

## 翻訳システム実用化のための翻訳作業環境

坂本義行  
(電子技術総合研究所)

有賀妙子 鷺尾 敦  
(ファコム・ハイタック株式会社)

機械翻訳システムを利用して翻訳を行う場合、どのような作業が必要となるだろうか。そして、その作業を効率よく行うための作業環境としてどのようなものが望まれるだろうか。科学技術庁機械翻訳プロジェクト (Muプロジェクト) における総合的な作業環境の提供を目的とした総合システムは、(1)即時翻訳・テキスト編集、(2)一括翻訳、(3)辞書編集、(4)入出力テキストエディタ、(5)会話制御、(6)イメージ辞書の機能を持ち、昭和61年3月より、日英、英日、英語論文では、使い勝手のよい翻訳システムについての考え方やシステム開発時に考慮した点について議論するとともに、システムの構成、機能を報告する。

## Working Environment for using the Machine Translation System

Yosiyuki Sakamoto

Taeko Ariga Atsushi Washio

Electrotechnical Laboratory (1-1-4, Umezono  
Sakura-mura, Niihari-gun, Ibaraki 305)

Facom-Hitac Ltd. Tsukuba branch (27-16-2,  
Takezono, Sakura-mura, Niihari-gun, Ibaraki 305)

How to use the machine translation system? What environment should be desired for the efficiency of the translation?

Through the development of Unified Operation Control System, we have been studying an useful system to operate, which is aimed to provide integrated environment for translation in the project of the Machine Translation System by the Science and Technology Agency of the Japanese Government (Mu project).

Both of the systems, the Japanese-to-English translation system and the English-to-Japanese translation system, were completed in march 1986. The Unified Operation Control Systems have the following functions: (1) the Interactive Translation and the Text Editing (2) the Batch translation (3) the Dictionary Editing (4) the Input/Output Text Conversion (5) the System Control and the Maintenance (6) the Image Dictionary.

In this paper, we discuss the problem how to use the system and the significant points at its developing. Moreover, we introduce the system configuration and its functions.

## 1. はじめに

機械翻訳システムを用いて、翻訳を行う場合、原文の入力、辞書の編集、そして翻訳、訳文の出力と数々の処理が必要となる。端末に向かって翻訳する際に、これら一連の処理がスムーズに行えないと、機械翻訳システムの有用性を半減し、機械翻訳システムを用いることが面倒になりかねない。機械翻訳システムの一般利用者がコンピュータの専門家ではないことから高いマンマシンインターフェイスが求められる。

科学技術庁機械翻訳プロジェクト (Muプロジェクト) の総合システムは、実際の翻訳処理を行う言語処理システムと一般利用者との間のインターフェイスシステムであるとともに、翻訳に伴って発生する数々の処理機能を高い操作性で提供することを目的として開発されてきた。

総合システムは、すでに報告した基本設計に基づき日英と英日両方向の翻訳システム開発を行い、両システムが61年3月に稼働しはじめた。ここでは、既報の基本設計に加えられた新しい考え、機能を中心に報告する。

## 2. 基本的考え方

### 2.1 翻訳システムにおけるマンマシンインターフェイス

ここでは、機械翻訳システムの使い易さについて、つまり利用者の立場にたったマンマシンインターフェイスという点について言及する。

将来的には、音声入力や文字認識技術を用い利用者が全くコンピュータ操作を意識しないでよいシステムが求められるだろうが、現実にはスイッチ一つというわけにはいかない。ある程度の操作は必要だという認識の上に立ち、いかにスムーズに会話をしながら処理を進めていけるかが、使い易さのポイントとなる。

#### (1) 機械翻訳システム利用時の作業の流れ

まず機械翻訳システムを利用して翻訳をするときに必要となる処理の流れを図2.1に示す。

翻訳したい文献をテストの翻訳する過程が、前処理である。翻訳対象の文献にあらわれるキーワードがその分野特有の訳語をもっていたり、未知語が多い場合に辞書の整備が必要になる。また、利用者によってはより正確な訳語を得たいという希望があり、そういう場合に文献特有の言い回しをシステムに合う形(システムが翻訳しやすい表現)に編集しなくてはならず、そのための準備がいる。

