidorn, G. E. : "Augmented Phrase Structure Grammar", in Schank and Nash-Webber (Eds.) : "Theoretical Issues in Natural Language Processing", Cambridge, Mass. 1975, 1-5.
- [Hendrix 77] Hendrix, G. G. : "Human Engineering for Applied Natural Language Processing", Stanford Research Institute AI Center, TR 139, Feb. 1977.
- [Hobbs 78] Hobbs, J. R. and Rosenschein, S. J. : "Making Computational Sense of Montague's Intensional Logic", *Artificial Intelligence*, 9, 1978, 287-306.
- [Hollan 75] Hollan, J. D. : "Features and Semantic Memory: Set-Theoretic or Network Model?", *Psychological Review*, Vol. 82, No. 2, 1975, 154-155.
- [Hutchins 78] Hutchins, W. J. : "Progress in Documentation-Machine Translation and Machine-Aided Translation", *Journal of Documentation*, 34, 2, June, 1978.

- [Kaplan 72] Kaplan, R. M. : "Augmented Transition Networks as Psychological Models of Sentence Comprehension", *Artificial Intelligence*, 3, 1972, 77-100.
- [Kintsch 72] Kintsch, W. : "Notes on the Structure of Semantic Memory", in Tulving and Donaldson (Eds.) : "Organization of Memory", Academic Press, 1972.
- [de Kleer 75] de Kleer, J. : "Qualitative and Quantitative Knowledge", AI-TR-352, MIT, 1975.
- [Kuno 65] Kuno, S. : "The Predictive Analyzer and a Path Elimination Technique", *CACM*, 8, 1965, 687-698.
- [Kuno 77] Kuno, S. and Kaburaki, E. : "Empathy and Syntax", *Linguistic Inquiry*, Vol. 8, No. 4, 1977.
- [Lakoff 76] Lakoff, G. : "Pronouns and Reference", in McCawley, J. D. (Ed.) : "Syntax and Semantics", Vol. 7, Academic Press, 1976, 275-335.
- [Lehnert 77] Lehnert, W. G. : "Human and Computational Question Answering", *Cognitive Science*, Vol. 1, 1977, 47-73.
- [Lehnert 78] Lehnert, W. G. : "The Process of Question Answering", Lawrence Erlbaum Associates, Publishers, 1978.
- [Lyons 68] Lyons, J. : "Introduction to Theoretical Linguistics", Cambridge Univ. Press, 1968.
- [Marcus 80] Marcus, M. : "A Theory of Syntactic Recognition for Natural Language", The MIT Press, 1980.
- [Martin 74] Martin, W. A. : "A System for Building Expert Problem Solving Verbal Reasoning", OWL-Notes MIT MAC, 1974.
- [Martin 78] Martin, W. A. : "Descriptions and Specialization of Concepts", LCS-TM-101, MIT, 1978.
- [McDermott 74] McDermott, D. V. : "Assimilation of New Information by a Natural Language Understanding System", AI TR-291, MIT, Feb. 1974.
- [Michie 77] Michie, D. (Ed.) : "Machine Intelligence", 8, Edinburgh Univ. Press, 1977.
- [Miller 76] Miller, G. A. and Johnson-Laird, P. N. : "Language and Perception", Harvard Univ. Press, 1976.
- [Minsky 68] Minsky, M. (Ed.) : "Semantic Information Processing", The MIT Press, 1968.
- [Minsky 75] Minsky, M. : "Framework for Representing Knowledge", in Winston (Ed.) : "The Psychology of Computer Vision", McGraw-Hill, 1975.
- [Minsky 76] Minsky, M. : "Computation : Finite and Infinite Machines", Prentice-Hall, Inc., 1976.
- [Nagao 76] Nagao, M. and Tsujii, J. : "Analysis of Japanese Sentences Using Semantic and Syntactic Information", *AJCL*, Microfiche 41, 1976.
- [Nash-Webber 77] Nash-Webber, B. and Reiter, R. : "Anaphora and Logical Form : On Formal Meaning Representations for Natural Language", *IJCAI*, 1977, 121-131.
- [Newell 73] Newell, A. : "Artificial Intelligence and the Concept of Mind", in Schank and Colby (Eds.) : "Computer Models of Thought and Language", Freeman and Company, 1973.

- [Newell 73] Newell, A. : "Productions Systems : Models of Control Structures", in Chase (Ed.) : "Visual Information Processing", Academic Press, 1973.
- [Nida 75] Nida, E. A. : "Componential Analysis of Meaning", Mouton & Co. N. V., 1975.
- [Nilson 74] Nilson, N. J. : "Artificial Intelligence", IFIP74, North-Holland, 1974, 778-801.
- [Norman 75] Norman, D. A. and Rumelhart, D. E. : "Explorations in Cognition", Freeman and Company, 1975.
- [Novak 76] Novak, G. S. : "Computer Understanding of Physics Problems Stated in Natural Language", AJCL, Microfiche 53, 1976.
- [Ogden 23] Ogden C. K. and Richards, I. A. : "The Meaning of Meaning", London : Kegan Paul, 1923.
- [Palmer 76] Palmer, F. R. : "Semantics", Cambridge Univ. Press, 1976.
- [Petrick 76a] Petrick, S. R. : "On Natural Language Based Computer Systems", IBM J. Res. Develop., July, 1976, 315-325.
- [Petrick 76b] Petrick, S. R. : "Summary of FBIS Seminar on Machine-Aided Translation", AJCL, Microfiche 46, 1976.
- [Pereira 80] Pereira, F. C. N. and Warren, H. D. : "Definite Clause Grammars for Language Analysis—A Survey of the Formalism and a Comparison with Augmented Transition Networks", Artificial Intelligence, 13, 1980, 231-278.
- [Plath 76] Plath, W. J. : "REQUEST : A Natural Language Question-Answering System", IBM J. Res. Develop., July, 1976, 326-335.
- [Pratt 73] Pratt, V. R. : "A Linguistic Oriented Programming Language", IJCAI3, 1973, 272-381.
- [Pratt 75] Pratt, V. R. : "LINGOL-A Progress Report", IJCAI4, 1975, 422-428.
- [Quillian 68] Quillian, M. R. : "Semantic Memory", in Minsky (Ed.) : "Semantic Information Processing", The MIT Press., 1968.
- [Raphael 75] Raphael, B. : "SIR : A Computer Program for Semantic Information Retrieval", in Minsky (Ed.) : "Semantic Information Processing", The MIT Press, 1968, 33-134.
- [Reddy 75] Reddy, D. R. (Ed.) : "Speech Recognition", Academic Press, 1975.
- [Reiter 78] Reiter, R. : "On Reasoning by Default", in Waltz (Ed.) : "Theoretical Issues in Natural Language Processing-2", ACM, 1978, 210-218.
- [Rieger 74] Rieger, C. : "Conceptual Memory", Stanford AIM-247, Jan. 1974.
- [Rieger 79] Rieger, C. and Small, S. : "Word Expert Parsing", IJCAI6, 1979, 723-726.
- [Riesbeck 74] Riesbeck, C. K. : "Computational Understanding : Analysis of Sentences and Context", Stanford AIM-238, May 1974.
- [Roberts 77] Roberts, R. B. and Goldstein, I. P. : "The FRL Manual", MIT AI Memo 409, Sept. 1977.
- [Rumelhart 72] Rumelhart, D. E., Lindsay, P. H. and Norman, D. A. : "A Process Model for Long-Term Memory", in Tulving and Donaldson (Eds.) : "Organization of Memory", Academic Press, 1972.
- [Rumelhart 73] Rumelhart, D. E. and Norman, D. A. : "Active Semantic Networks as a Model of Human Memory", IJCAI3, 1973, 450-457.
- [Rustin 71] Rustin, R. (Ed.) : "Natural Language Processing", Algorithmic Press, 1971.

- [Schank 73a] Schank, R. C., and Colby K. M. (Eds.): "Computer Models of Thought and Language", Freeman and Company, 1973.
- [Schank 73b] Schank, R. C.: "Identification of Conceptualizations underlying Natural Language", in Schank and Colby (Eds.): "Computer Models of Thought and Language", Freeman and Company, 1973.
- [Schank 73c] Schank, R. C. et al: "MARGIE: Memory, Analysis, Response Generation, and Inference on English", IJCAI3, 1973, 255-261.
- [Schank 74] Schank, R. C. and Rieger, C.: "Inference and Computer Understanding of Natural Language", Artificial Intelligence, 5, 1974, 373-412.
- [Schank 75] Schank, R. C.: "Conceptual Information Processing", North-Holland/American Elsevier, 1975.
- [Schank 77] Schank, R. C. and Abelson, R.: "Scripts Plans Goals and Understanding", Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1977.
- [Shapiro 78] Shapiro, S. C.: "Path-Based and Node-Based Inference in Semantic Networks", in Waltz (Ed.): "Theoretical Issues in Natural Language Processing-2", ACM, 1978, 219-225.
- [Shortliffe 76] Shortliffe, E. H.: "Computer-Based Medical Consultations: MYCIN", American Elsevier, 1976.
- [Simmons 73] Simmons, R. F.: "Semantic Networks: Their Computation and Use for Understanding English Sentences", in Schank and Colby (Eds.): "Computer Models of Thought and Language", Freeman and Company, 1973.
- [Steedman 77] Steedman, M. J.: "Time and Modality", Cognitive Science, Vol. 1, No. 2, 1977, 216-234.
- [Steinberg 71] Steinberg, D. D. and Jakobovits, L. A. (Eds): "Semantics", Cambridge University Press, 1971.
- [Stenning 78] Stenning, K.: "Anaphora as an Approach to Pragmatics", in Halle, Bresnan and Miller (Eds.): "Linguistic Theory and Psychological Reality", The MIT Press, 1978, 162-200.
- [Takahasi 72] Takahasi, H., Fujinura, O. and Kameda, H.: "A Behavioral Characterization of Topicalization: What is Topicalization? From a Computer System Point of View", Annual Bull. of RILP, No. 6, University of Tokyo, 1972, 93-110.
- [Tanaka 78] Tanaka, H., Sato, T. and Motoyoshi, F.: "EXPLUS—A Semantic Parsing System for Japanese Sentences", 3rd USA-JAPAN Computer Conference, 1978, 236-240.
- [Tanaka 79] Tanaka, H., Sato, T. and Motoyoshi, F.: "Predictive Control Parser: Extended LINGOL", IJCAI6, 1979.
- [Taylor 75] Taylor, B. H. and Rosenberg, R. S.: "A Case-Driven Parser for Natural Language", AJCL Microfiche 31, 1975.
- [Tulving 72] Tulving, E.: "Episodic and Semantic Memory", in Tulving and Donaldson (Eds.): "Organization of Memory", Academic Press, 1972.
- [Tulving 72] Tulving, E. and Donaldson, W. (Eds.): "Organization of Memory", Academic Press, 1972.

- [Waltz 76] Waltz, D.: "Natural Language Access to a Large Data Base", Computers and People, April, 1976, 19-26.
- [Waltz 77] Waltz, D. L.: "Natural Language Interfaces", SIGART Newsletter, No. 61, Feb. 1977.
- [Wilks 73] Wilks, Y.: "An Artificial Intelligence Approach to Machine Translation", in Schank and Colby (Eds.): "Computer Models of Thought and Language", Freeman and Company, 1973.
- [Wilks 74] Wilks, Y.: "Natural Language Understanding Systems within AI Paradigm—A Survey and Some Comparisons", Stanford AIM-237, Stanford Univ., Dec. 1974.
- [Winograd 71] Winograd, T.: "Understanding Natural Language", Academic Press, 1971.
- [Winograd 75] Winograd, T.: "Frame Representations", in Bobrow and Collins (Eds.): "Representation and Understanding", Academic Press, 1975.
- [Winston 77] Winston, P. H.: "Artificial Intelligence", Addison Wesley, 1977.
- [Woods 71] Woods, W. A.: "Experimental Parsing System for Transition Network Grammar", in Rustin (Ed.): "Natural Language Processing", Algorithmic Press, 1971.
- [Woods 75a] Woods, W. A.: "WHAT'S IN A LINK: Foundations for Semantic Networks", in Bobrow and Collins (Eds.): "Representation and Understanding", Academic Press, 1975.
- [Woods 75b] Woods W. A.: "Syntax, Semantics, and Speech", in Reddy (Ed.): "Speech Recognition", Academic Press, 1975.
- [Woods 77] Woods, W. A.: "Personal View of Natural Language Understanding", SIGART, No. 61, Feb., 1977, 17-20.
- [Yamanashi 72] Yamanashi, M.: "Lexical Decomposition and Implied Proposition", Chicago Linguistic Society, 8, 1972, 388-401.
- [Zampolli 77] Zampolli, A. (Ed.): "Linguistic Structures Processing", North-Holland, 1977.
- [Zloof 75] Zloof, M.: "Query by Example", AFIPS, 44, 1975, 431-438.
- [月刊言語 77a] 月刊言語: "特集・チョムスキー理論の展開", 3月号, 1977.
- [月刊言語 77b] 月刊言語: "特集・象は鼻が長い——日本語の分析", 6月号, 1977.
- [原田 69] 原田信一: "構文論における変項の制限", 英語教育, 8月号~12月号, 1969.
- [池上 75] 池上嘉彦: 「意味論」, 大修館, 1975.
- [池田 77] 池田尚志: "一般化された格構造による意味表現を用いた構文解析法について" 信学論 D, vol 60-D, 10, 1977.
- [井上 76] 井上和子: "変形文法と日本語 (下)", 大修館, 1976.
- [井上 78] 井上和子: "日本語の文法規則", 大修館, 1978.
- [石綿 75] 石綿敏雄: "日本語の生成語彙論的記述と言語処理への応用", 国研報告 54, 1975, 154-198.
- [石綿 79] 石綿敏雄 (編): "計算機による日本語談話行動の総合モデル化", 昭和 53 年度文部省科学研究費研究報告書 (課題番号 310702), 1979 年 1 月.
- [神尾 77] 神尾昭雄: "数値詞のシンタックス", 月刊言語, 6, 8, 1977, 83-91.
- [金田一 76] 金田一春彦 (編): "日本語動詞のアスペクト", むぎ書房, 1976.
- [北原 76] 北原保雄: "助動詞の相互承接についての構文論的考察", 国語学, 83, 1970, 32-59.



