

# 日英機械翻訳のための意味解析辞書

Semantic Analysis Dictionaries for Machine Translation

池原 悟<sup>\*\*</sup>      宮崎 正弘<sup>+</sup>      横尾 昭男<sup>\*\*</sup>  
Satoru IKEHARA      Masahiro MIYAZAKI      Akio YOKOO

+ 新潟大学: Niigata University

\*\* NTT情報通信処理研究所:

NTT Communications and Information Processing Laboratories

## あらまし

自然言語処理において意味処理を実現するには、意味処理の目的と範囲を明確にし、それに合わせて必要な知識を収集整理することが大切である。本論文では、言語過程説の立場から言語表現の意味を「表現」に結びつけられた「対象」と話者の「認識」の関係と捉え、「意味処理」を言語表現に用いられた言語規範の種類を判別する「意味解析」と、話者と対象世界とを関係づける「意味理解」の二つに分けた。そして、このうち日本語の「意味解析技術」を実現するために必要な言語知識を、さらに語彙に関する知識と用言を核とした表現構造に関する知識に分け、それぞれ単語意味辞書、構文意味辞書として収集整理した。

両者は約3,000種の単語意味属性による記述を介して相互依存の関係を持つ。文型の記述能力は、意味属性を精密化したことにより、従来の50~500種の意味属性分類の場合に比べて大幅に向上した。その効果は特に、和語系の用言と専門的用言の持つ文型において著しく、記述能力はいずれも数倍以上となった。また、これらの言語知識は機械翻訳において、用言及び用言に結びついた体言の意味による訳し分けに効果があるだけでなく、複合語解析や文脈処理、訳し分け易い日本語への自動書き換えなどの実現に有効であることが分かった。

## 1. まえがき

自然言語処理においては、対象知識や専門知識などの世界知識も必要であるが、その前に、処理対象とする言語に関する知識が大切である。処理のアルゴリズムや方式等は、知識の構造や適用の方法に依存する。実用的な言語処理を実現するには、膨大な量の情報を収集するだけでなく、それらの情報が相互に矛盾無く有効に働くよう体系的に整理することが重要である。

日本語単語の意味情報の扱いについては、古くから栗原、吉田等<sup>(1)</sup>の研究があり、最近では意味処理用ソーラスや<sup>(2)</sup>、EDRにおける概念辞書の研究<sup>(3)</sup>が行われているが、意味処理の内容とそれに必要な知識の範囲の関係や、意味処理の精度と知識の精度関係などを解明した例は見あたらない。また、意味処理においては単語の意味(語義)だけでなく表現の構造の持つ意味も重要となるが、表現の構造と意味の関係についてはまだ整理されたものがない。

本論文では、言語過程説の立場から、対象の認識と表現の過程に着目して、「意味処理」を「意味解析」と「意味理解」の2つのプロセスに分け、前者に必要な言語情報を言語知識として2種類の意味辞書にまとめる。

ここで、「意味解析」とは、言語知識に基づき、表現に使用された言語規範の解析を行う処理であり、「意味理解」とは、一般知識までを応用し言語表現と対象世界の結びつきまで解析する処理である。これに対応して意味処理に必要な知識は、前者に必要な言語規範に関する知識(言語知識)と、後者に必要な一般知識や専門知識、世界知識などの知識(言語外知識)に分けて考えることとし、「意味解析」のための言語知識を単語意味属性を介し

た単語意味辞書と構文意味辞書の関係として整理することを提案する。

構文意味辞書では用言を中心とする日本語表現の構造を意味の単位として、表現構造の持つ意味を英語側の表現に対応させて記述するが、その記述能力と単語意味属性の分解精度と関係を明かにすると共に、これらの言語知識の体系化の結果、実現が可能となった機械翻訳の機能について述べる。

## 2. 意味解析と言語知識

### 2.1 意味処理の内容

#### (1) 言語表現の意味

意味処理の内容を明確にするため、「言語表現の意味」について考える。言語過程説(時枝謙二<sup>(4)</sup>)によれば、言語は「対象」、「(話者の)認識」、「表現」の間の過程的複合体と考えることができる<sup>(4)</sup>。「対象」には物理的實在としての対象と観念的存在としての対象があるが、その何れもが話者の精神に反映し、認識が形成される。対象に対して形成された話者の認識は、それぞれの言語の持つ規範を介して、言語表現に結びつけられる。言語過程説を意味論の点で改良した三浦文法<sup>(5)</sup>によれば、「言語表現の意味は表現に結合された対象と、話者の認識の関係である」と説明される。すなわち、対象や認識そのものは意味の構成要素ではあるが、意味そのものではなく、意味はそれらの関係<sup>(1)</sup>とするものである。

#### (2) 意味解析と意味理解

言語表現が「対象」、「認識」、「表現」の順序の過程的構造を持つのに対して、言語理解はその逆の構造を持つ。聞き手はまず言語規範を手がかりに、表現に結びつけられた話者の認識のあ

<sup>1</sup> 意味を実体概念ではなく関係概念であるとする考え方は、状況意味論<sup>(6)</sup>とも通じるところがあるが、状況意味論では言語表現とその表現のおかれた場に関する「場の表現」を区別せず、むしろ場の表現の側から意味を説明しているのに対して、三浦文法では両者を明確に分けている。言語の個別科学を追求する立場からは、言語の個別性と特殊性に着目して、言語表現の意味を、記号など言語以外の表現についても共通する「場の表現」の意味とは区別して分析することが必要と考えられる。

