57－S004

文章情報データベース総合利用
調査研究報告書

昭 和 58 年 3 月

JIPDEC

財団法人 日本情報処理開発協会

この報告書は，日本自転車振興会から競輪収益の一部である機械工業振興資金の補助を受けて昭和５７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

近年国際情勢の変化は，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，これらの情報過程においては，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，初年度は，基礎的な調査研究を実施した。

第２年度として本年度は，基礎的な課題について引き続き調査研究を行うとともに海外先進事例の調査やデータベース作成・更新システムの研究開発を実施した。

今後は，各種調査研究と並行してシステム機能の整備・拡大を図り，より効果的，実用的な総合解析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５８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	渕　一博	(財)新世代コンピュータ技術開発機構研究所長
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
〃	矢田光治	電子技術総合研究所計算機室長
〃	所沢　仁	(財)日本エネルギー経済研究所所長付研究主幹
〃	小川芳樹	(財)日本エネルギー経済研究所研究部 第6研究室研究員
〃	小沢大二	国際協力事業団研修事業部次長
〃	木地三千子	日本経済研究センター特別研究員
〃	神尾達夫	日本経済新聞社データバンク局記事情報部次長
〃	知念利夫	日本経済新聞社データバンク局記事情報部
〃	小池康夫	㈱市況情報センター情報管理部長
〃	下岡克幸	日本貿易振興会企画部電子計算機室
〃	瀬谷重信	日本電信電話公社データ通信本部 第3データ部調査役
〃	山本恵美子	ソフトウェア研究会副会長
〃	木南公統司	㈱開発計算センターシステム管理部課長
〃	長谷川　亨	㈱ソフトウェア・リサーチ・アソシエイツ 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㈱開発技術第2部長
〃	竹内　憲	日本タイムシェア㈱システム開発担当事業本部長付
〃	小泉秀雄	日本ハネウェル・インフォメーション・ システムズ㈱マーケティング推進部次長
〃	渡辺龍雄	通商産業大臣官房情報管理課長
〃	藤森聿子	通商産業大臣官房情報管理課政策情報システム室長
〃	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
〃	佐藤安夫	通商産業大臣官房情報管理課 政策情報システム室企画係長

委員	石川敬子	通商産業大臣官房情報管理課 政策情報システム室電子計算機専門職
〃	栗川正仁	通商産業大臣官房情報管理課計画班第2係長
〃	市川　隆	(財)日本情報処理開発協会技術調査部長
〃	難波正之	(財)日本情報処理開発協会技術調査部次長
〃	宇野彰記	(財)日本情報処理開発協会開発部次長

「文章情報データベース定量化利用研究専門委員会」委員名簿

（順不同，敬称略）

主査	山本毅雄	図書館情報大学図書館情報学部教授
委員	石塚英弘	図書館情報大学図書館情報学部助教授
〃	石川徹也	図書館情報大学図書館情報学部助手
〃	神尾達夫	日本経済新聞社データバンク局記事情報部次長
〃	下岡克幸	日本貿易振興会企画部電子計算機室
〃	木南公統司	㈱開発計算センターシステム管理部課長
〃	長谷川亨	㈱ソフトウェア・リサーチ・アソシエイツ 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㈱開発技術第2部長
〃	竹内憲	日本タイムシェア㈱システム開発担当事業本部長付
〃	佐藤安夫	通商産業大臣官房情報管理課 政策情報システム室企画係長
〃	石川敬子	通商産業大臣官房情報管理課 政策情報システム室電子計算機専門職
〃	所沢仁	㈶日本エネルギー経済研究所所長付研究主幹
〃	野田信一郎	㈶日本エネルギー経済研究所第6研究室長
〃	小川芳樹	㈶日本エネルギー経済研究所第6研究室研究員
〃	難波正之	㈶日本情報処理開発協会技術調査部次長
〃	宇野彰記	㈶日本情報処理開発協会開発部次長

「機械翻訳システム研究専門委員会」委員名簿

（順不同，敬称略）

委員	矢田光治	電子技術総合研究所計算機室長
〃	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計算機室
〃	柿元俊博	富士通㈱科学エネルギーシステム開発部 第1科学システム課
〃	中瀬純夫	リソースシェアリング㈱開発技術第2部長
〃	山本恵美子	ソフトウェア研究会副会長
〃	知念利夫	日本経済新聞社データバンク局記事情報部
〃	小池康夫	㈱市況情報センター情報管理部長
〃	瀬谷重信	日本電信電話公社データ通信本部第3データ部調査役
〃	竹内　憲	日本タイムシェア㈱システム開発担当事業本部長付
〃	佐藤安夫	通商産業大臣官房情報管理課 政策情報システム室企画係長
〃	栗川正仁	通商産業大臣官房情報管理課計画班第2係長
〃	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

1. 事　業　概　要

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 実施経過

本事業の計画概要は図１－１に示すとおりであり，これまでの実施経過を以下に示す。

(1) 昭和５６年度

「文章情報データベース総合利用調査委員会」を設置して，本調査研究の基本計画を策定し，事業の推進とりまとめを行った。

また，委員会メンバーを中心とするワーキング・グループで以下のテーマに基づいて調査研究を実施した。

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

- キーワード自動抽出法の研究
- 入力データ作成上の課題
- 既存システム利用事例研究
- シソーラス辞書作成の研究
- 定量化利用方法論の研究
- 新記事情報利用システムの研究

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

- 実用可能性の研究
- 機械翻訳技術の研究
- 構文解析技術の研究
- 入出力インターフェースの研究
- ＬＣ－ＭＡＲＣの実験準備

なお，５６年度調査は，(財)日本エネルギー経済研究所への委託により実施した。

(2) 昭和５７年度

前年度と同様「文章情報データベース総合利用調査委員会」により，全

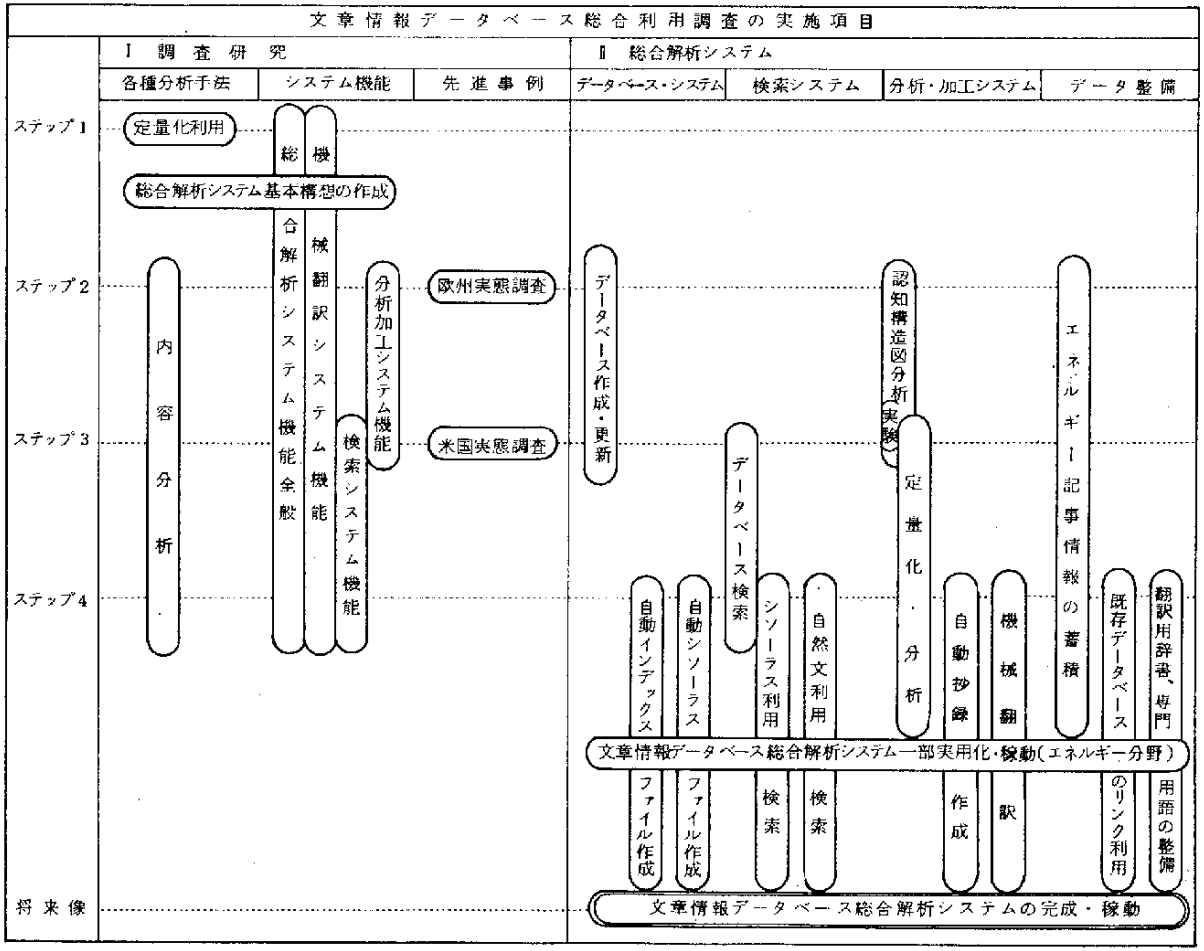

文章情報データベース総合利用調査の実施項目
Ⅰ 調査研究
Ⅱ 総合解析システム
各種分析手法
システム機能
先進事例
データベース・システム
検索システム
分析・加工システム
データ整備
ステップ1
ステップ2
ステップ3
ステップ4
将来像
定量化利用
総合解析システム基本構想の作成
総合解析システム機能全般
機械翻訳システム機能
検索システム機能
分析加工システム機能
内容分析
欧州実態調査
米国実態調査
データベース作成・更新
認知構造図分析実験
定量化・分析
エネルギー記事情報の蓄積
自動インデックス
自動シソーラス
データベース検索
シソーラス利用
自然文利用
自動抄録
機械翻訳
既存データベース
翻訳用辞書、専門
ファイル作成
ファイル作成
検索
検索
作成
訳
のリンク利用
用語の整備
文章情報データベース総合解析システム一部実用化・稼動(エネルギー分野)
文章情報データベース総合解析システムの完成・稼動

図1－1　計画概要

体の推進，統括を行った。

委員会の下に「文章情報データベース定量化利用研究専門委員会」並びに「機械翻訳システム研究専門委員会」の２つの専門委員会を設置して，研究を実施した。

「文章情報データベース定量化利用研究専門委員会」では，

- 国際紛争データと危機管理システムの研究
- General Inquirerにおけるデータ作成と解析
- 認知構造図手法の研究と適用実験

について事例研究を行うとともに，総合解析システムのうちデータベース作成・更新システムの開発を実施した。

つまり近年広く各界の注目を集めているエネルギー，カントリーリスク関連分野をモデルに記事情報の整備及びデータベースの構築を行った。

「機械翻訳システム研究専門委員会」では，

- 辞書構造の研究と作成上の課題
- 翻訳システム先進事例の研究
- 構文解析法と辞書の役割

について研究を行った。

また，①文章情報の有効利用法と機械翻訳のアルゴリズム，②文章情報解析システムにおける構文解析と用語データベースの役割をテーマに，欧州を中心とする先進事例の調査を行い，今後の参考に資した。

なお，システム開発に伴う作業の一部は(財)日本エネルギー経済研究所に委託して実施したものであり，５７年度本調査研究の経過概要は表１－１に示す通りである。

表１－１　昭和５７年度調査実施経過表

項目＼月	57/4	5	6	7	8	9	10	11	12	58/1	2	3
Ⅰ　実施計画の策定	←	—	→									
Ⅱ　委員会の開催												
• 文章情報データベース総合利用調査委員会				8○				25○				25○
• 文章情報データベース定量化利用研究専門委員会						22○			3○			23○
• 機械翻訳システム研究専門委員会								9○			2○	16○
Ⅲ　文章情報総合利用の研究												
• 国際紛争データと危機管理システムの研究				←	—	→						
• General Inquirer におけるデータ作成と解析							←	—	→			
• 認知構造図手法の研究と適用実験					←	—	—	—	—	—	—	→
• 辞書構造の研究と作成上の課題				←	—	—	→					
• 翻訳システム先進事例の研究							←	—	—	—	→	
• 構文解析法と辞書の役割								←	—	—	—	→
Ⅳ　海外調査員の派遣									↔ 欧米		↔ 欧州	
Ⅴ　総合解析システムの開発												
• データベース作成・更新システムの開発						←	—	—	—	—	→	
• エネルギー記事情報の整備					←	—	—	—	—	→		
Ⅵ　報告書の作成										←	—	→

1.3　本報告書の構成

本報告書は，次の５章からなっている。

まず才１章で事業概要をのべ，才２章では総合解析システム像と，その基幹となる諸機能につき，５６年度作成した基本構想を基として，本年度の経緯を踏まえた検討を加え，再度，基本構想をとりまとめた。

才３章は，本年度開発したデータベース作成・更新システムの概要と，カントリーリスク，エネルギー関連の記事情報の整備についてふれた。

才４章は，文章情報データベースの総合的な利用に資するため，文章情報の定量化による内容分析手法の検討と適用実験の結果および各種機械翻訳システムにおける構文解析や用語データベース（辞書）の構造等の事例研究，海外における先進事例調査等の成果を各テーマ毎にまとめた。

才５章は，これまでの成果を踏まえて問題点を抽出し，現状認識に基づく今後の方向性を考察して総合解析システム開発への具体的アプローチをまとめた。

2. 文章情報データベース総合解析システムの基本構想

2. 文章情報データベース総合解析システムの基本構想

2.1 総合解析システムの必要性

(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。世界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おり，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きている。このような状況のもとに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。

例えば，

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，MARKⅢ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。

等があげられ，こうした傾向はさらに増大すると予測される。そのため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，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の特色は情報需要の「総合化」傾向である。

従来の文章情報の検索は，データベース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，利用者がすべての情報にあたってみることは不可能になってきている。政治・経済・社会問題が顕著になる前にあらかじめ予知できるようにするには，データベースの情報から何らかの方法によって，見落しがちな変化やわずかな兆候を発見することが重要である。

一方，このような需要に対し，最近の情報解析技術の進歩，あるいはデータベースの普及等により，情報の総合的利用の条件が整いつつある。

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。大企業に偏よ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。また企業内では，特定の部門（たとえば調査部門）に集中しがちであった需要が，あらゆる部門に拡大してきている。このように専門家からいわゆる情報の“素人”にいたるまで情報を求める層が広がっている。

需要者層の広がりに伴って，情報の利用目的・利用方法は急速に拡大し，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，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) 総合解析システムの必要性

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おり，この傾向は今後さらに強まるものと予測される。

このため，需要者が求めている情報は質・量ともに拡大し，需要者は単にデータベースから検索して提供をうけるだけではなく，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り，より的確にニーズに対応するためには，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 ── の開発が必要になっている。

総合解析システムの主要な機能は，以下のとおりである。

① 文章情報を分析して内容を把握する解析能力
- 文章情報を定量化する機能
- 情報内容の時系列的分析ができる機能
- キーワード間の構造分析機能

② 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

③ 文章情報データベースと，数値データベースを組み合わせて分析する機能

④ その他
- 文章情報のなかから数値データ部分を抽出する機能

2.2 総合解析システムの概念像

(1) システムの目的

文章情報総合解析システムは，大量の文章情報をデータベースに蓄積し，これを検索・内容分析することによって，事態の客観的な把握・将来動向の予測・仮説の検証などに役立つデータを提供することを目的とし，政策の立案や決定，企業の経営計画の策定などに際して，有力なサポートシステムとして機能することをめざしている。

文章情報のデータベース化には，キーワードの付与をはじめとする文章の解析・加工作業が必要である。一般に，日本語文章情報においては，これらの作業はほとんど人手で行われているために，多くの時間・労力・コストがかかり，文章情報データベースの量的な不足と，アップ・ツー・デートな情報をタイムリーに利用できない原因となっている。また，文章情報の内容分析にいたっては，利用者個人の能力（読解力や分析力）に依存しているのが現状である。

総合解析システムは，文章情報利用における人間系作業依存の現状から脱して，機械による文章の解析・加工を行ってデータベース化を容易にするとともに，定量分析など内容分析を行い，さらに言語の異なる文章情報利用のため機械翻訳を行うシステムであり，これによって文章情報の高度な利用の促進に役立つことが期待される。

(2) システムの適用範囲

総合解析システムが取り扱う文章情報はその主題とする分野や言語を問わない，というのがあるべき姿である。すなわち，その主題が，政治・経済・外交・産業・社会のいずれの分野であっても，それぞれに即した解析や内容分析が実行できることが必要である。

また，外国語の文献の解析や，利用者が希望する言語での出力，日本語情報の海外への伝播などのための翻訳も，文章情報の総合利用のうえで欠かすことができない。したがって総合解析システムには，日本語はもとよ

り，英語をはじめとする主要な外国語の解析や，これらの言語を双方向に翻訳することが，究極的には求められる。

例えば以下のような分野での利用が期待できる。

① 企業経営・政策等意思決定への支援

- 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
- 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

② 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る。

(3) システム機能の概要

文章情報の総合解析は，概念的には，文章を構文的・意味論的に解析するプロセス（基礎解析）と，基礎解析の結果を分析・編集・合成して利用者が求める情報を，希望する形態，言語で出力するプロセス，の２段階の過程を経て達成される。後段のプロセスは，検索・内容分析と翻訳との２つのプロセスに大別され，利用者の選択によって，必要な処理が行われる。また，いずれのプロセスにおいても辞書や文法規則など知識ベースの助けが不可欠なことはいうまでもない。

すなわち，基礎解析，検索，内容分析，翻訳，知識ベースの５つが総合解析システムとして具備すべき機能である。

以下にそれぞれの機能の概要や要件を述べる。

基礎解析は，通常の文章を対象に形態素解析，構文解析等を行って，検索・内容分析や翻訳に使用する二次情報を作成する。二次情報は，一元的なデータ構造で蓄積され，検索や内容分析のためのキーワード等はもとより，機械翻訳に必要な中間言語的な要素等で構成される。したがって，基礎解析の解析結果如何が，システム全体の性能を左右することになるわけで，きわめて精緻・高度な処理能力が要求される。前述したように，キーワード付与などの作業は，現状ではほとんど人手を介しているが，これに要する時間やコスト，個人差による精粗のばらつきなどを考えると，機械による二次情報の作成は，文章情報利用促進のうえで，大きな役割を果たすものである。

これらの二次情報は，次のプロセスの効率化だけでなく，単語の意味の変化を蓄積するなど後述する知識ベースの基礎情報として，データベース化を図ることが望ましい。

検索機能は，一次情報・二次情報データベースから，利用者が求める情報を抽出する。適合率・再現率など検索効率の向上，使いやすさ，多様な検索手段などが求められる。検索された情報はアウトプットされ，または内容分析や翻訳へのインプットとなる。

内容分析は，基礎解析によって抽出されたキーワード・事実情報などを利用して定量分析などを行い，潜在的な事象や傾向を示すデータを取り出す役割を果たす。分析の方法は，単語の頻度分析をはじめ，多様な手法が考えられ，分析目的に合致した最適なものが選べる必要がある。分析の結果は，数値で示すばかりでなく，必要に応じて図やグラフで表わすことが望ましい。定量分析のほかには，単語を再編集してインデックスや抄録を作成する文章の要約化の機能も，内容分析の任務となろう。

翻訳は，二次情報データベースを利用して行われる。一般に，機械翻訳は解析→合成のプロセスをたどるが，総合解析システムでは，解析のフ

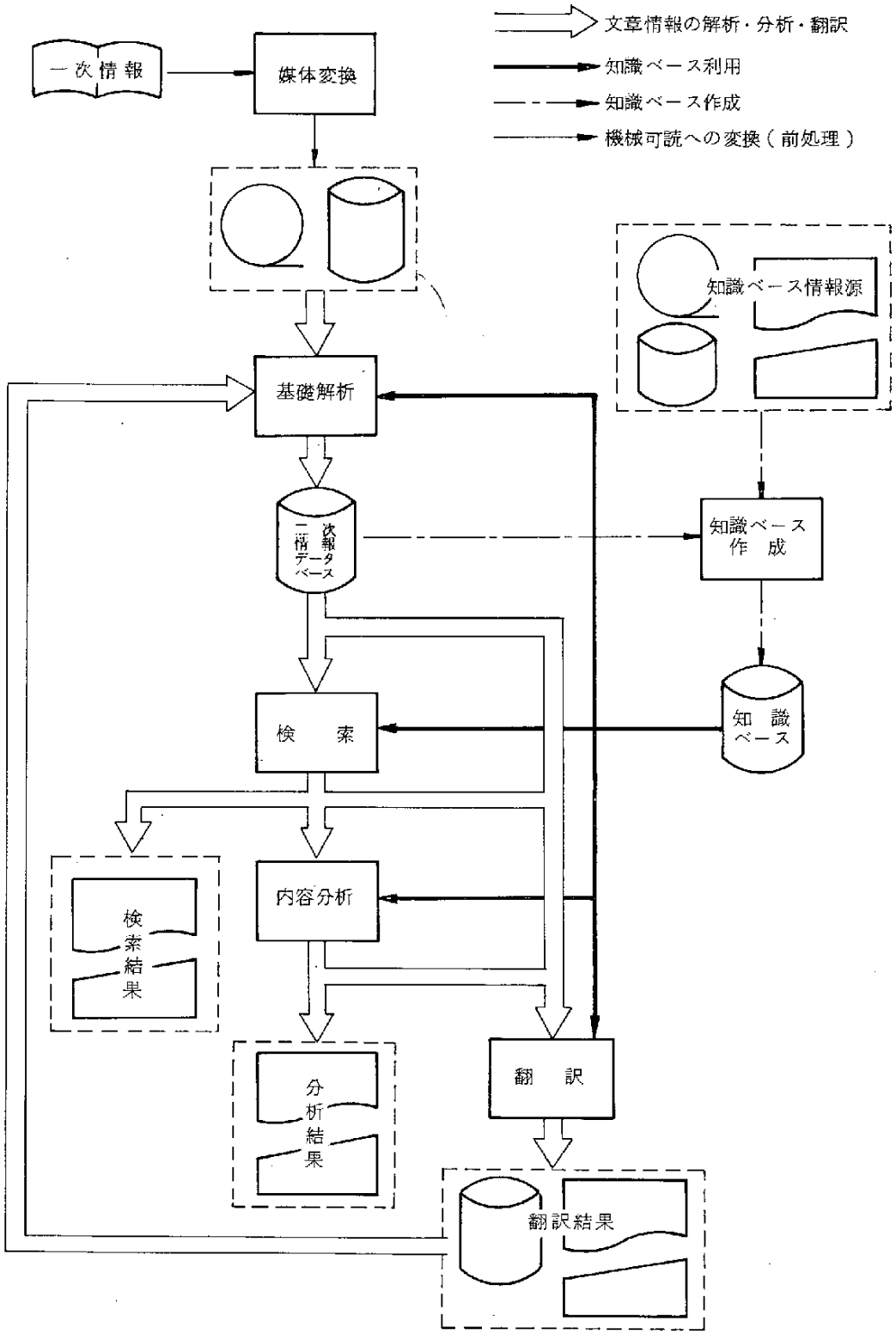


図２－１　総合解析システム概念図

ェーズは，基礎解析に委ね，翻訳は合成フェーズのみを担当することとする。これは文章情報の内容分析を行うための解析と，機械翻訳のための解析が，内容的に軌を一にすることにより，重復を避ける目的から，このような機能分担が望ましいと考えるからである。

以上の諸機能が実用上，有効に働くために，辞書や文法規則等の知識ベースが必要となる。特に総合解析システムが，対象とする文章情報の分野や言語を限定しない汎用的な機能を備えることをめざす以上は，それぞれの分野・言語に適応した知識ベースを個々に装備する必要がある。知識ベースの作成には，人手によるデータ入力をはじめ，既存の機械可読型データの入力や基礎解析の結果の活用など，様々な手段が考えられるが，日々更新される必要性からみて，自動増殖機能を具備することが重要である。

(4) システムの拡張性

以上述べてきたことは総合解析システムの理想の姿であり，これらの機能を完全に装備することは，現状の技術レベルでは困難であろう。しかし，文章情報総合利用の促進のため，本システムの開発は急務であり，技術の進歩を待っていたずらに時を過ごすことは避けるべきであろう。

総合解析システムの開発は，一度に高度なレベルのシステムの完成をめざすのではなく，現存の技術をもれなく取り入れたうえ，限定された機能のシステムを実現させることが，最善の道と考える。重要なことは，今後の技術進歩や研究開発の成果を逐次吸収できるよう，システムの拡張性に特に配慮することである。

2.3 基礎解析サブシステム

(1) 位 置 づ け

基礎解析の目的は日本語，英語等のソース言語で表現されている一次情報に言語解析，加工処理を行い，その結果を二次情報として蓄積し，次のステップである内容分析，検索あるいは翻訳などが効果的に行えるよう

にすることである。

基礎解析の対象となる一次情報は言語，データ属性等によって多種多様な形態が考えられる。

入力された一次情報の基礎解析処理フローは図２－２に示す処理要素で構成される。

- 形態素解析
- 構文解析
- 意味解析

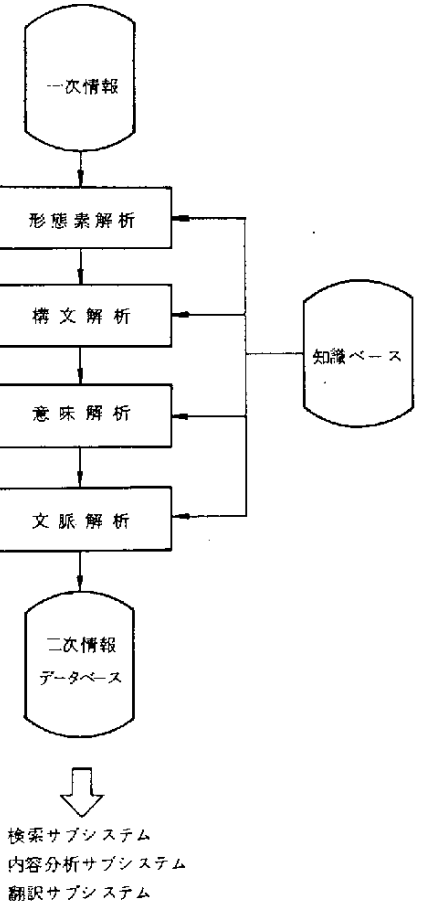


図２－２　基礎解析サブシステム

- 文脈解析

これらの一連の処理はシソーラス，専門語辞書，一般語辞書，不要語辞書などの自立語辞書や文法辞書などの知識ベースを参照しながら行われる。

二次情報の利用目的，すなわち内容分析や検索，翻訳の要求レベルによって一次情報の基礎解析の処理内容は異なるが，次の情報を二次情報として生成する。

- 文章ＩＤ
- 書誌的情報
- 本文
- キーワード
- 事実情報
- 機械翻訳用中間言語

(2) 機　　能

文章情報を一次情報として入力し，二次情報を生成する基礎解析処理は自然言語処理技術が基本となる。文章解析は大きく分けて形態素解析，構文解析，意味解析，文脈解析に分類できる。二次情報は，これらの各処理の結果として生成される。

形態素解析から文脈解析までのそれぞれの処理は，必ずしも完全に分離できるものでなく，相互に補完する関係にある。構文的な解析はかなり行われているが，文脈処理は高度な知能処理であるため，未だ研究の初歩的段階にある。

以下に，基礎解析サブシステムの構成技術と処理概要を示す。

(a) 形態素解析

形態素解析の処理内容は単語認識，品詞判定，複合語生成・分割等が主要なものである。日本語解析の場合，分ち書きの習慣のない言語であることから，最終的な語への分割は文の解析が終った時点で明確になる。英語のように分ち書きのある言語では，この種の問題点はない。

一連の解析処理では活用語辞書，付属語辞書，生成・分割情報辞書（複合語生成・分割）などの文法辞書をもとに，一般語辞書，専門語辞書，関連語辞書（シソーラス），不要語辞書，利用者辞書を用いて語の認識を行う。

(b) 構文解析，意味解析，文脈解析

形態素解析処理によって得られた文の各要素について，相互の文法的な構造を明確にすることが構文解析である。

形態素解析，構文解析処理における困難な問題点は「曖昧さ」の除去である。

意味解析とは，曖昧さを除去して文の各要素の意味構造を確定することである。

曖昧さの要因には多品詞語，多義語，異形同義語，省略，代名詞および指示詞等の使用，さらには統語論的曖昧さがある。

これらの曖昧さは本質的には，文脈解析での意味処理によって解決されるべきものであるが，このために多くの課題が現在研究の途上にある。

基礎解析の各処理過程で創成される各種の二次情報は情報検索，内容分析，翻訳処理の入力情報として使われる。図２－３は二次情報と利用内容の関係を示したものである。

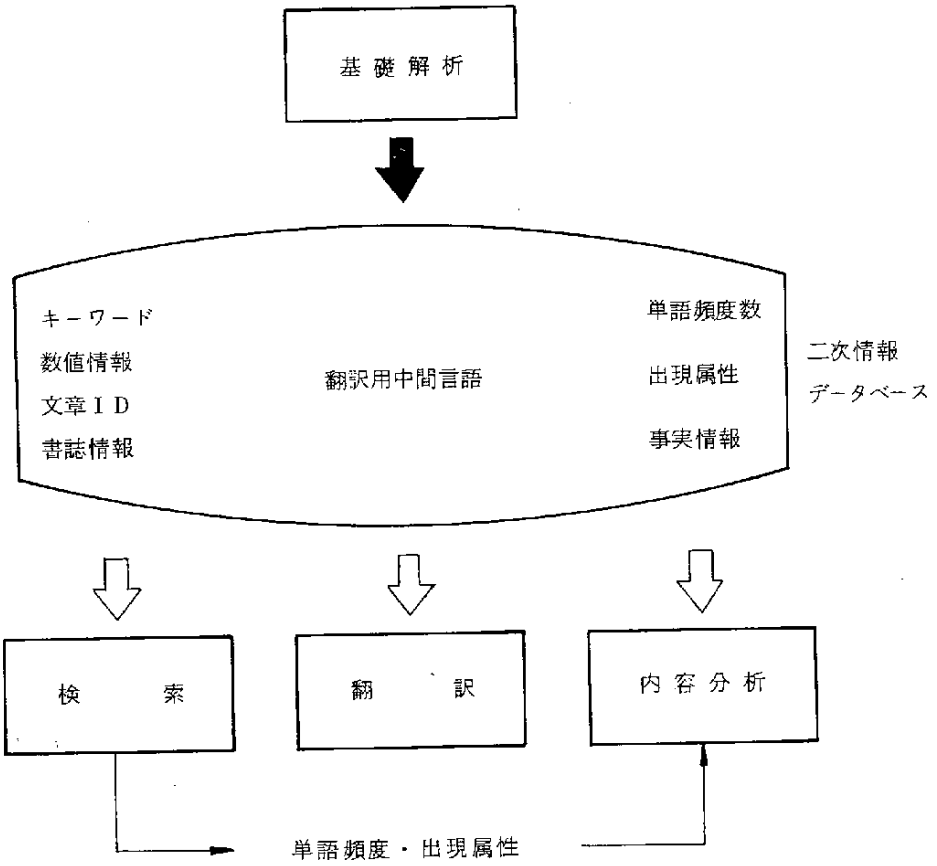


図２－３　基礎解析処理データの利用

例えば，キーワードの抽出結果は，情報検索分野から利用される。文章情報を入力し，単語認識，品詞判定，複合語生成・分割等の形態素解析処理に基づき，シソーラス，ストップワード（不要語）等の辞書を用いて該

当キーワードを抽出する。表２－２には，この処理技術の構成を示す。

表２－２　キーワード抽出の構成技術

技術項目（詳細）				英語	日本語
自然語の解析	人力テキスト編集 （誤字抽出／修正技術）			○	○
	形態素解析	単語認識		－	○
		品詞判定		○	○
		複合語	分割	－	○
			生成	○	○
		文法処理 （活用形，接辞，数詞，送り仮名）		○	○
	構文解析			△	△
	意味解析			△	△
自然語の加工	不要語の除去 （ストップ・ワード）			○	○
	関連語の処理 （同義語の統一，表記法の統一）			○	○
	出力テキスト編集 （人手による修正）			○	○
辞書	自立語辞書 ○ 一般語辞書 ○ 専門語辞書 ○ 関連語辞書 ○ 不要語辞書 ○ 利用者辞書			○	○
	文法辞書 ○ 活用語辞書 ○ 付属語辞書 ○ 分割情報辞書			○ －	○

(注)　○：必須，　△：場合によって必要，　－：不要

一方，翻訳処理では基礎解析の解析レベルにより各々の自然言語に依存した方式と自然言語に依存しないユニバーサルな方式の２つの方式がある。前者はトランスファー方式と呼ばれ，ターゲット言語文を生成するためにソース言語依存の内部表現からターゲット言語依存の内部表現に変換する。後者はピボット言語方式と呼ばれ，ユニバーサルな意味構造を持っておりソース言語，ターゲット言語の各々の個別言語に依存しないので，原理的には多言語間の翻訳が可能となる。従って，基礎解析では，解析レベルに応じて二次情報の生成内容が異なる。

2.4 検索サブシステム

(1) 位 置 づ け

文章情報データベースを有効に活用するために，まず検索効率の高い，使い易い情報検索機能を備えていなければならない。情報検索の対象は新聞の全文記事の一次情報に加え，抄録データ等をベースにした二次情報データベースも容易に検索できるようにする。

(2) 機　　能

各種のオンライン情報検索サービスの普及により，文章情報（文献情報）の検索はかなり一般化しつつある。現在一般に用いられている検索方法は，キーワードあるいはフリータームのマッチングによる語レベルの検索である。基本的には，次のような検索機能が考えられる。

① 検索効率の向上を図るためのシソーラス活用機能

② 全文検索，文字列検索，近接検索などと呼ばれるフリーテキスト・サーチの機能

③ 通常の遡及検索（Retrospective Serch）の他に，新しい情報が到着するごとに，あらかじめ登録してある質問に該当する情報を検索するＳＤＩ（Selective Dissemination of Information）の機能

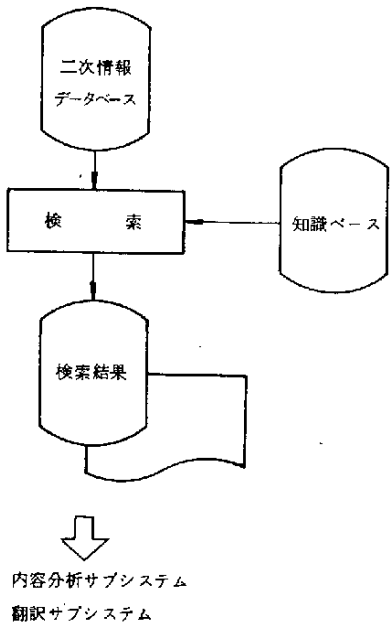


図２－４　検索サブシステム

2.5　内容分析サブシステム

(1)　位　置　づ　け

一次情報がそのままの形で利用者のニーズを満たす場合もあるが，大量の文章情報全体が示す意味を把握したい，あるいは要点のみを知りたい等の高度な処理を必要とする形で要求される場合もある。

内容分析は文章情報を定量化して利用したり，大量の文章をサマライズしたりする機能である。

(2)　機　　能

内容分析機能は，以下に述べるより高度な検索ニーズをサポートしたり，分析ニーズに対応する機能である。

前者は，キーワードの文中での使われ方や役割を分析し，主体，客体，時，場所，活動等の属性をキーワードに持たせて検索に利用する。また，名詞（キーワード）だけでなく，動詞，形容詞等との述語に対する関係を始め，各語間の関係を格構造の形でとらえ，検索に利用する。

さらに，文単位での意味内容が組み合されて，文章全体の意味内容が分析され，それに基づく検索が行われる。G. Salton による「SMART 情報検索システム」などは，その試みの１つであるが，本格的なシステムの実現には知識，推論レベルの高度な技術開発が必要である。

後者の内容分析のニーズに対応する機能としては大量の文章情報を有効に活用するため，文章情報を数量化して定量的に扱う機能と，報告書などの長文の文章を抄録などの形に集約化する機能が要求される。

定量化利用の第１ステップは，まず文章情報を数量化することである。

文章情報を数量化する最も基本的な方法は，文章中に出現する単語の頻度を分析する方法である。時系列的に出現頻度を分析する単純な方法から，語の同時出現を分析して相関を調べたり，さらには多次元の角度から語を数量化し表２－３に示す多変量解析など，各種の解析システムが考えられる。

単語の出現頻度分析だけでは，ある主題が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はできない。従って，文レベルの分析により述語とそれを修飾する格の内容をとりだすと同時に，主題に対する態度の強度等を定量化すること，さらには，各文の内容を主体又は客体ごとにまとめ，文章全体としての態度を定量化する機能(注)が必要である。

第２のステップは，政策決定のプロセスをモデル化し，数量化された文章情報をモデルに組込むことである。政策決定モデルはカントリー・リスク分析，エネルギー問題分析など，問題分野別に作成される。例えば，相手国の文書，メッセージが内容分析されてモデルにインプットされ，自国のとり得る政策を変数としてシミュレートし，その中から最適解を選んで政策を決定する。こうしたデシジョン・サポート・システムとしての機能が，総合解析システムには強く要求される。

