

## 日英機械翻訳における利用者登録語の意味属性の自動推定

・池原 悟<sup>(注)</sup> ・白井 諭 ・横尾昭男  
・Francis Bond \*\*小見佳恵

\*NTTコミュニケーション科学研究所 \*\*NTTアドバンステクノロジー

### あらまし

利用者が利用者辞書に単語を登録するとき、登録したい日本語見出し語とその英語訳語を与えるだけで、システムがシステム辞書の知識を応用して、名詞種別を自動的に判定し、名詞種別に応じた単語の意味属性を推定して付与する方法を提案する。実験結果によれば、本方式では専門家の付与した意味属性よりも多くの属性が付与されるが、50～80%の再現率が得られることが分かった。また、訳文品質に与える効果の評価では、人手で作成した利用者辞書を使用する場合と同等の訳文品質が得られることが分かった。その結果、利用者辞書作成への単語の登録において、最も熟練度の要求される単語意味属性付与作業を自動化できる見通しとなった。

### Automatic Determination of Semantic Attributes for User Defined Words in Japanese to English Machine Translation.

\*Satoru Ikehara, \*Satoshi Shirai, \*Akio Yokoo  
・Francis Bond and \*\*Yosie Omi

・ NTT Communication Science Laboratories  
\*\* NTT Advanced Technology Company

### Abstract

This paper proposes an automatic determination method of semantic attributes for user defined words. This method determines semantic attributes of the user defined words, both Japanese index word and the English translation of which are given by a user.

Applying this method to compiling of a user dictionary required for newspaper article sentences and sentences of software design documents has revealed the following. More semantic attributes are rendered by the proposed method in comparison to the Manual Method and between 50 to 80% recall factor were achieved. Further, the translation quality is on a level equivalent with that resulting from use of a manually compiled user dictionary. From the foregoing,

## 1. はじめに

機械翻訳システムを使用する時、利用者はシステム辞書に登録されていない単語や、登録されているが、訳語が不適切な単語に対して、利用者辞書を作成して使用することが多い。しかし、辞書に新しく単語を登録する際は、登録する語の見出し語、訳語の他に、文法的、意味的な種々の情報を付与する必要がある。高度な翻訳品質を狙ったシステムほど、利用者辞書にも詳細で正確な情報を必要としており、素人の利用者がそれらの情報を正しく付与するのは簡単でない。例えば、日英機械翻訳システムALT-J/Eでは、意味解析のため約3,000種の精密な意味属性体系を持っており<sup>(1)</sup>、利用者辞書の単語を登録する際は、各単語にこの意味的属性体系に従って意味的用法（一般に複数）を指定する必要がある。この作業は熟練を要し、一般の利用者には困難であるため、従来から自動化への期待が大きかった。

そこで本論文では、利用者登録語の特性に着目し、利用者が登録したい見出し語（名詞単語または複合名詞）に対して英語訳語を与えるだけで、システムがシステム辞書の知識を応用して、名詞種別を自動的に判定し、名詞種別に応じた単語の意味属性を推定して付与する方法を提案する。また、自動推定した利用者辞書を使用した翻訳実験により、方式の効果を確認する。

具体的には、名詞を対象に、与えられた見出し語と訳語から主名詞と名詞種別（一般名詞、固有名詞）を判定し、それぞれの場合に必要な単語意味属性を自動

推定する方法を示す。また、適用実験では、まず、本方式を、新聞記事102文とソフトウェア設計書105文の翻訳に必要な利用者辞書の作成に適用して、自動推定した単語意味属性と辞書専門家の付与した単語意味属性を比較し、精度の比較を行う。次に、これらの意味属性が翻訳結果に与える影響を調べるため、（1）意味属性のない利用者辞書を使用する場合、（2）自動推定した意味属性を使用する場合、（3）専門家が利用者登録語の見出し語と訳語を見て付与した意味属性を使用する場合、（4）正しい意味属性（専門家が翻訳実験により適切性を最終的に確認した意味属性）を使用した場合、の4つの場合について翻訳実験を行う。

