

英日機械翻訳システムSHALT

7J-4

の辞書編集機能

日本アイ・ビー・エム株式会社 サイエンス・インスティテュート

筧 義郎

I. 概要

英日機械翻訳システムSHALTにおける変換辞書編集機能について述べる。機械翻訳システムを実務に使用する際には、変換辞書をユーザーの適用業務に合わせるために、大量の修正、追加を行わなければならない。SHALTでは、辞書修正は翻訳処理の前と後に、それぞれ別のエディターを使って行なう。前処理では、翻訳対象となる文書中の未登録語を検出し、登録する。後処理では、翻訳結果を参照し、必要に応じて辞書レコードを修正する。

SHALTの英日変換部においては、品詞ごとにまとめられた変換辞書を使う。各辞書レコードは、見出し語、訳語、及び訳語選択のための付加情報からなっている。動詞の場合には、訳語に活用、アスペクトなどの情報も付いている。

SHALTにおいては、動詞辞書の記述が翻訳の質に重大な影響を及ぼす。各動詞の記述は、その動詞の主語、目的語となる名詞の規定などの情報を含んでいる。各種変換辞書の中で、動詞辞書の記述が最も複雑になっている。ここでは、その動詞辞書を例として、辞書エディターの機能を説明する。

II. 辞書レコードの例(動詞)

SHALTの変換辞書では、各辞書レコードはLISPのリストになっている。一例を図1に示す。

```

("accept"
 ( (SB (S ((AT LE LC UD HM)
           (Y1 "は")) )
   (DO (S ((LE IF LC VA AT)
           (Y1 "を")) )
   (P ("受け入れ")
       (V SHIMO1 NIL
          SHUNKAN TRANS) ) ) ) )

```

図1. 動詞辞書

'(SB (S) の部分がこの動詞の主語となり得る名詞を規定し、'(DO (S) の部分が目的語となり得る名詞を規定している。'(P (.) の部分には、訳語、活用、アスペクトなどの情報が含まれる。

'AT', 'LE'などは、名詞の意味マーカである。SHALTの意味マーカについては、文献1、2に説明されている。

III. 未登録語処理

翻訳対象となる文書中に未登録語があると、その語の訳語が定まらないだけでなく、周辺の語の訳語選択にも悪影響を及ぼす。そのため、翻訳プロセスに入る前に、未登録語を検出し、効率良く登録する機能が必要となる。SHALTにおいて、前処理の一部として、上記の機能を実現した。

前述のように、変換辞書に対し、適切に修正、または追加を行なうためには、SHALTに関する知識がかなり要求される。SHALTに関するこまかい知識のない操作員が、動詞辞書を編集することは困難である。そこで、こまかい訳し分けはあきらめ、必要最小限の情報を指定するようにした。動詞の場合、操作員は訳語、活用、自動詞か他動詞かの区別を指示する。残りの項目に関しては、デフォルト値が設定される。

次に、具体的な操作方法について述べる。まず、翻訳対象となる文書中の未登録語を検出し、表示する。操作員が特定の語を指定すると、図2の画面(品詞指定用)が表示される。未登録語を含む文は、ウィンドウの近くに表示される。(実際の画面では、未登録語は強調表示される。)操作員は、未登録語が含まれる文を見て、その語の品詞と訳語とを決定することができる。

Occasionally, your installation may have to write a special app to run in some operating system other than the production system machines, your installation can run the application without dec system to the other operating system. Use this approach when you out certain applications without converting them to the new of Your VH/SP system can be installed with a starter system of System Offering. VH/SP collects errors in its error , generates reports.

見出し語 : collect

品詞 : 1. 名詞
2. 形容詞
3. 副詞
4. 動詞

品詞の番号のPFキーを
押して下さい。

図2. 品詞指定用画面

ここで、操作員が品詞を指定すると、辞書項目を記入するための、図3の画面が表示される。(ここでは、記入した後の画面を示す。)

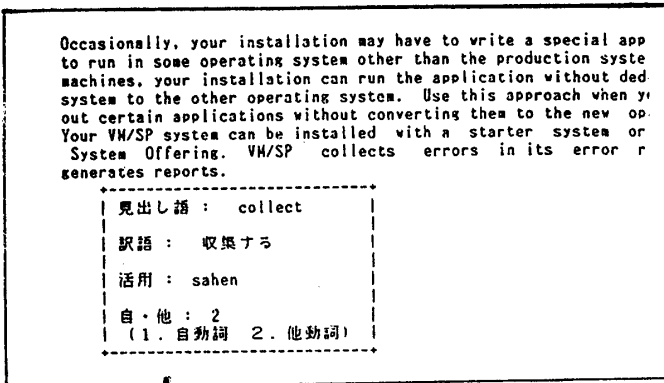


図3. 辞書項目記入用画面

この操作をすべての未登録語に対して行なう。もしも辞書項目の適切な指定ができないのであれば、それを登録しないでおくこともできる。その場合には、訳語として、英語の単語自身を使い、その他の情報については、デフォルト値を使う。