(注) 第４章 4.3 及び 4.4 で紹介されている“General Inquine”は，こうした機能を持つ総合的内容分析システムとして開発されている。

表２－３　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 量の推定	重回帰分析 正準相関分析	数量化分析１類	
2　型	標本の分類 質の推定	判別分析	数量化分析Ⅱ類	分割表の分析 クラスター分析
3　型	多変量の統合整理（減らす） 変量の分類，代表変量の発見（選定）	主成分分析 因子分析	数量化分析Ⅲ類 Ⅳ類	潜在構造分析
4　型	分析の検定 （test による justjfy）	分散分析		

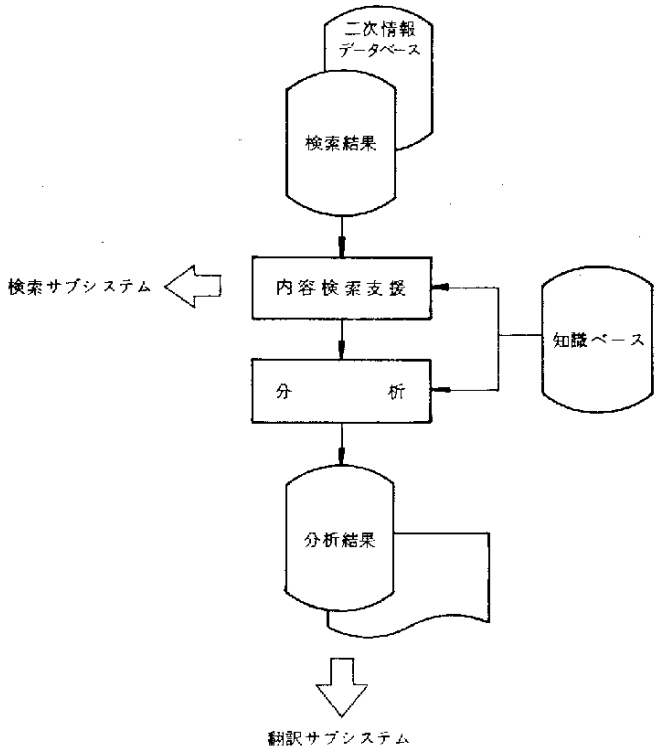


図２－５　内容分析サブシステム

2.6　翻訳サブシステム

(1)　位　置　づ　け

日本語の文章のみならず海外の文章情報も同様に処理を可能とするために，翻訳機能が必要である。機械翻訳のレベルを図示すると，図２－６の様になる。図の左半分が，入力文を解析する部分であり，右半分は，出力文を生成する部分である。図の様に，文の解析，生成はさらに多数のレベルに分割されている。

入力されたテキストに対して，まず形態素解析がなされる。形態素解析とは，単語認識，品詞判定，複合語分割，語形変化の認識などがなされる事を言い，その解析後の情報は，そのまま次の構文解析の入力となる。構文解析では，形態素解析された情報により，１つの文章の解析を行う。そ

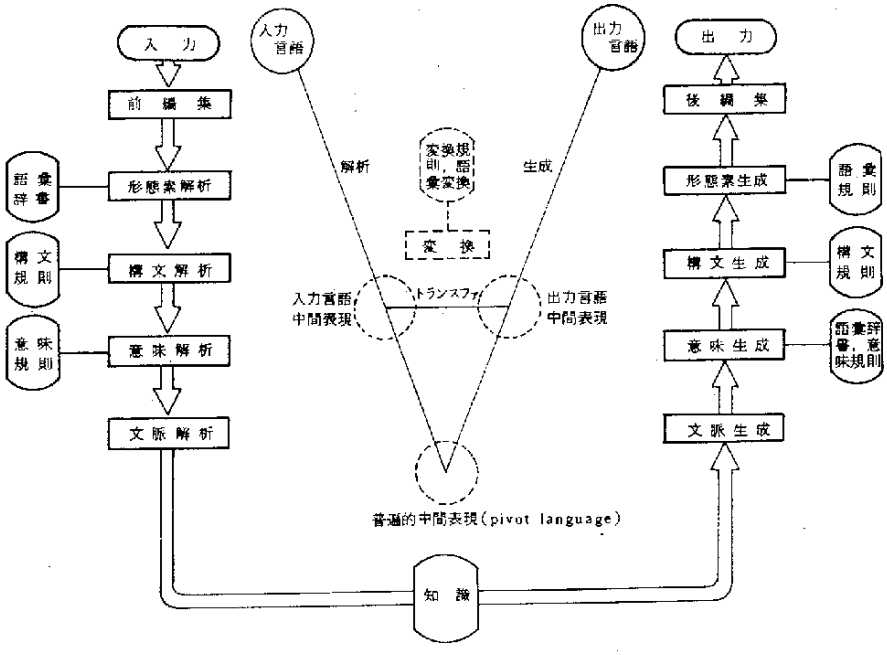


図２－６　機械翻訳のレベル

の手法には，ボトムアップ解析と，トップダウン解析の２つの方法が一般的に用られている。構文解析での問題点は，文法的に正しい幾通りもの構造が解析される事であり，このレベルでは，それらの構造から意味的に最適な構造を選び出す事は困難である。その選択は，次のレベルの意味解析に任される。意味解析は，１つの単語が，複数の品詞，複数の語義を持つ事に起因するあいまいさを是正するためになされる。次のレベルの文脈解析とこの意味解析については，まだ多くの研究を必要としており，今後どの様に発展していくかは現在のところ未知の部分が多い。

本システムでは，図２－６に示される左半分の解析部分は，すべて基礎解析でなされる。

解析の結果は，二次情報の中に格納される。翻訳では，その情報より，目的とする言語の生成が主な機能となる。本システムにおける機械翻訳システムの位置づけを図示すると図２－７の様になる。

以下，その概略について述べる。

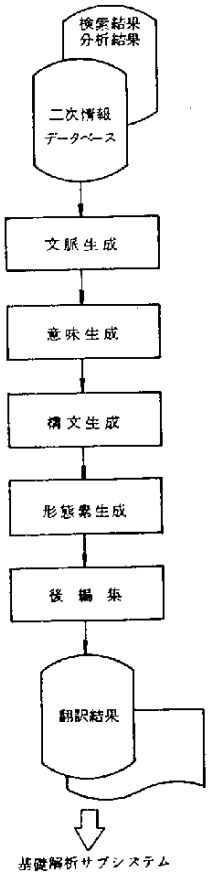


図２－７　翻訳サブシステム

(2) 機　　能

(a) 文の生成

文の生成には，文脈生成，意味生成，構文生成，形態素生成がある。まず，文脈全体から適切な主語，代名詞等の生成を行い，次に，ターゲット言語に適した意味表現の生成を行う。さらにターゲット言語の文法に添って構文の生成を行い，最後に形態素生成を行うという手順で進められる。翻訳においては，言語間におけるスタイルの違いによる読みづらさが問題となるが，文の生成でどの程度それが是正されるかは，現在のところ明らかでない。結果的には，生成後の文をそのまま利用する事は困難となる場合が多く，その後の編集に委ねられる所が大きい。

(b) 後　編　集

文の生成にも述べた様に，言語間のスタイルの違いは，機械翻訳システムの現在の技術レベルからはどうする事もできない。例えば，英語では関係代名詞を使った長々しい文章がかなり一般的に使用されるが，それを日本語に翻訳する場合には，複数の文章に分割し，かつ，その順序を入れかえる事によって，はるかに読み易くなる事が多い。

人間の手で翻訳を行う場合には，この様な事は，ごく一般的に行われているが，技術的に機械の実現は当分期待できない。そこで，人手による後編集が必要になってくる。

後編集で人間に委ねられる項目として，次のものがあげられる。

- ○誤訳の修正
- ○読みづらさの是正
- ○形式の整理
- ○数式，化学式，図表の挿入
- ○場合によっては注釈の挿入　　等

2.7　二次情報データベース

(1)　役　　割

一次情報は，基礎解析機能により，文章情報の高度利用が可能な内部形式に加工され，二次情報データベースに蓄積される。

二次情報データベースは，文章情報の多様な利用要求にかなう総合的な変換形式を整えていなければならない。多様な利用要求に対応するための内部情報構造は，共通な点も多いが，それぞれ固有な属性も要求される。

(a)　検　　索

書誌的情報，標題・見出し，抄録，本文等の蓄積情報及び検索を効率的に行うためのキーワードが必要である。

(b)　内容分析

①　内容検索

文章レベルあるいは文レベルでの意味内容の解釈を可能とする情報構造が要求される。

② 定量化分析

統計的分析手法による内容分析を行う場合には，文章や文中に出現するキーワード（単語・複合語）及びその出現属性が取り出されるようになっていなければならない。

(c) 翻　訳

文章情報データベースの検索・分析結果を，その利用目的に応じた言語で出力するために，任意のターゲット言語へ変換可能な中間言語形式で表現されていなければならない。

(2) 二次情報の構造

以上の要件を考慮すると，二次情報の蓄積構造は，概略次のとおりとなる。

(a) 文章ＩＤ

本システムに蓄積される順に付与される文章番号である。本システム内での個々の文章情報を一意に識別するために用いる。

(b) 書誌的情報

- 標題・見出し
- 出典
- 著者
- 発行（掲載）年月日
- 発行所　　などの原文情報

(c) 本　文

(d) キーワード

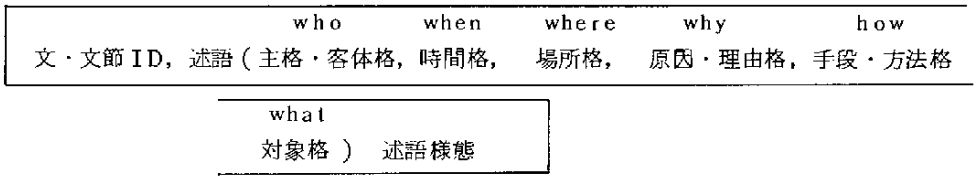
解析の過程で抽出した語のうち，文章全体の意味内容の表現に重要な役割をなすものと認定されたものに，文中での様態を付加した内部情報である。

(e) 事実情報

文章中の各文単位で述べられている事実・傾向などの意味内容を表現した内部形式の情報である。

意味表現方法は，現在実用化されているもの，あるいは研究レベルのものなどいくつか考えられるが，一般によく用いられている格文法解析に意味論的・語用論的な処理を加えたものが有効と思われる。

すなわち，意味内容を文中での述語と，それに係る格を５Ｗ１Ｈでとらえ，文節単位での統一的な構造で表現する方法をとる。



この構造では，文中での他の文節あるいは文章中での他の文との概念構造を導入することにより，はじめて事実情報の表現が達成されると思われる。従って，上記のそれぞれの格要素には語の見出しＩＤに，文中あるいは文章中で関係する構造データへの連結子（文・文節ＩＤ）も付加される。

述語様態は，その述語が文中で修飾される様態（時制，様相，法，肯定・否定など）の情報である。

この方法で内部表現する事実情報は，データの独立性・探索の一貫性を保持するために，方法論の確立したデータ構造としてのリレーショナル・モデルなどへの写像が考えられる。

上記の構造で文章情報を加工・蓄積することにより，意味内容に立入っての高度な検索・分析に対応することが可能である。更に，文章中の事実情報間の関連を文章レベルで探索することにより，文章全体の意味内容インデックスを作成する「自動インデックス機能」や文章内容の要約

を作成する「自動抄録機能」の実現へと発展させることができる。

(f) 機械翻訳用中間言語

標題・見出し，本文などは原文のまゝの蓄積の他に他言語への文章生成を可能とするために，中間言語形式で蓄えておく必要がある。機械翻訳の中間構造は，特定言語に依存しない普遍的な表現をも意識し，複数言語翻訳への拡張性を考慮しておく。また，前記の事実情報との構造上の差は予想されるが，意味的な把握の観点から事実情報データへの接続あるいは相互の構造の融合について検討しておく必要がある。

2.8 知識ベース

(1) 位置づけ

近年，自然言語処理において，処理の対象データやアルゴリズムから独立した独自の「知識要素」の設定の必要性が提起され，その重要度がますます強まってきている。本節では，総合解析システムにおいて「知識ベース」として設定されたその要素の役割りとそこで満たすべき機能に関してその概略を述べることとする。

知識ベースを独立して設定すべきとする点を次にかかげる。

① アルゴリズムとの独立

知識ベースが運用される時点において，先験的な情報（広義の「知識」）をプログラム／手続きから分離して扱う。

② 個別処理（目的）との独立

システムの総合性を確保する上で，個別の処理から独立した「知識」を設定する。

しかし，手続きには極めて知的な部分があり，そのかなりの部分は知識ベースと不可分なものであり，目的から離れた知識ベースの構築はなかなか困難である。そこで，独立性の確保には，高度な配慮が求められることに留意しつつ，むしろ，アルゴリズムと対象データの側から知識

ベースを増殖していく機能を考慮することとする。

次に，知識ベースとのかかわりから，総合解析システムの各要素の動きを粗描してみる。

基礎解析では，用意された語彙情報（辞書）を適用しながら，規則により定められた操作が施され，文章情報がコンピュータ内部表現に構造化される。この際，必要に応じて「知識」が援用される。

内容分析では，基礎解析と類似の操作により質問文の解析がなされ，回答探索，推論，モデル分析などの構造化された文章情報間の操作も，辞書，規則，知識を用いた操作により実現される。また，回答すべき内容に対して，知識を用いた規則適用と語彙の挿入を行って，回答文が生成される。

機械翻訳の場合も，必要とされる辞書，規則などは異なっていても，おおよそ以上と同様の操作がなされる。すなわち，ソース言語の解析においては，基礎解析で行われるような操作が，ソース言語からターゲット言語への移行に際しては，内容検索などでも必要となる。構造化された文章情報間の変換処理が，そしてターゲット言語の生成の為には，回答文の生成と同様な操作が行われる。

以下では，知識ベースを辞書，規則，知識表現の３つの要素に分解し，それに，それらを維持運用していくためのサポート系を加えた，各要素について述べる。なお，辞書，規則，知識表現の間の境界は必ずしも明確ではなく，その各々について他で代用するような場合，他と組合せてはじめて機能する場合，更にはアルゴリズム等でそれらの機能を大部分吸収する場合といった系の実現形態もありうる。

(2) 辞　　書

コンピュータによる自然言語処理のための辞書は，人間が参照するための伝統的な辞書と異なり，厳密な論理性をもっている必要がある。場合によっては，人間が用いている辞書とは，全く異った構成と構造をもつもの

であってもかまわない。しかし，文法理論の発展と共に，近年，辞書への一般的な考察が深まっており，少なくとも論理的な意味では，コンピュータ用辞書を特別扱いする必要はないのではないかと思われる。

以上を踏まえて，知識ベースにおける辞書に関する各事項について記す。

(a) 辞書の種類

論理的には，次の様な基準に対応して複数種の辞書が必要となる。

① 言語の別に対応した辞書

② 解析，生成，各内部操作の処理フェーズに対応した辞書

③ 適用目的の種類に対応した辞書

(b) 辞書の構造

個別記述と全体記述に分かれる。

① 語彙情報

「単語」の各々に関する構文情報，意味情報など。

② 共通情報

辞書のファイルとしてのヘッダ情報と辞書規則など。

(c) 登録語の問題

登録範囲，形態の扱い等により辞書内の項目，辞書の用い方等も異なってくる。

① 登録範囲

対象となりうる文章情報の全ての語が収納されていることを原則とするものと，一部のみを登録することで目的を達しうるような特別の場合。

合成語などを含めた膨大な専門語などを登録したものと，数万語程度の一般語を中心として対処する場合。

② 形態の扱い

活用，派生，熟語，合成語，同形異義（異品詞），異綴語に対する扱い。

(d) 構文情報

各語彙の構文情報は語自体の分類と，前後に出現しうる語の分類の2種類の内容をもつ。

① 構文カテゴリ

「動詞」，「名詞」といった，その語自体の「品詞」。

② 構文的文脈情報

語の文中での役割に応じて細分化をした構文情報。

他の語との関係で，前後に出現しうる語を構文情報で指定する。例えば，補語として形容詞をとることができるという情報である。しかし，この情報は構文カテゴリの細分類として実現する場合が多い。

(e) 意味情報

意味情報に関しては，必ずしも理論的に位置づけが定まっている訳ではない。ここでは，比較的浅いレベルで，かつ広く認められている「内在素性」，「選択制限」の他に，流動性をもたせて，「意味標記」を別アイテムとして採用する。

① 意味カテゴリ（意味標識）

語の内在的な性格を大きな枠でカテゴリ化したもの。

例えば，名詞が具象物か否か，動物か否かなどのカテゴリである。これらは統語素性とも見られる。

意味カテゴリは，次の選択素性と共に用いられるが，そのためのカテゴリ化にも，曖昧性が多く，また，際限なく分類が細かくなる危険がある。むしろ，出来る限り粗い分類とする方がよい。

内容分析などでは，分析のための属性値として，こうしたカテゴリが陽にあらわれるが，それらを辞書の意味カテゴリと同一視することは留保しておく必要がある。

② 意味的文脈情報（選択素性）

意味カテゴリなどを用いて，他の語との述語関係，格支配関係等を

記述したもの。

例えば，この動詞の主語は，人間または動作の主体となりうるものであることなど。これは，解析のあいまい度を解消するなどの効果がある一方で，フィルタとしての機能を重視しすぎると，典型的な文以外は扱えないような事態となる。

③ 意味記述

語の意味を論理的に表現したもの。

元来は意味標識の下に，または意味標識と同一視して考えられるものである。ここでは，①，②を合わせ，精密化したものとなろう。そして文の意味構造，更には事実表現などを組み上げる上では，そのもとなる情報が意味記述として記されている必要がある。

(f) 辞書規則

辞書における規則記述の自由度を許すことは，非常に強力な効果を生む。これは，文法規則とは厳密に区別する必要がある。

① 辞書全体の規則

派生語の形態，構文，意味生成のパターンを記述した規則など。

その他，辞書項目の形式記述など。

② 語彙毎の規則素性

その語には，どの規則が適用できるか（あるいは適用しなければならないか）などの情報のための項目。

(3) 規　則

辞書を補うような規則は，ここでは辞書に吸収されるものとする。またアルゴリズム（プログラム）を補うための規則も，ここでは除く。

(a) 構文構造に関する規則

文章の統語論的な構造（コンピュータ上での内部表現と考えてよい）を規定すると共に，実際に表出される文の形態と，内部表現（深層構造）との間の変換操作を規定するものである。ほとんどの場合は，句構造規

則またはそれに何らかの操作機能を付加したものを用いることになろう。構文規則だけによる解析には，数多くの曖昧度を生んでみたり，逆に解析の失敗を場当り的な規則を導入して救うことを強いられるといった問題がある。そこで，意味情報をも用いたり，意味解析規則などを並用するといった方策が必要となる。

(b) 意味に関する規則

意味に関する規則は，構文規則との関係において次のように分類される。構文規則の適用による変換操作と同時に意味構造と表層構造を結びつける意味規則と，内部的な構文構造に意味解釈を施して意味構造を作る。あるいはその逆のプロセスをたどるための意味規則とである。

意味構造の表現法には必ずしも確定したものはないが，多くのモデルは本質的に「拡張された述語とそのアーギュメント」といった枠組みで把えている。モデルの選択はむしろ，語彙や意味規則の記述の仕易さがその目安となる。

なお，操作のレベルでは，構文レベルと意味レベルとを分ける必要はない。両者を融合したものとして規則を記述するものもある。

(c) 語用論的な規則

情報検索，内容分析では，各々の文をそれだけで解析してみても，要求を満足するに十分な情報が得られない場合が多い。機械翻訳においても，文の間，文の外からの情報が必要となる事がある。そこで，それらのメカニズムを記述した規則が必要である。

しかし，現在こうした規則をプログラムから分離して独立した規則として表現する手法は定まっているとは言えない。現在は，知識ベースの中で開発されるであろうそれらの規則のための記述言語を準備しておくことが重要である。

(4) 知 識 表 現

規則と辞書の語彙情報とを用いて文の内部を操作するだけでは把えられ

ない情報内容がある。それらは別の形で表現された「知識」を用いることにより処理されることになろう。

(a) シソーラス

シソーラスは語の間の構造を表現した体系である。その汎用化は別の観点からは，概念構造の一般記述とも言いうるであろう。すなわち，シソーラスは，いくつかの概念単位として設定されたものの集合の構造記述となる。それらは，辞書における各語彙の意味記述の基ともなる。

しかし，この様な意味でのシソーラスは，現時点では研究的なものに止まるであろう。むしろ，各分野における専門用語の管理のために一般の検索用シソーラスと重複した形で作成・運用していくことが重要である。

(b) 事実データベース

総合解析システムにおいては，二次情報データベースを吸収して，知識を増殖していく様なシステム構造が望まれる。そのためには，文章解析の技術の高度化が必要となる。

事実情報は例えば，述語とそのアーギュメントの組を基とした事象レコードを，リレーショナル・モデルなどで構造化させることが出来る。

(c) 分析モデルベース

定量化分析などの基本的分析モデルを予め登録しておき，それを加工して用いられるような形態が考えられる。そのためには，そのモデル記述形式の確定が必要となる。

(d) 一般知識

高度な意味内容の操作を行うためには基本フレームとして，一般知識が知識ベースの中に表現されていることが必要である。

(5) 知識ベースのサポート系

総合解析システムの実用化に際し，知識ベースの作成維持の為には，体制とソール及びマン・マシン・インターフェースの設定が重要なキーポイ

ントとなってくるであろう。

開発や運用にあたっては，適用分野の専門家，担当者と言語を中心とする関連分野の専門家集団が中心となり，ＳＥサポートを後にひかえた体制が必要となる。

一方，基本的ツールとしても，一般的なハードウェア環境やソフトウェア環境の問題に加え，総合解析システムの各要素に関するモニタリング・システムとシミュレート・システムが特に必要となる。また，知識ベース等のための言語プロセッサも必要となろう。それは辞書，規則，知識いずれに関してもコンピュータの専門家でない人が十分に記述可能なものである必要がある。しかし，一方では，効率やシステムの融通性などの点から閉じられた系の中だけでしか使えないような目的向き言語であっても困る面がある。

以下では，知識ベースの各エレメントについて，その他の要因を簡単に述べる。

(a) 辞書の作成・更新のサポート

これは大量の人手を要する作業である。それは多数の専門家でもあり，多数の非専門的作業員でもある。それを補償する体制組織の問題が，ここでは最大の要因となる。それをサポートするためには，前記のツールの他にデータの収集・整備システム，データ分析システムその他のサービス・システムや辞書変換システムなどが必要である。

(b) 規則の作成・更新のサポート

はじめに述べた様なツールの他に，言語や，各適用分野の専門家が簡易に規則の作成や修正を行えるシステムが必要である。

(c) 知識表現に関するサポート

シソーラスに関しては(a)と同様なことが言える。また，その他の実験的な部分と実際の適用のインターフェースのためには，(b)におけると類似の作業となるであろう。

3. データベース作成・更新システムの開発

3. データベース作成・更新システムの開発

3.1 システムの目的

文章情報は非常に豊富な情報を内容としているが，情報密度は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，一般検索による利用もさることながら，特定のテーマに基づき文章情報を集約化して利用することが重要である。

その方法として文章情報から様々な情報を定量化し，さらに多変量解析等を適用していく方向が考えられる。このような文章情報定量化利用の有効な方法論が確認できれば，統計数値情報の分析結果は全く異なる視野に立つ分析が可能と考えられる。例えば，エネルギー問題，経済問題，カントリー・リスク問題といった分野などで大いに役立つであろう。

本調査研究で開発を検討している文章情報総合解析システム像は前章で述べたとおりである。そのモデルとしてエネルギー分野を取り上げた。本年度は上述のような問題意識に基づいて，キーワードを付したエネルギー関連記事情報のインデックスを収録する二次情報データベースの構築をめざし，データベース作成・更新システムを開発することを目的とした。また，モデル実験を行なうために必要な記事情報インデックスのデータ整備作業にも一部着手することとした。

3.2 情報の範囲

情報の収集は海外で発行されている主要５０紙誌の中からエネルギー問題を論及した主要記事を対象とする。

情報の内容は，上記主要記事から一定のルールに従ってコーディングしたものであり，大別すると次の３種類となる。

① データの所在に関する情報

② データの具体的内容に関する情報

③　データを入手し利用するための情報

なお，情報の項目及びその概要の一覧は，表３－１に示すとおりである。

表３－１　情報の項目一覧表

No.	項目名		備考
1	データ番号		10文字　（EBCDIC）
2	媒体フラッグ		1　文字　（EBCDIC）
3	ページ		3　文字　（EBCDIC）
4	抄録フラッグ		1　文字　（EBCDIC）
5	マイクロフィシュアドレス		9　文字　（EBCDIC）
6	予備		16文字
7	掲載年月日		漢字6文字（12文字）……XX　XX　XX （JIS）　年（西歴　下二桁）　月　日
8	媒体		漢字1文字（2文字）（JIS）
9	ページ		漢字3文字（6文字）（JIS）
10	記事インデックス		漢字40文字（80文字）（JIS）　外字"#(漢字コード)" XX……XX　（XXYYY） 記事インデックス　紙誌名略号　紙面の大きさ
11	記事インデックス内外字数		2文字　（EBCDIC）
12	外字コード		漢字10文字（20文字）（JIS）
13	予備		6文字
14	キーワード総数		2文字　（EBCDIC）
15	キーワード1		
	1	カテゴリー(大)	1文字　（EBCDIC）
	2	カテゴリー(小)	1文字　（EBCDIC）
	3	キーワード	漢字40文字（80文字）（JIS）　外字"#(漢字コード)"
	4	外字コード	漢字2文字（4文字）
	5	キーワードカナ	50文字
16	キーワード2		
	1	カテゴリー(大)	1文字　（EBCDIC）
	2	カテゴリー(小)	1文字　（EBCDIC）
	3	キーワード	漢字40文字（80文字）（JIS）　外字"#(漢字コード)"
	4	外字コード	漢字2文字（4文字）
	5	キーワードカナ	50文字

No	項目名		備考
	・・・・・・	（キーワードの繰り返し）	
44	キーワード 30		
	1	カテゴリー(大)	1文字 （EBCDIC）
	2	カテゴリー(小)	1文字 （EBCDIC）
	3	キーワード	漢字40文字（80文字）（JIS） 外字"#(漢字コード)"
	4	外字コード	漢字2文字（4文字）
	5	キーワードカナ	50文字

1レコード 4250バイト

3.3 システムの機能

本システムの機能としては，エネルギー記事情報インデックスの入力データ作成機能，データベースの作成機能，データベースの更新機能，データベースの修正機能の4つがある。

一般にデータベースの作成・更新は，一括してユーザの利用時間外に行われるのが原則であるから，システムの処理形態はバッチ処理とする。

各サブシステムとその機能については，図3－1，各サブシステム間の情報関連図は図3－2に示すとおりである。

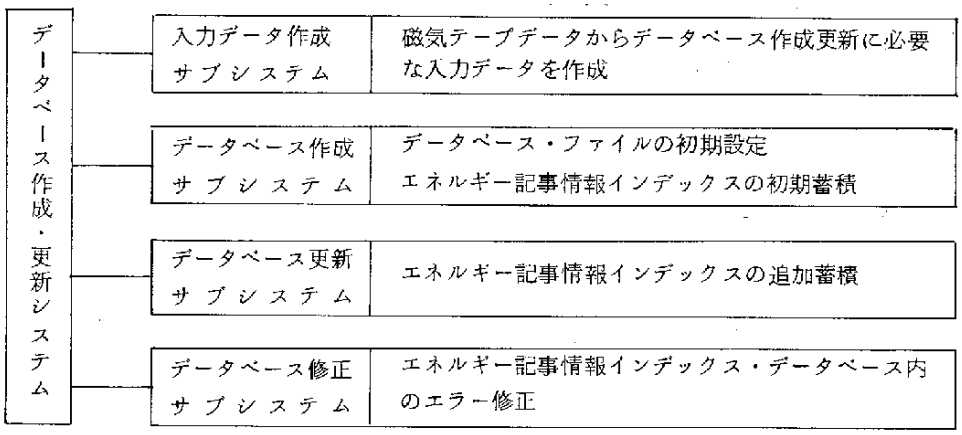


図3－1 データベース作成・更新システム

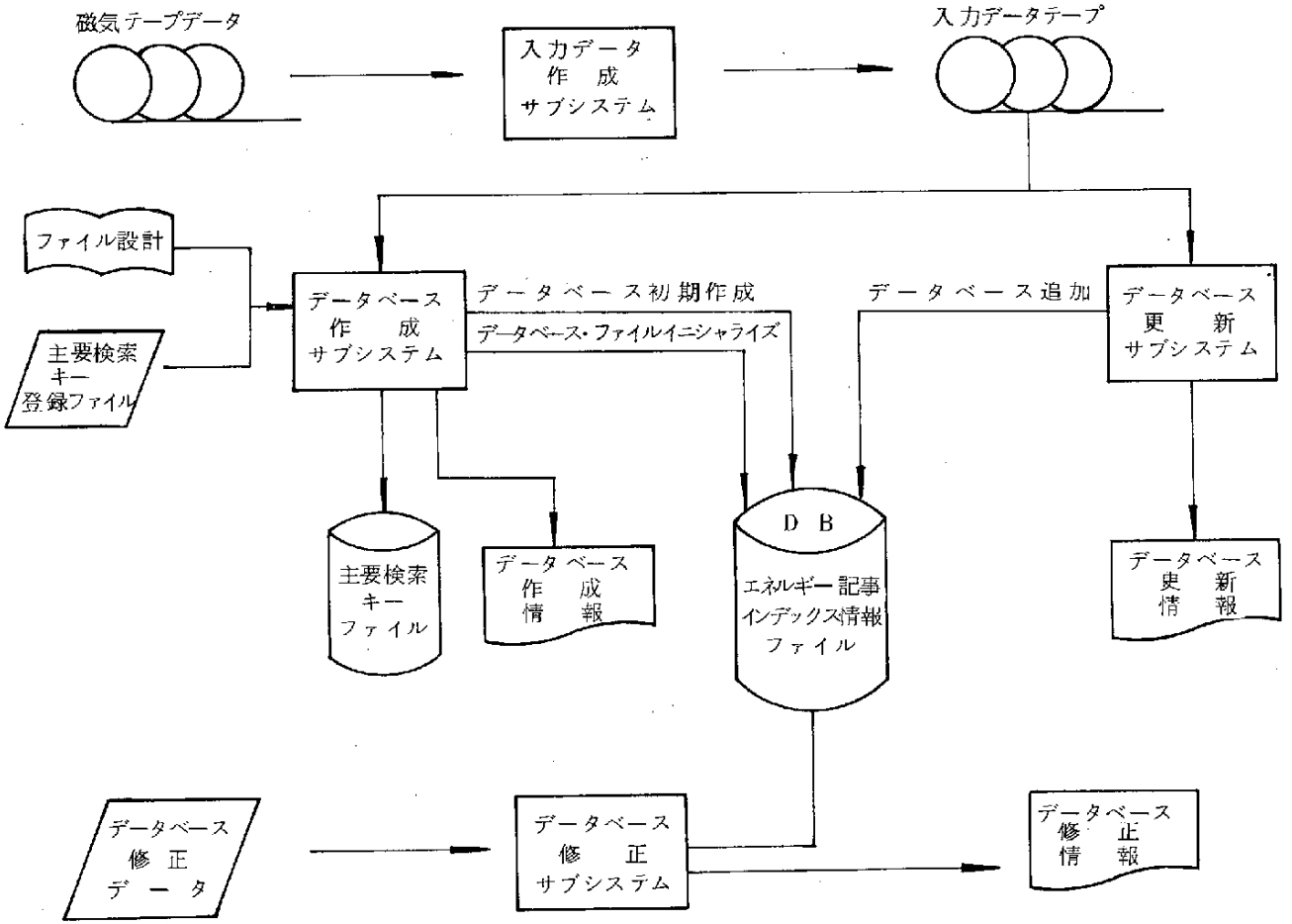


図３－２　サブシステム間の情報関連図

3.4　システムの概要

各サブシステム単位に，サブシステムの機能，入力情報，出力情報，作成ファイル等の概要について述べる。

データベース作成・更新システムの入出力情報の一覧は図３－３に示すとおりである。

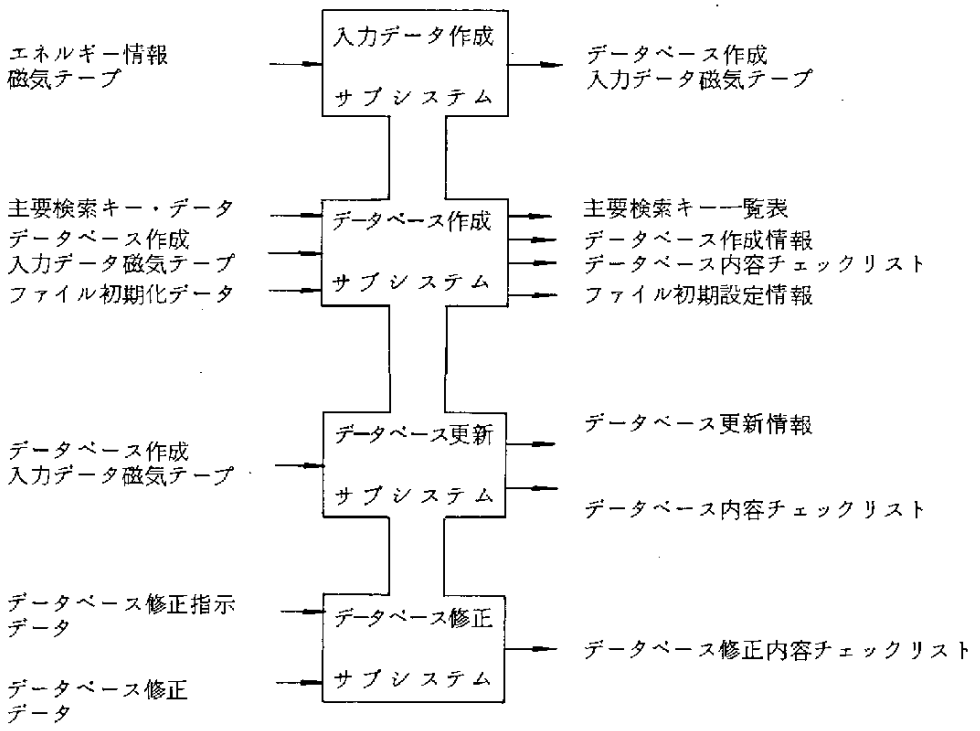
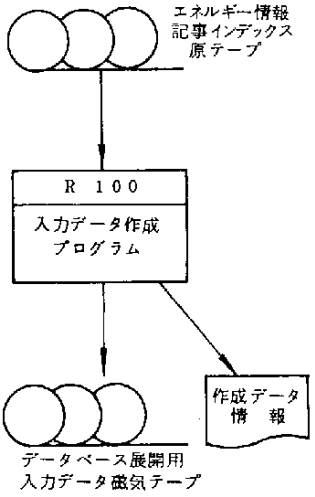


図３－３　データベース作成・更新システムの入出力

3.4.1 入力データ作成サブシステム

(1) 入力データ作成サブシステムの機能

処理フロー	ブロック名	機能
		当システムでは，エネルギー情報記事インデックス原テープを入力データとして，データベースへ展開するために必要な入力データ磁気テープを作成する。
	R 100	本プログラムで以下の作業を行なう。 ① 原テープはＥＢＣＤＩＣコードとＪＩＳ漢字コードの混在なのでコンピュータの機種に併せてコード変換する。 ② 単一項目と可変項目の情報を分離し識別する符号をつける。 ③ ４０字インデックスの中から紙誌名略号と記事行数を分離し紙誌名略号はキーワードの一つとする。 ④ 入力データに対する可能な論理チェックを行ないエラー情報等を出力する。

(2) 入力データ作成サブシステムへの入力情報

エネルギー情報記事インデックス原テープの内容は表３－１に示した通りである。一つのエネルギー記事について合計4250バイトの情報が含まれている。コードはEBCDICコードあるいはＪＩＳ漢字コードで表現されている。

(3) 入力データ作成サブシステムの出力情報

後で述べるようなデータベース作成システムあるいはデータベース更新システムで入力データとして使用するための必要情報を磁気テープに出力する。表３－１に示した１つのエネルギー記事に関する入力データから出力情報を抽出してファイルに書き出す。出力磁気テープの１レコードの内容を表３－２に示す。

表３－２　データベース展開用入力データ磁気テープの内容

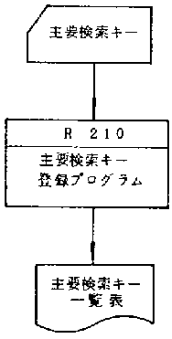
No.	項目名	備	考
1	単一項目識別情報	1文字	プログラムで付加
2	データ番号	10文字	表3-1 No.1
3	年　月	漢字4文字（8文字）	表3-1 No.7
4	日　付	漢字2文字（4文字）	表3-1 No.7
5	ページ	漢字3文字（6文字）	表3-1 No.9
6	記事内容インデックス	漢字40文字（80文字）	表3-1 No.10
7	記事行数	漢字5文字（10文字）	表3-1 No.10
8	紙誌名略号	漢字4文字（8文字）	表3-1 No.10
9	可変項目識別情報 キーワード1 分類カテゴリー	1文字 漢字40文字 2文字	プログラムで付加 表3-1 No.15の3 表3-1 No.15の1，2
10	可変項目識別情報 キーワード2 分類カテゴリー	1文字 漢字40文字 2文字	プログラムで付加 表3-1 No.16の3 表3-1 No.16の1，2
11	・ ・ ・ ・ ・		
n	可変項目識別情報 キーワードn 分類カテゴリー	1文字 漢字40文字 2文字	プログラムで付加 表3-1 No. M．3 表3-1 No. M．1，2

