

# 非線形言語モデルに基づく 意味的等価変換方式の実現に向けて

日本語重文複文に対する  
意味類型パターン辞書の研究開発

平成19年3月31日

池原 悟(鳥取大学)

## 全体の目次

第1章	研究の背景と手順	11
第2章	非線形言語モデルと文型パターン	29
第3章	パターン記述言語の設計	74
第4章	重文複文の文型パターン辞書の開発	121
第5章	意味的等価変換方式と文型パターンの意味類型化	183
第6章	パターン検索プログラム	224
第7章	適合パターンの絞り込みと翻訳への適用例	261
第8章	研究成果と今後の展望	283

## 全体の詳細目次

第1章	研究の背景と手順	11
1.1	意味処理と機械翻訳	12
	(1)従来の意味処理技術 (2)要素合成法からの脱却	
1.2	構造と意味の問題への挑戦	18
1.3	本研究の狙いと内容	21
1.4	研究の体制	23
1.5	研究の手順	25
1.6	パターン辞書の内容	27
第2章	非線形言語モデルと表現パターン	29
2.1	言語表現とその意味	30
	(1)言語表現の本質的特徴 (2)意味の定義 (3)意味と概念と (4)言語表現と概念の関係 (5)言語表現から見た概念の種類 (6)表現構造の持つ意味	
2.2	非線形言語モデル	51
	(1)言語表現の線形性と非線形性 (2)表現構造の意味の記述法 (3)線形要素の重要な特徴 (4)非線形言語モデル	
2.3	非線形表現構造のパターン化	69
2.4	第2章のまとめ	73

第3章	パターン記述言語の設計	76
3.1	パターンの役割と対象	77
3.2	パターンの設計条件	81
3.3	パターン記述要素の構成	83
	(1)記述要素の種類 (2)字面による記述、(3)変数による記述 (4)関数による記述、(4)記号による記述	
3.4	単語意味属性体系の構築	112
	(1)意味属性体系と意味辞書の種類 (2)用言意味属性体系 (3)名詞の意味属性体系	
3.5	第3章のまとめ	122
第4章	重文複文の文型パターン辞書の開発	123
4.1	重文と複文のパターン化	124
	(1)パターン化の対象 (2)文型パターンの生成手順 (3)単語レベル文型パターン化 (4)句と節レベルの文型パターン化	
4.2	文型パターン辞書の実験的改良	139
	(1)離散記号の付与基準の精密化 (2)述部語尾表現(特に時制)の汎化 (3)英語文型パターンの改良 (4)その他の主な改良	
4.3	文型パターンの記述例	151
4.4	静的特性の評価	156
	(1)文型パターン数 (2)パターン構成要素の分析 (3)線形要素の割合	

4.5	被覆率特性の評価	165
	(1)実験の条件 (2)評価パラメータ (3)統語的被覆率の飽和特性	
	(4)適合率と適合パターン数 (5)適合率R1と正解率P1・P2	
4.6	第4章のまとめ	182
<b>第5章 意味的等価変換方式と</b>		
<b>文型パターンの意味類型化</b> ……………186		
5.1	意味的等価変換方式の概要	187
	(1)理論的背景 (2)意味的等価変換方式	
	(3)本方式とパターン翻訳の違い	
5.2	重文・複文の意味類型化	205
	(1)真理項の設計条件 (2)真理項の構成、(3)統語構造による分類	
	(4)節間の意味分類体系、(5)節の意味分類 (6)分類コードの付与	
	(7)「統語・意味分類コード」の例	
5.3	日英構文対応関係の分析	221
5.4	第5章のまとめ	226
<b>第6章 文型パターン検索プログラム</b> ……………227		
6.1	パターンパーサ	228
	(1)機能条件と設計条件 (2)パターン記述要素の種類	
	(3)パターン照合基本機能 (4)パターン照合基本方式	

	(5)照合速度向上の方法 (6)パターンパーサの実装と評価	
	(7)高速化手法の評価 (8)パターン検索の例	
6.2	パターン意味検索プログラム	256
	(1)機能条件と設計条件 (2)意味検索プログラムの構成	
	(3)意味検索実行例	
6.3	第6章のまとめ	261
<b>第7章 適合パターンの絞り込みと</b>		
<b>翻訳への適用例</b> ……………264		
7.1	パターン検索結果の絞り込み	265
	(1)実験の目的と実験方法 (2)絞り込み実験の結果	
	(3)検索結果の例	
7.2	パターン翻訳実験の例	276
	(1)評価の狙いと方法 (2)パターン翻訳実験システムの試作	
	(3)実験結果と翻訳例	
7.3	意味的等価変換方式の効果	281
	(1)評価の目的と方法 (2)評価結果	
7.4	第7章のまとめ	285
<b>第8章 研究成果と今後の展望</b> ……………286		
	(1)研究の到達点 (2)生産物 (3)残された課題	
	(4)応用研究への期待	

## 補足説明

【補足説明1】	従来の言語意味論の分類	34
【補足説明2】	意味解析と意味理解	38
【閑話休題3】	意味を別の言語で定義する理由	59
【補足説明4】	従来の線形言語モデル	67
【補足説明5】	構文解析技術は不要となるか?	72
【補足説明6】	意味制約条件とは?	91
【補足説明7】	3段階に汎化する理由	132
【補足説明8】	自己例文・自己パターンとは?	149
【補足説明9】	パターンがカバーする入力文	155
【補足説明10】	クロスヴァリデーションとは?	165
【補足説明11】	パターン数と被覆率	180
【補足説明12】	関数と記号の汎化効果	181
【補足説明13】	文型パターン辞書圧縮の可能性	183
【補足説明14】	文型パターン辞書の圧縮	184
【補足説明15】	意味類型とは?	188
【補足説明16】	人間による翻訳のプロセス	199
【補足説明17】	方式限界について	204
【補足説明18】	意味類型化作業の困難さ	225
【補足説明19】	パーサ照合能力の限界	256
【補足説明20】	日英機械翻訳の2方式	263
【補足説明21】	日本文書き換えの2方式	264

## 閑話休題

<閑話休題1>	最近の統計翻訳	16
<閑話休題2>	機械翻訳方式の行き先は?	17
<閑話休題3>	「構造と意味」に関する研究	20
<閑話休題4>	概念と構造(認知言語学の場合)	43
<閑話休題5>	対象認識の言語依存性	49
<閑話休題6>	対象認識の個別性と共通性	50
<閑話休題7>	非線形な言語表現の必然性	52
<閑話休題8>	言語と複雑系の共通性	53
<閑話休題9>	線形性に関する従来の説明	54
<閑話休題10>	子供の言語獲得とパターン	68
<閑話休題11>	プラトンの問題(刺激の貧困)	74
<閑話休題12>	プラトンの問題への解答	75
<閑話休題13>	用言変数化の必要性	93
<閑話休題14>	非線形なパターン数は?	173
<閑話休題15>	大切なのは量か質か?	185
<閑話休題16>	市川亀久弥の研究活動	189
<閑話休題17>	共通見地 について	192

## 本研究の推進母体

事業主： 独立行政法人 科学技術振興機構 (JST)  
事業名： 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」  
<研究総括> 長尾 真 (独立行政法人 情報通信研究機構 理事長)  
<領域アドバイザー>  
植村俊亮 (奈良先端大学) 牛島和夫 (九州産業大学)  
後藤敏 (早稲田大学) 坂内正夫 (国立情報学研究所)  
諏訪 基 (身障者リハビリセンター研究所)  
所 眞理雄 (ソニー) 松田晃一 (NTT-AT IPシェアリング)

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)

9

## 研究課題

言語表現に用いられる対象認識の形式をあらかじめ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

### 具体的課題と実現の方法

#	具体的課題	実現の方法
課題1	日本語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	対訳コーパスを対象に線形要素をさまざまな方法で汎化し、文型パターン化する。 <文型パターン辞書の研究開発>
課題2	日本語重文複文に意味的に対応する複数の英語表現が発見できる仕組みを実現する	上記パターンを意味類型化し、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 <意味類型パターン辞書の研究開発>

10

## 第1章 研究の背景と手順

### 第1章の目次

- 1.1 意味処理と機械翻訳
- 1.2 構造と意味の問題への挑戦
- 1.3 本研究の狙いと内容
- 1.4 研究の体制
- 1.5 研究の手順
- 1.6 パターン辞書の内容

11

## 1.1 意味処理と機械翻訳

### (1) 従来の意味処理技術

単語の場合・・・多義語の語義解析

表現の場合・・・要素合成法 (単語の意味から合成)

<前提> Compositional Semantics

全体の意味は部分の意味から説明できる

### 問題点

- ・ 非線形 (Non-Compositional) な言語表現が多い
- ・ 単語分解の過程で全体の意味が失われる

## 従来の代表的機械翻訳方式

### ピボット方式

- <方法> 万国共通な普遍言語を仲介して翻訳
- <利点> 多言語翻訳に有利
- <欠点> 普遍言語が設計できない  
現実には、PIVOT=既存の自然言語

### トランスファー方式

- <方法> (1) 構文構造を変換する過程(構文処理)  
(2) 要素を翻訳する過程(意味処理)  
(3) 結果の合成
- <利点> 直訳型で見通しが立てやすい
- <欠点> 原文の意味が失われやすい

**表現構造と意味を一体化した処理が必要**

## (2) 要素合成法からの脱却

**要素合成法** = 従来の自然言語処理の方法

- ・ 言語表現の意味は線形であると仮定  
表現全体の意味を部分の意味で説明

要素毎の翻訳 + 結果の合成  
元の意味が再現されない

- ・ **多くの言語表現は非線形**  
従来の言語処理の限界

**目標: 全体の意味から部分の意味が説明できる  
表現全体の意味を掬い取るための仕組みが必要**

14

## 「構造と意味を一体化」した従来の方法

### 初期のパターン翻訳

- 問題点
  - ・ パターン作成コスト(被覆率が問題)
  - ・ パターン間の排他性の確保
- 現状: 100~200パターン(特殊目的用)

### 用例翻訳

- <方法> 対訳用例との類似性に着目
  - ・ 文法的、意味的に類似する要素を置き換えて翻訳
- <利点> 構文解析と意味解析の曖昧性を回避
- <欠点> **置換可能な要素は、用例毎に異なる**  
翻訳時に自動的に判定することは困難

**【解決策】** 予め置換可能な場所を発見しておく方法  
本質的にパターン翻訳と同じ

**再びパターン翻訳へ**

## < 閑話休題 1 > 最近の統計翻訳

暗号処理 音声認識へ適用(成功)  
英仏翻訳へ導入  
日英翻訳へ適用(ATRなど)  
= 特定分野(旅行会話文) =

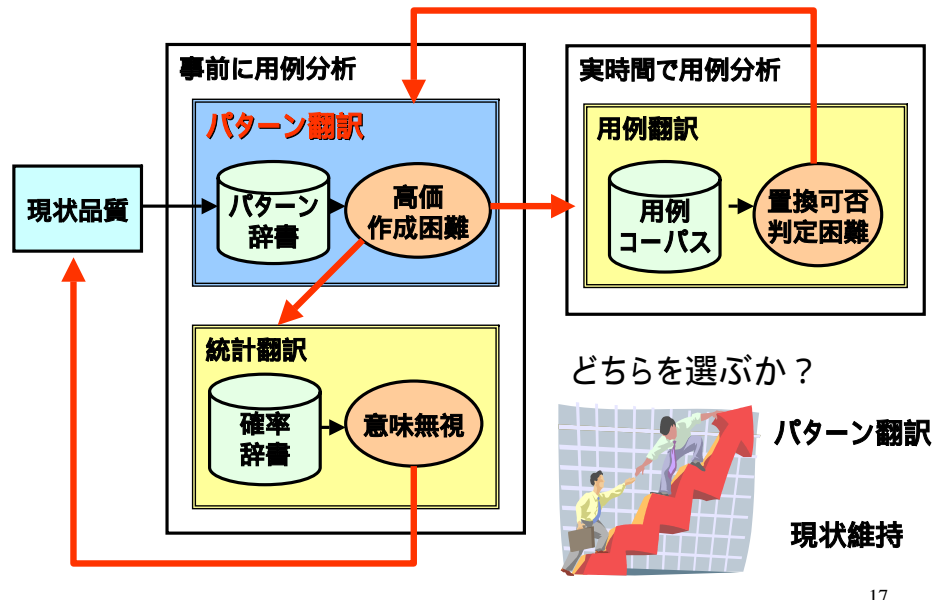
- <利点>
  - システム開発が容易
  - 短期間・低コスト**(人手作業が不要)

- <欠点>
  - × 膨大な対訳例文が必要  
20文字/文の場合 : 4億のバリエーション?
  - × 個別的な誤り対策は不可能 **改良が困難**
  - × 異なる言語族間への不適合性



16

## < 閑話休題 2 > 機械翻訳方式の行き先は？



## 1.2 構造と意味の問題への挑戦

### < 最初の試み > 多段翻訳方式 (池原他1987)

- (1) 客体的表現と主体的表現の**分離と融合**
  - ・客体的表現 = 客体化された対象概念を表現
  - ・主体的表現 = 話者の感情や意志の直接的表現
- (2) **客体的表現**に**結合価パターン**を適用  
表現構造の意味を目的言語に対応づけ

対象: 単文 (動詞と名詞の意味的な関係)

パターン数: 17,000 文型

名詞の意味属性 = **意味的用法**の分類

一般名詞: 2700 種類、固有名詞: 120 種類

名詞意味辞書: 40 万語

「日本語語彙大系」岩波書店1977

18

### < 残された課題 >

- (1) **複文(埋め込み)、重文(接続)の持つ非線形構造が扱えない**

< 表現の3分類 >

表現	現状
名詞句: 実体の表現	×
単文: 事象の表現	×
重文・複文: 事象間関係の表現	×

- (2) **非線形な主体的表現の意味が失われる**  
結合価パターン方式の限界

- (3) **訳し分け不可能**

< 翻訳家の翻訳 >

複数の表現から文脈などにより選択

< 機械による翻訳 >

決定論的な方法で、単一の訳文構造にみ翻訳

### < 閑話休題 3 > 「構造と意味」に関する研究

#### 部分の意味と全体の意味を関連づける方法

- (1) ゲシュタルト心理学 (20世紀初頭)

恒常仮説を否定

構造化の要因と法則を解明

- (2) 認知言語学の試み (Langacker 1987)

**認知文法** (CG: Cognitive Grammar)

生成文法の枠組みを否定

「対象構造が認識に反映する過程」に着目

< 問題点 > 深層格と認知格による意味規定

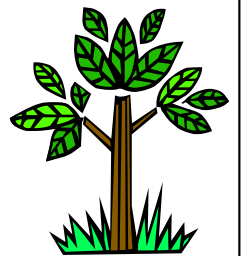
- (3) 構文文法の試み (Fillmore et.al 1988)

**構文文法** (CxG: Construction Grammar)

Lexiconの意味と統語構造をConstructionにより統合

< 問題点 > **部分と全体の意味的關係が不明確**

構造的上の意味の単位が決定できない **工学的適用困難**



20

## 1.3 本研究の狙いと内容

### < 表現意味解析の基本技術の確立 >

【目標】 **非線形な表現の意味を掘り取るための仕組みの実現**

全体の意味から部分の意味が説明できるようにしたい

【方法】 表現のパターン化 = 構造と意味の一体化

**非線形な言語表現に挑戦**

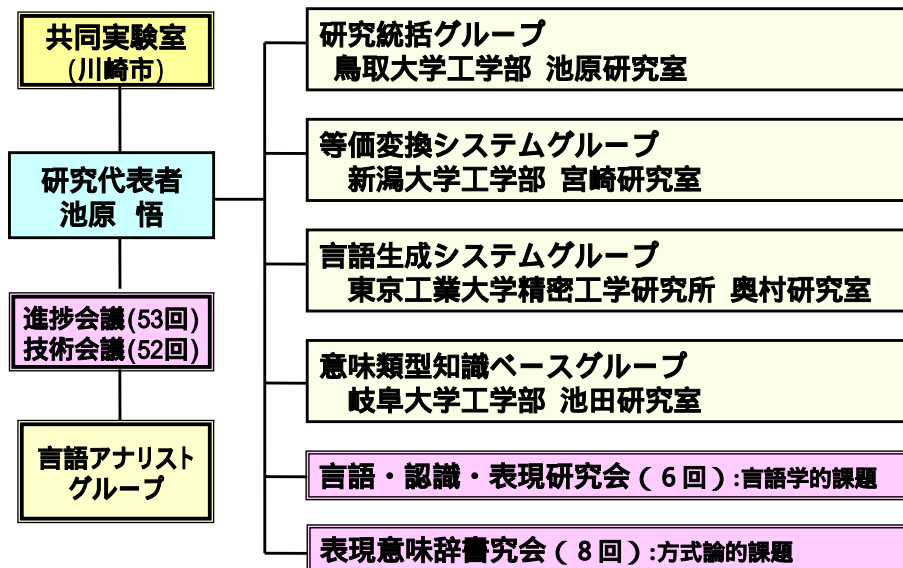
21

## 本研究の内容

- (1) **非線形言語モデル**の提案  
表現構造の持つ意味に着目
- (2) 文型パターン辞書の開発  
対象: **重文**と**複文**
- (3) **意味的等価変換方式**の提案  
言語表現を意味的に等価な別の表現に写像
- (4) **文型パターンの意味類型化**  
すくい取った意味による表現のグループ化

22

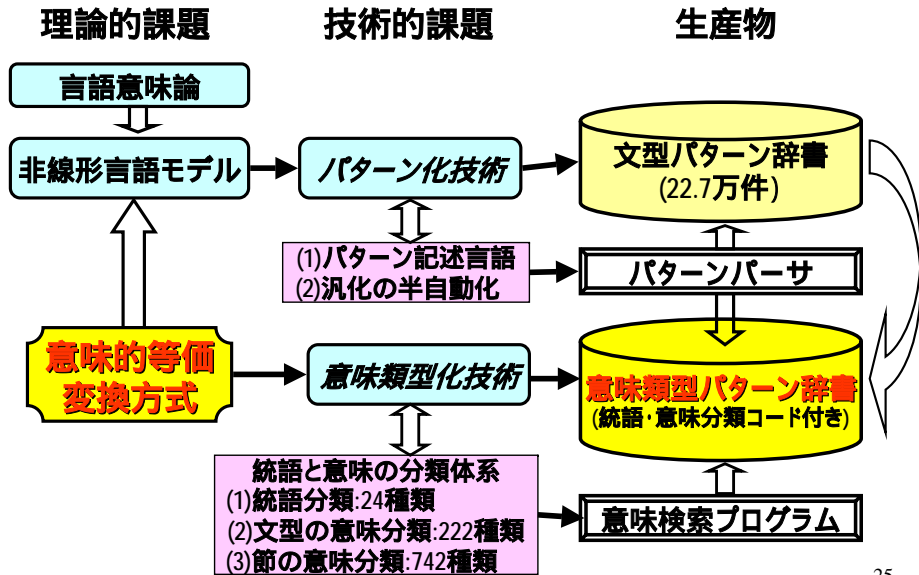
## 1.4 研究の体制



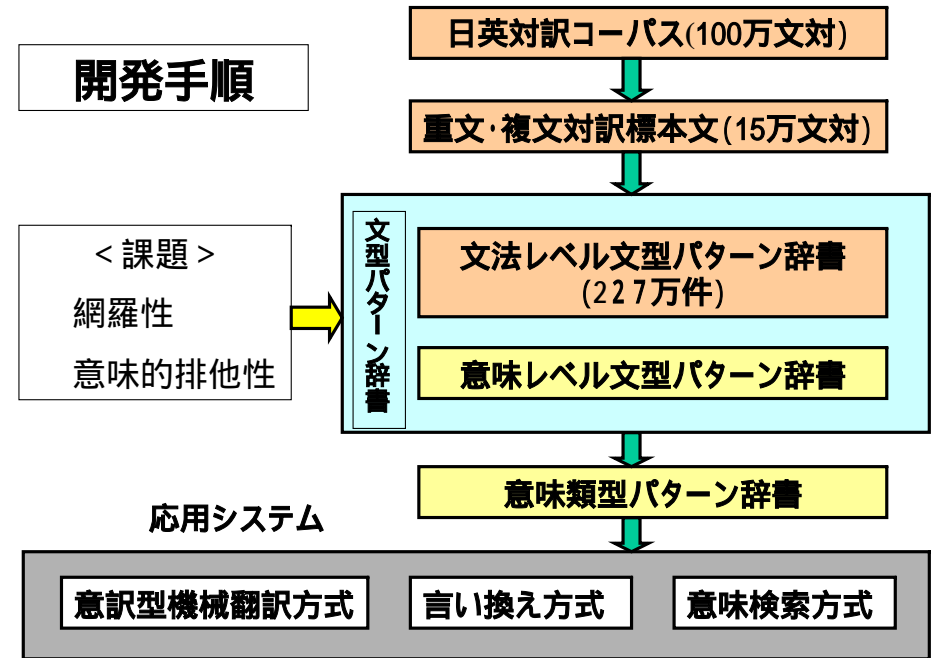
## 会議の開催と学会発表

分類	会議等の種別	開催回数	議題・発表数	参考
研究チーム 部内会議	進捗会議	52回	151件	進捗管理
	技術会議	53回	397件	技術検討会
	KB班会議等	4回	23件	個別設計会議
	年度末報告会等	5回	40件	
	合計	114回	611件	
研究チーム 主催の研究会	認識・言語・表現研究会	6回	44件	LACE(理論)
	表現意味辞書研究会	8回	51件	技術検討
	合計	14回	95件	
学会発表 (対外発表)	論文	--	28件	
	国際会議	--	20件	
	研究会・シンポジウム	--	25件	
	全国大会	--	103件	
	招待講演	--	4件	
	各種著作(図書・解説)	--	10件	
	合計		186件	
JST主催	シンポジウムなど	5回	--	

# 1.5 研究の手順



# 開発手順



# 1.6 パターン辞書の内容

## パターン辞書の構成

(0) 対訳用例...日英対訳文

(1) 日英パターン対

- ・日本語文型パターン + 英語文型パターン
- ・単語レベル / 句レベル / 節レベル

文型パターン

(2) 統語・意味分類コード

- ・統語分類コード
- ・意味分類コード
- ・接続キーワード

意味類型パターン

(3) その他...日本語 / 英語構文分類コードなど

## 意味類型パターンの例

標本文(日) : 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。  
 標本文(英) : The climate here suits me, so I would like to live here in old age.

