

HICATS/JEの辞書作成支援環境

6J-5

村上孝也、唐木武志、吉村紀久雄 梶博行
 ㈱日立製作所ソフトウェア工場 同システム開発研究所

1. はじめに

日立日英翻訳支援システム HICATS/JE (Hitachi Computer Aided Translation System / Japanese to English) は、辞書、文法、プログラムから構成される。辞書は、語に関する情報の集りであり、文法は、翻訳のための文法規則の集りであり、プログラムは、文法を翻訳実行する。

このうち、辞書の作成においては、大量の語に対して、いかに効率よく、かつ、いかに翻訳の質を高くするかが重要であり、辞書作成支援環境の整備が必要となった。

本稿では、HICATS/JEにおける辞書の作成に用いた辞書作成支援ツールについて述べる。

2. 辞書の概要

HICATS/JEでは、次の3種の辞書を用いる。

(1) 基本用語辞書

助詞、助動詞などの機能語、使用頻度の高い一般語や固有名詞、科学技術の各分野で共通に用いられる一般性の高い専門語など約5万語を含む。

(2) 科学技術専門用語辞書

JIS用語や学術用語など科学技術各分野の専門用語約25万語を含む。

(3) ユーザ用語辞書

ユーザ独自の用語を含む。

基本用語辞書は、標準提供で翻訳に必須な辞書であり、科学技術専門用語辞書は付加機能であり、任意の辞書である。また、ユーザ辞書は、ユーザ作成の辞書である。

辞書レコードの例を図1に示す。

3. 辞書作成支援環境

辞書作成においては、大量の語をいかに効率よく作成するか、いかに翻訳の質の高い辞書を作成するかが重要であり、次のようなツールを作成した。

- (1) 用語辞書保守
- (2) 専門用語辞書作成支援
- (3) 各種リストニング支援

4. 用語辞書保守

用語辞書保守は、用語辞書を作成、保守するための基本機能を実現するプログラムであり、次のような機

カイヨウ〔海洋〕

語ID : #0009945
 1. 品生語 (MF) 活用種 () 性 () 活用行 () 転成名詞 (0)
 意味 : SMC=PLAC

1. 訳選択 :
 1 品詞 (N) 英訳 : sea
 属性 : NC=C, NO=D
 反意語 () 発音 (C) 複数形 (1)
 2. 訳選択 :
 1 品詞 (N) 英訳 : ocean
 属性 : NC=C, NO=D
 反意語 () 発音 (V) 複数形 (1)
 2 品詞 (ADJ) 英訳 : oceanic
 反意語 () 発音 (V) 比/最 (9)

図1 辞書レコードの例

能を持つ。

(1) 対話辞書保守機能

メニュー、ガイダンス方式により、端末から用語辞書の登録、更新、表示を行う。

端末から即時更新、確認ができるため、少量の入力、修正に適する。

(2) バッチ辞書保守機能

ソース形式(図2 参照)の辞書データから、用語辞書(VSAM)を作成する。大量のデータを一度に入力するのに適する。この場合、確認はリストで行う。また、既存データの修正も可能である。多数の語に同様な修正を行う場合、ソース形式をプログラムで生成すると効率がよい。

```
#TLAA  海洋
#TLBA  カイヨウ
#J1AA  MF
#J1CA  SMC=PLAC
#G1AA
#E1BA  SEA
#E1GA  NC=C, ND=D
#G2AA
```

図2 ソース形式辞書データの例

(3) 印刷機能

辞書全体の印刷のほか、読みの範囲や更新日付の範囲、品詞などを指示して、一部分だけ印刷できる。

(4) 比較・複写機能

2つの辞書を比較して、その差異を印刷する。基準となる辞書からの変更箇所を確認したり、2つの辞書の語いの差などをチェックすることがで

きる。また、2つの辞書を比較結果にもとづいて統合し新しい辞書を作成することができる。

5. 専門用語辞書作成支援

専門用語辞書は、品詞が名詞、サ変動詞等に制限されており、また、大部分は、名詞又は名詞連語のため各語に付与すべきコードをある程度推定できる。専門用語辞書作成支援は、対訳データに、推定にもとづいてコードを付与し、専門用語辞書を生成するプログラムである。推定できない部分や推定が誤った部分を修正することにより、効率よく、専門用語辞書を生成できる。

6. 各種リスティング支援

語いや辞書データを調整して翻訳の質の向上を図るには、目的にあったリストが必要である。

(1) 未知語リスト

例文中に含まれた、その辞書に対する未知語を検出し、印刷する。これにより、その辞書の語いのかたよりや不足をチェックできる。また、ユーザ辞書作成時の語い候補の選択に利用できる。

(2) 調整用リスト

語いに付与されたコードに従って選択、ソートしたリストを出力することにより、コード付けのバラツキや矛盾を検出しやすくする。

7. まとめ

HICATS/JEで用いている辞書作成支援ツールについて述べた。辞書は今後とも、翻訳の質の向上のための調整と、語いの追加を行う必要があり、辞書作成支援環境としては、以下の検討が必要である。

- (1) 各ツール的高速化、使用リソースの削減
- (2) 辞書の学習機能(特に訳語の順序)
- (3) 既存辞書を使用した、辞書作成支援
(連語レコードに対するコード付与など)