（n：は表3-1 No.14キーワード総数）

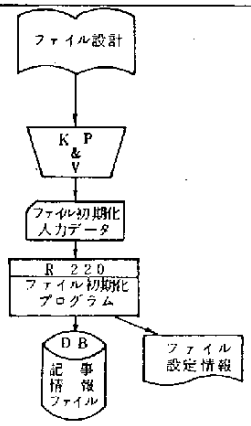
3.4.2 データベース作成サブシステム

(1) データベース作成サブシステムの機能

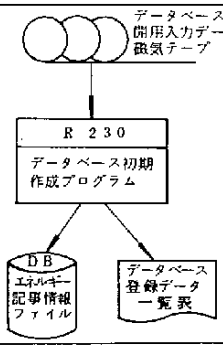
① 主要検索キー登録機能

処理フロー	ブロック名	機能
		当プログラムでは，本データベースの利用に当って極めて頻繁に使用するキーをあらかじめファイルに登録しておく。
	R 210	数字コードと漢字コードのペアでパンチされた入力カードを読み込み，登録済みのキーとのエラー・チェックを行いファイルに登録する。 登録終了後，数字コードと漢字コードを並行したキーワード一覧表を出力する。

② データベース・ファイル初期化機能

処理フロー	ブロック名	機能
		当プログラムでは，登録予想データの量，ファイル構造を十分検討したファイル設計結果に基づきファイル初期化入力データを作成する。
	R 220	入力データに基づき，コンピュータ上に磁気ディスク・ファイルとして必要なファイル構造とボリュームを持ったデータベース・ファイルを設定し，ファイル領域確保情報等を出力する。

③ データベース初期作成機能

処理フロー	ブロック名	機能
		入力データ作成サブシステムで作成したデータベース展開用入力データ磁気テープを入力データとする。
	R 230	前述の磁気テープを入力データとして読み込み，エネルギー記事情報ファイルの初期作成を行う。 作成後，登録データチェックのため一覧表を出力する。

(2) データベース作成サブシステムの入力情報

データベース作成サブシステムの入力情報は，入力データ作成サブシステムが出力する磁気テープファイルである。１レコードの内容と構成は表３－２に示した通りである。

また，主要検索キー登録のための入力データは，カード・イメージで図３－４に示すようなものである。

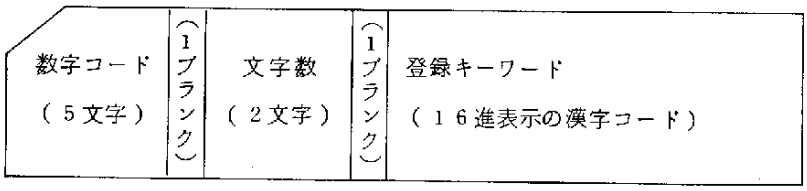


図３－４　主要検索キー登録のための入力データ

(3) データベース作成サブシステムの出力情報

データベース作成サブシステムは，登録データのチェックを行うために必要なデータベース登録データ一覧表を出力する。

その具体的な出力形式は図３－５に示す。

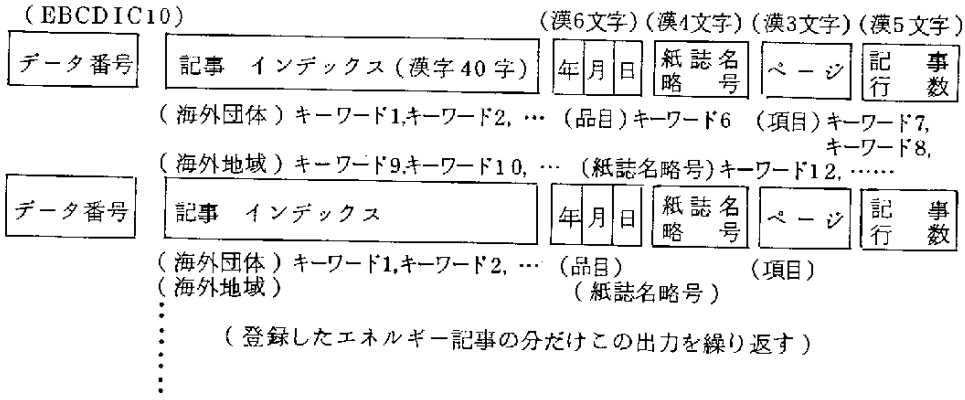


図３－５　データベース登録データ一覧表出力形式

(4) データベース作成サブシステムで作成するファイル

① 主要検索キー登録ファイル

主要検索キー登録ファイル（図３－６）は索引付順インデックス編

成（ＩＳＦ）として作成し，キーコード，キーワード名，キーワード名の文字数で構成する。

キーコード（５文字）	文字数（２文字）	キーワード名（漢字４０字以内（８０字以内））

図３－６　主要検索キー登録ファイル

② エネルギー記事情報インデックス・データベース・ファイル

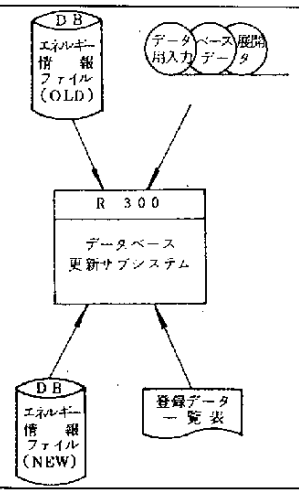
データベース作成サブシステムにより，データベースの初期作成が行われる。このデータベースファイルの構造は表３－３に示すようなものである。

表３－３　エネルギー記事情報インデックス・データベースのファイル構造

項目名	繰り返数	キー項目	単一項目	可変項目	圧縮の有無	備考
データ番号		○	○			主キーとする
年		○	○			漢字４文字
月			○			漢字２文字
日			○			漢字２文字
ページ			○			漢字３文字
記事内容インデックス			○			漢字４０文字
記事行数			○			漢字５文字
紙誌名略号		○	○			漢字４文字
キーワード名	<30	○		○	○	漢字４０文字以内
分類カテゴリー	<30			○	○	コード２文字

3.4.3　データベース更新サブシステム

(1)　データベース更新サブシステムの機能

処理フロー	ブロック名	機能
		入力データ作成サブシステムで作成したデータベース展開用入力データ磁気テープを入力データとする。
	R　300	前述の磁気テープを入力データとして読み込み，すでに作成されているエネルギー情報ファイルに追加登録する。作成後登録データチェックのため一覧表を出力する。

(2) データベース更新サブシステムの入力情報

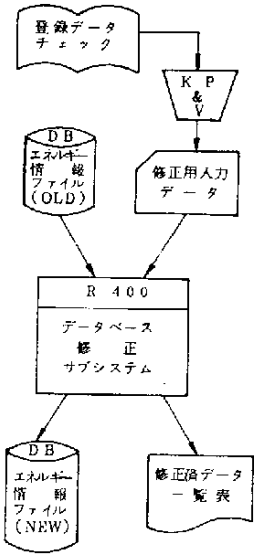
当サブシステムの入力情報は入力データサブシステムが出力する磁気テープファイルである。レコードの内容は表3－2に示されている。

(3) データベース更新サブシステムの出力情報

当サブシステムも登録データのチェックを行うために必要なデータベース登録データ一覧表を出力する。その具体的な出力形式は図3－5に示されている。

3.4.4 データベース修正サブシステム

(1) データベース修正サブシステムの機能

処理フロー	ブロック名	機能
		データベース登録データ一覧表をチェックして，修正箇所を決定した後，修正内容をカードパンチして修正用入力データを作成する。
	R 400	修正用入力データを読み込み，その指示に従ってデータベース内の情報を修正削除する。 データベース修正後，修正済のデータについてチェック用の一覧表を出力する。

(2) データベース修正サブシステムの入力情報

修正対象となる情報は以下の通りである。

① 記事インデックスの修正

② 年月日の修正

③ 紙誌名略号，ページ，記事行数の修正

④ キーワードの削除，追加

①～④までのどの修正であるかをオプションで選択できるようにオプションの入力情報を設定し，オプション・カードの後に必要な修正データの指示をデータ番号を指示して行えるように詳細設計する必要がある。

(3) データベース修正サブシステムの出力情報

修正済のデータについて図3－5に示したような出力内容でチェック・リストを出すとともに，読み込んだ修正用入力データをしかるべきフォーマットで打ち出し修正箇所がどこであったかを示すようにする。

3.5 コンピュータの機器構成

当システムで必要な機器構成を図3－7に示す。

中央処理装置，磁気ディスク装置（コンピュータシステムが使用するもの以外にデータベース用として1台），磁気テープ装置（2台），カード読取装置及びコンピュータシステムの制御・監視のためのオペレータ・ステーションが構成機器である。

さらに当システムにおいては，出力帳票に漢字を使用しているため，漢字プリンタ装置が必要である。

また今後開発する検索システムでは，できればオンライン検索が望ましく日本語処理ターミナルの設置が必要である。

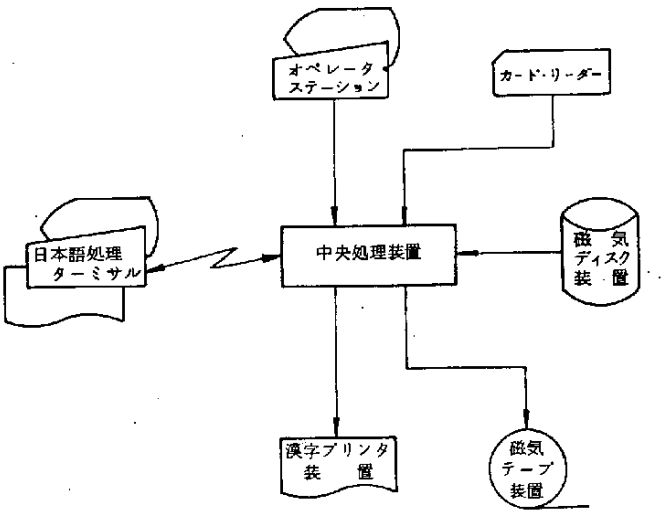


図３－７　コンピュータの機器構成

3.6　データ整備

3.6.1　データ整備の方法

抽出した記事について記事全体を把握し，各記事に対して内容を４０字以内にまとめた記事インデックスを作成し，掲載年月日，掲載紙面あるいはページ，紙誌名略号，記事行数を付加してデータを作成しコーディング作業を行った。さらに記事内容にふさわしいキーワードを海外会社，海外団体，項目，品目，地域等の大分類に従って選択し，検索あるいは定量化解析に利用するためキーワード付けを行った。これらのデータは，コーディング後パンチされ磁気テープ化される。こうして作成したエネルギー記事インデックスの例を表３－４に示す。

3.6.2　整備したデータの量

今回のデータ整備では，８２年４月から８３年１月までのＯＰＥＣ及びOPEC 13カ国に関連して重要なエネルギー記事の整備を行った。表３－５にサウジアラビアについて作成した記事インデックスを例示する。

記事インデックスを作成した件数はOPEC各国別に表3－6に示すとおりであり，総計1,179件である。

表3－6　記事インデックス・データの作成件数

国名	件数
アラブ首長国連邦	32
アルジェリア	75
イラク	93
イラン	147
インドネシア	64
エクアドル	14
カタール	8
ガボン	2
クウェート	61
サウジアラビア	192
ナイジェリア	74
ベネズエラ	33
リビア	42
OPEC	342
計	1,179

表３－４　エネルギー記事インデックスの例

★821103F015　（069面）　Y　0
国）　ベトロミン　品）　LPG　項）　価格動向F，　販売動向F，　対外交渉F，　商談F　地）　サウジアラビア
◇ベトロミン，11月末までにLPG契約の大部分決定か（APS）

★821108F031　（128面）　Y　0
国）　サウジアラビア政府　品）　製油所，　潤滑油　項）　エネルギー開発F，　工場建設F，　設備増設F　地）　サウジアラビア
◇サウジアラビア，潤滑油プラント・プロジェクトさらに拡張へ（OGJ）

★821110F019　（016面）　Y　0
品）　原油，　LPG　項）　生産動向F，　需給動向F　地）　サウジアラビア，　中近東，　東南アジア，　アフリカ
◇原油と随伴ガスとの密接な関係について――LPGの現在と将来の緊張を解明（PI）

★821119F006　（020面）　N　0
国）　サウジアラビア政府，　OPEC　人）　ヤマニ　品）　石油　項）　価格政策F，　産油国　地）　サウジアラビア
◇サウジ石油相，OPEC基準価格34ドル維持を言明（FT108）

★821119F011　（015面）　N　0
国）　サウジアラビア政府，　OPEC　人）　ヤマニ　品）　石油　項）　産油国，　価格政策F，　値下げF　地）　サウジアラビア
◇サウジ石油相，OPEC基準価格値下げの可能性を示唆（LT153）

★821122F009　（001面）　Y　0
国）　OPEC，　サウジアラビア政府　人）　ヤマニ　項）　産油国，　価格政策F，　エネルギー政策F　地）　サウジアラビア
◇ヤマニ石油相，サウジはOPECの現行基準価格34ドル支持を再確認（FT190）

★821122F011　（008面）　Y　0
会）　テキサコ，　アラムコ　品）　原油　項）　輸入F　地）　米国，　サウジアラビア
◇テキサコ，82年アラムコ原油の引き取り量を対前年比50％削減の意向（PIW）

表３－５　サウジアラビアに関するエネルギー記事インデックスの例

サウジアラビア

ヤマニ・サウジ石油相，メジャーの石油価格政策を近視眼的と決めつける（ＬＴ９８）	820401-4018(	-013)	
ナイジェリア，石油販売不振で苦境――値下げ要求でサウジとメジャーが対立（ＥＣＯ）	820403-4005(	-103)	*
サウジ石油相，ナイジェリア圧迫の石油各社制裁のためＯＰＥＣ会議召集（ＭＥＥＳ）	820405-4013(	-005)	
サウジのナイジェリア石油販売引き上げ工作に石油各社同意せず（ＷＪ１３７）	820405-4033(	-002)	*
サウジ，８１年の石油収入は前年比約２０％増と発表（ＷＪ３８）	820406-4005(	-031)	
サウジ，さらに日量５０万バーレルの減産を発表する模様（ＦＴ６０）	820407-4009(	-016)	
仏，サウジからのＧＧ石油輸入量を削減（ＡＰＳ）	820407-4016(	-001)	
英国産石油，サウジをしのぎ西独輸入石油のトップに躍り出る（ＡＰＳ）	820407-4017(	-001)	
米石油輸入，サウジ産減り北海・メキシコ産が台頭（ＡＰＳ）	820407-4018(	-002)	
米国メジャー，サウジの基準価格固守で減益の影響を受ける見込み（ＷＪ１０７）	820407-4022(	-007)	*
世界石油過剰の圧迫にもかかわらずサウジの国内支出用外貨は十分（ＷＪ１６４）	820407-4024(	-032)	*
ＯＰＥＣ原油需要，予想をやや上回る見込み――サウジの努力が奏効か（ＩＨＴ１０５）	820408-4018(	-009)	*
西側石油会社，サウジの警告でナイジェリア原油輸入契約解消撤回も（ＰＮ）	820412-4007(	-001)	
サウジ，アルジェリア圧迫のメジャーに最後通告出しＯＰＥＣ価格を防衛（ＮＷ）	820412-4018(	-042)	
アラムコのサウジ原油高値買い入れはサウジに補助金を与えているのと同じ事（ＢＷ）	820412-4023(	-023)	*
サウジ，石油値下がり阻止で更に５０万バーレル減産か――ＭＥＥＳ報道（ＦＴ２００）	820413-4001(	-018)	
フランスの８２年２月石油輸入状況――サウジからの輸入は２４．８％の大幅減（ＰＩ）	820415-4008(	-001)	*
サウジ，シリアによるイラクの石油パイプライン閉鎖に伴い減産を考慮（ＭＥＥＳ）	820419-4025(	-001)	*
サウジの８２年第１四半期原油生産，日量７９１万９０００バーレルに（ＭＥＥＳ）	820419-4026(	-007)	*
サウジの石油生産，８２年第１四半期は日量平均７５０万バーレル（ＭＥＥＳ）	820419-4033(	-007)	
サウジ，インド企業グループに総額１億５０００万ドルの住宅建設を委託（ＦＴ３６）	820421-4005(	-006)	
イラク，サウジ経由の新石油パイプライン建設急ぐ――シリア経由は放棄（ＷＪ１４４）	820422-4014(	-035)	*
サウジ含む４主要産油国，石油過剰のため輸出量削減（ＦＴ１７０）	820423-4001(	-001)	
サウジ石油相，ＯＰＥＣ価格維持に必要ならばさらに石油減産すると言明（ＬＴ８１）	820423-4016(	-021)	
サウジ通貨庁，８１年サウジ経済動向報告を発表（ＭＥＥＳ）	820426-4021(	-001)	*
エクソン，高価格のサウジ原油取引反映し第１四半期の純益２２．５％減（ＦＴ１１９）	820428-4008(	-001)	
サウジ，石油製品販売量を８４年までに日量７５万バーレルに拡大（ＡＰＳ）	820428-4014(	-002)	
石油過剰問題，ソ連・サウジなど産油国に深刻な影響をもたらす（ＷＪ３２４）	820429-4011(	-001)	*
サウジの安全は，周辺諸国の騒乱により脅かされるとみられる（ＡＷＪ３３８）	820429-4014(	-007)	*
ナイジェリア石油相，サウジを突然訪問――金融援助を要請（ＷＪ７０）	820430-4010(	-030)	*
米石油各社，高いサウジ原油取引に泣き国内原油売却価格の高水準に笑う（ＥＤ）	820501-4003(	-072)	
サウジ，日量７００万バーレルの石油生産で８２－８３年度予算の確保に自信（ＡＯＧ）	820501-4005(	-009)	
サウジ，製油所・潤滑油工場建設計画の一部を延期する模様（ＡＯＧ）	820501-4006(	-010)	
サウジの８２－８３年度予算，日量７５０万バーレルの原油生産見込む（ＭＥＥＳ）	820503-4018(	-008)	
米系アラムコ４社，第１四半期大幅減益――割高サウジ原油が足ひっぱる（ＰＩＷ）	820503-4022(	-002)	
サウジ，８２－８３年度は均衡予算を目指す――歳入は７．８％減（ＭＥＥＳ）	820503-4027(	-001)	*

4. 文章情報総合利用の研究

4. 文章情報総合利用の研究

4.1 国際紛争データに関するデータ整備とデータ作成上の問題点

国際政治学の分野では，分析素材の一つとして記事情報等の文章情報が早くから着目され，精粗まちまちな文章情報から有効な情報を導き出すための定量化利用が盛んに行われてきた。また，国際紛争・国際事件に関するデータベースの整備も，相当量進められてきた。このような先駆的な分野において，どのような考え方に基づいて，どのような方法で，どのような情報源からデータ整備が行われ，どのような問題点があったかを探ることは，文章情報の定量化利用システムを研究するにあたってきわめて重要であると考えられる。表４－１にこれまで国際政治学の分野で作成されてきた国際紛争・国際事件のデータベースを例示する。

表４－１　国際紛争・国際事件に関するデータベース

(1) CASCON

Lincoln P. Bloomfield & Robert Beattie, "Computers and policy-making : the CASCON experiment ", The Journal of Conflict Resolution, Vol. XV, 1971, pp. 33-46.

(2)Butterworth & Scranton/FACS

Robert Butterworth, Managing Interstate Conflict: Data with Synopses (Pittsburgh: University Center of International Studies, 1976)

L. Farris, H. R. Alker, Jr. , K. Carley and F. Sherman, "Phase /Actor Disaggregated Butterworth-Scranton Codebook "Working Paper, M. I. T. Center for International Studies, 1980.

(3) COPDAB

Edward E. Azar and T. Sloan, Dimensions of Interaction: A Source Book (Pittsburgh: International Studies Association, 1975)

Edward E. Azar, "The Conflict and Peace Data Bank (COPDAB) Project," Journal of Conflict Resolution, Vol. 24, 1980, pp. 143-152.

(4) CACI

CACI, "Analysis of Superpower Crisis Management Behavior : Data and Source Code Documentation ", 1980.

“CASCON“というデータベースは，MITのBloomfield らが中心となって，60年代から70年代初めにかけてデータ整備が行われたものであり，国際間の地域紛争について，新聞・資料等の情報ソースから整理したものである。データ内容としては，地域紛争名，時期区分，進行した局面，重要な因子などが盛り込まれている。紛争の局面を①Dispute（論争・争議－軍事力使用せず），②Conflict（論争・争議－軍事力使う用意がある），③Hostilities（システマチィックな論争），④Conflict（論争・争議－いざとなれば軍事力再使用），⑤Dispute（論争・争議－軍事力使用せず），⑥Settlement（解決）という6段階に分けて，各紛争がどの段階まで進行したかを示していることがデータベースの大きな特徴となっている。米国務省インド担当官など国際関係分野の専門家が新聞・資料等の情報ソースからデータ抽出を行っている。このデータベースの具体的な利用事例として，超大国の紛争介入度を見るための要因分析がパイロット・スタディとして行われた。データベース化された地域紛争について何が重要であったかという要因を書き出させ，要因の一般化を行って分析したものである。一般化された要因は，①Disputeの段階で144個，②Conflictの段陥で141個，③Hostilitiesの段階で197個で，これら要因の関連を検討することにより，超大国の介入度が分析されている。

“FACS“も60年代から70年代にかけて整備されたデータベースである。このデータベースでは，1945－77年まで370ケースの地域紛争を取り扱っている。各紛争に対して表4－2に示すような18の質問を設定し，その答えを数量化してデータ・整理を行っている。18の質問に対する解答は1枚のカードにパンチされるので，数量化データは370枚のカードに表現されているが，文章情報の形で詳しい事件の経緯を与えている点が“FACS“データの大きな特徴となっている。“FACS"データベースのために設定された18の質問は，国際紛争の諸特徴を数量化して取り扱うにあたり，どのような観点がキー・インフォーメーションを与えうるかを示す好例となって

表４－２　“FACS”データベース作成のための１８の質問

1. Fatalities: how many battle-related deaths were there?
2. Duration: how long did the conflict last (in years)?
3. Likelihood of Abatement: how likely were the parties to abate their claims by themselves (i.e., to lower the intensity of the conflict)?
4. Likelihood of Disappearance: how long would it likely take before the parties would let their claims lapse if left to themselves and without serious escalation?
5. Likely Degree of Spread: to what extent was the conflict likely to spread further to third parties?
6. Likelihood of Super Power War: how likely was it that the United States and the Soviet Union would engage in major war over this issue?
7. Type of Warfare: what was the degree and political intent of military operations?
8. Strategic Category: what was the strategic character of military operations (e.g., military subversion, guerrilla insurgency, confrontation, and so forth)?
9. Strongest Antagonist: what rank was the strongest party, as measured by the Cox-Jacobson (1973: 437-443) scale?
10. Power Disparity: how divergent were the parties in terms of their ranks as measured by (9)?
11. Degree of Spread: to what extent and military degree had the conflict involved third parties?
12. Type of Issue: was the substantive contention over decolonization, interstate and cold war, internal cold war, interstate/other, or internal/other issues?
13. Alignment of Parties: with respect to the cold war blocs, were the parties both members of the same bloc, opposing blocs, a bloc member against a nonaligned state, or both nonaligned?
14. Ethnic Factors: were there relevant ethnic factors? How strong?
15. Ideological Contention: were there ideological components to the dispute, and were they compatible or contending?
16. Past Relationship: if there were relevant historic sources of conflict between the parties, did they center around disputed boundaries, irredentist claims, or other general animosities?
17. Great Power Interests: if there were great power interests immediately relevant to the conflict, were they economic or political/strategic?
18. System Period: did the major events of the conflict occur in the unipolar period (1945-1947), bipolar (1948-1955), tricentric (1956-1962), or multicentric (1963-continuing)?[2]

いる。しかしながら，開始年月日・死傷者数といった客観的に特定できる質問に対して，例えば超大国戦争に発展しうるかいなかなどのような主観的因子によって多分に影響を受ける質問が含まれていることは，専門家でない人間によるデータ作成を難しくしていると言うことができる。

“FACS”データは，“CASCON”データと異なって進行した局面に関する情報は提供していないが，ＭＩＴのAlkerらにより進行する局面の情報をも加味した“FACS”データベースの作成が行われている。

これまでに述べた“CASCON”および“FACS”の両データベースは，進行した局面で言うと3.Hostilitiesの段階まで突入した地域紛争を取り扱

ったものである。これに対して，歴史上には③.Hostilities の段階まで進行しないで解決した紛争が数多く存在するので，この情報も加えるべきであるという主張を踏まえて誕生したのが，次の“COPDAB”データベースである。これは1948－78年までに発生した135カ国50万件にわたる国際紛争・国際事件をデータ化したものであるが，膨大な情報量を処理するため表4－3に示すように紛争年月日，紛争当事者，紛争の目標，出所，紛争行

表4－3　“COPDAB”データの内容

(1) the *date* (year, month, and day of the event)
(2) the *actor* (who initiated the event)
(3) the *target* (to whom the event was directed)
(4) the *source* (where the event description was gathered)
(5) the *activity* (the verbal and physical act which an actor initiated)
(6) the *scale value* (assessment of the degree of cooperativeness or conflictiveness of the event using COPDAB's international and domestic scales)
(7) the *event type* (describing the issues and content of the event)
(8) the *issue area* (information regarding the substance of the event)

動，紛争の度合を示す尺度，紛争のタイプ，紛争地域といったデータ内容に絞られ，比較的シンプルなものとなっている。“COPDAB”データベースの大きな特徴は，紛争の度合を示す尺度を1～15までの点数で与えていることである。他のデータ内容はある程度特定できるものであるが，この尺度の点数を決める作業には，やはり主観的な要素が大きく働くと考えられる。前2者のデータベースでは紛争内容の評価に関する情報がかなり細かく記述されているのに対して，“COPDAB”のデータベースでは尺度の情報にすべてが圧縮されているわけである。分析素材の提供に力点を置いてデータベースを見た場合“COPDAB”はやや難があると言えよう。

表4－1の最後に示されているCACI社のデータベースは，米ソ中三大国の国際紛争に対する対処の仕方を速やかに分析し，早期の段階における危機管理を行なうことを目的として作成されたデータベースである。米国防総省

がスポンサーとなって作成されたこのデータベースは，1945－80 年までの期間における米中ソ各国が関与した紛争・事件に関するデータを網羅している。表4－4に中国に関する場合のデータを示すが，米国・ソ連の場合も

表4－4　CACI データベースの内容

1.	Record identifier
2.	Initiation date of crisis
3.	Termination date of crisis
4.	Crisis duration
5.	Crisis location by JCS Region
6.	Geopolitical location
7.	Character of events
8.	Scope
9.	Level of violence
10.	Strategic confrontation
11.	Perceived threat to Communist Party/Regime/Movement
12.	Actor mix
13-22.	Key individual actor codes
	a = USSR
	b = U.S.
	c = Other western countries
	d = India
	e = Taiwan
	f = Japan
	g = Vietnam
	h = Other South East Asia
	i = Indonesia
	k = Korea
23.	Chinese verbal involvement
24.	Chinese physical involvement
25.	Geographic involvement
26.	Consolidated involvement
27.	Chinese objectives with respect to in-theater supported set
28.	Chinese objectives with respect to in-theater opposed set
29.	Chinese crisis management capabilities
30.	Crisis outcomes for the PRC
31.	Crisis outcomes for chinese clients/allies
32.	Crisis outcomes for politics of interest to PRC

ほぼ類似した内容で構成されている。データ内容は，危機の期間，地域，タイプ，度合，関与国，関与方法，関与目的，関与結果など多岐にわたっており，このデータに基づいて紛争を明確に特徴付けることができる。データのソースは，新聞・雑誌等を含む様々な資料であるが，米国の場合，認知した国際紛争・国際事件のすべてを取り扱うと膨大な数にのぼるので国防総省

のあるデスクを通過した事件のみを取り扱うという形で一種のスクリーニングがかけられている。ソ連と中国に関しては，上述のスクリーニングは不可能なので，それぞれの公式発表の文献で取り扱われた事件がデータ整備の対象となっている。取り扱い紛争事件数は，1949－80年の間で米国352件，ソ連444件，中国427件である。実際のデータは表4－4に示した各内容について数字で符号化して与えられており，データベースを利用して情報を定量的に取り扱うことは比較的容易である。CACI社では，このデータベースに基づいて危機管理システムを開発した。これは，発生しうる問題，各国の目的，各国の取り得る行動と判断を過去のデータを分析することにより設定し，ある紛争が発生した時どのような径路を取って外交の最終ゴールに向け各国が行動を取っていくかを推定しようというシステムである。ソ連の場合には，43個の発生しうる問題，59個の目的，69個の行動形式が整理された。CACI社により米ソ関係に関する多くの問題がこのシステムを用いてシミュレートされたが，データベースの整備は予算打ち切りのため1980年までで中止された。

第2次世界大戦後の超大国間の冷戦構造のもとで生じた共産圏に関する情報整備の必要性が，国際政治学の分野におけるデータベースの整備を逸早く促進せしめたと言うことができよう。しかしながら，このようにして整備された国際紛争のデータベースにもいくつか作成上の問題点が見受けられるので，次にそれらの諸点を整理してみよう。第1に，どのようなユーザを対象とするかによってデータベースの内容が大きく異なる可能性があるということである。専門の研究者がユーザとなる場合，情報加工度の少ない豊富な分析素材をデータベースから提供することが要求されるであろうが，一般ユーザの場合，“COPDAB”のデータベースのように細かい判断材料の提供よりも高度に集約化された評価が要求されるであろう。従ってデータベースのユーザがどのような対象となるかはデータベース作成にあたって十分吟味しなければならない。

第2に，本節で取り上げた国際紛争データベースのいづれも文章情報から整理され，組織立てられたデータベース情報を抽出する，いわゆる内容分析の作業を人力を介して行っていることである。内容分析を人力で行うことは，労力がかかり過ぎる点でまず大きな問題を抱えている。特にいくつかのデータベースで見られたように高度の判断を要求してデータの抽出を行なう場合は，専門家の労力を要することになり負担はさらに大きくなる。また人力による作業は，データ作成者の主観的判断が大きく反映しうる可能性があり，データの信頼性の面でも大きな問題を抱えている。この問題を避けるためには，例えば3～4人で同じ問題をコーディングさせ一致率が90％に上がるまでコーダーの訓練を行うといった方法も考えられるが，結局労力の問題に帰結される。このような観点から，内容分析を人力を介さず機械的に行う方法はきわめて興味深いものであるが，このシステム例については4.3節および4.4節で言及する。

第3は，定量化利用のためのデータ抽出方法の汎用性がどの程度保持できるかという問題である。データ抽出を行うための1次のソースである文章情報は一定の共通したルールで蓄積していくことが可能であるが，1次のソースから定量化利用のためのベースデータを抽出する段階では，利用目的に応じて抽出するデータも抽出方法も異なってくる可能性が高いわけである。データ作成作業を組織するにあたって十分配慮しなければならない点である。

第4は，データベース作成作業の継続性の問題である。すでに述べたように内容分析を人力で行うことは，膨大な労力を要することになり，これを継続するためには十分な予算的措置が必要である。CACI 社がめざしたデータベースとしてなかなかすぐれたものではあるが，予算措置の継続性が保証されなかったため1980年までをもって中止に至っている。従って，データベース化開始にあたっては，データベースの必要性を十分検討し，継続性を維持できるような体制でのぞむことが重要である。

さて，このようにいくつかの問題点を抱えてはいるが，データベースを整

備し文章情報の定量化利用をはかっていくことは，何も国際紛争・国際事件の領域に限らず様々な分野で必要と考えられる。例えば，米国あるいはヨーロッパとの貿易摩擦問題に関連して，これまでの国際間の経済紛争を整理したデータベースを持ち，早期対策を立案するための経済紛争管理システムを開発することは，大いに期待されるところであろう。エネルギー問題に関連しては，世界経済・日本経済に多大の影響力を持つ原油価格の形成に産油国消費国，メジャーがどのように関与するかを探れるような素材を提供するデータベースが期待される。このように様々な分野で文章情報をベースとしてより組織立てられた情報に変換し定量化利用を行うことが必要とされているわけであるが，これらの実現にはすでに述べたようにクリアしなければならないいくつかの障害がある。いずれにしてもデータベースの構築は巨額の投資を必要とするので，ユーザ・サイドからの情報利用の有効性を十分確認した上で順次根元へデータベース整備の体制を伸ばしていくことが重要と考えられる。

4.2　CACI社の国際紛争データを用いた中国監視システム

4.1節では，国際政治学の分野で作成されたいくつかの国際紛争・国際事件に関するデータベースを紹介したが，本節ではその定量化利用の事例として，CACI社のデータベースを用いた中国監視システムについて詳しく検討したい。このシステムは，現在平和安全保障研究所に在職している田中明彦氏が米国のＭＩＴに留学中に完成したものである。

中国監視システムは，いくつかの前提情報と過去の先例に基づいて，中国の国際紛争に対する介入態度を予知することを目指したものである。このシステムにおける情報処理の体系を図４－１に示す。中国監視システムでは，前提情報として世界友敵地図，外交フレーム，評価戦略，決定戦略を持っており，ある事件の発生をシステムに伝えるとその事件の属性・文脈の理解を行い，事件の特徴付けを行った後，過去の先例を当たり類似した事例を

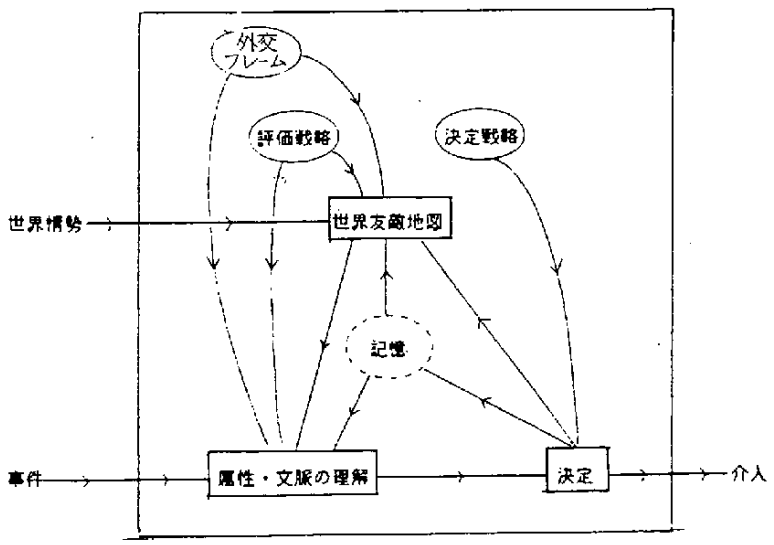


図４－１　中国監視システムにおける情報処理の体系

捜し出し，先例と前提情報に基づいて事件に対する態度決定を行うという仕組みになっている。

本システムで前提情報として設定された中国の外交フレームは，革命外交，現実主義外交という２つの見方があり，それぞれ表４－５に示すような特徴

表４－５　中国の外交フレームの２つの見方と特徴

	革命外交	現実主義外交
目的	世界革命，国内の革命の遂行	生存
国際政治の根本的な見方	紛争的（階級闘争）	紛争的（権力闘争）
国際政治の主体	階級（階級利益を代表する国家、国家内の集団、ＥＴＣ.	主権国家
友	社会主義、共産主義国；民族解放運動	敵の敵
敵	資本主義、帝国主義国；その手先	自国の安全を脅かす国
国際政治で重要な要素	経済的要因；プロレタリア階級への脅威	地政学的要因（勢力範囲など）自国への脅威

を持っている。評価戦略は，国際政治の主体を友か敵かに判断するのに困難を生じた場合に取る方法で，硬直な対応，柔軟な対応という２通りの見方を取ることができる。決定戦略は，複数の選択枝のうちからどれを選ぶべきか困難が生じた場合に取る方法で，攻撃的な選択，慎重な選択の２通りの見方を取ることができる。もちろん，中国の外交政策の見方としてこれ以外の様々な見方を取ることも可能なわけではあるが，本システムではこのような対比しうる２通りの見方を前提とすることにより，中国の紛争介入態度の分析にのぞんでいる。

中国監視システムを用いて分析した事例としては，外交フレーム，評価戦略，決定戦略それぞれ２通りの見方を組み合わせた８通りの見方のいずれが中国の外交態度に合致していたかを調べるため過去の歴史的な事件に対して行ったシミュレーションを挙げることができる。このシミュレーションは図４－２に示すようなフローで解析が行われ，上に述べた８通りの外交態

