61－S－002

自然言語のコンピュータ解析用辞書に関する調査研究報告書

昭和62年3月

JIPDEC

財団法人　日本情報処理開発協会

請求番号	PDEC 61-S-002	登録番号	
~~登録名~~	自然言語のコンピュータ解析		
書名	国語書に関する調査研究報告書		

所属	帯出者氏名	貸出日	返却予定日	返却日

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６１年度に実施した「自然言語のコンピュータ解析用辞書に関する調査研究」の成果をとりまとめたものであります。

はじめに

我が国の産業を取り巻く国際経済環境は複雑化してきており，政府，産業界では海外の産業経済動向，技術動向等について迅速かつ正確に把握する必要性が従来にも増して高まり，一方，海外の新聞情報，文献情報等は国際ネットワークを通じて即時に入手可能な状況になった。

このような国内外の自然言語による文章情報の蓄積，検索利用等の需要が増大する中で，自然言語を対象としたコンピュータによる解析を正確に行うニーズが高まっている。

その際に必要になるのが辞書であり，解析に用いる辞書の良否が解析の精度を決定するとさえ言われている。

現在，自然言語解析の応用技術として，自動抄録，機械翻訳，自然言語による質問応答等に関するシステムの研究開発が諸機関において進められているものの，解析に用いる辞書は諸機関における解析手法，解析レベル等の違いにより各々異なっており，システム間の相互利用は不可能な状況にある。

このような背景から，当協会では，辞書のシステム間相互利用を可能とすることを目的に辞書体系を整備し，効率的な辞書作成，維持体制等の確立を図るべく，昭和６０年度より３年計画で，自然言語のコンピュータ解析用辞書に関する調査研究を実施することとした。

６０年度においては，諸機関におけるコンピュータ解析用辞書について海外も含めた先進事例調査を行い，辞書の相互利用を可能とするための辞書項目の共通性，辞書の全体構成，品詞分類・意味分類等の体制について検討した。

２年目に当る本年度は，６０年度の調査結果，および，国内・海外の辞書更新機能を中心とした調査結果，等を踏まえて，コンピュータ解析用辞書整備のための基本構想を描こうと試みた。本報告書はその成果をとりまとめたものである。

本報告書が，関心のある方々に広く利用され，今後の自然言語処理および辞書研究の向上の一助として寄与できれば幸いである。

最後に，本調査研究にあたって，ご指導ご協力いただいた渕委員長，石川主査（前半），横井主査（後半）をはじめ，各委員並びに関係各位に感謝の意を表します。

昭和６２年３月

「自然言語解析用辞書調査委員会」　委員名簿

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　教授
委員	矢田　光治	㈱ＣＳＫ総合研究所　社長
委員	瀬谷　重信	日本電信電話㈱ＮＴＴデータ通信 事業本部営業統括本部 担当部長
委員	石川　徹也	図書館情報大学図書館情報学部　助教授
委員	横井　俊夫	(財)新世代コンピュータ技術開発機構・研究所　次長
委員	荻野　綱男	埼玉大学教養学部　助教授
委員	横山　晶一	通商産業省電子技術総合研究所 パターン情報部推論システム研究室　主任研究官
委員(前半)	山本恵美子	ソフトウェア研究会　(前)副会長
(後半)	坂本　光子	〃　(現)　〃
委員	中瀬　純夫	リソース・シェアリング㈱ＡＩ研究開発部　部長
委員	木南公統司	㈱開発計算センター　システム第２部　次長
委員	長谷川　亨	㈱ソフトウェア・リサーチ・アソシエイツ 第２システム本部システム企画室　室長
委員	竹内　憲	日本タイムシェア㈱第１システム事業部　部長付
委員	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　室長
委員	佐藤　安夫	通商産業大臣官房 情報管理課政策情報システム室　企画係長
委員	佐内　則夫	(財)日本情報処理開発協会　開発研究室　室長
事務局		(財)日本情報処理開発協会　開発研究室（担当：茂呂　知明）

「自然言語解析用辞書調査専門委員会」　委員名簿

主査（前半）	石川　徹也	図書館情報大学図書館情報学部　助教授
主査（後半）	横井　俊夫	(財)新世代コンピュータ技術開発機構　研究所　次長
委員	荻野　綱男	埼玉大学教養学部　助教授
委員	村木新次郎	国立国語研究所言語体系研究部第１研究室　室長
委員	田中　裕一	(財)新世代コンピュータ技術開発機構　研究所 第二研究室　研究員
委員	横山　晶一	通商産業省電子技術総合研究所 パターン情報部推論システム研究室　主任研究官
委員	寺野　隆雄	(財)電力中央研究所　経済研究所 情報システム部知識処理研究室　主査研究員
委員	中瀬　純夫	リソース・シェアリング㈱ＡＩ研究開発部　部長
委員	木南公統司	㈱開発計算センター　システム第２部　次長
委員	長谷川　亨	㈱ソフトウェア・リサーチ・アソシエイツ 第２システム本部システム企画室　室長
委員	竹内　憲	日本タイムシェア㈱第１システム事業部　部長付
委員	高橋　清一	㈱ＣＳＫ総合研究所ＡＩ商品開発部 知的ワードプロセッサ開発課　主任
委員	佐藤　安夫	通商産業大臣官房 情報管理課政策情報システム室　企画係長
委員	石川　敬子	通商産業大臣官房 情報管理課政策情報システム室　調査係長
事務局	(財)日本情報処理開発協会　開発研究室（担当：茂呂　知明）	

目　　次

1. 事業概要 …… 1

1.1 目的と背景 …… 1

1.2 事業計画と実施経過 …… 2

1.3 調査研究成果の概要 …… 4

2. 自然言語処理応用システムと解析用辞書の現状 …… 15

2.1 機械翻訳システム …… 15

2.2 文書作成支援システム …… 19

2.3 データベース・アクセスシステム …… 21

2.4 その他のシステムと今後の問題点 …… 22

3. 国内における機械翻訳システムの辞書更新の現状 …… 25

3.1 国内調査の概要 …… 25

3.2 国内調査結果 …… 29

3.2.1 国内調査結果の概要 …… 29

3.2.2 メーカ調査結果 …… 32

3.2.2.1 機械翻訳システム …… 33

3.2.2.2 辞書の構成・提供 …… 34

3.2.2.3 辞書更新処理方式 …… 35

3.2.2.4 辞書のメンテナンス・サポート体制 …… 36

3.2.2.5 ユーザへの希望 …… 37

3.2.2.6 メーカ調査集計表 …… 38

3.2.3 ユーザ調査結果 …… 50

3.2.3.1 機械翻訳システム …… 50

3.2.3.2 辞書の構成・評価 …… 52

3.2.3.3 辞書の更新 …… 54

3.2.3.4 辞書更新運用体制 …… 54

3.2.3.5 ユーザからの要望 …… 55

3.2.3.6 ユーザ調査集計表 …… 56

4. 海外における辞書の作成及び標準化動向 …… 69
4.1 海外調査の概要 …… 69
4.2 海外調査結果 …… 70
4.2.1 海外調査結果の概要 …… 70
4.2.2 機関別調査結果 …… 73
4.2.2.1 国立医学図書館（米国） …… 73
4.2.2.2 アメリカ保健機構（米国）…… 79
4.2.2.3 ホートン・ミフリン出版社（米国）…… 82
4.2.2.4 モントリオール大学（カナダ）…… 84
4.2.2.5 ウォータールー大学 …… 84
〔第2回 New OED センターコンファレンス〕（カナダ）
4.2.2.6 ロングマン出版社（イギリス）…… 87
4.2.2.7 オックスフォード大学出版部（イギリス）…… 88
4.2.2.8 EUROTRA プロジェクト …… 92
4.2.2.9 EURODICAUTOM プロジェクト …… 92
4.2.2.10 ボン大学 コミュニケーション・音声研究所（IKP）（西独）… 96
4.2.2.11 ジーメンス（西独）…… 98
4.2.2.12 INFOTERM（オーストリア）…… 98
4.2.2.13 ピサ大学および計算言語学研究所（イタリア）…… 99

5. 機械翻訳システムを例とした辞書及びその更新のあり方 …… 103
5.1 基本的考え方 …… 103
5.1.1 機械翻訳システムと辞書 …… 103
5.1.2 機械翻訳システムのユーザ …… 105
5.1.3 辞書の更新と作成をめぐる諸問題 …… 106
5.2 辞書の分類・構成 …… 108
5.2.1 機械翻訳システムにおける言語データ …… 108
5.2.2 管理上の辞書の構成 …… 110
5.3 辞書の作成 …… 113
5.3.1 基本辞書 …… 115

5.3.2　専門語辞書 …… 116
5.3.3　ユーザ辞書 …… 117
5.4　辞書の更新 …… 118
5.4.1　更新の必要性 …… 118
5.4.2　ユーザによる更新方法 …… 120
5.4.3　更新サポート機能 …… 123
5.4.4　基本辞書の更新 …… 125
5.5　現状の問題点とあるべき姿 …… 126
5.5.1　辞書の管理 …… 126
5.5.2　辞書の更新 …… 127
5.5.3　互換性 …… 129
5.5.4　基本辞書 …… 130
5.5.5　まとめ …… 130

6.　LDB(Lexical Database)の構想 …… 133
6.1　ＬＤＢとは …… 133
6.1.1　ＭＲＤとは …… 133
6.1.2　LDBを必要とする背景 …… 134
6.1.3　ＬＤＢの構想 …… 135
6.2　国内外におけるLDBの動向 …… 136
6.2.1　日本におけるLDBの動向 …… 137
6.2.2　海外におけるＬＤＢの動向 …… 138
6.2.3　日本におけるＬＤＢ関連研究事例 …… 138
6.2.3.1　国立国語研究所 …… 138
6.2.3.2　電子技術総合研究所 …… 142
6.2.3.3　日本電子化辞書研究所 …… 144

7. 今後の課題 …… 149

資料編 …… 151

資料Ⅰ．国内調査票 …… 151

資料Ⅱ．海外調査訪問機関 …… 203

1．事業概要

1. 事業概要

1.1 目的と背景

我が国の産業を取り巻く国際経済環境は複雑化してきており，政府，産業界では海外の産業経済動向，技術動向等について迅速かつ正確に把握し，かつ国内情報を海外に提供する必要性が従来にも増して高まっている。

一方，国内外の新聞情報，文献情報等は国際ネットワークを通じて即時に入手可能な状況になっている。

このような海外の自然言語による文章情報の蓄積，検索利用等の需要が増大する中で，自然言語を対象としたコンピュータによる解析を正確に行うニーズが高まっている。

その際に必要になるのが辞書であり，解析に用いる辞書の良否が解析の精度を決定するとさえ言われている。

現在，自然言語解析の応用技術として，情報蓄積面では自動インデクシング，自動抄録等に関して，また，情報の利用面では機械翻訳，内容検索，自然言語による質問応答等に関するシステムの研究開発が諸機関において進められているものの，解析に用いる辞書は諸機関における解析手法，解析レベル等の違いにより各々異なっており，システム間の相互利用は不可能な状況にある。

このため，本事業では，昭和６０年度より３ケ年計画でコンピュータ解析用辞書のシステム間相互利用を可能とすることを目的に，辞書体系を整備し，効率的な辞書作成，維持体制等の研究を行っている。

６０年度においては，諸機関におけるコンピュータ解析用辞書について海外も含めた先進事例調査を行い，辞書の相互利用を可能とするための辞書項目の共通性，辞書全体構成，品詞分類・意味分類等の体系について検討した。６０年度の調査研究成果は，「60−S002　自然言語のコンピュータ解析用辞書に関する調査研究報告書」（昭和６１年３月，財団法人日本情報処理開発協会）としてまとめられている。

本年度は，６０年度の調査研究成果を踏まえ，コンピュータ解析用辞書整備のための基本構想をとりまとめるべく調査研究を実施した。

1.2 事業計画と実施経過

(1) 事業計画

コンピュータ解析用辞書整備のための基本構想とりまとめに当たり，具体的事象から本調査研究を進めるため，本年度は下記の調査および検討を行うこととした。

① 国内調査

昨年度は特に自然言語解析のアプリケーションを限定せずに幅広くマシンリーダブルな辞書について調査を実施したが，近年，アプリケーションの一つである機械翻訳システムが商用システムとして世の中に出てきており，このような状況を踏まえて機械翻訳システムを例とした辞書について，運用面からの更新機能等のユーザの立場からの調査をさらに行い，辞書のあるべき姿を探る検討資料を得る。

なお，この調査は，機械翻訳システムのメーカ側とユーザ側に分けて行うこととした。

② 海外調査

海外におけるコンピュータ用辞書の作成，更新および標準化の動向を調査するため，調査員を，アメリカ，カナダ，欧州に派遣することとした。

③ 委員会における検討

調査研究を推進するために，当協会役職員及び外部有識者からなる「自然言語解析用辞書調査委員会」を６０年度に引き続き設置して，実施計画等を検討するとともに，「同・専門委員会」において，国内・海外調査の詳細検討と調査結果にもとづく辞書基本構想の検討を行うこととした。

(2) 実施経過

本年度は，辞書更新および辞書基本構想の調査検討に力点をおき，委員会および専門委員会における調査検討，国内の機械翻訳システムを例とした辞書更新の調査，海外の辞書作成・更新および用語も含めた標準化動向調査を行い，それらの調査結果を元にして，辞書更新のあり方，辞書整備基本構想としてのＬＤＢ（Lexical Database）の構想を試みた。

本年度の調査研究実施経過の概要を表１－１に示す。

表１－１　調査研究実施経過

委員会	回	開催日	議　　事
本委員会	1	61. 9. 4	①　委員紹介 ②　調査研究計画の検討
	2	61.12.17	①　調査研究の実施経過 ②　内外調査の実施状況報告
	3	61. 3.24	①　昭和６１年度調査研究報告書について
専門委員会	1	61. 7.11	①　調査研究の大枠について
	2	61. 9. 4	第１回本委員会と合同
	3	61. 9.18	①　調査研究計画 ②　調査研究方針
	4	61. 9.29	①　ヒアリング〔東芝㈱総合研究所　天野真家氏 ㈱市況情報センター　渡辺哲夫氏〕 〔機械翻訳システムの辞書とユーザインタフェース〕 機械翻訳システムのメーカとユーザの各々の立場から，辞書の役割，構成，更新を含めたメンテナンス，マンマシンインタフェースについて現状を解説。
	5	61.10.13	①　調査項目（メーカ／ユーザ／海外）の検討
			61.11. 1～62. 2.28　国内調査の実施
			61.11. 3～61.11.23　海外調査の実施
	6	61.12.17	第２回本委員会と合同
	7	62. 1.14	①　内外調査に基づく辞書基本構想検討項目について
	8	62. 2. 4	①　辞書基本構想の検討の枠組み ②　昭和６１年度報告書構成の検討
	9	62. 3. 3	①　昭和６１年度報告書（素案）の検討
	10	62. 3.24	第３回本委員会と合同

1.3 調査研究成果の概要

本報告書は，７つの章と資料編から成っている。ここでは，報告書の構成に沿って，本調査研究の昭和６１年度調査研究の概要を述べ，大略を把握しうるようにした。

(1) 自然言語処理応用システムと解析用辞書の現状（２章）

コンピュータで自然言語を処理する研究が発展してくるにつれて，自然言語処理を応用したシステムの開発が盛んに行われるようになってきた。自然言語処理に対する研究要素はまだ数多く残されているが，これまでの研究成果を用いても，ある程度までユーザの要請に応えうるシステムも現われてきた。そのようなシステムにとって重要な要素の一つに辞書がある。辞書の部分が充実していないシステムは，その価値が半減するといっても過言ではない。しかしながらユーザの立場から見た場合には，たとえメーカ側から見てどんなに完璧と思われる辞書が供給されても必ず不満が残る。それは，ユーザが，自分の使用条件に応じて，辞書に新たな語や情報を付け加えたり，逆に場合によっては削除したりする必要があるからである。

ここでは，ユーザ側から見た自然言語処理応用システムの現状，ならびにユーザの希望する，あるべき姿といったものを，機械翻訳システム，文書作成支援システム，データベース・アクセスシステム，等を例にとって概観している。

(2) 国内における機械翻訳システムの辞書更新の現状（３章）

自然言語解析用辞書について，効率的な構築・運用・維持等の体制について検討するため，国内の代表的なメーカ，ユーザを対象として現状及び将来望ましいと考えられる辞書の姿について調査を行った。自然言語解析用辞書を利用したシステムとしては「機械翻訳システム」がある。近年，機械翻訳システムは商用システム化が進み，一部では実用にも供されるようになっており，自然言語解析用辞書利用アプリケーションの典型的な利用例とみなすことができる。そこで，国内の商用機械翻訳システム提供企業（「メーカ」と称する）とその製品の導入企業・団体（「ユーザ」と称する）に対して，表１－２の調査項目に基づくアンケート及びヒアリングによる調査を行った。

表１－２　国内調査項目と調査対象企業・団体

	調　査　項　目	調 査 対 象
メーカ	1）辞書の種類と機能 2）各辞書のデータ項目 3）辞書のユーザへの提供方式 4）辞書更新に関するユーザへの提供メンテナンス機能 5）辞書更新に関するユーザへのメーカ側の支援体制	㈱ブラビス・インターナショナル 富士通㈱ 日本電気㈱ ㈱東芝 ㈱日立製作所
ユーザ	1）辞書更新の必要性，その程度及び内容 2）辞書更新に関するユーザ側の運用規定 3）提供されている辞書メンテナンス機能の評価 4）辞書メンテナンス機能のあるべき姿	本田技研工業㈱ ㈶電力中央研究所 ㈶日本原子力研究所 ㈱日本経済新聞社 ㈱QUICK （旧：市況情報センター）

国内調査結果のポイントは，以下に示すとおりである。

① 今回の調査において総じて言えることは，我が国においては機械翻訳システムはまだ試用の段階にあるものがほとんどであるため，機械翻訳システムの持つ辞書更新や編集機能についても，まだ実使用の経験から十分なことが言える段階ではない，ということである。いまだ，十分な市場が形成されているわけでなく，多くのユーザは試用の段階にあり，それぞれ工夫しながら実用への試みを続けているのが現状である。

② メーカ調査結果のポイント

- 今回の対象となったほとんどの機械翻訳システムは，英語－日本語間の翻訳をするもので，翻訳の対象としては，科学技術分野が多かった。
- 機械翻訳システムで使用するコンピュータは，汎用大型機，ワークステーション，パソコンであり，規模により想定するユーザ像に違いが見られる。

- 機械翻訳システムのメーカ提供辞書の更新機能については大体共通していて，辞書更新の範囲も大体ユーザ辞書を対象として，機能語以外の訳語，属性の変更であり，文法の変更は許さないというのが一般的である。また，共同開発の場合を除き，メーカから提供される基本辞書そのものを更新することはできないことが多い。
- メーカ側がある範囲の専門用語を対象とした専門語辞書を提供している場合はユーザ側の負担がかなり低減される場合もある。しかし，現在，機械翻訳システムの市場が十分形成されていない等のことから，メーカ側でユーザの実情に十分対応した辞書提供をしうるまでに至っていないため，ユーザ自身の辞書作成の負担は避けられないようである。
- 辞書更新のためのユーザインタフェースは，ディスプレイ画面に向かい対話的に行うものが代表的で，コマンドを入力する方式，メニュー方式，テンプレート方式がある。
- メーカ側で想定しているユーザ像としては，メーカ側で提供する辞書の他にユーザ自身が必要とする語彙・訳語についてはユーザ自身が辞書を作成して使用する，というものである。メーカ側はユーザに対して，現在の機械翻訳システムはユーザが自分で作りながら実用化していくものだという認識を期待しているようである。
- メーカ側の提供するサービス体制については，現在，メーカによってかなり違いがみられる。教育サポート，ＳＥサポート，苦情処理，ユーザ会等サービスの形態はいろいろある。概して，ユーザの増加に応じてサービス体制を整えつつあるというのが現状である。

③ ユーザ調査結果のポイント

- ユーザの典型的な姿としては，機械翻訳を適用したい分野を具体的に一つ持っており，メーカ提供の一般辞書と専門語辞書の上に，ユーザ辞書を構築しながら，実用化へ向けて試験的に使用しているというものである。ユーザの中である程度成功をおさめている例を見ると，機械翻訳の適用範囲，分野をかなり狭い範囲に限定している場合が見られる。

- 辞書更新については，かなり機械翻訳システムに通じており，文法の知識についてもある程度持っている人でないと難しいようである。
- ユーザの機械翻訳システム使用者は，(i)エンドユーザ(ii)翻訳者(iii)コンピュータシステムの研究開発担当者というタイプ分けができる。エンドユーザは，対象言語の知識の多少により，翻訳結果の質に対する要求が異る。翻訳者は，下訳を出させて翻訳の能力向上のために使用する。システム研究開発者は，翻訳サービスをエンドユーザに提供するために使用しているので，機械翻訳システム，文法，対象分野の知識を備えている。
- 辞書作成・辞書更新にかかわる運用体制についても，機械翻訳システム自体がまだ試用段階にあるので，運用体制と言う程の実績を持つところはまだ極く少数である。
- 辞書メンテナンス機能の問題点として，システムの仕組みを知らないと辞書更新が正しく正確に行われないこと，文法の知識がある程度必要であること，意味記述など相当の専門性が必要な項目もあり，エンドユーザの使用を困難にしている可能性があること，等を指摘することができる。
- 辞書作成・辞書更新の在り方については，ユーザの使用者の違いに応じて，意味が異なるようである。(i)エンドユーザの立場では更新機能はわかりやすく，簡便なことが望まれるが，(ii)翻訳者(iii)研究開発者の立場では，使い易さと同時に，必要なことを的確に更新できることが望まれる。

(3) 海外における辞書の作成及び標準化動向（４章）

海外における辞書の作成および標準化の動向を調査するために，田中裕一専門委員会委員を昭和６１年１１月３日から１１月２３日までの３週間にわたり，アメリカ，カナダおよびヨーロッパ諸国の関連機関に派遣した。

- 本海外調査の具体的調査目的は次のとおりである。
  ① コンピュータ用辞書の作成を行っている機関の調査
  ② コンピュータ用辞書の標準化の動向の調査
  ③ コンピュータ用辞書の更新の体制および更新用インタフェースの調査

④　ウォータールー大学でのコンピュータ用辞書に関する国際会議への出席

- 調査対象機関および調査日程は次のとおりである。

①　１１月４日：国立医学図書館（米国）

②　１１月５日：アメリカ保健機構（米国）

③　１１月６日：ホートン・ミフリン出版社（米国）

④　１１月７日：モントリオール大学（カナダ）

⑤　１１月１０日～１１日：ウォータールー大学（カナダ）

⑥　１１月１３日：ロングマン出版社（英国）

⑦　１１月１４日：オックスフォード大学出版部（英国）

⑧　１１月１７日：EUROTRA プロジェクト（ルクセンブルグ）

⑨　１１月１７日：EURODICAUTOM プロジェクト（ルクセンブルグ）

⑩　１１月１８日：ボン大学（西独）

⑪　１１月１９日：ジーメンス（西独）

⑫　１１月２０日：INFOTERM（オーストリア）

⑬　１１月２１日：ピサ大学（イタリア）

- 辞書更新体制についての態度は出版社と研究機関とで大きく異なる。出版社においては辞書は商品であり，その内容に関して責任を持つと同時に，その内容をユーザには公開しない。

一方，実際に日常的に運用されているシステムの場合は更新の体制が確立されていることが多い。更新はユーザ側でなく，システムの提供者（開発者）側で行われる。米国の国立医学図書館やＥＣ（ヨーロッパ共同体）のEURODICAUTOMのように大きな組織の大規模な辞書と，アメリカ保健機構のように小規模な組織の違いはあっても，辞書の更新は一つの組織内で一元的に管理され，その責任も明らかになっている。

研究機関において開発途中の辞書の場合には更新の体制はほとんど無きに等しい。

- 辞書更新インタフェースについても，組織による差が現れている。出版社の場合には「原稿に朱を入れる」感覚が残っているようであり，編集者は計算機の出力用紙上で校正し，それを専門のオペレータが計算機

に入力していた。その入力にはワークシート形式の単純なエディタが用いられているようである。IBMマシン上で動くNew OED用のエディタはカラー・ディスプレイを活用して記述項目の違いやフォントの違いを明示し，テキストの表示を大域的にも局所的にも切り換えることができる見事なものである。操作も非常に簡単で，編集室では出版社の辞書編集者が端末の前に座り，自らキーボードを操作していた。

また大規模な組織では更新のインタフェースは一般にさほど重んじられていない。これは更新の手続きがユーザからは切り離されているため，むしろ更新内容の妥当性の吟味等に重点が置かれ，更新の作業自体はやはり専門のオペレータに任せることになるからである。

アメリカ保健機構の機械翻訳システムはやや特殊である。この小規模なシステムはここで開発されたものだが，現在のところ機構内でのみ使われているので，辞書に規模，用法上の階層はなく（一方向の翻訳に対して）一種類で済んでいる。従って更新はユーザ辞書と同じであると思ってよい。

大学の辞書システムには更新のために便利なインタフェースを用意しているものが多かった。しかし，ボン大学には大規模な辞書があり，大型機上で検索，更新が一応できるようになっているが，このエディタは古色蒼然たる代物で，ほとんど使われていないようだった。人材不足が原因かと思われる。

- 大規模化および標準化への動きも，海外で活発化しつつある。

日本における計算機用辞書の環境と英語圏のそれとを比較すると，最も顕著な差はファイル化された辞書（MRD；Machine-readable Dictionary）が英語圏では，数多く出版されていることである。これは有名なLDOCE（Longman Dictionary of Contemporary English）を始めとして，ウェブスター，アメリカン・ヘリテージ等の冊子体辞書として既に広く使われているものが多く含まれている。現在，計算機用辞書の大規模化や標準化が声高に求められているのは，主に機械翻訳システムの開発を通じてこのようなMRDにおける経験が基になっているものと考えられる。

詳しくは後のＬＤＢ（Lexical Database）の章で述べるが，冊子体辞書の構造，内容に多く依存したＭＲＤに代るものとして，語に関する情報を計算機用のデータベースの形で提供する新たなＬＤＢなるものの構想がある。これは必然的に大規模化，標準化を伴うが，さまざまな組織でそのための検討が進められている。

例えば，ＥＳＦ（European Science Foundation ―ヨーロッパ諸国のNational Science Foundationの合同組織）が主体となって開かれたフォーラム"計算機用辞書の標準化"などはその動きを示すものである。また，計算機用辞書における文法事項の標準化は可能かという問題も（本委員会でも昨年度の調査事項に含めていたが）さまざまな言語理論の専門家を集めて検討しようとしているということである。

(4) 機械翻訳システムを例とした辞書及びその更新のあり方（５章）

国内外の辞書に関する調査結果を踏まえ，機械翻訳における辞書のあり方および辞書の更新について問題点とあるべき姿を示した。

- 管理の上での辞書の構成を，基本辞書，ユーザ辞書，グループ辞書，専門語辞書，プライベート辞書に分類して示した。（図１－１）
  - ・基本辞書 － 機械翻訳システム本体と共に供給される翻訳用辞書。
  - ・ユーザ辞書 － 利用者にとって所与の辞書体系（基本辞書や上位のユーザ辞書としての「グループ辞書」，やはり上位に設定された専門語辞書）に対し，その利用者にとって不足の語や不都合な辞書内容を調整した辞書。
  - ・グループ辞書 － 機械翻訳システム利用者の集まり（グループ）の限定された目的に沿って，基本辞書あるいは上位のグループ辞書を拡大・整備させた辞書を「グループ辞書」と称し，最終利用者が各人の責任の中で運営していくプライベート辞書と区別する。
  - ・専門語辞書 － 専門用語に代表される，分野毎の辞書データであり，主に個別の機械翻訳システムとは独立に作成

され，必要に応じて最小限の変換が為されるもの。

・プライベート辞書 ― システムの試験的な利用や，純粋に個人的な利用のために作成される辞書を正式な運用基準から除外する意味でプライベート辞書と呼ぶ。

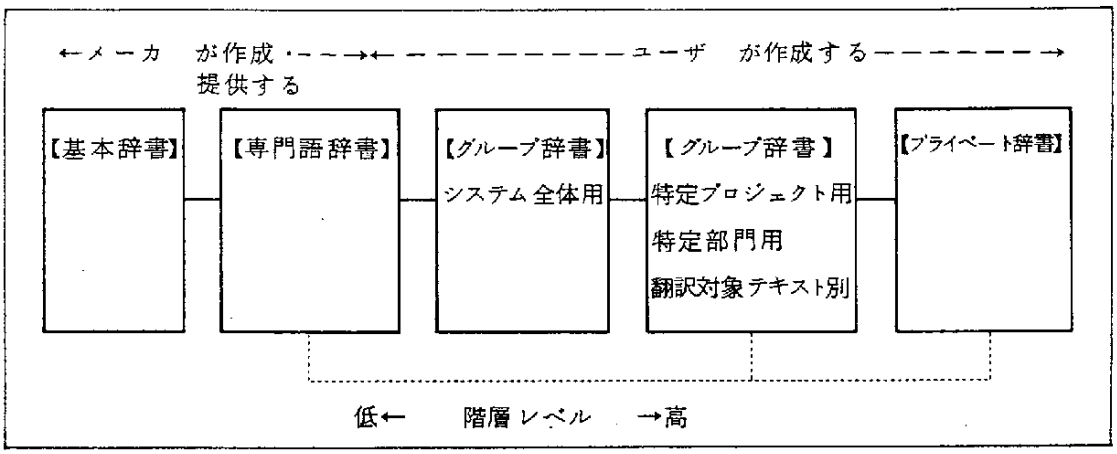


図１－１　辞書の分類

- 機械翻訳システムで作成される辞書の種類を，図１－１のように分類して，各々の辞書の作成について，概略，手順，組織，留意点を述べた。
- 辞書の更新については，機械翻訳システムの辞書更新の問題に焦点を当てて，まず，その必要性を論じ，ついで、ユーザの立場からの更新作業の特質とそのためのサポート機能について述べ，最後に，それとは逆に，機械翻訳システム開発者の立場から行われる辞書更新の問題について考察した。

ユーザによる辞書の更新は，組織的に実施しようとすると，時間・作業量の面で非常にコストがかかることが知られており，組織的な辞書の更新はグループ辞書単位以上の規模で実施することが望ましい。しかし，辞書更新の作業手順は，決定的なものは確立していない。

ユーザで辞書更新を効率的に行うために必要なサポート機能としては，バッチ型／対話型辞書更新，原文／訳文編集，文法情報・属性情報付加支援，既存辞書の内容参照，各種辞書の管理，原文・訳文に対する統計

分析，等の諸機能が考えられる。

機械翻訳システム開発者による更新に関する課題としては，文法情報との整合性の管理，翻訳例文の管理，複数利用者からの要請の調整，翻訳辞書のもととなった辞書との整合性管理，等が考えられる。

ところで，機械翻訳システムのユーザやメーカが辞書の作成・更新を遂次行ない，様々な辞書が数多く世の中に存在するようになると，ユーザ辞書やユーザ辞書インタフェースの整合性管理が非常に困難になってくる。そうなると，それらの整合性を保つ共通基盤となるものが望まれることになるが，後述するＬＤＢがその役割を担うものとして期待される。ＬＤＢは必らずしも，機械翻訳システムを含む自然言語処理システムのためのみを目的としているものではないが，ＬＤＢによってこれらシステムのための辞書の，メーカ間及びユーザ・メーカ間インタフェースを容易ならしめる共通基盤が提供されることになろう。

(5) ＬＤＢ（Lexical Database）の構想（６章）

- 計算機による自然言語処理を行うためには対象となる言語の辞書が必要である。このために冊子体の辞書を計算機で読むことのできる形に変換した機械可読辞書（ＭＲＤ；Machine-readable Dictionary）が従来から利用されてきた。しかし，深く広い自然言語処理のためにはＭＲＤでは不十分であり，最近では大規模な語彙に関するデータベース（LDB；Lexical Database）が求められている。
- ＭＲＤは自然言語処理システムの研究開発にとっても利用価値のあるものであるが，①語に対するデータが少ない，②構造化されていない，③冊子体辞書の持つ制約を逃れることができない，④記述が人間を対象としている，⑤語彙数が少ない，等の数多くの欠点がある。
- 以上のようなＭＲＤの欠点を克服するためには，計算機による自然言語処理のために適切な記述方法と十分な語彙を持ち，厳密に構造化されたデータベースが必要である。現在の自然言語処理技術，データベース／知識ベース技術などの水準は，工学的にそれを可能にしていると言える。ＬＤＢはこのような要請のもとに構想されているものであり，言語情報に関する統合的なデータベースである。ＬＤＢに望まれる資質とし

ては，①標準的なデータベース，②一元的な管理形態，③応用辞書の作成が容易，④機械用のものも人間のものも，高品質の辞書が作れる，⑤豊富な言語情報を有する，⑥広い応用分野を有する，等が考えられる。

- 自然言語処理用として電子化辞書（ＥＤ；Electronic Dictionary）というものがある。冊子体辞書からの機械可読化であるＭＲＤと異なり，電子化辞書は初めからコンピュータ処理を意図しているもので，コンピュータで自然言語（人間が日常読み書きする言語）を処理するために必要な語についての基本データを，コンピュータで扱うことのできる形に記述した辞書である。もっとも，自然言語の処理の深さにはいろいろのレベルがあり，それに応じて電子化辞書にもいろいろの種類があり，各機関で開発が試みられている。

ＬＤＢは，これら近い将来開発されるであろう各種電子化辞書を統合的にデータベース化し，さらに豊富な言語データを持つよう拡張し，語彙に関する情報を統一的に網羅的に提供するものと理解されよう。

- ＬＤＢを開発しようという動きは各国に見られる。日本における関連動向としては，国立国語研究所により言語データの整備・蓄積が言語学的観点から進められており，これはＬＤＢへの重要な基盤となりうる。

電子化辞書（ＥＤ）に関しては，電子技術総合研究所や㈱新世代コンピュータ技術開発機構，等で各種試作が試みられ，基本的な技術が蓄積された。これらの成果等を含め，昨年度に㈱日本電子化辞書研究所が設立され，本格的に，日本語・英語に関する電子化辞書の研究開発が着手された。これらは，情報処理技術的観点からのＬＤＢへの重要な基盤となる。

ＭＲＤに関しては，各出版社から手持ち冊子体辞書の機械可読化という形で多種のものが商品化されるようになっている。

海外では，ホートン・ミフリン出版社，オックスフォード大学出版局，ピサ大学などで，ＬＤＢ構想が検討されている。

このように，内外でＬＤＢへ向けての動きがあるが，それは自然言語処理技術の発達がＬＤＢを必要とするフェーズに差し掛かったことが大きな理由と考えられるが，それと同時にＭＲＤ等の利用経験とそれに伴

う資源の蓄積がＬＤＢの開発を可能にしていることも事実であり，このようなＬＤＢの開発は世界的な趨勢である。

(6) 今後の課題（７章）

前述した調査研究の中から，ユーザやメーカで辞書の作成・更新がさらに進められていくと，辞書の整合性管理が大変困難になるため，各種インタフェースを含めた共通化，標準化を計れるものの必要性が認識され，また，内外の各機関での動向はＬＤＢ（Lexical Database）の開発へ向かっていることも認識された。

このことから，本年度調査研究のテーマである「辞書整備基本構想」の基本概念としてＬＤＢを採用することとした。

したがって，「今後の課題」として，ＬＤＢを中心に据えて，(1)ＬＤＢの開発・維持・管理に関する技術的課題，(2)ＬＤＢの開発・維持・流通に関する体制上の課題，(3)ＬＤＢの波及効果に関する課題―を指摘する。

今後の課題

① ＬＤＢの開発・維持・管理に関する技術的課題
- 言語情報の体系化に関する技術
- 知識ベース，データベース技術
- ネットワーク技術
- 電子化辞書（ＥＤ），ＭＲＤの現状と動向

② ＬＤＢの開発・維持・流通に関する体制上の課題
- ＬＤＢの開発・維持に関する組織
- ＬＤＢの管理体系
- 辞書，ＬＤＢの流通の仕組み
- 産学官の役割分担

③ ＬＤＢの波及効果に関する課題
- ＬＤＢの利用発展性
- 言語情報の産業化

2．自然言語処理応用システムと解析用辞書の現状

2. 自然言語処理応用システムと解析用辞書の現状

コンピュータで自然言語を処理する研究が発展してくるにつれて，自然言語処理を応用したシステムの開発が盛んに行われるようになってきた。自然言語処理に対する研究要素はまだ数多く残されているが，これまでの研究成果を用いても，ある程度までユーザの要請に応えうるシステムも現われてきた。そのようなシステムにとって重要な要素の一つに辞書がある。辞書の部分が充実していないシステムは，その価値が半減するといっても過言ではない。しかしながらユーザの立場から見た場合には，たとえメーカ側から見てどんなに完璧と思われる辞書が供給されても必ず不満が残る。それは，ユーザが，自分の使用条件に応じて，辞書に新たな語や情報を付け加えたり，逆に場合によっては削除したりする必要があるからである。ただ現在のシステムでは，これらの辞書の実体が必ずしもユーザ側によく見えている訳ではないことは，次章の機械翻訳システムに対する調査結果からも推察することができる。ある程度まで辞書の内容に通暁していないと，辞書を拡張，修正していくことは困難になるのである。そこで，ここでは，ある程度理想的なユーザになって，ユーザ側から見た自然言語処理応用システムの現状，ならびにユーザの希望する，あるべき姿といったものを簡単に概観する。また逆に，自然言語処理システムの現状を列挙することによってもユーザ像が明らかになってくる。なお，自然言語処理の応用システムというと，何といってもまず第一に思い浮かぶのは機械翻訳システムであるが，実際のシステムに対する調査結果は，次章に詳しく述べられているし，辞書とその更新のあり方についても第5章で述べられるので，ここではシステムの必要性など全般的なことを記述するにとどめることにする。