つまりここは、翻訳対象文を翻訳システムへ少しでも近づかせる段階といえる。尚、この作業は必ずしも必要ではない。

翻訳したい文献の翻訳を実行し出力する過程がメイン処理である。

大きな流れとしては原文テキストの入力、翻訳の実行、訳文の編集、印刷あるいは、他のファイルへの出力である。

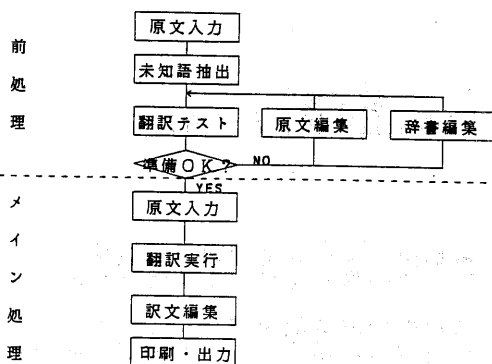


図2.1 翻訳作業フロー

以上から、機械翻訳システムに必要な機能は

- i 原文の入力 (一括入力、翻訳できるファイル形式への変換、文の切り出し、等)
- ii 未知語の抽出
- iii 辞書の編集
- iv テキストの編集
- v 印刷、他ファイルへの出力
- vi 翻訳の実行 (即時翻訳、一括翻訳)

となり、さらに各処理がスムーズに連携できている必要があろう。

#### (2) 使いやすさの基準

機械翻訳システムは発展途上のシステムである。開発・研究者が日々努力して、そのシステムの質、機能の向上を目指している。開発・研究者は、翻訳の各ステップの解析ツリーや他の情報とともに、システムがどのように処理を進めてきたのか、辞書の内容のどの部分に問題があったのか等、調査分析を行うためにシステムを利用する。

ところが、一般利用者にとってみれば、翻訳の途中経過などはどうでもよく、辞書編集にしてもあまり深いレベルまで求めないのが普通である。また、同じ一般利用者でも、ただ単に翻訳を行う者もいれば、なるべく質のよい訳を得ようと、辞書の整備や原文の手直しをする者もいる。このようにシステムに対する要求が異なる各々の利用者に対して使いやすいマンマシンインターフェイスを構築するのは、難しい問題である。

当システムは、主に一般利用者を対象者としつつ、言語処理システムの開発研究者にも必要な情報が提供できるhybridな翻訳システムの構築を目指した。

\* 本研究は、国の科学技術振興調整費による「日英科学技術文献の速報システムに関する研究」の一部として行ったものである。その遂行のために作業分科会を組織し、その審議のもとに研究を進めた。

### (3) 辞書の更新、保守性

辞書は訳文の質を大きく左右する要素の一つである。とくに言語処理システムの中核であるプログラムや文法規則を修正することが許されない一般利用者にとって、翻訳の質を上げる唯一の道が辞書の修正である。となれば、翻訳システムを使うにつれて、利用者は自分にあった辞書内容に作り変えて行くのは当然であろう。

反面、利用者の辞書修正が訳文の質を下げることもある。それは、システムが標準の辞書として提供する基本辞書は、出来るだけ汎用性を考慮して集められており、辞書内容も十分考慮されているが、利用者の辞書編集はいま翻訳している文のみについて考えがちで、他の場合に応用できるとは限らないからである。一方、利用者の辞書編集結果は、今後基本辞書を拡充していく上の重要な資料になりうる。私用辞書の有用なデータを基本辞書に取り込んでいけるような情報を残しておくことも必要であろう。

## 2.2 総合システムの開発に当たって

総合システムの開発に当たって、特に以下の点を考慮した。

### (1) 即時翻訳、テキスト編集

開発・研究者には、翻訳の結果のみでなく、途中経過を知りたいという要求がある。そのため、翻訳しながらその時々翻訳木や他の情報をトレースする機能、翻訳した結果を参照する機能を持たせる。

また、原文訳文を同時に編集できるテキストエディタは、従来の機能のほかに文を対象とした移動や複写等の機能を設け、より強力なものとする。さらに、エディタの機能のひとつとして、未知語抽出ができるようにする。これは前処理において辞書の整備をする際、必要となる機能である。