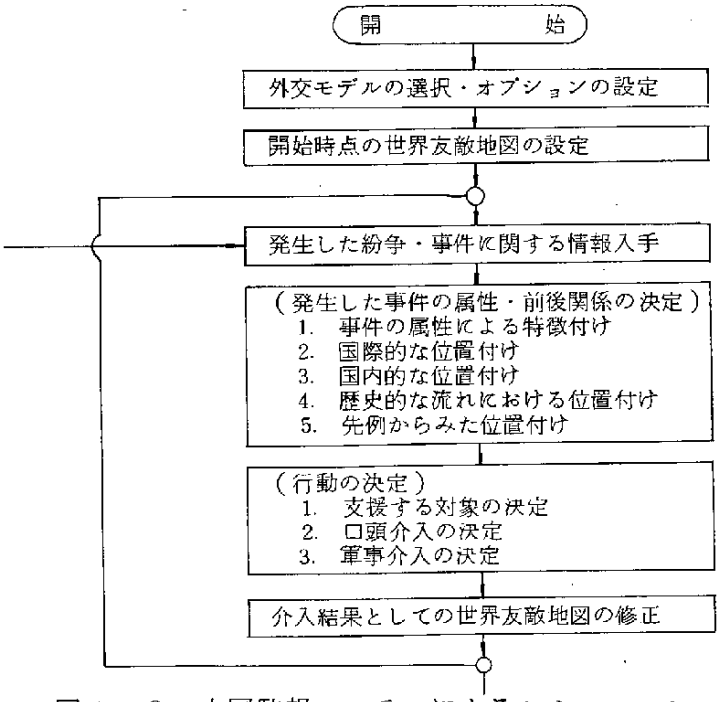


図４－２　中国監視システムによるシミュレーション・フロー

度を取った場合，各紛争・事件に対処した結果としていくつかの国との友敵関係がどのように変化するかが求められている。1949年以降発生した国際紛争を中国に関する8タイプの外交モデルで順次シミュレートしていったわけである。図4－3にシミュレーションの結果として得られた中国とインド

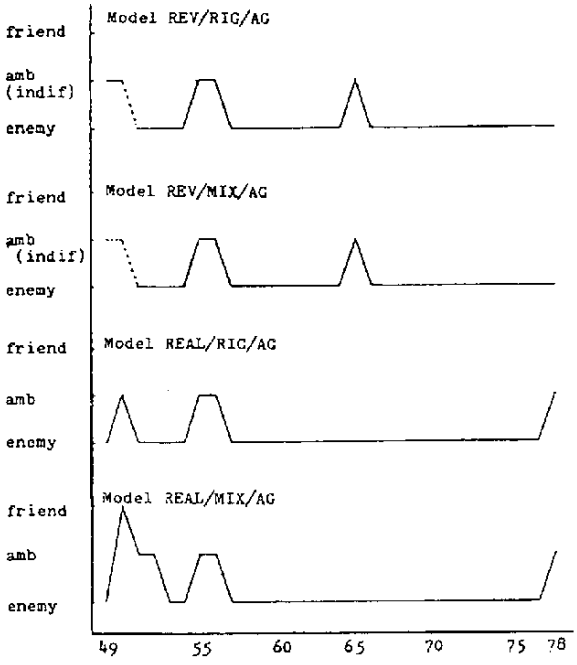


図4－3　中国とインドの友敵関係

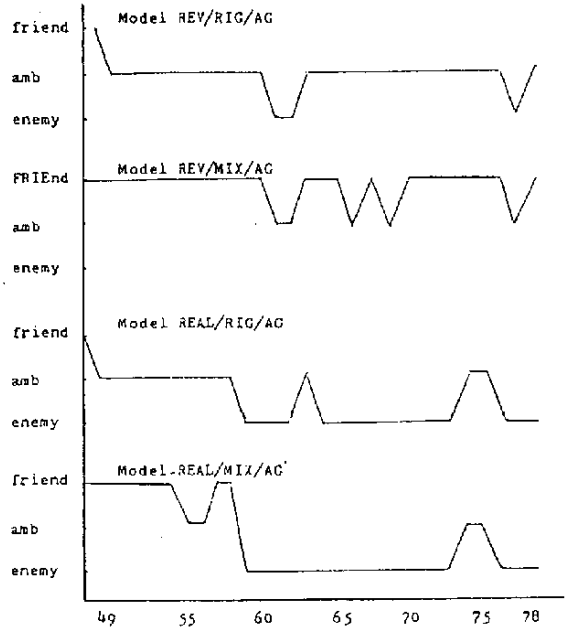


図4－4　中国とソ連の友敵関係

の友敵関係を，図4－4に同じく中国とソ連の友敵関係を示す。1949年から78年までの間に現われた歴史的事実を考え合わせると，中国とインドの友敵関係は図4－3の第4番目の関係が最も合致しているように考えられる。従ってインドに対する中国の外交態度は現実主義外交で硬直的な態度と柔軟な態度の入り混ったものであったと結論付けられる。一方中国とソ連との友敵関係は図4－4の第1番目で説明できる部分もあるし，第4番目で説明で

きる部分もあるということから，硬直的な革命外交と見える面もあるし，硬軟入り混った現実主義外交と見える面もあるということである。

このように中国監視システムは，文章情報を整理・集約したデータをベースとして中国外交の構造に関するきわめて重要な解析結果を提供してくれる興味深いシステムである。特に，中国のような情報がはなはだしくクローズされた国に対して，このようなシステムによる構造解析の利用価値ははかりしれないものがある。中国監視システムの場合の問題点は，システム機能をどこまで汎用化できるかということである。例えば，このシステムをアメリカ監視システムに直そうとした場合同じような構造を持つCACI社のデータがそのまま使えるとしても，外交フレーム，評価戦略，決定戦略といった前提情報が大幅に異なるためシステムのアルゴリズムは全く異なったものとなるであろう。仮りにこのような障壁を越えて無理な汎用化を進めると，各国の独自性が失われたシステムとなる可能性が高い。このような点を考慮すると，データベースあるいは，システムを考える際，どこまでは汎用化でき，どこからは個性を生かさねばならないかの判断がきわめて重要と考えられる。

中国監視システムは，ある紛争が発生した時，中国の介入態度を予知しようという機能ももちろん重要であるが，このシステムを通して中国の外交態度を構造的に組織的に整理できるという点もきわめて重要と考えられる。細かい解析のアルゴリズムや必要となるデータは，取り扱うテーマにより大分異なるかもしれないが，中国監視システムが採用している基本的な考え方は様々な分野へ有効に適用することができると考えられる。エネルギーの分野でも，例えばOPEC加盟国の動向などについて日々の情報からあまり組織立たない方法で現在は分析を行っているが，中国監視システムのようなデータベースと解析システムを整備してOPEC各国のエネルギー政策を構造的に理解することはきわめて重要である。エネルギーの分野では，国際政治学の分野のようにキーとなるベース情報としてどのような情報をデータベース化することが有効か，解析システムのアルゴリズムはどのようであるべきか，

といった諸点についての研究がまだ不十分である。従って，第1に中国監視システムのような有効なデータ利用の方法をいくつか確立し，その後それらをより効率的に運用するための本格的なデータベース整備へと拡張していくことが必要と考えられる。

4.3 General Inquirer のデータ作成法と辞書

4.1節で述べたように，文章情報から必要な情報を抽出して定量化利用のためのベースデータを作成していくいわゆる内容分析の作業を人力を介して行うことは，データベース整備の継続性を考慮した場合，きわめて大きな障壁となっている。1次の文章情報をソース・データとして，コンピュータにより必要な情報を機械的に抽出するシステムが開発できれば，このようなデータベース整備の作業ははかりしれない寄与を受けることになるであろう。1960年代に米国では，このようなシステムの先駆けとして，内容分析を総合的に行なう“General Inquirer”というシステムが開発された。当時のコンピュータの能力を反映して時代をしのばせる制約も多々あるが，このシステムがどのようなデータとどのようなアルゴリズムに基づいて開発されたかを検討することは，内容分析のシステム化を考えていく上できわめて重要である。本節では，まず“General Inquirer”の入力データ作成方法と辞書のあり方について検討する。

“General Inquirer”システムでは，コンピュータによる機械的な作業を可能とするため次のような処理を入力データ作成時に行っている。

① 固有名詞および代名詞の識別

代名詞が前の文章で一体何をさしているかを示すことは，一般の文章解析で常に大きな問題となるところである。“General Inquirer”システムの研究でも，代名詞の指示対象を自動的に識別するアルゴリズムの開発が試みられたが，“it”の解釈に難がありうまくいっていない。そこで次善の手段として入力データ作成時に代名詞のあとに（　　）を付して代名詞が

指しているものの情報を与えている。また固有名詞についても，それがどういう属性のものあるいは人であるかをコンピュータで自動的に解析することはきわめて難しい。“General Inquirer”では代名詞と同じ手法を用いて括弧付きで固有名詞がどんな属性を示すものかを与えている。

② 同じ単語が異なった意味で使われる場合の識別

文章中で複数の意味を持つ単語がどの意味で使用されているかを判断するには前後の意味関係を理解しなければならないので，きわめて難しい。“General Inquirer”では入力データ作成時にそのような単語のうしろに“／”を入れ簡単ないくつかのサブスクリプトを加えて，どの意味で使われているかを示す情報を与えている。例えば，Buffaloの場合，米国の都市名と動物名のどちらにも解釈されるわけであるが，Buffalo／Tとサブスクリプトを加えることにより米国の都市名であることが判然とする方法をとっている。

③ 構文の関係

“General Inquirer”が形成された当時，構文解析の研究もさかんに進められていたわけであるが，システム制作者は研究段階にある構文解析のアルゴリズムを組み込むことを将来の課題としている。それにかわって，入力データ作成時に表4－6のような構文関係を示すための数字サブスクリプトを与えている。

表4－6　構文関係を示す数字サブスクリプト

／1	主語の位置
／3	動詞の位置
／5	目的語の位置
／8	限定主語の位置
／9	限定動詞の位置
／0	限定目的語の位置

このような構文情報を付加するため，入力データ作成者は，複文構造を

持った文章をできるだけ基礎的なレベルの独立節まで分解し，それぞれの文を単独文として“＋”を挿入して分離する必要がある。また“I think”などのように発言内容の背語にかくれている発言主体を示す句は，限定詞と呼ばれ表４－６に示した数字サブスクリプトで特別に表示される。例えば，

John said to Mary that Boston needs rain

という文章は，

JOHN／8 SAID／9 TO MARY／0 THAT BOSTON／1
NEEDS／3 RAIN／5

とコーディングされ，コンピュータにより“THAT”が自動的に＋に変えられて文は２文に分離される。

主として，以上のような入力データ作成手続きが“General Inquirer”システムの入力データには必要である。図４－５にこのようなルールを適用して作成された入力データの例を示す。

“General Inquirer”では，上述のようにして作成された入力データの解読，また“General Inquirer”システムを用いていろいろな分析を行う際に辞書が重要な役割を担っている。“General Inquirer”の辞書は，内容分析用カテゴリー辞書と呼ばれているが，この辞書が例えば英語の標準辞書と異なる点は，システム利用者の社会科学理論に基づく関連性で語義分類がなされ，意味付けが行われているという点である。すなわち分析者が行いたい分析主題に応じて大分類であるカテゴリーが設定され，そのカテゴリーのいずれに属するかという形で単語や慣用句が整理されている。内容分析辞書は，分析者の理論の具体的な表示という性格を持つものである。

1960年代には，様々な分野で17個の内容分析用カテゴリー辞書が作成された。これらの辞書は，非常に特殊な分析を目指して限定されたカテゴリーのもとで整理された辞書と社会科学の一般理論の開発を目指してより汎用的なカテゴリーのもとで整理された辞書との２通りに分類される。Harvard

```
EXAMPLE OF SYNTAX EDITED TEXT.    LETTERS FROM JENNY.   LETTERS 146 AND 147.
                                    JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY.

14620 MY DEAREST GIRL.
14621 I/1 HAVE DECIDED/3 NOT TO SLEEP/3 ALL OF THIS WINTER. + AS I/1 DID
14621 LAST, BETTER DIE/3 AT ONCE. SO I/1 HAVE STARTED MY STUDY/3 OF GRECIAN/S
14621 HISTORY/S ALL OVER AGAIN. I/1 HAVE ALWAYS BEEN A STRONG/1 ADMIRER/1
14621 OF THE GREEKS/S, THEIR WARS/S, DRAMAS/S, LITERATURE/S AND ART/S. THIS
14621 HISTORY/S I/1 AM SO MUCH IREAC/31 STARTS OFF ON THE HELLENES/S. BUT
14621 THAT IS FAR ENOUGH BACK FOR ME. ROSS/1R GLORIED/3 IN GREEK/S ELIKE/31
14621 (ART/S1 SCULPTURE/S, + MANY A DAY WHEN IN CHICAGO/C I/1 HAVE CARRIED/3
14621 HIM/SR THROUGH THE ART/3 GALLERIES/T ON MY/1 BACK/1 + HIS/1R DRAWINGS/1
14621 WERE LATER HUNG/3 AT THE STUDENTS/S EXHIBITION/S, + HE/1R ATTENDED/3
14621 THE SCHOOL/S THERE. THE ART/1 INSTITUTE/1 IN CHICAGO/1 IS, IN MY
14621 OPINION, MUCH MORE/1 BEAUTIFUL/1 THAN THE METROPOLITAN (MUSEUM/11 HERE.
14622 I/1 INTEND/3 TO GO/3 TO THE METROPOLITAN/S (MUSEUM/S) OFTEN THIS
14622 WINTER + THEY (MUSEUM/1) HAVE/3 LECTURES/S, AND ,,GALLERY/S TALKS/S,,
14622 ALMOST EVERY DAY. MY/1 REASON FOR NOT GOING/3 MORE LAST WINTER IS
14622 BECAUSE + THEY ARE (LECTURE/3) MOSTLY DELIVERED BY WOMEN/1. + AND
14622 I/8 DO NOT BELIEVE/9 NATURE EVER INTENDED WOMEN/3 FOR THAT (SHOULD/3
14622 LECTURE/3) PURPOSE. THEY/1 (WOMEN/1) GRIN/1 AND LAUGH/1 TOO MUCH +
14622 I/1 (DISLIKE/31 CANNOT BEAR ,,SMILERS/S,, + WE HAVE DOZENS OF THEM HERE
14622 IN THIS PRISON, WHAT THEY PLAN TO DO ABOUT GRINNING AT US A MYSTERY
14622 TO ME.  LADY M.

14720 MY DEAREST GIRL.
14721 WHAT A PERFECTLY/1 LOVELY/1 SHIRTWAIST/1 (CLOTHING/11, IT CAME YESTERDAY
14721 YOU MIGHT HAVE HEARD MY ,,OH,, OF SURPRISE, THERE. IT IS MY TASTE
14721 EXACTLY, AND THE STRANGE PART IS (CLOTHING/1) THAT IT IS A PERFECT/1
14721 FIT/1. BECAUSE OF MY/1 GORILLA/1 LIKE LONG/1 ARMS/1 I/1 HAVE ALWAYS
14721 HAD TO BUY/3 MY/S SHIRTWAISTS (CLOTHES/S) TOO LARGE/S, + OTHERWISE
14721 THE SLEEVE/1 WOULD BE UP/3 TO MY/S ELBOW/S, + BUT THIS (CLOTHES/11 ONE
14721 IS JUST RIGHT/1.
14722 I/1 WENT/3 OUT EARLY ON CHRISTMAS (HOLIDAY/1) MORNING + SOON AS THE
14722 PRISON/1 DOORS/1 WERE UNLOCKED/3.  AND + (I/1) STAYED/3 OUT/3 ALL
14722 DAY. THE CHRISTMAS (HOLIDAY/1) SHOW/1 AT RADIO/1 CITY/1 WAS JUST ABOUT
14722 PERFECT/1, ALL EXCEPT THE ADMISSION PRICE. + IT ALWAYS MAKES ME/1 SORE
14722 (ANGRY/11 TO SEE/3 THE PRICES/S RAISED/S ON THE VERY DAY + WHEN PEN/1
14722 WANT/3 TO TAKE/3 THEIR/S CHILDREN/S, FAMILY/S AND FRIENDS/S TO THE
14722 GREATEST (HALF/T IN THE SHOW/T LINE THAT CAN BE PRODUCED. IF THOSE
14722 GRAFTERS WERE (CRIMINALS/1) REALLY CHRIST/1 LIKE/1 THEY WOULD LOWER/3
14722 THE PRICES/S + SO THE WHOLE/1 FAMILY/1 COULD/3 ENJOY/3 IT (SHOW/S1.
14722 IT MUST BE TERRIBLE/1 FOR A FATHER/1 TO HAVE TO LEAVE/3 OUT/3
14722 (BOY/S GIRL/S1 LITTLE/S JOHNNY/S OR MAY/S BECAUSE HE HAD COME TO THE
14722 END OF HIS DOLLARS. NO WONDER MEN/1 STEAL/3.
14723 I/1 WAS ALONE/1 ALL DAY, + (I/1) NEVER OPENED/3 MY/S LIPS/S TO A SOUL/7.
14723 WHAT SHALL I/1 WISH/3 YOU/S8 IN THE NEW.YEARS DOUBTLESS YOU/10 WILL
14723 GET/3 ALL/S THAT IS COMING TO YOU IN ONE WAY OR ANOTHER, + AND THAT
14723 MY/1 WISHES/1 WILL NOT HAVE A (NOTHING/1) THING TO DO WITH IT. I/8 CAN/9
14723 ONLY SAY AGAIN WITH THE IMMORTAL WILLIAM (POET/11 ,,SEE WHAT IS BEST,
14723 THAT BEST/S I/1 WISH/3 IN THEE/S8...  LADY M.
```

図４－５　構文及び代名詞を編集したテキスト例

Ⅲ Psychosociolosical Dictionary は，後者の代表例である。この辞書は社会心理学用として構築されたもので，83のカテゴリーのもと約3500項目の単語，慣用句が整理されている。カテゴリーの構成は人称代名詞を包括するものとして

SELF : I , me , mine , my , myself

SELVES : we , us , our , ours , ourselves

OTHER : you , your , yours , yourself , yourselves

が与えられている。he とか she とかいった男性代名詞，女性代名詞はカテゴリー化されず，man , woman , husband などのように性別で区分される名詞と共にMALE－ROLE（男性の役割），FEMALE－ROLE （女性の役割）という2つのカテゴリーに入っている。その他の役割指示としては，

JOB－ROLE（職業の役割，lawyer , magician , mayorなど）やNEUTER-ROLE（中性の役割，ladolescent , acquaintance , neighbor , foreignerなど）などがある。表4－7に示すようにこのような役割基準での分類の他，Harvard Ⅲ Dictionaryは制度基準と地位基準による分類を持っており，単語・慣用句はさらに細分化されることになっている。例えば，“doctor”という単語は，役割基準からはJOB－ROLEのカテゴリーに，制度基準からはMEDICALのカテゴリーに，地位基準からはHIGH－STATUSのカテゴリーに属している。Harvard Ⅲ Dictionaryを例として，内容分析用カテゴリー辞書の性格の一端を紹介したが，かなり一般性を持たせて作成してあるとはいえ，社会心理学の分野で有効に働らきうるようにするために設定されたきわめて特殊なカテゴリーの組み合わせで成り立っているようである。

表4－7　Harvard Ⅲ Dictionaryのカテゴリー基準

役割基準 （4）	MALE－ROLE(男性の役割)，FEMALE－ROLE(女性の役割) JOB－ROLE(仕事としての役割)，NEUTER－ROLE(中性の役割)
制度基準 （12）	ACADEMIC(学界)，ARTISTIC(芸術)，COMMUNITY(地域) ECONOMIC(経済)，FAMILY(家族)，LEGAL(法曹) MEDICAL(医療)，MILITARY(軍)，POLITICAL(政治) RECRETIONAL(レジャー)，RELIGIOUS(宗教) TECHNOLOGICAL(技術)
地位基準 （3）	HIGHER(高)，PEER(同等)，LOWER(低)

“General Inquirer”システムは，内容分析の作業をコンピュータを用いて自動的に行うシステムとして開発されたものであるが，すでに述べたように入力データ作成の過程でなお多くの人力による作業を必要としている。確かにこのようなデータ加工を加えれば，4.4節で述べるように，入力デー

タをコンピュータで処理し，様々な解析を行うことができるが，継続的に入力データ作成作業を行っていく上ではやや負担が大きいと言わねばならない。“General Inquirer”が開発された時代からはすでに２０年が経過しており，その後構文解析などの自動化についてはかなりの進歩があったわけであるが，自然文から直接内容分析を行うにはまだ溝が深く，“General Inquirer”と同じようにテキスト・データの加工を加えなければ確実な内容分析はできないと考えられる。

“General Inquirer”システムのために開発された辞書の中で Harvard Ⅲ Dictionaryなどは社会心理学の分野での一般化をめざしたものであると言われているが，かなり特異な色彩を持っているように感じられる。その点で内容分析用カテゴリー辞書の性格は，シソーラスの性格とも大部異なっているようである。すでに述べたように内容分析用カテゴリー辞書は分析理論が大きく反映された構成を取るわけであるが，はたしてどの程度汎用性を持たせることができるか疑問である。むしろ広範な課題を織り込まず分析課題ごとに必要な内容分析カテゴリー辞書を用意することが内容分析に関しては望ましい姿かもしれない。たとえある限られた分析課題についての内容分析カテゴリー辞書を作成するとしても，より一般的な分析理論に裏打ちされた辞書を作成することは膨大な作業を要すると考えられる。従って，システム利用の有効性をパイロット的に十分検証してから，このような基礎知識ベースの開発に大きな予算をかけて精力的に取り組むことが必要と考えられる。

4.4 General Inquirer の解析手法と応用例

4.3節では，“General Inquirer”データ作成方法と内容分析用カテゴリー辞書について述べたが，本節ではシステムの解析方法と応用例について述べる。“General Inquirer”システムによって入力データは，以下の手順で処理されていく。

① タギング：内容分析カテゴリーの割り当て

タグ付け操作は，このシステムによる内容分析の第１段階にあたっている。コンピュータは加工された入力データを始めから終わりまでスキャンし，文章を一定のルールでまず単語に分割する。その後，内容分析用カテゴリー辞書の中の単語との対応あるいは慣用句のチェックが行われ，各単語に対するタグ付けが行われる。もし構文関係の情報を単語が持っている場合は，タグとともに構文コードも蓄積されるようになっている。またタグ付けのプログラムはカテゴリーの指定と同時にタギングの頻度数の蓄積も行なっている。

② テキスト及びタグのリスティング

作成したタグのリスティングが次に行なわれる。テキストは一連番号を付して文章単位で左側にリストアップされ，その右側にはタグが列記されて出力する。このリストを点検することにより，分析者は単語および慣用句に正しいタグ付けが出来たか否かを確認することが可能である。またこのリストを点検することによって，これまで予測していなかった同時出現のパターンを発見することも多々ありうる。

図４－６には，大統領就任受託演説の一部について，タグ付け操作を行

```
        SENTENCE  25      TOTAL WORDS=  22      IDENTIFICATION 364

THE NEEDS WE SEEK TO FILL, THE HOPES WE SEEK TO          URGE            SELVES          ATTEMPT         SIGN-ACCEPT     SIGN-ASCEND
REALIZE, ARE NOT CUPS ALONE, THEY ARE THOSE OF OUR       URGE            SELVES          ATTEMPT         SIGN-ACCEPT     SENSE
PEOPLE.                                                  NOT             SELVES          QUAN-REF        OTHER           QUAN-REF
                                                         SELVES          LARGE-GROUP     COMMUNITY

        SENTENCE  26      TOTAL WORDS=   8      IDENTIFICATION 364

MOST AMERICANS WANT MEDICAL CARE FOR OLDER               QUAN-REF        OVERSTATE       NEUTER-ROLE     POLITICAL       URGE
CITIZENS.                                                MEDICAL         GUIDE           STRE-REF        HIGHER-STAT     NEUTER-ROLE
                                                         POLITICAL       PEER-STATUS

        SENTENCE  27      TOTAL WORDS=   4      IDENTIFICATION 364

AND SO DO I.                                             SELF

        SENTENCE  28      TOTAL WORDS=   9      IDENTIFICATION 364

MOST AMERICANS WANT FAIR AND STABLE PRICES FOR           QUAN-REF        OVERSTATE       NEUTER-ROLE     POLITICAL       URGE
FARMERS.                                                 GOOD            UNDERSTATE      IDEAL-VALUE     SIGN-STRONG     QUAN-REF
                                                         ECONOMIC        JOB-ROLE        TECHNOLOGICL

        SENTENCE  29      TOTAL WORDS=   4      IDENTIFICATION 364

AND SO DO I.                                             SELF

        SENTENCE  30      TOTAL WORDS=  12      IDENTIFICATION 364

MOST AMERICANS WANT A DECENT HOME IN A DECENT            QUAN-REF        OVERSTATE       NEUTER-ROLE     POLITICAL       URGE
NEIGHBORHOOD FOR ALL.                                    SOCIAL-PLACE    FEMALE-THEME    FAMILY          SOCIAL-PLACE    COMMUNITY
                                                         QUAN-REF        OVERSTATE

        SENTENCE  31      TOTAL WORDS=   4      IDENTIFICATION 364

AND SO DO I.                                             SELF
```

図４－６　テキスト及びタグリストの作成例（大統領就任受託演説）

い，リスティングした例を示す。

③　タグ・タリー手順

タグ・タリー手順は，各タグがドキュメント全体の中でどの位多く指定されているかを示す統計値を生み出すものである。このタグ・タリー手順は，分析者が特定の引用文やイメージに関心を持っている場合のために，単語カウントに置き換えることも可能である。

また，分析者が一つのドキュメント内部で複数の主題を研究したい場合のために，文章カウント手順も用いることができる。以上述べたタグ・タリー手順，単語カウント手順，文章カウント手順の３種の集計法を用いて頻度数に関する多くの情報を得ることができる。タグ・タリー手順では，構文における役割の相違を別々にカウントできるようにもなっている。タグ・タリー手順には，グラフ化機能も着いている。

④　検索および同時出現テスト

検索プログラムでは，調査者がキーパンチした１つあるいは複数の設問に基づいて，設問事項とマッチする文章を磁気テープ上のタグ付きテキスト内から捜す作業を行っている。タグ情報の与え方によってきわめて広範な検索が可能となっている。与えられる検索文は，

```
    +01S   +   43V   =   P
(SELF AS)  (COMMUNICATE)
(SUBJECT)  (AS VERB     )
```

といった形で与えられ，主語としてのSELF，動詞としてのCOMMUNICATEをタグとして持つ文が検索される。Pは検索結果処理のためのオプションの一つであり，Pの場合適合する文がその識別コード及び文番号と共にプリントされる。

以上のような手順を経て，ドキュメント内の情報は整理され，様々なタグで検索されたり，統計量の計算が行われたりする。“General Inquirer”の応用例としては，図４－７に示すような例を挙げることができる。

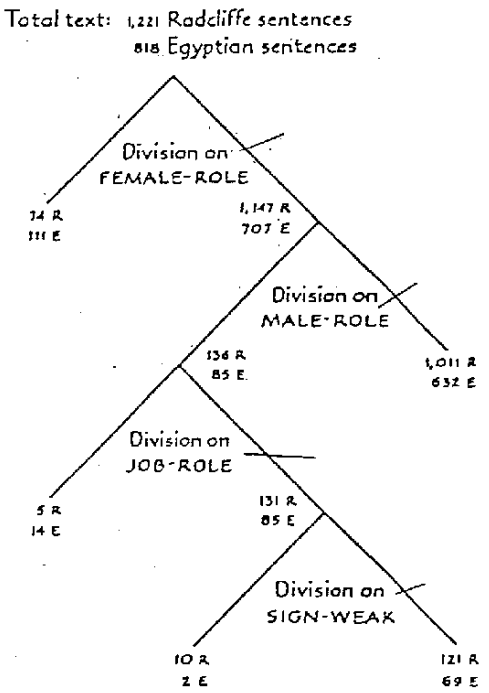


図４－７　General Inquirer によるトリー分析

これは，Radcriffe と Egyptian の未来自叙伝を入力データとして比較したものである。文章の総数は，Radcriffe$^{(R)}$ が 1221 件，Egyptian$^{(E)}$ が 818 件である。まず，それぞれの文を Female－Role のタグで分割した場合，Rの文では 74 文が，Eの文では 111 文が Female－Role のタグに帰属しているということである。帰属しなかった文を Male－Role で分割すると Rの 1147 件のうち 136 件が Male－Role のタグに帰属し，Eの 707 件のうち 85 件が Male－Role のタグに帰属している。また Male－Role に帰属したものを Job－Role で分割すると Rの５件と Eの 14 件がそれぞれ Job－Role のタグに帰属し，Rの 131 件と Eの 85 件が帰属しないことがわかる。このように文章を構成する文の内容をタグでもって分類していくことにより，文章の内容構造に関する有益な情報が得られるわけである。

以上述べてきたように，“General Inquirer” も文章情報の定量解析とい

う意味で大変興味深い情報を提供してくれる。このような解析システムがすでに２０年も前に検討され，開発されていたことは非常に驚異的なことである。入力データ作成後の解析のためのアルゴリズムは現在考えたとしてもさして変わらないものとなるであろう。すでに述べたように“General Inquirer”の難点は，入力データの加工作業のうち機械的にできない判断の部分を結局人力に依存していることと，内容分析用カテゴリー辞書の作成がやはり人力による大がかりな作業となることである。内容分析用カテゴリー辞書の作成は，言ってみれば分析課題に適合したキーワード体系を作る作業と似たようなものである。“General Inquirer”システムは，解析手法をできるだけ一般的な共通のアルゴリズムに整理し，システムで取り扱う主題の特殊性はできるだけ内容分析用カテゴリー辞書へ持ち込んでいると言うことができよう。従って，辞書の構成は取り扱う主題に合わせた片寄った内容のものとなることをある程度宿命付けられているように感じられる。どこまでの領域は共通の内容として盛り込むことができ，どこからはオリジナリティを生かさなければならないか，辞書をできるだけ汎用的に作成したいと考えた場合，充分考慮すべき問題である。

“General Inquirer”システム機能として本節で挙げられたものはどちらかというと，解析のためのベースとなるデータ（タグの頻度数など）を提供するものである。このベース・データをもとにして，さらに各種の解析を行うアプリケーション・システムが必要と考えられるが，その次元にいたると取り扱う主題によって有効かつ適切な解析手法は大きく異なってくるのではないかと思われる。

こうしたデータベース，知識ベース，内容分析システムの一般性，共通性という問題を考えた場合，１次のソースの蓄積としてのデータベース，１次のソースから内容分析によってベースデータを作り出すための知識ベースとシステムはある程度一般性・共通性を持たせることができるように考えられるが，ベースデータに基づいたアプリケーションはそれぞれの特殊性が出て

くるように思われる。いずれにしても，核となるデータベース，知識ベース，内容分析システムを構築するには，膨大な労力と膨大な費用を要するので，各方面から利用可能性について綿密な検討を加えることが必要である。

4.5　認知構造図手法におけるデータ作成法と解析法

文章情報の半定量化利用法として，最近認知構造図という手法が開発された。この方法は，分析対象者の発言あるいは文書等から一定の手続きで分析データを作成し，分析対象者の思考構造を明らかにし，その行動を予測しようというものである。分析データの作成作業は従来から行われている内容分析に近いものである。認知構造図による分析は最近国際政治学の分野で活発に行われているが，1973年の石油危機以来世界中の関心を集めてきたエネルギー問題に関しても適用可能性を持つ手法である。そこで本節では，まず，認知構造図のデータ作成方法および解析手法について検討を加えてみることにした。

認知構造図の一次情報となっているのは，新聞・雑誌など公刊された分析対象者の発言あるいは文書である。インタビュー記事などこれら一次の情報から一文一文を判断し，因果関係を述べた文を抜き出して，原因概念と結果概念を抽出して正負の関係を付けたものが，認知構造図手法の分析対象データとなる。表4－8に実際のデータ作成例を示す。表中①の文に対して，原因概念として抽出されたものは「各国とも，今後関税は非常に低くなる」で，結果概念として抽出されたものは「関税交渉としては東京ラウンドが今世紀最後のものになる」である。原因概念と結果概念の関係は正(+)である。このような原因概念，結果概念と関係とを該当する文からすべて抽出しておいて共通な概念に整理したものが認知構造図におけるコンセプトである。上述の原因概念はそのままであるが，結果概念は「関税はこれから先大きな交渉の対象にはならない」に整理されている。

このような文章を「原因概念／連鎖／結果概念」に分解して整理するわけ

表４－８　認知構造図用の抽出データ例

(本文)

牛場　ガットの交渉は元来関税交渉なんですが①関税交渉としては、あなたが言われたとおり、おそらく今世紀最後のものかもしれません。各国とも今後、関税は非常に低くなりますからね。今度の交渉の特徴は、いわゆる非関税障壁の問題で、これについてコード（規約）をつくるということです②非関税障壁というのは、一番はっきりしているのはクォータつまり量的輸入制限なんですけど、それ以外にもいろんなものがある。今度はだいぶその範囲が広がりまして、たとえば、いままではガットのルールにはなじまないと思われていた政府調達にまで取り組んでいこうということで、相当野心的な交渉をやったわけです。したがって全部が全部これで解決したという具合にはなかなかいかない。関税交渉は確かにもうあまりないと思いますけれども、そういうコードづくりの交渉はこれからも継続的に起こってくるんじゃないですか。

(抽出データ)

１３７	２２
１Ｂ：各国とも今後、関税は非常に低くなる。　／＋／	１Ａ：関税交渉としては東京ラウンドが今世紀最後のものになる。
１３８	１３９
１Ｃ：非関税障壁の規約の範囲がだいぶ広がった。（政府調達にも取り組む）　／－／	１Ｄ：非関税障壁の問題の全部が全部、解決する。

であるが，一次データからコーディング作業は，現在までの所，手作業で行われており，機械化できるような有効なアルゴリズムは見つかっていない。また，コーディング作業に関しては，文型を１６のパターンに分類し，それぞれのケースで概念の抽出・関係付けにどのような注意が必要かということが詳しくコーディング・ルールとしてまとめられている。表４－９に文型パターンを示す。コーダーは，このコーディング・ルールに従って，できるだけ自身の主観を混えず原文に忠実にデータ作成を行わなければならない。とはいえ，主観を混えずということは，きわめてむつかしいのでコーダー間で一定の一致率が得られるまでは訓練を行うことが必要であろう。

このようにして抽出されたデータを，分析者が考えている要素に整理したものがコンセプトである。コンセプトにはＰ－コンセプト，Ｖ－コンセプト，Ｃ－コンセプトの３種類がある。Ｐ－コンセプトは取り得る政策を示すコンセプトで，このコンセプトに影響を与えるコンセプトは存在しない。いわば

表４－９　コーディング・ルールで整理された16の文パターン

1. 単一原因概念／単一連鎖／単一結果概念
2. 単一原因概念／単一連鎖／複数結果概念
3. 複数原因概念／複数連鎖／単一結果概念
4. ～かまたは～の関係
5. 蓋然的な関係
6. 代名詞が原因概念あるいは結果概念である場合
7. 内容分析（文章構造だけでコード化不能の場合）
8. 逆転した因果関係
9. 効用関係
10. 複数主語／単一連鎖／単一目的語
11. 複数主語／単一連鎖／複数目的語
12. 単一主語／複数連鎖／複数目的語
13. 複数主語／複数連鎖／複数目的語
14. 関係の中の概念がもつ二重の役割
15. 連鎖的事象の主張のコード化
16. 無関係という関係の問題

外生変数である。Ｖ－コンセプトは，政策の効果を見るための指標で，Ｖ－コンセプトから他要素へは影響を与えないように設定されている。これら２つのコンセプトは分析者がシミュレーションを行なう際適宜設定することになる。最後のＣ－コンセプトは，社会システムの内部構造を規定するコンセプトである。コンセプト間の関係は，正・負・０の３種に分けて＋１，－１，０という値で設定することができる。その結果として，Valency Matrix というコンセプト間の関係を示すMatrix 表示が得られることになる。

Ｖ（ｉ，ｊ）＝｛＋１ インパクト正，－１ インパクト負，０ インパクト無｝

認知構造図は，このValency Matrixを出発点として解析的手法を用いて分析が進められる。図４－８に解析フローを示す。Valency Matrixを出発点として生のコンセプト間の関係から矛盾のないコンセプト間の関係を認知構造図として求め，Ｐ－コンセプトすなわち政策の影響をＶ－コンセプトで判断し，最も取り得る行動を判断しようというのが基本的な流れである。 Valency Matrix からは，まず最初にRechability Matrix が計算される。これは各コンセプトに影響を与えながらリンクを経由して到達しうるコンセプトがど