2.1 機械翻訳システム

すでに述べたように，現在自然言語処理の応用システムといったときにすぐ考えられるのは，機械翻訳システムである。これは，機械翻訳システムが，自然言語処理の応用として適した分野であるという，研究側からくる必然性だけではなく，社会的な方面からの種々の必要性があるからである。まず第一に，近年国際交流が活発になり，日本企業の海外進出，海外

企業の日本への進出が増大していることがあげられる。また，輸出，輸入など，ものの動きも増大している。このような国際交流の活発化に伴って，コミュニケーションの円滑化をはかるために，必然的に翻訳の需要は増大する。輸出入でも，たとえば電子機器の場合には，マニュアルなどの翻訳が必要になる。また，最近では，ものの動きだけではなく，情報の国際交流も盛んであり，こちらの面からの要請も多い。

このような大量の翻訳を迅速にこなすためには，人手にたよるだけではとても応じきれない。そこで機械翻訳によって，翻訳にさくコストを軽減する試みがなされている。機械翻訳には，もう一つ大きなメリットがある。それは，人間が大量の翻訳を手分けして行う場合に比べて訳語の統一がとりやすいということである。特に同じ文書の中で専門用語に不統一な訳語が与えられていると，せっかくの翻訳をうまく生かせないことになるからである。

機械翻訳には，メーカによって種々のやり方があるが，非常に簡略化した形で書けば，おおむね図２－１に示すような形で処理が進行しているものと思われる。この中で，特に辞書のところに注目してみると，解析，生成部分で辞書が分かれているのはもちろんであるが，その辞書の中が基本辞書とユーザ辞書に分けられる。基本辞書は，形態素解析，形態素生成，また文法部分で主として利用され，次のようなものを含む。

- 単語，合成語のつづりに関するもの
- 品詞，活用，係り受け属性に関するもの
- 上位概念や他の語概念との隣接関係を示すもの

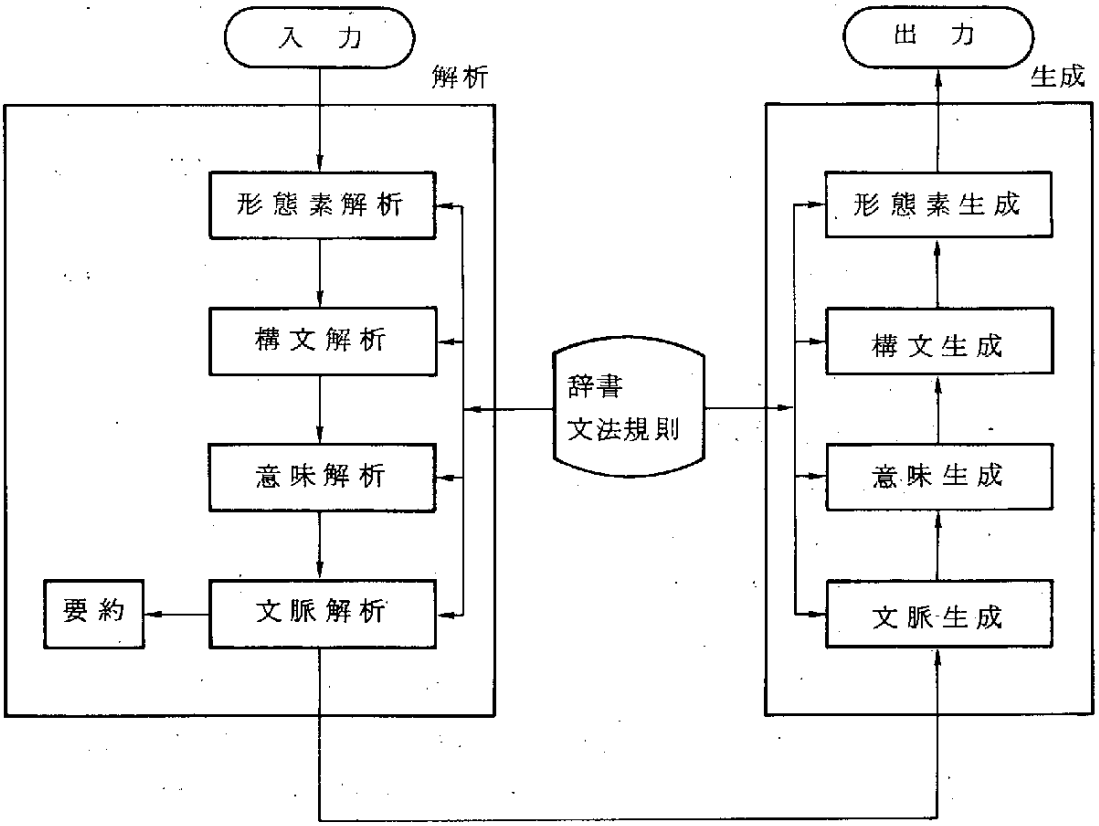


図２－１　機械翻訳システムの処理概要

一方ユーザ辞書は，機械翻訳システムのユーザが翻訳の効率アップ，品質向上のためにユーザ独自で登録するもので，これをさらに分けると，一般用語辞書，専門分野辞書に分けられる。前者は，もともとメーカから提供された辞書に拡張，修正を加えることが多い。市販の翻訳システムでは，実際に登録できる語彙を名詞に限っているケースが多いが，動詞・形容詞・副詞などまで解放しているものもある（次章の国内調査参照）。また，専門分野辞書は，メーカによっては大雑把な専門分野を分けてそれ向きの辞書を提供している例もあるが，細かい分野になると訳語が異なる場合や，ユーザによってはある語を社内的に，一般に用いられているのとは異なる語で呼んでいたり，逆に一般用語を特殊な意味に使用したりする場合があ

るので，最初からメーカ提供の専門分野辞書を購入してそれを用いるか，すべて自社開発するかに分かれる可能性がある。後者の場合には使用するシステムの内部的なアルゴリズムをある程度知っているか，専門分野辞書を作るための適切な支援システム（たとえばエディタ）がメーカ側から提供されている必要がある。エディタを含むユーティリティが豊富であれば，ユーザ辞書の修正，拡張にも益するところが多いと思われる。それも単なるエディタでなく，類似の単語を検索してそれを参照しながら新しい単語を付加していくといった，ユーザフレンドリーなものが望ましい。

機械翻訳システムには，バッチ方式で大量の翻訳を一度に行うものと，ユーザと対話的に翻訳処理を行うものとがある（もちろんその中間的な折衷型もある）が，現状では，前者の場合でもプレエディットないしはポストエディットが不可欠である。このようにコンピュータと人間（ここではこのような作業をする人間をオペレータと呼んでおく）が接する場面でも辞書の果たす役割は重要である。それは上記のような訳語の統一といった場面だけではなく，誤訳，翻訳不能，未定義語の出現などといった場合にも役に立つはずである。そのためには（翻訳専門のみならず，ある程度一般的知識を含んだ）辞書と有機的に結合した，機械翻訳分野におけるいわば翻訳支援システムとでもいったものが必要になる。以下は，現状というより，いささか理想論になるが，このような場面でのシステムや辞書のあり方といったものを述べることにする。なおこの内容は，次節に述べる文書作成支援システムとも種々の面でかかわりを持つものであり，共通した内容も多く含まれているので，後述するように，将来は結合することも望ましい。

機械翻訳の出力をオペレータが見る場合には（多くはネイティブな言語の出力を見る場合であろう），まず構文，意味的に正しい文かどうかといったことをチェックするであろう。そのときにオペレータにとって理解できない単語や構文が出てきたときには，当然システムに対して質問を発するはずである。その場合に，システム側としては，辞書に書かれた知識を援用して，そのような訳語が出てきた背景，そのような構文になった理由などをある程度説明できることが望ましい。構文や意味など，問題が文全

体に及ぶ場合には，次の候補を出力するとか，辞書だけでなく，他の部分の途中経過なども必要に応じて取り出せるのが理想的である。このような支援システムを作る，あるいは機械翻訳のユーティリティの一つとして持つためには，これまでの機械翻訳用の辞書のみでは不十分で，もっと大きなバックグラウンドの辞書というものを考える必要がある。現状では（おそらく将来も），構文解析や意味解析の部分は，実は辞書と密接に関連してお互いに補完しながら解析結果を作っているにもかかわらず，それがユーザサイドに解放されることはまずないので，機械翻訳の品質向上をユーザが図る場合には，どうしても辞書の部分にたよらざるを得ないが，もしこのような一般的知識までを持つ辞書というものが機械翻訳と密接に結び付いた支援システムという形で提供されれば，機械翻訳を行うユーザにとって益するところは大きいであろう。

2.2 文書作成支援システム

コンピュータで文書を作成するといった場合に，誰もが真っ先に頭に思い浮かべるのはワードプロセサであろう。日本語のワードプロセサが出現してからそんなに日がたっていないにもかかわらず，今日ワードプロセサのない職場は少ないといえるほど爆発的に普及した事実が，この分野での必要性を物語っている。しかしながら，ワードプロセサといったときに，自然言語処理とのかかわりを論じることは少ない。それは，現在のワードプロセサが，自然言語処理の最も重要な側面である，構文解析や意味解析の部分を，コストがかかったり，難しすぎるという理由で，いわば「避けて」通っているからであると思われる。ここではもう一段高いレベルから見た，文書の作成を支援するというシステムを（これもある程度理想像ながら），辞書という側面を中心に述べてみよう。

まず文書作成支援システムといった場合には，次のようなシステムが考えられる。つまり，文書を作成する場合に，用字・用語がある基準に従って統一されるようにシステムからアドバイスを受けることができ，また形式の決まった文書では，その形式を教えてくれるようなシステムである。これには，次のような各分野からの必要性がある。まず，新聞，雑誌など

のジャーナリズムや法令関係者にとっては，用字・用語に一定基準を保っていてしかも分かりやすい文書を受け取り，また作成する必要がある。またその他の企業においても，その企業独自の形式を持つ社内文書，一般的に形式が決まっているがその形式に従って書く事に熟練技術を要する商用の手紙文，契約書などは表記の形式がすぐに取り出せて，できれば必要な言い回しなどが参照できるシステムが望ましい。さらに研究分野では，学会投稿の論文などの形式を整えるために無駄な時間を費やすよりも，そのような形式が自動的に示され，フォーマッティングしてくれるようなシステムがあれば理想的である。

以上述べたようなシステムを使用した場合の効果としては，次のようなことが考えられる。

- 重要な文書の形式のチェック。
- 不要文字の削除，文字の抜けの追加。
- 同音異義語のチェック。
- 係り受けが複雑で分かり難い文の抽出，長すぎる文のチェック。

このように，形式，表記，用語を統一し，登録しておくことによって再利用を可能にすることは，多量の文書を効率よく処理することにつながり，作業の効率アップとなる。このようなシステムの処理のやり方は，概略図２－２に示すようなものになる。その中で辞書の果たす役割としては，上記のような，いわゆるスペルチェックのための辞書としての機能（英語などでは単語単位に区切って書かれるが，日本語のような言語では，単語の区切りが明確でないので英語に比べて難しい）や，文字のチェックなどがある。商用手紙文のうちで決まった文を登録しておいて出力するなどというのも，広義には「手紙辞書」と解釈できなくもない。前節の機械翻訳のような辞書の分け方をすれば，基本辞書，付属語辞書，活用語尾，接尾語辞書などが考えられるであろう。もちろん機械翻訳で用いられたような構文規制や意味規則が入っていることが望ましい。また，文書を処理する場合には，それに付随する図や表，数式，特殊文字（たとえば上つき，下つき，ルビなど）も処理できる機能が含まれている必要がある。

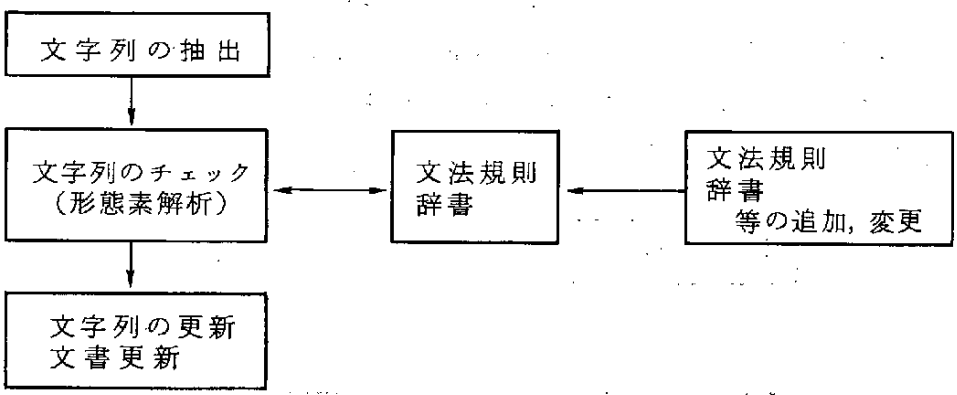


図2－2　文書作成支援システムの処理概要

2.3　データベース・アクセスシステム

近年専用回線を通じた大規模なネットワークが普及するにつれて，それを用いて中央においた大規模なデータベースにアクセスして必要な情報を瞬時に取り出そうとするシステムが発展しつつある。システムが大規模になるにつれて，このようなデータベースには，昔とは違って専門家でない人間が触れる機会も増えてきた。このような場合にインタフェースとして考えられるのが，適切なメニュー画面を持ったシステムか，あるいは自然言語処理でのアクセスである（両者を備えていればもっとよい）。現在のデータベースシステムへのアクセスは，一見自然言語風の問い合わせ言語を備えているが，キーワードの順序が変わったり，語の省略が起こると問い合わせの意味を解釈できなくなる。従って，このような変化にも柔軟に対応できる自然言語処理を応用したデータベース・インタフェースシステムが必要になる。この種の研究は，欧米では進んでいるが，日本ではまだこれからの分野である。それは，日本語での入出力が欧米の言語に比べて困難なのもその一因と思われる。しかしながら音声入出力も含めて，自然言語によるデータベースアクセスは今後需要が増大するものと考えられる。

データベース・アクセスシステムの処理の概要を図2－3に示す。このシステムは，機械翻訳システムの場合と違って，ユーザが直接システムと

対話する場合が多いので，リアルタイム性がより重要視される。従って，そのような要求に対処できる柔軟な辞書構成にしておかなければならない。また，辞書自身をデータベースの基本データとして用いる場合も数多く存在する。その場合には，辞書は，解析と検索という二重の役割を持つようになる。

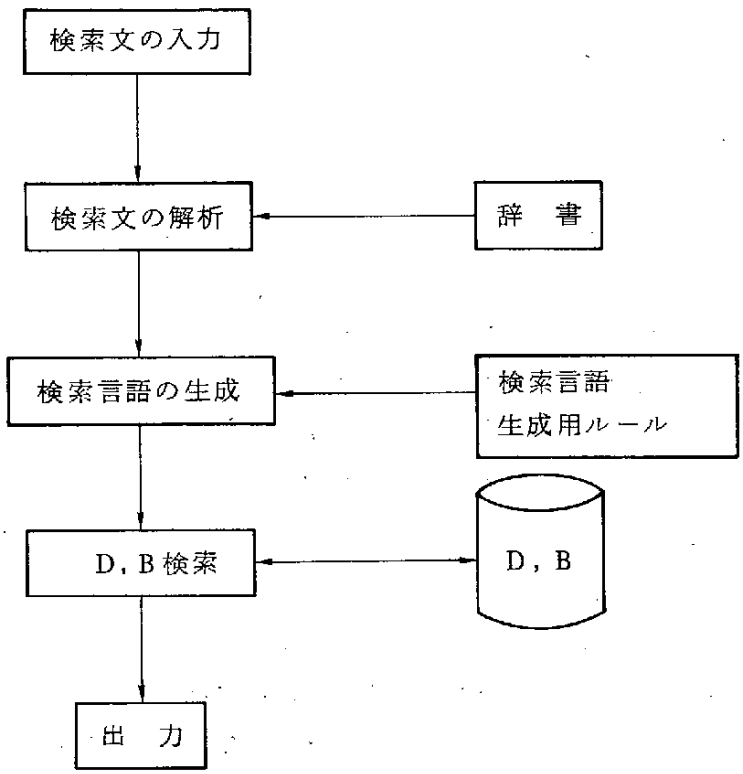


図２－３　データベース・アクセスシステムの処理概要

2.4　その他のシステムと今後の問題点

以上のように，自然言語処理の応用システムの代表例として，３つのシステムを取り上げて，それぞれのシステムの簡単な処理方法と，その中で辞書の果たす役割について述べた。いずれの場合にもシステムの中で辞書の果たす役割は類似しており，システムの中で重要な位置を占めている。従って個々のシステムを取り上げるよりも，あるシステムについて辞書の役割を論じた方が問題の焦点が明らかになるものと思われる。ここで述べ

たシステムの中で，機械翻訳システムは，現在最も普及し，需要の高いシステムである。このシステムを集中的に取り上げることは，上に述べた趣旨とよく合致するので，次章以下では機械翻訳システムを中心に述べることにする。機械翻訳システムは，他のシステムと同じように，その「使い心地」とでもいったものが辞書によって左右されるといっても過言ではない。実際のシステムの「使い心地」は，次章で示されるが，現状ではまだ改善すべき点は多い（それについては第5章で述べられる）。

自然言語処理を応用したシステムとしては，そのほかにも種々のものが考えられる。たとえば，前節に述べたデータベースアクセスのちょうど逆のものとして，データベース登録システムがある。これは広くとらえると，キーワードや抄録の自動作成システムにまでつながる話であるが，まだ研究をこれから進めようという段階である。またスペルチェックという面からは，ＯＣＲで入力した文字のチェックに自然言語処理を応用することも考えられる。さらに入出力として音声を用いる場合には，当然自然言語処理がかかわって来る。そのほかにもＣＡＩに自然言語処理を用いるとか，外国語教育に応用するなどが考えられる。また．当然のことであるが，これらの個別システムを有機的に結合したシステムも考えられる。たとえば，機械翻訳の入出力部に，文書作成支援システムを取りつけるなどといった具合である。

このように，自然言語処理のさまざまな応用例から辞書というものを見ていくと，同じ辞書という言葉で表現されるものでもその内容は非常に多岐にわたっていることが分かる。しかもそれらの辞書を作成するのに，莫大な時間と労力を要することも容易に理解されよう。しかもそのように多岐にわたった辞書でも，よく考えてみると，その根底には，ある共通性が浮かび上がって来る。それは，人間の持つ知識を従来の冊子体の辞書からさらに発展，洗練させた形でコンピュータの中に蓄えるということである。このようなことは，最近では世界各国で認識されはじめている。第4章における諸外国の例は，外国でのこのような認識をある程度裏付けてくれるであろう。また，この概念をさらに発展させると，大規模な知識をいわばデータベース的に蓄えておいて，必要に応じて自由にその知識を活用して

いくという，Lexical Database（LDB）の考え方にもつながっていくことになる（第6章参照）。

自然言語処理システムは，今後とも上に述べたような各分野において発展し，次第に理想的な姿に近づいていくものと思われるが，それを支えていくのがさまざまなアルゴリズムであるとともに辞書である。特に辞書は，ユーザ側からシステムを改善できる唯一の手段であり，本章ならびに他章で指摘された数々の問題点を突破することによって，ユーザ側の不満を次第に解消する主要な役割を担うものである。

3．国内における機械翻訳システムの辞書更新の現状

3. 国内における機械翻訳システムの辞書更新の現状

3.1 国内調査の概要

(1) 目　　的

自然言語解析用辞書について，効率的な構築・運用・維持等の体制について検討するため，国内の代表的なメーカ，ユーザを対象として現状及び将来望ましいと考えられる辞書の姿について調査を行った。自然言語解析用辞書を利用したシステムとしては「機械翻訳システム」がある。近年，機械翻訳システムは商用システム化が進み，一部では実用にも供されるようになっており，自然言語解析用辞書利用アプリケーションの典型的な利用例とみなすことができる。そこで，国内の商用機械翻訳システム提供企業（「メーカ」と称する）とその製品の導入企業・団体（「ユーザ」と称する）に対してアンケート及びアンケートに基づくヒアリングによる調査を行うことにした。

(2) 内　　容

機械翻訳システムのメーカとユーザの各々に対してアンケート調査及びヒアリング調査を実施した。調査項目の詳細については，巻末にあるアンケート調査票（メーカ用，ユーザ用）のとおりである。各調査票の調査項目は簡単にまとめると以下のとおりである。

① メーカに対する調査項目

1）辞書の種類と機能

2）各辞書のデータ項目

3）辞書のユーザへの提供方式

4）辞書更新に関するユーザへの提供メンテナンス機能

5）辞書更新に関するユーザへのメーカ側の支援体制

② ユーザに関する調査項目

1）辞書更新の必要性，その程度及び内容

2）辞書更新に関するユーザ側の運用規定

3）提供されている辞書メンテナンス機能の評価

4）辞書メンテナンス機能のあるべき姿

(3) 方　法

① 調査方法

機械翻訳システムのメーカ及びユーザに対して各々，アンケート調査及びヒアリング調査を行い，その結果をまとめた。メーカに対しては，商品として販売しているシステム（英→日，日→英又は両方）について公表可能な範囲においてアンケート調査票に回答を記入していただき，これを基にヒアリングを実施した。その他システムのデモンストレーションを拝見したものもある。ユーザに対しても同様の方法をとったが，実際にシステムの運用テストをさせていただいたものもある。

さらに，調査対象のシステムのユーザーズマニュアルを購入して（1社を除く）調査の参考とさせていただいた。

② 調査対象

調査対象企業・機関として機械翻訳システムのメーカ／ユーザ各々5社ずつ選定した。選定の基準は以下のとおりである。

〔メーカ〕 調査開始時点（昭和６１年１０月）に既に機械翻訳システムを商品として販売している企業で，主要なもの。

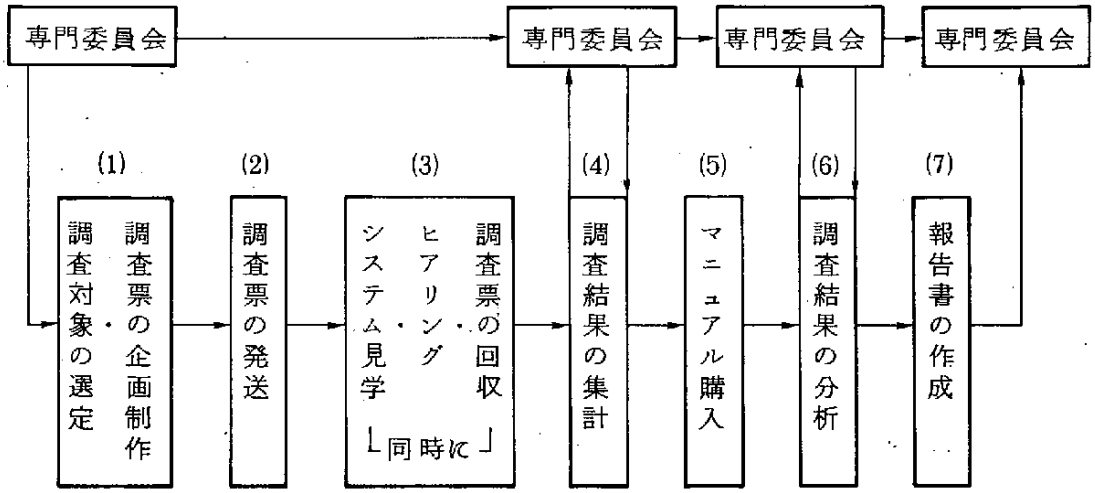
〔ユーザ〕 機械翻訳システムを導入している又は，実用化システムの運用を行っている企業・機関。

調査対象として選定した企業・機関及びシステムは，表３－１に示すとおりである。

表３－１　国内調査対象企業・機関一覧

	対象企業 / 対象システム等	A	B	C	D	E
メーカ	日→英・システム名称	1.ミディアムパック J／E 2.マイクロパック J／E	ATLAS－I	PIVOT	――	HICATS／JE
	発表日	1. 84.5 2. 85.6	84.9	85.5	――	86.4
	発売日	1. 84.6 2. 85.8	84.9	85.12	――	86.5
	英→日・システム名称	――	ATLAS Ⅱ	PIVOT	TAURAS	――
	発表日	――	84.9	85.5	85.5	――
	発売日	――	85.6	86.9	85.12	――
	ヒアリング実施日	86.12.4	86.11.28	86.12.1	86.11.28	86.12.2
	備考	マニュアル未入手				
	対象企業・機関 / 対象システム等	F	G	H	I	J
ユーザ	日→英・システム名称	マイクロパック J／E	ATLAS－I	ATLAS－I	日経－CSK 機械翻訳システム（共同開発）	HICATS／JE（共同開発）
	導入日	85.11	（試験導入）82		86.9	79
	英→日・システム名称	――	ATLAS Ⅱ	ATLAS Ⅱ	――	――
	導入日	――	85		――	――
	ヒアリング実施日	――	86.12.5	86.12.19	86.12.11	86.12.1
	備考	先方の都合によりアンケート調査票による回答のみ				

③ 調査フロー



④ 備　考

1）ユーザ的立場でメーカとの共同開発体制をとっている企業はユーザとみなした。

2）ユーザ等の協力により，システム操作や運用を見学したが，これは各システムの性能を比較する趣旨のものではない。

3）調査の参考にシステムのユーザマニュアルを購入したが，このユーザーズマニュアルも調査結果の分析以外の目的に利用したり，外部に公開することはない。

4）調査票等に関する資料原本は，外部に公開することはない。

以上，調査対象企業・機関の皆様には，貴重な時間を本調査にさいていただくとともに貴重なご意見，資料を提供いただきました。まことにありがとうございました。

3.2 国内調査結果

3.2.1 国内調査結果の概要

今回の調査において総じて言えることは，我が国においては機械翻訳システムはまだ試用の段階にあるものがほとんどであるため，機械翻訳システムの持つ辞書更新や編集機能についても，まだ実使用の経験から十分なことが言える段階ではない，ということである。しかしながら，前節の調査対象選定基準に示したように，対象としたメーカ／ユーザは現状では実使用に近い企業・機関であると考えられることから，現状把握のための資料を提供しえると考える。

以下に，今回の調査で主に目についたポイントを中心に概要をまとめてみる。

(1) メーカ調査結果の概要

- 今回の対象となったほとんどの機械翻訳システムは，英語一日本語間の翻訳をするもので，翻訳の対象分野としては，科学技術分野が多かった。使用コンピュータの規模としては，

  ① 汎用大型機

  ② ワークステーション

  ③ パソコン

  の3種類が考えられ，それぞれの規模によって想定するユーザ像にも，違いが見受けられるようである。

- 次に，機械翻訳システムが持つ，辞書関係については，メーカ提供辞書の更新機能については大体共通していると言える。辞書更新の範囲も大体，機能語以外の訳語，属性の変更であり，文法の変更は許さないというのが一般的である。辞書内容の更新は大体ユーザ辞書を対象とするもので，メーカから提供される基本辞書そのものを更新することはできないことが多い。更新のための手続きはディスプレー画面に向かい対話的に行うものが代表的である。

- 辞書更新のためのユーザインタフェースについては，いくつかの方式がある。大きく分けると，コマンドを入力する方式，メニュー方式，テンプレート方式がある。テンプレート方式のほうが記憶する量は少なく

て済むが，操作は煩雑という意見もあった。コマンド方式は必要なコマンドをいちいち覚えていなければならない。大量のデータについては，バッチ的に一括処理することも可能である。

- 次に，ユーザとの関係についてのべる。メーカ側で想定しているユーザ像としては，メーカ側で提供する辞書の他にユーザ自身が必要とする語彙・訳語についてはユーザ自身が辞書を作成して使用する，というものである。メーカ側である範囲の専門用語を対象とした専門辞書を提供している場合はユーザ側の負担がかなり低減される場合もある。しかし，現在メーカ側の対応は個々のユーザの実情を十分考慮している状態ではないため，ユーザ自身の負担は避けられないようである。

  メーカ側では，現在の機械翻訳システムはあくまでも翻訳の支援システム的なものであり，ユーザもそういうことを納得の上で，システムの特徴を理解して使って欲しいという意見が多かった。また，現在の機械翻訳システムはユーザが自分で作りながら実用化していくものだという認識を期待しているようである。メーカ側の提供するサービス体制については，現在，メーカによってかなり違いがみられる。教育サポート，ＳＥサポート，苦情処理，ユーザ会等サービスの形態はいろいろある。概して，ユーザの増加に応じてサービス体制を整えつつあるというのが現状である。

- ユーザが作成した辞書の活用の問題についても，意欲を持っているメーカは存在するが，時期尚早というのが実情のようである。ユーザと共同で実用化をはかっている所を別とすればユーザ辞書の作成はほぼ全面的にユーザの手にゆだねられており，辞書作成，辞書更新の教育サポートがあるのみというケースが多いようである。

(2) ユーザ調査結果の概要

- ユーザの典型的な姿としては，機械翻訳を適用したい分野を具体的に一つ持っており，メーカ提供の一般辞書と専門語辞書の上に，ユーザ固有のユーザ辞書を構築しながら，実用化へ向けて試験的に使用しているというものである。ユーザの中である程度成功をおさめている例を見ると，機械翻訳の適用範囲，分野をかなり狭い範囲に限定している場合が

見られる。適切な訳語を出力させるには，現状ではかなり分野をしぼる必要があるようである。また，翻訳文の質よりは，ある程度の量をある程度の質で訳してくれて，大体の意味が把握できれば十分という使い方の場合には，辞書がどうあるべきかに関しては，違う意見になるようである。

- 辞書更新については，かなり機械翻訳システムに通じており，文法の知識についてもある程度持っている人でないと難しいようである。現在のユーザというのは素人のエンドユーザというよりはむしろ，機械翻訳の導入を検討している研究開発者的な人々が多いようである。

  ユーザのシステムの使用者にも次のようなタイプがあるようである。

  ① エンドユーザによる使用

  a）対象言語の知識のない人が翻訳結果のみを利用する場合。この場合，意味的に正しく，かつ文法的な正確さもそれなりに要求される。

  b）翻訳結果を利用するが，対象言語に関する知識をかなり持ち，必要に応じて翻訳結果の吟味を行うことが可能である場合。

  c）対象言語に関する知識をかなり持ち，原文を読む手間を省くために機械翻訳を使用し，必要な情報を見つけたら，原文にもどって読解する場合。この場合は，少々の間違いはあっても，大局的に必要な情報がそれなりの精度で訳されていれば十分である，ということになる。

  ② 翻訳者による使用

  翻訳の下訳を出させて，翻訳の能率をあげるために使用する。

  ③ 研究開発者による使用

  将来のエンドユーザによる使用あるいはエンドユーザに対する翻訳サービスに向けて，研究開発者自身が使用している場合。この場合のユーザは機械翻訳システムについてもある程度詳しく，文法の知識も備え，翻訳対象分野についても知識を持っている。

- 辞書作成・辞書更新の在り方については，このようなユーザ像の違いに応じて，意見が異なるようである。①エンドユーザの立場では更新機能はわかりやすく，簡便なことが望まれるが，②③の立場では，使い易

さと同時に，必要なことを的確に更新できることが望まれる。

- 機械翻訳の使用レベルとしては，ほとんどは実験的に導入しているところが多く，すぐに実用化できるとは考えていないところが多い。辞書作成の手間は覚悟はしているが，実際，相当かかるというところが多く，現在のメーカ提供の辞書で満足というところはないようである。辞書更新についても，失敗を繰り返しながらも地道に取り組みを続けているという状態である。
- 機械翻訳システムの使用状態としては，バッチ処理的な使い方と，対話的な使い方がある。現状では，機械翻訳システムの運用にはかなりの知識が要求されるのでエンドユーザの使用法はバッチ処理的な使い方に近くなってしまうようである。対話的な使用形態で使いこなすには相当の訓練が必要という印象を受けた。
- 辞書更新についても，どのようにやったらうまくいくかというのは，担当者個人のノウハウ的な部分が強く，試行錯誤的に行っている傾向が強いようである。辞書作成・辞書更新にかかわる運用体制についても，機械翻訳システム自体がまだ試用段階にあるので，運用体制と言う程の実績を持つところはまだほとんどない，というのが現状である。
- また，個々に作成された辞書の活用についてもまだ真剣に検討できるほどユーザの層は厚くなっていないようである。
- 辞書更新まわりの問題点としては，システムの仕組みを知らないと辞書更新が正しく正確に行われないこと，文法の知識がある程度必要であること，意味記述など相当の専門性が必要な項目もあり，エンドユーザの使用を困難にしている可能性があること，等を指摘することができる。

まだ使用現場から十分なことが引き出せるほど，市場が形成されているわけではない。多くのユーザは試用の段階にあり，それぞれ工夫しながら実用への試みを続けているのが現状である。

3.2.2 メーカ調査結果

本節では，機械翻訳システムを開発しているメーカに対する本調査の

結果の詳細を記述する。今回，調査対象となったのは，現在市販されている機械翻訳システムのメーカ５社である。なお，資料として個々のメーカの各項目別の回答の集計表が表３－２にあるので，参照願いたい。以下，なるべく集計表の順（アンケート用紙の順）に記述していくが，内容によって一部変更してある箇所もある。

3.2.2.1　機械翻訳システム

①　今回，調査対象となった機械翻訳システムが対象としている言語は，英日または日英（一社のみ日韓も対象としている）である。また，翻訳の対象としている分野は限定されており，主に科学技術分野となっている。つまり，翻訳の対象分野は限定されており，一般の文書をすべて訳せるようなシステムではない。

稼動マシンは，パーソナル・コンピュータ，ワークステーション，大型汎用計算機など多岐にわたっている。ただし，各システムは，いずれか一種のマシン上でしか稼動しない。

利用方式は，いずれのシステムとも，バッチ方式及び会話方式の両者が可能となっている。多量の文書を一括して翻訳する場合にバッチ方式を利用し，少量で即時性を必要とするときに会話方式を利用する。利用形態は，パーソナル・コンピュータ及びワークステーション上のシステムでは個人利用，大型汎用計算機上のシステムではセンタ利用となっている。

②　システム導入にあたって，ユーザが行わなければならない準備事項として，

(i)　要員の訓練

(ii)　ユーザ辞書の作成

がある。

機械翻訳システムは，原文（訳される文章）の入力に始まり，原文の編集（プレ・エディット），翻訳，訳文の編集（ポストエディット），出力を行うトータルなシステムである。その他に辞書の作成，更新，管理などの機能を持っている。そのため，システムを使いこなすには高度な技術が必要とされる。特に重要なのは，プレ・エディットと辞書の更新である。翻訳の能力を最大限まで引き出すためには，この２つの操作に習熟しなけ

ればならない。そのためには，機械翻訳の仕組みにまで立ち入って理解する必要（メーカはここまでは望んでいないが，実際に調査員が使用してみた実感である）があり，高度な技術が必要となる。従って，要員の訓練は必須のものである。

翻訳（機械翻訳に限らず，人手による翻訳でも）にあたっては，ユーザ独自の用語や訳語といったものが必ず存在する。1ユーザについても，別のジョブでは別の用語，訳語が必要である。従って，システム導入時，または，ジョブの開始時にユーザ辞書を作成しなければならない。

③ 保守体制はメーカによりまちまちであるが，販売形態によって対応が異なるようである。パーソナル・コンピュータやワークステーション上で稼動するシステムは，販売方式は売り切りとなっている。その場合，保守体制は電話による応対やバージョン・アップでの対応となっている。大型汎用計算機上のシステムはレンタル販売を行っており，通常のＳＥによる保守体制を採っている。

④ 機械翻訳システムは，翻訳を支援するシステムであるとメーカ側は主張している。さらに，現状では，処理の対象となる分野もかなり絞られたものとなっている。これに対し，ユーザは機械翻訳システムを使えばどんな文書でも翻訳できると考えており，両者の意識にはかなりの隔たりがあるようである。

3.2.2.2 辞書の構成・提供

① 辞書の構成は，基本的に，

(i) 基本辞書

(ii) 専門語辞書[*1]

(iii) ユーザ辞書

（＊1：専門用語辞書と呼ばれる場合もあるが，本報告書では専門語辞書と呼ぶことにする。ただし，集計表ではそれぞれの呼称で記述している。）

からなっている。ただし，専門語辞書は持たないシステムもある。また，ユーザ辞書を複数同時に使用できるシステムもある。

これらの辞書は優先度を持っており，優先度の高い順に単語を検索にい

く。優先度はユーザ辞書が最も高く（ユーザ辞書が複数使用可能な場合は，その中の優先度はユーザが決定できる。），専門語辞書，基本辞書の順に低くなる。これにより，同一の単語でもユーザ独自の使い方ができる。

