59－S002

文章情報データベース総合利用
調査研究報告書

昭和60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５９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，事前に兆候を把握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，これらの情報処理過程においては，なるべく人手を介さない，自動的な機械処理が可能な形が望まれている。

こうしたことから，本事業においては，文章情報データベースを効果的に利用する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，各種調査研究並びに機能開発等を実施している。

初年度は基礎的な調査研究を実施し，第２年度はこれに基づく課題を研究するとともに海外先進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。第３年度はひきつづき各種機能研究や海外調査を行うとともにデータベース検索システム，定量化分析モデルシステムの開発，文章情報の各種解析処理実験等を行った。

最終年度として本年度は，これまでの実施成果を総合解析システムの一つの流れで機能させることに留意して，この為の中核機能となる基礎解析システムと，これにより生成される二次情報データベースの実験開発を重点的に行った。また，これによる文章情報の効果的な利用につき，検索，内容分析，翻訳等の各面から検討し，望ましい総合解析システムの構築に向けた今後への展望と課題をまとめた。

最後に，４カ年にわたる本調査研究の推進にご指導，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６０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	東京工業大学工学部情報工学科助教授
〃	矢田　光治	知識情報研究所所長
〃	所澤　仁	(財)工業開発研究所主任研究員
〃	湯浅　俊昭	(財)日本エネルギー経済研究所第4研究室長
〃	小澤　大二	国際協力事業団沖縄国際センター所長
〃	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 大阪本社総合情報部
〃	大倉　健治	(株)市況情報センター開発第2部長
〃	海老原悦夫	日本貿易振興会海外事業部施設課
〃	瀬谷　重信	日本電信電話公社関東電気通信局データ通信部長
〃	山本恵美子	ソフトウェア研究会副会長
〃	木南公統司	(株)開発計算センターシステム管理部課長
〃	長谷川　亨	(株)ソフトウェア・リサーチ・アソシェイツ開発本部開発第5部長
〃	中瀬　純夫	リソースシェアリング(株)開発技術第2部長
〃	竹内　憲	日本タイムシェア(株)第1システム事業部長付
〃	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(株)マーケティング推進部次長

委　員	渡辺　龍雄	前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
〃	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政策情報システム室技術係長
〃	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
〃	(前) 市川　隆	(財)日本情報処理開発協会技術調査部長
〃	(後) 瀬戸　良洋	(財)日本情報処理開発協会技術調査部長
〃	(前) 難波　正之	(財)日本情報処理開発協会技術調査部次長
〃	(後) 山川　浩二	(財)日本情報処理開発協会技術調査部資料編集室長

「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	石川　徹也	図書館情報大学図書館情報学部助教授
〃	松井　賢一	(財)日本エネルギー経済研究所エネルギー計量分析センター研究主幹
〃	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
〃	海老原悦夫	日本貿易振興会海外事業部施設課
〃	木南公統司	(株)開発計算センターシステム管理部課長
〃	長谷川　亨	(株)ソフトウェア・リサーチ・アソシェイツ開発本部開発第5部長
〃	中瀬　純夫	リソースシェアリング(株)開発技術第2部長
〃	竹内　憲	日本タイムシェア(株)第1システム事業部長付
〃	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
〃	(前) 難波　正之	(財)日本情報処理開発協会技術調査部次長
〃	(後) 山川　浩二	(財)日本情報処理開発協会技術調査部資料編集室長

「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	知識情報研究所所長
〃	田中　穂積	東京工業大学工学部情報工学科助教授
〃	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	日本経済新聞社大阪本社総合情報部
〃	大倉　健治	㈱市況情報センター開発第２部長
〃	瀬谷　重信	日本電信電話公社関東電気通信局データ通信部長
〃	竹内　　憲	日本タイムシェア㈱第１システム事業部長付
〃	池田　泰則	コンピュータサービス㈱システム機器販売事業部特機販売部次長
〃	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・アソシェイツ開発本部開発第５部長
〃	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政策情報システム室技術係長

目　　次

1. 事　業　概　要

1. 事業概要

1.1 目的と背景

コンピュータや通信技術の進展に伴い，各種情報利用は，広く社会・生活面の深部に至るまで浸透し，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や高度情報通信システムの進展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の一途にある。ニューメディアの台頭も目覚しく，益々増大，多様化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っている。

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のに必要となる各種機能を備えた総合解析システムを開発することを目的として，調査研究を実施する。

1.2 事業計画と実施経過

本事業の全体計画と実施概要は図1.2－1に示すとおりであり，4ヵ年にわ

たる調査研究並びにシステム開発等の実施経過を年度別に示すと次のとおりである。

(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。

本事業初年度にあたり，文章情報データベースの総合的な利用のために必要となる各種分析手法や，総合解析システム開発に際し具備すべき諸機能につき基礎調査を実施した。また，総合解析システムとして，あるべき姿と現状で実現可能なシステム形態を並行して検討しつつ，次年度以降のシステム開発に向けて基本構想を作成した。

具体的には，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イスの研究
- ＬＣ－MARCの実験準備

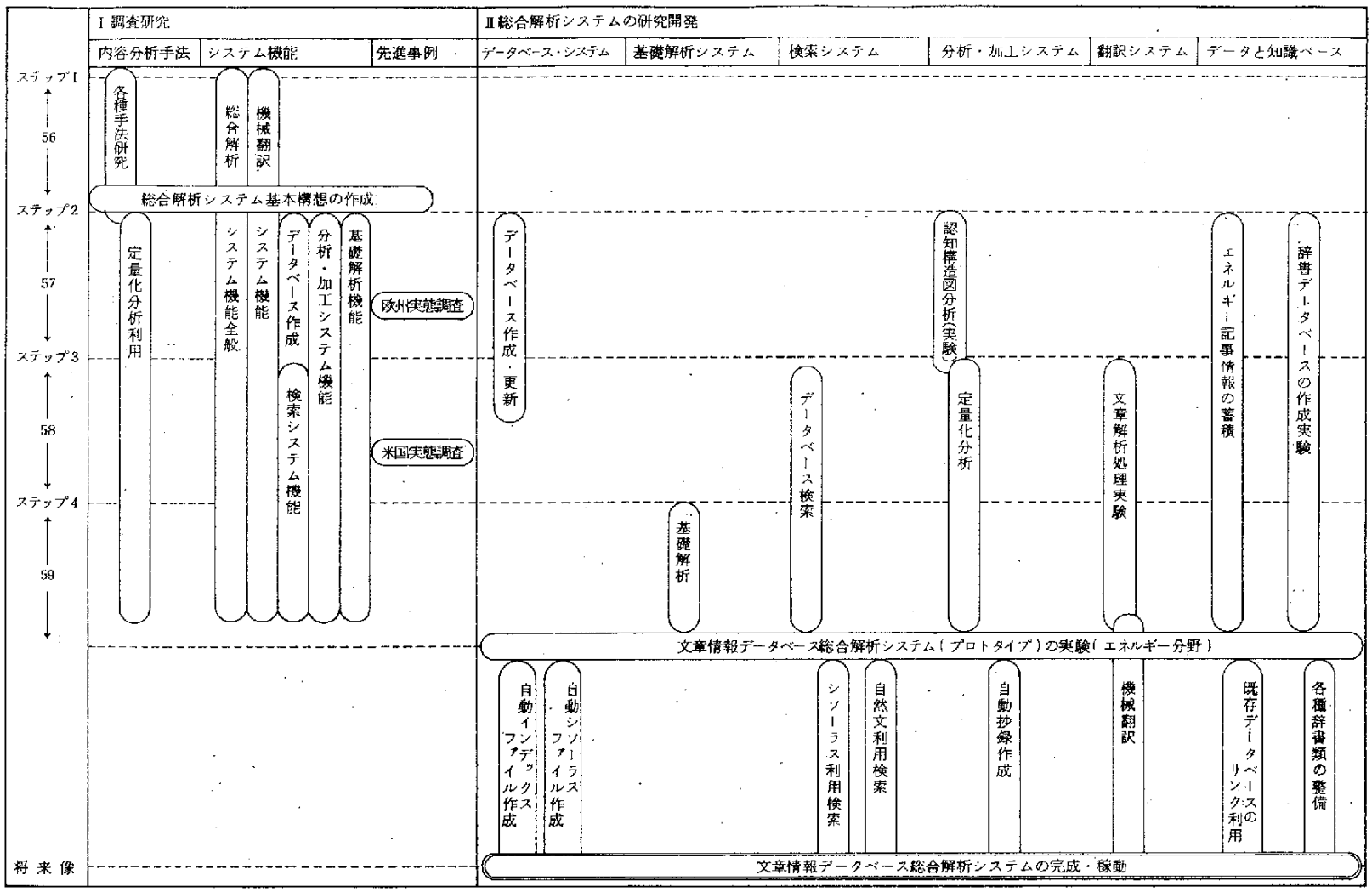


図1.2－1　全体計画と実施概要

(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，次の研究を実施した。

「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。

(3) 昭和５８年度

５６，５７年度に引き続き委員会並びに２つの専門委員会を設置して調査研究を推進した。

「文章情報データベース定量化利用研究専門委員会」においては，

- 記事情報インデックスのキーワード出現頻度数分析
- キーワード自動抽出システムの現状

○　文章情報内に出現する自然語の頻度分析

について事例研究を行うとともに，５７年度開発したデータベース作成・更新システムに付加する検索システム並びに定量化分析のモデルシステムの開発を実施した。また，エネルギー記事情報の本格整備を図り，データベースを構築した。

「機械翻訳システム研究専門委員会」においては，翻訳アルゴリズムや文法規則，辞書等の研究を次の実験を通して行った。

○　PROLOGを利用した文章解析処理の実験

○　エネルギー分野における用語収集実験

また，米国の研究機関を中心に自然言語処理技術の動向やデータベース作成処理の実態調査を行い参考に資した。

(4)　昭和５９年度

これまでと同様，委員会並びに２つの専門委員会を設置し，本事業最終年度としてこれまでの成果を１つにまとめる方向で調査研究並びにシステム開発を推進した。特に，総合解析システムの中核機能として，検索，内容分析や翻訳に使用される二次情報の生成を行う基礎解析を中心に機能を研究，整理した。

また，５８年度開発した定量化分析モデルシステムにつき基礎解析機能との連動に係わる分析機能の拡張を検討した。

並行してモデルデータベースの充実を図り，これまでに蓄積したエネルギー記事情報に付加する同種データを，５９年１２月分まで整備して，エネルギーデータベースに構築した。

これら５９年度作業と過去３ヵ年の研究開発成果を基に，範囲を限定してプロトタイプ型の総合解析システムの開発実験を行い，今後への展望と課題をとりまとめた。

なお，５９年度本調査研究の経過概要は表1.2－1に示す通りである。

表1.2－1　昭和59年度調査実施経過表

項目 ＼ 月	59/4	5	6	7	8	9	10	11	12	60/1	2	3
委員会の開催												
・総合委員会					2 ○							26
・定量化専門委員会						28 ○				19		
・翻訳専門委員会							26 ○					
実施計画の策定	←	―	→									
文章情報総合利用の研究			←	―	―	―	―	―	―	―	―	→
総合解析システムの開発												
・エネルギー記事情報の整備					←	―	―	―	―	→		
・二次情報データベース構造の研究					←	―	―	―	―	→		
・基礎解析サブシステムの開発					←	―	―	―	―	―	→	
・定量化分析モデルの拡張								←	―	―	→	
・開発システムの実験・評価											←	→
報告書の作成											←	→

1.3　報告書の構成

本報告書は次の３章からなっている。

まず第１章事業概要で事業主旨，４ヵ年にわたる調査研究の実施経過を概説した。

第２章総合解析システムの構築においては，まず基本構想と，それに基づく開発アプローチで，研究開発を段階的に実施してきた経緯を述べた。更にこれをうけた開発実験として，これまでの個々の成果を総合的な１つの流れにまとめる方向で，５９年度実施したプロトタイプ型の総合解析システムの開発につき実験内容を詳述した。また，総合解析システムの機能とその利用では，総合解析システムとして具備すべき主要な機能につき，その目的，機能，構造を解説するとともに考えられる利用のイメージを例示した。

第３章展望と課題では本事業所期の目的と実施内容，成果を照らして残された課題を整理し，各所における先端的な研究状況も勘案して今後への展望をまとめた。

なお，４ヵ年を通して，各年度の調査研究成果を表1.3－1の報告書にまとめてある。

表1.3－1　調査研究報告書

５６　Ｒ０１１	文章情報データベース総合利用調査研究報告書
５７　Ｓ００４	〃
５８　Ｓ００３	〃
５９　Ｓ００２	〃

2. 文章情報データベース総合解析システムの構築

2. 文章情報データベース総合解析システムの構築

2.1 総合解析システムの基本構想

2 1.1 背景ー情報需要の変化ー

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。

2.1.2 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　　等

2.1.3 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。（図2.1－1参照）

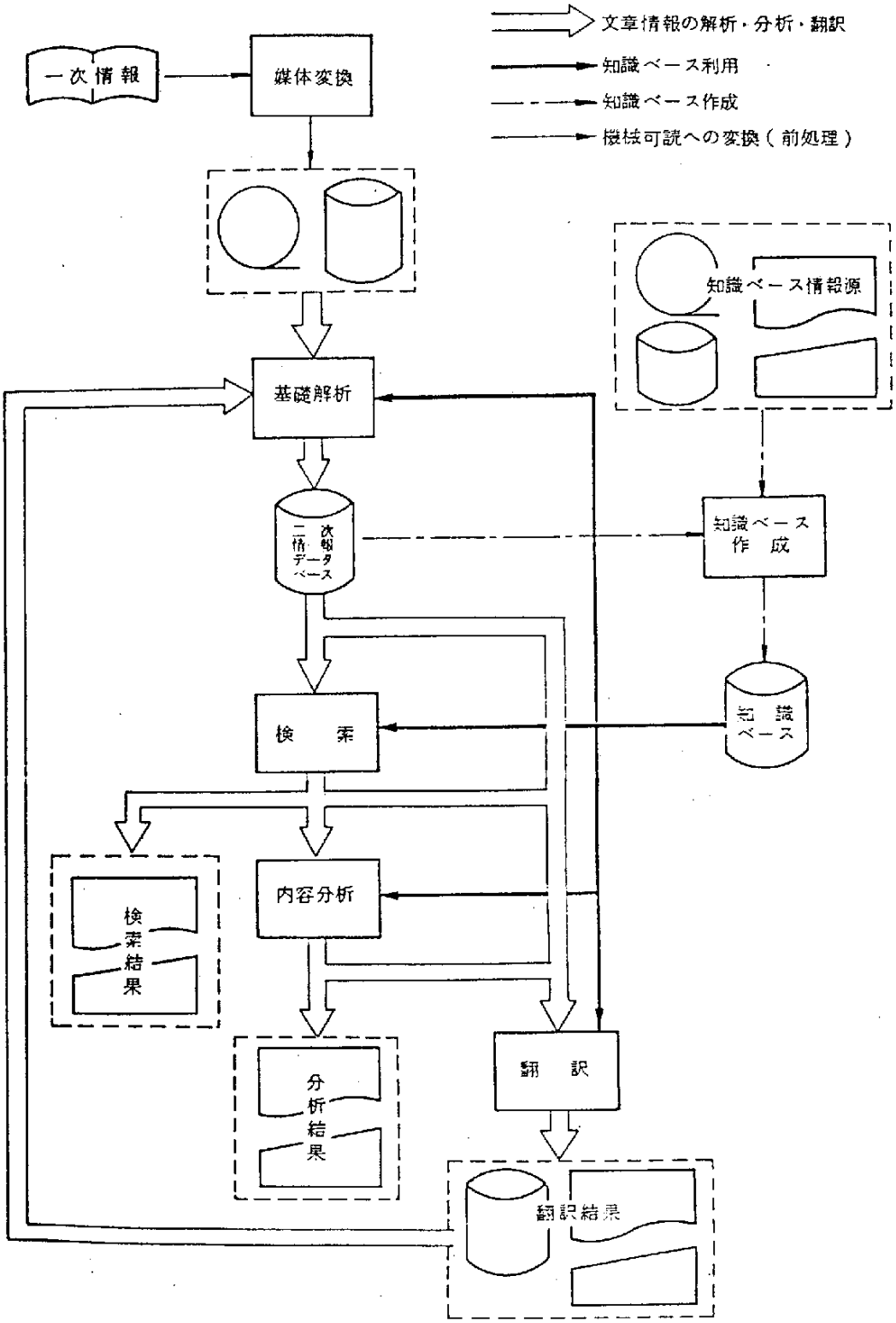


図 2.1－1　総合解析システム概念図

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。

2.2 総合解析システム開発のアプローチ

前節で述べた基本構想に基づき，各種分析手法やシステム機能につき一貫して基礎研究を進める一方，実現可能なところから順次システム機能の研究開発を実施した。

調査研究の一環としては海外も含めた形での先進事例調査を行い，実データを用いた文章解析や情報内容分析の実験等により，システム開発への参考に資した。

システム開発にあたっては，まず，関心度，実用性等を勘案してエネルギー分野に対象を絞ってモデルデータベースを構築した。

これと並行して，総合解析システムとしてのあるべき姿を常に考察しつつ，現状の技術水準を踏まえた主要機能の開発を実施した。

以下これまでに開発したシステム，実験や研究を行ったシステム機能，モデルデータベースにつき，総合解析システムにおける位置づけの観点から手順を追って実施状況を振り返ってみる。

2.2.1　データベース作成・更新サブシステムの開発（５７年度）

本サブシステムで取り扱うデータは海外のエネルギー関連情報を基に，専門家により日本語で作成したインデックスである。

従って，データベースからみると総合解析システム概念図の二次情報データベースに該当するものであり，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本サブシステムを開発した。

情報の内容は，大別して，データの所在，具体的内容，入手利用の為の情報の３種類であり，システムの機能は入力データ作成，データベース作成，更新，修正の４つがある。

（５７年度報告書　Ｐ３９～５１に詳述）

2.2.2　データベース検索サブシステムの開発（５８年度）

本サブシステムは概念図の検索部分のベースになるものであり，データベース作成・更新サブシステムを活用して開発した。

システム機能は１次検索，論理検索，２次検索，頻度検索の４つであり，処理形態は，応答形式のインタラクティブな処理とした。

（５８年度報告書　Ｐ９～２３に詳述）

2.2.3　定量化分析サブシステムの開発（５８年度）

文章情報総合解析システムの内容分析機能につき，基礎的な一手法として情報内容の定量化による分析をとりあげ機能研究，分析実験の結果を検討して本サブシステムを開発した。

対象とする情報は検索サブシステムの頻度検索で得られるキーワードの出現頻度数や生のキーワードの出現頻度に加工を加えた基礎情報である。

システム機能としては，データ準備のためのキーワード頻度数の検索，保存，データ加工，生データ及び加工データグラフ出力機能をまず用意した。これは検索サブシステムに組み込む機能とも言える。分析機能としては，分類のためのクラスター分析，次元縮小のための主成分分析，相関関係や予測のための回帰分析をとりあげて考えた。このうち，キーワードを基にした頻度の時系列・クロスセクション分析を可能とするものであり，さらに必要に応じ主成分分析，回帰分析をも可能とするものである。

また，処理形態は応答形式のインタラクティブな処理とした。

（５８年度報告書Ｐ２４～３５に詳述）

なお，５９年度，基礎解析機能開発に伴う分析機能の拡張につき実験を行った。（2.3，2.4.4を参照）

2.2.4　翻訳機能の実験（５８年度）

日本語の文章のみならず海外の文章情報も同時に処理を可能とするために，翻訳機能が必要である。機械翻訳のレベルは，入力文の解析と出力文の生成である。文の解析，生成はさらに多数のレベルに分割される。

このうち，入力文の解析部分はすべて基礎解析でなされ，解析の結果は二次情報データベースの中に格納される。翻訳では，解析の情報より目的とする言語の生成が主な機能となる。

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，人手による後編集に委ねられる所が大きい。

機械翻訳機能については一部実用システムが開発されているものの，自然言語処理技術や人工知能の研究に関連するものとして厳密には未だ研究途上にあると言える。

そこで，総合解析システムにおける翻訳機能としては，５８年度においてPROLOG言語を利用した文章解析処理の立場から英日機械翻訳実験を試みた（５８年度　Ｐ１３８～１７３に詳述）ほか，５９年度エネルギー記事情報の解析処理による事実表現の観点から翻訳機能の研究（2.3，2.4.5参照）を進めるにとどめた。

2.2.5　基礎解析機能の開発実験（５９年度）

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，形態素解析，構文解析，意味解析，文脈解析の各処理要素で構成される。

これらの一連の処理は相互に補完しつつ，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

このうち本文情報の先頭文を対象に解析を行い，事実情報として編集することにより，次節2.3の総合解析システムの実験を試みた。

これにより，記事原文から自動的に二次情報を作成し，検索，分析に供するとともに翻訳機能ともリンクした基礎解析機能の実現が期待できる。（2.3，2.4.1を参照）

2.2.6　二次情報データベースの作成実験（５９年度）

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。

① 検索のためには．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。

② 内容分析のためには，文章レベルあるいは文レベルでの意味内容の解析を可能とする情報構造が要求される。

③ 統計的分析手法による内容分析（定量化分析）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。

④ 翻訳のためには，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。

二次情報データベースの蓄積構造は，2.3と2.4.2で詳述するとして，二次情報として生成される情報のうち，ここでは特に事実情報と機械翻訳用中間言語につき，基本的な考え方を示す。

事実情報とは，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，意味内容に立入っての高度な検索・分析に対応できる有効な方法を採用すべきである。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。

機械翻訳用中間言語とは，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二次情報である。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。

2.2.7 知識ベースの研究

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とそこで満たすべき機能に関して述べる。

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。

しかし，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。そこで独自性の確保には，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

こととした。

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。

知識ベースは，辞書，規則，知識表現の３つの要素に分解できる。但し，そ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。

① 辞書は，言語別，内部操作の処理フェーズ別，目的別の複数種必要となる。その構造は，語彙別の構文（品詞，文脈等）情報，意味（内在素性，選択制限，意味標記等）情報，と辞書ファイルとしてのヘッダ情報，辞書規則などの共通情報に分かれる。また登録範囲が一般語か専門語か，活用，派生，熟語，合成語，同形異義，異綴語等の形態の扱いにより，辞書内の項目，用法が異なってくる。この他文法規則とは厳密に区別した辞書用の規則も必要である。

② 規則には，辞書を補うもの，アルゴリズム（プログラム）を補うもの，

構文構造に関するもの，意味に関するもの，語用論的なものがあり，辞書を補う規則は辞書に吸収されるものとする。こうした規則をプログラムから分離独立して表現する手法は定まっているとは言えず，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と考える。

③ 知識表現は，規則と辞書の語彙情報だけでは促えられない情報内容を処理する際に用いられる別の形で表現された「知識」である。

語の間の構造を表現した体系であるシソーラス，事実データベース，分析モデルベース，一般知識が考えられる。

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，体制とツール及びマン・マシン・インターフェイスの設定が重要なキーポイントに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。

2.2.8 エネルギーデータベースの構築（５７～５９年度）

海外で発行されている雑誌等の主要５０紙誌を情報源として，第２次石油危機以後の１９７９年１月から１９８４年１２月までの６ヵ年分を対象に，

エネルギー関連の記事情報を抽出してインデックスを作成し，更にキーワード付けを行った。

データ整備の方法は，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分析に利用するための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し，磁気テープ化した。

なお，データベースに登録したデータ件数は，総計３６,９０８件であり，年別の内訳は次のとおりである。

年	件　数
７９年	9,９０３
８０年	8,６５４
８１年	7,９６３
８２年	6,０４２
８３年	3,５４６
８４年	８００
計	３６,９０８

また，作成した記事インデックスの例は，データベースの検索出力例に示すとおりである。

出力例１　記事インデックス検索結果

期　間：８２年４月～８３年３月
キーワード：サウジアラビア *価格政策F *原油　全34件

記事インデックス	年月日	紙誌名	紙面・ページ	行数	記事番号
サウジのアブドゥラ皇太子、ＯＰＥＣ加盟国に協定遵守を要請	830323	FT	008	81	8303234009
サウジアラビア、アラビアン・ライトは2月1日実施で３０ドル	830317	LT	001	68	8303174011
サウジアラビアとクウェート、原油値下げで大幅財政赤字	830304	LT	015	112	8303044010
湾岸産油国、４ドルの値下げを認めないなら７ドルの値下げも辞せず	830228	LT	017	90	8302284007
湾岸５大産油国、２月２３日に原油価格引き下げで合意	830224	LT	001	60	8302244015
湾岸５大産油国、価格引き下げで合意――ＯＰＥＣ会議で提案か	830224	IHT	001	115	8302244012
湾岸諸国石油相、リヤドで会合し原油値下げで合意――値下げ幅未定	830224	WJ	003	141	8302244003
湾岸４大産油国、バーレル当たり４ドル以上の値下げを検討	830223	IHT	001	104	8302234005
湾岸４大産油国石油相、少なくともバーレル当たり４ドルの値下げを検討	830223	WJ	003	124	8302234004
サウジのヤマニ石油相、８３年２月１０日基準原油価格の維持を断念	830211	FT	001	122	8302114006
サウジのヤマニ石油相、原油価格の引き下げ不可避を公式に認める	830211	LT	019	98	8302114001
サウジとナイジェリア、原油値下げで合意か――各産油国待機戦術	830207	PN	001		8302074012
中東湾岸産油国の原油引き取り範囲が狭まる	830117	PIW	002		8301174008
ＯＰＥＣ原油価格、野放し状態へ――国別生産割当に失敗	830110	BW	017		8301104005
サウジのヤマニ石油相、値下げについてジュネーブでアラムコ代表と会談	830105	FT	014	48	8301054005
アラムコ代表とヤマニ・サウジ石油相、価格・生産量について合意に達せず	830105	WJ	002	38	8301054004
アラムコ、ヤマニ・サウジ石油相に原油値下げを要請	830105	LT	013	51	8301054003
サウジが値下げか減産を行わねば、原油引取量を削減とアラムコが警告	830104	AWJ	002	60	8301044006
米国石油メジャー、サウジ原油の価格問題でヤマニ石油相と会談予定	821231	FT	001	96	8212314001
サウジのヤマニ石油相、ＯＰＥＣマーカー原油価格３４ドルを守ると明言	821207	WJ	029	46	8212074004
サウジ国王、アルジェリア訪問――価格ディファレンシャルの協力で合意	821201	APS	091		8212014003
ロイヤル・ダッチ・シェル、原油価格値下げの影響を評価――アナリストに語る	821124	PN	001		8211244001
ヤマニ、原油価格維持を再確認――インドネシアは値下げを辞せず	821123	FT	015	236	8211234006
ヤマニ、原油スポット価格と為替レート考慮するが、具体的な対策記述せず	821122	IHT	009	30	8211224028
サウジ、ＬＰＧ価格決定で柔軟姿勢――原油価格連結方針にも幅広い選択	820927	PIW	003		8209274011
サウジ石油相、アフリカ原油に合わすならＯＰＥＣ基準価格も値下げと警告	820921	FT	004	40	8209214012
ＯＰＥＣ専門委、サウジのアフリカ原油値上げ要求を支持	820906	IHT	013	79	8209064021
サウジの原油価格連動ＬＰＧ価格、８３年には湾岸協力会議も基準として採用	820818	APS	001		8208184014
サウジ、原油価格引き下げよりも販売政策の変更考慮――アラムコ外に開放	820816	MEES	001		8208164021
サウジ、原油増産よりも値下げ検討――増産のうわさは根拠なし	820816	PIW	001		8208164020
イラン勢力復活でＯＰＥＣ内部に異変――強硬派の地位優勢に原因	820721	WJ	033	230	8207214013
ＯＰＥＣ、サウジのアフリカ原油値上げ強硬主張めぐり内部分裂	820710	FT	026	164	8207104002
サウジ、アフリカ原油のディファレンシャルの引き上げを要望	820705	MEES	001		8207054012
サウジ、バーレル３４ドルの基準価格固守の構え――減産強化も辞さず	820505	APS	001		8205054017

出力例２　記事インデックス検索結果（全キーワード付き）

期　間：８２年４月～８３年３月
キーワード：サウジアラビア *価格政策F *原油
全34件中の12件を記載

出力形式：記事インデックス　年月日　紙誌名　紙面・ページ　行数　記事番号　　全キーワード

サウジのアブドゥラ皇太子、ＯＰＥＣ加盟国に協定遵守を要請　830323　PT　006　81　8303234009
（海外団体）　サウジアラビア政府　ＯＰＥＣ　（品目）　原油　（項目）　国際協定Ｆ　エネルギー政策Ｆ　価格政策Ｆ　生産計画Ｆ　石油会議　産油国
（海外地域）　サウジアラビア　（紙誌名略号）　ＦＴ

サウジアラビア、アラビアン・ライトは２月１日実績で３０ドル　830317　LT　001　68　8303174011
（海外団体）　サウジアラビア政府　（品目）　原油　（項目）　値下げＦ　価格政策Ｆ　エネルギー政策Ｆ　（海外地域）　サウジアラビア　（紙誌名略号）　ＬＴ

サウジアラビアとクウェート、原油値下げで大幅財政赤字　830304　LT　015　112　8303044010
（海外団体）　クウェート政府　サウジアラビア政府　（品目）　原油　（項目）　オイルマネー　値下げＦ　価格政策Ｆ　財政Ｆ　（海外地域）　サウジアラビア
（海外地域）　クウェート　（紙誌名略号）　ＬＴ

湾岸産油国、４ドルの値下げを認めないなら７ドルの値下げも辞せず　830228　LT　017　90　8302284007
（海外団体）　クウェート政府　サウジアラビア政府　カタール政府　アラブ首長国連邦政府　（品目）　原油　（項目）　産油国　価格政策Ｆ　エネルギー政策Ｆ　値下げＦ
（海外地域）　クウェート　サウジアラビア　カタール　アラブ首長国連邦　（トピック）　ペルシャ湾　（特別人名）　中東産油国　（紙誌名略号）　ＬＴ

湾岸５大産油国、２月２３日に原油価格引き下げで合意　830224　LT　001　60　8302244015
（海外団体）　イラク政府　クウェート政府　サウジアラビア政府　カタール政府　アラブ首長国連邦政府　（品目）　原油　（項目）　エネルギー政策Ｆ　価格政策Ｆ
（項目）　値下げＦ　産油国　（海外地域）　クウェート　カタール　アラブ首長国連邦　イラク　サウジアラビア　（トピック）　ペルシャ湾　（特別人名）　中東産油国
（紙誌名略号）　ＬＴ