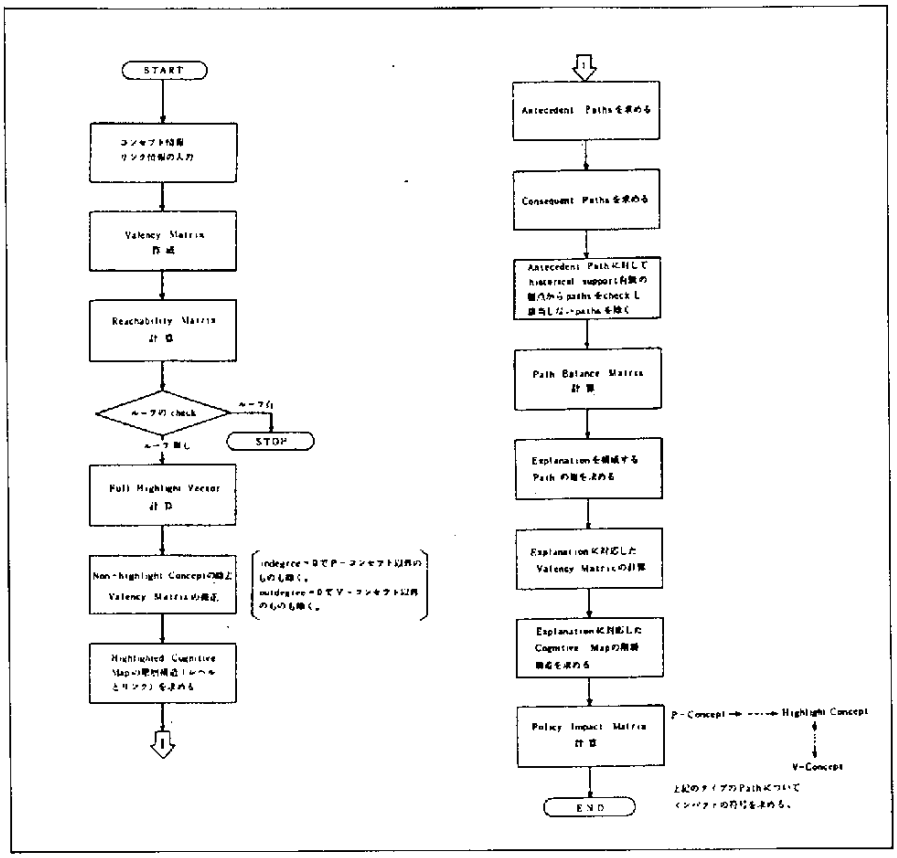


図４－８　認知構造図システムの解析フロー

れであるかを示すMatrix である。値は１（到達可能）か０（到達不能）で示される。

$$R(i,j) = \begin{cases} 1 & : \text{到達可能} \\ 0 & : \text{到達不能} \end{cases}$$

コンセプトｉとコンセプトｊに関して両方向から到達可能な場合は，ループが発生している。ループの発生は階層グラフの解析に大きな困難を生ずるので，一応Ｖ－Matrixからは，この経路をはずして解析する。

次に要素の数が多い場合は，考えている問題のkeyまたは核となるコンセプト（ハイライト・コンセプト）を設定して対象となる要素を整理する必要

がある。そのためにはReachability Matrixから次に示すようなフル・ハイライト・ベクトルを作成する。

$$H(i) = \begin{cases} 1：指定されたハイライト・コンセプト \\ 2：ハイライト・コンセプトに到達可能なコンセプト \\ 3：ハイライト・コンセプトが到達可能なコンセプト \\ 0：ハイライト・コンセプトに無関係なコンセプト \end{cases}$$

このフル・ハイライト・ベクトルの値からハイライト・コンセプトには全く無関係であると評価されたコンセプトは分析の対象から除外する。

このような形で要素を簡素化した後，P－コンセプトからV－コンセプトにいたる様々な経路を決定しなければならないが，その経路の選択基準となる2つの情報がまず決定される。第1はコンセプトの階層レベルを示す情報であるが，次のような方法で決められる。

① 自分に入ってくる矢印を持たないコンセプト（レベル1）

② レベル1のコンセプトとリンクを消した後，自分に入ってくる矢印を持たないコンセプト（レベル2）

③ コンセプトがすべてなくなるまで操作を繰り返す。

第2は，Indegree，Outdegree，Total degreeの計算で

$$id(i) = \sum_{j=1}^{n} |V(i,j)|$$ コンセプトiに直接影響を与えるコンセプトの数

$$od(i) = \sum_{j=1}^{n} |V(j,i)|$$ コンセプトiによって直接影響を受けるコンセプトの数

$$td(i) = id(i) + od(i)$$

によって求められる。Total degreeの情報は，コンセプトの認知の中心性を示す重要な情報となっている。

さて，これだけの準備が整った段階でいよいよ経路を決定していくわけで

あるが，経路の１つとしてまず，Antecedent Paths が求められる。これはハイライト・コンセプトを終点とする様々な経路である。パスの選択は以下のようにして行われる。

① 外部から指定されたハイライト・コンセプトを終点として設定する。

② ハイライト・コンセプトに直接影響を与えるコンセプトを抽出し，この中から td の高いものを選択する。

③ 選択したコンセプトについて，上の手順を繰り返し，さかのぼれなくなるまで探索する。

④ このようにして一つのパスが決定したら，パスにユニークな関係を Valency Matrix から消去する。

⑤ 上記①から④の手順を繰り返して，見付けられるパスがなくなるまで探索する。

⑥ 可能な antecedent paths に対して Historical Support の有無をチェックする。Historical Support の全く無い場合はパスから落す。

Consequent Path は，ハイライト・コンセプトを始点とする様々な経路であるが，パスの探索方法は，Antecedent Path と同様の方法を用いる。ただし，Historical Support のチェックは行わない。

このようにして求められたパスの中には，図４－９に例示するように，通

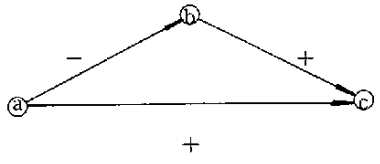


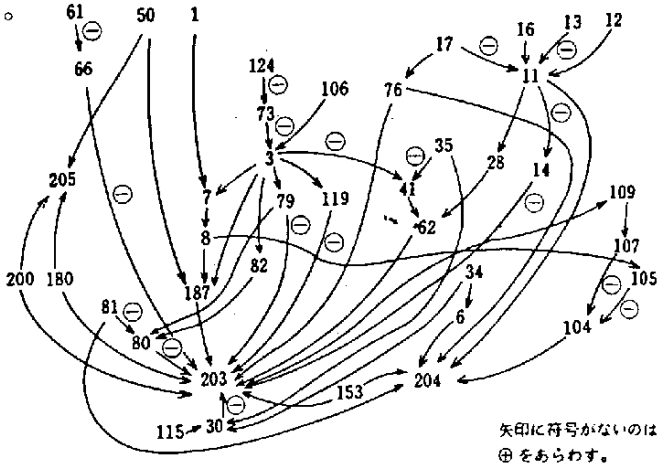
図４－９　Path 間の imbalance

る経路によってインパクトの符号が異なる状況が発生する。この状態は，imbalance と呼ばれる。すべてのパスについて imbalance が生じていないかをチェックした結果として，Path Balance Matrix が作成される。

$$B(i,j) = \begin{cases} 1 : \text{Pathi と Path j に imbalance 無} \\ 0 : \text{Pathi と Path j に imbalance 有} \end{cases}$$

Path Balance Matrix をチェックして，imbalance のない経路だけを選んだものが，Explanation と呼ばれるものであり，分析のための認知構造図の土台となる。Explanation で用いられている要素とパスについてのみ，Valency Matrix を設定して階層グラフを求めると，各Ｐ－コンセプトが各種のＶ－コンセプトにどのようなインパクトを与えるかを決定することができる。

図４－１０は，牛場信彦氏が日米間の貿易摩擦について，インタビューに答えた記事や論文として書いたものから，上述の方法に従って作成した認知構造図である。



1. ケネディ・ラウンド	35. 日本人の考え方が変らない	106. 日本商品の競争力の高さ
3. 日本の黒字	41. 日本のイメージ・アップ	107. 日本の複雑な流通機構
6. 非関税障壁の撤廃要求	50. 南北問題の解決	109. 日本の伝統的な流通機構
7. 出し払い交渉（出血）	61. 日本と後進国との関係のよさ	115. 日本社会のわかりにくさ
8. 先進国間の新秩序	62. 日本の自己発言の機会	119. 日本の立場の弱さ
11. 東京ラウンド	66. 日本の孤立化	124. 米国のインフレ
12. アメリカの主導権	73. 円安	153. 日米の農産物交渉
13. 73年の石油危機	76. 日米関係の密接化	180. 総合的見地からの日本外交
14. 保護主義	79. 米国で日本がunfairと思われていること	187. 世界貿易の秩序ある発展
16. ストラウス	80. 米国からの貿易不均衡是正の圧力	200. 日本と新興工業国の自由貿易を原則にした共存共栄
17. （米国）政府より議会の方がつよい	81. 交渉の成功の顕示	203. 日本の効用
28. 規格・関税評価・輸入手続のコードを決める	82. 日米間の問題	204. アメリカの効用
30. 日本が閉鎖社会だという批判	104. 米国の商品輸出の増大	205. 世界の効用
34. 日本に物が売れない		

図４－１０　牛場信彦氏の認知構造図

牛場氏の国際経済に対する考え方が構造的に表現されていて大変興味深い。

さて，認知構造図の応用範囲であるが，次のような利用法が考えられる。

① 一人の政策決定者の認知構造を明らかにする。

② 与えられた認知構造図に基づいて，一定の状況の下での政策決定者（国家）の行動を予測する。

③ いく人かの政策決定者（国）の一定の問題についての認知構造図を作成し，比較する。

④ ある国（政策決定者）の認知構造図の時間的な変化を追うことによって対外政策の構造的な変化を解明する。

このように認知構造図は幅広い利用法が考えられ，国際関係の中でも，外交問題，経済問題，エネルギー問題と非常に多様な分野で使用しうる可能性を持っている。

しかしながら，認知構造図手法を利用するにあたっては，次のような限界があることを踏まえた上で使用することが必要である。

① 原文に現われた主張がホンネかタテマエかという問題。ただし，主張がたとえタテマエであっても，行動者は拘束されることを考えると必ずしも現実離れした結果とはならない。

② 政策決定者の因果主張が客観的にみて妥当か否か。

③ 組織あるいは交渉といった因子が含まれていない。

④ 発言の文脈とシミュレーションの時設定するシナリオの対応関係が必ずしも明確でない。

⑤ シミュレーションの妥当性を吟味する方法が十分確立されていない。

⑥ モデルのセンシティビティ・テストの手法が十分開発されていない。

また，認知構造図のデータ作成は，大量の情報処理を行うにはあまりに人的労力と時間を費やすものである。認知構造図の手法が各種の分析に有効であることが確認されたならば，認知構造図解析をより組織的に行えるようデータ作成面でもデータ処理面でも有効な機械的手法を確立することが必

要と考えられる。

4.6　認知構造図手法のエネルギー分野への適用実験

4.5節では，認知構造図の解析手法を詳しく紹介し，日米貿易摩擦（自動車輸出問題）について，牛場信彦氏の発言を分析した例を示した。この認知構造図の分析手法は，エネルギー分野における様々な問題に対しても，興味深い情報を提供しうる可能性を持つ有力な手法の一つと考えられる。

そこで，図４－８に示した解析フローに沿って4.5節で示した解析原理に基づくミニコン用の認知構造図解析システムを開発し，適当な題材を選んでエネルギー分野への認知構造図手法の適用実験を試みた。本節では，解析結果を示すとともに，実験の結果わかった様々な問題点について検討を加えたい。

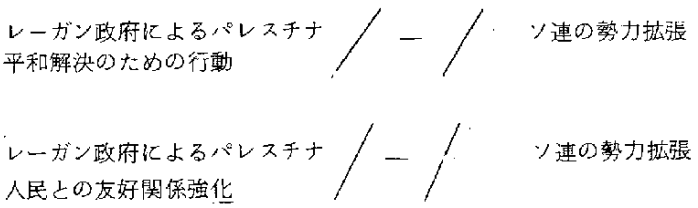
認知構造図解析のための原材料としては，サウジアラビアのヤマニ石油相が，１９８１年２月にベルギーのルーベン大学で講演を行った後，記者団の質問に応じたインタビュー記事（Middle East Economic Survey 1981年２月５日号）を使用した。この記事は，単に石油問題だけでなく，発足まもない米レーガン政権への期待，パレスチナ問題，エネルギー問題，オイルマネー問題，イラン・イラク戦争などサウジアラビアを取り囲む広範な問題について，ヤマニ石油相の言及がなされている点が特徴となっている。4.5節で示した認知構造図データのコーディング・ルールに従って，できるだけ客観性を保つように注意しながらデータ作成を行った。図４－１１に部分的にではあるが，コーディングの結果を示す。

このようにして，一文一文から抽出した原因概念と結果概念を整理して求められたのが表４－１０に示すコンセプト群とリンク情報である。パレスチナ問題，エネルギー問題，石油問題，オイルマネー問題，イラン・イラク戦争などに関連した４１個の認知コンセプトとサウジアラビアの効用，消費国の

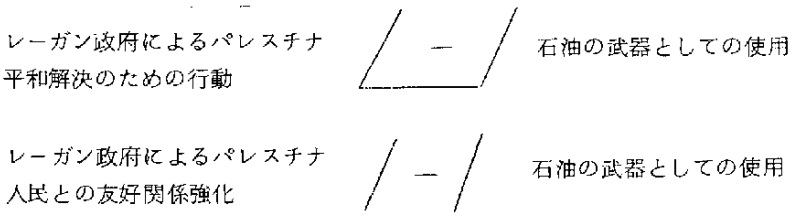
Q: What initiatives are you hoping to see from the new Reagan administration in the US, and would you be willing to use the oil weapon again this year if you are disappointed in his initiatives?

A: Let us concentrate on the first part of your question.

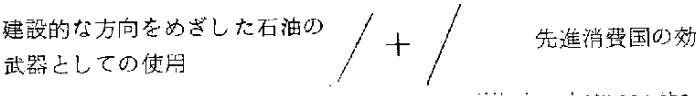
We hope that the Reagan administration will do what should be done in order to bring about a peaceful settlement in the area which will force the Israelis to leave the occupied territories, give the Palestinians their homeland and their state and their own flag, and ensure the safety and security of all the people in the area.
We also hope that the Reagan administration will strengthen their friends in the area; and we hope that this administration will be an active one to take decisions when they are needed, so that we don't see the Russians gaining ground every day.



Now, there are some strong indications that these hopes are not a mirage, that there might be something in the air to realize what we need. And therefore I don't think we need to talk about the oil weapon, which is something that we hate to discuss.



We want to keep using our so-called oil weapon in a positive manner, the way I described in my speech today, in a constructive way, to help the consumers rather than to let them suffer.



Q: Do you think that we are reaching a better equilibrium between the oil producing and consuming countries and between the developed and developing countries?

図４－１１　認知構造図のコーディング例

A: I think that what the major consumers did between 1979 and today is very remarkable. I think that if you carry on and continue what you are doing right now, you will definitely avoid any unnecessary hardship, any disaster which might arise from an energy crisis. You already reduced your consumption in 1980 by something more than 7.5% compared to 1979, and you are supposed to reduce further your consumption in 1981 compared to the low level in 1980. Your investments in alternative sources of energy are remarkable. Your efforts to conserve on oil are most appreciated, and we hope you carry on.

先進国によるエネルギー消費の縮小化の実施と継続 / − / エネルギー危機のもたらす災害と不必要な苦境

先進国による代替エネルギー源への投資の実施と継続 / − / エネルギー危機のもたらす災害と不必要な苦境

Now, our worry is that you might in the future, in 1982 when we will have a surplus in the supply of oil which means that the price of oil will be floating in the market at a lower level, you might relax.

石油の供給過剰 / + / 石油価格の低下

石油の供給過剰 / + / 先進国によるエネルギー対策活動のゆるみ

表４－１０　ヤマニ石油相の発言から抽出したコンセプトとリンク情報

番号	コンセプトの内容	リンク情報
C1	米国のパレスチナ平和解決への行動	+C3,−C7,−C8
C2	米国とパレスナ人民との友好関係強化	−C7,−C8
C3	パレスチナ問題の解決	+V3,+V5
C4	エルサレムの復帰	+V3
C5	南北の対話	+C35
C6	アラブとイスラエルの争い	−C3,+C7
C7	石油の武器としての使用	+C3,+C5,+C11
C8	ソ連の勢力拡張	−V1,−V4
C9	消費国による省エネルギーの実施	+C16,+V4,+V5
C10	消費国による代替エネルギーの開発	+C16,−C32,+V4,+V5
C11	石油危機の発生	+C9,+C10,+C12,+C18
C12	消費国による資源探査活動	+C16
C13	石油の供給過剰	+C14,+C17
C14	石油価格の低下	+C17
C15	石油価格の上昇	+C9,+C10,+C12,+C18
C16	エネルギー危機の解決	+V1,+V4,+V5
C17	エネルギー対策のゆるみ	+C15
C18	石炭・原子力の増加	+C16,−C32
C19	サウジの西欧への投資	+C31,+V4
C20	ヨーロッパの国際的重み	+C3,+C21
C21	米国への圧力	+C1,+C2
C22	主要消費国のスポット買い控え	+C14,+C23
C23	石油市場の秩序	+V1,+V2,+V4
C24	イラン・イラク戦争の発生と継続	+C11,−C23,+C26
C25	イラン・イラクの原油生産・輸出の再開	+C13
C26	中東における政治事件の発生	+C11,−C23
C27	OPECの存続	+C23,+C35,+V2,+V5
C28	発展途上国との友好関係	+V1,+V2,+V5
C29	石油の安定供給	+C28,+V1,+V2,+V4,+V5
C30	イラン革命	−C23,+C24,+C26

番号	コンセプトの内容	リンク情報
C31	金融市場の安定	+C41, +V1, +V5
C32	主要エネルギー源としての石油の立場・継続	+C5, +C27
C33	オイルダラーの集積	+C19, +C27, +C35, +C40
C34	サウジ自身の必要性を越えた石油生産	+C28, +C33, +C38, +C39, −V1, +V4, +V5
C35	発展途上国の成長	+V5
C36	サウジの巨大な原油埋蔵量	+C37
C37	サウジの石油需給調整機能	+C29, +C38, +V2
C38	サウジのOPEC内外における役割	+C33, −V1, +V2, +V4
C39	消費国のサウジに関する関心	+C22
C40	サウジの太陽エネルギーへの投資	+C16, +V1
C41	石油の価値安定	+V1, +V2
V1	サウジアラビアの効用	
V2	OPECの効用	
V3	イスラム世界の効用	
V4	消費国の効用	
V5	世界の効用	

効用など5個の価値コンセプトが抽出されている。これらコンセプト間のリンク情報を図示したものが図4－12に示すヤマニ石油相の発言による認知構造図である。ただし，図4－12では後で述べる政策コンセプトも挿入してある。図をみるとわかるように色々なコンセプトが複雑に絡みあっていて一目見ただけでは，流れを理解するのはきわめて難しい。この情報が認知構造図解析システムの入力データとなるものである。

開発したシステムは4.5節で述べた解析原理をアルゴリズム化したもので，FORTRANで書かれており，ミニコンでも使用できるように配列等の設定に注意を払っている。開発したプログラムのテスト・ランは，4.5節で述べた文献に出ている簡単な例題と牛場信彦氏の認知構造図を使用して解析結果が再現されることで確認した。

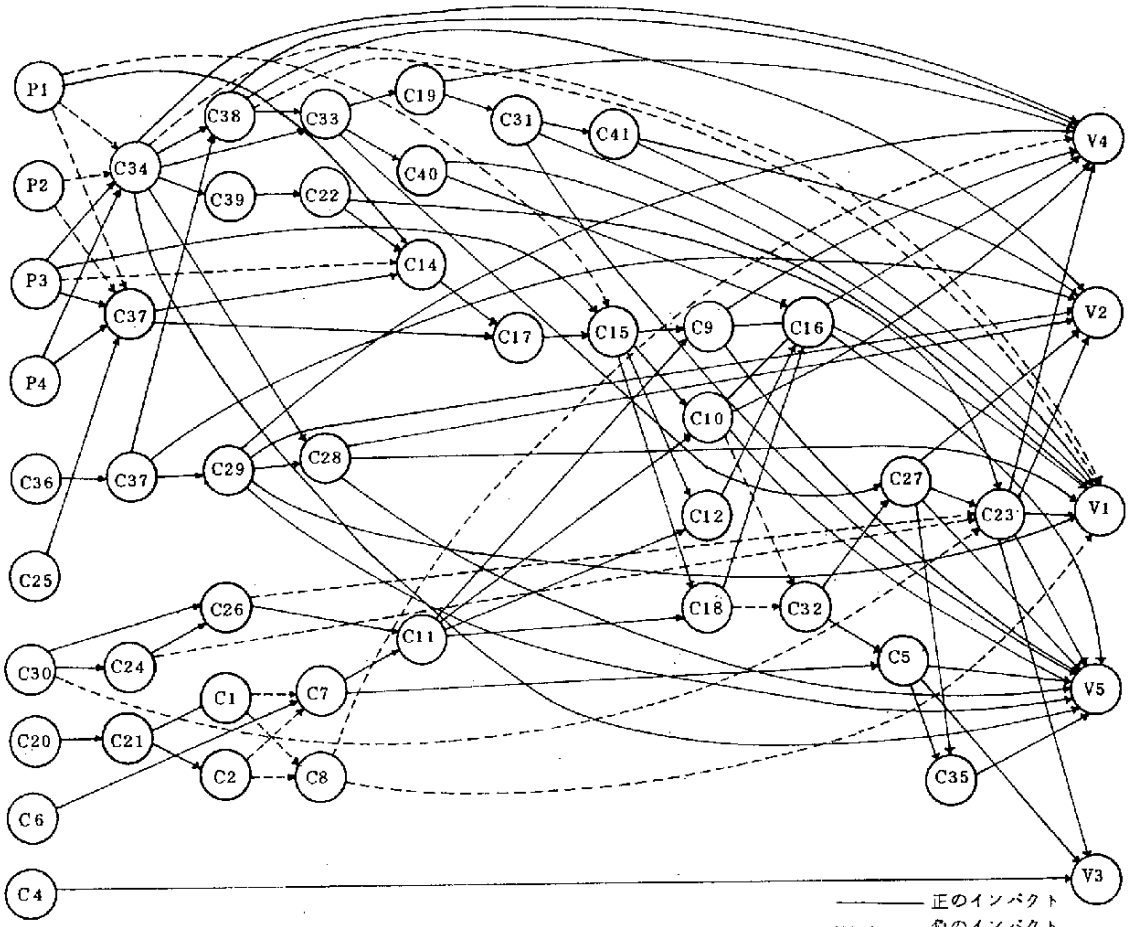


図４－１２　ヤマニ石油相の発言による認知構造図

次に，図４－１２に示した認知構造図を解析するシミュレーション分析の実験を行った。シミュレーションを行うための前提としては，世界的な原油供給過剰と値下げ圧力という事態に直面した場合を想定して，サウジアラビアの取り得る政策として，

P１：原油を減産し，原油価格を下げる。

P２：ただ原油の減産のみを行う。

P３：おかまいなく原油を増産し，石油価格を上げる。

P４：原油を増産する。

の４政策を考え，それぞれの政策コンセプトは，次のようなインパクトを認知コンセプトに対して与えると考えた。

P１：－C13,+C14,－C15,－C34

P２：－C13　　－C34

P３：+C13,+C15,－C14,+C34

P４：+C13　　+C34

また，ハイライト・コンセプトとしては，次の３つを設定した。

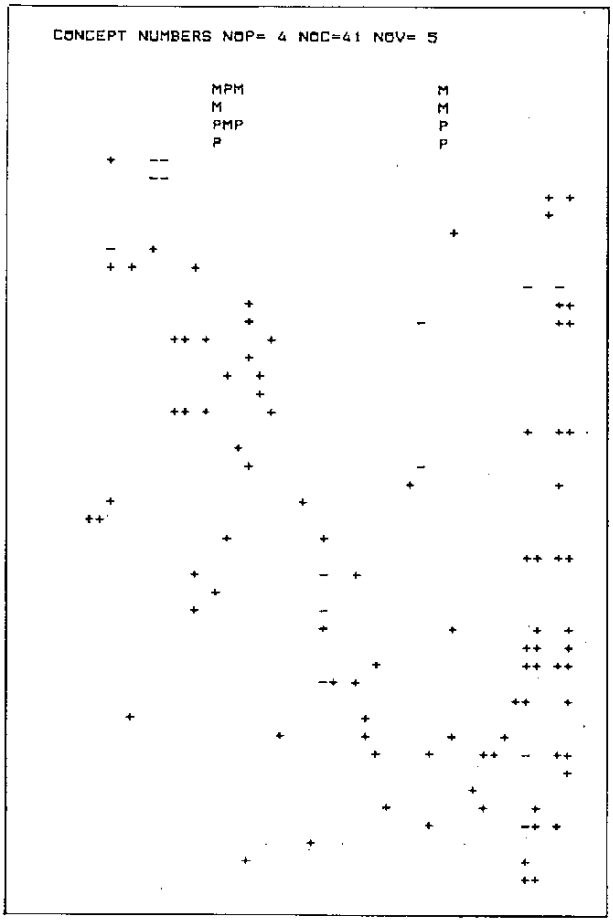
C16：エネルギー危機の解決

C23：石油市場の秩序

C29：石油の安定供給

開発した認知構造図解析システムの入力データは，コンセプトの数，コンセプトのリンク関係，ハイライト・コンセプトの数と番号である。図４－13にコンセプト間のリンク情報の入力データを示した。これは図４－１２の認知構造図をいわば行列表示したものである。この入力データをもとに，次のようなステップで解析は進められる。

① Valency Matrix，Historical Support Matrix の計算



P：正のインパクト（歴史的裏付け有）
M：負のインパクト（歴史的裏付け有）
＋：正のインパクト
－：負のインパクト

図４－１３　リンク情報の入力データ

② Reachability Matrix の計算

③ 与えられたコンセプトに対するハイライト・ベクトルの計算と無関係なコンセプトの除去（図４－１４）

```
HILIGHT VECTOR AND CONCEPT

          5        10        15        20
20  2 2 2 2 2 2 0 0 0 2 2 0 2 2 2 2 2 2 2 1
27  2 2 2 2 2 2 0 0 0 2 2 0 0 2 2 0 2 2 2 0
33  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

          25        30        35        40
    2 2 0 2 2 2 0 2 2 2 2 0 0 2 2 0 0 2 2 0 2
    2 2 0 2 2 2 1 2 2 2 2 0 0 2 0 2 2 2 2 0 2
    0 0 0 0 0 0 0 0 0 0 0 0 3 1 0 0 0 0 0 0 2

          45        50
    2 2 2 2 0 3 0 0 3 3
    2 2 2 0 0 3 3 0 3 3
    2 0 0 0 0 3 3 0 3 3

ELIMINATION OF NON-RELATED CONCEPTS
CONCEPT NO.  7 HAS BEEN ELIMINATED
CONCEPT NO.  8 HAS BEEN ELIMINATED
CONCEPT NO.  9 HAS BEEN ELIMINATED
CONCEPT NO. 10 HAS BEEN ELIMINATED
CONCEPT NO. 12 HAS BEEN ELIMINATED
CONCEPT NO. 23 HAS BEEN ELIMINATED
CONCEPT NO. 24 HAS BEEN ELIMINATED
CONCEPT NO. 29 HAS BEEN ELIMINATED
CONCEPT NO. 34 HAS BEEN ELIMINATED
CONCEPT NO. 35 HAS BEEN ELIMINATED
CONCEPT NO. 39 HAS BEEN ELIMINATED
CONCEPT NO. 40 HAS BEEN ELIMINATED
CONCEPT NO. 45 HAS BEEN ELIMINATED
CONCEPT NO. 48 HAS BEEN ELIMINATED
```

図４－１４　ハイライト・ベクトルとコンセプトの除去

④ Indegree，Qutdegree，Total degree の計算

⑤ 階層グラフの計算（図４－１５）

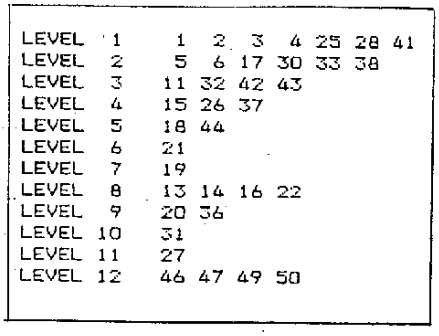


図４－１５　コンセプトの階層レベル

⑥ Antecedent Path の計算（求まったパスは145個）

⑦ Consequent Path の計算（求まったパスは14個）

⑧ Path Balance Matrix の計算

⑨ Explanation の決定（図４－１６）

⑩ Policy Matix の計算（図４－１７）

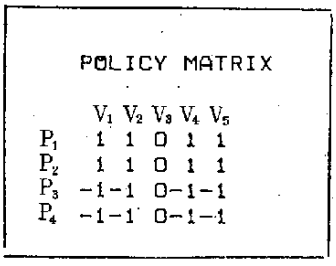
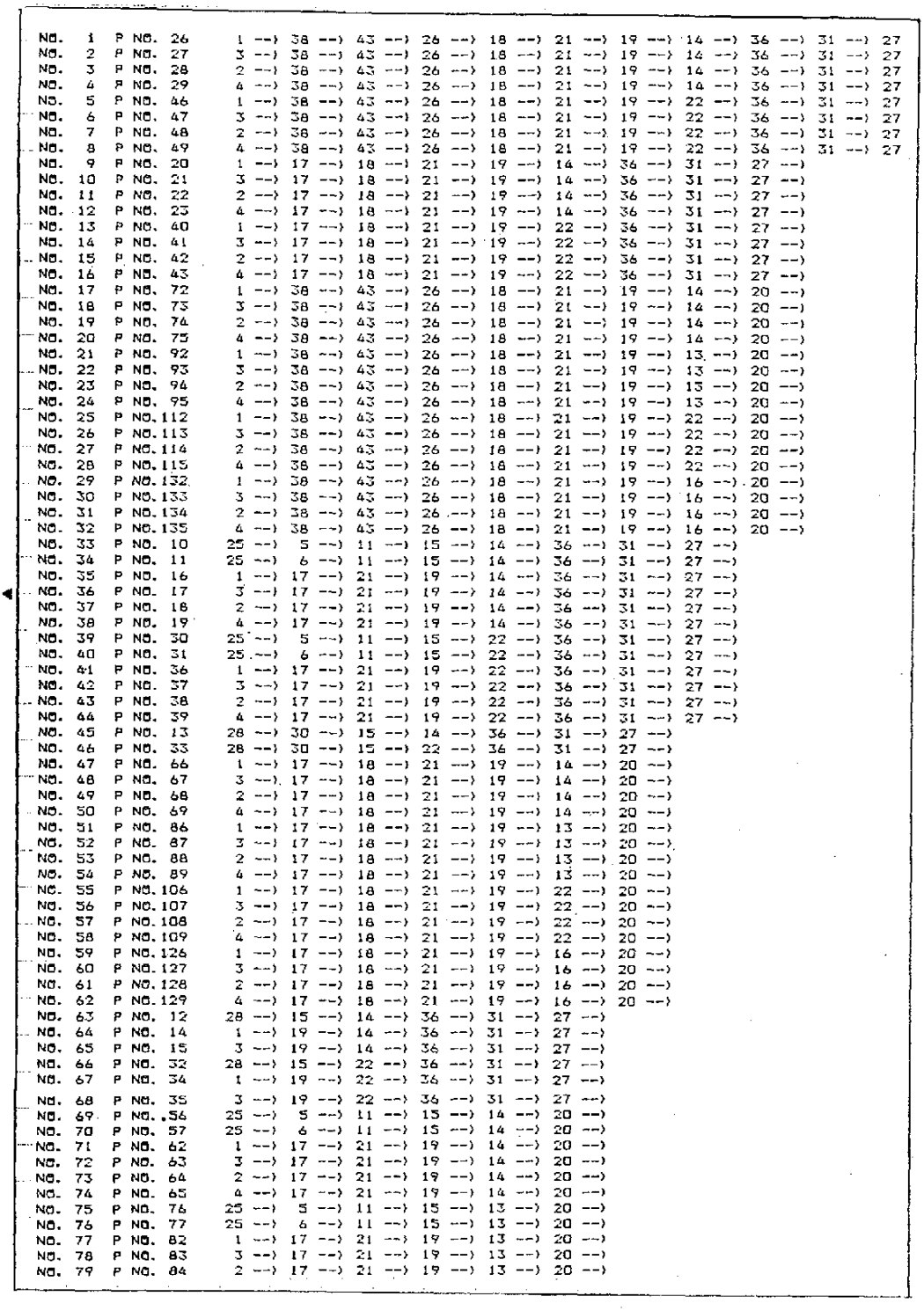


図４－１７　Policy Matrix

図４－１６に示した Explanation が，このシミュレーションの結果求められた考え方の経路であり，これを図４－１２と同じような形で図示したものが図４－１８である。もとの入力に用いた認知構造図に比べるとかなり単純化されているが，それでもまだ複雑である。この認知構造図を決定する過程では，Antecedent Paths と Consequent Paths を別々の経路として分離して考え，Path Balance Matrix を求めてパス間のインバランスを取り除いた Explanation を決定しているが，実はこの方法では完全にインバランスを



NO. 1 P NO. 26 1 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27
NO. 2 P NO. 27 3 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27
NO. 3 P NO. 28 2 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27
NO. 4 P NO. 29 4 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27
NO. 5 P NO. 46 1 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27
NO. 6 P NO. 47 3 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27
NO. 7 P NO. 48 2 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27
NO. 8 P NO. 49 4 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27
NO. 9 P NO. 20 1 --> 17 --> 18 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27 -->
NO. 10 P NO. 21 3 --> 17 --> 18 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27 -->
NO. 11 P NO. 22 2 --> 17 --> 18 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27 -->
NO. 12 P NO. 23 4 --> 17 --> 18 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27 -->
NO. 13 P NO. 40 1 --> 17 --> 18 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27 -->
NO. 14 P NO. 41 3 --> 17 --> 18 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27 -->
NO. 15 P NO. 42 2 --> 17 --> 18 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27 -->
NO. 16 P NO. 43 4 --> 17 --> 18 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27 -->
NO. 17 P NO. 72 1 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 14 --> 20 -->
NO. 18 P NO. 73 3 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 14 --> 20 -->
NO. 19 P NO. 74 2 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 14 --> 20 -->
NO. 20 P NO. 75 4 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 14 --> 20 -->
NO. 21 P NO. 92 1 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 13 --> 20 -->
NO. 22 P NO. 93 3 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 13 --> 20 -->
NO. 23 P NO. 94 2 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 13 --> 20 -->
NO. 24 P NO. 95 4 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 13 --> 20 -->
NO. 25 P NO.112 1 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 22 --> 20 -->
NO. 26 P NO.113 3 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 22 --> 20 -->
NO. 27 P NO.114 2 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 22 --> 20 -->
NO. 28 P NO.115 4 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 22 --> 20 -->
NO. 29 P NO.132 1 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 16 --> 20 -->
NO. 30 P NO.133 3 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 16 --> 20 -->
NO. 31 P NO.134 2 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 16 --> 20 -->
NO. 32 P NO.135 4 --> 38 --> 43 --> 26 --> 18 --> 21 --> 19 --> 16 --> 20 -->
NO. 33 P NO. 10 25 --> 5 --> 11 --> 15 --> 14 --> 36 --> 31 --> 27 -->
NO. 34 P NO. 11 25 --> 6 --> 11 --> 15 --> 14 --> 36 --> 31 --> 27 -->
NO. 35 P NO. 16 1 --> 17 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27 -->
NO. 36 P NO. 17 3 --> 17 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27 -->
NO. 37 P NO. 18 2 --> 17 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27 -->
NO. 38 P NO. 19 4 --> 17 --> 21 --> 19 --> 14 --> 36 --> 31 --> 27 -->
NO. 39 P NO. 30 25 --> 5 --> 11 --> 15 --> 22 --> 36 --> 31 --> 27 -->
NO. 40 P NO. 31 25 --> 6 --> 11 --> 15 --> 22 --> 36 --> 31 --> 27 -->
NO. 41 P NO. 36 1 --> 17 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27 -->
NO. 42 P NO. 37 3 --> 17 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27 -->
NO. 43 P NO. 38 2 --> 17 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27 -->
NO. 44 P NO. 39 4 --> 17 --> 21 --> 19 --> 22 --> 36 --> 31 --> 27 -->
NO. 45 P NO. 13 28 --> 30 --> 15 --> 14 --> 36 --> 31 --> 27 -->
NO. 46 P NO. 33 28 --> 30 --> 15 --> 22 --> 36 --> 31 --> 27 -->
NO. 47 P NO. 66 1 --> 17 --> 18 --> 21 --> 19 --> 14 --> 20 -->
NO. 48 P NO. 67 3 --> 17 --> 18 --> 21 --> 19 --> 14 --> 20 -->
NO. 49 P NO. 68 2 --> 17 --> 18 --> 21 --> 19 --> 14 --> 20 -->
NO. 50 P NO. 69 4 --> 17 --> 18 --> 21 --> 19 --> 14 --> 20 -->
NO. 51 P NO. 86 1 --> 17 --> 18 --> 21 --> 19 --> 13 --> 20 -->
NO. 52 P NO. 87 3 --> 17 --> 18 --> 21 --> 19 --> 13 --> 20 -->
NO. 53 P NO. 88 2 --> 17 --> 18 --> 21 --> 19 --> 13 --> 20 -->
NO. 54 P NO. 89 4 --> 17 --> 18 --> 21 --> 19 --> 13 --> 20 -->
NO. 55 P NO.106 1 --> 17 --> 18 --> 21 --> 19 --> 22 --> 20 -->
NO. 56 P NO.107 3 --> 17 --> 18 --> 21 --> 19 --> 22 --> 20 -->
NO. 57 P NO.108 2 --> 17 --> 18 --> 21 --> 19 --> 22 --> 20 -->
NO. 58 P NO.109 4 --> 17 --> 18 --> 21 --> 19 --> 22 --> 20 -->
NO. 59 P NO.126 1 --> 17 --> 18 --> 21 --> 19 --> 16 --> 20 -->
NO. 60 P NO.127 3 --> 17 --> 18 --> 21 --> 19 --> 16 --> 20 -->
NO. 61 P NO.128 2 --> 17 --> 18 --> 21 --> 19 --> 16 --> 20 -->
NO. 62 P NO.129 4 --> 17 --> 18 --> 21 --> 19 --> 16 --> 20 -->
NO. 63 P NO. 12 28 --> 15 --> 14 --> 36 --> 31 --> 27 -->
NO. 64 P NO. 14 1 --> 19 --> 14 --> 36 --> 31 --> 27 -->
NO. 65 P NO. 15 3 --> 19 --> 14 --> 36 --> 31 --> 27 -->
NO. 66 P NO. 32 28 --> 15 --> 22 --> 36 --> 31 --> 27 -->
NO. 67 P NO. 34 1 --> 19 --> 22 --> 36 --> 31 --> 27 -->
NO. 68 P NO. 35 3 --> 19 --> 22 --> 36 --> 31 --> 27 -->
NO. 69 P NO. 56 25 --> 5 --> 11 --> 15 --> 14 --> 20 -->
NO. 70 P NO. 57 25 --> 6 --> 11 --> 15 --> 14 --> 20 -->
NO. 71 P NO. 62 1 --> 17 --> 21 --> 19 --> 14 --> 20 -->
NO. 72 P NO. 63 3 --> 17 --> 21 --> 19 --> 14 --> 20 -->
NO. 73 P NO. 64 2 --> 17 --> 21 --> 19 --> 14 --> 20 -->
NO. 74 P NO. 65 4 --> 17 --> 21 --> 19 --> 14 --> 20 -->
NO. 75 P NO. 76 25 --> 5 --> 11 --> 15 --> 13 --> 20 -->
NO. 76 P NO. 77 25 --> 6 --> 11 --> 15 --> 13 --> 20 -->
NO. 77 P NO. 82 1 --> 17 --> 21 --> 19 --> 13 --> 20 -->
NO. 78 P NO. 83 3 --> 17 --> 21 --> 19 --> 13 --> 20 -->
NO. 79 P NO. 84 2 --> 17 --> 21 --> 19 --> 13 --> 20 -->

