

意味的等価変換方式のための 重文複文の統語的意味的分類体系について

池原悟[†] 阿部さつき^{††} 竹内奈央^{†††} 徳久雅人[†] 村上仁一[†]

[†] 鳥取大学工学部 〒680-8552 鳥取市湖山町南4-101

^{††} NTTアドバンステクノロジー株式会社 〒210-0007 川崎市川崎区駅前本町12-1 川崎駅前タワービル

^{†††} 言語アナリスト 〒700-0082 岡山市津島中1-3 RA-202

Email: '{ ikehara/ murakami/ tokuhisa@ike.tottori-u.ac.jp } "satsuki.abe@ntt-at.co.jp "'nao@spa.nifty.com

あらまし 従来の要素合成法の限界を超える意味処理の基本方式として、最近、与えられた言語表現を、類推の原理によって、同じ意味の異なる表現に写像する「意味的等価変換方式」が提案された。写像は、真理項と呼ばれる表現パターンの表す概念を介して実現される。そこで、本研究では、日本語重文と複文の統語的意味的分類体系を構築し、既に関与されている22.7万件の文型パターンの真理項としてそれらの分類コードを付与した。キーワード 意味的等価変換、重文複文、意味分類、統語分類、文型パターン

Analogical Mapping Method and Syntactic-Semantic Categorization of Japanese Compound and Complex Sentence Patterns

Satoru Ikehara[†] Satsuki Abe^{††} Nao Takeuchi^{†††} Masato Tokuhisa[†] Jin'ichi Murakami[†]

[†]Faculty of Engineering, Tottori University, Minami 4-101, Koyama-cho, Tottori city, 680-8552 Japan

^{††}NTT Advanced Technology Co.LTD., 12-1, Ekimaehonmachi, Kawasaki city, 210-0007 Japan

^{†††}Language Analyst, RA-202, naka 1-3, Tsusima, Okayama city, 700-0082 Japan

Email: '{ ikehara/ murakami/ tokuhisa@ike.tottori-u.ac.jp } "satsuki.abe@ntt-at.co.jp "'nao@spa.nifty.com

Abstract As one of the method to overcome the conventional method based on Compositional Semantics, Analogical Mapping Method that maps linguistic expressions into other expressions with the same meaning by the principle of analogy has been proposed. In this method mappings are performed via the concepts (called as True Item) represented by expression structures. Then this paper built a syntactic and semantic category system for Japanese complex and compound sentences and gave their codes to 220,000 sentence patterns included in the sentence pattern dictionary.

Key Words Analogical Mapping Method, Complex and compound sentence pattern, Semantic categorization

1. まえがき

機械翻訳では、さまざまな翻訳方式が研究されてきた [1, 2, 3] が、日英言語のような言語族の異なる言語間での翻訳は難しく、期待されるだけの訳文品質を得るに至っていない。最近、統計翻訳方式 [4, 5, 6] が注目されているが、異なる言語族間での翻訳では、困難さが予想される。実用システムの多くは、依然としてトランスファー方式が中心で

ある。

トランスファー方式は、統合構造と意味を分離して翻訳する方法である。解析の過程で原文の意味が失われることが多く、正しい翻訳を困難にしている。これに対して、パターン翻訳方式 [7, 8] は統語構造と意味を一体的に扱う方式であり、高品質の翻訳が期待されるが、意味的な排他性を考慮した大規模なパターン辞書を開発することが困難である。この問

題を解決する方法として、あらかじめパターン辞書を準備する必要のない用例翻訳方式[9、10、11、12]が期待されたが、置換可能な要素は各用例の持つ全体の意味との関係で決まるため、その判定を自動化することは難しい。

これに対して、最近、「意味的等価変換方式」が提案された[13]。また、言語表現の非線形性に着目した言語表現モデルが提案され、日本語の非線形な重文と複文の構造を対象に大規模な文型パターン辞書(22.7万パターン)が開発された[14]。「意味的等価変換方式」は、原言語の文型パターンと目的言語の文型パターンそれぞれを意味類型化(意味的に分類)し、両言語間で同一の意味を持つ文型パターングループを対応づける方法である。文型パターン間の写像は、表現パターンの表す概念(真理項と呼ばれる)を介して実現される。