## 2. システム辞書と利用者辞書

### 2.1 ALT-J/Eの意味辞書の構成

ここでは、機械翻訳システム側であらかじめ用意された辞書をシステム辞書、利用者が作成して使用する辞書を利用者辞書と呼ぶ。日英機械翻訳システムALT-J/Eのシステム辞書と利用者辞書および単語意味属性の関係を図1に示す。

#### (1) 意味辞書の種類

ALT-J/Eでは意味解析を実現するため、これらの辞書に単語意味属性を使用した意味情報が登録されるようになっている。意味情報を記載された辞書は意味辞書と呼ぶ。現在、実装されている意味辞書は単語意味辞書と、構文意味辞書の2種類からなる。単語

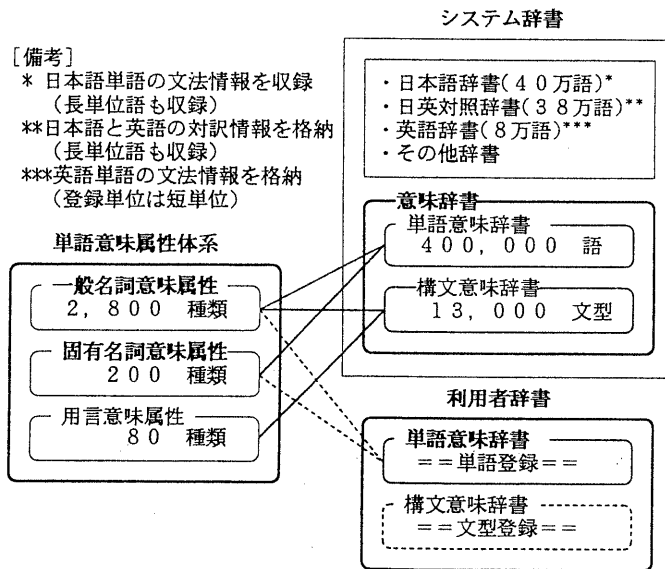


図1. ALT-J/Eの意味属性体系と意味辞書

意味辞書は日本語単語の意味的用法を記述した辞書（日本語解析用の辞書40万語と訳語決定用の辞書38万語が含まれる）であり、構文意味辞書は、用言毎の日本語文型とそれに対応する英語文型（13、000文型）を収録した辞書である。システムがあらかじめ用意したこれらの単語または文型が不足した時、もしくは不適切なときは、同種の辞書を利用者が利用者辞書として作成して使用する。

### (2) 単語意味属性の種類

ALTの単語意味属性には一般名詞意味属性（2,800種）、固有名詞意味属性（200種）、用言意味属性（100種）の3種類がある。固有名詞意味属性は、一般名詞意味属性の一部を取り出して、複合語解析の観点から詳細化したものであり、属性名数は一般名詞意味属性の数より少ないが、分類精度は詳細である。

単語意味辞書の一般名詞には一般名詞意味属性（一般に複数個）が、固有名詞には一般名詞意味属性と固有名詞意味属性の両者（いずれも複数個）が付与される。用言意味属性は構文意味辞書に登録された文型パターンの主用言に付与される。

## 2.2 利用者登録語の特性

本論文では、名詞（単語名詞または複合名詞）の利用者辞書への登録を考える。通常の機械翻訳システムでは、一般語（一般名詞）についてはほぼ漏れなくシステム辞書に収録されるが、専門用語や固有名詞などは余り収録されていない場合が多い。ALT-J/Eの場合は、新聞記事で使用される語を中心に多数（延べ50万語）の固有名詞、専門用語なども収録されているが、全てを網羅することは不可能であり、必ずしも十分とは言えない。