基本辞書とは，いわゆる一般語の辞書であり，メーカから提供される。基本辞書は，一般にシステムに含まれるものとみなされている。基本辞書には，機能語や文法的に重要な単語など文法に直接関わる単語が含まれている。そのため，通常ユーザには解放されていない。語彙数は，どのシステムとも約5万語と共通している。

専門語辞書は，特定の専門分野でのみ一般に使用される単語から成っている。語彙数は，1万5千語から25万語までというようにシステムによってまちまちである。専門語辞書も現在はメーカから提供されている。ただし，システムからみた扱いはユーザ辞書と同列に扱われる。メーカからの提供もシステムとは別料金になっている場合もあり，やはり，システムとは切り離されたものとして扱われている。

ユーザ辞書は，ユーザで作成するものであり，メーカはそのためのサポートを行ったり，ツールを提供するだけである。完全にユーザ側で使用，管理が任されており，ユーザ辞書の変更による訳文の品質などの責任はユーザにあるユーザ辞書はユーザによりさまざまな利用形態が考えられるが，詳細は「3.2.3ユーザ調査結果」の中で述べることにする。

② メーカ提供辞書（基本辞書，専門語辞書）の権利に関しては，どのメーカともプログラム・プロダクトの一部と見なしており，システムに含まれるものであると回答している。従って，契約内容も一般のプログラム・プロダクトと同じになっている。ただし，専門語辞書については，システムとは別扱いにしているメーカもある。

3.2.2.3 辞書更新処理方式

① ユーザで更新処理を行えるのは，基本的にはユーザ辞書だけである。基本辞書や専門語辞書を更新したいときは，ユーザ辞書にその単語を登録して変更しなければならない。

② 更新範囲

更新処理の対象となる単語は，品詞別にみると，システムにより異なる

が，最大で名詞，形容（動）詞，動詞，副詞であり，最小では，名詞だけしか操作できないシステムもある。助詞や助動詞などの機能語及び文法は解放されていない。機能語や文法は互いに関連しあっており，さらに，システムに依存しているためである。また，多義語，複合語，慣用句などもすべて辞書に登録することによってその処理が可能となる。

更新内容は，基本的には，見出し語，品詞，訳語，活用語の活用形である。さらにシステムによっては，名詞に対しては意味属性，動詞に対しては格属性などを解放しているものもある。更新内容が多いほどきめ細かな訳が可能となるが，その分ユーザに負担がかかる。

③ 更新ツール

更新は，専用の辞書エディタや辞書ユーティリティを用いて行われる。エディタやユーティリティの機能としては，辞書項目の追加，変更，削除であり，どのシステムともほぼ共通している。処理方式は，コマンド入力によるもの，メニューによるもの，テンプレート方式などがあり，それぞれに工夫がこらされている。一般的に，パーソナル・コンピュータやワークステーション上のシステムは対話形式で行われ，汎用計算機上のシステムでは対話，バッチの両者が可能である。また，今後備えていく機能としては，学習機能などが挙げられている。

3.2.2.4 辞書のメンテナンス・サポート体制

① メンテナンス

ユーザ辞書のメンテナンスはユーザに任されているので，「3.2.3 ユーザ調査結果」で述べることとし，ここではメーカ提供辞書に関して述べる。メーカ提供辞書（特に専門語辞書）のメンテナンスはテスト用の例文を用意しているメーカもあるが，主に社内外からの苦情や新規文書の翻訳を試みることによって行われている。辞書の拡張に関しても，苦情や新規文書の他，ユーザ辞書や他システムの辞書も利用されている。ただし，ユーザからのフィードバックについては組織だっては行われてはいない。

システムや基本辞書のバージョン・アップによるユーザ辞書の整合性は各メーカとも保証している。辞書のバージョンアップは定期的，または不定期に行われているが，1年に2回から3回くらい行われる。これは，新

規用語の追加などによるものよりも，システムのバグ対応などによるものが多く，このことから，現在，機械翻訳システムは発展途上であることが窺える。

② サポート体制

ユーザ辞書の構築に対するメーカ側の支援体制は，

- 通常のＳＥサポート
- ツールの提供
- 教育

などが代表的なものである。ただし，機械翻訳システムが市場に出回ってから間がないため，ユーザが少ないうちは教育を行えるが，ユーザが増えてくると教育ができなくなるのではないかというメーカもある。

また，メーカによっては，

- 辞書作成の請け負い
- 辞書作成会社の紹介

などをしているところもある。今後，機械翻訳システムが普及していくためには，専門語辞書が充実し，ユーザ辞書を作成する手間を省くことが重要なことである。そのためにも，ユーザ辞書の受託や辞書作成会社などは注目される動きであり，ニュー・ビジネスの対象になる可能性を秘めている。

辞書の更新に対する支援体制は辞書構築に対するそれとほぼ同じである。システム（メーカ提供辞書を含む）のバージョン・アップに対しては，メーカ請け負い，または，自動更新ツールの提供で対処している。その他の支援としては，ユーザ会やパソコン通信による講座などがあり，興味深い動きをしている。

3.2.2.5 ユーザへの希望

ユーザ側からはユーザ作成辞書を買い上げて欲しいとか，ユーザからの意見をフィードバックして欲しいなどという声がある。しかし，現状では，メーカ，ユーザともそのような体制にはない。また，システム自身が発展途上であり，ユーザも未熟である。このようなことから，ユーザ作成辞書からのフィードバックは今後の問題であるとしている。

ユーザからは，「なぜ訳せないのか」，「ニーズに合った辞書が欲しい」などという声があり，機械翻訳システムに過大な期待を持っている。メーカは機械翻訳システムをあくまで翻訳支援と位置付けており，プレ・エディット，ポスト・エディット，ユーザ辞書の作成を前提としている。これに対し，ユーザには文を入力すればすぐ訳文が現れると思っているユーザもいるようで，両者の意識にはかなりのギャップがあるようだ。

3.2.2.6　メーカ調査集計表

今回の機械翻訳システムの辞書に関する国内調査のメーカ集計表を表３－２に示す。

なお，表３－２中で，調査票にあった会社概要は省略し，メーカ名はＡ～Ｅで示した。

表3－2　　機械翻訳システムの辞書調査集計表　（メーカ向け）

No.〔 1/11〕

メーカ / 調査項目	A	B	C	D	E
Ⅱ機械翻訳システムについて 1．システムの名称	①ﾏｲｸﾛﾊﾟｯｸ J/E ②ﾐﾃﾞｨｱﾑﾊﾟｯｸJ/E ③ﾏｲｸﾛﾊﾟｯｸ J/K	①ＡＴＬＡＳ－Ⅰ ②ＡＴＬＡＳⅡ	ＰＩＶＯＴ	機械翻訳システム	日英翻訳支援システム ＨＩＣＡＴＳＪ／Ｅ
2．対象言語	①日→英 ②日→英 ③日→韓	①英→日 ②日→英	日→英、英→日	日→英、英→日	日→英
3．対象分野	①、②、③ 技術文献、マニュアル、新聞記事等	①、② 科学技術文書	技術分野	科学技術文献 技術マニュアル	科学・技術分野
4．稼動環境 (1) 計算機の機種	VAX シリーズ、NEC PC98ｼﾘｰｽﾞ・N-5200ｼﾘｰｽﾞ、HITACHI-816、IBM5560/5550/5540	ＦＡＣＯＭ Ｍシリーズ Ｓ３０００シリーズ	ＡＣＯＳ４５０以上	ＵＸ－７００	Ｍ－２４０Ｈ以上
(2) ＯＳ	ＶＭＳ（ＶＡＸ） ＭＳ－ＤＯＳ（ＰＣ）	①ＯＳⅣ／Ｆ４ＭＳＰ ＯＳⅣ／Ｘ８ＦＳＰ ＯＳ１Ｓ／Ｓ ②ＯＳⅣ／Ｆ４ＭＳＰ	ＡＣＯＳ４	ＵＮＩＸ	ＶＯＳ３
(3) プログラミング言語	Ｃ言語、アセンブラ言語	①アセンブリ言語 ②アセンブラ	Ｃ、ＣＯＢＯＬ	Ｃ	①文法規則 …… ＧＤＬ ②システム …… ＰＬ／Ｉ
(4) 必要とするメモリ容量	６４０ＫＢ（ＰＣ）	①４ＭＢ （ﾏﾝﾏｼﾝｲﾝﾀﾌｪｰｽ部含む） ②４０００ＫＢ	４ＭＢ	４ＭＢ	５ＭＢ
(5) 必要とするメモリ容量	２０ＭＢ	①２０ＭＢ以上 ②６０ＭＢ以上	１００ＭＢ	９２ＭＢ	①４０ＭＢ（基本辞書） ②７０ＭＢ （専門用語辞書）
5．利用方式	バッチ方式、会話型	①、② バッチ方式、会話型	バッチ方式、会話型	バッチ方式、会話型	バッチ方式、会話型
6．想定するユーザ形態	個人利用、エンドユーザー センタ利用(VAX)、翻訳者	①、② センタ利用、翻訳者	個人利用、センタ利用 翻訳者	個人利用（ＷＳ型） 翻訳者、エンドユーザー	センタ利用 企業内利用

表3－2（つづき）　機械翻訳システムの辞書調査集計表　（メーカ向け）

調査項目＼メーカ	A	B	C	D	E
7．想定する運用体制	教育されたオペレークによる利用	①、② 専門用語辞書要員の訓練などを予め実施した上で専任組織により運用する	集中管理、バッチ中心 対話は、コスト問題あり	機械支援翻訳であり完全自動ではない。又、ユーザ辞書はユーザサイドで作る	運用保守部署が存在して機械翻訳向けの文書の作成・編集の指導を行い、ユーザ辞書の作成・保守を行う。
8．保守体制	電話による対応 （使用法、障害） ﾊﾞｰｼｮﾝｱｯﾌﾟによる対応 （ｼｽﾃﾑの有料交換）	①、② 担当ＳＥ経由での標準障害処理体制	事業部で苦情処理		開発部署で対応する。
9．ユーザの準備事項	ハード条件の整備 ユーザ教育の習得	①、② ・専門用語辞書の開発 ・要員の訓練 ・入出力インタフェースの整備	ユーザ辞書 要員訓練	ユーザ特有の単語が必ずあるので、それは準備して頂く	・ユーザ毎の文書に則した文章作成編集ノウハウのまとめ、既存文書の編集 ・ユーザ独自の用語、訳語を含むユーザ辞書の作成
10．ユーザのシステムへの介入	文章の修正機能 辞書の追加、修正機能	①辞書内容変更 原文に対する構文指示記号 ②？	不可 ユーザ辞書は変更可		不可 基本辞書でも訳語順の変更は可
11．システムの販売 (1) 販売方式	直販、特約店販売	レンタル販売	システム、基本辞書、ユーザ辞書の3本立てでレンタル	売り切れ（辞書を含むソフトのみ可）	プログラムプロダクト（専門用語辞書は別価格、レンタル方式）
(2) 販売部署	システム営業部	営業部門	事業部、C&C VAX サービス	電算機事業部	情報システム営業本部

表3－2（つづき）　機械翻訳システムの辞書調査集計表　（メーカ向け）

No.（ 3/ 11）

調査項目＼メーカ	A	B	C	D	E
Ⅲ辞書の概要 1．名称	J／Eパック辞書	①基本辞書 ②ATLAS Ⅱ日本語基本辞書 ATLAS Ⅱ英語基本辞書	なし		①基本用語辞書 ②科学技術専門用語辞書
2．作成日		①、② 非公開	現在まで3年半	5年 （但し、研究期間は除く）	①1年半
3．機種	日英	①基本辞書 ②？	日本語辞書、英語辞書	一般辞書・専門用語・ユーザ用語	①＋②＋ユーザ辞書
4．機能	日本語の解析及び英語の生成	①基本辞書 ②？		意味分析（一般）	日英機械翻訳時の各語の情報含む。
5．役割	？	①基本辞書 ②？		解析、トランスファ	文法とともに機械翻訳に用いる
6．規模（語彙数）	約6万語	①、② 約5万語	一般　5万語 専門　1万5千語	一般　5万語 専門　max 25万語 ユーザ　max 5万語	①5万語 ②25万語
7．辞書パターン	コア辞書、ユーザ辞書	①、②	一般辞書、ユーザ辞書		最大3つの辞書使用可
8．使用コード	JIS	JEF	JIS	JIS	KEIS、EBCDIC
9．辞書構成	社内機密	①、② メーカー提供の基本辞書を含め最大5個の辞書を任意の優先順位で連結して使用	基本辞書、ユーザ辞書		基本辞書1つと科学技術専門用語辞書を含む最大2つの辞書、計最大3つまで可
10．データ項目	社内機密	①マニュアル参照 ②見出し、概念記号、文法属性		見出し、品詞、デフォールト訳語性、数、意味マーカ、語彙文法その他	見出し、品詞、訳語、保守情報

表3－2（つづき）　機械翻訳システムの辞書調査集計表　（メーカ向け）

No.〔 4/ 11〕

メーカ / 調査項目	A	B	C	D	E
Ⅳ辞書のユーザへの提供方式 1．方式	システムに含まれる	①？ ②	2種の辞書込みの料金	プログラムと一括	プログラムプロダクト
2．媒体	磁気テープ、ハードディスク、フロッピーディスク(15MB)	①、② 磁気テープ	磁気テープ（100MB）	フロッピーディスク(9MB)	磁気テープ
3．形態	オブジェクトコード	①、② データ	オブジェクトコード	オブジェクトコード	オブジェクトコード
4．バージョンアップ版の提供	有り	①、② ― 有り	有り	有り	有り
(1) 種類	コア辞書・専門辞書	①、② 基本辞書	システム辞書	一般辞書、専門辞書	①基本用語辞書 ②科学技術専門用語辞書
(2) 頻度	2回／年	①、② 数回／年	かなり多い	不定	2回／年
(3) 規模	？	①、② 約5万語	安定しない情報について	全体	各辞書の改訂版を提供
5．辞書に関する権利	契約有効期間、使用	①、② レンタル期間中、当該センタにおいてのみ、ユーザが使用権を有する	はっきりしていない		・著作権、財産権に関する法的保護 ・プログラムプロダクトの一般契約
6．もれ防止対策	社内機密	①特になし（信頼関係） ②特に成し。辞書ファイルの構成を公開していない。当社のエディタを通してのみ見れる。	良い方法はない	コピー防止策がとってある	・法的保護及び契約による ・ユーザに開示しない部分はドキュメントに記述しない

表3－2（つづき）　機械翻訳システムの辞書調査集計表　（メーカ向け）

No.〔 5/ 11〕

調査項目＼メーカ	A	B	C	D	E
Vユーザへの開放について					
1．開放範囲	辞書更新プログラムの提供	①4．以外すべて ②辞書エディタで見せる	ユーザ辞書のみ	ユーザ辞書	①ユーザ辞書 ②基本用語辞書の訳語順
2．開放条件	名詞、形容詞、動詞、副詞	①なし ②特になし		特になし	品質の責任はユーザが持つ
3．開放内容	見出し語、訳語、属性、活用形を指定させる	①登録内容すべて ②自立語については、訳語を見せる。特に名詞は意味分類を、動詞は格属性を見せている。	客にトレーニングしているのは、名詞のみ。文法はさわらせない。	名詞・動詞・形容詞・副詞について、見出し・品詞・訳語を開放	①名詞・サ変動詞・数詞接尾語 ②基本用語訳語順 ③文法ルールはだめ
4．開放できない部分・理由	部分：助詞・助動詞・接続詞 理由：解析及び生成のルールに大きい影響を与えるため	①部分：機能語など1000語 理由：文法ルールと密接な関係がありユーザによる変更に適さないため ②付属語（機能語）は開放していない	部分：文法 理由：整合性がくずれてしまう。保守が大変	部分：機能語及びルール 理由：システムに依存しており、一般ユーザには書けないため解放していない。	部分：①機能語 ②3の①以外の内容語 ③文法との関連の強い項目 理由：①文法との依存性が強く副作用大 ②システムの改良に対する自由度を保つ

表3－2（つづき）　機械翻訳システムの辞書調査集計表　（メーカ向け）

No.（ 6/ 11）

調査項目＼メーカ	A	B	C	D	E
Ⅵ辞書作成（メーカ側）					
1．収集法、選択基準	社内機密	①、②　非公開			
(1)基本語彙			市販辞書から		頻度順、分類別、品詞別
(2)専門語彙			社内にもともとあったものから、特殊な訳語、頻度の高いものを選んだ。		市販の対訳辞書
2．辞書化の手順	社内機密	①、②　非公開	見出しを選び、対訳をつけ品詞別に分け、辞書化し、ベリファイする。		
(1)基本語彙					人手によりコード入力しバッチ入力。対話方式で修正実験により調整
(2)専門語彙					対訳ﾃﾞｰﾀ に対しプログラムで標準値を設定、対話又はバッチで修正。実験で調整
3．複合語等への対処法	社内機密				
(1)複合語		①、②　辞書に登録	辞書に入れる	複合語として辞書に登録。もし登録してなかったら自動合成する。	①名詞連鎖については連語として登録 ②派生語は中心語レコードに登録
(2)慣用句		①、②　辞書に登録	辞書に入れる	語彙文法により慣用句の認定と意味の決定を行う	中心語のレコードに登録
(3)多義語		①辞書に登録。意味カテゴリによる訳語自動選択、対話的にユーザ選択 ②単語間の意味関係による選択。	部分的に並列処理し、うまくマッチすると解析の段階で1つにしぼる。	意味分析により一意に決定する。	複数の訳語（グループ）を登録
4．派生的な意味・特殊な意味	社内機密	？	分野情報、語用情報	語彙文法により、脈絡、用法、の詳細な記述を行い、これにより対処する。	特に重要な語については、個別文法で対処。派生語の扱いについては3（1）を参照

表3－2（つづき）　機械翻訳システムの辞書調査集計表　（メーカ向け）

No.〔 7/11〕

調査項目＼メーカ	A	B	C	D	E
Ⅶ辞書メンテナンス（メーカ側）					
1．内容チェック	社内機密	①、② 非公開			
(1)頻度			頻繁		年2回程度の予定
(2)方法					目視によるチェック、翻訳実験による問題点の抽出。
(3)手順			社内外の苦情にもとづく		チェック項目の決定、目視によるチェック、対話又はバッチによる修正、確認。
2．テスト用例文の種類と数	社内機密	①、② 非公開	全部で150ページ程 100ページがコンピュータマニュアル		＊文法・辞書の基本項目が二千例文 ＊その他各種チェックが二万例文
3．修正・維持・拡張方式	社内機密	①、② 非公開	気づいたらやる。 外部からの苦情から。	実際のマニュアル等を訳し訳文を検討し、文法あるいは語彙文法を修正・拡張している。	語学文献、見直し、実験
4．辞書のデータ拡張	他のシステム辞書の利用、ユーザ所持辞書の利用。 翻訳辞書や拡張ユーザ辞書の他、アプリケーションを利用する。	①、② 必要語を独自に収集	ユーザ所持辞書。 翻訳辞書や拡張ユーザ辞書の他アプリケーションを利用する。	上記3による。	各種情報の利用。 市販辞書の対訳情報の利用
5．専門語辞書の複数利用	ある	①、② ？	ある 効果：新しい訳語を得ること。	ある	ある 適訳ができる。未知語の減少

表3－2（つづき）　機械翻訳システムの辞書調査集計表　（メーカ向け）

No.〔 8/ 11〕

調査項目＼メーカ	A	B	C	D	E
Ⅷ辞書内容の変更（ユーザ側）					
1．辞書による変更	可	①可　②？	可	可	可
(1)変更	可	①可	可	可	可
・内容	一般動詞、名詞、形容詞、副詞に関して訳語、属性の変更	①　Ⅴ参照		Ⅴ参照	Ⅴ参照
・方法	辞書更新プログラムによる	①辞書エディッタまたは辞書ユーティリティ		対話的にできるユーティリティが用意されている	対話 バッチ
(2)追加	可	①可	可	可	可
・内容	上記と同様	①(1)と同じ		同上	ユーザ辞書に追加
・方法	上記と同様			同上	対話、バッチ、対話情報よりプログラムで作成
(3)削除	可	①可	不可（リプレースのみ可）	可	可
・内容	前記(1)と同様	①(1)と同じ		同上	ユーザ辞書から削除できる
・方法	前記(1)と同様			同上	対話、バッチ
(4)その他の機能					辞書の印刷、複写、比較／ユーザ辞書作成時に標準値設定／出典情報による訳語の選択
2．ユーザによる訳し分け	不可	①？	不可（一度辞書にもどって変更）	可（ユーザ辞書の入れ換え可。学習機能有り）	可－文型や訳語の選択（ポストエディット、ユーザ辞書、基本辞書の訳語選択）
3．マニュアル	有り	有り	有り	有り	有り
4．ユーザメンテに関するユーザからの要望	・ユーザが辞書を簡単に管理できる手段の提供 ・ユーザへの訪問を希望する声が多い	①、②　非公開	「なぜ訳せないのか」 「ニーズに合った辞書が欲しい」 過大な期待がある。		簡易にユーザ辞書の作成、保守を行う機能
5．今後備える機能	社内機密	①、②　未定	今は改良の段階		辞書の学習機能等検討中

表3－2（つづき）　機械翻訳システムの辞書調査集計表　（メーカ向け）

No.〔 9/ 11〕

調査項目＼メーカ	A	B	C	D	E
Ⅸ辞書更新ツール				（Ⅷに同じ）	
1．辞書ツール	有り	①有り　②？	有り	有り	有り
・種類	会話型の辞書更新プログラム				①用語辞書保持 ②ユーザ辞書作成支援
・機能	辞書の内容の変更・追加・削除	①Ⅶ参照 ②変更・追加・削除		登録、削除、リスト出力	①追加・更新・削除・印刷・複写・比較 ②対訳情報利用、出典情報訳語順変更
・方法	簡単な質問形式	①、② …マニュアル参照		画面上で対話的に行う	対話又はバッチ
・特徴	専門的知識を必要としない	①、② …マニュアル参照	モード2つ ― オンラインエディッタ ― カード流す		メニュー、ガイダンス方式による操作

表3－2（つづき）　機械翻訳システムの辞書調査集計表　（メーカ向け）

No.〔10/11〕

調査項目＼メーカ	A	B	C	D	E
X辞書構築（ユーザ側）					
1．構築にあたって利用できる機能	・専門辞書登録機能 ・専門辞書管理機能	①IX参照 ②マニュアル			IX－1参照
2．バージョンアップに対する整合性	社内機密	①、②整合性を保証する	名詞しか作らせない	ユーザ辞書は、必ず機能的にサブセットになっており上方に対する互換性あり	①公開項目を安定した範囲にとどめる ②語彙管理情報を辞書中に入れる
3．構築への支援体制	メーカ提供の分野別辞書の充実 ユーザ専門辞書の開発受託	①、② 通常のＳＥ支援	教育 （客が少ないうちは可）	ユーティリティが用意されている。	①利用の手引 ②ツールの提供 ③辞書作成会社の斡旋
4．ユーザからメーカへのフィードバック	ケースバイケース	①、② 直接のフィードバックなし	良い辞書を作っている所からはある。	基本的には行わない	現在は大部分のユーザが試用の段階であり、ユーザ作成辞書からのフィードバックは今後の問題

表3－2（つづき）　機械翻訳システムの辞書調査集計表　（メーカ向け）

No.〔11/11〕

メーカ / 調査項目	A	B	C	D	E
X1 更新に関する支援体制					
1．辞書更新支援体制	辞書更新プログラムの提供 ユーザ教育	①X参照 ②通常のSE支援	教育のみ	ユーティリティが用意されている	項番X－3に同じ
2．ユーザの習熟度への対処	ユーザ講習会 PC通信による講座	①、②　？	過大評価してもらわない		①説明会 ②SEによる支援 ③辞書作成会社の紹介
3．バージョンアップ支援体制	メーカ請け負い 自動更新ツール	①、②　？	メーカ側請け負い	ツールの提供	・メーカ側改訂版の提供 X－2参照
4．その他支援	ユーザ会 News 講習会 PC通信による講座	①、②　検討中	特にない ユーザ辞書の買い取りの希望はある。		①SEによるユーザのフォロー ②辞書作成会社の紹介

3.2.3　ユーザ調査結果

本節では，ユーザに対する本調査の結果の詳細をまとめて記述する。なお個々の項目に対する回答を資料として表３－３に集計表としてまとめてあるので，参照願いたい。以下，なるべく集計表の順（アンケート用紙の順）にしたがって記述していくが，内容によって一部変更してある箇所もある。

3.2.3.1　機械翻訳システム

①　今回調査対象となったユーザ５社のうち，３社はメーカで販売している機械翻訳システムをそのまま導入している。残り２社のシステムはメーカと共同で開発されている。

対象言語は，いずれも英日または日英である。ハードウェアは，パーソナル・コンピュータ，ワークステーション，汎用大型計算機となっており，メーカから提供されている形態をすべて網羅している。

②　導入から実運用まで

導入から実運用までの作業工程は，おおよそ次のようになっている。

(i)　システム導入

システム導入では一般の計算機システムと同じように，システムのセットアップやインストールなどが行われる。

(ii)　ユーザ辞書の構築

翻訳（機械翻訳に限らず，人手による翻訳でも）にあたってはユーザ独自の用語や訳語といったものが必ず存在する。それらの用語や訳語を辞書に登録する作業がユーザ辞書の構築である。

(iii)　要員の教育・訓練

ここでいう要員とはシステムの運用，管理者ではなく，エンド・ユーザのことである。機械翻訳システムではプレ・エディットや辞書の更新，管理など高度な技術を必要とする作業があるため，要員の訓練は必須である。

導入から実運用までに要する期間は，年のオーダーとみられる。しかし，調査対象となった５社のうち，実運用をしているのは１社だけであったので，正確な数値は出てこなかった。

③　運用体制

運用体制に関しては，実運用を行っている１社以外は，現在の体制（試用期間中の体制）について回答してもらった。実運用にいたっていないところでは，現在，テスト中または準備段階ということで，数名の要員しか割いていない。また，体制も特に組織だったことはしておらず，ＴＳＳなどにより各個人が自由に使用している。実運用をしているところでは，システムの管理者をも含めて，約４０人で運用している。

④　評　　価

システム全般について，人手翻訳と比較してみると，全ユーザが人手翻訳より手間がかかると答えている。その理由としては，

- プレ・エディット，ポスト・エディットに時間がかかる。
- ユーザ辞書が未完成であり，辞書登録をたびたび行わなければならない。

などが挙げられる。具体的に数量的な回答を与えてくれたところでは，人手翻訳の1.5倍から２倍くらいの時間がかかっているようだ。

プレ・エディットに関する質問では，プレ・エディットは行わないと答えているところが２社あった。これは，キーワードだけでも日本語にして目につき易くするなどの目的で機械翻訳システムを使っているからである。プレ・エディットを行っているユーザはすべて，プレ・エディットは手間がかかると答えている。また，プレ・エディットの結果，原文が不自然になってしまうという不満もある。

ポスト・エディットについては，人手翻訳とあまり変わらないという意見と手間がかかるという意見があった。いずれにせよ，校正者は原文を読まなければならない。

システムの完成度については，その尺度がはっきりとしていないため，ユーザにより様々な回答が寄せられた。平均してみると，完成度はあまりよくないという評価である。ただし，メーカと共同開発しているユーザは，もともと目標がはっきりしていたため，その目標に達しているということで一応満足していると回答している。導入効果に関する質問では，まだ実運用に至っていないため，効果を考えるのはこれからのことであるとして

いるユーザ以外は，なんらかの効果はあったとしている。

全体としての満足度もユーザそれぞれ，人それぞれで評価が異なっている。現在のハードウェアなどの環境を考慮するとまあ満足であるという考えと，実用レベルからみるとまだ実用にはならないという考えが代表的である。調査をしていての印象では，現状ではまだ実用レベルまでは達していない。使い勝手が悪いなどという意見はあるものの，今後の発展に期待するところが大きいようであった。

3.2.3.2　辞書の構成・評価

①　辞書の構成は，

(i)　基本辞書

(ii)　専門語辞書 *1

(iii)　ユーザ辞書

（＊1：専門用語辞書と呼ばれる場合もあるが，本報告書では専門語辞書と呼ぶことにする。ただし，集計表ではそれぞれの呼称で記述している。）

となっている。ただし，専門語辞書を使用していない場合もある。また，ユーザ辞書は複数同時に使用可能なシステムもある。これらの辞書は優先度を持っていて，優先度の高い順に単語を検索にいく。通常，ユーザ辞書が優先度が最も高く，専門語辞書，基本辞書の順で低くなる。

基本辞書，専門語辞書はメーカから提供される。基本辞書には，いわゆる基本語彙が収められている。専門語辞書には，特定の専門分野で一般的に使用されている語が収められている。（詳細は，3.2.2メーカ調査を参照）

ユーザ辞書を複数持つことができるシステムの場合にはさまざまな使い方がある。一般的には階層化されることが多いが，同列に扱う場合もある。複数の辞書を同列に扱う場合，辞書の優先度を決定することが重要となる。

②　ユーザ辞書

ユーザ辞書が階層化される場合には，

(i)　グループ辞書

(ii)　プライベート辞書

に分けられる。（名称は各ユーザにより異なるが，ここではこの名称を用いる。）

グループ辞書は複数の人間が共通に使用する辞書である。一企業内またはプロジェクトで共通に使用されるものである。社内用語や，特定のプロジェクトやジョブで共通に使用される用語が収められる。システム導入時やプロジェクト，ジョブの開始時にグループ辞書は作成される。グループ辞書はさらに階層化される場合もある。

プライベート辞書は，各個人が使用するもので，個人のノウハウの蓄積やテスト用に使われることが多い。この辞書の作成，管理，更新などは各個人に任されている。

以上が一般的な辞書の構成であるが，これらの他に一過性の辞書を用いているところがある。これは実運用をしているシステムで使われているものであるが，特定の一つまたは複数の文に対してだけ良い訳が出るようにするためのものである。

③ 辞書の評価

次に，辞書に対する評価を見てみる。メーカから提供される辞書に対しては，全体として不満が多い。理由としては，

- 訳語が不適切である
- 文法体系が不自然である
- 翻訳の専門家の知識が吸収しきれていない

などが挙げられる。

ユーザ辞書についても不満が多い。その理由は，

- 辞書がまだ充実されていず，語彙数が少ない

ということである。実運用を行うためには，まず，ユーザ辞書が充実されなければならない。

以上のことから，基本辞書（メーカ提供辞書）は量よりも質が期待されており，ユーザ辞書は反対に，語彙数が豊富であることが望まれている。

④ 辞書の権利

ユーザ辞書に対する権利に関する質問では，権利はユーザにあるべきであるとすべてのユーザが主張している。その理由としては，

• 辞書は一種の著作物である

• 辞書には部外秘に当たる事項も含まれている

• 辞書の構築は高度な作業である

などといったことが挙げられている。

3.2.3.3　辞書の更新

① 辞書の更新の必要性はほとんどのユーザが，その必要性は高いと答えている。特に，ユーザ辞書の更新の必要性が高い。（現状で更新できるのはユーザ辞書に限られている。）更新が必要な理由は訳文として適切な訳（つまり，日本語〔英語〕として通じる訳）を得るためということを挙げている。実際に更新される頻度を見ても，ユーザ辞書の方が頻度が高い。（基本辞書が更新可能な場合）

② 更新の内容としては，削除されることは少なく，追加されることが多い。よく追加される語は，専門語，複合語，固有名詞，社内用語などである。辞書に登録する用語の選定，訳語の決定，専門家の意見を取り入れて行なわれている。用語の選定，訳語の決定は，その時の都合と汎用性を考慮して登録するか否かが決められる。

辞書と対を成すものとして文法がある。文法はシステムの動作を左右するため，通常は非公開である。しかし，共同開発の場合はこの限りでない。共同開発のシステムでユーザに文法の更新が許されている場合には，文法の更新も行われる。従って，文法の更新に対する潜在的ニーズは存在するものと思われる。

ユーザ辞書を更新し，充実させていくことは，より適切な訳を得るためであり，実際，その効果は現れている。しかし，その反面，ユーザ辞書全体の管理が難しくなるという弊害もある。

3.2.3.4　辞書更新運用体制

今回調査対象となった5社のうち，実運用を行っているのは1社のみで，他は実運用への準備階段またはテスト使用の段階であった。そこで，回答は必ずしも実運用に対するものとはなっていない。

① 運用規定

辞書の更新を組織的に行っているのは2社であり，他は必要に応じ，エ

ンド・ユーザが直接行っている。組織的に辞書の更新を行っているところでは，運用マニュアルもきちんと整備されている。運用規定の基本的考えは，次のようなものである。

(i) 辞書の更新は複数の人間で可否判定を行い，リリース前に実験を行う。ただし，プライベート辞書は自由に使用してよい。

(ii) 文法的に重要な語は特定管理者しか変更できないようにしている。

② 用語の収集

基本辞書に属するような用語は，市販されている一般の辞書を用いている。（共同開発のシステムで，基本辞書を更新できる場合）

ユーザ辞書に関する用語は，既に社内に蓄積されている用語，専門分野の用語集などを基に用語の収集を行っている。

③ 更新ツール

辞書更新は，システムに組み込まれている辞書エディタ，辞書ユーティリティを用いて行われている。これらのツールに関する評価は，コマンド形式の場合はコマンドを覚えなければならない，メニュー方式では，入力量が多いなどの不満がある。しかし，これらは機械翻訳システムに限られたことではなく，一般の計算機システムのインタフェースに共通する問題点であろう。

メーカから提供されている更新ツールは，会話方式，バッチ方式の両者が使用できるものが多いが，ユーザでは会話方式のみ，または会話方式が多く使われているようである。

3.2.3.5 ユーザからの要望

最後に，機械翻訳システム及びそこで使われる辞書などについて，利用や将来に対する期待などを自由に述べてもらった。

① 辞書については，

- メーカから提供される基本辞書が充実されるべきだ

という声が多い。具体的には，

- 汎用性が高いこと
- 知識をルールで使い分けられること

が挙げられている。

② 辞書のメンテナンス機能については，マニュアル・レスであることや，学習機能（ポスト・エディットなどの内容を辞書に反映させること）などを望んでいる。ただし，これらは辞書のメンテナンスに特有の問題ではなく，現在の計算機システムすべてに共通することのように思われる。

③ 公的機関に対する要望もいくつか挙がっている。それらを列挙すると以下のとおりである。

- 公的な機関で共通に開発した方がよい。ユーザ個々に行うと手間がかかるし，質の向上も望めない。
- 民間主導で行い，行政はそのバックアップにとどめるべきだ。
- まだ方法論が定まっていない現状では，多元的なバックアップが必要である。

これらについては，各ユーザそれぞれの考え方があり，共通の認識はまだ得られていないようである。

3.2.3.6 ユーザ調査集計表

今回の機械翻訳システムの辞書に関する国内調査のユーザ集計表を表３－３に示す。

なお，表３－３中で，会社・団体名は，Ｆ～Ｊで示した。

表3－3　機械翻訳システムの辞書調査集計表（ユーザ向け）

No.〔 1/12〕

ユーザ / 調査項目	F	G	H	I	J
Ⅱ導入している機械翻訳システムについて 1. (1) システム名	①ＡＴＬＡＳ－Ⅰ ②ＡＴＬＡＳⅡ	①ＡＴＬＡＳ－Ⅰ ②ＡＴＬＡＳⅡ	社内コードのみ （名称なし）	日経－CSK コンピュータライズド・トランスレーションシステム	マイクロバック J/E
(2) 対象言語	①英→日 ②日→英	①英→日 ②日→英	英→日	日→英	日→英
(3) メーカー名	①富士通 ②富士通	①富士通 ②富士通	日立製作所	日経と CSKの共同開発	ブラビスインターナショナル
(4) 販売部署			非売品	――――	?
2. (1) 導入目的	①原子力分野の翻訳、世界中の原子力情報の取り扱い EPRIの研究情報翻訳 ②研究用のデータベースを構築 米国等の海外に情報提供するため	テスト用	海外の経済情報の日本語サービス	日経本紙の朝夕刊株式解説欄を機械翻訳して英文日経テレコンでオンラインサービスするため	和文英訳の能力アップ並びに経費削減
(2) 導入形態	①レンタル ②レンタル	無償でテスト	共同開発	共同開発	買取り
(3) 導入部署	①、②、 OSC 電研コンピュータシステム		開発第二部	（日経）データバンク局 （CSK） 技術開発事業本部 MTプロジェクト	品質部品質企画センター