図4－16　Explanationを構成するパス群

NO. 80	P	NO. 85	4	-->	17	-->	21	-->	19	-->	13	-->	20	-->	
NO. 81	P	NO. 96	25	-->	5	-->	11	-->	15	-->	22	-->	20	-->	
NO. 82	P	NO. 97	25	-->	6	-->	11	-->	15	-->	22	-->	20	-->	
NO. 83	P	NO.102	1	-->	17	-->	21	-->	19	-->	22	-->	20	-->	
NO. 84	P	NO.103	3	-->	17	-->	21	-->	19	-->	22	-->	20	-->	
NO. 85	P	NO.104	2	-->	17	-->	21	-->	19	-->	22	-->	20	-->	
NO. 86	P	NO.105	4	-->	17	-->	21	-->	19	-->	22	-->	20	-->	
NO. 87	P	NO.116	25	-->	5	-->	11	-->	15	-->	16	-->	20	-->	
NO. 88	P	NO.117	25	-->	6	-->	11	-->	15	-->	16	-->	20	-->	
NO. 89	P	NO.122	1	-->	17	-->	21	-->	19	-->	16	-->	20	-->	
NO. 90	P	NO.123	3	-->	17	-->	21	-->	19	-->	16	-->	20	-->	
NO. 91	P	NO.124	2	-->	17	-->	21	-->	19	-->	16	-->	20	-->	
NO. 92	P	NO.125	4	-->	17	-->	21	-->	19	-->	16	-->	20	-->	
NO. 93	P	NO.140	1	-->	38	-->	42	-->	37	-->	44	-->	20	-->	
NO. 94	P	NO.141	3	-->	38	-->	42	-->	37	-->	44	-->	20	-->	
NO. 95	P	NO.142	2	-->	38	-->	42	-->	37	-->	44	-->	20	-->	
NO. 96	P	NO.143	4	-->	38	-->	42	-->	37	-->	44	-->	20	-->	
NO. 97	P	NO. 9	41	-->	42	-->	37	-->	31	-->	27	-->			
NO. 98	P	NO. 59	28	-->	30	-->	15	-->	14	-->	20	-->			
NO. 99	P	NO. 79	28	-->	30	-->	15	-->	13	-->	20	-->			
NO.100	P	NO. 99	28	-->	30	-->	15	-->	22	-->	20	-->			
NO.101	P	NO.119	28	-->	30	-->	15	-->	16	-->	20	-->			
NO.102	P	NO.136	1	-->	38	-->	37	-->	44	-->	20	-->			
NO.103	P	NO.137	3	-->	38	-->	37	-->	44	-->	20	-->			
NO.104	P	NO.138	2	-->	38	-->	37	-->	44	-->	20	-->			
NO.105	P	NO.139	4	-->	38	-->	37	-->	44	-->	20	-->			
NO.106	P	NO.144	41	-->	42	-->	37	-->	44	-->	20	-->			
NO.107	P	NO. 58	28	-->	15	-->	14	-->	20	-->					
NO.108	P	NO. 60	1	-->	19	-->	14	-->	20	-->					
NO.109	P	NO. 61	3	-->	19	-->	14	-->	20	-->					
NO.110	P	NO. 78	28	-->	15	-->	13	-->	20	-->					
NO.111	P	NO. 80	1	-->	19	-->	13	-->	20	-->					
NO.112	P	NO. 81	3	-->	19	-->	13	-->	20	-->					
NO.113	P	NO. 98	28	-->	15	-->	22	-->	20	-->					
NO.114	P	NO.100	1	-->	19	-->	22	-->	20	-->					
NO.115	P	NO.101	3	-->	19	-->	22	-->	20	-->					
NO.116	P	NO.118	28	-->	15	-->	16	-->	20	-->					
NO.117	P	NO.120	1	-->	19	-->	16	-->	20	-->					
NO.118	P	NO.121	3	-->	19	-->	16	-->	20	-->					
NO.119	P	NO. 55	28	-->	30	-->	27	-->							
NO.120	P	NO.157	33	-->	32	-->	50	-->							
NO.121	P	NO.158	33	-->	32	-->	46	-->							
NO.122	P	NO.159	33	-->	32	-->	47	-->							
NO.123	P	NO. 54	28	-->	27	-->									
NO.124	P	NO.145	41	-->	33	-->									
NO.125	P	NO.146	27	-->	50	-->									
NO.126	P	NO.147	27	-->	46	-->									
NO.127	P	NO.148	27	-->	49	-->									
NO.128	P	NO.149	27	-->	47	-->									
NO.129	P	NO.150	20	-->	50	-->									
NO.130	P	NO.151	20	-->	46	-->									
NO.131	P	NO.152	20	-->	49	-->									
NO.132	P	NO.153	33	-->	50	-->									
NO.133	P	NO.154	33	-->	46	-->									
NO.134	P	NO.155	33	-->	49	-->									
NO.135	P	NO.156	33	-->	47	-->									

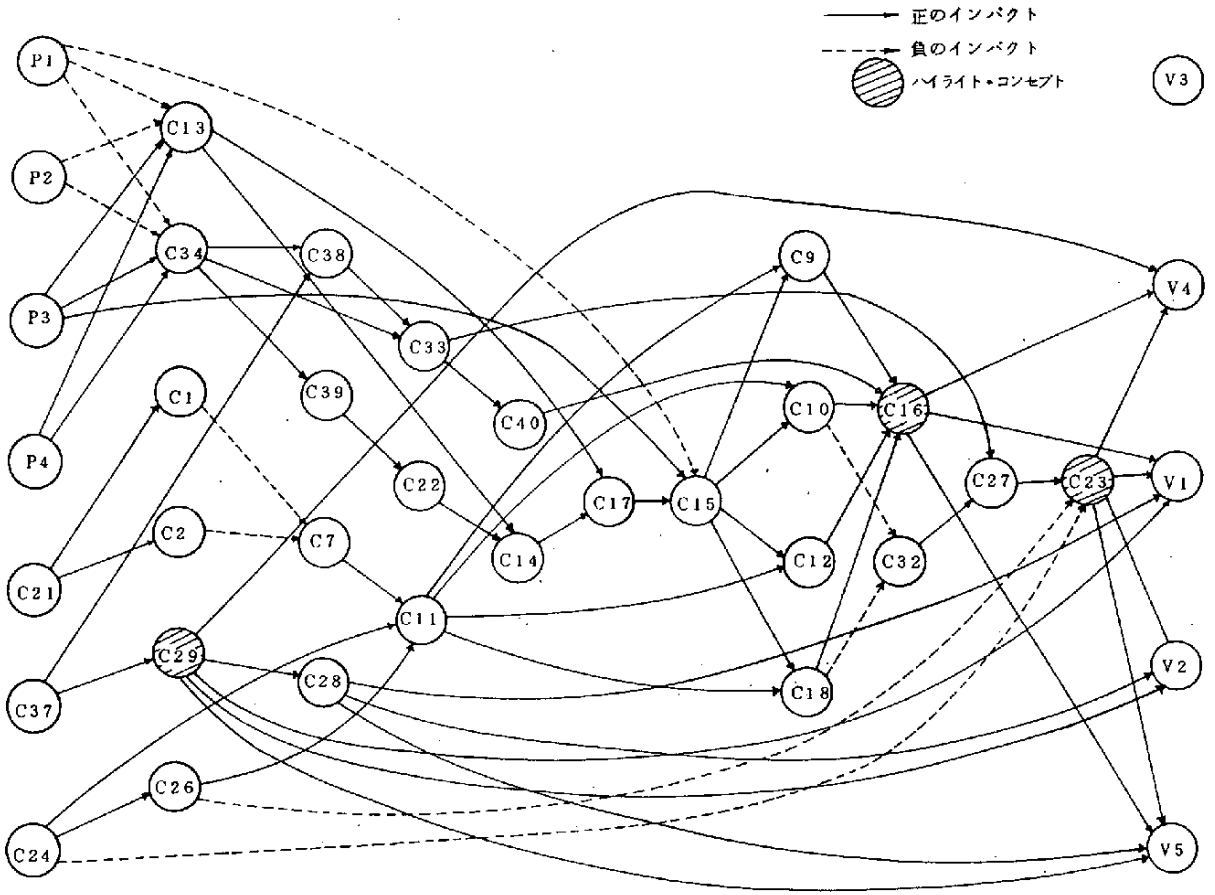


図４－１８　ヤマニ石油相の認知構造図に基づくシミュレーション

A PATH, C PATH 分離ケース

取り除くことができない。例えば，図４－１８でP１ →C１３→C１７→C１５→C１０ →C３２→C２７→C２３→V５の経路とP１ →C１３→C１７→C１５→C１０→C１６→V５の経路とでは付号がまったく逆になることに容易に気付くことができる。このようなインバランスが存在しうるのはハイライト・パスの前後でパスを分けて考えているためである。

そこで，今回開発したシステムでは，ハイライト・コンセプトを通して，Antecendent Path と Consequent Path を結合し，政策コンセプトから効用コンセプトにいたる経路をつなげておいて，そうしたパス間のインバランスを取り除くというアルゴリズムで計算しなおしてみた。その結果認知構造図は，図４－１９に示すようにきわめて見やすいものとなった。

このような方法で確かにパス間のインバランスは取り除けるわけであるが Antecedent Path と Consequent Path の結合を考えた場合，組み合わせはハイライト・コンセプトごとにみた経路数の掛け算で効いてくるので，総経路数が膨大な量となってしまう。今回の場合は，分離ケースで１４５＋１４＝１５９個に対して，結合ケースは４９４個にのぼってしまった。今回システム開発面で苦心しなければならなかったのは，このような大量の行・列を持つマトリックスとミニコンのコアの限界に対する工夫であった。

さて，このようにして求められた認知構造図に基づいて，各政策の効用に対するインパクトを見ると表４－１１のようになる。

表４－１１　政策の効用に与えるインパクト

	サウジアラビアの効用	OPECの効用	イスラムの効用	消費国の効用	世界の効用
原油の減産と原油の値下げ	＋	＋	０	＋	＋
原油の減産	＋	＋	０	＋	＋
原油の増産と原油の値上げ	－	－	０	－	－
原油の増産	－	－	０	－	－

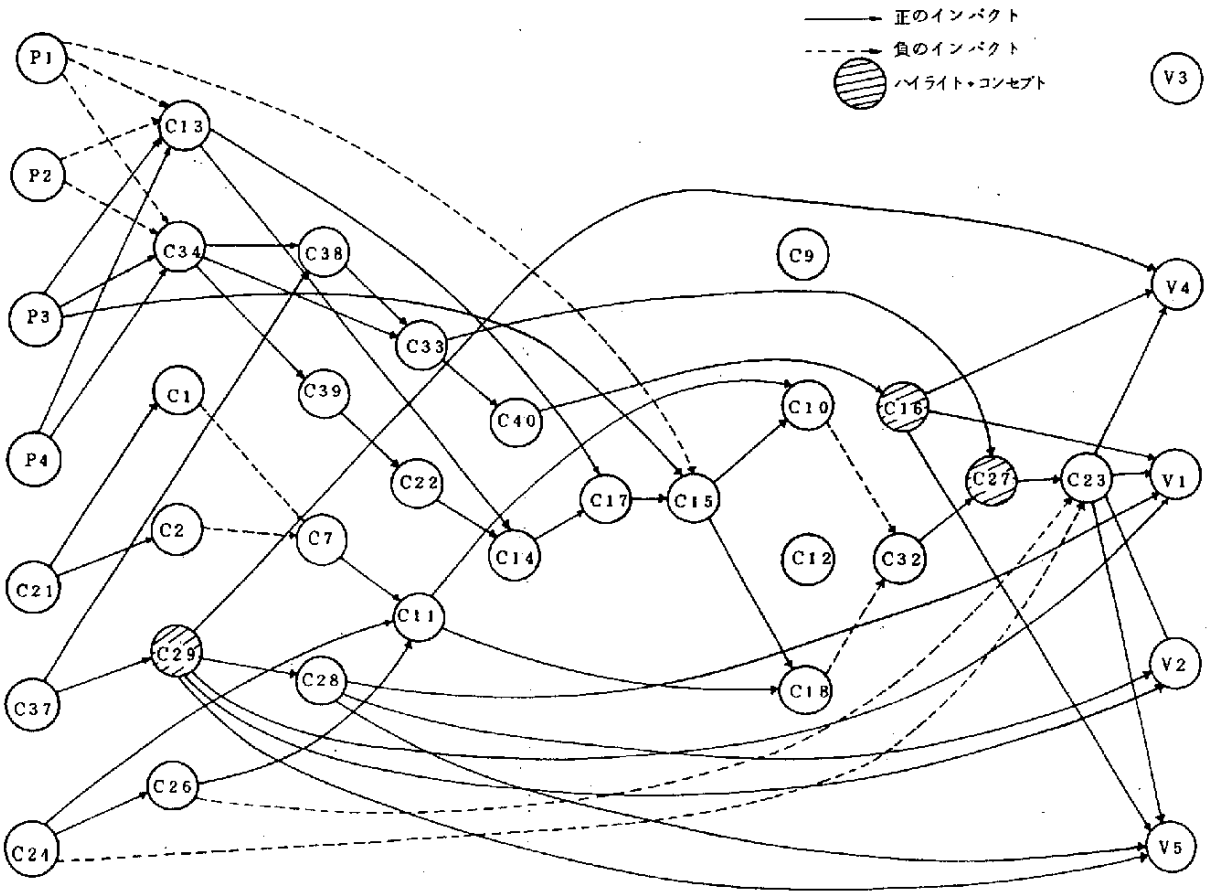


図４－１９　ヤマニ石油相の認知構造図に基づくシミュレーション

A PATH , C PATH 結合ケース

単純にみると，このインパクトの状況から，サウジアラビアが様々な事態に対して取り得る政策はどれかといった推測もできるわけであるが，このインパクトが決まった経路を順次たどってみると必ずしも論理の流れは十分納得のいくものではない。

このような状況にいたった問題点としては，入力データ作成上の問題点とシステム・アルゴリズム上の問題点とに分けて，次のように整理される。

(a) 入力データ作成上の問題点

① Policy コンセプトをどの認知コンセプトと関連付けるかという問題。

② コンセプトそのものの中に正負の概念が入る場合，まず認知コンセプトとして採用するか，次に採用した場合リンクをどのように決めるかという問題。

③ 効用のコンセプトをどのような範囲，レベルで求めるかという問題。

④ コーダーの主観をいかにして取り除くかという問題。

(b) システム・アルゴリズム上の問題点

パスの優先順位を決めるため，コンセプトのTotal degree の合計量を用いているが，コンセプトを数多く通るまわりくどい経路を最優先していないかという問題。

今回の例の場合，入力データ作成作業の過程をもっと吟味すると改善の余地は大きいと考えられるが，上述のような問題に関してワーキング・グループで検討した結果，一般的な対策として次のような点に留意する必要があると考えられる。

① コーダーの主観を取り除くためには，テーマに応じて抽出するコンセプトを標準化し，因果関係（リンク情報）の判断もできるだけルール化することがのぞましい。また，コーダーと分析者は別の人間である方が良い。

② PからVまで通したパスで考えるならば，V－コンセプトは，「ＸＸの効用」といった大きな概念まで設定するのではなく，もう少し細かい次元で考えた方が良い。

③ コーダーは一定の一致率が達成できるようになるまで訓練が必要と考えられる。

④ パスの優先順位の決定は，今回のアルゴリズム以外の方法を見出すことが必要と考えられる。例えば，Historical Support Matrix に歴史的な頻度確率の情報を入れて，それで Weight 付けしてパスの順位を決めるといった方法が考えられる。

⑤ ハイライト・コンセプトを1個ずつ指定して，各ハイライト・コンセプトに対する経路とインパクトを求め，総合的に比較評価するのも一つの方法である。

さて，今回の実験結果は，深く吟味するといろいろな面で多くの問題があるが，入力データ作成時の標準的な方法を確立できれば，かなり改善の余地はあると考えられる。4.5節でも述べたように，認知構造図の手法は，本質的な意味でいくつかの限界を持っているが，その限界を十分認識して使えば原理的にはかなりいろいろな問題に応用できると考えられる。もし，コンピュータ可読な媒体から構文解析，意味解析などの手法を通して，自動的に概念の抽出を行うことができ，インパクトの判断を下すことができるとすれば，コーダーの学習効果による主観的判断を避け，きわめて客観的なデータ作成も可能になると考えられる。このような研究への強いインセンティブを与えるためにも，認知構造図を幅広い分野で利用し，その有効性を確立していくことが必要と考えられる。

4.7　海外における機械翻訳先進事例

文章情報データベースの総合的な利用のためには，国内のみならず，海外情報の有効活用を図る必要がある。これら情報の総合解析システムの研究開発にあたって，言語をより機械的な処理により翻訳して利用する方法を考えることは，さけて通れない関門であると言える。

歴史的にいくつかの言語を用いている国や国際機関等においては，異言語による障害により迅速に，高度に対応する必要上からも，コンピュータによる翻訳の研究が，早期より進められてきた。これら先進的な研究事例を調査し，本プロジェクトのめざすシステム開発の参考とするべく，本年度は欧州を中心に機械翻訳システムと辞書の活用をテーマとする調査を実施した。その成果を機関別に4.7.1～4にとりまとめた。

4.7.1　EUROTRA計画の概要

1980年から検討が行われて来た，多言語間翻訳システムの開発を目的とするEUROTRA計画が11月4日EC理事会で正式に承認された。この計画立案は1978年に始められ80年にはEC委員会の承認を得たものの理事会側から異論が出て，延び延びになっていたものである。従来の自動翻訳の研究やEUROTRA計画策定の過程でグルノーブル大学，ザールブリュケン大学，マンチェスター大学等各国の大学や研究機関の協力をあおいできたが，今後はEC組織がシステムの開発を行うことになった。計画のあらましは5年半の期間に約2,600万ドルの予算で多言語間翻訳システムのパイロットモデルを開発する。予算の内訳は基本部の開発に1,600万ドル各国が負担する周辺部が1,000万ドルとなっている。

対象言語はEC加盟各国言語であるが，将来はアラビア語，日本語の追加の可能性がある。パーサーは意味論解析レベルまでを目標とし，ポストエデットは10%以下を目指す。まず，中核となるソフトウエアを確立し，その上に各言語モデルを周辺に追加する形になると思われる。

EUROTRA計画の目的は，

① ヨーロッパ共同体の公式言語を対象とした機械翻訳システム（EUROTRA）を開発する。

② 計画完了時には限定された分野と限定されたテキストのカテゴリーを対象とするプロトタイプのシステムが作られていること。なお，このシステムはこの計画のあと開発が予定される実用ベースシステムの基礎になりえること。

の2点であり，計画は次の3つのフェーズに分けられている。

(1) 準備フェーズ （2年間 200万ECU）

①第1段階

- ACPM(Advisory Committee on Programme Management)の設立。
- プロジェクトとその組織ならびに関連各国とセンターの責任の明確化。
- 計画実施の方法論の明確化。
- 関連センターで行われる言語関係の研究および対象とするテキストの領域およびカテゴリーの詳細な計画書の準備。
- 各関係者の実際の貢献度に応じた権利の配分と研究成果の発表に関する取決めの定義。

②第2段階

- 言語モデルおよび各処理プロセス（解析，変換，生成）のコンポーネント開発方法の詳細スペックの準備
- いろいろな処理過程（解析，変換，生成，モニタリング，機能，テキスト管理）を処理できるプログラムおよびEUROTRAシステムの基本ソフトウェアに関する詳細かつ拘束的なスペックの準備。
- 語彙データベース（辞書）の詳細なスペックの準備。
- 金銭および関係団体の他の貢献を含めた関係団体との契約についての準備。

以上の第2段階が終わった時点で先に挙げたスペックに対するACPMの意見が述べられるであろう。これは言語研究をより早く進めるためで，その結果できるだけ早くソフトウェア作成に関する入札の募集を広範囲に発行するためでもある。

(2) 基礎および応用言語研究のフェーズ （2年間 850万ECU）

準備フェーズが完了し，ACPMならびにCrestからの助言を受けたあと，2つの部分に分けられた第2のフェーズを実施する。

①基礎言語研究

- ECの公式言語に関する解析，生成および各言語間の変換のための言語モデルの開発。この研究ではある限定された分野の文献と約2,500エントリーからなる語彙をベースとする。
- 上で述べた語彙の辞書準備
- 各処理過程に関し，コンピュータ処理に最も適する言語処理方法の研究

②EUROTRAシステムのための基本ソフトウェアの作成

- 第1フェーズで作成されたスペックのソフトウェア作成に関する入札の招待状発行
- ACPM の助言のあと，できるだけ早い時期に入札に対する応募とEUROTRAシステムの基本ソフトウェア作成機関の選定に関するEC委員会による厳しい検査。
- 選ばれた機関による基本ソフトウェアの開発。なお基本ソフトウェアは，次の条件を満たすこと。
  - ⓐ 言語データとその方法を表現できる高級言語であること。
  - ⓑ ユーザとシステムの間で会話ができる高級言語であること。またこのシステムには，いろいろなモジュールを追加できること。
  - ⓒ データベースの管理及びテストならびに高級言語をコンパイルするためのユーティリティソフトウエア。

ソフトウエアの最初のバージョンは，関係センターによって定義された言語モデルの開発ならびにマシンテストに使う予定である。このソフトウエアの開発は，言語研究を有効にする為に前もって必要な条件である。

EUROTRAの商用システム開発すなわち商用で翻訳するのに必要な効率と信頼性を持ったものの開発は，この計画が完了したあとから着手する。

(3) 言語モデルと評価の確定フェーズ （1年半 550万ECU）

第2フェーズ終了後すなわち，最初の言語モデルのシステマティックなテストを行うことができた時，ACPM,Crest（the Scientific and Technical Research Committee),CIDST（the Committee for Information and Documentation in Science and Technology) Cetilから意見を受け,以下の形に集約されるであろう。

- できるだけ信頼性の高い，言語モジュールを作るために言語モデルを適用する。モジュールは予備使用に適合するものである。
- 特定分野のテキスト文，言語モデル，語彙の基礎を急速に拡大する。そして，テキストについては，複雑なものへと拡張を図る。
- できるだけ早く選んだ分野をカバーする辞書（各言語について2万エントリー）の見直しと拡張。
- システムの技術的評価，経済性の評価。
- 工業レベル（実用）のシステム開発と商用ベースの展開に関するプロポーザルの準備。

以上，EUROTRA計画の具体的内容は第1フェーズで検討されるものであるが，EC本部担当者が持っているシステムのイメージは次の様なものであった。

パイロットモデルとしては，使用分野を限定し，辞書は約2万語のものを考えている。これは，限定した分野では十分な語彙数と考えている。コ

ンピュータはマイクロコンピュータクラスのもので16～32万bitレジスター，メモリー1Mバイトで外部記憶は，20～50Mバイト相当を考えている。また，OSはUNIXで言語はC言語を使用したいと言っている。使用コンピュータはまだ決まっていない。校正率10%以下を目指している。

しかし，校正は分野によって違い，例えば正式文書は現在でも2～3回の校正が行われている。開発体制は，中核となる人員が8名で言語学，ソフトエンジニア，マネージメントの担当に分かれる。また各国語担当として，言語学関係の人が12名，ソフトウエア担当として16名が予定されている。翻訳対象としてECがいちばん関心を持っている分野は，鉄鋼，農業，エレクトロニクス関係で，特にエレクトロニクスについては日本の情報が一番ほしいとのことであった。

4.7.2 ECにおけるシストランの利用

ECでシストランを導入したのは1976年頃であったが本格的な使用を始めたのは1981年3月以降である。現在，英語→仏語，仏語→英語，英語→イタリア語の3つの言語ペアーしか翻訳していないが，英語→独語，仏語→独語の言語ペアーについても昨年4月にインプリメントし，来年3月か4月に稼動する予定である。ECとしてはEUROTRA計画とは別に，シストランシステムをそれなりに利用していく姿勢で，処理量はしだいに増えている。対象分野はECでの要求が多分野であるため，一般的なものになっている。ECにおける機械翻訳の考え方は「機械翻訳システムは使えるか」ではなく，「どのようにECの翻訳機関へシステムを適合させていくか」という観点を重視している。

(1) シストランシステムの利用状況

ECでは英語→仏語，仏語→英語，英語→イタリア語の言語ペアーについて，月間400～1,000 ページ（1ページ250ワード）処理している。英語→仏語のシステムはフランスの航空機会社SNIASの航空機マ

ニュアル翻訳や西独カールスルーエの原子力センターで使われていた特定分野用システムであったものをＥＣの要求である多分野用に一般化したものである。

(2) 対 象 分 野

ＥＣが翻訳対象としている分野は，経済，農業，原子力，金属工学，機械工学，情報処理，エレクトロニクス，航空宇宙と多分野にわたっていて，専門分野に限定していない。

担当者によると，機械翻訳にとって重要なのは，対象分野よりも，ドキュメントの種類であるという。機械翻訳で対象としているドキュメントは，会議に使う技術ドキュメント，マニュアルレポート，メンテナンスマニュアル，データベースの抄録等であり，これらは，内容がわかれば良いというレベルのものである。ＥＣの公式文書は人間の翻訳であっても，更に数回の校正が入るくらいの質の高さが要求され，こうした文書や出版物のように質の高い翻訳を要求されるものは対象としていない。

(3) 機械翻訳の生産性とコスト

機械翻訳システムによる翻訳の生産性は，翻訳者の心理的影響が大きく作用している。最初の機械翻訳導入時は，機械拒否といった心理的抵抗があり，翻訳者の半数しか利用していなかったが，しだいに機械翻訳が理解されていった。これはユーザのニーズを組み入れるため，翻訳者と共同で開発していったためである。機械翻訳（ポスト・エディト含めず）のコストは，１ワード当り0.4～0.35 ベルギーフランである。また，人間による翻訳では，１日１人当り５～８ページであったのが機械翻訳では25ページ（粗い訳の場合。出版物などのように正確に訳すためには10ページ）となっている。担当者は，機械翻訳ではコストも重要であるが，それ以上に翻訳スピードを重視しなければならないという。出版しなければならないものは別として，内容が早くわかれば良いという

レベルのもののように，スピーディな翻訳を要求するものには，機械翻訳が必要である。例えば，ヨーロッパでは，翻訳者が得られないために会議ができないケースがある。西独の原子力センターでは，会議資料をシストランシステムで翻訳し，会議に利用している。

また，フランスのCNRSは，データベース関係の翻訳が多いので機械翻訳を利用している。

(4) 中央処理装置とソフトウエア

最初IBM で次にシーメンスのコンピュータを使ったが，現在はＩＢＭ370／158 を使い従来の２～３倍の能力となり，１時間当り，35万ワード処理できる。ユーティリティを含め，プログラムは，アセンブラ言語で書かれており，約100Kステップである。このうちアナリシス部はマクロ，アセンブラーで２Kステップである。プログラム（ロジック）と辞書と比較するとロジックの方に問題が多いが，現在のシステムは翻訳が各段階に分かれており，それぞれのチェックが可能になっているため，どこで問題があるか発見が早く，システムのメンテナンスが容易となった。またＩＢＭマシンでは，クロスリファレンスリストが出力されるため，シーメンスコンピュータよりメンテナンスがしやすい。

(5) システムのメンテナンス

シストランシステムのメーカは世界で４つの組織がある。それらは米国のWTC（World Translation Company）カナダのWTCC（World Translation Company Canada），西独のSystran Institute，シストラン・ジャパンである。通常ベーシック部分のコーディングは，これらの４つの組識と契約して行い，ＥＣはそれをチェックする。現在は，Systran Institute と連絡をとりながら，システムの改良を行っている。こうした外部からの協力によりＥＣ側として得るものは多く，特に新しい言語分野での開発には，このような方法は有効である。ＥＣにおけるマシントランスレーター関係の要員は25人である。

要員構成は言語学と言語解析者が23人，ソフトウエア要員が２人であり，現在５言語ペアーについて改良を行っている。

(6) 辞　　書

辞書は単語辞書，イディオム辞書，文脈辞書の３つから構成される。辞書規模は 約13万エントリーで，英語について見ると，単語辞書８万エントリー，イディオム辞書２万エントリー，文脈辞書１万エントリーである。これらの辞書作成のために５～６人で６年半の年月を費やした。ＷＴＣからは最初6,000~7,000エントリーしか供給されず，あとはＥＣで作成したものである。辞書作成の中で，ターゲット言語の意味作成が一番困難であり，やさしい単語で１日当り150ワード，複雑な単語で１日当り40ワードである。後者の単語の方がコンテキストの中では有効な働きをする。メンテナンスは現在１～２カ月に１回更新する。

(7) 入力と校正

入力と校正はオンラインで接続されたワング社のワードプロセッサーＯＩＳ12のディスプレイスクリーン上で行っている。

ＯＩＳ12にはディスプレイが最大９台まで接続できる。米国空軍では，入力用にＯＣＲを使っているが，ミスタイプ等を考慮して，ＥＣではワードプロセッサーによる入力を採用している。１日１人約30ページの割合で利用されている。校正もワードプロセッサーのスクリーン上で行っているが，校正時間は，人によってまちまちである。翻訳が10~20ページ約15分で行われ，校正は１時間に２ページくらいである。ユーザの必要性に合ったポストエディティングができれば，１時間５ページの割でできると期待される。

(8) システランシステム全体の動き

システランシステムの中で最も進んでいる言語ペアは，仏語→英語，英語→仏語，英語→イタリア語，ロシア語→英語（米語）である。英語→仏語に関しては生の翻訳文で90％をフランス人が理解できるという。

英語→スペイン語，英語→ポルトガル語はXeroxで使われており，特殊分野ではあるが十分に満足のいく成果が得られている。英語→独語は近く英語→仏語レベルにまでもっていく予定であり，仏語→独語，独語→仏語も来年末までにかなり高いレベルへ持っていく予定である。日本語→英語は'83年6月に仏語→英語レベルまで持っていくらしい。英語→アラビア語，英語→ペルシャ語はまだ動くレベルではない。一般的に言語の開発の中で，労力の80％はパーサーであり，ジェネレーションは容易であると考えている。

4.7.3　グルノーブル工科大学における研究

早くから言語理論を尊重した立場を維持して自動翻訳の研究を始め，現在も着実な歩みを続けている機関のひとつに，グルノーブル大学翻訳研究所が挙げられる。ここは1960年代初期に，このテーマを取り上げ，学究的態度を一貫して守りながら，欧州を中心に全世界に影響力を及ぼしてきた。

当初は一般的な中間言語を想定した翻訳システムを研究していたが，1973年から考え方を変え，トランスファー方式による文章解析システムをARIANE78 と呼ばれるプロジェクトに集大成した。中間言語研究時代にCETA（1961〜1971年）と呼ばれた研究は，トランスファー方式に変更したものを機にGETA（1971年〜）と改められた。

ARIANE78はここ3年間順調に研究が進み，フェーズ4から，フェーズ5に移っている。対象言語は仏⇄露を中核に英，独，ポルトガル，スペイン，マレーシア，日本，中国の各国語で，非常に幅広く取り上げて実験してきている。ユニークなのは中国語をローマ字式に入力し，日，独，露，仏各国語へのトランスファーを試みた事で，小規模とはいえ，注目される。こうした実験が多国間言語自動翻訳システムという概念を生み，世界における自動翻訳研究に大きな影響を与える事になる。現在は英語，仏語の精密な分析を進めると同時に，仏→英，仏→スペイン，仏→ポルトガル語の関係を研究している。

(1) ARIANEの思想

ARI ANEは現在では当り前だが，システムのソフトウエア部分と各国言語データ部分をはっきり分けた事に特色があり，一定のルールにのっとって文法と辞書をつくれば，解析，合成がどんな言語でも可能だという考え方をとっている。入力した文章の表層構造を最終段階までなるべく崩さず処理を進め木構造のノード（節）につぎつぎ情報をぶら下げていき深層構造を表現しようというものである。

これについて，「ノードにつく変数がふくれ上がり，複雑になる」，「新しい変数の付加が全体の見直しにつながる場合もある」という批判がある。

ARIANEは多国言語翻訳システムとして，最も進んだものと認められるが，現在は一段文法レベル段階のシステムで，実際の言語をあてはめ，きめ細かい操作をするには不便だとの見方もある。この点GETAは次のステップで工業化プロジェクトに発展させ（後述），多くの辞書の挿入をし，いろいろな試みを行ってみたいとしている。

(2) ARIANE 78 のあらまし

「機械翻訳システムに関する海外調査報告書」（日本電子工業振興協会）に詳しく紹介されているので，ここでは機能構成（図4—20）作業上のファイル構成（図4—21）ソフトウエアツール（図4—22）を掲げるのにとどめる。

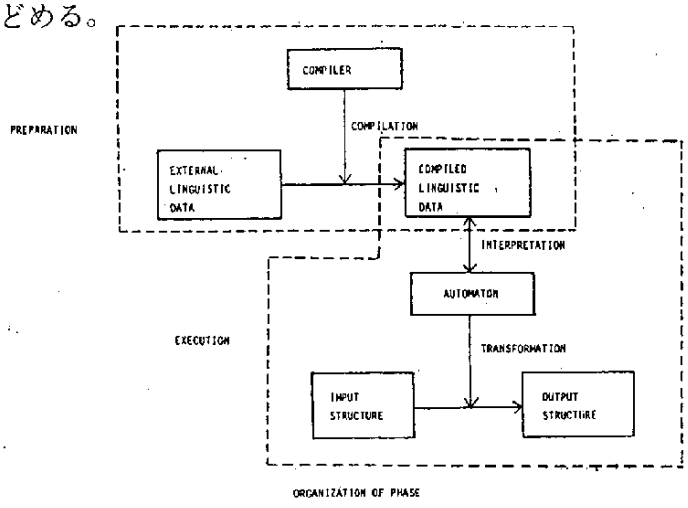


図4—20 ARIANE'78の機能構成

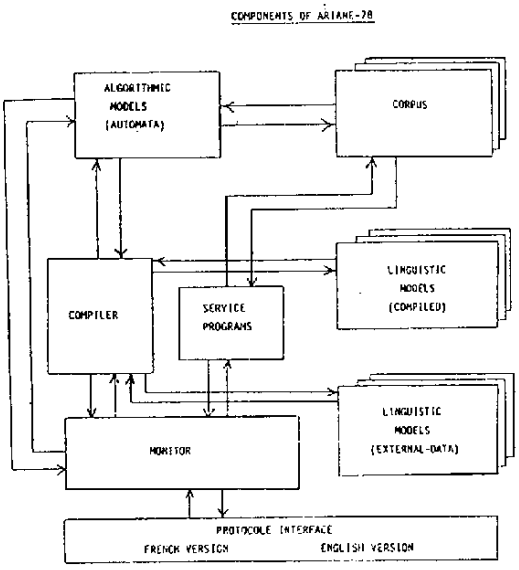


図４－21　作業上のファイル構成

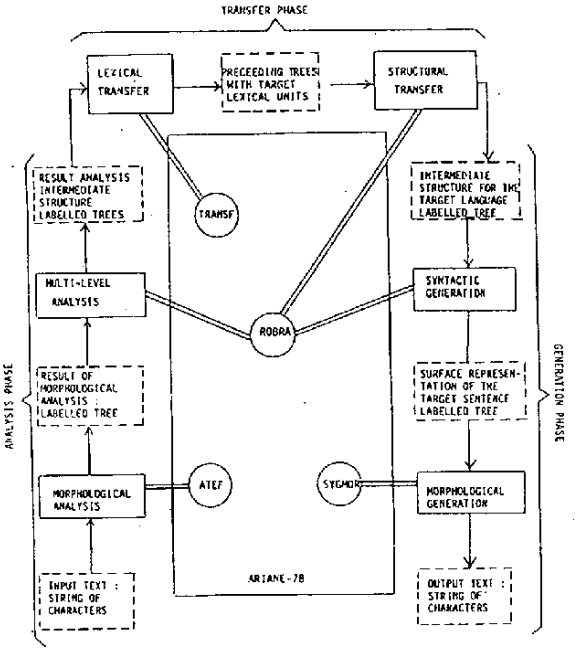


図４－22　ソフトウエア・ツール

システムは順次改良されてきており，より多くの辞書を挿入できる状況にある。これまでは，大量の用語を必要としない特定の分野を対象にテキストを入れ実験してきた。その結果，文章構造が似たものなら，書かれている内容の分野が異なっていても，かなり満足できる分析が出来た。たとえばエレクトロニクスの文献と高校の化学の教科書は文の構造が類似的なのでうまくいった。基本語2,000～3,000語を多く含むテキストの処理ならば大むね成果を得られる。