り方を知り、次にその認識を通じて対象のあり方を知る。この過程を言語処理の立場から見ると、言語理解を以下の2つのステップに分けることができる。すなわち第1は単語の語義や構文の使い方に関する言語上の多くの規範の中から、話者の使用した規範を特定するステップであり、第2は特定された言語規範の組み合わせから対象世界を再構成すると共に、それに対する話者の認識を追体験するステップである。ここでは、前者を「意味解析」、後者を「意味理解」と呼び、両者を合わせて、「意味処理」と呼ぶことにする。

## 2.2 意味解析のための言語知識

### (1) 言語知識と言語外知識

「意味解析」を実現するには、それぞれの言語の持つ規範(約束)を収集し整理することが必要である。また「意味理解」では、聞き手側が話者の世界をシミュレートするだけの世界知識を持つことが必要であるため、計算機内に、一般知識や対象分野に関する専門知識などを有する世界モデルの構築が必要となる。ここで、意味解析に必要な情報を言語知識、意味理解に必要な情報を言語外知識とすると、言語知識とは言語規範に関する知識、言語外知識とは常識や専門知識などのことである\*1。

機械翻訳においては原文の対象とする分野は多彩な場合が多く、あらかじめ一般知識や対象分野の専門知識まで集大成して用意するのは困難であるため、意味理解型翻訳の実現は容易ではない。しかし、最終的な翻訳結果を理解するのは目的言語側の人間であるので、原言語側で使用された言語規範を目的言語側の言語規範に対応させ、置き換えることにより、かなりの程度の翻訳が実現できると期待される。そこで、以下では「意味解析」のための言語知識について考える。

### (2) 言語知識の種類

言語が表現する対象は、実体、属性、関係の3者からなる客体の世界と話者自身を示す主体に分けられる。このうち、主体も実体も含め、ここでは言語の表現対象として、実体、属性、関係の3者を考える。次に、「関係」には、実体間の関係、属性間の関係、実体と属性の関係などがあるが、何れも対象の構造を立体化するものであり、概念化\*2された後は言語表現上、実体(観念的実体)概念として扱われる。そこで、言語表現上の対象世界の構成要素を、実体と属性の2つで捉えることにする。

実体と属性の言語上の表現方法について見ると、物理的実在であるか、仮想的存在であるかを問わず、実体は何れも言語表記上、

概念化の過程を経て名詞として表現される。また、属性は動的なものや静的なものに分けられ、動的な属性は動詞で、静的な属性は形容詞で表現されるが、まとめれば言語表現上は用言に対応する。

以上から、言語解析のための知識(言語知識)として、名詞と用言に関する約束(語義と用法)の体系化を考える。

## 3. 言語知識の体系化

### 3.1 単語意味属性と単語知識

#### (1) 単語意味属性の体系

話者の認識の結果として得られた実体の概念は表現上、単語(名詞)に対応づけられる。単語と概念の対応関係は必ずしも一対一とは限らず、複数の概念が一つの単語に対応づけられることも多い。通常、単語の意味と言われているものは、このような単語と概念の対応関係に関する約束のことであり、厳密には語義である\*3。

単語レベルでみると、意味解析は言語表現中の単語がどのような約束(語義)で使用されたかを調べることであるが、単語辞書に記載された語義情報だけでそれを決めることは困難であり、現実の表現上での用法に関する知識が必要と考えられる。

そこで、概念化の過程と概念を単語に対応させる方法について考えてみると、これらは、対象とする実体の見方、捉え方に大きく依存し、同一の対象でも見方、捉え方によって使用される単語に違いが生じることが分かる。例えば、妻が夫を表現するとき、夫婦の関係で見れば「夫」となり、家の関係で見れば「主人」、恋人の関係で見れば「彼」、一人の人間で見れば「山田太郎」などとなる。また逆に、一つの単語を一つの語義で使ったとしても、その表す概念はさまざまである。例えば、単語「学校」は「人が集まって教育を受けるところ」と言う語義を持つが、現実の表現で使われたときは、「子供の集まりとしてみた学校」や、「教育の機関としてみた学校」、「ある地域の特定の場所としてみた学校」などがある。このような情報は通常の語義の範囲を越えるが、通常の辞書上は用例などとして表現されていることを考えると、やはり言語知識(言語規範)の一部と考えられ、計算機処理に可能な形式に整理する必要がある。

そこで、概念化された対象と単語との対応関係を、対象の見方捉え方に着目して整理する基準として、単語意味属性を考える。すなわち、ここで言う単語意味属性とは対象を概念化する際の視点を整理したものであり、話者から見れば単語の用法に相当する。

\*1意味理解に必要な情報としては、この他、文脈、段落の情報や発話の状況に関する情報が考えられるが、ここではシステムであらかじめ用意すべき情報を考える。

\*2 概念化：認識のプロセスの一つで、対象の持つ個別的、特殊の側面を捨象すること。対象に対して、取り上げる側面と切り捨てる側面が取捨選択される。すなわち、抽象化・一般化と個別化・特殊化を調和させるのが概念化であり、その結果得られるのが、認識の単位としての概念である。