従って、通常、利用者辞書に登録される語は、(1)原文に現れた専門用語や固有名詞、利用者固有の技術用語で、システム辞書に登録されていないため未知語となった語、もしくは(2)システム辞書に登録されているが、訳語が適切でない語の2種類に大別される。後者の単語意味属性は既にシステム辞書に登録されているため、通常改めて登録する必要はないのに対して、前者は登録語が複合名詞である場合が多く、その構成要素は単語として既にシステム辞書に登録されている場合が多い。このようにシステム辞書は、多くの場合、利用者辞書登録語と関係する情報を持つ場合が多いので、その情報を利用すれば、多くの利用者登録語の意味属性は自動付与できると期待できる。

## 3. 意味属性推定の方法

利用者登録語の日本語表記と英語訳語が与えられた

とき、機械翻訳システムに装備されたシステム辞書の情報を使って、登録語の意味属性を推定する方法を図2に示す。

利用者登録語の単語意味属性を推定する手順は、主名詞の判定、名詞種別の判定（固有名詞、一般名詞）の判定、固有名詞意味属性の推定（固有名詞の場合）、一般名詞意味属性の推定（一般名詞、固有名詞双方の場合）の手順からなる。

### 3.1 主名詞の判定方法

利用者辞書に登録される見出し語は、単一の名詞又は複数単語から構成される複合名詞のいずれかであるのに対して、訳語は単一の単語、名詞連続の複合語、名詞句のいずれかとする。見出し語、訳語を構成する単語のうち、中心的な意味を担う名詞を主名詞と呼ぶ。

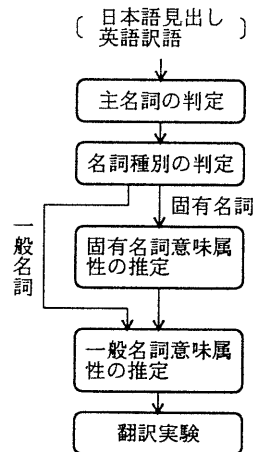


図2. 意味属性自動推定の手順

通常、登録語の単語意味属性は主名詞の単語意味属性と一致する。また、システム辞書の中に利用者辞書登録語の見出し語または訳語の一致する語が存在する可能性に比べて、利用者辞書登録語の主名詞が存在する可能性は高い。従って、主名詞に着目すれば、登録語の意味属性を推定できる可能性が大きい。

日本語名詞は語形変化しないため、システム辞書の見出し語と利用者登録語の主名詞を含む部分とを直接比較し、システム辞書内から必要な情報を引き出すことができる。これに対して、英語名詞は複合語内などで屈折による語形変化を伴うことがあるため、主名詞を含む部分とシステム辞書の英語訳語を直接比較することはできない。そこで、ここでは、システム辞書の訳語との比較が可能となるよう、利用者登録語の英語訳語に対して主名詞を抽出する。

[英語主名詞の判定手順]

- (1) まず、訳語が単語一語で構成されるときは、その語を主名詞とする。
- (2) 次に、訳語が2語以上の語から構成されている場合は、まず、訳語の全体が、システム辞書に登録されているか否かを調べ、登録されている場合は、訳語全体を主名詞とする。
- (3) 登録されていない場合は、名詞句(訳語)を構成する単語の中から主名詞を推定す。この場合、英語訳語は名詞連続複合語または修飾語や句を伴った名詞句で構成されていると考えられる。前者の場合は、最後の名詞が主名詞になるのに対して、後者の場合では、修飾語句は主名詞の前方だけでなく後方に来る場合のあることを考慮する必要がある。通常、後方修飾は前置詞、関係詞等で導かれることを考慮して、以下の方法で主名詞を選定する。