- [国立国語研究所 51] 国立国語研究所：“現代語の助詞・助動詞一用法と実例”，秀英出版，1951。
- [国廣 80] 国廣哲彌(編)：“日英語比較講座第2巻文法”，大修館，1980。
- [久野 73] 久野暉：“日本文法研究”，大修館，1973。
- [久野 78] 久野暉：“談話の文法”，大修館，1978。
- [栗原 73] 栗原俊彦：“自然言語の機械処理”，情報処理，14，4，1973，267-281。
- [黒田 76] 黒田成幸：“日本語の論理・思考”，岩波講座日本語1—日本語と国語学，岩波書店，1976。
- [松下 30] 松下大三郎：“標準日本口語法”，白帝社，1930。
- [三上 60] 三上章：“象は鼻が長い”，くろしお出版，1960。
- [水谷 73] 水谷静夫：“機械処理のための日本語文法”，電気学会誌，93，11，1973，995-999。
- [桃内 78] 桃内佳雄：“指示連体詞「その」の文脈的機能に関する一考察”，信学技報，AL78-31，1978。
- [元吉 77] 元吉文男，田中穂積：“未定義語を含む文の構文解析”，情報処理学会 18 回全国大会予稿，No. 95，1977。
- [長尾 74] 長尾真，辻井潤一：“自然言語処理のためのプログラミング言語 PLATON”，情報処理，9，1974，654-661。
- [長尾 76] 長尾真，辻井潤一，田中一敏：“意味および文脈を考慮した日本語文の処理”，情報処理，1，1976，10-28。
- [長尾 77] 長尾真，辻井潤一：“自然言語処理プログラム”，情報処理，18，1，1977。
- [長尾 79] 長尾真，田中穂積(編)，“日本語情報処理特集号”，情報処理，20，10，1979。
- [奥津 74] 奥津敬一郎：“生成日本文法論”，大修館，1974。
- [奥津 78] 奥津敬一郎：“『僕ハウナギダ』の文法”，くろしお出版，1978。
- [雨宮 77] 雨宮真人，島津明：“日本語の意味解釈過程”，電子通信学会論文誌，J60-D，8，1977，633-640。
- [青木 74a] 青木晴夫：“「意味」と言語の構造—文章論への動き”，月刊言語，3，9，1974，793-802。
- [青木 76b] 青木晴夫：“言語理論”，国語学，105，1976，24-35。
- [電総研 78a] 電子技術総合研究所：“LISP User's Manual”，EPICS-5-ON-4，Mar. 31，1978。
- [電総研 78b] 電子技術総合研究所推論機構研究室：“拡張 LINGOL”，Oct.，1978。
- [淵 77] 淵一博，田中穂積：“日本語理解システムへのアプローチ”，パターン情報処理システム講演会論文集，電総研 PIPS-No. 16，1977，145-161。
- [淵 79] 淵一博：“自然言語と計算機——その内的関連を探る——” in 石綿敏雄(編)：“計算機による日本語談話行動の総合モデル化”，昭和 53 年度文部省科学研究費研究報告書(課題番号 310702)，1979 年 1 月，17-20。
- [淵 76] 淵一博：“70 年代の人工知能”，数理科学，9，1976，14-18。
- [藤崎 79] 藤崎哲之助 et al.：“データベース照会システム「ヤチマタ」と名詞句データ模型”，情報処理，20，1，1979，77-84。
- [藤田 76] 藤田毅，鶴丸弘昭，首藤公昭，吉田将：“日本語文の標準形について”，電子通信学会技報，AL 76-41，1976。
- [小野 78] 小野寛明：“名前と名指し関係”，計数国語学，11，4，1978，138-151。

- [坂井 79] 坂井秀寿：“日本語の文法と論理”，勁草書房，1979。
- [佐久間 40] 佐久間鼎：“現代日本語法の研究”，厚生閣，1940。
- [佐藤 76] 佐藤泰介，田中穂積，淵一博：“VISUALIZER—自然言語理解システムの立場からみた機械による空間の把握”，電子通信学会技報，AL 76-11，1976。
- [佐藤 78] 佐藤泰介：“意味構造からの日本語文生成プログラム”，情報処理学会計算言語研究会 17-1，1978。
- [渋谷 79] 渋谷政昭，藤崎哲之助，鷹尾洋一：“日本語によるデータベース照会”，in [長尾 79]。
- [柴谷 78] 柴谷方良：“日本語の分析”，大修館，1978。
- [白井 80] 白井英俊，横尾英俊：“統語処理と意味解釈を同時に行うシステムについて”，情報処理学会，計算言語学研究会資料 21，1980。
- [首藤 77] 首藤公昭，鶴丸弘昭，吉田将：“日英機械翻訳のための述部処理システム”，電子通信学会論文誌，J60-D，10，1977，830-837。
- [首藤 80] 首藤公昭：“文節構造モデルによる日本語の機械処理に関する研究”，福岡大学研究報，第 45 号，1980。
- [鷹尾 76] 鷹尾洋一，藤崎哲之助，間下浩之，諸橋正幸：“日本語によるデータ・ベースの照会”，IBM 東京サイエンティフィック・センター，Dec.，1976。
- [高橋 77] 高橋太郎：“文中にあらわれる所属関係の種々相”，国語学，103，12，1975，1-17。
- [滝浦 77] 滝浦静雄：“時間—その哲学的考察”，岩波新書，960，1977。
- [田中 76] 田中穂積：“談話の理解とメモリ・モデル（その 1-3）”，電総研彙報，第 40 卷，7，9，10，1976。
- [田中 77] 田中穂積，佐藤泰介，元吉文男：“自然言語処理のためのプログラミング・システム—拡張 LINGOL について—”，電子通信学会論文誌，J60-D，12，1977，1061-1068。
- [田中 78a] 田中穂積：“日本語の意味構造を抽出するシステム EXPLUS について”，電子通信学会論文誌，J61-D，8，1978，548-556。
- [田中 78b] 田中穂積，内田ユリ子：“意味抽出における概念階層の利用と手続き付加”，情報処理学会人工知能と対話技法，8-1，1978。
- [田中 79] 田中穂積：“計算機による自然言語の意味処理に関する研究”，電子技術総合研究所研究報告 No. 797，1979。
- [鶴丸 76] 鶴丸弘昭，藤田毅，首藤公昭，吉田将：“日本語の機械処理”，電子通信学会技報，AL 76-43，1976。
- [寺津 78a] 寺津典子：“T. Reinhart (1976) : The Syntactic Domain of Anaphora”，メモ，1978。
- [寺津 79b] 寺津典子，山梨正明：“日本語における照応現象について（その一）”，in 石綿敏雄（編）：“計算機による日本語談話行動の総合モデル化”，昭和 54 年 1 月。
- [寺村 71] 寺村秀夫：“一タの意味と機能—アスペクト，テンス，ムードの構文的位置づけ”，in “言語学と日本語問題”，くろしお出版，1971。
- [安井 78] 安井稔：“言外の意味”，研究社，1978。
- [安井 79] 安井稔：“新しい聞き手の文法”，研究社，1979。
- [山田 78] 山田尚勇：“日本語テキスト入力法の人間工学的比較”，日本語情報処理シンポジウム報告集，1978。
- [山梨 79] 山梨正明：“自然言語の意味構造—含意関係からの再考察” in 長尾真（編）：“機械処理関係シンポジウム・辞書ファイル委員会活動報告書”，昭和 54 年 3 月。
- [吉田 78] 吉田将，首藤公昭，藤田毅：“日本語文における曖昧さと関係表現について”，電気四学会連合大会，昭 53・10，6-79~82。

- [Aho 72] Aho et al : The theory of parsing, translation and compiling Vol. 1, Préntice-Hall, Inc. (1972).
- [Hopcroft 69] Hopcroft et al. : Formal language and their relation to automata, Addison-Wesley (1969).
- [Kuno 65] Kuno, S. : The predictive analyzer and a path elimination technique, CACM, 8, (1965).
- [Zwicky 65] Zwicky et al : The MITRE syntactic analysis procedure for transformational grammars, AFIPS, FJCC, Vol. 27 (1965).
- [Woods 70] Woods, W. A. : An experimental parsing system for transition network grammar, Natural language processing, in Rustin (ed.), Algorithmic Press (1970).
- [Woods 78] Woods, W. A. : Semantics and quantification in natural language question answering, Advances in computers, Vol. 17, Academic press (1978).
- [Heidorn 78] Heidorn, G. E. : Augmented phrase structure grammar, Theoretical issues in natural language processing, in Schank and Webber (eds.), Cambridge, Mass (1978).
- [Winograd 72] Winograd, T. : Procedures as a representation for data in computer programs for understanding natural language Academic Press (1972).
- [Pratt 75] Pratt, V. R. : LINGOL-A progress report 41JCAI (1975).
- [Grishman 76] Grishman, R. : A survey of syntactic analysis procedures for natural language, AJCL, microfiche 47 (1976).
- [Thompson 75] Thompson et al. : Practical natural language processing : The RIEL system as a prototype, Advances in computers, Vol. 13, Academic Press (1975).
- [Pereira 80] Pereira, F. C. N. and Warren, H. D. : Definite clause grammars for language analysis A survey of the formalism and a comparison with augmented transition Networks, Artificial Intelligence, 13 (1980), 231-278.
- [奥野 74] 奥野敬一郎 : 生成日本文法論, 大修館 (1974)
- [安井 78] 安井稔 : 新しい聞き手の文法, 大修館 (1978)
- [北原 70] 北原保雄 : 助動詞の相互承接についての構文論的考察, 国語学, 83 (1970).
- [西村 78] 西村他 : 日本語基本文法単文篇, 電総研い報 No. 783 (1978).
- [吉田 72a] 吉田 将 : 二文節間の係り受けを基礎とした日本語の構文分析, 信学論 D, Vol. 55-D, No. 4 (1972).
- [吉田 72b] 吉田他 : 日本語の機械処理——構文分析から意味分析へ——, 信学論 D, Vol. 55-D, No. 6 (1972).
- [長尾 76a] 長尾他 : 意味および文脈処理を用いた日本語文の解析——名詞句・単文の処理——, 情報処理, Vol. 17, No. 1 (1976).
- [長尾 76b] 長尾他 : 意味および文脈処理を用いた日本語文の解析——文脈を考慮した処理——, 情報処理, Vol. 17, No. 1 (1976).
- [石綿 76] 石綿敏雄 : 変形とその逆探知を含む構文分析, 電子計算機による国語研究は, 国研報告 59 (1976).
- [石原 74a] 石原他 : D-tree モデルとそれに基づく英日機械翻訳の為の言語分析について, 信学論 D, Vol. 57-D, No. 7 (1974).

- [石原 74b] 石原他：D-tree モデルに基づく一つの英日機械翻訳システムおよび実験，信学論 D, Vol. 57-D, No. 7 (1974).
- [雨宮 77] 雨宮他：図形世界を話題とした質問解答システム，情報処理，Vol. 18, No. 8 (1977).
- [島津 78] 島津 明：日本語の構文・意味解析——質問応答システム MSSS 78 での試み——，信学会，AL 研究会，AL 78-94 (1978).
- [池田 77] 池田尚志：一般化された格構造による意味表現を用いた構文解析法について，信学論 D, Vol. 60-D, No. 10 (1977).
- [池田 78] 池田尚志：語の構造・意味記述と文の構造・意味分析，信学会，AL 研究会，AL 78-32 (1978).
- [諸橋 76] 諸橋他：ヤチマタにおける擬似日本語，情報処理，CL 研究会，CL-8 (1976).
- [Fillmore 75] Fillmore, C. J. 著，田中春美他訳：格文法の原理——言語の意味と構造——，三省堂(1975).
- [Hays 64] Hays, D. G. : Dependency theory : a formalism and observations, Language, 40, 4 (1964).
- [Kay 71] Kay, M. : The MIND system, Natural Language Processing, in Rustin (ed.) Algorithmic Press (1971).
- [Colmerauer 78] Colmerauer : Metamorphosis Grammars, in Natural Language Communication with Computer, Lectures Notes in Computer Science, Springer Verlag (1978).
- [淵 76] 淵 一博：定理証明としてみた Earley/Pratt のパージング・アルゴリズム，情報処理学会 17 回全国大会 (1976).
- [西岡 76] 西岡，その他：導出原理を用いた構文解析システム信学会 AL 研究会 AL 76-74 (1976).
- [首藤 79] 首藤，その他：日本語の機械処理のための文節構造モデル信学論 D Vol. J 62-D, No. 12, (1979).
- [Schank 79] Schank et al : Parsing Directly into Knowledge Structures, 6th IJCAI (1979).
- [Rieger 79] Rieger et al : Word Expert Parsing, 6th IJCAI (1979).
- [Andreka 76] Andreka, H. and Nemeti, I., "The Generalised Completeness of Horn Predicate-Logic as a Programming Language", D. A. I. Research Report No. 21, University of Edinburgh, March, 1976.
- [Barr 79] Barr, A., "Meta-knowledge and Cognition", IJCAI-6, pp. 31-33, 1979.
- [Baudisch 80] Baudisch, A., Seese, D., Tuschik, H.-P. and Weese, M. : Decidability and Generalized Quantifiers, Akademie-Verlag, Berlin, 1980.
- [Chang 73] Chang, C. L. and Lee, C. T. : Symbolic Logic and Mechanical Theorem Proving, New York, Academic Press, 1973.
- [Clark 78] Clark, K., "Negation as Failure", in 'Logic and Data Bases' Plenum Press, pp. 293-322, 1978.
- [Davis 77a] Davis, R. and Buchanan, B. G., "Meta-Knowledge : Overview and Application," IJCAI-5, pp. 920-927, 1977.
- [Davis 77b] Davis, R. and King, J., "An Overview of Production Systems", in 'Machine Intelligence 8' E. Elcock and D. Michie (eds.) pp. 300-332, 1977.