\*3 従来、単語の意味(語義)を扱う方法として、意味標識(semantic marker)、意味素(semantic primitives)、意味素性(semantic feature)などを用いる方法が考えられている<sup>(7)・(8)</sup>。その違いは必ずしも明確でないが、大きくみて、単語の意味をさらに要素に分ける立場と、単語の意味を対象の持つ特徴で捉える立場に分けられる。本論文では、部分的表現の意味は全体の表現の中で決まることを基本的な立場としており、その点で、単語の意味についても還元論的立場(顕微鏡的、モンタージュ的とも言う)はとらない。すなわち、名詞の語義で表される概念は、認識の単位(第1回概念化の結果)として一定のまとまりを持ったものであり、意味の要素に分けられないものとする。そのような意味で、ここでは、対象の見方を示す用語として「意味属性」を用いる。「意味属性」の名称もやはり名詞を用いて表現されるが、そのときの名詞は一語一義で使用する。

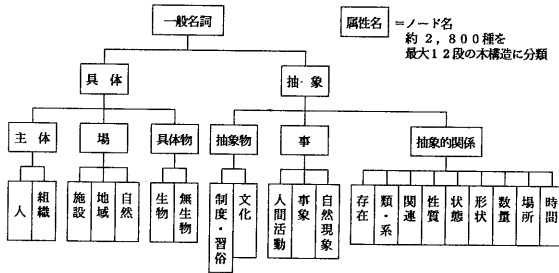


図1. 一般名詞意味属性体系の一部 (上位4段まで)

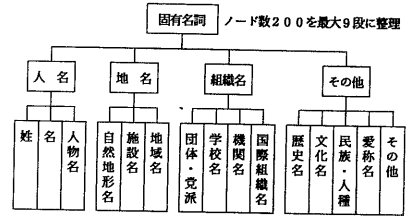


図2. 固有名詞の意味属性 (上位3段まで)

例えば、単語「東京」は「行政区画」の一つ、「駅・港」の一つ、「地域」の一つなどとして取り上げるために使われるから、「東京」の意味属性は「行政区画」、「駅・港」、「地域」などとなる<sup>1)</sup>。

また、意味属性の分解精度は用途に依存する。ここでは、日英機械翻訳のための意味解析を考え、日本語の単語意味属性を体系化する。

### (2) 意味属性の分解精度

対象の見方捉え方は言語によっても異なるため、意味解析の対象言語毎に単語意味属性の体系が必要となる。また、細かく見れば、対象に対する見方、捉え方は単語毎にすべて異なるから、見方、捉え方の数は単語数以上に存在することになる。そのため、語義をすべて正しく扱うには単語数以上の意味属性の分解精度が必要となるが、実際の計算機処理では現実的でない。日英機械翻訳では、英語での訳し分けが可能な程度に意味属性の分解精度があれば良いため、実験的に意味属性の分解精度を決定することが考えられるが、それも容易ではない。

そこで、ここでは日英機械翻訳における分解精度としては、普通の人が日常的に使いこなしている語意や文字数程度が必要と考えて、意味属性を約3,000に分類する<sup>2)</sup>。ただし、個々の属性名の相互関係などについては、実験的に改良している。

### (3) 意味属性体系

以上の考えに基づいて、下記の通り、単語意味属性を一般名詞意味属性と固有名詞意味属性に分けて整理した。

(a) 一般名詞意味属性体系……12段の木構造で、2,800ノード

(b) 固有名詞意味属性体系……9段の木構造で、200ノード

各意味属性体系の上位ノードに一部を図1、図2に示す。また、トップノードから見た意味属性の深さと意味属性名の数の関係を図3に示す。これより、一般名詞意味属性は8~9段の付近で最

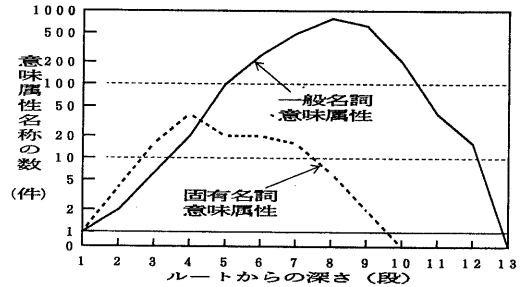


図3. 意味属性名称の数の分布

単語見出し	読み	標準標記	品詞	(文法情報等)
東京	とーきょう	東京	固有名詞	.....

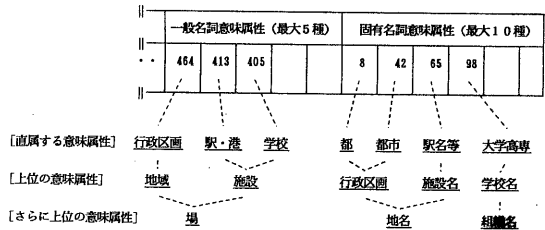


図4. 日本語意味辞書の形式 (例)

も横に広がり、10段以降は収束するのに対して、固有名詞意味属性は4段目で横の広がりが最大となるが、そのままは収束せず7段目以降で収束している。これは、固有名詞の性質を反映したもので、文中では人名関連の名詞が「姓」、「名」などのように並列的に使用されるのが多いのに対して、地名関連の名詞は「県」、「郡」、「町」、「字」などのように上位から下位に向かってネ

<sup>1)</sup> 従来、語義を分類するものとして、類語辞典<sup>(9)</sup>や分類語彙表<sup>(10)</sup>などがあるが、計算機処理では語義だけでなくその語義で表される対象概念の見方捉え方の知識が必要である。また、計算機処理を前提とするものに概念を用法で分類しようとしたものや<sup>(9)</sup>、シソーラス化したもの<sup>(12)</sup>があるが、ここでは対象の持つ特殊性を捨象する立場から概念化の視点を分類する。

