

**戦略的創造研究推進事業 CREST**  
**研究領域「高度メディア社会の生活情報技術」**

**研究課題「セマンティック・タイポロジーによる  
言語の等価変換と生成技術」**

## **研究終了報告書**

研究期間 平成13年12月～平成19年3月

**研究代表者：池原 悟**  
**(鳥取大学工学部知能情報工学科 教授)**

# 目 次

1	研究実施の概要	5
1.1	はじめに	5
1.2	研究の概要	5
	(1) 文型パターン辞書の研究開発	
	(2) 文型パターンの意味類型化	
	(3) パターン検索プログラム	
1.3	研究成果	7
1.4	今後の課題	8
	(1) 変数意味属性の適正化	
	(2) パターン照合方式との整合性	
2	研究構想及び実施体制	9
2.1	研究構想	9
	(1) 研究の課題と方法	
	(2) 研究実施計画	
2.2	実施体制	10
	(1) 研究統括グループ	
	(2) 3グループ	
3	研究実施内容及び成果	12
3.1	意味的等価変換方式の提案	13
3.1.1	意味的等価変換方式の原理	13
	(1) 等価的類推思考の原理	
	(2) 意味的等価変換方式の原理	
3.1.2	文型パターンの意味類型化	15
	(1) 意味類型化の狙い	
	(2) パターン翻訳の場合	
	(3) 意味類型化されたパターン翻訳の場合	
3.2	非線形言語モデルの提案	16
	(1) 線形表現と非線形表現	
	(2) 言語表現のモデル	
	(3) 言語モデルの基本的特徴	
	(4) 言語モデルと文型パターン	
3.3	パターン記述言語の開発	19
3.3.1	パターン化の原則と枠組み	19
	(1) パターン化の原則	
	(2) パターン記述の枠組み	
3.3.2	変数化する線形要素	20
	(1) 線形要素の変数化の原則	
3.3.3	任意化とグループ化	21
	(1) 必須要素と任意要素	
	(2) 表現要素のグループ化	
3.3.4	語順の扱い	22
	(1) 順序任意要素指定記号	
	(2) 位置変更可能要素指定記号	

3.3.4	文型パターン記述例	22
3.4	単語意味属性体系の構築	23
3.4.1	対象とするパターン変数	24
	(1) 変数の意味的制約条件の付与 (2) 従来の意味属性体系の問題	
3.4.2	用言の意味属性体系	24
	(1) 用言の意味属性体系への要求条件 (2) 意味属性体系の作成手順	
	(3) 用言意味属性体系の概要	
3.4.3	名詞の意味属性分類	29
	(1) 名詞の意味属性分類の目的と方針 (2) 名詞意味属性体系の構成	
3.5	文型パターンの作成方法	30
3.5.1	パターン化の方針と汎化のレベル	30
	(1) 文型パターン化の方針 (2) 汎化の3レベル	
3.5.2	単語レベルの文型パターン化	32
	(1) 原文任意要素の削除 (2) 自立語の変数化 (3) 述部語尾表現の関数化	
	(4) パターン任意要素の指定 (5) 表現要素のグループ化 (6) 各種加工	
3.5.3	句レベルの文型パターン化	35
	(1) 句の変数化 (2) 丁寧表現の標準化 (3) 名詞から名詞句への拡張	
	(4) 動詞から動詞句への拡張 (5) 機能語の拡充	
3.5.4	節レベルの文型パターン化	36
	(1) 変数化する節 (2) 変数化する節の範囲	
3.5.5	文型パターンの例	36
3.6	文型パターン辞書の実験的改良	37
3.6.1	離散記号付与基準の再検討	38
	(1) 離散記号の詳細化 (2) 離散記号付与基準 (3) 離散記号挿入の作業基準	
3.6.2	述部語尾表現の汎化基準の見直し	39
	(1) 時制・相・様相情報の汎化の必要性 (2) 汎化基準の概要 (3) 汎化の効果	
3.6.3	英語パターン記述法の改良	41
	(1) 英語パターンの不良の分析 (2) 英語構文情報の付与	
	(3) 英語パターンの改良 (4) 見直しの効果	
3.6.4	高頻度適合パターンの削除	43
	(1) 削除するパターンの検出方法 (2) 意味的な正しさの判定方法	
	(3) 対象パターン数と効果の見積もり	
3.6.5	各種不良チェックによる改良	44
	(1) 自己例文への適合性の検証 (2) 変数適合表現のチェックと改良	
	(3) 字面表記に着目した改良 (4) 離散記号を利用した不良の発見と改良	

(5) 単文適合パターンの改良作業	(6) 局所言い換え辞書利用法の検討
(7) 個別的表現の点検と改良	
3.6.6 不適合試験文に着目した改良作業	46
(1) 狙いと意義	(2) クロス照合実験と結果の分類
(4) 作業支援環境の実現	
3.7 文型パターン辞書の特性	48
3.7.1 作成された文型パターン数	48
(1) 作成された文型パターン数	(2) 汎化された要素の割合
3.7.2 被覆率の評価	49
(1) 被覆率評価の方法	(2) 適合率の飽和特性
(3) 適合率と正解含有率	
(4) 意味的被覆率	
3.8 文型パターンの意味類型化	52
3.8.1 統語構造によるパターン分類	52
3.8.2 意味によるパターン分類	52
(1) 意味分類の方法	(2) 節間の意味分類体系
(3) 節の意味分類体系	
3.8.3 統語的意味的分類コードの付与	55
(1) 半自動化の方法	(2) 意味類型パターンの例
3.8.4 日英構文の対応関係の分析	59
3.9 パターン検索プログラム	61
3.9.1 パターンパーサ	61
(1) 基本アルゴリズム	(2) 照合結果の記憶
(3) モジュールの構成	
(3) 高速化の効果	(6) 検索例
3.9.2 意味検索プログラム	66
(1) 機能と構成	(2) 意味検索の実行例
(3) 本プログラムの利用方法	
3.10 翻訳簡易実験システムの試作	69
3.11 研究開発年表	70
(1) 線表	(2) 主な検討課題と検討結果
(2) 途中生産物とそれらの相互関係	
3.12 研究成果の今後期待される効果	77
4. 研究参加者	78
4.1 研究統括グループ	78
4.2 等価変換システムグループ	79
4.3 生成システムグループ	79

4.3	意味類型知識ベースグループ	80
5	招聘した研究者等	80
6	成果発表等	80
6.1	原著論文発表	80
6.2	その他の著作物	82
6.3	学会発表	83
6.3.1	招待講演	83
6.3.2	口頭発表	83
6.3.3	ポスター発表	92
6.3.4	プロジェクト主催の研究会発表	92
6.4	特許出願	96
6.5	受賞等	96
6.5.1	受賞	96
6.5.2	新聞報道	97
6.6	その他特記事項	97
7	研究期間中の主な活動	97
7.1	ワークショップ・シンポジウム	97
7.2	技術会議など	98
8	結び	101

## 1 研究実施の概要

### 1.1 はじめに

従来の自然言語処理では、言語表現の線形性を仮定した要素合成法が基本となっているが、現実には非線形な表現が多く、表現を単語に分解する過程で全体の意味が失われることが問題であった。この問題に挑戦するため、本研究では、「意味類型論（セマンティック・タイポロジー）」と「等価的類推思考の原理」の2つの観点から言語の「意味的等価変換方式」を提案し、それを実現するため、日本語重文複文を対象に意味類型パターン辞書を作成した。

「意味的等価変換方式」は、表現構造の持つ意味に着目し、与えられた言語表現を意味的に等価な別の表現に変換するもので、以下の2つのステップから構成される。

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」（有田潤1987）の考えに基づくもので、意味のまとまる表現構造をパターン化し、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」（市川亀久弥1963）を言語表現に適用するもので、与えられた表現構造を言語共軛な概念（真理項と呼ぶ）を介して、他の表現構造に写像する。異なる言語間の表現に変換する場合は、機械翻訳の技術となり、同じ言語内で変換する場合は、「言い換え技術」となる。

本方式は、意味的に非線形な表現構造を分解不能な単位とすることによって全体を線形近似に持ち込むものである。意味処理の基本技術として多くの応用が期待される。

### 1.2 研究の概要

本研究の主たる目標は、重文複文を対象とする「意味類型パターン辞書」を研究開発することである。「意味類型パターン辞書」は、「文型パターン辞書」を意味類型化したものである。そこで、「文型パターン辞書」を試作し、それを意味的に類型化することによって「意味類型パターン辞書」を開発した。その過程で、それらの品質を評価し、さまざまな改良を実施するため、「パターン検索プログラム」を試作した。

図1.2.1に「意味類型パターン辞書」を実現した過程を示す。

#### (1) 文型パターン辞書の研究開発

重文複文の表現を意味的にほぼ網羅するような表現意味辞書を実現することを目標に文型パターン辞書を研究開発した。

まず、言語表現とそれを構成する要素の線形性と非線形性の定義を明確にし、非線形言語モデルを提案した。この言語モデルは、「すべての言語表現は、0個以上の線形要素と一つ以上の非線形要素から構成される」とするものである。

次に、このモデルに基づいて、文型パターン記述言語を設計した。この言語は、「字面」、「変数（17種類）」、「関数（10種類145関数）」、「記号（10種類）」の4種類の要素を用いて文型パターンを記述するものである。線形要素は、原則として字面以外

の要素で記述される。線形要素の中でも変数の変域を的確に指定することが重要である。そこで、重文複文のパターン記述の観点から用言と名詞に対する新しい意味属性体系を構築した。用言意味属性体系は、日本語用言を意味的用法によって4段366種類に分類するもので、これに基づく約6,000語の用言意味辞書を作成した。また、名詞意味属性体系は、日本語名詞を意味的用法によって4段364種類に分類するもので、これに基づいて約6万語の名詞意味辞書を作成した。

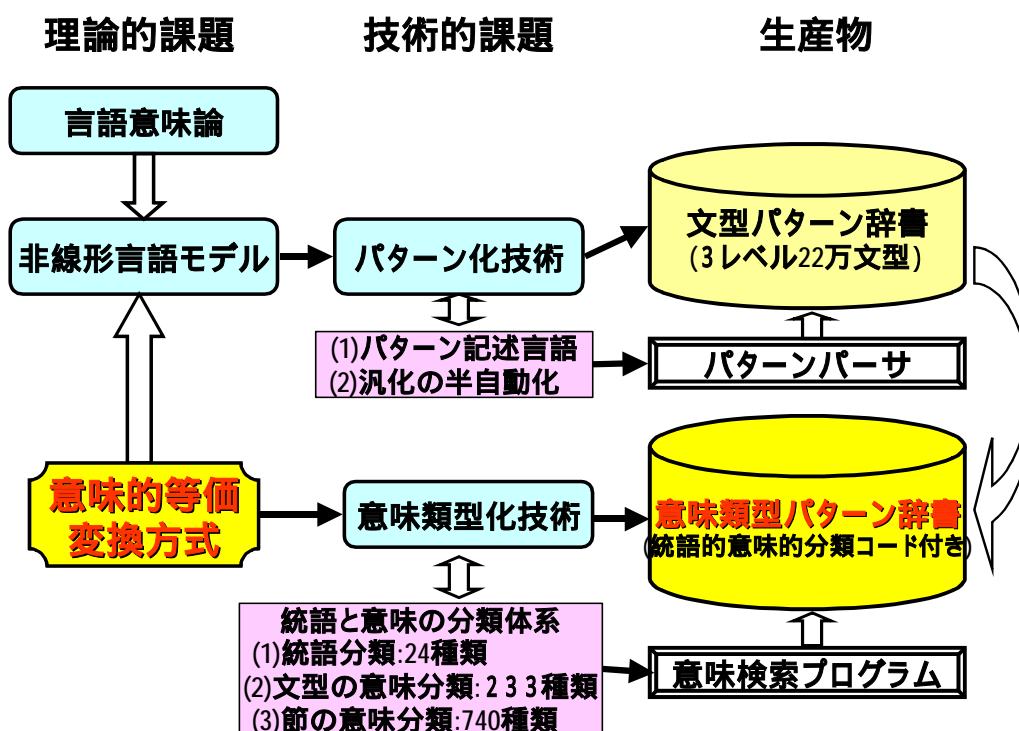


図1.2.1 研究の概要

ところで、パターン化の対象となる表現は、1つ以上の非線形要素を持つ表現である。汎化対象となる線形要素の半自動的な判定方法を検討し、文型パターン作成手順の半自動化を図った。文型パターン辞書作成の具体的手順は以下の通りである。

まず、約30種類のドキュメントから比較的標準的な日英対訳例文約100万件を収集した。その中から、述部2つ及び3つの重文と複文を抽出して形態素解析を行い、解析誤りを人手で修正の上、タグ付きの対訳コーパス（15万件）を作成した。

次に、この対訳コーパスを対象に半自動的な汎化手順を適用して、単語レベル（12.2万件）、句レベル（8.0万件）、節レベル（2.6万件）の文型パターン辞書（合22.7万件）を作成した。なお、品質向上のため、作成の過程では、汎化誤りの検出と修正、パターン記述言語の機能拡張など、さまざまな改良と工夫を行った結果、実用的な水準の被覆率（統語的被覆率98.5%、意味的被覆率79.5%）を達成することができた。

#### (2) 文型パターンの意味類型化

前項で作成した文型パターン辞書の意味類型化を行うため、重文複文の統語的構造に

関する分類体系と意味の分類体系を構築した。また、すべての文型パターンに統語的分類コードと意味分類コードを付与した。

このうち、統語的分類は、構成要素となる節の關係に着目したもので、24種類に分類した。これは、意味類型化では必須な情報でなく、今後のパターン利用のための参考情報とするものである。

これに対して、意味分類コードは、文型パターンを意味類型化（意味的なグループ化）するために必須のものである。第1の分類体系として、重文複文全体を構成する複数の節の意味的關係に着目して、「節間意味分類体系（4段227種類）」を開発した。第2の分類としては、個々の節の意味を表す「節の意味分類体系」も使用することとした。但し、節の意味は、述部の意味属性分類に従うものと考え、用言述部の節の意味は用言の意味属性体系を、体言述部（だ文）の節の意味は、名詞の意味属性体系を用いて分類することとした。これにより、節の意味は合計5段742種類に分類される。

ところで、文型パターン全体に含まれる節の数は26.7万件であり、節間の關係は15万カ所存在する。節の意味コードはほぼ機械的に付与できるが、節間の意味判定は単純ではない。表現の形式と意味の対応規則（必ずしも1対1ではない）の分析ツールを作成し、人手による判定作業を支援した。

### （3）パターン検索プログラム

意味類型パターン辞書の被覆率を評価し、問題分析と改良に役立てるため、「パターンパーサ」と「意味検索プログラム」を作成した。

「パターンパーサ」は、入力日本文と各パターンの構成要素を比較して、適合するパターンを発見するものである。ネットワーク文法を階層化したアルゴリズムを基本としている。さまざまな工夫を行っており、パターン記述言語の変更に簡単に対応できる。また、12.2万件の標本文と22.7万パターン間の照合実験（クロス照合実験）を約30時間（入力文当たり1秒以下）で実施できる。

また、「意味検索プログラム」は、入力文の意味コードを判定し、それと同じ（又はその配下の）意味コードを持つ文型パターンを検索するものである。

前者は入力文と一致する構造のパターンのみが検索されるが、後者は、構造の異なるパターンも抽出されるため、日本語書き換えなどへ応用できる。

### 1.3 研究成果

本研究の最終的な生産物を表1.3.1に示す。これらは、重文複文の文型パターン辞書と意味類型パターン辞書を構成するものであるが、それに含まれる意味属性体系や単語意味辞書など、今後のさまざまな意味処理に使用できる。

また、この表に含まれないが、日英対訳コーパス(100万文)や英語構文体系(83分類)、重文複文と英語構文の意味的対応表などの途中生産物も従来にない知的な資産であり、今後の多くの研究に役立つと期待される。



表1.3.1 最終的な研究生産物の構成

#	分類	項目	内容説明
1	研究報告書類	「研究成果報告書」	研究の目的、方法、成果など全体を解説
2		「意味類型パターン辞書開発技術報告書」	本研究の過程で行った技術的問題点の検討資料を収集したもの
3	意味類型パターン辞書	「日英意味類型パターン辞書ファイル」と「ファイル説明書」	文型パターンに意味類型情報を付与したファイル。単文レベル、句レベル、節レベルの文型パターン(合計約 22.7 万件)を収録
4		「日英意味類型圧縮版パターン辞書ファイル」と「ファイル説明書」	上記パターン辞書を重要なパターン約 5 万件に圧縮したもの
5	パターン記述言語	パターン記述言語仕様書	パターン記述言語規則の詳細を定義したもの。「節間意味分類体系表(4段 227 分類)」を含む
6		「用言意味属性体系」と「用言意味辞書」	用言変数の意味的制約条件と節の意味分類(4段 366 分類、約 6000 語)
7		「名詞意味属性体系」と「名詞意味辞書」	名詞変数の意味的制約条件と節(だ文)の意味分類(4段 364 分類、語:約 6 万語)
8	パターン検索プログラム	「パターンパーサ」の「プログラムファイル」と「プログラム説明書」	要素比較によって入力文と文型パターン辞書を照合するプログラムと「アルゴリズム説明書」、「機能仕様書」、「操作説明書」など
9		「意味検索プログラム」の「プログラムファイル」と「プログラム説明書」	入力文と同じ意味の文型パターンを検索するプログラム。および、その機能、動作、操作法などの説明書
10	付属品	形態素タグ付き重文複文対訳コーパス	重文複文の対訳データの日本語側に形態素情報を付与したもの。(収録対訳文数:約 15 万件)

#### 1.4 今後の課題

目標とする重文複文の文型パターン辞書の開発とその意味類型化は、ほぼ達成することができたが、残された問題も多い。今後の主要な課題は、以下通りである。

##### (1) 変数意味属性の適正化

パターン辞書の問題として残された最大の課題は、パターン変数に付与された意味的制約条件(意味属性)の適正化の問題である。本研究では、動詞、名詞に対して新しい意味属性体系を構築し、変数に対して、それに基づく制約条件を付与したが、この作業は、コストの関係から機械的に可能な範囲に止まっている。そのため、変数に対して、不要な意味属性を含む複数の意味属性を付与したままになっているものが多く、また、適正なレベルまでの汎化も行われていないままである。

##### (2) パターン照合方式との整合性

文型パターン辞書の被覆率は、パターンパーサの能力にも依存している。特に、句変数や節変数に対するパターンパーサの照合アルゴリズムに依存するところが大きい。今後、パターン記述方式との整合性を高めることで、文型パターン辞書の意味的な被覆率はさらに向上することが見込まれる。

### (3) 適合パターン絞り込み技術の研究

入力日本に対して、通常多くのパターンが適合し、その中には、意味的に不適切なものも含まれる。そのため、適切なパターンを選択する方法を検討する必要がある。実験では、多変量解析の方法が有力であるが、現状では、変数の意味的制約条件の絞り込みがまだ甘いと見られるため、これに関する改良も必要と思われる。

## 2 研究構想及び実施体制

### 2.1 研究構想

#### (1) 研究の課題と方法

本研究は、自然言語処理における従来の要素合成法の限界を克服することを目指しており、重文と複文を対象に以下の目標を達成しようとするものである。

研究目標	あらかじめ、言語表現に用いられる対象認識の形式を網羅的に収集しておく方法として、意味のまとまる範囲の非線形な表現構造を取り出し、その意味を対応づけた言語知識ベースを作成する方法を実現する。
------	--

これは、言語表現の処理において、「非線形領域を分解不能部分として抽出し、単位化することによって全体を線形近似に持ち込むこと」であり、非線形問題に対して、「局的写像を介して大局写像を得る方法」を実現しようとするものである。

具体的には、線形要素と非線形要素を識別する方法を考案することにより、従来の結合価文型パターンする方法では扱うことのできなかつた「重文と複文の非線形な表現形式を文型パターン化すること」、また、得られた文型パターンを意味的に分類する手法を開発し、「意味的等価変換に必要な意味類型パターン辞書を開発すること」である。研究の課題とそれを実現する方法は表2.1.1のようにまとめられる。

表2.1.1 研究の課題と方法

番号	研究の課題	研究の方法
課題1	日本文(重文複文)の表現を意味的にほぼ網羅するような表現意味辞書を開発すること。但し、日本語表現の意味は別の言語(英語)で記述するものとする。	対訳コーパスを対象に線形要素をさまざまな方法(変数化、関数化、任意化など)で汎化し、日英対訳文型パターン化する。 <b>&lt;文型パターン辞書の研究開発&gt;</b>
課題2	日本文(重文複文)に意味的に対応する複数の英語表現を発見する仕組みを実現すること。	日英文型パターンを意味的に類型化し、マッピング関数によって意味類型間の対応関係を決定する。 <b>&lt;意味類型パターン辞書の研究開発&gt;</b>

#### (2) 研究実施計画

本研究の中心的な課題は、文型パターン辞書を開発し、その意味類型化を行うことである。この課題を達成するには、意味解析の応用面からの検討による設計条件の検討と意味類型辞書の適用性に関する実験的な検討が必要である。これらの事情をふまえ、本研

究は、表2.1.2に示す計画で進めることとした。

但し、この線表は途中で見直し、大幅に変更することとなった。それは以下の事情による。すなわち、研究開始時点では、重文複文5万件を対象に文型パターン辞書を試作する計画であったが、約3倍の対訳標本文を入手することができた。そのため、汎化作業の効率化について再検討することが必要となり、開発手順を見直したものである。最終的に実施した線表については、3.11節で示す。

表2.1.2 計画線表

分類	H13年度	平成14年度	平成15年度	平成16年度	平成17年度	H18年度
知識ベース	標本のDB化	・パターン記述方式の研究 ・第1次文型パターン化(品詞レベル)	・パターン記述言語の設計 ・第2次文型パターン化(意味レベル) ・真理項抽出 ・シソーラス化	・第3次文型パターン化(節レベル) ・選択条件付与	・改良	・図書編纂
等価変換		・実験システム第1版	・実験システム第2版	・実験システム第3版	・総合実験と改良	・デモシステム化
言語生成		・言換え方式予備検討	・言換え方式の検討	・言語生成方式検討	・総合実験と改良	・デモシステム化

## 2.2 実施体制

本研究は、実用的な規模の意味類型パターン辞書を研究開発することを目標としており、方式や技術的な課題に関する研究チームのほかに、言語アナリストとプログラマによる開発チームを結成することが必要である。

そこで、本研究では、図2.2.1に示すような研究チーム(4グループ)から編成することとし、研究統括グループの配下に開発チームを設けることとした。開発チームは、企業内に設けたチームで、委託や発注によって維持できる体制とし、開発作業の円滑化を図るため、川崎に共同実験室を設けた。また、研究開発全体の進捗を管理するため「進捗会議」を設け、研究途上発生したさまざまな技術的問題を検討すると共に、特に、開発チームの作業過程で発生した問題の解決を図るため、「技術会議」を設けた。

また、本研究を推進するに当たって、研究メンバー以外の言語学、言語処理の研究者との間で、理論的問題を中心に研究発表を行い意見を交換する場として、「言語・認識・表現研究会」と「表現意味辞書研究会」を開催した。

なお、「進捗会議」、「技術会議」は、研究メンバー全体を対象とするクローズドな会議であるのに対して、「言語・認識・表現研究会」と「表現意味辞書研究会」は、研究メンバー外の研究者をも対象としたオープンな研究会である。

研究グループの役割は以下の通りである。

### (1) 研究統括グループ

各グループの事前検討結果をベースに年度半期毎の作業項目を決定してその作業手順を詳細化し、外部企業に試作発注を行う。また、試作の過程で発生するさまざまな問題

については、その解決策を検討し遅滞なく作業が進むよう指示すると共に、必要に応じて他の研究グループへの検討依頼を行う。なお、定例（月1～2回）の進捗会議と技術会議を主催し、外部発注作業を含む全体の進捗管理と技術的問題の検討を行う。

## （2）3グループ

各グループが担当する観点から、意味類型パターン辞書開発に関連する問題について事前の検討を行い、開発作業に反映させることを基本とする。

個別的には、「意味類型知識ベースグループ」は、パターン記述方式に関する事前検討と意味類型化のための意味分類体系に関する事前検討を行うと共に、作成されたパターンの品質について、さまざまな観点から分析検討を行い、改良修正作業に反映する。「等価変換システムグループ」は、意味類型パターン辞書を使用した機械翻訳実験システムの検討並びに試作を行い、本研究の効果について実験的な評価を行う。また、「言語生成システムグループ」は、意味類型パターン辞書の言語生成への応用方法について検討すると共に、文型パターン辞書開発の過程で、言い換え技術の研究結果に基づき、言い換え辞書の構築を行うなど被覆率向上対策についての検討を行う。

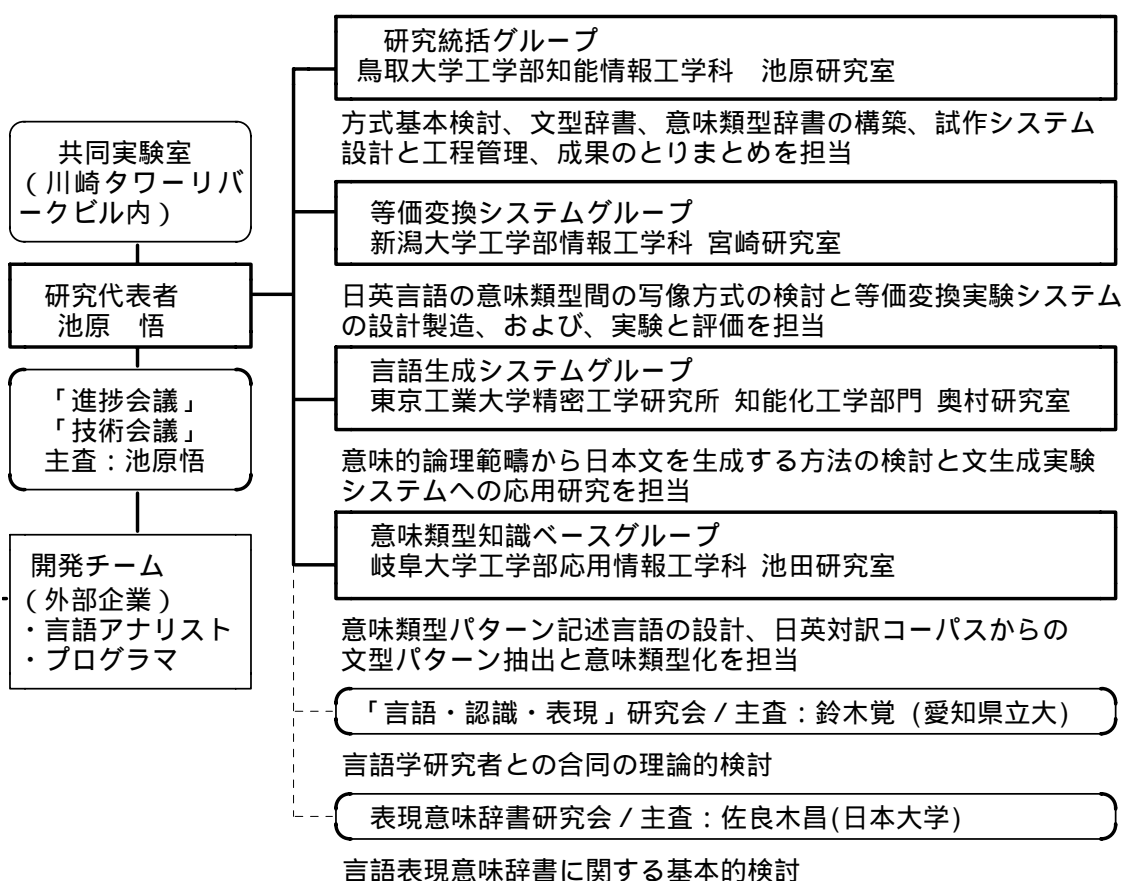


図2.2.1 研究実施体制

### 3 研究実施内容及び成果

本研究では、言語に対する「意味的等価変換方式」を提案し、それを実現するため、重文と複文を対象に、「文型パターン辞書」を研究開発した。また、それを意味類型化することによって「意味類型パターン辞書」を試作した。

本研究では、文型パターン辞書の被覆率を如何に向上させるかが大きな課題である。そこで、これらを研究開発する過程で、さまざまな実験評価を行い、それに基づく改良を行った。本章では表3.1の構成に従って、その概要を述べる。

表 3.1 第 3 章の内容構成

3章の目次の構成	記載内容の概要
3.1 意味的等価変換方式の提案	言語意味解析のための基本技術として提案した方法の概要を述べる。本研究で研究開発した「意味類型パターン辞書」は、この方式の実現を目指すものである。
文型パターン辞書の研究開発	日本語重文複文の非線形な表現構造をパターン化(合計 22.7 万件)する過程と結果を以下の7節に分けて記述する。
3.2 非線形言語モデルの提案	言語表現のパターン化の基本として提案した言語モデルについて述べる。
3.3 パターン記述言語の開発	言語モデルに基づき、線形な表現要素を汎化してパターン記述するために開発した言語の概要を述べる。
3.4 単語意味属性体系の構築	パターン記述で使用する変数の意味的制約条件を指定するために開発した用言と名詞の意味属性体系の概要を述べる。
3.5 文型パターンの作成方法	大規模な日英対訳標本から文型パターンを作成する方法とそれを半自動化する方法について述べる。
3.6 文型パターン辞書の実験的改良	前節の方法で試作した文型パターン辞書(初版)に対する実験的改良の方法と内容について述べる。 (注)本研究開発で最大の研究コストを要した部分である。
3.7 文型パターン辞書の特性	作成した文型パターン辞書の静的、動的な特性の分析結果を述べる。
3.8 文型パターンの意味類型化	「文型パターン辞書」の各パターンを意味類型化するために行った統語的意味的分類コード付与について述べる。また、そのために作成した各種分類体系の概要についても述べる。
3.9 パターン検索プログラム	要素照合型の「パターンパーサ」と意味検索型の「意味検索プログラム」について述べる。
3.10 翻訳簡易実験システムの試作	意味類型パターン辞書を機械翻訳に応用した場合の効果を調べるために試作した実験システムの概要を述べる。
3.11 研究開発年表	主な研究課題とその結果、途中生産物とその相互関係を図表にまとめる。
3.12 研究成果の今後期待される効果	意味類型パターン辞書の応用が期待される研究分野などについて述べる。

### 3.1 意味的等価変換方式の提案

#### 3.1.1 意味的等価変換方式の原理

言語では、同様の内容を表現するにもさまざまな表現の仕方があり、意味と表現は1対1の対応関係を持たない。従って、どの表現とどの表現は意味が同じか、また、どの表現とどの表現は意味が異なるかが判定できるようになれば、言語意味処理は成功したとすることができる。このような考えから、本研究では、言語意味処理の基本技術として、「意味的等価変換方式」を提案した。

この方式は、思想的には、「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」と言う「セマンティック・タイポロジ（意味類型論：有田潤）」と「独創的な発想の背景には等価的類推思考が存在する」という「等価的類推思考の原理（市川亀久弥）」を背景に提案するものである。与えられた言語表現と同じ意味を表す別の表現を発見する方法を実現しようとしている。

この方式を、「与えられた言語表現と同じ意味の別の言語の表現を発見すること」に適用すれば、「機械翻訳方式」となり、「同じ言語内の別の表現を発見すること」に適用すれば、「言い換え方式」となる。また、同一の言語内で同じ意味の表現を探すことは、意味による情報検索の基盤技術ともなる。

#### (1) 等価的類推思考の原理

最近、人工知能の分野では、汎用的な規則による推論能力に比べて、類推能力の重要性が注目されるようになってきた。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、「覚える」、「思い出す」、「一般化する」、「特殊化する」等の機能が有機的に結合したものとされている。

かつて市川亀久彌は、「人間の独創性は、類推思考の能力にある」として、「等価的類推思考の原理」を提案した[市川1963]。これは創造的問題解決プロセスのモデルと言えるものである。このモデルによれば、類推は、式(1)に示すように、「異なる系での異なる現象が、ある一定の条件Cの下で共通点を持つこと」を前提として成り立つ知的な活動で、「ある系 上の事象A（ソース）が与えられたとき、条件Cの下で なる共通性を持った事象系 上の事象B（ターゲット）を思い浮かべること」だと定式化される。

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

但し、C：条件， $\mathcal{A}$ ：共通見地，A：系 上の事象A，B：系 上の事象B

ここで、共通見地 $\mathcal{A}$ は、意識されるときもあれば、意識されないときもある。 $\mathcal{A}$ が発見できるか否かは、条件Cに強く依存している。独創的なひらめきは、思いもよらないような条件Cと共通見地 $\mathcal{A}$ を同時に発見することによってなされる場合が多いとされている。いずれにしても、 $\mathcal{A}$ の存在しない事象の思いつきは、類推とは言えない。

ところで、翻訳では、原言語の表現の意味から、それを表すのに適した目的言語の表現を思い浮かべることが必要であるから、翻訳も類推であり、式(1)を前提としていると

考えられる。

式(1)を言語に適用した場合、その意味は、以下の通りとなる。すなわち、 $A$ 、 $B$ 、 $C$ を言語上の表現、言語上の表現とし、 $C$ の条件下で、それらの持つ意味に着目すると、両表現間に同一な、もしくは類似な意味が存在することである。言い換えれば、言語上の表現 $A$ が言語の表現に翻訳できるには、言語の表現の中に、同じの概念を表すような表現 $B$ が存在しなければならない。これは、異なる言語間での翻訳が成り立つための条件である。

## (2) 意味的等価変換方式の原理

前節で述べたように、異なる言語間で、無限とも言える言語表現を意味的に正しく対応づけることは無理である。そこで、式(1)を前提とした「言語の等価変換方式」を提案する。この方式は、「意味類型」間に対応をとるもので、原言語の表現形式は、真理項と呼ぶ概念（意味）を介して、一つ以上の目的言語の形式に写像（決定論的な変換とは区別）される。本方式の原理を式(2)に示す。

$$A \quad P(A) \quad C(\quad) \quad B \quad (2)$$

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 $B$      $\rightarrow$  : 変換、 $\mapsto$  : 写像  
 $\mathcal{C}$  : 「論理意味範疇」（真理項（概念）の集合）  
 $P(A)$  : 表現 $A$ の形式を求める関数、  
 $C(\quad)$  : 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している。

の場合は翻訳方式となり、 $C = A$ 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 $C(\quad)$ 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。

「意味類型」は、「複合概念」を表す言語表現の形式の集合である。これに対して、「真理項」は、「複合概念」を定義するための言葉（例えば、「肯定判断」、「否定判断」、「因果関係」、「比較」など）である。単一の真理項で表せない意味類型は複数の真理項を使用して定義される。「真理項」を用いた「意味類型」間の意味的等価変換の仕組みの例を図3.1.1に示す。

この図は「因果関係」の概念を表す形式の一部を示したものである。原言語の形式は真理項を介して目的言語に写像される。

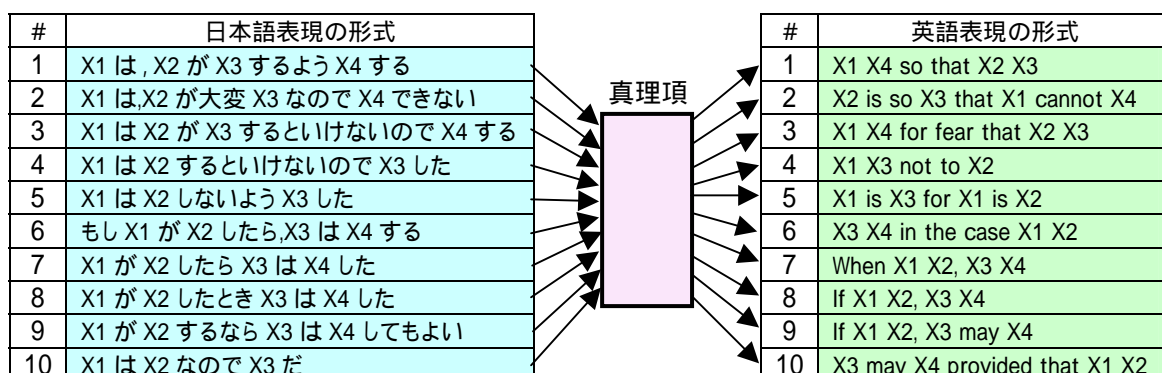


図3.1.1 真理項を介した形式の写像

### 3.1.2 文型パターンの意味類型化

意味的等価変換方式を実現するには、人間の思考形式とも言える言語表現の形式を取り出し、「意味類型化」することが必要である。「意味類型化」とは、同一と見なせる意味を表す表現形式をグループ化することである。そこで、本節では、機械翻訳への適用を例に文型パターンを意味類型化することの意義について述べる。

#### (1) 意味類型化の狙い

意味類型化は、文型パターンの意味を表す真理項を体系化し、各文型パターンに真理項を付与すること、また、付与された真理項を用いて文型パターンを意味分類することによって実現される。

文型パターンは、表現の意味をすくい取る網の目であるが、このような表現の意味のグループ化ができると、異なる表現構造をもった言語表現でも、それが同じ意味を持つかどうかを判定することができるようになる。これを機械翻訳に適用すると、入力文に対するきめ細かな訳し分けが可能となる。

従って、機械翻訳において、意味的等価変換方式の狙いは、入力文に対する訳し分け機能を実現することであるが、従来のパターン翻訳方式でもパターン数が増大するにつれて同様の効果が期待できる。そこで、以下では、従来のパターン方式と意味的等価変換方式における訳し分け機能の違いを図3.1.2の例に基づいて述べる。

#### (2) パターン翻訳の場合

大規模な文型パターン辞書では、入力文に対して複数のパターンが適合する場合が増加する。適合するパターンは、入力文が統語構造と意味が同じ原言語のパターンであるが、適合した原言語のパターンに対応する目的言語のパターンの統語構造が必ずしも同じではないから、異なる統語構造の訳文候補が得られると期待される。その例を図3.1.2に示す。

図3.1.2は、入力文に対して、パターンJA-1とJA-2が適合する場合を示している。パターンJA-1とJA-2の統語構造は同じ(統語構造JA)である。これに対して、対応する目的