単(日) : /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)/V7(NY:0200,KR:0003,IY:3210,IY:3240,IY:6970).tai。

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

句(日) : /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,KR:0003,IY:3210,IY:3240,IY:6970).tai。

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

節(日) : /ytcfkCL1(NY:0501,NY:0506,NY:0701,NY:0703,NY:1500,NY:2502,NY:3002,IY:5110,IY:6930,IY:6940).teiru#5(.genzai|.kako)^rentaiので!CL3(NY:0200,IY:3210,IY:3240,IY:6970)。

節(英) : CL1, so CL3^past.

1-副 / 4110,5930,5940-2210,2240,5970 / Fub100 / -ので / 接続詞相当 / so

# 第2章 非線形言語モデル と表現パターン

## 第2章の目次

- 2.1 言語表現とその意味
- 2.2 非線形言語モデル
- 2.3 非線形表現構造のパターン化
- 2.4 第2章のまとめ

## 2.1 言語表現とその意味

非線形性 = 「表現の構造」と「意味」の関係の問題

### 本節の課題

- Q1 言語とは何か？
- Q2 意味とは何か？
- Q3 概念はどのようなものか？
- Q4 言語表現は概念を表すか？
- Q6 表現構造は概念を表すか？

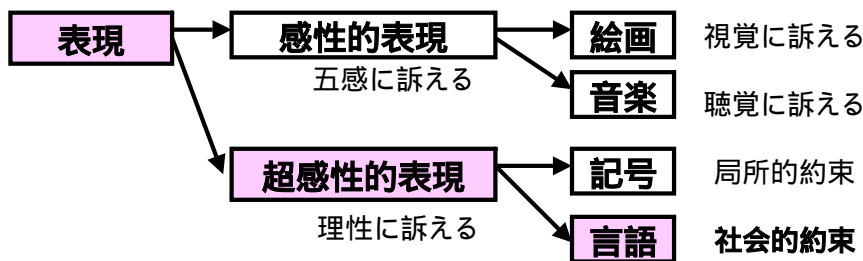
### 結論

表現構造の持つ意味は、それに包含される言語表現に共通する概念、すなわち上位概念である。

## (1) 言語の本質的特徴(三浦つとむ)

### Q1 言語とは何か？

言語規範(社会的規範)に支えられた理性的表現



社会的規範の特徴: あらかじめ定義しなくても  
使用できること

## 言語規範の種類と従来の言語処理

### 【意味的な約束】

- (1) 単語と語義の関係の約束 : 単語意味辞書
- (2) 表現と意味の関係の約束 × : 表現意味辞書

### 従来の言語処理

### 【文法的な約束】

- (3) 単語の並び(語順)の約束 : 文法書

<注> (3)は(2)に共通する部分を取り出したもの

➡ 本来(2)が優先する



## (2) 意味の定義

### Q2 意味とは何か？

言語過程説：「**関係意味論**」(三浦つとむ1967)

「言語表現の意味」  
= 「**言語表現**」とそれに対応づけられた「**話者認識**」の関係

<特徴>

- ・「実際に使用された表現」に対してのみ定義
- ・意味は、それぞれの**表現に固有で客観的**  
**解釈によって変化しない**
- ・言語表現の一部でも変更すると  
**意味(話者の認識との関係)は変わる**

言語処理上問題

解決策を考える

33

【補足説明1】従来の言語意味論の分類(池原1991)

言語実体	対象	話者	表現	聞き手
過程的構造	対象	認識	表現	理解
意味論	<b>関係意味論</b> ・三浦つとむ(言語過程説)			<b>解釈意味論</b> ・グライス ・横光利一 ・ラッセル ・ヴィトケン ・シュタイン
	<b>対象意味論</b> ・ベリンスキー ・蔵原 ・ブルーム ・フィールド ・ハヤカワ	<b>認識意味論</b> ・サアル ・ロック ・ルフェーブ ル ・山田孝雄	<b>形式意味論</b> ・ソシュール ・チョムスキー	
		<b>折衷意味論</b> ・橋本進吉 ・時枝誠記(言語過程説)		
				<b>実体意味論</b>

34

## (3) 意味と概念

### Q3 概念とはどのようなものか？

<概念とは>

対象の持つ**特殊性**を**普遍性**の側面から取り上げた  
**認識の単位**

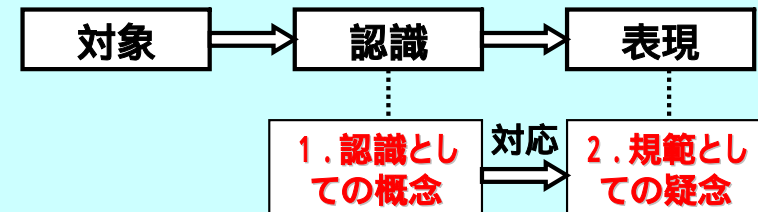
話者が、対象の持つ個別的特徴を捨象し  
必要不可欠の要素を**統一体**で捉えたもの

<概念の性質> **分割不能** **非線形**

すべての概念を人工言語で記述することは原理的に困難

33

### 言語における概念の二重性(三浦つとむ)



#### 1. 認識としての概念

話者の認識の中で形成された概念のこと

- … 言語規範としての概念を介して表現される
- … 必ずしも正しく表現できるわけではないし、表現できないものもある

#### 2. 言語規範としての概念

各言語において規範として成立した概念のこと

- … 単語や表現とそれによって表される内容との関係の規則

本研究の対象

## 本研究で対象とする意味について

本研究では、言語表現表される概念と意味と称す。従って、「認識としての概念」、「規範としての概念」の2種類の意味が存在する。

### 意味処理過程の区別 (池原1992)

「意味理解」……「本来の意味」が対象

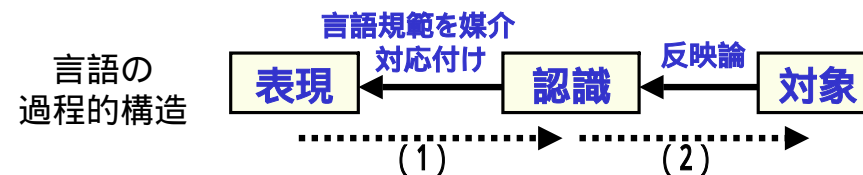
「意味解析」……「規範としての意味」が対象

本研究の  
対象

<本検討の狙い> 「意味解析」の処理の実現

37

## 【補足説明 2】 意味解析と意味理解 (池原1991)



### 言語意味処理の2ステップ

(1) **意味解析**……話者の使用した約束を特定すること  
辞書上の意味の中から、表現上で使用された約束を同定  
表現と概念の関係に関する言語知識ベースが必要  
**本研究の課題**

(2) **意味理解**……聞き手の認識の追体験  
聞き手の世界モデルとの対応付け  
話者と共通する世界知識が必要



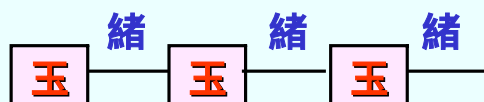
世界知識に関する知識ベースが必要 **将来の課題**

38

## (4) 言語表現と概念の関係

### 主体的表現と客体的表現

本居宣長：「玉の緒の理論」の指摘



ポイント：「玉」と「緒」は、対象のとらえ方が異なる点を指摘

【参考】ポールロワイヤルが200年前に同種の区別を提案

時枝誠記：「言語過程説」の提案

「玉」……「客体的表現」：名詞、動詞など自立語  
= 概念化された客体の表現

「緒」……「主体的表現」：助詞、助動詞など付属語  
= 概念化されない主体の感情や意思の表現

## 表現要素と概念化の関係

### Q4 言語表現は概念を表すか？

言語で表現されるのは、概念化された認識のみか

判断	学説	主体的表現	客体的表現
No	時枝文法	概念化なし	概念化あり
Yes	三浦文法	概念化あり	概念化あり

### 対時関係による説明 (池原2004)

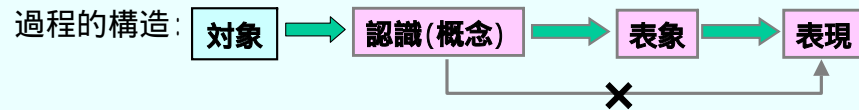
「客体化」とは：対象が話者以外であることの認識  
話者と対象との間に対時関係を発生

「主体」の感情・意志は、対時関係なしに概念化できる

「主体的表現」：概念化された対象に対する話者自身の認識  
話者自身の内的な行為で対象との対時関係は不要

## 構造から見た概念と表現

【場合1】「概念は単一の単語のみ表現される」とする場合  
句、節、文などの表現は、複数の「概念」がそれぞれ、単語(表象)を得て  
構造化されたもの



<問題点>この図式では、構造の情報は、表現にまで媒介されない

【場合2】「概念は、句、節などでも表現される」とする場合

認識は「概念」を介して直接「表現」対応づけられる



<結論> **概念化された認識のみが言語で表現される**  
(注) 但し、概念化されたすべての認識が表現できる訳ではない

## <閑話休題 4 > 概念と構造(認知言語学の場合)



従来の説明

統一体 + 分割不可能  
構造を持つかどうか不明

認知言語学の説明

概念を**構造体**として**側面**に  
初めて光を当てた



- (1) スキーマ(認知の構造)の具体化により  
プロトタイプを説明
- (2) プロトタイプからの拡張(メタファ/メトノミー  
/イメージスキーマ)によりカテゴリーを説明

42

## (5) 言語表現から見た概念の種類

**単一概念** = 単一の単語で表現できるような概念  
内部構造を意識しないまでに抽象化

**複合概念** = 複数単語の表現(句、節、文など)に  
よって表される内部構造を意識した概念

### 両者の相対性

#### 1. 概念形成を支援する言語の仕組み

- (1) 複数の「単一概念」から「複合概念」を形成
- (2) その後、単語を与えると「単一概念」に移行

#### 2. 慣用熟語の場合

- 故事に基づく概念 = 複合語(単一概念)を使用
- ・ 単語で表現できる程度に概念化が進行
  - ・ 定義なしに複雑な内容を表現

言語規範の成長

## 複合概念を単一概念化する方法

### 自然言語の場合

- (1) 新語の意味を**文中で定義**して使用
- (2) 一度述べた内容(複合概念)を**代名詞**(それ、これ、等)で表現
- (3) 「~すること」等のような**体言化**
- (4) 複雑な内容をthat節により**先行詞に縮退化**

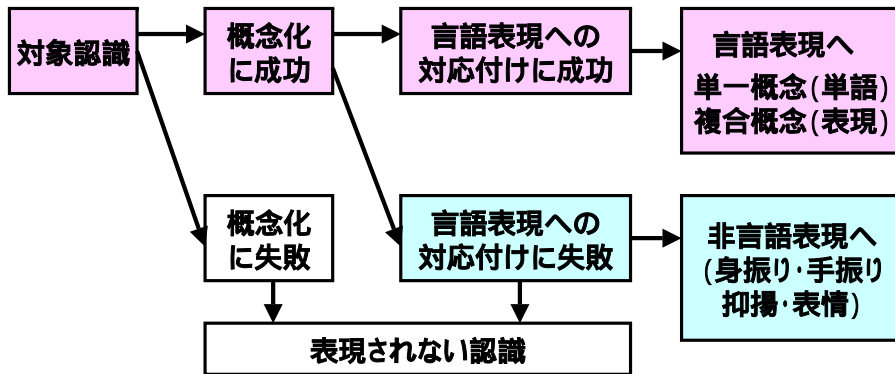
### プログラミング言語の場合

**「関数」、「手続き(サブルーチン)」、「マクロ」**

- ・ 定義の段階: 計算手続きを意識し、種々の単一概念を  
組み合わせて記述
- ・ 定義後の使用: 計算手続きは意識せず単一のラベルで参照

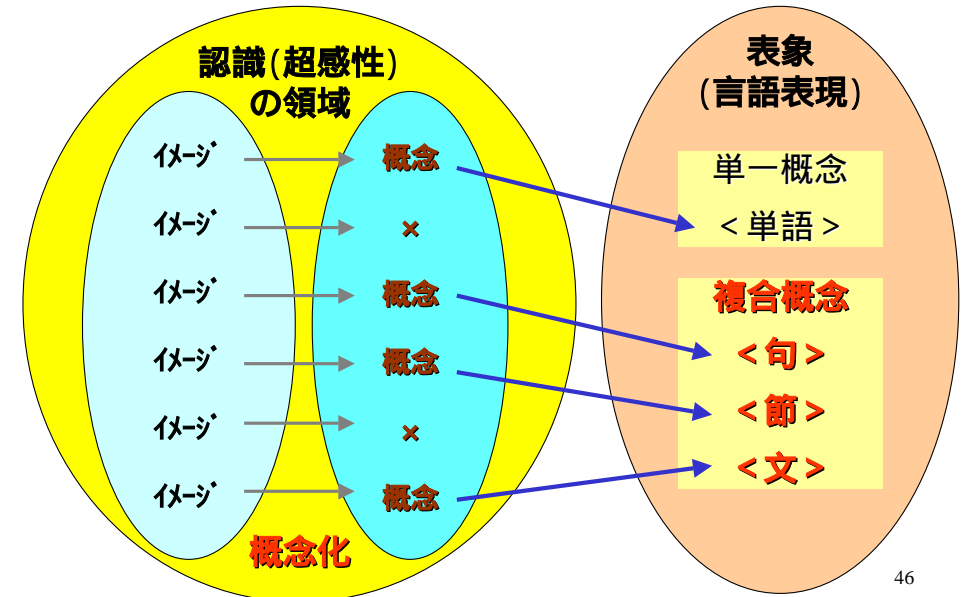
44

## 認識 / 概念 / 表現の関係



45

## 認識とその表現



46

## (6) 表現構造の持つ意味

### Q5 表現構造は概念を表すか？

<結論> 表現構造の持つ意味は、言語表現で表された**複合概念の上位概念**である

### 言語規範としての意味のまとめ

- (1) 単語の意味………単一概念の表現
- (2) 表現の意味………複合概念
- (3) 表現構造の意味………抽象化された複合概念

### 【原理】

表現構造が変わらなければ、その意味も変わらない

## 概念の階層構造と内部構造 (池原2003)

區別	概念の階層構造	概念の内部構造の記述
上位概念		<p>&lt;句型パターンによる記述&gt;</p> <p>「N1はN2よりADJ3だ。」</p>
下位概念		<p>&lt;文による記述&gt;</p> <p>「太平洋は日本海より広い。」</p> <p>「富士山は大山より高い。」</p>

複数単語からなる表現で表された概念にも**上位概念**が存在  
**「言語表現の汎化」 上位概念を表す表現への縮退**

## < 閑話休題 5 > 対象認識の言語依存性

言語の成立 **対象認識の枠組み**として使用(林語堂1960)

- ・話者の対象認識の形成能力を増大
- ・対象認識の制約条件としても働く

ヘーゲル:「直感的に統括される類型」

- ・念把握で使用される形式の存在を示唆

関口文法:「**意味形態論**」(関口存男1980)

- ・世界の把握作用が形式を伴う

時枝誠記:「**言語過程説**」(時枝誠記1941)

- ・対象認識方法が言語表現の構造に反映



49

## < 閑話休題 6 > 対象認識の個別性と共通性

【個別性】

緑なす平原に住む民族

**自然との一体性**が基調…多神教の自然観

砂漠に住む民族

**自然と人間との戦い**が基調…一神教の自然観

・主体と客体の区別の明確化と対象を立体的把握(外山1992)

< 語彙の粒度精密化の例 >

- (1) 漁労民族: 魚の語彙
- (2) 遊牧民族: 家畜の語彙
- (3) 砂漠地方: 砂の状態
- (4) 寒冷地方: 氷雪の状態

【共通性】

1. 人間としての共通性

・**地球の重力の下**で生活(衣食住)…(レイコフ1986)

・科学的な対象認識…自然法則の把握が基礎

2. 共通認識の拡大

・民族間の文化的な共通基盤の拡大



50

## 2.2 非線形言語モデル

### 言語表現の意味

< 前提1 > 言語表現の意味は概念(複合概念)である

< 前提2 > 概念は分解不能である

< 結論 > 従って、言語表現の意味は分解不可能、すなわち非線形である

【注意】 **1単語置換しても意味が変わる**

### 表現構造の意味

< 前提 > 表現構造の意味は、抽象化された複合概念である

< 結論 > 従って、表現構造の意味も非線形である

【注意】 **一部の要素を置換しても、構造の意味は変わらないことがある(抽象化された概念)**

## < 閑話休題 7 > 非線形な言語表現の必然性

(1) 対象の多次元性と表現媒体の一次元性

表現内容 = 空間(三次元) + 時間 + 感情や意志

表現媒体 = 空気の粗密波か文字列(一次元)

**例外の多い複雑な規則 = 非線形**

(2) 表現構造の保守性(頑固さ)

部分的な変更 体系に矛盾を生じやすい

(3) 社会的に開かれたシステム

社会集団の中で自然発生した  
典型的な複雑系のシステム

新しい思想や考え 既存の枠組みの変更と拡張

慣用表現など**非線形な構造が発達**



52

## < 閑話休題 8 > 言語と複雑系の共通性

### 1. 自己組織化された表現規則の存在

- ・ 個体間の単純な相互作用から全体として想像のつかない規則性が発生  
その生成過程の説明は困難

### 2. 進化に伴うエントロピーの減少

- ・ 比喩や類推の必要な表現の発生  
**エントロピー増大**
- ・ 繰り返し利用などで規則化される  
**エントロピー減少**



### 3. 初期値依存性

- ・ 初期値のわずかな違いが、次第に拡大  
言語族の違い、方言の成長など
- ・ 構造的な規則は頑固

53

## < 閑話休題 9 > 線形性に関する従来の説明

### 1. 「フレーゲの原理」(論理式)の場合



- ・ 結合表現の意味は、その諸部分の意味の関数

「各要素の真理値の任意の組が与えられると  
論理式全体の真理値は完全に決定される」

- ・ 同値の要素と置換しても全体の意味は変わらない

< 現実の言語表現 > **要素置換は意味が変わる**

### 2. Compositional Semantics の場合

- ・ **全体の意味が部分の意味の和で表される**
- ・ 「重ね合わせの原理」が保証される

< 現実の言語表現 > 分割不能の概念を表すから、  
**単語や句の意味に分割できない**

54

## (1) 言語表現の線形性と非線形性

### 対象認識の枠組み

筆者が思い浮かべる**フレームワーク**の構成

他の要素に置換すると全体の意味が損なわれる  
要素

「**非線形要素**」

他の代替要素に置換しても全体の意味は損なわ  
れない要素

「**線形要素**」

< 備考 > 要素：部分的な文字列。単語とは限らない

55

## 線形/非線形の定義

**言語表現**： 2単語以上の意味のまとまる範囲

**表現要素**： 言語表現内の部分的文字列

[定義1]： **線形な表現要素**

代替要素があり、それに置き換えても全体の  
意味(抽象化された複合概念)が変化しない要素  
それ以外は非線形要素

[定義2]： **線形な表現**

線形要素のみからなる表現

1つでも非線形要素があれば非線形表現

56

## (2) 表現構造の意味の記述法

### < 表現構造の意味 >

表現構造の持つ意味は、言語表現で表された  
**複合概念の上位概念**である

#### < 補足 >

- (1) **言語規範としての概念**であり、話者の認識として形成された概念そのものではない
- (2) **抽象化された複合概念**である
- (3) 適切な単語が存在すれば「**単一概念**」化する