<sup>2)</sup> 語彙統計<sup>(11)</sup>によれば、日常、普通の人が使用する単語(3,000語でカバー率80%)や英会話でよく使われる単語の数は2000~3000語とみられること、基本英語(Basic English)の単語は1000語以下であること、また、通常使いこなされる漢字は2000~3000程度であることなどから、単語意味属性項目数は3,000項目を目標とした。本研究の初期に500分類程度を試行したが、その程度では記述不能な用言文型が続出したため精度を上げた。この問題は4章で述べる。なお、名詞は数が多いため意味属性を介した処理を考えるが、用言はその数が個別に扱える範囲であるので、後に述べるように、直接処理の対象とする。

ストして使用されることが多いなどにより、固有名詞のタイプごとに必要とする分類の深さが異なるためである。

#### (4) 単語意味辞書

一般語10万語、固有名詞20万語、専門用語7万語、その他3万語の約40万語の日本語辞書に含まれる名詞約30万語に対して、前述の意味属性を付与し、単語意味辞書を作成した。単語意味辞書項目の例を図4に示す。図では単語「東京」に対して、3つの一般名詞意味属性と4つの固有名詞意味属性が付与されているが、一般名詞意味属性は用言との関係での使われ方を示すのに対して、固有名詞意味属性の方は複合語内での語の使われ方を見たものであり、前者に比べて段数は少ないが、より細かいレベルの属性となっている。

単語当たりの意味属性の数は平均約2件である。語数からみれば多数を占める漢語の意味属性の大半が1であるのに対して、和語の意味属性数はかなり多く、用法が多彩であることを反映している。

### 3.2 表現構造知識と構文意味辞書

#### (1) 構造化された認識の表現

名詞が実体概念を表現するに対して、用言は単独で実体の持つ性質、運動、変化などの属性の概念を表すが、これらの属性は実体に付属したものであるため、表現の上では何らかの実体概念との関係で用いられることが多く、立体的な対象世界を一次元的な言語表現に結び付けるための枠組みをも与えている。

すなわち、対象世界の持つ立体的な構造は、話者の認識過程において、実体間の関係や実体と実体の持つ属性との関係などとして捉えられ、言語表現上、単語間の関係として対応づけられる。このうち、実体と属性の関係、属性を介して関連づけられる実体間の関係は、用言を用いて構造化される。

実体と属性の関係に関する認識においても、話者の視点に沿って個別性と特殊性の捨象が行われ、用言の種類と用法が決まる。例えば、人がある地点間で移動する事象を「来る」と見るか、「行く」と見るかは、話者の視点に依存する。ここでは、このような実体と属性の関係の捉え方を用言の用法、すなわち、名詞と用言の関係として整理する。

#### (2) 表現構造の持つ意味の単位化

表現に結合された対象と認識の関係が表現の意味であるとする立場からみれば、表現の構造は意味の一部といえる。そこで、用言に関する言語知識の扱いにおいては、用言を介して表現される

構造を意味の単位として整理する\*1。

意味の単位、すなわち意味の分解精度としては日英機械翻訳の立場から、日本語の各用言に対して英語側の訳語が決まる範囲で記述する方針とし、格要素の意味属性は訳し分けのできる最小限の深さで記述する。また、文型の抽象化レベル（一般性、個別性）に着目して、以下の2つのタイプに分けて整理する\*2。

#### a) 一般文型

用言の字面をキーとする意味的結合パターンによる表現。用言の字面と1つ以上の格要素から規定される。ただし、格要素は名詞と格助詞から構成されるが、その内、名詞は意味属性が規定される\*3。

#### b) 慣用文型

一つ以上の格要素が意味属性でなく、直接名詞の字面で規定される外はa)と同じ。但し、パターンの適用条件が格要素に対する制約条件として記述されることがある。

#### (3) 構文意味辞書の作成

以上の考えに従って、三省堂新明解国語辞書等から収集した動詞、形容詞、形容動詞の見出し語に対して、市販の和英辞書（研究社新和英、小学館プログレス、ライトハウス等）を参照し、訳語と用例を分析して、用言の持つ構造と意味の関係を構文意味辞書にまとめた。また、これと比較するため、新聞の情報産業関連記事で使用される専門的な用言についても一部構文意味辞書を作成した。慣用的表現については、「日本語慣用句の意味と用法（明治書院）」に記載された日本語表現を中心に、上記の和英辞書を参照して用例を分析し、同様の構文意味辞書を作成した。対象とした用言数と作成した文型数は以下の通りである。

一般文型	一般語	動詞……4	600語/7,900文型
		形容詞……260語/	600文型
		形容動詞……900語/1,	300文型
	専門語	……………180語/	200文型
慣用文型	……………	2,	000語/3,000文型

#### (4) 用言の種類と文型数の関係

用言の種類と文型の数の関係を図5に示す。また、用言当たりの文型数の多いものを付表1に示す。これらの図表より以下のことが分かる。

①動詞の場合、一用言当たりの文型数は、漢語系の動詞（1.4件/語）に比べて和語系の動詞（2.3件/語）の方が多い。この関係は形容詞（2.1件/語）と形容動詞（1.5件/語）についても同様である。