表3－3（つづき）　機械翻訳システムの辞書調査集計表（ユーザ向け）

No.〔 2/ 12〕

調査項目＼ユーザ	F	G	H	I	J
3．導入から実用運用での作業内容と所要時間・工数 (1) 内容	①共同開発的なので表わせない ②これから導入	試用段階	基礎研究→プロトタイプシステム→小規模翻訳実験（辞書、評価）→大規模翻訳実験（辞書、評価）→改良	情報提供、訳文監修、翻訳実験、評価	1.導入(40H) 初期立上げ、オペレーション習得 2.辞書の構築 基本辞書(50H) ユーザ登録(150H) 自動車用語(20H) 3.ﾃｷｽﾄｴﾃﾞｨｯﾄ の習熟 マニュアルで自習(360H) 講習会(9H) 他との情報交換(10H) ※但し、実運用には至っていない
(2) 所要時間・工数	――		約8年	85.10 ～ 86.6 （開発） 86.6～ 86.9 （運用準備）	
4．現在のシステムの概要 (1) 計算機の機種	FACOM M-200, M-360R		HITAC M240-H	DEC マイクロVAX	IBM 5550 Gタイプ
(2) OS	F4 / NSP		VOS3 (SP21)		日本語DOS
(3) プログラミング言語	①PL／I ②SPL アセンブラ （非公開）	アセンブラ	PL／I		C
(4) 要員	2～3人		非公開	日経 8名 CSK 34名	1名
(5) 予算	?		同上	日経 0 CSK	10万円／月
(6) 翻訳量	アブストラクト 3000件／年		同上	6000語／h 4000語／h×MIPS	無回答

表3－3（つづき）　機械翻訳システムの辞書調査集計表（ユーザ向け）

No.〔 3/ 12〕

調査項目＼ユーザ	F	G	H	I	J
5. 機械翻訳システムと結合して作成利用しているツール類がありますか (1) ある・なし	ある	ない	ある	ツール1: 形態素解析用 ツール2: 構文解析用 ツール3: 構文トランスファ用 ツール4: 英文生成用	ない
(2) ツール名、内容	PCからのファイル転送 データベースとの連携 ODMの中でアトラスを動かす		KWOC KWIC 入力文分析：単語数、長さ 頻度、辞書引き 原文入力、訳文出力 辞書保守：更新チェック分析		
6. 機械翻訳システムの運用体制について	まだ不確立・とりあえずTSSで利用	試用段階	まだ実用化されていないので開発部門がTSSで利用	日経側にチェック用の端末CSK側にエディット作業用の端末とシステム本体がありそれぞれ役割分担している。	まだ実用でない
7. 機械翻訳システムの評価 (1) 人手翻訳との比較	比較できない	著しく手間がかかった	手間がかかった ☆校正しようとすると人手翻訳よりややこしい	著しく手間がかかった ☆人手（ベテラン） 1H/1記事 プリ、ポストエディット、機械翻訳 1.5 ～2H/1記事	手間がかかった ☆(1)プリエディット に時間がかかる (2)特殊用語辞書が未完成なので、特殊用語をその都度辞書登録するため
(2) プリエディット	①やらない ②未	著しく手間がかかった とても不満	行わない	比較出来ない ただし、入力は平易にするようにしている	手間がかかった 不満 正しい英文を出力するために不自然な日本語になる場合がある
(3) ポストエディット	①比較できない	現在は行っていない	手間がかかった 不満 ☆校正しようとすると、人手翻訳よりややこしい ポストエディットの要・不要のために結局人間が全文読む必要があるため	普通 ☆新聞は正確さと速さが必要 あまり英文を要求しない	普通

表3－3（つづき）　機械翻訳システムの詳細調査集計表（ユーザ向け）

No.〔 4/12〕

ユーザ / 調査項目	F	G	H	I	J
(4) 完成度	①初期のワープロ程度	とても悪い	普通	とても良い ☆完成度をどのような尺度で測るかによるが、当初の目標としてはまずまず	悪い ☆複雑な日本文になると訳文はまだまだ
(5) 導入効果	①効果あり ☆これまで英文でしか読めなかったものが、とりあえず日本語で読める	普通 ☆効果を考えるのはこれからのこと 今は見守っていく段階	まだ実運用していない	効果あり 日経がMTサービスしていることのPR効果も含めて	まだ効果を測るに至らず
(6) 全体としての満足度	①人によって異なる	とても不満 今は実用にならない	まだ、実際に運用していない	まずまず満足	満足 ☆パソコンレベルであれだけ出来ればまあまあである

表3－3（つづき）　機械翻訳システムの辞書調査集計表（ユーザ向け）

No.〔 5/12〕

調査項目＼ユーザ	F	G	H	I	J
Ⅲ運用している辞書の実態 1. (1) 種類	①基本辞書 ②専門用語辞書	①、② 一般辞書と科学技術用語辞書	①基本辞書 ②グループ辞書 ③プライベート辞書	①基本＋②分野＋③社名 辞書　辞書　辞書	
(2) 機能			①システム全体の共通辞書 ②グループ共通辞書 ③プライベート辞書	①基本語彙及び基本文法情報 ②経済分野用辞書 （90%名詞） ③会社名データベース （正式＋略号）	
(3) 設計			③→①の順で強さに階層あり		
(4) 規模	①5万語 ②5万語		7万語	①5万語 ②3万語 ③2万語	
2. 辞書の構成内容	①ユーザー辞書 専門用語辞書 基本用語辞書	→マニュアル	tree構造	単語ベース 若干のイディオムと格言あり	BRAVICE方式の辞書は、ユーザーが会話形式でYES又はNOで登録するためその構成内容はユーザには よく解からない
3. ユーザへの開放範囲 (1) 基本辞書	②1でできたことが必ずしも出来ない	①ユーザ辞書， ②専門辞書も解放	全面解放（カーネル以外） （ほとんど当社で作成、 カーネル部は日立で）		辞書の構成がユーザに不明な位だから、解放はされていない
(2) 専門語辞書		科学用専門辞書から原子力分野をとりだして編集したものを使用している	一般辞書で翻訳実験を行ない、辞書チューニングで、一般的なものは一般辞書，ローカルなものは利用者辞書を直す またチューニングテストも利用者辞書で行なう	株式用語 経済用語 株式格言 } 8～10万語	基本辞書と同じ

表3－3（つづき）　機械翻訳システムの辞書調査集計表（ユーザ向け）

No.〔 6/ 12〕

ユーザ / 調査項目	F	G	H	I	J
4．一般辞書、ユーザ辞書の使い分け	①かさねて使う	連結して使用	――	この2つはシステムとして使う その他に使いすての一過性の辞書もある	使い分けはしていない
5．専門辞書の分野分け	まだはっきりしていないが電力，経済等9分野程を考えている	適した分野の辞書を選択して構成	株式分野（限定している）	分野は株式関係に限定	自動車技術 社内用語
6．専門辞書の複数利用成果	ある どっちを上にするかという判断が必要	ない	ない	ない	ない

表3－3（つづき）　機械翻訳システムの辞書調査集計表（ユーザ向け）

No.〔 7/ 12〕

ユーザ / 調査項目	F	G	H	I	J
Ⅳ辞書の更新 1．ユーザによる更新の必要度	かなり必要	とても必要 更新なしで直訳のようになって	とても必要 該当分野の日本語として通用するレベルの訳を得るために必要	かなり必要 株式銘柄は1回／1日定期更新	あまり必要ない 基本辞書6万語を外ど更新する必要はない
2．更新頻度	現在はセッションのたび				
(1) 一般辞書		まあまあする	まあまあする	まあまあする	あまり必要ない
(2) 専門辞書		まあまあする	かなりする	非常に多い	?
(3) その他辞書		（なし）		なし 担当｛更新内容…日経 更新作業…ＣＳＫ	
3．内容 (1) よく追加される語	専門用語、複合語	専門用語、複合語	専門用語、複合語、新語	固有名詞	ホンダ用語
(2) よく削除される語	複合語	あまりしない	不要な訳、古い語	なし	削除しない
(3) 訳語の決定	担当者の判断、専門家の見解	その時の都合と汎用性を考慮して、汎用性のないものはいれない	テキストコーパスのKWICリストを見て決定	専門用語については日経のノウハウにもとづいて決定	専門家にきいて決定
(4) 訳語の統一	あまり必要ない	統一を考える程つっこんで使っていない	まあまあ必要	あまり必要ない （訳語はバラエティーに富む方が望ましい）	全く必要ない 1つの和単語には1つの英単語が対応するシステムであり、訳語を統一する必要はない
4．テスト用例文	研究用アブストラクト	原研レポート 470文使って26文不受理 問題のあるもの100文ほど	全体のソース として1万数千 実テストはこの中から2～3千 テスト内容によりさらに数百しぼり込む	●文（複合文…接続の並列等） ●句（文の構造） 10,000文	不明
5．文法の更新	不可	不可	まあまあする （辞書文法の整合性を保ちルールを強化するため）	まあまあする	不可

表3－3（つづき）　機械翻訳システムの辞書調査集計表（ユーザ向け）

No.〔 8/ 12〕

ユーザ / 調査項目	F	G	H	I	J
V辞書に対する評価 1．メーカー提供辞書	一般辞書…やや満足 専門辞書…やや不満	かなり不満 語数は豊富だが訳語が不適切	――――	とても不満 理由ー ●自然な文法体系ではない ①品詞概念の未熟 ②語の転用処理が未熟	普通
2．ユーザ辞書	やや不満 （語数が少なく役立たない）	ユーザ辞書は200語くらいでほとんどメーカー提供のものにたよっている	普通 語数としては充分 経済一般としてもカバー率高い 専門辞書としてはチューニング中	●句構造文法が日本語の論理に合わない ●翻訳のエキスパートの知識で吸収しきれない	とても不満 自社用語の辞書をもっと充実する必要がある。 (1) より適切な訳語へ更新 (2) 自社用語をふやす
3．メーカー提供辞書の解放	できるだけ解放してほしい	すでにかなり解放してもらっている		開発者とユーザ同一 共同開発したものは両者に属す 専門用語の著作権は日経	解放する必要なし
4．ユーザ辞書の充実 (1) 効果	訳の質の向上		分野に合った訳が出る		訳文がより適切になった
(2) 弊害	全体の管理難しくなる				なし
5．ユーザ拡張部の権利	ユーザにある	ユーザ	ユーザにある	開発者とユーザ同一	ユーザにある
理由	辞書を作るのと同じだから	作った人に権利がある	チューニング，評価等、高度な作業の結果であるから		(1)一種の著作物として保護されるべきである (2)部外秘の部分もあるのであくまでユーザに権利があるべきである
6．翻訳以外への使用	ない	ない	ない	ない	ない

表3－3（つづき）　機械翻訳システムの辞書調査集計表（ユーザ向け）

ユーザ / 調査項目	F	G	H	I	J
VI辞書更新運用体制（ユーザ）	――――		担当専門委員		組織だった運用体制定めていない
1．運用体制		直接ユーザがやる 組織的にはやっていない		●株式銘柄の更新 …1回／月 ●その他…発生に応じて随時　数語／日	（実際に翻訳していく中で辞書を構築していく）
2．一般辞書の用語	いじらない	収集はしてない			
(1) 収集に用いた資料			一般辞書複数	市販辞書	
(2) 収集方法			――――	頻度順	
(3) 辞書修正の理由			――――	文法の体系が異なる	
(4) 翻訳辞書化の手順			①見出し語抽出 ②辞書形式に記述 ③パンチ ④辞書へ登録（バッチ）	●文法的な変換 ●訳語の決定 ●動詞については、格（日本語）と前置詞の対応等補足	
3．専門辞書の用語		収集、辞書化はしていない			
(1) 収集に用いた資料			経済辞書、専門用語集	日経に蓄積させている知識用語を利用	図面用語集（社内刊）
(2) 収集方法					
(3) 辞書修正の理由			専門家にみてもらう 頻度分析をする		
(4) 翻訳辞書化の手順	専門用語を標準辞書で試して、専門家の判断によって辞書に登録		2に同じ		BRAVICEのシステムに従って一語一語登録する
4．辞書更新の運用 (1) 運用規定	なし	簡単なきまり程度 （所内で公式に発表してない．うちわのみで使っている）	基本辞書の更新は複数人間で可否判定 リリース前に実験 個人辞書は自由	文法的に重要な語は、特定管理者しか変更できないようにツールができている	特になし

表3－3（つづき）　機械翻訳システムの辞書調査集計表（ユーザ向け）

No.〔10/12〕

調査項目＼ユーザ	F	G	H	I	J
(2) 運用マニュアル	なし	1ページ程度のもの	ある 管理，バックアップ， マージ，辞書手引	ある	なし
5．更新時の入力伝票のサンプル	キーワード，訳，分野， キーワードが入っている文献のタイトル	まだない	見出し語，更新内容 品詞，品詞の中身 直した理由	なし	なし
6．入力方法 (1) 方式	バッチ方式	会話方式	バッチ方式、会話方式	会話方式	
(2) 手順	マニュアル通り	マニュアルに従って	フィジカル		1.初期メニューで「辞書更新」 2.ソース語入力 3.「登録しますか？」 4.Y 5.品詞の決定 6.訳語入力 7.訳語属性の決定 8.ソース語と訳語との「つながり」 9.改行キーで登録完了
(3) 修正操作の手順	対話型	マニュアルに従って			1.初期メニューで「辞書更新」 2.ソース語入力 3.登録済 4.品詞属性を変える 5.改行キーで完了

表3－3（つづき）　機械翻訳システムの辞書調査集計表（ユーザ向け）

調査項目＼ユーザ	F	G	H	I	J
Ⅶ辞書メンテナンス 1．解放度	①　普通 ②　あまりない	自由に使える 特にⅡは正式な商品でないので自由にやれる	自由に使える	前出（ⅤのⅢ参照）	やや自由に使える （一般的な単語）
2．利用者辞書エディッタ		ある	ない（専門家向け）	ある	ない
(1)　特徴		Ⅰは簡単であるが，コマンドを知らないとできない Ⅱはメニュー方式なのでコマンドを知らなくともできる ただし、入れる量は多い			
(2)　使い勝手	やや使い易い	やや使いにくい （ⅠもⅡも）	やや使いにくい	非常に使い易い	
3．メンテナンスのオリジナルツール	なし	なし	ある （統計管理、必要語リストアップ、見出し語コピー）	単語／修正ツール 熟語単語／修正ツール …熟語の登録・修正 単位＝句（文法的な構造）	不明
4．会話式修正とバッチ式修正の比率	会話方式のみ	会話方式のみ	会話方式の方が多い	会話方式のみ	会話方式のみ

表3－3（つづき）　機械翻訳システムの辞書調査集計表（ユーザ向け）

ユーザ / 調査項目	F	G	H	I	J
Ⅶその他 1．他システムとの辞書内容変換	全くしない	全くしない	全くしない	全くしない	全くしない
2．ユーザ辞書のメーカーへのフィードバック	行っている （専用のクレーム用紙をメーカーに出す）	ユーザ側にできる体制がない		日々の運用そのものがフィードバック	メーカー、ユーザともにそのような体制にない
3．理想		一般辞書として完璧なものがメーカーから出ること しかしユーザ辞書は必要			
(1) 辞書体系	？		基本辞書が各分野の知識まで持っていてそれをルールで使いわけられる	基本用語の辞書がしっかりしていて汎用性が高いこと	不明
(2) メンテナンス機能	？	マニュアルなしでできること	ポストエディット等の内容を学習する	充実した基本辞書上で専門用語の追加をする程度	不明
4．他機関との連携	メーカー以外なし 共同開発はやっていない	今はなし 辞書を使わせてもらうなどは今後あり得る	日立との共同開発	なし	社内のみ
		その他 ・訳語選択あると便利 ・かっこを入れるとリージョンオーバーするなど、システムがまだこなれていない ・公的な所で共通に開発できる所はした方がよいユーザが別々にやると手間もかかるし、質も落ちてしまう		その他 ・機械翻訳はあくまで人間が行う翻訳のサポートである ・辞書の構築について 行政が介入するのではなく民間にまかせた方がよい 行政はそのバックアップでよい ・もっと多元的なバックアップがあった方が良い 方法論が定まっていない時に集中してしまうのはよくないのではないか	

4．海外における辞書の作成及び標準化動向

4. 海外における辞書の作成及び標準化動向

4.1 海外調査の概要

海外における辞書の作成および標準化の動向を調査するために，田中裕一専門委員会委員を北米およびヨーロッパ諸国の関連機関に派遣した。期間は昭和６１年１１月３日から１１月２３日までの３週間である。

(1) 目　　的

本海外調査の目的は次のとおりである。

① コンピュータ用辞書の作成を行っている機関の調査

② コンピュータ用辞書の標準化の動向の調査

③ コンピュータ用辞書の更新の体制および更新用インタフェースの調査

④ ウォータールー大学でのコンピュータ用辞書に関する国際会議への出席

(2) 調査対象機関

調査対象機関および調査日は次のとおりでである。

① １１月　４日：国立医学図書館（米国）

② １１月　５日：アメリカ保健機構（米国）

③ １１月　６日：ホートン・ミフリン出版社（米国）

④ １１月　７日：モントリオール大学（カナダ）

⑤ １１月１０～１１日：ウォータールー大学（カナダ）

⑥ １１月１３日：ロングマン出版社（英国）

⑦ １１月１４日：オックスフォード大学出版部（英国）

⑧ １１月１７日：EUROTRAプロジェクト（ルクセンブルグ）

⑨ １１月１７日：EURODICAUTOMプロジェクト（ルクセンブルグ）

⑩ １１月１８日：ボン大学（西独）

⑪ １１月１９日：ジーメンス（西独）

⑫ １１月２０日：INFOTERM（オーストラリア）

⑬ １１月２１日：ピサ大学（イタリア）

4.2 海外調査結果

4.2.1 海外調査結果の概要

(1) 辞書更新体制

本項と次項で述べる辞書の更新の問題は，対象となる辞書の性格，その辞書を扱う組織の性格によって全く違ったものとなる。辞書が一つのデータベースをなしていて半固定的なものの場合，それに対する更新は辞書の改訂と同じ意味をもつ大がかりなものとなる。一方，応用システムのユーザが自分専用に用意した辞書は，ユーザ個人の責任で自由に更新ができる。システム開発者によって応用システムに組み込まれた辞書の場合には，更新の意味はこれらの中間的なものとなるだろう。

このことから当然であるが，辞書更新体制についての態度は出版社と研究機関とで大きく異なる。出版社においては辞書は商品であり，その内容に関して責任を持つと同時に，その内容をユーザには公開しない。従って更新の体制はユーザからは見えない。通常の冊子体辞書の編集・増補・改訂体制とほぼ同じであると言ってよいだろう。例外としてはオックスフォード大学出版部のNew OEDがあるが，この場合は辞書のデータベース化それ自身がウォータールー大学での研究対象であるので，その更新（この場合は入力されたデータに対する校正）も研究者に対してオープンになっている。

一方，実際に日常的に運用されているシステムの場合は更新の体制が確立されていることが多い。更新はユーザ側でなく，システムの提供者（開発者）側で行われる。米国の国立医学図書館やECのEURODICAUTOMのように大きな組織の大規模な辞書においては，辞書の更新を専門的に管理する委員会が存在する。更新の要求は委員会の審議を経て受け入れられるか却下される。またアメリカ保健機構のように小規模な組織の場合には更新の担当者と更新手順が定められているが，その内容については担当者個人に任されている。いずれにしても辞書の更新は一つの組織内で一元的に管理され，その責任も明らかになっている。

研究機関において開発途中の辞書の場合には更新の体制はほとんど無きに等しい。特に今回の調査では大学で開発中の辞書が多かったので，規模

も小さく，１～２人の担当者で適当に判断しているという印象を持った。（その一方で更新のインタフェースには凝ったものが多かった。）

(2) 辞書更新インタフェース

ここでも上記のような組織による差が現れている。出版社の場合には「原稿に朱を入れる」感覚が残っているようであり，編集者は計算機の出力用紙上で校正し，それを専門のオペレータが計算機に入力していた。その入力にはワークシート形式の単純なエディタが用いられているようである。

New OEDはここでも例外である。詳しくは機関別調査結果の項で述べるが，IBMマシン上で動くNew OED用のエディタはカラー・ディスプレイを活用して記述項目の違いやフォントの違いを明示し，テキストの表示を大域的にも局所的にも切り換えることができる見事なものである。操作も非常に簡単で，編集室では出版社の辞書編集者が端末の前に座り，自らキーボードを操作していた。

また大規模な組織では更新のインタフェースは一般にさほど重んじられていない。これは更新の手続きがユーザからは切り離されているため，むしろ更新内容の妥当性の吟味等に重点が置かれ，更新の作業自体はやはり専門のオペレータに任せることになるからである。

アメリカ保健機構の機械翻訳システムはやや特殊である。この小規模なシステムはここで開発されたものだが，現在のところ機構内でのみ使われているので，辞書に規模，用法上の階層はなく（一方向の翻訳に対して）一種類で済んでいる。従って更新はユーザ辞書と同じであると思ってよい。更新を行うのはポスト・エディットを担当する人間の翻訳者であり，計算機には素人であるので，非常に丁寧なインタフェースが提供されている。なお，計算機はIBMマシンであるが，翻訳前後の文書の処理，ポスト・エデイット，辞書の更新等の処理には端末として用いられているWangのワークステーションのローカルな機能を利用している。

大学の辞書システムには更新のために便利なインタフェースを用意しているものが多かった。計算機科学科等の学生がアルバイトで作ったものもある。計算機は大部分がIBMのPCであり，小回りの効く点を活かして

相当凝ったものを作っていた。なお，ボン大学には大規模な辞書があり，大型機上で検索，更新が一応できるようになっているが，このエディタは古色蒼然たる代物で，ほとんど使われていないようだった。人材不足が原因かと思われる。

なお，辞書の更新時に他のシステム，例えばKWIC等，を利用したいという要求が起こることは自然であろうが，それを考慮したものはほとんどなかった。IBMのPC程度ならば（あるいは大型機の端末とワークステーション程度ならば）2台並べておく方が簡単だということらしい。日本では1台のワークステーション上でマルチ・ウィンドウを用いて何通りもの仕事を同時に進めるということが流行しているが，そのようなシステムは皆無であった。

(3) 大規模化および標準化への動き

日本における計算機用辞書の環境と英語圏のそれとを比較すると，最も顕著な差はファイル化された辞書（MRD）が数多く出版されていることである。これには有名なLDOCE（Longman Dictionary of Contemporary English）を始めとして，ウェブスター，アメリカン・ヘリテージ等の冊子体辞書として既に広く使われているものが多く含まれている。現在，計算機用辞書の大規模化や標準化が声高に求められているのは，主に機械翻訳システムの開発を通じてのこのようなMRDにおける経験が基になっているものと考えられる。

詳しくは後のLDB（Lexical Database）の章で述べるが，冊子体辞書の構造，内容に多く依存したMRDに代るものとして，語に関する情報を計算機用のデータベースの形で提供する新たなLDBなるものの構想がある。これは必然的に大規模化，標準化を伴うが，さまざまな組織でそのための検討が進められている。例えば，ESF（European Science Foundation ― ヨーロッパ諸国のNational Science Foundationの合同組織）が主体となって開かれたフォーラム"計算機用辞書の標準化"（資料編のプログラム参照）などはその動きを示すものである。また，計算機用辞書における文法事項の標準化は可能かという問題も（本委員会でも昨年度の調査事項に含めていたが）さまざまな言語理論の専門家を集めて検討しよ

うとしているということである。

4.2.2　機関別調査結果

4.2.2.1　国立医学図書館（米国）

(1) 名　　称

National Library of Medicine

National Institute of Health

Public Health Service

U. S. Department of Health and Human Services

(2) 住　　所

8600 Rockvill Pike, Bethesda, MD 20894, U.S.A.

(3) 訪問日時

11月4日（火）　10：30～17：00

(4) 対 応 者

Richard K. C. Hsieh, Assistant Director, International Programs

David Kenton, Information Systems

Roy Rada, Editor IM, Chief MeSH, Research Fellow

Atsuko Craft, Index Section

Alexa McGay（MeSH担当の言語学者）

Peri Schuyler（Unified Medical Languageを担当）

Larry Kingsland（エキスパートシステムおよびAIを担当）

(5) 組　　織

米国保健福祉省（U. S. Department of Health and Fuman Services）の一機関である。ワシントンDCの北郊，ベセスダに広大な敷地を占めるメディカル・センターの一画にあり，一般に公開されている図書館とLister Hill Centerと呼ばれる研究施設とを持つ。150年の歴史を持ち，現在は数百人がそこに働いている。

(6) 概　　要

① MEDLARS

"MEDLARS（Medical Literature Analysis and Retrieval System）"とはNLMが提供している医学・生化学関係の多数のオンライン

DBの総称である。これには次のような21種のDBが含まれる。

AVLINE, BIOETHICSLINE, CANCERLIT, CANCERPROJ, CATLINE, CCRIS, CHEMLINE, CLINPROT, DIRLINE, HAZARDOUS SUBSTANCES DATA BANK, HEALTH, HISTLINE, MEDLEARN, MEDLINE, MeSH VOCABULARY FILE, NAME AUTHORITY FILE, POPLINE, PDQ, RTECS, SERLINE, TOXLINE.

（この中にはNLM以外で開発されたDBも含まれている）

MEDLARSに採録される記事は世界36ケ国で発行される3,500種以上の専門誌から集められる。1966年以降の分がDB化されているが，現在カレントファイルとしてアクセスできるのは最近3年間分である。

これに対するアクセスは非常に容易で，専門家でなくとも家庭のパソコンから呼び出すことができる。（そのためのツールが発売されている。）また，図書館のホールの一角に専用のターミナルが7～8台置かれていて，図書館利用者が使えるようになっている。オンラインサービスは国外のユーザもいるので，年中無休24時間稼働である。機種はIBM 3081－Kである（近々2台になる）。図4－1にMEDLINEの出力例を示す。

なお，MEDLINEを始めとするいくつかのDBについては，JICSTが磁気テープを持って日本でのオンラインサービスを行っている。

1 MEDLINE
Fisher B
Adjuvant therapy for breast cancer: a brief overview of the NSABP experience and some thoughts on neoadjuvant chemotherapy.
Recent Results Cancer Res 1986;103:54-68
Unique Identifier: 86288654

Medical Subject Headings:

Age Factors
Antineoplastic Agents, Combined--THERAPEUTIC USE
Breast Neoplasms--DRUG THERAPY--SURGERY
Clinical Trials
Combined Modality Therapy
Comparative Study
Doxorubicin--ADMINISTRATION & DOSAGE
Female
Fluorouracil--ADMINISTRATION & DOSAGE
Human
Lymphatic Metastasis
Melphalan--ADMINISTRATION & DOSAGE--THERAPEUTIC USE
Methotrexate--ADMINISTRATION & DOSAGE
Middle Age
Receptors, Estrogen--METABOLISM
Support, Non-U.S. Gov't
Support, U.S. Gov't, P.H.S.
Tamoxifen--ADMINISTRATION & DOSAGE

2 MEDLINE
Fisher B; Fisher ER; Redmond C
Ten-year results from the National Surgical Adjuvant Breast and Bowel Project (NSABP) clinical trial evaluating the use of L-phenylalanine mustard (L-PAM) in the management of primary breast cancer.
J Clin Oncol 1986 Jun;4(6):929-41
Unique Identifier: 86226429

Medical Subject Headings:

Actuarial Analysis
Age Factors
Breast Neoplasms--DRUG THERAPY--PATHOLOGY--SURGERY
Cell Nucleus--PATHOLOGY
Clinical Trials
Combined Modality Therapy
Female
Human
Lymph Nodes--PATHOLOGY
Mastectomy
Melphalan--THERAPEUTIC USE
Middle Age
Neoplasm Recurrence, Local
North America
Placebos
Random Allocation
Support, Non-U.S. Gov't
Support, U.S. Gov't, P.H.S.

図４－１　MEDLINEの出力例（端末のハードコピー）

② MeSH

MeSH（Medical Subject Heading）とは，MEDLARSに採録された文献中に現れる専門語を集めたシソーラスである。基本的に約15,000語がディスクリプタとして分類されて登録されているが，化学用語約40,000語がそれに加えられている。各ディスクリプタは同義語を代表し，それを含めた総数は数十万語に及ぶ。これにより，概念の上位 ― 下位関係を検索できるほか，キーワードの標準化の機能も果たしている。

MeSHはオンラインDBになっている（上記参照。操作例については図4－2参照。）。また，同時に印刷もされていて，

a) MeSH－Tree Structures

b) Permuted MeSH

c) MeSH－Annotated Alphabetic List

d) MeSH－Supplementary Chemical Records

```
USER:

------------------------------------------------------------------------
1
MESH HEADING                    Herpes Zoster
DESCRIPTOR TYPE                 1
DESCRIPTOR CLASS                1
MESH TREE NUMBER                C2.256.466.423
MAJOR REVISION DATE             770315
RECORD ORIGINATOR               C MED ; M CAB ; M CAB
BACKWARD XREF                   SHINGLES 1 00000000 0000000 740330 ;
                                ZONA 1 00000000 0000000 740330 ;
                                ZOSTER 1 00000000 0000000 740330
ANNOTATION                      caused by HERPES ZOSTER VIRUS (see
                                VARICELLA-ZOSTER VIRUS)
RECORD TYPE                     D
MED83 POSTINGS                  159 ; *102
MED80 POSTINGS                  350 ; *248
MED77 POSTINGS                  313 ; *204
MED75 POSTINGS                  258 ; *156
MED71 POSTINGS                  383 ; *235
MED66 POSTINGS                  378 ; *278
SS 2 /C?
```

図4－2　Online MeSHの操作例（端末のハードコピー）

という形でNTIS㈱から発行されている。価格は各$20程度である。a)は15,000のディスクリプタを木構造で分類した形のまま印刷したものである。最上位概念の数は76で，全体は森の形をしている。 b)は複合語に対するもので，先頭語以外からも検索できるようになっている。c)は注釈付でアルファベット順に並べたもの， d)は化学用語40,000語に対するものである。図4－3にシソーラスの例を示す。

〔MeSHの更新〕

MeSHセクションは各専門分野の専門家を所内にかかえている。彼らは専門誌に目を通し，現場の専門家とディスカッションを行うことにより，用語を集め，定義を与え，他の用語との関係付けを行う。ここで集められた用語は，勧告の形で上部の委員会（委員長：Roy Rada，委員は数名）に上げられ，そこで採択の可否が決定される。

更新は毎週のように起こる。化学用語に関しては週に4件ということもある。オンラインDBはそれに伴って順次更新されるが，印刷物のほうは年に1回（10月末）改定される。

なおDBの更新作業は専門のオペレータが行うので，特にそのためにインタフェースを作ったということはない。

㈱ NTIS

National Technical Information Service

U. S. Department of Commerce

5285 Port Royal Road, Springfield, VA 22161

なお，日本からの注文は下記のManaging Dealerを通じなければならない。

Attn：Mr. H. Hori, NTIS Manager

〒100 東京都千代田区大手町2－3－6

タイムライフビル

三菱総合研究所

C4 - DISEASES-NEOPLASMS

NEOPLASMS

•NEOPLASMS	C4			
APUDOMA	C4.70			
CHORISTOMA	C4.131			
CYSTS	C4.182			
BONE CYSTS	C4.182.89	C4.557.499.	C5.316.7..	
JAW CYSTS	C4.182.89.530	C5.500.470		
NONODONTOGENIC CYSTS	C4.182.89.530.660	C5.500.470.		
ODONTOGENIC CYSTS	C4.182.89.530.690	C5.500.470.		
BASAL CELL NEVUS SYNDROME •	C4.182.89.530.690.150	C4.557.117. C4.133.[7..]	C5.316.99.	C5.500.470.
DENTIGEROUS CYST	C4.182.89.530.690.310	C5.500.470.		
PERIODONTAL CYST	C4.182.89.530.690.790	C5.500.470.	C7.465.714.	
RADICULAR CYST	C4.182.89.530.690.790.820	C5.500.470.	C7.465.698.	
BRANCHIOMA	C4.182.117			
DERMOID CYST	C4.182.201	C4.557.337.		
EPIDERMAL CYST	C4.182.254			
ESOPHAGEAL CYST	C4.182.281	C5.306.316		
FIBROCYSTIC DISEASE OF BREAST	C4.182.289	C19.246.370		
KIDNEY, CYSTIC	C4.182.394	C12.777.419.		
KIDNEY, POLYCYSTIC	C4.182.394.420	C12.777.419.	C14.331.809.	
KIDNEY, SPONGE	C4.182.394.586	C12.777.419.		
MEDIASTINAL CYST	C4.182.444	C23.304.612.		
MESENTERIC CYST	C4.182.473	C6.772.628		
MUCOCELE	C4.182.511			
OVARIAN CYSTS	C4.182.612	C13.371.56.		
POLYCYSTIC OVARY SYNDROME	C4.182.612.765	C13.371.56.		
PANCREATIC CYST	C4.182.640	C6.689.500		
PANCREATIC PSEUDOCYST •	C4.182.640.692	C6.689.500.		
PAROVARIAN CYST	C4.182.668	C13.371.56.		
PILONIDAL CYST	C4.182.710			
RANULA	C4.182.766	C7.465.780		
SYNOVIAL CYST	C4.182.867			
THYROGLOSSAL CYST	C4.182.902			
URACHAL CYST	C4.182.946			
NEOPLASM REGRESSION, SPONTANEOUS	C4.495	C4.805.670		
NEOPLASMS BY HISTOLOGIC TYPE (NON MESH)	C4.557			
CARCINOID TUMOR	C4.557.112			
MALIGNANT CARCINOID SYNDROME	C4.557.112.524			
CARCINOID HEART DISEASE •	C4.557.112.524.329	C14.280.129		
CARCINOMA	C4.557.117			
CARCINOMA 256, WALKER	C4.557.117.45	C4.619.45		
CARCINOMA, BASAL CELL	C4.557.117.110			
BASAL CELL NEVUS SYNDROME •	C4.557.117.110.150	C4.557.99. C4.133.[7..]	C5.316.99.	C5.500.470.
CARCINOMA, BASOSQUAMOUS	C4.557.117.141			
CARCINOMA, BROWN-PEARCE	C4.557.117.172	C4.619.124		
CARCINOMA, DUCTAL	C4.557.117.234			
PAGET'S DISEASE, EXTRA-MAMMARY	C4.557.117.234.500			
PAGET'S DISEASE OF BREAST •	C4.557.117.234.666			
CARCINOMA, EHRLICH TUMOR	C4.557.117.265	C4.619.149		
CARCINOMA IN SITU	C4.557.117.327			
CARCINOMA, KREBS 2	C4.557.117.358	C4.619.234		
CARCINOMA, MUCINOUS	C4.557.117.390	C4.557.574.		
KRUKENBERG'S TUMOR	C4.557.117.390.500	C4.557.574. C13.371.56.	C4.588.322.	C4.588.945.
CARCINOMA, OAT CELL	C4.557.117.421			
CARCINOMA, PAPILLARY	C4.557.117.452			
CARCINOMA, SCIRRHOUS	C4.557.117.483			
LINITIS PLASTICA •	C4.557.117.483.515			
CARCINOMA, SQUAMOUS CELL	C4.557.117.490			
BOWEN'S DISEASE •	C4.557.117.490.318	C17.302.137		
CARCINOMA, TRANSITIONAL CELL	C4.557.117.514			
LEUKEMIA	C4.557.337			
LEUKEMIA, EXPERIMENTAL	C4.557.337.372	C4.619.531		

• INDICATES MINOR DESCRIPTOR

図４－３　MeSH — Tree Structures, 1986（冊子体シソーラス）の頁の例

4.2.2.2　アメリカ保健機構（米国）

(1) 名　　称

Pan American Health Organization

Pan American Sanitary Bureau

Regional Office of the World Health Organization

(2) 住　　所

525, 23rd Street, N. W., Washington, DC 20037, U.S.A.