湾岸５大産油国、価格引き下げで合意――ＯＰＥＣ会議で提案か　830224　IHT　001　115　8302244012
（海外団体）　クウェート政府　イラク政府　アラブ首長国連邦政府　カタール政府　サウジアラビア政府　（品目）　原油　（項目）　産油国　値下げＦ　石油会議
（項目）　エネルギー政策Ｆ　価格政策Ｆ　（海外地域）　サウジアラビア　イラク　アラブ首長国連邦　カタール　クウェート　（トピック）　ペルシャ湾
（特別人名）　中東産油国　（紙誌名略号）　ＩＨＴ

湾岸諸国石油相、リヤドで会合し原油値下げで合意――値下げ幅未定　830224　WJ　003　141　8302244003
（海外団体）　サウジアラビア政府　カタール政府　アラブ首長国連邦政府　イラク政府　クウェート政府　（品目）　原油　（項目）　産油国　値下げＦ　石油会議
（項目）　エネルギー政策Ｆ　価格政策Ｆ　（海外地域）　サウジアラビア　カタール　アラブ首長国連邦　イラク　クウェート　（トピック）　ペルシャ湾
（特別人名）　中東産油国　（紙誌名略号）　ＷＪ

湾岸４大産油国、バーレル当たり４ドル以上の値下げを検討　830223　IHT　001　104　8302234005
（海外団体）　クウェート政府　アラブ首長国連邦政府　カタール政府　サウジアラビア政府　（品目）　原油　（項目）　産油国　値下げＦ　価格政策Ｆ　エネルギー政策Ｆ
（海外地域）　アラブ首長国連邦　カタール　サウジアラビア　クウェート　（トピック）　ペルシャ湾　（特別人名）　中東産油国　（紙誌名略号）　ＩＨＴ

湾岸４大産油国石油相、少なくともバーレル当たり４ドルの値下げを検討　830223　WJ　003　124　8302234004
（海外団体）　クウェート政府　サウジアラビア政府　カタール政府　アラブ首長国連邦政府　（品目）　原油　（項目）　産油国　値下げＦ　エネルギー政策Ｆ　価格政策Ｆ
（海外地域）　サウジアラビア　カタール　アラブ首長国連邦　クウェート　（トピック）　ペルシャ湾　（特別人名）　中東産油国　（紙誌名略号）　ＷＪ

サウジのヤマニ石油相、８３年２月１０日基準原油価格の維持を断念　830211　FT　001　122　8302114006
（海外団体）　サウジアラビア政府　（品目）　原油　（項目）　価格政策Ｆ　エネルギー政策Ｆ　値下げＦ　（海外地域）　サウジアラビア　（紙誌名略号）　ＦＴ

サウジのヤマニ石油相、原油価格の引き下げ不可避を公式に認める　830211　LT　019　98　8302114001
（海外団体）　サウジアラビア政府　（品目）　原油　（項目）　価格政策Ｆ　エネルギー問題Ｆ　値下げＦ　（海外地域）　サウジアラビア　（紙誌名略号）　ＬＴ

サウジとナイジェリア、原油値下げで合意か――各産油国持続政府　830207　PN　001　8302074012
（海外団体）　ナイジェリア政府　サウジアラビア政府　（品目）　原油　（項目）　産油国　価格政策Ｆ　二国間会議Ｆ　値下げＦ　（海外地域）　ナイジェリア
（海外地域）　サウジアラビア　（紙誌名略号）　ＰＮ

出力例３　キーワード出現頻度検索結果

期　　間：８２年４月～８３年３月
キーワード：サウジアラビア＊価格政策Ｆ＊原油
全３１８種中の７６件を記載

NO.	キーワード	頻度数	NO.	キーワード	頻度数	NO.	キーワード	頻度数	NO.	キーワード	頻度数
1.	サウジアラビア	273	2.	サウジアラビア政府	169	3.	エネルギー政策Ｆ	109	4.	石油	95
5.	原油	85	6.	産油国	77	7.	価格政策Ｆ	68	8.	ＯＰＥＣ	51
9.	ＦＴ	49	10.	ＷＪ	36	11.	ＭＥＥＳ	35	12.	ペトロミン	32
13.	生産計画Ｆ	32	14.	値下げＦ	32	15.	米国	32	16.	ＰＩＷ	28
17.	ヤマニ	27	18.	ＡＰＳ	26	19.	アラムコ	23	20.	価格動向Ｆ	23
21.	生産動向Ｆ	23	22.	石油会議	22	23.	ＩＨＴ	21	24.	ＬＰＧ	19
25.	エネルギー開発Ｆ	19	26.	ＬＴ	18	27.	イラン	18	28.	クウェート	17
29.	輸出Ｆ	17	30.	イラク	16	31.	対外交渉Ｆ	15	32.	メジャー	13
33.	英国	13	34.	輸入Ｆ	13	35.	ＡＯＧ	12	36.	ＰＮ	12
37.	オイルマネー	12	38.	フランス	12	39.	需給動向Ｆ	12	40.	製油所	12
41.	エネルギー問題Ｆ	11	42.	ナイジェリア	11	43.	ファハド	11	44.	中近東	11
45.	アラブ首長国連邦	10	46.	クウェート政府	10	47.	価格交渉Ｆ	10	48.	石油化学	10
49.	フランス政府	9	50.	販売戦略Ｆ	9	51.	イラク政府	8	52.	カタール	8
53.	カタール政府	8	54.	ペルシャ湾	8	55.	企業業績Ｆ	8	56.	工場建設Ｆ	8
57.	中東産油国	8	58.	ＡＷＪ	7	59.	アラブ首長国連邦政府	7	60.	減産Ｆ	7
61.	合弁開発Ｆ	7	62.	財政Ｆ	7	63.	借款Ｆ	7	64.	商談Ｆ	7
65.	西独	7	66.	中東和平	7	67.	日本	7	68.	物流機整備Ｆ	7
69.	予算Ｆ	7	70.	エタノン	6	71.	サウジアラビア基礎産業公社	6	72.	ダウ・ケミカル	6
73.	軽質油	6	74.	石油業界Ｆ	6	75.	戦争Ｆ	6	76.	販売動向Ｆ	6

出力例４　キーワード月別頻度検索結果

検索キーワード：　エネルギー外交F

7901	7902	7903	7904	7905	7906	7907	7908	7909	7910
24	34	48	31	22	34	21	27	30	47
7911	7912	8001	8002	8003	8004	8005	8006	8007	8008
30	31	31	23	47	36	27	27	24	17
8009	8010	8011	8012						
11	25	21	23						

検索キーワード：　原油

7901	7902	7903	7904	7905	7906	7907	7908	7909	7910
197	225	286	181	174	202	154	104	99	176
7911	7912	8001	8002	8003	8004	8005	8006	8007	8008
131	211	159	147	138	126	146	124	104	99
8009	8010	8011	8012						
103	78	78	101						

出力例５　キーワード国別頻度検索結果

検索キーワード：　産業政策F

米国	英国	日本	西独	フランス	スウェーデン	イタリア	ベルギー	オランダ	スペイン
41	67	1	13	27	0	2	2	2	0

検索キーワード：　石油

米国	英国	日本	西独	フランス	スウェーデン	イタリア	ベルギー	オランダ	スペイン
1995	647	83	263	183	0	60	11	50	14

2.3　総合解析システムの実験

2.3.1　目的と範囲

今回の実験は文章情報総合解析システムとして，次の様な機能の実現を試みるものである。

① 多様な検索・分析ニーズに対応できる柔軟な文章情報の構造表現を設定する。

② その様な，文章情報の構造表現を機械的に作成する。

これまでの本委員会において総合解析システムにかかる研究，実験，開発を行ってきた中で，データベース作成・検索機能，定量化分析機能，翻訳機能，基礎解析機能を統合する要素として，基礎解析機能における二次情報データベース作成機能の充実の必要性がより明らかになってきた。そこで，今回，文章解析機能の拡張により，定量化分析，検索機能との結合をめざした，記事情報よりの事実表現抽出の実験を行った。

どれほど簡単な文章であっても，それが認識され，他の行為，他の認識に供される為には，多くの知識を援用し，複雑な適用メカニズムが起動される。また，機械処理により，このようなことの一部でも行うためには，その文章に対する内部表現とその利用の基本モデルが完全に確定することが必須である。しかし，一方で，文章情報の利用に関する論理的な枠組は，殆んど明示されていないというのが現状であろう。そこには，人間の"勘"といったレベルの論理化しえない経験的メカニズムが用いられており，個別に提起される論理的な方法論にしても，コンピュータ上で実現しようとして，その覆いを取り払っていくと，どうしてもアドホックなメカニズムになり，汎用化にも実用化にも問題を残すことになってしまう。これは，いかに明解に見えても，現実の文章情報の利用のメカニズムが本質的に論理化しえない要素から成り立っているという背景があるのであろう。そこで，そのような論理化しえない部分を含むエレメントを橋渡しする仮想のモデルをあらゆるレベルで設定することが必要となってくる。そうすると，それらのモデルは，どれだけそこでの目的に適うかが唯一の評価の基準となる。その内部が問われるのは，

そこに含まれる下位のモデルに対し，同様の評価がなされる場合だけになる。

今回の実験では人間が機械的に文章を処理したときに作るであろう，そして機械的な処理で用いたいであろう構造を第1近似として設定し，それを実現するメカニズムを考察した。結論としては，多少の出入りはあるにしても，アドホックな扱いを前提とすれば実現は常に可能であるが，逆にそのアドホック性を避けることは，現段階では殆んど不可能であろうということである。問題はアドホックであることのコストを容認可能なレベルまで引き下げることであり，そのためにも，"現場における"このような処理の経験の蓄積が必要となる。勿論，そうした実用システムでは，一般的モデルのそのままの適用は何ら益をもたらさないが，その間を結ぶ，想像力，連想力，構想力にあたるリンクを常に残しておくことは肝要であろう。次節以下では，そのためのプロトタイプとしての意義を果すよう，データの分析，モデルの設定，実験を行い，その経過と結果を提示する。

今回の実験は，基礎解析機能の中で，特にこれまで余り明らかになっていない，文章の表現している内容，事実を利用サイドの為に抽象するフェーズを扱う。そのような観点で総合解析システムを図示すると，図2.3－1のようになる。中央の「事実抽出」のモジュールを中心としての実験を行った。ここで，図のデータの流れは次の様になる。

① 原テキストが，一定のレベルまでの解析処理を経て，比較的ユニバーサルな解析構造（ここでは構文構造）に変換される。ここでは，形態素解析，構文解析までを必須とし，意味解析，文脈解析は仮定していない。

② 解析された構造は，一方で，翻訳処理に渡され，ターゲット言語生成が行われる。ここでは，解析として統語論レベルまでしか仮定していないので，シンメトリカルには，構文生成，形態素生成を行い，かつ相当にアドホックにはなるが，意味処理が必要に応じてちりばめられるであろう。

③ 解析された構造に対して，内容の検索，内容を用いた分析の為に，事実抽出のフェーズが設けられる。ここで，解析処理としては，意味処理

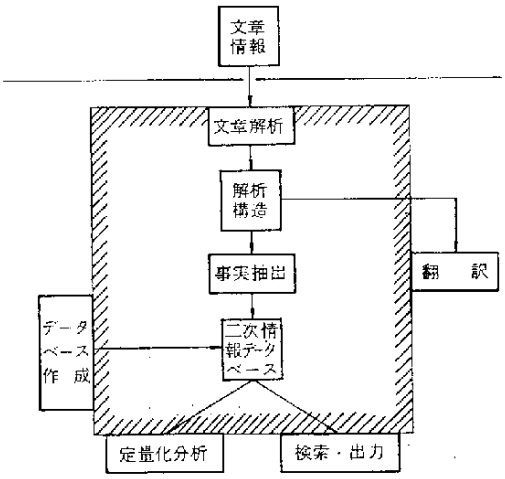


図 2.3－1　総合解析システムにおける実験の範囲

以降を一体とした要約，抽象化の処理が行われ，二次情報データベースが作成される。

④　二次情報データベースに対し，内容に踏みこんだ検索・分析までが行われる。

以上の流れを過程して，事実抽出のプログラム（この中には，知識ベースも一体となっている）の開発を行ったが，それを中心として，今回の実験に関して，次節以下で，サンプルテキストの分析，解析処理に関する分析，事実表現の説明，事実抽出メカニズムの説明及び二次情報データベースとその利用に関して述べることとする。

2.3.2　サンプルテキストの分析

本実験では，エネルギー関連の英文記事データに対し，その叙述する「事実」を，適切かつ簡素に中間表現に変換し，有効利用の可能性を検討することを目的とした。その為にいくつかの仮定を前提とした。

まず，各記事を，本文のはじめの１文で代表させることにした。これは，１つには新聞記事等の場合，できる限り文頭で簡潔に事実を述べ，次に，それを補足する形を繰り返して詳細におとす，という形態をほぼ原則としていることを根拠としている。実際に，この実験を行うにあたって分析した２種類４４記事のデータに関して考察すると，解説記事の様な少数のものを除いては，専門家による記事内容インデックスに表現された情報が，先頭文に，何らかの形では見られることが判った。

先頭文だけを今回の対象としたもう１つの理由は，主に，技術的な制約に

よる。今回の実験は，自然言語処理技術を，文章情報の総合利用の為に適用することを一つのテーマとしている。文章情報の高度な総合利用技術を実現しつつ，基礎解析機能の自動化を図る場合，自然言語処理技術の活用は，必須の要件となってくる。しかし，自然言語処理技術を適用する為には，文字列処理に限った場合に比べ，不連続な一定の高さの障壁を越えることが，自然言語処理技術及びその適用・利用技術の双方で要求される。その１つの要件は，記事や記事の集合を対象としている場合においても，それを構成する各文が，コンピュータ・ロジックとしては，何らかの形で"完全に"解析されていなければならないということである。これは，統語論的な解析や意味解析という用語で表現される形式で，研究開発されている技術であるが，一部実用化されている一方で，未だ解決すべき問題も残されているのが現状である。そして，この様にして解析された文の表現を，いかにして文の有機的な並びとしての記事，あるいは記事集合全体の表現にまとめあげるかということは，また別の技術課題である。文に対する解析という側面からも，こうした文脈処理の意識が必要であるということは，かねてより主張されており，その為の研究も進んでいる。また，テキストの全体的把握という観点からも，ヒューリスティックなアプローチを中心として，いくつかの試行がなされている。しかし，本実験では，文脈の意識は，それらの研究開発成果によることとして，具体的には，文を単位とした処理によって記事の内容表現にかえることとした。

以上のことから，ここでは，先頭文を対象とした基礎解析処理によって記事全体の内容表現にあたるものを作ることを試みた。

次に，総合解析システムとしては，不特定かつ大量の文章データを扱うことを目指しているが，ここでは，分野・範囲を限定し，少量の記事を対象とした。基礎解析の基本アルゴリズムの確立の為には，サンプル・データを用いた高度かつ詳細な検討・分析と試行の繰り返しが必要であり，それを総合的に行うためには，大量のデータから出発することは困難であるという事情があった。しかし，一方，それは大量データに対する総合解析システムの利

用の段階を見極め，示唆するに十分な普遍性をもつ必要がある。そこで，少量データへの絞り込みが無機的なものであってはならないことになる。すなわち，少数の例外を含みつつも，全体で何らかの利用側から見たシナリオの中で閉じており，逆にそれらのシナリオから見ても，ほぼ完結していると見なすことができるという範囲を選択するということである。

ここでは，あらかじめ2種類のデータを選択し，その検討を行ったのち，その一方について実験を行った。当初選ばれた記事データは，次の様なものである：

① 海外の主要５０紙誌から選ばれたエネルギー記事データの中で，「米国」，「石炭」，「エネルギー政策」のキーワードを有するもの。このうちとくに，８2年7月末から８3年6月上旬までの約１年分の１０記事。

② 同上データから，「サウジアラビア」，「原油」，「価格政策」を有する８２年４月から８３年３月までの１年分３４記事。

これらの記事は，５８年度本委員会において開発されたデータベース検索システムを用いて検索されたものである。

プロトタイプ・システムの構築にあたっては，②を具体的なサンプル・データとした。それらの記事の先頭文，専門家により付与された記事インデックス及び機械処理という観点での概説を行ったものが資料1,2である。

(1) 米国，石炭，エネルギー政策に関する記事（資料１参照）

全体として，記事インデックスと見比べると，先頭文だけで記事全体をほとんど表現していることがわかる。

ここで選ばれたものの中で7月３０日～８月１１日の間の４記事は，同一のトピック（石炭スラリー法案の採択に関する記事）を扱っている。逆に，ここでの記事は，キーワード検索によって得られたものであるので，８２年８月３１日付の記事に見るように，ポーランドの話題であって，米国は単にその競合国の一つという意味で表わされているものもある。逆に言えば「米国」の，「石炭」に関する，「エネルギー政策」といった検索ができないか，というのが今回の実験の重要なテーマとなっている。

(2) サウジアラビア，原油，価格政策に関する記事（資料２参照）

ここに見られる記事を通して，この期間の動きは，大筋として，次の様な内容になっている。すなわちサウジアラビアを中心とする原油価格がバーレル当り３４ドルで固定し，種々の憶測，アフリカ原油やＬＰＧと関連する問題などにもかかわらず，価格維持が表明されていた時期から値下げへの動き，値下げの協定，そして協定に従った値下げを要求するまでにいたったという流れである。したがって，このサンプル・テキストについては，このような大局的なシナリオを背景として分析し，処理系を試作することが有意義である。

しかし一方で，これらの記事の場合，解説記事のように先頭文が，記事全体を代表しえないものなどもいくつか見られる。

2.3.3 文の解析

本実験では，原テキストから二次情報データベースを作成するまでの過程を２段階に分け，前段階において構文の解析を行って原文を構文構造に変換した後，その構文構造からテキストの（今回の実験では文の）内容表現を作成することとした。このことより，文の解析のために発生する諸問題から，内容表現作成フェーズの分析を解放することが出来た。ここで仮定した構造解析は，統語論的な範囲に限定し，文脈自由型文法の形式の構文構造を設定した。このようなレベルの構文解析プログラムとしては，既に多数の方式のものが開発されており，本委員会においても，５８年度Prolog言語を用いた文章解析実験の中で開発を行っている。

本実験で用いたサウジアラビア，原油，価格政策の３４記事の構文構造の一部を図2.3－2に示す。なお，１行文字数制約の関係から，インデンテェイションが乱れている部分がある。ここでは，構文構造をProlog 言語のためリスト形式で表現した。各ブラケットが木構造のノードに当るものとし，終端記号（木構造のリーフ，ノード）と非終端記号の別で次の規約により読むことができる。

① 終端記号

第1要素が「noun」，「vi」，「prep」といったシンタクティック・カテゴリーにあたるリーフ・ノードのラベル。

第2要素はそれに対応する文中の語。これは英語原形で表現する。解析用辞書は，このような語形をキーとする。屈折形の処理は，ロジックとして行ってもよいし，あらかじめ屈折形で辞書を展開してもよい。

第3要素以降は任意情報でここでは内容表現の作成の処理に必要な情報などを格納するものとする。

（例1）

〔noun，＼Saudi Arabia／，organ〕

シンタクティック・カテゴリーがnounである語Saudi Arabiaで，辞書にこの語が『組織』（＼organ／）である旨の情報が与えられている。

（例2）

〔vt，warn，past〕

他動詞warnで，構文解析の結果，この動詞が過去形で使われていることが判り，＼past／を補っている。なお，構文解析の負担を軽減するためには，たとえばこのwarnという語であれば，この語が＼SVOO／型の文型をとりえて，しかも間接目的語として名詞節をとりうるといった情報も辞書中に記述しておく必要があろう。そして，これと連係して，その間接目的の名詞節中の if 型の副詞節の内容を否定する内容を，主語にあたる主体が望み，要求しているといった内容の抽象化を後のフェーズが行うことができる。このような情報は解析用辞書と連係しており，統合的に「知識ベース」に収納することが望ましいが，一方，どちらかといえば普遍性を希求する辞書記述中に，多少とも特殊的なこのような記述を係記することは，好ましくない面も多かろう。このことは，前例での＼Saudi Arabia／に与えた＼organ／というカテゴリーについても程度の差こそあれ，通用する問題であり，この意味で，辞書データベースの構成，内容に関しては更なる検討が必要である。

```
['1-8303234009FT',
[s, [np, [np, [adj, 'Crown Prince', occup],
              [noun, 'Abdullah', name]],
         [pp, [prep, of],
              [np, [noun, 'Saudi Arabia', organ]]]],
    [vp, [vt, warn, past],
         [np, [np, [adj, other],
                   [noun, member, plu]],
              [pp, [prep, of],
    [np, [noun,  'the Organization of Petroleum Exporting Countries (opec)',
                        organ]]]],
         [nouncl,
             [intro, that],
             [s, [np, [det, the],
                      [noun, kingdom, organ]],
                 [vp, [verb, be, past],
                      [adv, perfectly],
                      [adj, capable],
                      [pp, [prep, of],
                           [np, [vp, [vt, win, gerund],
                                          [np, [det, an],
                                               [adj, oil],
                                               [noun, price],
                                               [noun, war]]]]]],
                 [advcl,
                      [intro, if],
                      [s, [np, [prn, 'they']],
                          [vp, [vt, fail, present],
                               [toinf, [intro, to],
                                       [vp, [vt, observe, inf],
         [np, [np, [det, the],
                   [noun, pact]],
              [pp, [vt, conclud, ptp],
                   [adv, 'last week', time],
                   [pp, [prep, in],
                        [np, [noun, 'London', place]]]]]]]]]]]]]]]].


['5-8302244015LT',
[s, [np, [np, [noun, 'Saudi Arabia', organ]],
         [conj, and],
         [np, [adj, four],
              [adj, other],
              [noun, 'Gulf states', organ]]],
    [vp, [vp, [vi, agree, past],
              [adv, yesterday, time],
              [toinf, [intro, to],
                      [vp, [vp, [vt, cut, inf],
                                [np, [adj, their],
                                     [adj, oil],
                                     [noun, prices]]],
                                [pp, [prep, by],
                                     [adj, 'at least'],
                                     [adj, '10'],
                                     [noun, 'per cent']]]]],
         [conj, and],
         [vp, [vi, threaten, past],
              [toinf, [intro, to],
                      [vp, [vt, make, inf],
                           [np, [adv, even],
                                [adj, bigger],
                                [noun, reduction, plu]]]]],
         [advcl, [intro, unless],
                 [s, [np, [np, [adj, other],
                               [noun, member, plu]],
                          [pp, [prep, of],
    [np, [noun, 'the Organization of Petroleum Exporting Countries']]]],
                      [vp, [vp, [vi, fall, present],
                                [pp, [prep, into],
                                     [np, [noun, line]]]],
                           [pp, [prep, with],
                                [np, [det, their],
                                     [noun, plan]]],
                           [toinf, [intro, to],
                                   [vp, [vt, avert, inf],
                                        [np, [det, an],
                                             [adj, international],
                                             [adj, oil],
                                             [adj, price],
                                             [noun, war]]]]]]]]]].
```

図 2.3 − 2　構文構造の例

（例３）

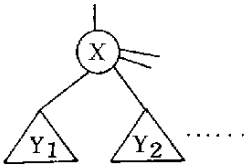
〔 adj, ＼Saudi Arabia／, poss, organ 〕

第１の例に加えて，所有格の 's のために，＼poss／が付加されている。ここで第１要素がnoun でなくadjとなっているのは，構文構造のタイプの単純化のためであり，このことについては後述する。

② 非終端記号

第１要素がそのノードの非終端記号を表わし，第２要素以降が，その「子」の構造を表わす。

〔 X_1, Y_1, Y_2, …… 〕



（例１）

〔 s, 〔 np, …… 〕, 〔 vp, …… 〕 〕

主語の名詞句と動詞句とから文が成っているという文脈自由型規則に対応している。

（例２）

〔 '1-8303044010LT／, 〔 s, …… 〕, 〔 s , …… 〕 〕

テキスト（本実験についてはその先頭の文）に対して構造のルートが対応し，記事の情報と文構造とから構成される。記事の情報は，連番と，ハイフンに引き続く１０桁の記事番号，紙誌名コードからなる。連番は対象テキストの最新分からふってある。また，記事番号の先頭６桁はその記事の発行日である。

（例３）

〔 np, 〔det, an 〕,
〔adj, international 〕,
〔adj, oil〕,
〔adj, price 〕,
〔noun, war 〕〕

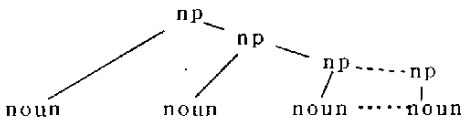
名詞連続においては，主名詞以外は＼adj／を与えている。なお，このような名詞，形容詞の並びの中の構造については，特に構文解析で扱っておくことは期待していない。

このような不定個の並びを子にもつような構造は，構造の見通しの悪さを避ける為に必要である。たとえば，次の様な規則を考える。

np → noun

np → noun・np

このときの名詞連の構造は次の様になって，上位ノードからの見通しは極めて悪くなる。



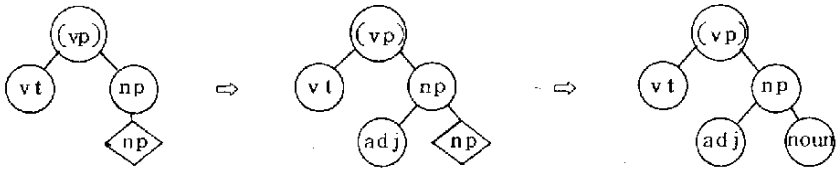
これを避けるためには，次の様な規則が必要である。

np → (noun)$^{+}$

ここで＋印付きの括弧は１回以上の繰り返しである。この実現には規則として，想定できる繰り返し個数までの展開した規則を書き下すか，構文解析プログラムで上記記法をメカニズムとして取り込めばよい。

一方，構文解析の過程の木構造表現（以下パージング・トリーと呼ぶ）が構文解析結果（ここでは，統語論的な処理を主としていることもあり，以下，構文構造と呼ぶ）と同一であると仮定する理由は何もない。そこで，パージング・ルールにくくりつけられた補強メカニズム・ルールで構文構造を作成していくといった類のアルゴリズムにすれば，上記の問題に関しても対応することができる。たとえば，np → noun・npというパージング・ルールに対しては，「adjという終端ノードと np に対応する仮ノードを並列に，現行仮ノードに置換しなさい」，np → noun には「noun で置換しなさい」というルールでよい。それでは，非終端ノード npはどこで生まれるかと言うと，上位ルール，たとえば，vp → vt・np のルールで，npノードと，その下の仮ノードが作られていればよい。

これらの構文解析の実働システムへのインストールに際しては，いくつかの考慮すべき課題がある。短い基本的パターンの文に対しては，ど



のような解析アルゴリズムをとっても，誤った解析の出る確率は相対的に少ないし，処理時間は必要な記憶容量も，殆んど問題にならないであろう。しかし，実データは，ここで扱った文にも見られるように，３０，４０語を越えるものも普通にあり，文型としても種々のルールを要求する各種パターンがあり，更に，名詞連，and等の接続構造，挿入，副詞節，形容詞節，前置詞句等々構文的曖昧度（アンビギュイティ）を生む多数の要因を含んでいる。こうした事情から，たとえば，忠実な構文解析をバックトラックを行っていった時は，簡単に処理時間の爆発に直面する。このことは，次のことからも明らかであろう。

簡単にする為に，X→X・Xというルールだけを考え，Xのn個の列を解析してみよう。すると，$\frac{1}{n}\binom{2n-2}{n-1}$ 個の解析方法があることがすぐに判る。また，記号例のn個の部分で局所的な解析方法がK個ずつあってもK^{n}個の解析方法があることも判る。いずれにしても，指数オーダーのアンビギュイティを生む。ここで問題なのは，単に文全体のアンビギュイティだけではなく，陽の目を見ることなく埋もれてしまうべき，中間的な試行に関しても，同様な事情があることである。この結果，処理時間への爆発への対応は，入力記号列の長さを小さく止められる場合や，ルールとして排他的なものだけで十分であるとか正規型規則で十分であるといった場合のような特別なケースを除けば，必須の考慮である。

この問題の解決のために，種々のテクニックが使われている。たとえば，ルール記述（あるいは，その枠組）での考慮である。上で例として出したX→X・Xといったルールを実際に記述する人はいないであろう。また，結果に反映しない（失敗する）解析過程を，できる限り早く判定し，それを行わない，あるいは中断することによって，無意味な処理をで

きる限り除去する方法も種々開発されている。更に，バックトラックを行わない方法は広く行われている。先のX→X・Xのルールの場合でも過程を記憶していけば，最小限記憶すべき情報は$\frac{n(n+1)}{2}$個にすぎず，パージング・トリーを記憶する為にも，また，処理ステップとしてもn^3のオーダーで十分なことは，簡単なアルゴリズムで確かめられる。多くの構文解析プログラムは，必要であれば入力文の特性等も考慮して，上記の各要素を組み合せて作成してある。

構文解析のプログラム実現のためのもう1つの重大な障壁は，自然文のもつ多数の解析方法（アンビギュイティ）の中から正しい解を得る方法である。これには，語レベルでのセマンティック・マーカーから頻度情報までの種々のレベルの情報，規則に関する情報などに基づいた局所的評価，大局的評価などが用いられている。しかし，これはよりヒューリスティックなアプローチが必要なテーマであり，その実現のために必要な辞書や規則の準備などのための負担との相互関係でも評価すべきものであろう。

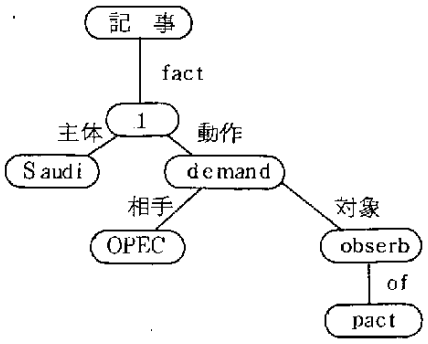
2.3.4 事実表現

事実表現とは，本システムの二次情報データベース過程構造である。文章情報の内容表現あるいは事実表現といわれるものは，文の構文構造の場合に比較して，いわく有効性をもち，支持されている枠組は，まだ確立されているとはいえない。まして，実働システムにただちに採用できる状態で準備されている体系は存在していない。したがって，一定の観点（すなわち，業務の目的による要求定義）の下で必要十分な構造であり，かつ準備しうるリソースにより作成可能なものを個々に検討し，実現しているという努力が必要であろう。そのことは，例えば，「数値データを利用するためのデータベース」とだけ言っても，具体的なデータ構造やシステム構成は規定することはできず，「どのような数値データを，どのように入手するのか」，「そこで扱う数値データ自体のもつ構造はどのようなものか」「そこでのデータ利用のモ