IV. 専任操作員用辞書エディター

ここでは、SHALTの翻訳アルゴリズムについての知識を有し、さらに個々の動詞の属性を正確に判定できる専任操作員が使うことを前提とした辞書エディターについて説明する。

このエディターは、翻訳結果の後編集で判明した辞書の不備を直すためのものである。現在は、後編集機能と辞書編集機能とが関連していないので、翻訳結果と辞書レコードを別々に修正しなければならない。この点を改良し、後編集と同時に、辞書レコードがある程度自動的に修正される機能を、目下検討している。(参考文献4.)

辞書レコードを修正する場合には、まず見出し語と品詞を指定する。すると、図4の画面が表示される。訳語が複数個ある場合には、その中から修正すべきものを指定する。

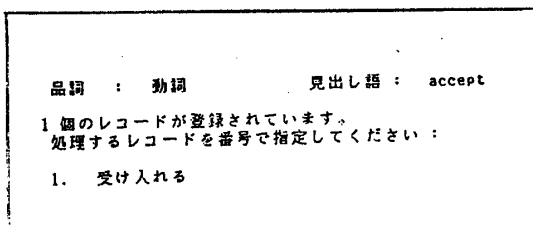


図4. レコード選択用画面

修正すべき訳語を選択すると、次のような、レコードの全項目を含む画面が表示される。ここで各項目を必要に応じて修正、追加する。なお、この画面は、II. で示したレコードを処理したものであり、'SB'等の記号の意味は、II. で述べた通りである。

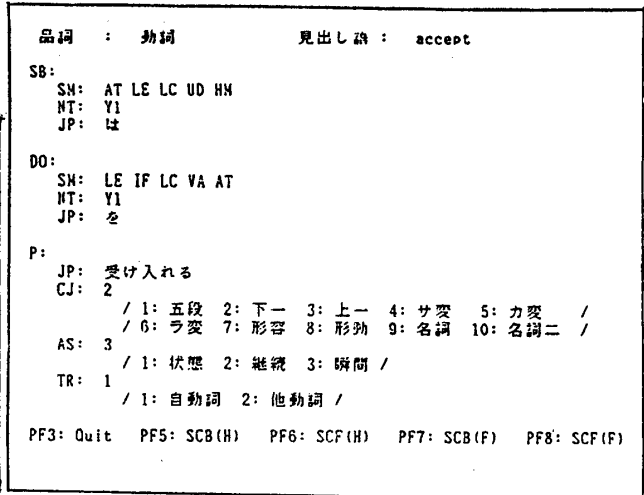


図5. 修正用画面

修正、追加が終わると、エディターがLISP形式に変換し、辞書に書きこむ。

V. まとめ

現在、III.、IV. で述べた2種類の辞書エディターが使用可能である。未登録語用のエディターは、SHALTに関する知識があまりない操作員でも使えるという利点はあるが、辞書レコードは訳語部分のみが規定されているので、こまかい訳し分けの指定はできない。逆に、専任操作員用エディターの場合は、辞書レコードのすべての部分を操作員が指定しなければならない。結果としてこまかい訳し分けが期待できるが、SHALTに関するある程度の知識が要求される。

実際の運用を考える場合、この二つの中間の機能を持つエディターが、使いやすさ、翻訳結果に与える影響の両方からみて、望ましいと思われる。すなわち、辞書レコードの項目を可能なかぎり指定するが、その際に、システムが適切なガイダンスを与えるので、SHALTに関する知識は必要としないエディターの開発が今後の課題である。

VI. 参考文献

- 堤、他 「機械翻訳システムSHALTの概要および英文解析とトランスファー」 第31回情報処理学会全国大会
- 堤、他 「英日機械翻訳システムSHALTにおける単純名詞句の翻訳方法について」 自然言語処理研究会資料56-1
- 笈、他 「英日機械翻訳システムSHALTにおける英日変換辞書の構成」 第32回情報処理学会全国大会
- 堤 「簡単な学習機能を備えた機械翻訳のためのエディター」 第33回情報処理学会全国大会