57

## < 計算機から見た「意味」の記述 >

- (1) いかなる書き方も単なる記号 **どんな書き方でもOK**
- (2) 人工言語では記述不可能 **自然言語を使用**

機械翻訳への応用を前提に  
**原言語表現の意味を目的言語の表現で記述**

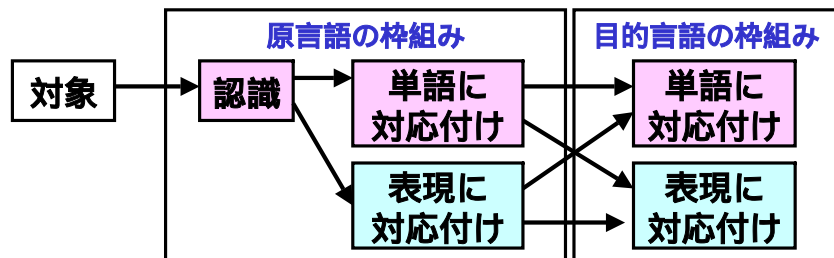
- < 例1 > 原言語表現: 「私は彼に手紙を送る」  
意味の記述: 「I send him a letter」
- < 例2 > 原言語表現: 「猿も木から落ちる」「弘法も筆の誤り」  
意味の記述: 「Any body can make a mistake」

< 目的言語表現の**多義性**の問題 >  
翻訳結果の意味を理解するのは目的言語側の人間  
**あまり問題にはならない**

### 【補足説明3】 意味を別の言語で定義する理由

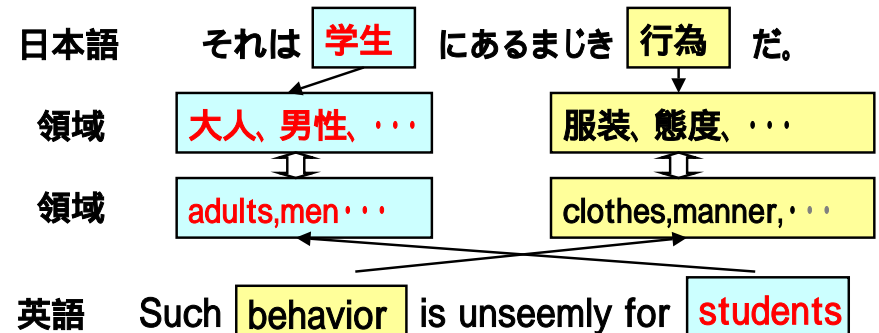
記号による定義方法      単語による定義と同等  
… 単一概念レベルで近似度が悪い

言語表現による定義方法  
… 複合概念レベルで近似度が高い



59

### 線形要素の例



#### < 線形要素の判定条件 >

- (1) 原言語表現と目的言語表現の中で**対応関係を持つ要素**であること
- (2) それらを意味的対応関係のある別の要素に置き換えても、目的言語の表現との**意味的対応関係**が保たれること

### (3) 線形要素の重要な特徴

#### < 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  
線形要素の数や範囲は言語ペアに依存

同族言語の場合：線形要素の範囲が大きい  
例) 日本語 / 朝鮮語、英語 / 仏語

異なる言語族の場合：線形要素は少ない  
例) 日本語 / 英語、日本語 / 中国語

言語による翻訳の難易性の違いを反映

61

#### < 特徴2 > 線型要素の相互独立性

線型要素の条件

- = 代替要素に置き換えても全体の意味が変わらないこと
- = 各線型要素毎に成立することが必要

複数の線型要素 相互独立であること

#### 線型要素汎化の指針

【備考】 厳しい条件であり、今後、線形要素の定義の見直しが期待される

(改良案の例) 代替要素を要素集合間の直積の部分集合を使用して定義する方法

62

#### < 特徴3 > 代替要素の変域の存在性

置き換え可能な要素が線形要素  
任意の要素への置き換えを意味しない

文法的・意味的な変域(値域)が存在  
線形要素の変域間には相互関係あり

例) 「私は彼に手紙を送る」 = 「I send him a letter.」

「私は」 「彼女は」、 × 「リンゴは」  
「手紙を」 「贈り物を」、 × 「東京湾を」  
「送る」 「さしあげる」、 × 「落とす」

相互関係を考慮した変域指定が必要

63

#### < 特徴4 > 要素分解の仕方への依存性

表現要素の線形性  
表現全体の線形性

要素の範囲に依存  
要素分割の方法に依存

例) 「私は彼に手紙を送る。 = I send him a letter」

< 全体が線形となる要素分解 >

「私は(I)」、「彼に(him)」、「手紙を(a letter)」、  
「送る(send)」に分割

< 全体が線形とならない要素分解 >

「私」、「は」、「彼」、「に」、「手紙」、「を」、「送る」  
に分割 「は」、「に」、「を」は置き換え不能  
(日本語側では、「は」、「が」、「に」、「へ」は可能)

汎化で工夫できる余地が大きい

64



## < 特徴5 > 線形要素と非線形表現の同時性

### 線形、非線形の区別

表現の部分と全体の意味的關係で定義

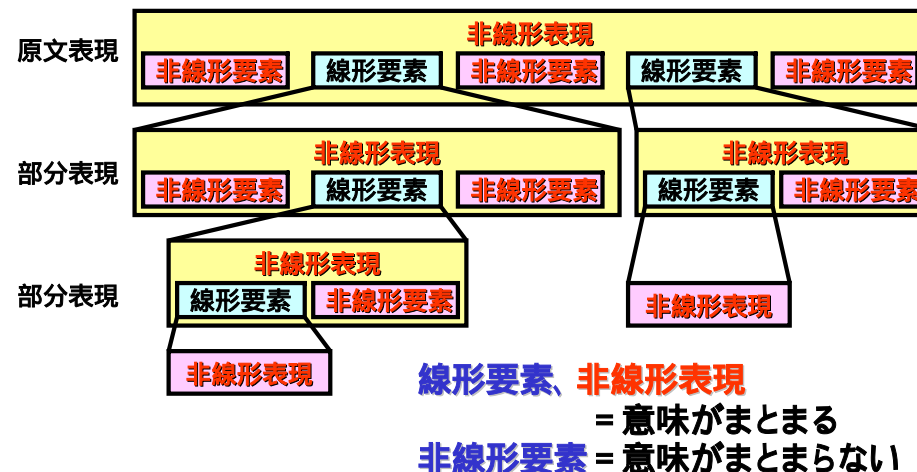
### 線形要素とは

- ・ 表現全体に対して線形である要素のこと
- ・ **要素自身の線形性を意味しない**
- ・ **線形要素の構造は非線形であっても良い**

非線形表現は、線形要素を介した  
**再帰的な構造**を持つ

### 非線形な言語表現モデルの構成原理

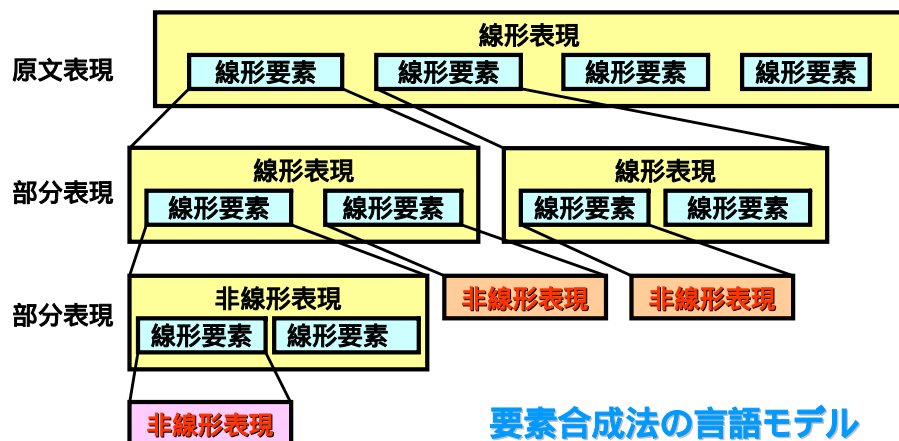
## (4) 非線形言語モデル



言語表現 = 非線形要素と線形要素の複合体

非線形表現 = 構造的な意味の単位 **パターン**へ

## 【補足説明4】 従来の線形言語モデル



言語表現 = 線形要素の複合体  
(線形要素は、単語、句、節など)

## < 閑話休題 10 > 子供の言語獲得とパターン

- (1) 単語と概念の対応關係の知識を獲得する段階  
・ たまたま複数単語の表現を使ったとしても  
単語単位への**文節化の意識なし**
- (2) 表現の部分的共通性から分節化  
に気づく段階  
・ 表現の一部で、**置き換え規則を獲得**
- (3) 表現全体が**分節化できることに気づく段階**  
・ 表現能力は急速に拡大 誤った表現を頻発
- (4) むやみに線形規則が使えないことを知る段階  
・ 線形規則の**意味的な適用範囲の知識**を獲得



## 2.3 非線形表現構造のパターン化

意味のまとまる構造の単位を抽出  
パターンで記述

なぜパターンか？

- (1) 非線形表現は非線形要素と線形要素の配列  
**語順任意性が少ない**
- (2) 非線形要素は表現構造のキーとなりやすい  
**字面表記が適切**

69

## 表現パターン方式の意義

1. 人間の思考パターンに近い構造を持つ
2. 非線形な言語表現の中の線形要素が縮退されており、無限の表現が有限の表現構造で説明できる
2. 語や表現の意味と用法との関係を表現でき、意味的曖昧性の解消に役立つ

70

## パターン化の対象表現

各レベルの**非線形表現**を形式化

- (1) 重文/複文レベル…事象関係の表現
- (2) 単文レベル…事象の表現
- (3) 句レベル…実体、属性の表現  
(複合語を含む)

## 言語表現への適用の順序

原則: **大域的構造から適用**

- (1)
- (2)
- (3)の順

71

## <補足説明 5> 構文解析技術は不要となるか？

すべてのレベルの非線形な表現に対して、  
意味的に排他的なパターン化されたとき

1. 意味解析のための前段技術  
としての構文解析技術は不要
2. 構文情報が必要なときは、各パターンに  
**構文情報を付与**しておき、それを使えばよい



**→ 従来の構文解析技術は不要**

72

## 2.4 第2章のまとめ

### (1) 言語表現と表現構造の意味

- ・ 言語表現の意味 = 言語規範としての概念
- ・ 表現構造の意味 = 抽象化された**複合概念**

### (2) 非線形性に着目した言語表現モデル

- ・ **線形 / 非線形の定義**と言語表現モデル
- ・ **原言語表現の意味**を目的言語で記述

### (3) 非線形表現のパターン化

- ・ 語順任意性が少ない
- ・ 字面表記の必要性

### < 閑話休題 11 > プラトンの問題 (刺激の貧困)

#### 子供はなぜ3~4歳にして言葉をマスターできるのか

人間は、外界との接触が短く個人的で限定されているにも拘わらず、どうしても多くのことを知っているであろうか

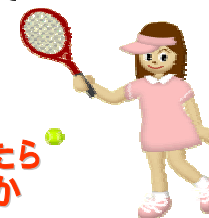
人間は有限で不完全なデータからどのようにして無限の言語表現をまなぶのか (チョムスキー)

#### 自然言語処理の基本問題

記憶容量が有限の計算機で、どのようにしたら**ほぼ無限とも言える言語表現**を処理できるか

従来の自然言語処理は、表現の多様性の問題で失敗  
**この問題に答えられる言語モデルが必要**

< 参考 > 人間はこんなに沢山の経験を積みながら、なぜこんなに愚かなのか (オーウエルの問題)



74

### < 閑話休題 12 > プラトンの問題への解答

#### (1) チョムスキーの言語モデル

**深層構造**から**意味を変えない変形規則**で生成

意味を変えない変形規則の矛盾  
意味を変える変形規則へ 意味の二元論へ  
その後、生得的な言語能力の存在を仮定  
変形規則を廃して「普遍文法」へ

< チョムスキーの偉さ > オーウエルの問題に挑戦

#### (2) 本研究の言語表現モデル

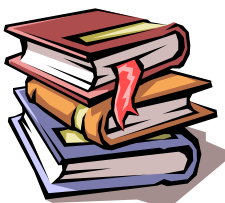
**有限の非線形表現の線形結合**で合成

説明 1: 非線形な表現は記憶しないでは使えないが、有限だから記憶し使用できる

説明 2: 表現の多様性は、**有限な非線形表現の結合**から生まれる **幾何級数的な増大**

< 課題 > **果たして、非線形な表現構造は有限か?**

(参考) 統計翻訳 = プラトンの問題から逃避



75

## 第3章 パターン記述言語の設計

### 第3章の目次

- 3.1 パターンの役割と対象
- 3.2 パターンの設計条件
- 3.3 パターン記述要素の構成
- 3.4 単語意味属性体系の構築
- 3.5 第3章のまとめ

76

# 3.1 パターンの役割と対象

## 文型パターンとは？

- ・意味のまとまる表現の単位に、その構造を形式化
- ・原言語と目的言語の構造を意味的に1対1に対応づけ

### (1) 言語表現の過程

概念化された対象認識を表現する枠組み

### (2) 言語理解の過程

意味をすくい取るための網の目

意味と構造を一体化した処理の実現

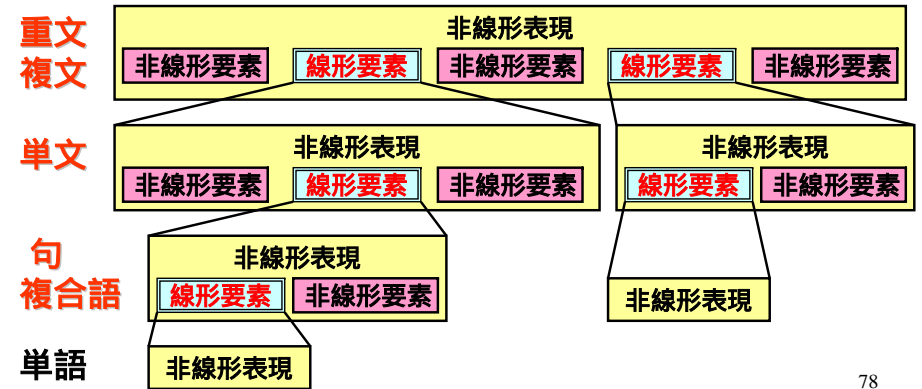
## 工学的な意義

要素合成法からの脱却

全体の意味から部分の意味の説明を可能とする

## パターン化の対象 = 意味のまとまる非線形な表現 =

- (1) 句(名詞句、副詞句、形容詞句)・・・実体と属性の表現
- (2) 単文……………事象の表現:実体と属性の関係
- (3) 重文、複文……………事象間の関係の表現



## パターン化の方法

### (1) 原言語の意味を目的言語で定義

対訳用例の線形要素を汎化

### (2) 非線形な表現構造が意味の単位

非線形な表現構造を文型パターン化

### (3) 結合価パターン方式の反省

非線形な主体的表現を汎化しないこと

## 線形要素の種類

### 日英対訳例文を対象とする場合

- (1) 英語側に**対応する要素**があるもの
- (2) 英語側に**対応する要素はないが、削除**しても英語表現は変化しないもの
- (3) 元の対訳表現には**ないが、それを挿入**しても、対訳関係が変化しないもの
- (4) 語順など**構造的な変更**を行っても英語側の構造が変化しないもの

## 3.2 パターンの設計条件

- (1) 文法的・意味的に正しい入力表現が前提  
パターンに誤り検出能力は不要
- (2) 汎化レベルに応じたパターンの記述  
個別的で具体的なパターン……優先的に使用  
汎用性の高いパターン……補助的に使用

汎化レベル	汎化の内容
単語レベル	線形要素とみなせる <b>自立語を変数化</b> したパターンのレベル
句レベル	線形要素と見なせる <b>名詞句, 動詞句を変数化</b> すると共に, 単語レベルで作成されたパターンの単語変数を可能な限り句の変数に置き換えたパターンのレベル
節レベル	線形要素と見なせる <b>節を変数化</b> したパターン

81

- (3) 日英表現パターン相互の整合性  
英語表現がユニークに決定できること  
・目的言語側変数 すべて原言語側で定義
- (4) 入力表現との照合の単純性と無曖昧性  
形態素解析情報を使用した照合  
解析多義の残る部分は, 字面情報で照合  
**パターン照合時の入力文意味解析は不要に**  
<理由> 従来の構文解析や意味解析の限界を克服
- (5) 表記と表現の揺らぎの吸収
- (7) 体系的で分かり易く拡張性があること  
・多数の作業者の動員 作業の均質性

82

## 3.3 パターン記述要素の構成

### (1) 記述要素の種類

「字面」、「変数」、「関数」、「記号」の4種類

#### 記述の原則

- (1) 非線形要素                      字面 (2種類)
- (2) 線形要素
  - ・自立語的線形要素              **変数** (17種類)
  - ・付属語的線形要素              **関数** (157種類)
  - ・構造的線形要素                **記号** (10種類)

83

### 必須要素と任意要素

#### 「必須要素」

それがないと、英語との意味的対応関係が失われる

#### 「任意要素」

それがなくても元の英語との意味的対応が保たれる

- (1) **元の表現にない要素** 「原文任意要素」  
パターンに新たに挿入しても対応する英語パターンを使用して英文生成ができる要素
- (2) **元の表現にある要素** 「パターン任意要素」  
英語側に対応要素がなく、無視しても良い要素  
英語側に対応要素があり、双方から削除できる要素

## (2) 字面による記述

### 出現形字面

日本語パターン、英語パターン共に原文通りの字面による指定

### 終止形字面(日本語)と原形字面(英語)

字面に関数付与を可能とするための記述法

機能: 字面部分の標準形を指定する

用途: 様相時制の任意化で使用する

#### <使用例>

日本語の場合: **行く** [.masu]

「行く」、「行きます」の入れにも適合

英語の場合: **is** #3(^present | ^past)

適合条件に応じて、「is」または「was」が生成される

## 「字面」の例

日本語 例文 気が向くと、彼女はお茶をいれてくれる。

終止形字面

原文字面 = 全角文字列

/y \$<sub>1</sub> /<sup>f</sup> VP<sub>2</sub> ^syushi と \$<sub>1</sub> ^ {/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)

英語

例文 She serves us tea when she is in the mood.

原形字面

原文字面 = 半角文字列

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) .

## (3) 変数による記述

線形な自立語的要素(単語、句、節)を変数化

### (1) 変数化の対象

- ・表現全体に対する線形要素
- ・それ自身は非線形構造を持っていてもよい

### (2) 対応する目的言語表現

- ・両言語間で要素の対応関係がとれること
- ・同一の文法的属性を持つ必要はない

例) 名詞Nの原言語要素に対応する目的言語要素が動詞Vの場合  
目的言語側は、関数N(V) (動詞の名詞化)などで記述

### (3) 変域の指定

- ・文法的な変域 変数名
- ・意味的な変域 意味属性

## 変数の種類と用法

分類	種類	文型照合での代入(バインド)値
単語変数	無活用語	人力表現中の該当する単語または複合語の標準形がバインドされる
	活用語	該当する単語の基本形(現在形)がバインドされる
	ANY変数	品詞によらず任意の単語にバインドされる
句変数	無活用の句	該当する名詞句または副詞句がバインドされる
	活用のある句	該当する句、但し、活用は基本形がバインドされる
節変数	- - -	命題レベルの単文(核文)の部分がバインドされる

## 変数一覧(17種類)

分類	変数種別
単語変数	Nn:名詞一般+複合名詞、 TIME n:時間の名詞、 NUMn:数詞、 NDn:用言性名詞、 RENn:連体詞、 GENn:限定詞、 Vn:動詞、 AJn:形容詞、 AJVn:形容動詞、 ADVn:副詞、 ANY:任意
句変数	NPn:名詞句、 VPn:動詞句、 AJPn:形容詞句、 AJVPn:形容動詞句、 ADVPn:副詞句
節変数	CLn:節 (但し、核文の現在形)
<注> すべての変数には、パターン内の通し番号nが付与される。	

89

## 変数の意味的制約条件

### CREST独自の意味属性体系

名詞:370種類、用言:370種類

<現状では以下の2種類を併用>

- (1) 日本語語彙大系の意味属性体系  
名詞:2800種類、用言:36種類
- (2) 類語大辞典の意味分類  
名詞+動詞:1000種類

## 【補足説明6】 意味的制約条件とは？

### 従来の方法 = 語彙に着目

・上位概念名または配下の概念に共通する属性で定義

(例) 意味標識、意味素、意味素性

**意味的制約条件に向かない**

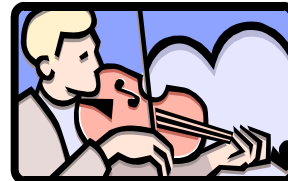
<理由> 古典的なカテゴリー論(概念=属性の束)に立脚  
表現上の意味=辞書的な意味(語彙)

### 本研究の方法 = 用法に着目(池原他1993)

・「意味的用法」を「意味属性」として体系化

(例)「学校」 組織、場所、建物、...