(3) 訪問日時

11月5日（水）　9：30～12：00

(4) 対 応 者

Lee Ann Schwartz, Machine Translation Program

Muriel Vasconcellos, Chief, Terminology and Machine Translation Program

(5) 組　　織

独立した組織だが，世界保健機構（WHO）の下部機関（regional office）の役割も果たしている。担当は，南北両アメリカ大陸の諸国である。ワシントンDCの市内の2つのビルに本部があり，そこに数百人の所員がいる。また，各地の支部やオフィスにも数百人の所員がいる。組織内での公用語は英・西・葡・仏であるが，英・西が主に使われている。

機械翻訳セクションはその内部にあり，Muriel Vasconcellosを長とする6人で構成されている。

(6) 概　　要

① 機械翻訳システムの開発の経緯

1970年代より機械翻訳（MT：Mechine Translation）システム開発の構想はあった。Georgetown 大学の協力を求め，そこで作られたGAT（Georgetown Automatic Translator ― 露英）の基本的な構想を参考にすることにした。1980年よりスペイン語→英語のMTシステムSPANAMの開発を，1985年より英語→スペイン語のENGSPANの開発を始めた。また，現在は英語→ポルトガル語のENGPORTを開発中である。ENGSPANに対してはUSAID（U.S. Agency for In-

ternational Development）から２年間の資金援助を受けた。開発に携わった人間は６人であるが，その他にGeorgetown 大学の学生が協力した。

② 機械翻訳システムの概要

ＭＴシステムはＩＢＭ−４３８１の上で動いているが，Wang OIS／140の端末上の欧文ワードプロセッサを用いて（そのコマンドの一つとして）実行させるようになっている。入力用にＯＣＲも使うことができる。プリエディットは行わず，翻訳者（従来から１０人位いる）によってポストエディットを行う。

今までのところ，このＭＴシステムは内部で使われているだけだが，外部機関（nonprofit institution）での使用を検討中である。現在，ＵＳＡＩＤにあるＩＢＭ−３０８１に移植作業中である。１９８７年1月には"Consultative Group on International Agricultural Research"での使用を予定している。

ＭＴシステムの詳細はＡＣＬ，Vol.11，Nos.2−3に発表されている通りである。解析はＡＴＮパーザにより行われる。プログラムはＰＬ／Ｉで書かれ，翻訳速度は約１８０,０００（words／hour）である。ＥＮＧＳＰＡＮによる翻訳例を図４−４に示す。

③ 辞書の概要

辞書の一レコードは，原言語・目標言語の単語対，識別番号，品詞，および専門分野コードよりなる。原言語・目標言語の単語対を識別するために１２桁の識別番号が用いられる。上位６桁は原言語の単語を識別し，下位６桁はそれに対応する（一般には複数個の）目標原語の単語を識別する。例えば一対一に対応するとき下位６桁は０００００であり，多品詞語のとき名詞の単語対には０００１０がつき，動詞の単語対には０００２０がつく。

当初，ＳＰＡＮＡＭの辞書は市販の辞書（西英辞典）から単語対を抽出して作られた。それに更新・追加を施した結果，現在は約６万語のエントリーを持つ。ＥＮＧＳＰＡＮの辞書はＳＰＡＮＡＭの辞書を逆変換することによって原型が作られた。現在は約４万語のエントリーを持つ。

Each participant, who should have actual experience in using microcomputers in one of these areas of education or training, will be asked to prepare a demonstration or paper describing that experience.	OK TU	A cada participante, que debe tener experiencia real en empleo de microcomputadores en una de estas áreas de educación o capacitación, se le pedirá que prepare una demostración o documento que describe esa experiencia.
We request that each potential participant send biographical information and a 100-word abstract of the demonstration or paper, so that we can select those who will make the greatest contribution to a useful exchange of information at the symposium.	OK TU	Solicitamos que cada participante potencial envíe información biográfica y un resumen analítico de 100 palabras de la demostración o documento, para que podamos seleccionar los que harán la contribución mayor a un intercambio útil de información en el simposio.
The degree of behavioral impairment was found to be roughly proportional to the treatment dose level.	OK	El grado de deterioro conductual se halló que era aproximadamente proporcional al nivel de dosis de tratamiento.
Laboratory studies have shown marijuana to impair perceptual and perceptualmotor functions important to driving.	OK TU	Los estudios de laboratorio han revelado que la marihuana deteriora funciones de percepción y perceptomotrices importantes a conducción.
The task of hiring and assigning staff to perform the work is one which must be completed prior to training.	OK	La tarea de contratación y asignación de personal para realizar el trabajo es una que se debe completar anterior a capacitación.
Laboratory studies (e.g., roentgenograms, coagulation studies, and blood tests) are useful in delineating the nature and extent of current trauma and in defining the presence of previous trauma.	OK	Los estudios de laboratorio (por ejemplo, roentgenogramas, estudios de coagulación, y pruebas sanguíneas) son útiles en delinación de la naturaleza y grado de trauma actual y en definición de la presencia de trauma anterior.
WHO recommends that attention first be paid to immunizing children and pregnant women already using existing health facilities.	OK TU	*La OMS recomienda que la atención primero se preste a inmunización de niños y mujeres embarazadas ya empleando establecimientos de salud existentes.
The theoretical design we propose should permit us to study various lifestyles within the macroscopic political, economical and cultural context of the nation.	OK	El diseño teórico que nosotros proponemos debe permitirnos que estudiemos diversos estilos de vida dentro del contexto macroscópico político, económico y cultural de la nación.
The mothers were asked to bring their children to the vaccination center.	OK TU	A las madres se les pidió que trajeran a sus niños al centro de vacunación.
This approach brings immunization services to communities with difficult access to health services.	OK	Este enfoque les trae servicios de inmunización a las comunidades con acceso difícil a servicios de salud.
The proposal must include a list of the facilities which are to be made available.	OK	La propuesta debe incluir una lista de las instalaciones que se harán disponibles.

図４－４　ENGSPANによる翻訳例

④ 辞書の更新

単語の追加登録は主にポストエディタの仕事である。実際，翻訳誤りの大部分は未登録語に起因するので，ポストエディタが辞書の更新を行うのは非常に効率がよい。一日の仕事の終りに1時間，辞書の更新のための時間がとってある。更新のためのプログラムは，ユーザ・インタフェースの良いものであるという。

ＵＳＡＩＤでは辞書の更新を専門に行うためのパートタイムの人間を雇う予定である。一人は英語を母国語とし，もう一人はスペイン語を母国語とする。この二人のコストとして，登録に2ドル／語，修正に0.25ドル／語を想定している。

4.2.2.3 ホートン・ミフリン出版社（米国）

(1) 名　称

Houghton Mifflin Company

(2) 住　所

One Beacon Street, Boston, MA 02108, U.S.A.

(3) 訪問日時

１１月６日（木）　１１：００～１５：００

(4) 対応者

Howard Webber Vice President and Publisher

David R. Replogle, Executive Vice President

Mary Sutherland, Director of Marketing

Mary Ellen Hoke, International Account Manager

Joanne Dudek, Technical Support Division

(5) 組　織

アメリカの大手出版社であり，広い分野の本を出版している。本社はボストン市の中心にあるビルの28階から32階を占める。

辞書の分野では"American Heritage Dictionary"や"Roget's-Ⅱ Thesaurus"の出版で知られている。現在，辞書部門（Reference Division）では言語ＤＢの設計・開発といった活動中心である。また，1985年には社内にビジネス・ソフト部門を新設し，言語ＤＢを用いた各種アプ

リケーション・システムの開発を行っている。

(6) 概　　要

① MRD vs LDB

MRD（Machine-readable Dictionary）は冊子体の辞書をその形式のままで電子化したものである。現在，この種のものは多数世にでているが，実験的な用途はともかく，現実の応用範囲は限られ，もう役目を終ったと言って良いと考えている。それに対して，今必要とされているのはLDB（Lexical Database）の開発である。これは言語情報をDB化したものであり，この情報を抽出して新しいDB（さらには冊子体の辞書まで）を作ったり，自然言語アプリケーション・システムに組み込んだりすることができる。ここでは社内のMRDをもとにして，積極的にLDBの開発を進めている。

② 活　　動

最近の主な活動には

- American Heritage DictionaryによるDBの開発
- 米・英・仏・独・西・伊語のスペル・チェッカ，コレクタの商品化
- Roget's-Ⅱ Thesaurusによる類語生成プログラムの商品化
- 文法修正プログラムの商品化
- OEDのDBの設計
- 辞書の改訂
- 英語の大規模LDB
- 新しい頻度データの調査
- 意味ネットワークによるAmerican Heritage Dictionaryの記述の表現

などがある。商品化したDBやソフトは，ライセンス契約により国内外の多数のメーカの製品に組み込まれている。

③ 辞書の更新（改訂）

辞書は応用システムに組み込まれて用いられるので，ユーザによる更新は全く考えていない。社内では通常の冊子体辞書の改訂作業と同様の

手続きで行っている。言語の変化のスピードが速くなっているので，更新の頻度は非常に多い。例えば，スペルの上でも，この数年で"data base" → "database"や"on-line" → "online"などと変化しているので，これらを許容するように修正が必要になる。

4.2.2.4　モントリオール大学（カナダ）

(1) 名　　称

University of Montreal

(2) 住　　所

3150, Jean-Brillant, CP 6128, succ. A,
Montreal, Quebec, Canada, H3C 3J7

(3) 訪問日時

11月7日（金）　13：45～16：00

(4) 対 応 者

Jean Baudot, Professor, Department of Linguistics and Philology
Jean-Yves Morin, Department of Linguistics and Philology
Andre Clas, Editor／Director, "Meta"
Richard Kittredge, Department of Linguistics and Philology

(5) 組　　織

モントリオール大学の言語学科を訪問した。このグループではBaudot教授を中心に，専門語辞書，機械翻訳，知識表現などの研究をしている。

(6) 概　　要

① 機械翻訳

カナダ政府からの機械翻訳への補助金が昨年で打ち切られており，機械翻訳の研究は現在全く行われていない。

② 専門語辞書

専門語は小規模な検索システムを実験中。専門語は複合語の扱いが難しいので，特別なインデックスを工夫した。

4.2.2.5　ウォータールー大学〔第2回New OEDセンターコンファレンス〕（カナダ）

(1) 名　　称

Second Annual Conference of the

University of Waterloo Centre for the New OED

(2) 住　　所

University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1

(3) 訪問日時

１１月１０日（月）～１１日（火）

(4) 対 応 者

Gaston H. Gonnet, Department of Computer Science（プログラム委員長）

Frank W. Tompa, Department of Computer Science（プログラム委員）

(5) 概　　要

今回のコンファレンスのテーマは"Advances in Lexicology"であり，発表の題目は次のとおりである。

① "Morpheme-based Lexical Analysis"

Manfred Gehrke, Hans-Ulrich Block（ジーメンス）

② "The Role of Etymology and Word Length in English Word Formation"

Frank Anshen, Mark Aronoff（ニューヨーク州立大，Stony Brook）

Roy Byrd, Judith Klavans（IBMワトソン研究センター）

③ "Thesaural Knowledge Representation"

Sally Yeates Sedelow, Walter A. Sedelow Jr.（アーカンサス大）

④ "Explicit and Implicit Information in Dictionaries"

Beryl T. Atkins（コリンズ出版社）

Judy Kegl（プリンストン大）

Beth Levin（MIT）

⑤ "Building a Lexical DB by Parsing Webster's 7th New Collegiate Dictionary"

Thomas Ahlswede, Martha Evens, Kay Rossi（IIT）

Judith Markowitz（ナビスター・インタナショナル）

⑥ "A Project for a Bilingual Lexical Database System"

Nicoletta Calzolari（ピサ大）

Eugenio Picchi（イタリア計算言語学研究所）

⑦ "New OED Status Report"

Timothy Benbow, John Simpson（オックスフォード大学出版部）

Gaston Gonnet, Frank Tompa（ウォータールー大）

⑧ "Towards an Objective Analysis of the Age and Origins of the English Vocabulary"

Peter Davies（OED編集委員）

⑨ "The Lexicon and Commonsene Reasoning : A Case Study"

Alexander Nakhimovsky（コルゲート大）

⑩ "A Conceptual Dictionary for Contextually Based Structure Selection"

Fabrizio Sebastiani, Giacomo Ferrari（ピサ大）

Irina Prodanof（イタリア計算言語学研究所）

⑪ "Why Dictionaries Should List Case Structure"

Graeme Hirst（トロント大）

②や⑧はNew OEDに含まれる語源の記述を利用しての研究であり，この辞書に特徴的なものと言える。⑦はウォータールー大学で進められているNew OEDのデータベース化作業の現況報告である。これ以外は特にNew OEDに限らない計算機用辞書のまわりの研究の発表であった。

興味深い発表からいくつか紹介する。③はロジェのシソーラスを用いて意味空間を数学的に構築しようというものである。シソーラスに暗示的に含まれている情報は意味空間を作るのに適しており，人間用のシソーラスは機械処理には向かないという批判には根拠が無いと主張している。④は辞書は情報をすべて明示的に記述すべきかという議論である。多くの冊子体の辞書では意味論的あるいは語用論的な情報を明示的に書かずに，例文をいくつかあげることで済ませている。構文的な情報も例文で表現されることがある。そのような傾向は計算機処理にとってマイナスだが，明示的

に書くことは可能かという問題を工学的，辞書学的，心理学的などさまざまな角度から扱っている。⑪もこれに関連する話題である。⑨は自然言語理解における時に関する推論を行うための知識表現の研究である。ここでは過程，状態などを表す動詞の意味論が扱われている。

なお，①については以下のジーメンスの項で，⑥と⑩についてはピサ大学の項で，⑦についてはオックスフォード大学出版部の項でそれぞれ説明する。

4.2.2.6　ロングマン出版社（英国）

(1) 名　　称

Longman Group Limited

(2) 住　　所

Longman House, Burnt Mill, Harlow, Essex CM20 2JE, England

(3) 訪問日時

１１月１３日（木）　１０：００～１４：００

(4) 対 応 者

Michael Rundell, Managing Editor, ELT Reference

Nikolai Dejevsky, Publisher, Dictionaries and Reference Division

(5) 組　　織

辞書の出版で知られるイギリスの大手出版社である。ロンドンから汽車で３０分のハーロウという町に，辞書編集部門がある。（本社はロンドン市内）

辞書編集部門にはELT（English Language Teachers Section）と呼ばれる外国人用の辞書の編集部と，一般英語国民用の辞書の編集部とがある。例えば，LDOCE（Longman Dictionary of Contemporary English）の編集は前者が行い，Roget's Thesaurusの編集は後者が行っている。

(6) 概　　要

① MRDの開発

社内では編集用にほとんどの辞書をMR化しており，校正・改訂等の作業はすべて計算機の出力用紙の上で行われている。計算機はVAXである。端末は８台あり，朱の入った原稿をもとにオペレータがエディタ

で修正している様子が見られた。

商品としてはLDOCEのMR（Machine-readable）版があるが，それ以外の辞書をDB化するなどして発売する予定は今のところない。また，辞書を組み込んだ応用システムなどの開発を社内で行うことは考えていないという。

4.2.2.7　オックスフォード大学出版部（英国）

(1) 名　　称

Oxford University Press（略称：OUP）

(2) 住　　所

Walton Street, Oxford, OX2 6DP, England

(3) 訪問日時

１１月１４日（金）　１５：００～１７：００

(4) 対 応 者

Edmund Weiner, Editor, New Oxford English Dictionary

Timothy J. Benbow, Director, New Oxford English Dictionary

John A. Simpson, Co-Editor, New Oxford English Dictionary

(5) 組　　織

学術書を専門とする英国の大手出版社である。ロンドンから汽車で１時間のオックスフォードに本社およびすべての部門がある。New OED のDB化の作業のうち，校正はすべてここの編集部が担当している。

(6) 概　　要

① New OED プロジェクト

New OEDのDB化は１９８６年９月から１９８９年８月までの３年計画である。作業の内容とその分担は次の通りである。

- ハードウェアおよびソフトウェア・ツール提供：英国IBM
- 入力（キーボーディング）：ICC（Reed Int'l の子会社，アメリカ）
- DB化および応用プログラム作成：ウォータールー大学（カナダ）
- 校正：OUP

英国 IBMは計算機と１０数台の端末を無償で提供した。さらに，同

社のREXというエディタを加造してこのプロジェクト用のエディタLEXを作るため，プログラマー人を半年間無償でOUPに派遣した。

米国のICCは最大120人程度のタイピストを使って，350,000,000文字におよぶNew OED全16巻のテキストを18ケ月で入力しようとしている。この費用は2.2百万ドルになる。来年6月にこの入力作業が終るが，その時点でMR化だけを行ったNew OEDがCD-ROMの形でICCから売り出される。

ウォータールー大学ではMR化されたテキストのDB化を行っている。これは辞書構造記述言語"Goedel"を用いる。これは辞書をパースするための有限状態オートマトンを生成するもので，今のところはさまざまなDBを作る学生の実験用に使われている。（図4－5，図4－6参照）現在DB化されているのは"J"の項と"M"の一部だけであるが，そのデータを使ったKWICシステムが同大学でのコンファレンスに展示されていた。

OUPにおける校正作業はすべてIBMの端末上のエディタで行われる。これはテキスト中の記述項目の違いやフォントの違いをカラーで区別して表示するなど，ユーザ・インタフェースが非常に良く，出版社の編集者でも容易に扱えるように作られている。

② その他の辞書について

OUPでは"Oxford Advanced Learner's Dictionary of Current English"（いわゆる"Hornby"）と"Concise Oxford English Dictionary"（いわゆるCOD）もMR化して発売している。

```
dictionary  :=  entry +;
entry       :=  hwgp etym ?  senbank * sen *;
senbank     :=  sen + quot +;
quot        :=  qdate qsource qtext;
...


EtymG  :=  {  etym   :=  lang ?  (text delim lang) * text;
              lang   :=  {'OF.', 'AF.', 'F.', 'Pg.', ...  };
              text   :=  char * - lang;
              delim  :=  {' ', '.', ...  }
           }


firstlang  :=  proc(x)
               E := etym  in x;
               lang  in  ( E  reparsed by  EtymG );
               string( " )
               end;


earliestquot  :=  proc(x)
                  q :=  every qdate  in x;
                  if  size q > 0
                      then min (year  mapped onto  q)
                      else 0
                      fi
                  end;


AngloNorman  :=  NewOED  where
                 firstlang(X) ∈ {'AF.', 'OF.'}
                     or firstlang(X) = 'F.' and earliestquot(X) < 1500;
```

図 4 − 5　Goedel 言語による記述例

```
        University of Waterloo
Pat 1.0 (Oct/86) Text searching system

# search for possible variant spellings of Shakespeare
 1: >> Shak and not ( "Shaks. " or Shakespeare )
                        1:  12 matchings
 2: >> pr 1
 198385, ..e and shoulder-shake+10 machinaments and fortifications. </R></qt...
2173511, ..busy with your Shake-H bags and Goose-greens, that I thought you ...
3405058, .. Crown; it can shake hands with *man-selling. </R></qtxt></quot><...
3186126, ..be inclined to shake off the French yoke and to co-operate with u...
3682189, ..me after it is shaken. </R></sen> <qbank><quot><qdat>1839 </qdat>...
 521858, .. use+10 shook (shaken), madding +15 etc.+16 . </R></qtxt></quot><...
 660575, ..noise, that it shaketh the rings in the doores of the inhabitants...
4657220, ..s+16 together, shaking the Bees out of one Hive on the Mantle whe...
4295917, ..ch fall off by shaking, we called Manna, (</R><I>i. Thuris</I><R>...
3699915, .. is a kinde of shaking off the yoake. </R></qtxt></quot><quot><qd...
 768283, ..0 with Ossian, Shakspeare,+10 &c.; or, as the maggot takes him, a...
3259246, ..ick+16 mangled Shakspere. +63 And he patched the mammock'd plays ...

 2: >> ( little or young ) fby ( child or boy or girl )
                        2:  13 matchings
 3: >> pr 2
4390564, ..t) to a little boy. </R></sen> <qbank><quot><qdat>1823 </qdat><sr...
4382103, ..><R>The little boy+10 was chid by his mother for not being more m...
2195978, .. little vulgar Boy+13 I said, +17 What make you here?+18 </R></qt...
3240699, ..for her little boys. </R></qtxt></quot><quot><qdat>1863 </qdat><s...
 240397, ..nes. +63 Young boys, called *machine-minders,+10 tend them. </R><...
3623572, ..e, Agaynst the boystrous winter stormes. </R></qtxt></quot></qban...
3331126, ..r a young male child. +63 Similarly </R><lem face=I>young man </l...
2066858, ..ttle </auth><I>Child of Stafferton </I><R>i. 12 </R></srce><qtxt>...
2190896, ..ttle </auth><I>Child of Stafferton </I><R>xii. 151 </R></srce><qt...
3464160, ..s amongst men, children, dogs, and horses+10 ; and by this means,...
3165809, .. made by young children; similar words for +17 mother+18 exist in...
3211526, .. used by young children was probably common, </R><I>mamma+11 </I>...
3208990, ..ively by young children, who are in many countries taught to use ...


 3: >> {PrintLength 145}
 3: >> lrep Steele
                The longest repetition is 157 chars long
                        3:  2 matchings
 4: >> pr 3
4074562, ..at><srce><auth>Steele </auth><I>Spect. </I><R>No. 153 +36 4 </R></src
e><qtxt><R>It is thus in the Life of a Man of Sense, a short Life is suffici...

3390211, ..at><srce><auth>Steele </auth><I>Spect. </I><R>No. 153 +36 4 </R></src
e><qtxt><R>It is thus in the life of a *Man of Sense, a short Life is suffic...


 3: >> pr ( lrep {180}) and not ( lrep {200} )
                The longest repetition is 200 chars long
                The longest repetition is 200 chars long
1709658, .. </lab></sen> <qbank><quot><qdat>1375 </qdat><srce><auth>Barbour </au
th><I>Bruce </I><pna>ii. </pna><R>486 </R></srce><qtxt><R>Bot always, as A m...

4254986, ..r. </R></sen> <qbank><quot><qdat>1375 </qdat><srce><auth>Barbour </au
th><I>Bruce </I><pna>ii. </pna><R>486 </R></srce><qtxt><R>Bot always, as a m...


 3: >> done
used 5.02 cpu seconds
```

図４－６　New OED　DBに対する検索例

4.2.2.8　EUROTRAプロジェクト（ルクセンブルグ）

(1) 名　　称

EUROTRA Project

(2) 住　　所

EUROTRA, Commission of the European Communities

Bat. Jean Monnet, L−2920, Luxembourg

(3) 訪問日時

11月17日（月）　9：00～10：00, 11：00～12：00

(4) 対 応 者

Sergei Perschke, Multilingual Action Programme

Peter Lau, Multilingual Action Programme

Loll N. Rolling, Multilingual Action Programme

(5) 組　　織

ECの一部局である。ルクセンブルグの町外れのヨーロッパ・センターにある"Jean Monnet"ビルディングにはECの技術部門がまとめられている。

(6) 概　　要

日本における電子化辞書の開発計画をEUROTRAの技術委員会（座長：A. Zampolli）で発表した。質問事項は言語理論に依存する文法事項の辞書内部での取扱い，概念辞書の構造・表現，既存の辞書を使わないことの理由，等であった。

4.2.2.9　EURODICAUTOMプロジェクト（ルクセンブルグ）

(1) 名　　称

EURODICAUTOM Project

(2) 住　　所

Terminology Bureau, Commission of the European Communities

Bat. Jean Monnet, L−2920, Luxembourg

(3) 訪問日時

11月17日（月）　10：00～11：00

(4) 対 応 者

Jacques Goetschalckx, Adviser and Head of the Terminology Bureau

Alain Reichling, Terminology and Information Applications

(5) 組　織

ＥＣのTerminology Bureauにおけるプロジェクトである。前項のEUROTRAと同じ建物の隣の棟にある。

(6) 概　要

人間の翻訳者のために作られた7ケ国語の専門語ＤＢである。専門分野はＥＣで扱う文書に含まれるあらゆる分野にわたる。1960年代から開始され，中断を含み現在にいたっている。

ＥＣには翻訳者を抱える40程のセクションがあるが，それぞれのセクションに1台ずつの端末が置かれている。特に複合語の検索が問題となるので，機械に素人のユーザでも容易に検索ができるようなインタフェースを提供している。（図4－7参照）

```
%C E2E2 PLEASE LOGON AT V219H00
##########################################################################
%  E2E3 LOGON ACCEPTED FROM LINE V219H00 /AM29013 AT 1048 ON 86-11-17, TSM
ASSIGNED
%(MSG) %  %  E229 LU326 LOGGED ON LINE V219H00 /AM29013
%***   WELCOME TO ECHO   ***
%MESS  NEW DB VERSION : 28-10-86          COPYRIGHT CEC :104839 : ***1097*F
 PRESS L FOR TERMINOLOGY OR X FOR ABBREVIATION
*l
 TYPE CODE OF SOURCE LANGUAGE
 DE GERMAN       DA DANISH       EN ENGLISH      FR FRENCH
 IT ITALIAN      NL DUTCH        PT PORTUGUESE   ES SPANISH
*en
 TYPE CODE(S) OF TARGET LANGUAGE(S) WITH SINGLE SPACE BETWEEN
 (FOR EXAMPLE: DE NL) OR A FOR ANY LANGUAGES
*fr
 SOURCE LANGUAGE     :EN
 TARGET LANGUAGE(S)  :FR
 SUBJECT CODE        :
 PRESS Q OR ANOTHER COMMAND
*q mainframe computer
 SEARCH IN PROGRESS
                                          DOC =    1  PAGE =   1
 BE= BTL    TY= MIE82  NI= 0000426  DATE = 860429  CF= 2
 EN VE  mainframe
    DF  The main portion of the computer,including the CPU,memory,and
        peripheral interfaces,Also:A large computer.
    DF  Partie principale de l'ordinateur comprenant l'unite de
        traitement,la memoire et les interfaces aux peripheriques.
        Egalement:gros ordinateur.
         PRESS C TO CONTINUE OR GIVE ANOTHER COMMAND
*r
                                          DOC =    1  PAGE =   1
 BE= BTL    TY= MIE82  NI= 0000426  DATE = 860429  CF= 2
 EN VE  mainframe
    DF  The main portion of the computer,including the CPU,memory,and
        peripheral interfaces,Also:A large computer.
 FR VE  unite centrale
    DF  Partie principale de l'ordinateur comprenant l'unite de
        traitement,la memoire et les interfaces aux peripheriques
        Egalement:gros ordinateur.
    RF  Sybex,1981,p.105
         PRESS C TO CONTINUE OR GIVE ANOTHER COMMAND
*c
                                          DOC =    2  PAGE =   1
 BE= BTL    TY= AFN78  NI= 0002626  DATE = 851024  CF= 3
 EN VE  1)computer;2)data processing machine;data processor
    DF  a data processor that can perform substantial computation
        without intervention by a human operator during a run
    DF  machine de traitement de l'information susceptible d'
        effectuer les calculs importants comprenant de nombreuses
        operations arithmetiques ou operations logiques,sans
        intervention d'un operateur humain au cours de l'execution
         PRESS C TO CONTINUE OR GIVE ANOTHER COMMAND
*t
 SEARCH IN PROGRESS
                                          DOC =    1  PAGE =   1
 BE= BTL    TY= P2E70  NI= 0001613  DATE = 860724  CF= 4
 EN VE  1,maximum computed flood
    DF  the largest momentary discharge believed possible from a
        consideration of meteorological condition and snow cover on
        the watershed,It presupposes simultaneous occurrence of all
        possible natural contingencies favourable to high floods
    DF  plus forte crue qui puisse survenir en un lieu donne,en
        supposant reunies simultanement toutes les causes
        productrices de la precipitation maximum et de l'ecoulement
         PRESS C TO CONTINUE OR GIVE ANOTHER COMMAND
*q main frame computer $
 SORRY, NO ANSWER.
```

図４－７　EURODICAUTOMの対話例（端末のハードコピー）

```
*x
 PRESS Q FOR ABBREVIATION
*q eec
   1 EN Y*  EEC          *Eurocontrol Experimental Centre
   2 EN Y*  EEC          *Engineered Electronics Company
   3 EN Y*  EEC          *Engineering Equipment and Credit Company Ltd.
   4 EN N*  EEC          *Evaporative Emission Control
   5 EN N*  EEC          *English Electric Company Ltd.
   6 EN N*  EEC          *enteropathogenic Escherichia coli
   7 EN Y*  E.E.C.       *European Economic Community
   8 EN N*  EEC          *English Electric Computers
  PRESS Q  (QUESTION),    =n (n :LINE NUMBER)
       L (TERMINOLOGY), Tn (n : TABLE NUMBER) OR E (END)
*=7                                          DOC =    7        BTBABR30000465
    SI  Communaute Economique Europeenne
    SI  Europaeische Wirtschaftsgemeinschaft
    SI  Comunita Economica Europea
    SI  European Economic Community
    SI  Europese Economische Gemeenschap
    SI  Det europaeiske oekonomiske Faelleskab
    SI  Comunidad Economica Europea
 PT SI  Comunidade Economica Europeia
  PRESS Q  (QUESTION),    =n (n :LINE NUMBER)
       L (TERMINOLOGY), Tn (n : TABLE NUMBER) OR E (END)
*e
```

図４－７（つづき）　EURODICAUTOMの対話例（端末のハードコピー）

4.2.2.10　ボン大学コミュニケーション・音声研究所（ＩＫＰ）（西独）

(1) 名　　称

Institute for Communication Research and Phonetics - University of Bonn

(2) 住　　所

Poppelsdorfer Allee 47, D−5300 Bonn, Federal Republic of Germany

(3) 訪問日時

１１月１８日（火）　９：３０～１２：００

(4) 対 応 者

Winfried Lenders, Professor

Jan Brustkern

(5) 組　　織

Lenders 教授を中心として数人のスタッフで小規模な開発を進めている。

(6) 概　　要

ＩＢＭ／ＰＣ上で，９ケ国語間のトランスファ辞書を編集するシステムＭＵＬＩや，各種テキストに対するＫＷＩＣシステムＢＹＵが動いている（図４－８参照）。後者は売り物である。また大型機上，既存の９種のＭＲＤ（３０万語）を同時に編集できるシステムもあるが，統合化は行われていない。

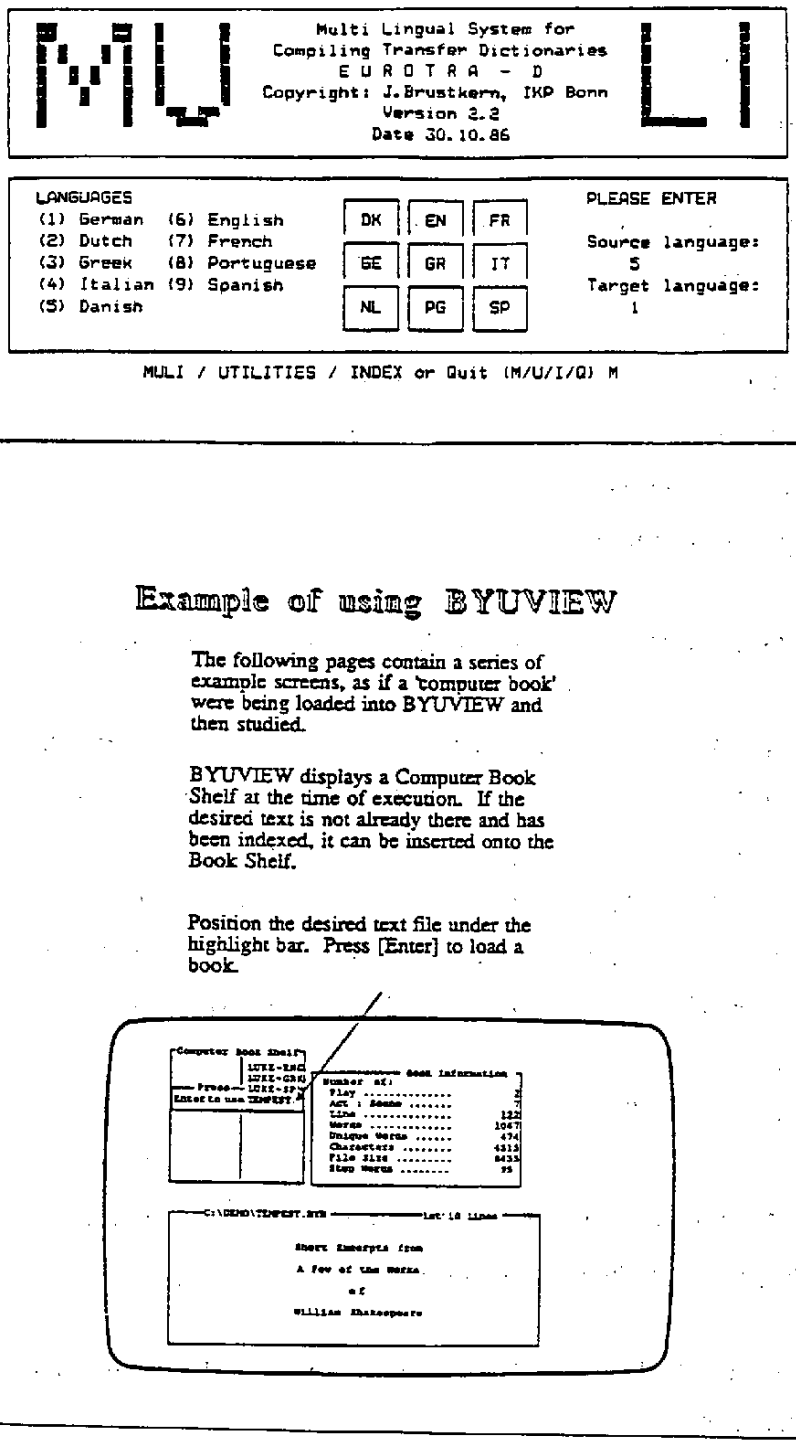


図４－８　ＭＵＬＩとＢＹＵの画面例（端末のハードコピーの使用説明書）

4.2.2.11　ジーメンス（西独）

(1) 名　　称

Siemens AG

(2) 住　　所

Otto-Hahn-Ring 6, D−8000 Munich 83, Federal Republic of Germany

(3) 訪問日時

11月19日（水）　13：00～15：00

(4) 対 応 者

Manfred Gehrke

Werner Remmele

(5) 組　　織

自然言語処理研究を行っているジーメンスの一部門である。

(6) 概　　要

形態素をベースとした辞書を開発中である。ドイツ語では複合語が非常に豊富であり，かつ自由に作れるので，このような辞書が必要になる。辞書の作成および更新を容易にするために，語の記述（1万程度）および語構成の規則（少数）はLFGによって記述する。計算機上にはATNのプログラミング環境が既にあるので，LFGの規則はそれに変換して用いることになる。このシステムはINTERLISP−Dで書かれ，現在実験中である。

4.2.2.12　INFOTERM（オーストリア）

(1) 名　　称

International Information Centre for Terminology(Infoterm)

Austrian Standards Institute

(2) 住　　所

A−1021 Vienna, Heinestrasse 38, POB. 130, Austria

(3) 訪問日時

11月20日（木）　9：30～13：00

(4) 対 応 者

Christian Galinski, Director of Infoterm, Secretary of ISO/TC 37
Helmut Felber, Ex-Director of Infoterm, UNESCO and Infoterm Consultant

(5) 組　織

オーストリア規格協会の一部門である。

(6) 概　要

西ドイツのTrier大学と共同で"Association for Terminology and Knowledge Transfer"という組織を作っている。Terminologyの研究は新組織に移行する予定。1987年9月にTrierで国際会議を開く。("International Congress on Terminology and Knowledge Engineering" 9月29日～10月1日)

4.2.2.13　ピサ大学および計算言語学研究所(イタリア)

(1) 名　称

Department of Linguistics, University of Pisa
Institute of Computational Linguistics, CNR

(2) 住　所

Via della Faggiola 32, I-56100 Pisa, Italy

(3) 訪問日時

11月21日(金)　14:00～21:30

(4) 対応者

Antonio Zampolli, Professor and Head of ILC
Nicoletta Calzolari
Irina Prodanof
Amedeo Cappelli
Lorenzo Moretti
Fabrizio Sebastiani
Carlota Nicolas

(5) 組　織

ピサ大学の言語学科と，イタリア政府のCNR(National Research Council)の一研究機関であるILC(Institute for Computational

Linguistics）とが一体化している。A. Zampolliは両者の長である。

CNRは(1)大学に補助金を出す，(2)国立研究所（ＩＬＣを含む）を運営する，(3)ＮＳＰ（National Strategic Project）を運営する，という３つの仕事をしている。ＮＳＰは軍と関係があるわけではなく，自然言語処理（代表者：A. Zampolli）や地熱エネルギー利用，など２０位のテーマを取り上げて補助金を出すプロジェクトである。

(6) 概　　要

① ＩＬＣで研究されているテーマ

ここでは自然言語処理の諸テーマを広く研究している。

1) 人文科学のためのテキスト処理

人文科学の分野におけるテキストの機械処理（頻度調査，KWIC，語彙調査，統計処理，…）を目的とする。ＩＬＣはそのためのソフトを持っており，イタリア国内の１００を越える研究室からの依頼に応えている。その代償としてＭＲ化されたテキストのコピーがＩＬＣに残るので，現在ではホメロス，アリストテレスからダンテ，ジョイスに至るまで10,000種にのぼるテキストが蓄積されている。

また，欧米の研究機関とも協力体制を取っている。相手先にはオックスフォード大学，フランス古文書館，ボン大学，プリンストン大学などがある。

2) 言語学研究用対話型ワークステーションの開発

テキストＤＢ，単言語ＬＤＢ，２言語ＬＤＢ，パーザ，知識表現，等をワークステーション上に統合化しようとしている。イタリアの辞書学者たちによってイタリア語版ＯＥＤを作ろうというプロジェクトがあり，それにも用いる予定である。

3) 古 典 語

ラテン語の形態素のためのＭＲＤを開発した。

4) スペイン語のmorpho-syntactical analyser

曖昧性を除くためにエキスパート・ルールを導入した。

5) ＬＤＢの開発（計画中）

イタリア語のＬＤＢとテキストＤＢとの結合，多言語のＬＤＢの開

発を行おうとしている。これは計算機科学者，言語学者，辞書学者のみならず，人文科学者やその他の一般ユーザまでもが利用できるようにする。そのためには各分野に対応するインタフェースを充実することが重要である。

また，MAGMA－Lispの上に知識表現言語KL－MAGMAを作り，それによる知識表現の研究も行っている。これも将来LDBの記述に利用される予定である。

② ヨーロッパ諸国にわたる活動

12ケ国で構成されるECに対し，21ケ国で構成されるCE（Council of Europe）という組織がある。後者の加盟国のMinistry for Researchの決議により"Scientific Network"を組織することになった。その中のcomputational lexicology and lexicographyの分野においては次のことを協議している。