言語パターンの統語構造は必ずしも同じでなく、統語構造EXとEYである。従って、一つの入力文に対して、統語構造の異なる2種類の訳し分け候補が得られるが、同じ意味の統語構造EZのパターンが存在してもそれを発見することはできない。

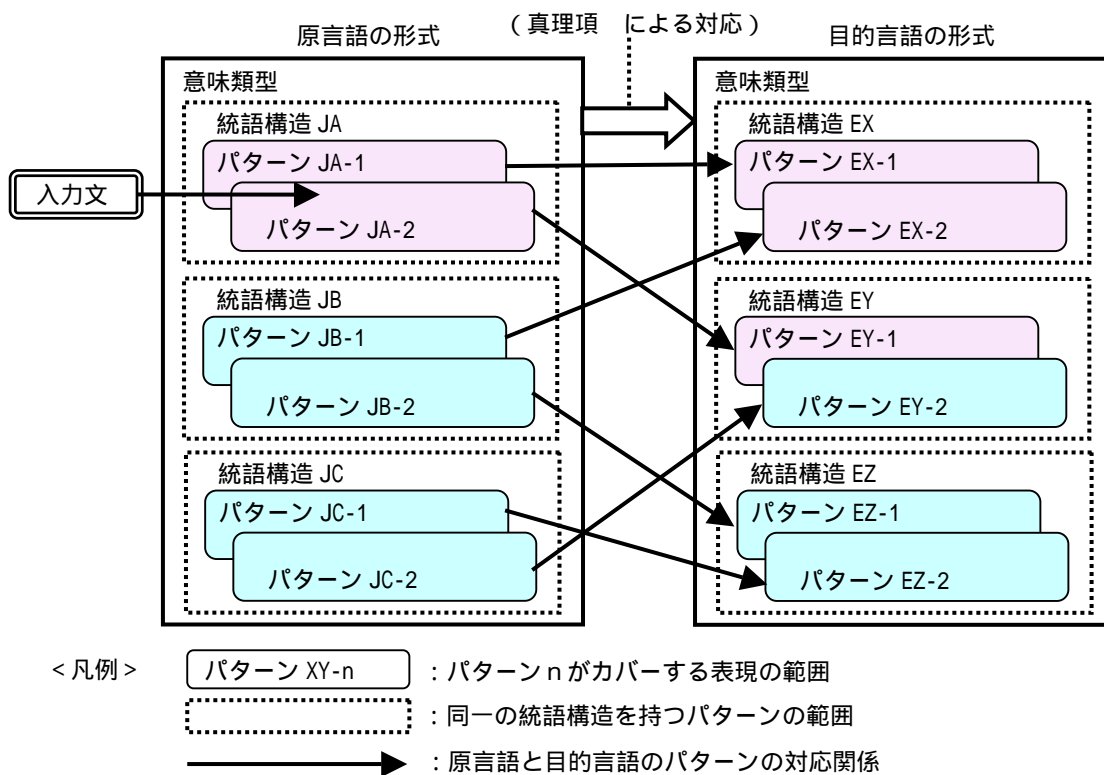


図 3.1.2 パターンの翻訳方式と意味的等価変換方式の違いの例

### (3) 意味類型化されたパターン翻訳の場合

これに対して、同一の意味を持つパターンがグループ化され、真理項を介して対応づけられている場合を考える。図3.1.2は原言語の統語構造JAのパターンJA-1、JA-2が統語構造JBのJB-1、JB-2、および、統語構造JCのJC-1、JC-2と同じ真理項の意味で使用されることを示しており、それらに対応する目的言語のパターンのすべてから訳文候補が生成できる。また、新たに得られる訳文候補は統語構造EZのパターンだけではなく、統語構造EXのパターンEX-2、統語構造EYのパターンEY-2も含むことが分かる。

以上の例から分かるように、従来のパターン翻訳方式では、訳し分けに使える表現形式がすべて得られる保証はないのに対して、意味的等価変換方式では、異なる統語構造の訳文がもれなく得られるだけでなく、同じ統語構造内の表現でも、微妙なニュアンスの異なる訳文候補も得られると期待される。

## 3.2 非線形言語モデルの提案

「意味類型パターン」は「文型パターン」を意味類型化したものである。「意味類型パターン辞書」を開発するためには、まず、「文型パターン辞書」を実現することが必要である。そこで、本研究では、認知論的な立場から、「パターン化すべき対象表現を

明確化すること」、「表現要素の線形性と非線形性を区別すること」を基本に、「非線形言語モデル」を提案し、それに基づいてパターン化の方法を明確にした。

「非線形言語モデル」は、要素合成法に立脚した従来の言語処理の限界を克服することを狙ったものである。本節では、このモデルの基本的な枠組みについて述べる。

(1) 線形表現と非線形表現

文型パターンは、話者が対象に対する認識を表現するための枠組みである。このような枠組みの構成要素を「他の要素に置き換えると全体の意味が損なわれるもの（非線形要素）」と、「他の代替要素に置き換えても全体の意味は損なわれないような要素（線形要素）」の2種類に分類する。

【言語表現の線形性と非線形性の定義】

一つ以上の代替要素が存在し、その要素に置き換えても「表現全体の意味」が変化しないような要素を「線形要素」と定義する。

また、「線形要素」のみから構成される言語表現を「線形な表現」、1つ以上の非線形要素を有する言語表現は「非線形な表現」と定義する。

【言語表現の意味について】

上記の定義を実際の文に適用するには、「表現全体の意味」を明確にすることが必要である。ここでは、三浦つとむの「関係意味論」の立場から、「言語規範としての概念」を意味と考え、表現構造の表す意味は「抽象化された複合概念」とするが、計算機から見たら、いずれの記述方法も記号でしかないから、排他性が保証できる体系であれば処理上の問題は発生しない。そこで、意味（規範としての概念）を記述する方法として、異なる言語表現を用いることとする。これに従い、本研究では、日本語表現の意味を英語表現を用いて定義することとする。

<線形要素の典型例>

図3.2.1に日英対訳文の例を示す。原文は、「何かの事象の直後、誰かが何かの行為をする」と言う「事象間の関係」（「抽象化された複合概念」）を表す日本語の表現で、その意味は英語表現で定義されている。個別概念である「彼女」、「大学」などは、それを代替可能な要素が存在し、それを置き換えても「抽象化された複合概念」（英語表現構造）は変化しないから、これらは、線形要素である。

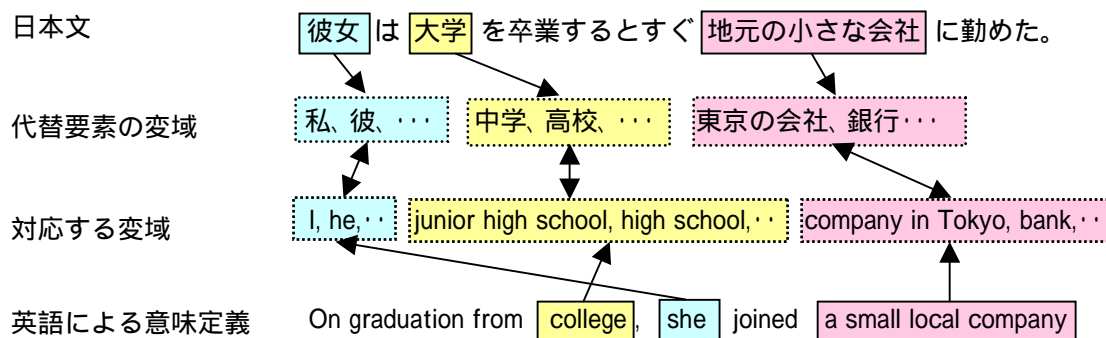
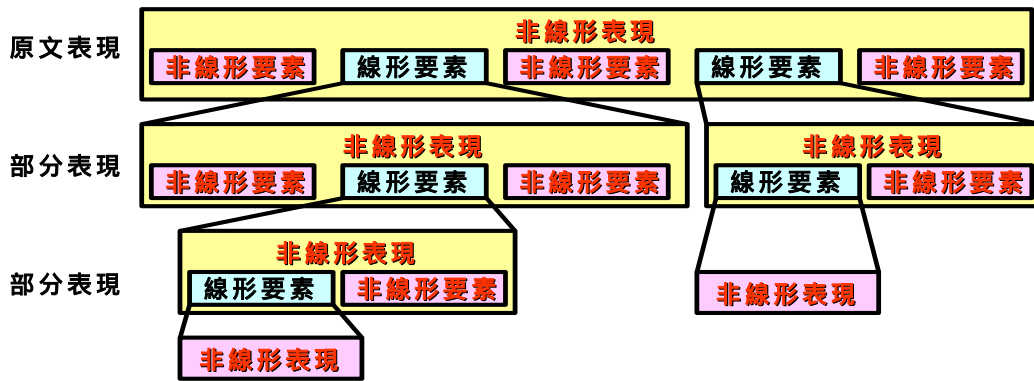


図 3.2.1 線形要素の例

(2) 言語表現のモデル

定義により、言語表現は、一般に図3.2.2の例のように非線形要素と線形要素から構成される(参考として従来の線型言語モデルを図3.2.3に示す)。この図では、「要素」はある「表現」の部分を表す用語であること、「要素」はそれ単独で見たとき「表現」でもあることに注意が必要である。但し、「表現」は意味のまとまる単位に対して使用しているため、「要素」が常に「表現」であるとは限らない点にも注意が必要である。



3.2.2 非線形言語モデル

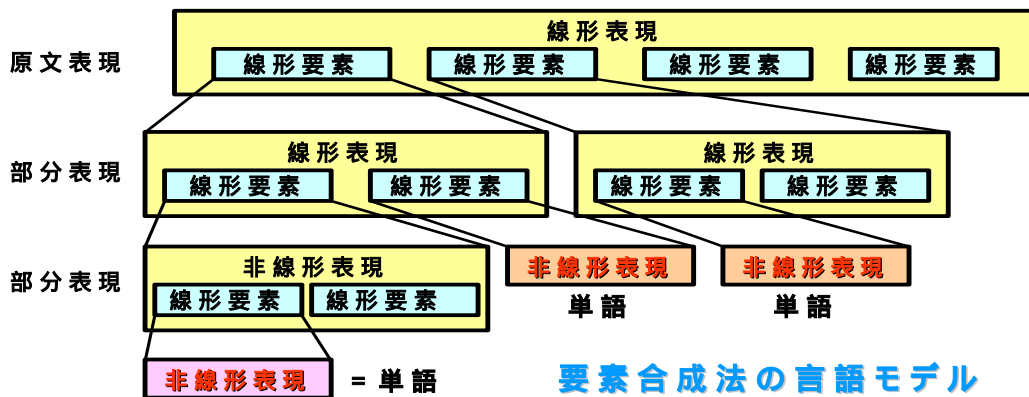


図3.2.3 線形言語モデル(従来の言語モデル)

ここで重要な点は、線形要素は、「それ自身も言語表現であるが、線形表現だとは限らないこと」である。図の例では、原文表現の「線形要素」は、いずれも「非線形要素」と「線形要素」からなる「非線形表現」である。このように、非線形表現に含まれる線形要素を取り出していくと、最終的には線形要素を持たない非線形表現に帰着する。帰着した非線形表現は、単一の単語の場合(単語はそれ以上分解できない)もあるが、置き換え可能要素を持たない慣用句のような場合もある。従って、一般に言語表現は、最終的に1つ以上の非線形表現と0個以上の線形要素に分けることができる。

(3) 言語モデルの基本的特徴

上記の言語モデルは、ラングとパロールを基本としたソシュールの言語モデルやチョムスキーの二元論的な言語モデルとは本質的に異なり、工学的な判断基準が得られる点に特徴がある。本言語モデルの基本的な特徴は、以下の5点にまとめられる。

#### <特徴1> 線形要素の言語ペア依存性

言語表現の意味を別の言語表現で定義しているため、線形要素の数や範囲は言語ペアに依存する。同族言語の場合は線形要素の範囲が増大し、異なる言語族間では減少することが予想されるが、これは翻訳の難易性を反映している。

#### <特徴2> 線型要素の相互独立性

線型要素は代替要素に置き換えても全体の意味が変わらないことが条件である。この条件は各線型要素毎に成立しなければならないから、一つの表現構造の中に複数の線型要素がある場合は、それらの線型要素間の関係は相互独立でなければならない。

#### <特徴3> 代替要素の有限性

線形要素は置換可能だと言っても何に置き換えても良い訳ではない。置き換え可能な範囲は、文法的、意味的に制限される。

#### <特徴4> 線形要素の相対性

線形要素のみからなる表現が線形な表現であるが、要素の範囲は任意に決められるから、表現全体の線形性、非線形性は要素の選び方に依存する。従って、汎用的な文型パターンを得るには、線形要素が多くなるように工夫すればよい。

#### <特徴5> 線形要素と非線形表現の循環性

線形な要素だと言っても、それは表現全体から見たときの性質である。その要素自身が線形な表現だとは限らない。

### (4) 言語モデルと文型パターン

図3.1.4の言語モデルで大切なのは、要素分解の各段階で出現する非線形表現が「構造的に意味のまとまる単位」だということである。要素分解によって元の意味を失わないようにするためには、各段階の非線形表現に対して意味をすくい取る仕組みが必要である。そこで、本研究では、非線形表現の持つ以下の性質に着目して文型パターンを使用する。

非線形な要素は、助詞、助動詞など、字面表記される場合が多い。

各線形要素と非線形要素の出現順序は固定的な場合が多く、任意性が少ない。

具体的には、重文複文を構成する要素を、単語、句、節の3段階に分け、それらの線形な要素を汎化することによって文型パターンを作成することとした。

## 3.3 パターン記述言語の開発

日本語表現の意味を英語表現を用いて定義することについてはすでに述べた。この典型的な例は日英対訳文である。そこで、日英対訳例文を汎化することによってパターン対を作成することを前提に、パターン記述言語を研究開発した。本節では、パターン記述言語の概要を述べる。

### 3.3.1 パターン化の原則と枠組み

#### (1) パターン化の原則

定義1から、パターンは線形要素と非線形要素の2種類の要素からなり、線形要素で

あるのは、以下の2つの場合である。

日本語側の表現要素に対して英語側に対応する要素がある場合

英語側にそれに対応する要素はないが、日本語側のその要素を削除しても対応する英語表現は変化しない場合

そこで、与えられた日英対訳表現を対象に、これらに該当する部分表現を抽出して汎化することによってパターンを作成する。

## (2) パターン記述の枠組み

日英対訳パターンを記述するため、

汎化の程度に応じた記述能力を持つこと

パターン間で意味的な排他性が得られること

大規模な対訳コーパスから半自動的にパターンが作成できること

を目標にパターン記述言語を設計した。パターンの記述に使用される要素は、表 3.3.1 に示すように字面、変数、関数、記号の4種類である。

表現要素のうち、自立語的な要素(単語、句、節)では、線形な要素は変数を使用して記述し、非線形な要素は、字面によって記述する。これに対して、付属語的な要素(助詞、状動詞など)では、線形な要素は関数を用いて記述し、非線形な要素は字面または関数で記述する。

関数には、非線形な要素を指定するもの(字面の代わりに使用される)と線形な要素を指定するものがあり、両者は区別して使用される。例えば、表現には現在時制でしか使用できないもの(非線形)や過去形でも使用できるもの(線形)がある。このような、時制、相、様相などの表現の線形性、非線形性は、使用する関数によって識別される。また、表記の揺らぎや語順の任意性など、特殊な線形要素は、記号を用いて記述する。

### 3.3.2 変数化する線形要素

#### (1) 線形要素の変数化の原則

線形な単語、句、節を変数化する。要素の線形性を判定するための原則は以下の通りである。

変数化する対象は、パターン全体から見たときの線形要素である。それ自身が非線形表現であっても良い。

変数化する要素と対応する英語表現の要素が必ずしも同一の文法的属性を持つ必要はない。

あってもなくても意味の変わらない要素(それも線形要素)、また、構造上の線形要素(語順の任意性など)は、記号で表示する。

#### (2) 制約条件の付与

特徴3に基づき、日本語パターン側の変数に対して、意味的に置き換え可能な範囲を名詞意味属性、動詞意味属性、副詞意味属性などを用いて指定する。

表 3.3.1 文型パターンと記述言語の構成要素

分類		機能説明
字面	日本語は全角文字、英語はアルファベット半角小文字	
記号類	離散記号	原文任意要素(6種類に分類)が存在可能な場所を示す。
	文節境界記号	原文任意要素が存在してはいけない文節境界を意味する(通常は省略)。
	要素選択記号	代替可能な要素(複数)を指定する。
	対応型要素選択記号	日英での要素の対応関係を保ったまま、対応する順序を指定する。
	訳出要素選択記号	日本語側に対応する要素がないとき、生成する要素を指定する。
	要素挿入記号	英語側の要素の中に副詞などの別の要素を挿入することを意味する。
	任意要素記号	あってもなくても良い要素(1つ以上)を指示する。
	順序任意記号	指定された複数の要素の出現順序が任意であることを示す。 例){彼女を   彼は   駅まで}送る。
	位置変更可能要素指定記号	位置を移動して良い要素と移動可能な位置を示す。 例)#1{誰が見ても}彼の作品は(#1)第1級のものだ。
	補完要素記号	省略された主語と目的語の補完を意味する。
変数類	単語変数	線形な名詞、動詞などの自立語を意味する。文型照合では、人力文中の該当する単語(標準形)がバインドされる。(11種類)
	句変数	線形な名詞句、動詞句などを意味する。文型照合では、該当する句がバインドされる。(5種類)
	節変数	線形な連体節、連用節、引用節を意味する。
関数類	品詞変換関数	単語変数に代入された値の文法属性を他の文法属性に変換する。(11種類)
	字面関数	関数名に字面を使用する関数で、引数表現が関数名で指定された字面を含むことを表す。引数には関数名部分を除く表現がバインドされる。(例)「外は大変暑かったので」は、「大変(CL1.past)ので」との照合に成功、CL1 = 「外は暑い」となる。(1種類)
	要素抽出関数	句、節変数に代入された値から主語、目的語の部分を取り出す。
	語形関数	動詞や動詞句、節の活用形などを指定する。(日本語側 11種類、英語側 20種類)
	語形任意化関数	用言字面の標準形を指定する。様相時制の任意化で使用する。
	様相関数	時制、相、様相、態などの表現を指定する。(42種類)
	同値型グループ関数	英語文型に影響を与えない同種の意味の関数をグループ化する。(8種類)
	対応型グループ関数	日英で対応関係を持つ関数を対応関係を保ったままグループ化する。(12種類)
	構文生成関数	一つ以上の句や節から英語文構造を合成する。(12種類)
	マクロ関数	変数を含む表現構造を指定し、それを上位の変数にバインドする。(1種類)

### 3.3.3 任意化とグループ化

#### (1) 必須要素と任意要素

パターン化では、「必須要素」と「任意要素」を明確に区別する。「必須要素」は、日本語パターン内にその要素がないと全体の意味が変わってしまい対応する英語パターンが決定できなくなるもの言う。これに対して、「任意要素」は日本語パターン内にその要素がなくても対応する英語パターンは変化しないものを言う。

このうち、「任意要素」は、さらに「原文任意要素」と「パターン任意要素」に分類する。「原文任意要素」は、日本語パターンでは、その位置と内容が指定されるが、英語表現側で指定する必要のないものである。また、「パターン任意要素」は、英語パターン側に対応する要素を指定する必要があるものである。例えば、訳語挿入位置など、英語側に指定がないと英語表現の生成が困難となるものである。

### (2) 表現要素のグループ化

線形要素には、表記上の揺らぎのように、変域が特定の表現に限定されるものが存在する。このような線形要素では、選択記号を使用し変域を具体的に指定する。

### 3.3.4 語順の扱い

言語表現では、語順や出現する位置を変更しても表現全体の意味は変わらないような要素が存在する。このような要素も線形要素である。以下の2種類の記号を使用して指定する。

#### (1) 順序任意要素指定記号

格要素など語順を変更してもパターンの意味は変化しない要素をグループ化する。

#### (2) 位置変更可能要素指定記号

副詞(副詞的表現を含む)など出現する位置が変わっても全体の意味が変わらない要素について、変更可能な位置を指定する。

### 3.3.4 文型パターン記述例

以上で述べた文型パターンの記述要素について、字面と変数の使用例を図3.3.1に、関数の使用例を図3.3.2に、また、記号の使用例を図3.3.3に示す。

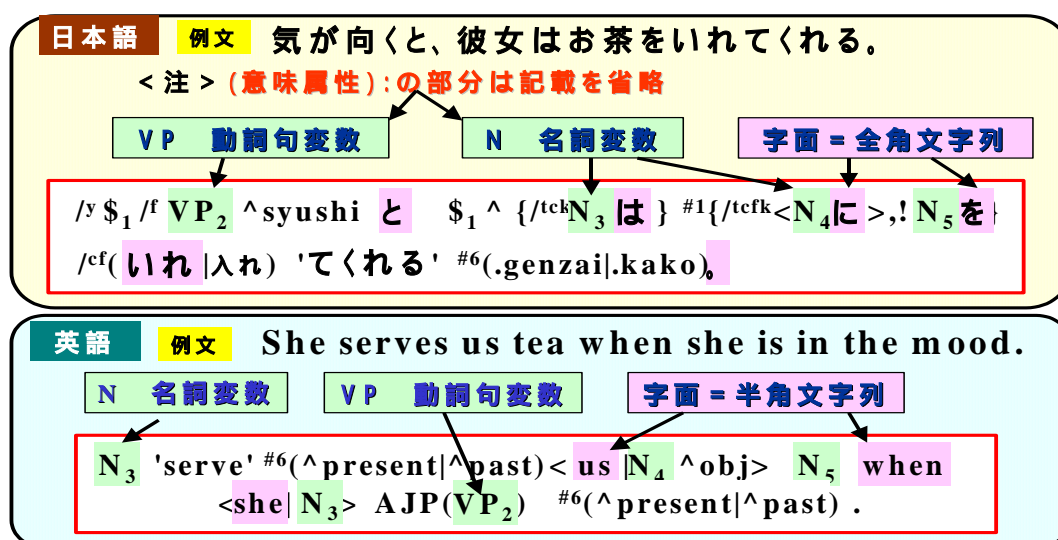


図 3.3.1 字面と変数の使用

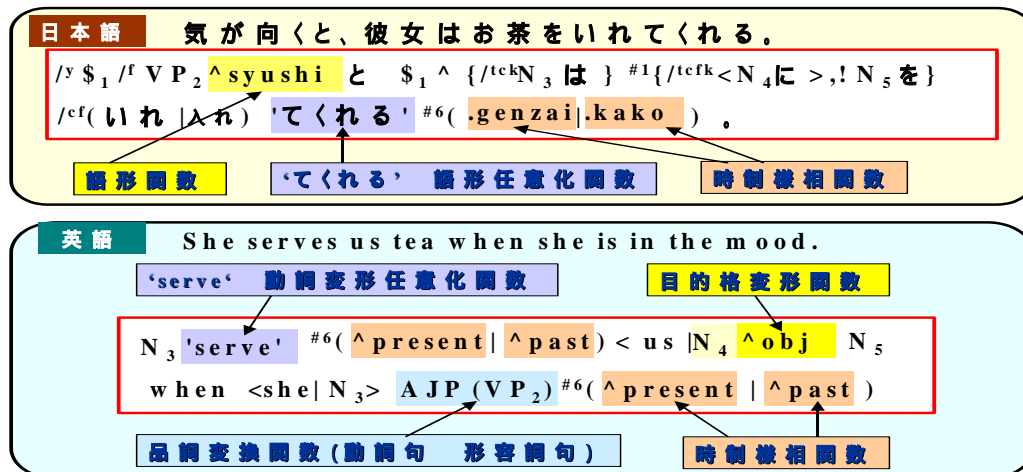


図 3.3.2 関数の使用例

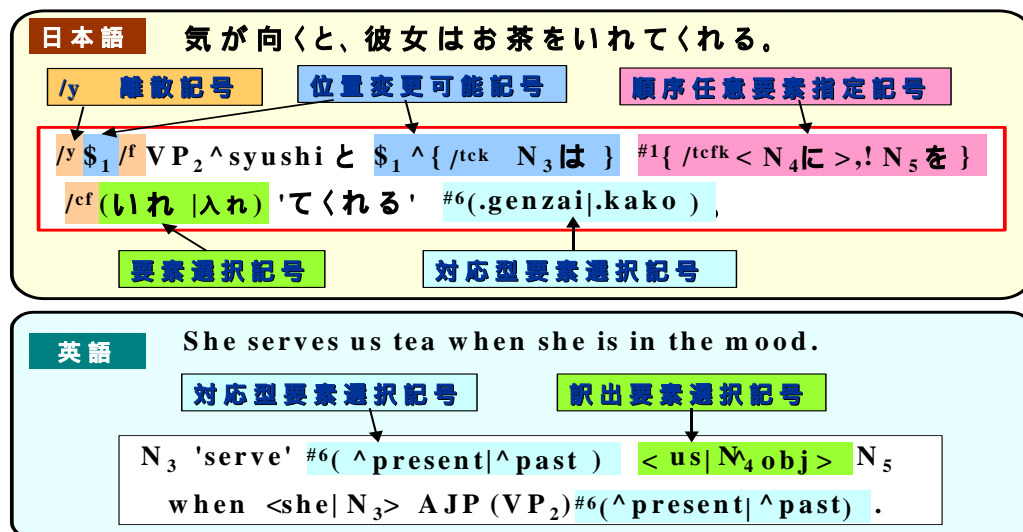


図 3.3.3 記号の使用例

### 3.4 単語意味属性体系の構築

重文複文を対象とする文型パターン化では、体言だけでなく用言も変数化の対象としており、変域を意味的に指定するための意味属性の分類体系とそれによる単語意味辞書が必要である。この体系は、表現パターンの意味的な排他性を実現する上できわめて重要な役割を果たすもので、古典的なパターン方式が成功していないのは精密な体系を持たなかったことが大きな原因とも言える。

また、従来、単語の意味を属性の束と考えた分類法などが行われてきたが、線形要素の変域を指定する観点からは、単語そのものの意味を分類するのではなく、「意味的な用法」（意味属性と呼ぶ）を分類体系化することが必要である。これは、既に「日本語彙大系」で提案された方法である。本研究では、これを踏襲し、重文複文の線形要素の変域を指定するための単語の意味属性体系を開発した。



### 3.4.1 対象とするパターン変数

#### (1) 変数の意味的制約条件の付与

文型パターンで使用される変数は、以下に示すような「単語変数」、「句変数」、「節変数」の3種類である。

- ・「単語変数」： 名詞変数（複合名詞を含む）、時詞変数、数詞変数、  
用言性名詞変数、動詞変数、形容詞変数、形容動詞変数、  
副詞変数、連体詞変数、限定詞変数、ANY変数
- ・「句変数」： 名詞句変数、動詞句変数、副詞句変数、  
形容詞句変数、形容動詞句変数
- ・「節変数」： 節変数

これらのうち、単語変数の意味的制約条件としては、それぞれの品詞の単語の意味属性体系が必要とされる。但し、名詞変数、サ変名詞変数、数詞変数の3種類は、通常、名詞の意味属性が適用される。

これに対して、句変数の意味は、それぞれの句の中核となる単語の意味で代表することができる。また、節変数の意味は、その節を構成する述部の主用言（だ文の場合は体言）で代表される。従って、変数の意味的制約条件を付与するには、単語の意味属性体系があればよい。

#### (2) 従来の意味属性体系の問題

単文に対する結合価パターンの場合と比較して、重文複文のパターンを記述する上で問題となるのは、用言意味属性体系と名詞意味属性体系である。結合価パターンでは、用言は字面で指定されているため、それと共に起する名詞の意味的用法の分類精度が問題であったのに対して、重文複文のパターンの場合、パターン数と被覆率の関係から、用言の変数化が必須でより本質的である。新しい分類体系を作成することが必要である。

これと比べると、名詞変数の意味制約の方は二次的な役割に留まると考えられる。用言の意味的用法が詳細に指定されるため、名詞の意味的用法の分類は、単文の場合に比べて比較的少ない数で良いと考えられる。そこで、本研究では、用言と名詞について新しい属性体系を構築し、それらに基づく単語意味辞書を試作した。以下、それらについて述べる。

### 3.4.2 用言の意味属性体系

#### (1) 用言の意味属性体系への要求条件

- ・統語的構造に考慮した意味属性体系の必要性

用言は、それが表す個々の動的、静的属性の違いに依存して、それぞれ固有の結合価を持ち、これは、文型構造をかなりの程度規定する機能を持っている。

例えば、「が格」と「を格」を伴う文型パターンにおいて、述部用言が変数化されているとする。この場合、用言変数の変域に、「を格」を取ることはできない用言を加えることは適切でない。すなわち、与えられた単文では、使用された動詞の結合価と整合する格要素や副詞などの要素が共起しているのが普通であり、その制約を無視して用言

の部分を書き換えることは適切でない。

また、可能動詞、使役動詞などについても事情は同じである。例えば、可能表現の意味で述部が「Vできる」と変数化されたパターンを考えると、変数Vの変域に可能動詞（結べる、出せる、など）を当てはめることは適切ではない。

これらのことから、用言変数の変域は、文型構造からの制約を受け、与えられた単文の構造に適した用言の範囲に限定される。従って、用言の意味属性分類では、同一の分類項目内に、統語的に異なる構造を要求するような用言が混ざらないような体系とすることが望まれる。

・絞り込みに必要な意味分類の粒度

変数の意味的制約条件を用いないとき、入力日本語と文型パターン辞書全体との照合で適合する文型パターン数は、数千件程度に上ることが予想される。しかし、対象とする重文複文の文型パターンは2つ以上の用言（述部）を持つ。そこで、パターン絞り込みに必要な用言の分類数をnとすると、nは平均的に見て、数十のオーダーで良さそうである。

最も細かいレベルでの意味属性数をどの程度とするかについては、特に決め手となるような基準はなさそうであるが、文型照合実験では、1万件以上の文型パターンが適合する入力文も存在する。そのような場合を考え、用言の意味属性数は最大数百程度とする。

以上から、用言意味属性体系はおおよそ以下のような構成とすることを目標とした。

分類の深さに意味的な粒度をほぼ対応させる。

中間的な深さでの分類数を60程度とする

最新の深さでの分類数合計を数百程度とする

## (2) 意味属性体系の作成手順

用言変数の意味的制約条件に使用することを念頭に、前述の条件に近いと考えられる既存の分類体系をベースに、個々の用言を分類する過程で不足する分類項目とその配置を見直し、段階的に体系を組み直した。以下、その具体的な方法を示す。

・寺村の分類体系に対する分類名の見直し

用言変数の変域を表す意味属性としては、統語構造を意識した体系が望まれることを述べたが、この観点から見ると単文の意味分類として提案された寺村の分類が注目される。そこで、まず、意味属性分類の観点から分類項目名を見直して新しい分類名を付与し、それをベースに拡張を図ることとした。

・外国人のための日本語用例集の参照

用言変数の変域指定に使用する目的から見ると、意味属性体系は、「統語的な性質を反映していること」と合わせて、「該当する単文が使用される場面をも反映していること」が望まれる。この観点から、「外国人のための日本語用例集」の中の動詞編と形容詞編の中から用例に使用された用言（数百件）を取り出し、それが使用される場面との関係を表にまとめた。この表を元に分類体系全体を組み直した。この段階で、寺村の分類は大きく変化した。特に、「形状既定」（形容詞、形容動詞）の意味属性はかなり細

かいものとなった。

以上から、統語的構造と場面を意識した用言意味属性体系の骨格ができてきた。

・ IPAL辞書からの組み入れ

上記の分類体系を用いてIPAL辞書に収録された動詞と形容詞を分類し、問題点を分析すると共に分類の詳細化を行った。IPAL辞書では、各単語に対して意味分類のためのキーワードが付与されている。これらのキーワードは、新しい分類項目の新設や既存の分類項目名の改良などで参考となった。

・ 「類語大辞典」収録語からの追加

分類項目の充足性の確認と登録語数の拡大を狙って、「類語大辞典」からの登録語の収集を行うこととした。しかし、設計思想が異なるため、既存の分類項目名の改良の参考とはならず、未登録の語を探して該当する分類項目に組み込むに止まった。

・ 「日本語語彙大系」利用した登録語の拡充

分類項目の体系化がほぼできたと思われるので、不足する語彙を「日本語語彙大系」から取り込んで分類体系に組み込んだ。

「日本語語彙大系」の名詞意味辞書には、2700分類からなる詳細な一般名詞意味属性体系に基づき、名詞化した動詞（サ変名詞、連用形名詞）についても意味コードが付与されている。この情報を参考に、分類項目の拡充と収録用言の拡充を行った。但し、名詞化した用言と用言そのものの意味属性分類は、部分的な類似点はあるものの同じではない。そのため、名詞の意味属性分類と用言の意味属性分類は必ずしも対応しない部分が多く存在する。そこで、追加収録した語を個別にチェックした。

以上の結果、形容詞、形容動詞に関する分類項目を大幅に補強することができ、用言意味辞書に収録した全体の用言数は、ほぼ倍増した。

(3) 用言意味属性体系の概要

作成した分類体系の分類の深さと分類項目数の関係を表3.4.1に示す。またその詳細を付表2に示す。これらの表に示されるように新しい体系は、5段の木構造で表現されるもので、「一般名詞意味属性体系」と比べて大幅にフラットになっている。

表 3.4.1 用言の意味属性体系の分類数

#	1段	2段	3段	4段	5段	合計
1	動 詞	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		地域社会生活の行為	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	9	48
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	性状既定の表現	6	34	31	72
	合計		55	182	123	371

<注> 「形容詞文」は、形容動詞文（な型形容詞文）を含む

### < 動詞の分類 >

「動詞」を「人間の感覚と知性に関する表現」、「人間生活の場に使用される表現」、「動的属性自体の論理的構造に関する表現」の3つの観点から捉えて、以下の通り分類した。

#### ・人間の感覚と知性に関するもの

物理的で感性的な表現から理性的で超感性的な表現へまでを以下の2つに分類した。

「知覚と情緒の表現」          「知的な行為の表現」

このうち、**は**、物理的感覚から直感的情緒間での表現を表す。主体自身の感覚、感情を表すだけでなく、主体から客体に対する感覚や感情を表す表現も含む。これに対して、**は**、理性と知性による行為行動の表現で脳の活動を代表する。

#### ・生活の場面に関するもの

家庭生活から一般社会での活動と行為を以下の3種類に分類した。

「日常生活の行為」          「地域社会生活の行為」          「社会的活動の行為」

身近な家庭生活から、地域社会、一般社会の順に配置した。現実の生活や社旗的活動の行為は、場面に密着したさまざまな属性を組み合わせた意味で使用されるものが多い。これらは、原始的な動詞が社会の発展に伴って分類進化したものと考えられ、(1)の「感覚と感情」や「裸の動詞」といえるような次項(3)の分類にはなじまないものが多い。そのような動詞の多くはこの分類の中に吸収される。

#### ・動的属性を持つ論理的構造に関するもの

動詞本来の意味を分類する観点から、以下の4種類の分類を設けた。

「現象事象の表現」          「変化の事象の表現」

「移動行為の表現」          「対物行為の表現」

このうち、**は**、対象が人為的ではなく、自発的に存在したり活動したりする現象や事象の表現である。物理的な現象や社会的な辞書も含まれる。これに対して、**は**、人為的であるか人為的でないかを問わず、対象物の形状や性状が変化することを表す表現である。**は**、対象物の移動を表す表現で、この場合も人為的であるか人為的でないかは問わない。最後の**は**、物理的客対物に対する働きかけの表現で、客体は調節的な影響を受ける。

以上の動詞文の意味分類(合計9分類)は、動詞変数の変域を最も汎化した場合に使用することを考え、意味分類体系上、同一のレベルに配置した。

### < 形容詞の分類 >

形容詞(形容動詞を含む)の意味属性分類を表3.4.2に示す。

「性状既定」表現する形容詞文は、「存在既定」、「感覚感情既定」、「人の性状既定」、「ものの性状既定」、「事象の性状既定」、「政治経済社会」の6種類に分類した。言語表現では、動的属性でもそれを話者が静的に捉えれば形容詞によって表現し、静的な属性と見られるものでも動的な側面に着目すれば、動詞が用いられる。動詞の意味属性分類と重複するような分類項目が必要となるため、上記の6種類の項目は、動詞の場合以上に、さらに細かい分類が必要となった。

表3.4.2 形容詞の意味属性分類

<b>存在既定</b>	<b>物の状態既定</b>	<b>政治経済社会</b>
<b>感覚感情既定</b>	物の形状既定	安全危険混乱
感覚の既定	物の性状既定	幸運不運
感情的性状既定	量既定	様態
情緒的状態既定	<b>事象の性状既定</b>	様相
気配趣きリズム	評価既定	経済的活動
<b>人の性状既定</b>	関係	
外見	種類形式	
姿	特徴	
性状	内容既定	
気質	力勢い	
人柄	分離統合	
性格	緊張緩和	
技量能力	計算論理真偽	
肯定評価	位置空間既定	
否定評価	時間と順序	
立場	程度既定	

以下、6種類の分類項目の内容を示す。

・「存在既定」

存在既定を表す形容動詞は多少存在が、形容詞は「ない」くらいである。少ない分類項目であるが、重要な意味を表すため、高いレベルの分類項目とした。

・「感覚感情既定」

この項目は、動詞の「知覚と情緒の表現」に相当する。「感覚の既定」、「感情的性状既定」、「情緒的性状既定」、「気配趣きリズム」の4種類に細分類した。

・「人の性状既定」

人の性状既定を行なうための形容動詞は実に沢山存在する。肯定的に見るよりも否定的に見る言葉が多いようである。人間は、よくもこんなに人のことをこと細かく批評しているものかと驚く位である。「外見」、「姿」、「性状」、「気質」、「人柄」、「性格」、「技量能力」、「肯定評価」、「否定評価」、「立場」の10種類に分類した。