デルはどのようなものか」等々の前提があって，はじめて具体化の道筋を設定できるということからも，十分に示唆されるであろう。ただし，文章情報というとき，定形化・定量化されていない情報を通常は意味し，コンピュータが出現するはるか以前から，長い歴史の中で確立されてきたモデル化・抽象化の方法論により定量化，定形化された結果の数値データ等を扱う場合とは本質的に異なった問題を有していることには留意しておく必要がある。そこで，文章情報に対するデータ構造は，むしろ余儀なくではあるが，柔軟な構造を設定する余地があるとも考えることができる。

本実験においては，記事から事実内容を抽出し，それをネットワーク表現，特に木構造として表現することとした。一般的に，意味表現としては，フレーム構造，述語論理表現，セマンティック・ネットワーク，フレーズ・マーカー表現など，何らかの意味で述語支配に類する構造を用いるものが，現在多方面で用いられている。ここでも，その考え方を意識しつつ，一方で，現行データベースで使われているキーワード，記事インデックスとの距離感を少なくするものとした。

（例）



構造のエントリーは記事とし，ルート・ノード（ソース・ノード）として表現する。記事のノードから張られる木構造が，その記事の内容を表現する。逆に，各ノード（あるいは部分木）からアークをさかのぼることにより，それを含む記事が得られる。

各ノードは１つの概念を表わす「キー・ターム」とする。これは通常のキーワードにあるように，固有名詞の類から，行動を表わすものまでを対等に

含む。本実験で用いたノードのキー・タームの一覧を図2.3－3に示す。各ノードは，それぞれ任意の格を属性値として持つことができる。それは，そのノードの子を指すアークとして表現し，格の種類をアークに，その値はアークの指す子ノードと，子ノードが支配する構造で表わすものとした。本実験で用いられた格コードの一覧を図2.3－4に示す。

本実験においては，各記事は木構造で表現されており，したがって，異なる記事が同じノードや部分構造を共通に支配することはない。また，同じ記事中でも，同じノードが異るノードに支配されるような構造はない。それらの場合は，同じ形の部分構造が重複して出現する。ある事象をこの事実表現の集りの中から検索する場合は，その事象を表わす部分構造を事実表現集合中から，パターン・マッチングにより，とり出すことになる。もし，事象が1つのキーワードで，1ノードとして表現されていれば，キーワード検索と同じことになる。

（例1）

〔article, 3-8303044010LT,　〔fact, 1.〔agent, ……
　　〔act, ……
　　〔cause, ……
　　〔fact, 2.〔agent, ……
　　〔act, ……
　　〔cause, ……

連番3，記事番号8303044010の紙誌名コードLTの記事が2つの事実表現（格コード＼fact／）をもつことを示している。この記事の場合，実際には＼act／と＼cause／が共通であるが，独立の表現とした。

（例2）

〔fact, 1,〔agent, Saudi,〔occup, prince,〔name, Abdallah〕〕〕,
　〔act, demand,〔obj1, OPEC〕,
　　〔act, observe,〔obj2, pact〕〕〕〕

1つのfactのもつ基本要素は主体（＼agent／）と動作（＼act／）とした。

/*verbal*/
[accept, 受諾]
[adopt, 採用]
[agree, 同意]
[backdate, さかのぼって実施]
[complain, 不満表明]
[continue, 継続]
[cut, 値下げ]
[decrease, 減少]
[defend, 維持]
[demand, 要請]
[deny, 否定]
[devision, 分裂]
[discuss, 会談]
[drop, 下落]
[face, 直面]
[inevitable, 不可避]
[insistence, 主張]
[minor casualty, 小さな損害]
[no intension, 意図無し]
[observe, 順守]
[open, 開放]
[proposal, 提案]
[quote, 引用]
[respond, 対応]
[response, 反応]
[raise, 上昇]
[support, 支持]
[under no pressure, 圧力なし]
[unite, 結束]

/*etc*/
[African crude, アフリカ原油]
[budget deficit, 財政赤字]
[investor, 投資家]
[LPG pricing option, ＬＰＧ価格オプション]
[new structure, 新構造]
[oil price, 原油価格]
[oil production, 原油生産]
[oil revunue, 原油収入]
[price differential, 価格差]
[pact, 協定]
[waiting, 待機戦術]

/*org*/
[Algeria, アルジェリア]
[Arab oil producers, アラブ産油国]
[Gulf states, 湾岸諸国]
[Kuwait, クウェート]
[OPEC, ＯＰＥＣ]
[Saudi, サウジアラビア]
[the Gulf Co-operation council, 湾岸諸国会議]
[U.S. oil companies, ４大アメリカ石油会社]
[Venezuela, ベネズエラ]

/*occup*/
[expert, 専門委]
[oil minister, 石油相]
[prince, 王子]
[representative, 代表]

/*name*/
[Abdallah, アブダラ]
[Al-Sabah, アルサバ]
[Berti, ベルティ]
[Mark Twain, マーク・トゥエイン]
[Yamani, ヤマニ]

図2.3－3　用いたキーターム一覧

/*格*/ /*日本語*/ /*ゴミ*/
[article, 記事［ＩＤ，］]
[fact, 事実，]
[agent, 主体，が]
[act, 動作，]
[obj1, 相手，に対して]
[obj2, 対象，を]
[occup, 役職，である]
[name, 名，というもの]
[cause, 原因，の原因で]
[purpose, 目的，の目的で]
[of, of，の]
[with, with，と]

図2.3－4　用いた格コード一覧

agentは抽象化を図るため，個人というよりも，その属する組織を中心とすることにした。そのため，一般の修飾構造（「((サウジの）皇子である）Abdallah」）とは逆の構造となっている。

fact に対して，理由（`cause'），目的（`purpose'）などの格も与えられることがある。

また，こうした事実表現を誰が語ったかという情報や，どのような観測が行われているかということを，どのように表現すればよいかという問題も考察すべき課題である。記事情報に関しても，多くの場合，発言者，発表者，観測者，それに紙誌自体の立場を含め，種々の考慮すべき情報があるが，ここでは，そこに立ち入ることは避けた。

事実表現は，格コードを適切な日本語送りに置き換えて読むことにより，記事インデックスに類似したものになることも考慮した。しかし，agentに関して述べたような問題もあり，「左から右へ，末端からルートへ」という読み方だけでは，「サウジの皇子である Abdallah，OPEC に対し協定の遵守を要請」という日本文にはならない。しかし，これは簡単なメカニズムで実現可能である。

2.3.5 事実抽出のメカニズム

(1) 処理の概要

構文解析の結果得られた文の構文的構造を表現する構文解析木を変換し，その要素のみを抽出した形（事実表現）に変換を行う。ここでの処理にはこういったリストを扱う処理が非常に簡潔に記述できるプログラミング言語Prolog を用いる。入力は，構文解析木をリスト構造にして表現したもの（例を図 2.3－5に示す）である。このリスト構造に，構造パターンごとの処理，シソーラス等の知識を使った処理を施し，その出力として均一でかつ重要点を抽出した事実表現（図 2.3－6 に図 2.3.－5を変換した結果を示す）を得る。

(2) 全体の処理の流れ

全体としての処理の大まかな流れは，次の４つのフェイズに分かれる。

① 前処理－後の処理を簡単にするため，名詞節の後置修飾をする形容詞節，分詞を取り除く。

② 主体抽出－先頭の名詞句に注目し，動作の主体となる概念の抽出をする。

③ 動作抽出－主動詞に注目し，各動詞ごとに用意してあるパターンに従って動作，動作対象などを決定する。

④ 表現の標準化－決定されたリストの各要素を標準表現リストにより統一的表現に変換する。現在は，用語の統制語化に限っている。

この流れを制御するプログラムを図2.3－7に示す。（なお，ここで用いるProlog の構文規則として，以下のプログラムにおいて，大文字は変数を表わし，文字列として空白や特殊記号を含むときは，引用符(♥)でくくるものとする。）

図2.3－7は解析木を読み込み，続いて処理をする述語tran を，ファイルの終わりを検出するまで繰り返し呼ぶものである。述語readsantは指定されたファイル（ここではtree1. ptという名前のファイル）から解析木を一つ読み込み，それをsent (T)として（Tは解析木を表わすリスト）assert し，データベースに登録する。readsentに続く節にこの解析木を結合させ，tran にそれをひきわたすわけである。述語tran は，入力ファイルの終わりのときに偽となり，not（tran(X)）が真となり，次の述語へ実行が進む。次の行では，ファイルのクローズ等の後始末を行い実行を終了する。

述語tran は，まず入力ファイルの終了かどうかをチェックする。一連の変換は上から順に，

イ．読み込まれてassert してある入力表現のヘッダと解析木の分離

ロ．前処理のフェイズ

ハ．主体抽出のフェイズ

```
['1-8303234009',
[s, [np, [np, [adj, 'Crown Prince', occup],
              [noun, 'Abdullah', name]],
         [pp, [prep, of],
              [np, [noun, 'Saudi Arabia', organ]]]],
    [vp, [vt, warn, past],
         [np, [np, [adj, other],
                   [noun, member, plu]],
              [pp, [prep, of],
     [np, [noun,  'the Organization of Petroleum Exporting Countries (opec)',
                        organ]]]],
         [nouncl,
             [intro, that],
             [s, [np, [det, the],
                      [noun, kingdom, organ]],
                 [vp, [verb, be, past],
                      [adv, perfectly],
                      [adj, capable],
                      [pp, [prep, of],
                           [np, [vp, [vt, win, gerund],
                                          [np, [det, an],
                                               [adj, oil],
                                               [noun, price],
                                               [noun, war]]]]],
                 [advcl,
                      [intro, if],
                      [s, [np, [prn, they]],
                          [vp, [vt, fail, present],
                               [toinf, [intro, to],
                                       [vp, [vt, observe, inf],
         [np, [np, [det, the],
                   [noun, pact]],
              [pa, [vt, conclud, ptp],
                   [adv, 'last week', time],
                   [pp, [prep, in],
                        [np, [noun, 'London', place]]]]]]]]]]]]]]].


['2-8303174011',
[s, [s, [np, [adj, 'Saudi Arabia', poss, organ],
             [adj, '$5 a barrel'],
             [adj, oil],
             [adj, price],
             [noun, cut]],
        [vp, [vt, backdate, future, passive],
             [pp, [prep, in],
                  [np, [adj, two],
                       [noun, stage, plu]]],
             [pp, [prep, to],
                  [np, [np, [art, the],
                            [noun, beginning]],
                       [pp, [prep, of],
                            [np, [adj, 'last month', time]]]]]],
  [s, [np, [prn, it]],
      [vp, [vt, report, past, passive],
           [adv, yesterday, time],
           [advcl, [intro, as],
                   [s, [np, [np, [det, the],
                                 [noun, price]],
                            [pp, [prep, of],
                                 [np, [adj, 'North Sea'],
                                      [noun, oil]]]],
                       [vp, [vi, continue, past],
                            [toinf, [intro, to],
                                    [v, be, inf],
                                    [adv, 'in limbo']],
                            [pp, [prep, after],
                                 [np, [np, [adj, 'this week', poss],
                                           [np, [noun, agreement]]],
                                      [pp, [prep, by],
 [np, [noun, 'the Organization of Petroleum Exporting Countries', organ]]]
                                 ]]]]]]]]].
```

図2.3－5　入力の構文解析木の例

```
[article,1-8303234009FT,[fact1,[agent,Saudi,[occup,prince,[name,Abdallah]]],
                                [act,demand,[obj1,OPEC],
                                            [act,observe],[obj,pact]]]]

[article,2-8303174011LT,[fact1,[agent,Saudi],
                                [act,backdate,[obj2,cut,[of,oil price]]]]]
```

図 2.3 − 6　作成された事実表現の例

```
/*  all sentences  */
?- repeat,
   readsent('treel.pt'),
   sent(X),
   not(tran(X)),

!,tell('conrep.p'),told,see('treel.pt'),seen,retract(sent(X)),!,fail.
```

図 2.3 − 7　メインプロセス

```
/*  shapeup.p  delete  adj cluse,  adverb  etc.  */
shapeup([],[]) :- !.
shapeup([np,[np|A],
            [adjcll_]],
        [np|A]) :- !.
shapeup([np,[np|A],
            [pal_]],
        [np|A]) :- !.
shapeup([A,B|C],[A,B|C]) :- atom(B),!.
shapeup([A,B|C],[A|D]) :- shapeup1([B|C],D).
shapeup1([],[]) :- !.
shapeup1([A|B],[C|D]) :- shapeup(A,C),shapeup1(B,D).
```

図 2.3 − 8　前処理のプログラム

ニ. 動作抽出のフェイズ

ホ. 表現の標準化フェイズ

ヘ. 画面への表示

ト. ファイルへの出力

チ. assert してある文の消去

の処理を行う。なお，文が2つ以上の文からなっているとき，及び主語が共通で2つ以上の動詞句を持つときは，この処理の途中にくり返し，及びそれらをまとめる処理が加わる。各フェイズの処理過程の詳細を，図2.3－5に構造を示した2例を用いて，以下の各節で説明する。

(3) 前　処　理

ここでは，後の処理を単純にするため不要な要素の切り捨てを行う。様々な要素をこの段階で切り捨てることが可能だが，今回は当面の処理を考えて，名詞の節の後置修飾をする形容詞節，分詞を取り除くにとどめる。重要な概念のみの抽出を考える限り，こういった要素は，不要である。また，これらを切り落とすことによって，Prologによるパターンマッチ処理の効率の低下を防ぐことができる。

図2.3－8に，この処理をするプログラムを示す。

例えば，〔np,〔np,Ⓐ〕,〔pa,Ⓑ〕〕というパターンを〔np,Ⓐ〕というリストに変えてしまうことによって後置修飾する分詞の部分〔pa,Ⓑ〕の部分がもともとなかったかのごとく以下の処理ですることができる。この処理をリストすべてにわたって行うわけである。

(4) 主体抽出

このフェイズで事実表現の主体の抽出を行う。記事の先頭の文を取ったということから，ほとんどすべての対象文で，事実表現の主体となるべきものが，先頭の名詞句（倒置の構文を除いて，通常，主語となる名詞句）に出現している。この事実を考慮して，ここでの処理は次の順に行う。

イ. 先頭の名詞句を捜し出す。

ロ. そこから主体を捜し出す。

なお，２つの段階では，主名詞句，動作の対象となるような表現が表われる場合がある場合，これもひろい出す。図2.3－9にそのプログラムを示す。(a)の述語 findagent は前記イ，ロに対応する述語pick，nporganを呼び出し処理する。

pick は，第1引数（np)を親とするサブトリーを第2引数（Sent）の木構造から捜し出し，第3引数（NP)に返す。

また，(b)のnporganは，NPのパターンに応じて，主体を捜し出し，それをAgentに返す。この例では，

＼Crown Prince Abdnllah of Saudi Arabia／が，役職名，人名，of，国名（組織名）というパターンであることに注目し，

〔agent Saudi Arabia〔occup, Crown Prince〔name, Abdallah〕〕〕という表現を返す。主体を含む句の表現は多様であり，様々なパターンを用意する必要がある。様々なパターンを用意すれば，用意するほどマッチングのための時間もそれに比例してかかることになるが，今回の実験ではその問題に立ち入ることはしなかった。

２番目の文章の先頭の名詞句Saudi Arabia's $5 a barrel oil price cut のように動作の対象となるoil price cutがここに現われるような場合はfindagentの３番目の引数Additionalに，この情報を渡し，次のフェイズで利用する。図の最下段がこのパターンのための節である。ここで，この場合＼Saudi Arabia's／と組織が所有格で使われていることに注目し，（意味マーカー的なこれらの情報を用いて）主体（Agent）と，動作対象（Dousa）を引き出す。

(5) 動作抽出

このフェイズでは文の主動詞に注目し，各動詞ごとにパターンを用意して主体以外の事実表現を求める。この処理をするプログラムを図2.3－10に示す。この述語verb handlingでは，まず最初に述語pick main verb が変数sentを通して文（実際には動詞を含むようなサブトリー，例えば，不定詞句などでも良い）を受けとり，変数Verbに主動詞を，VP

```
/*  findagent.p    find agen  */
findagent(Sent,[agent|Agent],Additional) :-
    pick(np,Sent,NP),nporgan(NP,Agent,Additional).
```

(a)　主体抽出

```
/*  nporgan.p  */
/*  for 1     Crown Prince Abdullah of Saude Arabia     pattern  */
nporgan(NP,Agent,[]) :-
    equal(NP,[np,[np,[adj,Occup,occup],
                     [noun,Name,name]],
                 [pp,[prep,of],
                     [np,[noun,Organ,organ]]]]),
    equal([Organ,[occup,Occup,[name,Name]],Agent).

/*  for 1.subclause    other members of Opec     pattern  */
nporgan(NP,Agent,[]) :-
    equal(NP,[np,[np,[adj,other],
                     [noun,member,plu]],
                 [pp,[prep,of],
                     [np,[noun,Organ,organ]]]]),
    equal(Organ,Agent).

headnoun(NP,Out)  :-
    pick(np,NP,NPi),
    equal(NPi,[np,[det,the],
                  [noun,Out]]).

/*  for 2  Saudi Arabia's $5 a barrel oil price cut  pattern  */
nporgan(NP,[Agent],Dousa) :-
    equal(NP,[np,[adj,Agent,poss,organ|_]|_]),
    pick(noun,NP,[noun,Dousa]).
```

(b)　名詞句各パターンの処理

図 2.3 − 9　名詞句の処理

```
/* verbhandling.p */

verbhandling(Sent,Infsubj,Outrep) :-
    pickmainverb(Sent,Verb,VP),
    P=..[Verb,VP,Infsubj,Outrep],
    call(P).
```

(a) 動詞処理のトップレベル

```
/* pickmainverb.p */
pickmainverb(Sentorphrase,Verb,VP) :-
    pick(vp,Sentorphrase,VP),
    equal(VP,[vp,[vt,Verb|_]|_]).
pickmainverb(Sentorphrase,Verb,VP) :-
    pick(vp,Sentorphrase,VP),
    equal(VP,[vp,[vi,Verb|_]|_]).
```

(b) 主要動詞の抽出

```
/* eachverb.p */
/* warn */
warn(VP,[],Outrep) :-
   equal(VP,[vp,[vt,warn|_],
                Npobj2,
                Nounclause]),
    /*      dislist(['Nouncl=',Nounclause]),  */

   pick(advcl,Nounclause, [advcl,[intro,if],Sent2]),
    /*      dislist(['Sent2=',Sent2]),  */

   pick(vp,Sent2,[vp,[vt,fail|_],
                    [toinf,[intro,to],
                            Vpdousa]]),
    /*      dislist(['Vpdousa=',Vpdousa]),  */

   nporgan(Npobj2,Agent,Additional),
   verbhandling(Vpdousa,[],VConrep),
   equal(Outrep,[act,demand,[obj1,Agent],
                            VConrep]).

/* observe */
observe(VP,[],Outrep) :-
   equal(VP,[vp,[vt,observe|_],
                Obj]),
   headnoun(Obj,Obj2),
   equal(Outrep,[act,observe,[obj2,Obj2]]).

   /* backdate */
   backdate(VP,Obj,[act,backdate,[obj2,Obj]]) :-
       equal(VP,[vp,[vt,backdate|[_,passive]]|_]).
```

(c) 各動詞に関する記述

図2.3－10 動作抽出の処理

にその主動詞から始まる動詞句を返す。2つ目の述語は，新しい述語の定義であり，これが次に呼ばれる。この例文の場合まずVerb にこの文の主動詞warnがはいり，warn(VP, Infsubj, Outrep) という節で呼ばれることになる。このように主動詞によってまったく別の述語が呼ばれ，主動詞ごとに独立な処理が行われる。図2.3－10の(c)に，これで呼ばれる述語warnの定義及び，この後の処理で呼ばれるobserveの定義が示されている。

述語warn は動詞（VP）と主語を表わす節からの情報（Infsubj；この場合は何もなし）を受け取り，事実表現の部分（Outrep ）をくり返す。ここでは，＼warn, 組織， if ～ fail to 動詞句 ／，というパターンに注目して変換を行っている。ここで注意したいのは，fail toに続く動詞句において，再帰的にverb handling を呼び出し，＼observe the pact／の部分を変換していることである。この過程で，述語observe が，先のwarn のときと同様にして呼ばれるわけである。一方，名詞句の扱いにおいても，主体抽出のときに用いた述語nporganを用いている。このように統一的に扱うことによって，全体が見通しよく構成できる。

2番目の文の場合は，主動詞が＼Will be backdated ／（受身の未来）である。図2.3－10の(c)にこれを処理する述語backdate を示した。ここでは，backdateの目的語cut（値下げ）を主語の処理（ find agent）から受け取っている。このcutは後でシソーラスと参照され〔cut,〔of, oil price〕〕（原油値下げ）と変換される。このような記事では，ただcutと出た場合は，原油価格のcut（値下げ）と解釈するわけである。

(6) 表現の標準化

最後に標準表現リストを参照し，様々な形で表現されている表層の形を統一的に書き替える。処理としては，ここまで得られた事実表現のリストの各要素を順にたどり，標準表現リストを見て，必要ならばあらかじめ決めた表現に置き替えるというものである。図2.3－11にそれを行うプログラムと標準表現リストの一部を示す。この例ではOPEC がフルに表わ

れているものをOPECに，Saudi Arabiaを Saudiに，Crown Princeをpriceに変えている。標準表現リストにマッチしないものはそのまま出力する（最後の節で処理される）。

(7) その他の述語の処理内容

図2.3－12はサブトリーの抽出のための述語pick,リストのプリティ・プリンタpplist 及び equalの定義である。これらは，今まで見てきたように，様々なところで使われる。

(8) まとめと問題点

事実表現抽出を，具体的処理内容で見たように，原則として，主体は最初の名詞句，動作は主動詞を中心にするという方法で行った。また名詞句・動詞句の処理は汎用な記述により，文全体の主語，動詞句だけでなく，例えば目的語，名詞節の中，不定詞句といった部分構造に対しても，同一の処理が行えるように構成した。これらのことにより，全体の処理が見通しよく，また簡潔に記述できた。

動詞の処理を各動詞ごとに記述することによって，プログラムの部品化が自然に行える。上のレベルの述語（プログラム）では，主動詞に応じて，その述語を呼び出すだけでよく，細かいことにとらわれず処理を依頼するだけで済む。また，それぞれの動詞に対する辞書記述と見ることもできる。ある動詞があるパターンで出現したときには，こういった事実を表現するといった知識の表現になっている。これらの知識を増加させることにより，ある程度，同種のほかの記事に対する事実抽出も可能であろう。

全体の処理の流れはすっきりと記述できたが，個々の処理は，アドホックなものにならざるを得なかった。名詞句からの主体抽出では，辞書引きの段階で付加可能とした若干の意味マーカと，その名詞句の構成パターンを頼りに，それぞれの文を見ながら処理内容を検討した。その為，例えばいくつかの名詞句の処理は同一な方法（述語）として記述できたが，ほとんどは，そのパターンのためだけの記述を数多く用意することとなった。これは単純に処理効率だけを見ても，非常に多くのパターンのマッチング

```
/*  checkthesau.p  */
checkthesau([A,B1|C1],[A,B2|C2]) :-
    atom(A),!,
    thesau(B1,B2),
    checkthesaul(C1,C2).
checkthesau(A,B) :-
    checkthesaul(A,B).
checkthesaul([],[]).
checkthesaul([A1|B1],[A2|B2]) :-
    checkthesau(A1,A2),
    checkthesaul(B1,B2).
```

(a)　表現の標準化処理

```
/*  thesau.p   thesaurus  */
thesau('the Organization of Petroleum Exporting Countries (opec)', 'OPEC') :- !.
thesau('the Organization of Petroleum Exporting Countries', 'OPEC') :- !.
thesau('Opec', 'OPEC') :- !.
thesau('Crown Prince', 'prince') :- !.
thesau('Saudi Arabia', 'Saudi') :- !.
thesau(X,X).
```

(b)　標準表現リストの例

図 2.3 － 11　表現の標準化

```
/* X pattern , Y sent */
pick(X,[X|Y],[X|Y]).
pick(X,[_|Y],A) :- pick1(X,Y,A).
pick1(X,[Y1|Y2],A) :- pick(X,Y1,A).
pick1(X,[_|Y],A) :- pick1(X,Y,A).

/*  nolet    number of letters needed to print an atom X  */
nolet(X,N) :-
    name(X,Y),length(Y,N).

/*  pplist.p  pretty printer of a list  */

pplist(X) :- pplistn(X,0),nl.

pplistn(X,N) :-
    atomic(X),!,write(X).
pplistn([X|Y],N) :-
    Y=[],!,write('['),M=N+1,pplistn(X,M),write(']').
pplistn([X|Y],N) :-
    atomic(X),!,write('['),write(X),write(','),
    nolet(X,M1),M2=N+M1+2,pplistn1(Y,M2),write(']').
pplistn([X|Y],N) :-   /*  X list  Y non nil  */
    write('['),M=N+1,pplistn(X,M),write(','),nl,tab(M),
    pplistn1(Y,M),write(']').
    pplistn1([],N) :- !.
    pplistn1([X|[]],N) :-
        !,pplistn(X,N).
    pplistn1([X|Y],N)  :-
        atomic(X),!,pplistn(X,N),write(','),
        nolet(X,M1),M2=M1+N+1,pplistn1(Y,M2).
    pplistn1([X|Y],N)  :-   /*  X list  */
        pplistn(X,N),write(','),nl,tab(N),pplistn1(Y,N).

    /*  equal  */
    equal(X,Y) :- X = Y.
```

図 2.3－12　その他の述語

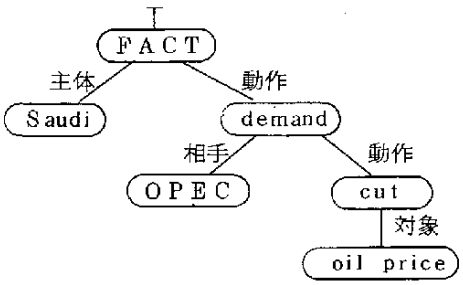
を必要とするため非能率である。また，方法論として考えた場合，このような方法で一般的な場合に応用できるという保証はなく，大きな問題として残される。

2.3.6 データベースによる実現と利用

事実抽出のプログラムにより作成された事実表現は，そのままの形でProlog 言語上で検索・分析処理にゆだねることができる。しかし，大量の事実表現を効率的に蓄積し，多様な検索・分析要求にこたえる為には，DBMS（データベース・マネジメント・システム）の利用が必要となる。ここでは，５８年度に開発したデータベース検索システムにおいても用いられた汎用ＤＢＭＳのＩＮＱを設定した。

事実表現を忠実に表現し，利用にゆだねることは，現存する汎用DBMSの枠組にそぐわない面が相当にある。事実表現の基本構造でなく，その下で表現される各種の論理的構造が，例えば，述語に対して５Ｗ１Ｈといった形で模型化されてしまえば，データベースのデータ構造もその形で定義することができる。しかし実際には，構造の入れ子があり，また，５ＷＩＨといったお仕着せで対応するまでには，本実験が固まったものでもないので，このような標準化は今回は採用しなかった。そこで，逆に，論理的，アルゴリズム的に生の形で低いレベルの表現を行うこととした。すなわち，事実表現の各ノードを INQ ファイルの１レコードとして表現し，その子ノードへのリンクをそのレコードのπの不定繰り返しとして表現した。

例えば，次の構造は６個の INQ レコードとなる。



（ファイル構造）

記事の文イメージ等を格納したARTICLEファイルと，事実表現を格納したNETWORKファイルとからデータベースを構成し，記事番号により，結合することができるようにした。

(1) INQファイルARTICLEのデータ構造記述

（INQ／FDL）

```
02  ARTICLE-REC.
  03  ARTICLE-NO                PIC 9(10) PKY.
  03  SOURCE                    PIC X(4).
  03  TEXT(N).
    04  LINE.                   PIC X(64).
```

ARTICLE-NO ：10桁の記事番号

SOURCE ：4桁の紙誌名コード

LINE ：原文の文字列イメージ

表示上の問題などから，見やすい1行単位に分割しておく。

(2) INQファイルNETWORKのデータ構造記述

（INQ／FDL）

```
02  NODE-REC.
  03  NODE-NO                   PIC 9(8) PKY.
  03  NODE-TYPE                 PIC X.
  03  ARTICLE-NO                PIC 9(10).
  03  TERM-ENGL                 PIC X(32).
  03  TERM-JPN                  PIC N(16).
  03  ARCS(N).
    04  CASE-CODE               PIC X(16).
    04  CASE-JPN                PIC N(8).
    04  CASE-JSFX               PIC N(8).
    04  DESTINATE-ND            PIC 9(8) BY CASE-CODE.
    04  DESTINATE-EN            PIC X(32) BY CASE-CODE.
```

NODE-NO ：ノード間のリンクをとる為に準備したユニークな番号。ノード間のリンクをプログラムでのみ行う場合は不要である。

NODE−TYPE　：構造の最上級ノードのとき'F'，それ以外はブランクとなる。

ARTICLE−NO　：ARTICLE ファイルとの結合キーとなる記事番号。NODE−TYPEが'F'のときのみ，意味をもち，それ以外は0が入る。

TERM−ENGL　：NODE−TYPEが'F'以外のとき，そのノードに対応するときキー・タームの文字列。

TERM−JPN　：TERM−ENGL に対応する日本語。

ARCS　：そのノードに対する子ノードへのアークの本数だけの繰り返しをもつ。

CASE−CODE　：そのアークのラベルコード。

CASE−JPN　：CASE−CODEに対応する意味。

CASE−JSFX　：CASE−CODE に対応する日本語送り。単純なケースの場合，子ノードの日本語の後にこれを並べると，日本語の格文節になる。

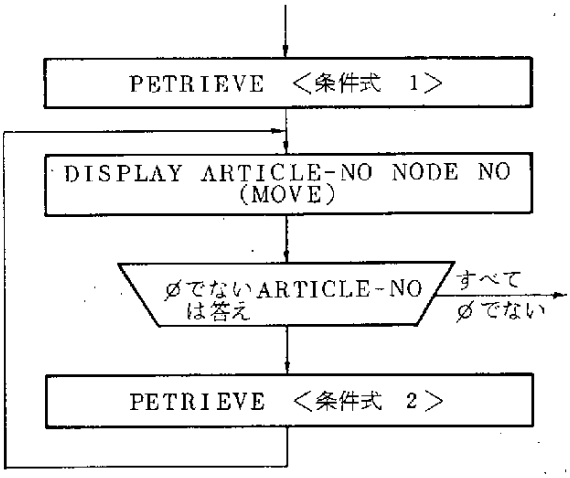
DESTIN−NODE　：子ノードのノード番号。NODE−NO の項で述べたように，プログラムでのみユーザとのインターフェイスをとるならば，これはINQファイルの内部的なアドレス番号で代替できる。

