

機械翻訳総合システムの基本設計

天田光治 (電子技術総合研究所) 長尾 真 (京都大学)

1. はじめに

日英科学技術文献の速報システムにおける翻訳プログラムや翻訳辞書を、有機的に結合して実用するための機械翻訳作業環境を、総合システムとして設計したので報告する。

翻訳プログラムや翻訳辞書を補いながら実用するためには、前編集や後編集が有効な手法であり、これらを充分とり入れることが必要とされている。このため本システムでは、これらの手法を活用するしかけを考慮するとともに、応用分野をいくつか想定し、メニュー方式などを採用している。たとえば、文献検索における機械翻訳の活用や、文書作成における高度利用である。このようなアプローチについては、すでにいくつかの成果を得ている。

機械翻訳作業システムの普及は、ハードウェアのつくりにも依存することが大きいので、大型計算機、ミニコンピュータ、パーソナルコンピュータにおいてもインプリメントできるように設計した。また、ビットマップディスプレイやイメージ文書ディスプレイの活用も期待し、マルチウィンドウやイメージ辞書の取扱いについても考慮している。

また、専門分野別用語の蓄積も重要である。このため、各ユーザが定義して使用した用語については、ユーザ毎の個別辞書の作成を可能にし、これらを自動的に吸い上げる機能をもっている。つまり、辞書の階層構造をとることによって辞書の蓄積をはかれるようにしている。

以下に、機械翻訳総合システムの基本設計の概要を示す。

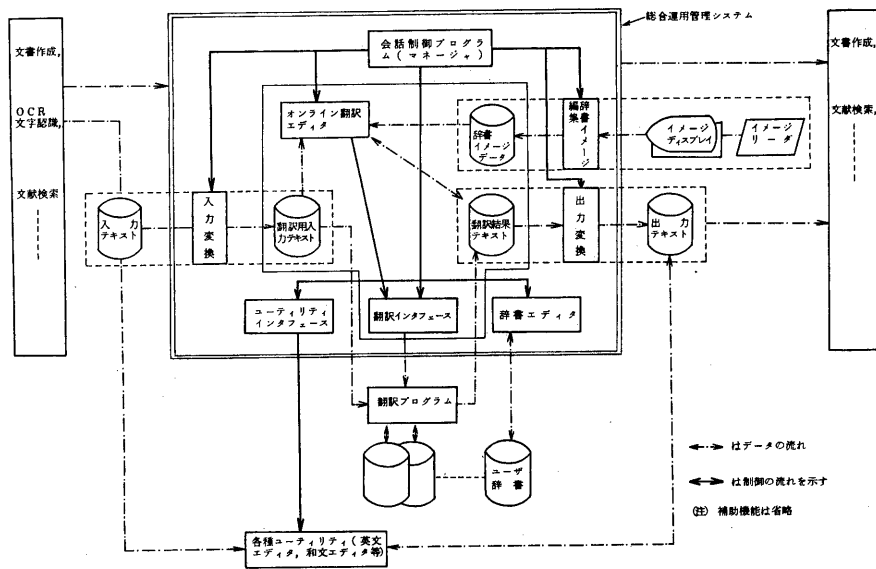


図1 総合運用管理システムの概念図

2. システムの分析

2.1 ユーザ要件

ユーザが使用する「日英科学技術文献の速報システム」を有機的に統合し、共通資源の一元管理、各種ファイルの対話的な編集機構、各コンポーネント間の主要なインタフェースの標準化などにより、翻訳システムの多面的な利用を可能にするためのシステムの全体的な方式を設計する。

2.2 速報システム内の位置付け

総合運用管理システムは、京都大学と電子技術総合研究所で開発している翻訳プログラムと翻訳用基本辞書とJICSTで開発している翻訳用分野別辞書を使用して翻訳サービスを行うために、速報システム全体の管理及び情報検索システムや文書作成等のアプリケーションと翻訳システムとの結合を行う。総合運用管理システムの概念図を図1に示す。

2.3 利用形態

総合運用管理システムは以下の2通りの利用方法がある。

- (1) 文献検索システム作成/使用時の翻訳サポートは、データベース作成時に表題や抄録等を翻訳する、選択的配布(SDI)利用時に表題や抄録等を翻訳する、遡及検索(RS)で検索結果の表題や抄録等を翻訳するなどである。
- (2) 文章作成(論文作成)は、英文エディタや和文エディタで作成した英語文章や日本語文章を翻訳する。