- [Deliyanni 79] Deliyanni, A. and Kowalski, R. A., "Logic and Semantic Networks", CACM, vol. 22, no. 3, pp. 184-192, Mar. 1979.
- [Dowty 78] Dowty, D. R., "A Guide to Montague's PTQ", Indiana University Linguistics Club, Dec. 1978.
- [Fahlman 79] Fahlman, S. E., "NETL: A System for Representing and Using Real-World Knowledge," MIT Press, Cambridge, Mass., 1979.
- [Fahlman 80] Fahlman, S. E., "Design Sketch for a Million-Element NETL Machine," The 1st Annual National Conference on Artificial Intelligence, Aug. 1980.
- [Feigenbaum 71] Feigenbaum, E. A., Buchanan, B. G. and Lederburg, J., "On Generality and Problem Solving: A Case Study Using DENDRAL Program", in 'Machine Intelligence 6', 1971.
- [Findler 79] Findler, N. V. (ed.): Associative Networks, Academic Press, 1979
- [Gallin 75] Gallin, D.: Intensional and Higher-Order Modal Logic, North-Holland, 1975.
- [Gentner 80] Gentner, D. and Collins, A. M., "Knowing about Knowing: Effects of Meta-Knowledge on Inference", Memory and Cognition, 1980.
- [Hendrix 79] Hendrix, G. G., "Encoding Knowledge in Partitioned Networks", in N. V. Findler (ed.) 'Associative Networks' pp. 51-92, 1979.
- [Hewitt 72] Hewitt, C., "Description and Theoretical Analysis (using schemata) of PLANER, a Language for Proving Theorems and Manipulating Models in a Robot", MIT, AI-TR-258, 1972.
- [Hughes 68] Hughes, G. E. and Cresswell, M. J.: An Introduction to Modal Logic, London, 1968.
- [Janas 79] Janas, J. M. and Schwind, C. B., "Extensional Semantic Networks: Their Representation, Application, and Generation", in N. V. Findler (ed.) 'Associative Networks' pp. 267-302, 1979.
- [McCarthy 80] McCarthy, J., "Circumscription-A Form of Non-Monotonic Reasoning", Artificial Intelligence, vol. 13, no. 1, 2, pp. 27-39, Apr. 1980.
- [Montague 74] Montague, R.: Formal Philosophy, R. Thomason (ed.) Yale Univ. Press, 1974.
- [Moore 75] Moore, R. C., "Reasoning from Incomplete knowledge in a Procedural Deduction System", MIT, AI-TR-347, Dec. 1975.
- [Nilsson 71] Nilsson, N. J.: Problem-Solving Methods in Artificial Intelligence, New York, McGraw-Hill, 1971.
- [Nilsson 80] Nilsson, N. J.: Principles of Artificial Intelligence, Tioga Publishing Co., 1980.
- [Partee 75] Partee, B., "Montague Grammar and Transformational Grammar", Linguistic Inquiry 6, 2 1975.
- [Post 43] Post, E., "Formal Reductions of the General Combinatorial Problem", American Journal of Mathematics, 65, pp. 197-268, 1943.
- [Reiter 80] Reiter, R., "A Logic for Default Reasoning", Artificial Intelligence, vol. 13, no. 1, 2, pp. 81-132, Apr. 1980.

- [Roberts 77a] Roberts, D. B. and Goldstein, I. P., "The FRL Primer", MIT AI memo 408, 1977.
- [Robinson 65] Robinson, J. A., "A Machine Oriented Logic Based on the Resolution Principle", J. ACM, 12, pp. 23-41, Jan. 1965.
- [Sowa 80] Sowa, J. F., Book Reviews in American Journal of Computational Linguistics, vol. 6, no. 2, pp. 110-112, April-June, 1980.
- [辻井 79] 辻井潤一, "プロダクションシステムとその応用", 情報処理学会誌 vol. 20, no. 8, pp. 735-743, 1979.
- [Waterman 78] Waterman, D. A. and Hayes-Roth, F.: Pattern-Directed Inference Systems Academic Press, 1978.
- [Winograd 72] Winograd, T.: Understanding Natural Language, New York, Academic Press, 1972.
- [Winograd 80] Winograd, T., "Extended Inference Modes in Reasoning by Computer Systems", Artificial Intelligence, vol. 13, no. 1, 2, pp. 5-26, Apr. 1980.
- [Zadeh 65] Zadeh, L. A., "Fuzzy Sets," Information and Control 8, pp. 338-353, 1965.
- [Zadeh 79] Zadeh, L. A., "A Theory of Approximate Reasoning", in 'Machine Intelligence' 9, pp. 149-194, 1979.
- [Yngve 62] Yngve, V.: "Random generation of English sentences," 1961 International Conference on Machine Translation of Language and Applied Language Analysis. National Physical Laboratory No. 13, 1962.
- [Sakai 65] Sakai, T., Nagao, M.: "Sentence Generation by Semantic Concordance," ICCL, 1965.
- [Friedman 69] Friedman, J.: "Directed Random Generation of Sentences," CACM vol. 12 No. 1, 1969.
- [Self 75] Self, J.: "Computer Generation of Sentences by Systemic Grammar," AJCL, Vol. 12-5 microfiche 29, 1975.
- [Clippinger 75] Clippinger, J. H.: "Speaking with Many Tongues: Some Problems in Modeling Speakers," TINLAP1, 1975.
- [Snell 79] Snell, B. M. (ed): "Translating and The Computer," North-Holland, 1979.
- [Thompson 77] Thompson, H.: "Strategy and Tactics: A Model for Language Production," CLS apr. 14-16, 1977.
- [Weizenbaum 66] Weizenbaum, J.: "ELIZA-A Computer Program for the Study of Natural Language Communication Between Man And Machine," CACM Vol. 9 No. 1, 1966.
- [Winograd 70] Winograd, T.: "Procedures as a Representation for Data in a Computer Program for Understanding Natural Language," Ph. D. Thesis MIT, 1970.
- [Shortliffe 76] Shortliffe, E. H.: "Computer-Based Medical Consultations: MYCIN," American Elsevier, 1976.
- [Scott 77] Scott, A. C. et al.: "Explanation Capabilities of Production-Based Consultation Systems," CSD report No STAN-CS-77-593 Stanford univ, 1977.

- [Codd 78] Codd et al.: "Rendezvous Version 1: An Experimental English-Language Query Formulation System for Casual Users of Relational Data Base," IBM Research Lab., 1978.
- [McKeown 80] McKeown, K. R.: "Generating Relevant Explanations: Natural Language Responses to Questions about Data Base Structure," proc. of NCAI, AAAI, 1980.
- [Simmons 72] Simmons, R. & Slocum, J.: "Generating English Discourse from Semantic Networks, CACM Vol. 15 No. 10, 1972.
- [Slocum 75] Slocum, J.: "Speech Generation from Semantic Nets," AJCL microfiche 33 : 63, 1975.
- [Slocum 78] Slocum, J.: "Generating a Verbal Response", Understanding Spoken Language, Elsevier North-Holland, 1978.
- [Shapiro 76] Shapiro, S. C.: "Generation as Parsing from a Network into a Linear String," AJCL microfiche 33 : 45, 1976.
- [Goldman 74a] Goldman, N. M.: "Computer Generation of Natural Language from a Deep Conceptual Base," STAN-AIM-247 Stanford univ., 1974.
- [Goldman 74b] Goldman, N. M. "Sentence Paraphrasing from a Conceptual Base," CACM Vol. 2 No. 18, 1974.
- [Goldman 75] Goldman, N. M.: "The Boundaries of Language Generation," TINLAP-1, 1975
- [Wang 80] Wang, J.: "On Computational Sentence Generation from Logical Form," ICCL-8, 1980.
- [McDonald 75a] McDonald, D.: "Preliminary Report on a Program for Generating Natural Language," IJCAI-4, 1975.
- [McDonald 75b] McDonald, D.: "A Framework for Generation Grammars for Interactive Computer Programs," AJCL, microfiche 33 : 4, 1975.
- [McDonald 80] McDonald, D.: "A Linear-time Model of Language Production: Some Psychological Implications," ACL-18, 1980.
- [佐藤 79] 佐藤泰介: "意味構造からの日本語文生成プログラムについて," 計算言語学 17-1, 1979.
- [横山 78] 横山品一, 田中穂積: "意味構造からの日本語文の再構成," 第 19 回情報処理全国大会, 1978.
- [Nagao 80] Nagao, M. et al.: "A Machine Translation System from Japanese into English" ICCL-8, 1980.
- [Boitet 80] Boitet, Ch. & Nedobekine, N.: "Russian-French at GETA: Outline of The method and Detailed example," ICCL-8, 1980.
- [Vauquois 79] Vauquois, B.: "Discussion about Software Tools for M. T. and Natural Language Processing", 情報処理学会, CL 研究会 18-2, 1979.
- [JIPDEC 80] "第 5 世代の電子計算機に関する調査研究報告書——言語の機械翻訳とその関連技術," 日本情報処理開発協会, 1980.
- [Meehan 75] Meehan, J. R.: "Using Planning Structures to Generate Stories," AJCL Vol. 12 No. 5 microfiche 33 : 78, 1975.
- [Kahn 79] Kahn, K. M.: "Making Aesthetic Choices," AI # 513, MIT AI Lab., 1979.

- [Schank 73b] Schank, R.C. & Colby, K.M.(ed.) "Computer Models of Thought and Language," W. H. Freeman and Company San Francisco, 1973.
- [Hutchins 71] Hutchins, W. J.: "The Generation of Syntactic Structures from a Semantic Base," North-Holland, 1971.
- [Huber 76] Huber, F.: "On the Generation of English Sentence" IEEE Trans. of Computers 25 : 90-91, 1976.
- [Gradner 79] Gradner, A. et al.: "Natural Language Understanding," STAN-CS-97-754 Stanford univ., 1979.
- [Bruce 75] Bruce, B. C.: "Generation as a Social Action," TINLAP-1, 1975.
- [Appelt 80] Appelt, D. E.: "Problem Solving Applied to Language Generation," ACL-18, 1980.
- [首藤 73] 首藤公昭: "専門分野を対象とした日英機械翻訳について," 情報処理 Vol. 10 No. 9, 1973.
- [石原 78] 石原公宏, 田町常夫: "日本語終端表現生成システムの作成と実験," 信学論 D Vol. J61-D, No. 9, 1978.
- [長尾 81] 長尾, その他: "計算機マニュアルを対象とした日英機械翻訳システムにおける英文の生成," 情報処理学会 計算言語学研究資料 25, 1981.
- [Bobrow 77] Bobrow, D. G. et al.: "GUS, A Frame-Driven Dialog System," Artificial Intelligence 8, 1977.
- [佐伯 75] 佐伯哲夫: "現代日本語の語順", 笠間書院, 1975.
- [佐伯 76] 佐伯哲夫: "語順と文法", 関西大学出版, 1976.
- [久野 73] 久野 暉: "日本文法研究", 大修館, 1973.
- [久野 78] 久野 暉: "談話の文法", 大修館, 1978.
- [国広(編) 80] "日英比較講座 第2巻 文法", 大修館, 1980.
- [井上 79] 井上 和子: "新しい情報・新しい情報" 言語, Vol. 8, No. 2 大修館, 1979.
- [井上 80] 井上 和子: "格助詞をめぐって", 言語, Vol. 9, No. 2 1980.
- [安藤 80] 安藤 貞雄: "日本語の「ハ」と「ガ」の機能について", 言語, Vol. 9, No. 7, 1980.
- [尾上 81] 尾上圭介: "「象は鼻が長い」と「ぼくはウナギだ」", 言語, Vol. 10, No. 2, 1981.
- [大野 77] 大野 晋: "日本語の文法を考える", 岩波新書, 1977.
- [Clark 78] Clark, K.: "Negation as Failure," in 'Logic and Data Bases, Plenum Press, pp. 293-322(1978).
- [Davis 80] Davis, M.: "The Mathematics of Non-Monotonic Reasoning," Artif. Intell. 13, 1/2, pp. 73-80 (April 1980)
- [Doyle 78] Doyle, J.: "Truth Maintenance System for Problem Solving," MIT AI-TR-419 (1978)
- [Doyle 79] Doyle, J.: "A Glimpse of Truth Maintenance," 6th IJCAI, pp. 232-237 (1979)
- [Doyle 80] Doyle, J. and London, P.: "A Selected Descriptor Indexed Bibliography to the Literature on Belief Revision," SIGART Newsl., 71, pp. 7-23 (April 1980)
- [Hayes 73] Hayes, P. J.: "The Frame Problem and Related Problems in Artificial Intelligence," in 'Artificial and Human Thinking' Elithorn, A. and Jones, D. (eds.), Jossey-Bass Inc., pp. 45-59(1973)