\*1 従来、要素合成法を基本とする立場からは、表現の構造と意味は別ものと考えられることが多く、構造の持つ意味は考慮の対象外となっていた。そのため、構造と意味の一体化した慣用表現などは例外的扱いとされることが多かったが、翻訳の現場では、むしろ、「翻訳とは慣用句から慣用句への書き換えだ」とする人もいるくらいである。慣用表現に限らず、表現構造に関する言語知識の体系化は機械翻訳にとって重要な課題である。

\*2 このようにして用言毎に整理された体系を使用すれば、表現構造が意味単位毎に捉えられるため、構文解析上の曖昧性が減少する。また、機械翻訳において、日本語の表現構造と対して英語の構造を記述しておけば、意味解析の終了した時点で、同時に英語の表現構造も決まっていることになるため、改めて変換過程を持たなくて済む利点がある。

\*3 記述形式はValentzパターン<sup>(9)</sup>と同等であるが、格要素に対して深層格の立場はとらない。すなわち、各パターンは構造の持つ意味を単位化したしたものと考え、各パターンの意味がその部分である格要素の意味から合成されるとは考えない。

②慣用文型はそのほとんどが和語系の用言に対するものである。一用言当たりの文型数は、下記の通り何れも一般文型よりも慣用文型の方が多い。

	一般文型	慣用文型
動詞	2.3件/語	5.2件/語
形容詞	2.1件/語	11.6件/語

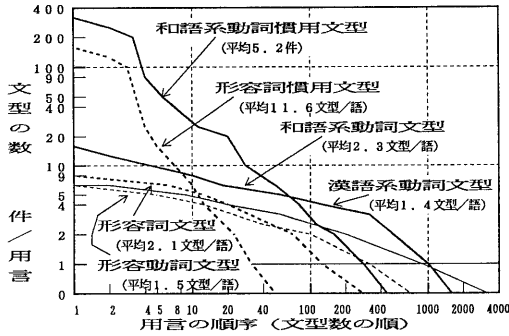


図5. 文型の種類とその分布

#### 4. 意味属性の分解精度と記述能力の関係

構文意味辞書の作成においては、一語当たりの文型数の多い用言の文型記述能力が問題となる。例えば、動詞「掛ける」の持つ文型(付表2)において、一般文型に記述された格要素の意味属性を見ると、5段以上深いものが60%、7段以上深いものがなお40%もあるなど、深い位置の意味属性が使われているものがかなり存在する。また、14の文型のいずれを見ても深さ4以下の範囲で書けるものはないことが分かる。これは動詞「掛ける」を訳し分けるには、それだけ細かな意味属性記述が必要であることを示している。

そこで、以下では種々の用言について、文型と意味属性の深さの関係について考察する。

##### 4.1 意味属性の分解精度

図3から意味属性の深さに対する意味属性名の数の累積分布を求めると図6を得る。この図より属性名称の数と段数の関係に着目して、以下の3つの場合について考える\*1。

場合	属性名称の数	相当する段数	記事
Case.1	30~50件	約4段	従来の標準的な例
Case.2	約500件	約6段	EDR辞書相当
Case.3	3,000件	8~9段	本研究

但し、「相当する段数」は意味属性が最も横に広がった付近の

段数を示す。

#### 4.2 言語知識記述の精度

##### (1) 単語意味属性の記述の深さ

単語意味辞書に付与した意味属性の深さを図7に示す。但し、用言の意味属性はその用言が名詞化したときの意味属性を示す。図より、以下のことが分かる。

①一般名詞意味属性、固有名詞意味属性、用言の意味属性ともに属性名の数は広い範囲に分布しており、意味属性がまんべんなく使用されていることが窺える。

②Case1, Case2の意味属性分解精度ではかなり限られた範囲でしか名詞の意味属性が記述できない。

単語辞書では各単語の意味的な分解精度を上げるため、なるべく細かいレベルで属性を付与しているが、①は、単語ごとに見たとき、上位レベルと下位レベルの属性が混ざって使用されていることをも示している。

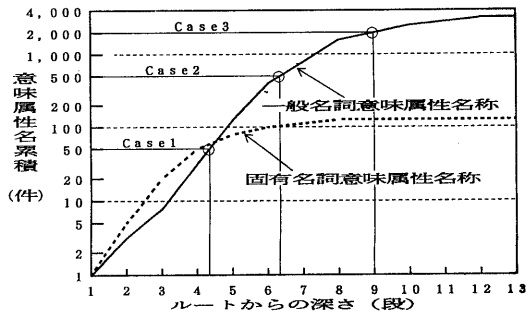


図6. 意味属性名称の数の累積分布

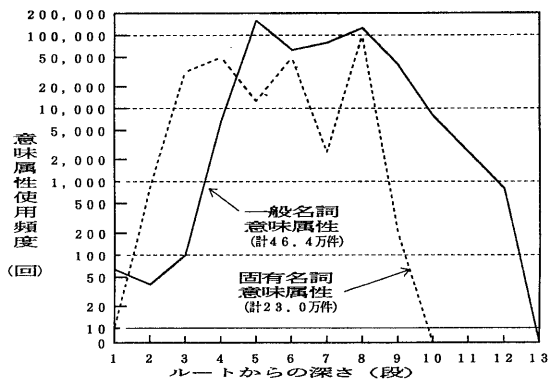


図7. 意味属性の使用頻度(単語意味辞書)