2.4 総合運用管理システムに必要な機能

ユーザ要件及び翻訳システムの現状を考慮して総合運用管理システムの機能を以下の通りに定める。

- (1) 翻訳入力文章の前編集は、翻訳する前に入力文章に手を入れて翻訳しやすくする。
- (2) 翻訳結果文章の後編集は、翻訳後に、翻訳に失敗した場合や言い回しがおかしい場合に翻訳結果に手を入れて完全な文章を作る。
- (3) ユーザ辞書の作成・修正は、ユーザ毎の私用辞書を作成したり修正したりする。
- (4) 文章毎の辞書の指定は、文章毎にその文章の翻訳に使用する辞書及びその検索順序を指定する。
- (5) 英文エディタ、和文エディタ、データベースシステム(文献検索システム)等とのインタフェース機能は、翻訳システムの多面的な利用を可能にするために、英文エディタや和文エディタ等で作られた文章を翻訳できる形式に、あるいは翻訳結果を英文エディタや和文エディタ等で取り扱える形式にする。
- (6) 一括翻訳は、英語や日本語の文章をファイル単位で一括して翻訳する。動作形態としてはバッチ処理が主体となる。
- (7) ファイル管理は、「日英科学技術文献の速報システム」で使用する共有資源を管理する。

3. システム方式

3.1 システム構想

- システム分析をもとに総合運用管理システムをサブシステムに分割する。
- (1) 会話制御プログラムは、総合運用管理システムを統括管理する。端末とのやりとり(入出力)を制御するものでメニュー方式、フルスクリーン・マルチウィンドウ方式をとる。
 - (2) 入力変換は、英文エディタや和文エディタで作成された文章や文献検索システムで検索された文章等を翻訳用の文章形式にする。
 - (3) 出力変換は、翻訳結果を英文エディタや和文エディタで取り扱える形式やあるいは文献検索システム用のデータの形式に変換する。
 - (4) オンライン翻訳エディタは、英語の文章や日本語の文章を対話的に翻訳することのできるフルスクリーンエディタで、原文と訳文が画面の上下に分割し、対訳表示する。また、用語の内容(訳や例文)をイメージデータの形式で画面に表示する。
 - (5) 一括翻訳は、英語や日本語の文章をファイル単位で一括して翻訳する。
 - (6) 辞書エディタは、ユーザ用の私用辞書を作成・修正するフルスクリーンエディタである。用語を対話的に編集(作成、修正)したりする他に、辞書内での用語の複写・移動等も行える。
 - (7) 端末定義は、総合運用管理システムで使用する端末の定義を行う。これは、TTY接続端末の特殊コード(ローカルデリート、ローカルインサート等)、TTY接続端末の擬似PFK(プログラムファンクションキー)、文字種(カタカナ、英小文字)、PFKの機能コードなどを定義できる。
 - (8) 翻訳アロファイル定義は、翻訳する文章ごとに使用する辞書名及びその辞書の検索順序、オンライン翻訳エディタで文章から文を切り出す際に使用する区切り文字を定義できる。
 - (9) イメージ編集は、イメージディスプレイとイメージスキャナを用いてオンライン翻訳エディタで使う辞書イメージデータ(用語)を作成したり削除したりする。
 - (10) ファイル管理は共用辞書、文法規則(ルール)、ソースプログラム、翻訳基準データを一元管理する。

3.2. ハードウェア環境

3.2.1 前提条件

総合運用管理システムは汎用大型計算機システムミニコンピュータ、ワークステーション、パーソナルコンピュータなどを使用する。

3.2.2 ワークステーションの概念

総合運用管理システムでは使用するハードウェアを分散処理の概念(ワークステーション、サーバ、プロセッサ)でグループ分けする。以下に個々のワークステーション、サーバ、プロセッサ及びそのハードウェア構成を示す。

- (1) 翻訳ワークステーションはユーザが翻訳サービスを受けるところで、文字処理翻訳ワークステーションは文献検索や文書作成の行えるワークステーションで、日本語ディスプレイ、日本語ハードコピーからなる。イメージ処理翻訳ワークステーションは、文字処理翻訳ワークステーションの機能の他に、イメージ辞書作成や文字認識(イメージ)を行えるワークステーションでイメージディスプレイ、イメージスキャナ、ハー

ドコピーからなる。

- (2) 清書出力サーバは、翻訳結果を清書出力するところで日本語ラインプリンタからなる。
- (3) ファイルサーバは、翻訳入力文章や翻訳結果を保存するところで、DASDおよび光ディスク、レーザカードからなる。
- (4) 翻訳プロセッサは、実際に翻訳処理を行うところでCPU、文法規則等が格納されているDASDおよび光ディスク、レーザカードからなる。
- (5) 辞書プロセッサは、翻訳に使用する辞書を辞書引きするところでCPU、翻訳用辞書が格納されているDASDおよび光ディスク、レーザカードからなっている。
- (6) データベースプロセッサは、科学技術文献の文献検索を行うところでCPU、文献検索システムで使用するシステムファイルが格納されているDASDおよび光ディスク、レーザカードからなる。

4. 編集ソフトウェア

4.1 目的

編集ソフトウェアは、オンライン翻訳エディタ、辞書エディタ、辞書イメージ編集からなり、よりよい翻訳結果を得るために文書(原文・訳文)、辞書、辞書イメージの編集を行うことを目的としている。

4.2 特徴

このサブシステムは、原文を与え、それを画面に表示し、かつ翻訳プログラムによって出力された訳文を画面に表示する機能を備え、それらの文章の画面上での、各種コマンドによる作成・修正を可能としている。また辞書イメージの表示・編集及び辞書の編集を対話的に行う機能を有している。そのために翻訳、辞書引き等を含む文章編集を画面上で実現するのに以下のような特徴を持っている。