- [Israel 80] Israel, D. J.: "What's Wrong with Non-Monotonic Logic," Proc. 1st Annual National Conference on Artificial Intelligence, pp. 99-101 (Aug. 1980)
- [McCarthy 69] McCarthy, J. and Hayes, P. J.: "Some Philosophical Problems from the Standpoint of Artificial Intelligence," in 'Machine Intelligence 4' Meltzer, B. and Michie, D. (eds.) Edinburgh University Press, pp. 463-502 (1969)
- [McCarthy 80b] McCarthy, J. "Addendum: circumscription and other non-monotonic formalisms," *ibid.*, pp. 171-172.
- [McDermott 80a] McDermott, D. and Doyle, J.: "Non-Monotonic Logic I," *ibid.*, pp. 41-72.
- [McDermott 80b] McDermott, D.: "Non-Monotonic Logic II," Research Report 174, Yale Univ. Dept. of Computer Science, (Feb. 1980)
- [Nilsson 80] Nilsson, N. J.: "Defaults and Contradictory Information," in 'Principles of Artificial Intelligence', pp. 408-411, Tioga Publishing Co., (1980)
- [Raphael 71] Raphael, B.: "The Frame Problem in Problem Solving System," in 'Artificial Intelligence and Heuristic Programming', N. V. Findler and B. Meltzer (eds.), Edinburgh University Press, pp. 159-169, (1971)
- [Reiter 78a] Reiter, R.: "On Reasoning by Default," 2nd Symposium of Theoretical Issues in Natural Language Processing, pp. 210-218 (July 1978)
- [Reiter 78b] Reiter, R.: "On Closed World Data Base," in 'Logic and Data Bases' Plenum Press, pp. 55-76 (1978)
- [Sandewall 72] Sandewall, E.: "An Approach to the Frame Problem and its Implementation," in 'Machine Intelligence 7', Ellis Horwood, pp. 195-204 (1972)
- [Sussman 70] Sussman, G., Winograd, T. and Charniak, E.: MICRO-PLANNER Reference Manual, MIT AIM-203 (1970)
- [Weyhrauch 80] Weyhrauch, R. W.: "Prolegomena to a Theory of Mechanized Formal Reasoning, Artif. Intell., 13, 1/2, pp. 133-170 (April 1980)
- [Boden 77] Boden, M., "Artificial Intelligence and Natural Man", Basic Books, 1977.
- [Cofer 76] Cofer, C. N. (Ed.), "The Structure of Human Memory", Freeman and Company, 1976.
- [Halliday 76] Halliday, M. and Hassan, R., "Cohesion in English", Longman, London, 1976.
- [Newell 73] Newell, A. et al.: "Speech Understanding Systems—Final Report of a Study Group", North Holland / American Elsevier, 1973.
- [Ogden 23] Ogden, C. K. and Richards, I. A., "The Meaning of meaning", London: Kegan Paul, 1923.
- [Searle 69] Searle, J. R., "Speech Acts: An Essay in the Philosophy of Language", Oxford Univ. Press, 1969.
- [Walker 78] Welker, D., "Understanding Spoken Language", North-Holland, 1978.
- [Frederiksen 75] Frederiksen, C. H., "Representing Logical and Semantic Structure of Knowledge Acquired from Discourse", Cognitive Psychology, 7, 1975, 371-458.
- [Hobbs 79] Hobbs, J. R., "Coherence and Coreference", Cognitive Science 3, 1979, 67-90.

- [Kahn 75] Kahn, K. M., "Mechanization of Temporal Knowledge", MIT MAC-TR-155, Aug. 1975.
- [Kuno 77] Kuno, S. and Kaburaki, E., "Empathy and Syntax", *Linguistic Inquiry*, Vol. 8., No. 4, 1977.
- [Lakoff 76] Lakoff, G., "Pronouns and Reference", in McCawley, J. D. (Ed.), "Syntax and Semantics", Vol. 7, Academic Press, 1976, 275-335.
- [Lehnert 77] Lehnert, W., "Human and Computational Question Answering", *Cognitive Science*, Vol 1, No 1, 1977, 47-73.
- [McDermott 74] McDermott, D. U., "Assimilation of New Information by a Natural Language-Understanding System", AI TR-291, MIT, Feb. 1974.
- [Moore 74] Moore, J. and Newell, A., "How Can Merlin Understand?", in Gregg (Ed.), "Knowledge and Cognition", Lawrence Erlbaum Associates, Publishers, Potomac, Potomac, Maryland, 1974.
- [Nash-Webber 78] Nash-Webber, B. L., "A Formal Approach to Discourse Anaphora", BBN Report No. 3761, May 1978.
- [Winograd 77] Winograd, T., "Understanding Natural Language", Academic Press, 1977.
- [Abelson 73] Abelson, R. P., "The Structure of Belief Systems", in Schank and Colby (Eds.), "Computer Models of Thought and Language", Freeman and Company, 1973.
- [Bien 76] Bien, J. S., "Multiple Environments, Approach to Natural Language", *AJCL*, Microfiche 54, 1976.
- [Bullwinkle 77] Bullwinkle, C., "Levels of Complexity in Discourse Anaphora Disambiguation and Speech Act Interpretation", *IJCAI* 5, 1977, 43-49.
- [Charniak 73] Charniak, E., "Jack and Janet in Search of a Theory of Knowledge", *IJCAI* 3, 1973, 337-334.
- [Charniak 77] Charniak, E., "Ms. MALAPROP, A Language Comprehension Program", *IJCAI* 5, 1977, 1-7.
- [Norman 73] Norman, D., "Memory, Knowledge and the Answering of Questions" in Solso, R. L. (Ed), "Contemporary Issues in Cognitive Psychology, The Loyola Symposium", V. II. Winston and Sons, Washington, D. C., 1973.
- [Phillips 79] Phillips, B.: "A Model for Knowledge and its application to Discourse Analysis", *AJCL*, Microfiche 82, 1979.
- [Rumelhart 75] Rumelhart, D. E., "Notes on a Schema for Stories", in Bobrow and Collins (Eds.), "Representation and Understanding", Academic Press, 1975.
- [Schank 75] Schank, R. C. et al, "SCRIPT, PLANS and KNOWLEDGE", *IJCAI* 4, 1975, 255-261.

- [Schank 76] Schank, R. C.: "The Role of Memory in Language Understanding", in Cofer, C. N. (Ed.), "The Structure of Human Memory, Freeman and Company, 1976, 162-189.
- [Sidner 78] Sidner, C., "A Progress Report on the Discourse and Reference Components of PAL", MIT AI Memo 468, 1978.
- Annual Bull. No. 6, Research Institute of Logopedics and Phoniatrics, University of Tokyo, 1972, 93-110.
- [Winograd 77] Winograd, T., "A framework for Understanding Discourse", Stanford AIM-297, April 1977.
- [Young 78] Young, R., "Text Understanding, A Survey", AJCL, Microfiche 70, 1973.
- [Tanaka 78] 田中穂積, 内田ユリ子  
"意味抽出における概念階層の利用と手続き付加", 人工知能と対話技法 8-1, 1978.
- [Terazu 79] 寺津典子, 山梨正明  
"日本語における照応現象について (その1)",  
文部省科学研究費特定研究(1)課題番号 310702
- [Terazu 80] 寺津典子, 稲田俊明, 山梨正明  
"日本語における照応現象について (その2)", 1979  
文部省科学研究費特定研究(2)課題番号 410202, 1980
- [Uchida 79] 内田ユリ子, 田中穂積  
"「ソノ」の分析と理解のモデル (その1)", 情報処理学会第20回全国大会報告書, 1979.
- [Alter 68] Alter, R., "Utilization of Contextual Constraints in Automatic Speech Recognition", "IEEE. Trans. vol. AU-16, No. 1, pp. 6-11 (Mar. 1968)
- [Bahl 76] Bahl, L., Baker, J., Cohen, P., Dixon, N., Jelinek, F., Mercer, R., and Silverman, H., "Preliminary Results on the Performance of a System for the Automatic Recognition of Continuous Speech", IEEE ICASSP' 76, pp. 425-429 (Apr. 1976).
- [Bahl 78a] Bahl, L. R., Baker, J. K., Cohen, P. S., Cole, A. G., Jelinek, F., Lewis, B. L. and Mercer, R. L., "Automatic Recognition of Continuously Spoken Sentences from a Finite State Grammar", Proc. of IEEE ICASSP' 78, pp. 418-421 (Apr. 1978)
- [Bahl 78b] Bahl, L. R., Baker, J. K., Cohen, P. S., Jelinek, F., Lewis, B. L. and Mercer, R. L., "Recognition of a Continuously Read Natural Corpus", Proc. of IEEE ICASSP' 78, pp. 422-424 (Apr. 1978)
- [Bahl 79] Bahl, L. R., Bakis, R., Cohen, P. S., Cole, A. G., Jelinek, F., Lewis, B. L. and Mercer, R. L., "Recognition Results with Several Experimental Acoustic Processors", IEEE ICASSP' 79, pp. 249-251 (Apr. 1979)
- [Baker 75] Baker, J., "The Dragon System - An Overview", IEEE Trans. Acoust. Speech Signal Process., vol. ASSP-23, No. 1, pp. 24-29. (Feb. 1975)
- [Baker 79] Baker, J. M., "Performance Statistics of the HEAR Acoustic Processor", IEEE ICASSP' 79, pp. 262-265 (Apr. 1979)
- [Barnett 73] Barnett, J. A., "A Vocal Data Management System", IEEE, Trans. AU, vol. AU-21, No. 3, pp. 185-188 (Jun. 1973)

- [Barnett 74] Barnett, J. A., "A Phonological Rule Compiler", pp. 188-192 in Erman (Apr. 1974).
- [Bates 74] Bates, M., "The Use of Syntax in a Speech Understanding System", IEEE Symp. Speech Recognition, pp. 226-233 (Apr. 1974)
- [Bates 75] Bates, M., "The Use of Syntax in a Speech Understanding System", IEEE Trans. Acoust. Speech Signal Process., vol. ASSP-23, No. 1, pp. 112-117. (Feb. 1975)
- [Beck 77] Beck, B., Neuberg, E. P., and Hodge, D. C., "An Assessment of the Technology of Automatic Speech Recognition for Military Applications", IEEE Trans. ASSP, vol. ASSP-25, No. 4, pp. 310-322 (Aug. 1977)
- [Christiansen 79] Christiansen, R. W. and Rushforth, C. K., "Word Spotting in Continuous Speech Using Linear Predictive Coding", IEEE ICASSP' 79, pp. 557-560 (Apr. 1979)
- [Cohen 75] Cohen, P. S., and Mercer, R. L., "The Phonological Component of an Automatic Speech Recognition System", pp. 275-320 in 'Speech Recognition' edited by Reddy (Academic pr.) (1975).
- [Cook 77] Cook, C. C. and Schwartz, R. M., "Advanced Acoustic Techniques in Automatic Speech Understanding", IEEE ICASSP' 77, pp. 663-666 (May 1977).
- [Cook 79] Cook, C., "Word Verification in a Speech Understanding System", IEEE ICASSP' 79, pp. 553-556 (Apr. 1979)
- [Denes 59] Denes, P. B., "The Design and Operation of the Mechanical Speech Recognizer at University College London", Journal of Brit. IRE, vol. 19, No. 4, pp. 219-234 (1959)
- [Diller 79] Diller, T. C., "Phonetic Word Verification", IEEE ICASSP'79, pp. 256-261 (Apr. 1979)
- [Erman 74] Erman, L. D., 'Contributed Papers of the IEEE Symposium on Speech Recognition' (IEEE Catalogue No. 74CH0878-9AE) (Apr. 1974)
- [Fant 70] Fant, G., "Automatic Recognition and Speech Research", STL-QPSR 1, No. 1, pp. 16-31 (1970)
- [Gilligly 72] Gilligly, J. J., "The TECHNOLOGY chess program," Artif. Intel., vol. 3, pp. 145-163 (1972)
- [Glave 78] Glave, R. D. and van der Giet, G., "The DAWID Speech Recognition System", IEEE ICASSP' 78, pp. 429-432 (Apr. 1978)
- [Goodman 76] Goodman, G., "Analysis of Language for Man-Machine Voice Communication", Ph. D dissertation (CMU), (1976)
- [Halle 62] Halle, M. and Stevens, K., "Speech Recognition: A Model and a Program for Reserch", IRE Trans. of Inf. Theory IT-8, pp. 155-159 (Feb. 1962).
- [Haton 76] Haton, J. P., and Pierrel, J. M., "Organization and Operation of a Connected Speech Understanding System at Lexical, Syntactic and Semantic Level", IEEE ICASSP' 76, pp. 430-433 (Apr. 1976).
- [Haton 79] Haton, J. and Morel, O., "Automatic Recognition of Connected Digits Sequences by Means of Segmentation and Dynamic Programing", IEEE ICASSP' 79, pp. 245-248 (Apr. 1979)
- [Hayes-Roth 76a] Hayes-Roth, F., and Lesser, V. R., "Focus of Attention in a Distributed Logic Speech Understanding System", IEEE ICASSP' 76, pp. 416-420 (Apr. 1976).