##### (2) 構文意味辞書の記述の深さ

構文意味辞書で使用されたすべての格要素について、それを規

\*1 意味属性の段数と名称の数の関係は、意味属性体系の作り方に依存するため、必ずしも対応関係がない。しかし、属性体系が目的に応じてバランス良く作られている場合は、文型内格要素の記述の精度は、相対的に属性体系の深さに対応すると考えられるから、ここではノードの深さをもって意味属性の分解精度の尺度とする。

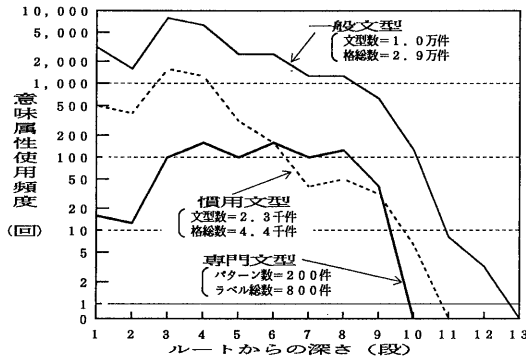


図8. 意味属性の使用頻度 (構文意味辞書)

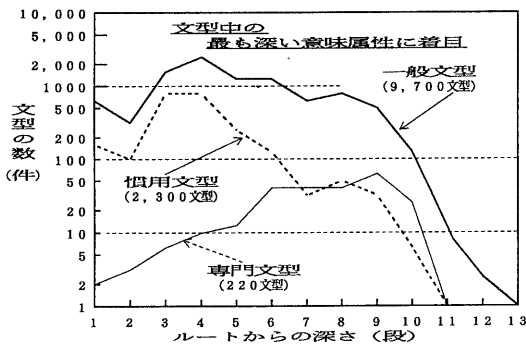


図9. パターン記述の深度分布

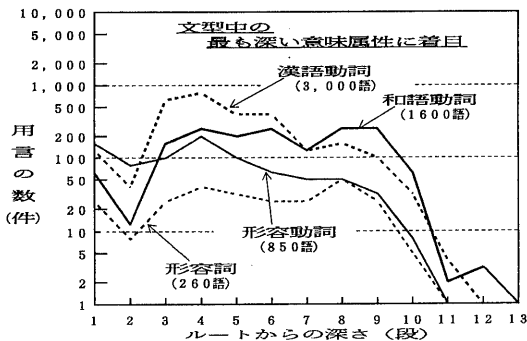


図10. 文型記述の精度 (一般文型)

定する意味属性の深さを調べ、図8に示す。次に、各文型毎にその文型を記述するには最低どのレベルの意味属性が必要かを調べるため、各文型を最も深い意味属性に着目して整理すると、図9の結果を得る。図9を用言の種類に着目して分解すると図10の

結果を得る。但し、図8、9の慣用文型の数は、字面以外の登録格要素の同一性に着目して縮退させ、集計した数値を示す。用言毎に文型を見たとき、その用言を訳し分けると最低必要な意味属性のレベルについては、図9、図10よりも分布若干が深い方にシフトする。これらの図から以下のこと分かる。

- ①文型格要素全体を見ると、意味属性は3段目が最も多いが、9段目まではなお高い使用頻度を持つ\*1。
- ②文型単位に見たときの記述の深度は、格要素単位に見たとき以上に、深い方に移動している。また、文型種別毎に見ると、専門文型、一般文型、慣用文型の順に深いレベルの意味属性が必要であることが分かる。
- ③和語系と漢語系の用言を比べると、和語系の用言の方が深いレベルの意味属性を必要としていることが多い。

これらの内、②は、慣用文型は一つ以上の格要素が字面で規定されているため、残りの格要素は比較的浅いレベルで規定できるのに対して、専門文型は一般文型と混同しないようなるべく厳密に記述する必要があるため、深いレベルの意味属性が必要であることを示している。

なお、意味属性無指定の格要素には第1段の意味属性(ワイルドカード)を付与している。

#### 4.3 言語知識記述能力について

図7～図10の結果から、Case1～Case3の場合についての記述可能な文型の範囲の相対値を求めると、表1の結果を得る。この表から以下のこと分かる。

- ①名詞単語にはなるべく下位の意味属性を付与していることもあり、Caseによる名詞の意味属性の記述精度の差は大きい。
- ②構文意味辞書では、文型の記述能力は、いずれの場合もCase1,2に比べてCase3の方が高い\*2が、その差は一般文型では和語系、全体からみれば専門文型において顕著である。

このうち、①からは、後に述べるように、用言の訳し分けよりも複合語解析や、名詞の意味による訳し分けで効果が期待できる。また、②からは特に意味的多義の多い和語系用言の訳し分けや、用言の専門的用法と一般的用法の訳し分けなどでの効果が期待できる。

#### 5. 意味解析の効果

上記で示した単語意味辞書と構文意味辞書を用いた日英機械翻訳実験システムALT-J/E<sup>(13)(14)(15)</sup>の実験的な試作によって、用言の訳し分けと、用言を介した名詞の訳し分けにおける効果が確認された。約5,000種の慣用文型を用いた翻訳実験では、慣用文型として登録された表現も、それが慣用的な意味で

\*1 構文意味辞書では一般文型、慣用文型共に、登録する文型の適用範囲をなるべく広く取れるよう、可能な限り上位の意味属性を付与する方針を取ったため、格要素全体では上位の意味属性が付与されたものが多いが、文型記述の可否は最も下位の意味属性を要求する格要素で決まる。