「鶏」 鳥、食料



### 最近の理論的な研究(黒田航2004)

・認知言語学(スキーマの具体化による概念化の説明)

を背景とする理論的な研究

「意味役割名」と「意味型名」を区別

91

## 「変数」の例

日本語 例文 気が向くと、彼女はお茶をいれてくれる。

<注> (意味属性): の部分は記載を省略

VP 動詞句変数

N 名詞変数

$/y \$_1 /{}^f VP_2 \wedge syushi$  と  $\$_1 \wedge \{ /{}^{tcl} N_3 \text{ は} \} \#1 \{ /{}^{tcfk} < N_4 \text{ に} >, ! N_5 \text{ を} \}$   
/cf(いれ|入れ) 'てくれる' #6(.genzai|.kako),

英語 例文 She serves us tea when she is in the mood.

N 名詞変数

VP 動詞句変数

$N_3$  'serve' #6(^present|^past) < us  $N_4$  ^obj>  $N_5$  when  
<she|  $N_3$ > AJP(VP<sub>2</sub>) #6(^present|^past) .

## < 閑話休題13 > 用言変数化の必要性

分類	述部用言の表記	用言の分類数	名詞意味属性	必要なパターン数
単文	字面	6000種類 (36分類)	2700分類 <粒度小>	2.5万件
重文複文	<b>変数</b>	370種類 (意味属性)	370分類 <粒度大>	5万~10万 (推定)

単文の場合: 変数化なし 密度 = 2.5万 / 6000 = 4パターン / 用言

重文複文の場合: 変数化必須

必要見込み10万パターンを仮定

・変数化しないと 6千×6千×10(KW) = 3600万  
でスペース過ぎる

・密度 = 1の時 10万 = 316件 により

数百文例程度が適当



93

## (4) 関数による記述

### 1. 要素の語形指定 「語形関数」(ハット関数)

用言の活用形など要素自身の屈折を指定

例) <sup>^</sup>rentai、<sup>^</sup>poss

### 2. 付属語類の指定 「時制様相関数」(ドット関数)

- ・線形 / 非線形な付属語類を指定
- ・意味解析の必要な関数は設けない<注>

(1) 線形要素型 例) .dantei  
代替可能な複数の字面を示す

(2) 非線形要素型 例) .teiru  
単一の字面を示す

<注> 表現パターンは意味解析のためのものであるから入力表現との照合時に意味解析を必要とする指定はしない

## 3. 関数のグループ化

(1) 「同値型グループ関数」 例) V2.desu  
英語表現に影響しない同値の関数をまとめて指定

(2) 「対応型グループ関数」  
日英対応関係のある関数を指定まとめて指定  
例) 日:V2.#3meirei# 英:V2.#3base#

## 4. 変数への適合条件の指定

(1) 「字面関数」 例) とても(CL1.kako)  
変数への適合条件として要素内に含まれる字面を指定

(2) 「マクロ関数」 例) CL1#(N2はAJ V3 V4)  
変数に適合する要素自身の構造を指定

95

## 関数の種類(全157種類)

#	種類	説明
1	様相関数	時制、相、様相、態などを指定。(日本語用42種、英語用23種)
2	語形関数	動詞や動詞句、節の活用形などを指定。 (日本語用13種、英語用20種)
3	語形任意化関数	用言字面の標準形を指定。様相時制の任意化で使用。
4	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。 (8種)
5	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 (24種)
6	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 (11種)
7	字面関数	句変数、節変数に代入される表現要素の含む字面を指定
8	マクロ関数	変数を含む表現を指定し、それを上位の変数にバインド
9	要素抽出関数	句変数、節変数の値から特定の部分を抽出(2種類)
10	構文生成関数	一つ又は複数の句や節から英文構造を合成(22種)

96



## 関数の記述形式とその意味

番号	種別	形式	意味
1	ハット関数	^関数名	付与された用言の語形を指定する
2	ドット関数	.関数名	指定された付属語の存在を意味する
3	シャープ関数	#関数名	同値型グループ関数
		#関数名#	対応型グループ関数
		#字面(引数)	引数に含まれる字面を指定する
4	変数関数	変数名(引数)	引数を変数名の品詞に変換する
5	要素抽出関数	関数名(引数)	引数から関数名で指定した要素を取り出す
6	構造指定関数	変数名^(引数)	<日本語> 変数にバインドされる表現の形式を引数で指定する
			<英語> 引数から関数名で指定する構文を生成する

97

## 時制・相・様相に関する関数(ドット関数)

### <日本語> (42種類)

.genzai:現在、.kako:過去、.Mirai:未来、.tekuru:てくる、.tearu:てある、.teiku:ていく、.teoku:ておく、.temiru:てみる、.nisuru:にする、.teiru:ている、.sase:使役、.tai:希望、.hitei:否定、.joutaihenka:状態変化、.youda:推定、.masu:丁寧、.you:意志、.dekiru:可能、.kinshi:禁止、.teyoi:許可、.teyaru:受益、.tekureru:被益、.meireigo:命令、.kaisi:開始、.Kamosirenai:推量、.teshimau:好ましくない結果、.souda:不確実な断定、.da:断定、.rareru:可能/受身/自発/尊敬、.gimu:義務、.darou:推量、.Nichigainai:必須、.suitei:推定、.desu:丁寧な断定、.utosuru:意思行為、.noda:断定、.Kanyu:勧誘、.hougayoi:良好、.sugiru:過剰、.yotei:予定、.Tekudasai:丁寧な命令、.All:任意

### <英語> (23種類)

.can:可能、.not:否定、.may:許可、.should:義務、.better:良好、.must:義務、.will:意志、.would:意志、.let:使役、.Please:丁寧、.Donot:否定、.too:付加、……

## 語形関数(ハット関数)

分類		種類(13種類 + 20種類 = 33種類)
日本語文型	活用形指定関数	^mizen:未然形、^renyo:連用形、^shuushi:終止形、^rentai:連体形、^katei:仮定形、^meirei:命令形
	助動詞内包指定関数	^desire:可能、^dues:肯定、^sugiru:過剰、^dare:推定、^sase:使役
	意味指定	^Sieki:使役、^kano:可能
英語文型	名詞変形関数	^obj:目的格変形、^poss:所有格変形、^abposs:独立所有格変形、pro:代名詞変形、^reflex:再帰代名詞変形、^pl:複数形^
	動詞変形関数	^base:原型、^present:現在、^past:過去、^future:未来、^prp:現在完了、^psp:過去完了、^prg:進行形、^passive:受け身、^ing:現在分詞、^ed:過去分詞、^grn:動名詞化変形
	形容詞変形関数	^st:最上級、^er:比較級

99

## 同値型グループ関数

#	分類	種類(合計8種類)
1	用言述部型	.#masu、.#noda いずれも.masu + .nodaの意味
2	体言述部型	.#desu、.#da いずれも.desu + .daの意味
3	推量型	.#daoru、.#desho いずれも、.darou + .deshoの意味
4	可能型	.#rareru、.#dekiru いずれも.rareru + .dekiru + /dekiruの意味

100

## 対応型グループ関数(24種類)

#	分類	日本語側	英語側
1	時制グループ	.#genzai# .#kako# .#mirai#	.#present# .#past# .#future#
2	真偽判断グループ	.#nichigainai# .#kamosirenai#	.#must# .#may#
3	価値判断グループ	.#gimu# .#hougayoi#	.#must# .#better#
4	希望・勧誘グループ	.#meirei# .#tekudasai# .#kanyu#	.#base# .#please# .#lets#
5	禁止否定依頼グループ	.#kinshi# .#hitei . tekudasai	.#donot# .#please.donot#

## 品詞変換関数(英語用、11種)

#	関数形	引数	機能
1	N(arg)	V, AJ, AJV	引数を名詞に変換
2	V(arg)	N, AJV	引数を動詞に変換
3	AJ(arg)	N, V, ADV	引数を形容詞に変換
4	ADV(arg)	N, V, AJ	引数を副詞に変換
5	NP(arg)	VP, AJP	名詞句に変換
6	VP(arg)	NP	名詞句を動詞句に変換
7	AJP(arg)	NP, VP	形容詞句に変換
8	ADVP(arg)	NP, VP, AJP	副詞句に変換
9	NP_pred(arg)	CL, VP	引数を述部に着目して名詞句に変形
10	NP_subj(arg)	CL, VP	引数を主語に注目して名詞句に変形
11	NP_obj(arg)	CL, VP	引数の目的語に注目して名詞句に変形

## 字面関数(日本語用)

記述形式: #字面(arg)

引数: VP, CL

機能: 引数表現が特定の字面を含むことを表す

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

## マクロ関数(日本語用、3種)

種類例: CL<sup>^</sup>(arg)、NP<sup>^</sup>(arg)、VP<sup>^</sup>(arg)

引数: 節構造, 名詞句構造, 動詞句構造を持つ文型

機能: 指定に適合した表現構造を関数名として使用されている変数の値とする

<例>

和文P: CL<sup>^</sup>(N1はADV2 AJV3).pastのでCL4.cannot.past,  
= N1はADV2 AJV3.pastのでCL4.cannot.past,

英文P: so\_that(#CL1, CL1)  
= N1 was so AJV1 that N2 VP1.

例文: その星は大変暗かったので、我々は見ることができなかった。  
The star was so dark that we could not see it.

## 要素抽出関数 (英語用)

記述形式: Subj(arg), Obj(arg)

引数: CL

機能: argの主語又は目的語を取り出す。

例) CL1 = We eat an apple のとき,  
subj(CL1) = we, obj(CL1) = an apple

105

## 構文生成関数 (英語用) (22種類)

#	分類	種類so.that(arg1,arg2),
1	態の変形	Active(arg)
2	程度を伴う因果関係	so.that(arg1,arg2), so_that(arg1,arg2), such_that(arg1,arg2), such.that (arg1,arg2), too_for_to(arg1,arg2,arg3), too_for_even.to (arg1,arg2,arg3), much.too_for_to (arg1,arg2,arg3), so_as.to(arg1,arg2,arg3)
3	目的	in_order_that (arg1,arg2,arg3)
4	時間的關係	no_sooner_than1 (arg1,arg2), no_sonner_than2 (arg1,arg2,arg3), scarcely_before1 (arg1,arg2), scarcely_before 2 (arg1,arg2), hardly_when1 (arg1,arg2), hardly_when2 (arg1,arg2)
5	不可避	cannot_without (arg1,arg2), never_without (arg1,arg2)
6	強調	inverse (arg1), inverse_as(arg1)
7	分詞構文	ptcle(arg1)
8	前置詞句	to_emotion(arg1)

## 「関数」の例

日本語 気が向くと、彼女はお茶をいれてくれる。

$/^{\nu} \$_1 / ^f VP_2 \wedge \text{syushi}$  と  $\$_1 \wedge \{ / \text{teki} N_3 \text{ は} \} \#1 \{ / \text{tefk} < N_4 \text{ に} >, ! N_5 \text{ を} \}$   
 $/ \text{cf} ( \text{いれ} | \text{入れ} ) \uparrow \text{'てくれる'} \#6 ( . \text{genzai} | . \text{kako} )$ 。

語形関数

時制様相関数

英語 She serves us tea when she is in the mood.

時制様相関数

目的格変形関数

$N_3 \text{'serve' } \#6 ( \wedge \text{present} | \wedge \text{past} ) < \text{us} [ N_4 \wedge \text{obj} ] N_5$   
 when  $< \text{she} | N_3 > \text{AJP} ( VP_2 ) \#6 ( \wedge \text{present} | \wedge \text{past} )$

品詞変換関数(動詞句 形容詞句)

## (5) 記号による記述

### 1. 要素選択型の表現

表記と表現の揺らぎを指定

- (1) 「要素選択記号」…例) #2(とかく | 兎角)  
英語パターンが変化しないもの
- (2) 「対応型要素選択記号」…例) #2(^meirei | .kanyu)  
英語側と対応関係のある言い換え表現を指定

### 2. 要素補完のための表現

省略可能な主語・目的語などを指定

- (1) 「補完要素記号」…例) #1 < N2は >  
英語側にあるが日本語側にはなくても良い要素を指定
- (2) 「訳出要素選択記号」…例) #2 < N1 | it >  
日本語側の要素の有無に応じて選択する表現を指定

### 3. 任意要素の指定

#### (1) 原文任意要素 「離散記号」

連体・連用修飾句など(6種類)の「原文任意要素」の位置を指定

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

#### (2) パターン任意要素 「任意要素記号」

日英パターンで対応関係を持つ要素で、双方から削除できる要素 例) #3[ADV4]

### 4. 語順と位置変更可能要素の指定

(1) 「順序任意要素指定記号」 例) #1 {N1が、N2で} 格要素などの要素をグループ化

(2) 「位置変更可能記号」 例) \$1 ^ {N1は}....\$1 副詞などの変更可能な位置を指定

109

### 記号一覧

#	分類	記法	機能
1	離散記号	/ = y, t, c, k, f	原文任意要素(種類)が存在しても良い場所を示す
2	文節境界記号	!	文節境界位置を示す(通常は使用しない)
3	要素選択記号	#n(arg1   arg2   ...)	代替可能な要素(異表記と異表現)の指定
4	対応型要素選択記号	#n(arg1   arg2   ...)	同上。但し、英語側と対応させて使用
5	訳出要素選択記号	#n<arg1   arg2>	訳出要素を指定。日本語側に適当な要素がないときの代替要素を指定
6	任意要素記号	#n[arg1]	入力表現中にあってもなくても良い要素
7	順序任意記号	#n{arg1   arg2   ...}	出現順序が任意の要素を指定 例){彼女を 彼は 駅まで}送る。
8	位置変更可能要素指定記号	対象要素:\$n^{ } 移動位置:\$n	位置変更可能な要素と変更可能位置の指示。
9	補完要素記号	#n <arg >	省略要素の補完方法の指示
10	要素挿入記号	#n{arg1}	群動詞中に副詞を挿入することなど

### 「記号」の例

日本語 気が向くと、彼女はお茶をいれてくれる。

/y 離散記号      位置変更可能記号      順序任意要素指定記号

/y \$1 /f VP2 ^syushi と \$1 ^{/tck N3は} #1{/tcfk <N4に>, ! N5を}  
/cf (いれ|入れ) 'てくれる' #6(.genzai|.kako),

要素選択記号

対応型要素選択記号

英語 She serves us tea when she is in the mood.

対応型要素選択記号

訳出要素選択記号

N3 'serve' #6(^present|^past) <us|N4 obj> N5  
when <she|N3> AJP (VP2)#6(^present|^past) .

### 3.4 単語意味属性体系の構築

#### (1) 意味属性体系と意味辞書の種類

変数の変域指定のため新規に開発

分類		意味属性体系	単語意味辞書
用言	動詞	4段295分類	6000語
	形容詞	3段71分類	
体言	名詞	4段364分類	6万語

## (2) 用言意味属性体系

### 設計条件

#### (1) 統語的依存性の考慮

- ・ 変域は、表現構造からの制約を受ける
- ・ 統語的に異なる構造を要求するような用言を排除例) 自動詞 / 他動詞、可能動詞、使役動詞

#### (2) 場面依存性の考慮

格要素の名詞と用言の意味との調和性

**対策:** 両者を考慮して分類体系化

<備考> 将来は、用言意味属性の2次元化する予定

#### (3) 分解能

- ・ 分類の深さに意味的な粒度をほぼ対応
- ・ 中間的な深さでの分類数を60程度
- ・ 最新の深さでの分類数合計が数百程度

## 意味属性体系と意味辞書構築の手順

- (1) 寺村の分類体系(1982)に対する分類名の見直し  
寺村の分類(統語構造に着目:約30分類)の見直し
- (2) 外国人のための日本語用例集(名柄1987)の参照  
動詞編と形容詞編の用例から、用言と場面の関係を分類
- (3) IPAL辞書(1987)からの組み入れ  
分類項目と対象用言の追加
- (4) 「類語大辞典」(柴田、山田2002)収録語からの追加  
分類代表語1000語からの補充
- (5) 「日本語語彙大系」(池原他1997)からの登録語の拡充  
・ 結合価パターン辞書(15,000件)見出し用言の抽出  
・ 名詞辞書(「こと類」の名詞、サ変名詞)からの見出し

114

## 動詞の意味属性体系

#	第1段	第2段	第3段	第4段	第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
合計	1	9	49	147	88	295

- <観点> (1) 感覚と知性 (# 1, # 2)  
(2) 生活の場面 (# 3 ~ # 5)  
(3) 論理的構造 (# 6 ~)

115

## 動詞の意味属性体系 (1 ~ 2 段)

知覚と情緒の表現	社会的活動の行為	変化の現象事象の表現
知覚感覚	組織	身体的変化
個人的感情	行社会議	量的変化
对人的感情	規則契約	物理的变化
知的な行為の表現	係争紛争	状態変化
思考と認識の行為	情報活動	接辞による表現
言語と表現の行為	企業活動	移動行為の表現
教育学習の行為	労働	出入りの移動
学術的活動	現象事象の表現	上下の移動
芸術趣味の行為	存在既定	前後移動
信仰尊敬の行為	開始終了	行き来の移動
日常生活の行為	発生消滅	遠近移動の事象
身体的動作	自然現象	各種の移動
家庭生活の行為	物理的事象	所有物の移動
個人的な行為	力学的事象	対物的行為
对人的な行為	物象	単一的物体への行為
地域社会生活の行為	社会現象	複合的物体への行為
風俗風習	抽象的事象	対象依存の行為
学生生活		創造破壊の行為
対人交流		
障害災害危機		

1段目(9種) + 2段目(45種) +  
3段目(101種) + 4段目(70種) = 295種

## 形容詞の意味属性体系

#	第1段	第2段	第3段	第4段	第5段	合計
1	形容詞	性状規定	存在規定	0	0	1
2			感覚感情規定	4	19	24
3			人の性状規定	10	0	11
4			物の性状規定	3	2	6
5			事象の性状規定	12	12	25
6			政治経済社会	5	0	6
合計	1	1	6	34	33	75

【動的属性を静的に捉えた語の場合】

動詞の意味分類と重複. 但し, より細かく分類

117

## 形容詞の意味属性体系(1~2段)

存在既定	物の状態既定	政治経済社会
感覚感情既定	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	事象の性状既定	様相
気配趣きリズム	評価既定	経済的活動
人の性状既定	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	

1段目: 6種  
2段目: 34種  
3段目: 31種  
合計: 71種

## (3) 名詞の意味属性体系

### 設計方針

1. 用言の意味分類体系とほぼ同等の粒度
2. 粒度と深さを対応させたフラットな構造

### 参考資料

- (1) 「日本語語彙大系」(池原他1997)の2700分類
- (2) 「講談社類語大辞典」(柴田、山田2002)

## 名詞の意味属性体系

#	第1段	第2段	第3段	第4段	第5段	合計
1	名詞	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	1	6	41	209	114	365

## 名詞意味属性体系(1~2段)

主体	抽象物	抽象的關係
人稱	知的生産物	存在
呼稱	創作物	類・系
擬人	言語	関連
組織	文書	性質
場所名	倫理	状態
施設名	宗教	形状
地域名	伝承	数量
陸地名	情報	場
海の名称	評価	時間
景観	制度	
宇宙名称	成敗・業績	
具体物	習慣	
動物(個体)	こと	
動物(部分)	精神	
植物(個体)	行為	
植物(部分)	事象	
自然物	自然現象	
人工物		

1段目: 6種  
2段目: 41種  
3段目: 209種  
4段目: 114種  
合計: 370種

## 3.5 第3章のまとめ

### パターン記述言語を設計

- <目的> 意味的に非線形な表現構造の記述
- <目標> 意味的な相互独立性と高い被覆率  
文型パターン辞書の半自動生成
- <構成> 4種類の記述要素(字面、変数、関数、記号)

= 意味属性体系と単語意味辞書の開発 =

- (1) 用言の意味属性体系(360分類、6000語)
- (2) 名詞の意味属性体系(370分類、6万語)