- (1) 文章の作成・修正において日本語文章と英語文章の両方を同じコマンドで統一的に扱える。
 - (2) 原文の部分的(文単位)翻訳が、対話的に行える。
 - (3) 辞書を対話的(メニュー方式)に作成・修正することができる。
 - (4) イメージディスプレイを用いて、辞書イメージの表示、編集ができる。
- オンライン翻訳エディタコマンドの例を表1に示す。

5. ファイル管理システム

5.1 目的

以下の各データ等について一元管理するシステムを設計する。

- (1) 辞書データ
翻訳システムで使用する辞書でかつユーザが共用する辞書、基本辞書、トランスファ用辞書、分野別辞書等の辞書で私用辞書は含まない。
- (2) 文法規則
翻訳システムが使用する文法規則の集合、ルールともいう。
- (3) ソースプログラム
翻訳プログラムであるLISPソースプログラム(コンパイルした形式のプログラムは除く)や総合システム(総合管理システム、ファイル管理システム、編集ソフトウェア)のソースプログラム(高級言語、LISP、

表1 オンライン翻訳エディタコマンドの例

コマンド名 (データ名)		機能
文章作成	表示・検索用コマンド	TOP 原文、訳文とも文章の先頭より表示する。
	BOTTOM 原文、訳文とも文章の最後より表示する。	
	UP 原文、訳文の表示を先頭文方向へ移動させる。	
	DOWN 原文、訳文の表示を最終文方向へ移動させる。	
	FIND 指定された文字列を検索し、その文字列の存在する文より表示する。	
	SENTENCE センテンスモードでの文章の表示をする。	
	TEXT テキストモードでの文章の表示をする。	
修正機能	文章作成・修正用コマンド	CHANGE 文章中の任意の文字列を他の文字列に変更する。
	D 文章中の文字(列)を削除する。	
	IA IB 文章中に文字(列)を挿入する。	
	R R RS 文章中の文字(列)を他の文字(列)に変更する。	
	C C A(B) A(B) 文章中の文字(列)を複写する。	
	M M A(B) MS A(B) 文章中の文字(列)を移動させる。	
	その他コマンド	CC 文章中に入力するコマンドの制御コード('1')を切り換える。
SW 編集対象文章の切り換えを行う。		
SAVE 編集用ファイルの内容を元のファイルへ格納する。		
辞書検索修正機能	< > 指定した用語を辞書エディタにより修正する。	
翻訳機能	T 指定した文の翻訳を行う。	
辞書イメージ表示機能	I 指定した用語の辞書イメージを検索し表示する。	
	IDEL 表示されている辞書イメージを消去する。	

アセンブラ)等。

(4) 翻訳基準データ

文法規則や辞書データを修正した時に翻訳プログラムのチェックを行うためのテストデータ。

5.2 方針

当システムは以下の方針で設計する。

- (1) プログラムはソースレベルで管理する。オブジェクトやロードモジュールについては一切管理しない。ロードモジュール等が破壊された場合は、ソースプログラムからコンパイル、結合編集してロードモジュールを再作成する方針とする。
- (2) ソースプログラム、文法規則、翻訳基準データ等のソース形式(EBCDIC文字などで記述されている)のファイルは、ライブラリ管理システムで管理する。
- (3) 辞書の管理は、当システム最大の関心事であり、分野別辞書のメンテナンス及び拡張の機能をセンタユーティリティ、辞書エディタで実現する。

6. おわりに

本報告は、昭和57年度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」の一環として行われた「総合システムに関する開発」の基本設計である。これは研究推進委員会（長尾 真委員長）の作業分科会メンバーである、梅垣邦一（川崎製鉄）、柏木 寛、田中穂積（電総研）、中井 浩（J-ICST）、荻 一博（ICOT）、高濃立夫（日本コンベンション）、吉野 真（三井物産）の名氏の参加によって行われた。なお本計画は、科学技術庁の小路情報室長、八日市谷室長補佐、および工業技術院の平田計画課長、山岸課長補佐、新井良一課員の推進のもとに行われた。なお本研究は、辻井潤一（京都大学）、田中 隆（電総研）、柿元俊博（富士通）の名氏の協力におうところが多い。

参考文献

1. 長尾 真、機械翻訳、情報処理、Vol.20、No.10、1979.
2. 昭和57年度科学技術振興調整費、日英科学技術文献の速報システムに関する研究、総合システムの開発に関する報告書、工業技術院総務部計画課、情報計算センター、電子技術総合研究所。
3. 長尾、辻井、矢田、柿元：科学技術論文表題の英和機械翻訳システム、情報処理学会論文誌、第23巻第2号、昭和57年3月。
4. N.Nagao, J.Tsujii, K.Yada, T.Kakimoto, An English Japanese Machine Translation System of the Titles of Scientific and Engineering Papers, COLING82, 1982.

附図 ワークステーション関連図