- LDBの標準化
- テキストDBの標準化
- 著作権問題

この中にはOEDプロジェクトのためのWGもある。メンバーは，Zampolli（座長），Weiner，Zimmerman，等である。また，来年1月末にLDB中の文法情報の表現についてのワークショップを開く予定である。諸言語理論の創始者ないしはそれに近い研究者を集め，英語の最頻動詞200語に対する記述を行って，共通の要素を取り出す，という内容を計画している。

また，これとは別に，ヨーロッパ各国のNational Research Councilが集まってできているESF（European Science Foundation）という組織がある。ここには社会科学，自然科学，人文科学の3つの委員会がある。ESFがスポンサーとなって，1988年にピザ大学でcomputational lexicology and lexicographyのインタナショナル・サマースクールを開く予定である。

5．機械翻訳システムを例とした辞書及びその更新のあり方

5. 機械翻訳システムを例とした辞書及びその更新のあり方

本報告書の3章および4章で述べてきたような国内外の辞書に関する調査結果を踏まえ，ここでは，機械翻訳における辞書のあり方を述べ，さらに辞書の更新について問題点とあるべき姿を示すことにする。

5.1 基本的考え方

5.1.1 機械翻訳システムと辞書

機械翻訳システムは，さまざまな情報を内蔵し，それぞれに応じた処理を行なっている。現在は，機械翻訳システムは大きく三つの部分に分けて考えられることが多い。個々の語についての情報を記録した「辞書」，多くの語に共通する規則を集めた「文法」，それに，それらを実際に利用して翻訳を行なう「処理プログラム」である。

これらの三者の中では，辞書のもつ重みがもっとも大きいものである。実際に機械翻訳システムを作成する段になってもっとも手間がかかるものであり，また，翻訳に際してその質を左右するものが辞書だからである。某社の試算では，機械翻訳システムの作成コストの9割を辞書が占めるという話である。

機械翻訳システムにより，辞書の記述には相当大きな差があるが，辞書には一般に次のような項目が記述され，利用されている。

- 原言語の語形・活用・派生など→語の特定
- 原言語の品詞・文型→語の統語的関係の解析
- 原言語の意味記述→意味的解析
- 目標言語の意味記述→原言語からの意味の受渡し
- 目標言語の品詞・文型→統語的な文の生成
- 目標言語の語形・活用・派生など→形態素・単語の生成

機械翻訳システムがどのようにこれらの情報を利用して翻訳を進めていくのか，その概略を述べてみよう。とりあえず，英日の例を想定しよう。まず，入力テキストからピリオド（＋空白）までを1文として切りだす。仮に，"I ate an apple." だったとする。次にそれぞれの単語を

辞書で調べる。ここで，"ate"が"eat"という動詞の過去形であるという情報は辞書の語形・活用・派生情報によってわかることである。次に，それぞれの文法関係を調べ；"I"が主語，"an apple"が目的語であることがわかる。これが語の統語的解析であり，このためには，辞書中に"I"や"an apple"が名詞（厳密には代名詞と名詞句）であること，"eat"は主語と直接目的語をとる動詞であることなどを書いておく必要がある。次に必要なのが意味解析で，"eat"が「食べる」意味である場合には（これがもっとも普通の意味であり，「吸う」とか「する」などの例は少ないとしてのことだ），その主語には「人間」がくるのが普通で，直接目的語には「食べ物」がくるのが普通であること，また，"I"は「人間」であり，"an apple"は「食べ物」であることなどをチェックすることになる。その結果，"an apple"は「大都会」とか「乳房」ではなく，「リンゴ」であることがわかる。意味解析は訳語の決定にこのように大きな役割を果たしている。次に英語の語順を日本語の語順に直す段階になるが，これは「文法」のルール（トランスファールール）によるので，ここでは関係ない。次に格助詞の挿入の段階で，これまた辞書の「食べる」のとる格助詞パターンによって，「＜主語＞が＜直接目的語＞を」というパターンが決定される。最後に，「食べる」を過去の形にするのにそれが下一段活用の動詞であることを辞書で調べた上で「食べた」の形にする。こうしたすべての操作をへた上で，「私がリンゴを食べた。」という文が出力されるのである。なお訳として「私は」と「私が」のどちらがいいかはもっと他の情報を使わないと一概には決定できない。

こうしてみてみると，翻訳の過程のうち，解析と生成は辞書の情報に大きく頼っていることが実感されよう。しかも，人間用の辞書ならば，ごく普通の常識として記述する必要のないことまでも，機械翻訳用の辞書には書いておく必要があるわけである。たとえば「食べる」の主語は「人間」であるとか，「リンゴ」は「食べ物」であるというようなことである。したがって，機械翻訳用の辞書は，形式や構成だけでなく，記述内容も人間用辞書とは別のものであると考えなければならない。人間

用の辞書ですら，現在はまだ形式化が充分でなく，読み手の常識に頼る部分が多い。ましてや，機械翻訳用の辞書が充分であるはずはない。

5.1.2　機械翻訳システムのユーザ

どのような機械翻訳が望ましいかについてはいろいろな考え方があると思うが，スイッチ一つで自動的にすべての文章が翻訳できる機械翻訳システムが望ましいとしても，これは遠い未来の話であり，現実にはあり得ない。

では，現在のレベルではどうなのだろうか。機械翻訳システムそのものがまだ半製品の状態であるといってよかろう。不完全といってもいいかもしれない。

そのような機械翻訳システムの現状を考慮するならば，機械翻訳システムの利用には辞書の更新という作業が必然的に伴うものであることは容易に理解されよう。機械翻訳システムの構成要素の中では，辞書の部分がもっとも独立性が高く，機械翻訳システムと分けて考えられる要素があるということで，文法や処理プログラムをいじるくらいならまず辞書をいじろうということになるのである。

機械翻訳では，ユーザが辞書を「チューニング」しているのだと考え，そうすることによって訳文の品質を著しく改善することができるのだとする積極的な考え方をとりたい。「更新しなければならない」のではなく「更新できる」と考えるのである。ユーザが手を加えられることが機械翻訳の特徴であると考えてもよい。つまりはシステムにflexibilityがあるというわけである。このようにして，機械翻訳システムの透明度を上げ，多様な要求に応ずることができる道が開けてくる。

また，辞書の更新機能をうまく利用して，用途にあった訳文を得ることが可能になる。この効果がはっきり現れる場合というと，たとえば，ある翻訳グループが仕事をはじめるときに，翻訳に先立ち訳語を一斉に作っておくような場合が考えられる。別の翻訳の仕事のときには，また別の辞書を用意するというわけである。

不完全なものを使うためには，ユーザが自由にいろいろいじりながら，あるいは試しながら機械翻訳システムを「使いこなす」ことが望ま

れる。「更新」にせよ「チューニング」にせよ，ユーザが辞書に触るとなると，機械翻訳のしくみとその中での辞書の作用や因果関係といったものをユーザが理解する必要がある。つまり，ユーザが翻訳システムの中を見られるようになっていることが必要である。もちろん，実際の翻訳プロセスをすべてそのまま見るというわけではなく，仮想の（ユーザから見た）フレームを通すのでもよい。

ここでいう「辞書の更新」には，原言語・目標言語間の単語の組を追加することだけでなく，それに付属する「翻訳用の各種の属性情報の定義の追加・修正」や，メーカ提供辞書の修正までが含まれるものである。

もっとも，このようなユーザによる辞書の更新の場合，メーカの作成したものの有効性については考慮しておくべき問題がある。たとえば，品詞分類・意味分類などをユーザが自由に設定できたり，あるいは少なくともメーカ提供の初期状態と一部入れ替えるようなことができる方がいいのか，できなくてもやむを得ないのか。これは機械翻訳システム自体のあり方にもかかわる大きな問題になるので，ここでは特に扱わないが，このような問題があるということは意識しておいた方がいいだろう。

いずれにせよユーザが辞書に追加しながら使うなら，はじめから大きな網羅的な辞書を作っておけばいいではないかという考えがあり得る。しかし，大きな網羅的な辞書は，現状の技術レベルでは，必ずしも良い結果を生まない。多義語の発生など，問題点を引き起こす方が多い。将来は，解析技術の進歩によって多義語の発生が小さくなるかもしれない。そうなってはじめて大規模な辞書というものが考えられるだろう。現状では，したがって，辞書に適切な「規模」というものがあると考えられ，また，辞書の更新作業が必要になるわけである。なぜ辞書の更新が必要になるかについては，「5.4.1　更新の必要性」でさらにくわしく述べることにする。

5.1.3　辞書の更新と作成をめぐる諸問題

さて，こうして辞書の更新が必要になれば，機械翻訳には，メーカが提供する翻訳用の辞書だけでなく，個別に利用するユーザ辞書が必須のものになってくる。すなわち，機械翻訳ではレベルの違う辞書を複数個

使用する必要があるのである。機械翻訳で使用する辞書にはどのような種類があるかについては，「 5.2　辞書の分類・構成 」でくわしく述べる。

このように大まかにメーカ提供辞書とユーザ辞書とに二分して考えると，そのそれぞれに対し，「 作成 」と「 更新 」の二つの問題が考えられる。

作成はゼロに対する追加・更新であると考えられる。したがって，作成と更新は同一の基準で考えられる面を持つ。ただし，同じく「 辞書の作成 」といっても，メーカによって作成・提供される基本辞書と，ユーザが作成するユーザ辞書とでは違いた問題点がある。ユーザ辞書の場合は，規模・手順・内容のいずれの面から見ても，確かに「 作成 」と「 更新 」とであまり大きな違いはない。しかし，メーカが提供する基本辞書の場合は，「 作成 」にあたってはユーザ辞書よりもはるかに規模が大きいものになり，また，その「 更新 」は，全ユーザに影響を与えることを考えると，なるべくならやらないですませたい性格のものであり，実際に慎重にやらざるを得ないものである。

現実的には，機械翻訳用の辞書を作成するのは機械翻訳システムのメーカであるから，辞書の作成にはメーカ側の視点が必要になってくる。また，更新というのはたいていはユーザの行なうことであり，ユーザの視点で議論する必要がある。そう考えると，作成と更新とは，メーカとユーザという「 視点 」の違いがある（ないしそのような違いが反映するものである）ということにもなるが，ここでは一応，前述の立場にしたがって，作成と更新の問題と，メーカ・ユーザの問題は分けて考えることにする。

辞書にはさまざまなものが考えられるので，その作成に当っては，それぞれの辞書によって考える事情が異なってくる。その詳細については「 5.3　辞書の作成 」で述べる。

辞書の更新は必要に応じてユーザが行うことが多いから，ユーザの観点から見て使いやすくなければならない。その際，どのようなユーザ像を描くかということも重要である。初心者にも熟練者にもある程度使い

やすいシステムを目指すのか，どちらか一方を対象と考えるのか。ユーザは，会社のような「グループ」なのか，それとも個人が中心なのか。

そのほかにもいろいろと考えておくべき問題は多い。ユーザによる辞書の更新はどうしたらいいかということであるが，詳細は「5.4.2　ユーザによる更新方法」と「5.4.3　更新サポート機能」で述べることにする。

更新が必要である以上，機械翻訳システムにはユーザがいじりやすいような辞書が備わっていることが望ましいことも明らかである。

また，メーカが提供する基本辞書の更新については，メーカ自身によってなされるものであるから，通常のユーザによる更新とは性格がやや異なるので，「5.4.4　基本辞書の更新」で詳しく論ずることにする。

5.2　辞書の分類・構成

5.2.1　機械翻訳システムにおける言語データ

機械翻訳システムの内部から見たとき，言語データはいろいろの形で蓄えられ，利用される。それは，利用者による言語データの管理や操作のフェーズでの資源構成とは必ずしも一致していない。ここでは，利用の側面からみた辞書の構成を述べる前に，機械翻訳システムによる扱い方からみたときの言語情報の構成について述べておく。

(1)　ロジック内部の言語データ

それぞれのシステムの基本アルゴリズムやデータ表現そのものは，そこで言語情報を扱う上での基本的な立場を決定してしまう。その意味では翻訳システムの「言語モデル」を利用者の観点に引き寄せて吟味することが必要であるが，現実にはその為の情報が十分に公開されているとは云えない。混乱を避けるためにも，ここで云う「言語モデル」が，実用システムにおいては決して不変性を持ったものではなく，言語理論等で云う言語モデルとはかなりの距離があることにも留意しておく必要がある。

こうしたレベルの情報は殆ど変更不可能であると云ってよいであろう。そこで，利用者に対して，ノウハウとか理論とか云うよりは立場としての「言語モデル」をいかにして保証していくかが機械翻訳システムの供給者

側の課題になっている。

機械翻訳システムの中では言語情報の一部が，多くの場合やむをえず，プログラム・ロジックの中に取り込まれてしまっている場合もある。例えば，「否定」の処理がロジックにより行われていると，「否定表現」の言い回しをコントロールしたいというときにも，プログラムそのものを変更することが必要になる。こうしたものは，「言語モデル」の設定を精密化していくことにより，ルールや更には辞書で扱う様にすることもある程度可能である。しかし，これを全面的に実現しても，「言語モデル」がより煩雑になってみたり，ルールや辞書に異質な記述が強いられるといったケースも少なくない。

(2) ルールの形での言語情報

多くの自然言語処理システムでは，プログラムを出来る限り言語情報から切り離す，あるいは出来うれば言語モデルからも独立でありたいとする立場を取る。そこでは，いろいろのレベルで抽象化された言語モデルとしてルールが設定され，一定のルール記述言語によってコーディングされる。そして，言語現象に対する深いレベルでの変更の多くはルールの書換えによって行われる。

市販のシステムでは，非常に単純なルールを除けば，システム利用者がルールにアクセスするのは不可能になっている。しかし，利用者にとっても言語処理に介入したい要因はいろいろと考えられる。例えば，入力テキストの書式，各文書に特有な形態素解析ルール，特殊な文体・文法や意味の扱いなどある程度習熟した利用者にとってはもどかしくなる事も少なくない。そういったニーズへの対応も必要になってくるであろう。実際，中核的な「言語モデル」からは外れるが，翻訳用辞書とは別に，そこで認識する略語の一覧を用意したり，定型的な表現パターンを外部ファイルとして与えたりといった簡単な「ルール」を利用者がコントロールすることがある。

(3) システム辞書

機械翻訳システム用の辞書は，語彙をベースとして，システムのいろいろな動作を規定する「規則素性」の集まりであると云うことが出来る。辞

書にも，ある程度の柔軟なルール記述を許している場合が多いが，データ量の問題からも，ユーザとのインターフェイスの取りにくさの問題からも，これを最小にする工夫が為される。機械翻訳システム辞書は，内部形式においても，ファイル構成上も，アルゴリズムに好都合なように編成されて用いられるのが普通である。

(4) 辞書インタフェース

機械翻訳システム用辞書をそれ自身の内容・ロジックに即してユーザが操作することが出来る事は望ましいが，現実的には，非常に高度の知識を要求されるとか，機械翻訳システムの限界として，必ずしも理論的にすっきりした形で辞書が構築されていないとかいった点から，ユーザから見た辞書の見え方としての辞書インタフェースが必要になってくる。言い替えると，内部的な翻訳アルゴリズムを問題にしない限り，翻訳用辞書とは，この辞書インタフェースのことを云っている。しかし，これを余りに安易にアルゴリズムから離してしまうと，機械翻訳システムの動作に対しユーザが関与できる余地が狭くなったり，ユーザの変更とシステムの動作とがかけ離れてしまう危険性を持つ。この点は，機械翻訳システムの設計者にとって，翻訳アルゴリズムの開発以上に困難な課題である。

5.2.2 管理上の辞書の構成

(1) 基本辞書

機械翻訳システム本体と共に供給される翻訳用辞書を機械翻訳システムの基本辞書と呼ぶことにする。これは，ユーザによりいろいろな形で変更が加えられて，翻訳処理に用いられる。但し，このユーザによる変更は，次に述べるユーザ辞書，グループ辞書という範疇で考えるのが妥当であろう。

基本辞書は，通常，システム供給者側で作成されるが，その際に用いられる「LDB」等の典拠辞書との関係についてはここでは述べない。

(2) ユーザ辞書

機械翻訳システムは，少なくとも現状の技術レベルから見る限り，一般の文書に対し完全な翻訳結果を与えるものではない。これを原文を読むためのきっかけや手だてとするというのも有効な活用法であるが，それも併

せて通常は，利用の側での条件に沿って様々に辞書をチューンアップしてゆく必要がある。その為に，基本辞書の変更データとしてのユーザ辞書が作られていく。

ユーザ辞書はその背景の辞書体系（基本辞書や上位のユーザ辞書としての「グループ辞書」，やはり上位に設定された専門語辞書）に対し，その利用者にとって不足の語や不都合な辞書内容を調整する。それは，そこでの業務においては一般的な用法であったり，特定の文書にだけ必要なものであったり，さらにはその利用者にとっての使いやすさや「趣味」の問題であったりと実に様々のパターンがありうる。そこで，基本辞書から末端のユーザ辞書に到る「縦の関係」（次の「グループ辞書」で述べる）の他に，選択対象としての同一レベルのユーザ辞書の間の流用，整合，併合，分割といった事象を含むシステム・利用者両方向からの管理の問題は重要なテーマである。また，環境としての上位辞書体系を切り替えて使う場合にも，上位辞書の更新に伴う問題と共に，このユーザ辞書管理の課題は複雑になる。

(3) グループ辞書

パーソナル・ユースといった，小規模な機械翻訳システム利用の場合は，機械翻訳システム基本辞書に対し単一レベルのユーザ辞書だけで運用していくことが出来るが，一般的には，システム管理者，（翻訳）グループ管理者といった形で，何レベルかの辞書管理体制を設ける必要がある。パーソナル・ユースを想定しても，機械翻訳システムの拡がりを考慮すると，「ユーザ会」，種々のメディア等を通した「共通のユーザ辞書」といったものが考えられる。そうしたグループ辞書は，基本辞書あるいは上位のグループ辞書に対し，より限定された目的に沿って，辞書を拡大し，整備する。そして，1階層末端に近いレベルのユーザに翻訳・辞書環境を提供する。この様な意識を持って管理される辞書システムをグループ辞書と呼び，最終利用者が各人の責任の中で運営していくプライベート辞書と区別する。

グループ辞書が提供する環境は，基本辞書への修正・追加結果にとどまらない。併せて，その「分野」に有効な専門語辞書や意味カテゴリー等の分類基準なども提供する必要がある。また，辞書データだけでなく，翻訳

システムにおいてルールとして扱われているものについても分野・特定目的への特化を行いたいという要求が増大していくであろう。

基本辞書から末端のユーザ辞書へ到る系列において，システムや辞書の改良にともなう，ユーザ側で運用中のシステム・辞書への影響の問題も重要である。また逆に，ユーザ側での改良を上位レベルで吸収することも必要であるが，そのための機構とそれにともなう管理上の混乱を避ける方途も重大な課題となっている。

(4) 専門語辞書

基本辞書をベースとして何階層かにわたり変更が加えられて末端ユーザに到ると云う一般語を中心とするパスの他に，専門用語に代表される，分野毎の辞書データがある。これは，主に個別の機械翻訳システムとは独立に作成され，必要に応じて最小限の変換が為されると考えるべきであろう。それらは，システム供給者に限らず，サードパーティーによって，あるいはユーザ側によって作成される。そして，メーカーからユーザまで，あらゆるレベルで機械翻訳システムに採用される。

専門語辞書は特定の機械翻訳システムアルゴリズムを想定して作成されるとは考えにくい。従って，その様な性格のデータに対応することを可能とする機構を，機械翻訳システム側で準備することになる。このとき，機械翻訳システムの為に，独自の情報を与える必要が出てくる。専門語としては相対的に少数であるが，用言系の語については，格支配の情報等はグループ辞書などに取り込むとしても，少なくとも活用型等の情報は必要であろう。ただし，これらについては，ある程度の標準化は可能であろう。一方，セマンティック・マーカ等については，一定の標準化があったとしても，元々分野によって全く異なった体系が必要なものであり，機械翻訳システムのメカニズムにこうした情報を外部から操作できる柔軟性が要求されてくるであろう。

専門語辞書はその規模から，個々のメーカーやユーザの手に余る部分も多い。そこで，公共の辞書作成の機構が重要な意味を持ってくる。しかし，それがどの様に大規模な辞書を作成したとしても，個々の業務で用いる場合には，決定的に不十分であったり，不適切であったりということは避け

られない。その意味では，個々人，個々の文書のレベルまでの様々の場で専門語辞書を作成する必要がある。それを保証するいろいろの仕組みを成立させなければならない。

(5) その他の辞書データ

基本辞書を作成し，更新するための典拠データとして，冊子体形式の辞書や用語集にかわって「ＬＤＢ」が利用できる状況になりつつある。一方，ユーザの側でも，翻訳用辞書とは別体系である独自の用語集・規約，更には関連するデータベースとの連携が見込まれる。

また例えば機械翻訳システムの結果から内容を理解したりリライトしたりするユーザは，辞書類から多岐の分野の専門知識までを利用しなければならない。これは，辞書などの作成システムにおいても特に問題となる。こうした面まで含めた機械翻訳システムにおける辞書の体系化の問題はこれからの課題である。

5.3 辞書の作成

機械翻訳システムではどのような辞書が必要であるかは前節までに述べてきたが，大きく，メーカで作成・提供するものとユーザが作成するものとに分けることができる。そして，具体的にどのような辞書を作成すべきかの論議は，使用する機械翻訳システム及びそのユーザの運用の仕方により異なる。前節までに述べてきた辞書の種類を，以下辞書作成の観点から概略述べてみる。

一般に，機械翻訳システムには必ずメーカがシステム本体に付属して基本辞書を作成・提供する。基本辞書の語彙が一般語に限定されている場合には，これとは別にメーカよりオプションとして，専門語辞書が作成・提供されることがある。

しかし，現状の機械翻訳システムはまだまだ発展途上であり，かつ，それに付属する辞書も実用上充分なものではない。従って，ユーザは翻訳の品質を少しでも実用に近付けるために，入力文の前編集や，翻訳文の後編集をしたり，あるいはメーカ提供辞書に何等かの拡張・修正をすることとなる。

ユーザが辞書を作成する方法としては，メーカ提供辞書に直接手を加える方法とユーザが別の辞書を作成する場合とに分けられる。前者は作業形態としては辞書の更新に相当するので，5.4で詳しく取り上げることとする。後者はそのシステムが基本辞書そのものをユーザが更新することを許してなくて，基本辞書を拡張・修正するために，ユーザ自身がユーザ辞書を別に作成する方式をとっているシステムにみられる例である。この場合には，基本辞書とユーザ辞書は階層化して同時に用いられる。

ユーザ辞書を複数個用いて多重階層化して利用できるシステムでは，辞書を駆使した様々な使い方が可能となろう。ユーザ辞書として作成される辞書は大きく次のように分けることができる。

まず，メーカから提供される基本辞書や専門語辞書の内容を拡張・修正するための辞書が考えられる。これは更に，システムの全体的な管理のための辞書，特定のプロジェクトや部門のための辞書，あるいは特定の翻訳対象テキストのための辞書等として作成されるであろう。ここではこれをグループ辞書と呼ぶことにする。また，システムの試験的な利用や，純粋に個人的な利用のために作成される辞書を正式な運用基準から除外する意味でプライベート辞書として分類することができる。

以上，機械翻訳システムで作成される辞書の種類を図5－1に示す。

ユーザ辞書（専門語辞書を含む）作成の方法は，ユーザ自身で作成する。メーカに作成を委ねる，あるいは辞書作成専門会社に委託する等に分けることができる。特に，辞書の構築は専門性を要し，かつ大変な労力，費用，

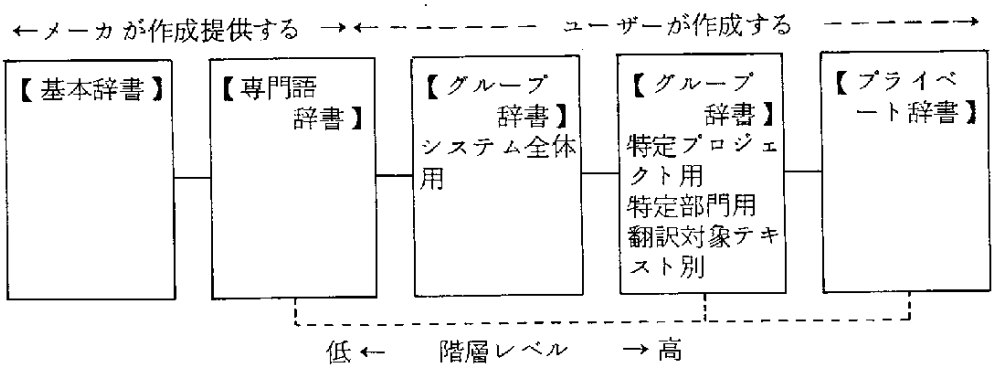


図5－1　機械翻訳システムで作成される辞書

期間がかかる作業であるので，今後は機械翻訳システムの普及と共に，"餅は餅屋"でサードパーティーとしての専門会社の支援を仰ぐケースが増えるものと思われる。

以下に，機械翻訳システムにおける辞書作成の現状と問題点について辞書の種類ごとに考察するが，ユーザの立場や視点からは，辞書作成と辞書更新とは同一の基準で考えられる面を持つので，多くは「5.4　辞書の更新」を考察してある。

5.3.1　基本辞書

基本辞書は機械翻訳システムのメーカがそのシステムの機能の一部として一般語を収録して，作成・提供するものである。従って，辞書の内容はそのシステムのアルゴリズムに依存した作成ロジックに基づき作成されており，見出し語とそれに関する形態・意味・構文・その他の言語属性情報および相手言語との単語（複合語）対情報などが含まれるが，システム間の共通性は少ないのが現状である。

辞書の作成は，言語学者，計算機科学者および入力作業者などの多くの人々の分担・協力の下に平行的に行なわれる。これらの作業者は，個人差による質の変動を極力避ける為に，詳細な作業マニュアルに基づいて辞書データの作成を行なう。又，前置詞や基本的な動詞など翻訳アルゴリズムと密接な関係のある重要な品詞の基本語については，少数の専門家により作成される。

辞書の作成は一般につぎの作業手順で行なわれる。

① 言語データの収集
② 言語データの分類・分析
③ 言語データの入力
④ 内部形式への変換
⑤ 校正・修正

ここで，言語データの収集は，既存の国語辞典や英和辞典などの冊子体辞書を参考にする場合が一般的である。

また，そのシステムが主な対象とする分野の文献調査により用語を収集し補完する場合も見受けられる。

収集した単語についてKWICなどにより，分類・分析し，各種言語属性マーカーや訳語が付与される。言語データの記述は，作業者が翻訳システムのアルゴリズムの詳細を知らなくても記述できるように表形式の標準様式が使用される。

所定の様式のワークシート等に記述された言語データは外注等を使い直接データエントリされるか，辞書エディタを用いて画面より対話形式で入力される。前者は一般的なデータエントリであるが，後者については言語学的知識を持つ人によって入力される場合がある。

入力された辞書データは，形式変換プログラムなどを用いて，そのシステムに都合のよい内部形式に変換され，辞書ファイルに収録される。

辞書ファイルの内容は，見易い形式に印刷又は画面表示され，翻訳システムのアルゴリズムとの整合性が検討され，校正される。

このように辞書の開発においては，各システムごとに多数の人々と長期間にわたり多大の費用を掛けて行なわれているが，正確な言語情報を大量に収集することを一企業に期待することは苛酷であるといえる。最近，国および国際レベルで機械翻訳システムにおいても応用できる網羅的で詳細な言語情報を収録する汎用の辞書づくりが議論され，かつ言語データベースとして実現されつつある。将来，これらの汎用辞書データベースが利用できればかなり効率的に高度な辞書開発が可能となろう。

その為には今後，国内外の協力により汎用の辞書のデータベース化が進展し，その利用技術やネットワーク化などの検討され，確立されることが大いに期待されるところである。

5.3.2　専門語辞書

専門語辞書は一般に機械翻訳システムのメーカ又はサードパーティーにより作成・提供されている。収録語数は２０～３０万語が，いくつかの大まかな分野に分類されている。一般に専門語は数100万語といわれており，まだまだ不足である。かつ，専門語の妥当な辞書がなく，企業ごとに必ずしも統一した用語が用いられていない等問題を複雑にしている。

現状の専門語辞書は，分野の区分が大雑把すぎるので，ユーザがそのまま利用することはない。一般には，メーカから提供されるユーティリ

ティ等で特定分野ごとにきめ細かに必要な部分を抽出して再構成し，それにユーザ自身で収集した専門語を加えて専門語辞書を作成する例が多くみられる。

今後は，学会や協会等が主体となり本格的な辞書作りが行なわれ，原子力技術，航空機工学等といった細分化された分野ごとの標準的なものを決定して行く方向が望まれる。

5.3.3　ユーザ辞書

機械翻訳システムの現状の技術では，付属している辞書は，一般に収録語数も5～6万語程度であり，内容的にも完全でない。翻訳は企業特有の用語を使用するケースが多く，かつより精度の高い訳文を得るために，ユーザ側で補完的に辞書を用意するのが一般的である。

ユーザ辞書は前述のように

- システム全体に対応したグループ辞書
- 特定のプロジェクト，部門，対象テキスト別に対応したグループ辞書
- プライベート辞書

などに分かれる。

いずれも，訳文の品質を改善するために，基本辞書や専門語辞書の内容を補完あるいは修正する（すなわちユーザによるメーカ提供辞書のチューニング）目的で作成され，メーカ提供辞書に被せて使用される場合が多い。従って；これらの辞書は，機械翻訳システムの導入後に，対象テキストの翻訳結果に基づいて経験的に作成されるものである。辞書は必要な数だけ作成して，翻訳実行時に対象テキストに合わせて随時切り替えて使い分けるやり方が翻訳システムをうまく利用する1つの方策といえる。

一般にユーザ辞書の作成は，各分野の専門家と翻訳システムの専門家とからなる組織により，次のような作業手順で体系的に行なわれているようである。（作業手順の事例を5.4で述べる）

＜訳文履歴情報に基づく原文調査＞→＜必要語の抽出＞→
＜訳語付け＞→＜委員会による承認＞→＜翻訳属性付け＞と辞書

への登録＞→＜テスト＞

ユーザ辞書の作成に当たっては，誰にも使いやすい，基本語辞書に対応する辞書エディタ，辞書ユーティリティと，必要語の収集ツールなど各種支援ツールが具備されている必要がある。

最後に，ユーザ辞書の運用に当たっては，規模拡大等による混乱を避ける為に，予め次の点を配慮しておく必要があろう。

- メーカ提供辞書のバージョンアップに追随できる基準で作成し，メーカもそれを保証する。
- メーカ提供辞書にフィードバックするためのユーザとメーカとの協力体制を敷いておく。

5.4 辞書の更新

本節では，機械翻訳システムの辞書更新の問題に焦点を当てる。まず，その必要性を論じ，ついで，ユーザの立場からの更新作業の特質とそのためのサポート機能について述べる。最後に，それとは逆に，機械翻訳システム開発者の立場から行われる辞書更新の問題について考察する。

5.4.1 辞書更新の必要性

現在，入手可能な機械翻訳システムは，完璧な翻訳を自動的に得たいという究極の願いからはいずれも不完全なものであり，今後，システムが完璧になるという見込みも近い将来においてはほとんど考えられない。したがって，機械翻訳システムの利用者はシステムが不完全なことを認識しながら，それでも翻訳業務が効率化できる問題を選んでシステムを使っていくという姿勢をとらざるをえず，また，開発者側も絶えず未完成な部分を改良しつつシステムを提供していかざるをえない。参考文献〔1〕

このような状況を認識すると，機械翻訳システムの性能の改善や，対象目的にあわせたチューニング作業を効率良く実施する方法・ツールを整備する必要があることがわかる。システムの構成要素中，これらの作業がもっとも容易に行えるのが翻訳辞書の部分であり，機械翻訳システムの実運用には，辞書の更新作業に重点をおくことが必要となる。

辞書の更新・チューニングを行うためには，原言語・目標言語間の単

語の組を追加修正すると同時に，機械翻訳システムの特性に応じた，翻訳用の文法情報・属性情報を付加・修正する必要がある。辞書を更新する目的は次の4点にある。

第一の目的はシステムの翻訳機能の誤り，もしくは，あらかじめ作成されている辞書の誤りを訂正することである。機械翻訳システムの辞書は，少ないものでも一万語，多いものでは十万語以上の見出しをもつ。これは，現在のエキスパートシステムの知識ベースの規模に比較すると百倍以上の大きさとなる。このようなエキスパートシステム程度の規模でも知識ベースの誤りをなくすことは難しく，そのための手法は確立されていない。さらに，一般にこれらの辞書の作成は人手によって行われる作業であり，作業中で誤りが皆無であるとはいえない状況である。これらの辞書の誤りはユーザ側で発見されることが多い。そのような場合には開発者に修正を依頼するのでは効率が悪く，ユーザ側で辞書を更新する必要が生ずる。

更新の第二の目的は，見出し語に対して対応する訳語の追加・修正・削除を行うことである。これは狭い意味での辞書の更新作業であり，ユーザが翻訳対象に合わせて機械翻訳システムをチューニングする際には必要となる。現在の機械翻訳システムは，見出し語に付加する文法情報・属性情報によって，複数の意味をもつ単語に対してある程度の訳しわけが可能となっているものが多い。このようなシステムに対しては，訳語と文法情報・属性情報とを対にして辞書に追加することによって翻訳の質を向上させることができる。一方，逆に，文法情報・属性情報が対象分野に対して不適切であった場合は，修正・削除の作業によって翻訳の質の向上させることが可能となる。特に複数の辞書を同時に参照できるような機械翻訳システムの場合は，前節で述べたプライベート辞書のレベルでの見出し語の削除機能が有用であることが多い。

更新の第三の目的は，専門用語に対する訳の充実である。これは，専門語辞書，あるいは，グループ辞書のレベルで整備・更新を行うべき作業である。特に，技術文書の翻訳を対象とする場合は，各分野の専門用語の翻訳さえ正確になされていればその分野の専門家が読むかぎりにお

いて，翻訳を誤解釈する可能性は著しく減少するし，また，機械翻訳システムの訳文に慣れていない場合でも，それを利用することに抵抗がなくなる傾向がある。また，専門用語は大部分名詞もしくは熟語であるので見出し語に複雑な文法情報・属性情報を追加する必要はあまりなく，人手をかけることさえ厭わなければ，それを更新する作業は不慣れなユーザでも比較的容易に作業を行うことができる。現在，機械翻訳システムの先進ユーザが実施している辞書の更新作業はほとんどこのレベルであるということができる。

第四の目的は，ユーザ側で必要な慣用句を登録することである。機械翻訳システムで採用している原理がどのようなものであるにせよ，よく使う慣用句に対しては適当な属性を付加して正しい訳語を与えておけば当然優れた品質の訳文が出力されし，また，辞書をひく段階で訳が確定するので翻訳速度も向上する。たとえば，"Time flies like an arrow" という文が慣用句としてよく使われるならば，これを名詞句と考えて，"光陰矢のごとし"という訳を与えておけば，"時バエは矢を好む"というような不適切な訳の生成を防ぐことができる。このような工夫は特に特殊な慣用句が多い，株式情報の翻訳などに有効である。

5.4.2　ユーザによる更新方法

ユーザによる辞書の更新は，組織的に実施しようとすると，時間・作業量の面で非常にコストがかかることが知られており，また，この作業を効率化するには，担当者・各分野の専門家・機械翻訳システム専門家からなる組織と，支援ツールの開発が重要とされている。これは，品質の良い翻訳文を作成するという本来の達成に非常な困難が伴うと同時に現在の機械翻訳システムを利用するためには，計算機利用技術・ソフトウエア技術的に工夫が必要なことになる。たとえば，機械翻訳システムを実運用するには，前節で述べたように，基本辞書，専門語辞書，グループ辞書，プライベート辞書を適切に組合せる必要があるが，組織的にそれらを利用しようとする各辞書を複数用意しておいて，対象分野の性質に応じて階層的に組合せなければならない（図5－2）。