- [Hayes-Roth 76b] Hayes-Roth, F. and Mostow, D. J., "Syntax and Semantics in a Distributed Speech Understanding System", IEEE ICASSP' 76, pp. 421-424 (Apr. 1976)
- [Hyde 72] Hyde, S. R., "Automatic Speech Recognition: A Critical Survey and Discussion of the Literature", in 'Human Communication: A Unified View', edited by E. E. David and P. B. Denes (McGraw-Hill, New York) (1972)
- [Itahashi 73] S. Itahashi, S. Makino and K. Kido, "Discrete-Word Recognition Utilizing a Word Dictionary and Phonological Rules," IEEE Trans. vol. AU-21, 3 (1973) pp. 239-249.
- [Jaschul 79] Jaschule, J., "An Approach to Speaker Normalization for Automatic Speech Recognition", IEEE ICASSP' 79, pp. 235-238 (Apr. 1979)
- [Klatt 72] Klatt D. H. and Stevens K. N., "Sentence recognition from visual examination of spectrograms and machine-aided lexical searching," in Proc. Int. Conf. Speech Commun. Processing, pp. 315-318 (1972)
- [Klatt 77] Klatt, D. H., "Review of the ARPA Speech Understanding Project", JASA, vol. 62, No. 6, pp. 1345-1366, (Dec. 1977)
- [Klovstad 74] Klovstad, J. W. and Mondschein, L. F., "The Caspers Linguistic Analysis System", IEEE Symp. Speech Recognition, pp. 234-240 (Apr. 1974)
- [Klovstad 75] Klovstad, J. W. and Mondschein, L. F., "The CASPERS Linguistic Analysis System", IEEE Trans. ASSP, vol. ASSP-23, No. 1, pp. 118-123 (Feb. 1975)
- [Kohda 79] Kohda, M., Nakatsu, R. and Shikano, K., "Speech Recognition in the Question-Answering System Operated by Conversational Speech", IEEE ICASSP' 79, pp. 442-445 (Apr. 1979)
- [Lea 73] Lea, W. A., "An Approach to Syntactic Recognition Without Phonemics", IEEE Trans. AU, vol. AU-21, No. 3, pp. 249-258 (Jun. 1973)
- [Lea 75] Lea, W. A., Medress, M. F. and Skinner, T. E., "A Prosodically Guided Speech Understanding Strategy", IEEE Trans. ASSP, vol. ASSP-23, No. 1, pp. 30-38 (Feb. 1975)
- [Lea 78a] Lea, W. A. and Shoup, J. E., "Recommendations for Advancing Recognition", 95th Meeting of ASA, (May 1978)
- [Lea 78b] Lea, W. A. and Shoup, J. E., "Gaps in the Technology of Speech Understanding", Proc. of IEEE ICASSP' 78, pp. 405-408 (Apr. 1978)
- [Lesser 75] Lesser, V. R., Fennell, R. D., Erman, L. D. and Reddy, D. R., "Organization of the Hearsay-II Speech Understanding System", IEEE Trans. Acoust. Speech Signal Process., vol. ASSP-23, No. 1, pp. 11-23. (Feb. 1975).
- [Levinson 79] Levinson, S. E. and Rosenberg, A. E., "A New System for Continuous Speech Recognition - Preliminary Results", IEEE ICASSP'79, pp. 239-244 (Apr. 1979)
- [Lowerre 76] Lowerre, B. T., "The HARPYP Speech Recognition System", CMU WPSR (Apr. 1976)
- [Mariani 79] Mariani, J. J., Lienard, J. S. and Renard, G., "Speech Recognition in the Context of Two-Way Immediate Person-Machine Interaction", IEEE ICASSP'79, pp. 269-272 (Apr. 1979)
- [Medress 78] Medress, M. F. et al, "Speech Understanding Systems-Report of a Steering Committee", Art. Intell. vol. 9, pp. 307-316 (1978).

- [Mercier 78] Mercier, G., Quinton, P. and Vives, R., "Dialogue Homme - Machine avec KEAL", *Reserches/Acoustique CNET*, vol. 4, pp. 187-206 (1977)
- [Miller 73] Miller, P. L., "A Locally-Organized Parser for Spoken Input", Ph. D thesis, MIT, (Mar. 1973)
- [Moor 77] Moor, R. K., "Evaluating Speech Recognizers", *IEEE Trans. ASSP*, vol. ASSP-25, No. 2, pp. 178-183 (Apr. 1977)
- [Nakagawa 76] Nakagawa, S., "A Machine Understanding System for Spoken Japanese Sentences", Ph. D thesis Kyoto Univ., (1976)
- [Nash-Webber 74] Nash-Webber, B., "Semantic Support for a Speech Understanding System", *IEEE Symp. Speech Recognition*, pp. 244-249 (Apr. 1974)
- [Nash-Webber 75a] Nash-Webber, B., "Semantic Support for a Speech Understanding System", *IEEE Trans. ASSP*, vol. ASSP-23, No. 1, pp. 124-129 (Feb. 1975)
- [Nash-Webber 75b] Nash-Webber, B., "The Role of Semantics in Automatic Speech Understanding", *Representation and Understanding (Studies in Cognitive Science)* Academic Press, pp. 351-382 (1975)
- [Newell 71] Newell, A., Barnett, J., Forgie, J., Green, C., Klatt, D., Licklider, J. C. R., Munson, J., Reddy, R. and Woods, W., "Speech-Understanding Systems: Final Report of a Study Group", (May 1971)
- [Newell 75] Newell, A., Cooper, F. S., Forgie, J. W., Green, C. C., Klatt, D. H., Medress, M. F., Neuburg, E. P., O'Malley, M. H., Reddy, D. R., Ritea, B., Shoup, J. E., Walker, D. E. and Woods, W. A., "Consideration for a Follow-on ARPA Reserch Program for Speech Understanding Systems", (Aug. 1975)
- [Newell 78] Newell, A., "HARPY Production Systems and Human Cognition", CMU WP-SR (Sep. 1978)
- [Niimi 78] Niimi, Y. and Kobayashi, Y., "A Voice Input Programing System Using BASIC-like Language", *IEEE ICASSP'78*, pp. 425-428 (Apr. 1978)
- [O'Malley 73] O'Malley, M. H., Kloker, D. R. and Dara-Abrams, B., "Recovering Parentheses from Spoken Algebraic Expressions", *IEEE, Trans. AU*, vol. AU-21, No. 3, pp. 217-225 (Jun. 1973)
- [Oshika 75] Oshika, B., Zue, V. W., Weeks, R. V., Nuc, H. and Aurbach, J., "The Role of Phonological Rules in Speech Understanding Research", *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-23, No. 1, pp. 104-112 (Feb. 1975).
- [Paxton 76] Paxton, W. H., "A Framework for Speech Understanding", Ph. D. Dissertation (Stanford University) (1976).
- [Quinton 76] Quinton, P., "A Syntactic Analyzer Adapted to Speech Recognition", *IEEE ICASSP '76*, pp. 453-456 (Apr. 1976)
- [Reddy 71] Reddy, D. R., "Speech Recognition: Prospects for the Seventies", *Proc. IFIP '71*, pp. I-5-I-13 (1971)
- [Reddy 72] Reddy, R. and Erman, R. and Neely, R. et al. "Working Papers in Speech Recognition—I—", WPSR CMU, (Apr. 1972)
- [Reddy 73] Reddy, D. R., Erman, L. D. and Neely, R. B., "A Model and a System for Machine Recognition of Speech", *IEEE Trans. AU*, vol. AU-21, No. 3, pp. 229-238 (Jun. 1973)

- [Reddy 76] Reddy, D. R., "Speech Recognition by Machine : A Review", Proc. IEEE, pp. 501-531 (Apr. 1976)
- [Reddy 77] Reddy, R., et al., "Speech Understanding Systems : Summary of Results of the Five-Year Research Effort at Carnegie-Mellon University", CMU, (1977)
- [Saitta 78] Saitta, L., "Fuzzy Semantic Network for a Speech Understanding System - An Experimental Study -", IEEE ICASSP'78, pp. 433-436 (Apr. 1978)
- [Sakai 77] Sakai, T. and Nakagawa, S., "A Speech Understanding System of Simple Japanese Sentences in a Task Domain", Trans. IECE Japan, vol. E60, pp. 13-20 (1977)
- [Skinner 76] Skinner, T. E., Kloker, D. R. and Medress, M. F., "Speech Recognition System for Connected Word Sequence", IEEE ICASSP'76, pp. 434-437 (Apr. 1976)
- [Smith 79] Smith, A. R., "Word Hypothesization in the HEARSAY II Speech System", IEEE ICASSP'79, pp. 54-552 (Apr. 1979)
- [Sondhi 78] Sondhi, M. M. and Levinson, S. E., "Computing Relative Redundancy to Measure Grammatical Constraint in Speech Recognition Tasks", IEEE ICASSP'78 pp. 409-412 (Apr. 1978)
- [Tappert 73] Tappert, C. C., Dixon, N. R. and Rabinowitz, A. S., "Application of Sequential Decoding for Converting Phonetic to Graphemic Representation in Automatic Recognition of Continuous Speech (ARCS)", IEEE Trans. AU, vol. AU-21, No. 3, pp. 225-228 (Jun. 1973)
- [Tappert 74] Tappert, C. C. and Dixon, N. R., "A Procedure for Adaptive Control of the Interaction Between Acoustic Classification and Linguistic Decoding in Automatic Recognition of Continuous Speech", Artificial Intelligence, vol. 5, pp. 95-113 (1974)
- [Vicens 69] Vicens, P., "Aspects of Speech Recognition by Computer", Stanford Univ. CS 127 MEMO A1-85, (Apr. 1969)
- [Vintsiuk 76] Vintsiuk, T. K., Gavriluk, O. N. and Shinkazh, A. G., "Phoneme-by-Phoneme Recognition of Speech Composed of the Words of Given Vocabulary", IEEE ICASSP'76, pp. 450-452 (Apr. 1976)
- [Vintsiuk 76] Vintsiuk, T. K., "Generative Grammars and Dynamic Programing in Speech Recognition with Learning", IEEE ICASSP'76, pp. 446-449 (Apr. 1976)
- [Walker 75] Walker, D. E., "The SRI Speech Understanding System", IEEE Trans. ASSP, vol. ASSP-23, No. 5, pp. 397-416 (Oct. 1975)
- [Walker 78] Walker, D. E., "Understanding Spoken Language", North-Holland (1978)
- [Weinstein 75] Weinstein, C. J., McCandless, S. S., Mondschein, L. F., and Zue, V. W., "A System for Acoustic-Phonetic Analysis of Continuous Speech", IEEE Trans. ASSP, vol. ASSP-23, No. 1, pp. 54-67 (Feb. 1975)
- [Weisen 74] Weisen, R. A. and Forgie, J. W., "An Evaluation of the Lincoln Laboratory Speech Recognition System", J. Acoust. Soc. Am., vol. 56, S27(A) (1974).
- [White 78] White, G. M., "Dynamic Programing, The VITERBI Algorithm, and Low Cost Speech Recognition", IEEE ICASSP '78, pp. 413-417 (Apr. 1978)
- [Woods 70] Woods, W. A., "Transition Network Grammars for a Natural Language Analysis", Commun. ACM, vol. 13, No. 10 (1970)
- [Woods 74] Woods, W. A. and Makhoul, J., "Mechanical Inference Problems in Continuous Speech Understanding", Artificial Intelligence, vol. 5, pp. 73-91 (1974)
- [Woods 75] Woods, W. A., "Motivation and Overview of SPEECHLIS : An Experimental Prototype for Speech Understanding Reserch", IEEE Trans. ASSP, vol. ASSP-23, No. 1, pp. 2-10 (Feb. 1975)