DESTIN−TERM　：その子ノードのTERM−ENGL の値。これは冗長な情報である。INQがサポートしている基本機能だけを用いては，ノード間のリンクをとった検索は簡便にはできない。そこで，ここに重複して情

報を入れることにより，「主体が何某で，動作として何々の値をもつもの」という１階の検索が容易にできるようにした。

（データの検索）

(1) キーワードによる検索



<条件式－１>　：　ORを用いたTERM−ENGL又はTERM−JPNに関する条件式

<条件式－２>　：　CASE−CODE NE SPACE AND DESTIN−NODE EQ（値）

<値>　：　ARTICLE−NOがØであるNODE−NOをORで結ぶ。

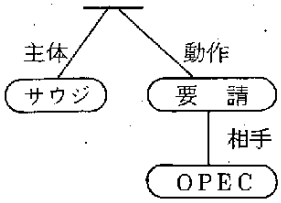
図2.3－13　キーワードによる記事の検索

いくつかのキーワードを含む記事を得るときには，図2.3－14の手順を，画面を見ながら行うか，プログラムにより同じことを行う。このような頻雑な手順となるのは，論理的な１レコード（１記事あるいは１つの事実構造）とINQファイルで実現したレコード単位の相違に起因する。実は，タイプ

Fのレコード以外にも，記事番号を入れておけば，1回の検索ですべてが終了するが，ここでは，キーワード検索は主目的ではないので省いた。

(2) 構造による検索

例えば，「サウジがOPEC に対し何らかの要請を行った記事が欲しい」という要求を考える。このとき，現在の枠組みでは次の様な構造をサーチすることになる。



これに対しては，次の4つのコマンドで記事を求める。

```
RETRIEVE TERM-JPN EQ 要請 AND ( CASE-CODE EQ OBJ2 AND DESTIN-TERM EQ OPEC )
DISPLAY NODE-NO    （この結果を次のコマンドによって求める。）
RETRIEVE ( CASE-CODE EQ AGENT AND DESTIN-TERM EQ SAUDI ) AND
 ( CASE-CODE EQ ACT AND DESTIN-NODE EQ 7 OR 18 OR 54 OR 99 )
DISPLAY ARTICLE-NO NODE-NO
```

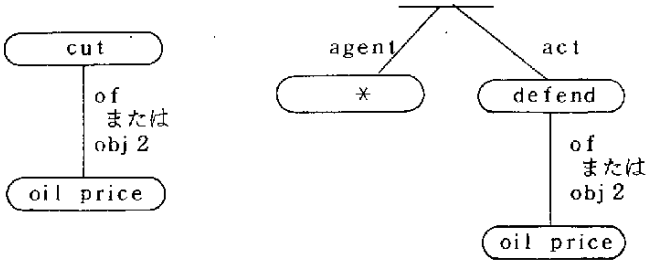
このように，様々な構造に対して検索を行うことができるが，その反面，極めて頻雑な操作が必要となる。そこで，想定されるいくつかのパターンについてサポートする検索インターフェイス・プログラムを準備することが望ましい。この機能は，また，ユーザの要求を解釈し，特定の内部構造を生成する機能をも含むであろう。

(3) データの利用

このような事実表現がデータベース中に実現されていくことにより，これまで不可能であったような，多様な，きめの細かい分析を行う展望がひらけてきた。ここでは，石油価格に関するサウジの及びサウジに対する動きの分析をシミュレートしてみよう。

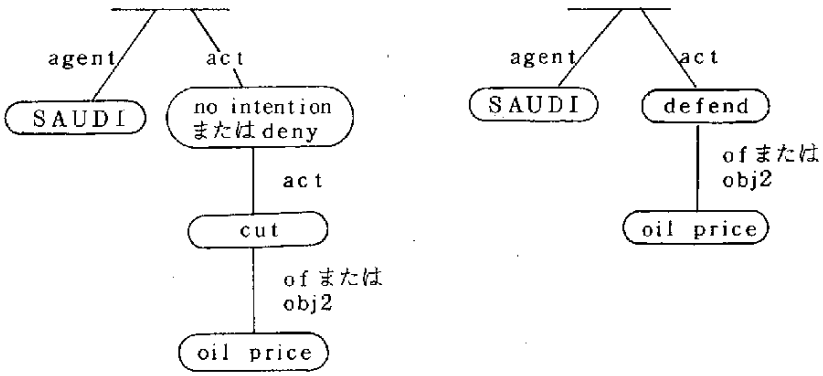
'oil price' という語は，作成された事実構造の随所に表われるが

（１５個所），その中の９件については，cutが上位ノードになっている。残りの中の４件がdefendであり，その２種類についてまず考える。

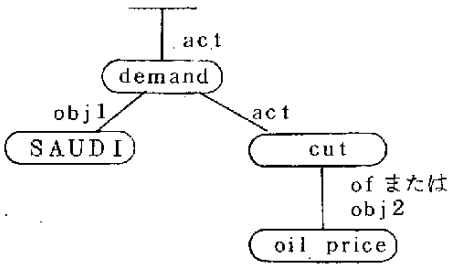


defendは，既に動作として完結し，主体も要求していると考えられるので，上図の構造に絞って差しつかえないであろう。これに対し，'cut of oil price' は何らかの動作と主体を要求すると考える。そこで，調べてみると，上位のノードには，backdate, demand, agree, inevitable, deny, no intentionという'act'があることが判る。ここで，backdateは評価が難しいとして除外すると（強いて言えば，値下げが確定している段階を表現している。また，今回の処理では時点の処理を省いたこともあり，意味付けが仕難い），次のシナリオが設定できる。

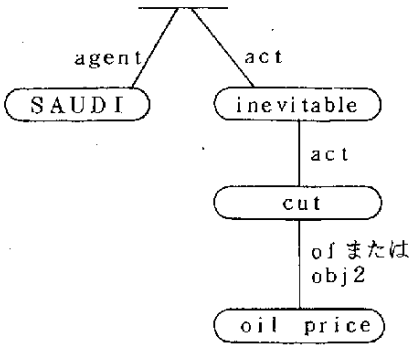
(a)　「サウジは値下げをしない」



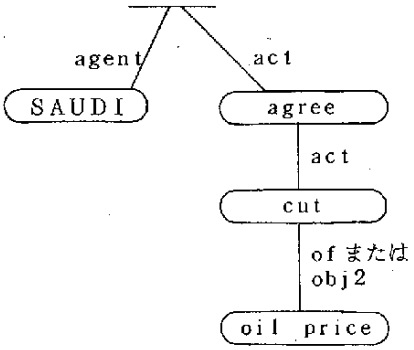
(b)　「サウジへの値下げ要求がある」



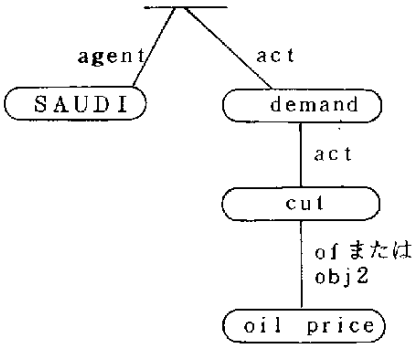
(c) 「サウジが値下げを不可避とみた」



(d) 「サウジが値下げに合意した」



(e) 「サウジは他国に対しても値下げを要求」



ここで，＼SAUDI／とあるのは，

Saudi C Gulf States C OPEC

という関係を用いることを意味し，事実上は，上記3タームの論理和と考えて差しつかえない。

各シーンを事実表現から得るのは，(2)で述べた構造による検索の場合

と同様の操作による。この結果，各シーンの見られる記事の日付けは次の様になる。

シーン (a) 825050, 820816, 821122（同一文で2回），821207

シーン (b) 830104

シーン (c) 830211（2記事）

シーン (d) 830224（3記事）

シーン (e) 830228

少数のサンプルを対象としているが，それでも抽出しえないものもある。更に，大雑把なシーンの設定に対し，もっと細かいニュアンスも加える必要があるかも知れない。更に事実表現作成のメカニズムから，シナリオの設定まで，余りにも結果からの作り上げが多く，アドホックに過ぎるように思われるかも知れない。しかし，この一連の操作，分析のモデルの垣間見させるもの，差し示すものは，これまで機械処理で見られなかった豊かな地平であることを察することは，同時に，容易である。例えば，極めて即時的なイメージを挙げてみても，上記のように，適切なシナリオを設定して，時系列での頻度分析，レーダ・チャート，空間的には国等による色分け，特異値の特定など多様な応用が考えられる。その為には，大量のデータに対応できる（知識ベースを含めた意味での）メカニズムと，シナリオの作成の様な過程でのユーザをサポートする方法論の確立が必要である。

2.3.7 サンプルテキストに対する事実抽出とその評価

今回の実験では，サンプル・テキストを評価する中で事実抽出のメカニズムを検討し，抽出結果の分析を行った。そこで，資料2の各テキスト毎にメカニズム作成の為の着眼点，問題点並びに結果の構造の評価に関して述べる。

① [article,34-8205054017APS,[fact,1,[agent,Saudi],
[act,defend,[obj2,oil price]]]]

'is determined to' の部分は次に続く不定詞で動作を記述するものととらえる。続く並列句は考慮に入れていない。

② [article,33-8207054012MEES,[fact,1,[agent,OPEC],
[act,discuss,[obj2,differential,[of,African crude]]]]]

こういったまわりくどい表現は，非常に扱いにくい。アドホックに扱わざるを得ない。先頭の名詞句に現われる組織ＯＰＥＣを切り出し，主体とする。また，それがかかる名詞（agenda）を動作とする。一般的にこのような処理を施すことは，非常に困難であろう。

③ [article,32-8207104002FT,[fact,1,[agent,OPEC],
[act,devision,[cause,insistence,[agent,Saudi],
[act,raise,[of,African crude]]]

〝be devided over ～〟の～の部分を原因（cause）と解釈する。〝Saudi's insistence〟は２番目の文と同じパターンで，「組織の所有格へ動作」に注目し変換した。

④ [article,31-8207214013WJ,[fact,1,[agent,Kuwait,[occup,oil minister,[name,al-Sabah]]],
[act,quote,[obj2,Mark Twain]]]]

こういった文は，扱えないとして省いた。

⑤ [article,30-8208164020PIW,[fact,1,[agent,Saudi],
[act,no intention,[obj1,raise,[of,oil production]]]],
[fact,2,[agent,U.S. oil companies],
[act,under no pressure]]]

先頭の分詞は無視する。〝make no effort〟を〝no intention〟と他の記事に出現する表現に統一する。２番目の節と１番目の節との関係は今のところ記述できない。

⑥ [article,29-8208164021MEES,[fact,1,[agent,Saudi],
[act,no intention,[act,cut,[of,oil price]]]]]

先頭の前置詞は考慮しない。～understandsは，それに続く名詞節から情報を得ることとする。

⑦ [article,28-8208184014APS,[fact,1,[agent,the Gulf Co-operation Council],
[act,adopt,[obj2,proposal,[of,Sauji]]]]]

情報が不定詞，前置詞句等に埋めこまれており，抽出がむずかしい。すべての不定詞・前置句を考慮に入れると，処理が非常に複雑になり，また，処理時間も爆発的に増大すると考えられる。

⑧ [article,27-8209064021IHT,[fact,1,[agent,OPEC,[occup,expert]],
[act,support,[obj2,raise,[of,oil price]]]]]

挿入されている分詞構文を考慮せずに処理を行う。ここでは，原油は原油でも高品質原油が対象となっているが，その情報は，この対象からはずした分詞構文にのみ存在する。どこまで処理を細かくするかむずかしい

問題である。

⑨
```
[article,26-820914012FT,[fact,1,[agent,Saudi],
                                [act,demand,[obj1,OPEC],
                                        [act,raise,[of,African crude]]]]]
```

これはすでに何回か出現したパターンである。このパターンが非常に多い。

⑩
```
[article,25-8209274011PIW,[fact,1,[agent,Saudi],
                                [act,continue,[agent,LPG pricing option],
                                        [act,open]]]]
```

despite以下の副詞節は，考慮せずに処理する。＼LPG pricing option／を主体とすることは問題かもしれないが，全体の枠組からこのような表示を選択した。

⑪
```
[article,24-8211224028IHT,[fact,1,[agent,Saudi],
                                [act,defend,[obj2,oil price]]],
                        [fact,2,[agent,Kuwait],
                                [act,defend,[obj2,oil price]]],
                        [fact,3,[agent,Saudi,[occup,oil minister,[name,Yamani]]],
                                [act,deny,[obj2,cut,[of,oil price]]]]]
```

主体が並列になっているため，最初の節で２つの事実表現になる。２番目の節が３番目の事実表現となる。２番目の節のotherwiseは，高度な知識がないと扱えないものだが，ここでは，この例文のためだけの処理を行った。

⑫
```
[article,23-8211234023FT,[fact,1,[agent,responce,[of,investor]],
                                [act,minor casualty]]]
```

今回は，処理からはずした。

⑬
```
[article,22-8211244001PU,[fact,1,[agent,OPEC,[occup,oil minister]],
                                [act,discuss]]]
```

この記事も予想である。＼事実／表現として，先頭の名詞節から主体動作を抽出した。全体的に「主体が，動作をした」という形でとらえているので，こういった記事の処理は，特別に扱わざるを得ない。

⑭
```
[article,21-8212014003APS,[fact,1,[agent,Algeria],
                                [act,accept,[obj2,price differential]],
                                [purpose,unite,[of,OPEC]]]]
```

＼algerian government／をアルジェリアと同一に扱い，＼for the sake of／をpurpose(目的)ととらえる。

⑮ [article,20-8212074004WJ,[fact,1,[agent,Saudi,[occup,oil minister,[name,Yamani]]],
[act,defend,[obj2,oil price]]]]

組織の所有格・役職・（同格で）名前のパターンが主語の部分に現われている。

⑯ [article,19-8212314001FT,[fact,1,[agent,Saudi,[occup,oil minister,[name,Yamani]]],
[act,discuss,[with,U.S. oil companies,[occup,representative]]]]]

観測を述べている記事であり，＼事実／表現としてはなじまない。そこで＼could result from／に続く名詞句から情報をひき出す。こちらは＼事実／である。

⑰ [article,18-8301054003LT,[fact,1,[agent,U.S. oil companies],
[act,discuss,[with,Saudi,[occup,oil minister]],
[purpose,demand,[act,cut,[obj2,oil price]]]]]]

「press～to 動詞句」を要請ととらえた。

⑱ [article,17-8301044006AWJ,[fact,1,[agent,U.S. oil companies],
[act,demand,[obj1,Saudi],
[act,cut,[of,oil price]]]]]

ここまでいくつか現われた，＼warn ～unless（あるいはif）／のパターンであり，要請（demand）としてとらえる。＼expect to／は次の不定詞に主動詞の役目を代行させた。

⑲ [article,16-8301054004WJ,[fact,1,[agent,U.S. oil companies],
[act,discuss,[with,Saudi,[occup,oil minister,[name,Yamani]]]]]]

「組織・meet・with・組織2」を「組織1が組織2と会談」ととらえる。2番目の節の主語は最初の節の主語と目的語を含んでおり，こういったことまで現在は処理できない。

⑳ [article,15-8301054005FT,[fact,1,[agent,Saudi,[occup,oil minister,[name,Yamani]]],
[act,discuss,[with,U.S. oil companies,[occup,representative]]]]]

名前の次に同格として，国名・役職が続くパターンから主体の表現を得る。＼continue talk／から会談（discuss）をとり，amid 以下の前置詞句は今回は対象としなかった。

㉑ [article,14-8301104005BW,[fact,1,[agent,Venezuela,[occup,oil minister,[name,Berti]]],
[act,complain]]]

＼be in a foul mood／により＼complain／(不満の表明）と変換した。OPEC に対する不満の表明であることはわかるが，これを抽出するのは困難である。

㉒ [article,13-8301174008PIW,[fact,1,[agent,Arab oil producers],
[act,be pressed]]]

記事の具体的内容は言っておらず，以前から関連のある記事を読んでいることを前提として書かれている。このような文章から，この記事が何を意味するかを抽出するのは極めて困難である。

㉓ [article,12-8302074012PN,[fact,1,[agent,waiting],
[act,continue]]]

〝Crude－price－cut waiting game〟といった比喩的な表現が使われており，今までの経過を熟知しているものを対象として書かれている。事実表現にそのような知識を反映することは現段階では不可能である。事実表現を利用するものの知識に頼らざるを得ない。

㉔ [article,11-8302114006FT,[fact,1,[agent,Saudi],
[act,abandon,[obj2,defence,[of,oil price]]]],
[fact,2,[agent,Saudi],
[act,inevitable,[obj2,cut,[of,oil price]]]]]

動詞句が２つありこれらを別々の事実として表現する。

㉕ [article,10-8302114001LT,[fact,1,[agent,OPEC],
[act,inevitable,[obj2,cut,[of,oil price]]]]]

「名詞・by・組織」というパターンに注目して，主体を組織，主名詞を動作（この場合は状態）の対象ととらえた。

㉖ [article,9-8302234004WJ,[fact,1,[agent,Gulf states],
[act,discuss]]]

〝met〟を会談として解釈した。amid～以下の前置句は考慮しないことにした。どのような前置句なら表現に入れるべきかは，むずかしい問題である。

㉗ [article,8-8302234005IHT,[fact,1,[agent,Gulf states,[occup,oil minister]],
[act,discuss,[obj2,respond,[obj2,drop,[of,oil price]]]]]]

〝held talks〟を会談（discuss）ととり，on～をその対象としてとらえた。respond，drop をそれぞれdiscuss，respondの対象ととらえることにした。動詞による表現と名詞による表現の変換は困難な問題である。

㉘ [article,7-8302244003WJ,[fact,1,[agent,Gulf states,[occup,oil minister]],
[act,agree,[obj2,cut,[of,oil price]]]]]

「役職・〝from〟・組織」のパターンに注目して，主体を抽出した。組織のところで国名を列挙してあるが，これは，シソーラスによって，湾岸諸国ととらえる。

㉙ [article,6-8302244012IHT,[fact,1,[agent,Gulf states],
[act,agree,[obj2,cut,[of,oil price]]]],
[fact,2,[agent,Gulf states],
[act,demand,[obj1,OPEC],
[act,accept,[of,new structure,[of,oil price]]]]]]

前の文と同様２つの事実表現として表わす。＼threaten to ～ to force／を前の文と同じように，要請としてとらえた。

㉚ [article,5-8302244015LT,[fact,1,[agent,Gulf states],
[act,agree,[obj2,cut,[of,oil price]]]],
[fact,2,[agent,Gulf states],
[act,demand,[obj1,OPEC],
[act,unite]]]]

動詞句が２つ並列しているので，２つの事実表現で表わす。＼threaten ～ unless～／のところを，直前の文と同じパターンで変換した。

㉛ [article,4-8302284007LT,[fact,1,[agent,Gulf states],
[act,demand,[obj1,OPEC],
[act,cut,[of,oil price]]]]]

「組織１・will・動詞句１・unless・組織２・動詞句２」のパターンを，組織１が組織２に，動詞句２を要求するととらえた。先頭の文に似たパターンである。

㉜ [article,3-8303044010LT,[fact,1,[agent,Saudi],
[act,face,[obj2,budget deficit]],
[cause,decrease,[of,oil revenue]]],
[fact,2,[agent,Kuwait],
[act,face,[obj2,budget deficit]],
[cause,decrease,[of,oil revenue]]]]

この文では，主体となる組織が２つ先頭の名詞句に現われている。後での参照を考えるとき，動作等を重複させて事実を２つ作った方が良いであろう。＼as a result of／を原因ととらえる。

㉝ [article,2-8303174011LT,[fact,1,[agent,Saudi],
[act,backdata,[obj2,cut,[of,oil price]]]]]

「組織の所有格＋名詞」のパターンを，主体を組織とし，名詞部分を次に続く動詞の対象としてとらえる。it was reported以下は，情報源を表わすが，全体とのバランスから事実表現に反映させることを見合わせた。

㉞ [article,1-8303234009FT,[fact,1,[agent,Saudi,[occup,prince,[name,Abdallah]]],
[act,demand,[obj1,OPEC],
[act,observe,[obj,pact]]]]]

「組織１・warn・組織２～if～fail to 動詞句」のパターンを，「組織１が組織２に，動詞句の動作をするように要請」と変換した。このようなパターンは，いくつかの記事に表われる。

2.4　総合解析システムの機能とその利用

2.4.1　基礎解析

(1)　位置づけ

基礎解析サブシステムの位置づけは，図 2.1－1 総合解析システム概念図に示されるように，本システムの全ての入口となっている。出版物や，手書き文書，テレックス，手紙等生の情報の状態では，計算機で処理することはできない。これらの情報は，計算機で判別できるように，つまりマシンリーダブルな形式となるように，一旦媒体変換される。媒体変換された情報は一次情報として計算機内の記憶媒体に格納される。基礎解析サブシステムは，この一次情報を入力として各個別のサブシステム，つまり検索サブシステム，内容分析サブシステム，翻訳サブシステムで容易に処理できる情報に加工する。この加工された情報は二次情報といい，二次情報データベースに格納されて，各サブシステムより利用される。各サブシステムでは，この二次情報データベースを利用することによって一次情報を直接扱う煩わしさより解放される。従って，基礎解析サブシステムの持つべき機能は，各サブシステムで必要とする共通的な機能を出来るだけ多く持つように考慮されている。

(2)　機　能

従来の情報検索システムにおけるデータベース作りは，人間の知的な助けなしに達成できるものではなかった。例えば，1つの学術論文に対して適切なキーワードを付したり，その要約を作ることは人間に任された仕事であった。

総合解析システムの目的の1つに，こういった人間の作業を機械化することがある。この作業の機械化を行うサブシステムが基礎解析サブシステムである。基礎解析サブシステムにおける機能構成を図 2.4－1 に示す。

一次情報である文章情報を入力し，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，文脈解析及び文脈解析結果による要約とに分類できる。

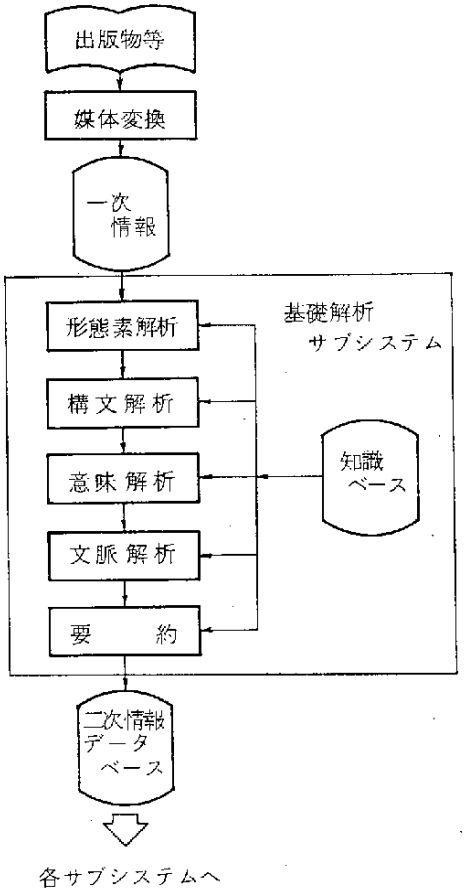


図 2.4 －1　基礎解析サブシステム構成図

二次情報は，これらの各処理の結果として生成される。形態素解析から要約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はなく，むしろ相互に補完する関係にある構文的な解析と，部分的な意味解析は，現在の技術ですでに実用レベルに達しているが，文脈解析と，それに基づく要約は，さらに研究を要するものである。

以下に基礎解析サブシステムの各構成機能の概略を示す。

(a)　形態素解析

形態素解析の主なる処理内容には次がある。

- 単語認識
- 品詞判定
- 複合語生成・分割等

英語等のようにすでに分かち書きのある言語においては単語認識は問題とはならないが，日本語においては，分かち書きがないため，その解析には高度の技術が要求される。

一連の解析処理では，活用語辞書，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，利用者辞書を用いて単語の認識を行う。また最近のユニークな研究として，解析のスピードアップを図るために，漢字とひらがなを区別して単語認識する方法がある。

このようにして単語が認識されれば，次に知識ベースである単語辞書よ

り，その単語のとり得る品詞の全てが抽出され，その単語に付されて次の構文解析処理へと渡される。

次に英文による形態素解析の例を図 2.4 － 2 に示す。

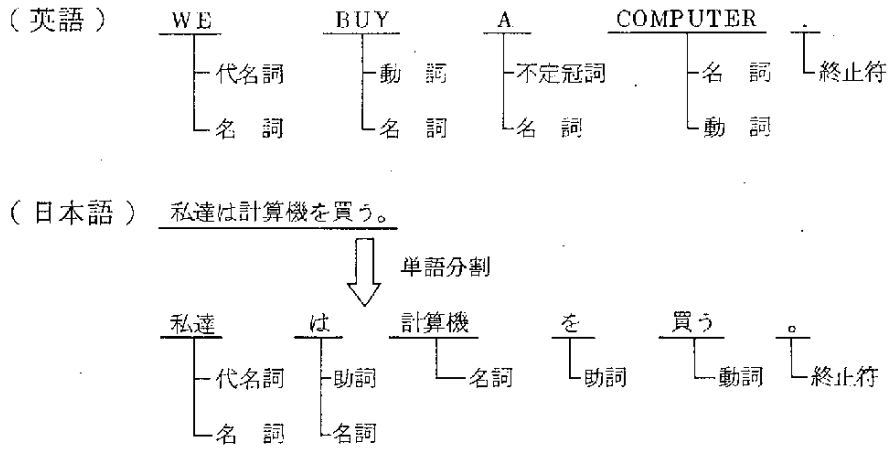


図 2.4 － 2　形態素解析の例

最も単純な二次情報データベースとして，形態素解析しただけの個々の単語をキーワードとして登録する形態も考えられる。この場合，全ての単語がキーワードとして登録されるため，検索結果にノイズが多くなり使いものにならないという心配がある。しかし，その反面検索漏れがないため，文章に忠実な検索結果が得られるという利点もある。ここで論じる二次情報データベースへの情報登録とは，文章の真意を登録することであり，形態素解析された結果を，さらに意味情報で絞って登録することである。次に構文を解析する機能について述べる。

(b)　構文解析

形態素解析によって得られた各単語要素について，それぞれの単語の文法的な役割を明確にし，文章としての構造を決定することが構文解析である。また同時に，形態素解析によって抽出されたそれぞれの単語の品詞を決定することも，この構文解析で行う。しかし，文章の構造の決定においては，一概に形式的側面だけでできるものではなく，意味的側面からのアプローチも必要としており，次のステップである意味解

析と連携して行わなければならないことが多い。

構文解析によって，各単語の品詞，格が決まり，その文においての単語の役割が明確になるため，その役割を利用した二次情報データベースに対する検索の意味精度は高くなる。

構文解析の例を，５８年度に行ったPROLOGによる翻訳実験より図2.4－3に示す。

（英文） WE BUY A COMPUTER.

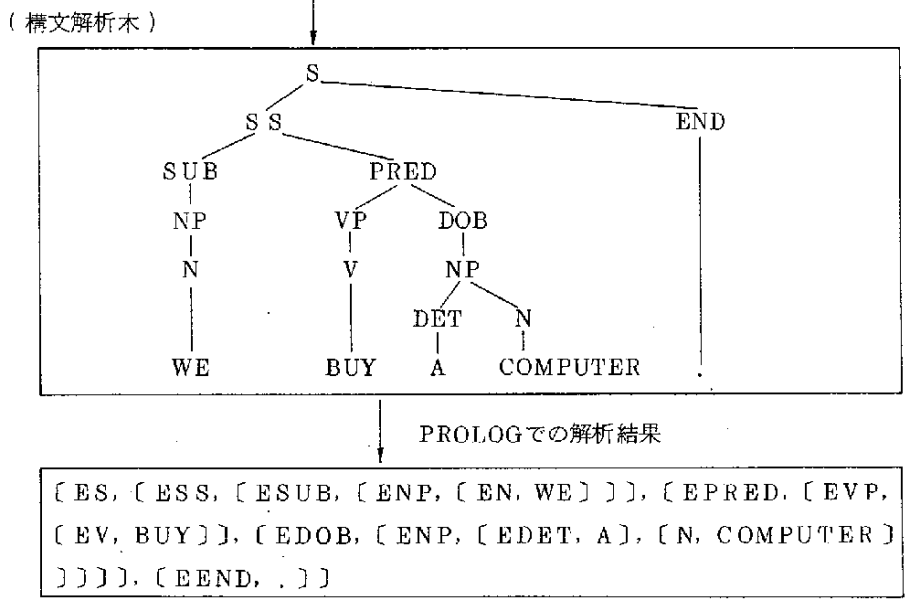


図2.4－3　構文解析の例

(c)　意味解析

構文解析を行っただけでは単に文法的な構造のみが決まったにすぎず，意味的な曖昧さが除去されたことにはならない。また逆に，意味が決まらなければ構文が決まらないことも多い。その良い例で引用されるのは次の文であろう。

TIME FLIES LIKE AN ARROW.

「光陰矢の如し」または「時バエは矢を好む」とも訳せる。しかし

我々が通常会話をしたりまたは文章を読んだり書いたりする上では，このような極端な誤解は起こり得ない。なぜならば文章の前後関係や，単語間との関連で意味を限定し得るからである。つまり，ここでの意味解析も次の２つの機能に分類できる。

・１文内での各単語の意味限定

・複数文にまたがった解釈による各単語の意味限定

① １文内での意味限定

１文内での範囲で曖昧さができる場合であり，文法規則による場合と，単語の意味の関連で意味決定できる。例えば，WE BUY A COMPUTER．を認識する場合において，BUY は「買う」という動詞と，「買物」という名詞の２つの品詞と意味をもつが，文法規則であるＳ＋Ｖ＋Ｏとの関連で，動詞と決定できる。

また，品詞は決定できたが，意味が多数あり，どの意味に決定すればよいかという場合がある。この場合の簡単な例に次がある。

WE HAVE LUNCH．

HAVE は動詞ではあるが，「持つ」，「食べる」，「飲む」，「経験する」，「だます」等，概念の全く異なる意味をあわせ持っている。しかし，この例の場合，目的語が「昼食」という名詞であるため，容易に「食べる」という意味である事が分る。また，目的語が「ボール」であるならば，「ボール」を食べることはできないため「持つ」という意味になる。このように単語の関連によって，多数の意味概念を各単語が持っている場合でも意味が限定され，曖昧さを除去できる場合も多い。しかし，１文内では限定できない場合には，次の複数文の意味関連による意味限定に頼らざるを得ない。

② 複数文の関連による意味限定

先の時バエの例のように，前後の文の解釈により突然時バエの話題になると話しの流れに添わなくなることが起る。このように前後

関係により文の意味が決定されることを，複数文の関連による意味限定という。基本的には，１文内の各々の単語の意味関連と同じであり，それが文の単位になったと考えればよい。

WE SAW MANY TIME FLIES AND AN ARROW.
TIME FLIES LIKE AN ARROW.

