6F－7　類語国語辞典を介した意味マーカー付与

田中英輝*　江原暉将**　有賀憲和***松田健生***

*ＮＨＫ放送技術研究所　**ＡＴＲ自動翻訳電話研究所　***カテナリソース研究所

１．はじめに

筆者らは英日自動翻訳システム[1]で動詞と名詞の共起意味選択制限による動詞訳語選択[2]に使用するための，意味マーカーの設定と，翻訳辞書英語見出し語に対する日本語名詞訳語への付与方法を研究している．一般に辞書の中で，名詞の占める割合は一番大きい．多数の語義に対して，設定された意味マーカーをひとつずつ付与していく作業は，非常に労力がかかる．また，意味マーカーに変更を生じた時の訂正もひとつずつ行わなければならない．そこで本稿では，角川類語国語辞典[3]を仲介して，半自動的に意味マーカーを日本語名詞訳語に付与する方法を提案し，その実験を行ったので報告する．

２．意味マーカーについて

意味マーカー体系は，具体物，抽象物，気象現象，補完マーカーの４つを柱とした上位下位分類の形をしており全部で８１個設定した．具体物の中には，ニュース分野に対応するため［ＣＲＩＭＩＮＡＬ］｛犯罪人｝を設けている．また抽象名詞はその分類の難しさを軽減するために，まず，動詞の派生形を持つか，形容詞の派生形を持つかで分類した後，意味的な分類を行っている．

３．マーカー付与の方法

意味マーカーの半自動的な付与は，図１に従い，類語国語辞典を仲介して，２つの対応表を作成した後行った．

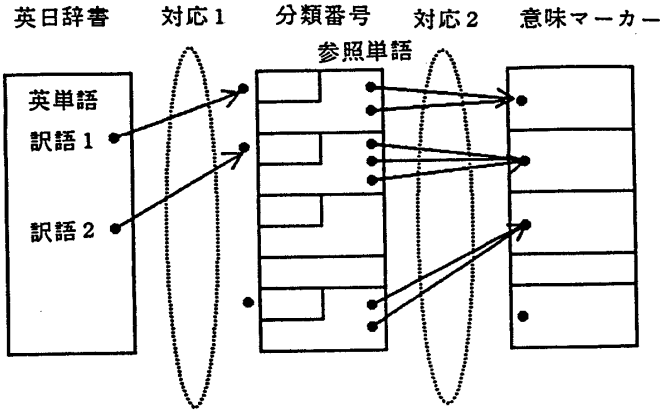

図１．意味マーカー付与方法

３．１．名詞訳語と分類番号の対応

図１の対応１の部分を以下の要領で作成した．

角川類語国語辞典は，細分類まで含めると，２７９４の分類番号で単語を分類している．日本語名詞訳語に対して，その語が収録されている類語国語辞典分類番号（以下，分類番号と略す）をすべて付与した．複合語については，主要構成語を用いて番号を付与した．分類番号が見つからない語については，新たな分類番号３３個を作り，それらを付与した．

３．２．分類番号と意味マーカーの対応

図１．対応２の部分を以下の要領で，４名の作業者で作成した．

（１）すべての分類番号に対して，以下の単語を，その分類番号の参照単語として抽出した；その分類番号の項目名（整数分類番号のみ）；その分類内の先頭の単語；中間付近の単語．これらの参照単語を分類番号に併記した作業ファイルを作成した．

（２）このファイルを４つに分割し，各作業者は，その内の２つについて（３）の作業を行った．

（３）各番号に付与された，参照単語（２個または３個）を見て，それに対応すると思われる意味マーカーを付与する．このとき，複数の意味マーカーの付与を許した．

この作業の結果，２７９４分類すべてに対して，２名の作業者による，意味マーカーとの対応データが得られた．この２名の対応の相違を各分類について厳密に計算したところ，１６００分類について相違がみられた．この相違は，抽象物の下の上位下位の間で多くみられた．

表１，表２に，同一ファイルに対する２名の意味マーカーの使用頻度の違いを示した．これは，各作業者が付与したマーカー上位４つとその割合を示したものである．これらの表によれば，作業者によって，使用する意味マーカーに違いがみられる．また表１，表２の作業者２に着目すると，意味マーカー使用に一定の傾向がみられる．

表１．作業者によるマーカー付与の差１

作業者１		作業者２	
CONACT	9.5%	ABSACT	13.5%
HUMAN	7.9%	CONACT	13.5%
HUMSTATIC	7.6%	NOWILLCHA	7.1%
ACTION	7.2%	HUMAN	6.6%

An Assignment of Semantic Markers to Nouns thorugh a Dictionary

Hideki TANAKA (NHK), Terumasa EHARA (ATR)
Norikazu ARUGA, Takeo MATSUDA (CLI)

表２．作業者によるマーカー付与の差２

作業者２		作業者３	
CONACT	14.1%	ACTION	17.6%
MONOACT	8.7%	HUMSTATIC	9.2%
ABSACT	8.4%	HUMAN	7.0%
HUMAN	6.4%	ADJECTIVAL	5.6%

３．３．名詞訳語と意味マーカーの対応

３．１．及び３．２．で得られた対応表を用いると名詞訳語と意味マーカーの対応が機械的にとれる．分類番号と意味マーカーの対応は，２人分得られており，今回は，それらの和集合を対応データとした．これらの対応データを用いて，ニューステロップ英語原稿１９８９年９月分の中で使用頻度の高かった１６個の動詞と共起した名詞１６５語の日本語訳語に自動的に意味マーカーを付与した．その結果を検討したところ，期待された意味マーカーは，ほぼ付与されていたが，不適切な意味マーカーも付与されていた．