### (2) 辞書体系

#### ① 処理用辞書とMTフォーマット

利用者が編集の対象とする辞書は、MTフォーマットと呼ばれ、訳語、文法的働き、意味といった通常の辞書内容が、解析、変換、生成の三種類の辞書に収められている。一方、言語処理システムが翻訳に用いる辞書は、訳語、文法的働き、意味といった通常の辞書内容のほかに局所的な文法規則として辞書ルールを各単語に付け加えている。また、処理効率を考え、翻訳の各過程ごとに使用する情報だけを集め、言語処理システムが参照しやすい内部表現で書かれている。これを処理用辞書と呼び、形態素解析、解析、変換、生成、形態素生成の五種類からなりいずれもMTフォーマットから生成されている。

#### ② 基本辞書と私用辞書

基本辞書は、システムが提供する汎用的な辞書であり、現在は

基本語と電気工学を主とする科学技術用語が収められている。これに対して、利用者自身が作成する辞書を私用辞書という。

汎用性、バランスを考えてつくられた基本辞書を各利用者が原文の特異性によって個々に修正してしまうと、基本辞書は一定の品質を保持し得なく成る。そこで利用者からは辞書をどんどん更新しているように見え、基本辞書そのものは変更されないよう、私用辞書の考え方を取り入れた。利用者が辞書修正した結果は、すべて私用辞書に格納される。これにより、基本辞書の汎用性が保たれる。そして私用辞書と基本辞書を組み合わせると、全体として最適一つの辞書に見える。

### (3) 辞書編集

利用者が辞書編集の対象とするMTフォーマットは、LISPのS式構造で記述されている。開発者やS式に慣れた者にとってS式のままで編集の方が使い易いが、一般利用者にとっては、編集しづらいであろう。そこで、S式で書かれた内容を表の形に展開して辞書修正する方法を採用した。

また、辞書体系のところでも述べたように、利用者が編集の対象とするMTフォーマットは三種類ある。そのため、一つの語を編集するときに三種類のMTフォーマットをいかに効率よく、正確に編集できるかは重要なポイントである。また、ある一つのMTフォーマットのみを修正したい場合に備えて、個々のMTフォーマットを編集する機能もある。そのため、三つのMTフォーマットを連続して編集する系統的辞書編集と個々のMTフォーマットを編集する個別辞書編集の機能を設ける。

## 3. システム構成

### 3.1 システム概要

言語処理システム、辞書システム、総合システムの関係と総合システムのサブシステム構成を図3.1に示す。

総合システムの各サブシステムの概要を次に示す。

#### (1) 即時翻訳・テキスト編集

原文訳文を対応付けて画面上に表示し、随意に編集できる。ここから言語処理システムの翻訳関数を起動し即時翻訳を行うことが出来る。インタラクティブに翻訳・編集を繰り返す作業は当機能で行う。