## 第4章 重文複文の文型パターン辞書の開発

### 第4章の目次

- 4.1 重文と複文のパターン化
- 4.2 文型パターン辞書の実験的改良
- 4.3 文型パターンの記述例
- 4.4 静的特性の評価
- 4.5 被覆率特性の評価
- 4.6 パターンパーサと翻訳実験
- 4.7 第4章のまとめ

## 4.1 重文と複文のパターン化

### (1) パターン化の対象

- 1) 句……未
- 2) 単文……「日本語語彙大系」  
結合価パターン辞書1.7万件開発済  
<問題点> 非線形な主体的表現の扱い  
非公開  
**再開発が必要**
- 3) 重文、複文……未  
現状の訳文品質が劣悪 **開発が急務**

## 対象文種別

**重文と複文、但し、述部数2または3**

述部4以上の文は、分解して処理

重文(文接続のある文): Complex Sentence

複文(埋め込み文のある文): Compound Sentence

分野依存のない標準的な日本文

専門的な表現パターンは、個別に任す

文脈によらず対訳関係が成立する文

記述文

意識された対訳例の文型パターン化は特に重要

125

## 対訳標本文

パラレルコーパス 100万文ペア

約30種類のドキュメントから収集

辞書や語学教育用の教科書,  
機械翻訳機能評価用の試験文等

標本文数: **128,713ペア**

日本語標本文

・平均文字数: 23.3字 / 文(最大148字 / 文)

・平均形態素数: 12.9個 / 文(最大63個 / 文)

英文訳文

・平均単語数: 10.3語 / 文(最大59語 / 文)

126

## 対象標本文の種類と数

文種別	述部の数	説明	日英対訳標本文数
文種別1	2	文接続1カ所を持つ文	57,235 (44%)
文種別2	3	文接続2カ所を持つ文	6,196 (5%)
文種別3	2	埋込み文1つを持つ文	46,907 (36%)
文種別4	3	埋込み文2つを持つ文	5,986 (5%)
文種別5	3	文接続と埋込文各1を持つ文	12,389 (10%)
- -	- -	合計	128,713 (100%)

127

## (2) 文型パターンの生成手順

形態素解析

解析誤りは人手で修正

線形要素の抽出

対訳辞書などを使用

線形要素の汎化

判定困難なものは人手判定

検証と改良

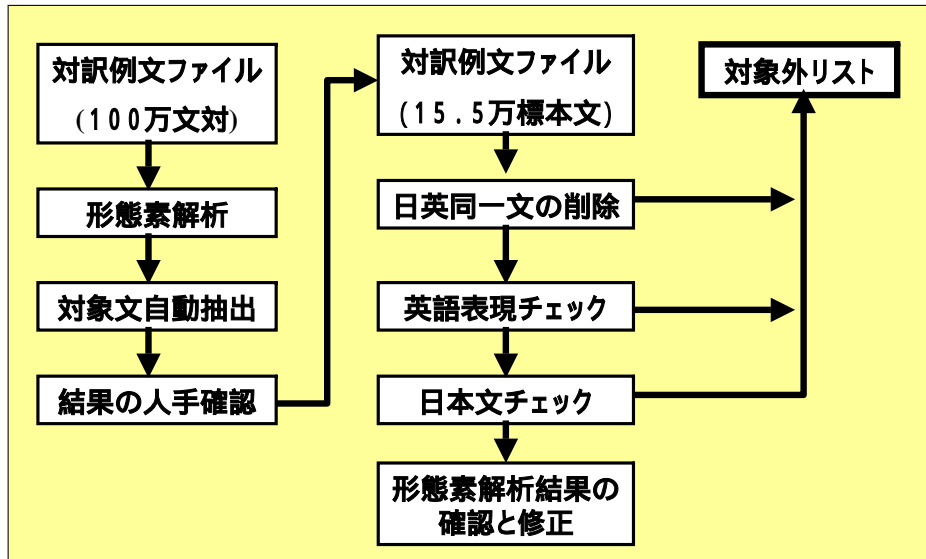
**半自動化の効果(初版)**

言語解析のためのマンパワー : **1 / 10**

128



## 文型パターン化の準備作業



## 3段階の汎化

言語表現の**文法的な構成単位**に着目

<単語レベル> 単語変数の例) **N3**, **V2**, **AJ3**, **ADV5**  
表現に含まれる名詞、動詞、形容詞、副詞などの**自立語**  
のうちの線形なものを変数化

<句レベル> 句変数の例) **NP3**, **VP2**  
名詞句、形容詞句、動詞句、副詞句など句のうちの線形  
なものを変数化

<節レベル> 節変数の例) **CL2**  
連体節、連用節などの節のうちの線形なものを変数化

130

## 汎化作業項目と作業基準

レベル	汎化の概要
単語 レベル	線形な自立語を単語変数化 (1)名詞、動詞、形容詞、副詞などの自立語の変数化 (2)文型上不要な要素の任意化。文型の骨組みとなる 要素の抽出 (3)字面要素のグループ化
句 レベル	線形な句を句変数化 (1)適用範囲を品詞から句へ拡大 (2)機能語の適用拡大(格助詞 格助詞相当語、等) (3)英語句生成関数の適用
節 レベル	線形な節を節変数化 (1)名詞節、副詞節、主節、従属節の1変数化による 重文複文の基本構造のパターン化 (2)日本語節から英語句への変換関数の使用 (3)その他英語構造生成関数の使用

## 【補足説明7】 3段階に汎化する理由

### 個別的なパターン

- <利点> 高い意味的排他性で  
高品質な意味解析が期待される
- <欠点> 適用範囲が狭い

### 汎用的なパターン

- <利点> 適用範囲が広い
- <欠点> 意味的排他性があまいで  
多義の発生が懸念される



### 組み合わせる方法

- ・ 低位レベルで適切なパターンが存在する場合  
それを使う
- ・ それが存在しない場合は  
一段高位に汎化されたパターンを使う

132

### (3) 単語レベル文型パターン化(初版)

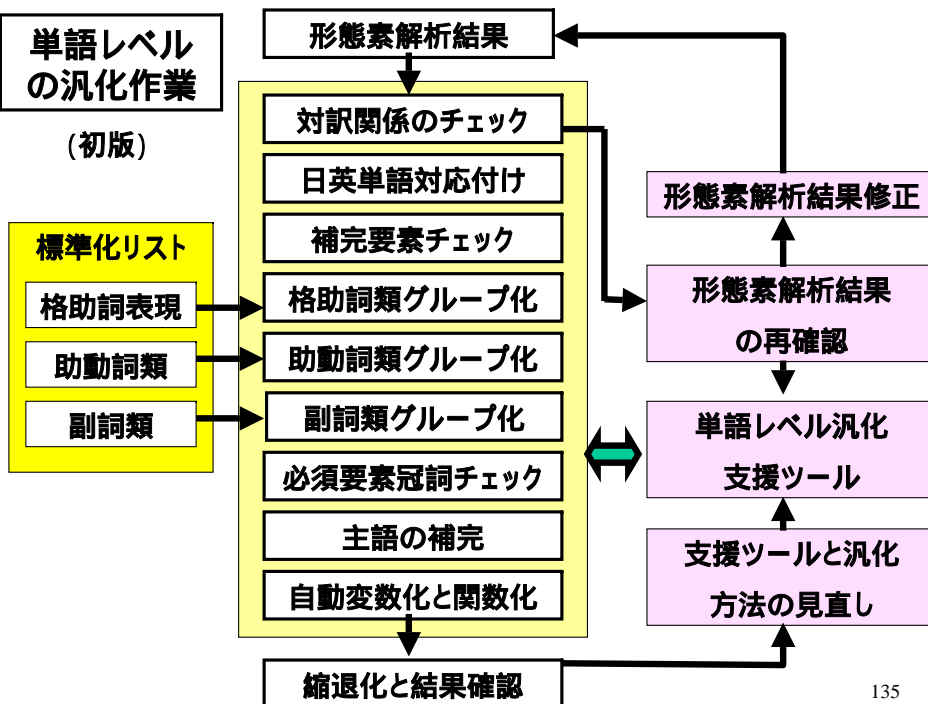
#	汎化規則の分類	汎化規則の内容	
1	原文任意要素の削除	原文任意要素を削除する	
2	自立語の変数化	名詞一般 ・格要素の名詞の変数化、・述語名詞の変数化 ・数詞の変数化(数詞として識別) ・複合名詞の単一名詞化	
3		複合名詞 述語単独動詞の変数化	
4		動詞一般 単独動詞連体形と連用形の変数化	
5		複合動詞 複合動詞の変数化	
6		形容詞・形容動詞	述語形容詞・形容動詞の変数化
7			単独形容詞形容動詞連体形、連用形の変数化
8	副詞	文修飾、用言修飾の副詞の変数化	
9	品詞変換関数の適用	品詞変換関数を使用する(表記法の例)V(N2)、N(V3)	
10	述部語尾表現	英語アスペクト情報の関数化	
11	の関数化	英語様相情報の関数化	

### 単語レベル文型パターン化(続き)

#	汎化規則の分類	汎化規則の内容	
12	パターン任意要素の指定	名詞句複合語 助詞結合型名詞句の単一名詞化	
13		名詞修飾語 単独の動詞連体形、形容詞、形容動詞連体形、連体詞の任意化	
14		用言修飾語	単独動詞連用形、単独形容詞、単独形容動詞連用形の任意化
15			副詞的用法の名詞(昨日、今日)の任意化
16		その他	英語に訳出されるその他線形要素の任意化
17			英語に訳出されない要素の任意化
18	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	
19	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。例)[N1]英語側のパターンは(N1 he)の形式で記述。	
20		冠詞の削除 英語パターン内の名詞の冠詞を削除。	

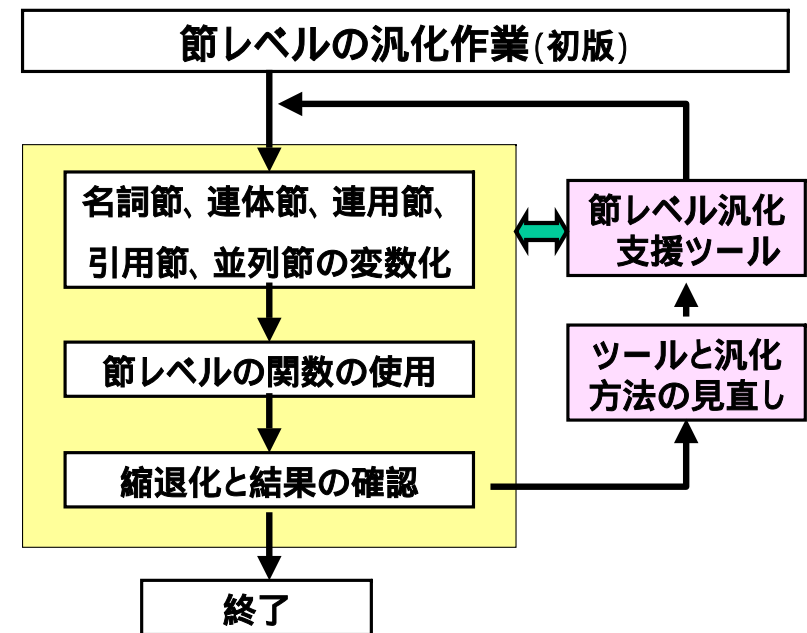
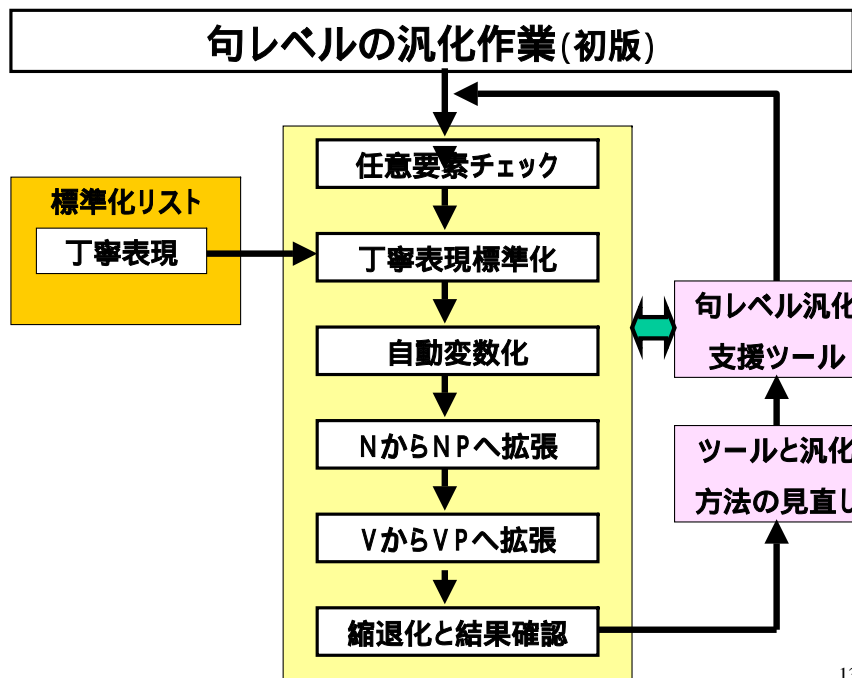
### 単語レベルの汎化作業

(初版)



### (4) 句と節レベルの文型パターン化(初版)

#	汎化規則の分類	汎化規則の内容	
21	句レベルの汎化	現在形変換 現在形でも使用される文型パターンを対象に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 <注> 英語訳出されない丁寧表現が対象	
23		適用範囲の拡大	NをNPへ拡張 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をVPへ拡張 名詞変数Vを名詞句変数VPに置換
25			機能語の拡張
26		述部の語尾を同種語尾パターンに置換	
27		同意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



## 4.2 文型パターン辞書の実験的改良

狙い: **被覆率向上**

- (1) 離散記号付与基準の精密化
- (2) 述部語尾表現(特に時制)の汎化
- (3) 英語パターン記述法の改良
- (4) その他の主な改良

### (1) 離散記号の付与基準の精密化

#### 1. 離散記号の使用目的

- ・原文任意要素への適用
  - (1) 被覆率の向上
  - (2) 4節以上の重文複文へも適合

#### 2. 経緯

初期の版: 1種類の離散記号(/: オールマイティ)例) /N1は、/N2を/V3。

改良版: **任意要素の種類を5種類に識別付与基準の精密化**

/y:連用節、/t:連体詞、/c:格要素、  
/k:形容詞連体形or連体詞、/f:副詞  
例) /yN1は、/ckN2を/fV3。

## (2) 述部語尾表現(特に時制)の汎化

例) 現在 ↔ 過去 ↔ 未来      肯定 ↔ 否定

関数の任意化と要素選択記号化

字面要素への関数付与

「対応型任意記号」と「対応型要素選択記号」の導入

「時制関数」の導入

「グループ関数」の導入 = “関数の要素選択記号化”

<同値型グループ関数>

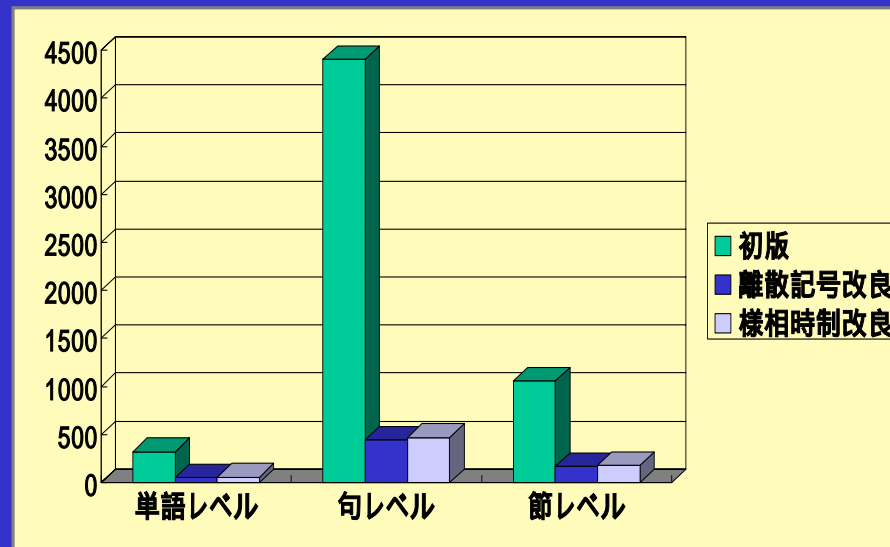
<対応型グループ関数>

その他の仕組み

- ・「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」

## 平均適合文型数の変化

「離散記号の細分化」と「様相時制の汎化」



## (3) 英語文型パターンの改良

日本語側が意味的に適切なパターンに適合しても英語側が使えないものが多い(初版では約半分)

例) 「彼は腕を組んで考えた。」 <群動詞の例>

/y#1{/tcfkN1は,/tcfkN2を}/cfV3(て|で)</yfN4を>!V5.kako.

N1 V5.past with N1^poss N2 V3^ed.

x He thought it over with his arms hold.

**英語パターン改良の必要性：大**

## 英語構文情報の付与

改良作業のための準備

英語表現の骨格が分かり易い記述法にすること

<対象> 文種別1 + 文種別3の標本文

<付与する情報>

- ・英語構文分類(83種類)
- ・日本語構文分類(178種類)

<参考>

- 語形変形関数ingの現在分詞と動名詞の区別  
現在分詞: V.ing、動名詞: V.grn (grnはgerundの略)
- 用言の用法(連体修飾用法と述部用法)の区別  
修飾要素: AJ、V、述部: 【AJ】、【V】
- 不定詞”to”と前置詞”to”の区別  
不定詞: ~ TO go there、前置詞: ~ to Yokohama

## 英語文型パターンの適正化

自己例文に適合しない英語文型パターン  
(初版15%以上:合計2万件)の分析と改良

### 副詞が挿入された完了形等の時制

副詞を持つ受身文や完了形の定義

受身を表すpassive関数の導入: V.passive = be V.ed

進行形を表すprog関数の導入: V.passive = be V.ing

挿入された副詞の記述方法導入: have 副詞 V.ed V.pft{副詞}

完了形の記述方法の明確化

関数の記述順序の明確化: 関数による時制/相/様相の複合表現

倒置関数inverseの導入(仮定法に対応)

Inverse(have 副詞 V.ed V.pft{副詞})

145

## 目的語が挿入された群動詞の記述

目的語が挿入された群動詞の英語表現の生成

英語文型パターンのVに目的語の範囲を明示

目的語が変数の場合 V2 N1-v2obj

目的語が字面の場合 V2 (it)-v2obj

群動詞の場合

例) N1 = it, V2 = carry outの場合

= V2 N1-V2obj carry it out

群動詞でない場合

例) N1 = it, V2 = fulfillの場合

= V2 N1-V2obj fulfill it

146

## 限定詞に関する新表記法

定冠詞、連体詞(「あの」、「この」等)に  
対応する英語表現が適切に定義できない。

「冠詞」、「指示形容詞」のいずれが省略されて  
いても、日英対応関係が成り立つようにする

### その他の記述法の改良

1) 前置詞の識別方法

2) 所有格代名詞の汎化方法

所有格代名詞と名詞との対応関係を記述

3) 省略された目的語に対する名詞の対応指示

英語目的語と名詞との意味的対応関係を記述

147

## (4) その他の主な改良

### 表記の揺らぎの吸収

標準表記への統一と選択記号の適正化

字面表記チェック 「原文表記 + 標準表記」

対象: 約3,000種類の表記

選択記号表示の漏れの吸収と表記の均質化

最大効果の推定 = 文型パターン数2倍に相当

### 任意記号の適正化

文型が単文化するような任意化は取りやめる

### 自己例文への適合性向上

自己例文に不適合な標本文(2万件)

148

## 【補足説明8】自己例文・自己パターンとは？

文型パターンは、日英対訳例文の汎化により生成

**自己例文**：各パターンから見て、当該パターンの元になった例文

**自己パターン**：例文から見て、当該例文から作成されたパターン



特別の場合を除き、自己例文は自己パターンに適合**パターン照合プログラムの動作検証に使用できる**

**他己例文**：自己例文以外の例文を言う

**他己パターン**：自己パターン以外のパターンを言う

オープン試験は他己パターンへの適合性で評価できる

## 4.3 文型パターンの記述例

### 3レベルパターンの例

- ・ 単語レベルパターンの例
- ・ 句レベルパターンの例
- ・ 節レベルパターンの例

## 節変数化の適用範囲の拡大

- ・ 文脈依存の省略要素を持つ節も汎化  
節and節、節but節、節or節の型（追加2,291件）
- ・ 命題レベル（当初版）から、  
様相時制を含む範囲へ（改良版）の拡大  
節レベルパターンが1万件増加

## 汎化不良パターン検出の半自動化

- ・ 変数適合表現の検証
- ・ 字面表記の適切性検証
- ・ 離散記号の利用
- ・ 単文適合性排除
- ・ 局所言い換え辞書の利用

## 単語レベル文型パターン

### 標本文

日本文：うっかりして定期券を家に忘れてきた。

英文： I was so careless as to leave my season ticket at home.

