

日英機械翻訳のための意味解析用の知識とその分解能

池原 悟[†] 宮崎 正弘^{††} 横尾 昭男[†]

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すること、また、それに合わせて必要な知識の種類とその分解能を見定めて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」の二つに分けることを提案した。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について、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（名詞語数約37万語）、構文意味辞書（文型数約1.3万文型）として収集整理することを試みた。両者は名詞の単語意味属性による記述を介して相互に結合されるため、それら辞書の記述能力は単語意味属性の分解精度で決まる。そこで、単語意味属性の分解精度と用言文型の記述能力との関係を調べると、従来の50~500種の意味属性分類では記述不能な用言の文型が頻出するのに対して、約3,000種に分類した場合は、一部の用言（補助用言など抽象度の大きいもの）を除いて、訳し分けに必要な文型がほぼ記述できることがわかった。文型記述能力向上の効果は、和語系の用言と専門用語の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、これによって記述された言語知識は機械翻訳において、用言と体言の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、訳しやす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わかった。

Classification of Language Knowledge for Meaning Analysis in Machine Translations

SATORU IKEHARA,[†] MASAHIRO MIYAZAKI^{††} and AKIO YOKOO[†]

In the translation process, various types of ambiguities arise, but this is regarded as a lacking the knowledge that is required. To overcome such ambiguity, there is a need to study what knowledge is lacking and to have such knowledge established as the rule or as a dictionary. Separation of meaning processing into meaning analysis and meaning understanding was proposed. And 2 major types of knowledge for meaning analysis are compiled into a form of dictionary for the purpose of solving ambiguity of sentence structure and of the meaning of individual words. The one is a semantic word dictionary (370,000 words) for knowledge related to semantic use of words. The other is a semantic structure dictionary (13,000 sentence patterns) for knowledge related to meanings of expression structures. As a means of describing these knowledge, the use of the meaning of words were compiled into a semantic attribute system with 3,000 attributes. The examinations shows that, in Japanese to English Machine Translation, precision level for some 3,000 types of semantic attributes are necessary for differentiation of translation for verbs. However, this is not sufficient for translation of nouns.

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な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¹⁾の研究があり、最近では意味処理用シソーラスや²⁾、EDRにおける概念辞書の研究³⁾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との関係などを解明した例は見当たらない。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。

[†] NTT 情報通信網研究所 知識処理研究部
NTT Network Information Systems Laboratories
^{††} 新潟大学工学部情報工学科
Department of Information Science, Faculty of Engineering, Niigata University

本論文では、言語過程説⁴⁾,⁵⁾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」の二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、実際の表現に使用された言語規範⁴⁾を調べる処理であり、「意味理解」とは、一般知識をも使用し、実際の言語表現と表現対象世界内の対象物や現象との結びつきを決定する処理である。この区分に従って、意味処理に必要な知識を、「意味解析」に必要な言語規範に関する知識（言語知識）と、「意味理解」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。

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⁴⁾を英語側の表現に対応させて記述する。構文構造を規定する情報は用言の見出し語とその用言の格として使用される名詞の意味的性質および助詞から構成されるため、構文構造の記述能力は、ほぼ名詞の意味属性の分解能で決まると言ってもよい。そこで、本論文では、単語意味属性の分類の深さに着目して、構文意味辞書に登録された個々の構文構造とそこに使用された単語意味属性のレベルとの関係から、単語意味属性の分解精度と構文構造の記述能力の関係を明らかにする。また、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その他の機能についても述べる。

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（時枝誠記）によれば、図1に示すように、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⁴⁾。「対象」には物理的実在としての対象と観念的

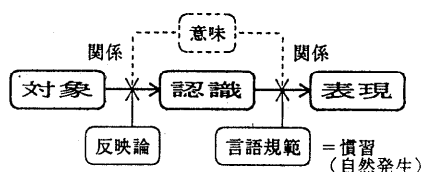


図1 言語過程説（時枝誠記）

Fig. 1 Constructive process theory.

存在としての対象があるが、そのいず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⁵⁾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³⁾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ものに対して、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、言語理解を以下の二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ともに、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、図2のように、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

⁴⁾ 単語の用法、文法など、それぞれの言語の持つ社会的な約束全体をいう。

⁵⁾ 表現の意味は単にそれを構成する単語の意味だけではなく、表現の構造の持つ意味からも構成される。表現を個々の単語に分解してしまい、単語ごとに訳した結果を持ち寄っても、元の表現の意味が復元できない場合が生じる。ここでは分解することによって失われる意味を表現構造の持つ意味と称する。

³⁾ 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⁶⁾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。

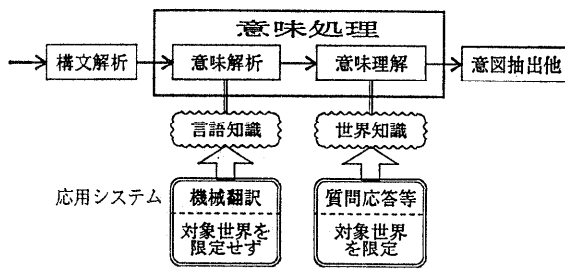


図 2 意味解析と意味理解
Fig. 2 Meaninganalysis and meaning understanding.