組織的な辞書の更新はグループ辞書単位以上の規模で実施することが

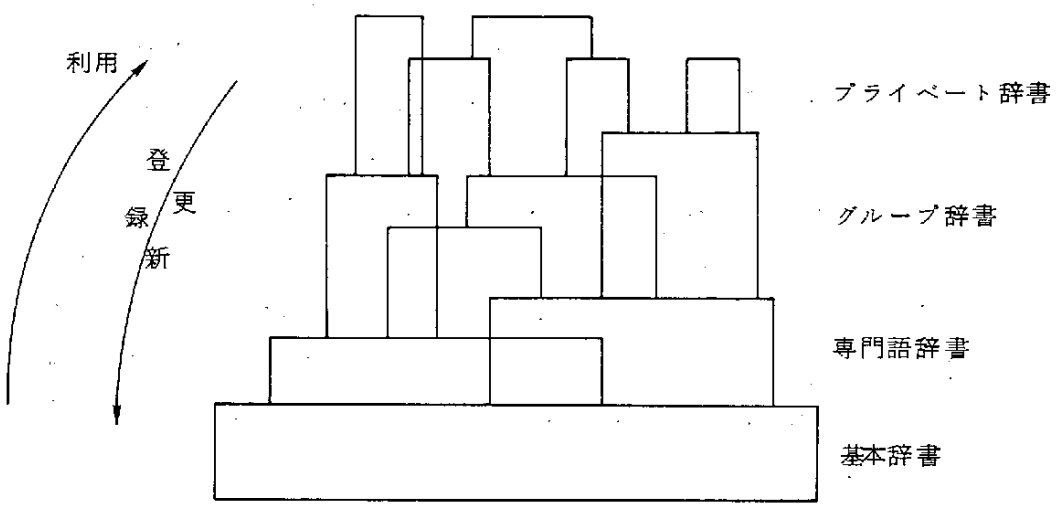


図５－２　ユーザにおける辞書体系

望ましい。そのための作業手順は，一般に，原文の調査，辞書更新に必要な語の抽出，訳語付け，専門家による承認，翻訳属性・文法情報の付加，テストというサイクルを何回か繰り返すことになる。ただし，この作業は，ユーザが導入した機械翻訳システムの機能に依存する部分も多く，現在のところ決定的な手順は確立されていない。そこで，以下では(財)電力中央研究所における作業手順参考文献〔2〕を紹介してユーザにおける辞書の更新方法に対する考察にかえることとする。

(財)電力中央研究所における機械翻訳システムの実用化の考え方は，前・後編集にはできるだけ手間をかけずに，電気事業に関連するさまざまな分野の大量の技術文書に対する下訳を迅速に生成するというものである。また，原文としてはすでに計算機上で蓄積されているもの，もしくは，マシンリーダブルなものを想定している。そのため，運用システムの整備が終了した時点においては，対話型で原文・翻訳文の編集を行うというシステムの利用よりも，むしろ，計算機で処理可能な形式で入手した原文をセンター側で一括して機械翻訳して提供する，もしくは，原文のデータベースを検索しながら必要な文書をエンドユーザが翻訳という利用が中心となる。

辞書の更新手順を図５－３に示す。原文は，上で述べたように計算機にデータベースとして蓄積されている。各レコードは，タイトル，アブストラクトといくつかのキーワードから構成される。そこで，まず，専門用語であるキーワードが正しく翻訳されるようにこの部分について辞書のチューニング作業を実施した。訳語の適・不適は各分野の専門家が判断した。こうして作成された辞書と基本辞書とを組合せてタイトル，アブストラクトについて翻訳実験を行い，辞書の精度を向上させる方針をとった。また，グループ辞書としては，電力分野，原子力分野など適当に９個に分類したものを最初に作成した。そして，これにはメーカ提供の専門語辞書も活用している。

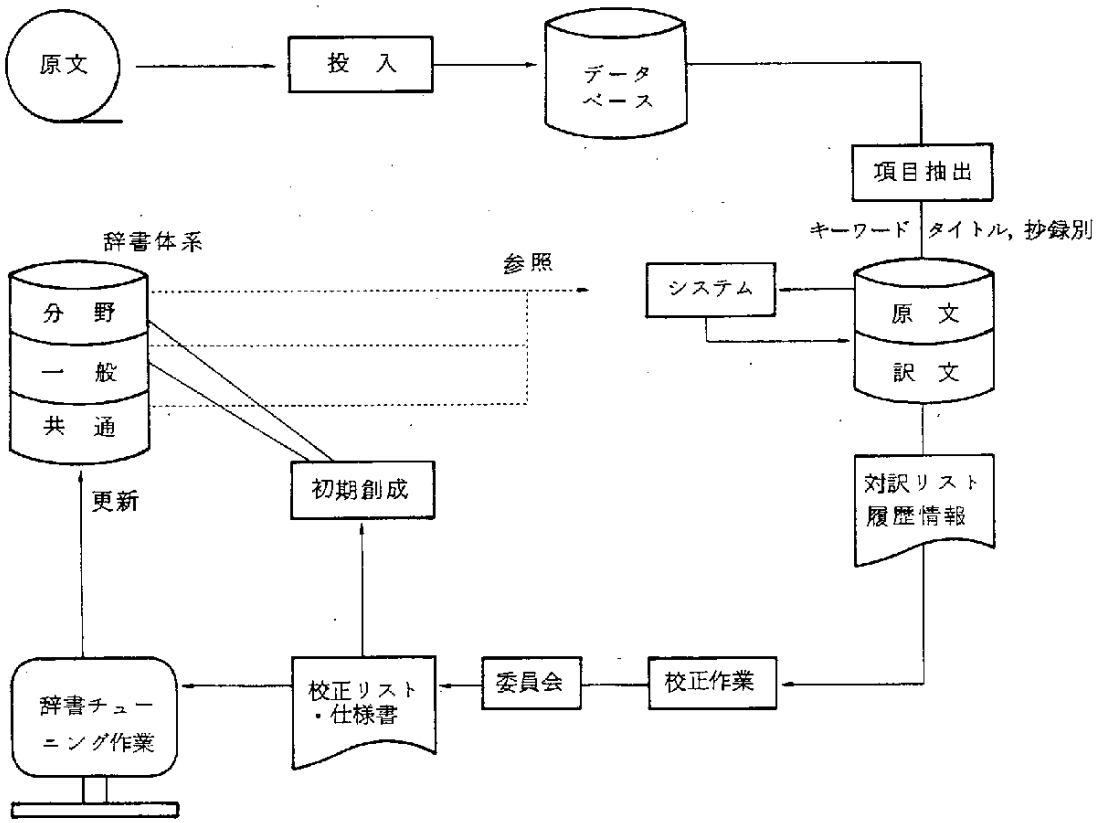


図５－３　(財)電力中央研究所における辞書更新作業例

5.4.3 辞書更新サポート機能

機械翻訳システムを利用する場合には，翻訳の質の向上とシステムの維持・改善の容易性とのトレード・オフの問題を考慮することが重要である。たとえば，慣用句を大量に登録しておくとその範囲においては訳文の質が向上するが，それからはずれた場合のシステムの柔軟性に欠けることとなる。一方，文法情報・属性情報を高度化すると扱える文章の範囲と訳文の品質は向上する。しかし，その分だけ，辞書・文法の修正・改善作業が困難になって，システムを維持しにくくなる。このトレード・オフを十分考慮しないと実用的な機械翻訳システムは実現できない。そのため，ユーザで辞書を更新する作業を効率化するためには，それをサポートする機能・ツールの充実が不可欠となる。その機能としては次のようなものが上げられる。

(1) バッチ型／対話型の辞書更新機能

組織的な辞書の更新を行う場合には，作業が大規模になるので，システムはバッチ型の辞書更新機能をもつことが望ましい。ただし，その後の更新やチューニングの作業には原文と翻訳文とを比較しながらのきめ細かい作業が必要になるので，以下に述べる原文・翻訳文の編集機能と組合せて利用できる対話型の辞書更新機能も必要である。

(2) 原文・訳文編集機能

原文と訳文とを同時に表示しながら両者の前編集・後編集が行える柔軟な専用エディタが必要である。このエディタは必要に応じて，翻訳辞書の内容アクセスでき，また，翻訳の途中結果の情報もユーザに分かりやすい形で表示できることが望ましい。

(3) 文法情報・属性情報付加の支援機能

原言語・目標言語間の単語の組を登録修正する作業は一般のユーザでも比較的容易に行えるが，翻訳の質を向上させるためには，さらに，辞書中の見出しに詳細な文法情報・属性情報を付加しなければならない。これは一般に複雑な作業であり，各機械翻訳システムの実現方法によって作業の内容も異なるものである。したがって，この作業を支援・容易化できるような，オンラインのヘルプ機能，デフォルト機能などが必要である。また，

この場合には，上と同様に，翻訳の途中結果の情報もユーザに分かりやすい形で表示できることが望ましい。

(4) 既存辞書の内容の参照・援用機能

辞書の更新を行う場合には，既存辞書の内容を参照してその一部を援用して新しい見出し語を作成するという作業がしばしば有用となる。たとえば，すでに辞書に存在する単語と同じような意味をもつ別の単語を新しく見出しとして登録する場合は，文法情報・属性情報にはあまり差がないのが普通であり，複雑な情報を再入力するよりも，参照・援用・修正の作業を中心とした方が効率が良い。そのための機能が必要である。

(5) 各種辞書の管理機能

辞書の更新作業はきわめて永続性の高い性質をもつ。そして，前節で述べたように，機械翻訳システムを実運用するには複数の辞書を同時に管理し，特定の部分の変更をすべてに適切に反映させる必要がある。これを支援する機能の実現が望まれる。

(6) 翻訳例文の管理機能

辞書の更新は，本来翻訳性能の向上を目指すものであるが，特定の文章を対象とした辞書の更新が他の文章に対しては性能を劣化させることがしばしば起こる。したがって，組織的に辞書の更新を実施したのちは，必ず，システムの性能劣化が生じていないかどうかをテストする必要がある。そのためには，標準的な翻訳例文を準備しておいて，更新前後の翻訳文の比較作業が実施されるのが普通である。これらの翻訳例文を効率良く管理し，その履歴を維持・保存する機能の充実が必要である。

(7) 原文・訳文に対する統計分析ツール

一括型の機械翻訳システム利用形態を前提とする場合は，入力文の特性に応じて，どのような品質の翻訳文が生成されるかを事前に予測できることが望ましい。しかも，予測に要するコストは，翻訳そのものにかかるコストに比較してなるべく安価におさえる必要がある。そのためには，入・出力文を人手で分析したり，システムの編集機能を利用したりすることは避けなければならない。それには，我々は，古くから行われている統計的な文の分析手法や人間向きに定義された読みやすさの尺度に基づいて入力

文の特性を評価し，それを提案された翻訳文の評価法と関連させることで，実用的かつ簡便な入力文の事前評価方式の開発をはかることは有用である。参考文献〔3〕

5.4.4　基本辞書の更新

5.4.1でも述べたように現在の機械翻訳システムは不完全なものであり，開発者側でも基本辞書の更新作業を絶えず続けてなくてはシステムの性能を維持していくことは難しい。更新に伴う問題点は前節までに論じたユーザ側の問題と共通するものが多いが，開発者に特有の課題としては次のようなものが考えられる。

(1)　文法情報との整合性の管理

基本辞書に登録された単語は機械翻訳の立場からは解釈の困難な基本的なものが多いため，その更新はシステムのもつ文法処理機能と深く関連する可能性が高い。したがって特に文法情報との整合性の管理には十分な注意を払う必要がある。

(2)　翻訳例文の管理

機械翻訳システム開発者側でも，前節で述べたような翻訳例文の管理を行わなければならないが，対象が基本辞書・専門語辞書であるために，いっそう，幅広い分野の微妙な例文を準備しておく必要があり，これらの管理には高度のデータベース機能が必要となる。

(3)　複数の利用者からの要請の調整

機械翻訳システムは専用システムでないかぎり，複数のユーザに提供され，その中で使われながら性能が改善されていく性質のソフトウェアである。ところが，ユーザの要求にはさまざまなものがあり，それらの間には互いに相反するものもあると考えられる。このような要請を基本辞書に反映させる場合にはそれらの間の調整・管理に注意を払わなければならない。

(4)　翻訳辞書のもととなった辞書との整合性の管理

開発者が基本辞書を作成する場合でも，まったくの最初から辞書を開発する例はまれで，ほとんどの場合そのもととなる（人間用の）辞書が存在している。ところが，この内容は時代の変遷とともに絶えず変化している。この傾向は，専門語辞書の更新に関連する作業においては特に著しい。そ

れにあわせて基本辞書・専門語辞書の内容を管理していくことは今後重要な課題となると考えられる。その作業を効率化するためには，標準的な言語情報データベースとしてLDB（Lexical Database)の開発が望まれている。

参考文献

〔1〕 情報処理：『特集：機械翻訳』 Vol.26, No.10, (Oct.1985)

〔2〕 寺野隆雄：『機械翻訳システムの評価とその利用方式』電力中央研究所報告・研究報告：584008, 昭和60年6月.

〔3〕 寺野隆雄（他）：機械翻訳文に対する事前評価方式，情報処理学会第34回全国大会7W－6, 1987年.

5.5 現状の問題点とあるべき姿

本章では，機械翻訳システムを例としてそこで使用されている辞書の問題を考えている。つまり，コンピュータ用の辞書ということであり，本節でもこれを中心に議論していく。しかし，機械可読用辞書を人間が使用するという問題も度外視することはできない。このことについては，本節の最後で述べる事にする。

5.5.1 辞書の管理

現在，機械翻訳システムのユーザの中で実際に業務の中でシステムを運用しているところが非常に少ないということが，本調査を通じて強く感じたことである。機械翻訳システムを導入しているユーザのほとんどが，いわゆる研究開発部門などであり，試用を目的としている。研究開発部門の特徴として，システムに積極的に介入しようとしている。つまり，機械翻訳システムそのものに精通している，あるいは精通しようとしている。従って，システムの使い勝手などにおいてもシステムに合わせようとする。反面，実運用に対する配慮に欠け，どのようにエンドユーザに提供して行くかという問題があまり議論されていない。エンドユーザを含めた運用体制が整っていないというのが現状である。

ここでは，実運用を行った場合の辞書の更新を中心とした運用体制に

ついて述べる。一般的にユーザが更新可能な辞書は大きく2種類に分かれる。一つは個人に属する辞書（プライベート辞書）で，その個人特有の言い回しやノウハウを収めたり，一過性の辞書として用いられる。プライベート辞書は各個人に管理が任されるべきであり，ここでは議論の対象としない。もう一つは，企業全体やグループなど複数の人間が使用する辞書（グループ辞書）であり，この辞書について議論する。

複数の人間が利用するグループ辞書については，その辞書の管理を行う専門の人間，または専門のグループ（以下，辞書管理者と呼ぶ）が必要であると思われる。辞書管理者は，

1) 辞書の更新，管理を行う。
2) 用語の選定，訳語の決定を行う。
3) 利用者の管理を行う。

などの任務を負う。

辞書の管理，訳語の決定（登録する語の決定を含む）などは高度な技術を必要とする。そのため，単なるエンドユーザの集まりやシステム管理者などの片手間でできる作業ではない。特に訳語の決定や用語の選定にはその分野の専門家の意見が重要であり，この種の人間を辞書管理者に加える必要がある。

さらに，グループ辞書は複数の人間が利用するものであるので，各利用者の合意を得る必要のある問題も生じるであろう。また，各個人の辞書に含まれているノウハウや用語などを吸い上げることもグループ辞書を充実させるためには欠くことができない。これらのことを実現するためにも辞書管理者が必要となる。

5.5.2 辞書の更新

機械翻訳システムではユーザ辞書はユーザで作成・更新することが前提となっているため，各システムとも辞書更新のためのユーティリティやエディタがメーカから提供されている。これらの辞書ユーティリティや辞書エディタの機能などに関してはどれもほぼ同じである。これらに関する問題点は，通常の計算機システムのインタフェースに関する問題点と同じと考えてよい。

特に問題点として挙げられることは，ユーザのレベルによって機能に

対する要望が異なることであろう。末端のユーザは辞書の詳細まで手を加えずに辞書の更新を簡単に行いたいと希望し，研究開発部門などではよりよい訳を得るためには多少煩雑でも辞書のいたるところを直したいと思っている。従って，ユーザのレベルに応じて機能が異なるような更新ツールが理想的であろう。管理の面からみても，一定水準の機能を確保するために，重要な語は特定の管理者だけしか操作できないような機能が必要である。

また，更新能力を向上させるために，何らかの学習機能といったものが必要であろう。これには，ポスト・エディットの結果が反映されるようなものがよいと考えられる。しかし，ポスト・エディットでは，文体を整えたり，微妙なニュアンスの違いを直すなどといった変更も行われるため，必ずしも辞書を更新するということにつながらない。そのため，ポスト・エディットと更新処理の接続には技術的に大きな問題を含んでいる。

次に，更新の頻度からみた問題点について考えてみる。

プライベート辞書は一過性の辞書としても使われ，全く個人の管理に任されるので問題はない。グループ辞書は，業務によって更新の頻度は異なり，場合によっては毎日更新されるであろう。しかし，グループ辞書は社会的には個人と見なせるため，そのグループ（企業）内で調整すれば良い。そのために，5.5.1で述べたような辞書管理者が必要となる。

問題となるのは専門語辞書である。専門語辞書は，現在はメーカから提供されているが，今後はサード・パーティーからも提供されていくであろう。いずれの場合にも，専門語辞書は学会や協会などが監修していくのがよいという意見がある。専門分野によっては新しい用語などの追加が頻繁に行われている分野もある。このような場合，メーカやサード・パーティーでの対応，さらには学会などの監修機関の対応などによって，対応が遅くなってしまう。しばらくの期間はユーザ辞書で対処できるが，専門語辞書の更新は必ず必要であり，そのための組織や更新方法に問題が残されている。

5.5.3 互換性

機械翻訳システムで使用される辞書はシステムに依存する部分が非常に多い。基本辞書はシステムの一部に位置づけられるものであり，ユーザでの更新もできないようになっているのでここでは議論しない。専門語辞書やユーザ辞書もシステムに依存しており，他のシステムと互換性がない。バージョン・アップに対してはメーカ側で互換性を保証しているが，機械翻訳システム自身が発展途上であるためその互換性に疑問も残る。また，同じメーカでも別システム（例，日英と英日）では，別の辞書を用意しなければならない。

辞書には，単なる単語の情報だけでなく，各種のノウハウが反映される。さらには，単語にまつわる各種の知識なども辞書に吸収され，辞書は，エキスパート・システムにおける知識ベースのような役割を果たすようになるであろう。従って，辞書に互換性がなく，ノウハウが蓄積されていかないということは，辞書そのものの進歩に取って大きな障害となり，コスト的にも無駄が生じる。ユーザの利益を守るために辞書の互換性は重要な問題となる。

また，専門語辞書の構築・更新は一ユーザだけの問題でなく，同じ専門分野を扱っている複数のユーザに関わる問題である。充実した専門語辞書を構築していくには1ユーザでは困難であり，複数のユーザが協力する必要があると考えるからである。更に，現在は，専門語辞書はメーカから提供されているが，今後は，学会や協会などの監修でサード・パーティーが提供する方向も考えられる。

このように，辞書の構築・更新には，メーカ，ユーザ，第三者機関が携わっているため，相互の協力体制が必要であると思われる。その際，議論のベースとなる辞書が必要になる。その際にも，辞書の互換性という問題を解決することが重要となってくる。互換性を保証する代わりに，共用の辞書を作り，その中からシステムにあった辞書を再構築していくという方法も考えられる。いずれにせよ，メーカ，ユーザ，第三者機関が辞書の構築・更新に関して協力できるような組織が辞書の発展には必要になるであろう。

5.5.4 基本辞書

機械翻訳システムにおける基本辞書は機能語などを含み，システムに大きく依存し，システムの機能，能力も左右するため，ユーザには解放されていない。また，文法も同様の理由からユーザには解放されていない。しかし，ユーザからは，基本語や文法も更新したいという要望がある。市販されているシステムでは文法の更新などはできないが，メーカとユーザの共同開発のシステムで文法が更新できる場合には，実際，その更新も行われている。

より良い訳文を得たい，自分のニーズにあった訳文を得たいなどというユーザの希望は絶えることがない。このようなユーザの多様なニーズに対応するためには，基本辞書や文法の更新も行えなければならない。このようなニーズに対しては，ユーザ辞書だけでは対処できないからである。

しかし，基本語や文法はシステムの機能，能力を左右するものであり，変更の仕方を誤るとシステムの機能を落とすことにもなる。また，基本辞書・文法はメーカのノウハウが含まれている部分である。従って，基本辞書や文法の更新はメーカとの連携が必要となる。

現実的には，ユーザの要望をメーカ側で吸収していくことになるだろう。しかし，現状では，システム自身が発展途上であることや，ユーザ側も運用体制が確立していないなどの理由により，組織的にメーカがユーザの意見をフィードバックするような体制は整っていない。今後は，システムが進歩するに従いユーザからメーカへのフィードバックの体制が整うことが期待される。

5.5.5 ま と め

以上論じてきたように，辞書の構築，更新，管理にはさまざまな問題が残されている。ここでは，機械翻訳システムを例として論じてきたが，ここで論じられたことは，他の全ての自然言語処理システムについても同様に考えられることである。

辞書の更新は，従来の計算機システムのデータの更新とは同列に扱えない。辞書の更新は，システムに与える影響が大きく，場合によっては

システムの機能や訳文の質などにも影響を与える。そのため，辞書の更新・管理には専門家グループが必要となってくる。この専門家グループも，辞書の種類により，一企業内のものや社会的なものなど様々なレベルのものが存在し，対応も異なってくる。

また，辞書の互換性の問題は，辞書そのものの発展や自然言語処理システムの発展にとって重要な問題である。辞書の互換性は，共用辞書の問題に置き換えることができるかもしれない。辞書の互換性，または共用辞書の問題は，メーカ，ユーザ，第三者機関の協力が必要となる。また，そのための組織的な問題も解決していかなければならない。

辞書は，その性質上，単なる単語の情報だけでなく，単語にまつわる各種の知識の集大成となるものである。さらに広い意味で辞書を考える場合には，計算機だけでなく人間も利用できるような情報を網羅することも大切な要因になってくるだろう。また，これらの情報はその検索，更新，管理などが簡単にできなければならない。このように考えていくと，辞書に含まれる情報をデータベース化するという方法が具体的には挙がってくる。

辞書の共通化，データベース化という条件を満たすものとして，次章で詳しく説明するLDB（Lexical Database）がある。LDBとは，あるひとつの言語のマスター辞書となるものであり，単語にまつわるあらゆる情報はデータベース化されている。自然言語処理応用システムの開発者は，LDBをもとに，それぞれの応用システムのための辞書を作成する。従って，LDBを仲介として各システムで使用される辞書は互換性が保たれることになる。

今後は，LDBのような共通の基盤上で辞書のあり方などを各者が議論できるような状況を作り出すことが必要であろう。

6．LDB（Lexical Database）の構想

6. LDB（Lexical Database）の構想

6.1 LDBとは

計算機による自然言語処理を行うためには対象となる言語の辞書が必要である。このために冊子体の辞書を計算機で読むことのできる形に変換した機械可読辞書（MRD；Machine-readable Dictionary）が従来から利用されてきた。しかし，深く広い自然言語処理のためにはMRDでは不十分であり，最近では大規模な語彙に関するデータベース（LDB；Lexical Database）が求められている。

6.1.1 MRDとは

従来，機械翻訳などの自然言語処理を行うにあたって，語に関する情報はシステム開発者が独自に収集しシステムの中に組み入れるという方法が多く取られてきた。そこでは市販の辞書から必要な情報を抽出し，システムに入力した上で編集・変換を加えるなどの作業が行われる。しかしそのような作業には人手もかかるし，他のシステムとの間で作業や情報の重複も多くなる。そこで，まず市販の辞書をファイル化して計算機で扱えるようにする試みが行われた。このようなファイルをMRD（Machine-readable Dictionary）と呼ぶ。これを利用するとデータ収集および入力の作業が不用になるので，大学・研究機関をはじめとして多くの組織で歓迎された。最近では辞書の印刷過程でいわゆる電算写植や電算鋳字が広く行われるようになったためにMRD化が容易になったが，それればかりではなく出版社自らがMRD化を行うことも可能になってきた。

日本におけるMRDの例としては，三省堂や岩波の国語辞典，電辞林という愛称で知られる英和・和英辞書（専門語についてはインタープレスのデータを利用），三修社の英独和科学技術用語辞典，また学研やインタープレスの多種のデータベースなどがある。英国圏でも同様の試みがあり，これまでにロングマン現代英語辞典（ロングマン），アメリカン・ヘリテージ英語辞典（ホートン・ミフリン），コンサイス・オックスフォード英語辞典（オックスフォード），ウェブスター英語辞典（ウ

ブスター）などがMRD化されている。さらに本調査でもたびたび触れているように，新オックスフォード英語辞典（New OED－全12巻および補遺4巻）を1989年までの3年でデータベース化する作業がウォータールー大学を中心として行われているが，その副産物としてのMRDが本年中に発売されるといわれている。

6.1.2　LDBを必要とする背景

上記のようにMRDは自然言語処理システムの研究開発にとって非常に利用価値の高いものであるが，一方で数多くの欠点がある。そのうちの主なものは次のとおりである。

①　構造化されていない

MRDの中には印刷書式のままでファイル化されたものがある。そこから所定の項目を検索することは直接にはできない。

②　間違いが多い

特に構造化されていないMRDには冊子体の辞書に含まれる誤植がそのまま受け継がれることが多い。また，各種括弧の使用法，注釈など凡例と異なる記法が随所に見られる。

③　記述が人間を対象にしている

当然のことではあるが，元にした辞書が人間の読者を対象にしたものであるので，記述はさほど厳密なものでなく，例などを使って読者の常識に訴えることにより逃げていることが多い。また語義も機械処理に向いた語の定義にはなっていない。

④　語彙数が少ない

New OEDを別にして，その他のMRDは机上版の辞書を元にしているため，一般に語彙は数万語のレベルである。しかし自然言語処理のためにはこの語彙では不十分である。

上の①，②の欠点はファイル化する際に厳密な構造化を行うことによりある程度補うことができる。しかし，原典となる辞書が人間を対象としたものである限り，③はどうしようもない。そこで人間用の記述とは別に，いずれにしても機械処理のための情報を用意する必要がある。

以上のようなMRDの欠点を克服するためには，計算機による自然言

語処理のために適切な記述方法と十分な語彙を持ち，厳密に構造化されたデータベースが必要である。現在の自然言語処理技術，データベース／知識ベース技術などの水準は，工学的にそれを可能にしていると言える。LDBはこのような要請のもとに構想されているものである。

しかし一方で，理論的にはLDBに関して困難な問題が多数あると言わざるを得ない。主な問題としては，機械処理のための語義の記述とはどういうものか，あるいは応用分野を選ばない汎用のデータベースの構築は可能か，などがある。これらの問題を現時点では保留にするが，大規模なデータベースを作る方を優先しようというのがいわゆる電子化辞書の開発の試みである。これについては6.2で述べる。

6.1.3 LDBの構想

上記のように，計算機による自然言語処理のための大規模なデータベースが現在さまざまな分野で求められている。ここではそれをLDBと呼ぼう。これはデータベースとして汎用であり，一つの標準を与えるものでなければならない。後の節で示すようにこのようなLDBの開発は世界的な趨勢である。

諸分野におけるLDBへの要請をまとめると次のようになる。

① 標準的なデータベース

一つの言語圏において標準的なLDBが一つ必要である。また費用効率から考えても，汎用のものが一つだけあれば良い。

② 一元的な管理形態

LDBの管理（作成・更新・増補・改訂）は権威者による委員会のような組織により一元化されるべきである。

③ 応用辞書の作成が容易

応用システムのための辞書（応用辞書）はLDBから必要な情報を抽出することにより容易に作成することができる。

④ 冊子体の辞書の作成が可能

冊子体の辞書の編集，出版もLDBを原典として利用することにより容易になる。

⑤ 広い応用分野

応用は言語学的，工学的なものに限らず，歴史，文学など人文学的な分野にも広げる。

このような要請を満すにはLDBの扱う対象は相当に広いものとなる。まず，扱う言語としてはmonolingual（単言語），bilingual（二か国語），multilingual（多言語）のものが応用に応じて必要である。また語彙の範囲としては一般語の他に諸専門分野に応じたそれぞれの専門語が必要である。さらに記述内容としては通常の辞書に含まれるような形態・構文情報，語義，用例などの他に，計算機による自然言語処理のための情報を持たなければならない。

LDBとしてNew OEDのような規範的なものを考える限りにおいては扱う言語は単言語である。しかし，その応用システムとして機械翻訳などを取ると二か国語さらにはそれ以上の情報が必要になる。多言語間での語の対応関係は一般には複雑なので，単言語のLDBを組み合せるだけでは多言語のLDBを得ることはできない。LDBとしてはこのような多言語の情報までを含めたものが望まれる。

専門語の扱いは特殊である。専門分野によっては化学のように語彙数が数百万にのぼるものもあり，その収集，管理はもはや言語やデータベースの問題とは言えず，その専門分野の問題となる。従って，データベースの形態としてはLDBの一部ではなく，専門語データベースとしてLDBに付随したものとなるべきであろうが，その利用者にとっては区別無く使えなければならない。

次に計算機による自然言語処理のための情報としてはシソーラスや知識表現が上げられる。前者は語義を表現するのに他の概念との関係を示すことによって行うものであり，後者は語義や語の用法を計算機によって扱うことの可能な数学的表現によって記述するものである。意味論的な処理を伴う高度な自然言語処理を行うためにはこれらの情報は必須のものである。

6.2 国内外におけるLDBの動向

LDBを開発しようという動きは各国に見られる。これは自然言語処理

技術の発達がLDBを必要とするフェーズに差し掛かったことが大きな理由と考えられるが，それと同時にMRDの利用経験とそれに伴う資源の蓄積がLDBの開発を可能にしていることも事実である。

6.2.1 日本におけるLDBの動向

言語情報の統合的データベースであるLDBは，その規模・標準性から，LDBそのものが日本で開発されている訳ではない。しかし，LDBへ向けての言語データの整備，MRD，電子化辞書の開発および研究は進められている。ここで，その動向に簡単に触れ，「6.2.3 日本におけるLDB関連研究事例」で詳しく紹介することとする。

言語データの整備については，国立国語研究所における調査研究がある。国立国語研究所では，現代日本語の書き言葉の使用状況についての情報をおさめる「総合辞書」の構想が進められつつあり，また，日本語英語の機械可読用シソーラスの作成に着手している。また，日本語の合成語に関する「造語成分の辞書」を拡充すべく研究が進められている。これらは，LDBに対して言語データの供給源となりうることが期待される。

電子技術総合研究所では，国内では最も早く，1977年に三省堂新明解国語辞典（第二版）をコンピュータ上にインプリメントしてMRDを作成している。電総研ではこのデータを各研究機関に配布し利用に供しているとともに，さらに，英英辞典，国語辞典，英和辞典などの各種市販辞書を機械可読化し，また，これらMRDを用いる自然言語処理システムなどのソフトウェアを開発している。現状では，複数MRDが混在している状況であるが，これら種々の媒体に分散している各種MRDの統合・整理とコンピュータ・ネットワークを介した開発・保守ツールの整備が将来進められれば，有力なLDBへの情報源となりえよう。

また，昭和61年4月には㈱日本電子化辞書研究所が設立され，電子化辞書の研究開発を開始した。この会社は国が7割，コンピュータ・メーカ8社が3割の割合で共同出資して設立されたもので，10年間で日本語と英語の電子化辞書を各20万語の規模で開発することを目的としている。ここでいう電子化辞書は前節のLDBの構想と比較するとその

標準性，管理機構の点で，必ずしもLDBに望まれる資質を満たしているとは言えないかも知れないが，内容としてはLDBそのものを目指すものである。

6.2.2　海外におけるLDBの動向

海外においても，本報告書の第4章で示したように多くのコンピュータ解析用辞書の開発の構想がある。その中でLDBに分類されるものとしては次のようなものがある。

①　ホートン・ミフリン出版社

アメリカン・ヘリテージ英語辞典を元にしたLDBの構想。語義を意味ネットワークを用いて記述することを目指している。

②　オックスフォード大学出版部

New OEDを元にしたLDBの構想。これは開発段階にある。

③　ピサ大学（イタリア）

イタリア語と英語のLDBの構想。ここでは文学的なテキストベースと組み合わせて，より広範なシステムを目指している。

また，LDBとは称していないが，EUROTRAの各機関で開発が進められている辞書も，その規模や内容からみてLDBに近いものになることが考えられる。

6.2.3　日本におけるLDB関連研究事例

6.2.3.1　国立国語研究所

(1)　語彙調査にもとづく「総合辞書」

国立国語研究所では，現代日本語の語彙の実態を把握するために，主として書き言葉を対象とした語彙調査を行ってきた。大規模な語彙調査としては，以下のようなものが挙げられる。

(a)　婦人雑誌の語彙調査

(b)　総合雑誌の語彙調査

(c)　現代雑誌90種の語彙調査

(d)　現代新聞の語彙調査

(e)　高校教科書の語彙調査

(f)　中学校教科書の語彙調査

このような調査の成果のうち，方法論あるいは理論に関するものとしては，

① 層分け標本抽出法による語彙調査の方法など統計学の応用

② 同じ語か異なる語かの判別を操作的に行う基準の作成

③ 語の基本度関数の提示

などを挙げることができる。

一方，語彙表としては，以下のようなものが報告されている。

① 五十音順による語の使用率

② 使用率順による語の配列とその分布

③ 語種（和語・漢語・外来語・混種語）別，品詞別，語彙成別などの量的構造の分析

④ 意味領域による語彙分類とそのコード化

⑤ 漢字の使用率ならびに表記の実態をしめす漢字表，語の表記の一覧表

こうした語彙調査の結果をまとめて，今後の語彙調査をふくめ自然言語処理に役立てようというのが「総合辞書」の構想である。「総合辞書」は，日本語の書き言葉の使用状況についての情報をおさめるもので，われわれが一般に目にする冊子体の辞書とはその内容が大きく異なっている。

ただし，この「総合辞書」の作成には，いくつかの問題があるようである。それは，(a)～(f)の調査における，調査単位が共通でないことである。ある調査では長い単位が，別の調査では短い単位が採用されていて，これらを単純にそのままのかたちで統合するわけにはいかない。また，同じ語か異なる語かの判別がすべての語彙調査で等しく行われたわけではない。つまり，それぞれの調査に参加した研究者の方針が異なり，単位や単語の認定に差がみとめられるのである。そのような問題はたいしたことではないという意見もあろうが，言葉を研究するものにとっては重要な問題のひとつである（日本語には分かち書きの習慣がなく，このことも安定した単位を得にくくしている要因である）。「総合辞書」の作成にむけて，こうした問題点を解決する必要がある。

理想をいえば，「総合辞書」には，既存の語彙調査の結果から得られた

情報に加えて，次のような情報を盛り込むことが望ましい。

- 類義語，対義語，上位語，関連語などの意味に関わる情報
- 語形変化をする単語の変化のタイプ（活用のタイプ），派生語・複合語などに関する形態論的な情報
- 動詞のように，広義の語形変化を発達させている単語群には，ウォイス，アスペクト，スタイル，待遇，テンス，ムードといった文法的カテゴリー

今日，このような研究が多くの研究者のあいだで進められており，その成果を辞書にとりこむ必要がある。さらに，動詞や形容詞における格支配の現象も，これらの品詞に属する単語がもつ統語的な特性として辞書に記述することが可能である。

(2) シソーラス

国立国語研究所では，機械可読用のシソーラスの作成に着手しており，現在「分類語彙表」（秀英出版）の増補が行われている。この仕事は「新明解国語辞典」（MT版，電子技術総合研究所）に，「分類語彙表」の分類コードを付し，これに「角川類語新辞典」（角川書店）の分類コード，「教育基本語彙」（牧書店）の学習段階をしめした記号などを補充するかたちで進められてきた。機械可読のシソーラスは，従来の冊子体の「分類語彙表」などに比べて，収録語数が格段にふえているうえに，単語の多義性を配慮し，一つの単語に，それぞれの意味に応じて複数のコードが付与されている。

日本語のシソーラスとは別に，英語のシソーラスも計算機に入力されている。Longman lexicon of contemporary English がそれである。（当初は，Roget's International Thesaurus の約10万語の入力が予定されたが，語数が多いことから，Longmanに変更された。）

以上の二つが完成すれば，日本語と英語の意味分類体の辞書がそろうことになる。両者を比較し，自然言語処理用の辞書の問題点を明らかにしようとするプロジェクトがある（中野洋「日英語彙データの収集・比較と機械辞書の作成」ー文部省特定研究(1)「言語情報処理の高度化の基礎的研究」の中のテーマ）

(3) 造語成分の辞書

日本語は，単語あるいは形態素を組み合わせて新たな合成語（語基と接辞からなる派生語と二つ以上の語基からなる複合語）を生産的に作ることができる言語である。それゆえに，辞書の語数を増やしても，なおかつ辞書に登録されていない合成語がしばしば現れるものである。専門用語としての，いわゆる術語に加えて，新聞・雑誌などにもその場かぎりの臨時的な合成語が頻出する。こうした合成語を計算機で処理するためには，合成語の構造パターンを明らかにしておく必要があると同時に，造語成分の辞書を拡充することが不可欠である。こうしたことから，合成語のデータを統一的な単位に分割し，それぞれの構成要素について，語種情報・形態情報・結合情報の三つの付加情報を記述した連接表が作成されている。図6－1にあげるのはその具体例である（石井正彦「語構成分析のための単位と付加情報について」）。