- [Woods 76a] Woods, W. A., Bates, M., Brown, G., Bruce, B., Klovstad, J. W. and Nash-Webber, B., "Uses of Higher Level Knowledge in a Speech Understanding System: A Progress Report", IEEE ICASSP '76, pp. 438-441 (Apr. 1976)
- [Woods 76b] Woods, W. A., Bates, M., Brown, G., Bruce, B., Cook, C., Klovstad, J., Makhouli, J., Nash-Webber, B., Schwartz, R., Wolf, J., Zue, V., "Speech Understanding Systems—Final Report (November 1974–October 1976)", BBN Report, No. 3438 (1976)
- [Woods 79] Woods, W. A. and Zue, V. W., "Dictionary Expansion via Phonological Rules for a Speech Understanding System", IEEE ICASSP '79, pp. 561-564 (Apr. 1979)
- [de Figueiredo 79] de Figueiredo, R. J. P. and Brzustowicz, J., "Techniques for Spectrogram Patterns Based on Dynamic Modeling", IEEE ICASSP '79, pp. 266-268 (Apr. 1979)
- [de Mori 75] de Mori, R., Rivoira, S. and Serra, S., "A Speech Understanding System with Learning Capability", Proc. 4th IJCAI, Tbilisi, USSR, pp. 468-475 (1975)
- [de Mori 78] de Mori, R., "Recent Advances in Automatic Speech Recognition", IJ CPR '78, pp. 106-124 (Nov. 1978)
- [de Mori 79(a)] de Mori, R. and Torasso, P., "Lexical Classification in a Speech Understanding System Using Fuzzy Relations", IEEE ICASSP '79, pp. 565-568 (Apr. 1979)
- [de Mori 79(b)] de Mori, R., Lesmo, L. and Poncini, M., "The Structure of a Lexicon for a Speech Understanding System", IEEE ICASSP '79, pp. 252-255 (Apr. 1979)
- [板橋 71] 板橋, 城戸「辞書と音形規則を利用した単語音声の認識」日本音響学会誌 27 巻, 9 号, pp. 473-482, (1971 年 9 月)
- [好田 78] 好田正紀, 中津良平, 鹿野清宏, 伊藤憲三「音声によるオンライン質問回答システム」日本音響学会誌 34 巻, 3 号 pp. 194-203 (1978 年 3 月)
- [坂井 63] 坂井利之, 堂下修司「会話音声識別装置」電子通信学会誌 46 巻 pp. 1696-1704, (1963 年 11 月)
- [坂井 76] 坂井利之, 中川聖一「不特定話者・連続音声向き単語音声の識別」情報処理 17 巻, 7 号, pp. 650-658, (1976 年 7 月)
- [鹿野 78] 鹿野清宏・好田正紀「会話音声の機械認識における言語処理」電子通信学会論文誌 J 61D, 4 号 pp. 253-260 (1978 年 4 月)
- [白井 78] 白井克彦, 荻田雅彰「音声波からの調音パラメータの推定」電子通信学会論文誌 J 61A, 5 号, pp. 409-416 (1978 年 5 月)
- [関口 77] 関口芳広・大輪一・青木憲一・重永実「フォートラン・プログラムの音声認識システム」情報処理 18 巻, 5 号, pp. 445-452 (1977 年 5 月)
- [関口 78] 関口芳広・重永実「日本語文章の音声認識システム」日本音響学会誌 34 巻, 3 号, pp. 204-213, (1978 年 3 月)
- [武谷 76] 武谷峻一・田町常夫「依存文法を用いた音声認識のための構文解析法について」電子通信学会誌 J59-D, 7 号, pp. 475-482, (1976 年 7 月)
- [千葉 78] 千葉成美「単語音声の認識」情報処理 19 巻, 7 号, pp. 669-674 (1978 年 7 月)
- [中川 78] 中川聖一「機械による会話音声の認識・理解研究の動向」情報処理 19 巻, 7 号, pp. 675-685, (1978 年 7 月)
- [中島 78] 中島隆之・鈴木虎三・大村浩・石崎俊・田中和世「デコンボリューションによる声道形の推定と適応型音声分析システム」日本音響学会誌 34 巻, 3 号, pp. 157-166 (1978 年 3 月)
- [中田 80] 中田和男「音声入出力の装置化と応用」計測と制御 19 巻 3 号, pp. 31-33 (1980 年 3 月)
- [中津 78] 中津良平, 好田正紀「会話音声の機械認識における音響処理」電子通信学会論文誌 J 61D, 4 号, pp. 261-268 (1978 年 4 月)



- [中津 79] 中津良平, 「ARPA 音声理解プロジェクトの解説」電子通信学会誌 62 巻, 3 号, pp. 334-337, (1979 年 5 月)
- [新美 75] 新美康永, 浅見俊幸「音声認識システムにおける言語情報の利用とその効果」電子通信学会論文誌 58D, 12 号, pp. 741-747, (1975 年 12 月)
- [新美 77] 新美康永, 小林豊, 浅見俊幸, 三木豊「Spoken BASIC 1 の認識システム」情報処理 18 巻, 5 号, pp. 453-459, (1977 年 5 月)
- [新美 79] 新美康永「音声認識」共立出版, 東京 (1979 年)
- [藤崎 72] 藤崎博也「音声認識の諸問題」日本音響学会誌, 28 巻, 1 号 pp. 33-41 (1972 年 1 月)
- [藤崎 78] 藤崎博也「音声研究の現状と将来」日本音響学会誌 34 巻, 3 号 pp. 117-121 (1978 年 3 月)
- [ALPAC 66] ALPAC, "Language and Machines, Computers in Translation and Linguistics", National Academy of Sciences, 1966.
- [Arthern 79] Arthern. P. J., "Machine Translation and Computerized Terminology Systems - A Translators View Point", in [Snell 79].
- [Bar-Hillel 60] Bar-Hillel. Y., "A Demonstration of the Nonfeasibility of Fully Automatic High Quality Translation", in [Bar-Hillel 64].
- [Bar-Hillel 64] Bar-Hillel. Y., "Language and Information, Selected Essays on their Theory and Application", Addison-Wesley: Massachusetts, 1964.
- [Borko 68] Borko. H (Ed.), "Automated Language Processing", John Wiley and Sons: New York, 1968.
- [Brinkman 77a] Brinkman. K. H and Thake. E., "TEAM: A program System for the solution of Terminological and Lexicographical Problems", in [Brinkman 77b].
- [Brinkman 77b] Brinkman. K. H et., "Machine-assisted Translation in West Germany", JPRS 68726, 1977.
- [Bruderer 77] Bruderer. H. E., "The Present State of Machine and Machine-assisted Translation", in [Comission of EC 77].
- [Bruderer 78] Bruderer. H. E., "Handbuch der maschinellen und maschinenunterstützten Sprachübersetzung", Verlag Dokumentation: München 1978.
- [Cauche 75] Cauche. J., "The ATEF and CETA System", Am. J. Comput. Linguist, microfiche 17, 1975.
- [Chafe 75] Chafe. W., "An Approach to Verbalization and Translation by Machine", Am. J. Comput. Linguist., microfiche 10, 1975.
- [Chaloupka 76] Chaloupka. B and Gnugnoli. G, "On Xonics System", in [Hays 76].
- [Chandiox 76] Chandiox. J., "METEO, An Operational System for the Translation of Public Weather Forecasts", in [Hays 76].
- [Chandiox 77] Chandiox. J., "Creation of a Second Generation System for Machine Translation of Technical Manuals", in [Commission 77].
- [Chevalier 78] Chevalier. M, Danserau. J and Poulin. G, "TAUM-METEO, Description du system", Universite de Montreal, 1978.
- [Chomsky 57] Chomsky. N., "Syntactic Structure", The Hague: Mouton & Co., 1957.
- [Chomsky 65] Chomsky. N., "Aspects of the Theory of Syntax", M. I. T. Press, 1965.

- [Commission of the EC 77] Commission of the EC, "Overcoming the Language Barrier, Third European Congress on Information Systems and Networks", 1977.
- [Damerau 76] Damerau. F. J, "Automated Language Processing", Annual Review of Information Science and technology, 1976.
- [Dostert 55] Dostert. L, "The Georgetown-I. B. M. Experiment", in [Locke 55a].
- [Dr. F 77] Dr. F. B, "The Lexicography Information System (LEXIS) of the Bundeswehr Language Service", in [Brinkmann 77b].
- [Elliston 79] Elliston. J, "Computer Aided Translation-A Business View Point", in [Shell 79]
- [Gerald 76] Gerald. M and Hutton. F, "Georgetown University MT System Usage, Nuclear Division, Union Carbide Corp., Oak Ridge", in [Hays 76]
- [Gleibach 65] Gleibach. S, "A New Normal-form Theorem for Context-free Phrase Structure Grammar", J. Assoc. Comput. Mach., 12-1, 1965.
- [Goetschalckx 77] Goetschalckx. J, "Terminological Activities in the European institutions, with Special Reference to EURODICAUTOM", in [Commission of the EC 77].
- [Goetschalckx 79] Goetschalckx. J, "Eurodicautom", in [Snell 79].
- [Goshwake 77] Goshwake. W, "SLUNT (Spoken Languages Universal Numeric Translation)" in [Commission of the EC 77].
- [Hutchins 78] Hutchins. W. J, "Progress in Documentation, Machine Translation and Machine-aided Translation", J. Document., 34-2, 1978.
- [Jordan 77] Jordan. S. R, Brown. A. F. R and Hutton. F. C, "Computerized Russian Translation at ORNL", J. Am. Soc. Inf. Sci., 1, 1977.
- [Kay 72] Kay. M and Bisbey. R, "The MIND Translation System---A Study in man-machine Collaboration", NTIS AD 749000, 1972.
- [Kay 73] Kay. M, "The MIND System", in [Rustin 73].
- [Kay 76] Kay. M, "The Proper Place of Man and Machines in Translation", in [Hays 76].
- [Kertesz 74] Kertesz. F, "How to Cope with the Foreignlanguage Problem: Experience Gained at a multidisciplinary Laboratory", J. Am. Soc. Inf. Sci., 1-2, 1974.
- [Knowels 79] Knowels. F, "Error Analysis of Systran Output-A Suggested Criterion for the internal Evaluation of Translation Quality and a Possible Corrective for System Design", in [Snell 79].
- [Krollman 77] Krollman. F, "User Aspects of an Automatic Aid to Translation as Employed in a Large Translation Service", in [Commission of the EC 77].
- [Kuno 67] Kuno. S, "Computer Analysis of Natural Languages", Proc. Symp. Appl. Math., 19, 1967.
- [Laurent 77] Laurent. J, "Utilization of the Technical Terminology Standardized at AFNOR (Association française de Normalization)", in [Commission of the EC 77]
- [Lippman 71] Lippman. E, "An Approach to Computer-aided Translation", IEEE Trans. Eng. Writ. Speech, 14-1, 1971.
- [Lippman 75] Lippman. E, "On-line Generation of Terminological Digests in Language Translation, An Aid in Terminology Processing", IEEE Trans. Prof. Commun., 18-4, 1975.

- [Locke 55a] Locke, W. N and Booth, A. D (Eds.), "Machine Translation of Languages", Greenwood Press: Connecticut, 1955.
- [Locke 55b] Locke, W. N and Booth, A. D, "Historical Introduction", in [Locke 55a].
- [Loh 76] Loh, S. C, "CULT (Chinese University Language Translator)", in [Hays 76].
- [Loh 77] Loh, S. C and Kong, L, "Computer Translation of Chinese Scientific Journals", in [Commission 77].
- [Loh 79] Loh, S. C and Kong, L, "An Interactive On-line Machine Translation System (Chinese into English)", in [Snell 79].
- [Lytle 76] Lytle, E, "Automatic Language Processing Project", in [Hays 76].
- [Maas 77] Maas, H. D, "The Saarbrücken Automatic Translation System (SUSY)", in [Commission of the EC 77].
- [Marchuk 77] Marchuk, Yu. N, "Machine Translation in the All-Union Translation Center in [Commission 77].
- [Martew'janov 77] Martew'janov, Yu. S, "Eliminating Formal Constraints of Language", in [Commission of the EC 77].
- [Masterman 79] Masterman, M, "The Essential Skills to be Acquired for Machine Translation", in [Snell 79].
- [Mathias 76] Mathias, J, "Chinese-English Translation Assistance Group", in [Hays 76].
- [Moscovoj 77] Moscovoj, V.A, "On the Elimination of Formal Word Order in the German Sentence", in [Commission 77].
- [Oubine 76] Oubine, I. I, "Machine Translation ---Moscow International Seminar", Am. J. Comput. Linguist., microfiche 51, 1976.
- [Pendergraft 68] Pendergraft, E. D, "Translation Languages", in [Borko 68].
- [Petit 77] Petit, A. J, "Objective Evaluation Criteria for Machine Translation", in [Commission of the EC 77].
- [Raskin 75] Raskin, V, "A Restricted Sublanguage Approach to High Quality Translation", Am. J. Comput. Linguist., microfiche 9, 1975.
- [Rozenveic 74] Rozenveic, V. J (Ed), "Machine Translation and Applied Linguistics", Athenaum: Frankfurt, 1974.
- [Rustin 73] Rustin, R (Ed), "Natural Language Processing", Algorithmic Press: New York, 1973.
- [Sager 79] Sager, J. C, "Multilingual Communication: Chairman's Introductory Review of Translation and the Computer", in [Snell 79].
- [Schultz 77a] Schultz, J, "A Terminology Data Bank for Translators-Methods of Interogation in the TEAM System", in [Commission of the EC 77].
- [Schultz 77b] Schultz, J, "Lexicography with TEAM---Automatic Dictionary Composition", in [Brinkman 77].
- [Tanke 79] Tanke, E. II, "Implementing Machine Aids to Translation", in [Snell 79].
- [Toma 76] Toma, P, "Systran", in [Hays 76].
- [Toma 77] Toma, P, "Systran as a Multilingual Machine Translation System", in [Commission of the EC 77].