そこで、本研究では、この方式の実現を目指し、日本語重文と複文の真理項となる統語的意味的分類体系を構築し、既に関与されている22.7万件の文型パターンに分類コードを付与した。付与した情報は文型パターンの意味的な検索に使用されるものであるから、入力文と文型パターンの構成要素を照合するパターンパーサの検索結果の絞り込みにも適用できる。本稿では、その可能性についても検討する。

2. 意味的等価変換方式の概要

(1) 方式概念

「意味的等価変換方式」は、表現構造の持つ意味(概念)に着目し、与えられた言語表現を意味的に等価な別の表現に写像するもので、以下の2つの理論を背景としている。

第1は、「人間の対象把握作用には、思考形式とも言うべきある種のフレームワークが存在し、それが言語表現に反映される」とする「意味類型論(セマンティック・タイポロジー)」[15]の考えで、意味のまとまる表現構造を形式化(パターン化)し、意味的に分類する。

第2は、「人間の独創性は、何らかの共通点を背景とした類推思考から生まれる」とする「等価的類推思考の原理」[16]で、これを言語表現に適用し、与えられた表現形式を言語共軛な概念(真理項)を介して、他の表現形式に写像する。

(2) 意味的等価変換方式の原理

「意味的等価変換方式」の最大の特徴は、「意味を

媒介とした類推」を基本としている点である。従来の変換が決定論的であるのと異なり、類推は非決定論的である。また、表現全体の意味を媒介としている点で、部分的な要素間の意味的類似性を基本とする用例翻訳とも異なる。

類推は、あるものを他の似たものに例えることによって思考する形式を持つ。一般的な規則を用いない点に特徴があり、「一般化する」、「覚える」、「思い出す」、「特殊化する」などの機能が有機的に結合したものとされている。

「意味的等価変換方式」は、人間による翻訳もこのような類推を背景としていることに着目し、異なる言語表現間の変換を(1)のようなプロセスで実現しようとしている。

$$A \quad P(A) \quad C(\quad) \quad B \quad (1)$$

但し、記号の意味は以下の通りである。

A : 言語 上の表現 A B : 言語 上の表現 B
 : 変換、 : 写像
 : 「論理の意味範疇」(真理項(概念)の集合)
 P(A) : 表現 A の形式を求める関数、
 C() : 真理項から言語 の形式を求める関数

この式は、言語 上の表現 A を言語 上の表現 B に変換する手順を示すもので、 の場合は、翻訳方式となり、 = の場合は、同一言語内での言い換え方式となる。変換の手順は以下の通りである。

<ステップ1> 原言語の表現形式の抽出

与えられた言語 の表現 A に対して、意味類型知識ベースから形式 P(A) を求める。

<ステップ2> 真理項への写像

形式 P(A) の属す意味類型の真理項 (「概念」のこと) の組を求める。

<ステップ3> 目的言語の意味類型の決定

得られた真理項の組に対応する言語 の意味類型 C() を求める。

<ステップ4> 目的言語生成

意味類型 C(B) に属す形式の一つを選び、それを使用して表現 B を生成する。

ここで、「真理項」は、表現形式の表す概念を表す言葉であり、与えられた表現は、真理項を介して、一つ以上の他の表現形式に写像される。

「真理項」を用いた「意味類型」間と「意味類型内」での意味的等価変換の仕組みの例を図1に示す。翻訳では、真理項を介して、原言語の表現構造が、

#	日本語表現の形式		#	英語表現の形式
1	X1はX2がX3するようX4する	真理項 の集合	1	X1 X4 so that X2 X3
2	X1は、X2が大変X3なのでX4できない		2	X2 is so X3 that X1 cannot X4
3	X1はX2がX3するといけないのでX4する		3	X1 X4 for fear that X2 X3
4	X1はX2するといけないのでX3した		4	X1 X3 not to X2
5	X1はX2しないようX3した		5	X1 is X3 for X1 is X2
6	もしX1がX2したら、X3はX4する		6	X3 X4 in the case X1 X2
7	X1がX2したらX3はX4した		7	When X1 X2, X3 X4
8	X1がX2したときX3はX4した		8	If X1 X2, X3 X4
9	X1がX2するならX3はX4してもよい		9	If X1 X2, X3 may X4