\*2 但し、EDRの概念辞書における用言と名詞の関係は、関係性に意味を持たせた2項関係としてみたものであり、意味の単位の文構造を記述しようとする構文意味辞書とは情報の質に差がある。

表 1. 意味属性精度と言語知識記述能力の関係

比較項目 意味属性 の分解精度	意味属性の ノード総数 (一般名詞)	ノードの 実効的な 深さ	名詞語義の記述可能範囲 (付与可能な意味属性 の割合)		格要素または文型の記述可能な範囲(注)						
			一般名詞	用言性名詞	一般文型				慣用文型	専門文型	
					和語動詞	漢語動詞	形容詞	形容動詞			合計
Case. 1 (従来標準例)	30~50	約4段	3%	21%	[31%]	[57%]	[38%]	[67%]	5.2% [5.2%]	8.1% [8.0%]	1.1% [1.0%]
Case. 2 (EDR相当)	約500	約6段	2.2%	2.5%	[5.9%]	[8.6%]	[6.0%]	[8.4%]	7.8% [8.1%]	9.5% [9.5%]	3.6% [3.4%]
Case. 3 (ALT-J/E)	2,800	8~9段	100% (仮定)	100% (仮定)	100% (仮定)	100% (仮定)	100% (仮定)	100% (仮定)	100% (仮定)	100% (仮定)	100% (仮定)

(注) 記述できる格要素の数の比率を示す。但し、[ ]内の数値は文型毎の最深の意味属性に着目して集計した文型の割合を示す。例えば、[60%]は60%の文型の記述が可能であることを示す。

現実の表現において、慣用的な文型と一般文型との調和は、訳文品質の向上を期待させる。

また、従来、前編集の項目と考えられていた以下の機能についても、実験的な試作の結果、実現のめどを得た。

(1) 日本語自動書き換え型翻訳機能

英語に直接対応する表現がなくて直訳できない表現や直訳できても英語としてみれば不適切な表現を翻訳し易い日本語に自動的に書き換えて翻訳する。書き換えの規則の適用条件が精密に書けるようになり、副作用の心配が減少したため、効果的な書き換えが可能となった。

(2) 複合語翻訳機能

複合語(名詞連続型)には一般名詞を結合したもの、接頭、接尾辞を持つもの、サ変名詞を含むもの、固有名詞を含むものなどがあり、その構造は多彩である。単語意味辞書はこれらの複合語を構成する単語間の意味的な関係の解析において、大きな役割を持つことが分かった。

また、サ変名詞などの用言性名詞を含む複合語では、他の名詞が格要素となる場合が多く、その解析においては構文意味辞書が有効である。

(3) 文脈処理による文要素補完型翻訳機能

用言を中心とする表現構造が意味の単位として捉えられるため、用言間の関係としての文脈の追跡が可能となった。具体的には、省略された格要素の補完処理が実現された。

6. あとがき

「意味処理」を「意味解析」と「意味理解」に分け、意味解析に必要な知識(言語知識)を約3,000種の単語意味属性を媒介とする単語意味辞書、構文意味辞書の2つの辞書にまとめた。

その結果、従来の50~500種の意味属性を使用する場合に比べて、文型の記述能力が大幅に向上した。その効果は和語系用言の持つ文型や専門文型について著しく、前者では2~3倍、後者では3~10倍の文型の記述能力の向上を得た。その結果、これらの文型を記述するには、3,000種程度の意味属性の分解精度は必要なが分かった。

また、これらの言語知識は、用言及び用言と結びついた体言の意味の訳し分けに効果があるだけでなく、複合語解析や文脈型処理、翻訳し易い日本語への自動書き換えなどの翻訳新機能の実現に有効であることが分かった。

今回整理できた言語知識(言語規範)は客体化された概念と語義との間の関係知識と、属性概念と実体概念間の関係知識の2種類であり、実体間の関係知識については残されている。その意味で、対象の1.5回概念化の範囲にとどまっていると言える。

今後、実体間の関係知識も加えて2回概念化までを体系化するには、単語意味属性の多次元化やソーラスの併用などが考えられる。また、長文に取り組むには用言間の関係知識についても整理することが必要と考えられる。

[謝辞] 最後に、共に議論を進めてきた白井、小倉両主任研究員、中岩研究主任他、翻訳研究グループの皆様へ感謝する。

==参考文献==

- (1) 栗原、吉田、鶴丸、藤田、「言語と試行のシミュレーション」、情報社会科学講座No.4、学習研究社、昭52
- (2) 田中穂積、仁科喜久子、上位/下位関係ソーラス I S A M A Pの作成 [I]、[II]、情処研報、Vol.87, No.84
- (3) 技術報告:「概念辞書」第3版、日本電子化辞書研究所、1991
- (4) 時枝誠記:「国語学原論」、岩波書店、1941
- (5) 三浦つとむ:「認識と言語の理論、Vol.1~3」、昭草書房、1967
- (6) J.Barwise and J.Perry.:Situation and Attitudes, J. of Philosophy, Vol.78, pp.668-691
- (7) 高松忍、西田富士夫、動詞パターンと格構造に基づく英日機械翻訳、信学誌、Vol. J-64-D, No. 9, PP.815-822
- (8) 長尾真:「言語工学」、昭晃堂、昭58
- (9) 大野晋、浜西正人「角川類語新辞典」角川書店、昭56年
- (10) 国立国語研究所、「分類語彙表」、秀英出版、昭和39年
- (11) 林大監修「角川小辞典NO.9:図説日本語」角川書店、昭57年
- (12) 水谷他、「文法と意味I」、朝倉日本語新講座No.3、朝倉書店、1983.9
- (13) 池原、宮崎、白井、林、「言語における話者の認識と多段翻訳方式」、情処論、Vol.28, No.12, pp.1269-1279, 1987
- (14) S. Ikehara, "Multi-Level Machine Translation System", Future Computer Systems, Vol.1, No.3, pp.261-274, 1989
- (15) S. Ikehara, M. Miyazaki, S. Shirai, A. Yokoo, "An Approach to Machine Translation Method based on Constructive Process Theory", Review of ECL, Vol.37, No.1, pp.39-44, 1989