ユ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⁵⁴。

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、図3に示すように、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、主体も実体に含め、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、実体と属性の関係の3種があるが、概念化⁵⁵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

⁵⁴ 「意味理解」に必要な情報としては、このほか、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。

⁵⁵ 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。

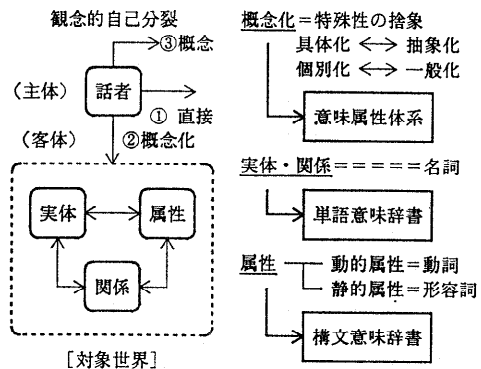


図 3 対象世界の概念化
Fig. 3 Conceptualization of the object world.

構成要素を、実体と属性の二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、実体はいずれも言語表現上、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、「意味解析」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⁵⁶。

⁵⁶ 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^{57,58}。その違いは必ずしも明確ではないが、大きくみて、単語の意味をさらに細かな要素に分ける立場と、単語の意味を対象の持つ特徴の総体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

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わかる。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」、恋人の関係であれば「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」という語義を持つが、現実の表現で使われたときは、「子供の集まりとして見た学校」や、「教育の機関として見た学校」、「ある地域の特定の場所として見た学校」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。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」などとなる^{*7}。

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異なるため、「意味解析」の対象言語ごとに、異なった単語意味属性の体系が必要と考えられる。また、細かく見れば、対象に対する見方、捉え方は単語ごと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よいと、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彙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^{*8}。

(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、木構造の形に整理する。

(a) 一般名詞意味属性体系 (12 段約 2,800 ノード)

一般名詞意味属性は名詞全般の意味属性を記述するために使用される。

(b) 固有名詞意味属性体系 (9 段約 130 ノード)

一般名詞意味属性のうち、固有名詞に該当する部分をより詳細化した意味属性で構成される。

固有名詞を含む複合語等の解析では、固有名詞について、一般名詞意味属性より細かい精度の意味属性の分解能が必要になるため、部分的に細分化した別の属性体系として作成する。一般名詞意味属性が文型辞書とのパターンマッチを中心とする意味解析で使用されるのに対して、固有名詞意味属性は固有名詞を含む表現（複合語＝複合名詞）の解析に閉じた範囲で使用される。

各意味属性体系の上位ノードの一部を図 4、図 5

りを持った総体であり、意味の要素に分けられないものと考えられる。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される。

^{*7} 従来、語義を分類するものとして、類語辞典⁹⁾や分類語彙表¹⁰⁾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¹¹⁾、シソーラス化したもの¹²⁾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。

^{*8} 語彙統計¹³⁾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 2,000~3,000 語とみられること、基本英語 (Basic English) の単語は 1,000 語以下であること、また、通常使いきなされる漢字は 2,000~3,000 程度であることなどを参考に、単語意味属性項目数は 3,000 項目を目標とした。本研究の初期には 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。

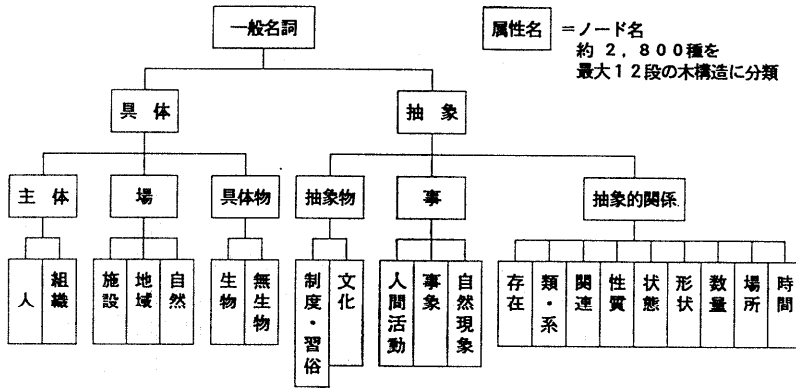


図4 一般名詞意味属性体系の例 (上位4段まで)
Fig. 4 Example of meaning attributes for general nouns.