### 文型パターン

日本語： #1 [N1(G4)は] / V2(R3003)て / N3(G932)を / N4(G447)に / V5(R1809).tekita。

英語： #1 < N1 | I > was so AJ(V2) as to V5 #1 [N1\_poss] N3 at N4.

## 句レベル文型パターン

### 標本文

日本文：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。

英文：The conclusion is wrong in that it is based on a false premise.

### 文型パターン

日本語：NP1(G1022)は / V2(R1513).ta / N3(G2449) に / V4(R9100).teiruのだから / N5(N1453).dantei。

英語：NP1 is AJ(N5) in that it V4 on AJ(V2) N3.

## 節レベル文型パターン

### 標本文

日本文：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。

英文：It is significantly toxic so that great caution must be taken with its use

### 文型パターン

日本語：CL1(G2492).tearuので、N2(G2005)に当たっては / VP3(R3901).gimu。

英語：so+that (CL1, VP3.must.passive with subj(CL1)\_poss N2)

## 【補足説明9】 パターンがカバーする入力文

離散記号により適合範囲を拡大

重文複文を広くカバーするため

但し、  
単文には適合  
しない筈

X印を不良  
パターンの  
検出に利用



入力文の種別			適合するパターン種別				
種別	接続 文数	埋込 文数	文種 別1	文種 別2	文種 別3	文種 別4	文種 別5
単文	0	0	X	X	X	X	X
重文	1	0		X	X	X	
	2以上	0			X	X	X
複文	0	1	X	X			
	0	2以上	X	X			X
複重文	1	1		X		X	
	2以上	1				X	
	1	2以上		X			
	2以上	2以上					

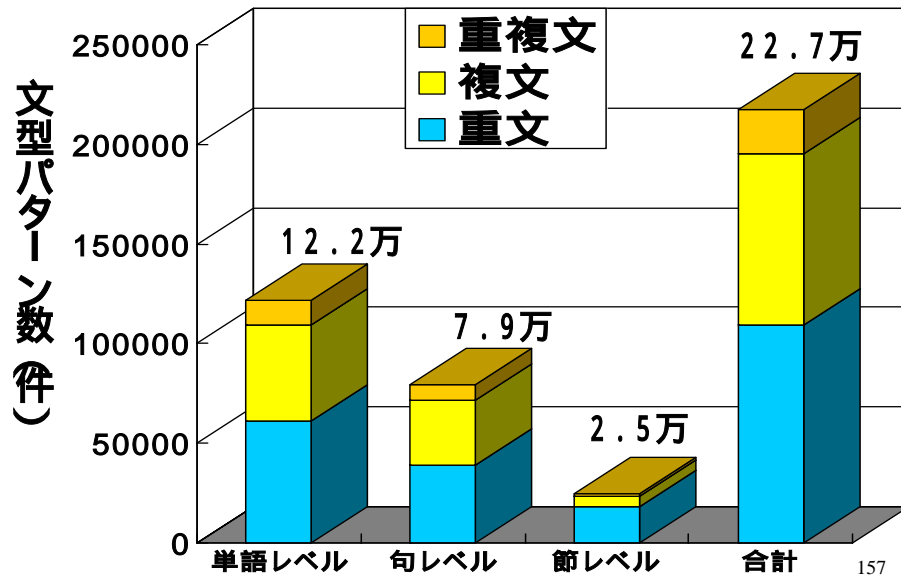
## 4.4 静的特性の評価

### (1) 文型パターン数

文種別		単語レベル	句レベル	節レベル	合計
重文	文種別1	55,508	36,002	17,859	109,369
	文種別2	5,663	3,241	314	9,218
	小計	61,171	39,243	18,173	118,587
複文	文種別3	42,485	28,040	4,998	75,523
	文種別4	5,638	4,009	780	10,427
	小計	48,123	32,049	5,778	85,950
複重文	文種別5	12,610	8,146	1,524	22,280
合計		121,904	79,438	25,475	226,817

- (1) 重文と複文が中心で複重文は少ない
- (2) ほとんどの対訳文から単語レベルを生成

## 文型パターン数



## (2) パターン構成要素の分析

### 品詞毎の出現回数

#	種別	形態素数	
		延度数	異り語数
1	名詞	406,896	56,861
2	本動詞	214,026	10,324
3	補助動詞	51,918	271
4	形容詞	30,491	915
5	形容動詞	17,646	2,562
6	副詞	35,873	3,191
7	連体詞	29,596	731
8	接続詞	3,146	77
9	感嘆詞	147	60
10	接頭辞	1,068	110
11	接尾辞	1,749	336
12	助動詞	165,251	236
13	助詞	465,811	349
14	記号	121,555	32
-	合計	1,574,613	76,055

## 変数の使用頻度と変数化の割合

	単語レベル		句レベル	
N	323,725 (79.6%)		VP	69,901 (49.2%)
ND	19,583 (38.7%)		NP	37,227 (12.9%)
TIME	11,095 (68.8%)		AJP	2,590 (13.9%)
NUM	5,470 (42.4%)		AJVP	1,359 (11.9%)
REN	5,872 (19.8%)		ADVP	282 (1.1%)
GEN	16,316 (55.1%)			
V	117,397 (54.9%)			
ADV	16,435 (45.8%)		合計	111,359
AJ	13,607 (44.6%)		節レベル	
AJV	13,218 (74.9%)		CL	39,718 (23.7%)
ANY	205 (--%)			
合計	542,833		合計	39,718

## 関数と記号の使用頻度(単語レベル)

関数(上位10件)		記号	
^rentai	44,851	離散記号(17種)	776,265
.kako	53,722	選択記号	144,882
.genzai	14,120	文節境界記号	72,531
.teiru	11,779	補完要素記号	63,194
.hitei	7,599	任意記号	53,712
.reru	6,140	位置変更	36,633
.masu	4,428	対応型選択記号	14,120
.youy	2,917	順序任意記号	13,523
^katei	2,810		
.rareru	2,054		
合計(48種)	174,283	合計	1,178,860



### (3) 線形要素の割合

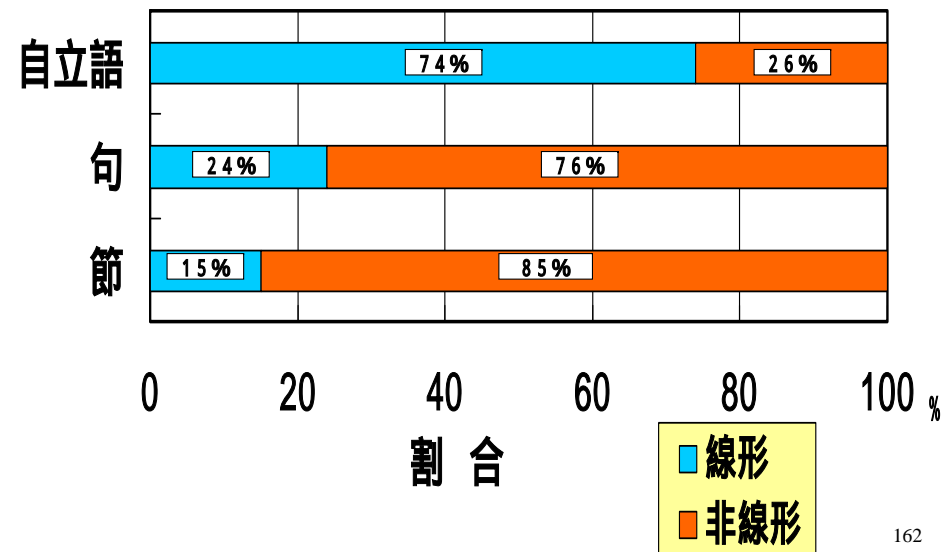
表現要素	出現頻度	変数化の数	線形要素の割合
自立語	734,528	542,833	73.9%
句	463,636	111,359	24.0%
節	267,601	39,718	14.8%

線形要素は、自立語でも、74%、  
線形な句は、24%  
節は、15%しか線形でない

従来のように、単文に分解して訳して合成しても、対訳例文のような訳文は生成不能

161

### 線形要素の割合



162

### 変数と関数による線形要素指定

文型パターン当たり

文型パターン種別	変数の使用頻度	関数の使用頻度	記号(離散記号を除く)
単語レベル	4.45個 (単語変数)	1.45個	2.19個
句レベル	1.40個 (句変数)	--	--
節レベル	1.56個 (節変数)	--	--

変数化、関数化される線形要素は、文当たり約8カ所  
挿入される離散記号は、文当たり約10カ所

163

### 線形なパターンと非線形なパターン

線形な標本文 >

字面を含まない文型パターン

- ・単語レベル = 15件
  - ・句レベル = 401件
  - ・節レベル = 155件
- 極めて少ない**

線形要素のない標本文 >

汎化できなかった標本文

- ・初期段階で削除 624件(対象外とした)
  - ・字面だけの文型パターン 単語レベルで302件
- 対象標本12.9万件中の割合 = 0.72%

**殆どすべてに線形要素がある**

殆どの文型は非線形だが、ほぼ線形要素を持つ  
文型パターン翻訳の可能性あり

## 4.5 被覆率特性の評価

### (1) 実験の条件

意味的制約条件 無視  
適合した文型パターンからの選択方式は  
今後の研究課題

テスト用入力文 1万文または200文  
対訳標本文の日本語からランダムに選択

実験方法 クロスバリデーション  
自己パターンへの適合は除外して評価する

**オープンテストの一種**

### 【補足説明10】クロスバリデーションとは？

#### 交差検定の一つの方法

n個のサンプルのn-1個を用いて  
モデルを構築(学習)し、残った  
1個を用いて評価する。  
この操作をn回繰り返す方法



少ない標本数で精度良い評価をするための方法  
人工知能のシステムの評価でよく使用される

本研究では、標本量は多いが、試験文も多数必要  
となるため、この方法を使用している

166

### (2) 評価パラメータ

**<適合率 R1>** (文単位に見た適合率) **統語的被覆率**

$$R1 = \frac{\text{1つ以上の適合文型があった入力文の数}}{\text{テスト入力文の数}}$$

入力文の一部の要素が文型パターンになくても良い

**<文型一致率 R2>** (文字単位に見た適合率)

$$R2 = \frac{\text{文型パターンでカバーされる  
入力文全体の文字数の最大値}}{\text{テスト入力文全体の文字数}}$$

入力文のすべての要素が文型パターンにある

**<平均適合文型数 (N = N1 + N2)>**

適合パターンがある入力文当たりの適合パターン数

完全一致文型数 = N1

部分一致文型数 = N2

**<正解率 P1>** **<ランダム選択方式>**

P1 = 適合文型が意味的に正しい確率

**<累積正解率 P2>** **<最適候補選択の限界>**

P2 = 各入力文に対する適合文型の中に意味的に  
正しいものが1つ以上含まれる確率

統語的被覆率...適合率R1で評価

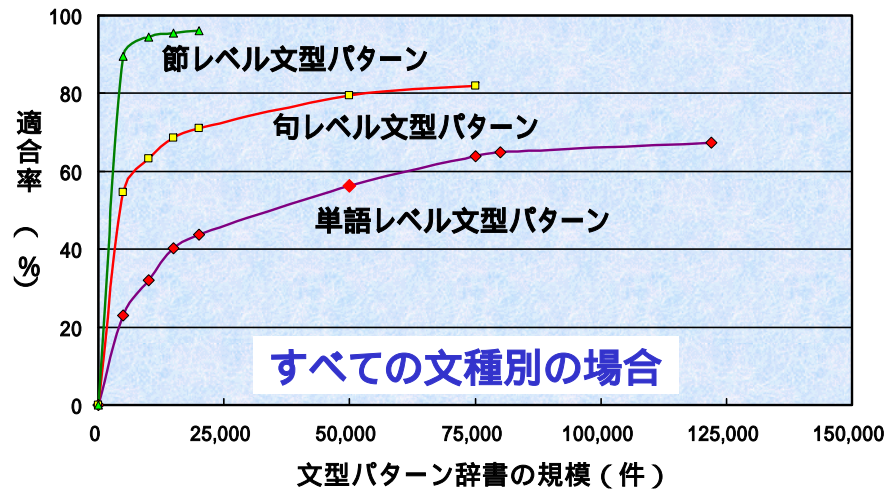
意味的被覆率...C = R × P2で評価

167

168

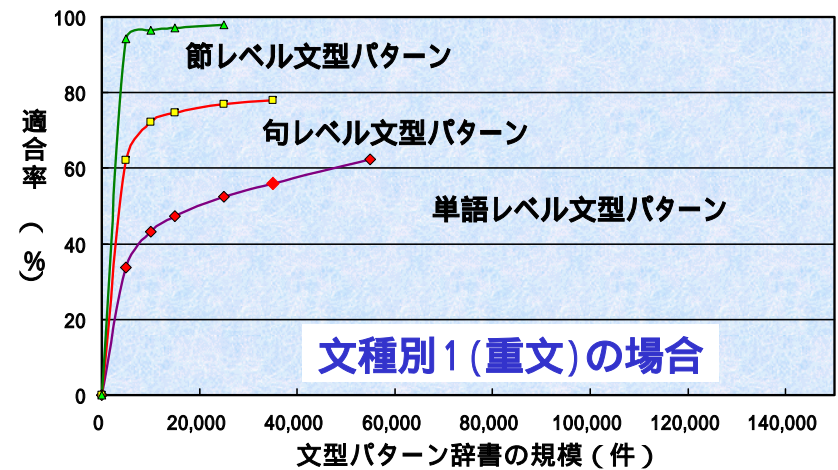
### (3) 統語的被覆率の飽和特性

文型パターン数と統語的被覆率R1の関係

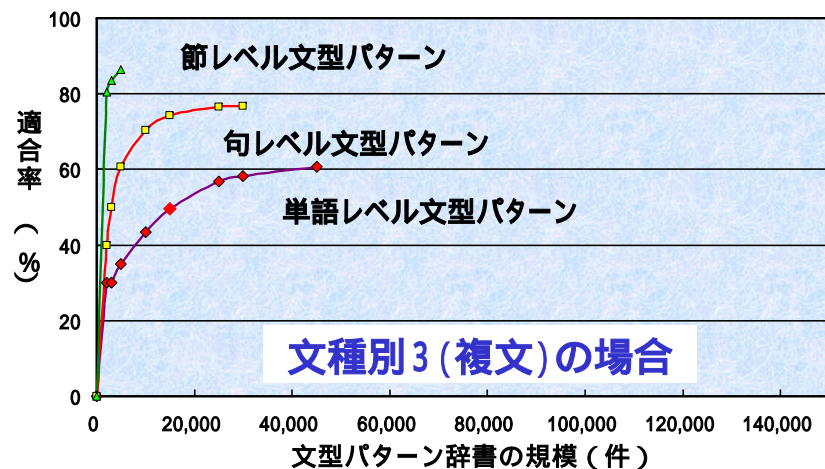


網羅性を得るためには、数万件以上のパターンが必要<sub>169</sub>

文型パターン数と統語的被覆率R1の関係



文型パターン数と統語的被覆率R1の関係



### 統語的被覆率R1の飽和特性について

- (1) 単語レベル, 句レベルの文型パターン  
数万件で飽和傾向
- (2) 文型パターン数が1万件以下では  
有効な被覆率は得られそうにない
- (3) 現状(10万件程度)以上に増加させても  
被覆率の向上はあまり期待できそうにない

➡ 汎化の方法がポイント

## < 閑話休題14 > 非線形なパターン数は？

非線形パターンは覚えないと使えない

(1) 人間はどれだけのことを覚えて、使っているか？

種類	個人の記憶	社会の記憶	
		社会人一般	専門的集団
漢字	3000文字	6000字	1万字
単語	3万語	10万語	500万語 固有名詞
単文パターン	2.5万件		
重文複文パターン	数万件？		



(2) プラトンの問題に答えられるか？

個人の記憶程度でのカバー率90%程度か？

歴史的、社会的蓄積の全体がカバーされなくても良い

175

## (4) 適合率と適合パターン数

文型パターンの適合率(テスト入力文:1万文)

レベル	適合率R1	文型一致率R2
単語レベル	62.6 %	45.2 %
句レベル	82.7 %	72.3 %
節レベル	96.4 %	91.0 %
混合レベル	97.8 %	92.7 %

R1: 任意要素への適合を含む

R2: 任意要素への適合を含まない

174

## 適合文型パターン数

(入力文1文に対する適合文型パターン数)

文型パターン種別	完全一致 文型数N1	部分一致 文型数N2	述べ適合 文型数N
単語レベル	8.4 件	54.7 件	63.2 件
句レベル	288.0 件	232.3 件	520.3 件
節レベル	211.2 件	1030.4 件	1241.7 件
混合レベル	507.6 件	1317.4 件	1825.2 件

汎化性が高く適合頻度の高いパターンが数百件程度存在する模様 代替パターンの存在するものは削除が望まれる

より強力な適合条件(意味的制約条件)の必要性

175

## (5) 適合率R1と正解率P1,P2

レベル	適合率 (R1)	正解率 (P1)	累積正解 率(P2)	意味的被覆率 (R1xP2)
単語レベル	62.6 %	26.8 %	64.4 %	36.7 %
句レベル	82.7 %	43.2 %	74.8 %	55.0 %
節レベル	96.4 %	32.0 %	69.4 %	67.0 %
Total	97.8 %	-----	-----	78.0 %

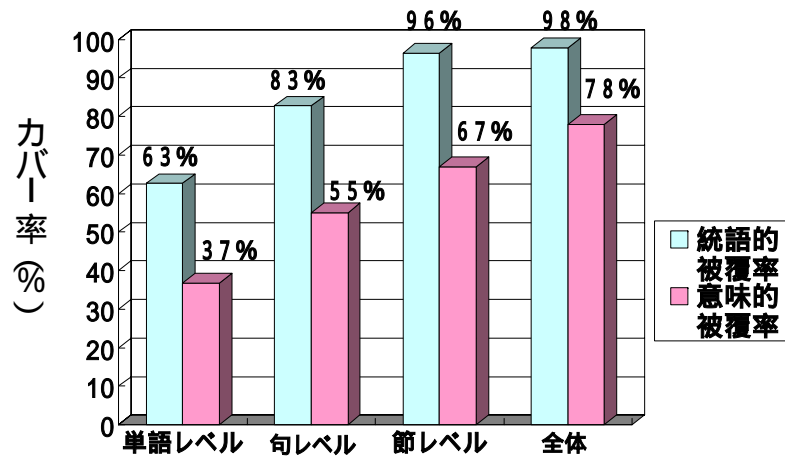
入力文には平均的に多数のパターンが適合し、意味的に不適切なものも多い

統語的には、全パターンで98.5%の入力文をカバー

意味的には、全パターンで、約80%の入力文をカバー

176

## 統語的被覆率R1と意味的被覆率C



レベル	単語レベル	句レベル	節レベル
平均適合パターン数	63件	520件	1242件

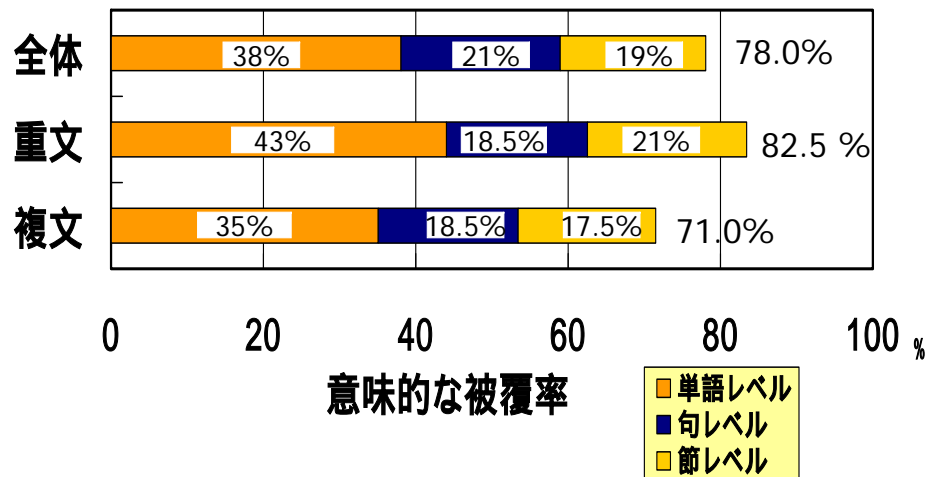
177

## 3レベルのパターンの適用順序について

- (1) **単語レベル**のパターン検索を行い、適切なパターンが存在すれば、それを使用する
- (2) 適切なパターンが存在しなければ、**句レベル**のパターン検索を行い、適切なパターンが存在すれば、それを使用する
- (3) 適切なパターンが存在しなければ、**節レベル**のパターン検索を行い、適切なパターンが存在すれば、それを使用する
- (4) 適切なパターンが存在しなければ、パターン方式を諦め、**従来の要素合成法**で処理する