この例では，「私達は，多くの時バエと矢を見た」という文があるため，「時バエは矢を好む」と訳すのが正しい。

IT IS TEN YEARS SINCE WE GOT HERE.
TIME FLIES LIKE AN ARROW.

「私達がここに着いてから１０年が経つ」という文と組合せれば当然「光陰矢の如し」となる。意味解析には，この他ＡＮＤの組合せの解釈，関係代名詞のかかり受け，さらに代名詞解釈や省略文の解釈等解決すべき多くの課題を残している。

(d) 文脈解析

意味解析と同様に，複数文との関連によって文の持つ意味が確定されたり，関係代名詞のかかり受けも一種の文脈解析であるともいえる。しかし，ここでいう文脈解析とは，文章全体の意味の流れを把握することであり，さらにその実現技術は高度のものとなる。例えば，文章には起承転結があり，どれが結論であるか等を認識することがあげられる。

文章情報の定量化分析を行う際，文章の中の１文ずつを単位にして分析を行う事はあまり意味がない。従って記事情報であるならば，１記事１文として定量的に取扱うことが要求される。

基礎解析サブシステムの出力である二次情報データベースには，当

然文章情報の定量化分析を可能とするような情報も含まれるべきである。そのためには，文脈解析によって，文章全体の意味の流れを把握することがそのための基本となる。

(e) 要　約

先に述べたように，文脈解析によって文章の意味の流れが把握され，それらを１つの文にまとめることが要約となる。要約は，数多くある情報の中から相対的に重要である情報を選ぶことであり，計算機処理が最も不得意とするところである。機械的にその文章の中で一番多く出現した単語をその文章の中心とすることもできるが，最後の結論は全く異なる意味を論じている場合も多い。

前節にも述べたように，本システムは試験開発においては，要約に対して，まともに取組むことを避け，その文章の中での第一文を要約文とした。

要約した文をどのように二次情報データベース内に格納するかということは，次に続く内容分析サブシステムの機能を限定するため，充分なる議論の上で決められるべきことである。

今までの議論では，次のようにすることが提案されている。

- 要約文を構文解析しそのままの形で二次情報データベースに格納する。
- 要約文を構文解析／意味解析し，５Ｗ１Ｈの情報に集約して格納する。
- ５Ｗ１Ｈでは情報が不足しているため，さらに数種の意味カテゴリを追加する。　等

５Ｗ１Ｈに機械的に集約してしまうことは，これより定量化分析するソフトウェアの開発コストは少なくてすむが，情報の欠落を伴うため，木目細かな定量化分析は困難となる。逆に，情報をあまり集約せず，そのままの形であると，定量化分析サブシステムの開発コストが増大する。つまり，この相反する２つの要求を，どこでトレードオフ

するかがシステム化の重要な鍵でもある。

(3) 二次情報データベースとの関連

基礎解析を行った結果は，二次情報データベースに格納される情報は，それを利用する各サブシステムの処理が容易にしかも効率良く行われることを考慮して決定される。基本的には基礎解析の各機能の処理結果から得られる情報である。その主なものを次に示す。

(イ) 形態素解析処理

- 各文の区切り
- 記　号
- 単　語
- 単語の品詞
- 複合語
- ページの区切れやラインの単位（改頁や改行の度合）
- 大文字，小文字，見出し文字（文字の形態）
- 文章の位置（例えば新聞の第何面）

(ロ) 構文解析処理

- 格
- 文構造

(ハ) 意味解析

- 単語の意味
- シソーラスを用いた意味カテゴリ

(ニ) 文脈解析

- 代名詞の意味
- 文章の意味構造の把握（例えば起承転結）

(ホ) 要　約

- 文脈解析からの結論文の抽出
- 文章意味カテゴリへの対応づけ（例えば5W1H）

(ハ) その他

その他の情報としては，その文章の中からは抽出できない情報を，人為的または機械的に外から与えるものである。例えば次のものが考えられる。

- 文章入力年月日，および時間（機械が自動的に付与することが可能）
- 文章あるいは記事番号（自動付与も可能）
- 出典（入手先）
- 著　者

等があげられる。

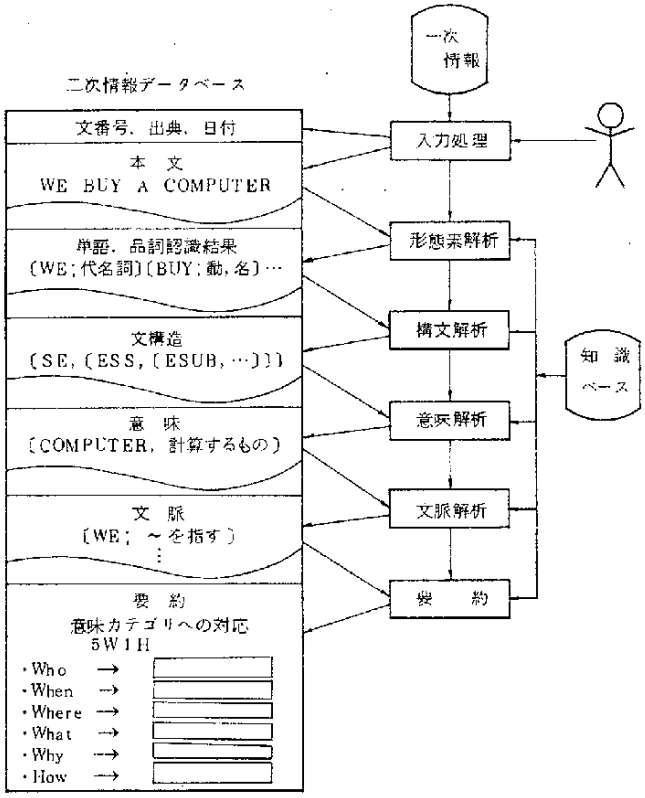


図2.4－4　二次情報データベースとの関連

以上の情報と，基礎解析サブシステムとの関連を図示すると図2.4－4のようになる。

図2.4－4に示されるように基礎解析サブシステムでの処理結果は二次情報データベースに格納される。これらの結果は，データベースに格納されるため，種々のシステムからの共有および多様な利用が可能である。例えば，形態素解析されて認識された単語は，そのままキーワードとして，従来の情報検索と同様な検索を可能とする。

また，構文解析，意味解析，文脈解析された結果は，そのまま翻訳サブシステムの入力となり，翻訳サブシステムでの解析処理を不要とする。翻訳サブシステムでは，単に目的語への生成をのみ行えばよいことになる。さらに要約された情報は，内容分析サブシステムに渡され，効率の良い分析を可能とする。このように従来人手を多く要していた文章情報に対する前処理が，基礎解析サブシステムによって機械化され，今後大量の文章情報の処理を可能にしていくものと思われる。

2.4.2 二次情報データベース

(1) 位置づけ

二次情報データベースに，日々刻々に発生する国内外の新聞・雑誌などの文章情報を収集し，その中から的確・迅速に必要な情報を得たり，潜在的な傾向や徴候を見いだすことを目的としたデータベースシステムを確立するための，文章情報データの総合的な格納・蓄積形態である。

文章情報を一次情報のままでマイクロフィルムや光ディスク等に移し替えて蓄積利用することも可能であるが，日々蓄積されるに従ってやがて膨大な量となり人手による処理能力の限界に達してしまうであろう。そこで，大量な蓄積データの中から必要なときに必要な情報を得るためには，コンピュータにより情報内容を色々な角度から利用できるように高度な内部形式に圧縮加工して二次情報データベースとして蓄積しておく必要がある。

文章情報の圧縮・加工の方法としては，大きく次の3つの方法について研究，開発あるいは実用化されてきた。

① 索引づけ…文章の意味内容を代表するいくつかの重要な語句をキーワードとして選択し，索引とする。

② 抄録化…文章の意味内容を要約して文（章）で表現する。

③ 事実抽出…文章の中で述べられている重要な事実・傾向などを抽出し内部形式で意味表現する。

この内①，②の作業については一般には専門家（インデクサーなど）に

より手作業で行われているが，方法論は確立しているものの人手による作業では膨大な情報量に対して多大な費用・期間を要し，かつ個人差などによる精粗のばらつきの問題があり，コンピュータによる支援・自動化が望まれるところである。

コンピュータによる文章情報の圧縮・加工の方法は，これまでに多くの試みがなされてきたが，ここにきてコンピュータのメモリ，スピード，コストパフォーマンスが飛躍的に向上し，加えて関連技術の急速な進展により実用化の可能性が高くなってきたといえる。更に，近年文章情報がコンピュータ出力で提供されるようになり，かつＯＣＲの普及と合わせて，一次情報のコンピュータ入力が容易になってきたといえる。コンピュータによる自動索引，自動抄録の技術は今だ完成されたものの，このようにコンピュータ可読となった一次情報を基に，現在利用できる技術を可能な限り取り入れ，コンピュータの壮大な力を利用すれば実用可能性が高いといえる。また③については最近の機械翻訳システムの開発などにみられる高度な自然言語処理技術を応用して，一次情報を構文解析／意味解析し，その結果を基に事実データを抽出し意味表現の形に変換し，二次情報データベースに蓄え，文章情報の高度利用を図ることも重要である。このようなコンピュータによる一次情報の圧縮・加工の方法の実現は，初めは限定した手法で，しかも限られた適用分野については，人手による作業をコンピュータが支援する方法（又はその逆）を採り，将来技術の進展をみながら段階的にコンピュータ側に移行していくやり方が現実的であり，肝要と思われる。一次情報の圧縮・加工工程の機能については 2.4.1 基礎解析の項で詳しく述べられている。

以上のように，収集された一次情報はデータの利用に備えて，基礎解析機能によって一元的なデータ構造に圧縮加工されて二次情報データベースに蓄積される。

(2) 二次情報データベースの構造

二次情報データベースに格納する情報構造は，大量文章情報の蓄積利用

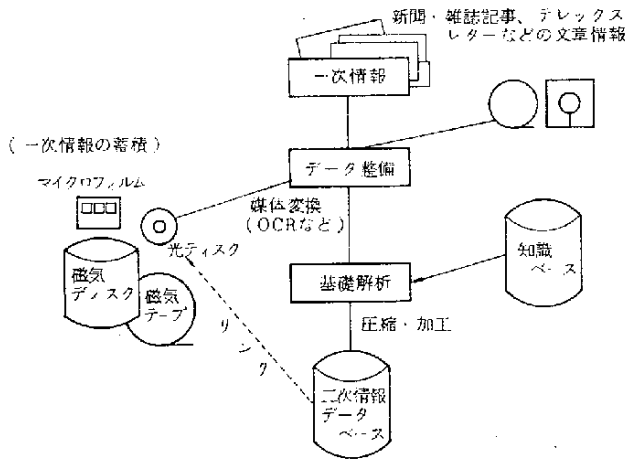


図2.4－5　一次情報から二次情報データベースへの変換

が目的であるので，一次情報の意味内容を可能な限り正確に反映するものであり，かつ利用者の文章情報に対する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。

一次情報から圧縮加工された二次情報の意味内容の正確度は基礎解析機能の性能に負うところが大きいが，情報の利用の側面からみると次に示すように様々な属性が要求され，これらの属性を一元的に表現できる総合的な情報構造を具備する必要がある。

(a)　検索利用の側面から

書誌的情報，検索を効率的に行うためのキーワード，検索結果の内容表示のための抄録，本文，及び一次情報の保管アドレスなどが要求される。

(b)　定量化分析利用の側面から

統計的分析手法などによる内容分析のためには，文章や文中に出現する重要な語句（キーワード）及びその出現属性が取り出される情報構造が要求される。

また，文章や文を対象とした意味論的分析のために文章レベルや文レベルでの意味内容を可能とする情報構造も要求される。

(c) 翻訳利用の側面から

文章情報データベースの検索・分析結果を利用目的に応じた言語に翻訳生成するために，言語解析済みの中間言語構造に変換されていることが好ましい。

更に，二次情報をこのように多面的に利用するために，二次情報のデータ構造に対応した辞書や諸規則などの知識ベースを装備する必要がある（2.4－6参照）。

以上の要件を考慮した二次情報データ構造を概略次に示す。

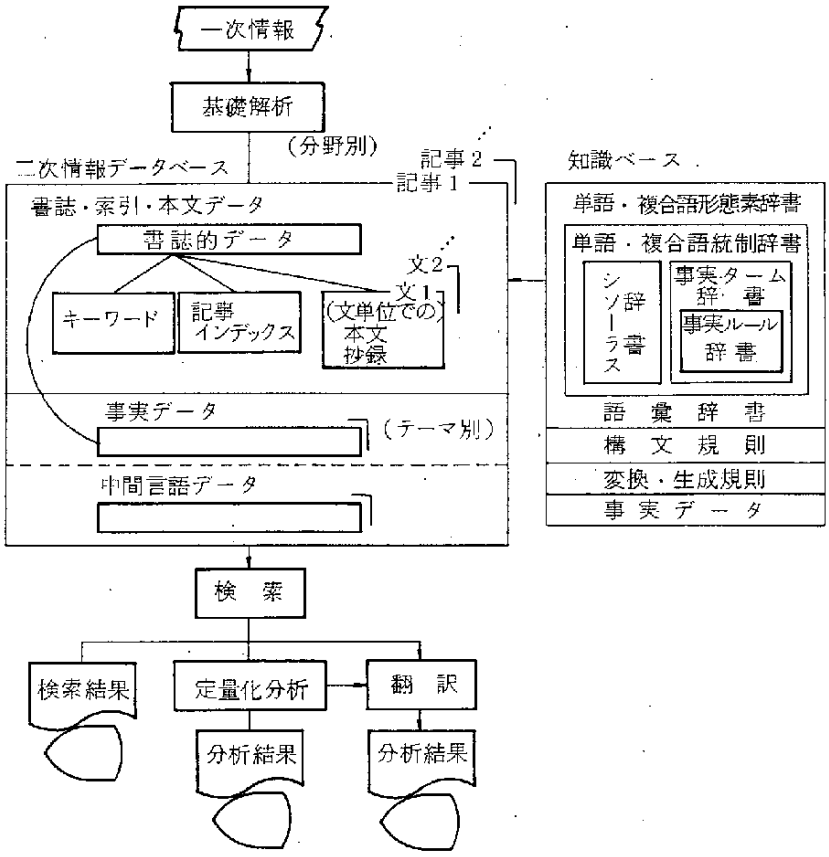


図 2.4－6　総合解析システムからみた二次情報データベースの概念

(3) 二次情報データの構造

二次情報データは，大きく次の4つの構造に分けられる。

- 書誌的情報
- 索　引
- 本　文
- 事実データ

これらの構造による二次情報データベースを総合利用の観点から示すと次のようになる。

(a) 書誌的情報

下図のように確定した属性の情報であるが，非定型な一次情報の中では一様でないのでデータ整備工程で人手により入力される。

イ	ロ	ハ	ニ	ホ	ヘ	ト	
データ番号	発行日付	紙誌名（略号）	紙誌内ページ番号	記事行数	入力日付	一次情報保管アドレス	

イ．データ番号

文章情報を二次情報データベースに格納する際に付与される整理番号であり，他の記事との識別あるいは一次情報との対応などに用いる。データ番号はコンピュータで自動的に決定する方が重複などを防止できるが，一次情報との対応をきちんとしておく必要がある。

ロ．発行日付

情報が記事として掲載された紙誌の発行日付である（テレックスであれば受信日付）。

ハ．紙誌名（略号）

情報が記事として掲載された紙誌名の略称である。表2.4－1にその一覧を示す。

表2.4－1　紙誌名一覧

略号	正式名	国籍	略号	正式名	国籍
FER	Far Eastern Economic Review	香港	IGR	International Gas Report	英国
AFR	Australian Financial Review	オーストラリア	PEN	Petroleum News	香港
AOG	Arab Oil and Gas	レバノン	MER	Manthly Energy Reviw	米国
APS	Arab Press Service	キプロス	MEED	Middle East Economic Digest	英国
A W	Alchol Week	米国	MEES	Middle East Economic Servey	キプロス
AWJ	Asian Wall Street Journal	〃	N F	Nuclear Fuel	米国
B D	Biomass Digest	〃	NUW	Nucleonics Week	〃
B M	Bulletin Mensuel (Comite Professionel du Pitrola)	フランス	N W	Newsweek	〃
			OAB	OAPEC News Bulletin	クウェート
B W	Business Week	米国	OBG	Oil Buyer's Guide	米国
CEN	Chemical & Engineering News	〃	OGJ	Oil and Gas Journal	〃
CHW	Chemical Week	〃	OPB	OPEC Bulletin	オーストリア
C N	Coal News	〃	P D	Prauda	ソ連
C W	Coal Week	〃	P E	The Petroleum Economist	英国
CWI	Coal Week International	〃	P I	Petrole Information	フランス
ECO	Economist	英国	PIW	Petroleum Intelligence Weekly	米国
E D	Energy Daily	米国	P N	Platt's Oilasam News Service	〃
E G	Ekonomicheskaja Gazeta	ソ連	P P	Platt's Oilgram Price Service	〃
EOP	Europe－Oil Prices	英国	P T	Petroleum Times	英国
E P	Europe	ベルギー	S C	Science	米国
EUR	Energy Users Report	米国	S F	Syn Fuels	〃
E W	Electrical Utility Week	〃	T W	Time	〃
F A	Foeign Affairs	〃	USN	U.S. News and World Report	〃
F T	Financial Times	英国	WFM	World Financial Market	〃
GER	Gas Energy Reviw	米国	WGR	World Gas Report	英国
GOI	German Oil Information Service	西独	W J	Wall Street Journal	米国
H B	Handelsblatts	〃	W O	World Oil	〃
IHT	International Herald Tribure (New York Times Washington Post)	フランス	PTP	Petroleum Time Price Report	英国
			E E	European Energy	〃
IPT	International Petroleum Times	英国	OET	Oil and Energy Trends	〃
L M	Le Monde	フランス	EUW	Electric Utility Week	アメリカ
L T	The Time	英国	P R	Petroleum Review	香港
AMS	APT Monthly Statistical Report	米国	AMS	API Monthly Statistical Report	米国
PSM	EIA Petroleum Supply Monthly	〃	I O	IEA Oil Market Report	フランス

ニ．紙誌内ページ番号

情報が紙誌のどのページに掲載されていたかを示す。（開始ページ）

ホ．記事行数

情報の量の目安を示すもので，掲載された紙誌面で実際に何行を占めていたかを表示する。

ヘ．入力日付

情報が二次情報データベースに格納された日付を表わす。

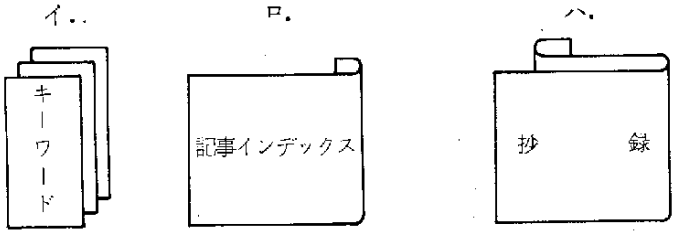
ト．一次情報保管アドレス

一次情報が保管されている媒体のアドレス及び媒体内のアドレスを表わす。二次情報データベースを検索の結果，該当する情報について一次情報も参照したいという要求に答えるためのものである。

更に，文章情報の種類や利用目的に応じて前記以外にも，タイトルなどの確定情報が必要となろう。

(b) 索引（インデックス）

文章情報の蓄積・検索効率を高めるために一次情報としての文章の意味内容を圧縮・加工したもので，次の3つに分けられる。



イ．キーワード

一次情報の文章を構成している語句の中で文章の重要な意味内容を代表している語句をいくつか選択して，それを統制語によってその文章の属性値としたものである。キーワードは統制語辞書あるいはシソーラス辞書によって体系化されているので，その語句に対応して付けられたコードで表記した方が処理効率的に好ましい。

以下に，選択されたキーワードの例を示す。

＜キーワードの例＞

米アパラチア産の需要は天然ガス価格規制撤廃がカギ――90年までに倍増も　811019　CW　004
（海外団体）　米国政府　（品目）　石炭　天然ガス　（項目）　規制撤廃F　需要見通しF　エネルギー政策F　価格動向F　（海外地域）　米国　（抵抗名略号）
米連邦エネルギー規制委決定、石炭市場・投資などの定義が不明確との批判も　811019　EW　009
（海外団体）　米国政府機関　（品目）　石炭　（項目）　エネルギー政策F　（海外地域）　米国　（抵抗名略号）　EW
インドネシア，外国企業と共同で石炭開発――狙いは石油埋蔵力の温存　811012　AWJ　003　150
（海外会社）　アトランティック・リッチフィールド・インドネシア　アジップ　コンジンク・リオ・チント・オブ・オーストラリア　ブリティッシュ・ペ
（海外会社）　○エタ・エクスプロレーション　（海外団体）　インドネシア政府　（品目）　石炭　石油　（項目）　エネルギー政策F　エネルギー
（項目）　埋蔵F　（海外地域）　オーストラリア　英国　インドネシア　米国　（抵抗名略号）　AWJ
米会計検査院、緊急時における石炭への転換計画の実効性について批判　811005　CW　002
（海外団体）　米国政府機関　（品目）　石炭　（項目）　エネルギー政策F　（海外地域）　米国　（抵抗名略号）　CW
レーガン米政権のエネ省廃止転換局廃止案、石炭転換計画企業に大きな打撃　811005　EW　001
（海外団体）　米国エネルギー省　米国政府　（品目）　石炭　（項目）　エネルギー政策F　政府財政援助F　（海外地域）　米国　（抵抗名略号）
米内務省、カーター前政権が禁止した国立公園近くの石炭開発を再検討　810922　WJ　009　69
（海外団体）　米国内務省　（品目）　石炭　（項目）　エネルギー開発F　環境政策F　エネルギー政策F　（海外地域）　米国　（抵抗名略号）
米EPA、石炭発電プラントなどの排ガス規制基準変更の審査時間を半減へ　810914　CW　002
（海外団体）　EPA　（品目）　石炭　排ガス　（項目）　公害規制F　エネルギー政策F　（海外地域）　米国　（抵抗名略号）　CW
米関係筋、石炭スラリー輸送問題で対立――決定はホワイトハウスに一任　810914　CW　001
（海外団体）　米国政府　（品目）　スラリー輸送　石炭　（項目）　エネルギー政策F　（海外地域）　米国　（抵抗名略号）　CW

• エネルギーデータベースより
• [　　] 枠内がキーワード

ロ．記事インデックス

一次情報の文章内容から重要な事実や観測情報を１行の文で表現したものである。二次情報データベースを検索した結果 該当した一群の記事を一覧表示して概観する場合に有効である。また(d)で述べる事実データの内部表現に対応した表層表現ともいえるものである。

＜記事インデックスの例＞

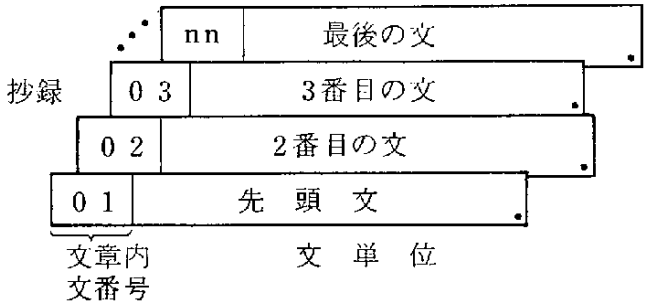
サウジ、ＬＰＧ増産のため投資と中質原油の生産割合を７０対３０へ　830330　APS　04
サウジアラビア政府、予算縮小でプロジェクト延期を検討　830325　ED　04
サウジのアブドゥラ皇太子、ＯＰＥＣ加盟国に協定遵守を要請　830323　FT　00
原油の値下がりで、サウジは投資計画を３億４４００万ドル削減　830323　WJ　033
サウジ・インタナショナル・バンク、８２年の利引利益は３１％増　830322　LT　020
西独、石油業者の８３年原油輸入計画――北海原油の輸入減少　830318　EOP　004
ＢＰ、ナイジェリア原油や基準原油価格をみて北海原油価格の微調整を示唆　830318　FT　016
サウジアラビア、アラビアン・ライトは２月１日現在で３０ドル　830317　LT　001
サウジアラビアの狙い、原油価格の値下げに耐えられるか　830315　FT　004
アラブ「友好国」のサウジアラビアへの批判高まる　830315　WJ　029
サウジアラビア、開発支出削減を計画　830314　MEES　003
サウジアラビア、ジュベールの４プロジェクトの定額式行う　830314　MEES　004
ＯＰＥＣロンドン会議の狙いはサウジの生産政策修改革　830314　FT　001
湾岸アラブ諸国、石油収入減少を平静に受け止める　830309　IHT　001
サウジアラビア、ＯＰＥＣ会議の会意でイランに接歩か　830307　IHT　009
サウジアラビア、日本の国債市場での動きは慎重　830307　FT　018
サウジアラビアとクウェート、原油値下げで大幅財政赤字　830304　LT　015
サウジ首相、国王と皇太子の反目説否定――軍事費支出は米国一辺倒から離脱　830302　APS　097
湾岸産油国、４ドルの値下げを認めないなら７ドルの値下げも辞せず　830228　LT　017
４ドルの値下げも認めないなら、湾岸諸国は７ドルの値下げも辞さない　830228　FT　001

記事インデックス　（エネルギーデータベースより）

ハ. 抄　録

解説文，論文記事など比較的量の多い文章情報から重要な意味内容を要約し，数行の文で表現したものである。最近，論文などでは，抄録を必ず掲載している紙誌もあるが，抄録がない場合にはデータ整備工程で専門家により作成されることになろう。なおコンピュータによる自動抄録機能に関して現状では技術的に未完成であり，自動化よりもむしろ専門家の作業をコンピュータ側で支援する方式の検討が現実的であろう。

入力された抄録は，基礎解析工程の形態素解析機能により文単位に分割されて文章内文番号が付されて二次情報データベースに格納される。



また一次情報としての抄録文は基礎解析サブシステムでのキーワード抽出機能の入力の対象にもなる。

二次情報データベースに格納された抄録データは，検索結果該当した個々の記事内容の概要把握用に表示出力される。必要があれば他言語への翻訳出力の対象ともなる。

(c) 本　文

新聞記事など一文章当りの量が比較的少ないものについては本文全体が基礎解析サブシステムの形態素解析機能により文単位に分割されて文章内の文番号が付されて，二次情報データベースに格納される。

一次情報としての本文は抄録文と同様，基礎解析工程でのキーワード抽出機能の入力の対象となる。

二次情報データベースに格納された本文データは，検索結果該当した個々の記事内容の詳細検討のために，一次情報をそのまま編集した形式で表示出力される。必要であれば目的言語への翻訳出力の対象ともなる。

(d) 事実データ

文章情報データベースをキーワードやシソーラスにより確率的に検索して利用するだけでなく，一次情報の意味内容から重要な事実や観測を内部構造に変換し，分析利用するための構造である。

すなわち，文章あるいは文が述べている事実や観測を次の例に示すように，主体・動作・対象などの概念関係を内部表現で構造化したものである。一般にこのような意味表現の枠組としては，フレーム，意味ネットワーク，論理，プロダクション表現など，何らかの意味で述語論理に基づき，概念を階層構造や概念同志の参照関係などを記述する構造を用いている。ここでも，現在多方面で用いられている述語論理を意識しつつ，現行のデータベース構造になじみやすい，次のようなノードとそれを互いに結ぶアークとから成るネットワーク表現のデータ構造としている。

- ○構造のエントリな記事（データ番号）をルートノードとする。
- ○記事ノードから分岐したアーク及びノードで構成するネットワークにより事実内容を表現する。
- ○各ノードは1つの概念を表わすキータームとして，それをリンクしているアークの指示する格の種類の属性値（データ）を持つこ

とができる。

① ノードとその属性値

各ノードはアークに指された種類の概念の格で以下に例示するような属性値をもつ。これらの属性値は通常のキーワードと基本的に同じであり，統制語や意味概念により概念的に整理された属性値である。各ノードはアークによる指示により任意の格を属性値として持つことができる。

（例）ノードの属性値

[demand，要請]
[observe，順序]
[backdate，さかのぼって実施]
[face，直面]
[unite，結束]
[avert，回避]
[accept，受諾]
[discuss，会談]
[respond，対応]
[inevitanble，不可避]
[continue，継続]
[open，開放]
[raise，上昇]
[support，支持]
[adopt，採用]
[no intension，意図無し]
[devision，分裂]
[insistence，主張]
[cut，値下げ]
[decrease，減少]
[drop，下落]
[observation，]
[delay，遅れ]

[udi，サウジアラビア]
[EC，OPEC]
[uwait，クウェート]
[lf states，湾岸諸国]
[ell-placed source，消息筋]
[rab oil producers，アラブ産油国]
[enezuela，ベネズエラ]
[.S.oil companies，4大アメリカ石油会社]
[lgeria，アルジェリア]
[Shell，シェル石油]
[the Gulf Co-operation council，湾岸諸国会議]

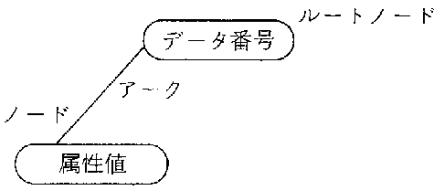
なお，ノードがもつ属性値として上記のような対象，事象，概念の他に，属性値を決定するための手続き（プログラム）の場合も考えられよう。

② アークの種類

アークはあるノードが支配する別のノードを指示し，その概念を表わし，事実データの構造を形成する。アークの種類としては次のようなものが考えられる。

（例）　アークの種類

[agent，主体，が]
[act，動作，]
[obj1，相手，に対して]
[obj2，対象，を]
[occup，役職，である]
[name，名，という名の]
[cause，原因，の原因で]
[purpose，目的，の目的で]
[obs，観測，と観測]
[speaker，話者，によると]
[time，時，に]
[value，値，]
[of，of，の]
[on，on，について]
[with，with，と]
[and，and，及び]
[mod，修飾，]



以下に，本文から抽出した事実データの例を示す。

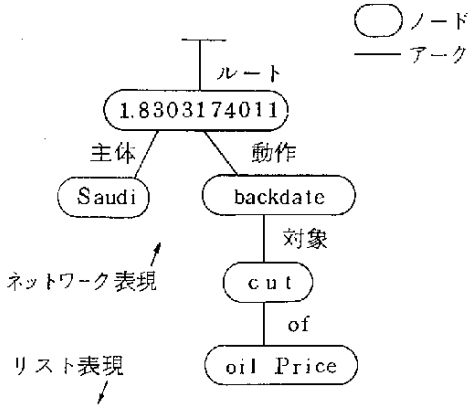
このように事実データの構造は事実表現のための固定的なものとしてとらえるのでなく，むしろ適用分野や利用モデルなどの対象を中心に決定されるものとすべきであろう。

（例1）

記事インデックス：サウジアラビア，アラビアン・ライトは2月1日実施で30ドル

本　文：

Saudi Arabia's $5 a barrel oil price cut will be backdated in two stages to the beginning of last month, it was reported yesterday as the price of North Sea oil continued to be in limbo after this week's agreement by the Organization of Petroleum Exporting Countries.