一方，文体の著しく異なるもの，特にスラングが登場するような文章の解析はまだまだである。

テキストのモデルは相当大量の文章で試してみないと駄目で，2～3ページのものでは不十分，本なら一冊全て翻訳しないと良否の判定ができない。やはり地道にテキストを選びマシンを動かし，駄目なものは捨てていくという作業を反復させながら構文の似かよったものについて，不透明部分をクリアーにしていく研究を推進し，ソフトウエアを発展させていきたいとしている。

テキストを入れた結果がよくなかった場合，①言語理論上間違っている，②文の意味があいまいであるという二つのケースが考えられる。現在，この修正の研究に力を入れており，今のところ人間がミスを見つけているが，これをマシンにやらせ自動的に辞書へ反映させていく研究を行っている。

以上のことから，ARIANE78は① 構造の似た実用的な文章の処理はうまくいく，②辞書は分野別につくっておき選択使用するという方向で実用化をめざした研究が進められている。

(3) ARIANE X

ARIANE78の実績を踏まえた工業化を目指すプロジェクトであり，多分「ARIANE 84」になる見通しである。これまで文法レベルの解析，トランスファーに傾注してきたエネルギーをいよいよ実用化に結集

させようというもので，基本的考え方は「78」を踏襲する。今後は用語辞書の構築，研究に重点が移行することになると思われる。

辞書作りは1日30語～40語程度処理できるとみられており，よく使われている言葉9,000語より，特定分野の用語6万語のほうが作業上簡単だという。現在は仏⇄露が最も進んでおり，8,000語用意されている。ただし商業ベースに乗せるには最低限5万～6万語ぐらいは必要とされている。

ARIANE Xの機種レベルはマイクロコンピュータ程度で，どのメーカのものにも適応できるものにする。A語→B語のペア開発に要するマンパワーは約1,200人/月を見込んでいる。A→Bが出来ればA→C，A→Dと進むにつれ労力は逓減していくとの事である。

(4) 各国の研究機関への影響

グルノーブル大学の自動翻訳研究の主柱であるVauquois教授は世界中の研究機関に関係があり，人脈を形成している。

米国には15の研究室があり，ユタ州のブリンガムヤング大学，テキサス大学オースチン分校の各研究室と密接な関係をもつ。カナダのモントリオール大学が研究室をつくった時，要員を派遣している。またブラジルのサンパウロ大学，マレーシアのペナン大学，バンコクのタイ大学とも人的交流があり，特に後2者にはARIANE 78を提供しており，研究上深い関係をもっている。

4.7.4 ハーバード大学における研究

コンピュータによる自然言語の自動構文解析の方法はいくつかあり，その各々は長所や短所を持っている。それらの構文解析方法の一つとして，予測的解析方法がある。この構文解析方法は1961年よりHarvard大学のComputation LaboratoryでKuno Susumu とÖtlinger Anthony G.によって研究されてきたものである。

(1) 予測的解析方法の概要

予測的解析方法は，与えられた文に対して，その標準形文法Gs に合致するすべての可能な構文の解析をするものである。標準形文法の規則は，　　$Z \rightarrow CY_1 \cdots Y_m$ （$m \geqq 1$ θ）

なる形をしていて，ここでZ，Ymは中間記号で，Cは終着記号である。入力連鎖$C_1 \cdots C_n$ の解析は，初期記号Xをもつ Pushdown Store（PDS）ではじめられる。連鎖の解析中のC_K においてZ_K をPDSの最高部にある中間記号とする。もし文法の中で規則，

$$Z_k \rightarrow C_k \ W_1 \cdots W_m \quad (m \geqq \theta 1)$$

がみつけられたら，Z_k はC_k によって満たされた（fulfilled）と呼ぶ事にする。そうするとPDSの中のZ_k はW_m を一番下とする新しい中間記号$W_1 \cdots W_m$（空にもなりうる）の列におきかえる。入力連鎖は，最後の終着記号の処理によりPDSが空になればwell－formed である。一方この中間記号Z_k は終着記号C_k をもつ規則が文法の中にみつからなければ，解析の道はすぐ前のC_{k-1}の枝分れする点にもどる。つまりPDSの内容は再構成され，次のとりうる道がたどれる。与えられた入力連鎖に対して，すべての解析の道をすべてし尽すまで処理を連続して行う。

例：

標準文法G_S を，

① $X \rightarrow a$

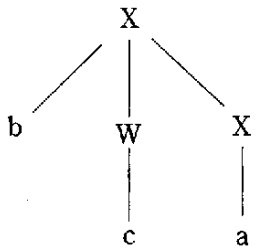
② $X \rightarrow bX$

③ $X \rightarrow bWX$

④ $W \rightarrow c$

⑤ $W \rightarrow cW$

とし，入力連鎖を“ｂｃａ”とすると，



が構文解析の結果となる。

予測的解析方法は，凝った解析の必要のない時，ある制限された目的のためには十分である。大量の資料を極めて能率よく処理することができ，またより洗練された解析を与えてくれるようなシステムは現実にないという理由で，予測的解析法は実用的な道具である。予測的解析法が役にたつ現実の領域は，文の完全な解析というよりむしろ，ある種の曖昧さ（ambiguity）の探索にある。

(2) 予測的解析法によるモデル

予測的解析方法で作成したモデルは，

辞書（変化形も含む）……30,000 語

文法規則……700 種

を用いている。

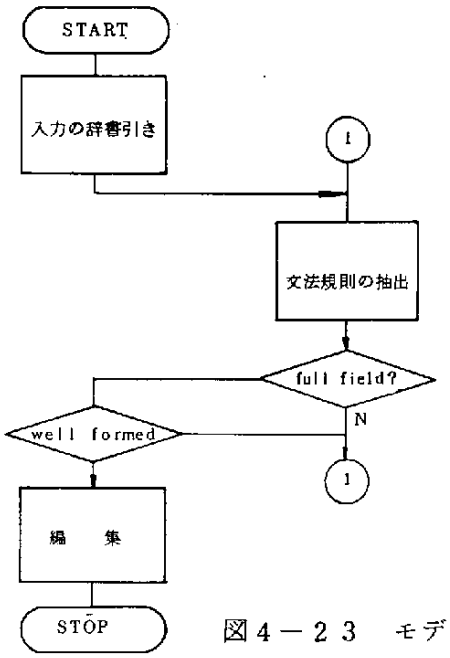


図4－23　モデルの処理フロー

モデルの詳細については，「構文分析法その１，その２」（情報管理 Vol 11 No 9.No 12 ）を参照されたい。

ここでは，モデルの3 version における処理時間の比較を記載しておく。

表４－12　モデルの処理時間の比較

Sentence Number	Number of Words	Number of Analyses	1963-FJCC (min.)	NEW SHARE (min.)	w/s Syntactic Analyzer (min.)
1	17	1	0.0	0.0	0.0
2	18	1	0.1	0.1	0.1
3	25	5	1.1	0.8	0.1
4	35	12	9.0	2.2	0.1
5	16	40	0.1	0.1	0.2
6	17	4	0.1	0.1	0.0
7	14	4	0.0	0.0	0.1
8	16	3	0.1	0.0	0.0
9	23	7	0.2	0.0	0.1
10	23	31	1.5	0.4	0.1
11	30	118	7.6	3.2	0.3
12	25	72	0.5	0.5	0.2
13	32	16	9.8	3.9	0.1
14	38	94	42.4	13.2	0.2
15	25	136	7.2	0.7	0.3
16	27	1	1.7	0.3	0.1
17	30	17	1.2	0.3	0.2
18	20	71	2.7	1.8	0.2
19	29	2	1.5	0.0	0.1
20	20	16	0.7	0.1	0.1
			Total 87.5	27.7	2.6

注１：1963—FJCCとNEW SHAREの相異

NEW SHAREではIBM/SHREシステムがNEW/SHAREとなり約３倍の効率となったことと一般化SHAPER TESTの機能が追加された。

注２：wfs. Syntactic Analyzer

NEW SHAREにrepeatitive path eliminator（くり返しpath の消去）の機能を追加したもの。

(3)　今後の検討課題

予測的解析法とそのモデルについて述べてきたが，モデルの改善あるいは翻訳システムへ応用するときの課題は以下のとおりである。

①処理を早くするために，規則の確率（文法のhit数），品詞の確率の導入を図る。

②大量の文章（例：1冊の本）を訳すときなどは，文章の性格を規則の1つとして入れておく。

③Semanticカテゴリーを導入するときは，たくさん導入しないで，10ぐらいにとどめる。

④数式が扱えるようにする。

4.8　機械翻訳における構文解析法

機械翻訳システムは,コンピュータによる自然言語処理技術のいわば集大成といえるものであり，自然言語処理と一体不可分の関係にある。従って，機械翻訳システムで行っている様々な手法は，自然言語処理において開発されてきた手法を踏襲したものであるし，機械翻訳を行う過程で生じる問題点も，その多くは，そのまま自然言語処理における問題点となりうるものである。

本節では，機械翻訳システムが，実際にどのようにして翻訳を行っているかを概観するために，構文解析という言葉をかなり広い意味で捕えて，辞書（レキシコン）部分を除いたあらゆる部分での手法を概説する。まず，機械翻訳における様々なシステムを分類し，その概要を述べた後，狭い意味での構文解析が，各々のシステムでどのように行われているかを見る。辞書の部分については，4.9節で改めて述べる。

4.8.1　機械翻訳方式の分類

(1)　世代分けによる分類

コンピュータを，その構成デバイス（真空管，トランジスタ，IC，LSI）によって第1，第2世代などと分けるように，機械翻訳システムに対しても，その目標のレベル，手法の違いなどによって世代分けが行われている。年代的には，機械翻訳に対して否定的な見解を出して，その後しばらくの間機械翻訳研究の沈滞をもたらす原因となったALPACレポート（1965年）以前に作られたシステムがほぼ第1世代に属し，その後70年代後半から復活してきた研究の線に沿ったものが第2世代ということができる。

表4－13に，このような世代分けに従って現在稼動中または企画中のシステムを分類し，その特徴を簡単に述べたものを示す。

表から明らかなように，今日稼動中のシステムの中で，第1世代に属するものとしてあげられているシステムは，いずれも商業ベースに乗って販売・供給されているものである。特徴のところに示したように，第1世代

表４－１３　機械翻訳システムの世代分け

世代	特徴	現在稼動または企画中のシステム
1	・言語学的記述の部分とプログラムが分離せず，どちらかと言えば経験的，ad hoc なシステム ・２カ国語間に限定	SYSTRAN LOGO ALPS Weidner　など
2	・言語学的記述とプログラムとの分離 ・文法・辞書をきちんと作る ・多言語間も可	ARIANE78(仏グルノーブル大) TAUM-METEO TAUM-AVIATION （以上 仏モントリオール大） EUROTRA（EC）　など
3	・意味記述を用いる ・文脈処理をとり入れる	（将来の目標）

のシステムは，内容に理論的裏づけが乏しく，どちらかといえば経験主義的なシステムである。しかもどのような分野の翻訳要求にも応じられる訳ではなく，辞書のエントリ部分などはユーザ側で（かなり長期間にわたって）開発し，保守しなければならない。また，核になる部分は（ハードウェア，ソフトウェアともに）通常企業秘密となっているか，契約上変更が許されないようになっている場合が多いために，システムの大幅な変更は不可能である。

以上のような欠点はあるが，第１世代のシステムには商用としての長い歴史があるために，文書翻訳に関しては，既にかなりの実績があり，データの蓄積も行われている。その評価によれば，自然言語をある程度人手を用いて前処理し，いわば自然言語のサブセットを作った結果を入力し，翻訳された結果を再び人間が手直しすれば十分に使用に耐えうるという報告がある。また一方では，前処理・後処理を含めた一語あたりの翻訳コストは，人間の翻訳者が全て行った場合のコストより高いという報告もある。いずれにしても商用のシステムは今後さらに評価・検討が必要である。また，これまで実績を積んできたのは，英⇆独，英⇆仏，英⇆露などの印欧語族内の間での翻訳のみであり，日本語，アラビア語などと英，仏語のように，全く言語体系の異なる語に対しては，商用システムの発売を予告（SYSTRAN）するか，計画中（Weidner）の段階にとどまっており，

真の評価は今後の問題である。現在盛んに研究が進められている第２世代以降のシステムとの競合が注目される。

第２世代システムの特徴を要約すれば，表の通りであるが，第１世代との大きな違いは，1965年以降発展してきた言語学上の成果（たとえば生成文法など）をとりいれて（言語学者がシステムの構築に直接参加している場合も多い），理論的裏づけをきちんとしたシステムが多いことである。

第２世代のシステムは，現在研究中のものが多く（実用に供されているものはTAUMMETEOのみである）その手法もさまざまである（手法による分類については，次節で述べる）が，共通している点は，複雑な翻訳過程を一気に行うのではなく，元の言語と目標言語との間にいくつかのレベルを設定しているということである。これを中間的表現と呼んでいる。中間的表現にも手法によってさまざまな表現法・表現形態があることはもちろんであるが，その代表的なものを簡単に図示すると図４−２４のようになる。

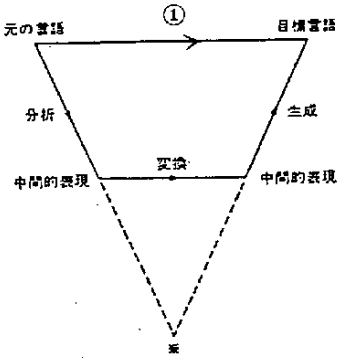


図４－２４　第２世代のシステム

この図は，第２世代のいわば基本理念を説明する時によく用いられる図である。まず元の言語を分析することによって，元の言語に対応した中間的表現を作る（分析過程）。次にその中間的表現を目標言語の中間的表現に変換する（変換過程）。こうして得られた表現から，目標言語を作

り出す（生成過程）。最終的には，元の言語の中間的表現と目標言語の中間的表現とが同じものになるような分析が得られれば，変換過程が不要になり，しかもすべての言語に共通な中間的表現（図の※部分で，inter - ligua とか，pivot language と呼ばれている）を得ることができる。この図の上で第1世代のシステムを表現すると，①と書かれたパスを通るのが，第1世代のシステムであるということができる。

(2) 手法による分類

機械翻訳システムの世代分けによる分類を前節で述べ，さらに基本的な手法についても簡単に触れた。本節では，第2世代システムを中心として，種々の機械翻訳システムに見られる手法について概観する。各々のもっと詳しい記述は次節で行う。

表4−14に主な手法を分類して掲げる。この分類は，かなり恣意的なも

表4－14　機械翻訳システムの手法

手　　法	要　　約
構文パターンマッチング	文法上の係り受けのパターンを定めておいてマッチしたものを変換して翻訳する。
トランスファ方式	元の言語から目標言語への変換部分を木構造の変換で行う。
融 合 方 式	変換・生成部分を一気に行う。
モンタギュー文法	元の言語を分析した結果を論理式の形で表し，それを目標言語の論理式に変換した後，生成する。
概念依存を用いた方法	どのような言語にしても共通の概念構造を作りそれを介して翻訳を行う。

ので，必ずしも明確に分れるものではない。たとえばパターンマッチングは，各システムのいずれかのフェーズに必ず含まれている方法であるし，融合方式，モンテギュー文法を用いた方式は，いずれもトランスファ方式と見ることもできる。ただ，中間的表現に落す場合にモンテギュー文法を用

いて論理式にしたり，トランスファという過程を陽に用いないで変換・生成を行う（融合方式）という違いがあるに過ぎない。

表には，ごく簡単な要約をつけたが，次節においては，もう少し実際の翻訳に即して，各々の方式を代表するシステムについて見ていくことにする。

4.8.2 種々の構文解析技法

(1) 構文パターンマッチングによる方式

自然言語処理においては，あらかじめシステム側で用意した語彙項目や文法規則と入力された文とを，あらゆる段階でマッチングしなければならない。従って，自然言語処理全般にわたってパターンマッチングが行われていると言っても過言ではない。

ここで述べる構文パターンマッチングとは，あらかじめいくつかの文型パターンを用意しておいて，それと合うように入力文を変形するやり方で，表題文の翻訳システムに代表される。[1]大量の自然言語を処理するためにもっと原始的な逐語処理を用いた例もある[2]が，ここでは上記のものを中心にとりあげる。なお，このシステムを工業技術院計算センター（RIPS）で稼動した結果については，既に報告されている[3]し，56年度報告書にも一部紹介されているので，ここでは簡単な紹介にとどめる。

この方式の特徴は，ほとんどが名詞，名詞連続と，それを修飾する形容詞，動名詞などから構成されるという論文表題の性質に着目して，それに即した処理をしていることである。処理は5段階（最後の生成部を入れて6段階）からなっており，まず第1段階で辞書引きと熟語の処理を行う。次に，第2段階でローカルな and（名詞句の中に埋めこまれている and）の処理を行う。第3段階では遷移網（transition network）によって，名詞連続を1つの名詞に縮退させる。第4段階では，第3段階までで現われてきた骨格パターンを文型パターンとマッチングさせて，第3段階で単

語の意味記述と，特定の単語に固有な文型パターンを用いて訳出語順を決定する。このようなやり方で翻訳した例を図4−25に示す。

入力：Industrial and Scientific Techniques for Measuringfield Effect Mobility

Step 1：Industrial and Scientific Techniques for Measuring　Effect Mobility

Step 2：Scientific Techniques for Measuring　Effect Mobility

Step 3：Techniques for Measuring　Mobility

Step 4：<文型パターン>　[n　prep　ing　n]

訳出結果：電界効果移動度測定のための工業的及び科学的手法

図4−25　論文表題英和翻訳システムの翻訳ステップ[1]

このやり方を，普通の文の翻訳に用いる場合に，種々の問題点があることは明らかであろう。一般に，構文解析を行う場合にパターンマッチングだけで行って，文法規則で生成的な処理を行わないやり方では，パターンにない文型や，未定義語などが入力された時に対処が難しいし，多くの言語データを扱おうとすれば，それだけパターンを増やす必要があるという欠点がある。従って，論文表題のように，ある程度目標のはっきりしたシステムに使用すべき手法であると言える。

(2)　トランスファ方式

既に述べたように，第2世代のシステムでの基本的な理念は，元の言語の中間的表現と目標言語の中間的表現を介しての分析・変換・生成と言うステップであった（図4−24参照）。このようなアプローチをとる限り，変換部分というのはどのようなシステムでも不可欠であり，従ってトランスファ（変換）というのはかなり一般的な概念であるが，ここでは，トランスファという考え方を最も明確にシステムにとり入れているARIANE−78[4]（4.7節参照）の内容を紹介する。このシステムは，既にいくつかの紹介もなされている[5][6]ので，ここではそれらに基づいて説明する。

ARIANE−78の流れを図4−26に示す。

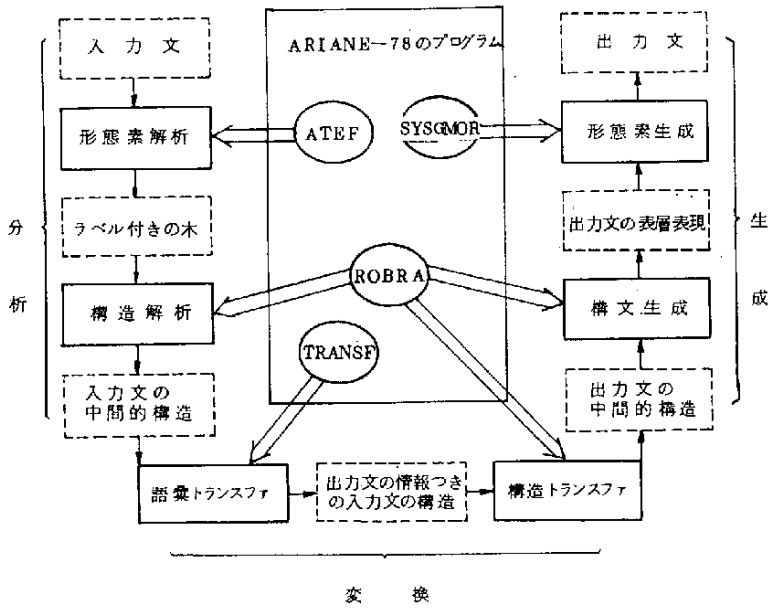


図４－26　ARIANE－78の処理の流れ

まん中の大きな四角で囲まれた部分がARIANE－78 のプログラムの名前を示す。この図から分るように，ARIANE－78 では，それぞれの解析ステップが明確に区別されているのが特徴である。図で，実線で囲んだのがプログラム処理部分，破線で囲んだのがその前後の入出力を表わしている。

ATEF，TRANSF，SYGMORの部分は，それぞれ形態素解析用辞書，2言語間の単語対照辞書，形態素生成辞書の形をとっているが，これらの部分については，4.9節でもう少し詳しく扱うことにする。

このシステムの各過程では，常に1本の木構造が処理の対象になる。通常，文を構文解析する場合には，文脈自由文法（context free grammar）を用いると，解釈の異なりによって多くの木構造が生成される。多くの解析システムでは，この木構造の生成時に，種々の制限条件をつけたり，生成される木構造を評価して順位をつけたりするが，このシステムでは，入

力文を形態素解析した結果が，まず入力単語とその辞書情報のついた平らな木構造に変換され，以後この木構造の各節点に種々の情報を付加しながら木構造を変換して出力文にまで変換する。他のシステムで，いくつかの木構造となって現われる解釈の曖昧性も，ここでは節点についた別々の解釈の枝によって表わされる。

木構造には，通常のシステムで付されている品詞，格パターンなどと共に，語や句がその木構造の中でどのような役割を果すかという，機能的な性質も，属性－属性値対の集合として表現されるようになっている。即ち異質なレベルの情報を一つの木構造に集約して表現している。この記述レベルは，次の4つである。

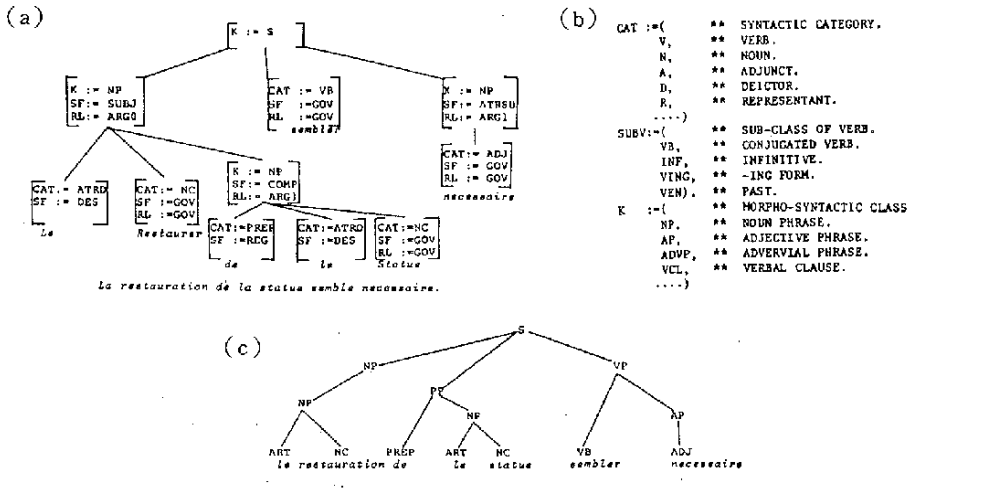
① 句のまとまり方の種類（Syntactic Category － K）

② 句の構文的役割（Syntactic Function － SF）

③ 句の意味的関係（Sematic Relation － RS）

④ 句の論理的関係（Logical Relation － RL）

図4－27[a]に「この像は修理が必要と思われる」という意味の仏文に対



(a) 木構造[6]　　(b) 属性と属性値[5]

(c) 形態構文レベルの木構造を(a)から抜き出したもの[6]

図4－27　構文解析の例

する構文解析木を示す。各節点に付された情報は，上から順にほぼ上記①～④のレベルに対応している。即ち一番上が形態構文レベル，一番下が深層構文レベルと見ることができる。図4－27(b)には各節点の一番上に付された要素の値の例を示す。この部分のカテゴリのみを抜き出して木構造を作ると，ちょうど通常の構文解析を行ったのと同じような木が生成されていることが分かる（図4－27(c)）。

解析，生成，トランスファといった重要な部分を受けもっているのが，ROBRAというプログラムである（図4－26参照）。ROBRAには，木構造の節点に付与される属性 ― 属性値対の宣言やそれらのチェックなどの機能などもあるが，主な機能は木構造の変換規則の定義である。木構造の変換規則は，適用条件と適用後の構造記述の部分に分かれている。条件部では根の条件，木の幾何学的形状，各節点の属性値の条件，節点と節点の間の属性値の一致条件を調べる。これらの条件に合致すると，構造記述の部分に書かれている木構造に，もとの木構造から変換し，さらに節点の生成や消滅，節点の属性値の変更などを行う。図4－28には，英語の入力文から日本語の出力文を作った例を示す。

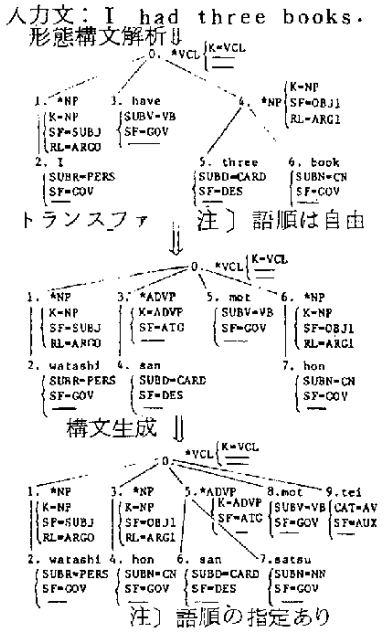


図4－28　英日翻訳例と木構造の変化の様子[5]

このようにトランスファ方式では，木構造を変換することによって，元の言語から目標言語を生成していく。グルノーブルのシステムのように，かなり深層のレベルまでを取りこんだ木構造を変換していくものもあればテキサス大学のシステムのように，比較的浅いレベルの木を作って，格フレームなどを用いて変換していくものもある[7]が，いずれもトランスファという概念が基本になっている。

これらのシステムは，意味素性を階層的につけていくことがやりにくいとか，木の変換規則の数が多くなったり複雑になったりしすぎるといった欠点はあるものの，機械翻訳の一つの方向を示すものとして注目される。

(3) 融 合 方 式

機械翻訳で問題になることの一つに，訳し分けの問題がある。たとえば，

I have a book.

I have a game.

I have a cold.

I have breakfast.

という文は，構文的には全く（あるいはほぼ）同じであるが，訳す際にはそれぞれ「持つ」，「する」，「ひく」，「食べる」としなければならない。このような問題は単純な構文解析や，表層レベルでの木の変換だけでは対処できない。従って，これを解決するためには，どうしても「意味」の問題に踏みこまざるを得なくなる。

木構造の節点に種々の条件を付与するのは，これに対する一つの解決法である（前節参照）が，最近「融合方式」と呼ばれる新しい手法が提案された[8]。以下では，この手法について述べる。

融合方式の手順は，次の通りである。

① 元の言語を構文解析して木構造を作る。

② 構文解析木上の文法規則に付加された意味規則を，木の下部（枝）から上へ送る。

③　構文解析木上の各節点で，下位から送られてきた情報に対してプログラムを起動する（これをユニット間会話という）。

④　このとき，意味的な条件を満たすものがあれば，そこを埋めていく（条件を付けて待っているものをslot，それを満たすものをfillerという）。英日翻訳の場合，助詞などもこの時点で自動的に付加される。

⑤　節点上で語順の変更が必要な場合には，ここで語順の変更を行う。

⑥　最上位節点（根）に至るまで②から⑤の手順を繰返し，最終的に意味解析がなされる。この時に翻訳文が合成される。

図4−29に，動詞と名詞句から動詞句を作るという文法規則の例，図4−30に，このような文法規則から作った構文解析木上に，意味規則を適用して得られた意味構造の例を示す。この図から分るように，得られた意

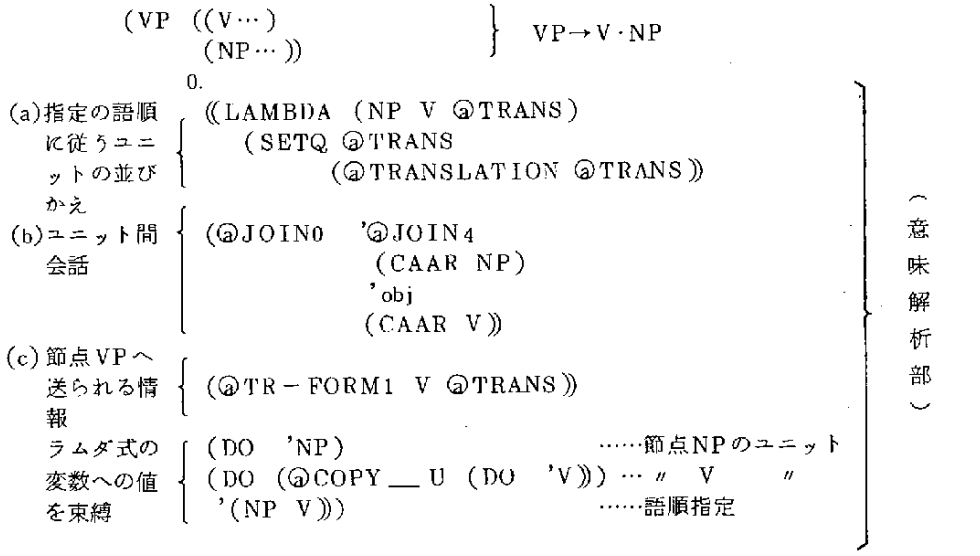


図4−29　文法規則の例[8]

味構造を順にたどっていくと，ほぼ完全な翻訳文が得られている。

この方式の特徴は，次の点である。

①　木の変換を陽に行わない。

（入力文：He opens the door with a key）

図4－30　意味解析結果（意味構造）[8]

② ある単語に対して，訳し分けが必要な場合でも，辞書項目は1つしか立てず，その意味のslotをプログラムで選択することによって翻訳を行う（これについては4.9節でもう少し詳しく述べる）。

③ slotとfillerの関係には，上位 — 下位概念なども許されているので，意味素性の階層的な記述が可能である（たとえば前節のARIANE－78ではこのようなことは困難である）。

④ 文脈自由文法（条件つき）を用いているので複数の構文解析木が生成されるが，意味解析によって多くの木がリジェクトされる。

⑤ 意味解析の終了とともに翻訳文が合成される。即ち，解析・変換・生

成というステップが，融合した形で行われる。

以上のようにこのシステムは，意味の段階に一歩踏みこんだものとして注目される。しかしながら翻訳の過程で，もっと陽に木の変換を用いた方がよい場合もあるので，今後の検討が必要になろう。

(4) モンテギュー文法

意味解析を行うシステムでは，意味構造をどのようにして表現するかが問題となる。前節で述べた融合方式では，slotとfillerという関係を用いて，構文解析木上でこの関係が満たされるかどうかを定める方式をとっていたが，本節では文を論理式の形に変換することによって，そこに意味解釈を含ませながら翻訳を行うシステム[9]について述べる。

モンテギュー文法は，自然言語に対する意味的・構文的なモデルを与える試みとして提案されたものである。モンテギュー文法では，入力された文を，辞書項目に関する「翻訳」規則（ここでの翻訳とは，言語間の翻訳という意味ではなく，論理式へ変換するという意味である）と，構文規則に関する「翻訳」規則を用いて論理式に変換する。このうちで辞書項目に関する規則を論理式への変換ではなく，目標言語への変換としたものが，ここで述べるシステムである。

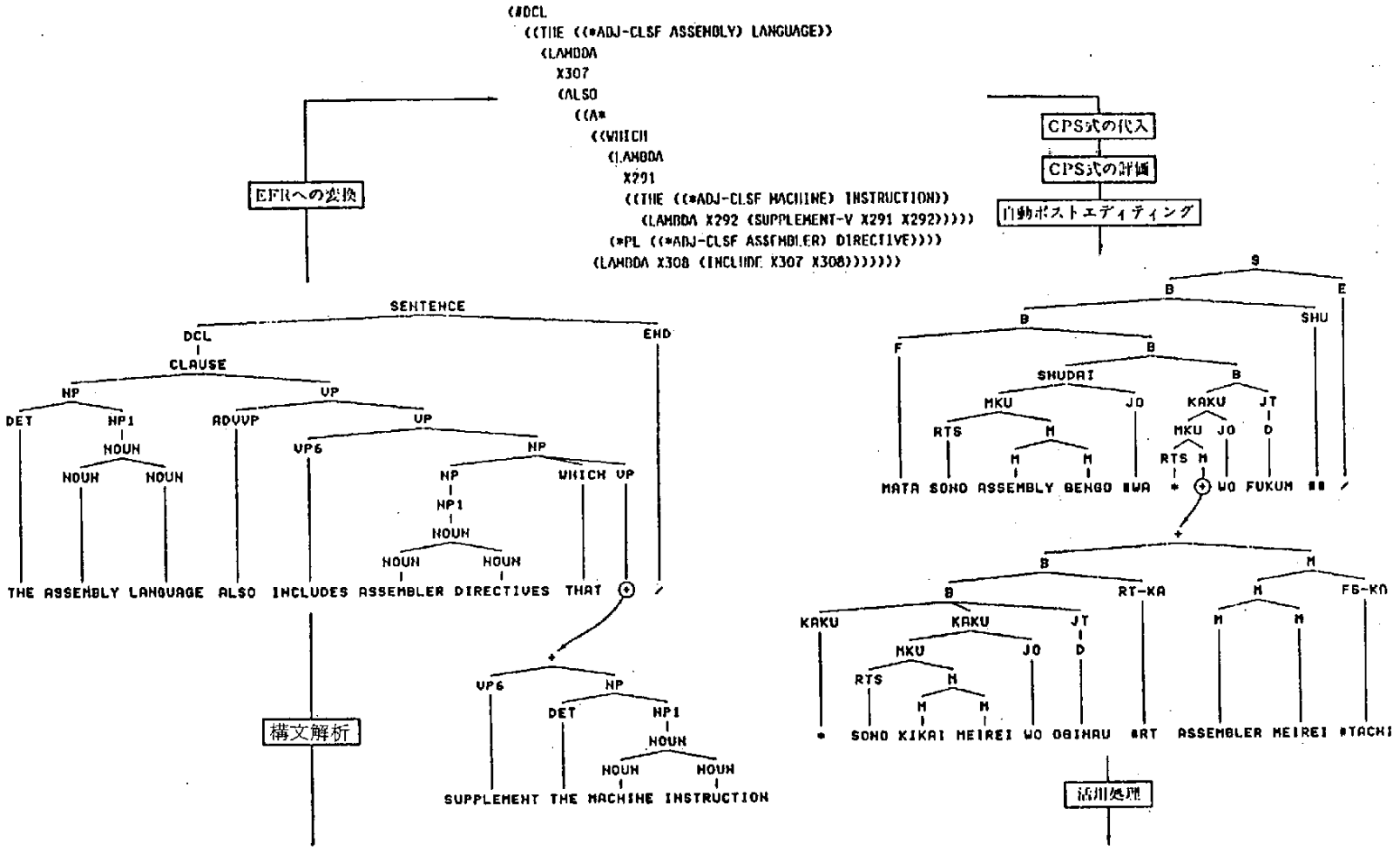
システムは，次のような手順で動かされる（図4−31参照）。

① 入力文を構文解析して木構造を作る。

② 木構造を構文規則に関する「翻訳」規則を用いて論理式に変換する。このシステムでは，変換された表現をEFR（English oriented Formal Representation）と呼んでいる。