図1．真理項を介した形式の写像

目的言語の表現構造に写像されるが、同一言語内での車窓は、言い換えとなる。

(3) 機械翻訳への適用方法

ところで、意味的等価変換方式は、要素合成法が適用できない意味的に非線形な表現を対象にしている。線形な表現の場合は、要素合成法に基づく従来の方法が適用できるから両者を併用すれば、図2に示すようなハイブリッド型の新しい機械翻訳システムが構成できると期待される。

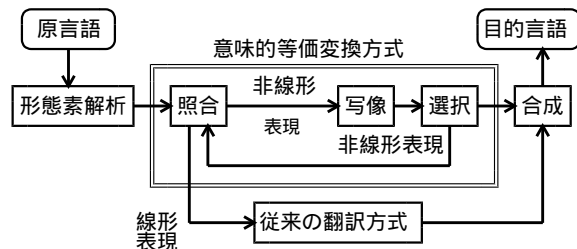


図2．言語の等価変換方式の構成

図2で、意味的等価変換方式内の閉ループは、レベルの異なる非線形表現について変換を繰り返し適用することを意味する。例えば、重文のパターンに含まれる節、句などの要素が、非線形な表現構造を持つ場合など、適合したパターン内に含まれる線形要素が非線形な表現構造を持っている場合である。

3．真理項の付与対象と設計条件

3.1 対象とする文型パターン

最近、「非線形言語モデル」に基づき、2つまたは3つの述部を持つ非線形な重文と複文を対象に表1に示すような大規模な規模な日英対訳パターン辞書が開発された[14]。

各文型パターンは、日英対訳例文を標本に、それらに含まれる線形な要素を下記の3段階に分けて汎化することによって得られたものである。

表1．文型パターン数

文種別	説明	単語レベル	句レベル	節レベル	合計
文種別1	接続1つ	55,509	36,063	17,641	109,513
文種別2	接続2つ	5,664	3,239	313	9,216
文種別3	埋込文1つ	42,492	28,451	5,038	75,981
文種別4	埋込文2つ	5,638	4,028	781	10,447
文種別5	接続1と埋込1の文	12,510	8,175	1,530	22,315
合計	-	121,913	79,956	25,603	227,472

- ・単語レベル：線形な自立語(名詞、動詞、形容詞、副詞など)を変数化したレベル
- ・句レベル：線形な句(名詞句、形容詞句、動詞句、副詞句など)を変数化したレベル
- ・節レベル：線形な節を変数化したレベル

そこで、以下では、これらの文型パターンに対して、意味的等価変換方式の実現に必要な真理項を付与するため、重文と複文の統語分類体系と意味分類体系を構築する。

3.2 真理項の設計条件

(1) 言語に共通する真理項の問題

一般に、同一の対象を見る場合でも、言語によって捉え方は異なり、それが表現に反映されるため、各言語で表現される概念は必ずしも共通とは言えない。これに対して、真理項は、原言語の表現形式と目的言語の表現形式の意味を仲介するための概念で

あり、原言語表現と目的言語表現に共通な概念であることが必要である。

この矛盾は、文型パターン対辞書では、表現レベルでの近似により解決されている。すなわち、原言語の各パターンに対して、目的言語のパターン1つが対応づけられているから、両者の真理項は等しいと考えることができる。そこで以下では、原言語の表現形式の表す概念を体系化して真理項とする。

(2) 真理項の意味分解能の問題

真理項は、同一の意味を持つ表現をグループ化できるだけの分類精度が要求されるが、表1で示されるように膨大な数の表現形式を単一の真理項の体系で意味分類することは、難しい。また、対象とする文型パターンは、大きく5種類の構造に分類されるが、その違いを考慮することも必要である。