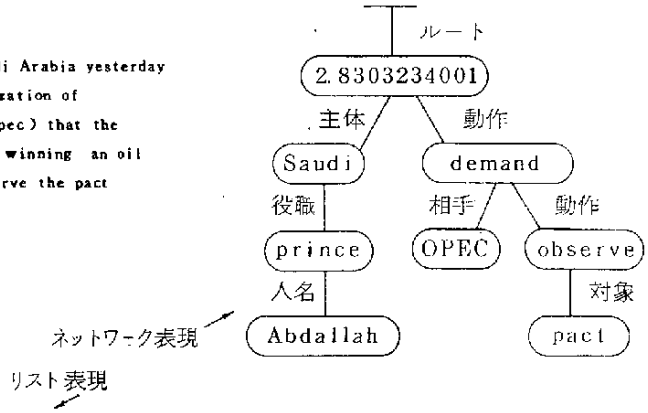
1.8303174011〔〔主体，サウジアラビア〕，
〔動作，さかのぼって実施，〔相手，OPEC〕，
〔対象，値下げ，〔of，原油価格〕〕〕〕

（例２）

記事インデックス：サウジのアブドゥラ皇太子，OPEC 加盟国に協定遵守を要請

本文：

CROWN PRINCE Abdullah of Saudi Arabia yesterday warned other members of the Organization of Petroleum Exporting Countries (Opec) that the Kingdom was " perfectly capable of winning an oil price war " if they failed to observe the pact concluded last week in London.

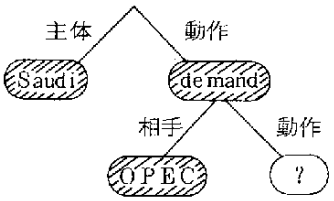


2.8303234001〔〔主体，サウジアラビア，〔役職，王子，〔人名，Abdullah〕〕〕，〔動作，要請，〔相手，OPEC〕，〔動作，守る，〔対象，協定〕〕〕

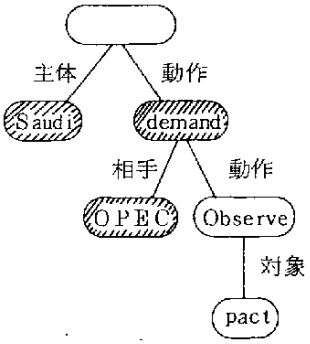
③　事実データの検索の仕方

事実データの検索は，検索式で指定する部分構造と事実データ中の事実構造の集合とのパターン照合により行われ，一致した事実構造のデータが取り出される。

例えば，前記の例２の事実構造に対して「サウジアラビアがＯＰＥＣに対して要請した事柄（動作）は何か？」の質問に対して

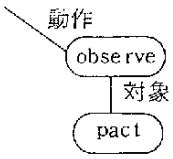


質問の概念



事実データ

質問概念のすべてのノード，アーク（////// 部分）が一致するので，目標とする？の対象として，事実データから，



すなわち"協定の遵守"が回答として取り出される。

2.4.3 検　索

文章情報は，数値情報より広い範囲の情報を持っており，その様な多くの情報の中から必要な情報を取り出すことは，全ての情報を見ることにもなりかねない大変な作業である。その為に，それらの情報をキーワード付け，抄録化等の加工をし，それらを利用することによって，判断材料として有効な情報をできるだけもれなく，しかも早く取り出すことが必要になる。今，述べた情報の加工は，基礎解析で行われ，その結果が二次情報データベースにキーワード付けした情報，抄録化した情報，事実情報，記事の書誌索引情報として出力される。

検索システムの役割は，この二次情報データベースを入力し，文章情報データベースの利用者が必要な情報をより迅速に，容易に抽出できるようにすることである。その為に，準備するとよい機能として以下の５つが考えられる。

- キーワード検索

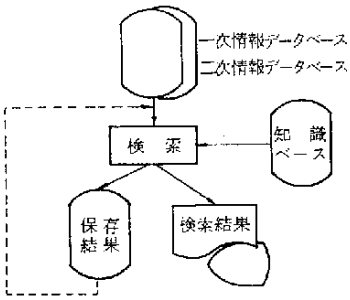
・フリーテキスト検索

・シソーラス活用検索

・事実検索

・頻度検索

これらの機能は，それぞれでも情報の抽出に役立つが，いくつかの機能を合せて利用することによって，検索もれを少なくしたり，調査分析したいテーマの客観的な把握，将来動向の予測，仮説の検証等を行うのに有効なデータも提供できる。また，検索結果を保存する機能をもたせて，再検索の時の作業効率を高める必要もある。



次に，前述の５つの検索方法の機能と利用イメージについて述べる。

(a)　キーワード検索

文章情報に付加された書誌的情報や本文の抄録を示すために付加された項目に対応したキーワードを利用者が入力して，必要な文章情報を取り出す検索方法である。この方法は，大量の文章情報の中から利用者が必要としている情報の量を絞り込む際に有効である。キーワード間の論理検索を考慮することによって，さらに効率よく検索できるこの検索では，抽出した情報の主体，目的，対象等の内容に関する事柄については，利用者が独自に分析する必要がある。例として，５８年度の「世界エネルギー情報のＯＰＥＣ加盟国を中心にした分析」で使用したキーワードを掲げておく。

表 2.4 − 2

	エネルギー資源	エネルギー関連	生産関連	価格関連	経済関連	軍事・政治	採鉱・設備	産油国関連
1	石　油（石油）	エネルギー政策（エ政）	生産計画（生計）	価格政策（価政）	経済動向・経済見通し・経済計画（経動）	兵器・国際紛争・戦争（戦争）	石油採鉱・油田（油田）	OPEC（OP）
2	原　油（原油）	エネルギー外交（エ外）	生産動向（生動）	価格動向（価動）	経済外交・経済協力（経外）	反政府運動・クーデター・ゲリラ・過激派暴動（反政）	製油所・パイプライン（製油）	石油会議（会議）
3	石油製品（製品）	エネルギー開発（エ開）	生産見通し（生見）	価格見通し（価見）	インフレ（イン）	政権交代・政治・政治体制（政治）		産油国（産国）
4	天然ガス・LNG・LPG（ガス）	エネルギー問題（エ問）	増　産（増産）	値上げ（値上）	オイルマネー（マネ）			石油消費国（消国）
5	電力・原子力（電力）		減　産（減産）	値下げ（値下）	対外債務・借款（債務）			石油業界・メジャー（業界）
6	石　炭（石炭）			価格交渉（価交）	産業政策（産政）			
7				プレミアム販売（プレ）				
8				商品市況・市場動向（市況）				

（）内はキーワードの略号

(b)　フリーテキスト検索

利用者が文章情報上のキーワード項目とは無関係にキーワードを設定して検索したり，検索もれを少なくしたい時に利用するものであって，全文検索，文字列検索といわれるものである。これは，単語（複合語，派生語も含む）のレベルで具体的な項目について調査・分析したいときに便利である。例えば，ある特定の商品に関することを調査したい場合にはその商品名で検索する。

(c) シソーラス活用検索

知識ベースに用意されたシソーラスを使って，利用者が設定したキーワードとそれに関連した情報を取り出す検索方法であり，論理検索のひとつである。キーワード検索より幅の広い範囲での情報の収集を可能にし，内容分析を行う為には，有効な検索方法である。例としては，最近の石油製品の価格傾向を調査したいとき，石油製品のシソーラスを利用することが考えられる。

(d) 事実検索

利用者が検索項目をキーワードとしてではなく，主体，目的，対象，時制，述語等の事象の事実表現（例えば５Ｗ１Ｈ）に近い形式で与えることによって，その事象に関連したり，それに近い事実情報を取り出そうとする検索方法である。例えば，レーガン大統領が，日米通信摩擦について発言したことがあるか，また，いつ頃かというようなことを知るのに便利である。今年度作成したプロトタイプシステム（詳細は，2.3参照）は，この検索を可能にしたものである。

主体として「サウジアラビア」,対象として「ＯＰＥＣ」,目的として「協定（価格）の維持」，動作として「要請」を入力することによって，「サウジアラビアがＯＰＥＣに協定の維持を要請した。」という８３年３月２３日のフィナンシャルタイムスの記事を取り出すことができる。

さらに，将来的なこととしてある事象の予測を行うことも可能である。それは，その事象をいくつかの事実データの時間的組み合せとして表現（シナリオ化）し，最新の二次情報データベースの事実データとマッチングすることで可能となる。

(e) 頻度検索

キーワード検索によって抽出された結果に検索項目の出現頻度数を加味して出力し，それらの統計情報として，簡単なクロス集計データ，主成分分析，回帰分析結果も合わせて出力するようにしたものである。例としては，５８年度実施した「世界エネルギー情報によるキーワード出現

頻度分析」があげられる。

2.4.4 定量化分析

(1) 位置づけ

二次情報データベースに蓄積した大量の情報群が示す意味や，その中から潜在的な徴候あるいは傾向を把握したり，将来動向の予測・仮説の立証のための素材を得るなど，文章情報の高度な利用要求に対して，文章情報を数量化して統計的解析手法などを用いて内容分析する機能である。

(2) 機　　能

定量化分析の機能を表 2.4 − 3 に示す。文章情報の数量化分析の方法として，第 1 番目に文章中に出現する語句（キーワード）の頻度を分析する方法である。これはキーワードの出現頻度を検索機能で求め，時系列的に出現頻度を分析したり，キーワードの同時出現を分析して相関を調べたり，更に多次元の角度からキーワードを数量化し，表 2.4 − 4 に示す多変量解析手法により分析する機能などがある。

この場合，キーワードの出現頻度属性だけを基に解析するのでは，主題に対する強度（＋，−）や事実か観測かの判別があいまいであり，分析結果の精度を落すことになるので，このような場合，主題に関連する事実データで補足する必要があろう。

定量化分析の第 2 番目の手法としては，意味決定，政策決定などのプロセスをモデル化し，そのモデルに数量化した文章情報を組み込み，問題点の解析手段を見い出すシミュレーション手法がある。

この場合の手法では，カントーリスク分析，通商問題分析，エネルギー問題分析など問題分野別にモデルが作成され，相手からの文書や発言をモデルに入力し，当方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立案や決定，企業の経営計画の策定などを行うための有力な意志決定支援システムとして機能することをめざしている。

（昭和５７年度報告書 4.2 −「ＣＡＣＩ社の国際紛争データを用いた中国監視システム」等を参照のこと。）

表2.4－3　定量化分析システムの機能

サブシステム名	機　能　名	機　能　内　容
頻　度　検　索	1. 時系列頻度数	・検索対象となるファイルは，オリジナルなデータベースの他，検索機能で絞り込んだ2次保存ファイルも含む。 ・時系列の期種は月次・四半期・年次・期間・期種と指定して時系列のキーワード出現頻度数を検索する。
	2. クロス頻度数	・検索対象となるファイルにおけるキーワードの共出現頻度数を検索する。
	3. 全キーワード頻度数	・検索機能で絞り込んだ2次保存ファイルを対象として含まれる全キーワードと頻度数をリストアップする。 ・2つの保存ファイルを比較して新しいキーワードの出現，頻度数の変化を分析する。
頻度数データ加工	1. 計算機能	・伸び率，構成比，頻度数間の四則演算をマトリックスの行・列単位で行い，加工データを作成する。
	2. マトリックス部分抽出	・分析用にオリジナルのマトリックスから任意の行あるいは列を抽出し，部分マトリックスを作成する。
	3. マトリックス結合	・異なった検索で得られた頻度数マトリックスから，行あるいは列の内容に注意して結合マトリックスを作成する。
頻度数データ保存	1. データ保存	・分析用の頻度数生データあるいは加工データを作業中の中間ファイルあるいは保存用ファイルに蓄積する。
定　量　化　分　析	1. 主成分分析（次元縮小）	・時系列あるいはクロスの出現頻度数に基づいて，キーワード間の独立な主成分を求め，キーワードの出現頻度数情報をより少ない成分で説明する。
	2. クラスター分析（分　類）	・クロスの出現頻度数に基づいて，各種キーワードをカテゴリー化，分類する。
	3. 回帰分析（相関関係・予測）	・時系列あるいはクロスの出現頻度数分析に基づいてキーワード間の相関関係を分析する。
グ　ラ　フ　出　力		・キーワードの出現頻度数の生データ，加工データを折れ線グラフ，棒グラフ，レーダーチャートなどで出力する。

これらのモデルに入力する定量化データの作成は人手による非常に緻密な作業を要するが，二次情報データベースに格納されている事実データを検索，分析して作業の素材として利用することが考えられる。

以上，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し，分析結果が意味する傾向などを視覚でも分かるようにしておくことが望ましい。

表 3.4－4　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Ⅰ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		

次に，前述の統計的手法による各種の定量化分析システムの機能について概要を述べる。（詳細については昭和58年度報告書2.3参照）

(a)　キーワード出現頻度数の時系列分析

キーワードの出現頻度数を大小順に並べて話題の推移を時系列に分析し，様々な事象の徴候的な変化を大づかみにとらえることにより，より詳しく調べるためのポイントを得るための機能である。変化の様子を直観的に，容易に把握できるように図2.4－7に例示するようにレーダチャートなどのグラフで出力される。

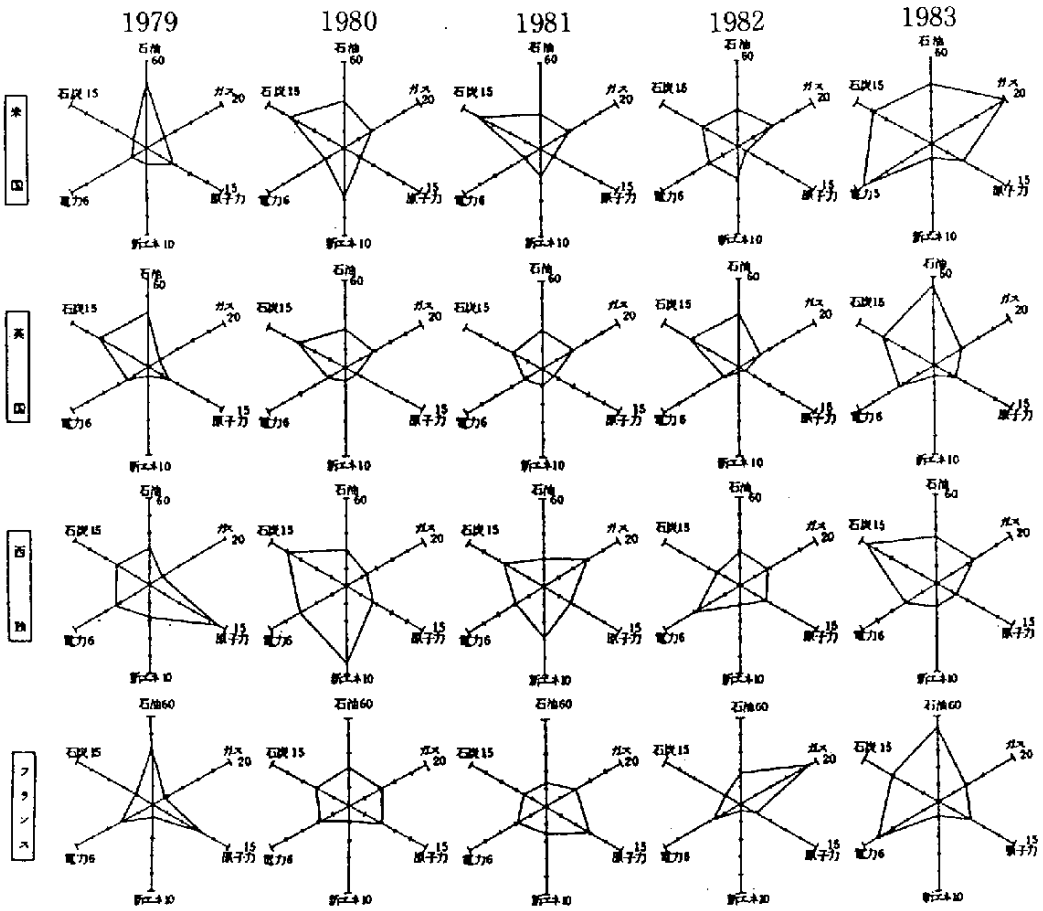


図2.4－7　キーワード出現頻度数の時系列分析

—先進国のエネルギー源別記事出現の時系列変化—

（エネルギー記事インデックスデータより）

(b) キーワード出現頻度数のクロス比較

キーワードの出現頻度数をある一定期間にしぼって，主体別に横並びに比較し，主体別の話題の特徴を分析するための機能である。出力は図2.4－8に例示するように，グラフ表示により，主体別の特徴を容易につかむことができる。

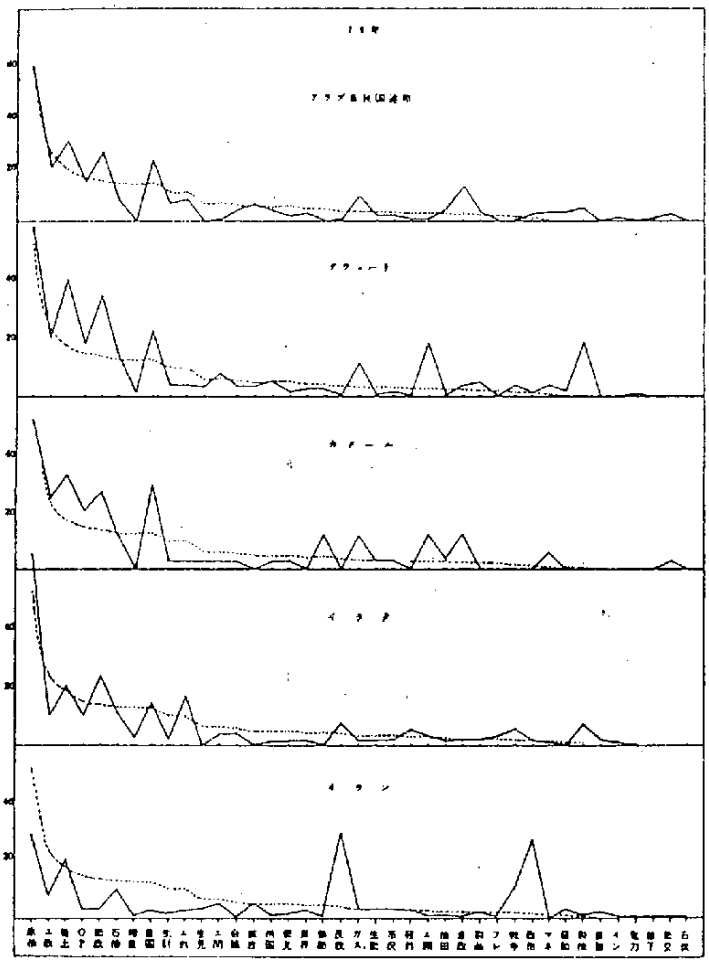


図2.4－8　キーワード出現頻度のクロス比較
—サウジアラビアとその他のＯＰＥＣ加盟国との
キーワード出現状況の比較—
（エネルギー記事インデックスデータより）

(c) キーワード出現頻度数の主成分分析

互いに何らかの相関性が考えられるキーワードを主成分として選び，その時系列出現頻度あるいはクロス出現頻度数を分析して中心軸となるキーワードを見通すことにより，全体の特徴を把握する機能である。

価格関連のキーワード

COMPONENT	EIGENVALUE	%	ACC.%	COPONENT	1	2	3	4
1	3.408	30.983	30.983	1 価格交渉	−0.545	0.512	0.357	0.233
2	2.687	24.429	55.412	2 価格提案	−0.1	0.008	0.916	−0.114
3	1.64	14.912	70.324	3 価格改革	0.381	0.785	0.113	−0.395
4	1.07	9.727	80.051	4 価格動向	−0.278	0.774	−0.055	0.204
5	0.763	6.94	86.991	5 価格見通し	0.77	0.235	0.239	−0.354
6	0.61	5.547	92.538	6 値 上 げ	0.93	0.192	−0.022	0.267
7	0.343	3.117	95.655	7 値 下 げ	−0.59	0.595	−0.024	−0.309
8	0.251	2.278	97.934	8 プレミアム販売	0.391	0.457	0.386	0.604
9	0.156	1.419	99.353	9 価格管理	0.109	−0.498	0.592	−0.261
10	0.056	0.506	99.859	10 価格カルテル	0.691	0.426	−0.306	−0.226
11	0.015	0.141	100	11 価格政略	0.657	−0.352	0.07	0.172

生産関連のキーワード

COMPONENT	EIGENVALUE	%	ACC.%	COPONENT	1	2	3	4
1	3.495	31.77	31.77	1 生産管理	−0.218	−0.238	0.843	−0.065
2	2.004	18.22	49.99	2 生産計画	0.503	0.617	0.042	−0.219
3	1.803	16.388	66.378	3 生産動向	−0.482	−0.072	−0.145	0.726
4	1.303	11.841	78.219	4 生産見通し	0.765	−0.314	0.409	0.064
5	0.737	6.696	84.915	5 生産費用	0.694	0.571	0.184	0.073
6	0.505	4.591	89.506	6 増 産	0.84	−0.281	0.072	−0.104
7	0.39	3.546	93.052	7 減 産	0.525	−0.009	−0.173	0.744
8	0.302	2.745	95.797	8 生 産 性	0.463	0.71	0.062	0.199
9	0.262	2.386	98.183	9 操業開始	−0.083	−0.038	0.838	0.258
10	0.133	1.206	99.39	10 操業再開	0.607	−0.613	−0.155	0.116
11	0.067	0.61	100	11 操業停止	0.568	−0.332	−0.321	−0.173

原油関連のキーワード

COMPONENT	EIGENVALUE	%	ACC.%	COPONENT	1	2	3
1	4.797	39.971	39.971	1 O P E C	0.677	0.621	−0.01
2	3.154	26.284	66.255	2 産 油 国	0.289	0.874	−0.086
3	1.023	8.524	74.779	3 石 油	0.591	0.441	0.122
4	0.959	7.992	82.771	4 原 油	0.908	−0.118	0.085
5	0.747	6.226	88.997	5 製 油 所	−0.45	0.704	0.348
6	0.619	5.162	94.159	6 油 田	0.802	−0.156	−0.098
7	0.272	2.267	96.427	7 石油産業	−0.003	−0.664	−0.053
8	0.195	1.621	98.048	8 石油業界	0.728	−0.359	0.362
9	0.126	1.046	99.094	9 石油パイプライン	0.529	−0.211	0.713
10	0.049	0.408	99.5	10 石油プラント	0.786	−0.048	−0.315
11	0.036	0.299	99.799	11 石油会社	0.16	0.803	−0.05
12	0.024	0.2	99.999	12 石油消費国	0.883	−0.118	−0.326

経済関連のキーワード

COMPONENT	EIGENVALUE	%	ACC.%	COPONENT	1	2	3	4
1	4.06	27.064	27.064	1 対外債務	−0.153	0.696	−0.344	−0.447
2	3.9	25.999	53.062	2 オイルマネー	0.228	−0.282	0.182	−0.746
3	1.73	11.531	64.593	3 インフレ	0.93	0.064	−0.084	0.047
4	1.363	9.085	73.678	4 経済成長	0.781	0.117	0.249	−0.005
5	1.067	7.112	80.79	5 経済動向	−0.303	0.875	0.128	0.185
6	0.799	5.328	86.118	6 経済見通し	0.295	0.761	−0.057	0.213
7	0.567	3.783	89.901	7 経済外交	0.519	−0.463	0.348	−0.295
8	0.515	3.432	93.333	8 経済協力	0.502	−0.228	−0.66	−0.009
9	0.35	2.334	95.667	9 経済計画	0.284	−0.45	−0.63	−0.002
10	0.249	1.658	97.325	10 経済政策	0.817	0.181	0.302	0.139
11	0.175	1.17	98.494	11 国 有 化	−0.043	−0.526	−0.239	0.498
12	0.128	0.851	99.346	12 借 款	0.039	0.709	−0.578	−0.269
13	0.058	0.389	99.734	13 財政政策	0.267	0.746	0.174	0.178
14	0.026	0.175	99.91	14 金融政策	0.561	0.354	0.009	−0.152
15	0.013	0.09	99.999	15 産業政策	0.819	−0.072	−0.156	0.207

図 2.4 − 9　時系列データによる主成分分析結果

（エネルギー記事インデックスデータより）

(d) キーワード出現頻度数のクラスター分析

相互に依存しているキーワードを類似度ごとに分類し，主体のカテゴリー化を行うための機能である。類似度の距離を定義する方法として，最短距離法，最長距離法，メジアン法，重心法，群平均法などがある。

(e) キーワード出現頻度の回帰分析

主体となるキーワードの出現頻度数と互いに依存する様々なキーワードの出現頻度数との相関分析を行い，特定の事象の予測式をみいだすための機能である。

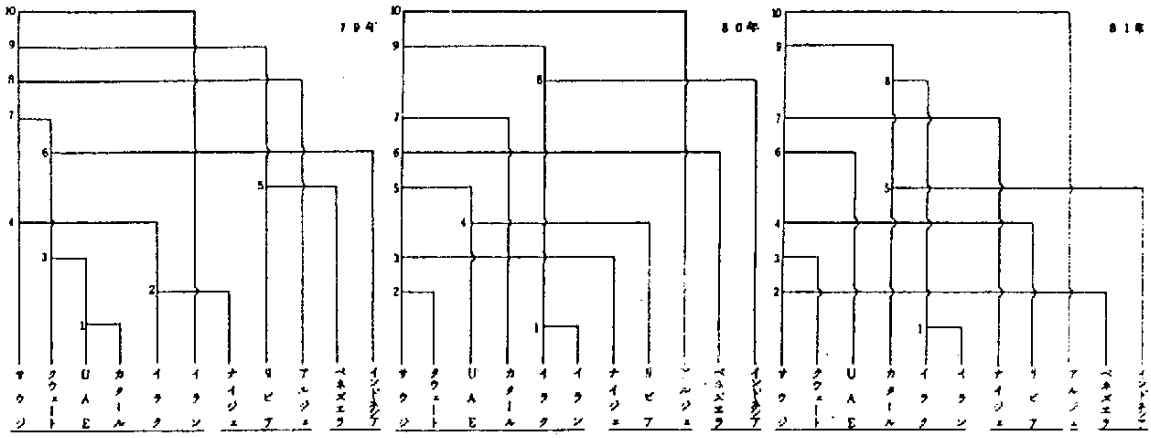


図 2.4－10　クラスター分析（最長距離法）結果
—記事出現にみるＯＰＥＣ加盟国の類似度—
（エネルギー記事インデックスデータより）

(f) 事実データによるキーワードの補足

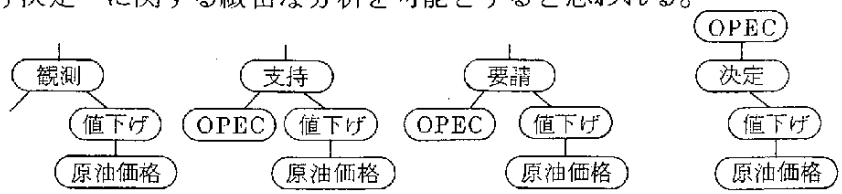
キーワードによる出現頻度だけを分析したのでは，主題に対する正確性，精密性を欠く場合があるが，これを事実データで補足することができる。

例えばエネルギー問題に関連して，「ＯＰＥＣの原油価格の値下げ決定」に関して分析したい場合を考えるとする。

キーワード“原油価格”&“値下げ”による検索だけでは値下げ決定に至るまでの経緯や値下げ決定後のその影響を分析することはできない。

一方，事実データでは，次の様な（値下げ）-（原油価格）の部分構造をもついくつかの事実データの目標構造で検索することにより“原油価格値

下げ決定 ”に関する緻密な分析を可能とすると思われる。



2.4.5 翻　訳

(1) 位置づけ

翻訳サブシステムの位置づけは，図2.1－1総合解析システム概念図に示されるように，基礎解析サブシステムによって作成された二次情報データベースを入力する形で構成される。

機械翻訳処理には，コンピュータによる文章情報処理技術のほとんどが含まれるといって過言でない。本システムにおいても翻訳サブシステムは基礎解析サブシステムと密接に関連し合い，互いに不可分の関係にある。一般に機械翻訳処理の手法を分類すると表2.4－5のようになる。

表2.4－5　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	元の言語から目的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わし，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概念構造を作り，それを介して翻訳を行う。

（57年度報告書より）

(a) 構文パターンマッチングによる方式

構文パターンマッチング方式は，あらかじめいくつかの文型パターンを用意しておいて，入力文とそれとを比較し，合致した文型パターンに従って翻訳をする方式である。この方式は，表題文の翻訳に代表されるように，パターンの数が限定される分野に向いている。通常の自然文に

対してこの方式を適用することは困難である。なぜならば，順列組合せの数だけのパターンを用意しなければならないからである。

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。

(b) トランスファ方式

トランスファ方式とは，元の言語で表現された文の分析を行い，その結果に従って目的言語に従った文を生成する方式をいう。この方式のことを一般に第2世代の機械翻訳システムといい，現在の機械翻訳システムの多くは，何らかの形でこの方式をベースとしている。

分析の結果は，木構造の形で表現され，文がどのような構造になっているかということが一目で分る。例えば，WE BUY A COMPUTE-Rという文は，基礎解析の項で示した図2.4－3の形で表現できる。このような一連の分析のことを構文解析処理といい，機械翻訳処理における基本的技術となっている。

(c) 融合方式

融合方式とは，構文解析が単に構造的な側面をのみ扱うのに対し，さらにそれに加えて意味的側面をも加味して翻訳処理を行う方式である。翻訳処理が，解析・変換・生成のステップを融合した形で行われるためこのように呼ばれる。

この方式の特徴は，次のとおりである。

① 木の変換を陽に行わない。

② ある単語に対して，訳し分けが必要でも辞書項目は1しか立てず，その意味のslot(注)をプログラムで選択することによって翻訳を行う。

③ slotとfiller(注)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。

④ 文脈自由文法（CFG, context free grammar）を用いる場合，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

（注）構文解析木で，条件を付けてもっているものをSlot，それを満たすものをfillerという。

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。

(d) 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表現するかが問題となる。前項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されるかどうかの方式をとっていた。ここで述べる方式は，文を論理式の形に変換することによって，そこに意味解析を含ませながら翻訳を行う方式である。モンテギュー文法による方式は，本方式によるものである。

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。

(e) 概念依存を用いた方法

概念依存を用いた方法とは，元の言語を解析し，最終的な意味表現である概念表現に落とし，その概念表現から目的言語を生成しようとするものである。この方法の特徴は，概念表現という言語によらないユニバーサルな表現をベースとするため，いかなる多言語間翻訳も可能とすることにある。

以上翻訳サブシステムの位置づけと，現在考えられている翻訳手法の概略について述べた。さらに翻訳手法の詳細については，５７年度報告書の4.8節を参照されたい。次に，本システムにおける翻訳システムの機能について述べる。

(2) 機　　能

本システムにおける翻訳サブシステムの機能構成を図示すると図2.4－１１のようになる。

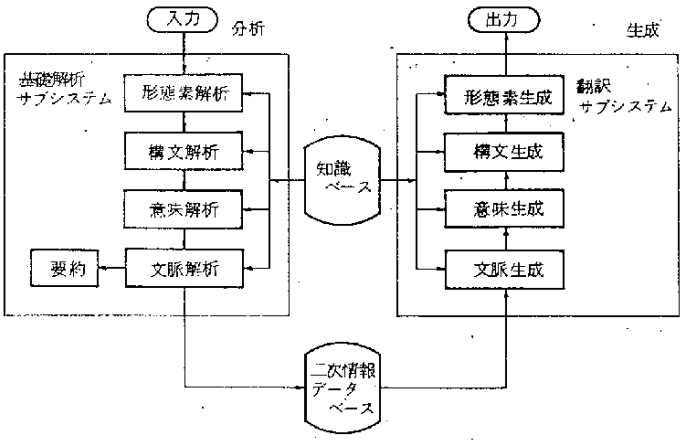


図2.4－11　翻訳サブシステムの機能構成

機械翻訳の機能を大別すると，分析部分と，生成部分に分けられる。分析部分は，図のように基礎解析システムにそのまま含まれる。従って翻訳サブシステムでの機能は，生成機能に限定される。次にその機能の概略について述べる。

(a)　文脈生成

二次情報データベースに格納された元の言語の文脈解析情報により目的言語の文脈生成を行う。各言語には各言語での特徴がある。例えば，英語では主語を省略することは，命令形や口語による省略形を除けば，まれであるといえる。しかし日本語においては，「私」という主語を省略することの方が多い。

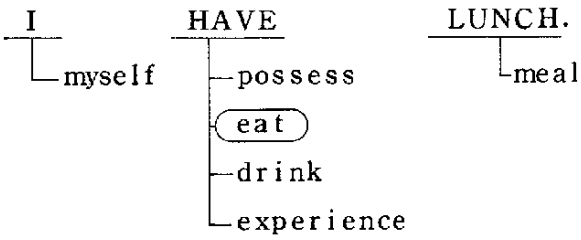
またその方が読み易い自然な文章でもある。このように文脈生成処理では，目的に合わせた適切な処理を行う。

また，文章によっては，結論を先にもって来るなど，読み易くするための文章の編成がえ等も，文脈生成処理で行われるであろう。

(b)　意味生成

意味生成処理とは，すでに分析されている元の言語の意味に対して，目的言語における適切な意味に生成することをいう。意味解析処理では，その文や，単語の持つ意味が限定されるため，目的言語の意味を決定す

ることは容易である。



例のように元の言語の上で，「HAVE」が「eat」つまり「食べる」という意味に限定されているため，翻訳する際には，その意味に対する目的言語のことばを生成すれば良いこととなる。

意味生成は，単にことばだけの問題ではなく，諺のように，目的言語での表現の習慣にも依存する。

It's no use crying over spilt milk.