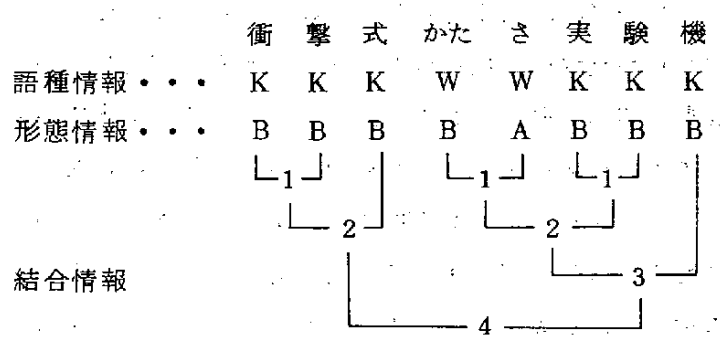


図6－1　連接表（例）

なお，語種情報というのは，形態素の出自による区別であり，和語系（W），漢語系（K），外来語系（G），混種（H）からなり，形態情報とは，語基（B）と接辞（A）の区別である。語基とは，一般に，相対的に自立できる形態素であるのに対して，接辞は自立できない形態素をさす。ただし，漢語系のものについては，自立性がなくても，一文字を語基とみなしているようである。結合情報の数字は結合のレベルを表している。

このような三つの付加情報は，たとえば，「衝撃式かたさ試験機」の場合には，"KB1KB2KB4WB1WA2KB1KB3KB"のような文字列とし

て書き表すことができる。合成語のデータにこうした分析をほどこし，そこから，語基や接辞をおさめた造語成分の辞書が得られるものと期待される。

以上，国立国語研究所で目下進められている三種類の機械可読辞書を紹介した。これらは，いずれも研究途上にあり，完成されたものではない。また，LDBそのものと言い得るかどうかは検討を要するが，LDBへの重要な基盤となりうると考えられる。

6.2.3.2　電子技術総合研究所

電総研では，国内では最も早く，すでに1977年に，三省堂の新明解国語辞典（第二版）をコンピュータ上にインプリメントした。その後このデータは，各研究機関に配布されて用いられている。同じ辞典の新しい版（第三版）についてもデータがほぼ完成しているが，種々の事情から校正が完了していないのでまだ一般には公開されていない（マニュアルだけは印刷物になっている）。その他にも，研究室予算で入力した和英辞典，他からのデータを研究に利用している英英辞典，国語辞典，英和辞典などがある。これらはいずれも市販の辞書をコンピュータ上で扱えるようにしたものである。辞書の性質から当然であるが，コンピュータ上である程度の容量が必要なので，中にはもとの媒体である磁気テープのままで利用が進んでいないものもある。

この中で，最もよく利用しているものは，入力の時から一貫して自主開発してきた新明解国語辞典であり，見出し語に関する種々の情報，見出し語を辞書として用いる自然言語処理システムなどのソフトウェアが揃っている。このようにデータの周りのソフトウェア環境を整えるためには，データそのものの何倍かの容量を必要とする。しかもデータそのものが正確でないと，ソフトウェア環境を整える上で障害となることがある。新明解国語辞典のデータは，最初の磁気テープの段階から約4万2千カ所の修正が施されているので，この点の問題はなかった。しかしながら，容量の問題など，以下に述べるさまざまの問題点のため，電総研の現状は，まだLDBまで至っておらず，複数のMRDが混在しているという状態である。以下では問題点をあげると

ともに，将来像についても触れることにする。

まず第一に，前述の容量の問題がある。通常の辞書は，だいたい10～20MBの容量でその本文を入力することができるが，それをきちんとしたデータベースの形で使用しようとすると，少なくとも50～200MBくらいの容量が必要である。もちろんこれには，見出し語などのファイル情報，保守のためのプログラムなどが含まれている。最近ではオフコンでもかなり大容量のものが増えてきたが，それでもこれは結構大きな容量である。また大型機を共用している場合には，(占有でなければ) この容量を確保するのは難しい。またこれだけの容量を確保したとしても，周りの環境を整えておかないとその後の保守が難しい(保守の問題については後述)。

第二に問題となるのはファイルの収容形態である。データは，その規模が大きくなればなるほど容量をコンパクトに収めることが重要になる。しかしながらそのデータへのアクセスの効率を考えると，あまりコンパクトにし過ぎるとアクセス時間がかかって使い勝手が悪くなる。さらに拡張性についても考慮する必要がある。これらの互いに矛盾を含んだ要求を満たすためには，ファイル校正について十分に考慮する必要がある。ところが，通常ファイルはコンピュータシステムに依存したものが多いので，システムが変わると全面的に変更しなければならないことがある。また，大容量ファイルを，たとえば順編成で扱うなどは非現実的であるが，システムに備わったエディタは，大抵順編成ファイルしかサポートしていない。また，ランダムファイルにすると，物理的にファイル容量を食うようなシステムも多い。ハッシングなどの技法も使用しているが，関数の定め方，拡張性などでの問題も多い。

第三に，すでにたびたび触れてきたように，保守の問題がある。前項でも触れたように，辞書向きのエディタというのは，市販のコンピュータに備えられた(日本語)エディタでは不十分である。したがってどうしても専用のツールを開発する必要がある。電総研でも簡単なツールはいくつか開発しているが，一般に応用できるような，辞書開

発ツールというものはまだ作成していない。保守（特に誤り訂正），拡張には，それに対応できる（エディタよりややシステム化された）ツールが必要である。

では，辞書の開発，保守ツールとしてどのようなものがよいかを含めて将来像をどのように描いているかを以下で少し述べてみよう。まず最初にやるべき（やらなければならない）ことは，種々の媒体に分散しているデータの統合，整理である。上に述べた各種の辞典は，(磁気テープのままであるものも含めて）種々のコンピュータ上に置かれている。しかもファイル形態もそれぞれ異なり，コードもまちまちである。したがって統合，整理といっても容易なことではない。ここでいう統合，整理とは，必ずしも同じ物理ファイル上にあることを意味しない。それは最近のネットワーク技術の発達によって，他のコンピュータをファイルサーバとして，必ずしも自分のシステム上にデータが存在しなくても高速アクセスが可能になってきているからである。したがって将来的には，ワークステーションのようなものをメインとして，各辞書のデータそのものはどのコンピュータ上にあってもよいが，それらのデータに柔軟にアクセスできるシステムやプログラムを，メインにすえておくようなシステムをねらっている。具体的には，マルチウィンドウの機能などをフルに活用して，ある辞書を見たり修正したりするのに常に他の辞書やシステムを参照することができるようなシステムが理想である。しかしながら現実の予算と人手（このようなシステムにはマンパワーとそれに伴う予算が必要である）からみて早急に実現する見通しは現在のところあまりない。

6.2.3.3　日本電子化辞書研究所

昭和６１年４月に国（（特認法人）基盤技術研究促進センター）と民間８社（富士通㈱，日本電気㈱，㈱日立製作所，シャープ㈱，㈱東芝，沖電気㈱，三菱電気㈱，松下電器産業㈱）の共同出資で㈱日本電子化辞書研究所が設立された。これは大規模で広範囲な自然言語処理のために言語情報を大量に集積した大規模辞書の研究開発を目的とした組織である。

この電子化辞書の構成は図６－２に示されている。マスタ辞書は日本語と英語を扱うが，概念辞書は語によって表される概念そのものを扱うためにそのような区別はない。

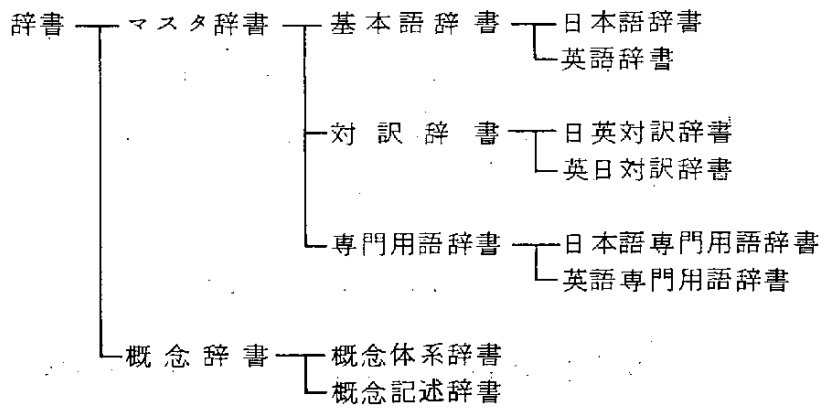


図６－２　辞書の構成（日本電子化辞書研究所）

マスタ辞書は言葉が表している概念を定義し，その表層表現である単語の文法的特性を記述するもので，基本語辞書，対訳辞書，および専門用語辞書からなる。基本語辞書は日本語辞書と英語辞書，対訳辞書は日英対訳辞書と英日対訳辞書であり，それぞれ２０万語の語彙項目からなっている。これらの辞書の作成については，語の選定から記述，形式化までを全く新規に行う。

辞書の内部に格納される情報としては，見出し語情報，品詞情報，意味情報，訳語情報，そのほかの情報がある（図６－３）。これらの項目は，人間が読んで理解するためのものとは異なり，コンピュータプログラムからのアクセスが容易なように設定されている。内部の記述を統一することを含め，コード化を徹底して行っている。

語義についても，例で示すのではなく，明確に定義する思想である。また語義文については，係り受けの曖昧さを解消するために，カッコにより係りを明示した形式となっている。

この単語辞書は，辞書標準インタフェースに従って，ワークステーション上の辞書ファイルとして構築される。日本語辞書，英語辞書，

項　　目	内　　容
見出し語情報	見出し語，重要度，表記法，異表記，単語識別番号
品 詞 情 報	異表記，品詞，活用（変化），構成語，補足説明
意 味 情 報	異表記，語義，概念体系への対応付け，表層格パターン，構成語，補足説明
その他の情報	用法上の注意，専門分野，類語，反意語，発音
訳 語 情 報	訳語区分（同義，類義，上位，下位語），訳語

図６－３　辞書項目の内容

それらの語義概念間の対応関係，上位－下位関係も項目単位のリンク情報として表現される。辞書の参照，検索のためのシステムはUNIX上に実装されている。

専門用語辞書については，一専門分野１０万語単位で辞書を作成する予定である。辞書の内容は基本語，対訳辞書に準ずるが，名詞が中心のものと考えられている。

概念辞書はマスタ辞書で定義された概念の体系化と関連付けを行うためのもので，具体的にはマスタ辞書で定義された概念を上位－下位，全体－部分関係等によって分類し，概念の体系化を図る（概念体系）。また体系化された概念に対し，属性を定義するための記述を行う（概念記述）。概念体系と概念記述とは一体となって，個々の概念の意味，概念相互間の関係を表現する。従ってその構造，内容，表現の上で両者は密接に関連し合うため，研究開発は相互にフィードバックを行いながら進められる。また１２０万概念以上に及ぶ膨大な概念世界を扱うため，研究開発ツールとして強力なデータ管理システムが用いられる。

辞書を開発する支援システムとして，ワークステーション上に辞書作成支援プログラムが用意されている。これは従来，手書きで行われていた辞書の校正，編集作業をワークステーション上の専用エディタで効率良く行うものである。作業はワークステーションに，辞書の記

述項目に従ったワークシートが表示され，その項目を入力，修正するような形となる。この支援プログラムは各種のユーティリティを持ち，フォーマットや内容のチェックに対しても強力なサポートが可能となっている。

辞書を作成するのは数年間にわたる作業になる。この間は，図６－４に示すように相互に干渉し，整合をとる方式がとられる。この方式に従うと，辞書作成のサイクルは次のようになる。まず，日本語辞書の各概念に対して対訳付けを行い，その対訳英語に対して英語辞書の概念とのリンク付けを行う。この概念が英語辞書に無い場合には，新たに英語辞書の見出しを設ける。英語辞書からも日本語辞書に対して同様のリンク付けが発生する。これを繰り返すことにより基本語辞書と対訳辞書の概念間のリンクを完全にすることを図る。

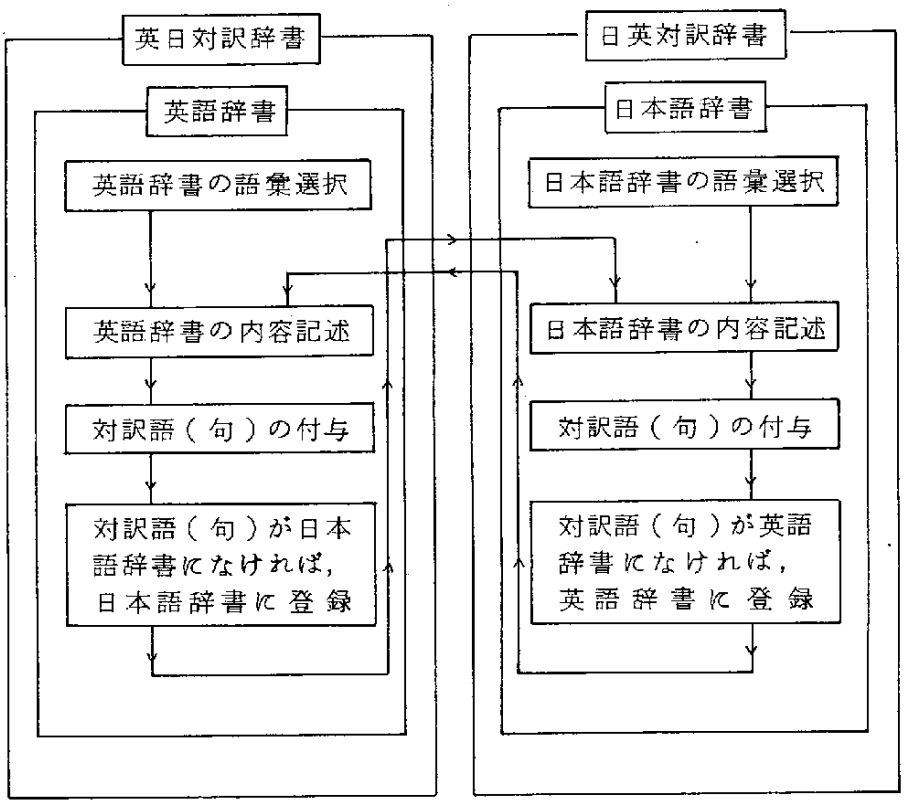


図６－４　辞書作成の方法

昭和６１年度末現在で研究開発は次のように進んでいる。マスタ辞書は，基本語辞書の見出し語１７万語の選択を終え，そのうちの２万２千語に関する内容記述が行われた。日本語辞書，英語辞書の記述内容について，機械翻訳をはじめとする自然言語処理一般に適合できるよう検討を行い，その内容および記述形式を定めて執筆・入力作業が進められた。また記述内容を検証するため，豊富な用例を収集して分析をしている。概念の体系化に関しては，次年度からの本格的な研究開発に備え，基本的な概念について分類を検討し，試作を行った。概念記述に関しては，既存の深層格の調査を踏まえて概念記述に必要な概念間の関係を洗い出し，文解析という面から見た必要な情報（知識）の選択を行っている。また，データ管理システムとして，各辞書を格納するための辞書管理システムの作成，およびワークシートからマスタ辞書を作成するためのユーティリティ，マスタ辞書の校正支援のための辞書エディタ・ウインドウシステムの設計，試作等を行っている。

7．今後の課題

7. 今後の課題

「自然言語のコンピュータ解析用辞書に関する調査研究」事業の第2年目に当る本年度の目標は，コンピュータ解析用辞書整備構想のための基本構想をとりまとめることであった。

このため，自然言語処理応用システムと解析用辞書の現状を概観し，機械翻訳システムの辞書更新についての国内調査，辞書の作成及び標準化動向を把握するための海外調査を実施した。

さらに，それらの調査結果をもととして，辞書及びその更新のあり方を考察し，辞書の分類・構成，作成，更新および現状の問題点とあるべき姿を検討した。

この調査研究の中から，ユーザやメーカで辞書の作成・更新がさらに進められていくと，辞書の整合性管理が大変困難になるため，各種インタフェースを含めた共通化，標準化を計れるものの必要性が認識され，また，内外の各機関での動向はLDB(Lexical Database)の開発へ向かっていることも認識された。

このことから，本年度調査研究のテーマである「辞書整備基本構想」の基本概念としてLDBを採用することとした。

したがって，「今後の課題」として，LDBを中心に据えて，(1) LDBの開発・維持・管理に関する技術的課題，(2) LDBの開発・維持・流通に関する体制上の課題，(3) LDBの波及効果に関する課題ーを指摘する。

今後の課題

① LDBの開発・維持・管理に関する技術的課題
- 言語情報の体系化に関する技術
- 知識ベース，データベース技術
- ネットワーク技術
- 電子化辞書(ED)，MRDの現状と動向

② LDBの開発・維持・流通に関する体制上の課題
- LDBの開発・維持に関する組織
- LDBの管理体系

- 辞書，ＬＤＢの流通の仕組み
- 産学官の役割分担

③　ＬＤＢの波及効果に関する課題

- ＬＤＢの利用発展性
- 言語情報の産業化

資　料　編

資料Ⅰ．国内調査票

資料Ⅱ．海外調査訪問機関

資料Ⅰ. 国内調査票

自然言語解析用辞書

アンケート調査票

（メーカー向け）

財団法人　日本情報処理開発協会

－目次－

Ⅰ　貴社の概要

Ⅱ 機械翻訳システムについて

Ⅲ 辞書の概要（機械可読辞書）

Ⅳ　辞書のユーザーへの提供方式

Ⅴ　辞書のユーザーへの解放について

Ⅵ　辞書作成（メーカー側）

Ⅶ　辞書メンテナンス（メーカー側）

Ⅷ　辞書内容の変更（ユーザー側）

Ⅸ　ユーザー用辞書更新ツール

Ⅹ　辞書構築（ユーザー側）

ⅩⅠ　辞書更新に関するユーザーへのメーカー側の支援体制

I　貴社の概要

1．貴社名＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

2．所在地　〒□□□-□□＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

TEL（　　）

II 機械翻訳システムについて

1．システムの名称＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

2．対象言語（英日・日英　等）＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

3．対象分野＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

4．稼働環境

（1）計算機の機種＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）ＯＳ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（3）プログラミング言語＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（4）必要とする主メモリ容量＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（5）必要とする外部メモリ容量＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

5．利用方式

○バッチ方式　○会話型　○その他＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

6．想定するユーザー形態

（1）○個人利用　○センター利用　○その他＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）○翻訳者　○エンドユーザー　○その他＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

7．ユーザー側の運用体制をどのように想定していますか。

8．保守体制はどうなっていますか。

9．システムの実働のために、ユーザーごとに準備しなければならない事柄をお書き下さい。

10．ユーザーのシステムへの介入をどの程度考慮していますか。

11．システムの販売について

（1）販売方式＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）販売部署＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

Ⅲ 辞書の概要（機械可読辞書）

1．名称＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

2．作成日（作成期間）＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

3．種類＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

4．機能＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

5．役割＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

6．規模（語い数）＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

7．ユーザー運用上の辞書パターン（システム辞書・センタ辞書・グループ辞書・ユーザー辞書　等）＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

8．使用しているコード

○ＪＩＳ　○シフトＪＩＳ　○その他＿＿＿＿＿＿＿＿＿＿＿＿＿＿

9．辞書構成（体系）についてお書き下さい。

１０．データ項目について簡単にお書き下さい。

Ⅳ　辞書のユーザーへの提供方式

1．ユーザーへの提供方式＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

2．提供媒体

○磁気テープ　○ハード・ディスク（容量＿＿＿＿ Mバイト）
○フロッピー・ディスク（容量＿＿＿＿ Mバイト）
○その他＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

3．提供形態

○ソース・コード　○オブジェクト・コード　○その他＿＿＿＿＿＿＿＿

4．バージョン・アップ版辞書の提供がありますか。

○ない
○ある
　└ ある場合

（1）種類＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）頻度＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（3）規模＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

5．辞書に関する権利はどのようになってますか。

6．辞書内容の漏れ防止対策について、お書き下さい。
（具体的な方式まで書かなくとも、コンセプト程度はお書き下さい）

V　辞書のユーザーへの解放について

1．解放範囲（どのように解放しているか）＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

2．解放条件＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

3．解放内容はどのようになっていますか。
（機能語・品詞・見出し語・訳語・ルール・パラメータ・コード体系）

4．解放できない部分については、その部分と理由をお書き下さい。

Ⅵ　辞書作成（メーカー側）

1．基本語い、専門語い別に収集法、選択基準についてお書き下さい。

（1）基本語い

収集法

選択基準

（2）専門語い

収集法

選択基準

2．基本語い、専門語い別に、辞書化の手順をお書き下さい。

（1）基本語い

（2）専門語い

3．複合語、慣用句、多義語への対処方法についてお書き下さい。

（1）複合語

（2）慣用句

（3）多義語

4．重要語の派生的な意味や、特殊な意味についての対処方法についてお書き下さい。

Ⅶ　辞書メンテナンス（メーカー側）

1．作成済みの辞書内容のチェックの頻度、方法、手順についてお書き下さい。

（1）頻度＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）方法

（3）手順

2．テスト用例文の種類と数についてお書き下さい。

3．辞書・文法の修正、維持、拡張方式についてお書き下さい。

4．辞書のデータ拡張に対する考え方

○他システムの辞書を利用する

○ユーザー所持辞書を利用する

○翻訳辞書や拡張ユーザー辞書の他アプリケーションを利用する

○その他＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

5．専門辞書を複数利用したことがありますか。

○ない

○ある

└─ その成果についてお書きください＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

Ⅷ　辞書内容の変更（ユーザー側）

1．ユーザーにより、辞書内容の変更ができますか。

○できない

○できる

└ ユーザー・メンテナンスの各機能について以下にお答え下さい。

（1）変更

○できない

○できる

└ 内容

方法

（2）追加

○できない

○できる

└ 内容

方法

（３）削除

○できない
○できる
　└ 内容

方法

（４）その他の機能

２．ユーザーによる訳し分け

○できない
○できる
　└ 内容

方法

３．マニュアル

○ない
○ある

4．ユーザー・メンテナンスに対する、ユーザーからの要望がありましたらお書き下さい。

5．今後備えていく機能がありましたらお書き下さい。

Ⅸ　ユーザー用辞書更新ツール

1．ユーザー用辞書ツールがありますか。

○ない

○ある

└ 種類＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

機能

方法

特徴

X　辞書構築（ユーザー側）

1．ユーザーが辞書（ユーザー辞書／専門辞書）構築に当たって、利用できる機能の詳細をお書き下さい。

2．ユーザー構築辞書と将来のシステム・バージョン・アップに対する整合性について、どの程度配慮していますか。

3．ユーザー辞書構築への支援体制をお書き下さい。

4．ユーザー作成辞書から、メーカー提供辞書へのフィードバックは、どの程度おこなわれていますか。

XI　辞書更新に関するユーザーへのメーカー側の支援体制

1．辞書更新支援体制についてお書き下さい。

2．ユーザーの習熟度への対処方法についてお書き下さい。

3．辞書バージョン・アップ時の支援体制（メーカー側請け負い・Ｋｉｔによる自動更新ツール提供　等）

4．その他の支援（ユーザー会、Ｎｅｗｓ、ユーザー辞書の買取り／販売支援等）がありましたらお書き下さい。

自然言語解析用辞書
アンケート調査票

（ユーザー向け）

財団法人　日本情報処理開発協会

－目次－

Ⅰ　貴方について

Ⅱ　導入している機械翻訳システムについて

Ⅲ　運用している辞書の実態

Ⅳ　辞書の更新について

Ⅴ　辞書に対する評価

Ⅵ　辞書更新に関する運用体制（ユーザー側）

Ⅶ　辞書メンテナンス

Ⅷ　その他

Ⅰ　貴方について

1．団体名＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

2．所在地　〒□□□-□□＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿TEL(　　)＿＿＿＿＿＿＿＿＿＿

Ⅱ　導入している機械翻訳システムについて

1.（1）システム名＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）対象言語（英日・日英　等）＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（3）メーカー名＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（4）販売部署＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

2.（1）導入目的＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）導入形態

○買取り　○リース／レンタル　○その他＿＿＿＿＿＿＿＿＿＿

（3）導入部署＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

3．導入から実運用までの作業内容と所要時間・工数

作業内容	所要時間・工数

4．現在のシステムの概要

（1）計算機の機種＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）ＯＳ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（3）プログラミング言語＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（4）要員＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（5）予算＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（6）翻訳量＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

5．機械翻訳システムと結合して作成利用しているツール類がありますか。

○ない

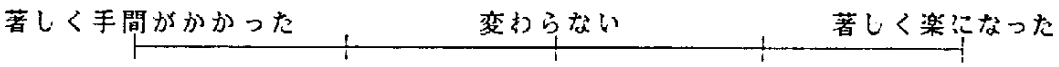
○ある

ツール名	内容

6．機械翻訳システムの運用体制についてお書き下さい。

7．機械翻訳システムの評価について、適当なところに印をつけて下さい。

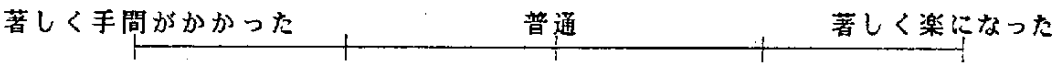
（1）人手翻訳との比較



その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

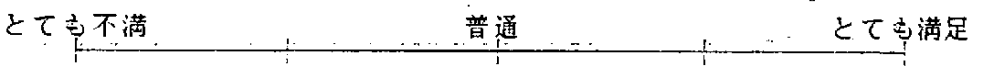
＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）プレエディト



その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

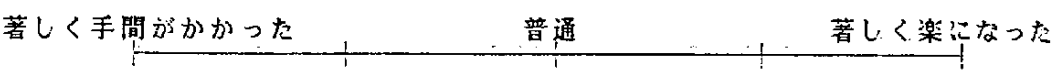
＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿



その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（3）ポストエディト



その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

とても不満　　普通　　とても満足

その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（４）完成度

とても悪い　　普通　　とても良い

その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（５）導入効果

全く効果無し　　普通　　とても効果有り

その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（６）全体としての満足度

とても不満　　普通　　とても満足

その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

Ⅲ　運用している辞書の実態

1．（1）種類＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）機能＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（3）役割＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（4）規模（語い数）＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

2．辞書の構成内容（単語・イディオム・品詞・意味情報・構文情報　等）についてお書き下さい。

3．メーカー提供辞書のユーザーへの解放範囲と利用の実際について、基本辞書・専門辞書別にお答え下さい。

（1）基本辞書

（2）専門辞書

4．一般辞書・利用者辞書の使い分けはどのようにされてますか。

5．専門辞書の分野分けは、どのようにおこなってますか。

分野	内容

6．専門辞書を複数利用したことがありますか。

○ない

○ある

└ 複数利用したことで、どのような成果がありましたか。

Ⅳ　辞書の更新について

1．ユーザー側で、どの程度辞書の更新が必要ですか。

全く必要ない　　まあまあ必要　　とても必要

その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

2．辞書の更新頻度

（1）一般辞書

全くしない　　まあまあする　　非常に多い

その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）専門辞書

全くしない　　まあまあする　　非常に多い

その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（３）その他の辞書＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

全くしない　　まあまあする　　非常に多い

その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

３．内容

（１）よく追加される語の種類（複数解答可）

○専門用語　○複合語　○固有名詞　○その他＿＿＿＿＿＿＿＿

（２）よく削除される語の種類（複数解答可）

○専門用語　○複合語　○固有名詞　○その他＿＿＿＿＿＿＿＿

（３）辞書の更新時に、訳語をどのように決定しますか。

（4）辞書の更新時に、訳語を統一する必要があると思われますか。

全く必要ない　　まあまあ必要　　とても必要

その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

4．テスト用例文の種類と数についてお書き下さい。

5．一般辞書更新時に、同時に、文法を更新することがありますか。

全くない　　まあまあある　　よくある

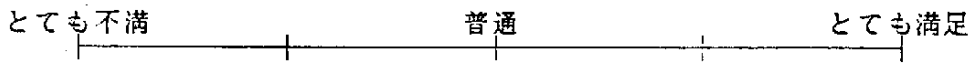
どんな場合ですか＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

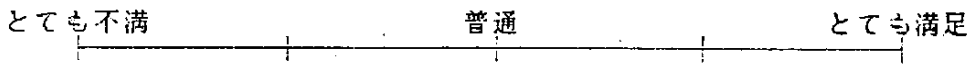
V　辞書に対する評価

1．メーカー提供辞書に対する評価



その理由についてお書き下さい。

2．ユーザーで現在使用している辞書の評価



その理由についてお書き下さい。

3．メーカー提供辞書をどの程度解放してほしいですか。

4．ユーザー辞書を充実することで、何らかの効果・弊害が起こりましたか。

（1）効果

（2）弊害

5．辞書の、ユーザーで拡張した部分の権利はどこにあるべきだと思いますか。

○メーカー　○ユーザー　○その他＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

その理由をお書き下さい。

6．辞書を翻訳以外の用途に使用したことがありますか。

○ない

○ある

└（1）どのような用途に使用しましたか。

（2）このとき、何らかの不都合が起きましたらお書き下さい。

VI　辞書更新に関する運用体制（ユーザー側）

1．運用体制について、（要員、工数などを含めて）以下にお書き下さい。

2．一般辞書の用語の収集・辞書化

（1）収集のために使用した資料

（2）収集方法

（3）辞書修正の理由

（4）翻訳辞書化の手順

３．専門辞書の用語の収集・辞書化

（１）収集のために使用した資料

（２）収集方法

（３）辞書修正の理由

（４）翻訳辞書化の手順

4．辞書更新の運用について

（1）運用規定についてお書き下さい。

（2）運用のマニュアルがありますか。

○ない

○ある

└ サンプルをお書き下さい。

５．辞書更新時の入力伝票のサンプルをお書き下さい。

6．入力方法

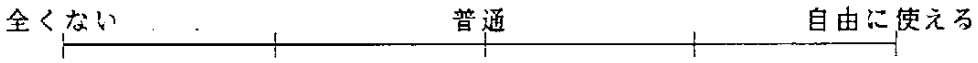
（1）　○バッチ方式　○会話方式　○その他＿＿＿＿＿＿＿＿＿＿＿＿

（2）登録操作の手順について詳細にお書き下さい。

（３）修正操作の手順について詳細にお書き下さい。

Ⅶ　辞書メンテナンス

1．辞書メンテナンス機能は、メーカーからどの程度解放されてますか。



解放内容についてお書き下さい。

2．利用者用辞書エディタがありますか。

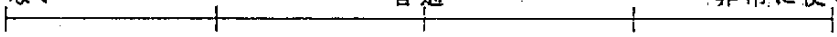
○ない

○ある

└─（1）エディタの特徴についてお書き下さい。

（２）エディタの使い勝手はどうですか。

使えない　　普通　　非常に使い易い



その理由＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（３）エディタに対する要望がありましたらお書き下さい。

3．メンテナンスのためのオリジナルのツールがありますか。

○ない

○ある

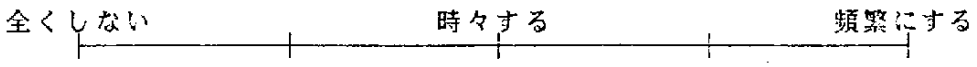
ツールの種類	機能

4．会話方式による辞書修正と、バッチ方式による辞書修正のおよその比率はどのくらいですか（語数で計測した場合）。

会話方式のみ　　同じくらい　　バッチ方式のみ

Ⅷ　その他

1．他システムとの辞書内容の交換をすることがありますか。



（1）交換形態

○他システムの辞書を利用する。
○他システムと辞書を相互に交換しあっている。
○他システムに辞書を提供している。
○その他＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

（2）辞書交換は、どのような場合にしていますか。

2．ユーザー辞書内容が、メーカー提供辞書にフィードバックされますか。

○フィードバックされるような体制が、メーカー側にない
○フィードバックされるような体制が、ユーザー側にない
○フィードバックする体制はあるが、ユーザー側ではやっていない
○フィードバックをおこなっている
　└ どの様にフィードバックがおこなわれているか、お答え下さい。

3．ユーザーとして考える理想的な辞書体系とメインテナンス機能について、お書き下さい。

（1）理想的な辞書体系

（２）理想的なメンテナンス機能

４．他機関との連携がありましたらお書き下さい。

資料 II. 海外調査訪問機関

1. National Library of Medicine
8600 Rockvill Pike, Bethesda, MD 20894, U.S.A.

Richard K. C. Hsieh, *Assistant Director, International Programs*
Tel: (301) 496-6481

David Kenton, *Information Systems*
Tel: (301) 496-1351

Roy Rada, *Editor IM, Chief MeSH, Research Fellow*
Tel: (301) 496-1495, ARPA: Rada@NLM

Atsuko Craft, *Index Section*
Tel: (301) 496-3261

Alexa McGay, *(Linguistics)*

Peri Schuyler, *(Unified Medical Language)*

Larry Kingsland, *(Expert Systems and AI)*

2. Pan American Health Organization
525, 23rd Street, N.W., Washington, DC 20037, U.S.A.

Muriel Vasconcellos, *Machine Translation Division*
Tel: (202) 861-4339

Lee Schwartz, *Machine Translation Division*
Tel: (202) 861-4338

3. Houghton Mifflin Company
One Beacon Street, Boston, MA 02108, U.S.A.
Telex: 940959 Houghton BSN

Howard Webber, *Vice President and Publisher*
Tel: (617) 725-5030

David R. Replogle, *Executive Vice President*
Tel: (617) 725-5032

Mary Sutherland, *Director of Marketing*
Tel: (617) 725-5232

Mary Ellen Hoke, *International Account Manager*
Tel: (617) 725-5232

Joanne Dudek, *Technical Support Division*
Tel: (617) 725-5232

4. **Université de Montréal**
3150, Jean-Brillant, CP 6128, succ. A, Montréal, Qué., Canada, H3C 3J7
Jean Baudot, *Professeur Agrégé, Département de Linguistique et Philologie*
Tel: (514) 343-7068
Jean-Yves Morin, *Département de Linguistique et Philologie*
André Clas, *Editor / Directeur, "μετα"*
Tel: (514) 343-7047
Richard Kittredge, *Département de Linguistique et Philologie*
Tel: (514) 343-7932

5. **University of Waterloo Centre for the New OED**
University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1
Tel: (519) 885-1211 ext 6183, Telex: 069-55259
CSnet: oxford3%watdcs@waterloo.csnet
Gaston H. Gonnet, *Department of Computer Science*
Frank W. Tompa, *Department of Computer Science*

6. **Longman Group Limited**
Longman House, Burnt Mill, Harlow, Essex CM20 2JE, England
Tel: (0279) 26721, Telex: 81259 Longmn G
Michael Rundell, *Managing Editor, ELT Reference*
Nikolai Dejevsky, *Publisher, Dictionaries and Reference Division*

7. **Oxford University Press**
Walton Street, Oxford, OX2 6DP, England
Tel: (0865) 56767, Telex: 837330 Oxpres
Timothy J. Benbow, *Director, New Oxford English Dictionary*
Edmund Weiner, *Editor, New Oxford English Dictionary*
John A. Simpson, *Co-Editor, New Oxford English Dictionary*

8. **EUROTRA Project**
EUROTRA, Commission of the European Communities
Bat. Jean Monnet, L-2920, Luxembourg
Tel: (00352) 4301-2885, Telex: 3423/3446 comeur lu
Sergei Perschke, *Multilingual Action Programme*
Peter Lau, *Multilingual Action Programme*
Loll N. Rolling, *Multilingual Action Programme*

9. EURODICAUTOM Project

Terminology Bureau, Commission of the European Communities
Bat. Jean Monnet, L-2920, Luxembourg

Jacques Goetschalckx, *Adviser and Head of the Terminology Bureau*
Tel: (00352) 4301-2389

Alain Reichling, *Terminology and Informatiion Applications*
Tel: (00352) 4301-2335

10. University of Bonn

Institut für Kommunikationsforschung und Phonetik - Universität Bonn
Poppelsdorfer Allee 47, D-5300 Bonn, Federal Republic of Germany

Winfried Lenders
Tel: (0228) 73-5638

Institut für angewandte Kommunikations- und Sprachforschung e.V.
Poppelsdorfer Allee 47, D-5300 Bonn, Federal Republic of Germany

Jan Brustkern
Tel: (0228) 73-5645, 73-5620

11. Siemens AG

Otto-Hahn-Ring 6, D-8000 München 83, Federal Republic of Germany
Telex: 52109 SIE D

Manfred Gehrke
Email: unido!ztivax!gehrke.uucp

Werner Remmele
Tel: (089) 636-44237

12. International Information Centre for Terminology (Infoterm)

Austrian Standards Institute
A-1021 Vienna, Heinestrasse 38, POB. 130, Austria
Tel: (0222) 26-7535 ext 309, Telex: 115960 onorm a

Christian Galinski, *Director of Infoterm, Secretary of ISO/TC 37*
Helmut Felber, *Ex-Director of Infoterm, UNESCO and Infoterm Consultant*

13. University of Pisa

Dipartimento di Linguistica, Università di Pisa
Istituto di Linguistica Computazionale del CNR
Via della Faggiola 32, I-56100 Pisa, Italy
Tel: (050) 502082, BITnet: SISTEMI@ICNUCEVM.BITNET

Antonio Zampolli
Nicoletta Calzolari
Irina Prodanof
Amedeo Cappelli
Lorenzo Moretti
Fabrizio Sebastiani
Carlota Nicolas

禁無断転載

昭和62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 (432) 9372 (代表)

印刷所　山陽株式会社
TEL (591) 0248(代表)

61-S-002