- ① 訳語中に in, on, with などの前置詞また that, which などの関係詞(ストワード)があるか否かを調べ、ある場合は、その語以下の語を削除する。
- ② 次に、残った英語全体に対して英語辞書引きを行い、辞書内に一致する語があれば、それを主名詞とする。
- ③ 一致する語のないときは、前方から一語ずつ落としながら(修飾語を外しながら)、残った語に対して英語辞書引きを行い、辞書と一致した語(または語の組)を主名詞とする。外せる修飾語がなくなった時は、残った語を主名詞とする。

### 3. 2 名詞種別の判定方法

前に述べたように、一般名詞には、一般名詞意味属性を付与すればよいのに対して、固有名詞には一般名詞意味属性と固有名詞意味属性の両方を付与することが必要である。そのため、利用者登録語が固有名詞か一般名詞かの判定を行う必要がある。この判定は、利用者にとって比較的容易な作業であるが、利用者の負担を少しでも削減することを狙って、自動化の方法を考える。

日本語表現では、一般名詞と固有名詞は通常、表記

上区別されないのに対して、英語表現では、固有名詞の先頭文字は大文字で書かれる点に特徴がある。そこで、登録された単語の英語側の表記に着目し、訳語が1単語の時は、先頭文字1文字が大文字の場合は固有名詞とし、それ以外は一般名詞とする。複数の単語から構成される訳語の時は、各単語の先頭一文字が大文字の場合は、固有名詞とする。訳語にすべて大文字からなる単語が含まれる場合は、それ以外の名詞がすべて固有名詞の時は全体を固有名詞とし、それ以外は一般名詞とする。

### 3. 3 意味属性推定の方法

利用者登録語の見出し語、訳語、訳語の主名詞と、システムに既に準備されている日英対照辞書の内容を比較して、利用者登録語の単語意味属性を推定する。利用者登録語が一般名詞の場合は、日英対照辞書に登録された一般名詞を検索の対象として、一般名詞意味属性を推定するのに対して、利用者登録語が固有名詞の場合は、日英対照辞書に登録された固有名詞を検索の対象として、一般名詞意味属性と固有名詞意味属性を推定する。

以下、利用者登録語の見出し語から意味属性を推定する方法と訳語から推定する方法を示すが、これらの手順は一般名詞意味属性の場合と固有名詞意味属性の場合に共通である。

#### (1) 見出し語(日本語表記)から推定する方法

日英対照辞書を検索し、利用者登録語の見出し語と一致する見出し語が日英対照辞書の登録語にある場合は、既に登録された訳語が適切でないため、訳語を変えるのが利用者辞書登録の目的であると考えられるから、単語意味属性の変更は必要ないと判断できる。そこで、日英対照辞書に記載された単語意味属性を利用者登録語の単語意味属性とする。

利用者登録語の見出し語と一致する見出し語が日英対照辞書の登録語にない場合は、利用者登録語の後方からの最長一致法で、再度、日英対照辞書を検索する。カタカナ語を除き、2文字以上が、日英対照辞書の見

表1 利用者辞書の例

日本語見出し (利用者付与)	英語訳 (利用者付与)	意味属性 (推定)
治療	cure	《治療》
放射線治療	radiotherapy	《治療》
手当て	treatment	《治療》
医療	medical treatment	《治療》
数値制御ロボット	numerical controlled robot	《産業機器》
照明付き机	desk with lighty unit	《家具》

表2 システム辞書(日英対照辞書)の例

日本語見出し	英語訳	意味属性
治療	treatment	《治療》
制御ロボット	controlled robot	《産業機器》
机	desk	《家具》
...	...	...