178

## パターン種別とカバー範囲



句型パターンの適用順序

単語レベル

句レベル

節レベル

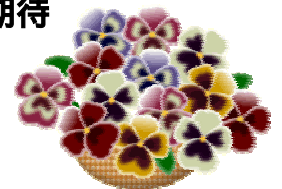
179

## 【補足説明11】 パターン数と被覆率

### 嬉しい誤算から生じた試行錯誤

- ・当初計画: 標本文5万件(20万件から抽出)
- ・実際: **標本文15万件**(100万件を入手)

比較的簡単な汎化で、被覆率はあがると期待  
被覆率は上がらず、試行錯誤を開始  
さまざまな汎化を実施



### 最終的な結果

- ・意味的なカバー率**約80%を達成**(改良の余地あり)
- ・被覆率向上の鍵は汎化方法
- ・パーサの判定能力も重要
- ・被覆率はパターン数5万件程度で飽和
- ・**パターン数は5万件程度に圧縮可能**

180

## 【補足説明12】 関数と記号の汎化効果

変数化のみの場合との比較（単語レベルの場合）

#	要素の種類	R1の向上効果	パターン拡大率	参考
1	関数を使用	2.2倍*	10倍	*.kako:1.5倍
2	記号を使用	3.4倍*	15倍	*離散記号:2.8倍
3	両者を使用	<b>7.6倍</b>	80倍	

(注1) 変数、関数を使用しない場合との比較実験  
(注2) パターン拡大率は推定値

【考察】 (1) 関数、記号による汎化効果は大  
(2) 記号による汎化効果の方がより大



181

## 4.6 第4章のまとめ

- (1) 文型パターン辞書を開発  
 <対象> 述部2または3の重文と複文  
 <規模> 単語 / 句 / 節レベルの**合計22.7万件**
- (3) 線形要素の割合  
 自立語 = 62% 句 = 22% 節 = 9.6%  
 非線形要素が多い  
 要素合成法では殆ど訳せない
- (4) 被覆率  
 ・ 統語的な被覆率 98%  
 ・ **意味的な被覆率 80%** 実用化が期待

182

## 【補足説明13】 文型パターン辞書圧縮の可能性 被覆率の低下なしにどこまで圧縮できるか？

### 1. クロス照合実験による削除

(1) 12万件の入力文に不適合パターンを削除

**合計8.5万件程度に縮退**

(2) 上位レベルのパターンを削除

下位レベルが適合する場合、その上位レベルを削除

**合計6万件程度に縮退**

### 2. 高頻度適合パターンの削除

代替できるパターンがあるもの **2000件程度削減**

### 3. 包含されるパターンの削除

包含されるパターン(英語構造が同じであることを)を削除

**5000件程度削減**

**最終的なパターン数 = 5万件程度？**



183

## 【補足説明14】 意味類型パターン辞書の圧縮

<条件> 被覆率を低下させないこと

- <手順>
- (1) クロス照合試験による絞り込み(自動)
  - (2) 包含関係による削除(半自動)
  - (3) オープン試験による復活(自動)
  - (4) 日英構文対応関係分析に基づく調整(人手)

<結果>

辞書種別	原本辞書		圧縮版辞書		圧縮率
	パターン数	適合率	パターン数	適合率	
単語レベル	121,904	62.6%	20,040	62.6%	16.3%
句レベル	79,438	82.7%	20,167	79.6%	25.4%
節レベル	25,475	96.4%	9,636	94.4%	37.8%
合計	226,817	97.8%	49,843	96.9%	20.0%

184

## < 閑話休題15 > 大切なのは量か質か？

嬉しい誤算から生じた試行錯誤と  
それによって得られた教訓

< 第1ステップ > : **初版の構築** C = 20~30%

- ・当初計画の3倍の標本量が得られたため
- ・変数中心の簡易な汎化で高い被覆率を期待

< 第2ステップ > : **汎化の高度化** C = 60~70%

- ・被覆率が低く試行錯誤
- ・高度な関数と記号類を充実

**知識化**

< 第3ステップ > : **詳細な改良** C = 80%~

- ・緻密な改良とデバックによる相乗効果

【結果】飽和特性（Cは、ほぼ5万パターンで飽和）は変化なし  
結局、Cは、汎化内容と作業品質に伴って向上

【教訓】量<sup>重要</sup>だが、質はそれ以上に重要

統計翻訳・用例翻訳・・・質を忘れて、量に頼っていないか？



185

## 第5章 意味的等価変換方式と 文型パターンの意味類型化

### 第5章の目次

- 5.1 意味的等価変換方式の概要
- 5.2 重文・複文の意味類型化
- 5.3 日英構文対応関係の分析
- 5.4 第5章のまとめ

## 5.1 意味的等価変換方式の概要

### (1) 理論的背景

(1) **意味類型論**： 有田潤 1987

- ・認識の枠組み = 母国語言語の構造  
翻訳の基本原則

➡ **パターン化と意味によるグループ化**

(2) **等価的類推思考の原理**： 市川亀久彌 1960

- ・科学的発明の思考 = 類推の原理

➡ **意味分類コードによるパターン間の写像**

187

### 【補足説明15】 意味類型とは？

有田潤(有田1987)

- ・意味類型は具体的言語表現の一段奥にある**思考形式**のごときもの
- ・その性質上各自の母語をもってこれを考え、理解し、また整理することができる
- ・もし諸言語に共通のものがあって、それを取り出すのが**普遍的な言語理論の仕事**であるとするなら、意味類型の考え方こそ、その可能性を開く唯一のものであろう
- ・**自動翻訳機**がもし本当に自然言語を翻訳しうるようになれば、その原理もおそらくここに求められるに違いない



188





# 類推思考と翻訳

## 「類推とは」

あるものを似たものに例えることにより思考する形式を持つ  
 <特徴> **非決定論的** 一般的な規則を使用しない( 推論)  
 <仕組み> 複合的機能  
 「覚える」、「思い出す」、「一般化する」、「特殊化する」

## 「翻訳とは」

ある言語の網の目で掬い取って表現された話者の認識を別の言語の網の目で掬い直し改めて表現すること

### <方法>

1. 言語表現の意味類型化  
表現構造をもれなく形式化し、形式と概念との関係を対応づける
2. 意味類型間の写像( 変換)  
**原言語と目的言語の「意味類型」を概念を介して対応付ける**

# (2) 意味的等価変換方式

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

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

A : 言語 上の表現、 B : 言語 上の表現  
 : 変換、 : 写像、 : **真理項 (概念の集合)**  
 P(A) : 表現 A の「形式」を求める関数  
 C( ) : 真理項からある条件下で言語 の形式を求める関数

の場合 : 翻訳方式  
 = の場合 : 同一言語内の言い換え方式

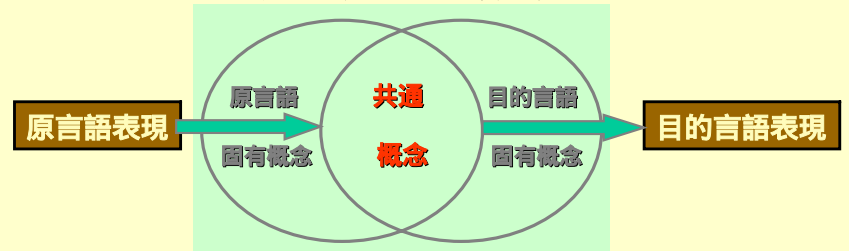
## 「真理項」を用いた「意味類型」間の意味的等価変換

#	日本語意味類型パターン	論理的意思範疇	#	英語意味類型パターン
1	X1は、X2がX3するようX4する	真理項の集合	1	X1 X4 so that X2 X3
2	X1は、X2が大変X3なのでX4できない		2	X2 is so X3 that X1 cannot X4
3	X1はX2がX3するといけなのでX4する		3	X1 X4 for fear that X2 X3
4	X1はX2するといけなのでX3した		4	X1 X3 not to X2
5	X1はX2しないようX3した		5	X1 is X3 for X1 is X2
6	もしX1がX2したら、X3はX4する		6	X3 X4 in the case X1 X2
7	X1がX2したらX3はX4した		7	When X1 X2, X3 X4
8	X1がX2したときX3はX4した		8	If X1 X2, X3 X4
9	X1がX2するならX3はX4してもよい		9	If X1 X2, X3 may X4
0	X1はX2なのでX3だ		0	X3 may X4 provided that X1 X2

<レベル1>	<レベル3>
比較	漠然、一般、くどい、重々しい、反事実、限度、所属、原因、譲歩、性質、継続、推量、感慨、関連、嗜好、状況、状態、引用(説明)、対句的、代替、定義づけ、適正(忠告、禁止、勧誘、命令)、伝聞、転換、結果、決定、限定、使役、事実、時点、自動的、主題、十分、同時、判明、反復(習慣)、比例、頻度、不定、付帯状況、並列、意志、可能性、可能、受身、関係、許可、えん曲、試行、選択、能力、量、条件、応答、過去、継続、概数、起点、完了、疑問、逆接、経過、経験、断定、割合、軽視、家庭、感嘆、期待、必要、やわらげ、目的、外見、否定、肯定
<レベル2>	
同一、比喩、同関係、同級、添加、同様、以上以下、換言、比較級、対比、倍数、相違、割合、折一、最上級、複数	

# 意味類型と機械翻訳

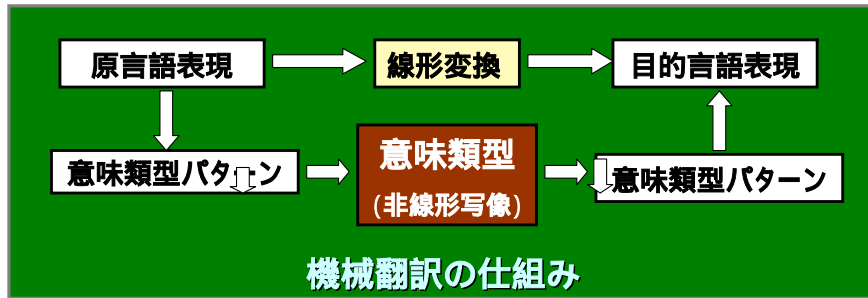
## 概念を介した翻訳の原理



## 概念を介した表現の対応付け

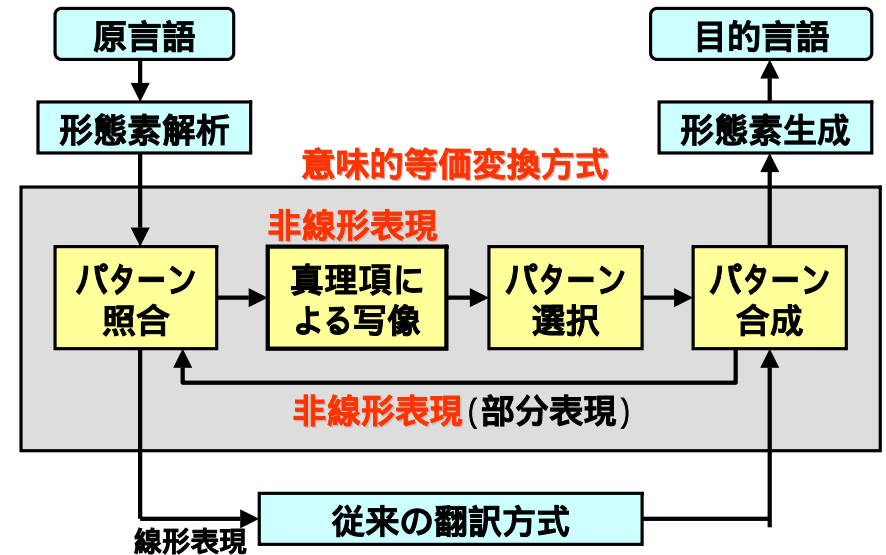
単語レベルの対応付け = **単一概念**を介した対応付け  
 表現レベルの対応付け = **複合概念**を介した対応付け  
**構造的書き換え規則**による従来の機械翻訳は不適切

# 意味類型を介した機械翻訳

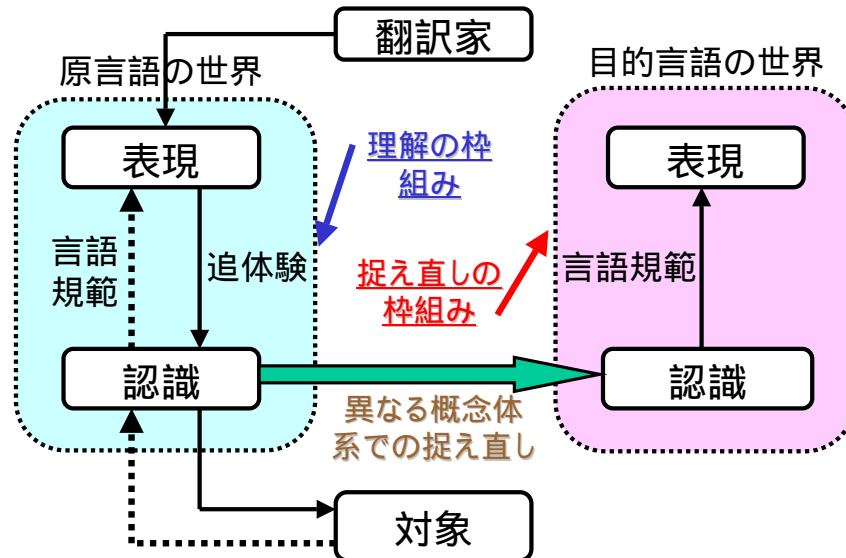


- < 場合1 > ある実体または事象の認識において、原言語と目的言語の双方に対応する概念が存在する場合  
**意味類型を介した写像**
- < 場合2 > ある実体または事象の認識において、原言語には対応する概念があるが目的言語にはない場合  
**下位概念を用いた説明**

# 翻訳システムの構成例



## 【補足説明16】 人間による翻訳のプロセス



## (3) 本方式とパターン翻訳の違い

### 意味的等価変換方式の狙い

意味類型化による文型パターンのグループ化と真理項を介したグループ間の意味的対応付け  
= 入力文に対する**訳し分け機能**を実現

### 従来のパターン翻訳方式

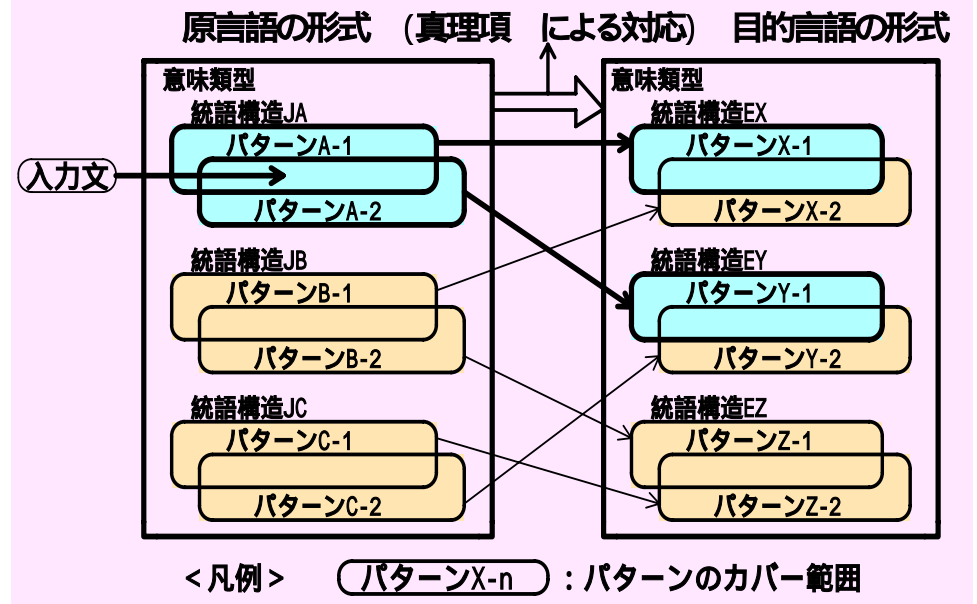
パターン数増大 複数のパターンに適合  
= 同様の効果あり

両者の違いは？

## 意味類型化による複数の訳文候補の生成

日本語表現の形式	真理項	英語表現の形式
1 N1はN2のN3にならないようにV2.	予断的 行為	1 N1 V2 for fear that N1 should V(N3) N2.
2 N1はN2のN3にならないようAD1にする.		2 N1 V2 for fear that N1 should V1.
3 N1はN1されるのを恐れてV2.		3 N1 V2 for fear N1 should be V1+ed.
4 N1はN1がV1するといけないのでV2.		4 N1 V2 for fear that N1 would V1.
5 N1はN2がV1するとまずいからV2.		5 N1 V2 for fear that N2 would V1.
6 N1はN2にV1とV1ないと思ってV2.		6 N1 V2 for fear N1 would V1 N2.
7 N1はN2のN3になるといけないのでV1.		7 N1 V2 for fear of V1+ing
8 N1はN1がV1しないようにV2する.		8 N1 V2 for fear of N2.
9 N1はN2がV1しないようにV2する.		9 N1 V2 just in case N1 V1.
10 N1はN2のN3をV1しないようV2.		10 N1 V2 so that N1 may not V1.
11 N1はN2がN3をV1しないようV2する.		11 N1 V2 so that N1 would not V1.
12 N1はN2をV1しないようにV2.		12 N1 V2, lest N1 should V1.
13 N1はN2がV1しないようV2.		13 N1 V2 lest N2 should V1

## パターン翻訳と意味的等価変換の違い



## 複数の目的言語表現形式からの選択

表現形式記述の情報 = 真理項の設計で吸収  
**訳し訳けの情報** = 実際の翻訳時点でのみ得られる情報

### (1) 文内情報

- 形式化の過程で失われた情報
- 元の例文にはなかった要素の持つ情報
- 例) 表現要素の意味性質、形式に挿入された各種修飾要素など

### (2) 文間情報

- 前後の文から得られた情報
- 例) 選択する表現形式が代名詞の意味

### (3) 文外情報

- 利用者が予め指定しておける情報
- 例) フォーマル/インフォーマルの区別
- 書き言葉/話し言葉の区別、場面情報

## 【補足説明17】 方式限界について

### 本方式の適用対象

- 「直接的言語表現」: 言語規範に直接媒介された表現
- × 「間接的な言語表現」: 比喩(直喩を除く)など



### 比喩認知の枠組み(3つの基本要素)

- 「たとえるもの(趣意tenor)」
- 「たとえられるもの(媒体vehicle)」
- 「たとえの根拠(ground)」

「類推思考の前提」同じ関係が成立

比喩理解 = 与えられた同一言語の表現 A と B  
 に対して、共通見地 を求める問題

限界突破の鍵も類推原理にある

## 5.2 重文・複文の意味類型化

### (1) 真理項の設計条件

(1) 両言語に**共通な概念**であること

【対策】対訳パターンの意味的關係に着目

原言語の表現形式の表す概念を体系化  
それを対訳關係にある目的言語の形式と  
共通する真理項とする

(2) 形式の意味に対応した**粒度**をもつこと

【対策】翻訳はあくまで近似 近似度の向上

真理項の階層關係による形式の指定  
複数の真理項を用いた形式の指定

### (2) 真理項の構成

重文・複文・・・事象間の關係の表現

主節と從屬節の關係の意味分類  
= **節間の意味分類**

単文・・・・・・・・事象の表現

節の意味分類

206

### 分類体系の構成

(1) 統語構造による分類(24種類)

(2) **意味による分類**

・ **重文複文の意味分類体系(222種類)**

➡ **新規に研究開発**

・ 節の意味分類体系(735種類)

(3) 参考情報

・ 節間キーワード(約500種類)

・ 日本語構文分類(178種類)

・ 英語構文分類(83種類)

207

### (3) 統語構造による分類

文型	述部	説明	文種	第1分類	第2分類
重文	2	接続1	種別1	連用節 + 主節 (1タイプ)	2種
	3	接続2	種別2	連用節 + 連用節 + 主節 (入れ子型 + 並列型; 2タイプ)	8種
複文	2	埋込み1	種別3	連体節 + 名詞 + 述部 (1タイプ)	2種
	3	埋込み2	種別4	連体節 + 名詞 + 連体節 + 名詞 + 述部 (1タイプ)	4種
複重文	3	接続、 埋め込み各1	種別5	連用節	4種
				連体節 + 名詞 + 述部 + 主節(1タイプ)	4種
				主節	4種
				連用節 + 連体節 + 名詞 + 述部(1タイプ)	4種
合計			5	7	24種