付表 1. 文型の多い用言 ( : 付表2を参照のこと。)

用言の種類		見出し語 ( ) は文型数
慣用文型	和語系動詞 (20文型以上)	する(314), なる(240), ある(199), 掛ける(66), 付ける(60), 付く(47)取る(37), 掛かる(33), 出す(30), 入れる(26), 出る(25), 入る(24), 利く(23), 立つ(22), あげる(21), あがる(20), 打つ(20)
	漢語系動詞	要する(2), 決する(2), 白黒する(2)
	形容詞(20文型以上)	無い(172), 良い(123), 悪い(104), 高い(26),
	複数パターン持つ 慣用表現 (再掲)	4件 気に入る(7), 気に食う(5), 網を張る(4), 以上 気持ちを込める(4) 心を打つ(4) 3件 合点が行く, 申し訳が立つ, 手が入る, 音を上げる, 余裕がある手を引く, 油を注ぐ (その他:合計34件)
一般文型	和語系動詞(10文型以上)	出る(16), 掛ける(14), 下りる(14), 掛かる(11), 出す(11), 流れる(10), 下ろす(10)
	漢語系動詞 (6文型以上)	7件 運用する 6件 接する, 発する, 期する, 勧誘する, 開通する, 整理する, 成立する
	形容詞 (6文型以上)	7~ 甘い(9), 薄い(9), 厳しい(9), 痛い(8), 怪しい(8), 明るい(7) 6件 深い, 荒い, 危ない, おかしい, 大きい, 細かい
	形容動詞	6件 結構だ, 細かだ, 不定だ, 険悪だ

付表 2 構文辞書の例 (用言“掛ける”の場合)

用言	掛ける	( )内の数字は意味属性の深さ(段数)を示す。				
慣用文型	格種別	格要素となる名詞				
	に格 (17文型)	お目, べてん, 気, 計略, 策略, 手塩(注1), 手塩(注2), 尻目, 心, 天秤, 秤, 鼻, 方略, 謀略, 魔術, 魔法, 飾り, 畏 (注1) (人) が (人) を ~ : ~ bring up ~ with tender care (注2) (人) が (生物) を ~ : ~ tame ~				
一般文型	を格 (49文型)	ストップ, プレーキ, プレッシャー(注1), プレッシャー(注2), 圧力, 鎌, 颯, 気合い, 疑い, 疑念, 疑惑, 局所麻酔, 局部麻酔, 金, 嫌疑, 攻勢, 催眠術, 雑巾, 思い, 歯止め, 歯止め, 時間, 手間, 手間暇, 手数, 集合, 心配, 水, 声, 全身麻酔, 梯子, 電話, 謎, 売り込み, 拍車, 発破, 負担, 磨き, 魔術, 魔法, 麻酔, 無線電話, 命, 迷惑, 目, 容疑, 輪 (注1): (人工物, 人) が (人工物) に ~ : ~ apply pressure to ~ (注2): (主体) が (主体) に ~ : ~ put pressure upon ~				
	一般文型	N 1 (が格)	N 2 (を格)	N 3 (に格)・その他	英語文型	
主体(3)		美術(6), 時計(8), 縄・鎖(9), 鏡(9), 衣料(6), 像・書画(9),	住居(6), 枝(8)		N1 hang N2 on N3	
主体(3)		橋(7)	場(3), 場所(4)		N1 build N2 over N3	
主体(3)		-----	* (任意) 数量金銭(8), 時間(4)]		N1 spend 数量 on N3	
人(4)		腰(8)	椅子(9)		N1 sit down on/in N3	
人(4), 機械(6)		数(5)	数(5), 価格(8)		N1 multiply N3 by N2	
主体(3)		機械(6)	-----		N1 start N2	
主体(3)		勢い(8), 攻勢(9)	主体(3)		N1 make N2 upon N3	
人(4)		眼鏡(8)	-----		N1 wear N2	
主体(3)		錠・鍵(9)	住居(6), 車(9), 箱(9)		N1 lock N3	
主体(3)		通信機器(8), 音楽(6), 応用電子機器(8)	-----		N1 play N2	
合計		主体(3)	調味料(8), 液体(7), 薬品(6)	人(4), 具体物(3), 火(9)		N1 pour N2 on N3
14文型		主体(3)	布(8)	人(4), 身体(6), 家具類(8) (へ/に)		N1 spread N2 on/over N3
		主体(3), 機械(6)	物理現象(8)	具体物(3) (へ/に)		N1 apply N2 to N3
	主体(3)	縄・鎖(9), 装身具(8)	具体物(3)		N1 tie N2 around N3	

[注] 実際は解釈において、格の出現順序も評価の対象となる。[表現の構造と意味の関係]