４．問題

不適切な意味マーカーは大別して次の２つの理由で付与されていた．

（１）不適切な語義の分類番号の仲介．（図２）

例えば，図３に示すようにようにｂａｃｋの訳語｛背｝に対して，［ＬＥＮＧＴＨ］｛長さの概念｝というマーカーが付与されていた．類語国語辞典では，｛背｝は１０４ｂ，１２３ｂ，６０１ｄの３カ所に登録されている．このなかで，１２３ｂは｛背丈｝の意味であり［ＬＥＮＧＴＨ］が付与されていたが，ｂａｃｋの語義として使用するのは不適切である．

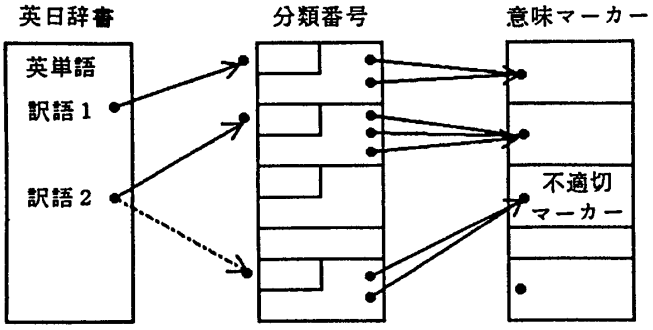

図２．不適切語義の仲介

ｂａｃｋ

｛背後｝：　［ＲＥＬＡＴＩＯＮ］

｛背｝　：　［ＲＥＬＡＴＩＯＮ］

［ＬＥＮＧＴＨ］

［ＬＩＦＥＰＡＲＴ］

図３．ｂａｃｋに付与された意味マーカー

（２）類語国語辞典とマーカーとの分類不整合．（図４）

図５に示すように，ａｄｖａｎｔａｇｅの訳語｛強み｝に対して［ＨＵＭＡＮ］｛人間｝という不適切なマーカーが付与されている．類語国語辞典の中で，｛強み｝は，１３５｛強弱｝１７２ｅ｛長所　優れているところ｝，６７０ｃ｛屈強　力が優れて強いさま｝の３カ所に登録されている．このうち６７０ｃの分類中に，｛精鋭｝という単語も収録されている．分類番号と意味マーカーの対応作業の時の参照単語の一つとして，｛精鋭｝が選ばれたため，６７０ｃに［ＨＵＭＡＮ］も付与されてしまった．類語国語辞典の一つの分類番号が，我々の意味マーカーの複数個を含んだために発生した誤りである．

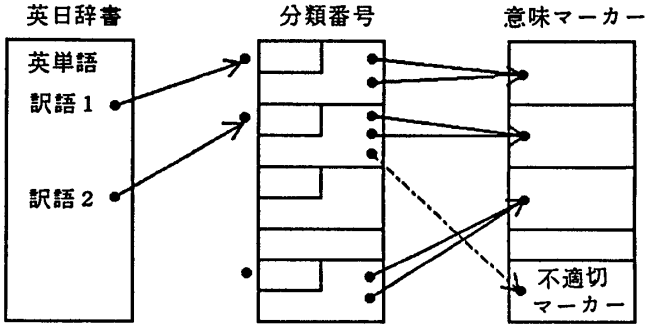

図４．分類の不整合

ａｄｖａｎｔａｇｅ

｛強み｝：［ＨＵＭＡＮ］｛人間｝，

［ＷＩＬＬＰＲＯＰ］｛形容詞派生を持つ名詞，意志物の属性｝，

［ＮＯＷＩＬＬＰＲＯＰ］｛形容詞派生を持つ名詞，無意志物の属性｝，

［ＨＵＭＳＴＡＴＩＣ］｛人間に関する抽象概念｝

図５．ａｄｖａｎｔａｇｅに付与された意味マーカー　（一部）

５．まとめ

角川類語国語辞典を仲介して，日本語名詞訳語に，意味マーカーを半自動的に付与する手法を提案し，実験を行った結果，少ない手間で必要なマーカーを付与することができた．しかし今回の実験は，作業者が付与したすべてのマーカーを使用し，また，名詞訳語が属する分類番号すべてを利用したため，不適切なマーカーも含む結果となった．今後，この不適切な意味マーカーを削除しなくてはならない．不適切な分類番号を仲介したために付与されたマーカーは，人手で英語の訳語として適切な分類番号だけを選択するか，和英辞典を利用して，自動的に選択することで削除することが考えられる．[4]いずれにしても削除の手間は，新たにマーカーを付与する手間より少ないと思われる．また，類語国語辞典と意味マーカー分類の不整合に関する問題，作業者によるマーカー付与傾向の差の取り扱いは，今後検討していきたい．

【参考文献】

（１）相沢，他　「衛星放送ワールドニュースの英日機械翻訳」　情報処理学会第４０回全国大会

（２）中瀬　「英日機械翻訳システムにおける解析手法について」　情処研資　NL69-7,1988

（３）大野，浜西　「類語国語辞典」　角川書店

（４）田中，他　「翻訳辞書からの中間概念の自動抽出に関する基礎的考察」　情処研資　NL72-3,1989