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6に示す。これより、一般名詞意味属性は8~9段の付近で最も横に広がり、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、そのままでは収斂せず7段目以降で収斂している。これは「人名」、「組織名」のノードが4~5段目で収斂するのに比べて、「地名」のノードが深くなっているためである。

(4) 単語意味辞書

一般語 12 万語、固有名詞 20 万語、専門用語 5 万語、その他 3 万語の合計約 40 万語の日本語辞書に含まれる名詞約 37 万語を対象に、前述の意味属性を付与し、単語意味辞書を作成した。

一般名詞には一般名詞意味属性のみを付与するのに対して、固有名詞には一般名詞意味属性と固有名詞意味属性の双方を付与する。これは、固有名詞も単独で格要素として使用される場合、文型辞書と直接パターンマッチが図れるようにするためである。

単語意味辞書の項目の例を図7に示す。図では単語「東京」に対して、二つの一般名詞意味属性と四つの固有名詞意味属性が付与されている^{*)}。

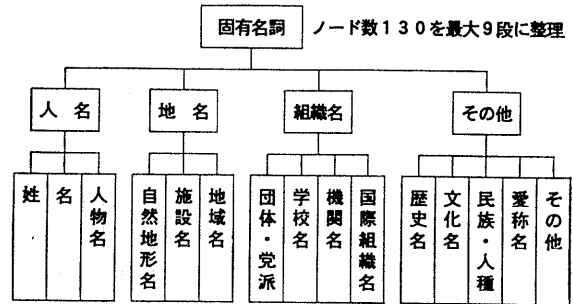


図5 固有名詞の意味属性体系の例 (上位3段まで)
Fig. 5 Example of meaning attributes for proper nouns.

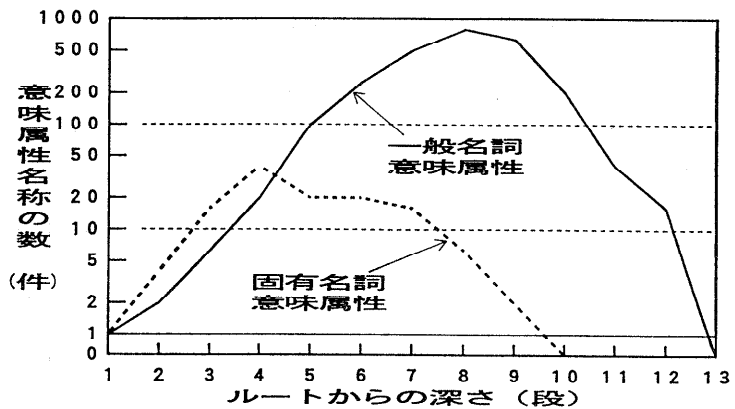


図6 意味属性名称の数の分布の例
Fig. 6 Distribution of meaning attributes.

^{*)} 図では、単語「東京」の固有名詞意味属性として、「駅名」、「大学高専」などが付与されているが、これは固有名詞辞書項目として、実在する主な駅や学校などの知識を持たせようとしたものであり、「東京」+「駅」などの複合固有名詞の解析に使用する。これに対応し

て、一般名詞意味属性としても、「行政区画」、「駅・港」、「学校」が付与されている。これは、例えば、「列車は東京(駅)を出た。(leave)」、「船は東京(港)を出た。(depart)」、「彼は大学は東京を出た。(graduate from)」等での訳し分けに使われる。

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

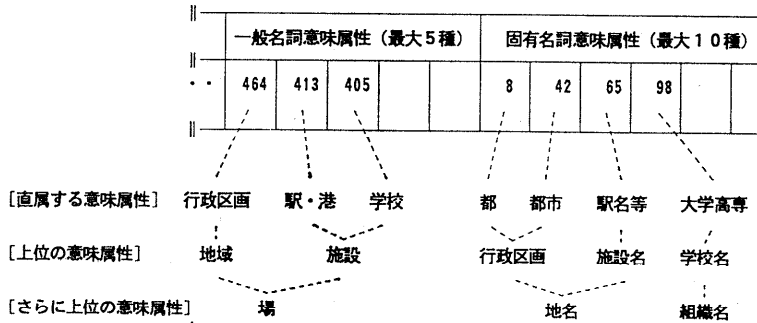


図 7 単語意味辞書の形式 (例)
 Fig. 7 Format of semantic word dictionaries.