③ EFRの英単語に，日本語への翻訳を論理式で記述したものを代入する。これをこのシステムでは，CPS式（Conceptual Phrase Structure formula）と呼んでいる。

④ CPS式の上で論理演算と，木構造の演算を行って日本語の構文解析木に直す。



□□ 入力文 □□

THE ASSEMBLY LANGUAGE ALSO INCLUDES ASSEMBLER DIRECTIVES THAT SUPPLEMENT
THE MACHINE INSTRUCTION ／

■■ 出力文 ■■

またそのアセンブリ言語はその機械命令を補うアセンブラ命令を含む

図４－３１　モンテギュー文法を用いた翻訳システム[9]

⑤ 得られた訳文を自動修正する。

⑥ 活用処理を行って訳文を生成する。

また，訳語の選択は，名詞句などに意味的なクラスに関する情報を持たせることによって行っている。このシステムで特筆すべき点は，未定義語が出てくると，それを対話的に処理できるようになっていることである。これは，機械翻訳における辞書の役割とも密接にかかわってくることであるので，4.10節で改めて取り上げることにする。しかしながら，構文解析（制限つきの文脈自由文法で行っている）の部分で多くの木が出力されることなどの問題点もある。木の選択というのは既に何度か述べてきたように，自然言語を構文解析した時に必ず生じてくる問題点であるが，ここでは，人とコンピュータの対話によって，そのような問題もある程度まで解決しようとしている。

(5) 概念依存を用いた方法

これまで述べてきたものとかなり異なったアプローチの仕方をするものに，概念依存理論（Conceptual Dependency Theory）を用いた翻訳の方法がある。この理論は，本来，物語や質問応答に対する理解を行うためのモデルとして提案された。[10] しかしながら，そこから抽出された概念表現（Conceptual Representation）は，言語によらないユニバーサルなもの（即ち前記図4−24の※にあたる）であり，従って概念表現を媒介とすれば，いかなる多言語間翻訳も可能となる。

この考え方では，まず，言語理解を，一連の単語列を形式の整った概念構造に写像する過程であるとみなす。そしてそれによって言語の奥底にある概念の階層構造を生成する。概念の範疇としては，次のようなものを認める。

PP：概念的な名詞類

PA：PPを記述する状態

ACT：PPが他のPPになす行動。次のようなものから成り，現在11個提案されている。

ATRANS：所有のような抽象的概念の移行（たとえばgive（与える）は，何かを誰かにATRANSすることである）。

PTRANS：対象の物理的場所の移動（go（行く）は自分をある場所にPTRANSすることである）。

PROPEL：対象に物理的力を加えること（push（押す）など）。

LOC：空間の座標上の場所

T：時間

AA：ACTのある場面の修飾。

VAL：状態に対する値。

そしてさらにこのようなACTを組み合わせた，概念表現よりもさらに高次な意味表現形式としてMOPs（Memory Organization Packets）が提案されている。[11] MOPは，ある目標を達成するためのSCENE（いくつかの行動をまとめて表現した記憶形態）から成るものであるが，ここでは詳細は略す。

さて，上述のような概念が，ある文からどのようにして得られるかを以下に解説する。[12]

プログラムは，制御構造とデータ構造を持っており，短期記憶（Short－Term Memory，略称STM）の入る概念リスト（CONCEPT－LIST，略称C－LIST）と，要求を起動するデータ構造を入れる要求リスト（REQUEST－LIST，略称R－LIST）という2つのリストを持っている。そして次のような手順で制御が行われる。

① 入力文を左から右に読んで，次の辞書項目（単語，熟語）を見つける。なければ終了。

② 新しい項目に関する要求をR－LISTの中へロードする。

③ R－LIST中の活性化した要求を考察する。

④ ①へ戻る。

この時要求というのは，上の②と③で次のようにしてなされる。

イ．概念構造をC－LISTに加える。

入力文 "Fred gave Sally a book"

↓ ①

"Fred"

↓ ②

REQUEST −
TEST : 真
ACTIONS : C−LISTに構造(PP CLASS (HUMAN) NAME(FRED))
を加える。

↓ ③

(PP CLASS (HUMAN) NAME(FRED))

↓ ①

"gave"

↓ ②

REQUEST −
TEST : 真
ACTIONS : C−LISTに次の構造を加える。
(ATRANS ACTOR (NIL) OBJECT(NIL)
TO(NIL) FROM(NIL)
TIME(PAST))

REQUEST −
TEST : C−LIST中にATRANSに先だって人間についての記述があるか？
ACTION : それをACTORとFROMに入れよ。

REQUEST −
TEST : C−LIST中にATRANSのあとに人間についての記述があるか？
ACTION : それをTOに入れよ。

↓ ③

ACTORとFROMのスロットを埋める。TO，OBJECTはまだ埋まらないのでR−LISTに残す。

↓

概　念：
(ATRANS ACTOR (PP CLASS (HUMAN) NAME (FRED))
OBJECT (PP CLASS (PHYSICAL-OBJECT)
TYPE (BOOK) REF (INDEF))
TO (PP CLASS (HUMAN) NAME (SALLY))
FROM (PP CLASS (HUMAN) NAME (FRED))
TIME (PAST))

図４－３２　概念の抽出例

ロ．概念構造内のスロットを他の概念構造で満たす。

ハ．他の要求をする。つまり，要求をＲ－ＬＩＳＴに加える。

ニ．今行った要求を終了する。

実際の文に適用した例を図4－32に示す。図中の①～④は上の番号と対応している。このようなことを各文，また各エピソードについて行って物語の構造を抽出する。そして，こうして抽出された概念構造を逆向きにたどって生成していけば，どのような目標言語に対しても「翻訳文」が生成できることになる。この時，因果関係などを推論したり，目標言語の語順などを定めたりする必要があるが，前者は階層化された概念関係をたどる

(a)

```
M-MO =
 CONCEPT M-MOCK-TRIAL
 ACTOR   HUM11 =
           CONCEPT TERRORIST
           ORG     OBJ3 =
                     CONCEPT TERRORIST-ORG
                     MEMBERS HUM11
           GENDER  MALE
           TYPE    GUERRILLA
           WEARING OBJ0 =
                     CONCEPT CLOTHING
                     TYPE    SUIT
                     COLOR   OLIVE-COLORED
 OBJECT  HUM6 =
           CONCEPT   PERSON
           NUMBER    AT-LEAST 60
           RESIDENCE COUNTRY
 SCENE2  ACC1 =
           CONCEPT ACCUSE
           OBJECT  HUM6
           BAD-ACT UND1 =
                     CONCEPT UNDESIRABLE-ASSISTANCE
                     JUDGER  HUM11
                     OBJECT  OBJ3 =
                               CONCEPT GOVERNMENT
                     ACTOR   HUM6
           ACTOR   HUM11
           IR-FROM UND1
 SCENE4  EXE0 =
           CONCEPT EXECUTE
           ACTOR   HUM11
           PLACE   LOC0 =
                     CONCEPT CITY
                     NAME    SAN PEDRO PERULAPAN
           OBJECT  HUM6
           IR-FROM UND1
 SCENE1  UND1
 SCENE3  TRY0 =
           CONCEPT TRY
           OBJECT  HUM6
           ACTOR   HUM11
```

(b)

```
At least 60 peasants were executed by a firing squid of men wearing
olive-colored uniforms in San Pedro Perulapan, about 25 kilometers
east of San Salvador, authorities there said.
According to the same sources, the victims were tried and then
executed in the town plaza by guerrillas who accused them of
collaborating with the government.

sukunekutomo 60 nin no noumintachi ga, seifu ni kyoryokushita
node, terososhiki ni sokusu oribuiro no fuku o kita geriratachi
wa, sono noumintachi o saiban-ni-kaketa.  soshite, sono
geriratachi wa, san-pedoro-perurapan toiu machi de sono
noumintachi o shokeishita.
```

図４－３３　(a)物語の概念表現
(b)物語の入力（の英訳）と日本語出力（下）[11]

ことによって，また，後者はスロットを取り出す順番を指定することによって行うことができる。図4−33(a)に物語の概念表現，(b)に入力文（スペイン語の原文を人間が英訳したもの）と対応する日本語文（(a)を用いて生成したもの）を示す。この例で分るように，入力文と出力文との間には，概念があるだけで，構文的な関係は何もない。従って，原文の構造というものが必ずしも保存されている訳ではないことに注意されたい。

(6) まとめ

本節では，構文解析そのものというよりは，むしろ，機械翻訳システムの全体像というものを，各システムについて紹介してきた。これらのシステムの記述を見れば分るように，機械翻訳では，まだ多くの難問が山積している。各システムの内容の解説のところで，そのシステム固有の問題点について簡単に触れたが，機械翻訳全般にまたがる問題については4.10節で辞書と関連した文脈の中で述べることにする。いずれにせよ，「完全な機械翻訳」（何をもって「完全」とするかは種々議論の分れる所である）を達成するのは，非常に困難な問題である。

4.9 機械翻訳における辞書の役割

4.8節において既に見てきたように，機械翻訳においては，構文解析のみならず，辞書の役割が非常に重要である。しかも，その辞書が，構文解析などの処理と表裏一体となって，機能的に有効なシステムを構成していることが必要である。ここでいう「辞書」とは，我々が日常用いている「〜辞典」という意味での辞書（dictionary）ではなく，機械処理ができるという意味での語彙項目（lexicon）や，分類語彙(thesaurus)にむしろ近いものである。

このような辞書類をどのようにして作成し，また，どのような形態に作り上げるかについては種々議論がある。実験室レベルでは，手作業で，いわば思いつき的に作っていくことも結構役に立つし，言語学者の協力のもとに，チームを作って行うようなこともできる。また，既存の辞書（dictionary）

から，機械処理に用いることのできる語彙項目を作成する方法もある。大量のデータを扱う文献検索，ジャーナリズムの分野では，独自にシソーラスを作っているところも多い。さらに，人手での翻訳においても，翻訳の質の向上や，翻訳者による用語の異同を防ぐために，シソーラスを作る動きが世界的に活発である。このようなシソーラスは，現在はまだ普通のデータベースとして使用されるに止まっていて，機械翻訳のシステムと直接結びついている訳ではないが，将来はこれらを結びつける方向も考えられる。

そこで，4.9.1においては，用例辞書のデータバンクの例や，市販の辞典をデータベース化した例などについて簡単に触れることにする。また，4.9.2では，4.8.2で紹介したそれぞれの機械翻訳システムで，辞書がどのように作られているかを簡単に概観する。辞書の構造についても簡単に述べるが，詳しい考察は4.10節で行う。

4.9.1　用例・辞書のデータベース

(1)　用語データバンク

機械翻訳ならぬ人間による翻訳で最も大きな問題になるのは専門用語の翻訳である。特に，エレクトロニクスなどのように進歩の著しい分野では，翻訳用語を定めるのも大変な作業であるが，日進月歩の用語に翻訳者がついていけなくなる恐れがある。また，専門用語辞典にしても，出版された時には既に掲載されていない用語ができるという状態である。

そこで，特に多言語間の翻訳問題が深刻化しているＥＣなどにおいて，専門用語のデータバンクを作る動きが活発になっている。表4—15にこのような用語データバンクの例を示す。各データバンクは，いずれも十数万から数十万という項目を備えており，オンラインでかなり知的な検索（たとえば語尾を自動的におきかえたりする）ができるようになっている。このようないわば「翻訳支援システム」とでも言うべき形態のものは，「完全な」機械翻訳に対して難問が山積している限り，機械翻訳に対する人間の対話的修正などを補う形で，今後ますます重要になっ

てくると思われる。

表4－15　用語データバンクの例.[19]

開発機関	システム名	内容
ブラッセル自由大学	DICAUTOM	EC石炭鉄鋼委員会で使われていたデータバンク
EC	EURODICAUTOM	DICAUTOMを発展させたデータバンクで，各言語間の用語の対応，定義，使用例を要領
西独連邦言語局	LEXIS	航空，通信，電子，軍事等の分野の300万語（一方向では150万語）における専門用語集
モントリオール大学	TERMIUM	EURODICAUTOMと類似したシステム
シーメンス	TEAM	用語や慣用句に対する訳語，使用例，定義の生成
フランス規格協会	NORMATERM	ISO(国際標準規格)にのっとった規格・標準のシソーラス(英，仏)

(2)　電　子　辞　書

翻訳のためには上記のような用語データバンクを作るのも一つの方法であるが，多大な人手と労力を要するし，分野をある程度限定しないと作業が収束しない恐れもある。そこでもう一つの方策として，既成の辞書（dictionary）をデータベース化して，それを自然言語処理や機械翻訳に応用することが考えられる。計算機上で対話的に校正などが行えれば，データの質の向上につながるばかりではなく，その成果を辞書の方へフィードバックすることによって，出版に与える利益も大きいと思われる。現に，最近では，コンピュータ編集を標榜した辞書も現われており[15]，今後，我が国でもこのような動きが盛んになるであろう。

我が国では，陽にコンピュータ編集をうたっているものはないが，既にいくつかの辞書はそのような形の編集を行っており，また研究用に供されているものもある[16]。このようなデータベースを作成しておくと，たとえば逆引きを作成したり（図4－34），品詞辞書を作ったりするのが容易になるばかりではなく，辞書自体を言語データと見なして言語学的な研究もできる[17]し，さらに進んで辞書の語に意味マーカなどをつけて，自然言語処理の補助手段とすることもできる[18]。これらのいわゆる

「電子辞書」は，データ構造を工夫したり，プログラムに種々の機能を付け加えることによってさらに応用範囲が広がる可能性があり，前節に述べた用語データバンクと共に将来が注目される。

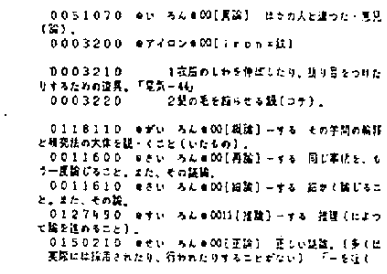


図4－34 “IRON”と入力した時の逆引き辞書の出力[16]

4.9.2 種々の翻訳システムにおける辞書

(1) 構文パターンマッチングによる方式(注)

論文表題の翻訳システムにおいては，形態素処理の部分を省いてあったために，辞書の項目としては「－ing」形などをつけた形も登録しておかなければならなかった。辞書の内部構成が実際にはどのようになっているかは，56年度報告書に記述があるので詳細は省くが，以下に述べる他のシステムとそれほど違わない構成を持っている。また，このシステムでは構文パターンマッチングの際に名詞の意味カテゴリを考慮することによってある程度意味処理も行えるようになっている。

(2) トランスファ方式

トランスファ方式の機械翻訳システムでは辞書がかなり重要な役割を持つ。構文解析を比較的表層部分だけで済ませているシステムにおいても辞書にそれだけ負担がかかってくるし，意味の部分まで深く扱おうとすると，こんどは辞書の意味記述を詳しくしなければならないからである。ここではARIANE－78[5][6]と，テキサス大学の機械翻訳システム[7]をとり上げる。

(注) 4.9.2の各項目は4.8.2の各項目に対応している。

ARIANE−78では，形態素解析部と，語彙トランスファ部，それに形態素生成部にそれぞれ辞書を持っている（図4−35，ただし生成部の辞書は略）。これらは，それぞれ，ATEF，TRANSF，SYGMORと呼ばれるプログラムを起動した時に参照される（図4−15参照）。ただし語彙トランスファの場合で，木の形を変換する必要性が生じた時は，ROBRAの方で行う。

(a)
```
DID          ==INVPT(PNITO, DO      ).
DISPLAC      ==V2   (PN1   , DISPLACE).
DISPLACEMENT==DVN1 (PN1   , DISPLACE).
```
注）形態素および構文フォーマットには，言語学者によって次のような解釈が与えられている。
```
INVPT==CAT=V,SUBV=VB,TENSE=PRET,NUM=SIN/PLU.
V2   ==CAT=V,SUBV=VB,VEND2.
DVN1 ==CAT=N,SUBN=CN,DRV=VN,NUM=SIN,NEND=1.
PNITO==VL1=N,VL2=TO,SEM=PROC.
PNI  ==VL1=N,SEM=PROC.
```

(b)
```
'INCLUDE'==$VERB/0(1,2)/0:'$CHOICE',*AGSEM;
                          1:'HOUGANS',$GN;
                          2:'GANYUUS',$CHEMI;
          $NOUN/0(1,2)/0:'$CHOICE',*AGSEM;
                          1:'HOUGAN',$GN;
                          2:'GANYUU',$CHEMI.
```
注）
```
$CHOICE ** POLYSEMY.
*AGSEM  ** AGREEMENT BY SEMANTIC CLASS.
$GN     ** GENERAL USE.
$CHEMI  ** CHEMICAL USE.
```

図4−35　ARIANE−78における(a)形態素解析辞書および(b)語彙トランスファ辞書[5]

図4−35を参照すれば分るように，形態素解析用の辞書は次のような形式を持っている。

＜文字例＞＝＝＜形態素フォーマット＞（＜構文フォーマット＞＜語彙ユニット＞）

また，語彙トランスファ用辞書は次のような形をとる。

＜元の言語の語彙ユニット＞＝＝＜条件＞／＜目標言語木構造＞／＜目標言語語彙ユニットと構文情報＞

さらに図には示していないが，形態素生成用辞書は次の形をしている。

＜語彙ユニット又は変数＞＝＝＜条件＞／＜割付け＞／＜文字例＞

上の表記法からも分るように，これらの辞書項目は，言語学者にとって比較的記述しやすい形になっている。実際に，この部分は，言語学者が記述しており，それによってシステムの質の向上に役立っている。

また，テキサス大学の翻訳システムでは，格フレームを中心とした構文解析を行っているが，その辞書項目部分の記述は，構文解析部では，訳文がうまく生成できるように種々の特徴値を入れた表現となっているのが特徴である。たとえばout putという語には，文法カテゴリは名詞語幹，異形態はoutput のみ，語尾変化は単数ではナシ，複数ｓ，性はunmarkedなどの特徴値がつけられている。

このシステムのトラスファ用の辞書は図4－36 に示すようなもので，ごく簡単な対訳と特徴値から成っている。この辞書項目記述からもある程度推察されるように，ARIANE—78などのように，深層格までふみこむことをなるべくせずにシステムを構築しようとしているのがこのシステムの特徴である。しかしながらこのシステムでも言語学的部分とプログラムとはかなりよく分離されており，将来辞書を大幅に改善することもできそうに思われる。

```
[TRANSFER-LEX
( ON                    (AUF) PREP (PRCOM NIL))
( OUTPUT                (AUSGABE) NST)
( THE                   (DER) DET (KD DET))
( GO                    (GEHEN) VST (PX NIL) (PF FIN INF PAPL))
( MAGNETIC-TAPE         (MAGNETBAND) NST)
( AFTER                 (NACH) PREP (RO TMP))
( HOUR                  (STUNDE) NST)
]
```

図4－36　テキサス大学の翻訳システムにおけるトランスファ辞書の例[7]

(3) 融　合　方　式

融合方式では，前述のように，木の陽な変形を行わないで，解析，変換，合成を一気に行う。従って融合方式では，トランスファ方式のように，構文解析や語彙トランスファの部分などで別々の辞書を用いるのと

は異なって，辞書項目は一種類である。

辞書の記述には，意味表現言語が用いられる。英語動詞 open についての記述を図4—37に示す。最初のVは動詞，次がいわば形態素辞書にあたる。3行目は，構文解析の節点上に送る情報である。そして4行目以下が意味表現部分である。まず，すぐに分るように open に対して「開く（ku）」という訳語が対応している。通常の場合は，これが訳語になる。part-of，self は，はそれぞれ全体と部分，上位と下位の関係を表す。また，sf は意味素性（semantic feature)を表し，ここでは open が行為をする動詞であることが示されている。さらに，それ以下は，前にも説明したslotの部分が続く。これ以下が，いわば，訳語の選択を行う部分になっている。このslotが適切なfillerで埋まると，求める訳文が生成される。その時，図から分るように，元々の訳語「開く」ではなく，「開ける」という訳語となる部分は，slotの記述の中に，そのような書きこみがなされている。

以上のように，融合方式においては，一般の辞書とかなり似たような記述になっている。特に訳語の選択にそのような傾向が見られる。機械翻訳システムの中には，このような訳し分けに対してそれぞれ別の辞書項目をたてるやり方をとっているものもあるが，辞書作成の上からは，余り項目数を増さない方がよいと思われる。ただ，語彙が増加するにつれて記述が爆発的に増加する恐れがないか，また辞書項目の記述をある程度対話的に行えないかなどといった問題点も今後検討される必要があろう。

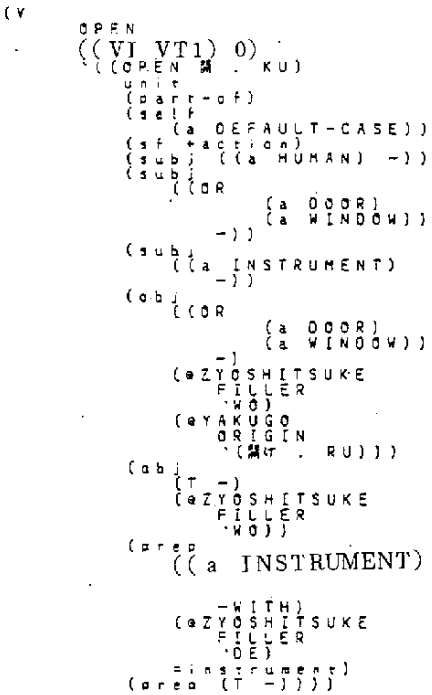


図4－37 融合方式におけるopenの辞書項目例[8]

(4) モンテギュー文法

モンテギュー文法による機械翻訳システムでは，最初の構文解析部分のところは，通常の句構造解析と同じであるため，辞書項目は品詞付けとほぼ等価になっている（図4－38(a)）。しかしながら，訳語にあたる部分は，トランスファ方式などの場合とは違って，辞書項目を論理式に変換する過程として表されるために，辞書項目の形ではなく，むしろ変換規則として与えられている（図4－38(b)）。この図から分るように，冠詞や動詞などは論理式を含んだ表現（λはLISPなどでも用いられるラムダ表現のオペレータを表す）に変換されるが，名詞は対応する訳語がつけられている。矢印の右辺にクラスなどの条件をつけることによって訳語の選択もできるようになっている。このように，このシステムでは，翻訳の核部分を，辞書ではなく，規則で記述するというのが特徴である。

また，ミニコン上で稼動しているシステムであるために，辞書に対し

て余り大きな容量が取れない。その防止策として，辞書項目や変換規則を対話的に修正する機能を付けているのもこのシステムの大きな特徴である。なお，対話的修正については4.10節でさらに述べる。

(a)
```
辞書項目：DET (determiner, 限定詞):
          {a, no, the, …}
        NOUN (noun, 名詞):
          {student, textbook, …}
        VT (transitive verb, 他動詞):
          {has, …}
```

(b)
```
no⇒λpλq[¬＋否定 (q(どの p も))]
a⇒λpλq[q(一つの p)]
textbook⇒教科書
student⇒生徒
have⇒λuλx[u がｘをもつ]
```

図4－38　モンテギュー文法を用いた翻訳システムにおける
(a)　構文解析用辞書項目　(b)　辞書項目の翻訳規則[9]

(5)　概念依存を用いた方法

概念依存を用いた機械翻訳システムでは，概念を抽出する過程と，概念から訳文を生成する過程とが全く分離されているので，各々の過程で用いる辞書は全く別になる。

概念を抽出する過程では，構文解析というよりは，むしろパターンマッチング的に文の解析が行われる。従って辞書の記述もslotを開けておいて，それを埋めるという，むしろプログラム的な記述で書かれている（図4－39）。

また，訳文を生成する過程では，概念表現の部分での対応語をつけるといった記述のしかたで，むしろ対訳辞書の形で行っているようである。いずれにせよ，概念の抽出がきちんと行われていれば，生成部ではむしろ文の順序づけや形態素処理（活用語など）にプログラムの中心的部分が注がれているのは当然であろう。

```
(DEF-WORD JACK
 ((ASSIGN *CD-FORM* '(PERSON (NAME (JACK)))
          *PART-OF-SPEECH* 'NOUN-PHRASE)

 ;GOT is a verb that means someone ATRANSed something to the subject.
 ;GOT looks for a noun phrase to fill the object slot.

(DEF-WORD GOT
 ((ASSIGN *PART-OF-SPEECH* 'VERB
          *CD-FORM* '(ATRANS (ACTOR ?GET-VAR3) (OBJECT ?GET-VAR2)
                            (TO ?GET-VAR1) (FROM ?GET-VAR3))
          GET-VAR1 *SUBJECT*
          GET-VAR2 NIL
          GET-VAR3 NIL)
  (NEXT-PACKET
     ((TEST (EQUAL *PART-OF-SPEECH* 'NOUN-PHRASE))
      (ASSIGN GET-VAR2 *CD-FORM*)
```

図4－39　概念を抽出するための辞書項目[12]

(6) ま と め

種々のシステムについて，その辞書項目がどのような記述になっているかを見てきた。それによれば，構文解析の部分を簡単にすれば（たとえば単なるパターンマッチングで済ませるなど），その分だけ辞書の部分に負担がかかってくるし，構文解析の部分を詳細にして，意味解析などを含ませた場合にも，辞書項目にその部分を書きこむ必要があるので，やはり辞書が複雑になる。将来のシステムでは，ますます辞書の記述が詳細，複雑化することは避けられない。また，語彙が増加すれば，もちろん辞書項目数も増大していく。従って，メンテナンスしやすいような辞書構造を作ったり，対話的に辞書を作る試みが必要となるが，それについては次節で論じる。

4.10 辞書構造の一考察

辞書項目を作る場合に留意すべき点は，既に56年度報告書にも指摘されているように，翻訳に要する十分な語数を含んでいることと，拡張性に富むことであるが，その他にも次のような点が肝要である。

- 修正が容易にできること－これは拡張ということともつながるが，たとえば辞書項目のある語の内容を書きかえたいとか，意味素性を全体にわたって変更したいとかいうときに，容易に修正できるか否かといった点である。

- 作成された項目の形が見やすいこと—たとえばARIANE−78では，言語学者が，言語学流のフォーマットで記述した辞書項目が，ほぼそのままの形で内部表現に書き換えられるという。いわば，計算機の非専門家がシステムの修正に参与できる形になっている。また，融合方式では，英和辞典にかなり近い形式での記述が，そのまま内部表現となっている。このように見やすい項目作成をしておくと，修正も容易になるのは論をまたない。
- アクセスがしやすいこと—辞書項目が大規模化してくると，主記憶上に辞書を置いておくことが困難になる。仮想記憶などで大きい記憶領域を（見かけ上）用いることができても，やはり効率のよい辞書項目のマッチングが必要である。そこで，データベースなどで盛んに用いられている種々の手法を用いて，辞書項目の検索を容易にするというやり方が行われている。

また，辞書項目と直接関連はないが，辞書項目に含まれていない語が出てきた時にどうするかという問題がある。パターンマッチングによる構文解析では，このような状況に対処しにくいし，通常の構文解析システムでも，いったいどの部分が未定義語になっているのかを判断するのは非常に困難を伴う問題である。これも今後の課題の一つであろう。

4.10.1では，辞書項目の修正について，辞書の構成とともに論ずる。また，4.10.2では，二次記憶上でどのように辞書を構成したらよいかについて簡単に触れる。さらに，4.10.3では，辞書項目の構成について論ずる。

4.10.1 辞書項目の修正

辞書項目がシステムのプログラム部分と分離しているもの（ほとんどのシステムはそうであるが）では，外部のエディタ（通常計算機システム上でサポートされているような）を用いて項目を修正し，再びシステムに組み込むのが普通である。しかしながら多くのシステムでは，新しい辞書項目の追加や，古い辞書項目の削除，置き換えなどは，システム内の機能（言

語がＬＩＳＰの場合には関数）を用いて修正が可能になっている。さらに，システムが処理を実行中に項目の修正追加を対話的に行って正しい翻訳を出すシステムもある。

ところが，稼動中に修正，追加を許すシステムでは，いわば人手の介入を認めたことになるという意見もある。また，辞書項目へのアクセスを速やかに行うために，項目自身をコンパイルしてしまうようなシステムでは，修正そのものが困難であるという観点から，この機能のないものも多い。しかしながら完全な機械翻訳が困難な現状では，未定義語が出現したり，構文解析できない文が入力された時には，何らかの対話的機能があった方がよいと思われる。この機能とは，たとえば，「～という語が辞書の中にありませんか」というような問を発して，それに自然言語で答えさせるといったものである。ただ，前述のように，未定義語を見出すのはむずかしい（特に日本語のように分かち書きをしない語では）し，この質問応答機能が余り煩雑に起動されるようになると，何のための機械翻訳か分らなくなる恐れがある。機械翻訳とは全く分野を異にするが，今後このような人間一機械の自然言語による質問応答システムの研究（特にデータベースに対する問合せという観点からは多くの研究がある）の成果を取り入れて，機械翻訳システムのユーティリティの一つとする方向が必要になると考えられる。

4.10.2　二次記憶上の辞書項目の構成

辞書が大規模化，複雑化してくると，それをうまく管理して迅速なアクセスが実現できるようにする必要性が生じる。このような大規模なデータを扱う問題というのは，昔から情報検索の問題として扱われてきており，さらに近年は，データベースに対するアクセスの研究として発展している[21]。ここでは自然言語処理や機械翻訳との関連でごく簡単に触れることにする。

まず，よく用いられるのはハッシングの手法である。ハッシュ（hash：

ごたまぜにする）という語からも分るように簡単な計算式で計算された位置に，実際の辞書項目を置いておくというやり方をする。たとえばaに1，bに2などと割り当てておいて，ストリング列の和を計算式にすると，abなら3，abcなら6という位置に配列される。うまい計算式（通常これをハッシュ関数と呼んでいる）を選ぶと，ほとんど一回で，所望の項目を得ることができる。同じところに異なる項目が重なった場合（たとえば上の例ではab，ba，cはいずれも同じ場所になる）には，たとえば後から入ったものを1つ後ろ送りするとか，もう一度関数を作り直すといった方法がある。

また，ストリング列を木構造に並べていく方法もある。たとえば，abcというストリングは，aという根を見つけて，bという節点に進み，さらにcまで枝をたどるというやり方で辞書項目を見出す。辞書項目のコンパイルの仕方をこのように行っているシステムもある。

ストリングレベルではなく，辞書項目の順番のレベル（各辞書項目をそれぞれ1レコードと見た時のレコード順）で木構造的な索引を編成するのが，B氏木（B－tree）である。B氏木の特徴は，拡張性に富むことで，内容をかなり大幅に変更しても，木構造自体をそんなにいじらなくても済む。また，木自体がバランスがとれていて余りいびつな格好にならない。最近では，B氏木のこのような特徴を生かし，さらに機能を付け加えて辞書が検索できるようにしたシステムも作られている[22]。

このように，種々のファイル形態を工夫して，二次記憶上に大規模な辞書を置くことにより，システムの大規模化がはかれると考えられるが，その作成をサポートする種々の機能が必要なこともちろんである。

4.10.3 辞書項目の構成

これまで見てきた多くのシステムでは，辞書項目は，次のような形をしていた。

（単語　文法カテゴリ　構文情報　意味情報）

この順序，内容はシステムによって異なっているし，構文情報と意味情

報の間に値を入れる欄を設けて構文解析木上で評価し，点の高い木を優先的に出力させて曖昧性を減らす試みも見られる。しかしながら，辞書項目と文法規則という形でシステムを構成する限り，今後ともこの基本形に大きな変化はないと見てよい。

機械翻訳に残された問題は数多くあるが，たとえばその一つである談話（discourse)の問題などは，文法規則の上で扱われる可能性が大きいと考えられる。ただしその時に，構文，意味情報とは異なる何らかの情報を辞書項目に付加する必要があるかも知れないが，まだ推測の域を出ない。

5. 今後の課題

5. 今後の課題

本プロジェクトの主要な課題は「データベース化された大量の文章情報を分析し，そこに潜在する徴候的情報，基調的変化を的確かつもれなく把握する」ことをコンピュータの活用により可能とすることである。

文章情報においては，数値と異なり1つの情報が多くの意味を持っている事が特徴である。しかし，対象とする分野を限定する事によって，多様な意味を一意的に特定することが可能になってくる。このような点からエネルギー分野をモデルに，上記課題にアプローチしていくこととした。

本年度は，海外で発行されている主要50紙誌の中からエネルギー問題を論及した主要記事を対象に，データの整備を開始するとともに，データベース構築のための作成・更新システムの開発を行った。

また，内容分析の検討素材として，既存の研究事例（国際紛争に関する研究，General Inquirer の内容分析，認知構造図分析等）の調査を行うとともに，認知構造図分析手法のエネルギー分野での適用実験を行い，必要な機能及び入力データを明らかにしてきた。

さらに，内容分析のための基礎技術であり，かつ，最近再び活発化してきた機械翻訳に具体的に応用されている自然言語処理技術について，海外の先進事例の調査を行うとともに，最近の技術動向を把握し，現在の水準を明らかにしてきた。

以上の結果を踏まえ，総合解析システムの基本構想として，本来あるべき姿を検討し，具備すべき機能を明確にした。

しかし，現在の自然言語処理技術の水準からみて，それらの全てを本調査研究で解決することは不可能であると考える。

本プロジェクトでは，既存の可能な限りの利用技術を駆使し，理想的なシステムとは言わないまでも，目的に合ったシステムの実用化をめざして検討を進めている。

今後は，各種基礎技術の調査研究を引き続き行うとともに，以下の点について具体的に着手することとする。

① データベースに蓄積するデータの整備

② 検索機能の設計開発

③ 定量化分析モデルの設計開発

④ 文の解析・変換・生成実験

⑤ 文法及び辞書の作成実験

参考文献（4.8～4.10節）

(1) 長尾・辻井・健部：技術論文　表題の英和自動翻訳の試み，情報処理学会計算言語学研究会資料　17－2（1979）

(2) Y．Arens：Using Language and Context in the Analysis of Text，Proc．7th IJCAI，pp 52－57（1981．8）

(3) 長尾・辻井・矢田・柿元：技術論文　表題の英和自動翻訳を利用した文献速報システム，情報処理学会才23回全国大会　2G－7（1981．10）

(4) C．Boitet & N．Nedobejkine：Russian－French at GETA：Outline of the Method and Detailed Example，Proc COLING－80，pp 437－446（1980）

(5) 岡田直之：ヨーロッパにおける自然言語処理の現状 ―― グルノーブル大学における機械翻訳を中心として ――，情報処理学会自然言語処理研究会資料　30－2（1982）

(6) 辻井潤一：ヨーロッパにおける計算言語学の様子，情報処理学会自然言語処理研究会資料　31－3（1982）

(7) 首藤公昭：テキサス大学における機械翻訳，情報処理学会自然言語処理研究会資料　28－6（1981）

(8) 田中・元吉・安川：意味表現言語SRLの機械翻訳への応用，情報処理学会自然言語処理研究会資料　31－5（1982）

(9) 西田・堂下：モンテギュー文法に基づく機械翻訳への新しいアプローチ，bit，vol．15，№3，pp 231－248（1983）

(10) R．C．Shank：The Structure of Episodes in Memory，in〔Bobrow et al，1975〕

(11) 石崎俊：意味情報を用いた日本語の生成 ―― イエール大学における自然言語処理 ――，情報処理学会自然言語処理研究会資料　34－9（1982）

(12) L . Birnbaum & M . Selfridge : Conceptual Analysis of Natural Language , in [Shank etal 1981]。

(13) R . C . Shank & C . K . Riesbeck (eds) : Inside Computer Understanding —— Five Prograss Plus Miniatures , Lawrence Erlbaum Associates (1981)

(14) D . G . Bobrow & A . Collins (eds) : Representation and Understanding —— Studies in Cognitive Science , Academic Press (1975)

(邦訳) 渕一博監訳：人工知能の基礎 —— 知識の表現と理解，近代科学社 (1978)

(15) Longman Dictionary of Contemporary English , Longman (1978)

(16) 横山晶一：国語辞典データベース化の準備，電総研彙報 vol . 41 №11 , pp 855 − 863 (1977)

(17) 横山晶一：国語辞典データベースの基本データ，計量国語学 vol . 13 №3 , pp 120 − 134 (1981)

(18) A . Michels : Exploiting a Large Dictionary Data Base , Université de Liége (1982)

(19) 池田尚志：機械翻訳システム，〔田中他 1981 〕pp 166 − 201 (1981)

(20) 田中穂積他：自然言語処理と言語理論，電総研調査報告第205号(1981)

(21) 植村俊亮：データベースシステムの基礎，オーム社 (1979)

(22) 日高・稲永・吉田：拡張Btreeによる日本語言語辞書の作成，情報処理学会自然言語処理研究会資料 33 − 8 (1982)

(23) 長尾真：機械翻訳，電子通信学会誌 vol . 65 №4 , pp 386 − 392 (1982)

昭和58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丁目5番8号
機械振興会館内
Tel（434）8211（代表）

印刷所　株式会社　タケミ印刷
東京都千代田区神田司町2－16

57-S 004

原本（持出厳禁）

受付No.	F-18
受付日	
作成課	