#### (2) 一括翻訳

翻訳処理をバッチ処理で行う機能である。

#### (3) 辞書編集

辞書の内容を対話的に編集する。編集結果は、すべて私用辞書に反映される。

#### (4) 入出力テキスト変換

外部日本語エディタ、英文エディタおよび文献検索システムとのインタフェースにあたり、それら固有のファイル形式・文章形式の入出力を可能とする。

入力変換は、原文ファイルの文章を一文づつ切り出し、翻訳ファイルへ格納する。

また、出力変換は、翻訳ファイル上の訳文を指定された訳文ファイルへ格納する。

#### (5) 会話制御

言語処理システムと総合システム間の制御や会話処理制御、一括翻訳依頼、ライブラリ管理機能を持つ。

#### (6) イメージ辞書

市販の英和辞典の電子化辞書システムで、参照、表示、印刷の機能を持っている。当機能は、オプションで総合システムとは独立に起動する。

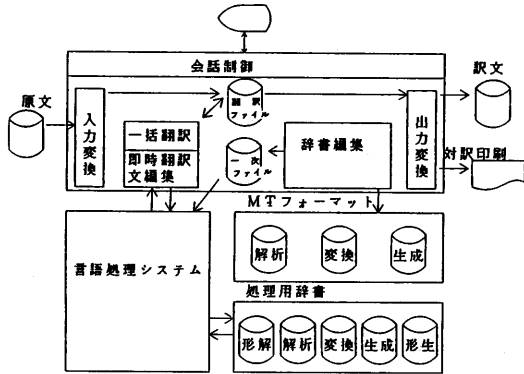


図3.1 システム関連図

### 3.2 言語処理システムとのインターフェイス

総合システムは、翻訳システムの核である言語処理システムと利用者を結ぶマンマシーンインターフェイスの性格を持つ。

言語処理システムの翻訳関数や辞書作成関数等は、UTILISPで記述されている。一方、総合システムは、プログラムの開発効率、運用管理、保守等を考慮しPL/Iを用いて開発を行った。総合システムは、翻訳を行ったり、辞書の作成を行う際に、言語処理システムを呼び出さなくてはならない。UTILISPには、PL/Iとのインターフェイスはあるが、PL/IからUTILISPを外部サブルーチンとして呼び出す機能はない。そこで両者を別タスクとして別々に起動し、同期をとりながら、並列的処理を行う方式を採用している。

### 4. システムの機能

総合システムは、すでに示したように大きく五つの機能を持つ。ここでは、すでに報告した基本設計の段階から機能強化改善された点を中心に機能を説明する。

#### 4.1 テキスト編集と即時翻訳

翻訳をインタラクティブに行い、原文訳文もその場で編集できるのが当機能で翻訳エディタと呼んでいる。翻訳エディタで重要なのは、いかに文章編集の行いやすい編集機能を持つかという点と、いかにスムーズに言語処理システムに翻訳処理を渡して結果をもらうかということである。

##### (1) テキスト編集

文章の編集で重要なものの一つは、テキストエディタの表示構造と機能である。原文訳文がいかに見やすくなっているか、コマンドの入力域はテキスト編集の際に効率が良い位置にあるかなどは、エディタ画面の設計におうところが多い。

##### ① 文章の表示

文章は、文単位に文番号付きの形式で表示している。原文を上部に訳文を画面下部に表示し文番号で対応付けを行って見やすくしている。翻訳や編集の際、文番号は大きな動きをする領域である。ここでは、各種文コマンドが入力でき、また文の分割や結合削除などの文番号変更に及ぶような編集の際にも原文と訳文の対応をとっている。

##### ② 禁則処理

表示の際、和文、英文として行頭または、行末に来てはいけない文字を制御する機能でこれにより和文、英文が見やすく、作業者に違和感のない表示が可能となった。とくに次の4点に注意を払った。

- i 一行に収まらない単語があった場合。
- ii 一行に収まらない単語の途中にハイフンがある場合。
- iii 日本語中に英文がある場合。
- iv 日本語中の英文が一行に収まらない単語をもっている場合。

##### ③ 文の編集

英文、和文ともに文章表示域の文字列を直接編集できる。入力には、アルファベットまたは、カナをタイプインしてファンクションキーによるカナ漢字変換による。以下の機能を持つ

- i 文字の置換、挿入、削除
- ii 文の移動、複写、削除、繰り返し
- iii 文の結合、分割
- iv 文字列の検索・変更
- v 対訳文の印刷

## (2) 即時翻訳

翻訳を会話的に進めていけるのが即時翻訳である。翻訳エディタには、そのほか、辞書の充実や翻訳過程のチェックのための翻訳トレース表示機能がある。

### ① 即時翻訳

翻訳したい文章の文番号に翻訳コマンドを入力するだけで言語処理システムに翻訳の依頼ができ、翻訳終了後は、訳文表示域の対応する個所に訳文が表示される。

### ② トレース付き翻訳

開発・研究者にとって入力した文に対する翻訳結果も重要であるが、なぜこう訳したのかという経過がより重要である。そうしたとき翻訳がどうすすんでいるか、翻訳ツリーはどのようになっているのか、辞書内容はどのように参照されているか等を知るために当機能を作成している。

### ③ 未知語抽出

文の中に未知語があると翻訳が正しく行えない。翻訳の前にあらかじめ未知語を洗いだしておけば、便利である。そこで指定された文を形態素解析し、その中に未知語があれば、訳文表示域に未知語を表示する機能を作成している。

形態素解析辞書には登録されていても、処理用辞書のいずれか一つにでもその語が登録されてなければ、やはり正しい翻訳結果は得られない。それぞれの辞書に存在するか否かをチェックするために、翻訳後に未知語を抽出する機能も設けている。

### ④ 辞書エディタの呼び出し

翻訳作業中、辞書の内容を直接修正、参照したいことが多い。翻訳エディタをいったん終了した後、辞書エディタを起動するのではなく、翻訳エディタから直接辞書エディタを呼び出す機能を有している。