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大半が、意味属性を一つしか持たないのに対して、和語は二つ以上の意味属性を持つことが多く、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のに対して、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、実体側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見れば、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^{*10}。

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、以下の二つのタイプに分けて整理する^{*11}。

a) 一般文型

用言の字面をキーとする意味的結合価パターンによる表現。用言の字面と一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、そのうち、名詞は意味属性が規定され

^{*10} 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。

^{*11} このようにして用言ごとに整理された体系を使用すれば、表現構造が意味単位ごとに捉えられるため、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。

る¹²。

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ほかは、a)と同じ。ただし、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¹³。

(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典等から収集した動詞、形容詞、形容動詞の見出し語約 6,000 語に対して、市販の和英辞書（研究社新和英、小学館プログレッシブ、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（現状では約 10,000 文型）にまとめた。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）等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（現状では約 3,000 文型）を作成した。また、このほかに、新聞の情報産業関連分野を対象とした若干の専門文型を作成した。

これらの意味辞書はまだ作成の過程にあり、文型の追加、統合を実施中であるが、文型のタイプや単語意味属性体系の文型記述能力など、意味辞書構成上の特徴や問題点は、おおよそ把握できる規模にある。

(4) 用言の種類と文型数の関係

現状での用言の種類と文型の数の関係を図8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わかる。

- 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）のほうが多い。また、形容動詞（1.5 件/語）に比べて形容詞（2.1 件/語）のほうが多い。

¹² 記述形式は Valentz パターン^{9),12)}と同様である（詳細は文献13）参照）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。

¹³ 例えば、「怠ける (idle away one's time)」という意味での「油を売る」の「油」は、「オイル」や「石油」などの単語に置き換えることができないため、意味属性を指定して一般化することができない。ここでは、このように、固定的に決められる名詞要素（格要素）を一つ以上持つ文型を慣用文型に分類している。その種類と形式の詳細は文献14）を参照。

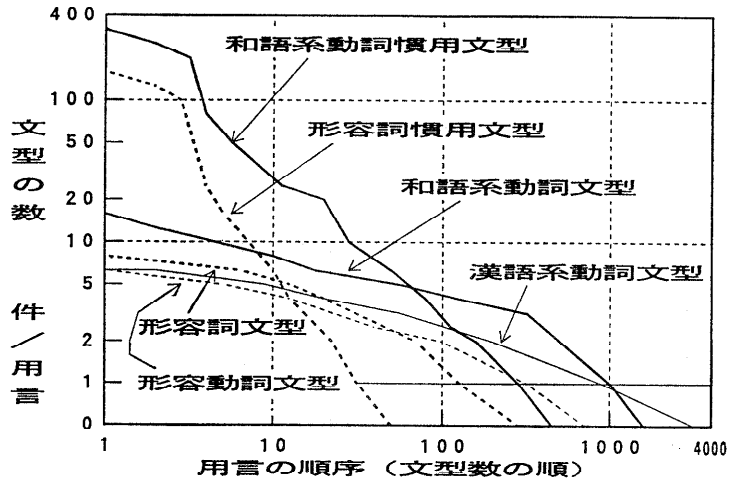


図8 文型の種類とその分布 Fig. 8 Types of semantic structure and their distribution.

- 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、下記の通りいづれも一般文型よりも慣用文型のほうが多い。

	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

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のがかなりたくさん存在する。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わかる。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な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6から意味属性の深さに対する意味属性名数の累積分布を求めると図9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、以下の三つの場合について考える¹⁴。

¹⁴ 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、必ずしも対応関係にない。し

場合	属性名称の数	相当する段数	記事
Case.1	50件	約4段	従来の標準的な例
Case.2	500件	約6段	本論文の例
Case.3	3,000件	8~9段	

ただし、「相当する段数」は意味属性体系内のノードの深さを示す。

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10に示す。図より、以下のことがわかる。

①一般名詞意味属性、固有名詞意味属性ともに属性名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わかる。

②Case 1, Case 2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。

(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、それを規定する意味属性の深さを図11に示す。次に、各文型ごと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12の結果を得る。図12を用言の種類に着目

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。

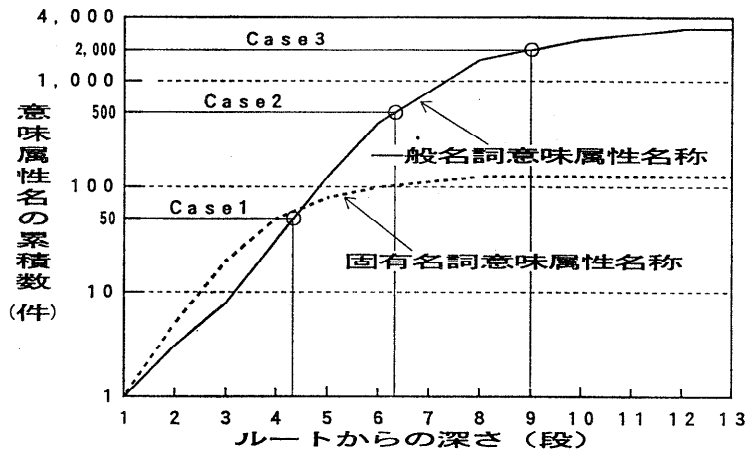


図9 意味属性名称の数の累積分布
Fig. 9 Cumulative distribution of meaning attributes.