出し語と部分一致（カタカナ語の場合は単語単位で一致）すれば、日英対照辞書の意味属性を利用者登録語の意味属性とする。

例えば表1で、利用者登録語の「治療」、「放射線治療」は、システム辞書(表2)に「治療」があるので、意味属性は〈治療〉となる。

#### (2) 訳語（英語表記）から推定する方法

日英対照辞書の訳語の中に、利用者登録語の訳語と一致する語がある場合は、その訳語に対応する見出し語は、利用者登録語の見出し語と同義語であると考えられるので、日英対照辞書に登録された意味属性を、そのまま利用者登録語の意味属性とする。

利用者登録語の訳語と一致する訳語が日英対照辞書の登録語にない場合は、(1)の場合と同様、再度、日英対照辞書を検索する。その中に、利用者登録語の主名詞もしくは主名詞を含む訳語部分が、日英対照辞書の訳語にあれば、日英対照辞書の意味属性を利用者登録語の意味属性とする。但し、利用者登録語と日英対照辞書の訳語が同一の主名詞を持つ場合でも、語形が異なる場合があるので、主名詞は可能な語形変化（単数複数など）をさせながら、照合を行う。

例えば、表1で、利用者登録語の「手当」、「医療」は、その訳語（または主名詞訳語）「treatment」がシステム辞書(表2)にあるので、意味属性は〈治療〉となる。

以上の方法では、システム辞書には一般に複数の意味属性が付与されていること、日本語表記だけでなく英語表記からも意味属性が抽出されるため、一般に一語に対して複数の意味属性が抽出されることになる。利用者辞書は特定の翻訳対象に対して指定して使用されるため、用語の用法に限られる特徴がある。従って、実際の用法が意味属性として与えられていれば、それ以外の用法が多少付与されていても、副作用は少ないと期待される。そこで、意味属性としては、得られた意味属性すべてを登録する。但し、同一の単語意味属性が重複して抽出された場合は、重複を取って登録する。

## 4. 意味属性推定精度の評価

### 4.1 実験の条件

表3に示すような新聞記事文とソフトウェア設計書の日本語に対して前章の方法を適用し、自動推定の精度を評価する。具体的には、以下の3つの場合に分けて、得られた単語意味属性を比較評価する。

#### (1) 自動推定方式による場合

与えられた見出し語、訳語のペアに対して、前章の方法で単語意味属性を付与する。

#### (2) 人手付与方式の場合

単語意味属性体系に精通した辞書担当のアナリストが、与えられた見出し語、訳語のペアを見て、単語意味属性を付与する。

#### (3) 最適意味属性の場合

(2)で作成した利用者辞書を使用して対象文の翻訳実験を行い、その結果を見て意味属性の修正追加を行う。最終的に翻訳結果が最適となるまでこの作業を繰り返して、意味属性を定める。この方法で得られた意味属性を、最適値と仮定する。

表3 実験対象文の特性

項目		新聞記事	設計書
対象文数(文)		102文	105文
平均文字数(文字/文)		42.2	40.2
平均単語数(単語/文)		21.2	16.0
利用者辞書登録語数	一般名詞	28語	98語
	固有名詞	49語	7語
	合計	77語	106語
利用者登録語を含む文数		53文	93文

### 4.2 名詞種別自動判定精度

前章の3種類の意味属性付与方式で得られた名詞種別の判定精度を表4に示す。

新聞記事の場合、自動判定方式では、利用者登録語全体77語のうち、判定の正しかった名詞は一般名詞27語、固有名詞45語の合計72語で、正解率は93.5%であった。人手付与方式では、一般名詞27語、固有名詞46語を正しく判定し、正解率は94.8%であった。これに対して、設計書の場合は、自動判定法の正解率89.5%、人手付与方式の正解率は97.1%であった。

自動判定で、一般名詞を誤って固有名詞と判定した語は、「郵政大臣」、「中部圏」、「GE」、「IGS」、「汎用GS」などであった。逆に、固有名詞を誤って一般名詞と判定したのは、「PC9800」、「VOS3.2」、「X.25プロトコル」などであった。

以上から、文書の種類によって多少の差はあるが、自動判定方式で人手判定方式と大差のない結果が得られることがわかった。

判定に失敗した約10%の名詞について考えると、固有名詞には固有名詞意味属性のほか一般名詞意味属性も付与することになっているため、一般名詞を固有名詞と判定した語（新聞記事1語、設計書語8語）の場合は、一般名詞意味属性も付与されることになり、訳文品質へ