## 4.2 一括翻訳

作業者は一文一文翻訳結果を確認しながら翻訳作業を進める場合、即時翻訳を利用しその場で訳文の編集をしたり辞書修正を行う。一方、辞書修正及び、訳文の編集は後でまとめて行うから、多量の文章を一括して翻訳したいという要求も多いだろう。これを実現するのが一括翻訳である。一括翻訳は、言語処理システムをバッチで起動する。多量の文章の翻訳がバッチ処理で行われている間に、利用者は前に翻訳した文章の編集などの別の仕事ができる。

## 4.3 辞書編集

辞書編集を容易にするためのツールを辞書エディタと呼ぶが、どれだけ辞書内容や記法を熟知しているか、どれだけエディタの操作に慣れているか等により最適なエディタの形や機能に大きな

差が生じる。総合システムでは、利用者のレベルを考慮して二つのタイプの辞書エディタを作成した。利用者は、自分に適したエディタを選んで、自由に編集作業を行える。

### (1) 辞書編集データの流れ

システムで用いられる辞書の体系については、2.2ですすでに説明した。図4.1にこれら辞書の関連と辞書編集の際の辞書データの流れを示す。尚、一次ファイルは修正された内容を辞書作成関数に渡すための作業ファイルである。

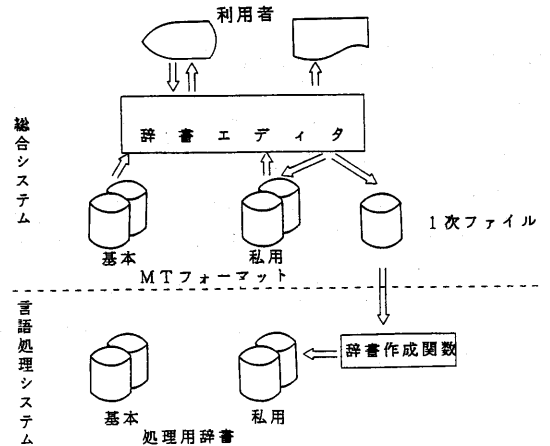


図4.1 辞書編集作業におけるデータの流れ

### (2) 表形式辞書エディタ

編集の対象となるMTフォーマットは、図4.2にあるようにLISPのS式の構造をもつ。この表現は一般利用者には理解しにくく、また、自由な形式であるが故にかえって辞書について詳しい知識がいる。

```
( (SEQ 1)
  (J_LEX 翻訳する)
  (J_CAT 動詞)
  (USAGE
    ( (J_SURFACE_CASE
      (J_SURFACE_CASE1 が)
      (J_SURFACE_CASE2 を)
      (J_SURFACE_CASE3 に))
      (CASE-PATTERN V1)
      (J_DEEP_CASE
        (J_DEEP_CASE1 主体)
        (J_DEEP_CASE2 対象)
        (J_DEEP_CASE3 終状態))
      (CONDITION ((E_LEX translate)))
    )
    (A
      (E_LEX translate)
      (E_CAT V)
      (E_SURFACE_CASE
        (E_SURFACE_CASE1 SUBJ)
        (E_SURFACE_CASE2 OBJ)
        (E_SURFACE_CASE3 = int.o))
      (CASE-PATTERN V1)
      (E_DEEP_CASE
        (E_DEEP_CASE1 AGT)
        (E_DEEP_CASE2 OBJ)
        (E_DEEP_CASE3 GOA))
      (CORRESPONDENCE
        (CORRESPONDENCE1 主体)
        (CORRESPONDENCE2 対象)
        (CORRESPONDENCE3 終状態))))))
```

図4.2 MTフォーマットの内容

また入力すべきデータが多く、かつ辞書内容の記入パターンが辞書の種類、品詞ごとに異なるため、一般利用者が編集作業するには、かなり労力を必要とする。そこで一般利用者にも編集しやすいよう開発したのが表形式辞書エディタである。当エディタは、S式構造の内容を表の形に展開し、その上で編集された結果を再びS式構造に変換してMTフォーマットに蓄積している。(図4.3)