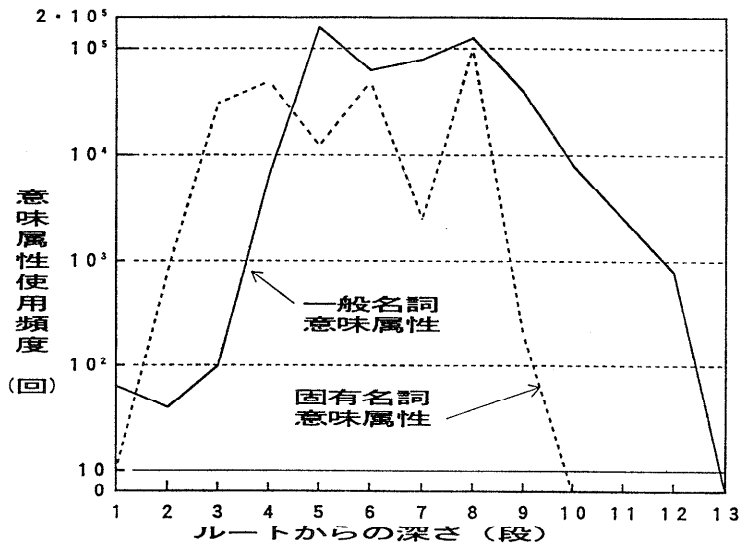


図10 意味属性の使用頻度 (単語意味辞書の例)
Fig. 10 Frequency in use of a meaning attribute at the semantic word dictionary

して分解すると図13の結果を得る。ただし、図11、図12の慣用文型は、字面以外の登録格要素の同一性に着目して縮退させたものを示す。これらの図から以下ことがわかる。

①文型内の格要素に使用される意味属性は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^{*15}。

*15 構文意味辞書では一般文型、慣用文型ともに、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、

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、深いほうに移動している。また、文型種別ごとに見ると、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わかる。

③和語系と漢語系の用言を比べると、和語系の用言のほうが深いレベルの意味属性（9～10 段目まで）を必要としている。

これらのうち、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。

上記の結果から、9～10 段目までの意味属性を使用すれば、13,000 文型は、ほぼ問題なく記述できることがわかる。この場合、意味属性の9～11 段目は、図9から意味属性分類数 2,000～3,000 項目に相当するから、用言の訳し分けに必要な文型辞書を記述するのに必要な名詞の意味属性分類精度は、2,000～3,000 種類と判断できる。

4.3 言語知識記述能力について

図10～図13の結果から、Case 1～Case 3 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。表では、前節の結果に基づき、Case 3 (意味属性のノード数 2,000～3,000 の場合) の文型記述能力を100 と仮定している。この表から以下のことがわかる。

- 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 Case 1, 2 に比べて Case 3 のほうが高い^{※16}が、その差は一般文型では和語系、全体から見れば専門

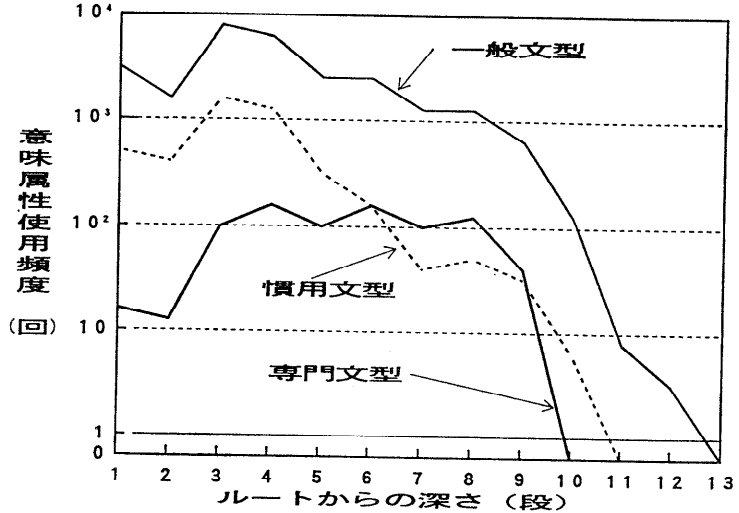


図11 意味属性の使用頻度 (構文意味辞書の例)
Fig. 11 Frequency in use of a meaning attribute at the semantic structure dictionary.