そこで、本研究では、5種類の文種別に共通して使用できるよう以下の条件で真理項を体系化する。

- (1) 係り受け関係を持つ2つの節を対象に、関係の意味を分類体系化する。
「節間の意味分類」
- (2) 重文複文を構成する個々の節の意味を分類体

系化する。「節の意味分類」

(3) 各真理項は包含関係に基づき階層化する

(4) 各文種別のパターンには、節の構成に応じて複数の真理項を付与する。

従って、意味分類で使用する真理項は、「節間の意味分類」、「主節の意味分類」、「従属節の意味分類」の3つである。但し、この意味分類は同じ体系に従う。

(3) 統語分類情報について

方式上、真理項の体系としては、意味分類体系があればよいが、前項で述べたように、付与される真理項の種類と数は、対象とする文型パターンの統語的構造に依存している。

そこで、文型パターンの意味検索の対象範囲が指定できるよう意味的な分類に加えて、統語的な分類も行おう。

4. 統語的意味的分類体系の構築

真理項として付与するために構築した統語的意味的分類体系について述べる。

表2. 文型パターンの統語的分類

文種別	統語構造	文型の分類
文種別1	連用節 + 主節	副詞節 + 主節 並列節 + 主節
文種別2	連用節 + 連用節 + 主節 < 入れ子型 >	副詞節 + 副詞節 + 主節 副詞節 + 並列節 + 主節 並列節 + 副詞節 + 主節 並列節 + 並列節 + 主節
	連用節 + 連用節 + 主節 < 並列型 >	副詞節 + 副詞節 + 主節 副詞節 + 並列節 + 主節 並列節 + 副詞節 + 主節 並列節 + 並列節 + 主節
文種別3	連体節 + 名詞 + 述部	補足節 + 述部 名詞修飾節 + 名詞 + 述部
文種別4	連体節 + 名詞 + 連体節 + 名詞 + 述部	補足節 + 補足節 + 述部 補足節 + 名詞修飾節 + 名詞 + 述部 名詞修飾節 + 名詞 + 補足節 + 述部 名詞修飾節 + 名詞 + 名詞修飾節 + 名詞 + 述部
文種別5	連用節 連体節 + 名詞 + 述部 + 主節 < 第1節埋め込み >	副詞節 補足節 + 述部 + 主節 副詞節 名詞修飾節 + 名詞 + 述部 + 主節 並列節 補足節 + 述部 + 主節 名詞修飾節 + 名詞 + 述部 + 主節
	主節 連用節 + 連体節 + 名詞 + 述部 < 第2節埋め込み >	主節 副詞節 + 補足節 + 述部 主節 副詞節 + 名詞修飾節 + 名詞 + 述部 並列節 + 補足節 + 述部 並列節 + 名詞修飾節 + 述部
分類数	7分類	24分類

4.1 統語的分類体系

対象とする2つまたは3つの述部を持つ重文と複文の統語構造は、大きく見て、表2のように5種類に分類され、パターンを構成する節の種類と係り受け関係に着目すると、全体で24通りに分類される。

意味検索の範囲指定を行うためには、最低限、文種別レベルの分類で良いが、細かな指定ができるようすべての文型パターンに対して最下位のレベルまで分類することとした。

4.2 節間の意味分類体系

「節間の意味分類」は、主節と従属節の意味的關係を分類するものである。主節と従属節は2つの節間で定義する。

本研究では、このような2節間の意味分類体系として、益岡・田窪の研究[17]と実際の文型パターンに関する用例分析の結果に基づき新しい体系を作成した。

表3. 節間意味分類の分類数

#	第1段	第2段	第3段	第4段	合計
1	補足節	名詞節	5	5	11
	補語相当節	引用節	4	9	14
2	名詞節 連体修飾節	補足語修飾節	2	0	3
		内容節	0	2	3
		縮約的修飾節	0	7	8
		機能的表現	4	0	5
		の型接続	0	3	4
		その他	0	0	1
3	副詞節 副詞的連用節	時	3	19	23
		因果関係	4	10	15
		条件譲歩	6	13	20
		付帯状況・様態	3	11	15
		逆接	0	11	12
		目的	0	8	9
		程度	0	11	12
		前提	0	6	7
		手段	0	7	8
		二者間計	0	13	14
		相関	0	3	4
		主観感情	0	12	13
		場面	0	4	5
その他	5	4	10		
4	並列節	順接的並列	7	2	10
	並列型連用節	逆説的並列	0	0	1
計	4種	24種	43	162	233

増岡は、従属節を文法的に、「名詞節」、「連体節」、「連用節」、「並列節」の4種類に分類した後、それらの特徴や意味を説明している。そこで、まず、この研究結果を元に約100種類からなる意味分類体系を作成した。引き続き、約1000件の文型パターンを対象に、節間の意味を「時間的關係を表すもの」、

「空間的位置關係を表すもの」、「論理的關係を表すもの」、「心理的關係を表すもの」の4種類の観点から分類を詳細化し、約150種類からなる分類体系を作成した。これを12万件の文型パターンに適用する過程で、さまざまな問題点を吸収し、最終的に表3で示すような4段階の階層構造で233分類からなる新しい意味分類体系を作成した。