- [Troike 76] Troike. R. C, "The Future for MT", Am. J. Comput. Linguist., microfiche 51, 1976.
- [Vauquois 73] Vauquois. B, "Intermediate Language for Automatic Translation", in [Rustin 73].
- [Vauquois 79a] Vauquois. B, "An Actual Large Scale Project of MT---The European Community", Comput. Linguist. (Inf. Process. Soc. Japan), 18, 1979.
- [Wang 76] Wang. S. Y, "Chinese-English Machine Translation Project on Linguistic Analysis", in [Hays 76].
- [Weaver 55] Weaver. W, "Translation", in [Locke 55a].
- [Weisenborn 77] Weisenborn. J, "The Role and Form of Analysis in Machine Translation ---The Automatic Analysis of French at Saarbrücken", in [Commission 77].
- [Wilks 73b] Wilks. Y, "The Stanford Machine Translation Project", in [Rustin 73].
- [Wilks 79] Wilks. Y, "Machine Translation and Artificial Intelligence", in [Snell 79].
- [Young 78] Young. M. E, "Machine Translation (A Bibliography with Abstracts)", NTIS Document PS-78/0448, 1978.
- [Zachary 78] Zachary. W, "Machine Aids to Translation---A Concise State of the Art Bibliography", Am. J. Comput. Linguist., microfiche 77, 1978.
- [Zarechnak 76] Zarechnak. M, "Russian-English System, Georgetown University", in [Hays 76].
- [久野 65] 久野 暁, 「ハーバード大学の言語自動処理研究」, 計量国語学, 32, 1965.
- [長尾 79] 長尾 真, 「ヨーロッパにおける機械翻訳の現状」, 計量言語学研究会資料, 20, 1979.
- [奥津 81] 奥津敬一郎: "せしめたしるこ", 国語学, 10, 2, 1981, 18-26

略語: IULCP=Indiana University Linguistics Club Publication.

LA=Linguistic Analysis.

LI=Linguistic Inquiry.

NELS=Papers from the *n*th Annual Meeting, North Eastern Linguistic Society.

SEL=Studies in English Linguistics.

ODL=Outstanding Dissertations in Linguistics. (ed. by J. Hankamer)

Andrews, A. 1978. "Remarks on *To* adjunction," LI 9-2.

Bresnan, J. 1975. "Comparative deletion and constraints on transformations," LA 1-1.

———. 1976. "Evidence for a theory of unbounded transformations," LA 2-4.

———. 1977. "Variables in the theory of transformations," In Culicover et al. (eds.) (1977).

Chomsky, N. 1957. Syntactic structures. The Hague: Mouton.

———. 1965. Aspects of the theory of syntax. Cambridge, Mass.: MIT Press.

———. 1969. "Deep structure, surface structure and semantic interpretation," In R. Jakobson and S. Kawamoto (eds.) 1969. Studies in general and oriental linguistics: presented to Shiro Hattori on the occasion of his sixtieth birthday. Tokyo: TEC Co. Also in Chomsky (1972a).

———. 1970. "Remarks on nominalization," R. A. Jacobs and P. S. Rosenbaum (eds.) Readings in English transformational grammar. Waltham, Mass.: Ginn. Also in Chomsky (1972a).

———. 1972a. Studies on semantics in generative grammar. The Hague: Mouton.

———. 1972b. "Some empirical issues in the theory of transformational grammar," S. Peters (ed.) 1972. Goals of linguistic theory. Englewood Cliffs, N. J.: Prentice-Hall. Also in Chomsky (1972a).

———. 1973. "Conditions on transformations," S. Anderson and P. Kiparsky (eds.) 1973. A festschrift for Morris Halle, New York: Holt, Rinehart and Winston. Also in Chomsky (1977a).

———. 1975a. Reflections on language. New York: Pantheon.

———. 1975b. "Questions of form and interpretation," LA 1-1. Also in Chomsky (1977a).

———. 1975c. The logical structure of linguistic theory. New York: Plenum.

———. 1976. "Conditions on rules of grammar," LA 2-4.

———. 1977a. Essays on form and interpretation. New York: Elsevier North-Holland.

———. 1977b. "On wh-movement," in Culicover et al. (eds.) (1977).

———. 1979a. "Markedness and core grammar," unpublished, MIT.

———. 1979b. "Pisa lectures (1)-(3)," (Lectures at 1979 GLOW Symposium, Pisa).

———. 1979c. "Principles and parameters in syntactic theory," (To appear in N. Hornstein and D. Lightfoot (eds.) Explanations in linguistics.)

———. 1980a. "On Binding," LI 11-1.

———. 1980b. Rules and representations. New York: Columbia University Press.

———. 1980c. "On opacity," in S. Greenbaum, G. Leech, and J. Svartvik (eds.) 1980. Studies in English linguistics: for Randolph Quirk. London: Longman.

———. 1980d. "On the representation of form and function," unpublished, MIT.

- Chomsky, N. and H. Lasnik. 1977. "Filters and control," LI 8-3.
- . 1978. "A remark on contraction," LI 9-2.
- Culicover, P., T. Wasow, and A. Akmajian. eds. 1977. *Formal syntax*. New York: Academic Press.
- Dresher, B.E. and N. Hornstein. 1979. "Trace theory and NP movement rules," LI 10-1.
- Emonds, J. 1970. *Root and structure-preserving transformations*. Ph. D. dissertation, MIT. IULCP.
- . 1976. *A transformational approach to English syntax*. New York: Academic Press.
- . 1977. "Comments on the paper by Lightfoot," in Culicover et al. (eds.) (1977).
- Erteschik, N. 1973. *On the nature of island constraints*. Ph. D. dissertation, MIT. IULCP.
- Piengo, R. 1974. *Semantic conditions on surface structure*. Unpublished Ph. D. dissertation, MIT.
- . 1977. "On trace theory," LI 8-1.
- Freidin, R. 1978. "Cyclicity and the theory of grammar," LI 9-4.
- 長谷川欣佑. 1977. 「変形の性質とチョムスキーの文法体系」(to appear in 寺沢芳雄編『英語の構造と変化』研究社)
- . 1980a. 「最近のチョムスキー理論批判」『言語』9-11
- . 1980b. 「変形適用への一般制約」『英語青年』126-6.
- 原口庄輔. 1980. 「コア・グラマーとはなにか」『言語』9-11.
- Helke, M. 1971. *The grammar of English reflexives*. Ph. D. dissertation, MIT. ODL.
- Higginbotham, J. 1979. "Pronouns and bound variables," NELS 9.
- . 1980. "Anaphora and GB: some preliminary remarks," NELS 10.
- Horn, G. 1979. "Unbounded rules and island phenomena," *Occasional Papers in Linguistics* 5, University of Massachusetts.
- Ike-uchi, M. 1979. "Remarks on surface filters," SEL 7.
- Iwakura, K. 1979. "On surface filters and deletion rules," LA 5-1.
- Jackendoff, R. 1977a. *X syntax: A study of phrase structure* (Linguistic Inquiry Monograph 2) Cambridge, Mass.: MIT Press.
- . 1977b. "Constraints on phrase structure rules," in Culicover et al. (eds.) (1977).
- Jaeggli, O. 1980. "Remarks on *to* contraction," LI 11-1.
- 堀田優. 1976. 『変形文法理論の軌跡』. 東京: 大修館.
- Koster, J. 1978a. *Locality principles in syntax*. Dordrecht: Foris Publications.
- . 1978b. "Conditions, empty nodes, and markedness," LI 9-4.
- Lightfoot, D. 1976. "Trace theory and twice moved NP's," LI 7-4.
- . 1977. "On traces and conditions on rules," in Culicover et al. (eds.) (1977).
- May, R. 1977a. *The grammar of quantification*. Unpublished Ph. D. dissertation, MIT.
- . 1977b. "Logical form and conditions on rules," NELS 7.
- 長原幸雄・稲田俊明・池内正幸. 1976. 「痕跡理論(1)-(3)」『英語青年』121-(10~12).
- 中村捷. 1979. 「核文法と束縛条件(1), (2)」『英語青年』125-1.
- 大島新. 1977. 「痕跡理論をめぐって」『言語』6-3.
- 太田朗・堀田優. 1974. *文法論Ⅱ*. 東京: 大修館.

- Pesetsky, D. 1979. "Complementizer-trace phenomena and the Nominative island condition," NELS 9.
- Postal, P. M. and G. K. Pullum. 1978. "Traces and the description of English complementizer contraction," LI 9-1.
- Pullum, G. K. and P. M. Postal. 1979. "On an inadequate defense of trace theory," LI 10-4.
- Reinhart, T. 1976. The syntactic domain of anaphora. Unpublished Ph. D. dissertation, MIT.
- Riemsdijk, H. van. 1978. A case study in syntactic markedness: The binding nature of PP. Dordrecht: Foris Publications.
- Rizzi, L. 1978. "Violation of the wh-island constraint and the subadjacency condition," Montreal Working Papers in Linguistics, 11.
- Ross, J. 1967. Constraints on variables in syntax. Ph. D. dissertation, MIT. IULCP.
- Sag, I. A. 1976. Deletion and logical form. Ph. D. dissertation, MIT. ODL.
- Selkirk, E. 1972. The phrase phonology of English and French. Unpublished Ph. D. dissertation, MIT.
- Stilling, J. T. 1977. Remarks on core grammar. Unpublished Ph. D. dissertation, University of Massachusetts.
- Taraldsen, K. T. 1978. "On the NIC, vacuous application and the *that*-trace filter," IULCP.
- Terazu, N. 1975. On trace theory. Unpublished M. A. thesis. Tokyo University.
- . 1977a. "Some remarks on trace theory," Studies in English Literature (『英文学研究』) 53, English Number.
- . 1977b. "On the internal structure of semantic component," SEL 5.
- 寺津典子. 1976. 「痕跡理論について(上)(下)」『英語文学世界』11-6, 7.
- . 1977. 「修正拡大標準理論における意味部門の内部構造について」『富山大学人文学部紀要』創刊号
- . 1978. 「"Filters and Control" について」『中部英文学』5.
- . 1979. 「修正拡大標準理論における削除部門の諸問題について」『富山大学人文学部紀要』3.
- . 1980. 「最近のチョムスキー理論のめざすところ—一文法と空所化」『言語』9-11.
- Wasow, T. 1972. Anaphoric relations in English. Unpublished Ph. D. dissertation, MIT.
- . 1979. Anaphora in generative grammar. Ghent: E. Story-Scientia.
- Anderson, S. and P. Kiparsky (1973, eds.) *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston.
- Austin, J. (1962) *How to Do Things with Words*, London: Oxford University Press.
- Bach, E. (1968) "Nouns and Noun Phrases," in Bach and Harms (eds.), 90-122.
- Bach, E. and R. Harms (1968, eds.) *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston.
- Bedell, G. et al. (1979, eds.) *Explorations in Linguistics*. Tokyo: Kenkyu-sha.
- Bever, T. et al. (1976, eds.) *An Integrated Theory of Linguistic Ability*. New York: Crowell.
- Chafe, W. (1967) "Language as Symbolization," in *Language* 43, 57-91.
- . (1970) *Meaning and the Structure of Language*. Chicago: Chicago University Press.
- Chomsky, N. (1955) *The Logical Structure of Linguistic Theory*. 1975-version, New York: Plenum Press.

- Cole, P. (1978, ed.) *Syntax and Semantics*, Vol. 9, New York : Academic Press.
- and J. Morgan. (1975 eds.) *Syntax and Semantics*, Vol. 3, New York : Academic Press.
- Davidson, A. (1975) "Indirect Speech Acts and What to Do with them," in Cole and Morgan (eds.), 143-85.
- Davidson, D. and G. Harman (1972, eds.) *Semantics of Natural Language*. Dordrecht, Holland : Reidel.
- Fiengo, R. (1974) *Semantic Conditions on Surface Structure*. Unpublished Ph. D. Dissertation, M. I. T.
- Fillmore, C. (1966) "A Proposal Concerning English Prepositions," in *Georgetown Monograph Series*, Vol. 19, 19-33.
- (1968) "The Case for Case," in Bach and Harms (eds.), 1-88.
- (1971a) "Types of Lexical Information," in Steinberg and Jakobovits (eds.), 370-92.
- (1971b) "Some Problems for Case Grammar," in *Georgetown Monograph Series*, Vol. 24, 35-56.
- and T. Langendoen (1971, eds.) *Studies in Linguistic Semantics*. New York : Holt, Rinehart and Winston.
- Fodor, J. D. (1977) *Semantics*. New York : Crowell.
- Frazer, B. (1975) "Hedged Performatives," in Cole and Morgan (eds.), 187-210.
- Geis, M. and A. Zwicky (1971) "On Invited Inference," in *Linguistic Inquiry*, Vol. 2, 561-66.
- Gordon, D. and G. Lakoff (1975) "Conversational Postulates," in Cole and Morgan (eds.), 83-106.
- Green, G. (1975) "How to Get People to Do Things with Words," in Cole and Morgan (eds.), 107-41.
- Grice, H. P. (1975) "Logic and Conversation," in Cole and Morgan (eds.), 41-58.
- Gruber, J. (1965) *Studies in Lexical Relations*. Indiana Linguistic Club.
- (1967) *Functions of Lexicon in Formal Grammar*. Santa Monica, Calif. : Systems Development Corporations.
- Harnish, R. (1976) "Logical Form and Implicature," in Bever *et al.* (eds.), 313-91.
- Horn, L. (72) *On the semantic Properties of Logical Operators in English*. Unpublished Ph. D. Dissertation, UCLA,
- Huddleston, R. (1976) *An Introduction to English Transformational Syntax*. London : Longman.
- Jackendoff, R. (1968) "Quantifiers in English," *Foundations in Language*, Vol. 4, 422-42.
- (1972) *Semantic Interpretation in Generative Grammar*. Cambridge, Mass : M. I. T. Press.
- Jacobs, R. and P. Rosenbaum (1970, eds.) *Readings in English Transformational Grammar*. Boston : Ginn and Co.
- Kachru, B. *et al.* (1973, eds.) *Issues in Linguistics*. Urbana : University of Illinois Press.
- Karttunen, L. (1971) "Implicative Verbs," in *Language*, Vol. 47, 340-58.