の影響は殆どないと期待される。しかし、逆に、固有

表4 名詞種別の判定結果 ( )内の数は、正しい判定の数を示す。

標本判定方法 属性種別	新聞記事			ソフトウェア設計書		
	自動判定	人手判定	最適解	自動判定	人手判定	最適解
一般名詞 意味属性	31語 (27語)	30語 (27語)	28語	93語 (90語)	99語 (97語)	98語
固有名詞 意味属性	46語 (45語)	47語 (46語)	49語	12語 (4語)	6語 (5語)	7語
合計	77語 (72語)	77語 (73語)	77語	105語 (94語)	105語 (102語)	105語
判定正解率	93.5%	94.8%	100%	89.5%	97.1%	100%

名詞を一般名詞と判定した語（新聞記事4語、設計書3語）には、固有名詞意味属性が付与されないで、その語が複合語構成要素として使用された場合、影響がでると考えられる。

#### 4.3 意味属性自動推定精度

単語別にみたときの自動推定とアナリスト付与の結果を表5、付与された意味属性全体の数とその内訳を表6に示す。アナリストの付与した意味属性が正解であると考えたときの適合率と再現率は、表6から表7の通り求められる。これらより以下のことが分かる。

- ①意味属性自動推定のアルゴリズムは、システム辞書の情報を手がかりに働くため、利用者登録語の

全てに意味属性が付与されるとは限らない。これに対して、実験結果では、意味属性付与の必要な単語延べ238語に対して、意味属性が自動推定された語数は211語であり、その割合(88.7%)は大きい。これは利用者登録語に関連する語の情報が、既にシステム辞書に豊富に存在することを示している。

- ②単語毎に見たとき、正解以外の余分の意味属性が付与された語も多いため、適合率はあまり高くないが、再現率を見ると、新聞記事の場合はいずれも80%以上、設計書の場合でも60%程度を得ている。従って、3.3節に述べた理由から自動

表5 単語別にみた単語意味属性の自動付与品質

比較項目	属性種別	新聞記事		ソフトウェア設計書	
		一般名詞 意味属性	固有名詞 意味属性	一般名詞 意味属性	固有名詞 意味属性
属性付与の必要な語数		77語(100%)	49語(100%)	105語(100%)	7語(100%)
付与された属性の内訳	属性が付与された語数の合計	73語(94.8%) 77語(100%)	47語(95.9%) 47語(95.9%)	88語(83.8%) 100語(95.2%)	3語(42.9%) 5語(71.4%)
	全属性が正解	38語(49.4%) 44語(57.1%)	42語(85.7%) 42語(85.7%)	3語(2.9%) 50語(47.6%)	2語(28.6%) 1語(14.3%)
	余分に付与	21語(27.3%) 9語(11.7%)	0語(0.0%) 0語(0.0%)	41語(39.2%) 11語(10.5%)	0語(0.0%) 1語(14.3%)
	一部付与不足	4語(5.2%) 8語(10.4%)	0語(0.0%) 0語(0.0%)	18語(17.1%) 27語(25.7%)	0語(0.0%) 2語(28.6%)
	全てが誤り	10語(13.0%) 16語(20.8%)	5語(10.2%) 4語(8.2%)	26語(24.8%) 12語(11.4%)	1語(14.3%) 1語(14.3%)
自動付与されなかった語数	3語(3.9%) 0語(0.0%)	3語(6.1%) 3語(6.1%)	17語(16.2%) 5語(4.8%)	4語(57.1%) 2語(28.6%)	
属性付与の必要ない語数	0語	28語	0語	98語	
属性付与された語数	0語 0語	1語 1語	0語 0語	8語(8.2%) 1語(0.1%)	