4.3 節の意味分類体系

節は単一の事象を表現したもので、その内容は単文に相当するため、節の意味分類は単文の意味分類に従うものとした。但し、時制、相、様相など助動詞で表現される意味は、重文複文全体の構造の持つ意味として扱うこととし、節の意味は、命題レベルの単文の意味として分類することとした。

ところで、単文は、動詞文、形容詞文(形容動詞文を含む)、「だ文」(名詞述語文)の4種類に分類できる。そこで、動詞文と形容詞文の意味は用言の意味属性分類に従い、「だ文」の意味は名詞の意味属性分類に従うものとして、用言と名詞に対して新しい意味分類体系を構築した。節(単文)の意味分類体系の分類数を表4に示す。

表4の動詞文の意味分類体系は、寺村の30分類[18]、中岩の36分類[19]、講談社類語大辞典[20]、IPALの日本語基本動詞辞書[21]などを参考に詳細化した。また、表5の名詞文の意味分類は、宮崎の一般名詞意味属性体系[19]、講談社類語大辞典[20]などを参考に編成した。いずれも、分類の意味の粒度がレベル毎に均一化するよう、フラットな構造に体系化した。

表4. 単文の意味分類体系の分類数

#	1段	2段	3段	4段	5段	合計
1	動詞文	知覚と情緒の表現	3	15	26	45
2		知的な行為の表現	6	17	20	44
3		日常生活の行為	4	17	10	32
4		地域社会生活行為	4	8	0	13
5		社会的活動の行為	7	23	9	40
6		現象事象の表現	9	29	11	50
7		変化の事象の表現	5	7	6	19
8		移動行為の表現	7	16	0	24
9		対物行為の表現	4	15	8	28
10	形容詞文	性状既定の表現	6	34	34	75
11	だ文	主体	4	26	16	47
12		場所名	6	23	0	30
13		具体物	6	40	0	47
14		抽象物	12	27	0	40
15		判断推定の表現	4	38	98	141
16		抽象的關係	9	55	0	65
計	-	16	96	390	238	740

<注>「形容詞文」は、形容動詞文(な型形容詞文)を含む

また、これらの意味分類に基づいて、用言6,000語、名詞6万語を対象に単語意味属性辞書を構築した。

5. 分類コードの付与

5.1 パターン種別と分類コードの関係

文型パターンは、すでに述べたように、単語レベル、句レベル、節レベルの3種類から構成されるが、これらは汎化のレベルの違いによるものであるので、包含関係にあるすべてのレベルに対して同じ真理項を付与した。

文型パターンの統語構造と意味分類コードの関係を表5に示す。このうち、第1欄から第3欄までは、統語的な構造によって文型パターンを分類するものであり、第4欄から第8欄まで5種類のコードは、文型パターンを意味分類するためのものである。なお、第9, 第10の欄は、節間の接続に用いられた特徴的なキーワードを登録したもので、検索結果の絞り込みなどで使用することを目的としている。

5.2 分類コード付与作業の半自動化

統語的意味の分類コードを付与する文型パターン辞書は膨大であり、人手で付与することは容易ではない。本研究では、以下で述べる方法で付与作業を半自動化した。

まず、統語分類コードは、標本文の構文解析情報が利用できるので、それを利用して付与し、誤りは人手で修正した。

次に、節の意味分類コードは、単語意味分類辞書

を用いて節の述部を構成する用言や名詞の意味属性を調べることにより半自動的に付与した。

これに対して、節間の意味分類コードは、従来の形態素解析や構文解析の情報だけで判断することはできないが、表現の形式とそれによって表される意味には一定の関係が認められる。そこで、以下に示す方法で、パターンの持つ表現形式を機械的に分類し、意味分類コードを付与した。

まず、重文の場合は、「接続助詞や機能語」、「主節と従属節の述部の統語的属性」に着目し、複文の場合は、「内の関係と外の関係」、「外の関係の時は、「先行詞の意味または意味分類(機能語の時は字面など)」に着目して、パターン分類用のテンプレート約90種類を作成し、各テンプレートに適合するパターンとそのパターンの表す意味(通常複数)との関係をまとめた。次に、テンプレート照合プログラム作成してパターンを分類し、各テンプレートに適合したパターングループ毎に、人手で意味分類コードを詳細化した。