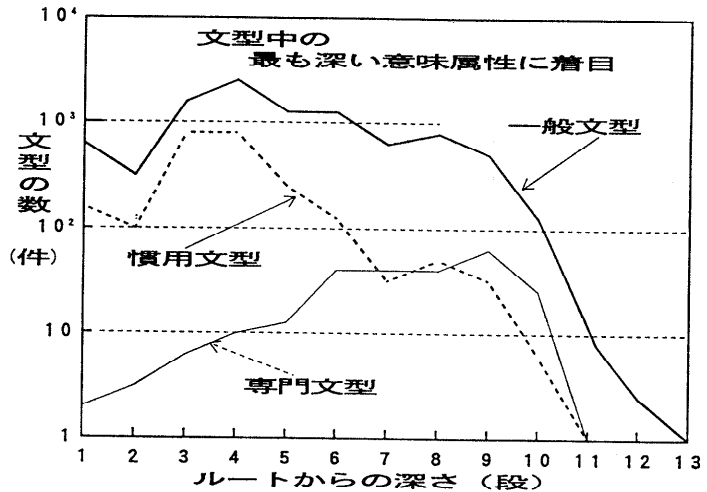


図12 パターン記述の深度分布例
Fig. 12 Distribution of description depth of sentence structure.

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。しかし、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。

※16 ただし、EDR の概念辞書における用言と名詞の関係は、関係子に意味を持たせた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には情報の質に差がある。

文型において顕著である。

このうち、①からは、後で述べるように、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言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/E¹⁵⁾⁻¹⁷⁾ 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 3,000 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使われているか、それとも文字通りの意味で使われているかの判断が容易になり、副作用の心配が減った。慣用的表現の多出する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。付図 1 に動詞「掛ける」訳し分けの例を示す。

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やす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わかった。

また、サ変名詞などの用言性名詞を含む複合語で

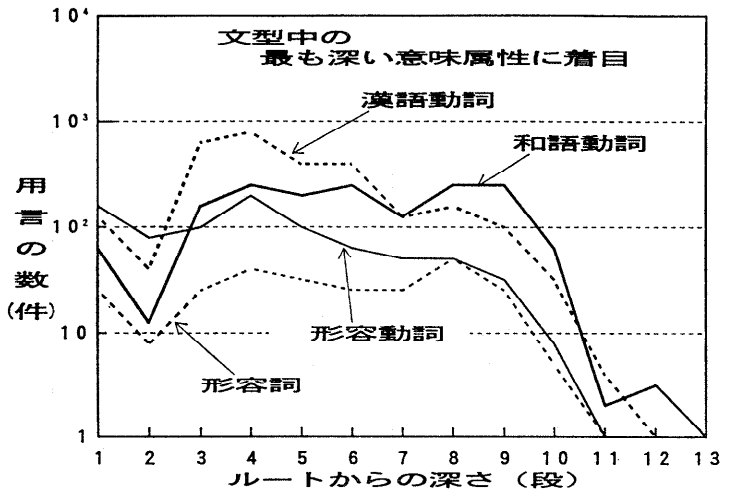


Fig. 13 文型記述の精度 (一般文型の例) Distribution of description depth of general sentence structure.

表 1 意味属性精度と言語知識記述能力の関係* Table 1 The relation between classification degrees of meaning attributes and description capabilities of language knowledge.

比較項目	比較する場合	場合 1	場合 2	場合 3	
意味属性のノード総数		約 50	約 500	約 3,000	
ノードの実行的な深さ		約 4 段	約 6 段	約 9 段	
単語意味辞書 名詞語義の記述 可能範囲	一般名詞	3%	22%	100% (仮定)	
	用言性名詞	21%	25%	100% (仮定)	
構文意味辞書 格要素または文型 の記述可能な範囲	一般文型	和語動詞	[31%]	[59%]	100% (仮定)
		漢語動詞	[57%]	[86%]	100% (仮定)
		形容詞	[38%]	[60%]	100% (仮定)
		形容動詞	[67%]	[84%]	100% (仮定)
	合計	52% [52%]	78% [81%]	100% (仮定)	
・記述可能な格要素の割合を示す ・ただし、[]内の数値は記述可能な文型の割合を示す。	慣用文型	81% [80%]	95% [95%]	100% (仮定)	
	専門文型	11% [10%]	36% [34%]	100% (仮定)	

* 本論文の意味属性で記述できる意味辞書の記述可能範囲を 100% とした場合の相対評価を示す。

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、構文意味辞書の二つの辞書にまとめる方法について提案した。

名詞 37 万語の単語意味辞書と 1.3 万文型の構文意味辞書を記述する実験を行った結果、単語意味属性の分解精度と用言文型の記述能力の関係では、従来の約 50 種または約 500 種の意味属性分類では、記述不能の文型が続出するのに対して、約 3,000 種に分解精度を上げた場合は、機能動詞などの一部の用言を除いて、1.3 万文型が問題なく記述できることがわかった。この場合の文型の記述能力向上の効果は、和語系用言の持つ文型や専門文型について著しく、記述能力の差は、前者では 2~3 倍、後者では 3~10 倍に向上していることがわかった。