・「ものの性状既定」

ものの性質や状態を表す形容詞、形容動詞を、「物の形状既定」、「物の性状既定」、「量既定」の3種類に分類した。

・「事象の性状既定」

現象や事象の内容に応じて、「評価既定」、「関係」、「種類形式」、「特徴」、「内容既定」、「力勢い」、「分離統合」、「緊張緩和」、「計算論理真偽」、「位置空間既定」、「時間と順序」、「程度既定」の12種類に分類した。

このうち、は、単一の事象に対する肯定、否定、優劣などさまざまな判断に関する表現を示す。また、は、複数の対象の関係であり、部分全体、独立従属などの関係を分類したものである。

・「政治経済社会」

「安全危険混乱」、「幸運不運」、「様態」、「様相」、「経済的活動」の5種類に分類した。これらの中には、社会的な事象や現象の状態やその見方を表す数

多くの形容動詞が含まれるが、これ以上の分類は行わなかった。

### 3.4.3 名詞の意味属性分類

#### (1) 名詞の意味属性分類の目的と方針

既に述べたように、重文複文の文型パターンでは、名詞の意味は必ずしも決定的なものではないが、用言の意味をきめ細かく分類するだけでは、必ずしも十分な制約条件とならない場合が考えられる。すなわち、用言変数に対していかに細かな制約条件を与えたとしても、用言の字面を決定する以上の制約はあり得ないから、単文の意味を用言の字面以上に特定することはできず、用言字面の持つ意味的な多義は絞り込めないことである。この問題は、結合価パターンの場合と同種の問題である。すなわち、重文、複文の場合においても、格要素となる名詞の意味的制約は重要で、用言の意味、すなわち単文の意味を限定する必要がある。

そこで、本研究では、「日本語語彙大系」で定義された「一般名詞意味属性体系」をベースに、以下の方針で新しい意味属性体系を作成した。

<方針1> 現在作成中の用言の意味属性体系とほぼ同等の粒度の分解能とする。

<方針2> 実験的に意味的粒度に着目した文型パターン選択の最適化が行えるよう、粒度と深さを対応させたフラットで比較的浅い木構造で構成する。

#### (2) 名詞意味属性体系の構成

本研究で作成した名詞の意味属性分類体系の概要を表3.4.3に示す。

表3.4.3 名詞の新しい意味属性体系

		第2段	第3段	第4段	第5段
名詞	主体	人称		8	6
		呼称		8	10
		擬人		3	0
		組織		7	0
		施設名	7	0	
	場所名	地域名	9	0	
		陸地名	7	0	
		海の名称	0	0	
		景観	0	0	
		宇宙名称	0	0	
	具体物	動物(個体)	6	0	
		動物(部分)	7	0	
		植物(個体)	3	0	
		植物(部分)	8	0	
		自然物	6	0	
		人工物	9	0	
	抽象物	知的生産物	3	0	
		創作物	4	0	
		言語	0	0	
		文書	0	0	
倫理		7	0		

こと	宗教	0	0	
	伝承	0	0	
	情報	0	0	
	評価	0	0	
	制度	6	0	
	成敗・業績	4	0	
	習慣	3	0	
	精神	12	13	
	行為	9	59	
	事象	12	4	
	自然現象	5	27	
	存在	0	0	
	類・系	0	0	
	関連	10	0	
	性質	4	0	
	状態	3	0	
	形状	7	0	
	数量	8	0	
	場	13	0	
時間	10	0		
計	6種	41種	209種	114種
累計		47種	256種	370種

日本語語彙大系の一般名詞意味属性体系が、最大12段、分類総数2700件であったのに対して、新しい分類体系では、5段のフラットな構造で分類総数は364種類である。大半の文型パターンでは、第3段（50分類程度）で意味的に適切なパターンが絞り込めることを狙っているが、特に「こと」類の名詞では、かなり細かい分類としている。これは、こと類の名詞は動詞的属性を表すもの、すなわち名詞化した用言に相当するものが多いことを反映している。

### 3.5 文型パターンの作成方法

前節までに示したパターン記述言語を用いて、2つまたは3つの述部を持つ対訳標本（重文、複文15万件）を対象に、単語レベル、句レベル、節レベルの文型パターン辞書を試作した。文型パターン辞書の試作は、大きく見ると、初版を試作する過程とそれを実験的に改良していく過程に分けられる。本節では、初版を試作するまでの過程について述べる。

本研究における言語知識ベースの開発において、第1の課題は、いかなる表現意味辞書を開発すればよいかを明らかにすることであるが、開発すべき表現意味辞書は、人手作業のみで実現するにはあまりにも膨大であり、機械化による支援なしに実現することは難しい。従って、如何にしたらそれを実現することができるかは、むしろそれ以上に重要な課題とすることができる。

そこで、文型パターン化の作業項目とそれらの作業を半自動化する方法について検討を行い、「文型パターン作成手順書」を作成した。文型作成手順は、各種既存の解析ツールをベースに新しいツールを加え、半自動的な作業手順を実現するものである。

本節では、初版として、いかなる文型パターンをどのような方法で作成したかについて述べる。

#### 3.5.1 パターン化の方針と汎化のレベル

以下で述べるパターン化の方針と汎化の方法は、いずれも第1段階で実施したものである。汎化作業では、実験的検証を経てさまざまな改良を行ってきたが、それについては、3.6説で述べる

##### (1) 文型パターン化の方針

文型パターン化における個別的な方針を示す。

##### ・作業自動化の可能性の追求

膨大な対訳例文から文型パターンを能率良く作成するため、対訳例文の形態素解析と構文解析情報を使用し、文型パターン作成作業の半自動化を目指す。そのため、機械的な変数化と関数化ができるよう、文型パターン作成に先立って元となる対訳標本文（15万文、単語数180万）は形態素解析し、解析誤り（5000カ所程度）は人手で修正した。

##### ・文型パターン照合の容易性

従来の構文解析で解消困難な構文多義の問題を解決することも文型パターン翻訳方式の目標の一つである。そこで、入力文と文型パターンの照合では構文解析は使用せず、

入力文の形態素解析結果を使用することとし、文型パターンは入力文の形態素解析結果との照合において曖昧さの発生しない方法で記述する。

これに伴い、形態素解析で文法的、意味的に解釈の確定しない表現要素については文法的、意味的な指定は行わず、字面もしくは字面の指定される関数を用いて文型パターンを記述するものとした。

表3.5.1 文型パターン化作業項目とその内容

#	汎化規則の分類		汎化規則の内容	
	単語レベルの汎化	原文任意要素の削除	原文任意要素を削除する	
2		自立語の変数化	名詞一般	・格要素の名詞の変数化, ・述語名詞の変数化, ・数詞の変数化(数詞として識別), ・複合名詞の単一名詞化
3			複合名詞	述語単独動詞の変数化
4			動詞一般	単独動詞連体形,と連用形の変数化
5			複合動詞	複合動詞(主動詞部分)の変数化
6			形容詞・	述語形容詞・形容動詞の変数化
7			形容動詞	形容詞形容動詞連体形, 連用形変数化
8			副詞	文修飾, 用言修飾の副詞の変数化
9			品詞変換関数	品詞変換関数を使用する
10		述部語尾表現の関数化		英語アスペクト情報の関数化
11				英語様相情報の関数化
12		パターン任意要素の指定	名詞修飾語	動詞, 形容詞, 形容動詞の連体形と連体詞の任意化
13			用言修飾語	動詞, 形容詞, 形容動詞の連用形の任意化
14			その他	副詞的用法の名詞(昨日, 今日)の任意化
15				英語に訳出される他の線形要素の任意化
16				英語に訳出されない要素の任意化
17		表現要素のグループ化		同種の格助詞相当語や副詞をグループ化
18		各種加工	主語の補完	英語主語に相当する名詞が日本語側がない時, 日本語パターンに補う
19			冠詞の削除	英語文型内の変数された名詞の冠詞を削除
20	句レベルの汎化	句の変数化	名詞句, 動詞句, 形容詞句, 形容動詞句, 副詞句の変数化	
21		適用範囲の拡大	現在形変換	現在形でも使用される文型を対象に past() を削除
22			丁寧表現の標準化	丁寧表現をフラットな表現に変更する <注> 英語訳出されない丁寧表現が対象
23			N NP 拡張	名詞変数 N を名詞句変数 NP に置換. 不要となったパターン任意要素を削除
24			V VP 拡張	名詞変数 V を名詞句変数 VP に置換
25		機能語の拡張		格助詞を同種の格助詞相当語に置き換え
26				述部の語尾を同種語尾パターンに置き換え
27				同じ意味の副詞等をパターンに追加
28	節レベルの汎化	節の変数化	名詞節, 連体節, 連用節, 引用節, 並列節の単一変数化	
29		節レベル関数の使用	日本語節を英語句に変換する関数の使用	
30			英語節構造を指定する関数の使用	



#### ・変数書き換えによる文型パターンの汎化

句レベルの文型パターン化では、対訳標本文中の線形な句（既に単語変数化された表現を含んでも良い）を変数化するが、より汎化の範囲を拡大するため、単語レベルで変数化した単語変数についても句変数に置き換え可能なものを探して句変数化する。

例えば、名詞を表す単語変数Nを名詞句変数NPに置き換えることができれば、より適用範囲の広い文型パターンが作成できたことになる。

#### ・離散記号による文型パターンの汎化

原文任意要素の定義から分かるように、機械翻訳で使用される文型パターンは、パターン要素のすべてが入力文に含まれていることが条件となるが、逆に入力文には、文型パターンに定義されない要素を含むことも認められる。

そこで、文型パターンがより多くの範囲の入力文に適合するようにするため、入力文と文型パターンの照合を制御するための離散記号を使用し、可能な限りこの記号を使用した文型パターン化を行う。

#### ・字面レベルでの表記の揺れの吸収

日本語は表記の揺れが多く、このことが、文型パターンとの照合で大きな問題となる。この問題を解決するため、入力文と文型パターンとの照合の段階で形態素解析プログラムの持つ標準表記認定機能を利用することを前提に、原則として標準表記を用いた文型パターン記述を行うこととするが、同時に、格助詞、助動詞、副詞（いずれも相当語を含む）等について、可能な範囲での汎化を行う。

### （2）汎化の3レベル

前章で述べた原則と方針に従って、下記に示す単語レベル、句レベル、節レベルの文型パターンを順に作成する。

単語レベル・・・線形要素の自立語のみを変数化した文型パターン

句レベル・・・線形要素の句まで変数化した文型パターン

節レベル・・・線形要素の節まで変数化した文型パターン

これらの文型パターンのレベルは、使用される変数の種類に着目した分類である。各レベルのその他の作業項目の詳細を表3.5.1に示す。また、半自動化の方法を図3.5.1に示す。

### 3.5.2 単語レベルの文型パターン化

すべての対訳例文に対して、単語レベルの文型パターンを1パターンずつ作成する。文型パターン化の作業は、「原文任意要素の削除」、「自立語の変数化」、「述部語尾表現の関数化」、「パターン任意要素の指定」、「表現要素のグループ化」、「各種加工」の6種類から構成される。以下、それぞれの具体的内容について述べる。

#### （1）原文任意要素の削除

以下の2の条件を満たす要素を「原文任意要素」と判定し、原文から削除する。

対応する文要素が英文にあり、日英双方の文からそれらを取り去っても対訳関係が成り立つこと。

着目する文要素を単独で英訳することができ、英訳文に組み込めること。訳語選択、訳語位置決定の困難な要素は、対象外。

また、日本語と対応関係を持たない英文中の要素も原文から削除する。なお、 の条件は満たすが、 の条件を満たさないものが、「パターン任意要素」である。

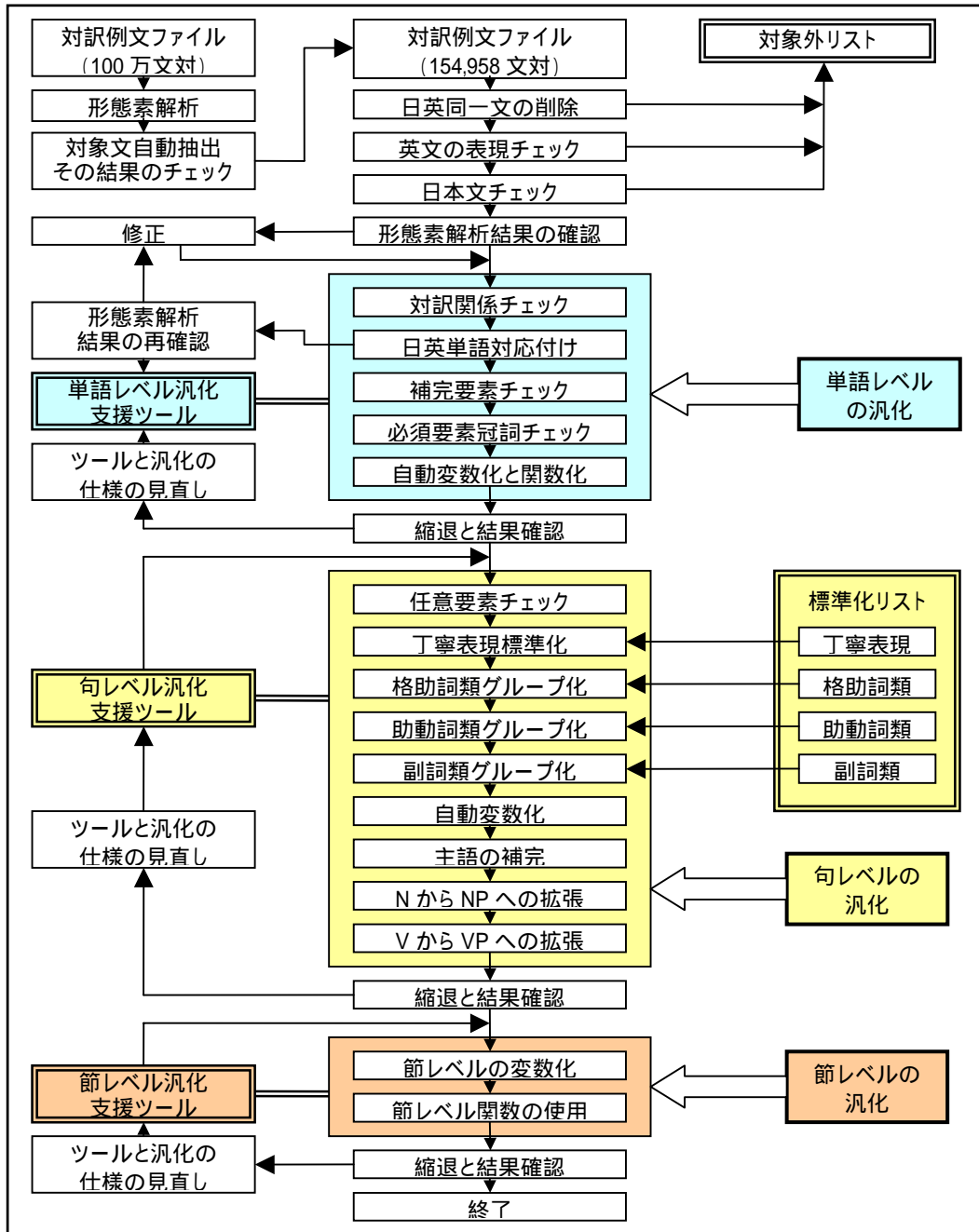


図 3.5.1 句型パターン作成手順

## (2) 自立語の変数化

語彙的な用法で使用された線形な自立語を以下の順序で変数化する。なお、機能的用法で使用された自立語は変数化しない。

まず、日本語表現内の自立語に対応する訳語が英訳文中の同一の品詞の語（複合的な表現でも良い）として存在し、対訳辞書などによってその対応関係が決定できるものを日英同一の変数（変数名と変数番号共に同一）に置き換える。

残された日本語表現内の自立語のうち、線形なものを変数化する。この場合、英語側に意味的に対応する自立語がなくても良い。また、英語側の自立語で、日本語側に対応する自立語を持たないものは変数化しない。

#### ・品詞変換関数の適用

日本語側の文型パターン要素が、英語文型パターンで文法属性の異なる自立語に対応するものを対象に「変数関数」を使用した文型パターン化を行う。対象となる表現部分は日本語側のN、Vがそれぞれ英語側でV、Nになるもので、英語側の要素を変数関数 $V(N)$ 、 $N(V)$ で表現する。

#### （３）述部語尾表現の関数化

動詞、形容詞、形容動詞に接続する助動詞および助動詞と助詞の連鎖表現（連語）を「形式指定関数」を用いて書き換える。書き換える際の注意点は以下の通りである。

この段階では、線形な時制、相、様相の情報の任意化は行わない。

従って、日本語側、英語側の文型パターンにおいて、これらの情報は、「形式指定関数」を使用するが、それが無いときは、関数を新設するか字面関数を使用する。

#### （４）パターン任意要素の指定

日英対訳文型パターンを見比べ、下記の線形要素を「パターン任意要素」として指定する。但し、英語文型パターンでこの要素が指定されたときは、対応する日本語文型パターンに同一の番号を持つパターン任意要素が指定されていなければならない。

単独の動詞、形容詞、形容動詞の連体形と連体詞

単独の動詞、形容詞、形容動詞の連用形

英語に訳出されるその他の線形要素

英語に訳出されない要素

#### （５）表現要素のグループ化

助詞、助詞相当語、または、副詞などの字面のうち、同一の意味で異なる表記を持つものを対象に、選択記号を用いて置き換え可能な表現として指定する。指定される要素は必須要素とパターン任意要素を共に含んでよい。

#### （６）各種加工

前節までで得られた日英文型パターンに対して下記の処理を行う。

##### ・主語の補完

対訳例文において英語パターンで主語となる名詞が日本語側では記述されていない（ゼロ代名詞化されている）場合、日本語側パターンのゼロ代名詞の部分にパターン任意要素記号を使用して主語を補完する。

##### ・冠詞（a、the）の削除

英文生成過程では、名詞に対する冠詞の必要性とその種類および数を決定することが重要であるが、これらは英語訳語に依存して決まることが多いため文型パターンで厳密

に定義することは困難である。そこで、英語パターンを使用した英文生成では、形態素調整処理において名詞変数の直前に冠詞のついていない名詞変数について冠詞と数の表現の必要性を判定するための処理を起動することとする。

従って、英語文型パターンに対して、文の意味上、定まった冠詞を要求する文型では、変数Nの前の冠詞を字面のまま残し、それ以外の場合は、冠詞はすべて削除する。

### 3.5.3 句レベルの文型パターン化

句レベルの文型パターン化では、「句の変数化」、「現在形変換」、「丁寧表現の標準化」、「名詞から名詞句への拡張」、「動詞から動詞句への拡張」、「機能語の拡充」の6種類の汎化作業を行う。

#### (1) 句の変数化

変数化する句は線形な「名詞句」、「動詞句」、「形容詞句」、「形容動詞句」、「副詞句」である。以下、それぞれ変数化する句の範囲を示す。

##### <名詞句の変数化>

格助詞と名詞の連鎖した表現を名詞句変数化の対象とする。また、動詞、形容詞などの単独用言の連体形により修飾された名詞の表現では、英語側に意味的に対応する部分があるものについて、修飾語と名詞の範囲を名詞句変数化の対象とする。それ以外の名詞句は変数化の対象としない。1つ以上の格要素を持つ用言によって修飾された名詞の表現は名詞句とはせず、「動詞句(形容詞句)+名詞」と解釈する。

なお、名詞句の変数化では、形容詞の部分や一部の字面や単語変数として残し、残りの部分を変数化しても良い。

##### <動詞句の変数化>

主格(が格、は格)を除く1つ以上の格要素を伴った用言について、動詞とその配下にある格要素すべての範囲の表現を動詞句変数化の対象とする。なお、主格を含む場合は、格要素と動詞を含む全体を節と解釈する。

##### <その他の句の変数化>

形容詞句、形容動詞句の変数化は、動詞句の場合に準じる。また、副詞句の場合は、名詞句の場合に準じる。

#### (2) 丁寧表現の標準化

英語文型パターンに影響のないことを確認して、日本語文型パターンの丁寧表現をフラットな表現に変更する。例えば、「~する(V)、~します(Vます)」、「~しない(V。not)、~しません」では、いずれも前者に縮退させる。

#### (3) 名詞から名詞句への拡張

修飾部を持たない名詞変数Nのうち、句に書き換えて良いものを選びNPに書き換える。また、「AJ1 N2」や「N1のN2」等のように連体修飾部を持つ名詞は、全体をNPに置き換える。後者の置き換えでは、パターン任意要素として任意化された助詞結合型名詞句の主名詞を除く部分や連体詞、単独形容詞の連体形、単独動詞の連用形を含む名詞はすべてNPに置き換える。

#### (4) 動詞から動詞句への拡張

以下の条件に従って、動詞変数Vもしくは動詞句を動詞句変数VPに置き換える。

英語側の主語に対応しない格要素を1つ以上を持つ動詞句であること。

英語側の該当する表現に主語がないこと(英語側に主語のある表現では、既に単語レベルで日本語文型パターンの任意要素として、主語が補われている)。

英語側も対応する句変数を使用されること(関数引数として使用されても良い)。

なお、日英の該当する動詞とそれを修飾するすべての格要素をまとめて一つの句変数に汎化すること。すなわち該当する格要素の一部を残してはいけない。

具体的な手順は以下の通りである。まず、修飾部を持たない動詞変数Vのうち、句に書き換えて良いものを選び、VPに書き換える。格要素や副詞に修飾された動詞の場合は、英語側の文型パターンとの対応関係を明示しなければならない要素部分を除きVPに置き換える。また、すでにパターン任意要素に指定されている副詞、単独形容詞の連用形、単独動詞の連用形は、修飾先の動詞と共にVPに書き換える。なお、時制、相と様相に関する関数は削除しない。複合動詞も単語レベルでの汎化のままとする。

#### (5) 機能語の拡充

ここまでに得られた文型パターン対の日本語文型パターン側を対象に、格助詞、格助詞相当語の部分と述部の語尾表現で置き換え可能な単語もしくは連語を考えて、要素選択記号( | | ··· )の形式のパターンとする。追加する単語、連語は、パターン全体の意味を変えないもので、通常よく使われる範囲の語でよい。

### 3.5.4 節レベルの文型パターン化

節変数CLを用いた英語文型パターンの汎化を行う。

#### (1) 変数化する節

以下の条件を満たす線形な名詞節、連体節、連用節、引用節、並列節を変数化の対象とする。

用言とそれに付随する格要素の範囲の表現(「パターン任意要素」が含まれても良い)のうち、英語表現の主語に対応する格要素を持つこと。

英語側に対応する節形式の要素が存在し、その意味は日本語側の節の意味に対応すること。

意味的対応関係は、文脈等の情報に依存しないこと。また、節内に英語文型全体を左右する要素を持たないこと。

#### (2) 変数化する節の範囲

変数化する範囲は、「命題命名のレベル」の単文(核文とも言う)とする。すなわち、節変数に代入される値は、用言(体言述部の場合は体言)とそれに付随する格要素、副詞要素の範囲とし、時制、相、様相の情報を示す関数類はそのまま残す。

### 3.5.5 文型パターンの例

文型パターンの記述例を表3.5.2に示す。以下、例に使用された変数、関数などの意味

を簡単に説明する。

表3.5.2 作成された文型パターンの例

レベル	言語	文型パターン	例文
単語 レベル	日本語	# < N1(G4)は > / V2(R3003)て / N3(G932)を / N4(G447)に / V5(R1809).tekita.	うっかりして定期券を家に忘れてきた。
	英語	I was so AJ(V2) as to V5 #[N1_poss] N3 at N4.	I was so careless as to leave my season ticket at home.
句 レベル	日本語	NP(G1022)1 は / V2(R1513).ta / N3(G2449)に / V4(R9100).teiru のだから / N5(N1453).dantei	その結論は誤った前提に基づいているのだから誤りである。
	英語	NP1 is AJ(N5) in that it V4 on AJ (V2) N3.	The conclusion is wrong in that it is based on false premise.
節 レベル	日本語	CL1(G2492).tearu ので、N2(G2005)に当たっては / VP3(R3901).gimu.	それは極めて有毒であるので、使用に当たっては十二分に注意しなくてはならない。
	英語	so+that(CL1, VP3.must.passive with subj(CL1)_poss N2)	It is significantly toxic so that great caution must be taken with its use.

< 単語レベルのパターンの説明 >

- ・ N1、N2、N3：「名詞変数」、V2、V5：「動詞変数」
- ・ (G4)、(R3003)：変数に対する意味的制約条件を意味属性番号で指定
- ・ #1[・・・]：省略しても良い要素（「パターン任意要素」）
- ・ /：入力文にはなくても良い要素（「原文任意要素」）
- ・ .tekita：述部語尾を指定するための関数（「語形指定関数」）
- ・ AJ(V2)：動詞変数V2の値を形容詞化したもの（「品詞変換関数」）
- ・ N1\_poss：N1の値を所有格に変形する（「語形指定関数」）

< 句レベルのパターンの説明 >

- ・ NP1：「名詞句変数」

< 節レベルのパターンの説明 >

- ・ CL1：「節変数」
- ・ so+that (・・・、・・・)：so that構文を生成する「構文指定関数」
- ・ subj(CL)：節変数の値から主語を抽出する「要素抽出関数」

### 3.6 文型パターン辞書の実験的改良

文型パターン化における最大の技術目標は、統語的・意味的に高い被覆率を得ることである。当初、5万件程度の重文複文を対象に汎化を行う計画であったが、15万件に上る対訳標本文を得ることができたため、比較的単純な汎化を行うレベルで実用的な被覆率が達成できることを期待し、高度な汎化作業については見合わせてきた。これは、標本文数の増大による汎化作業の増大を抑制することも狙いとしていた。

しかし、作成した文型パターン辞書（初版）の統語的な被覆率は、30～40%程度、意

味的な被覆率はその半分程度であり、目標とする被覆率は得られていないことが判明した。そこで、ペンディングとしていた汎化項目を含め、さまざまな改良項目について効果の予測検討を行い、コストの割に効果の大きいと見られる項目から順に改良を実施した。本節では、実施した検討と作業項目の概要を述べる。

### 3.6.1 離散記号付与基準の再検討

#### (1) 離散記号の詳細化

文型パターン辞書の初版では、任意の要素の省略を認める記号として、1種類の離散記号(オールマイティ)を設けていた。これは、離散記号の細かな区別は、膨大な作業を必要とするため、見合わせていたものである。

しかし、実験結果によれば、挿入可能要素の種類を指定しない方法では、あまりにも多くの不適切なパターンに適合してしまうことが明らかとなった。そこで、これを見直して、挿入可能な要素の種類が識別できるよう6種類の離散記号に分け、その任意の組合せを認めることとし、その付与基準を詳細化した。新たに導入した離散記号の種類は、以下の通りである。

/y: 連用節、/t: 連体節、/c: 格要素、/v: 動詞連体形、  
/k: 形容(動)詞連体形・連体詞、/f: 副詞

#### (2) 離散記号付与基準

新たに設けた離散記号の付与基準は、おおむね以下の通りである。

##### ・節(/y、/t)の挿入力所

おおむね、表3.6.1の通りとする。

表3.6.1 節の挿入許容箇所

挿入節	文型の直前	主節の末尾	連用節の直前	連用節の直後	連体節の直前	連体節の直後	節の内部	名詞句の直前	名詞句の内部
連用節 / y		x				x	x	x	x
連体節 / t	x	x	x	x	x	x	x		x

##### ・格要素(/c)の挿入箇所

文型パターン内に存在する格要素の前後に別の格要素の挿入を認める。但し、同一格要素内に別の格要素が挿入されてはならない。また、格要素のないところに格要素を挿入してはいけない。

##### ・連体修飾要素(/v)の挿入箇所

挿入可能な連体修飾要素は、連体詞、「Aの」、「Aと」など名詞句構成要素、形容詞連体形、動詞連体形のいずれかとし、文型パターン内の名詞句の直前への挿入を認める。同一名詞句内への挿入は認めない。連体修飾要素の挿入許容箇所を表3.6.2に示す。

表3.6.2 連体修飾要素の挿入許容箇所

挿入可能要素		名詞句直前	名詞句内	単独名詞直前	備考
連体修飾要素	連体詞		×		
	名詞 + 助詞		×		連体修飾となる助詞
	形容詞連体形		×		
	動詞連体形		×		

・連用修飾要素 ( /k、 /f )

挿入可能な連用修飾要素は、形容詞連用形、副詞、副詞句のいずれかとした。挿入可能な位置は表3.6.3通りとする。

表3.6.3 連用修飾要素の挿入許容箇所

挿入可能要素		文頭	述部の直前	形容詞直前	動詞の直前	備考
連用修飾要素	形容詞連用形					単独形容詞
	副詞(句)					

( 3 ) 離散記号挿入の作業基準

その後の実験から、上記の改良によって不適切な文型パターンへの適合は大幅に制限されることが分かったが、入力文の必須要素が離散記号に適合してしまうなど、まだ、様々な改良点が存在することが判明した。「離散記号」は、「時制様相情報の汎化の有無」と並んで適合文型数を大きく増減させる(10倍程度)パラメータであり、その設定の仕方の良否は被覆率特性に大きな影響を与える。

そこで、さらに以下の基準を設け、付与した離散記号を再度見直した。

文型パターンが4節以上の文にも適合するようにするため、節離散記号は線形要素としての節の挿入可能な場所にものみ付与する。

副作用を押さえるため、格要素、形容詞句、副詞句等の省略記号は、節レベルの文型でカバーできると判定できるところには入れない。

離散記号で解釈される要素は任意化可能な線形要素であるので、任意化不可能な要素が離散記号で解釈される現象を防止する。

連続した要素が別々の離散記号で解釈されるような現象を防止するため、離散記号で解釈される要素の前後は必須要素とする。

節の離散記号は、その適用により単文化しない場所に挿入する。

格要素と適合するような位置に連用節離散記号を付与しない。

必須格要素と適合するような位置に格要素離散記号を付与しない。

節間キーワードに適合するような離散記号を排除する。

意味のまとまらない断片的な要素にも適合するような離散記号を排除する。

3.6.2 述部語尾表現の汎化基準の見直し

( 1 ) 時制・相・様相情報の汎化の必要性

文型パターンの初版では、日本文の述部語尾に含まれる時制・相・様相の表現の大半



は、すでに関数化した。しかしこれは、汎化したものではなく、殆どの場合、非線形要素を関数化したものである。

初版の段階で、時制・相・様相の汎化を見合わせたのは、英語には時制の一致の原則が存在するなど、「日英での対応関係が個別的で汎化規則は単純ではないこと」、また、「文型パターンパーサの動作が複雑となること」などを考慮したものであった。

これに対して、その後の予備的な実験によれば、時制・相・様相に相当する関数が完全に汎化できると仮定した場合、適合率は大幅に向上（単語レベルで20%）する見込みである。現実には、述部語尾が完全に汎化できることはあり得ないことであるが、半自動的な処理が可能な範囲での汎化を行ったとしても、適合率はかなりの程度向上する可能性が期待できそうである。

そこで、時制、相、様相の情報を汎化するため、文型パターンの記述の仕組みについて、さまざまな新機能を追加し、文型パターンの汎化を行った。

## （2）汎化基準の概要

新しく設けた汎化基準の概要は以下の通りである。

### 関数の任意化と要素選択記号化

任意要素の指定と「要素選択記号化」は、変数を含む要素、もしくは、字面要素を対象としていたが、「述部語尾表現」を汎化するため、関数のみの「任意化」と「要素選択記号化」の機能を導入した。

### 字面要素への関数付与

関数は変数に対して付与されることを基本としてきたが、関数を使用した汎化を促進するため、字面に対しても関数付与が可能となるよう機能拡張を行った。

### 「対応型任意記号」と「対応型要素選択記号」の導入

「文型任意記号」と「要素選択記号」は、いずれも当該要素の有無もしくは選択が英語文型パターンに影響を与えず、英語文型パターン側に対応する任意要素を持たない要素を対象としてきた。そのため、日本語の表現で省略可能な要素に対応して省略可能性が存在する場合についての汎化は行っていなかった。

これに対して、時制等の汎化ができるようにするため、日本語側の選択要素に対応して英語側の要素を選択できるよう機能拡張した。すなわち、従来の「任意記号」と「選択記号」の機能を拡張した記号として、「対応型任意記号」と「対応型要素選択記号」を導入した。

### 「時制関数」の導入

時制を汎化するため、日英間ですべての時制に関する関数セットを設けた。

### 「グループ関数」の導入

日本語では、「N。」、「Nだ。」、「Nである。」などのように、どの記述の場合も英語文型パターンが同じになるような表現が存在し、これらを相互に適用可能とする効果は大きいと考えられる。従来は、関数に対する要素選択記号の適用は行っていなかったが、関数の要素選択記号化を認め、このような表現の選択を可能とした。

しかし、出現頻度の高い「です/ます調」や「だ/である調」の表現を常に要素選択号化することは煩雑であるし、文型パターンが読みにくくなることが問題である。この問題を解決するため、以下の2種類のグループ関数を新設した。

<同値型グループ関数>

どの関数に適合しても英語表現は変化しないような関数を集めたもの。現在の肯定を意味する述部語尾に対して「用言述部用」と「体言述部用」の2種類を設けた。

<対応型グループ関数>

日本語文型パターンと英語文型パターンで1対1の関係にある複数の関数をグループ化したもの。このタイプの関数としては、「時制グループ関数」、「肯定否定グループ関数」、「命令勧誘グループ関数」の3種類を設けた。

その他の仕組み

その他の文型パターン記述法としては、新規に「英語文型における完了形の記述方法」、「受動態の表現の記述方法」、「be動詞・have動詞と助動詞の時制表現」の記述法を導入した。

(3) 汎化の効果

以上による新しい汎化は、半自動化を考慮して導入したものであるが、現実にはかなりの人手作業を必要とした。しかし、残念ながらクロス照合実験(クロスヴァリデーションの方法)では、期待していたほどの効果は得られなかった。クロス照合実験は、パターン作成に使用した標本文12万件を入力文に使用した実験である。すべての入力文に対して、それを標本として作成されたパターンが存在し、それには必ず適合するから、適合率は、それ以外に適合したパターンによって評価される。

その後の分析によれば、その理由は、以下の通りと考えられる。すなわち、入力文に使用した標本文は、いずれも典型的な日本文であるが、そのような日本文では、時制・相・様相の表現も、かなりの程度、定型的であるため、この部分を汎化しても適合する入力文は、大幅には増えないということである。

このように、クロス照合実験では、労力の割に期待したほどの効果はなかったが、より一般の日本文を対象とする場合は、効果がある可能性がある。今後の評価を待つこととしたい。

### 3.6.3 英語パターン記述法の改良

(1) 英語パターンの不良の分析

入力文に適合した文型パターンが意味的に適切であるなら、対応する英語パターンによって意味的に適切な英文が生成できるはずである。このことに着目して、適合したパターンの意味的な適切性を評価したところ、英語パターンの記述方法にさまざまな問題があることが分かった。すなわち、意味的に適切と判断できるパターンでも英語パターンの不適切な記述が原因で正しい英文が生成できないものが多数存在する。

そこで、この問題の大きさを調べるため、対訳例文の英文と文型パターン辞書の英語文型パターンとの照合実験を行ったところ、英語文型パターンの約15%は自己例文に適

合しないこと、また、英語パターン記述法の改良によって、適合文型の意味正解率は最大20%程度向上する可能性がある」と推測されることが分かった。

そこで、「英語表現の骨格が分かり易い記述法にすること」を目標に、英語文型パターンの記述法を体系的に見直すこととした。しかし、新しい記述法を英語パターンに適用するには、膨大な量の作業が発生する。そこで、これらの作業が半自動的に実施できるようにするため、まず、英語構文に対する分類体系を作成し、各パターンに対して英語構文分類コードを付与した。

### (2) 英語構文情報の付与

英語構文分類情報を付与したのは、重文複文の基本ともなる文種別1、文種別3のパターンである。分類の基準を表3.6.4に示す。

### (3) 英語パターンの改良

前項で構文分類した情報を利用して、英語パターンの大幅な見直し作業を実施した。記述を改良した主な項目を以下に示す。

- ・ 節の変数化基準の見直し、
- ・ 語形変形関数ingの現在分詞と動名詞の区別
- ・ 修飾要素としての形容詞 / 形容動詞 / 動詞 (分詞) と述部の用言の区別
- ・ 不定詞 "to" と前置詞 "to" の区別、
- ・ 副詞が挿入された完了形の時制
- ・ 限定詞に関する新表記法の導入、
- ・ 目的語が挿入された群動詞の記述
- ・ その他記述法 (前置詞、所有格代名詞、省略目的語等)

英語パターンに分類コードが付与されたことにより、改良項目と改良パターンの関係が機械的に調べられるようになり、作業のコストと作業品質が向上したが、特に下記の点での効果は大きかった。

表3.6.4 英語構文タイプ

文種別	分類		詳細項目 (英語構文の訳出タイプによる分類)	該当件数
文種別 1	KWでの 絞込みが 可能	誤りの抽出 が少ない	従属接続詞(前置詞、関係詞としては不使用)	6,020
			慣用表現	434
			複合前置詞	384
			英語特有表現	1,029
			あいさつ表現	83
	誤り抽出 が多い	複数分類 が必要	等位接続詞(andを含む文)	8,107
			従属接続詞 after, until, before, till, since, while, as など前置詞にもなるタイプ、when など関係詞となるタイプ	5,867
			前置詞	2,689
			述部一体型 (モダリティなど)	1,318
	KWでの絞込みが困難	誤り抽出 が多い	付帯状況(withを含む文)	2,361
			不定詞(toを含む文)	4,804
英語的な表現			800	
		その他	19,422	
文種別 3	KWで絞込みが可能 (複数分類が必要)	That 節(文頭の That 以外に that を含む文)	5,014	
		関係詞節・疑問詞節(what, who, where, when, why, how, whose, whom, whatever, whoever, wherever, whenever を含む)	7,988	
		KWでの絞込みが困難		32,679

英文の骨格となる箇所に対する誤った変数化が容易にチェックできた。

汎化可能なレベル（単語・句・節）の判定が容易になり、過汎化が防止された。

なお、節レベルの汎化では、従来、文脈依存の省略要素を持つ節は汎化の対象外としていたが、省略要素の補完問題は英文生成に任すこととし、節内基本構造が保たれる範囲まで節変数化を行った。その結果、節レベルのパターン数が大幅に増加し、被覆率も向上した。