6. 文型パターン絞り込みへの応用例

既に述べたように、文型パターンに付与された真理項は、入力文に適合した文型パターンの絞り込みにも利用できる。本章ではその例を示す。

(1) 実験の方法

実験では、入力文に対して、「パターンパーサ」によって検索した文型パターンを「意味検索プログラ

表5. 文型パターン分類コードの種類例

種別	統語構造の分類			節の意味分類			節間意味分類		節間KW分類		
	欄番号	第1欄	第2欄	第3欄	第4欄	第5欄	第6欄	第7欄	第8欄	第9欄	第10欄
付与情報	文種別番号	統語構造第1分類	統語構造第2分類	第1従属節の意味分類	第2従属節の意味分類	主節の意味分類	第1節間の意味分類	第2節間の意味分類	第1節間KW	第2節間KW	
文種別1	文種別=1	- -	副詞型 並列型	連用節の意味分類	- - - -	主節の意味分類	連用節と主節の関係	- - - -	節間KW	- - -	
文種別2	文種別=2	入子係型 並列係型	4種類 4種類	第1連用節の意味分類	第2連用節の意味分類		第1連用節との関係	第2連用節との関係	節間KW	節間KW	
文種別3	文種別=3	- -	補足型 名詞修飾型	連体節の意味分類	- - - -		埋め込み節との関係	- - -	節間KW	- - -	
文種別4	文種別=4	- -	- -	第1連体節の意味分類	第2連体節の意味分類		第1埋め込み節との関係	第2埋め込み節との関係			
文種別5	文種別=5	従属節埋込型 主節埋込型	4分類 4分類	連体節の意味分類 連用節の意味分類	連用節の意味分類 連体節の意味分類		連体節との関係 連用節との関係	連用節との関係 連体節との関係	節間KW	節間KW	

ム」によって絞り込むこととする。前者は、入力文と各文型パターンの構成要素間の包含関係を調べ、適合する文型パターンを検索するプログラムである。また、後者は、入力文と各パターンの意味分類コードを比較して、入力文の意味に適合する文型パターンを検索するプログラムである。

パターンパーサによる検索では、さまざまな照合条件が指定できるが、ここでは、意味的に適切なパターンをなるべく漏らさないようにするため、パターン要素に対する意味的な制約条件は適用しないこととする。

また、「意味検索プログラム」でもさまざまな検索が可能であるが、ここでは、入力文の節間意味コードと節の意味コードが完全に一致する文型パターンのみを検索することとする。

(2) 絞り込み実験の結果

入力文「私は友人を頼って上京した。」を対象に検索結果を表6に示す。

まず、パターンパーサによる検索では、表6に示すように多数のパターンが検索された。その中から、入力文との字面の適合数の多い順（正解率が高いと見られる順）に、上位5件を<参考例1>に示す。 、 ×の印は、入力文の翻訳への適用性の程度を示す。パターンパーサでは、通常、多数のパターンが抽出されるが、この例からも分かるように、上位のパターンでも意味的に不適切なものが多い。

これに対して、意味検索プログラムによって検索されたパターンは23件である。その一部（5件）を<参考例2>に示す。これらの結果から、適合パターンは16件に絞り込まれる。その上位5件のパターンを<参考例3>に示す。

文型パターン数はかなり絞り込まれ、意味的に適切な文型パターンの割合がかなり向上していることから、絞込の効果は、かなり期待できそうである。

<参考例1>パターンパーサ検索結果（上位5件）

- 【1】 /y\$1/tcfkN1を/cfV2(て|で)\$1^y/tckN3は /cf(V4.kako)ND4をした。 N3 V(V4)ND4.past when N3 V2.past N1.

- <例>話を聞いて彼は逆上した。He went wild when he heard that.
 ×【2】 /y\$1/tcfkN1を/cfV2(て|で)\$1^y/tckN3は /cf(V4.kako)ND4をした。 N3 be.past V(V4)ND4.past to V2 N1.
 <例> それを聞いて私は安心した。I was relieved to hear it.
 ×【3】 /y\$1/tcfkN1を/cfV2(て|で)\$1^y/tckN3は /cf(V4.kako)ND4をした。 V2^gm N1 V(V4)ND4.past.
 <例> 死体を見て彼は茫然自失した。
 Seeing the dead body freaked him out.
 ×【4】 /y\$1/tcfkN1を/cfV2(て|で)\$1^y/tckN3は /cfV4.kako。 With N(V2) of N3^poss N1 N3 V4.past.
 <例> 手を振って彼女は立ち去った。
 With a wave of her hand she went away.
 ×【5】 /y\$1/tcfkNP1を/cfV2(て|で)\$1^y/tckN3は /VP4.kako。 N3 V2.past NP1 to VP4.
 <例> 盲目という悪条件を克服して彼は偉大な学者になった。
 He overcame the handicap of blindness to become a great scholar.