この文の意味は，「こぼれたミルクを悔やんでみても仕方のない事だ」という意味であるが，日本語では習慣的に「覆水盆に返らず」と訳すのが正しい。このように意味生成とは，目的言語に最適なことばを生成することでもある。

(c) 構文生成

元の言語における文法構造と，目的言語における文法構造とは異なることが常である。しかし，元言語の構造が分れば，目的言語での構造を生成することは容易にできる。

そのためには，元の言語での構造が，目的言語での構造にどの様に対応するかというルールを知識ベースとして持つ必要がある。

次に構文生成の例を示す。

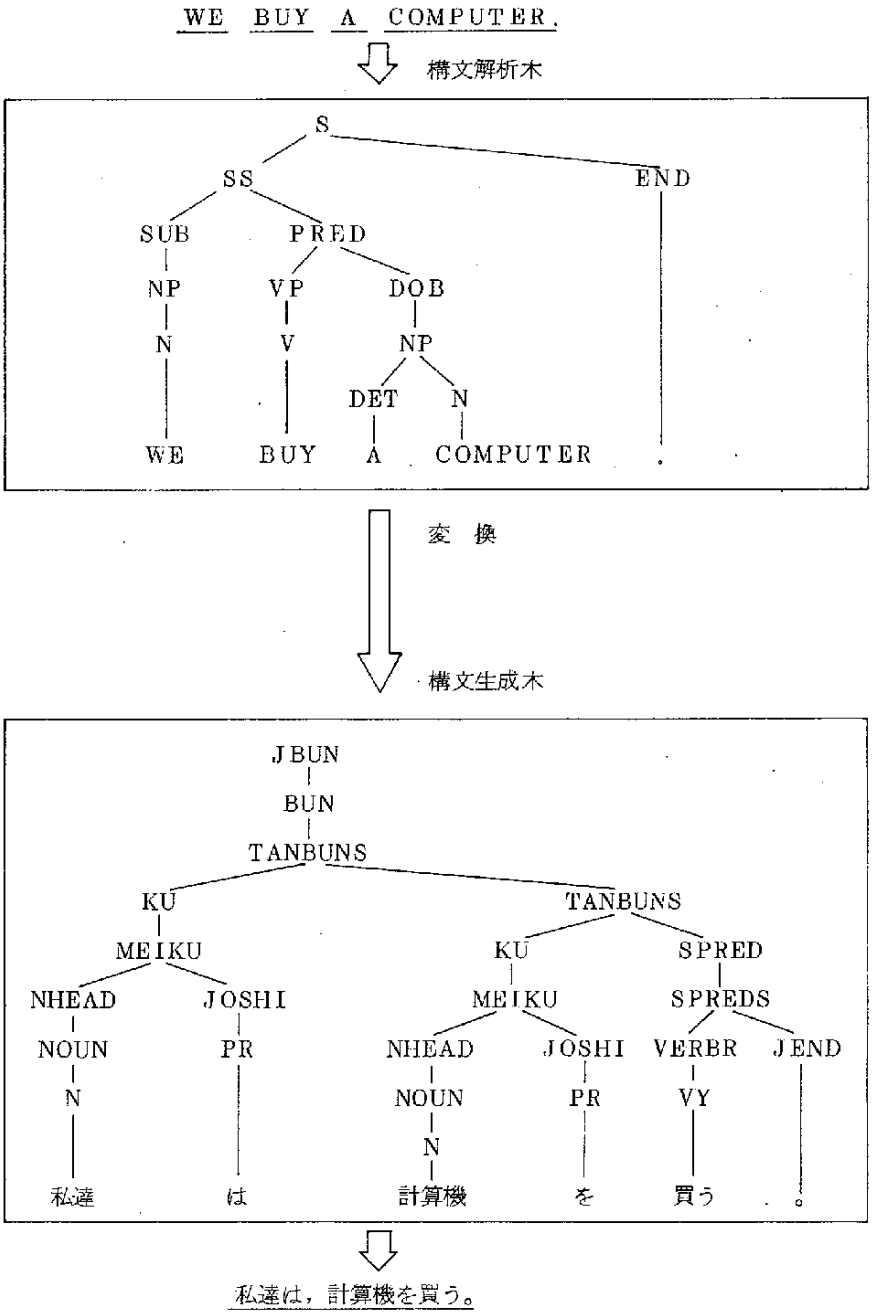


図 2.4 − 1 2　構文生成の例

(d) 形態素生成

形態素生成とは，構文生成によって生成された目的言語の生成結果を，文章としての形式になるように生成を行う処理をいう。日本語であるならば，各単語を詰めたり，また句読点を打ったりすることなどがあげられる。英語ならば，語と語のスペースをあけることや，文の最初の文字を大文字にしたりすること等がある。

いずれにしても，形態素生成処理は，目的言語の性質に合わせて木目細かに行われる必要がある。

(3) 翻訳サブシステムの利用例

基礎解析サブシステムによって解析された結果入力として，翻訳サブシステムは目的言語の生成を行う。

翻訳サブシステムでの処理イメージをつかむために，５８年度報告書のPROLOG言語を利用した文章解析処理の実験結果を例にして述べる。

実験の結果は次の意味を表現している。

＞：－GO……解析と生成実行指示

？［　　英　　文　　］……翻訳したい英文の入力

〔リスト構造になった英文〕

＊＊＊＊ THE ENGLISH PARSING TREE ＊＊＊＊＊＊＊＊＊

構文解析木の出力（英文）
基礎解析サブシステムでの処理結果

＊＊＊＊ THE JAPANESE PARSING TREE ＊＊＊＊＊＊＊＊

構文生成木の出力（日本文）
翻訳サブシステムでの処理結果

［　　日　本　文　　］……翻訳結果

次の英文における処理結果を示す。

① We buy a computer.

② The conference denies the sovereignty of Northumbria.

③ The institution gives the doctor a electroencephalograph.

④ The impressionism directs the trend of art.

⑤ The poverty cause the revolution which destroy the existence.

⑥ The bureaucrat controls the agency which collects the information.

⑦ With the advancement, the computer gives the society convenience.

①の結果

```
> :-GO.
? WE BUY A COMPUTER.
[WE.BUY.A.COMPUTER]
********** THE ENGLISH PARSING TREE **********
[ES.[ESS.[ESUB.[ESUB].[ENP.[ENP1.[ENP2.[ENP3.[EN.WE]]]]]]].[EDES.[EPRED.[EPRED]
.[EV.[EVT.BUY]].[EDOB.[ENP.[ENP1.[EDET.A].[ENP2.[ENP3.[EN.COMPUTER]]]]]]]]]]]]
********** THE JAPANESE PARSING TREE **********
[JBUN.[BUN.[TANBUNS.[KU.[MEIKU.[NHEAD.[NOUN.[N.拓道]]].[JOSHI.[PR.は]]]].[TANBU
NS.[KU.[MEIKU.[NHEAD.[NOUN.[N.計算機]]].[JOSHI.[PR.を]]]].[SPRED.[PREDRS.[VERBR
.[VY.買う]].[JEND.. ]]]]]]]
********** THE GENERATED JAPANESE SENTENCE **********
拓道は計算機を買う
>
```

②の結果

```
> :-GO.
? THE CONFERENCE DENY THE SOVEREIGNTY OF NORTHUMBRIA.
[THE.CONFERENCE.DENY.THE.SOVEREIGNTY.OF.NORTHUMBRIA]
********** THE ENGLISH PARSING TREE **********
[ES.[ESS.[ESUB.[ESUB].[ENP.[ENP1.[EDET.THE].[ENP2.[ENP3.[EN.CONFERENCE]]]]]]].[
EDES.[EPRED.[EPRED].[EV.[EVT.DENY]].[EDOB.[ENP.[ENP1.[EDET.THE].[ENP2.[ENP3.[EN
.SOVEREIGNTY]]]].[EPOSTMOD.[EOFP.[EOF.OF].[ENP.[ENP1.[ENP2.[ENP3.[EN.NORTHUMBRI
A]]]]]]]]]]]]]]
********** THE JAPANESE PARSING TREE **********
[JBUN.[BUN.[TANBUNS.[KU.[MEIKU.[NHEAD.[NOUN.[N.協議会]]].[JOSHI.[PR.は]]]].[TAN
BUNS.[KU.[MEIKU.[NHEAD.[RENKU.[NHEAD.[NOUN.[N.ノーサンブリア]]].[RJOSHI.[NO.の
]].[NHEAD.[NOUN.[N.主権]]]].[JOSHI.[PR.を]]]].[SPRED.[PREDRS.[VERBR.[VY.否定す
る]].[JEND.. ]]]]]]]
********** THE GENERATED JAPANESE SENTENCE **********
協議会はノーサンブリアの主権を否定する
>
```

③の結果

```
> :-GO.
? THE INSTITUTION GIVE THE DOCTOR THE ELECTROENCEPHALOGRAPH.
[THE.INSTITUTION.GIVE.THE.DOCTOR.THE.ELECTROENCEPHALOGRAPH]
********** THE ENGLISH PARSING TREE **********
[ES.[ESS.[ESUB.[ESUB].[ENP.[ENP1].[EDET.THE].[ENP2.[ENP3.[EN.INSTITUTION]]]]]].
[EDES.[EPRED.[EPRED1.[EV.[EVT.GIVE]].[EIOB.[ENP.[ENP].[EDET.THE].[ENP2.[ENP3.[E
N.DOCTOR]]]]]].[EDOB.[ENP.[ENP].[EDET.THE].[ENP2.[ENP3.[EN.ELECTROENCEPHALOGRAP
H]]]]]]]]]]]
********** THE JAPANESE PARSING TREE **********
[JBUN.[BUN.[TANBUNS.[KU.[MEIKU.[NHEAD.[NOUN.[N.公共団体]]].[JOSHI.[PR.は]]]].[T
ANBUNS.[KU.[MEIKU.[NHEAD.[NOUN.[N.医者]]].[JOSHI.[PR.に]]]].[TANBUNS.[KU.[MEIKU
.[NHEAD.[NOUN.[N.脳波記録装置]]].[JOSHI.[PR.を]]]].[SPRED.[PREDRS.[VERBR.[VY.与
える]].[JEND.. ]]]]]]]]]
********** THE GENERATED JAPANESE SENTENCE **********
公共団体は医者に脳波記録装置を与える
>
```

④の結果

```
> :-GO.
? THE IMPRESSIONISM DIRECT THE TREND OF ART.
[THE.IMPRESSIONISM.DIRECT.THE.TREND.OF.ART]
********** THE ENGLISH PARSING TREE **********
[ES.[ESS.[ESUB.[ESUB].[ENP.[ENP1.[EDET.THE].[ENP2.[ENP3.[EN.IMPRESSIONISM]]]]]]
].[EDES.[EPRED.[EPRED1.[EV.[EVT.DIRECT]].[EDOB.[ENP.[ENP1.[EDET.THE].[ENP2.[ENP
3.[EN.TREND]]]].[EPOSTMOD.[EOFP.[EOF.OF].[ENP.[ENP1.[ENP2.[ENP3.[EN.ART]]]]]]]]
]]]]]]
********** THE JAPANESE PARSING TREE **********
[JBUN.[BUN.[TANBUNS.[KU.[MEIKU.[NHEAD.[NOUN.[N.印象主義]]].[JOSHI.[PR.は]]]].[T
ANBUNS.[KU.[MEIKU.[NHEAD.[RENKU.[NHEAD.[NOUN.[N.芸術]]].[RJOSHI.[NO.の]]].[NHEA
D.[NOUN.[N.向き]]]].[JOSHI.[PR.を]]]].[SPRED.[PREDRS.[VERBR.[VY.指導する]].[JEN
D.. ]]]]]]]
********** THE GENERATED JAPANESE SENTENCE **********
印象主義は芸術の向きを指導する
>
```

⑤の結果

```
> :-GO.
? THE POVERTY CAUSE THE REVOLUTION WHICH DESTROY THE EXISTENCE.
[THE.POVERTY.CAUSE.THE.REVOLUTION.WHICH.DESTROY.THE.EXISTENCE]
********** THE ENGLISH PARSING TREE **********
[ES.[ESS.[ESUB.[ESUB].[ENP.[ENP1.[EDET.THE].[ENP2.[ENP3.[EN.POVERTY]]]]]].[EDE
S.[EPRED.[EPRED1.[EV.[EVT.CAUSE]].[EDOB.[ENP.[ENP1.[EDET.THE].[ENP2.[ENP3.[EN.R
EVOLUTION]]]].[EPOSTMOD.[ERELWHICH.[EWHICH.WHICH].[EPRED.[EPRED1.[EV.[EVT.DESTR
OY]].[EDOB.[ENP.[ENP1.[EDET.THE].[ENP2.[ENP3.[EN.EXISTENCE]]]]]]]]]]]]]]]]]]]]]
********** THE JAPANESE PARSING TREE **********
[JBUN.[BUN.[TANBUNS.[KU.[MEIKU.[NHEAD.[NOUN.[N.貧乏]].[JOSHI.[PR.は]]]].[TANBU
NS.[KU.[MEIKU.[NHEAD.[EN.[KU.[MEIKU.[NHEAD.[NOUN.[N.実在]].[JOSHI.[PR.を]]]].[
EMPRED.[PREDRE.[VERBR.[VY.破壊する]]]]].[NHEAD.[NOUN.[N.革命]]].[JOSHI.[PR.を
]]]].[SPRED.[PREDRS.[VERBR.[VY.の原因となる].[JEND.. ]]]]]]]
********** THE GENERATED JAPANESE SENTENCE **********
貧乏は実在を破壊する革命をの原因となる
>
```

⑥の結果

```
> :-GO.
? THE BUREAUCRAT CONTROL THE AGENCY WHICH COLLECT THE INFORMATION.
[THE.BUREAUCRAT.CONTROL.THE.AGENCY.WHICH.COLLECT.THE.INFORMATION]
********** THE ENGLISH PARSING TREE **********
[ES.[ESS.[ESUB.[ESUB].[ENP.[ENP1.[EDET.THE].[ENP2.[ENP3.[EN.BUREAUCRAT]]]]]].[
EDES.[EPRED.[EPRED1.[EV.[EVT.CONTROL]].[EDOB.[ENP.[ENP1.[EDET.THE].[ENP2.[ENP3.
[EN.AGENCY]]]].[EPOSTMOD.[ERELWHICH.[EWHICH.WHICH].[EPRED.[EPRED1.[EV.[EVT.COLL
ECT]].[EDOB.[ENP.[ENP1.[EDET.THE].[ENP2.[ENP3.[EN.INFORMATION]]]]]]]]]]]]]]]]]]]
********** THE JAPANESE PARSING TREE **********
[JBUN.[BUN.[TANBUNS.[KU.[MEIKU.[NHEAD.[NOUN.[N.官僚式の人]].[JOSHI.[PR.は]]]].
[TANBUNS.[KU.[MEIKU.[NHEAD.[EN.[KU.[MEIKU.[NHEAD.[NOUN.[N.通知]].[JOSHI.[PR.を
]]]].[EMPRED.[PREDRE.[VERBR.[VY.集める]]]]].[NHEAD.[NOUN.[N.働き]]].[JOSHI.[PR
.を]]]].[SPRED.[PREDRS.[VERBR.[VY.支配する].[JEND.. ]]]]]]]
********** THE GENERATED JAPANESE SENTENCE **********
官僚式の人は通知を集める働きを支配する
>
```

⑦の結果

```
> :-GO.
? WITH THE ADVANCEMENT,THE COMPUTER GIVE THE SOCIETY CONVENIENCE.
[WITH.THE.ADVANCEMENT,..THE.COMPUTER.GIVE.THE.SOCIETY.CONVENIENCE]
**********  THE ENGLISH PARSING TREE **********
[ES.[ESS.[EPREPP.[EPREPP].[EPREP.WITH].[ENP.[ENP].[EDET.THE].[ENP2.[ENP3.[EN.AD
VANCEMENT]]]]]]].[ECOM..].[ESUB.[ESUB1.[ENP.[ENP].[EDET.THE].[ENP2.[ENP3.[EN.CO
MPUTER]]]]]]].[EDES.[EPRED.[EPRED].[EV.[EVT.GIVE]].[EIOB.[ENP.[ENP1.[EDET.THE].
[ENP2.[ENP3.[EN.SOCIETY]]]]]].[EDOB.[ENP.[ENP1.[ENP2.[ENP3.[EN.CONVENIENCE]]]]
]]]]]]
********** THE JAPANESE PARSING TREE **********
[JBUN.[BUN.[TANBUNS.[KU.[FUKUKU.[NHEAD.[NOUN.[N.前進]]].[JOSHIF.[PREP].とともに
]]]].[TANBUNS.[KU.[MEIKU.[NHEAD.[NOUN.[N.計算機]]].[JOSHI.[PR.は]]]].[TANBUNS.[
KU.[MEIKU.[NHEAD.[NOUN.[N.社会]]].[JOSHI.[PR.に]]]].[TANBUNS.[KU.[MEIKU.[NHEAD.
[NOUN.[N.便利]]].[JOSHI.[PR.を]]]].[SPRED.[PREDRS.[VERBR.[VY.与える]].[JEND..]
]]]]]]]]
********** THE GENERATED JAPANESE SENTENCE **********
前進とともに計算機は社会に便利を与える
>
```

2.4.6 知識ベース

文章情報総合解析システムは，文章情報の巾広い情報を集約化し，利用者が判断しやすい形に変換するシステムであり，それを可能にする機能として，基礎解析，検索，内容分析（定量化分析），翻訳などがある。これらの機能をサポートし，効果的に働くために用意されたのが知識ベースである。従って，知識ベースは，各機能が使い易い構造や情報を持っていることが必要である。総合解析システムが，多くの人から利用され，利用者が望むかたちでの情報提供，加工および分析が行われるためには，各機能の拡張性も重要であるが，それらをサポートする知識ベースの拡張性もそれ以上に重要である。また，知識ベースには，基礎解析結果を反映（二次情報データベースの事実データの特別なケース）させたり，内容分析結果を反映させることによって知識ベースの質的向上をはかることが考えられる。そのためには，知識ベース作成，更新などのマン・マシン・インターフェイスが使い易いことも重要である。

以上，述べてきたように知識ベースは，各機能との結びつきが大きく，この構造（機能）を明確化することは困難であるが，概略としては，辞書，規則，知識表現の３つに分けられる。この３つの構造と前述の４つの機能のかかわりを以下に述べてある。

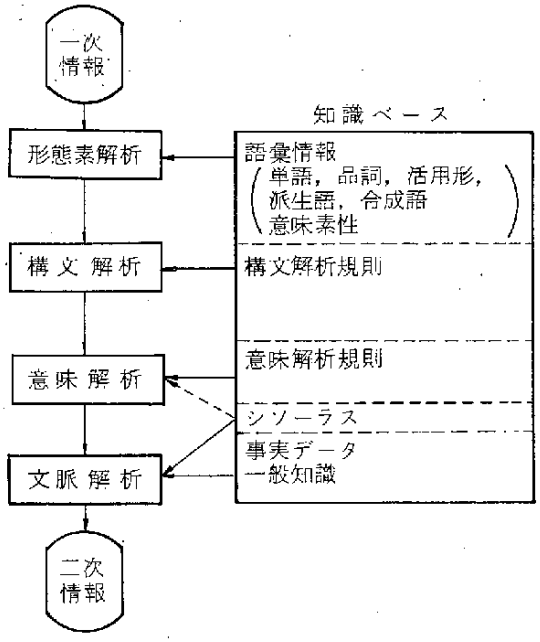


図2.4.－13　知識ベースと基礎解析の関係

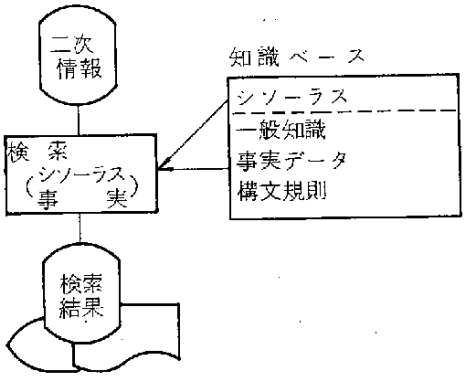


図2.4－14　知識ベースと検索の関係

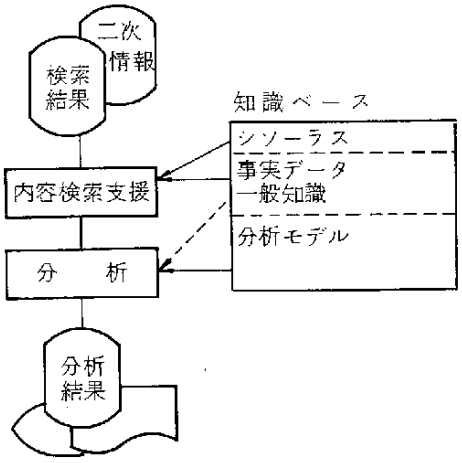


図2.4－15　知識ベースと内容分析の関係

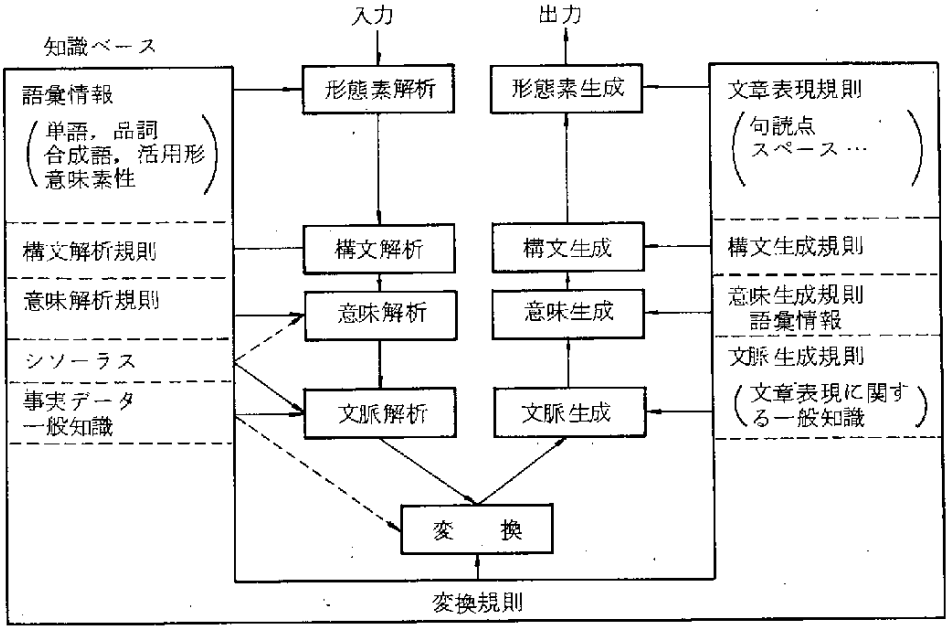


図 2.4－16　知識ベースと翻訳の関係

(1) 辞　　書

辞書は，主として基礎解析，翻訳で利用される。その情報は大きく分けて3つに分けることができる。

○形態情報……単語，合成語等の綴りに関するもの

○文法情報……品詞，係り受け属性等

○意味情報……上位概念や他の語概念との隣接関係の可能性を示す。

(2) 規　　則

規則も辞書と同様に基礎解析，翻訳で利用されることが多い。規則には構文解析用（例えば，文脈自由文法による書き換え規則）に関するもの，意味解析用（例えば，各用語間の関係を示した格文法）のものがある。また，辞書の隣接情報では補えないような語用論的なもの，単位文間の接続関係を解析する規則が考えられる。例えば，動物と run, walk は結びつくが，植物と run，walk は結びつかないといった情報がこれにあたる。

以上、辞書と規則について述べたが，この2つの部分の境界については明確にすることが難しく，異論の発生するところである。

(3) 知　　識

ここでいう知識とは，シソーラス，一般的知識（常識），事実データ，分析モデル等のことであり，文章の真の意味をとらえたり，意図することを把握したり，あるいは，可能ならば事態の予測ができるようにするために用意する情報である。例えば，レーガンが大統領になった日付を登録しておき，当選前の発言は個人的なものとし，当選後の発言（記事）は，アメリカとしての発言として扱うといったことが考えられる。二次情報データベースの事実データを正確に作成したり，質の高い翻訳ができるようにするためには，この知識の充実が重要であるが，どの程度のものを用意すればよいかということは，まだ研究段階のいきをでていない。

最後に，知識ベースの３つの構造と今年度実施したプロトタイプシステムの関係について図示しておく。

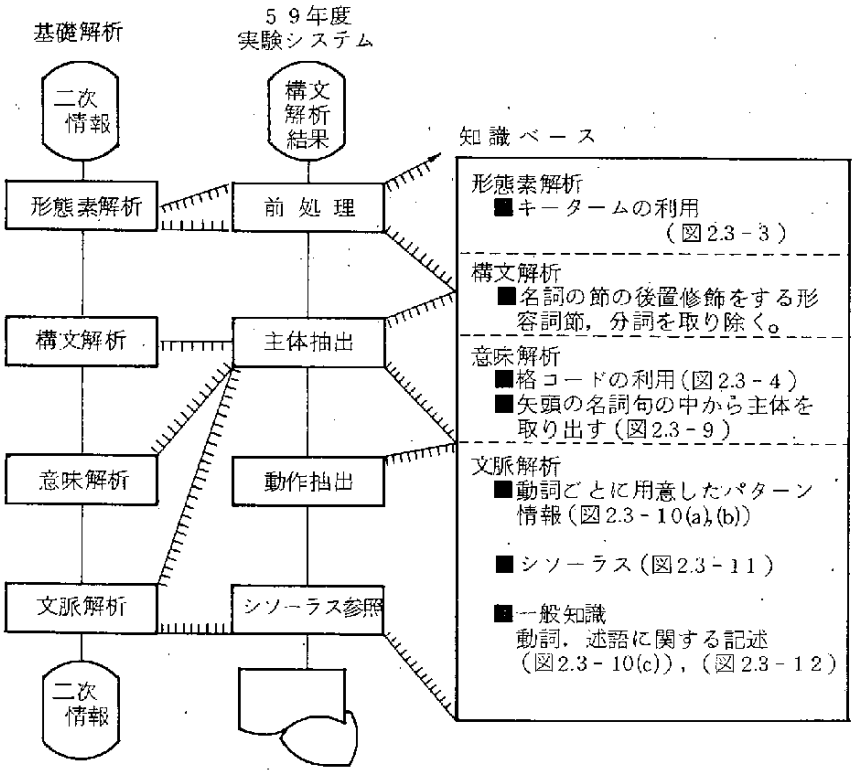


図2.4－17　総合解析システムと知識ベース

3. 展望と課題

3. 展望と課題

本事業の開始当時，文章情報の本格的利用の為の基盤は，技術レベルからも，その活用を求めるニーズとしても熟しつつあった。それに対し，商用の機械翻訳システムが各方面から発表され，人工知能の周辺技術が，当時標準装備化しつつあった段階を思わせる程の形で普及しようかという現状を見ると，それらの基盤は，ますます拡大し，活性化しているといえよう。その一方で，文章情報に正面から取り組んだ形での本格的な文章情報総合利用という意味では，底辺の拡大という事はあるものの，目覚ましい展開は見られていない。その中で，本委員会は，文章情報に関して，技術の側面からも，利用の側面からも，正面から取り組み，総合解析システムという形での具体像を構築すべく調査，研究，開発の努力を払ってきた。その結果，文章情報に対する処理技術と利用の技術とを統合した形での総合解析システムが，諸機関においてインプリメントされ，稼働されていくことの可能性と必然性とを明らかにする事ができた。同時に，総合解析システムの統合イメージを踏まえることにより，文章情報の処理技術と利用技術の双方について，その深化をはかる為の１つの礎を提供することともなった。