#### （４）見直しの効果

見直しの効果について個別には評価できていないが、目視確認による評価では、英語パターンの不良が大幅に減少した。また、さまざまな改良を含む評価では、被覆率が大幅に向上しているが、英語パターンの見直しは、特に、適合パターンの意味的な正解率向上で効果があったと推測される。

### 3.6.4 高頻度適合パターンの削除

適合頻度の高いパターンは、汎化されすぎていて適合しても余り役に立たないものが多く、その存在は最尤パターンの選択を困難にしていると考えられる。そこで、代替されるパターンが存在することを条件に、適合頻度の高い文型パターンを削除した。

以下、その方法を示す。

#### （１）削除するパターンの検出方法

パターン $P_i$ に対して、代替パターンが存在するということは、

$P_i$ に意味的に正しく適合した入力文 $S_i$ に対して、適合するパターン $P_n(i, n)$ が存在すること。この場合、 $S_i$ と $P_n$ の日本語文型は同じ。

また、適合したパターン $P_n$ の英語パターンが、 $P_i$ の英語パターンと同じ構文構造を持つこと

の２点を意味する。これより、削除可能なパターンを以下の手順で判定する。

まず、全パターンと全標本文との間でクロス照合実験を行い、文型パターンと適合した入力文の対応関係をマトリックス上にまとめた。但し、その際、自己パターンへの適合は無視する。この結果を基に、適合した入力文の数 $N$ の多い順に、そのパターンを削除しても良いかどうかを以下の方法で調べた。

<手順１> 着目したパターン $P_i$ に適合した入力文 $S_j$ を調べ、その適合は意味的に正しいか否かを調べる。

<手順２> 意味的に正しい場合は、その入力文 $S_j$ に適合する他のパターン $P_k$ を順に調べ、一つでも正しいパターンに適合していれば、その入力文 $S_j$ はOKと判定し、次の入力文 $S_{j+1}$ の検査に進む。

<手順３> 手順２で一つも正しいパターンに適合していない時は、その入力文 $S_j$ はNOと判定し、パターン $P_{i+1}$ の検査に進む。この場合、着目したパターン $P_i$ は削除せず、そのまま残す。

<手順４> 手順２で総ての $S_j$ に対して、意味的に正しいパターンが一つ以上存在するときは、着目したパターン $P_i$ を削除する。

<手順5> 上記の手順をある適当なNになるまで繰り返したのち、残ったパターンを取り出す。

#### (2) 意味的な正しさの判定方法

上記の手順では、入力文と適合したパターンが意味的に正しいかどうかを判定する必要がある。一般に、入力文と適合したパターンが、意味的に正しいかどうかは、通常は、意味的に正しい英文が生成できるかどうかである。正しい英文は一通りとは限らないから、厳密には、人間が見て判定するしか方法はないが、すべての英語例文は、既に構文分類が行われているから、この情報を使用して、当該パターンPiと同じ英文を生成できるかどうかを自動的に判定した。その方法は、以下の通りである。

手順1では、検査対象パターンPiで対応する英語パターンは1つに固定されているから、意味的に正しいかどうかは、適合した入力文がそのパターンを使って訳せるかどうかということである。

そこで、日英構文対応表を使用して、適合した入力日本文に対応する英語構文の分類コード（一般に複数個存在する）を調べ、当該パターンPiの英語パターンの構文分類コード（自己例文の英文の構文分類コード）がその中に含まれていれば、意味的に正しいと判定した。

手順2以降で、代替パターンの有無を検査するときは、適合した入力文を当該パターンPiと同じ英語パターンの構造で訳しても良いかどうかを判定基準となる。

そこで、日英構文対応表を使用して、Sjに適合したパターンPjnの英語パターンが、Piの英語パターンと同じ構文構造を持つかどうか判定し、構文コードが一致するとき意味的に正しいと判定した。

#### (3) 対象パターン数と効果の見積もり

チェック対象とする文型パターンは、クロス照合実験において、単語レベルの場合300文程度、句レベルの場合6000文程度、節レベルの場合4000文程度以上に適合するパターンとした。予備的な実験によれば、入力文と適合したパターンは常に意味的に正しいと仮定したとき、単語レベルのパターンで削除されるものは約7000件である。

### 3.6.5 各種不良チェックによる改良

既に述べたように、被覆率の向上を狙って、さまざまな照合実験を繰り返し行って問題点を分析すると共に、問題のありそうな部分についての手によるチェックなどを行い、問題表を作成した。また、それらの問題表に基づき、作業体制を組んで改良と実験的検証を行った。本節では、それらの中で、比較的まとまった作業として実施したものについて、その概要をのべる。

#### (1) 自己例文への適合性の検証

各パターンは、対訳標本文を汎化することによって作成しているから、どのパターンにも元となった例文（自己例文と呼ぶ）が存在し、照合プログラム（パターンパーサ）の解析能力を超える表現を含まない限り、その例文との照合では必ず適合するはずである。このことを利用して、各パターンの汎化不良を検出し、改良作業を行った。

この作業と合わせてパーサの機能改良を行ってきた結果、単語レベルのパターンで適合率72.1%、節レベルでは98.0%の適合率が得られるまでになった、

#### (2) 変数適合表現のチェックと改良

対訳例文の汎化では線型要素を変数化したが、入力文と自己パターンを照合させると、入力文の変数化した部分が予想通り自己パターンの変数に適合しない場合がある。その原因として、対訳文の汎化の際の形態素解析の結果とパターンパーサが使用している形態素解析の結果が一致しないことも考えられるが、汎化を誤っている可能性もある。

そこで、すべての変数を対象に、該当する箇所(約5000カ所)を調査し、間違っただけの汎化が行われていると認められた箇所を修正した。但し、形態素解析結果の違いによる場合はパターンは変更しなかった。

#### (3) 字面表記に着目した改良

文型パターンの縮退化を考えると、字面で表記された非線形要素が手がかりとなる。字面表記が共通しないパターンが9万件近く存在するが、これらは、縮退化できないことを意味する。汎化不足の要素を検出するために、字面文字列内に汎化可能な自立語が残っていないかを調べた。本方式による改良の手順は以下の通りである。

全パターンから字面の部分を取り出してリストを作る。

リストに収録された文字列の全入力文(12万文)内での出現頻度を調べる。

出現頻度の低いもの(最低は1:自己例文には出現するから)を抽出する。

抽出された字面を持つパターンについて、汎化できる部分が字面で残っていないかを調べ可能なものを汎化する。

#### (4) 離散記号を利用した不良の発見と改良

各文型パターンは離散記号を削除しても自己例文に適合することが保証されなければならない。このことを利用し、以下の2種類の方法で文型パターンの不良箇所を発見し修正した。

##### 【第1の方法】自己例文適合性の検証

離散記号を削除した自己例文との照合実験を行って不適合となった文型パターンを対象に不良の原因を調べ、適切な改良を行った。

##### 【第2の方法】異文種別パターン不適合性の検証

離散記号が存在する場合、文型パターンは同一の文種別の入力文以外にも適合することがあるが、離散記号が存在しない場合、入力文は自分と同じ文種別の文型パターンにしか適合しないはずである。このことに着目し、離散記号を削除したパターンに対する文種別間のクロス照合実験を行って不良パターンを抽出し、抽出した不良パターンに対して適切な改良を行った。

#### (5) 単文適合パターンの改良作業

本研究で試作した文型パターンは、重文複文を対象としている。これに対して、クロス照合実験において重文複文に適合した文型パターンを見ると、主節の部分にしか適合しないパターンがかなり存在する。これは、従属節の部分は、離散記号で解釈されたことを意味するから、そのようなパターンは、従属節のない文、すなわち単文にも適合す

ることになる。しかし、単文に対しては、単文用のパターン辞書を作成すればよいので、本研究の文型パターンは単文に適合する必要はない。このようなパターンは重文複文の意味処理に役立たないだけでなく、むしろパターン選択を困難にする原因となる。

そこで、単文に適合する文型パターンを実験的に検出し、本来の目的とする重文と複文にしか適合しないように改良した。適合する原因は従属節の部分が離散記号に吸収されているため、従属節部分が任意化されているためだと推定される。そこで、任意化された要素に従属節が適合しないよう改良した。また同時に、連用節、連体節の離散記号が適切かどうかを確認する。また、場合によってはパターンを削除した。

単文に適合するパターンは、適合頻度の高い文型パターンに多い。この作業は意味的に不適切な適合を減少させる点でも効果があった。

#### (6) 局所言い換え辞書利用法の検討

表記の揺らぎの吸収は、当初から取り組んできた課題であるが、文型パターンのひとつひとつを細かく検討することは工数上、大変難しい。これに対して、最近、言語表現の言い換え技術に関する研究が盛んになってきている。そこで、既存の日本語言い換えシステムを利用する方法、「文型パターン辞書」と「言い換え表現辞書」を併用する方法について検討した。

しかし、いずれの方法も、文型パターン辞書そのものの品質向上とは言えず、文型パターン辞書の利用技術と考えられるため、急ぐ必要はないと判断される。そこで、この検討は打ち切り、表記上の揺らぎの問題では、「標準表記への統一作業」と「選択記号の適正化作業」を行った。「選択記号の適正化作業」は既に選択記号化された表現をリストアップし、選択記号化されていないものの中からそれらと一致するものを検出し、選択記号化の可能性を検証するものである。

#### (7) 個別的表現の点検と改良

個別的な点検と改良に関しては、大小細々とした点検と改良を実施した。その例を下記に示す。

- 句動詞（群動詞）の表現方法の点検（8000件）
- 再帰代名詞を持つ英文のパターン化の点検（2000件）
- 慣用表現となりやすい身体的表現を持つパターンの点検
- 結果を示すSVOC文型における過汎化の有無の点検
- 機能語表現の過汎化の有無の点検

### 3.6.6 不適合試験文に着目した改良作業

前節までで述べたように、文型パターンの被覆率向上の対策として、さまざまな方法を検討し実施したが、適合するパターンが存在しない入力文を対象に、なぜ適合するパターンが存在しないのかが分かれば、対策の有無は容易に検討できる。しかし、クロス照合実験で適合パターンの存在しない4万件の入力文（単語レベルの場合）の一つ一つに対して、膨大なパターン辞書に、なぜそれに適合するパターンが存在しないのかを知ることとはとても困難な話である。

そのため、この問題についてはペンディングとしてきたが、個別的な改良の効果が次第に得られなくなってきたため、最後の段階で挑戦することにした。

(1) 狙いと意義

この作業は、「従来の改良とブラッシュアップは果たしてどの程度限界に達したか」を明らかにし、限界に達していないと判断される部分について、見逃されている問題点を分析して対策を講じることを目的とする。限界に達していると思われる場合は、現状のパターン記述方式のどの部分からその限界が生じているかを明らかにした上で、それを突破するための方策の有無を検討し、可能な範囲での対策を講じようとするものである。以下、本作業の手順について示す。

(2) クロス照合実験と結果の分類

12万件の例文を入力文に使用したクロス照合実験を行い、その結果に基づき入力文とそれに対応する自己パターンのペアを図3.6.1に示すような4つのグループに分類する。

分類した例文とそれに対する自己パターンは、表3.6.5で示すような性質を持つことが予想される。

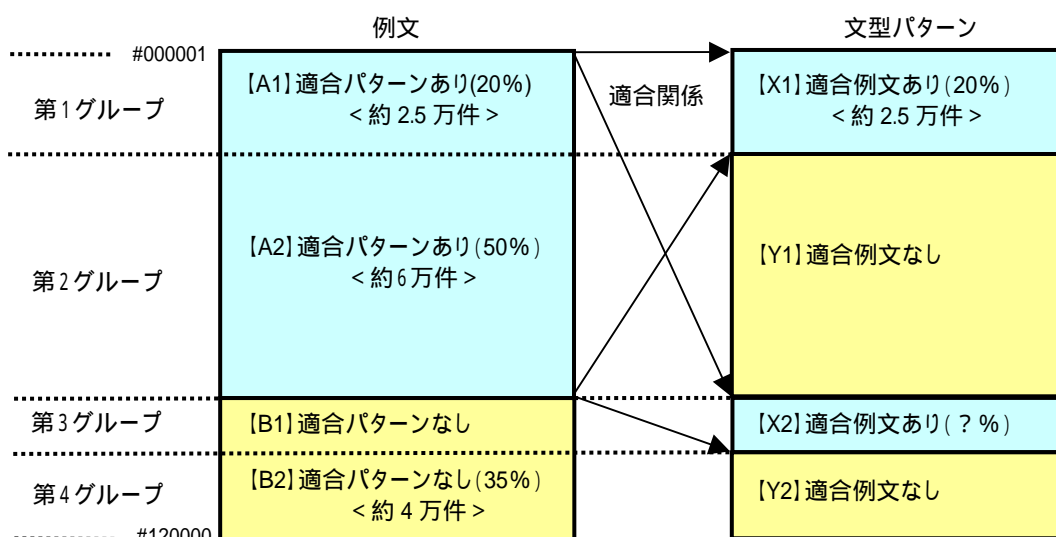


図 3.6.1 試験例文と文型パターンの適合関係 (単語レベル文型パターンの場合)

表 3.6.5 ペアグループの特徴

ペアグループ	推定される特徴
[A1 - X1] (第1グループ)	他己例文に適合するパターンを持つグループで、例文は標準的であり、パターンも適切な汎化が行われていることが推定される。
[A2 - X2] (第2グループ)	他己例文に適合する例文であるので、入力文は比較的標準的だと推定されるが、他己例文が適合しないので、うまく汎化されているかどうか疑問が残る。
[B1 - Y1] (第3グループ)	例文は他己パターンに当たらないので、あまり標準的な例文でなく特殊な点があると推定されるが、そのパターンには他己例文が適合するので、おかしい。実際にこのグループに該当するものは無いかも知れないが、もしあれば、入力文の特殊な部分が無理に汎化され、適合してはいけないような入力文に適合するようになってしまっている可能性がある。
[B2 - Y2] (第4グループ)	適合パターンが無いから、再現性の少ない特殊な部分を持った例文であると考えられる。またその例文から作成したパターンは他の例文には適合しないから、特殊な部分が非線形要素として適切に処理されている可能性が高いと推定される。



### (3) 実験的検討の方法

表1の結果に基づき、第3と第4のグループを検討の対象とする。第3のグループに該当するものは少ないと思われるが、存在すれば問題点を検討し、可能な範囲での改良を行う。以下では、第4グループの分析方法を示す。

第4グループについて、まず、例文が、(a)全体が特殊な言い回しの構文であるか、(b)特殊な部分を持つ例文であるか、それとも、(c)通常の変現であるかを調べ、それぞれの場合に応じて以下の処理を行う。

#### (a)全体が特殊な言い回しの構文の場合

この場合は、そもそもまれにしか出現しない例文であるから、適合するパターンが無くてもやむを得ないものとして、そのままとする。

#### (b)特殊な部分を持つ例文の場合

この場合は、特殊と思われる部分を取り出して、特殊性の種類によって分類し、以下のように処理する。

##### 共通的な特殊性

分類した特殊性の種類の内、パターン記述方式の改良で共通して解決が期待できるものについて、改良を実施する。

##### 個別的な特殊性

特殊と見られる部分を通常の変現に書き換えて、照合実験を行い、適合するパターンの有無を調べる。適合するパターンがあるときは適合したパターンを改良して(選択記号で特殊な部分を併記するなど)元の例文でも適合するようにする。

#### (c)通常の変現の場合

特殊な部分を持たない通常の変現であれば、それがなぜ他己パターンに適合しないかを検討する。通常の変現であるのに他己パターンに適合しないのなら、たまたま、そうなたただけで、パターンとしては重要である可能性が高い。このようなパターンは、オープン試験文に対して効力を発揮する可能性が高いので、マークしておき、辞書圧縮版の作成でも残すようにする。

### (4) 作業支援環境の実現

上記の処理では、混合パターン辞書の場合、数千件(最大1万件程度)、また、単語パターン辞書の場合、1.5万件(最大3万件)に上る入力文の部分書き換えと、パターン辞書との照合実験を繰り返す必要がある。この作業を効率化するため、人間の指示に従って実験を繰り返し、改良効果を自動的に検証する仕組みを準備した。

## 3.7 文型パターン辞書の特性

### 3.7.1 作成された文型パターン数

#### (1) 作成された文型パターン数

得られた文型パターンの種類と異なりパターン数を表3.7.1、図3.7.1に示す。

表 3.7.1 文型パターンの種類と異なりパターン数

文種別	説明	単語レベル	句レベル	節レベル	合計
文種別1	文接続1カ所を持つ文	55,509	36,063	17,941	109,513
文種別2	文接続2カ所を持つ文	5,664	3,239	313	9,516
文種別3	埋込み文1つを持つ文	42,492	28,451	5,038	85,981
文種別4	埋込み文2つを持つ文	5,638	4,028	781	10,447
文種別5	文接続と埋込文各1つを持つ文	12,610	8,175	1,530	22,315
	合計	121,913	79,956	25,603	227,472

単語レベルのパターンでは、日英対訳標本文の汎化によって縮退したパターンは少なく、殆どすべての標本文から異なるパターンが得られた。これに対して、句レベルのパターン化では、約4割の標本文において線形な句が存在せず、汎化できたパターン数は8万件程度に減少した。さらに、節レベルで汎化できた標本文は全体の2割程度形要素であることを示している。

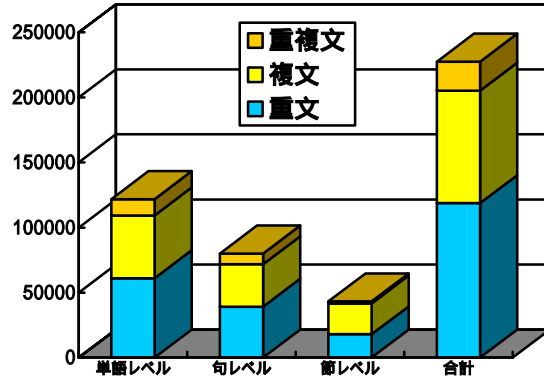


図 3.7.1 パターン数

表 3.7.2 線形要素の割合

要素種別	全要素数	変数の数	線形な割合
単語(自立語)	763,968	472,521	62 %
句	463,636	102,000	22 %
節	267,601	25,603	9.6 %

(2) 汎化された要素の割合

各レベルの汎化において変数化された要素数を表3.7.2に示す。

汎化された線形要素の割合 (図3.

7.2) は、単語レベルでは、62% (47万語 / 76万語)、句では、22%であったのに対して、節ではきわめて少なく4.3%であった。

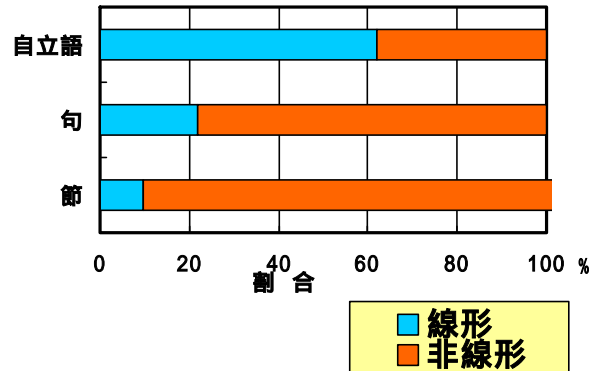


図 3.7.2 非線形要素の割合

非線形要素は、それを取り出して翻訳し、元の文に組み込んでも意味的に適切な翻訳結果は得られない。文型パターンの元となった対訳例文では、日本文が重文と複文であるのに対して、英語訳文の多くは単文となっている。また、単文化する方法は、実にさまざまである。上記の結果から、従来のような要素合成法では対訳例文に示されるような質の良い翻訳はできないことがわかる。

3.7.2 被覆率の評価

(1) 被覆率評価の方法

入力文とのパターン照合実験によって、パターン辞書の被覆率特性を評価した。実験

の条件は以下の通りである。

変数の意味的な制約条件は無視する

入力文に適合する文型パターンの中から意味的に適切なパターンを選択する方法は別の課題と考え、ここでは、得られた文型パターンの潜在的な能力として被覆率を評価する。

実験はクロスヴァリデーションの方法とする

文型パターン作成用の例文の中からランダムに1万文を抽出し、入力文として使用する。但し、当該入力文から作成された文型パターン（自己パターン）へは必ず適合するため無視し、それ以外のパターンへの適合のみを評価する。

ところで、入力文に対して適合する文型パターンは通常複数存在し、それらがすべて意味的に正しいとは限らない。そこで被覆率を以下のパラメータによって評価する。

- ・適合率（R）：入力した文のうち、1文型パターン以上に適合した入力文の割合で、構文的な被覆率を意味する。
- ・正解含有率（P）：1入力文に適合した文型パターンの中に意味的に正しく、入力文の翻訳に使えるパターンが1つ以上存在する割合
- ・意味的被覆率（C）：任意の入力文に対して、文型パターン辞書の中に意味的に適合するパターンが1つ以上存在する割合（R × P）

## （2）適合率の飽和特性

単語レベル、句レベル、節レベルの文型パターンの数と適合率の関係を図3.7.3に示す。この図から、適合率の飽和傾向が顕著である。横軸の文型パターンを使用頻度順に並べ替えると、飽和の速度は5倍程度速くなる。単文をほぼ網羅するために必要な結合価パターン数は、約2.5万件であると推定されている（Shirai et.al. 1995）。重文、複文の場合も、ほぼ網羅するために必要なパターン数も同じオーダーに収まることが期待される。

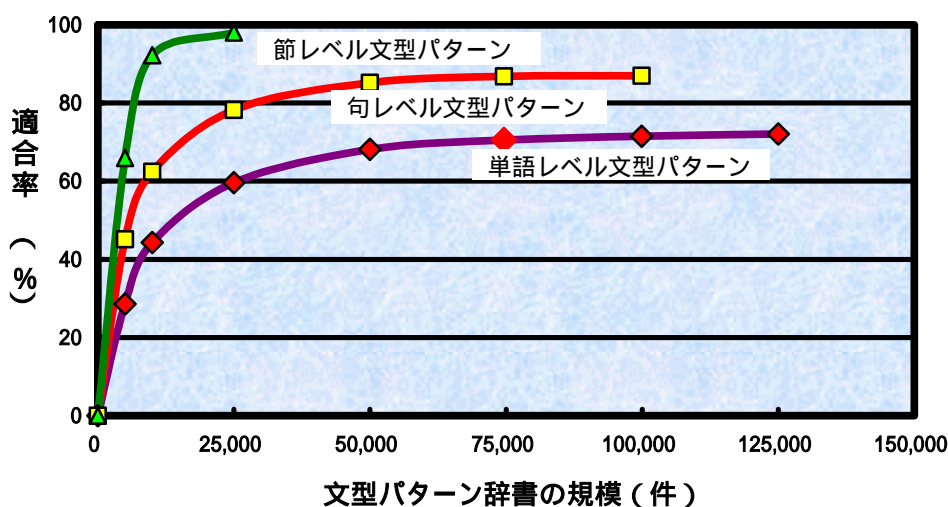


図 3.7.3 文型パターン適合率の飽和特性

### (3) 適合率と正解含有率

次に、文型パターン辞書全体の適合率と正解含有率を表3.7.3、図3.7.4に示す。この表から、文型パターン辞書全体では、構文上、入力文の90%以上をカバーすることが分かる。しかし、意味的に不適切で翻訳に使用できないパターンに適合する場合も多く、それらを除いたときの意味的な被覆率Cは70%である。

パターンの種別では、句レベルの文型パターンが適合率、意味的被覆率共に最大である。節レベルの文型パターン数は単語レベルの1/10しかないが、その割に被覆率は高い。パターン毎の汎用性の点から見ると、節レベルの文型パターンは、単語レベルと比べて10倍以上広い範囲をカバーしている。

表3.7.3 文型パターン辞書の被覆率特性

文型パターンのレベル	平均適合文型数	適合率(R)	正解含有率(P)	意味的被覆率(C)
単語レベル	44 件	72.1 %	76.1 %	55.0 %
句レベル	358 件	87.0 %	80.4 %	70.0 %
節レベル	507 件	98.0 %	72.4 %	71.0 %
合計	- -	98.5 %	- -	79.5 %

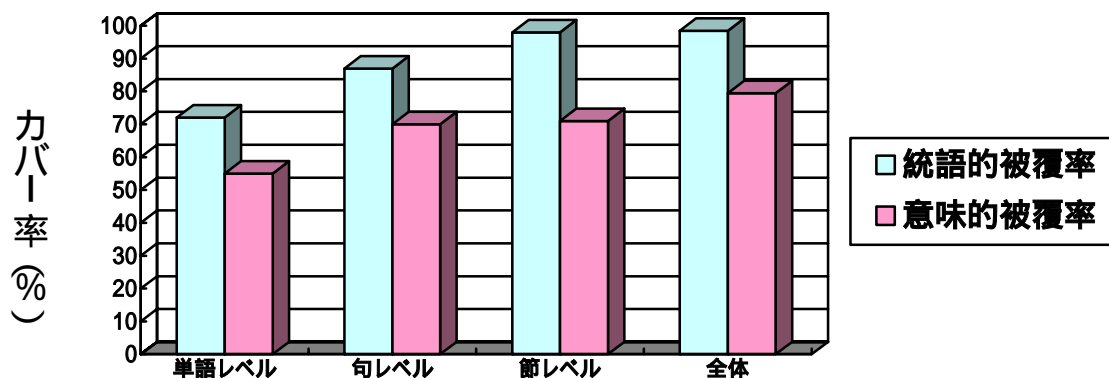


図3.7.4 統語的カバー率と意味的カバー率

### (4) 意味的被覆率

3種類の文型パターンは、単語レベル、句レベル、節レベルの順に適用範囲が広くな

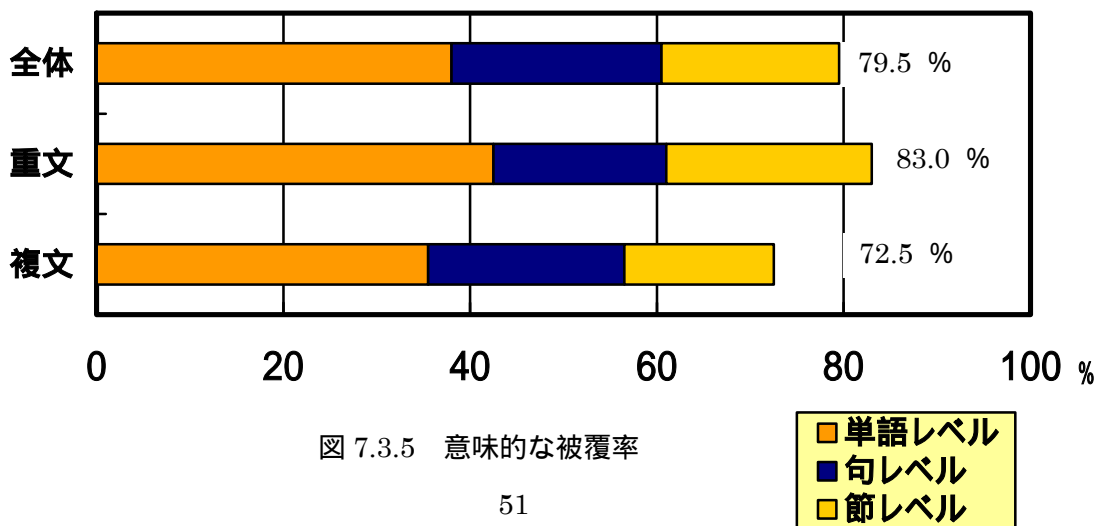


図 7.3.5 意味的な被覆率

るが、逆に意味の曖昧さが増大する。このことから、複数のレベルの文型パターンに適合する入力文の場合は、この順に意味的に適切なものを選んで使用するのが良いと思われる。そこで、この順に文型パターンを選択して使用するとき使用されるパターンの割合を図3.7.5に示す。なお、図では、最も基本的な文種別1と文種別3の文型パターン（いずれも述部数2）の場合についても示した。

図3.7.5から、文型パターン辞書を使用した機械翻訳では、重文・複文の約半分に単語レベルの文型パターン、残りの半分に句レベルと節レベルの文型パターンが使用され、最後の残りは、従来の翻訳方式が適用されることになると予想される。

### 3.8 文型パターンの意味類型化

文型パターンの意味類型化は、文型パターンを意味によって分類するものであり、各パターンに意味分類コードを付与することで実現されるが、本研究では、後の利用を考え統語的な構造による分類も行った。

#### 3.8.1 統語構造によるパターン分類

対象とする文型パターン辞書に収録された文型パターンは、2つまたは3つの述部を持つ重文と複文の持つ非線形な表現形式をパターン化したもので、その統語構造は、大きく見て、表3.8.1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。

そこで、統語構造による文型パターンの分類としては、この表に従って定めた分類コードをすべての文型パターンに付与することとした。

#### 3.8.2 意味によるパターン分類

##### (1) 意味分類の方法

既に3.1節で述べたように、言語表現間の意味的等価変換は、真理項を介して行われる。真理項は、文型パターンの表す意味、すなわち概念に相当する。また、意味類型化は、パターンを意味的にグループ化することであり、これを行うには、各文型パターンの表す概念を体系化することが必要である。

そこで、重文と複文の表す概念について考えると、単文は述部用言で代表される単一の現象を表す表現の枠組みだと考えられるのに対して、重文と複文は、複数の単文を組み合わせた表現であるから、事象間の関係を表すための枠組みだと言うことができる。従って、重文と複文の表す概念は、事象間の関係概念である。従来、このような関係概念を従属節の意味として分類された例（増岡）がある。本研究では、これを参考に「節間の意味分類体系」を作成した。

ところで、意味的等価変換方式において、真理項は原言語と目的言語の表現形式間の対応関係を決めるものである。大量の文型パターンの意味を対応づけるには、かなり細かな分類体系が必要とされる。そこで、本研究では、事象間の関係（節間）の意味分類だけでなく、それぞれの事象の意味、すなわち、重文を構成する単文の意味も真理項として使用した。

表3.8.1 文型パターンの統語的分類

文種別	説明	統語構造	文型の分類
文種別1	文接続1箇所を持つ文	連用節 + 主節 └──┬──┘ ↑	副詞節 + 主節      並列節 + 主節 └──┬──┘      └──┬──┘ ↑                    ↑
文種別2	文接続2箇所を持つ文	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	埋込み文1つを持つ文	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	埋込み文2つを持つ文	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	文接続と埋込み文各1を持つ文	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    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    副詞節 + 名詞修飾節 + 名詞 + 述部 └──┬──┘                    └──┬──┘ └──┬──┘ ↑                    ↑    ↑                    ↑    ↑                    ↑    ↑ 主節 並列節 + 補足節 + 述部    並列節 + 名詞修飾節 + 述部 └──┬──┘                    └──┬──┘ └──┬──┘ ↑                    ↑    ↑                    ↑    ↑                    ↑    ↑

以上から、本研究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、 と では、共に単文の意味分類体系を使用するものとする。以下、「節間の意味分類」と「単文の意味分類」を体系化した方法について述べる。

(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。接続構造を持つ文では、係り元の連用節が従属節であり、係り先の節が主節である。また、埋め込み構造を持つ文では、埋め込まれた節が従属節、埋め込み先の節が主節である。重文と複文の構造を併せ持つ文では、埋め込み關係から見たときの主節が、接続關係から見たときの従属節になることがある点に注意が必要で

ある。

本研究では、このような2節間の意味分類体系として、増岡の研究（増岡1992）と実際の句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 3.8.2 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(補語相当節)	2	7	13	23種類
2	名詞修飾節(連体節)	2	7	9	19種類
3	副詞節(副詞的連用節)	16	32	118	167種類
4	並列節(並列型連用節)	2	7	8	18種類
計	4種類	22種類	53種類	148種類	227種類

益岡は、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、この研究結果から、約100種類の分類用語を取り出し、意味分類体系を作成した。引き続き、約1000件の句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種類の観点から詳細に分類し、その結果を基に、表3.8.2に示すような150種類の項目で4段階の階層構造からなる新しい意味分類体系を作成した。

### (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。

表3.8.3 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	9	48
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	形容詞分	性状既定の表現	6	34	31	72
計	3	11	55	182	123	372
11	だ文	判断指定の表現	41	209	114	370

<注>「形容詞文」は、形容動詞文（な型形容詞文）を含む

表3.8.4 「だ文」(名詞)の意味分類

#	2段	3段	4段	5段	合計
1	主体	9	26	16	61
2	場所名	6	23	0	29
3	具体物	6	40	0	46
4	抽象物	12	27	0	39
5	こと	4	38	98	140
6	抽象的關係	9	55	0	64
計	6	41	209	114	370

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3.4節で作成した新しい意味分類体系を使用することとした。節(単文)の意味分類体系の分類数を表3.8.3に示す。このうち、「だ文」(体言述部)の場合の分類数を表3.8.4に示す。

### 3.8.3 統語的意味的分類コードの付与

文型パターンに表3.8.5に示すような文法的、意味的な分類コード10種類を付与した。各欄のコードの意味を表3.8.6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进行分类するものであり、第4欄から第10欄まで7種類のコードは、文型パターンを意味分類するための真理項の暫定案とするものである。本来、文型パターンの意味分類では、後者のみを使用する予定であるが、参考のため統語構造上の情報も付与している。

表3.8.5 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
コード	文種別	統語構造 第1分類	統語構造 第2分類	第1従属 節の意味	第2従属 節の意味	主節の 意味	第1節間 の意味	第2節間 の意味	第1節間 KW	第2節間 KW
文種別 1	文種別 = 1	= "0"	副詞型 並列型	連用節の 意味分類		主節の 意味分類	接続の タイプ		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 節の接続 タイプ	第2連用 節の接続 タイプ	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			埋込の タイプ		節間 KW	
文種別 4	文種別 = 4	0	0	第1連体 節の意味 分類	第2連体 節の意味 分類			埋込の タイプ		節間 KW
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 連用節の 意味分類			埋込の タイプ	節間 KW	節間 KW



表3.8.6 文法的意味的分類コードの種類と内容

欄番号	コード名	付与する方法
第1欄	文種別コード	文型パターンの文種別番号をそのまま付与。
第2欄	統語構造第1分類コード	節の意味を無視した節間の係り受け分類コードを付与
第3欄	統語構造第2分類コード	節の意味を考慮した節間分類コードを付与
第4欄	第1従属節意味コード	第1従属節の述部用言の意味分類コードを付与
第5欄	第2従属節意味分類コード	第2従属節の述部用言の意味分類コードを付与
第6欄	主節の意味分類コード	主節の述部用言の意味分類コードを付与
第7欄	第1節間意味分類コード	第1係り受け関係にある節間の意味コードを付与
第8欄	第2節間意味分類コード	第2係り受け関係にある節間の意味コードを付与
第9欄	第1節間KWコード	第1係り受け関係にある節間のKWを記入
第10欄	第2節間KWコード	第1係り受け関係にある節間のKWを記入

(1) 半自動化の方法

統語的意味的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工数的に不可能である。本研究では表3.8.7で示すような方法で半自動化した。

表3.8.7 統語的・意味的分類コード付与の方法

分類の種類	半自動的な分類の方法
統語構造分類	日英機械翻訳プログラム ALT-J/E の構文解析プログラムを使用して格文型パターンの元となっている用例文(日本文側)の構文解析を行い、節間の係り受け解析を求め、その結果に基づき「統語構造第1分類」と「統語構造第2分類」を行った。
節の意味分類	第1に、本研究で作成した節の意味分類体系とALT-J/Eの用言意味属性体系(36属性)との対応関係を表にまとめ、後者から前者へのコード変換表を作成する。次に、ALT-J/Eの意味解析を行い、適合した結合価パターン番号から各節の用言意味属性(ALT版)を求め、最後に、初めに作成した意味属性対応表を使用して、ALT用言意味属性から本計画の意味属性コードに変換する。
節間意味分類	第1に、節間関係の意味分類に対応する文型パターンを作成する。第2に、作成したパターンに対して、文型パターン作成の元となった用例(日本文)を入力文とする照合実験を行い、照合に成功した文型パターンの番号を求め、最後に、求めた文型パターン番号に対応する意味分類コードを節間関係の意味分類コードとする。

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ため、それを利用して付与した。また、節の意味分類コードは、単語意味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は判断することはできない。そこで、従属節に対する益岡・田窪の分類が、節間意味分類体系の基本としていること、また、既に、用言変数の意味的制約条件の参考情報としてすべての文型パターンに「類語大辞典」(くろしお出版)の1000種類の意味属性、及び、ALTシステムの36種類の用言意味属性を付与していることに着目して、これらの情報を用いて半自動化する方法を検討した。その結果に基づき、節間意味コードは、既に

付与されている情報から益岡・田窪の分類に写像すること、益岡・田窪の分類から本研究で開発した節間意味コードへ細分類することによって付与した。この過程で、は人手作業であるが、はほぼ機械的に実施可能であるため、全体の作業工数は大幅に削減することができた。

(2) 意味類型パターンの例

文型パターンに対する統語的意味的分類コードの付与方法について図3.8.1に示す。また、統語的意味的分類コードを付与した意味類型パターンの例を図3.8.2に示す。例によって汎化可能なレベル(単語レベル、句レベル、節レベル)は異なるが、統語的意味的分類コードは、レベルに共通している。

<b>日本語</b>		<b>文種別</b> 1 <b>接続が一箇所</b>	<b>統語分類コード</b> <b>副詞型</b>	<b>副詞節 + 主節</b>
従属節	/\$_1\$/VP <sub>2</sub> <sup>^</sup> syushiと = 気が向くと			
	意味分類 3140(全身動作),6000(現象事象の表現),6930(適合不適合)			
主節	\$_1\$ ^{/tckN <sub>3</sub> は} #1{/tcfk<N <sub>4</sub> に>,!N <sub>5</sub> を} /cf(いれ入れ)‘てくれる’#6(.genzai .kako), = 彼女はお茶をいれてくれる			
	意味分類 3450(受諾拒否), 6110(所属包摂), 8120(入り)			
	<b>第一連用節</b> 意味分類コード	<b>FUc201</b>	<b>事実認識依存関係</b>	<b>第一節間キーワード</b> と
<b>英語</b>		<b>英語構文情報</b> <b>接続詞相当</b>	<b>英語構文キーワード</b> <b>when</b>	
パターン	N <sub>3</sub> 'serve' #6(^present ^past) <us N <sub>4</sub> ^obj> N <sub>5</sub> when <she N <sub>3</sub> > AJP(VP <sub>2</sub> )#6(^present ^past).			
例文	She serves us tea when she is in the mood.			