<参考例2>意味検索プログラムの検索結果（5件）

- 【1】 /y<tkN1は>/cfいささかの/k知辺を/cf頼って /ytkN2へ/ (上っ|のぼっ|上ぼっ)た。 <|N1> went to N2, looking to a slight acquaintance for assistance.
 <例> いささかの知辺を頼って都へ上った。
 I went to town, looking to a slight acquaintance for assistance.
 【2】 /y\$1^y/tcfkN1は /ytkN2を/cf頼って \$1/ycf上京した。 N1 came to Tokyo from the country counting on N1^poss N2's help.
 <例> わたしはおばを頼って上京した。
 I came to Tokyo from the country counting on my aunt's help.
 【3】 /y<tkN1は>/tcfkN2/tck一人の/k知人を/cf頼って /ycf上京した。 <|N1> came to town, looking to an AJ(N2) friend for assistance.
 <例> たった一人の知人を頼って上京した。
 I came to town, looking to an only friend for assistance.
 【4】 /y<tkN1は>/tcfkN2を/cf頼って </ycfN3は>/cf(V4.kako)ND4をした。 <|N3> N(V4)ND4 where <|N1> could rely on <my|N1^pron^poss> N2.
 <例> 親類を頼って上京した。
 I went to Tokyo where I could rely on my relatives.
 ×【5】 /y<tkN1は>/tcfk雲の/k中を/tcfkN2に/cf頼って /ycfV3.kako。 <|N1> V3.past through the clouds with the help of N2.
 <例> 雲の中を計器に頼って飛んだ。
 We flew through the clouds with the help of the instruments.

<参考例3>絞り込みの結果（上位5件）

- 【1】 /y\$1^y/tcfkN1は /ytkN2を/cf頼って \$1/ycf上京した。 N1 came to Tokyo from the country counting on N1^poss N2's help.
 <例> わたしはおばを頼って上京した。
 I came to Tokyo from the country counting on my aunt's help.

表6 . パターンパーサとの併用によるパターン絞り込みの効果

#	検索プログラム	検索の条件	カウント方法	抽出された意味類型パターン数			
				単語レベル	句レベル	節レベル	合計
1	パターンパーサ	変数意味属性無視	延べパターン数	363 件	2,774 件	395 件	3,532 件
2	意味検索プログラム	真理項が一致	異りパターン数	15 件	8 件	0 件	23 件
-	絞込後の文型数	と に共通	延べパターン数(注)	6 件	10 件	0 件	16 件

(注) 機械翻訳では、適合の仕方に関する情報が必要となるため、「延べパターン数」でカウントする。

【2】 /y</tkN1は>/tcfkNP2を/cf頼って/ycf上京した。

<|N1> came to town, looking to NP2 for assistance.

<例> たった一人の知人を頼って上京した。

I came to town, looking to an only friend for assistance.

【3】 y</tkN1は>/tcfkN2を/cf頼って</ydfN3は>/cf(V4.kako)ND4をした。 <|N3> N(V4)ND4 where <|N1> could rely on <my|N1^pron^poss> N2

<例> 親類を頼って上京した。

I went to Tokyo where I could rely on my relatives.

【4】 /y</tkN1は>/tcfkNP2を/cf頼って/ydf上京した。

<|N1> came to town, looking to NP2 for assistance.

<例> たった一人の知人を頼って上京した。

I came to town, looking to an only friend for assistance.

× 【5】 /y</tkN1は>/tcfkN2を/cfV3(て|で)/ydf(V4.kako)ND4をした。

<|N1> N(V4)ND4 after having V3.past <my|N1^pron^poss> N2' consent by persuasion.

<例> 両親を説得して上京した。 I came to Tokyo after having obtained my parents' consent by persuasion.