前章に述べられた総合解析システム像は，それぞれの動機付けに基づいて，各機関において実現していくことが可能なステップにある。総合解析システム像の拡大という観点からも，一般的な研究ではなく，実務での鍛え上げが求められる。このようなシステムの導入に当って特に留意すべき点は，扱う対象のモデル化のフェーズが文章情報利用の成否を決するとさえ云いうることである。その機関あるいは業務の観点とともに，その対象を論理化し，抽象化することにより，システムの骨子となるモデルが設定されるが，これを，具体性と構想力とをもって，着実に進めていかなければならない。そして，これに基づいた柔軟なシステム構築により，総合解析システムが十全の形で実現されるであろう。

文章情報の総合利用の為には，更に解決してゆかねばならない多くの課題がある。以下ではその主要なものを提示することとする。

(1) 自然言語処理技術の定着

まず，文章情報の解析技術を，多方面で真に利用できる形で確立する必要がある。各個別の研究開発テーマやアプリケーションに対しては，強力な解析系も少なからず作成されているが，誰もが入手し，利用できる形となった解析系は見当らない。

自然言語処理において，辞書，規則，アルゴリズムといった要素は，一定の段階では互いの独立性を保つものの，その基本としては，不可分な一体の体系として存在する。自然言語処理技術の拡大と深化の為には，このような体系が万人に見られる形で確立することが必要である。文章情報処理にとって辞書は中核となる要素であり，一方，辞書の開発から管理まで，一機関で対処するには余りに膨大なものとなりうることから，公共性の観点から対応する必要があることは広く認められている。しかし，辞書も，アルゴリズムや規則から独立なものとして開発できる範囲は極めて限定されている。現在は，少しでも方式等が異ると，全く別の思想で，互換性のない辞書が開発されているといってよい。そこで，辞書＝規則＝アルゴリズムという知的な系を一体として把えた上で，各々の独立性，独自性を実現するといった観点が必要である。辞書の確立のためにも，いくつかのアプローチでの広範囲なプロジェクトが要請されるが，ここでも，辞書＝規則＝アルゴリズムといった有機的な技術項目と利用の側面，社会性といった観点の双方からの充分な分析をともなうことが必須である。

ここで，一方では，辞書開発の個別の作業が最終的に分業にならざるを得ず，その総合性から個別性までに到る方法論の確立は，文章情報の総合利用のあらゆる要素において求められている。

解析技術の中の各技術要素についても，諸々のテーマはあるわけであるが，ここでは，とくに文脈処理への取り組みについて提起しておく。文単位の処理では実用の範囲が自ら限られてしまうことは明らかである。文脈処理は，例えば構文処理などに比べると，理論的基盤は不鮮明であり，その分だけ，解析処理の中で「難問」の上位に位置づけられている。しかし，利用のサイ

ドからはより近い距離にあるということから，従来とは逆の方向から，自然言語処理技術へ向かうことにより，文脈処理技術の確立が進められるべきでもあろう。

(2) 文章情報処理の為のシステム技術の拡大

解析技術そのものについで，コンピュータ処理技術の課題としては，マン・マシン・インターフェイスの積極的な活用がある。本事業においては，この面への考慮は表面には出さず，処理系の内部自立性の下で調査・研究を進めてきたが，昨今の知識工学技術の発展と浸透に見られるように，マン・マシン・インターフェイスの活用は，システムの基本構造をさえ左右する程の段階に到っている。このような状況を踏まえて，各種基本アルゴリズムのレベルからのシステム・イメージ再構築が必要である。

その他システム技術としての様々の問題もあるが，ここでは文章情報を扱うためのデータベース技術の確立の必要性をあげる。現存するデータベース・マネージメント・システムの多くは，数値データや事務処理的アプリケーションへの指向が強く，文章情報を有機的に扱うに適したツールは殆んど存在しない。文章情報を扱うために，既存のＤＢＭＳやハイ・レベルのアクセス・メソッドなどを用いると，余りに効率が悪かったり，柔軟性を欠いていたりという事態に行き当たる。そこで，ロー・レベルから個別のデータ管理システムを構築するということになるが，これにはしばしば多大のコストが要求される。文章情報処理の普及が見通せる現在，こうした基本ツールの整備が強く求められるゆえんである。

(3) 文章情報の利用に関する技術の確立

総合解析システムに限らず，どのようなシステム概念においても，利用サイドからのシステマティックな分析の程度が，システム及びその有効性の質を決定する。特に，文章情報のような場合には，それ自体の論理化の方法論が，いくつかの限られた範囲を除けば，実務レベルでの歴史を持っていない。そのような状態における利用の側からの即目的なニーズへの処理技術の対応は，余りにも抽象的であったり矮小化されたりしたものとなりかねない。そ

こで，利用のサイドにおける問題の定式化のための方法論の確立の努力が，緊急な課題となってきているのである。他方で，このような努力は利用のサイドで扱う情報の全般に関する認識の問題にも関わり，狭義の総合解析システムのためというに留まらない豊かな付加価値をも産み出しうることにも留意する必要がある。

文章情報利用の為の具体的な分析技術として現在は，数値データを扱うための方法論を採用している場合が殆んどである。これまで，計算機から得られる分析に足るデータが，高々文字列レベル程度のものであったということから言えば，これはやむをえないことであったが，文章情報をより深いレベルで分析に供する可能性が見えてきた現在，利用のフレーム，処理のフレームと合わせて，新しい方法論の実現が望まれる。これは，検索の技術に関しても同様に包含されている課題である。

以上に概括した課題に関しては，本事業に様々な形で関わってきた諸機関，メンバーにおいて，総合解析システムの実現を含めて，深化し，解決していくことが望まれるが，同時に，次のステップとして，一歩進んだ形での，文章情報総合利用のための各層協働事業が更に展開されることが再び要請されてくるであろう。

資料1　サンプルテキスト（米国＊石炭＊エネルギー政策）

①年月日　820730

紙誌名　WJ

先頭文　The House Public Works Committee approved a bill that would make it easier to lay coal-slurry pipelines.

記事インデックス　米下院公共事業委，石炭スラリー・パイプライン建設促進法案を可決

概　説　『米下院公共事業委が石炭スラリー・パイプラインの建設を容易にする案件を採択した』と述べている。結果としての法案名との同定が問題である。

②年月日　820809

紙誌名　CN

先頭文　Coal pipeline supporters won a stunning victory last Thursday when the Snate Energy Committee voted 146 in favor of legislation (S 1844) providing federal eminent domain rights for interstate coal pipelines.

記事インデックス　米国上院エネルギー委，圧倒的多数でスラリー炭パイプライン法案を可決。

概　説　『石炭パイプラインの推進者たちが先の木曜日に圧倒的な勝利を得た』が主文で，従属文としてその時に起った事実を述べている。『上院エネルギー委が14対6で法案S1844を可決した』とし，法案の説明として，その法案が，『州にまたがる石炭パイプラインについて，連邦政府に用地強制収用権を与える』ものであるとしている。ここでは，主文の扱い，日時の特定といった問題や，"bill"と"legislation"の関係，法案名の同定といった問題もある。

③年月日　820810

紙誌名　WJ

先頭文　The Senate Energy Committee last week approved lagislation that could introduce much needed competition in the coal transportation market and might create further incentives to develop the nation's vast coal reserves.

記事インデックス　米上院エネルギー委の石炭パイプライン法採択で鉄道の石炭輸送独占に幕

概　説　『上院エネルギー委が先週，石炭輸送市場に必要であった競合を起こし，巨大な埋蔵石炭の開発をより刺激するであろう法案を可決した』としている。内容は十分であるが，法案名等は全く見当らない。また，説明部分を扱うとすると，相当に思いきった情報の抽出（構造の変換）を行う必要があろう。

④年月日　820811

紙誌名　CWI

先頭文　The US Senate Energy & Natural Resources committee is approving coal slurry pipeline legislation which proponents say is necessary to insure the attractiveness of US coal abroad.

記事インデックス　米国上院エネルギー・資源委，石炭パイプライン法案を可決。

概　説　『米上院エネルギー天然資源委が石炭スラリー・パイプライン法案を可決した』。ここで，委員会名の表現が違っていることに注意。そして，その法案は，『推進者の述べることによると，米国石炭の海外での競争力を保証するものである』。"which proponents say"といった構文的な難しさはおくとして，法案の説明の含意するものを抽出し，標

準化することが，本実験での最大の問題であることは，ここでも示されている。

また，以上の4文を合せ読むことで，この法案の内容・目的・影響といった豊富な情報が得られるが，これらの間での諸概念の同定の問題も見ることができる。

⑤年月日　820831

紙誌名　FT

先頭文　POLAND'S exports of coal this year, the country's main hard currency earner, have already exceeded all last year's shipments, but Poland is encountering stiff competition from the U.S. and other coal exporters in its traditional Western markets.

記事インデックス　ポーランド，石炭輸出で米国など西側輸出国との競争に直面。

概　説　『ポーランドの今年の石炭輸出量が — 石炭がこの国の主要な外貨獲得手段であるが — 既に昨年一年間の船積み量を越えた』という文と，『しかし（それにともなって），ポーランドは，西側市場において米国をはじめとする石炭輸出国とのきびしい競争に直面することになった』という文が並べられている。前文の主旨は，記事の日付も加味しないと，十分な情報とはならない。後文の方がここでの中心と思われるが，そこでは，例えば，前置詞句の係り先といった，構文解析の問題がある。

⑥年月日　820906

紙誌名　IHT

先頭文　In a sharply worded letter to James G. Watt, the interior secretary, the governors of nine western states have called his department's new coalleasing

policies a move " to once again centralize on the Potomac critical decisions affecting western states."

記事インデックス　米西部９州の知事が産炭地貸与計画に関し，ワット内務長官に抗議文書送る。

概　説　『James G. Watt 内務長官への抗議文書の中で，西部９州の知事たちが，その新しい産炭地貸与の方針を，西部諸州に関する重大な決定を連邦政府へ再度集中させようという動きであると言っている』。構文的には，同格の挿入句か，動詞" centralize "が前置詞句をはさんで，名詞目的語をとることに対するアンビギュイティの問題などがある。この様な文の場合，とくに含意をとることには困難がある。

⑦年月日　８２０９２７

紙誌名　ＣＮ

先頭文　The Synthetic Fuels Corp. agreed to consider 11 projects — including four involving coal — for financial assistance at its board meeting on Sept. 17.

記事インデックス　米国合成燃料公社，石炭含む１１プロジェクトに対する補助金検討を約束

概　説　『合成燃料公社が，９月１７日の役員会において，石炭関連の４つを含む１１のプロジェクトへの財政援助に合意した』。挿入句の部分の解析に問題がある。

⑧年月日　８２１００８

紙誌名　ＷＪ

先頭文　The Reagan administration indicated it will try to block any effort by Armand Hammer, chairman of Occidental Petroleum Corp., to build a pipeline to

carry liquefied coal from Siberia to Moscow.

記事インデックス　米政府，オクシデンタル社のシベリア石炭パイプライン建設を阻止する意向。

概　説　『レーガン政権は，液化石炭の輸送の為のシベリアからモスクワへのパイプラインの建設を行うためのOccidental Petroleum Corp.のArmand Hammer会長によるいかなる努力をも阻止を試みるであろうと示唆した』。肩書きの挿入句といった問題はあるものの，比較的扱いやすい。しかし，構文的なアンビギュイティは多く，名詞を含めた格支配，選択制限などが必要である。

⑨年月日　821025

紙誌名　CN

先頭文　Although the U.S. coal industry is facing " extremely trying times " over the short term, it is ultimately heading " toward an unprecedented era of progress, " NCA Chairman Stonie Barker Jr. said last week.

記事インデックス　パーカーNCA会長，米石炭業界の将来に明るい見通し — 実質年4％成長。

概　説　『米国石炭産業は，短期的には大変な試練の時にあるが，つきつめると，前例のない程の発展の時代へと向かっている』という叙述を『Stonie Barker Ⅱ世NCA会長が先週語った』という挿入文がうけている。これは，新聞記事等に多く見られる構文である。この文の中核の述語を何とするのか，更には，このような文を「事実表現」として有効にまとめられるのかといったことは，重大な問題である。

⑩年月日　830611

紙誌名　ECO

先頭文　The energy bust has provoked a more sedate round of mergers and takeovers than the energy boom.

記事インデックス　エネルギー戦略の崩壊で米国の合併や公開買付買収落ちつく。

概　説　直訳すると，『エネルギー戦略の崩壊（energy bust）は，energy boom（が起きた程度）に比して，より控え目な合併と買収の状態をもたらした』となるが，否定的表現の標準化といった処理で，整理はできる。むしろ，比較構文の構文構造と，このような描写的な表現に対する事実構造との設定の仕方に問題があろう。

資料2　サンプルテキスト（サウジアラビア＊原油＊価格政策）

①年月日　820505

紙誌名　APS

先頭文　The Saudi Government is determined to defend the maker price of $34.00／barrel and the OPEC position against market forces, or against foreign company pressures,even at the cost of lowering its oil production below 6 million b／d　(SP 69)。

記事インデックス　サウジ，バーレル34ドルの基準価格固守の構え ― 減産強化も辞さず。

概　説　記事インデックスの前半にほぼ対応する内容をもつ。ただし，数値関連の表現に対する解析処理が必要である。

②年月日　820705

紙誌名　MEES

先頭文　One issue that is likely to figure prominently on the agenda of OPEC's extraordinary ministerial conference in Vienna is that of the differential enjoyed by extra-light, low-sulfur African crudes above the marker Arabian Light crude.

記事インデックス　サウジ，アフリカ原油のディファレンシャルの引き上げを要望。

概　説　先頭文はOPECの閣僚会議で，アラビアンライトに対する高品質アフリカ原油のディフェンシャルが問題であったという内容で記事インデックスとは微妙な差がある。ちなみに記事のタイトルは，記事インデックスとほぼ同一である。

③年月日　820710

紙誌名　FT

先頭文　THE organization of Petroleum Exporting Countries was seriously divided yesterday over Saudi Arabia's insistence that the prices of premium African crude oils should be raised by $1.50 a barrel.

記事インデックス　OPEC，サウジのアフリカ原油値上げ強硬主張めぐり内部分裂。

概　説　具体的な要求値上げ額を示している他はほぼ記事インデックスと一致する。

④年月日　820721

紙誌名　WJ

先頭文　Sheik Ali Khallfa al-Sabah, Kuwait's oil minister and a leading figure of the Organization of Petroleum Exporting Countries, is fond of quoting Mark Twain whenever he is asked whether the OPEC crisis is terminal.

記事インデックス　イラン勢力復活でOPEC内部に異変 ― 強硬派の地位優勢に展開。

概　説　これは，OPECの最近の状況に関する概説記事で，この文はその序説に含まれている。むしろ，この様なデータは，ノイズとして処理する方法論が必要である。

⑤年月日　820816

紙誌名　PIW

先頭文　The latest round of rumors notwithstanding, Saudi Arabia is making no present effort to raise its

skidding oil production, and Aramco or Petromin offtakers are under no pressure to lift more than minimum cargoes.

記事インデックス　サウジ，原油増産よりも値下げ検討 — 増産のうわさは根拠なし。

概　説　先頭文では，値下げの可能性については述べていない。文頭の副詞句や後半の節を扱うと，混乱の可能性がある。

⑥年月日　820816

紙誌名　MEES

先頭文　Despite the recent spate of remors regarding the possbility of an oil price cut in Saudi Arabia, MEES understands that the Saudi authorities are not contemplating any reduction in the official price of Saudi crude（$34／B for Arabian Light）at present.

記事インデックス　サウジ，原油価格引き下げよりも販売政策の変更考慮 — アラムコ外に開放。

概　説　文頭の副詞句で，値下げの噂があることを示唆しているが，主文では（正確には節の中では）値下げがないと述べている。形式的な主文（"MEES understands that ……"）の扱いも問題である。

⑦年月日　820818

紙誌名　APS

先頭文　The new Saudi proposal to " tie the prices of LPG to the posted prices of Arabian Light crude oil f.o.b. Ras Tanura and Yanbu' on a BTU basis ",as the Aug.5 Petromin telex to its customers stated, is intended to become a standard posting formula for all the

member-states of the Gulf Co-operation Council as from the beginning of 1983.

記事インデックス　サウジの原油価格連動ＬＰＧ価格，８３年には湾岸協力会議も基準として採用。

概　説　記事インデックスの内容は，先頭文に含まれているが，枝葉を落とすことが必要である。

⑧年月日　８２０９０６

紙誌名　ＩＨＴ

先頭文　BAHRAIN — OPEC experts,meeting after Saudi Arabia warned that high-quality oil prices had to rise quickly if the kingdom was to continue defending the group's base tariff, have backed an increase of up to $2.30 a barrel, the Middle East Economic Survey said Sunday.

記事インデックス　ＯＰＥＣ専門委，サウジのアフリカ原油値上げ要求を支持。

概　説　挿入として，サウジによる高品質石油価格の速やかな値上げの主張があって，それを受けて，ＯＰＥＣがそのバーレル当り 2.3 ドルの値上げを支持するとなっている。ここで，挿入の中に見られる " warn " の構文は後でも見られるパターンである。" warn that A if B " ならＡを主張，要求しており，Ｂが否定であったり，" if " のかわりに，" unless " を用いているとＢをむしろ主張，要求している。

⑨年月日　８２０９２１

紙誌名　ＦＴ

先頭文　SAUDI ARABIA has warned other members of the Organization of Petroleum Exporting Countries that it might cut its oil price from the present $34 for

a barrel of Arabian Light " marker " unless the rate for African crudes is raised.

記事インデックス　サウジ石油相，アフリカ原油に合わすならＯＰＥＣ基準価格も値下げと警告。

概　説　記事インデックスは，先頭文とほぼ同じであるが，更に主旨を絞ると前の記事で述べたように，" unless "の後を主張しているとみられる。

⑩年月日　８２０９２７

紙誌名　ＰＩＷ

先頭文　Saudi Arabia is really Keeping its LPG pricing options wide open, despite its move to establish some linkage of gas prices to crude oil.

記事インデックス　サウジ，ＬＰＧ価格決定で柔軟姿勢 ― 原油価格連結方針にも幅広い選択。

概　説　記事インデックスとほぼ同じ内容を含むが，一方この先頭文は発言等を含む行為を述べた文ではない。

⑪年月日　８２１１２２

紙誌名　ＩＨＴ

先頭文　BAHRAIN — Saudi Arabia and Kuwait expressed firm support Sunday for OPEC's reference price of $34 a barrel, and the Saudi oil minister, Sheikh Ahmed Zaki Yamani, denied reports that his country thought otherwise.

記事インデックス　ヤマニ，原油スポット価格と為替レート考慮するが，劇的な対策配慮せず。

概　説　先頭文は（構造上の問題はおくと）簡潔に結論を述べている。記事インデックスに見られる情報は，専門知識を豊富に持ったシステムではじめて活用されるものである。

⑫年月日　821123

紙誌名　FT

先頭文　IN RETROSPECT the disappointing invester response to the flotation of Britoil may seem a minor casualty of unprecedented uncertainty in the oil market.

記事インデックス　ヤマニ，原油価格維持を再確認 ― インドネシアは値下げを辞せず。

概　説　これは英タイムズ誌の解説記事（"World oil market"の欄の"The Pressure on prices"というタイトルの署名付記事）であり，要点の抽出自体，相当な技術的困難さをともなう。とくに，先頭文を利用することは，この場合，妥当ではない。

⑬年月日　821124

紙誌名　PN

先頭文　New York 11/23 ― The outcome of the "$34/bbl debate" by OPEC oil ministers next month will determine the medium-term outlook for the crude market, Royal Dutch/Shell officials told security analysts here this week.

記事インデックス　ロイヤル・ダッチ・シェル，原油価格値下げの影響を評価 ― アナリストに語る。

概　説　この文から，機械的処理で利用しうる情報としては，来月ＯＰＥＣの会議が開かれ，バーレル当り３４ドルの規準価格に関する論議がなされるということであろうが，文の構成上前面には出ていない。

⑭年月日　821201

紙誌名　APS

The Algerian government has agreed in principle to soften its line on the issue of price differentials for the sake of restoring OPEC unity on prices and production.

記事インデックス　サウジ国王，アルジェリア訪問 ― 価格ディファレンシャルの協力で合意。

概　説　アルジェリアが価格ディファレンシャルでの妥協を行うことが，先頭文から判る。内容の利用可能性としては，記事インデックスと大差ない。

⑮年月日　821207

紙誌名　WJ

先頭文　Saudi Arabia's petroleum minister, Ahmed Zaki Yamani, reiterated his country's determination to defend the current price structure of the Organization of Petroleum Exporting Countries and to unhold the OPEC benchmark price at $34 a barrel.

記事インデックス　サウジのヤマニ石油相，ＯＰＥＣマーカー原油価格３４ドルを守ると明言。

概　説　基本的に，先頭文は記事インデックスと一致する。

⑯年月日　821231

紙誌名　FT

先頭文　A decisive rift in the Organization of Petroleum Exporting Countries' oil pricing structure could result from a meeting scheduled for early next week between Sheikh Ahmed Zaki Yamani, Saudi Minister of Oil, and representatives of the four U.S. shareholders in the operations of the Arabian American Oil

Company (Aramco).

記事インデックス　米国石油メジャー，サウジ原油の価格問題でヤマニ石油相と会談予定。

概　説　記事インデックスの内容を含む，豊富な情報をもつ。とくに価格構造変動の可能性を，処理可能な形式にすることは重要である。

⑰年月日　830104

紙誌名　AWJ

先頭文　NEW YORK—The four U.S. companies that pump and purchase most of Saudi Arabia's oil are expected to warn the kingdom that their petroleum liftings, which are already severely depressed by weak demand, will fall further unless the country reduces its prices or limits its production to help eliminate the persistent oil glut.

記事インデックス　サウジが値下げか減産を行わねば，原油引取量を削減とアラムコが警告。

概　説　"warn……unless"に着目すると，値下げ又は減産を要求していると抽象化できる。

⑱年月日　830105

紙誌名　LT

先頭文　The four American oil companies which lift the bulk of Saudi Arabia's oil production met Shaikh Ahmed Zaki Yamani, the kingdom's oil minister, yesterday to press him to reduce the price of Saudi Arabian oil.

記事インデックス　アラムコ，ヤマニ・サウジ石油相に原油値下げを要請。

概　説　先頭文は，値上げ要請の為，会談したとなっているが，内容は記事インデックスに近い。

⑲年月日　830105

紙誌名　WJ

先頭文　The four U.S. companies that operate Saudi Arabia's oil industry met yesterday with Saudi Oil Minister Sheik Ahmed Zaki Yamani in Geneva, Switzerland, but apparently didn't reach any definitive agreement on either the price or production level of Saudi crude.

記事インデックス　アラムコ代表とヤマニ・サウジ石油相，価格・生産量について合意に達せず。

概　説　先頭文に十分な情報がある。記事インデックスでは"but"以下を採用している。

⑳年月日　830105

紙誌名　FT

先頭文　SHEIKH AHMED ZAKI YAMANI, the Saudi oil minister, yesterday continued talks in Geneva with senior executives of four major oil companies, amid rising hopes within the industry that Saudi Arabia would agree in the near future to trim its prices.

記事インデックス　サウジのヤマニ石油相，値下げについてジュネーブでアラムコ代表と会談。

概　説　先頭文の文末の副詞句（"amid……"）の取り扱いで，結果の質が変わるであろう。

㉑ 年月日　830110

紙誌名　BW

先頭文　Venezuelan Energy & Mines Minister Humberto Calderón Berti was in a foul mood as he packed his bag before heading home from OPEC's yearend meeting in Vienna, a conference that was doomed from the start.

記事インデックス　OPEC原油価格，野放し状態へ — 国別生産割当に失敗。

概　説　最近の動向に関する全般的な解説記事であり，この先頭文は殆んど利用できない。

㉒ 年月日　830117

紙誌名　PIW

先頭文　Conflicting pressures on Arab oil producers groping to regulate pricing and production problems are tending to stretch out and delay decisive actions.

記事インデックス　中東湾岸産油国の原油引き取り範囲が狭まる。

概　説　この文に続く"But"に始まる文で，記事インデックスの内容が述べられている。

前記⑱においては重文の形で1文としているので拾われていた。

㉓ 年月日　830207

紙誌名　PN

先頭文　London 2/4 — The crude-price-cut waiting game goes on, with British National Oil Corp. reluctant to tip its hand and Libya refusing to quote a new price.

記事インデックス　サウジとナイジェリア，原油値下げで合意か — 各産油国待機戦術。

概　説　" crude-price-cut waiting game " が続いているという文で，記事インデックスの後半に対応している。

㉔年月日　830211

紙誌名　LT

先頭文　A cut in oil by the Organization of Petroleum Exporting Countries now inevitable, Shaikh Ahmed Zaki Yamani, Saudi Arabia's oil minister said yesterday.

記事インデックス　サウジのヤマニ石油相，原油価格の引き下げ不可避を公式に認める。

概　説　記事インデックスと先頭文は一致している。ここで，値下げがはじめて表明された。

㉕年月日　830211

紙誌名　FT

先頭文　SAUDI ARABIA publicly abandoned its defence of a $34-a-barrel reference price for oil yesterday and admitted that a cut in world prices was inevitable.

記事インデックス　サウジのヤマニ石油相，83年2月10日基準原油価格の維持を断念。

概　説　同様に，記事インデックスの内容を（" yesterday "など時制の問題はあるが）先頭文が比較的簡潔に表現している。

㉖年月日　830223

紙誌名　WJ

先頭文　Oil ministers of four Arab Persian Gulf states in the Organization of Petroleum Exporting Countries

met yesterday in Riyadh, Saudi Arabia, amid increasing indications that they were planning to slash at least $4 a barrel from the current $34 price for the benchmark grade.

記事インデックス　湾岸4大産油国石油相，少なくともバーレル当たり4ドルの値下げを検討。

概　説　主節は，会談の事実だけであるが，"amid"以下で値下げ幅の観測が見られる。

㉗年月日　830223

紙誌名　IHT

先頭文　RIYADH — The oil ministers of Saudi Arabia, Kuwait, Qatar and the United Arab Emirates held talks here Tuesday on how to respond to the sudden drop in world oil prices, and a well-placed source said they were considering cutting their oil prices by more than $4 a barrel.

記事インデックス　湾岸4大産油国，バーレル当たり4ドル以上の値下げを検討。

概　説　会談の事実と（信頼筋による）値下げ幅とを述べている。

㉘年月日　830224

紙誌名　WJ

先頭文　Oil ministers from Saudi Arabia, Kuwait, the United Arab Emirates, Qatar and Iraq ended talks in Riyadh, Saudi Arabia, yesterday with an agreement to cut their oil prices by an amount they refused to disclose.

記事インデックス　湾岸諸国石油相，リヤドで会合し原油値下げで合意 — 値下げ幅未定。

概　説　会談の結果で，明確な事実に近いことにも対応して，処理可能な情報量の多い先頭文となっている。

㉙年月日　830224

紙誌名　IHF

先頭文　RIYADH — Five Gulf oil-producing nations agreed in principle Wednesday to cut prices, and they threatened to use their economic might to force OPEC to accept a new price structure at emergency talks next week.

記事インデックス　湾岸5大産油国，価格引き下げで合意 — OPEC会議で提案か。

概　説　文の後半で" threaten "に対する後ろの" to "以下を処理することで記事インデックスを含む内容を得ることができる。

㉚年月日　830224

紙誌名　LT

先頭文　Saudi Arabia and four other Gulf states agreed yesterday to cut their oil prices by at least 10 percent, and threatened to make even bigger reductions unless other members of the Organization of Petroleum Exporting Countries fall into line with their plan to avert an international oil price war.

記事インデックス　湾岸5大産油国，2月23日に原油価格引き下げで合意。

概　説　既出の" but "による結合の場合と異なり，この場合は先頭文の前半が記事インデックスとしてとられている。また，後半は" threaten "を

"warn"の場合と同様に考えることが出来るが，ＯＰＥＣの他のメンバーに要求している内容の抽象化は容易ではない。

㉛年月日　８３０２２８

紙誌名　ＬＴ

先頭文　Bahain (Reuter) — The Gulf states, which on Saturday gave the Organization of Petroleum Exporting Countries on week to agree on new prices or face a possible price war, will cut tariffs by perhaps $7 a barrel unless Opec backs a $4 cut, according to the Middle East Economic Survey.

記事インデックス　湾岸産油国，４ドルの値下げを認めないなら７ドルの値下げも辞せず。

概　説　記事インデックスは，ＯＰＥＣに対する要請であることを明示していないが，ほぼ先頭文と同じである。ここでも，より抽象化するならば，"unless"以下を中心とする。

㉜年月日　８３０３０４

紙誌名　ＬＴ

先頭文　Bahrain (AP—Dow Jones) — Saudi Arabia and Kuwait are going to face large budget deficits as a result of falling oil revenues, according to diplomatic and banking sources in the Arabian Gulf.

記事インデックス　サウジアラビアとクウェート，原油値下げで大幅財政赤字。

概　説　記事インデックス中の「原油値下げで」という記述を除いて，先頭文で十分である。

㉝ 年月日　830317

紙誌名　LT

先頭文　Saudi Arabia's $5 a barrel oil price cut will be backdated in two stages to the beginning of last month, it was reported yesterday as the price of North Sea oil continued to be in limbo after this week's agreement by the Organization of Petroleum Exporting Countries.

記事インデックス　サウジアラビア，アラビアン・ライトは2月1日実施で30ドル。

概　説　先頭文では「先月始めにさかのぼる」5ドルの値下げを述べている。値下げ幅の違いは，2月始めと2月中旬以降の差である。ここで「先月始め」のような表現の処理は検討する必要があろう。

㉞ 年月日　830323

紙誌名　FT

先頭文　CROWN PRINCE Abdullah of Saudi Arabia yesterday warned other members of the Organization of Petroleum Exporting Countries (Opec) that the Kingdom was " perfectly capable of winning an oil price war " if they failed to observe the pact concluded last week in London.

記事インデックス　サウジのアブドゥラ皇太子，OPEC加盟国に協定遵守を要請。

概　説　ここでも" warn "の扱いによっては，記事インデックス相当の内容表現を得ることができる。

昭和60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

59－S002