表形式辞書エディタ 日英変換辞書 助詞 1>  
見出し語: 翻訳する  
コマンド --->

SEG	1	J SURFACE CASE1	J SURFACE CASE2	J SURFACE CASE3	J SURFACE CASE4
(1)	21				
(2)					
(3)					
(4)					
(5)					
CASE PATTERN		(1) (2) (3) (4) (5)			
J DEEP CASE1		J DEEP CASE2	J DEEP CASE3	J DEEP CASE4	
5 16					
(2)		J SURFACE CASE1	J SURFACE CASE2	J SURFACE CASE3	J SURFACE CASE4
(1)					
(2)					
(3)					
(4)					
(5)					
CASE PATTERN		(1) (2) (3) (4) (5)			
J DEEP CASE1		J DEEP CASE2	J DEEP CASE3	J DEEP CASE4	

図4.3 表形式辞書エディタ

① 系統的編集と個別編集

一つの単語を編集する際には、解析、変換、生成の三つのMTフォーマットを編集しなくてはならない。例えば、解析辞書を修正しても、他の辞書にその修正状況が反映されなければ、正しい翻訳結果を得ることができない。いくら各個人用の辞書とはいえ、翻訳結果に悪影響を及ぼすような辞書編集を利用者にさせてしまうようでは使いがたのよいシステムとはいえない。

そのため解析、変換、生成の順で連続的に辞書修正を行う系統的辞書編集を表形式辞書エディタで実現した。利用者は、辞書編集したい単語を指定しさえすればよい。エディタは、解析辞書を辞書引きし、編集が終わったら、辞書内容に対して必須項目が記述されているか、文字の属性は正しいか等のチェックを行う。解析辞書が終わったら次に変換辞書を同じエントリーで辞書引きし、編集が終了したら編集内容のチェックをする。生成辞書の編集では、変換辞書の訳語をエントリーとして辞書引きを行い、同様のことを続ける。訳語が複数ある場合は、すべての訳語について生成辞書の編集は繰り返される。最後にまとめて編集結果を保存する。

これに対して、一つの辞書内容のみを修正すればよいとわかっている場合がある。その場合に編集作業が簡単に行えるように、個々の辞書を編集する個別辞書編集も作成した。言語が進えば品

詞は必ずしも一致していない。例えば、編集対象として指定した語が日本語の助詞のように英語側に対応する品詞がないような語では、系統的編集はできない。このような品詞語を修正する場合にも個別辞書編集は使用できる。

② 編集機能

当エディタでは、さらに使い勝手をよくするため次のような編集機能を有している。

i 編集結果の反映

編集した結果を私用辞書に追加、更新、削除する。これは、基本辞書を一般利用者に修正させないためである。

ii 複写機能

コマンドによって他の見出し語の内容を編集画面上に複写することが出来る。これは、辞書内容が類似した見出し語を複写することによって辞書の新規追加作業を軽減するためのものである。(但し、表が辞書の種類と品詞によって異なるため、同種の辞書、同一の品詞に限られる。)

iii 参照機能

コマンドによって他の見出し語の内容を画面分割した下側に表示する。これはあるMTフォーマットの編集集中に別の種類のMTフォーマットを参照したり、同じMTフォーマット中で別の語を参照したい時に利用できる。

iv 戻り機能

系統的辞書編集のとき解析・変換・生成という順で編集が進むが、前の辞書に戻って修正をやり直したい場合がある。そのときこの機能によって戻ることが出来る。

v デフォルト機能

新しい語を登録する場合、辞書の入力項目が多いため慣れない利用者は戸惑ってしまう。そこで見出し語を指定さえすればとりあえず翻訳が進められるように、辞書内容の必須項目に省略値が自動的に取られるようにした。

vi 辞書内容管理機能

私用辞書の見出し番号という項目に自動的にシーケンシャル番号を振る。これは私用辞書データの管理に役立つ。

(3) S式辞書エディタ

表形式辞書エディタは、辞書内容を表の形にするので見やすいが、繰り返し項目や入力文字数などの制限があり自由度がない。またLISPに慣れた者にとって、ごく一部の修正にはS式に直接手を入れる方が便利なこともある。

このため、MTフォーマットのデータを直接画面に表示して編集するS式辞書エディタを当システムは設けている。

当エディタは、LISPのS式の構造で書かれている内容を、編集画面のデータ表示域上に見やすい形にして表示する。利用者はこの領域上で辞書内容を編集するのである。表の形からくる長さや繰り返し回数の制限がないため自由度の高い編集ができる。