5 . あとがき

意味的等価変換方式を実現するため、重文複文の統語的意味分類体系を構築し、既に関与されている文型パターン(22.7万件)に対して、分類コードを付与した。

対象とする日英文型パターンは、2つ又は3つの節から構成される重文と複文である。そこで、まず、「統語的な分類」としては、節の種類と構成に着目して3段階24種類に分類した。次に、「意味的分類」としては、「節間の意味分類」と「節自身の意味分類」を行った。「節間の意味分類」としては、従属節の種類に着目した益岡・田窪の分類体系を意味の観点から見直し、233種類に分類した。また、「節の意味分類」としては、述部の意味に着目して、動詞述部(295種類)、形容詞(形容動詞を含む)述部(75種類)、体言述部(370種類)の3種類に分類した(合計740種類)。その他、文型パターンには、これらの情報に加えて、節間キーワードも付与した。

付与された情報は入力文に統語的意味的に適合する文型パターン辞書の検索にも応用できる。入力文と文型パターンの構成要素を照合するパターンパーサの検索結果に適用した結果では、検索された多数の文型パターンの中から意味的に適切な候補を絞り込める可能性のあることが分かった。

本研究で構築した分類体系(真理項)は、まだ暫定的なものであり、今後、実験評価によって改良していく予定である。また、それと併行して、意味的等価変換方式への適用性も評価する予定である。

参考文献

- [1] 長尾真：自然言語処理，岩波書店，1996。
- [2] 池原悟：「機械翻訳」，in「言語情報処理」pp. 95-148, 岩波書店，1998
- [3] 田中穂積監修：「自然言語処理-基礎と応用」電子情報通信学会，岩波書店，1998
- [4] P. F. Brown, C. John, S. D. Pietra, F. Jelinek, J. D. Lfferty, R. L. Mercar and P. S. Roossin: A Statistical Approach to Machin Translation, Computational Linguistics, Vol. 16, No. 2, pp. 79-85, 1990
- [5] T. Watanabe and E. Sumita: Bidirectional Decoding for Statistical Machine Translation, Proceedings of COLING-02, pp. 1075-1085, 2002
- [6] S. Vogel, Y. Zhang, F. Huang, A. Tribble, A. Venugopal, B. Zhao and A. Waibel: The CMU statistical machine translation system. Proceedings of MT Summit IX, pp. 402-409. 2003.
- [7] K. Takeda: Pattern-based Machine Translation, the 16th COLING, Vol. 2, pp. 1155-1158, 1996.
- [8] H. Watanabe and K. Takeda: A Pattern-based machine translation system extended by example based processing, 17th COLING, pp.1369-1373, 1998.
- [9] M. Nagao: A Framework of a Mechanical Translation between Japanese and English by Analogy Principle, In A. Eithorn and R. Barneji (Eds.), Artificial and Human Intelligence, North-Holland, pp. 173-180, 1984.
- [10] E. Sumita, H. Iida, Experiments and prospects of Example-Based Machine Translation, 29th ACL, pp. 185-192, 1991
- [11] S. Sato: An example based translation and system, COLING-91, pp. 1259-1263, 1992.
- [12] Brown, R.D: Adding Linguistic Knowledge to a Lexical Example-Based Translation System, TMI 99, pp.22-32, 1999.
- [13] 池原悟：究極の翻訳方式の実現に向けて = 類推思考の原理に基づく翻訳方式 = =, AAMT Journal, アジア太平洋機械翻訳協会, No.33, pp.1-7 (2002.3)
- [14] 池原悟、阿部さつき、徳久雅人、村上仁一：非線形な表現構造に着目した日英文型パターン化、自然言語処理、Vol.11, No.3, pp. 70-95 (2004.7)
- [15] 有田潤：「ドイツ語講座II」南江堂, pp. 48-56, 1987.
- [16] 市川亀久彌 「独創的研究の方法論」(増補版), 三和書房, 1963
- [17] 益岡隆志、田窪行則：「基礎日本語文法」、くろしお出版(1992)
- [18] 寺村秀夫：「日本語のシンタックスと意味」くろしお出版(1982)
- [19] 池原悟、宮崎正弘、白井諭、横尾昭男、中岩浩巳、小倉健太郎、大山芳史、林良彦：「日本語語彙大系」岩波書店 1997.
- [20] 柴田武、山田進：「類語大辞典」、講談社(2002)
- [21] IPAL計算機用日本語基本動詞辞書、情報処理振興事業協会 技術センター、1987