<上段の数字>：自動推定方式の場合、<下段の数字>：人手付与方式の場合

表6 属性数から見た自動推定と人手付与の比較

標本 属性付与の方法 属性の種類	新聞記事			ソフトウェア設計書		
	自動推定 で付与した 属性数	人手付与 の属性数	最適解の 属性数	自動推定 で付与した 属性数	人手付与 の属性数	最適解の 属性数
一般名詞意味属性	194件 76+0件	110件 93+17件	127件 --	341件 67+20件	130件 74+17件	191件 --
固有名詞意味属性	46件 42+21件	43件 42+1件	48件 --	12件 2+0件	7件 1+2件	7件 --
合計	240件 116+21件	153件 135+18件	175件 --	342件 63+22件	137件 75+19件	198件 --

下段の数字の説明：nnn+mmm  
 nnn=付与された属性の内、最適解と一致する属性の数  
 mmm=自動付与された属性が最適値の近傍（上位または下位）にあるものの数を示す。

表7 自動付与した意味属性の適合率と再現率

意味属性種別	標本	新聞記事		ソフトウェア設計書	
		適合率	再現率	適合率	再現率
一般名詞 意味属性	自動付与方式	38.1% (49.0%)	58.3% (74.8%)	19.6% (25.5%)	35.1% (45.5%)
	人手付与方式	71.0% (84.0%)	73.2% (86.6%)	56.9% (70.0%)	38.7% (47.6%)
固有名詞 意味属性	自動付与方式	91.3% (同上)	87.5% (同上)	16.7% (同上)	28.6% (同上)
	人手付与方式	89.4% (91.5%)	87.5% (89.6%)	14.3% (42.9%)	14.3% (42.9%)
全体	自動付与方式	48.3% (57.5%)	66.3% (78.9%)	19.5% (25.2%)	34.8% (44.9%)
	人手付与方式	75.8% (86.0%)	77.1% (87.4%)	54.7% (68.6%)	37.9% (47.5%)

( )内の数字は、最適意味属性の近傍（上位下位）も正解とした場合を示す。

推定の効果は十分あると予測される。

- ③ソフトウェア設計書の場合、固有名詞の意味属性の精度かなり低い。しかし、この場合、固有名詞の数は少数であること、固有名詞でも一般名詞意味属性は付与されることから、訳文品質への影響は少ないと思われる。

使用した場合

- (場合2)自動推定方式により付与した意味属性を使用した場合  
 (場合3)人手付与方式により付与した意味属性を使用した場合  
 (場合4)最適意味属性を使用した場合

## 5. 訳文品質の向上効果

### 5.1 実験の条件

利用者登録語に対する意味属性自動推定の効果を調べるため、前章と同一の試験文（新聞記事102文、ソフトウェア設計書105文）を対象に、前章で得られた利用者辞書を用いて、翻訳実験を行った。実験は以下の4つの場合に分けて実施した。

(場合1)単語意味属性の付与されない利用者辞書を

### 5.2 実験結果

新聞記事102文中、利用者辞書への登録語を含む53文に対して、以下の3つの場合の翻訳結果を求めた。その結果を表8に示す。

これより、自動推定された単語意味属性を使用した場合、意味属性を付与しなかった場合に比べて、訳文合格率は6~13%向上し、人手付与方式の場合と同等の品質が得られることが分かった。この結果は、最適意味属性を使用する場合に比べて、翻訳率は2~3%

表8. 訳文品質の比較評価

訳文品質の比較 意味属性付与の方法		新聞記事		ソフトウェア設計書	
		訳文合格率	品質変化*	訳文合格率	品質変化*
場合1	意味属性付与無し	56.7%	±0.0%	65.7%	±0.0%
場合2	自動推定方式	69.6%	+16.7%	71.4%	+10.5%
場合3	人手付与方式	71.5%	+21.6%	71.4%	+15.2%
場合4	最適意味属性	72.5%	+25.5%	73.3%	+23.8%