図 3.8.1 統語的意味的分類コードの例

【凡例】		
XXnnnnnn	原文標本文番号	PJnnnnnn 句レベル日本語文型パターン
LJnnnnnn	日本語標本文	PEnnnnnn 句レベル英語文型パターン
LEnnnnnn	英語標本文	CJnnnnnn 節レベル日本語文型パターン
WJnnnnnn	単語レベル日本語文型パターン	CJnnnnnn 節レベル英語文型パターン
WEnnnnnn	単語レベル英語文型パターン	N-X-... 統語的意味的分類コードなど

### 単語レベルのみの例

AC000150-00 3 100

LJ000032-00:父は新しく買ったワープロを愛用している。

LE000030-00:Father loves using his new word processor.

WJ000032-00:/y\$1^{/tcfkN1(NI:48,NI:79,IM:11211,IM:11220)は}/cf新しく/f買った/fN2(NI:971,IM:13680)を\$/cfV3 (N Y:2400,KR:0603,KR:5400,IY:3330).teiru,

WE000031-00:N1 V3 N1^pron^poss new N2.

3-名 / 4730,5914 / -2330 / MSa100 / -ワープロ- / 形容詞修飾

### 単語レベル+句レベルの例

AC000051-00 4 110

LJ000012-00:これは人々に愛唱されている古い民謡の一つです。

LE000012-00:This is one of the old folk songs people love to sing.

WJ000012-00:/y#1{/tcfkN1(NI:1,NI:1235,IM:10000,IM:15000)は,/tcfkN2(NI:2606,NI:5,IM:11100,IM:16770)に}/cf(V3(N Y:3102,KR:0603,KR:1907).reru.teiru^rentai|ND3をされている)/fAJ4(NY:0501,KR:8800,IY:A5B0)^rentai /fN5(NI:1056, NI:1673,IM:14230,IM:15243)の/kN6(NI:2588,NI:2595,IM:16720,IM:16730).#da,

WE000012-00:N1 be N6 of AJ4 N5 N2 V(V3|ND3).

PJ000008-00:/y#1{/tcfkN1(NI:1,NI:1235,IM:10000,IM:15000)は,/tcfkN2(NI:2606,NI:5,IM:11100,IM:16770)に}/cf(V3(N Y:3102,KR:0603,KR:1907).reru.teiru^rentai|ND3をされている)/fNP4(NI:2588,NI:2595,IM:16720, IM:16730).#da,

PE000008-00:N1 be NP4 N2 V(V3|ND3).

4-名名 / 0 / 95B0 / A000 / MSa200,MSa100 MSa200,MSa100 / 連体形 / 連体形

### 単語レベル+句レベルの例+節レベルの例

AC000166-00 1 111

LJ000035-00:この気候はわたしに合っているので老後はここで暮らしたい。

LE000033-00:The climate here suits me, so I would like to live here in old age.

WJ000035-00:/y\$1^{/tcfkN1(NI:2611,IM:16810)の/kN2(NI:2359,NI:2360,IM:15420,IM:15421)は}/tcfkN3(NI:10, NI:15, NI:37,NI:48,NI:49,NI:8,NI:9,IM:11110,IM:11111,IM:11112,IM:11120,IM:11160,IM:11211,IM:11212)に\$/cfV4(NY:0501, NY:0506,NY:0701,NY:0703,NY:1500,NY:2502,NY:3002,KR:9200,IY:5110,IY:6930,IY:6940).teiru^rentaiので</ycN5は >|老後は/cfADV6(KR:9900)/fV7(NY:0200,KR:0003,IY:3210,IY:3240,IY:6970).tai,

WE000034-00:N2 ADV(N1) V4 N3^obj, so <|N5> would like to V7^base ADV6 in old age.

PJ000021-00:/ytcfkNP1(NI:2359,NI:2360,IM:15420,IM:15421)は!VP2(NY:0501,NY:0506,NY:0701,NY:0703, NY:1500, NY:2502, NY:3002,KR:9200,IY:5110,IY:6930,IY:6940).teiru#6(.genzai|.kako)^rentaiので</ycN3は>!VP4(NY:0200,K R:0003,IY:3210,IY:3240,IY:6970).tai,

PE000020-00:NP1 VP2#6(^present|^past), so <|N3> would like to VP4^base.

CJ000009-00:/ytcfkCL1(NY:0501,NY:0506,NY:0701,NY:0703,NY:1500,NY:2502,NY:3002,IY:5110,IY:6930,IY:6940).te iru#5(.genzai|.kako)^rentaiので!CL3(NY:0200,IY:3210,IY:3240,IY:6970).tai,

CE000009-00:CL1, so CL3^past age.

1-副 / 4110,5930,5940-2210,2240,5970 / FUb100 / -ので- / 接続詞相当 / so

図3.8.2 重文複文の意味類型パターンの例

### 3.8.4 日英構文の対応関係の分析

前節では、パターン対を意味類型化するため日本語パターンに付与した統語的意味的分類コードについて述べたが、文型パターンについてのさまざまな実験的改良を支援するための情報として、英語表現についての構文的な分類も行い、各英語パターンにその分類コードを付与した。表3.8.8に、日本語構文に対応する英語構文の分類を示す。

表3.8.8 日本語構文に対応する英語構文の分類

#	第1分類	第2分類
1	等位節・従属節相当表現	等位接続詞相当、従属接続詞相当、接続副詞、節を導く副詞、分詞構文
2	名詞節・関係節相当表現	名詞節、直接引用節、同格節、疑問節、関係節、仮主語(IT)構文
3	副詞相当表現	、副詞小辞相前置き表現、慣用的句表現、不定詞、付帯状況、前置詞、複合前置詞相当、副詞相当表現当
4	格要素相当表現	名詞化、不定詞、不定詞(目的語)、動名詞、動名詞(目的語)、補語、原型不定詞、仮主語(IT)構文
5	修飾要素相当表現	前置詞句による修飾、所有格による修飾、不定詞(形容詞用法)修飾、形容詞修飾、分詞修飾、後置修飾、その他修飾
6	モダリティ相当表現	
7	埋め込み先名詞 + 述部の一体化	従属節 従属接続詞相当表現、従属節 名詞節、従属節 同格節、従属節 不定詞、従属節 原型不定詞、従属節 動名詞、従属節 前置き表現、従属節 慣用的句表現、従属節 前置詞、従属節 複合前置詞相当、従属節 副詞、従属節 形容詞、従属節 補語、従属節 名詞化、従属節 その他の要素
8	日本語主節の他要素変換	主節 疑問節、主節 副詞、主節 名詞修飾、主節 前置詞句、主節 主語、Vのが AJ N be AJ
8	埋め込み節 + 被埋め込み名詞	名詞化
10	被埋め込み名詞の変換	被埋め込み名詞の省略、被埋め込み名詞 主語
11	英語特有構文・表現	SVOO文型、SVOO文型(無生物主語構文)、SVOC文型、SVOC文型(無生物主語構文)、制止・禁止表現、前置詞と呼応した表現、無生物主語構文、因果・程度の構文、瞬時を表す構文、比較の構文、対比・選択・累加の構文、命令形を用いた構文、否定の表現、反対語を用いた表現、限定表現、その他慣用的表現
12	記号を用いた表現	2文相当、補足・挿入、並列、同格・格言、対比
13	その他	挨拶
14	意識・個別的表現	
計	14	83

これらの分類情報は、パターン記述方法の改良項目や汎化不良項目が発生したとき、修正対象となるパターン対を絞り込むときの重要な情報として役立ち、パターン改良のための作業コストを大幅に削減することができた。

また、これらの情報は、日本語構文と英語構文の意味的な対応関係を分析するデータとしても有用で、パターン対辞書の圧縮版の作成にも大いに役立つものであった。

そこで、構文分類から見た日英構文の意味的な対応関係の一部を表3.8.9に示す。この表は、日本語の重文複文の構文を178種類に分類し、それに対応する英語構文を83種類に分類したときの日英構文の対応関係を示している。分類した対象は、重文または複文（接続と埋め込みが組み合わさった重複文を除く）の合計9.7万パターンである。

この表から、対応関係を持つパターン対の数を分析した結果を表3.8.10に示す。

表3.8.9 日英構文の対応関係

構文対応関係		英語の構文分類									
		1	2	3	4	5	6	...	82	83	
		等位接続 詞相当	従属接続 詞相当	接続 副詞	節を導く 副詞	分詞 構文	名詞 節		挨拶	個別的 意識	
日本語の 構文分類	1	こと類	2	20			2	825			531
	2	選択疑問節 +トウコト						131			
	3	疑問後疑問 節+トウコト									
	4	/型	4	63	1		3	131			379
	5	強調構文	1	27				28			129
	6	トコロ型	4	32				1			144
	7	疑問語疑問 詞トウコト型									
	8	節+ 格助詞型						8			69
	...										
	177	選択	7								4
178	逆説的並列	237	23			4				3	

表3.8.9 日英構文の対応関係の分析

対応する 文型パターン数	対応関係を持 つ構文の組	日本語から見たとき		英語から見たとき	
		分類数	平均対応構文数	分類数	平均対応構文数
1以上の場合	2023組	178分類	11.4種	83分類	24.4種類
10以上の場合	658組	110分類	6.0種	75分類	8.8種類
100以上の場合	163組	52分類	3.1種	46分類	3.5種類
1000以上の場合	17組	9分類	1.9種	13分類	1.3種類

この表から、以下のことが観察される。

- (1) 日本語構文178種類が、英語構文83種類のすべてと対応関係を持てば、 $178 \times 83 = 14,774$ 組の対応関係となるが、実際に対応関係が1パターン以上存在するのは、2023組(13.7%)であり、残りの86%は対応関係は存在しない。
- (2) 対応関係が、10パターン以上、100パターン以上、1000パターン以上存在する構文の組は、それぞれ、4.5%、1.1%、0.1%と、急激に減少する。
- (3) 日本語構文から見ると、1構文に対応する英語構文は、それぞれ、平均110構文、52構文、9構文であり、英語パターンから見ると、1構文に対応する日本語構文は、それぞれ、75構文、46構文、13構文である。
- (4) また、一つの日本語構文に対応する英語構文の数、及び、一つの英語構文に対応する日本語構文の数を比較すると、後者の方が多い。

(1)と(2)から、日英構文の間には強い対応関係が存在するが、それは1対1の

関係ではなく、通常、1対複数の関係であることが分かる。また、(4)は、英語の表現に比べて日本語の表現は、多彩であることを示すが、これは、文型の発達した英語構文と語順の自由度の高い日本語の特徴を反映している。

本節で示した日英構文の対応関係は、パターン辞書の圧縮方法を検討する過程で分析したものであるが、その結果は、下例に示すように、学習レベルに応じた語学教育の教材作成にも役立つと期待される。

【翻訳技法の用例集作成の例】

<初級レベル> 日本語構文9種類と英語構文13分類の対応関係から、17組の対訳例文を取り出し編集する。

<中級レベル> 日本語構文52分類と英語構文46分類の対応関係から、163組の対訳例文を取り出し編集する。

<上級レベル> 日本語構文110種類と英語構文75分類の対応関係から、658組の対訳例文を取り出し編集する。

### 3.9 パターン検索プログラム

任意の日本語入力文に対して、文型パターン辞書からそれに適合する文型パターンを検索するためのプログラムは、文型パターン辞書を改良していく上でも必須のツールである。このようなツールとして、以下の2種類のプログラムを開発した。

- (1) パターンパーサ
- (2) 意味検索プログラム

#### 3.9.1 パターンパーサ

入力日本文に対して、表現パターン辞書から、該当するパターンをすべて同時に抽出するための文型パターン照合プログラム(図3.9.1)を試作した。高速化のためのさまざまな実験的改良を行った結果、12万入力文×12万文型パターン規模の照合実験(クロス照合実験)を20時間程度で実施できる処理速度を実現した。これにより、パターン改良のための繰り返し実験が容易となった。

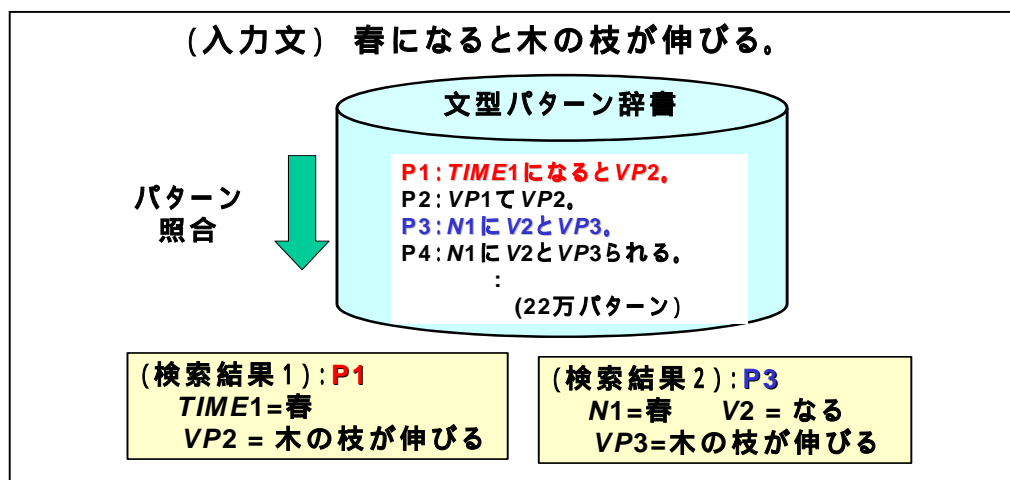


図3.9.1 パターンパーサの機能条件

(1) 基本アルゴリズム

照合の基本とするアルゴリズムは、構文解析アルゴリズム ATN(Augmented Transition Network)をベースとしている。全ての照合解を出力すること、および、1つの入力文に大量のパターンを適合検査することを考慮して、エージェントによる横型探索機構を設けた AB-ATN(Agent oriented Breadth-first ATN)を実現した。

文型パターンをATNで表現した例を図3.9.2に示す。

AB-ATN のアルゴリズムの概要は以下の通りである。

文型パターンを表すオートマトンの状態遷移により入力文と文型パターンの適合性を調べる。ノードは解析状態に対応しアークは文型パターンの記述要素に対応する。入力文を単語境界単位に参照し、アークに記述されている条件と照合する。適合すれば次ノードに状態遷移する。

アークから別のオートマトンを呼び出すことができる。その場合の遷移は、呼び出し先のオートマトンの照合に成功することがアークの条件となる。

エージェントはオートマトンの遷移経過を保持する。具体的には遷移経路およびアークの照合結果を保持する。非決定的な遷移をする場合にはエージェントは分裂する。

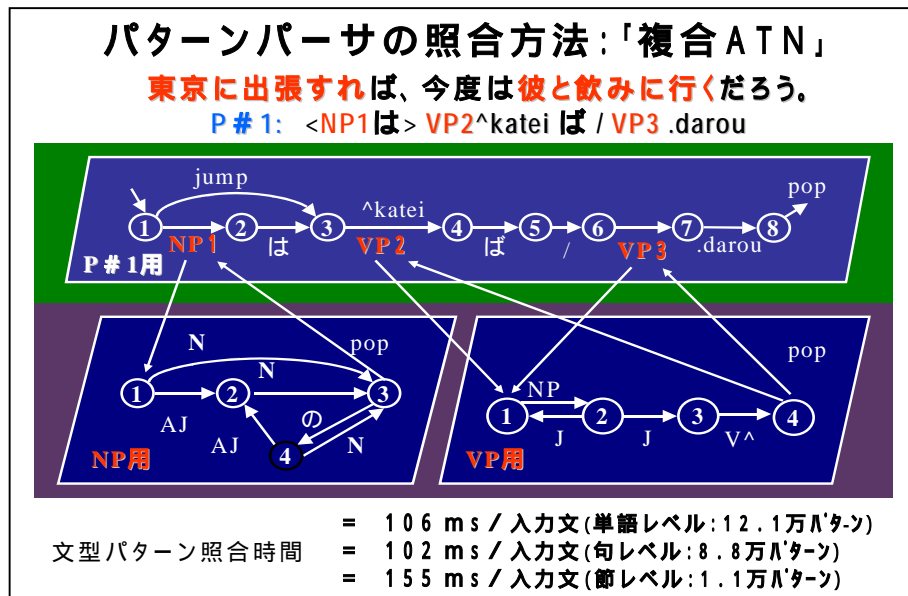


図3.9.2 照合アルゴリズム

全てのエージェントは、参照単語境界位置に同期してオートマトンを移動する。しかし、入力文の参照位置の先読みをした際、その数だけ休眠する。

複数のエージェントが同一のノードに存在する場合は、1つのエージェントに合体することができる。合体すると1つのエージェントがまとめて遷移経路および照合結果を保持し、残りのエージェントは消滅する。

同一文と複数のパターンを照合する際、参照単語境界位置とアークの条件の組に

より条件照合の結果はパターンによらず等しいことから、この組をキーに照合履歴を蓄え再利用することができる。

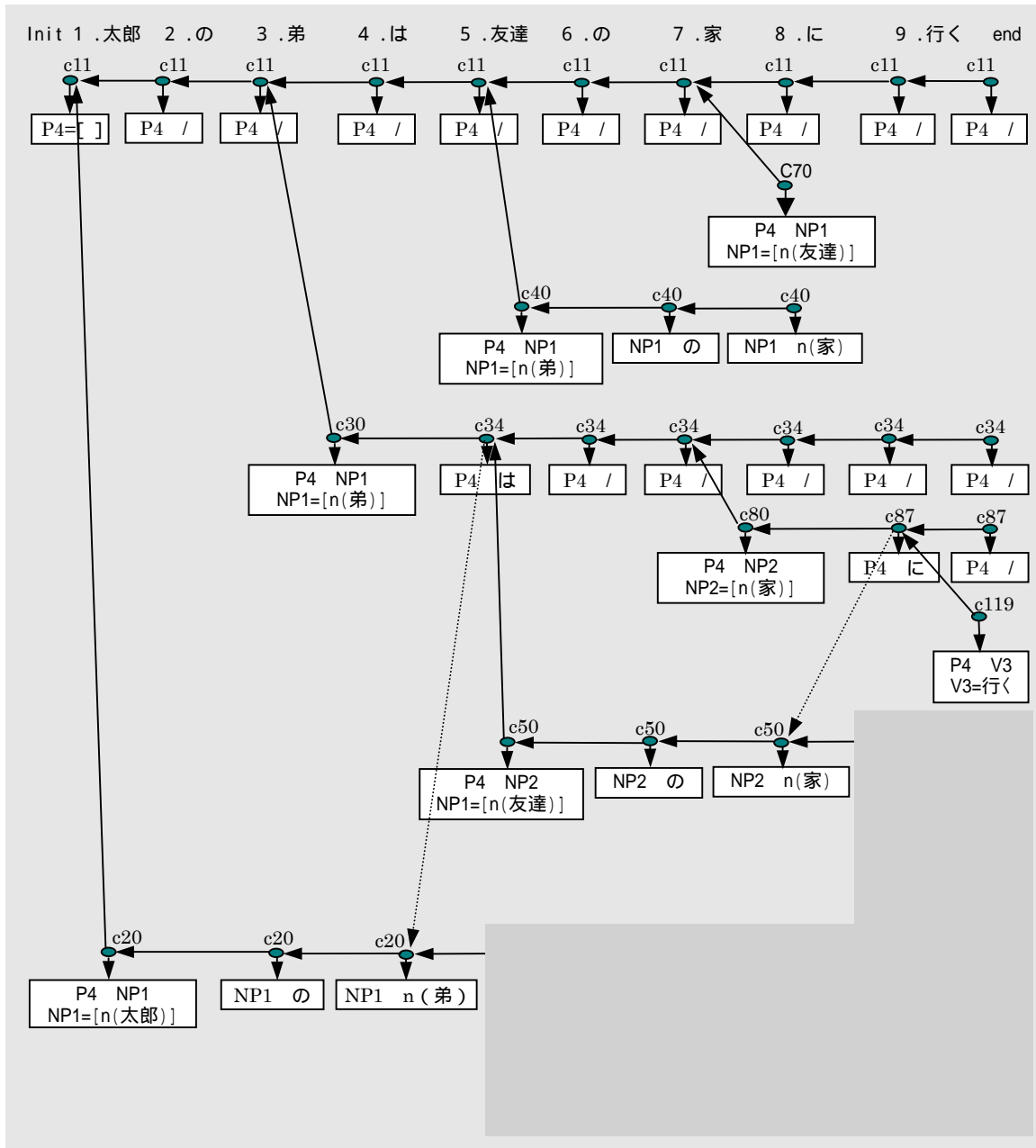


図3.9.3 照合結果格納の構造

(2) 照合結果の記憶

入力文とパターンの照合結果を記録する方法の例を図3.9.3に示す。

この図は、パターン「/NP1は/NP2に/V3」と文「太郎の弟は友達の家に行く」の照合結果を格納している様子を表す。横軸が処理時間の進む方向で、単語境界毎に並列に照合される。Cabはエージェント (aはエージェント番号, bはパターンのオートマトンのノードの位置) を表す。各時点でエージェントが照合処理を行ない蓄積しているデータへリンクしている。点線矢印はjoinにより追加されるリンク。影の部分はjoinにより省



略される部分を表す。

### (3) モジュールの構成

本プログラムは、「オートマトン生成系」、「入力文前処理系」、「絞込み系」、「照合本体系」で構成される。絞込み系と照合本体系は、連動して動作させるため、4つの系は、3つの実行コマンドから起動することになる。図3.9.4にモジュール構成図を示す。

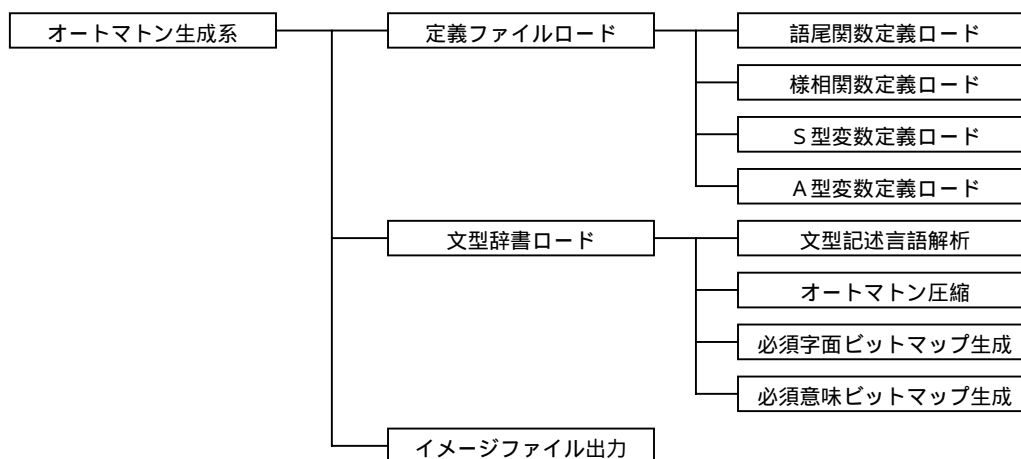


図 3.9.4 モジュール構成(1)

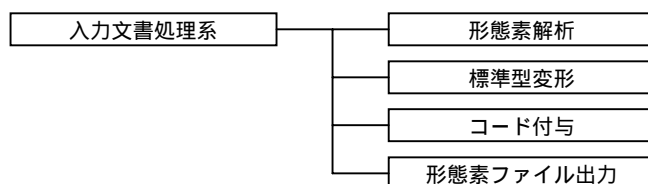


図 3.9.4 モジュール構成(2)

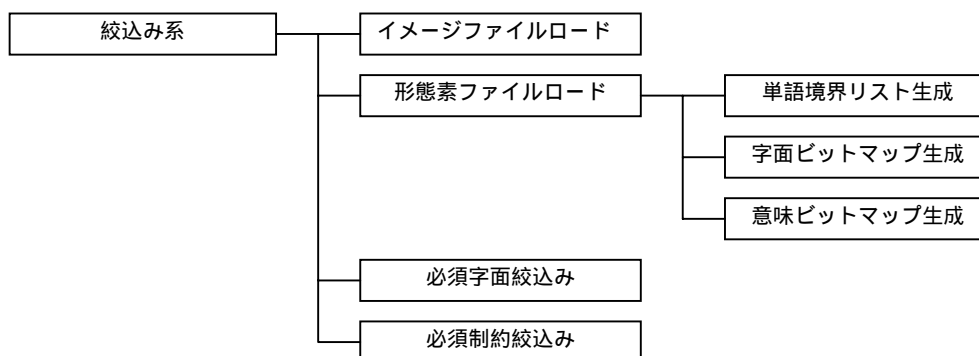


図 3.9.4 モジュール構成(3)

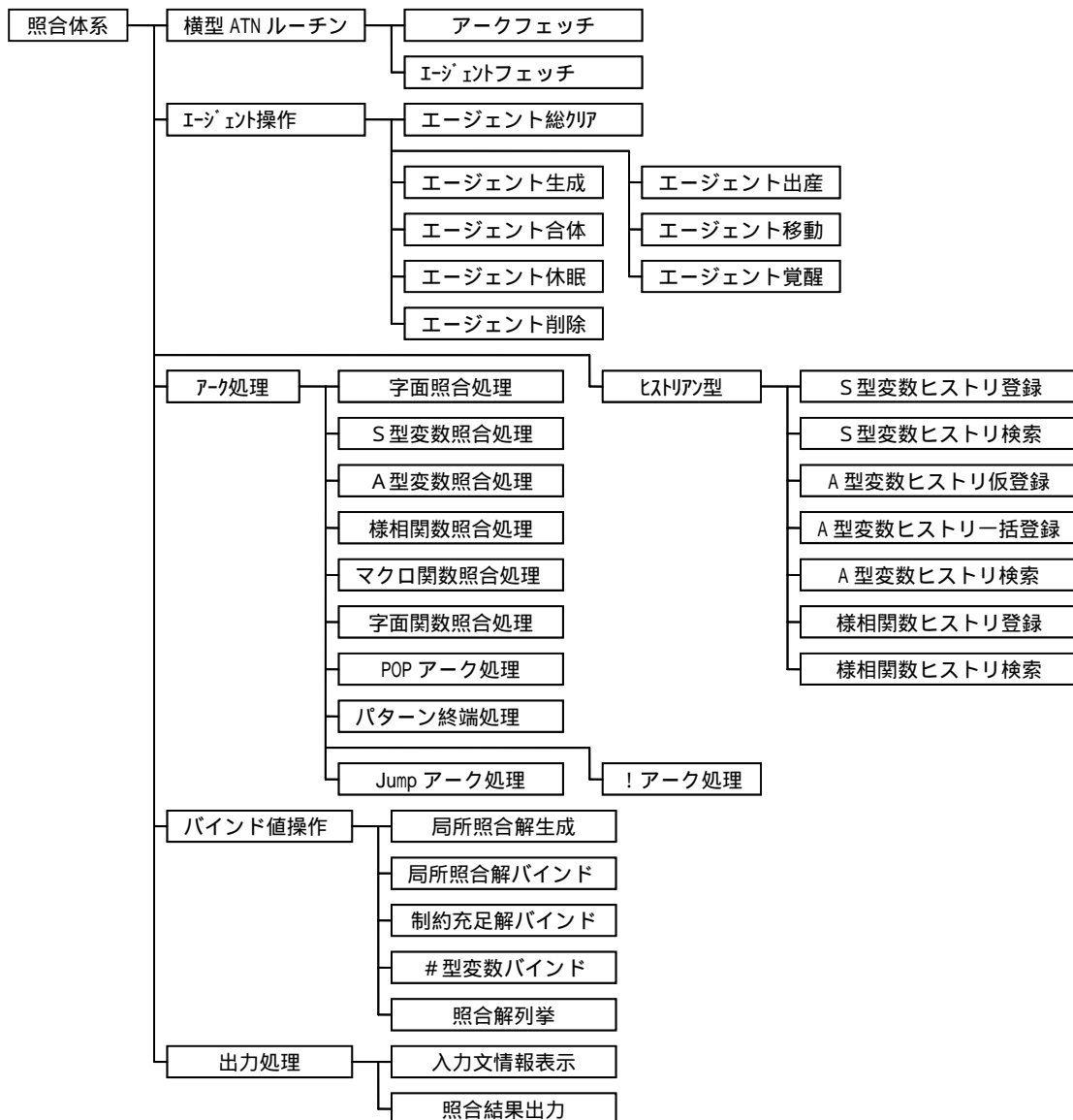


図 3.9.4 モジュール構成(4)

### (3) 高速化の効果

入力文とパターン間での字面チェック、照合過程でのエージェントの合体 (Join) 機能、及び、照合履歴の再利用の3種類の高速度化の効果を図3.9.5に示す。この図に示すようにパターンパーサの照合速度は大幅に向上し、クロス照合実験が可能となった。

< 検索時間の短縮 > 検索可能な入力文でのみの評価

- ・ 絞り込み前処理      6分の1以下
- ・ Join機能              9分の1以下
- ・ 履歴機能              38分の1以下

**複合的な効果 = 2,000倍以上の高速化**

- ・ クロス照合実験(注)…… 20~30時間
- ・ 1文 × 22.7万パターン…… 1秒以下

(注)12万文 × 22.7万パターンの総当たりの実験

いずれの高速化が無くてもクロス照合実験は不能

図3.9.5 高速化の効果

## (6) 検索例

パターンパーサによる文型パターン検索の例を図3.9.6に示す。

この画面は、単語レベルの文型パターン辞書(122,719パターン)では、入力日本語に適合する文型パターンに延べ21件(同一のパターンでも適合の仕方が異なるものは1件にカウント)適合することを示しており、上位の数件が表示されている。



図3.9.6 パターンパーサの検索画面

### 3.9.2 意味検索プログラム

「意味検索プログラム」は、意味累計パターン辞書を対象に、それに付与されている「統語分類コード」、「意味分類コード」、「節間KW(キーワード)」を組み合わせる意味検索を行うためのプログラムである。

(1) 機能と構成

本プログラムの構成を図3.9.7に示す。図で示すように、このプログラムでは検索すべき文型パターンの内容を指定する方法として、以下の2種類の方法を実現している。

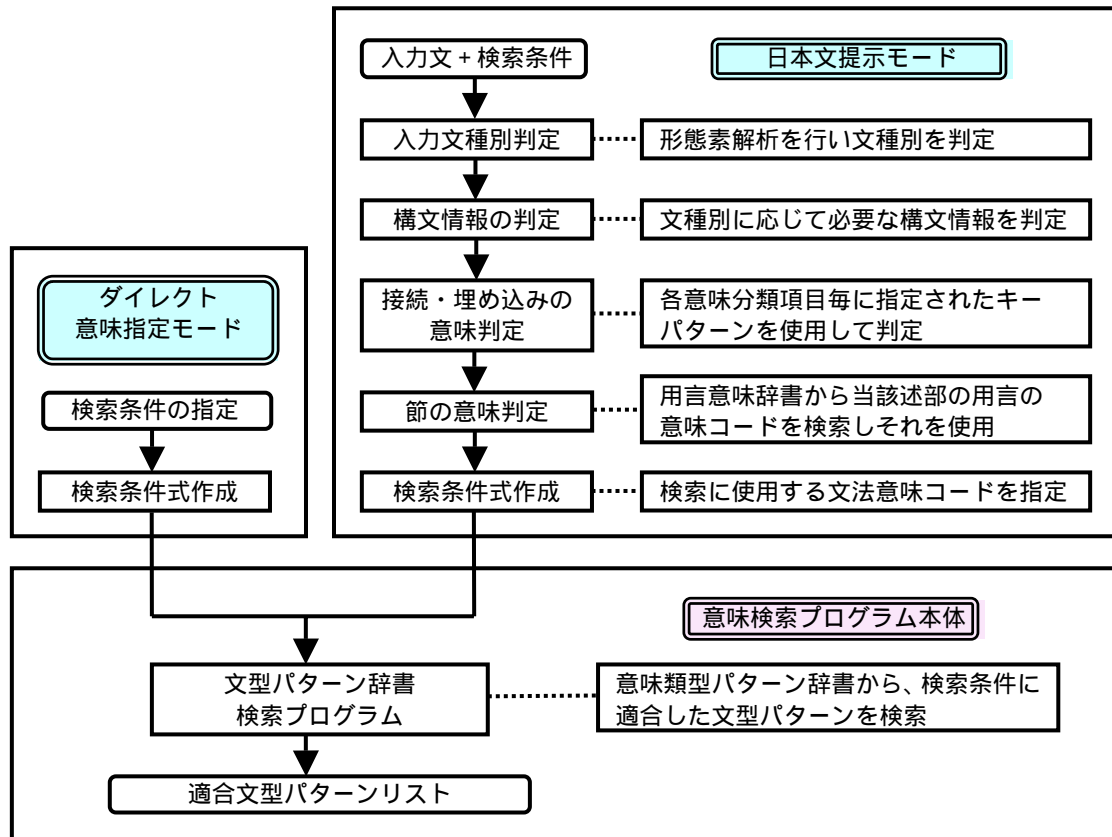


図3.9.7 文型パターン意味検索プログラムの構成

<ダイレクト指示モード>

この方式は、10種類の「文法意味分類コード」に対して検索キーとなる「分類名」もしくはその「コード」を指定する方法で、意味検索プログラムは、指定された条件のすべてのに適合する文型パターンを検索する。

<日本文提示モード>

この方式は、日本文と検索条件を入力とする方法である。検索条件として、10種類の「文法意味分類コード」のうちのいずれを使うか(1つ以上)を指定する。「意味検索プログラム」は、与えられた条件下で日本文と文法的、意味的に同じ文型パターンを検索する。

(2) 意味検索の実行例

意味検索プログラムの実行例を図3.9.8、図3.9.9に示す。図3.9.8は、検索用の入力文の解析結果に基づき、自動的に意味的な検索条件が設定された状態を示す。この検索条件は自由に変更できる。また、図3.9.9は、図3.9.8指定された検索条件に基づく検索結果を示す。

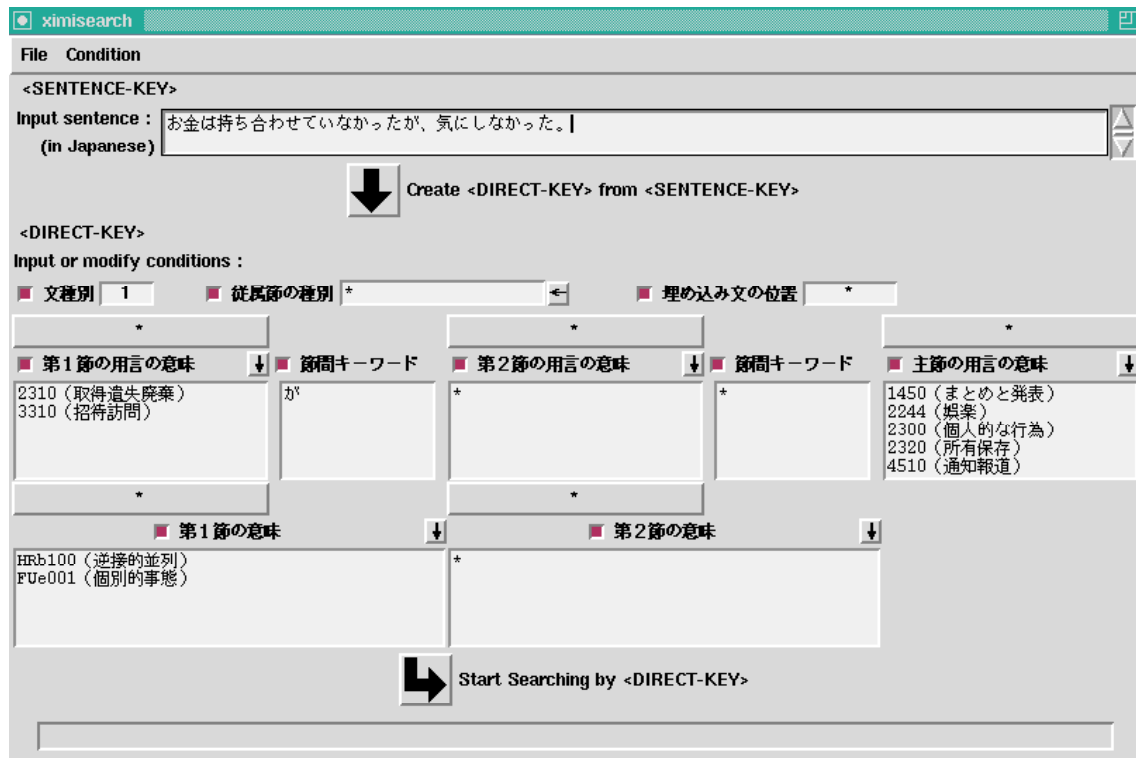


図3.9.8 意味検索プログラムの検索条件指定画面

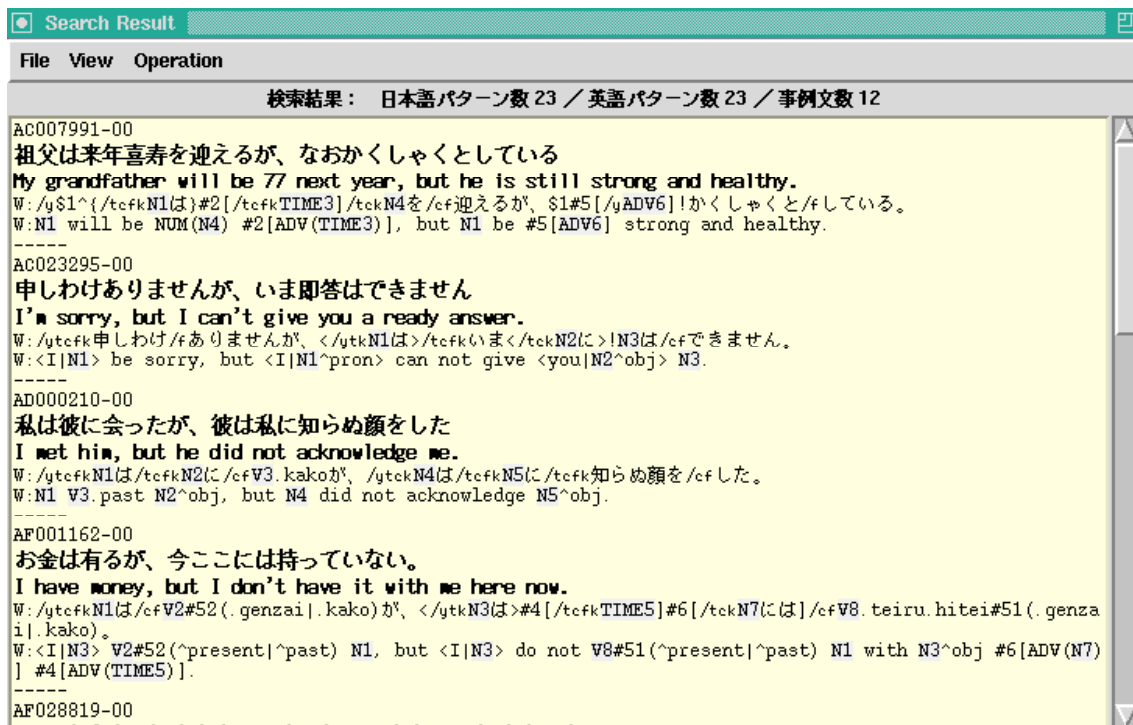


図3.9.9 意味検索プログラムの検索結果出力画面

### (3) 本プログラムの利用方法

#### ・人手翻訳支援への応用

このプログラムは、人手翻訳の支援にも使用できると期待される。すなわち、文型パターン照合プログラムによる文型パターン検索では、構成条件を厳密に満足する文型パターンのみが検索されるから、文型パターン辞書の中に入力文の翻訳に使用できる文型パターンや参考になる文型パターンが存在しても、それらが必ず抽出されるとは限らない。そのような場合のこのプログラムを使用すれば、入力文と意味的に類似の文型パターンが検索できる。