単語意味属性を介して記述されたこのような単語辞書と構文意味辞書の知識は、用言と体言の意味の訳し分けに効果があるだけでなく、複合語解析や文脈処理、翻訳しやすい日本語への自動書き換えなどのルール記述においても有効であることがわかった。

本論文では、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 種類を、言語知識（言語規範）として、辞書化することを提案したが、今後、提案した方法に沿って、言語知識の体系化を進め、翻訳実験を通じてその内容の充実と整合性の向上を図って行きたい^{*17}。

対象の概念化のレベルを第 1 回の概念化、対象間の関係を概念化を第 2 回の概念化と見ると、本論文の提案は 1.5 回概念化のレベルに相当する。名詞句の解釈などの曖昧性解消を考える場合は、第 2 回の概念化のレベルまで言語知識を体系化する必要があると考えられるため、今後、実体間の関係知識の分類法について検討を進めるとともに、文脈や長文の解析に必要な属性間の関係知識についても検討を進めて行きたい。

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、中岩研究主任ほか、翻訳研究グループの

皆様に感謝する。

参考文献

- 1) 栗原, 吉田, 鶴丸, 藤田: 言語と思考のシミュレーション, 情報社会科学講座, No. 4, 学習研究社 (1977).
- 2) 田中穂積, 仁科喜久子: 上位/下位関係シソーラス ISAMAP の作成 [I], [II], 情報処理学会研究会報告 87-NL-64 (1987).
- 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, 勁草書房 (1967).
- 6) Barwise, J. and Perry, J.: *Situation and Attitudes*, Cambridge MA: MIT Press. (1983).
- 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, 信学論, Vol. J-64-D, No. 9 pp. 815-822 (1981).
- 8) 長尾 真: 言語工学, 昭晃堂 (1983).
- 9) 大野 晋, 浜西正人: 角川類語新辞典, 角川書店 (1981).
- 10) 国立国語研究所: 分類語彙表, 秀英出版 (1964).
- 11) 林 大監修: 角川小辞典 No. 9: 図説日本語, 角川書店 (1982).
- 12) 水谷ほか: 文法と意味 I, 朝倉日本語新講座, No. 3, 朝倉書店 (1983).
- 13) 林, 奥, 石崎: 日英翻訳システム ALT-J/E における日英変換技術, 第 33 回情報処理学会全国大会, 6 J-3 (1986).
- 14) 奥: 日本文解析における述語相当の慣用表現の扱い, 情報処理学会論文誌, Vol. 31, No. 12, pp. 1727-1735 (1990).
- 15) 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, 情報処理学会論文誌, Vol. 28, No. 12, pp. 1269-1279 (1987).
- 16) Ikehara, S.: Multi-Level Machine Translation System, *Future Computer Systems*, Vol. 1, No. 3, pp. 261-274 (1989).
- 17) Ikehara, S., Miyazaki, M., Shirai, S., Yokoo, A.: An Approach to Machine Translation Method Based on Constructive Process Theory, *Review of ECL*, Vol. 37, No. 1, pp. 39-44 (1989).
- 18) 池原, 白井: 日英機械翻訳機能試験項目の体系化, NLC 90-43 (1990).
- 19) 黒橋, 長尾: 格フレーム選択における意味マーカと例文の有効性, NL-91-11 (1992).
- 20) Utsuro, T., Matsumoto, Y. and Nagao, M.: Lexical Knowledge Acquisition from Bilingual Corpora, *Proc. of the 14th International Conference on Computational Linguistics*, Nantes, France, pp. 581-587 (1992).

*17 最近、翻訳用例の中から翻訳したい文と類似の文型を探して翻訳する試み¹⁹⁾や、文型辞書自体を対訳集から半自動的に生成する方法²⁰⁾についての研究が見られる。今後、文型辞書の品質や訳文品質上の効果などの比較や辞書開発の効率化等の問題への取り組みも必要と考えられる。

付表 1 文型の多い用言の例 (「掛ける」の例の詳細は付表 2 を参照)
Appendix: Table 1 Declinable words which have many sentence structures.