\* 10点満点評価で1点以上、訳文品質に変化のあった文の割合を示す。

およばないが、最適意味属性を決定する繰り返し実験のコストを考えると、十分満足できる値である。

また、経験的に言って、機械システムの改良により10%の翻訳率向上を得ることは容易ではない。機械翻訳0前後%の翻訳率の向上は大きな効果といえる。

なお、新聞記事の場合、自動推定方式で訳文品質を向の実用レベルの品質は70~80%以上と考えられるから、訳文品質が50~60%の現状のシステムでは、1上できなかった3文を見ると、その原因は、名詞種別の判定誤りが1件、正解の意味属性の上位または下位の属性を選択したものが、それぞれ1件であった。本方式では、名詞の種別も自動判定しているが、誤りの例から見て、名詞種別と意味属性の単純な分類(上位2~3段程度)を利用者に依頼することができれば、これらの誤りは、ほぼ防ぐことができると推定される。

以上の結果、従来、利用者が利用者辞書を作成する際、最も熟練の必要な単語意味属性の付与作業を自動化できる展望が得られた。

## 6. あとがき

利用者辞書に登録する利用者登録語の見出し語(日本語)と訳語(英語)が与えられたとき、機械翻訳システムに既に存在する情報を利用して、その単語意味属性を自動的に推定する方法を提案した。また、本方式を新聞記事102文、ソフトウェア設計書105文の翻訳に必要な利用者辞書の作成に適用し、推定された単語意味属性の精度、最終的な翻訳結果に与える影響などを評価した。

その結果、自動推定された単語意味属性は、専門家が実験の繰り返しによって決定した意味属性(最適意味属性)の50~80%をカバーしていることが分かった。この値は、専門家が自動推定と同一の条件で人手付与方式により付与した意味属性のカバー率(50~90%)よりは若干(~10%)低いが、十分効果

の期待できる値である。

また、自動推定された単語意味属性を使用した翻訳実験では、意味属性を付与しなかった場合に比べて、訳文合格率は6~13%向上し、人手付与方式の場合と同等の品質が得られることが分かった。この品質は、最適意味属性を使用した場合に比べても、2~3%しか低下しない値であり、最適意味属性を決定する繰り返し実験のコストを考えると、十分満足できる値である。

これらの結果、従来、利用者が利用者辞書を作成する際、最も熟練の必要な単語意味属性の付与作業を自動化できる展望が得られた。今後は、対訳コーパスから、利用者辞書登録の必要な単語の見出し語と訳語を自動抽出し、利用者辞書全体を自動生成する方法について研究を進める予定である。

## 参考文献

- (1) J. Carbonell et. al. : JTEC Report on Machine Translation in Japan, Japanese Technology Evaluation Center, Loya College in Maryland, (January 1992)
- (2) 池原, 宮崎, 横尾: 「日英機械翻訳のための意味解析用の言語知識とその分解能」, 情処論, Vol.34, No.8, pp.1692-1704 (1993)
- (3) T. Utsuro, Y. Matsumoto and M. Nagao: Lexical Knowledge Acquisition from Bilingual Corpora, Proc. of the 14th International Conference on Computational Linguistics, Nantes, France, pp.581-587 (1992)
- (4) S. Ikehara: Multi-Label Machine Translation System, Future Computer Systems, Vol.1, No.3, pp.261-274 (1989)
- (5) S. Ikehara, M. Miyazaki, S. Shirai and A. Yokoo: An Approach to Machine Translation Method Based on Constructive Process Theory, Review of the Electotical Communications Laboratories, Vol.37, No.1, pp.39-44 (1989)
- (6) H. Nakaiwa and S. Ikehara: Zero Pronoun Resolution in a Japanese to English Machine Translation System using Verbal Semantic Attributes, Applied Computational Linguistics, (1992)