#### ・文型パターンパーサと併用した候補絞込への応用

上記の応用から簡単に類推できることであるが、文型パターンパーサは、変数の意味的な制約条件もチェックするものの、現実の入力文に対して、意味的に不適切な文型パターンを含む複数の適合文型が抽出される場合が多い。そこで、抽出された文型パターンに対するフィルターとしてこのプログラムを使用すれば、意味的に不適切な文型パターンを削除できることが期待される。

### 3.10 翻訳簡易実験システムの試作

意味類型パターン辞書の実用的な効果を確認するため、日英機械翻訳のための簡易実験システムを試作を開始した。このシステムは、変数、関数などの部分に対する英語を生成するもので、文型パターン辞書を使用することによって生成される英文の骨格を確認するためのものである。まだ部分的にしか完成していないため、総合的な評価結果は得られていないので、以下、本試作で検討中の概要を以下に示す。

#### ・文型パターン照合条件の最適化の検討

文型パターン辞書を機械翻訳に使用するには、文型パターン辞書から入力文に意味的に適合する文型パターンを抽出する必要がある。入力文に対して、文型パターン辞書から、なるべく意味的に適した文型パターンに絞って検索するための条件を検討中である。

#### ・最尤文型選択方式の検討

一般に入力文に適合する文型パターンとしてパターンパーサによって抽出される文型パターンは複数であるが、この中には、入力文に対して意味的に不適切な文型パターンの含まれるのが普通である。そこで、抽出された文型パターンの中から意味的に適切な文型パターンを選択する方法について検討した。

これによれば、多変量解析を使用した文型パターン選択方式が有望であることが判明したので、今後、それを改良し文型パターン選択プログラムを試作する。また、意味による文型パターン検索プログラムができたので、これによって、どれだけ入力文の意味に近い文型が検索できるかを明確化する。

#### ・変数にバインドされた表現の翻訳方法の検討

翻訳簡易実験システムは、変数にバインドされた表現の翻訳結果を作成して、該当する英語文型パターンの変数に代入するプログラムである。名詞句変数翻訳プログラム、動詞句変数翻訳プログラム等を作成した。

・文型変数に対する変換プログラムの試作

上記の結果に基づき、得られた文型パターンを使用して英文を生成するための簡易な実験プログラムを作成した。単語変数の値の変換では、日英対訳辞書を使用するが、句変数と節変数の値の変換では、日英機械翻訳プログラムALT-J/Eの機能を流用している。

・大域文型制約を用いた英文生成部の検討

日本語入力文に適合した日英文型パターン（複文、重文）から定まる英語文の大域的構造中に具体的な節、句、単語を埋め込み、判断／態／時制／相／様相のような主体表現的要素を融合して、必要な形態素調整などを行い、統語的に正しい英語文を生成する枠組を検討した。処理部を試作中である。

・任意要素を考慮した日英構造変換部の試作

構造化された日英文型パターン（複文、重文用）を用いて、構造化された入力文のうち日本語文型パターンに適合しない任意要素に対する翻訳を可能とするため、抽出された日本語任意要素に対応する英語任意要素の位置を推定する方法を検討した。処理部を試作中である。

また、複雑で長い入力文に対し、複数の構造化された日英文型パターン（複文、重文用）を適用し、入力文全体をそのような文型パターンと任意要素の組合せでカバーする方法を検討した。

3.11 研究開発年表

(1) 線表

当初は、年を追って汎化のレベルをあげる予定であったが、計画の3倍以上の標本文を入手することができたため、同時並行的に汎化を実施することとした。そのため、実施の線表は、当初計画とかなり異なるものとなった。表3.11.1に実施線表を示す。

表3.11.1 線表

年度	H13年度	H14年度	H15年度	H16年度	H17年度	H18年度	
意味類型知識ベース試作	日英対訳コーパス(45万件)	例文追加収集 追加例文(55万件) 作成(1万件)					
		標本文抽出(15万件)	被覆率実験 (基本文型)	文型パターン改良		改良	
		単語、句レベル汎化(22万パターン対)	節レベル汎化と記述法見直し	真理項体系化	意味類型化	意味類型化(2)	意味類型KB編集(インデックス作成等)
		意味類型化方式机上検討	意味類型化実験検討	意味類型化(部分試作)	改良		成果まとめ
日英変換システムの試作と実験		文型照合プログラム試作	文型照合デモプログラム試作	被覆率実験	実験システム試作	文型パターン翻訳実験	
			日本文解析モジュール試作	英文生成モジュール試作	システム設計	実験システム試作	意味類型翻訳実験
日本文生成		準備		日本文生成方式検討		まとめ	

(2) 主な検討課題と検討結果

各年度に取り組んだ主な検討課題とその結果の扱いについて表3.11.2に示す。

表3.11.2 主な検討課題と検討結果

年度	#	検討課題	検討の内容	その後の扱い
平成 13 年度	1	日英文型パターンの意味的対応方式の検討	比較構文に対するパターン化とその意味的な分類方法について先行的な検討	本計画の具体化で使用
	2	日英因果関係文型のパターン化事例検討	因果関係構文の文型パターン化について例題検討を行った。	文型パターン記述方式の検討に反映
	3	類推思考の原理に基づく機械翻訳方式	意味的等価変換方式を機械翻訳に適用する方法と新しい機械翻訳方式概念の検討	計画全体の具体化で使用
	4	日英対訳データ収集方法の検討	パターン化に適した対訳例文の収集方法を検討し、出版物の中から、約30種類のドキュメントを選択した	既所有の20万件に加え合計100万件の対訳データをファイル化
	5	重文複文の分類法と分類基準の詳細化	形態素解析結果を使用して、重文複文を選択抽出し、5種類の文種別の分類基準と判定方法を検討	文種別分類の半自動化を実現
	6	文型パターン照合アルゴリズムの検討	文型パターン辞書から入力文に適合する文型パターンをもれなく抽出するための基本アルゴリズムを検討	文法レベル文型照合プログラムの設計に継承
平成 14 年度	7	対象認識の概念化と機械翻訳方式の検討	概念と言語表現の関係に関する言語哲学的な問題点について検討した	「非線形言語モデル」の基礎理論となった
	8	意味類化と文型パターン化の検討	因果関係構文を対象に意味類型化の問題点を検討	意味類型分類体系の検討に継承
	9	形態素解析結果修正方法の検討	日本文の形態素解析の誤り特性を分析し、誤り箇所の発見と修正方法について検討	15万件の日本文の形態素解析誤りを修正
	10	言語表現の線形性と非線形性の定義と判定基準の検討	論理式の線形性の概念を参考に、言語表現の線形非線形要素と表現全体の線形非線形の定義と実際の判定方法を明確化	文型パターン化で汎化項目を選定する上の基本原理として使用
	11	文型パターン記述方式の事例検討	因果関係構文を対象に、文型パターン化の記述実験を行い、各汎化項目の適用頻度と汎化の効果について調べた	3段階の汎化レベルと各レベルでの汎化項目を決定
	12	文型パターン作成基準の検討	文型パターンの記述レベルの汎化項目の関係と各汎化項目の適用条件を明確化	文法レベルの文型パターンを作成
	13	日英文型パターン記述言語の設計	文型パターン記述言語(文法レベル版)を検討した	文型パターン記述言語仕様を作成
	14	半自動的な文型パターン生成方法の検討	修正された形態素解析結果を使用して文型パターンを半自動的に作成する方法の検討	文法レベル文型パターンを作成
	15	文型照合方式の検討	入力文と文法レベル文型パターンの照合方法を検討し、文型照合プログラムを設計した	文型照合プログラムを試作
	16	文型パターン辞書評価方法の検討	文型パターン辞書の被覆率を評価するための評価パラメータとその評価方法を検討	その後の評価の基本パラメータとして改良し、使用
平成 15 年度	17	接続表現意味体系化の検討	接続表現を例に文型パターンの意味分類をする方法について検討した	接続の意味分類体系表作成で使用
	18	文法レベル文型パターンの評価	文法レベル文型パターンの被覆率特性を評価分析し、文法レベル文型パターンの限界を明らかにした	意味レベル文型パターンの開発に取り組むことを決定



	19	意味レベル文型パターン記述方式の検討	変数に対する意味的制約条件の付与基準を決めた。また、2種類の要素順序記号の導入、離散記号へのレベルの導入と付与基準などを決めた	この基準に基づいて意味レベル文型パターン(初版)を作成
	20	文型パターンパーサのアルゴリズム設計	上記の仕様に基づいて、文型照合アルゴリズムを再設計し、文型パターンパーサの基本仕様書を策定した	この仕様書に基づき、業者によってJPPを試作
	21	文型パターンの文法的意味的分類法	文型パターンを文法的属性(3種類)、意味属性(5種類)、KW(2種類)に着目して分類する方法を検討した。	意味類型化の具体化に継承
	22	意味類型化のための概念体系化の検討	意味類型化に必要な論理意味範疇を体系化する方法について検討した	重文の意味分類体系の構築に継承
	23	パターン選択方式の検討	変数意味属性による方法、種々の特性を組み合わせた多変量解析の方法などを検討	多変量解析が最も有望 今後の継続検討とした
平成 16 年度	24	翻訳実験システムの検討	JPPによって抽出された文型パターン対の英語文型パターンを使用して翻訳実験を行うためのプロトタイプシステム構成法を検討	日英翻訳実験システムの試作に継承。
	25	文型パターン最適化手法の検討	変数の意味的制約条件を含むさまざまな制約条件の元に文型パターンの照合実験を行い最適な制約条件を検討。その結果、現状では単語レベルの再現率が低く用言の意味的制約の効果が小さいことが分かった	最適化は一旦中断し、新しい用言意味属性の開発と文型パターンの改良(時制様相の汎化など)に取り組むこととした
	26	様相時制情報の汎化方法に関する検討	文型パターンの様相、相、時制の情報について汎化の可能性と効果を検討し、汎化を行うための新しい文型記述方式を設計	この結果に基づいて述部語尾の汎化を実施
	27	文型間の意味的リンク発見アルゴリズム検討	日英文型パターン間で1対nの意味的な対応関係を半自動的に発見する方法として、3種類の方法を検討した	実施コスト高の見込みで実施は見合わせた
	28	接続の意味分類体系化の検討	従来の重文の意味分類体系を見直して従属節意味分類体系を作成し、その体系に基づいて文型に分類コードを自動付与する方法を検討した	複文重文の意味分類コードとして文型パターンに付与。意味類型化に使用
	29	単文の意味分類体系化の検討	日本語用言(動詞、形容詞、形容動詞)の新しい意味分類体系(400分類)を作成し、用言意味辞書を編集した	単文の意味分類コードとして文型パターンに付与。意味類型化でも使用
	30	名詞の意味属性体系の検討	名詞に対する新しい意味属性体系を構築し、名詞意味辞書を編集した	名詞変数の制約条件と「だ文」意味分類に適用
	31	英語文型パターン記述方式の再検討	文型パターン評価実験に基づき、英語文型パターンの記述方法の問題点と改良方法について検討した	検討結果に基づき、文型パターンの改良作業を実施
	32	文型パターン縮退方法の検討	文型パターンを意味的に縮退する方法について検討し縮退化用ツールを設計した	プログラム化しパターン辞書を縮退効果を調査
	33	文型汎化のための局所言い換え技術	日本語言い換え技術を用いたときのパターン被覆率向上効果を調べた	選択記号の充実に反映言い換え辞書化は中断
	34	用言意味属性二次元化の検討	用言自身が持つ相、様相情報の扱いと用言意味属性体系を二次元化の方法を検討	作業コストが取れないため、実施は見合わせた
	35	変数意味制約条件の絞り込み方法の検討	変数に付与された意味属性を半自動的に最適化する方法について検討した	作業コストが取れないため、一部のみ実施
	36	パターン汎化効果評価法の検討	記号、関数を使用したさまざまな汎化の効果を評価する方法を検討した	実際の評価に使用
平成 17	37	離散記号適正化の検討	離散記号と関数、変数の相互関係について検討した	文型パターンの改良作業に適用

年度	38	英語構文分類体系の検討	英語例文の特徴に着目して、英語パターンの構文を約 50 種類に分類し、分類コードを半自動的に付与する方法を検討した	対訳標本文の構文分類に使用。英語パターンの改良で効果が大
	39	節変数化拡張方式の検討	節変数の適用範囲を相、様相を含む範囲に拡張する方法について検討した	改良に適用。節レベルパターン数が大幅に増加
	40	英語パターン記述方式の問題検討	従来パターン記述言語は英語パターン記述能力があることが分かったため、改良方法を検討した	英語パターンの改良に適用
	41	自己例文不適合パターンの分析	自己例文に不適合なパターンの問題点を検討	パターンパーサとパターン不良分析で使用
	42	英語群動詞表記法の検討	英語群動詞のパターン記述法が大きな問題となってきたため、その方法を検討した	パターンとパーサの改良に適用
	43	文型パターン意味検索方式の検討	文型パターンを意味コード(真理項)によって検索する方法について検討し、意味検索プログラムを設計した	意味検索プログラムを試作し、実験に使用している
	44	各種不良パターン発見方法の検討	文種別から見たパターンの適合特性、離散記号の有無に着目した適合特性、変数適合要素の分析など、種々の不良を検討	パターンの改良作業に適用
	45	語形指定関数付与基準の検討	語形指定関数を変数以外に適用する方法について検討した	パターンの改良作業に適用
平成18年度	46	不適合試験文の解析方法の検討	不適合試験文を対象に、適合しない理由を推定分析する方法を検討した	パターンの改良作業に適用
	47	単文適合パターンの分析法の検討	重文複文のパターンは単文に適合しないはずであるが、適合する場合がかなりあるのでその理由を検討した	パターンの改良作業に適用
	48	高頻度適合パターンの縮退方法の検討	被覆率を下げないことを条件に、汎用性が高すぎるパターンを削除する方法を検討	パターンの改良作業に適用
	49	文型パターン辞書圧縮方式の検討	クロス照合実験、パターン間の包含関係、使用頻度特性などの情報を使用して、被覆率を下げないことを条件に辞書を圧縮する方法を検討した	文型パターン辞書の圧縮版作成に適用

### (3) 途中生産物とそれらの相互関係

本研究で作成した途中生産物の相互関係を図3.11.1に示す。また、それらの内容を表3.11-3に示す。生産物は、日英対訳コーパス、重文複文の対訳例文ファイル、文型パターンファイル、意味類型パターンファイル、パターン検索プログラムに関するものに分類される。このうち、途中生産物は意味類型パターンファイルに関するものが多くなっているが、これは、意味類型化の試行錯誤のためではなく、パターン辞書の被覆率向上のための実験的検討を重ねたことによる。

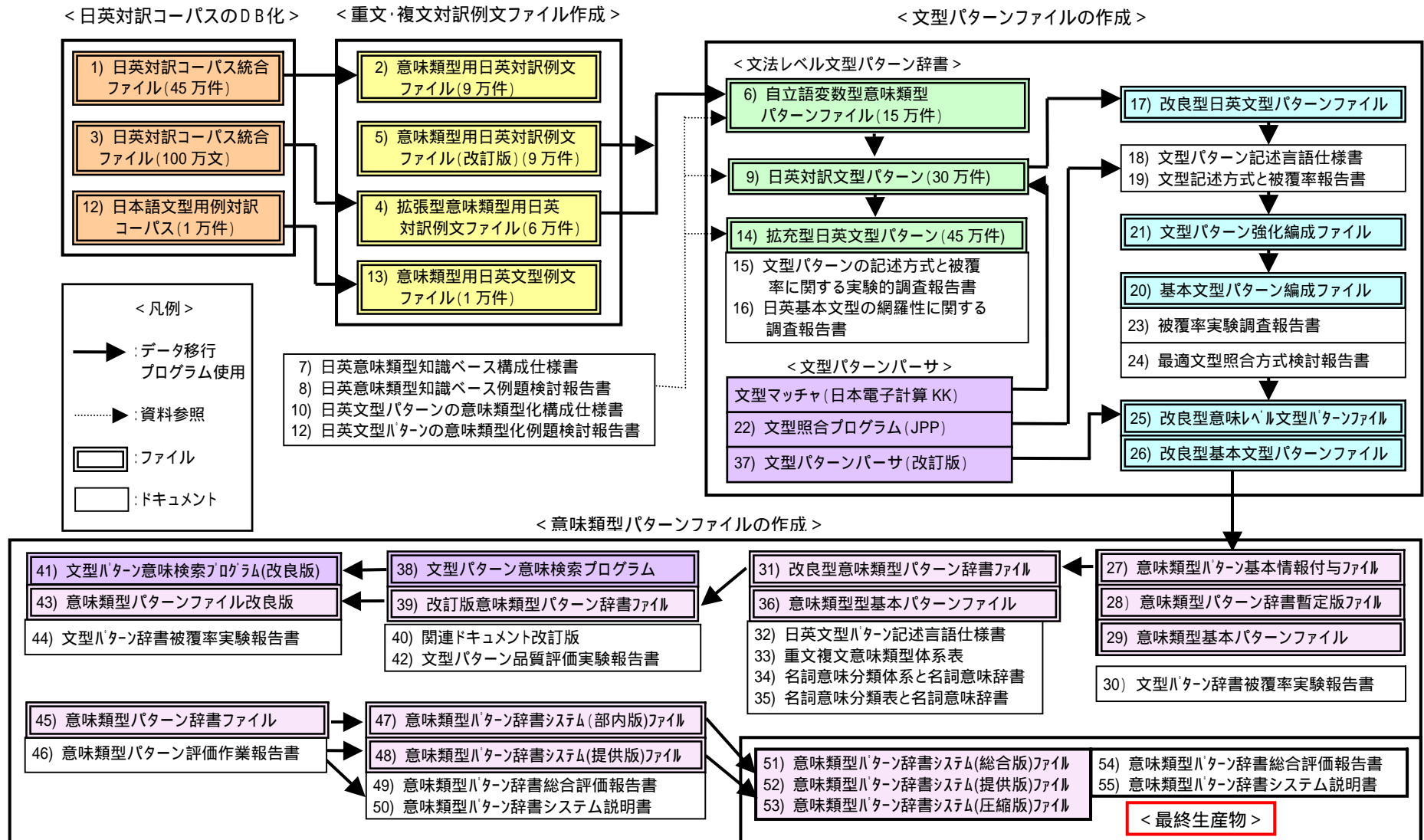


図 3.11.1 途中生産物の相互関係

表 3.11.3 途中生産物

時期	項目名	生産物	内容
平成 13 年度	日英対訳例 文ファイル	1)日英対訳コーパス統合 ファイル	24 種類のキュメントから日英対訳文を抽出し文字コード を統一した 45 万文の対訳コーパス。
		2)意味類型用日英対訳 例文ファイル	上記ファイルの中から5種類の重文・複文9万件を選別 して抽出した例文ファイル。
平成 14 年度 前期	日英対訳コ ーパス編成 ファイル	3)日英対訳コーパス統合 ファイル	日英対訳データ(約 55 万文)の文字コードをチェックし、 1)のファイルと統合したもので、100 万件の対訳例文を 格納。
		4)拡張型意味類型用日英 対訳例文ファイル	3)のファイルへの追加分 55 万文から、5種類の重文・複 文 6 万件を抽出したものを。
		5)意味類型用日英対訳 例文ファイル(改訂版)	2)のファイルに収録された日英対訳例文(9万文)を対 象に、不適切な英文を持つ例文レコードを削除
		6)自立語変数型意味類型 パターンファイル	4)と 5)のファイルを統合した 15 万件の日英対訳例文フ ァイルに対して、自立語等を変数化し縮退させたもの
		7)日英意味類型知識ベ ース構成仕様書	因果関係の意味を表現する重文複文約 500 件を対象 に意味類型化の記述実験を行った結果の報告書
		8)日英意味類型知識ベ ース 例題検討報告書	7)の作成に使用した例題、検討データ、分析結果など の途中結果を資料として整理したもの
平成 14 年度 後期	日英対訳文 型パターン 編成ファイル	9)日英対訳文型パターン ファイル	6)のファイルの対訳例文(15 万件)を単語レベル、句レ ベルで変数化、任意化 (15 万 × 8 = 120 万レコード収録)
		10)日英文型パターンの意味 類型化構成仕様書	因果関係の意味の日英対訳文型パターン 500 件を対象 に、意味類型化手法の記述実験結果をまとめたもの。
		11)日英文型パターン意味類 型化例題検討報告書	7)の作成に使用した例題、検討データ、分析結果など の途中結果を資料として整理したもの
	日本語文型 用例対訳編 成ファイル	12)日本語文型用例対訳 コーパス	日本語文型辞典から例文 1 万文余りを抽出し、翻訳家 による英訳文を付与。
		13)意味類型用日英文型 例文ファイル	12)のファイルから重文・複文を検査し抽出したもの (1 万文弱)。
平成 15 年度 前期	拡充型日英 文型パター ン編成ファ イル	14)拡充型日英文型 パターンファイル	15 万対訳文を対象に、節レベルの文型パターン化した データを追加すると共に、新しいパターン記述言語仕様 に従って書き換えたもの。
		15)文型記述方式と被覆率 報告書	14)のファイルを対象に照合実験を行いパターン記述方 式と被覆率の関係を調査。
		16)日英基本文型の網羅性 に関する調査報告書	13)のファイルの文を使用して、文型パターンの照合実 験を行い、「被覆特性」、「日本語内書き換え能力」、「記 述方式の改良項目」をまとめたもの。
平成 15 年度 後期	改良型日英 文型パター ン編成ファ イル	17)改良型日英文型 パターンファイル	14)のファイルに収録された 23 万件の文型パターン対の すべての変数(合計 100 万カ所)に意味的制約条件を 付与すると共に、「文型要素順序可変記号」の付与、不 適格文型パターンの意味的分離作業、文型パターンの 部分的適正化作業を行ったもの。
		18)文型パターン記述言語 仕様書	14)の文型パターンの記述方法を定めたもの。
		19)文型記述方式と被覆率 正解率の関係に関する 調査実験報告書	14)で改良した文型パターンの能力について、改良項目 毎の効果を実験的に評価すると共に、総合的な改善効 果を調べたもの。
	意味レベル 文型パター ン編成ファ イル	20)基本文型パターン編成 ファイル	17)のファイルから使用頻度の高い 2,000 文型を選択し、 ブラッシュアップ作業、類似文型の縮退と制約条件 の見直し修正作業、被覆率調査を行ったもの。
		21)文型パターン強化編成 ファイル	17)の文型パターンファイルに対して、表記と表現の ゆらぎを吸収、適用範囲の拡大と新変数を導入

	文型照合プログラム試作と被覆率の実験的調査	22)文型照合プログラム 23)被覆率実験調査報告書	3レベルの文型パターン辞書から、入力文に適合する文型パターンをすべてを発見し抽出するプログラム。 文型パターン辞書を対象に 22)のプログラムを適用して被覆率実験を行いその結果をまとめたもの。
平成16年度前期	意味類型パターン編成準備ファイル	24)最適文型照合方式検討報告書 25)改良型意味レベル文型パターンファイル 26)改良型基本文型パターンファイル 27)意味類型パターン基本情報付与ファイル	意味レベル文型パターンに様々な照合条件を設定、再現率と意味的適合率を共に最大とする条件を明確化 24)の分析で得られた各種問題点の中からコスト性能比の大きい項目を取り出し、22)の文型パターンを改良 2000件の意味レベル文型パターンファイルに対して、改良、類似文型の縮退と制約条件の見直し修正、意味による体系化、被覆率調査を行ったもの。 21)文型パターン 22.1万件に対して、10種類の統計的意味的分類コードを付与したもの。
平成16年度後期	意味類型パターン編成暫定版ファイル作成作業	28)意味類型パターン辞書暫定版ファイル	文型パターン辞書に「述部語尾表現の汎化」、「英語表現表記法の改良」、「用言変数意味属性付与」の3種類の作業を行い、「節の意味分類コード」を付与。
		29)意味類型基本パターンファイル	重文複文の意味分類(150分類)項目毎に基本的な文型パターン(各20件)を改良したもの。
		30)文型パターン辞書被覆率実験報告書	28)の文型パターン辞書の意味的な被覆率を総合的に評価した結果
	意味類型パターンファイルシステム作成作業	31)改良型意味類型パターン辞書ファイル	28)を対象に、節間意味分類コード付与、単文意味分類コード付与、離散記号付与基準の見直し、個別的改良
		32)日英文型パターン記述言語仕様書(改訂版)	31)のファイル記述に使用した最新の言語仕様書
		33)重複文意味分類体系表	深さ5段で約150種類からなる重文複文の意味分類体系表
		34)単文意味分類体系表と用言意味辞書	深さ5段で約400種類からなる単文の意味分類体系と約6000語の用言に対する意味コード辞書
		35)名詞意味属性表と名詞意味辞書	深さ5段で約400種類からなる名詞の意味属性体系と約1.6万語の名詞に対する意味コード辞書
	36)意味類型基本文型パターンファイル	29)を対象に各種の改良作業を行い、その効果を「文型パターンパーサー(改訂版)」を使用して評価	
	37)文型パターンパーサー(改訂版)	「文型パターンパーサー(jpp1)」に様相・時制などの述部語尾表現処理を組み込んだもの	
	38)文型パターン意味検索プログラム	「統語分類コード」、「意味分類コード」、「節間(KW)(キーワード)」を組み合わせる意味検索を行うためのプログラム	
平成17年度前期	意味類型パターンファイルシステム拡充版作成作業	39)改訂版意味類型パターン辞書ファイル	31)のファイルに、「拡散記号の適正化」、「時制・相・様相表現の記述改良」、「標準表記への統一」、「選択記号の適正化」、「文型任意記号の適正化」、「自己パターン適合性向上」、「英語構文情報付与作業」、「英語文型パターンの適正化」、「節変数化の起用範囲の見直し」を実施
		40)関連ドキュメント改訂版	「パターン記述言語仕様書」、「節間意味分類体系表」、「単文意味分類体系表」と「用言意味辞書」、「名詞意味属性表」と「名詞意味辞書」の改訂版。
		41)文型パターン検索プログラム改訂版	「文型パターンパーサー」と「文型パターン意味検索プログラム」の改訂版
		42)文型パターン品質評価実験報告書	「文型パターン辞書被覆率実験報告書」と意味的類似関係に着目して検索精度を評価した「意味的類似検索実験報告書」の2種類

平成 17 年度 後期	意味類型パターンファイル改良版作成作業	43)意味類型パターンファイル改良版	39)のパターン辞書に対し、(1)パターンの構文分類を行い、英語文型パターンの記述方法を大幅に改良、(2)従来の問題の中から効果の順に改良。(3)意味分類体系を見直しと新しい分類コードを付与
		44)文型パターン辞書被覆率実験報告書	43)のファイルで被覆率実験を行い、その結果から今後の改良項目の可能性を検討したもの
	意味類型パターンのブラッシュアップと評価作業	45)意味類型パターン辞書ファイル	39)のファイルに、(1)節間意味コードと用言意味属性の絞り込み、(2)過汎化パターンの見直し、(3)汎化バグ修正、(4)系統的不良の検出と修正、(5)文型パターン圧縮を行ったもの
		46)意味類型パターン評価作業報告書	45)のファイルについて、意味検索実験を行い、その結果をまとめたもの
平成 18 年度 前期	意味類型パターン辞書システム総合版編成作業	47)意味類型パターン辞書システム(部内版)ファイル	45)の内容をさらに改良し、最終版として集大成したもの。「パターン辞書(部内版)説明書」、「パターン辞書(部内版)」、「パターンパーサ(部内版)」、「意味検索プログラム(部内版)」、「対訳コーパスファイル」
		48)意味類型パターン辞書システム(提供版)ファイル	47)の内容を提供版として再編集。「パターン辞書(提供版)説明書」、「パターン辞書(提供版)」、「単語意味分類辞書ファイル(提供版)」、「パターンパーサ(提供版)」、「意味検索プログラム(提供版)」
		49)意味類型パターン辞書総合評価実験評価報告書	意味類型パターン辞書を総合的に評価したもので、以下のものから構成される。<1>文型パターン辞書の静的な特性評価、<2>文型パターン辞書の動的な特性評価、<3>意味的類似パターン検索実験評価
		50)意味類型パターン辞書システム説明書	「意味類型パターン辞書開発技術説明書」、「意味類型パターン辞書説明書」、「パターン記述言語仕様書」、「パターンパーサ説明書」、「意味検索プログラム説明書」、「重文複文対訳コーパスファイル説明書」
平成 18年 12月	意味類型パターン辞書システム改訂作業	51)意味類型パターン辞書システム(総合版)ファイル	45)のファイルに、個別的パターンチェックと改良、不適合試験文の解析と文型パターン改良、単文適合パターンの改良、高頻度適合パターンの改良の作業を実施
		52)意味類型パターン辞書システム(提供版)ファイル	49)の内容を提供版として再編集した。「意味類型パターン辞書システム(提供版)ファイル説明書」、「意味類型パターン辞書ファイル(提供版)」
		53)意味類型パターン辞書システム(圧縮版)ファイル	50)のパターンファイルを性能が低下しない範囲で圧縮したもの。
		54)意味類型パターン辞書総合評価報告書	19)と51)のパターン辞書を総合的に評価したもの。<1>静的な特性評価、<2>動的な特性評価、<3>意味検索実験評価
		55)意味類型パターン辞書システム説明書	48)の説明書に加筆修正を行ったもの。

### 3.1.2 研究成果の今後期待される効果

さまざまな分野で、バリアフリーの社会の実現が要請されているが、情報化社会であらゆる人が自由に使用できる言語メディアは母国語である。本研究による意味処理技術は、国際的なネットワーク社会の中で、日本語を使用した高度メディア社会の実現に大きく貢献するものと期待される。とりわけ、「計算機の言葉を人間が理解しなければならない社会」から、逆に、「計算機が人間に言葉を理解する社会」への突破口を開く点での意義は大きい。以下、言語教育や言語学、国語学などの研究での利用も含め、本研

究の成果の応用例を示す。

- (1) 意識型の高品質な「機械翻訳システム」や発想支援型の「文章作成支援システム」の実現
- (2) 意味による「情報検索」、「文書誤り検出訂正」、「音声認識・合成」の品質向上
- (3) 自然言語によるマンマシーンインターフェースの実現
- (4) 学生や社会人の英作文教育、外国人の日本語教育などでの利用
- (5) 日英言語表現分析など言語学、日本語学の分野への貢献

## 4. 研究参加者

### 4.1 研究統括グループ

方式基本検討、文型辞書、意味類型辞書の構築、試作システム設計と工程管理、成果のとりまとめ

氏名	所属	役職	研究項目	参加期間
池原 悟	鳥取大学工学部 知能情報工学科	教授	基本方式の検討	平成13年12月～ 平成19年 3月
田中 穂積	中京大学情報科学部 認知科学科	教授	基本方式の検討	平成13年12月～ 平成19年 3月
村上 仁一	鳥取大学工学部 知能情報工学科	助教授	基本方式の検討	平成13年12月～ 平成19年 3月
徳久 雅人	鳥取大学工学部 知能情報工学科	助手	基本方式の検討	平成13年12月～ 平成19年 3月
小林 昌博	鳥取大学工学部 知能情報工学科	専任講師	基本方式の検討	平成16年 4月～ 平成19年 3月
平川 秀樹	東芝研究開発センター	研究主幹	基本方式の検討	平成13年12月～ 平成19年 3月
山本 理絵	鳥取大学工学部 知能情報工学科	チーム 事務員	研究事務	平成14年 4月～ 平成19年 3月
吉岡 篤志	鳥取大学工学部 知能情報工学科	学生	名詞句への応用	平成15年 4月～ 平成18年 3月
堀田 波星夫	鳥取大学工学部 知能情報工学科	学生	音声への応用	平成15年 4月～ 平成18年 3月
前田 春奈	鳥取大学工学部 知能情報工学科	学生	照合方式の研究	平成15年 4月～ 平成18年 3月
小林 和晃	鳥取大学工学部 知能情報工学科	学生	汎化方式の研究	平成15年 4月～ 平成18年 3月
安井 敏	鳥取大学工学部 知能情報工学科	学生	名詞句への応用	平成15年 4月～ 平成18年 3月
片山 慶一郎	鳥取大学工学部 知能情報工学科	学生	圧縮法の研究	平成16年 4月～ 平成19年 3月
岡本 一輝	鳥取大学工学部 知能情報工学科	学生	音声への応用	平成16年 4月～ 平成19年 3月
山形 亮	鳥取大学工学部 知能情報工学科	学生	音声への応用	平成17年 4月～ 平成18年 3月
中村 聡	鳥取大学工学部 知能情報工学科	学生	意味検索方式	平成17年 4月～ 平成18年 3月
原 真一郎	鳥取大学工学部 知能情報工学科	学生	文型選択方式	平成17年 4月～ 平成18年 3月

揚 鵬	鳥取大学工学部 知能情報工学科	学生	中国語への応用	平成14年 4月 ~ 平成18年 3月
-----	--------------------	----	---------	------------------------

#### 4.2 等価変換システムグループ

日英言語の意味類型間の写像方式の検討と等価変換実験システムの設計製造、および、実験と評価

氏名	所属	役職	研究項目	参加期間
宮崎 正弘	新潟大学工学部 情報工学科	教授	日英等価変換方式	平成13年12月 ~ 平成19年 3月
フランス ・ボン	N T Tコミュニケーション科学 基礎研究所	主任 研究員	日英等価変換方式	平成15年 4月 ~ 平成18年11月
中岩 浩巳	N T Tコミュニケーション科学 基礎研究所	研究室長	日英等価変換方式	平成14年 8月 ~ 平成18年11月
荒木 哲朗	福井大学工学部 知能システム工学科	教授	日英等価変換方式	平成16年 4月 ~ 平成19年 3月
林 良彦	大阪大学大学院 言語文化研究科	教授	日英等価変換方式	平成16年 9月 ~ 平成19年 3月
久津見 毅	シャープ(株)	係長	日英等価変換方式	平成13年12月 ~ 平成19年 3月
古賀 勝男	(株)クロスランゲージ	代表 取締役	日英等価変換方式	平成13年12月 ~ 平成19年 3月
川辺 諭	新潟大学工学部 情報工学科	CREST 研究員	日英等価変換方式	平成14年 4月 ~ 平成18年 3月
武本 裕	新潟大学工学部 情報工学科	学生	日英等価変換方式	平成13年12月 ~ 平成19年 3月
藤田 和寛	新潟大学工学部 情報工学科	学生	日英等価変換方式	平成17年 4月 ~ 平成19年 3月
渡邊 南人	新潟大学工学部 情報工学科	D研究員	日英等価変換方式	平成14年10月 ~ 平成18年 3月
大山 芳史	N T Tアドバンス テクノロジー(株)	部長	日英等価変換方式	平成13年12月 ~ 平成15年 3月
白井 諭	A T R国際電気通信 基礎研究所	室長	日英等価変換方式	平成13年12月 ~ 平成14年 7月

#### 4.3 生成システムグループ

意味的論理範疇から日本文を生成する方法の検討と文生成実験システムへの応用研究

氏名	所属	役職	研究項目	参加期間
奥村 学	東京工業大学大学院 精密工学研究所	助教授	日本語生成方式	平成13年12月 ~ 平成19年 3月
乾 健太郎	奈良先端科学技術大学 院大学情報科学研究科	助教授	日本語生成方式	平成13年12月 ~ 平成19年 3月
山本 和英	長岡技術科学大学 電気系	助教授	日本語生成方式	平成14年 4月 ~ 平成19年 3月
山本 薫	東京工業大学大学院 精密工学研究所	CREST 研究員	日本語生成方式	平成16年 4月 ~ 平成17年 3月
阿部 修也	奈良先端科学技術大学 院大学情報科学研究科	学生	日本語生成方式	平成17年 7月 ~ 平成19年 3月