用言の種類	見出し	言 () は文型数
用言	和語系動詞	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), 他
用言	漢語系動詞	要する (2), 決する (2), 白黒する (2)
文型	形容詞	無い (172), 良い (123), 悪い (104), 高い (26),
文型	複数パターンを持つ慣用表現 (再掲)	4件 気にいる (7), 気に食う (5), 綱を張る (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, 他
用言	和語系動詞	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)
用言	漢語系動詞	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
文型	形容詞	7~ 甘い (9), 薄い (9), 厳しい (9), 痛い (8), 怪しい (8), 明るい (7)
文型	形容動詞	6件 渋い, 荒い, 危ない, おかしい, 大きい, 細かい
文型	形容動詞	6件 結構だ, 細かだ, 不定だ, 険悪だ

(*1)全体が3,000文型の時の例

(*2)全体が1万文型の時の例

兎が罠に掛かった。	A rabbit fell into its snare.
壁に絵が掛かっている。	A picture has hung on a wall.
彼にも容疑が掛かった。	He was also suspected .
彼女は催眠術に掛かった。	She was hypnotized .
彼は歯医者に掛かっている。	He consults a dentist.
その案件が会議に掛かった。	That matter was brought up at a meeting.
彼の足に波のしぶきが掛かった。	The splash of a wave splashed on his leg.
ストーブにやかんが掛かっていた。	A kettle had been placed on a heater.
テーブルにテーブルクロスが掛かっている。	A tablecloth is spread over a table.
彼女は花に水を掛けた。	She poured water on a flower.
母は神様に願を掛けた。	A mother prayed to God.
警官は泥棒に縄を掛けた。	A policeman tied up a thief with a rope.
豊臣軍は城に火を掛けた。	Toyotomi Army set fire to a castle.
私は彼らに迷惑を掛けた。	I caused them trouble.
彼は二階に梯子を掛けた。	He placed a ladder up to the second floor.
彼は梯子に腰を掛けている。	He is sitting down on a chair.
彼らは砂をふるいに掛けた。	They sifted sand.
建設省はここに橋を掛けた。	The Ministry of Construction built a bridge here.
彼女は毎日廊下に雑巾を掛けている。	She mops up a corridor every day.
彼女はプレゼントにリボンを掛けた。	She tied ribbon around a gift.
彼女は食卓にテーブルクロスを掛けた。	She spread a tablecloth on a dining table.
あの喫茶店はモダンジャズを掛けている。	That coffee shop is playing modern jazz.
映画離れに歯止めをかけることが必要だ。	It is necessary that the brake is about to be applied to a movie detached room.
私は試験にやまをかけた。	I took a chance in preparing for an examination.
富士通は全国に売り込みをかける。	Fujitsu Ltd. mounts a sales campaign in the whole country.
個数を単価にかけて、売上は合計を計算した。	A clerk multiplied a unit price by the number and calculated the total.

縄を枝から枝にかけて、その子供たちは遊んだ。

Those children **hung** a rope on a branch from a branch and played.

[注] テスト文は日英機械翻訳用機能試験文¹⁹⁾の3章「用言の訳し分け」の一部を使用

付図 1 動詞「掛かる」「掛ける」訳し分け効果の例 (ALT-J/E システム)

Appendix: Fig. 1 'Translations of Verb "kakaruru" and "kakeru".

付表 2 構文意味辞書の例 (用言“掛ける”の場合)
Appendix: Table 2 Example of a semantic structure dictionary.

用言	掛ける ()内の数字は意味属性の深さ(段数)を示す。			
構文型	格種別	格要素となる名詞		
	に格	お目, べてん, 氣, 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, 翼 (注1) (人) が (人) を ~: ~bring up ~ with tender care (注2) (人) が (生物) を ~: ~tame ~		
を格	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, 鐘, 腕, 氣合い, 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, 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 ~			
舟文型	N1 (か格)	N2 (を格)	N3 (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
	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
	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]

(平成4年6月7日受付)
(平成5年6月17日採録)

言語処理の研究に従事。現在、日英機械翻訳システムにおける日英構造変換処理や翻訳辞書の研究に従事。電子情報通信学会，人工知能学会各会員。



池原 悟 (正会員)

昭和19年生。昭和42年大阪大学基礎工学部電気工学科卒業。昭和44年同大学大学院修士課程修了。昭和44年日本電信電話公社に入社。現在、NTT 情報通信網研究所知識処理研究部勤務。この間、数式処理、トラヒック理論、自然言語処理の研究に従事。工学博士。昭和57年情報処理学会論文賞受賞。電子情報通信学会，人工知能学会各会員。



宮崎 正弘 (正会員)

昭和21年生。昭和44年東京工業大学工学部電気工学科卒業。同年日本電信電話公社に入社。平成元年より新潟大学工学部情報工学科教授。この間、大型情報処理装置DIPSの開発、計算機システムの性能評価法の研究、日本文音声出力システムや機械翻訳などの自然言語処理の研究に従事。現在、自然言語理解、機械翻訳などの研究に従事。工学博士。電子情報通信学会，人工知能学会各会員。



横尾 昭男 (正会員)

昭和32年生。昭和55年電気通信大学電気通信学部電子計算機学科卒業。同57年同大学院電子計算機学専攻修士課程修了。昭和57年日本電信電話公社に入社。現在、NTT 情報通信網研究所知識処理研究部勤務。この間、自然