また当エディタは、個別編集形式であり、各辞書間の整合性は、利用者に任されているので開発・研究者向けのエディタといえる。

#### 4.4 入出力テキスト変換

翻訳の対象となる文章は、外部で作成された文献、マニュアルなどマシーンリーダブルな文書である。また翻訳した結果を外部で流用することが多いだろう。当テキスト変換は、総合システムの動作環境で利用されるファイルと外部ファイルとのファイルインターフェイスである。

##### (1) 入力変換

入力変換は、原文ファイルから一文を切り出し、言語処理システムの要求する形で、翻訳ファイルの原文部に原文を格納する機能を果たす。以下に特徴的な点を示す。

##### ① 文の切り出し

この言語処理システムでは、文書から一文を切り出すのが翻訳の第一歩で、一文の認識を誤ると正しい翻訳はできない。そこで、入力テキスト変換は、利用者が指定した文区切り記号により文を判断し一文づつ切って格納する。

##### ② 翻訳するか否かの判断

本翻訳システムの利用対象の大きな柱として文献が上げられている。文献は翻訳対象となる表題・抄録の他に種々の書誌的事項を含む。これらは翻訳をしたとしても、意味のある翻訳結果が得られるかどうか不明である。かえって、言語処理システムに負荷をかける事になる。そのため、入力変換では抄録データに限って、項目ごとに翻訳するしないの判断を行い、各文にその旨の情報（翻訳フラグ）を付加することにした。翻訳エディタでは翻訳コマンドが入力されると、翻訳フラグがTの文だけを言語処理システムに渡す。

ところで、上のように翻訳しないと判断された書誌的項目でも、訳文には組み込みたい。しかし、言語処理システムを通らないので訳文は出ない。この点を解決するため、入力変換は翻訳しない（翻訳フラグをNILとをした）文については、同時に翻訳ファイルの訳文部にも同じ文を書き込み、原文と同じ情報がすべて伝わるようにした。

##### ③ 制御コード

外部エディタにより作成された文章にはフォーマットのための制御コードが含まれる。外部エディタの中でのみ意味を持つ制御コードを含む文を翻訳しても、正しい結果は得られない。そのため、入力変換では制御コードを翻訳対象から除外している。外部エディタの制御コードが変更されることも考え、制御コードのテーブルを外に持ち、テーブルにある制御コードを削除するようにしている。

##### (2) 出力変換

翻訳の結果は、翻訳ファイルに格納される。出力変換は翻訳ファイルの訳文だけを取り出し、外部のファイルへ出力するというファイルインターフェイスにあたる機能を持つ。また、翻訳ファイルの対訳を印刷させることもできる。

##### (3) 文字コードの変換

世の中には多くのコード系が存在する。システムの中でいくつかのコード系を持つとその処理が煩わしく、一つに統一したほうがよいであろう。また、日本語に関する翻訳システムでは漢字コードを必要とし、英文にしても日本語を翻訳した場合に未知語として日本語が出力される。そのため、当言語処理システムでは漢字コード系で処理を統一しており、テキスト変換はそれに対応している。

#### 4.5 会話制御・その他

利用者との対話処理を制御するマンマシーンインターフェイスの核となる部分であり、言語処理システムと総合システム間の制御、一括翻訳や辞書エディタなどのサブシステムを呼び出す機能を持つ。会話処理としてガイダンス機能やライブラリ管理機能も備えている。

マンマシーンインターフェイスとして有効な他の機能を挙げる。

##### (1) 辞書の創成・設定

利用者が個人用の辞書を持って利用したいとき使用する機能である。創成は、私用辞書を定義初期化する機能であり、設定は、翻訳や辞書編集で使用する私用辞書を指定する機能である。辞書は、MTフォーマットと処理用辞書の2種類があり、さらにそれぞれ3、5種の辞書がある。それを利用者が一々命名していたのでは大変であり、辞書管理の面からいっても面倒である。そのため、各MTフォーマット名を利用者が指定し、それに対応した処理用辞書名をシステムが自動的に命名するようにしている。