#### 4.4 意味類型知識ベースグループ

意味類型パターン記述言語の設計、日英対訳コーパスからの文型パターン抽出と意味類型化

氏名	所属	役職	研究項目	参加期間
池田 尚志	岐阜大学工学部 応用情報学科	教授	言語知識ベース 構成法	平成13年12月～ 平成19年 3月
佐良木 昌	日本大学	講師	言語知識ベース 構成法	平成13年12月～ 平成19年 3月
新田 義彦	日本大学大学院 総合社会情報研究科	教授	言語知識ベース 構成法	平成13年12月～ 平成19年 3月
柴田 勝征	福岡大学 理学部 応用数学科	教授	言語知識ベース 構成法	平成13年12月～ 平成19年 3月
井佐原 均	情報通信機構けいはんな 情報通信融合研究センター	研究室長	言語知識ベース 構成法	平成14年 8月～ 平成19年 3月
佐藤 暢雄	秋田大学 医科学情報学講座	教授	言語知識ベース 構成法	平成13年12月～ 平成19年 3月
滝沢 直宏	名古屋大学大学院 国際開発研究科	教授	言語知識ベース 構成法	平成13年12月～ 平成19年 3月
山田 節夫	N T Tサイバーソリュ ーション研究所	研究主任	言語知識ベース 構成法	平成16年 9月～ 平成18年11月
横尾 昭男	N T Tサイバースペ ース研究所	主幹 研究員	言語知識ベース 構成法	平成13年12月～ 平成16年 8月
荻野 孝野	日本システムアプリケ ーション	主任 研究員	言語知識ベース 構成法	平成13年12月～ 平成19年 3月
金田 重郎	同志社大学 工学部知識工学科	教授	言語知識ベース 構成法	平成15年10月～ 平成19年 3月

#### 5. 招聘した研究者等

なし

#### 6 成果発表等

##### 6.1 原著論文発表 (国内誌 22件、国際誌 5件)

1. Masato Tokuhisa, Jin'ichi Murakami, Satoru Ikehara: Construction and Evaluation of Text-Dialog Corpus with Emotion Tags Focusing on Facial Expression in Comics. KES 2006, Part III, LNAI 4253, B. Gabrys, R.J. Howlett, and L.C. Jain (Eds.), Springer-Verlag Berlin Heidelberg 2006 (to appear). 採録決定
2. 藤田篤, 乾健太郎, 松本裕治: 語彙概念構造に基づく言い換え生成 - 機能動詞構文の言い換えを例題に - . 情報処理学会論文誌 . 47(6), 2006 . (掲載予定)
3. Nguyen My Chau, 池田尚志: 日本語-ベトナム語機械翻訳における連体修飾構造の翻訳, 自然言語処理 Vol.12No.3, pp.145-182(2005.3)
4. ト朝暉, 浅井良信, 宇野修一, 池田尚志: 日中機械翻訳における連体修飾節構造の翻訳について, FIT2005(第4回情報科学技術フォーラム), p73-76(2005.9)
5. Katsuyuki Shibata: Dependence on Context in case of English-Japanese Machine Translation I-1, Fukuoka University Science Reports, 2005. vol.35, No.1, pp.75-79

6. Katsuyuki Shibata: Dependence on Context in case of English-Japanese Machine Translation I-2, Fukuoka University Science Reports, 2005. vol.35, No.2, pp.85-88
7. 謝軍, 今井啓允, 池田尚志: 日中機械翻訳システム jaw/Chinese における変換・生成の方法, 自然言語処理 Vol.11 No.1p.43 ~ 80(2004)
8. ト朝暉, 池田尚志: 日中機械翻訳における否定文の翻訳, 自然言語処理 Vol.11, No.3. July.2004. p.97 ~ 112 (2004.3)
9. 金出地真人、徳久雅人、村上仁一、池原悟: 結合価文法による動詞と名詞の訳し分け能力の評価、自然言語処理、Vol.11, No.3, pp. 149-164 (2004)
10. 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)
11. 池原悟、徳久雅人、竹内(村本)奈央、村上仁一: 日本語重文・複文を対象とした文法レベル文型パターンの被覆率特性、自然言語処理、Vol.11, No.4, pp. 147-178 (2004)
12. 藤田篤, 乾健太郎, 松本裕治: 自動生成された言い換え文における不適格な動詞格構造の検出、Detection of Incorrect Case Assignments in Automatically Generated Paraphrases、情報処理学会論文誌, Vol. 45, No. 4, pp. 1176-1187, 2004.04
13. Katsuyuki Shibata, Dependence on Context in case of English-Japanese Machine Translation I-1, Fukuoka University Science Reports, vol.35 No.1, pp.75-79, March (2005)
14. Yoshihiko NITTA, Extracting Semantic Role Sequence from Bilingual Corpora by Shallow Parsing, Nihon University Economic Review, Vol.73, No.4, Nihon University(2004-1)pp.1-12
15. 新田義彦, 外国語教育に自然言語処理を応用する方法の検討 (Applying Natural Language Processing Technique to Foreign Language Education), Nihon University Economic Review, Yoshihiko NITTA, Extracting Semantic Role Sequence from Bilingual Corpora by Shallow Parsing, Nihon University Economic Review, Vol.74, No.1, Nihon University(2004-4)pp.1-12
16. 新田義彦: 参照構造ネットワークによるアイデア・プロセッサの方式検討 (Idea Processing Method by Referential Network Structure), Nihon University Economic Review, Vol.74, No.2, Nihon University(2004-7)pp.1-17
17. 新田義彦, 自然言語処理研究の潮流概観 (A Survey of Recent Trend of Natural Language Processing Study), Nihon University Economic Review, Nihon University Economic Review, Vol.74, No.3, Nihon University(2004-10)pp.79-90
18. Yoshihiko NITTA, Applying Natural Language Processing technique to Language E-Learning, Nihon University Economic Review, Vol.74, No.4, Nihon University(2005-1)pp.1-16
19. 新田義彦, Web 学習における教材提示方法の検討 (How to Manipulate Teaching Materials in Web-Learning Environment), Nihon University Economic Review, Vol.75, No.1, Nihon University(2005-4)pp.1-20
20. Yoshihiko Nitta, Masashi Saraki & Satoru Ikehara, Natural Language Approach to Language E-Learning, in: K.Murase et al (eds.), Human and Artificial Intelligence, Vol.1, Fukui University (2004, Dec.5-6)pp.522-527

21. 池原悟, 村上仁一, 木本泰博: 単語意味属性を用いたベクトル空間法, 言語処理学会論文誌, Vol.10, No.2, pp.111-128 (2003.4)
22. 池原悟, 村上仁一, 桐沢洋: 意味的用法に着目した日本語名詞の英訳語選択について, 情報処理学会論文誌, Vol.44, No.5, pp.1343-1353 (2003.5)
23. 池原悟, 村上仁一, 桐沢洋: 意味的用法に着目した日本語名詞の英訳語選択について, 情報処理学会論文誌, Vol.44, No.5, pp.1343-1353 (2003.5)
24. 徳久雅人, 守谷有司, 池原悟, 村上仁一: 意味属性の共起による「AのB」型名詞句の翻訳規則, FIT 札幌, pp.87-89 (2003)
25. 謝軍, 今井啓允, 池田尚志: 日中機械翻訳におけるテンス・アスペクトの処理, 言語処理学会, 自然言語処理 Vol.10 No.4 p.177~200(2003)
26. ト朝暉, 謝軍, 池田尚志: 中日機械翻訳におけるとりたて表現の翻訳について—「も」「さえ」「でも」—, 自然言語処理学会論文 pp. 111-130 (2002.10)
27. 池原悟(鳥取大学): 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 (2002.5)
28. 池原悟, 村上仁一, 車井登: 日英機械翻訳のための日本語抽象名詞の文法的・意味的用法の分類, 自然言語処理, Vol.9, No.1, pp.117-134 (2001.1)

## 6.2 その他の著作物

1. 長尾眞監修: 「ヒューマン・インフォマティクス - 触れる・伝える・究める - デジタル生活情報術」、工作社、2005.6-10、ルポライターによる池原悟の研究紹介記事: 「思考や文章の本質に迫る」(pp.246-263)を掲載
2. 柴田勝征(著): 「US式和英翻訳システムの制作」、ラッセル社、2006.2-24、和英翻訳システムの作り方、考え方の背景、プログラム・ソースコードなどについて解説。(PP.1-301)
3. 滝沢直宏: 「コーパスで一目瞭然 - 品詞別 本物の英語はこう使う -」、小学館、2006.3
4. 池原悟: 意味類型論による言語の等価的変換方法の研究、TELECOM FRONTIER、SCAT-Technical Journal, No.48, pp.4-12, 2005 Summer
5. 新田義彦: Web学習における教材提示方法の検討 (How to Manipulate Teaching Materials in Web-Learning Environment), 日本大学経済学部、経済集誌、Vol. 75, No. 1, pp.1-20, Apr-2005
6. 池原悟: 「自然言語処理と言語過程説」、佐良木昌編: 「言語過程説の探求(第一巻)」、明石出版(2004), pp.331-408
7. 佐良木昌: 「正規表現とテキストマイニング」、明石書店(2003)
8. 池原悟: 「類推思考の原理に基づく言語の意味的等価変換方式」、鳥取大学情報処理総合センター広報<2003.1>
9. 池原悟: <AI - マップ> 自然言語処理の基本問題への挑戦, 人工知能学会誌, Vol.16, No.3, pp.522-430 (2001.5) <招待記事>
10. 池原悟: <AI - マップ> 「自然言語処理の基本問題への挑戦」へのコメントと回答, 人工知能学会誌, Vol.16, No.4, pp.538-549 (2001.7) <招待記事>

## 6.3 学会発表 (国際学会発表及び主要な国内学会発表)

### 6.3.1 招待講演 (国内会議 1件、国際会議 3件)

1. Satoru Ikehara: A Huge Lexical Database and a new Method for Machine Translation, EAJS Conference, Warsaw Poland (2003.8.27-28)
2. 池原悟 (鳥取大学) : 意味類型と類推思考に基づく言語の等価的変換と生成技術の研究、情報科学技術フォーラムFIT、JST-4(2002.9)
3. Satoru Ikehara: Toward the Realization of Typological Semantic Pattern Dictionary, MT Roadmap Workshop at TMI2002, 京阪奈Japan(2002.3)
4. Satoru Ikehara: Meaning Comprehension Using Semantic Patterns in a Large Scale Knowledge-Base, Proceedings of the PACLING'01, pp.26-35(2001.9)

### 6.3.2 口頭発表 (国内会議 128件、国際会議 18件)

1. Satoru Ikehara, Masato Tokuhisa, Jin'ichi Murakami : Pattern Dictionary Development based on Non-Compositional Language Model for Japanese Compound and Complex Sentences : Masashi Saraki, Masahiro Miyazaki and Takashi Ikeda, ICCPOL-06, The 21st International Conference on Computer Processing of Oriental Languages, (2006.12.17-19) 採録決定
2. 池原悟、阿部さつき、竹内奈央、徳久雅人、村上仁一 : 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、情報処理学会、自然言語処理研究会、2006-NL-176, pp.1-6 (2006.11.22)
3. 徳久雅人、村上仁一、池原悟 : 重文・複文文型パターン辞書からの構造照合型パターン検索、情報処理学会、自然言語処理研究会、2006-NL-176, pp. (2006.11.22)
4. Yoshihiko Nitta, Masashi Saraki and Satoru Ikehara : Building Carefully Tagged Bi-lingual Corpora to Cope with Linguistic Idiosyncrasy: LREC2006 (5th Conference on Language Resources and Evaluation), (2006.3)
5. Y.Nitta, M.Saraki and S.Ikehara : Building Carefully Tagged Bi-lingual Corpora to Cope with Linguistic Idiosyncrasy, Proc. of LREC06(5th Conference on Language Resources and Evaluation), ELRA(The European Language Resource Association), May-22-2006
6. Satoru Ikehara, Masato Tokuhisa, Jin'ichi Murakami and Masashi Saraki: Development of Semantic Pattern Dictionary for Non-linear Structures of Complex and Compound Sentences, Proceedings of the ESSLLI'05 Workshop on Empirical Challenges and Analytical Alternatives to Strict Compositionality, 2005.8, pp.79-98,(Edinburgh, Scotland)
7. Satoru Ikehara, Masato Tokuhisa, Jin'ichi Murakami, Masashi Saraki, Takashi Ikeda and Masahiro Miyazaki: Semantic Pattern Dictionary for Translating Non-linear Structures of Complex and Compound Sentences, PACLING05, pp.170-179, 2005.8 (Tokyo)
8. Masato Tokuhisa, Kumiko Endo, Yuya Kanazawa, Jin'ichi Murakami and Satoru Ikehara: Evaluation of Pattern Generalization Effect under Development of Pattern Dictionary for Machine Translation, PACLING05, pp.311-319, 2005.8 (Tokyo)
9. Zaohui BU, Yoshinobu Asai, Syuichi Uno, Takashi Ikeda : Translation of an Adnominal Construction in Japanese-Chinese Machine Translation—Japanese-Chinese "No" and Chinese "De" — , Proceedings of 2005 IEEE, International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE'05) (2005.10)

10. Tadahiro Matsumoto, Mayo Taniguchi, Akiji Yoshida, Nobuaki Tanaka and Takashi Ikeda : A Proposal of a Notation System for Japanese Sign Language and Machine Translation from Japanese Text to Sign Language Text, Pacific Association for Computational Linguistics, PAACLING2005, 218-225 (2005.8)
11. Atsushi Fujita and Kentaro Inui : A Class-oriented Approach to Building a Paraphrase Corpus. Proc. of the 3rd International Workshop on Paraphrasing (IWP2005), pp.25-32, 2005.
12. Atsushi Fujita, Kentaro Inui and Yuji Matsumoto : Exploiting lexical conceptual structure for paraphrase generation. In Proc. of the 2nd International Joint Conference on Natural Language Processing, pp. 908-919, 2005.
13. 池原悟、徳久雅人、村上仁一、佐良木昌、池田尚志、宮崎正弘：非線形な重文複文の表現に対する文型パターン辞書の開発、情報処理学会、自然言語処理研究会 2005-NL-170、pp.157-164 (2005.11/22)
14. 松本忠博、谷口真代、吉田鑑地、池田尚志：日本語-手話機械翻訳システムに向けて—テキストレベルの翻訳系の試作と簡単な例文の翻訳—、電子情報通信学会研究報告書 TL2004-43 WIT2004-67(2005-01)
15. 王軼謳、卜朝暉、宇野修一、浅井良信、池田尚志 日中機械翻訳における存在文および関連する問題について、情報処理学会、第 171 回自然言語処理研究会 2006-NL-171, p.95 ~ 102
16. 吉岡篤志、徳久雅人、村上仁一、池原悟:名詞句パターン辞書を用いた日英機械翻訳の試作、平成 17 年度情報関連学会・中国支部第 56 回連合大会、pp.305-306 (2005)
17. 安井敏、徳久雅人、村上仁一、池原悟:「A の B」型名詞句に対する連体修飾節の係り先の決定、言語処理学会第 12 回年次大会、E1-2, pp.125-128 (2006.3)
18. 小林和晃、村上仁一、徳久雅人、池原悟：単文文型パターンからの言い換えの抽出、言語処理学会第 12 回年次大会、P2-2, pp.200-203 (2006.3)
19. 揚鵬、村上仁一、徳久雅人、池原悟：結合価パターンを用いた日中機械翻訳方式の検討、言語処理学会第 12 回年次大会、P3-6, pp.264-267 (2006.3)
20. 原真一郎、村上仁一、徳久雅人、池原悟：日英機械翻訳における多変量解析を用いた最適パターンの選択、言語処理学会第 12 回年次大会、P3-7, pp.268-271 (2006.3)
21. 片山慶一郎、村上仁一、徳久雅人、池原悟：日本語文型パターンの縮退方法、言語処理学会第 12 回年次大会、E3-3, pp.568-571 (2006.3)
22. 小林和晃、村上仁一、徳久雅人、池原悟：選択記号による文型パターン汎化の効果、言語処理学会第 12 回年次大会、E3-4 pp.572-575 (2006.3)
23. 前田春奈、村上仁一、徳久雅人、池原悟：用言の意味属性を用いたパターン照合、言語処理学会第 12 回年次大会、E3-5, pp.576-579 (2006.3)
24. 吉岡篤志、徳久雅人、村上仁一、池原悟:日英対訳パターンを用いた名詞句翻訳、言語処理学会第 12 回年次大会、E3-6, 580-583 (2006.3)
25. 中村聡、村上仁一、徳久雅人、池原悟：「意味」による文型パターン検索方式の最適化、言語処理学会第 12 回年次大会、P4-11, pp.632-635 (2006.3)

- 26.田中努、徳久雅人、村上仁一、池原悟:情緒生起情報付き結合価パターン辞書の開発、言語処理学会第12回年次大会、E5-5, pp/1151-1154 (2006.3)
- 27.古塩貴行、徳久雅人、村上仁一、池原悟:テキストを対象とした情緒推定のための事象関係の解析に向けて、言語処理学会第12回年次大会、E5-6, pp.1155-1158 (2006.3)
- 28.徳久雅人、村上仁一、池原悟:文頭・文末表現を区別した確率言語モデルに基づく情緒推定、言語処理学会第12回年次大会ワークショップ、pp.17-20 (2006.3)
- 29.徳久雅人、松浦大樹、村上仁一、池原悟:漫画における表情に着目した情緒注釈付き対話コーパスの作成、電子情報通信学会総合大会、シンポジウム「円滑なコミュニケーション」(2006.3)
- 30.松田 聡史、山本 和英:機械翻訳における文型パターンの部分的利用.言語処理学会第12回年次大会,P3-4(2006.3)
- 31.武本裕、宮崎正弘:局所・大域制約による構造的曖昧性抑制機構をもつ日本語文パーザ、言語処理学会第12回年次大会、E1-5, pp.137-140 (2006.3)
- 32.斉藤仁史、宮崎正弘:英語文パーザにおける表層的・統語的情報を用いた構造的曖昧性解消機構、言語処理学会第12回年次大会、E1-6, pp.141-144 (2006.3)
- 33.佐藤弘幸、宮崎正弘:複合名詞構造化規則と表層的・統語的情報を用いた日本語複合名詞構造解析法、言語処理学会第12回年次大会、D3-8, pp.556-559 (2006.3)
- 34.藤田和寛、宮崎正弘:高品質な日英単文用例翻訳システムの枠組み、言語処理学会第12回年次大会、E3-1, pp.560-563 (2006.3)
- 35.山田武志、宮崎正弘:日英主体表現パターン対を用いた主体表現の日英翻訳とその有効性、言語処理学会第12回年次大会、E3-2, pp.564-567 (2006.3)
36. 浅井良信、宇野修一、田中祐樹、卜朝暉、Nguyen My Chau、池田尚志、機械翻訳システム jaw における連体修飾構造の翻訳 - 日中、日越を事例として -、言語処理学会第12回年次大会、P3-11, pp.284-287 (2006.3)
- 37.穆貴彬、宇野修一、福本真哉、浅井良信、卜朝暉、池田尚志、構造にずれのある表現に対する jaw による翻訳処理について、言語処理学会第12回年次大会、P3-10, pp.280-283 (2006.3)
- 38.王軼謳、卜朝暉、宇野修一、浅井良信、池田尚志 日中機械翻訳における存在文の翻訳処理について、言語処理学会第12回年次大会、P3-9, pp.276-279 (2006.3)
- 39.原田大樹、松本忠博、加藤三保子、原大介、池田尚志、日本語を援用した手話表記および Sign Writing による手話ビデオの書き取り実験、言語処理学会第12回年次大会、P6-10, pp.719-722 (2006.3)
- 40.石原吉晃、山田佳裕、松本忠博、池田尚志、1文字未知語を核とする未知語候補の抽出、言語処理学会第12回年次大会、P1-4, pp.165-168 (2006.3)
- 41.山田佳裕、高松大地、石原吉晃、水野智美、大口智也、佐藤芳秀、松本忠博、池田尚志、日本語文解析システム ibukiC/S について、言語処理学会第12回年次大会、P1-9, pp.185-187 (2006.3)

- 42.阿部修也, 乾健太郎, 松本裕治: 論理関係に基づく複文間の言い換え含意関係の認識と生成. 言語処理学会年次大会, P2-3, pp.204-207 (2006.3)
- 43.ト朝暉, 浅井良信, 宇野修一, 池田尚志: 日中機械翻訳における連体修飾節構造の翻訳について, FIT2005(第4回情報科学技術フォーラム), p73-76(2005.9)
- 44.Masato Tokuhisa, Tsutomu Tanaka, Satoru Ikehara and Jin'ichi Murakami: Emotion Reasoning Based on Valency Patterns - a prototype annotation of causal relationships -, International Series on Natural and Artificial Intelligence, Vol.1, Human Intelligence Systems, From Control to Autonomy, pp. 534-539 (2004)
- 45.Atsushi Fujita, Kentaro Furihata, Kentaro Inui, Yuji Matsumoto, and Koichi Takeuchi: Paraphrasing of Japanese Light-verb Constructions Based on Lexical Conceptual Structure. Proceedings of the 42th Annual Meeting of the Association for Computational Linguistics (ACL) Workshop on Multiword Expressions, Integrating Processing, pp. 9-16, Jul., 2004.
- 46.竹信伸介, 徳久雅人, 村上仁一, 池原悟: 動詞節に修飾された名詞句の係り受け解析, 情報処理学会, 自然言語処理研究会, 2004-NL-163, pp.83-89 (2004.9)
- 47.古塩貴行, 徳久雅人, 村上仁一, 池原悟: 情緒注釈付き対話コーパスの誤り分析, 人工知能学会 18 回全国大会, 2G3-02, (2004)
- 48.岡田敏, 村上, 徳久, 池原: 多変量解析による最適文型パターンの選択方式, 言語処理学会第 11 回年次大会, A 1 - 7 (2005.)
- 49.金澤佑哉, 徳久雅人, 村上仁一, 池原悟: 文型パターンにおける時制・相・様相表現の汎化とその効果, 言語処理学会第 11 回年次大会, A 1 - 8 (2005)
- 50.佐伯美香, 徳久, 村上, 池原: 感情表現に着目した日本語副詞の日英対照分析, 言語処理学会第 11 回年次大会, A 1 - 9 (2005)
- 51.石上真理子, 徳久雅人, 村上仁一, 池原悟: 結合価パターンを用いた動詞句の翻訳可能性の調査, 言語処理学会第 11 回年次大会, A 2 - 6 (2005)
- 52.遠藤久美子, 徳久雅人, 村上仁一, 池原悟: 文型パターンにおける任意要素の記述方法とその効果, 言語処理学会第 11 回年次大会, A 2 - 7 (2005)
- 53.西山七絵, 村上仁一, 徳久雅人, 池原悟: 単文文型パターンの構築, 言語処理学会第 11 回年次大会, A 2 - 8 (2005)
- 54.神野絵理, 徳久雅人, 村上仁一, 池原悟: 文型パターンにおける名詞句翻訳のためのパターン辞書の構築, 言語処理学会第 11 回年次大会, A 2 - 9 (2005)
- 55.徳久雅人, 村上仁一, 池原悟: 情緒推定の手がかりについて, 言語処理学会第 11 回年次大会, S 3 - 2 (2005)
- 56.藤田篤, 乾健太郎, 松本裕治: 語彙概念構造で表現できる言い換え, 言語処理学会第 11 回年次大会発表論文集, A6-1, pp. 1129-1132, 2005.03.
- 57.佐良木昌・新田義彦: 「シテ形接続の意味分類試案」, 電子情報通信学会 2005 年総合大会, 一般講演, 「基礎・境界」, A-13, 思考と言語 A-13-3
- 58.山田武史, 宮崎正弘: 日英主体表現パターン対を用いた様相表現の日英翻訳, 言語処理学会第 11 回年次大会, A3-1(2005).

- 59.上野圭介, 宮崎正弘: 構造変換対を用いた複合名詞の英日-日英翻訳, 言語処理学会第 11 回年次大会, A4-2(2005).
- 60.川辺諭, 宮崎正弘: 構造を含む生成規則を扱える拡張型チャートパーザ, 言語処理学会第 11 回年次大会, C4-4(2005).
- 61.吉田雄太, 宮崎正弘: 日本語を原言語とする機械翻訳における日本語等価変換機構, 言語処理学会第 11 回年次大会, A6-8(2005).
- 62.岩淵亮, 宮崎正弘: 英語構文解析における句動詞関連の曖昧性解消機構, 言語処理学会第 11 回年次大会, C6-1(2005).
- 63.宇野修一, 福本真哉, 田中伸明, 松本忠博, 池田尚志: 日本語から多言語への機械翻訳システム jaw, 言語処理学会, 第 11 回年次大会発表論文集 pp.538 ~ 541(2005.3)
- 64.谷口真代, 吉田鑑地, 田中伸明, 伊佐治和哉, 松本忠博, 池田尚志: 日本語—手話機械翻訳システム (jaw/SL) 構築の試みと翻訳実験, 言語処理学会, 第 11 回年次大会発表論文集 pp.526 ~ 529(2005.3)
- 65.伊佐治和哉, 山田佳裕, 石原吉晃, 高松大地, 松本忠博, 池田尚志: 文節構造解析システム ibukiC, 言語処理学会, 第 11 回年次大会発表論文集 pp.719 ~ 722(2005.3)
- 66.田中伸明, 卜朝暉, 浅井良信, 宇野修一, 伊佐治和哉, 松本忠博, 池田尚志: 日中機械翻訳システム jaw/Chinese における取り立て表現の翻訳処理, 言語処理学会, 第 11 回年次大会発表論文集 pp.189 ~ 192(2005.3)
- 67.浅井良信, 宇野修一, 卜朝暉, 福本真哉, 田中友樹, 田中伸明, Nguen My Chau, Ma Ngin Khaw Cing, Samantha Telijjagoda, 張穎, 吉田鑑地, 谷口真代, 伊佐治和哉, 松本忠博, 池田尚志: 機械システム jaw と多言語への翻訳実験, 言語処理学会, 第 11 回年次大会発表論文集 pp.233 ~ 236(2005.3)
- 68.My Chau Nguyen, Ma Nign Chaw Cing, 卜朝暉 Samantha Thelijjagoda, 池田尚志: 岐阜大・池田研究室での機械翻訳・自然言語処理研究の概要, 平成 16 年 11 月 12 日 大阪外国語大学 多言語処理と機械翻訳プロジェクトワークショップ(2004.11)
- 69.Yshihiko Nitta, Satoru Ikehara, Takashi Ikeda, Masahiro Miyazaki, Satoshi Shirai, Katsuyuki Shibata; Reevaluation of the Classic Machine Translation through the Pattern Translation, Proceedings of the PACLING-03, pp.1-4(2003)
- 70.謝軍, 今井啓允, 卜朝暉, 池田尚志: 一種日本用言後接成分的漢訳方法 (An Approach to Translate Japanese Modality into Chinese Expressions) (in Chinese), 20th International Conference on Computer Processing of Oriental Languages Shenyang, China, August 4-6, 2003 Proceedings, Advances in Computation of Oriental Languages, pp. 347 ~ 355(2003.8)
- 71.卜朝暉, 謝軍, 池田尚志: 日中機器翻訳の中文否定詞的位置確定法 (An Method for Deciding the Position of Chinese Negatives in Japanese-Chinese Machine Translation) (in Chinese), 20th International Conference on Computer Processing of Oriental Languages Shenyang, China, August 4-6, 2003 Proceedings, Advances in Computation of Oriental Languages, pp. 338 ~ 346 (2003.8)
- 72.Nguyen My Chau, Imai Yoshimasa, Ikeda Takashi :Translation of Adnominal Embedding Clause in Japanese-Vietnamese Machine Translation, 1st International Joint Conference Natural Language Processing, Companion Volume to the Proceedings



of the Conference , pp. 41 ~ 44(2004.3)

- 73.Samantha Thelijagoda, Yoshimasa Imai, Nayana Elikewala, Takashi Ikeda :Japanese-Sinhalese MT System(jaw/Sinhalese) and a Solution for the Translation of Japanese Function words after Predicate in to Sinhalese, 1st International Joint Conference on Natural Language Processing, Proceedings of Asian Symposium on Natural Language Processing to Overcome Language Barriers,,IJCNLP-04,Satellite Symposium , pp. 73 ~ 78(2004.3)
- 74.Tadahiro Matsumoto, Nobuaki Tanaka, Akiji Yoshida, Yoshimasa Imai, Takashi Ikeda :The First Step Toward a Machine Translation System from Japanese Text to Sign Language , 1st International Joint Conference on Natural Language Processing, Proceedings of Asian Symposium on Natural Language Processing to Overcome Language Barriers, IJCNLP-04, Satellite Symposium, pp. 61-66(2004.3)
- 75.衛藤純司, 池原悟, 池田尚志, 佐良木昌, 新田義彦, 柴田勝征, 宮崎正弘, 白井諭 : 意味類型構築のための接続表現の体系化について, 情報処理学会研究報告, 2003-NL-155, pp.31-38(2003. 5 /26/27 東工大)
- 76.柴田勝征 : 機械翻訳実習授業による言語観の獲得,電子情報通信学会・思考と言語研究会, 信学技報 TL2003-24, pp.1-5(2003-12)
- 77.池原悟 : 言語で表現される概念と翻訳の原理,言語と思考研究会, TL2003-25, pp.7-12 ( 2003.12 )
- 78.池原悟 : 非線形な言語表現と文型パターンによる意味の記述,情報処理学会,自然言語処理研究会,2004-NL-159, pp.139-146 ( 2004.1 )
- 79.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, 情報処理学会, 自然言語処理研究会, 2004-NL-159, pp.139-146 (2004.1.13)
- 80.池原悟, 徳久雅人, 村本奈央, 村上仁一 : 日本語重文・複文を対象とした文法レベル文型パターンの被覆率特性, 人工知能学会研究会 SIG-SLUE ( 2004.3/5 東大 )
- 81.佐伯美香, 徳久雅人, 村上仁一, 池原悟 : 副詞および形容詞による感情表現性の判定, FIT, 札幌,pp.117-118 ( 2003 )
- 82.田中康仁, 村上仁一, 徳久雅人, 池原悟 : 構造的類似文検索アルゴリズムを応用した日本語文型パターン抽出法, FIT, 札幌, pp.89-90 ( 2003 )
- 83.吉田真司,徳久雅人,村上仁一,池原悟 : 結合価パターンを用いた仮名漢字変換候補の選択, 自然言語処理学会第 10 回年次大会, pp.717-720(2004.3/16-18 東工大)
- 84.田中努,徳久雅人,村上仁一,池原悟 : 結合価パターンへの情緒生起情報の付与, 自然言語処理学会第 10 回年次大会, pp.345-348(2004.3/16-18 東工大)
- 85.徳久雅人,村上仁一,池原悟 : 文型パターンパーサの試作, 自然言語処理学会第 10 回年次大会, pp.608-611(2004.3/16-18 東工大)
- 86.岡田敏,村上仁一,徳久雅人,池原悟 : DP を使用した重文複文の日英翻訳の精度, 自然言語処理学会第 10 回年次大会, pp. 233-236(2004.3/16-18 東工大)
- 87.前田春奈,村上仁一,徳久雅人,池原悟 : パターンを使用した重文複文の日英翻訳の翻訳, 自然言語処理学会第 10 回年次大会, pp.237-240(2004.3/16-18 東工大)

- 88.道祖尾太祐, 村上仁一, 徳久雅人, 池原悟: N-gram を利用した日英対訳パターンの自動抽出, 自然言語処理学会第 10 回年次大会, pp.241-244 (2004.3/16-18 東工大)
- 89.藤本敬史, 村上仁一, 徳久雅人, 池原悟: 埋め込み文に対する日英翻訳規則の作成について, 自然言語処理学会第 10 回年次大会, pp.245-248(2004.3/16-18 東工大)
- 90.佐々木, 宮崎: 単語連鎖列書き換え規則を用いた形態素解析後処理とその評価, 言語処理学会第 10 回年次大会, A6-4 (2004.3)
- 91.須田, 宮崎: 大域的制約を利用した構造的曖昧さの抑止機構をもつ日本語文パーザ, 言語処理学会第 10 回年次大会, C3-1 (2004.3)
- 92.武本, 宮崎: 名詞間の接続強度と用例の係り受け情報を用いた「の」型名詞句構造解析, 言語処理学会第 10 回年次大会, C5-1 (2004.3)
- 93.孟, 川辺, 宮崎: 規則・用例融合型の複合名詞の日英翻訳方式, 言語処理学会第 10 回年次大会, B1-4 (2004.3)
- 94.川辺, 武本, 宮崎: 構造化された日英文型パターンを用いた線形・非線形融合型日英構造変換方式, 言語処理学会第 10 回年次大会, B1-6 (2004.3)
- 95.安藤, 川辺, 宮崎: 大域的句型制約を組み込んだ線形・非線形融合型英語文生成方式, 言語処理学会第 10 回年次大会, B1-5 (2004.3)
- 96.今井啓允, 池田尚志: オブジェクト指向言語のパラダイムを利用した機械翻訳エンジン jaw , 言語処理学会, 第 10 回年次大会発表論文集 pp.125 ~ 128(2004.3)
- 97.福本真哉, Ma Ngin Khaw Cing, 今井啓允, 池田尚志: ミャンマー語の述語部の文法分析と日本語 - ミャンマー語機械翻訳システム jaw/Myanmar について, 言語処理学会, 第 10 回年次大会発表論文集 pp.161 ~ 164(2004.3)
- 98.宇野修一, ト朝暉, 今井啓允, 池田尚志: 日中機械翻訳システム jaw/Chinese における連体修飾表現, 場所表現等の処理, 言語処理学会, 第 10 回年次大会発表論文集 p.640 ~ 643(2004.3)
- 99.伊佐治和哉, 山田将之, 池田尚志: 長単位の機能語を辞書に持たせた文節構造解析システム ibukiC , 言語処理学会, 第 10 回年次大会発表論文集 p.636 ~ 639(2004.3)
100. 田中伸明, 吉田鑑地, 松本忠博, 池田尚志: 日本語テキストから手話単語列への機械翻訳システム jaw/SL への試み, 言語処理学会, 第 10 回年次大会発表論文集 p.397 ~ 400(2004.3)
101. マニンコウシン, 福本真哉, 池田尚志: 日本語-ミャンマー語機械翻訳システム jaw/Myanmar における述語部構造の翻訳について, FIT2004 (第 3 回情報科学技術フォーラム) E-014, p.139 ~ 141 (2004.3)
102. 松本忠博, 田中伸明, 吉田鑑地, 谷口真代, 池田尚志: 日本語テキストから手話表記テキストへの機械翻訳の試み, 電子情報通信学会, 平成 16 年 9 月 FIT2004(第 3 回情報科学技術フォーラム) E-014, p.539 ~ 540 (2004.3)
103. 藤田篤, 乾健太郎: 「語彙的言い換えに必要な知識の部品化」, 情報処理学会台 149 回自然言語処理研究会 (平成 14 年 5 月)
104. 池原悟, 佐良木昌, 宮崎正弘, 池田尚志, 新田義彦, 白井諭, 柴田勝征: 等価的類推思

- 考の原理による機械翻訳方式、電子情報通信学会、言語と思考研究会 TL2002-34, pp.7-12(2002.12)
105. 道祖尾太祐、村上仁一、徳久雅人、池原悟：日英対訳パターンの自動抽出について、情報処理学会自然言語処理研究会、NL - 153-15、pp.113-118 (2003.1)
  106. 金出地真人、徳久雅人、村上仁一、池原悟：結合価文法による動詞と名詞の訳語選択能力の評価、情処研報、自然言語処理研究会、2003-NL-153-16, pp.119-124 (2003.1)
  107. 池原悟、宮崎正弘、佐良木昌、池田尚志、白井諭、村上仁一、徳久雅人：機械翻訳のための日英文型パターン記述言語、電子情報通信学会、言語と思考研究会 (TL 研) (2003.3)
  108. 衛藤純司、池原悟、池田尚志、佐良木昌、新田義彦、柴田勝征、宮崎正弘、白井諭：意味類型構築のための接続表現の体系化について、情報処理学会 NL 研究会、2003-NL(2003.5/26/27)
  109. 尾高智大、村上仁一、徳久雅人、池原悟：構造的類似文の自動検出と縮退方法、言語処理学会第9回年次大会 (2003.3)
  110. 岸井謙一、伊佐治和哉、高木優紀江、池田尚志：「文節解析システム ibukiB と大規模コーパス中の文節パターンの分布について、言語処理学会第9回年次大会 (平成15年3月)
  111. 川辺諭、武本裕、宮崎正弘：「線形・非線形融合型日英構造変換のためのパターン記述形式」、言語処理学会第9回年次大会 (平成15年3月)
  112. 武本裕、宮崎正弘、川辺諭：、「パターン対を用いた線形・非線形融合型日英構造変換方式」、言語処理学会第9回年次大会 (平成15年)
  113. 藤田篤、乾健太郎、松本裕治：「平易な表現への言い換えに必要なテキスト修正処理」、第65回情報処理学会全国大会特別トラック T16 言語バリアフリー (平成15年3月)
  114. 伊佐治和哉、岸井謙一、池田尚志：文節構造解析システム ibukiB について、FIT(情報科学技術フォーラム)2002,pp. 131-132 (2002.9)
  115. 高木優紀江、林誠悟、池田尚志：大規模コーパスにおける文パターンの分布調査、FIT(情報科学技術フォーラム)2002,pp. 165-166 (2002.9)
  116. 今井啓允、謝軍、池田尚志：パターン変換型機械翻訳システム j-aw について、FIT(情報科学技術フォーラム)2002,pp. 169-170 (2002.10)
  117. 謝軍、今井啓允、卜朝暉、池田尚志：日中機械翻訳における中国語語順の決定法について、FIT(情報科学技術フォーラム)2002,pp. 171-172 (2002.10)
  118. Nguyen My Chau , 今井啓允, 酒井美和, 池田尚志：日越語の構造比較と j-aw/Vietnamese による日越機械翻訳、FIT(情報科学技術フォーラム)2002,pp. 175-176 (2002.9)
  119. Samantha Thelijjagoda , 今井啓允, Nayana Elikewala , 池田尚志：Shinhalese Generation Rule and Machine Translation System j-aw / Sinhalese , FIT (情報科学技術フォーラム) 2002(2002.10)
  120. 池田尚志：Machine translation system j-aw and other NLP activities of our laboratory , 第2回中日自然言語処理シンポジウム論文集 pp. 135-141 (2002.10)