- Katz, J. (1972) *Semantic Theory*. New York : Harper and Row.
- (1973) "Interpretive Semantics Meets the Zombies," in *Foundations in Language*, Vol. 9, 549-95.
- (1977) *Propositional Structure and Illocutionary Force*. New York : Crowell.
- (1980) "Chomsky on Meaning," in *Language*, Vol. 56, 1-41.
- and J. Fodor (1963) "The Structure of a Semantic Theory," in *Language*, Vol. 30, 170-210.
- and P. Postal (1964) *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass.: MIT Press.
- Kempson, R. (1975) *Presupposition and the Delimitation of Semantics*. London : Cambridge University Press.
- Kimball, J. (1975, ed.) *Syntax and Semantics*, Vol. 4, New York : Academic Press.
- Kiparsky, P. and C. Kiparsky (1971) "Fact", in Steinberg and Jakobovits (eds.), 345-69.
- Lakoff, G. (1968) "Instrumental Adverbs and the Concept of Deep Structure," in *Foundations in Language*, Vol. 4, 4-29.
- (1970a) "Global Rules," in *Language*, Vol. 46, 627-39.
- (1970b) *Irregularity in Syntax*. New York : Holt, Rinehart and Winston.
- (1971a) "Presupposition and Relative Well-formedness," in Steinberg and Jakobovits (eds.), 329-44.
- (1971b) "On Generative Semantics," in Steinberg and Jakobovits (eds.), 232-96.
- (1972a) "Linguistics and Natural Logic," in Davidson and Harman (eds.), 545-665.
- (1972b) "Hedges : A Study in Meaning Criteria and the Logic of Fuzzy Concepts," in *Papers from the Eighth Regional Meeting of Chicago Linguistic Society*, 183-228.
- (1973) "Some Thoughts on Transderivational Constraints," in Kachru *et al.* (eds.), 442-52.
- (1977) "Linguistic Gestalts," in *Papers from the Thirteenth Regional Meeting of Chicago Linguistic Society*, 236-87.
- Lakoff, R. (1971) "If's, and's and but's about conjunctions" in Fillmore and Langendoen (eds.), 114-49.
- (1972) "Language and Context," in *Language*, Vol. 48, 907-27.
- McCawley, J. (1968) "Lexical Insertion in a Transformational Grammar Without Deep Structure," in *Papers from the Fourth Regional Meeting of Chicago Linguistic Society*, 71-80.
- (1970a) "Where Do Noun Phrases Come from?," in Jacobs and Rosenbaum (eds.), 166-83.
- (1970b) "English as VSO Language," in *Language*, Vol. 46, 286-99.
- (1972) "A Program for Logic," in Davidson and Harman (eds.), 498-544.
- Ota, A. (1977a) "Time Adverbials and Conversational Implicature," in *Sophia Linguistica*, Vol. 3, 1-4.
- (1977b) "Hitei no Ronri to Imi," in *Elec Bulletin*, No. 59, 6-21.
- and M. Kajita (1974) *Outline of English Linguistics : Grammar II*. Tokyo : Taishukan.

- Perlmutter, D. (1971) *Deep and Surface Structure Constraints in Syntax*. New York: Holt, Rinehart and Winston.
- Postal, P. (1970) "On the Surface Verb 'Remind'," in *Linguistic Inquiry*, Vol. 1, 37-120.
- (1971) *Cross-over Phenomena*. New York: Holt, Rinehart and Winston.
- Quine, W. (1960) *Word and Object*. Cambridge, Mass.: M. I. T. Press.
- Reibel, D. and S. Schane (1969, eds.) *Modern Studies in English*. New Jersey: Prentice-Hall.
- Ross, J.
- (1969) "Auxiliaries as Main Verbs, in Todd (ed.), 77-102.
- (1970) "On Declarative Sentences," in Jacobs and Rosenbaum (eds.), 222-72.
- (1974). "Three Batons for Cognitive Psychology" in Weiner and Palermo (eds.), 63-124.
- (1975) "Where to Do Things with Words," in Cole and Morgan (eds.), 233-56.
- Rutherford, W. (1970) "Some Observations concerning Subordinate Clauses in English," in *Language*, Vol. 46, 97-115.
- Sadock, J. (1974) *Toward a Linguistic Theory of Speech Acts*. New York: Academic Press.
- (1978) "On Testing for Conversational Implicature," in Cole (ed.), 281-97.
- Searle, J. (1969) *Speech Acts*. London: Cambridge University Press.
- Steinberg, D. and J. Jakobovits (1971, eds.) *Semantics*. London: Cambridge University Press.
- Todd, W. (1969, ed.) *Studies in Philosophical Linguistics: Series One*. Evanston, Ill.: Great Expectations.
- Weimer, W. and D. Palermo (1974, eds.) *Cognition and the Symbolic Processes*. Hillsdale, N. J.: Earlbaum.
- Weinreich, U. (1980) *On Semantics*. ed. by W. Labov and B. Weinreich, Philadelphia: University of Pennsylvania Press.
- Wilson, D. (1973) *Presupposition and Non-truthconditional Semantics*. Unpublished Ph. D. Dissertation, MIT.
- Yamanashi, M.
- (1977) *Generative Semantic Studies of the Conceptual Nature of Predicates in English*. Tokyo: Kaitaku-sha.
- (1979) "Detachability Hypothesis, Negation and Other Related Matters," in Bedell *et al.* (1979), 552-72.
- Yasui, M. (1978a) *Gengai no Imi*. Tokyo: Kenkyu-sha.
- (1978b) *Atarashi Kikite no Bimpo*. Tokyo: Taishukan.
- Zwicky, A. and M. Geis (1975) "Ambiguity Tests and How to Fail Them," in Kimball (ed.), 1-36.
- [Adler et al 74] K. Adler, B. Steffens: *Deutsch für die Mittelstufe-Texte und Übungen*, Max Hueber Verlag, 1974.
- [Dressler 73] W. Dressler: *Einführung in die Textlinguistik*, Max Niemeyer Verlag, 1973.
- [Dressler 78] W. Dressler(ed.): *Current Trends in Textlinguistics*, Walter de Gruyter, 1978.
- [Engel et al 78] U. Engel, H. Schumacher(Hrsg.): *Kleines Valenzlexikon deutscher Verben*, TBL Verlag Gunter Narr, 1978.

- [Helbig 71] G. Helbig(Hrsg.): Beiträge zur Valenztheorie, Mouton, 1971.
- [Helbig et al 78] G. Helbig, W. Schenkel(Hrsg.): Wörterbuch zur Valenz und Distribution, Leipzig, 1978.
- [Rickmeyer 77] J. Rickmeyer: Kleines japanisches Valenzlexikon, Helmut Buske Verlag, 1977.
- [Rieser 78] H. Rieser: On the Development of Text Grammar, in [Dressler 78].
- [Wahrig 78] G. Wahrig(Hrsg.): dtv Wörterbuch der deutschen Sprache, Deutscher Taschenbuch Verlag, 1978.
- [泉井 76] 泉井久之助: 言語研究とフンボルト, 弘文堂, 1976.
- [イヴィッチ 74] M. イヴィッチ, 早田, 井上訳: 言語学の流れ, みすず書房, 1974.
- [榎本 79] 榎本久彦: 結合価理論とその成果, 言語第8巻11号 pp. 46-53, 1979.
- [亀山 78] 亀山健吉: フンボルト, 中公新書 525, 中央公論社, 1978.
- [ド・ソシュール 72] ド・ソシュール, 小林訳: 一般言語学講義, 岩波書店, 1972.
- [トルベツコイ 80] トルベツコイ, 長嶋訳: 音韻論の原理, 岩波書店, 1980.
- [中島 80] 中島文雄: 英語の構造(上), 岩波書店, 1980.
- [ヘルビヒ 73] G. ヘルビヒ, 岩崎他訳: 近代言語学史, 白水社, 1973.
- [ホーンビー 77] A. S. ホーンビー, 伊藤訳: 英語の型と語法[第2版], オックスフォード大学出版局, 1977.
- [レイン 78] M. レイン編, 篠田監訳: 構造主義, 研究社出版, 1978.
- [Åquist 78a] Åquist. L., Guethner. F. & Rohrer. C., "Definability in ITL of some subordinate temporal conjunctions in English" in Studies in Formal Semantics ed. Guethner. F. North Holland pp. 201-221(1978).
- [Åquist 78b] Åquist. L., Guethner. F. "Fundamentals of a theory of verb aspect and events within the setting of an improved tense-logic" in Studies in Formal Semantics ed. Guethner. F. North Holland pp. 167-199(1978).
- [Åquist 78c] Åquist. L., "A system of Chronological Tense Logic" in Formal Semantics and Pragmatics for natural languages ed. Guethner. F. & Schmidt. S. J. Reidel pp. 223-254(1978).
- [Cornap 47] Carnap. R., "Meaning and Necessity" Chicago( 1947).
- [Cooper 76] Cooper. R., "Montague Grammar, Generative Semantics and Interpretive Semantics" in Montague Grammar ed. Partee. B. H. Academic Press pp. 311-362 (1976).
- [Cooper 78] Cooper. R., "Variable Bindings and Relative Clauses" in Formal Semantics and Natural Languages ed. Guethner. F. and Schmidt. S. J. Reidel pp. 131-169 (1978).
- [Church 51] Church. A., "A formulation of the logic of sense and denotation" in Structure, method and meaning ed. Henle. P. et al. (1951).
- [Dowty 76] Dowty. D. R., "Lexical Decomposition of Causative Verbs" in Montague Grammar ed. Partee. B. H. Academic Press pp. 201-245 (1976).
- [Frege 1892] Frege. G., "Über Sinn und Bedeutung" Zeitschrift für Philosophie und philosophische Kritik, vol. 100 pp. 25-50 (1892).
- [Gallin] Gallin. I., "Intentional Logic" North Holland

- [Guethner 78] Guethner. F., "Time Schemes, Tense Logic and the Analysis of English tenses" in *Formal Semantics and Pragmatics for Natural Languages* ed. Guethner. F. and Schmidt. S. J. Reidel pp. 201-222 (1978).
- [Kripke 63] Kripke. S. A., "Semantical analysis of modal logic I" in *Zeitschrift für mathematische logik und Grundlagen der Mathematik* vol. 9 pp. 67-96 (1963).
- [Montague 69] Montague. R., "On the nature of certain philosophical entities" in *The Monist* 53: pp. 159-194 (1969).
- [Montague 70a] Montague. R., "Pragmatics and intentional logic" *Synthese* vol. 22 pp. 68-94 (1970).
- [Montague 70b] Montague. R., "English as a formal language I" in *Linguaggi nella società e nella tecnica*, ed. Visentini. B. et al. (1970).
- [Montague 70c] Montague. R., "Universal grammar" *Theoria* vol. 36 pp. 373-398 (1970).
- [Montague 73] Montague. R., "The Proper treatment of quantification in Ordinary English" in *Approaches to natural language* ed. Hintikka. K. J. J. et al. Reidel pp. 221-242 (1973).
- [Schütte 70] Schütte. K., "Vollständige System modaler und intuitionistischer Logik" Springer (1970).
- [Smay 78] Smay. R. M., "Ambiguous coreference with quantifiers" in *Formal Semantics and Pragmatics for Natural Languages* ed. Guethner. F. and Schmidt. S. J. Reidel pp. 37-75 (1978).

## あ　と　が　き

本報告書は、本事業初年度として、文章情報データベースの総合的な利用に関する予備調査を実施し、総合解析システム開発のための基礎的事項について調査研究した成果をとりまとめたものである。

今後は、①構文解析、翻訳、辞書等の技術的問題の研究、②記事情報等による国際世論分析、エネルギー事情分析等の最終利用パターンからの問題設定、③適用データのリンク利用法の3つの側面から調査研究を推進し、文章情報データベース総合解析システムのあるべき姿を追究するとともに、産業経済分野で使用する実用的なシステムを開発、運用し、段階的な改良を加えてより完成度の高いシステムを作成していこうとするものである。

第2ステップとして、次年度は、本年度調査により洗い出された、文章情報データベースの効果的な利用手法や具備すべきシステム機能につき、引き続き検討を重ねるとともに、欧米の先進例を調査しシステム開発の参考に資することを目的に海外調査を実施する。

また、モデルデータベースとして、官公庁産業界等に広く注目を集めているエネルギー・カントリー・リスクに関する新データベースを構築し、総合解析システムのサブシステムとしてデータベース作成・更新システム並びに自動インデックスファイル作成システムの開発に着手する。

第3ステップとしては、前記2システムを運用体制にもっていくほか、検索システム（シソーラス利用、自然文利用）、分析、加工システム（定量化分析、自動抄録作成、自動翻訳）につき可能なところから開発に着手する。

このように徐々にシステム機能を整備、拡大しつつ、将来的には、諸機能を網羅した文章情報データベース総合解析システムの完成を目指すものであり、開発過程においては既存のシステム及びデータベースを積極的に取り入れ、より効果的、実用的なシステムを作成していくこととしたい。



——禁無断転載——

昭和57年3月発行

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3-5-8

機械振興会館内

TEL (434) 8211 (代表)

印刷所 山陽株式会社

TEL (591) 0248

56-R 011







原本 (持出嚴禁)

|         |           |
|---------|-----------|
| 受 付 No. | G-35      |
| 受付年月日   | 57. 4. 16 |
| 作 成 課   |           |

57. 4. 16