利用者は、MTフォーマットを辞書として意識しさえすればよく、処理用辞書は利用者からは見えなくなっている。

##### (2) TSS

総合システムを使用している最中にファイルの属性を調べたいとか、一括翻訳で流したジョブがどういう状態であるかなど調べたいときがある。総合システムを一端終わってまた起動するのは、作業効率上好ましくない。対話処理中にTSSのコマンドが発行できれば、対話処理システムとしての幅が広がる。そのため、TSSコマンドを総合システムから発行できる機能を付加している。

#### 4.6 イメージ辞書

##### (1) 目的

イメージ辞書とは、市販の辞書をディスク上にイメージデータ化した電子辞書で、弾力的な検索能力で辞書引きを行い、ディスプレイ上に辞書内容をイメージの形で表示あるいは印刷する機能を持つシステムである。辞書エディタで辞書を修正する場合、あるいは、訳文をチェックする場合などは、普通の市販の辞書を見たいことがある。イメージ辞書は、このようなときに、本のページをめくる代わりにイメージ化した辞書を計算機で引くためにつくられたものである。

(2) 構成と機能

イメージ辞書の構成を図4.4に示す。

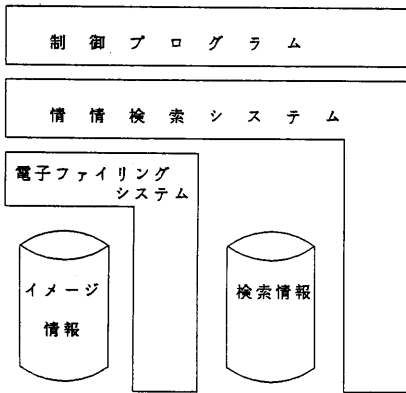


図4.4 イメージ辞書の構成

イメージ情報（一次情報）は、「大きな活字の新コンサイス英和辞典」（三省堂）の内容をページ単位で電子ファイリングシステム（ELF）のイメージデータ蓄積機能によってディスク上にデータベース化したものである。

検索情報（二次情報）は、イメージデータを検索するための各種情報を整理したデータベースで、対話型情報検索システム（FAIRS-I）のデータ蓄積機能で作成したものである。

イメージデータと検索情報とは、辞書のページをキーとして結合されており、ある二次情報がFAIRS-Iの機能で検索されたとき、それと結び付けられたイメージデータがELFの機能によって引き出され、表示・印刷されるのである。現段階では、イメージ辞書は、総合システムのオプション機能として別システムとして起動される。将来的には、一つのディスプレイ上で、翻訳や辞書編集、テキスト編集をしながら参照できるような形が

望まれる。

5. あとがき

総合システムは、日英・英日の両システムを提供している。英日システムの作成に当たっては、先に完成した日英システムのモジュールを積極的に共通利用するように務め、作成期間の短縮化と保守の容易性をはかった。

言語処理システムを含めた総合システム全体では、実用レベルの翻訳システムとして基本的な機能を有している。しかし、さらに実用性を考えた場合、今後検討すべき課題が多く残されている。以下に2点示す。

(1) 総合システムでは、現在、大型計算機上で8MBの領域を用いてシステムを動かしているが、さらに3倍程度の領域が実用化のために必要であり、またディスクの容量、アクセス速度の改善が求められる。

(2) 利用者の使い勝手の面から、マルチウインド機能をもち、単体として強力なエディタ機能、ファイル管理機能を有するワークステーション的なインテリジェント端末の開発とその有効利用が求められる。

最後に、本プロジェクトは過去4年間にわたって京都大学長尾真教授の指導のもと、科学技術庁、通産省の援助、各種委員会での専門の先生方の指導、助言、さらにソフトウェア、文法、辞書の作成を担当して下された京都大学の辻井潤一助教授、中村順一助手、JICSTの佐藤雅之氏ほか、多くの方々との協力により、研究、開発が進められたことを記し、感謝致します。

6. 参考文献

- (1) 坂本義行, 有賀妙子: Muプロジェクトにおける総合システムの設計  
自然言語処理研究会46-6 (1984)
- (2) 長尾真: 科学技術庁翻訳プロジェクトの概要  
自然言語処理研究会38-2 (1983)
- (3) 坂本義行, 有賀妙子, 鷲尾敦: Unified Operation Control System for Mu Machine Translation  
International Conference on STATISTICS OF THE ART IN MACHINE TRANSLATION in America, Asia and Europe and MT-SYSTEMS FAIR 投稿中 (1986)