121. Nguyen My Chau , 酒井美和 , 今井啓允 , 池田尚志 : 日越機械翻訳システム jaw/Vietnamese における名詞句の処理 , 言語処理学会第 9 回年次大会発表論文集 pp.437 ~ 440(2003.3)
122. Nayana Elikewala , Samantha Thelijjagoda , Takashi Ikeda : 日本語-シンハラ語における格助詞相当語の対応について , 言語処理学会第 9 回年次大会発表論文集 pp.473 ~ 476(2003.3)
123. Samantha Thelijjagoda , Yoshimasa Imai , Nayana Elikewala , Takashi Ikeda : Machine Translation System jaw/Shinhalese , 言語処理学会第 9 回年次大会発表論文集 pp.477 ~ 480(2003.3)
124. ト朝暉 , 謝軍 , 池田尚志 : 日中機械翻訳における否定文の扱いについて , 言語処理学会第 9 回年次大会発表論文集 pp.659 ~ 662(2003.3)
125. 岸井謙一 , 伊佐治和哉 , 高木優紀江 , 池田尚志 : 文節解析システム ibukiB と大規模コーパス中の文節パターンの分布について , 言語処理学会第 9 回年次大会発表論文集 pp.655 ~ 658(2003.3)
126. 今井啓允 , 高木優紀江 , 池田尚志 日本語からアジア諸言語への機械翻訳の試み , 情報処理学会第 65 回全国大会講演論文集(2003.3) 5-363 ~ 5-566
127. 廣瀬文男、池原悟、村上仁一：日本語で表現された対象物の位置的関係の解析、情報処理学会、NL 研究会、NL-146-10、pp.61-66(2001.11)
128. 的場和幸、池原悟、村上仁一：日英時間表現の意味と対応関係の解析、情報処理学会、NL 研究会、NL-146-9、pp.53-60 ( 2001.11 )
129. 斉藤健太郎、池原悟、村上仁一：意味類型を用いた日英文型パターンの対応付け、情報処理学会、NL 研究会、NL-147-1、pp.1-8 ( 2002.1 )
130. 金出地真人、池原悟、村上仁一：結合価文法による動詞の訳語選択能力の評価、情報処理学会第 63 回全国大会、6Y-04、2-267-268(2001)
131. 藤本敬史、表克次、池原悟、村上仁一：埋め込み文の日英翻訳方式について - 内と外の関係の判断方法 - , 情報処理学会第 63 回全国大会、6Y-03、2-265-266(2001)
132. 守谷有司、宮本健司、池原悟、村上仁一：意味的關係に着目した「A の B」型名詞句の翻訳規則、情報処理学会第 63 回全国大会、6Y-5、pp.2-269-270(2001)
133. 竹信伸介、藤井龍、池原悟、村上仁一：形容詞の修飾關係を使用した日本語名詞の訳語選択、情報処理学会第 63 回全国大会、6Y-02、pp.2-263-264 (2001)
134. 澤田康子、井上慎一、池原悟、村上仁一：接続助詞の意味解析、情報処理学会第 63 回全国大会、4H-02、pp.2-17-18(2001)
135. 谷口雄介、池原悟、村上仁一：日本語の構造的類似文検索アルゴリズムの検討、情報処理学会第 63 回全国大会、5Y-2、pp.2-247-248(2001)
136. 斉藤健太郎、池原悟、村上 仁一：日英文型パターンの意味的対応方式、自然言語処理学会第 8 回年次大会、pp.295-298 ( 2002.3 )
137. 沢田康子、池原悟、村上仁一、斎藤健太郎:日英因果關係構文のパターン化、自然言語処理学会第 8 回年次大会、pp.291-294 ( 2002.3 )

138. 的場和幸、池原悟、村上仁一：中間表現を用いた日英時間表現の意味的対応関係の解析、自然言語処理学会第8回年次大会、pp.287-290 (2002.3)
139. 藤本敬史、池原悟、村上仁一、表克次：複文における底の名詞と修飾部の内と外の関係の判断規則、自然言語処理学会第8回年次大会、pp.679-682 (2002.3)
140. 竹信伸介、池原悟、村上仁一：動詞節に修飾された名詞句の係り受け解析、自然言語処理学会第8回年次大会、pp.415-418 (2002.3)
141. 吉田真司、池原悟、村上仁一：入力文に対する結合価パターン対の選択方法について、自然言語処理学会第8回年次大会、pp.299-302 (2002.3)
142. 卜朝暉、謝軍、池田尚志：日中機械翻訳における取立て詞表現の曖昧さ解消、情報処理学会第63回(平成13年後期)全国大会講演論文集 pp. 2-275-276 (2001.9)
143. 謝軍、卜朝暉、池田尚志：日中機械翻訳におけるテンス・アスペクトの処理、情報処理学会第63回(平成13年後期)全国大会講演論文集 pp. 2-273-374 (2001.9)
144. 一ノ瀬友紀夫、池田尚志：文節の内部構造解析と出現頻度統計、言語処理学会、第8回年次大会論文集 pp. 108-111 (2002.3)
145. 卜朝暉、池田尚志：日中機械翻訳のためのとりたて詞の意味解析と課題、言語処理学会、第8回年次大会論文集 pp. 81-84 (2002.3)
146. 謝軍、池田尚志：テンス・アスペクトの日中機械翻訳処理、言語処理学会、第8回年次大会論文集 pp. 77-80 (2002.3)

6.3.3 ポスター発表 (国内会議 0件、国際会議 0件)  
なし

6.3.4 プロジェクト主催の研究会発表 (国内会議 147件、国際会議 0件)

---

平成18年12月9～10日 「認識・言語・表現」研究会 明治学院大学(東京)

---

1. 白井諭：従属節の意味と係り受け(仮題)
2. 佐良木昌：感情表現と感情論理
3. 徳久雅人：情緒生起の原因を表す用言の分布について(仮題)
4. 池原悟：意味的等価変換のための重文と複文の統語構造と意味の分類
5. 「荻野孝野：動詞の結語価の体系的研究
6. 柴田勝征：英和翻訳システムの機能アップ
7. 佐良木昌、宮澤織江：動詞接続句辞書の実装による英訳改善
8. 松本忠博・池田尚志：手話の表記法についての試み

---

平成18年3月21日 H17年度末報告会 日本教育会館(東京)

---

1. 池原悟：意味類型パターン辞書研究開発の進捗状況
2. 阿部さつき：意味類型パターン辞書開発の実際
3. 宮崎正弘：文型パターンと用例の機械翻訳への応用
4. 乾健太郎：論理関係に基づく複文間の言い換え・含意の計算モデル
5. 山本和英：大規模テキストを用いた2文接続関係の同定
6. 村上仁一：意味類型パターン辞書開発のための個別的課題とその取り組み
7. 池田尚志：パターン翻訳規則を用いる日本語から多言語への翻訳システム
8. 佐良木昌：シテ型接続の意味に関する研究
9. 柴田勝征：文型パターンの翻訳システムへの実装方法

---

平成17年12月24-25日 「認識・言語・表現」研究会 一碧荘（熱海）

---

1. 宮崎正弘、武本裕：三浦文法に基づく日本語文解析の試み
  2. 柴田勝征：機械翻訳のための言語基礎理論試案
  3. 池原悟：非線形性に着目した言語表現モデルと文型パターン
  4. 池田尚志，松本忠博，原田大樹：手話のテキスト化と機械翻訳について
  5. 佐良木昌：時枝文法における接続助詞『て』の分析をめぐって
  6. 池原悟：文型パターン辞書の開発
- 

平成17年12月24日 第5回言語知識ベース班会議 一碧荘（熱海）

---

1. 佐良木昌：日英対訳パターン記述の校訂作業について
- 

平成16年12月18-19日 「認識・言語・表現」研究会 電通共催生協会館（東京）

---

1. 池原悟：意味類型論に基づく言語の等価変換研究の現状と今後の計画
  2. 佐良木昌：作用性用言反撥の法則の現代文意味解析への適用
  3. 池田尚志：日本語 - 手話機械翻訳システムの試みと手話表記法について
  4. 渡邊南人：Machine Translation Methods for Basic Verbs with the Use of Core Concepts and Recognition Primitives
  5. 柴田勝征：英和機械翻訳における文脈依存性の解決
  6. 新田義彦：FSA（有限状態オートマトン）による解析と生成について
- 

平成16年9月11-13日 言語知識ベース研究会 久住高原やまなみ荘（大分県）

---

1. 池原悟：CREST研究の目標と今後の進め方
  2. 柴田勝征：“・・・して・・・する”の和文英訳を実装する3つの方法
  3. 衛藤純司：論理文法試案
  4. 滝沢直宏：現代英文における定型表現および構文の研究
  5. 池田尚志：日本語テキストから手紙への機械翻訳に関する試み
  6. 白井諭：英語を解する任意言語間対訳辞書の自動作成
  7. 中岩浩巳：システムと操作者の対話に基づく翻訳例一般化手法の評価
  8. 中岩浩巳：COLING参加報告
- 

平成16年3月30日 H15年度末報告会 電通共催生協会館（東京）

---

1. 池原悟：全体の研究状況報告
  2. 宮崎正弘：グループの状況報告 等価変換システム
  3. 乾健太郎：グループの状況報告 言語生成システム
  4. 池田尚志：グループの状況報告 意味類型知識ベース
  5. 衛藤純司：意味類型と日英翻訳への応用
  6. 徳久雅人：重文複文に対する文型パターン辞書のカバー率
  7. 阿部さつき：意味レベル文型パターン化の方法について（1）
  8. 竹内奈央：意味レベル文型パターン化の方法について（2）
- 

平成16年3月5-6日 第4回言語知識ベース班会議 福岡大学セミナーハウス（福岡県）

---

1. 池原悟：2005年度CREST研究計画と展望
  2. 佐良木昌：シテ形接続の意味分類体系試案
  3. 佐良木昌：シテ形接続パターン作成要領について
  4. 柴田勝征：“：.して...する”の和文英訳の3つの実装方法（その2）
  5. 白井諭：入力文と結合価パターン対辞書の照合に関する一手法
  6. 村上仁一：単文文型パターン化予備実験結果と多変量解析の文型パターン選択
  7. 池原悟：日本語単文の意味分類体系の検討
- 

平成15年3月29-30日 言語知識ベース研究会 日本教育会館（東京）

---

1. 中岩浩巳：動詞意味属性体系について
2. 佐良木昌：因果関係表現の分類と選択基準
3. 宮崎正弘：品詞別シソーラスの作成
4. 武本裕：Webブラウザから利用する意味属性体系検索ツール
5. 武本裕：パターン対を用いた線形・非線形融合型日英構造変換方式
6. 池原悟：言語における対象認識の概念化と機械翻訳
7. 池原悟：話者の認識を媒介する概念と言語表現の意味類型について
8. 森本英之：言語表現の非線形性とは

---

平成15年3月22日 H14年度末報告会 NTT弓町クラブ（東京）

---

1. 池原悟：研究全体の状況報告
2. 宮崎正弘：グループの状況報告 等価変換システム
3. 藤田和寛：グループの状況報告 言語生成システム
4. 池田尚志：グループの状況報告 意味類型知識ベース
5. 阿部さつき：言語KB編纂の方法と体制（1）
6. 竹内奈央：言語KB編纂の方法と体制（2）
7. 村上仁一：対訳コーパスの収集状況
8. 徳久雅人：文型パターン化の方法と文型照合方式
9. 衛藤純司：意味類型化方法の例題検討状況

---

平成15年3月20-21日 「認識・言語・表現」研究会 パノラマイン山中湖（山梨県）

---

1. 佐良木昌：述語の論理と発見の論理 - 西田幾多郎・シルバニアリティ・市川亀久彌
2. 池原悟：言語における対象認識の概念化と翻訳の原理について
3. 安井剛, 徳久雅人, 池原悟, 村上仁一：比喩の自動判定に向けて - 既存の比喩文との意味ベクトル比較による場合
4. 佐伯美香：副詞と形容詞に着目した日本文の感情表現性の調査
5. 池原悟：最近の認知言語学と言語過程説
6. 渡邊南人, 宮崎正弘：The Description of Core Concepts of Basic Words in Japanese and English based on their Recognition Primitives
7. 池田尚志：機械翻訳プロジェクトjaw、解析システム ibuki、自動点訳システムibukiTen
8. 衛藤純司：意味類型の構築に向けて - 重文の意味類型

---

平成15年3月13-14日 第3回言語知識ベース班会議 城崎荘（兵庫県）

---

1. 佐良木昌：KB班研究会議プログラム
2. 池原悟：文型パターン辞書の配列の順序と目次の構成に関する1案
3. 宮崎正弘：品詞別シソーラス（名詞・動詞・形容詞・副詞）の意味類型としての適用性
4. 衛藤純司：意味類型体系と日英翻訳への適用例
5. 佐良木昌：して型接続の意味分類とBLOG日英複数対訳コメント
6. 白井諭：従属句の意味と係り受け
7. 徳久雅人：基礎的重複文構造の含有率についてのトップダウン調査
8. 池田尚志：複文の意味関係に関する考察

---

平成14年11月30-12月1日 「認識・言語・表現」研究会 NTT弓町クラブ（東京）

---

1. 池原悟：認識と言語表現を媒介する概念と意味類型について
2. 村上仁一：英日対訳データベースの現状
3. 村上仁一, 池原悟, 徳久雅人：日本語英語の文対応データベースの作成
4. 衛藤純司：意味類型論の諸問題
5. 柴田勝征：US式和英翻訳システムのパターン実装実験
6. 九津見毅：英日MTにおける代名詞翻訳の改善
7. 徳久雅人：タスク指向対話のための情緒推定 - 勧誘の場合
8. 渡邊南人：英和辞書の諸問題

---

平成14年11月9-10日 言語知識ベース研究会 日本教育会館（東京）

---

1. 池原悟：等価的類推思考の原理による機械翻訳方式
2. 徳久雅人：日英文型パターンにおける節の変数化条件について
3. 池原悟：意味類型と類推思考の原理に基づく言語の等価的変換と生成技術の研究
4. 衛藤純司：意味類型記述の諸問題 / 意味類型の記述 / 論理範疇リスト
5. 宮崎正弘：科研費研究の計画について
6. 川辺諭：日英文型パターン記述言語に関して
7. 宮崎正弘：実装中の日英翻訳実験システムの構成
8. 柴田勝征：US式和英翻訳システム。CRESTデモ版
9. 森本英之：CRESTのあり方についての私見

---

平成14年8月21-22日 第2回言語知識ベース班会議 プランナール三朝（鳥取県）

---

1. 衛藤純司：意味類型・表現パターンに関する例文分析（中間報告）
2. 佐良木昌：対訳例文の分析報告
3. 新田義彦：シテ形接続をめぐって
4. 柴田勝征：US方式翻訳システムによる例文の翻訳パターンの実装メモ
5. 川辺諭、竹本裕、宮崎 正弘：日英コーパス加工法の検討
6. 池原悟：英語文型パターンの選択方法に関する問題
7. 池原悟：日英文型パターンの記述言語と作成基準
8. 池原悟：日英パターン評価のための実験システムについて
9. 徳久雅人：日本語パターンパーキングの机上シミュレーション

---

平成14年3月29-30日 言語知識ベース研究会 日本教育会館（東京）

---

1. 池原悟：平成14年度研究計画大綱
2. 宮崎正弘：平成14年度翻訳版研究計画
3. 池田尚志：平成14年度知識ベース班研究計画
4. 衛藤純司：意味類型を抽出するためのパターン化作業の検討
5. 佐良木昌：意味類型例題検討資料の作成試案
6. 宮崎正弘：意味類型としての概念ソーラスの半自動構築法
7. 新田義彦：意味類型パターンとその変換方法

---

平成14年3月16日 第1回言語知識ベース班会議 大垣ソフトピア（愛知県）

---

1. 佐良木昌：意味類型パターン化作業の基礎作業としての知識記述に向けて
2. 衛藤純司：意味類型を抽出するためのパターン化作業の点検

---

平成14年1月26日 CREST研究プロジェクト発足会 第一オカモトヤビル（東京）

---

1. 田中穂積：内外の情勢：言語処理と言語資源について
2. 池原悟：研究提案の経緯と研究テーマの概要
3. 池原悟：研究の全体計画と当面の計画
4. 宮崎正弘：最近の研究状況 等価変換システム班
5. 奥村学：最近の研究状況 言語生成システム班
6. 乾健太郎：最近の研究状況 言語生成システム班
7. 池田尚志：最近の研究状況 知識ベース班

---

平成13年12月8-9日 言語知識ベース研究会 電通共催生協会館（東京）

---

1. 宮崎正弘：H13年度科研費研究総括と来年度計画
2. 池原悟：戦略的基盤研究の計画大綱
3. 池原悟：類推思考の原理による言語の意味的等価変換方式の提案
4. 新田義彦：意味類型パターン時靴言語の検討課題
5. 柴田勝征：英和对訳文例の実装化とパターンの抽出について



6. 佐藤暢雄：情報構造と表現へ意識の多様性
7. 乾健太郎：言い換えエンジンKURAの概要と今後の発展の方向性
8. 衛藤純司：言い換えエンジンKURA用データの作成

---

平成13年8月29-30日 言語知識ベース研究会 トークイン琵琶湖（滋賀県）

---

1. 池田尚志：意味類型記述言語の設計状況
2. 柴田勝征：英語表現構造パターンの抽出状況
3. 新田義彦：表現構造パターン記述言語の設計状況
4. 池原悟：日本語表現構造パターンの抽出状況
5. 池原悟：等価変換理論の計算論的検討
6. 斉藤健太郎：意味類型に基づく比較構文のパターン翻訳方式
7. 衛藤純司：対訳DBと大域タグの検討

---

平成13年8月28日 「認識・言語・表現」研究会 トークイン琵琶湖（滋賀県）

---

1. 池原悟：自然言語処理の基本問題への挑戦
2. 佐良木昌：市川亀久弥の類推の理論と等価変換理論
3. 宮崎正弘：話者の認識と概念の工学的扱い—より高度な意味処理に向けて
4. 荻野孝野：深層格パターンの検討
5. 的場和幸：日英時間表現の意味と対応関係の解析
6. 廣瀬文男、池原悟、村上仁一：日本語の位置的关系表現の意味解析
7. 佐良木昌：翻訳の方法
8. 池原悟：セマンティック・タイポロジーによる翻訳の原理

---

平成13年6月9-10日 言語知識ベース研究会 NTT弓町クラブ(東京)

---

1. 池原悟，宮崎正弘，佐良木昌：セマンティック・タイポロジーによる言語の等価変換と生成方式の研究
2. 金出地：結合価パターン辞書を用いた動詞の訳語選択の精度
3. 谷口雄介：日本語の構造的類似文検索アルゴリズムの研究
4. 白井諭：日本語の”格要素＋動詞”と英語の動詞1語との対応付け

6.4 特許出願（国内 0件、海外 0件、その他 0件）

なし

6.5 受賞等

6.5.1 受賞

（1）テレコムシステム技術賞 平成14年3月25日受賞

（受賞者）池原悟、宮崎正弘、白井諭、横尾昭男、中岩浩巳、小倉健太郎、  
大山芳文、林良彦

（2）情報処理学会中国支部功績賞 平成16年5月14日受賞

（受賞者）池原悟

（8）鳥取大学学長表彰・研究功績賞 平成17年2月21日受賞

（受賞者）池原悟

（4）人工知能学会業績賞 平成18年6月8日受賞

（受賞者）池原悟

## 6.5.2 新聞報道

### 機械翻訳の新技術研究

日本海新聞、平成13年11月15日

### 仏の日本語研究者が講演

日本海新聞、平成14年5月15日

### コンピュータ翻訳などを評価

日本海新聞、平成18年7月25日

### 池原教授 人工知能学会業績賞受賞

朝日新聞、平成18年8月??日

## 6.5.3 その他

なし

## 6.6 その他特記事項

本研究の成果を実装したシステムはまだ無いが、機械翻訳システムのメーカにおいて、重文パターンの一部を切り出して利用するための検討が行われている。また、情報検索に関する研究プロジェクトにおいて、意味的な情報検索システムに応用するため、導入される見込みとなっている。

なお、本研究の成果は、言語学、言語教育、および翻訳家養成などでの利用が見込まれるため、図書版として編集し出版する計画である。但し、このままの規模で図書化することは難しいため、パターンを洗練して全体的に圧縮するための検討を実施中である。

## 7. 研究期間中の主な活動

### 7.1 ワークショップ・シンポジウム

年月日	名称	場所	参加者	概要
平成14年 1月26日	CRES 研究プロジェクト研究企画会議	東京オカモトビル	30名	研究計画全般の提案と討論(発表6件)
平成13年 8月??日~8月??日	「言語・認識・表現」研究会	トークイン琵琶湖	28名	次世代の機械翻訳方式と言語知識ベース(発表11件+パネル)
平成13年 6月9日~6月10日	表現意味辞書研究会	NTT弓町クラブ	20名	ドイツ語学者有田潤先生の講演とその他の研究発表(講演+発表6件)
平成13年 8月28日~8月30日	表現意味辞書研究会	トークイン琵琶湖	15名	表現意味辞書関連研究の発表と討論(発表9件)
平成13年 12月8日~12月9日	表現意味辞書研究会	電通共済生協会館	14名	表現意味辞書関連研究の発表と討論(発表8件)
平成14年 3月29日~3月30日	表現意味辞書研究会	日本教育会館	16名	芝原宏治先生の講演(言語とアナロジーの問題)と研究発表(講演+発表7件)
平成14年 11月30日~12月1日	「言語・認識・表現」研究会	NTT弓町クラブ	23名	宮下真二の英語学に関連する発表ほか(発表13件)
平成14年 11月9日~11月10日	表現意味辞書研究会	日本教育会館	15名	表現意味辞書関連研究の発表と討論(発表17件)
平成15年 3月22日	CREST 研究年度末報告会	NTT弓町クラブ	35名	平成14年度の研究概要の発表と次年度計画(発表9件)
平成15年 3月29日~3月30日	表現意味辞書研究会	日本教育会館	15名	表現意味辞書関連研究の発表と討論(発表10件)

平成15年 9月20日～9月21日	「言語・認識・表現」 研究会	パラマイン 山中湖	23名	言語本質論、感性表現などに関連 する発表(発表9件)
平成16年 3月30日	CREST 研究年度末報告会	電通共済 生協会館	32名	平成15年度の研究概要の発表と 次年度計画(発表6件)
平成16年 9月11日～9月13日	表現意味辞書 研究会	久住高原 やまなみ荘	16名	滝沢先生の講演(コーパス言語学)と 意味分類体系の研究発表など (講演+発表8件)
平成16年 12月18日～12月19日	「言語・認識・表現」 研究会	電通共済 生協会館	30名	亀井秀雄先生の講演(三浦つとむの 広がり)ほか(講演+発表9件)
平成17年 12月24日～12月25日	「言語・認識・表現」 研究会	一碧荘 (熱海)	16名	言語学研究の発表など (発表6件+パネル討論)
平成18年 3月21日	CREST 研究年度末報告会	日本教育 会館	40名	平成17年度の研究概要の発表と 次年度計画(発表10件)
平成18年 12月9日～12月10日	「言語・認識・表現」 研究会	東京	?名	言語学研究の発表など (発表13件)
平成19年 ?月?日	CREST 研究年度末報告会		?名	平成18年度の研究概要の最終総括 (発表?件)

## 7.2 技術会議など

年度	開催月日	会議名	報告議題数	場所(注)	参加者
平成 14年度	3月16日	第1回知識ベース班会議	議題 3件	大垣ソビア	5名
	4月20日	第1回 進捗会議	報告 1件	川崎会議室	14名
		第1回 技術会議	議題 7件		
	5月17日	第2回 進捗会議	報告 1件	川崎会議室	14名
		第2回 技術会議	議題 2件		
	6月1日	第3回 進捗会議	報告 5件	川崎会議室	16名
		第3回 技術会議	議題 5件		
	6月21日	第4回 進捗会議	報告 2件	川崎会議室	15名
		第4回 技術会議	議題 13件		
	7月6日	第5回 進捗会議	報告 4件	川崎会議室	14名
		第5回 技術会議	議題 5件		
	7月19日	第6回 進捗会議	報告 3件	川崎会議室	15名
		第6回 技術会議	議題 10件		
	8月3日	第7回 進捗会議	報告 3件	川崎会議室	17名
		第7回 技術会議	議題 6件		
	8月21～22日	第2回知識ベース班会議	議題 10件	フロンタル三朝	18名
	8月27日	第8回 進捗会議	報告 3件	川崎会議室	13名
		第8回 技術会議	議題 6件		
	9月7日	第9回 進捗会議	報告 3件	川崎会議室	16名
		第9回 技術会議	議題 4件		
	10月26日	臨時 進捗会議	報告 2件	川崎会議室	16名
		臨時 技術会議	議題 5件		
	11月16日	第10回 進捗会議	報告 3件	川崎会議室	14名
		第10回 技術会議	議題 8件		
12月7日	第11回 進捗会議	報告 2件	川崎会議室	14名	
	第11回 技術会議	議題 5件			
12月20日	第12回 進捗会議	報告 2件	川崎会議室	18名	
	第12回 技術会議	議題 7件			

	1月25日	第13回 進捗会議	報告 2件	川崎会議室	16名
		第13回 技術会議	議題 5件		
	2月15日	第14回 進捗会議	報告 3件	川崎会議室	16名
		第14回 技術会議	議題 6件		
平成 15年度	4月26日	第1回 進捗会議	報告 3件	川崎会議室	17名
		第1回 技術会議	議題 9件		
	5月24日	第2回 進捗会議	報告 2件	川崎会議室	18名
		第2回 技術会議	議題 8件		
	6月21日	第3回 進捗会議	報告 2件	川崎会議室	18名
		第3回 技術会議	議題 10件		
	7月19日	第4回 進捗会議	報告 2件	川崎会議室	16名
		第4回 技術会議	議題 8件		
	9月6日	第5回 進捗会議	報告 4件	川崎会議室	17名
		第5回 技術会議	議題 6件		
	10月9日	臨時企画会議	議題 3件	川崎会議室	5名
	11月1日	第6回 進捗会議	報告 2件	川崎会議室	13名
		第6回 技術会議	議題 11件		
	11月20日	臨時企画会議	議題 2件	川崎会議室	5名
	11月29日	第7回 進捗会議	報告 2件	川崎会議室	15名
		第7回 技術会議	議題 5件		
	12月20日	第8回 進捗会議	報告 2件	川崎会議室	18名
		第8回 技術会議	議題 5件		
	1月24日	第9回 進捗会議	報告 2件	川崎会議室	19名
		第9回 技術会議	議題 8件		
2月21日	第10回 進捗会議	報告 2件	川崎会議室	15名	
	第10回 技術会議	議題 4件			
3月13~14日	第3回知識^ - 2班会議	議題 7件	城崎荘	18名	
平成 16年度	5月15日	第1回 進捗会議	報告 1件	川崎会議室	19名
		第1回 技術会議	議題 5件		
	6月19日	第2回 進捗会議	報告 1件	川崎会議室	17名
		第2回 技術会議	議題 11件		
	7月21日	第3回 進捗会議	報告 2件	川崎会議室	19名
		第3回 技術会議	議題 10件		
	8月21日	第4回 進捗会議	報告 3件	川崎会議室	14名
		第4回 技術会議	議題 4件		
	9月18日	第5回 進捗会議	報告 4件	川崎会議室	18名
		第5回 技術会議	議題 8件		
	10月16日	第6回 進捗会議	報告 3件	川崎会議室	16名
		第6回 技術会議	議題 5件		
	11月27日	第7回 進捗会議	報告 5件	川崎会議室	16名
		第7回 技術会議	議題 20件		
	12月11日	第8回 進捗会議	報告 4件	川崎会議室	15名
		第8回 技術会議	議題 9件		
	1月29日	第9回 進捗会議	報告 4件	川崎会議室	17名
		第9回 技術会議	議題 8件		
	2月19日	第10回 進捗会議	報告 2件	川崎会議室	15名

		第10回 技術会議	議題 5件		
	3月26日	第11回 進捗会議	報告 3件	川崎会議室	16名
		第11回 技術会議	議題 5件		
	3月5~6日	第4回知識^ - 2班会議	議題 6件	福岡大	14名
平成 17年度	5月28日	第1回 進捗会議	報告 3件	川崎会議室	17名
		第1回 技術会議	議題 1件		
	6月15日	臨時企画会議	議題 1件	川崎会議室	7名
	7月12日	第2回 進捗会議	報告 2件	川崎会議室	17名
		第2回 技術会議	議題 11件		
	9月17日	第3回 進捗会議	報告 3件	川崎会議室	15名
		第3回 技術会議	議題 4件		
	11月12日	第4回 進捗会議	報告 2件	川崎会議室	14名
		第4回 技術会議	議題 7件		
	12月17日	第5回 進捗会議	報告 3件	川崎会議室	14名
		第5回 技術会議	議題 8件		
	12月24~25日	第5回知識^ - 2班会議	議題 4件	一碧荘(熱海)	5名
	1月28日	第6回 進捗会議	報告 3件	川崎会議室	18名
		第6回 技術会議	議題 9件		
	2月18日	第7回 進捗会議	報告 4件	川崎会議室	16名
		第7回 技術会議	議題 4件		
3月18日	第8回 進捗会議	報告 4件	川崎会議室	14名	
	第8回 技術会議	議題 7件			
平成 18年度	4月22日	第1回 進捗会議	報告 5件	川崎会議室	16名
		第1回 技術会議	議題 7件		
	6月24日	第2回 進捗会議	報告 3件	川崎会議室	16名
		第2回 技術会議	議題 12件		
	7月29日	第3回 進捗会議	報告 4件	川崎会議室	16名
		第3回 技術会議	議題 8件		
	8月26日	第4回 進捗会議	報告 4件	川崎会議室	18名
		第4回 技術会議	議題 6件		
	9月23日	第5回 進捗会議	報告 4件	川崎会議室	16名
		第5回 技術会議	議題 8件		
	10月22日	第6回 進捗会議	報告 5件	川崎会議室	15名
		第6回 技術会議	議題 5件		
	11月11日	第7回 進捗会議	報告 8件	川崎会議室	16名
		第7回 技術会議	議題 8件		
12月2日	第8回 進捗会議	報告 件	川崎会議室	名	
	第8回 技術会議	議題 件			
1月??日	第9回 進捗会議	報告 件	川崎会議室	名	
	第9回 技術会議	議題 件			
- -	合計	進捗会議 52回 技術会議 53回 その他会議 7回	報告 151件 議題 397件		

(注) 川崎会議室：本研究で借用した東京実験室のビル内の会議室

## 8 . 結び

言語表現では、個々の単語が意味を持っているだけでなく、表現の構造も意味を持っている。この問題を扱う方法が分からなければ、結局、本格的な意味処理は実現できないだろう。この問題は、自然言語処理の研究を開始した当初から抱いていた問題である。この研究プロジェクトは、この問題に挑戦するための願ってもないチャンスであった。

今まで、実用的な自然言語処理の実現を目指してさまざまなシステムの研究開発を行う中で、この問題への取り組みを続けてきた。はじめの取り組みは、単文構造の持つ意味を結合価パターンの形式で定義する方法である。日本語用言6000語の持つ表現構造の意味を17,000パターンで記述することにより、単文構造の持つ意味は、ほぼ捉えられるようになった（「日本語語彙大系」として1997に出版）。

しかし、それだけでは片手落ちであり、日本語表現の意味を捉えるには、重文や複文の意味が捉えられるようにすることが必要である。単文は、事象、現象に対する話者の認識を表す枠組みであるのに対して、重文複文は、現象や事象間の関係に関する認識を表す枠組みである。両者がそろふことによって、ようやく本格的な言語表現の意味処理が可能となる。

また、日本語語彙大系は、表現としての意味を掬い取るという点で大きな問題を残していた。それは、結合価パターン方式では、助詞や助動詞を用いて表される主体的表現の部分は分離して処理しなければならないため、それらの持つ非線形な意味が失われるという問題である。この問題を解決し、重文複文を対象とした表現意味辞書を研究開発したいと考え方式的な検討をしている過程で、本プロジェクト実施の機会が与えられた。

研究を開始するに当たって、多くの方々のご協力を頂き、研究計画を作成していたが、開始の直後から大きな変更が発生した。そのため、計画線表を急遽見直したが、予想以上の試行錯誤を繰り返すこととなった。その発端は、パターン辞書作成の元となる日英対訳標本文の増加である。当初、それまでに準備してきた20万件の対訳文から重文複文5万件を抽出して使用する予定であったが、実際には、100万件の対訳文を入手することができ、標本として使用できる重文複文は15万件となった。

嬉しい誤算であるが、そのままでは、実行不能な作業量となる恐れがある。作業の大幅な機械化を狙ってさまざまな検討を行い、開発手順も見直した。特に、汎化の程度については、パターン数が予想以上に増大する見込みとなったので、比較的簡単な汎化を行えば、実用上十分な被覆率が得られるだろうと期待し、高度な汎化は見合わせることにした。

これらの結果、作業の機械化は順調に進んだ。大規模なデータを能率良く汎化することができ、文型パターン辞書の初版は順調に作成することができた。しかし、用例数が3倍にも増加したにもかかわらず、被覆率を上げることは予想以上に困難で、その後、さまざまな試行錯誤を繰り返すこととなった。最終的には、様相時制の汎化など、当初見合わせていた汎化だけでなく、汎化のためのさまざまな工夫を行った。これらの汎化は、

当初の予想を遙かに超える内容のものであり、そのため、当初設計していたパターン記述言語の記述能力は大幅に増強され、その仕様書は分厚いものとなった。また、当初、実用性が期待できるパターン辞書の意味的な被覆率として、60%～70%の達成を目標としていたが、最終的な被覆率は80%となった。

振り返ってみると、当初は量の問題に目を奪われていたように思われる。結局のところ、量によって質の問題をカバーすることはできず、量と質が伴って初めて実用的な能力が生まれるということであろう。

なお、実現した被覆率は限界とは言えず、まだ改良の余地を残している。従来、パターン方式では、被覆率の高いパターン辞書を構築することが困難と見られてきたが、被覆率80%のパターン辞書が実現できたこと、また、それ以上の被覆率が実現できる可能性があることから見て、この方式の可能性と実用的価値が明確になったと思われる。

ところで、本文でも述べたが、パターン数と被覆率の関係をみると、被覆率は、数万件のパターン数でほぼ飽和する傾向を持ち、それ以上パターン数を増やしても容易に向上しない。被覆率を「向上させるには、第1に汎化の方法を高度化すること、第2にパターンパーサとの整合性を向上させることが重要である。

これは、20万件のパターンを作成することによって分かったことであり、当初の予定通りの5万件のパターン作成では分らなかったことと思われる。また、主体的表現の意味の問題を解決する方法として本研究で提案した「非線形言語モデル」は、汎化作業の過程で、その価値が試され、意義を確認することができた。

以上の通り、本研究の最大の難関は、汎化の方法と効果の推定、および限られた研究費の中での作業品質確保のための手段と方法にあったと言えるが、予想以上の試行錯誤は無駄ではなく、今後のこの種の研究開発に対して重要な知見を残すことができたと思う。この過程で取り組んだ問題やその対策についての検討資料もまとめて残せるようにしたい。

意味類型化は、この影響を受け、取り組みが遅れた。元々、意味類型化の方法については、試行錯誤を予想していたが、その通りとなった。まず問題となるのは、文型パターンの意味分類体系をいかに構築するか点である。この問題では、「意味類型知識ベースグループ」を中心に、研究開始直後から検討を進めた。しかし、作成した意味分類体系に基づいて、大量のパターンの一つ一つの意味を判断する作業は、容易ではない。さまざまな方法を検討したが、最終的には、意味判定が機械的に支援できるよう、分類体系を見直すことにより、この問題をクリアした。

また、パターン辞書開発の過程で被覆率実験を行うため、パターンパーサを開発した。クロス照合実験（12万入力文×22.7万パターンの総当たりの照合実験）が可能な処理速度を実現すること、頻繁に発生するパターン記述言語仕様の変更に対応できるようにす

ることを条件に、さまざまな工夫を行った結果、高速で安定したパーサを実現することができた。高速化の効果は大きく、処理速度は2000倍以上になった。この結果、1回のクロス照合実験が20～30時間（1入力文当たり1秒以下）で実施できるようになり、実験的改良効果の評価だけでなく、不良パターンの検出や絞り込みでも役立つものとなった。

このパーサは、文型パターン辞書を用いて入力文の意味解析を行うためのものであるが、文型パターンに構文情報を付与しておけば、照合結果から構文解析結果も得られる。従来、解析精度が問題とされてきた構文解析プログラムは不要となるもので、技術的な価値が高い。

以上の通り、紆余曲折は大きかったが、本研究では、当初目的としていた意味類型パターン辞書を実現することができた。また、生産性は、開発した意味類型パターン辞書の規模と研究費から見て、非常に高いものとなった。これは、すでに、さまざまな解析ツールなどの準備ができていたこともあるが、研究の過程で多くの新しい方法を考案したこと、スキルを持つ言語アナリストグループに支えられ、それらを組織的に実施することができた効果が大きい。

従来、単語の意味を定義した辞書は多いが、表現の意味を定義した辞書は、世界にも例を見ない。既に開発済みの「単文の表現意味辞書（日本語語彙大系）」と組み合わせれば、日本語表現のほぼ全体がカバーされる。日本語意味処理の基盤技術として、さまざまな応用が期待される。

なお、この研究では、研究メンバーの方々のほかにも、「言語・認識・表現研究会LACE」や「表現意味辞書研究会」などを通じて、多くの方々のご協力を頂いた。方式についての理論的な基礎は、これらの研究会での議論によって固められた。また、企業への試作発注を通じて、多くの言語アナリストの方々にご協力を頂いた。膨大で緻密な作業は、これらの方々の方なしにはなし得なかったものである。終わりに、これらご協力を頂いた多くの方々に深謝する。