208

# (4) 節間の意味分類体系

## 新しい意味分類体系を作成

< 第1ステップ > : **従属節の分類** (益岡・田窪1992)  
 名詞節、連体節、連用節、並列節の意味  
 意味的分類用語90種類を抽出し体系化

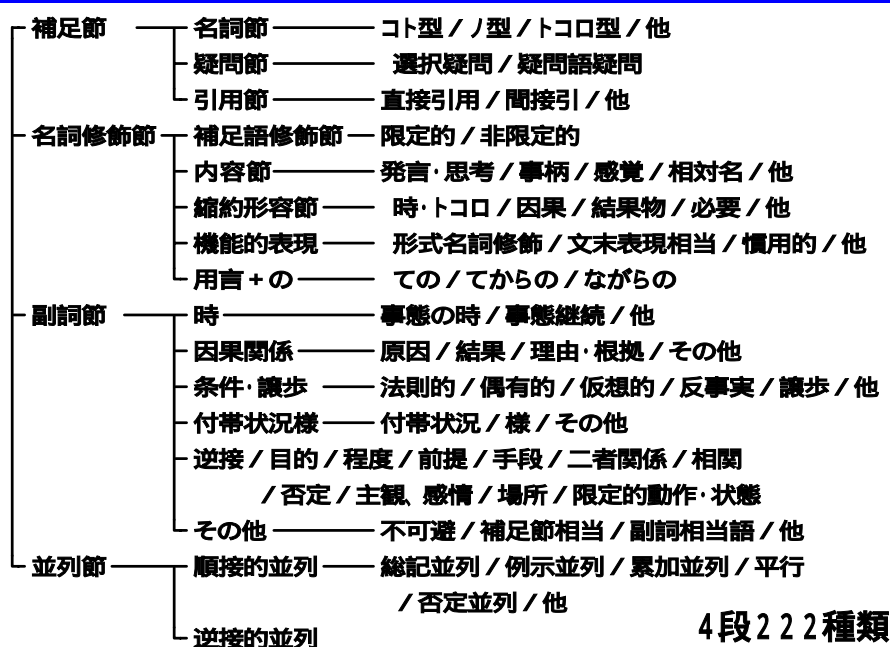
< 第2ステップ > : **文型パターン1000件の分析**  
 「時間的關係」、「空間的位置關係」、  
 「論理的關係」、「心理的關係」  
 の4種類の観点から詳細に分類

< 第3ステップ > : **統合再編成**  
 222種類の項目で4段階の階層構造

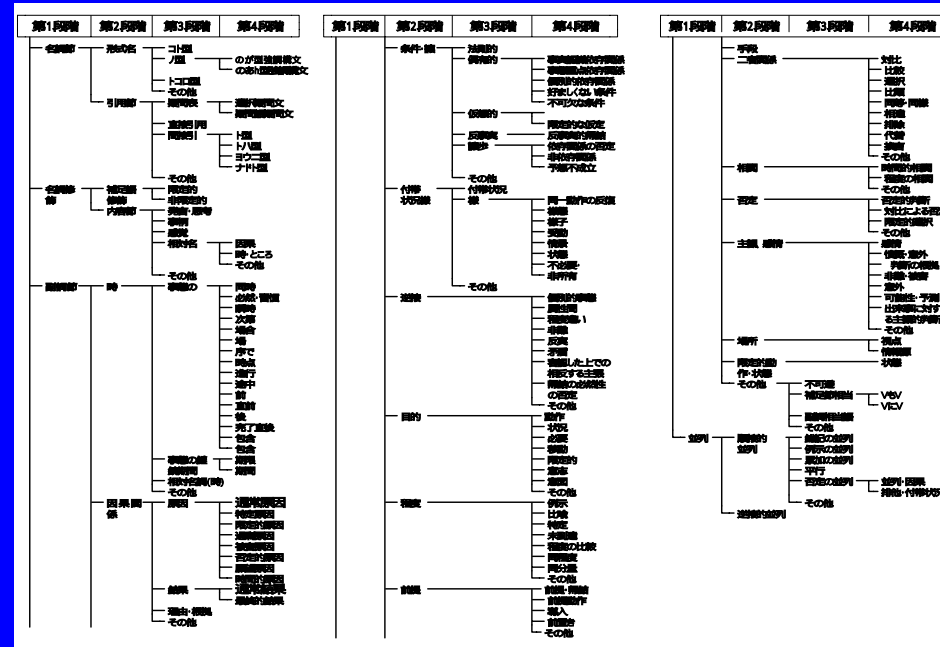
# 節間意味分類の数

#	第1段	第2段	第3段	第4段	合計
1	補足節 (補語相当節)	3	8	17	29種
2	名詞修飾節 (連体節)	6	6	12	25種
3	副詞節 (副詞的連用節)	16	17	121	155種
4	並列節 (並列型連用節)	2	6	4	13種
計	4種	27種	37種	154種	222種

## 節間意味分類体系 (1~3段まで)



## 節間意味分類体系 (全体)



## (5) 節の意味分類

単文(命題レベル)の意味分類に従う  
= 3分類 (合計735種類)=

### (1) 動詞文の節

動詞の意味分類(295種類)

### (2) 形容詞文(形容動詞を含む)の節(性状規定)

形容詞と形容動詞の意味分類(75種類)

### (3) 「だ文」(名詞述語文)の節(判断措定)

名詞の意味分類(365種類)

213

## 節の意味分類体系

分類	第1段	第2段	第3段	第4段	第5段	合計
動詞節	1	9	49	147	88	295
形容詞節	1	1	6	34	33	75
名詞述部節	1	1	41	209	114	365
合計	3	11	96	390	235	735

### 命題レベルの節の意味を分類

= 述部の主用言または主名詞の意味 =

- 【考慮した点】
- (1) 同義語・類義語の吸収
  - (2) 統語的情報の考慮
  - (3) 場面依存の意味分類
  - (4) 多義語への対応

214

## (6) 分類コードの付与

### 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
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
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
第4欄	第1従属節意味分類コード	第1従属節述部用言の意味分類コード
第5欄	第2従属節意味分類コード	第2従属節述部用言の意味分類コード
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
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意味コード
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意味コード
第9欄	第1節間KW	第1係り受け関係にある節間のKW
第10欄	第2節間KW	第1係り受け関係にある節間のKW

215

## 「パターン種別」と分類コード」の関係

欄の番号	第1欄	第2欄	第3欄	第4欄	第5欄	第6欄	第7欄	第8欄	第9欄	第10欄
真理項種別	統語分類		節の意味		節間の意味		KW			
意味分類コード	文種別	統語第1分類	統語第2分類	主節	第1従属節	第2従属節	第1節間	第2節間	第1節間	第2節間
パターン種別	種別1	統語第1分類コード	統語第2分類コード	主節	連用節	- -	接続	- -	KW	- -
	種別2			主節	連用節	連用節	接続	接続	KW	KW
	種別3			主節	連体節	- -	埋込	- -	KW	- -
	種別4			主節	連体節	連体節	埋込	埋込	KW	KW
	種別5			主節	連体節	連用節	埋込	接続	KW	KW
種別5	主節	連用節	連体節	接続	埋込	KW	KW			
分類総数	5	7	24	742	742	742	222	222	-	-

216

## 分類コードの付与対象

- (1) 単語レベル文型パターン・・・121,904件
- (2) 句レベル文型パターン・・・79,438件
- (3) 節レベル文型パターン・・・25,475件

但し、句レベル、節レベルのパターンは  
単語レベルのパターンをさらに汎化したもの

分類コードは、単語レベル文型パターン  
からコピーできる

217

## 分類コードの付与方法

1. **統語分類コードの付与**  
形態素解析、構文解析の結果に基づく自動判定
2. **節の意味分類コードの付与**  
単語意味辞書と構文解析結果から述部用言、  
体言の意味属性を付与  
= 命題レベルでの単文の意味分類
  - (1) 動詞と形容詞述部の節  
用言意味辞書(6000語)を作成  
機械的に付与
  - (2) 体言述部の節  
名詞意味辞書(6万語)を作成  
機械的に付与

## 3. 節間意味分類コードの付与

= 半自動化 =  
表現の形式と意味の関係に着目

- (1) 形式分類用テンプレートの作成  
統語的分類用テンプレート約90種類を作成
- (2) テンプレートと意味(複数)の対応付け
- (3) テンプレートを使用した用例の分類  
用例分類によりパターンを分類
- (4) 意味コードの付与  
パターングループ毎に付与(人手)

219

## (7) 「統語・意味分類コードの例

日本語	文種別 1 接続が一箇所	統語分類コード	副詞型	副詞節 + 主節
従属節	/v\$ <sub>1</sub> /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	\$ <sub>1</sub> <sup>^</sup> {/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)			
	第一連用節 意味分類コード	FUc201	事実認識依存関係	第一節間キーワード と
英語	英語構文 情報	接続詞相当	英語構文 キーワード	when
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.			

# 5.3 日英構文対応関係の分析

< 本来の目的 > 実験的改良作業の支援情報

標本原文に分類コードを付与

改良対象パターンの半自動的抽出

< 分類数 >

1. 日本語表現構造の分類

分類対象: 文種別1と3 = 9.7万文

分類数: 178種類

2. 対応する英語表現構造の分類

分類数: 83種類

< 対応分析 >

日英構文対応関係(178 × 83 = 14,774)を分析

221

## 日本語重文複文に対応する英語表現の分類

#	大分類(14分類)	中分類(83分類)	小分類
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	98
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	74
3	副詞相当表現	副詞小節前置き表現、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	400
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	11
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	57
6	述部相当表現	モダリティ相当表現	70
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	45
8	主節の他要素変換	主節 疑問節、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	28
9	埋め込み節 + 被埋め込み名詞 名詞化		1
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	2
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、同時を表す構文、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	117
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	11
13	その他	挨拶	2
14	意識・個別的表現		1
計	14分類	83分類	916

## 日英の表現で対応関係を持つ標本の数

(日本語178分類 × 英語83文型: 分類対象 = 9.7万文ペア)

構文対応関係		英語の構文分類									
		1 等位接続 詞相当	2 従属接続 詞相当	3 接続 副詞	4 節を導く 副詞	5 分詞 構文	6 名詞 節	...	82 挨拶	83 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	ノ型	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	

## 日英構文対応関係の分析結果

構文対応関係の可能性: 178分類 × 83分類 = 14,774組

対応する パターン数	対応関係を持つ 構文の組	日本語 英語		(注)英語 日本語	
		分類数	平均対応 構文数	分類数	平均対応 構文数
1以上	2023組(13.7%)	178分類	11.4種	83分類	24.4種
10以上	658組(4.5%)	110分類	6.0種	75分類	8.8種
100以上	163組(1.1%)	52分類	3.1種	46分類	3.5種
1000以上	17組(0.1%)	9分類	1.9種	13分類	1.3種

### 基本文型パターンの選択方法の例

100件以上出現する構文の組(1000組)

× 代表的パターン10件 = 5万パターン



## 【補足説明18】 意味類型化作業の困難さ

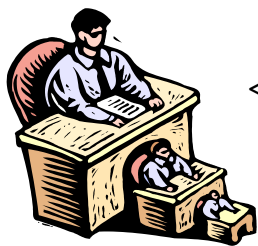
- ・熟練者が必要・・・重文複文の意味の判定
- ・作業量が膨大・・・12万パターンが対象

見積もり: <初版の試作>

30分 × 12万件 = 6万時間 = 30人年

30人年 × 500万円 / 人年 = 1.5億円

<実験と改良> ほぼ同上



- ・対策: 半自動的な意味判定ができるような分類体系を作成
- ・結果: 意味的等価変換方式の検討のための実験環境を実現

225

## 5.4 第5章のまとめ

- (1) 意味的等価変換方式を提案  
意味類型論 + 等価的類推思考の原理
- (2) 真理項の設計と意味類型化
  - ・統語分類(24種類)
  - ・節間意味分類体系(222種類)
  - ・節の意味分類体系(740種類)
- (3) 日英構文の対応関係の分析  
日本語構文(178分類) × 英語構文(84分類)  
語学教育の教材への活用が期待

226

## 第6章 パターン検索プログラム

### 第6章の目次

- 6.1 パターンパーサ
- 6.2 パターン意味検索プログラム
- 6.3 第6章のまとめ

227

## 6.1 パターンパーサ

### (1) 機能条件と設計条件

#### 機能条件

入力表現の構造に適合するパターンを検索

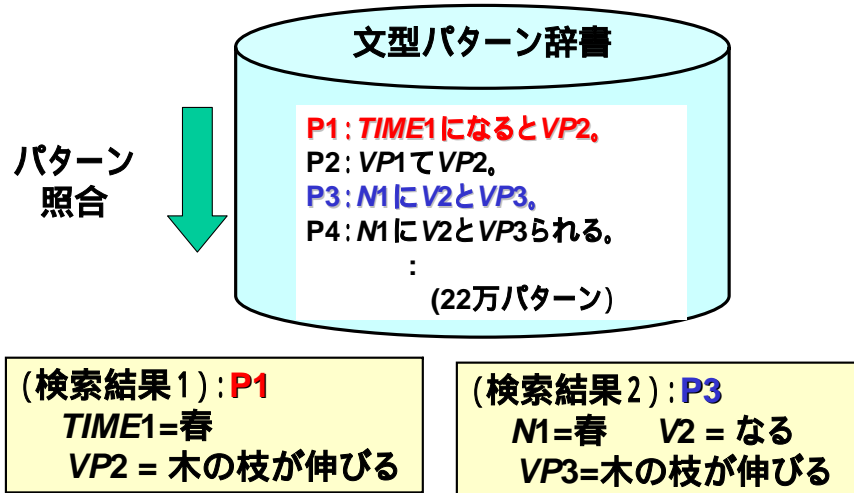
#### 設計条件

- ・大規模(20万件)のパターン辞書が対象
- ・複雑なパターン構造を高速に検索できること  
クロス照合実験(注) 24時間  
(注)テスト文12万件 × 20万パターンの照合
- ・適合するパターンを全て求めること
- ・バインド値(代入値)を全て求めること

228

## パターンパーサの機能条件

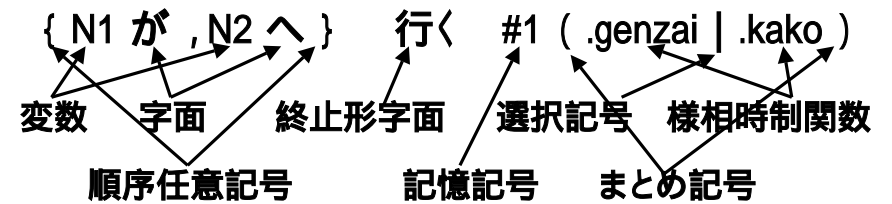
(入力文) 春になると木の枝が伸びる。



## (2) パターン記述要素の種類

< 4種類のパターン記述要素 >

- ・ 字面: 非線形要素の記
- ・ 変数: 線形要素の記述
- ・ 関数: 付属語要素の記述と語形指定
- ・ 記号: 構造上の線形性の記述



230

## 変数の意味的制約条件のチェック

【日本語パターン】

/y\$1^\{cf NPIは\}/tcfkああ/f見えても\$1/y AJ2.

意味的制約条件 入力要素の意味的適合性を検査

{NP1(IM:12100, IM:13660), AJ2(IY:A5B0)}

【英語パターン】

NPI be still AJ2 despite appearance.

< 適合する入力文の例 >

日本文: あの建物 は ああ見えても 新しい。

英文: That building is still new despite appearance.

231

## (3) パターン照合基本機能

{ N1が, N2へ } 行く #1 (.genzai | .kako)

(入力文) 学校へ 太郎が 行った

バインド値: M1 = 太郎 N2 = 学校 #1 = .kako

語形関数の場合

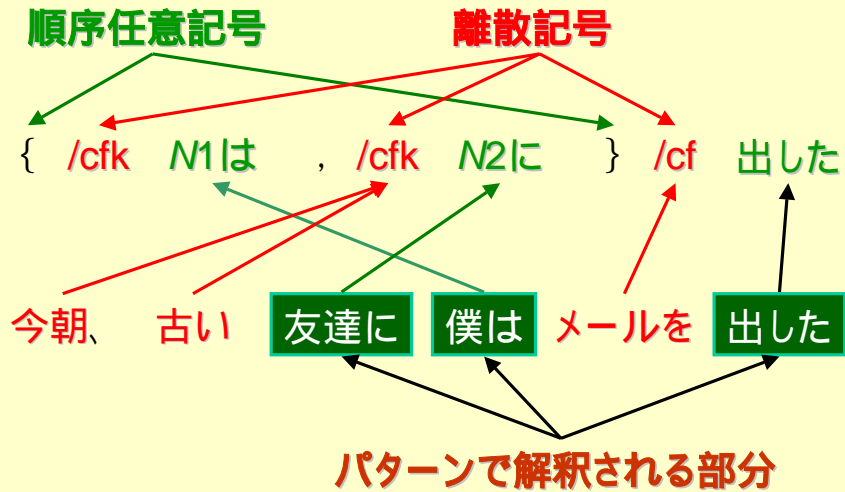
制約 = 述部の用言は「連用形」であること  
 CL1 ^renyou N2 に乗りました。

新橋へ行き ゆりかもめ に乗りました。

連体形にはマッチしない

新橋へ行く ゆりかもめ に乗りました。

## 順序任意記号と離散記号の処理

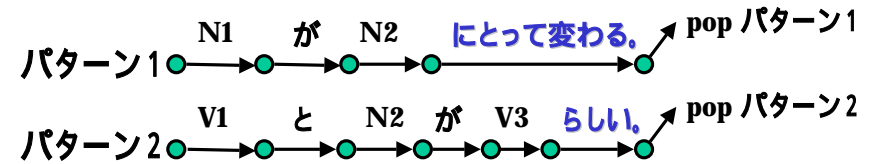


233

## (4) パターン照合基本方式

### ATNによる照合

1パターンから1ネットワークを作成



### アークの構造

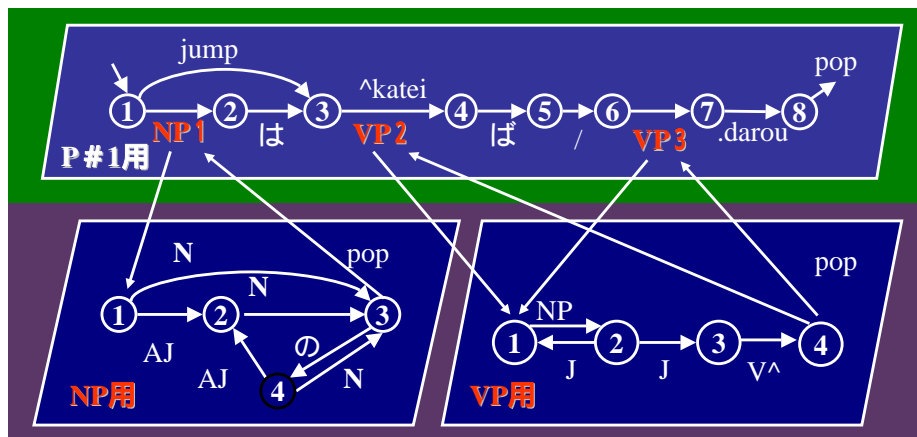
- ・ 任意記号 ... jump アーク
- ・ 選択記号 ... 分岐
- ・ 移動可能記号 ... パターンの展開
- ・ 順序任意記号 ... フラグ付き, 分岐

234

## サブネットワークによる階層型の照合

東京に出張すれば、今度は彼と飲みに行くだらう。

P#1: <NP1は> VP2^katei ば / VP3 .darou



235

## サブネットの構成規則

#	種別	説明	規則数
1	離散記号	/y, /t, /c, /l, /k及びこれらの組み合わせ(例 /kct)の解釈	50
2	様相関数	.kako, .Teiruなど、時制、相、様相の解釈	247
3	節変数の解釈	節変数CLとそれに様相、接続等が接続した表現の解釈	114
4	句変数の解釈	NP(7種)、VP(4種)、AJP(4種)、AJVP(4種)、ADVP(1種)の解釈	20
5	単語変数の解釈	N(11種)、ND(1種)、TIME(4種)、NUM(4種)、V(60種);活用形変形含む、AJ(20)、AJV(53)、ADV(14)、REN(5)に関する解釈	202
6	その他	助詞相当表現、修飾表現、格要素などの解釈	84
-	合計	- - -	717

236

## パターン解釈のためのネットワークの規模

項目	説明	規模
ネットワーク数	パターン解釈のためのネットワーク(サブネットワークを含む)全体数	157種類
アークの数	ノード間の遷移を定義したアークの数	5,061本
シンボルの数	ネットワーク内で使用された記号の種類数	323個
終端記号	ネットワークの終端を表現するもので、辞書や品詞コードに直結する変数や関数の終端を表す	166個

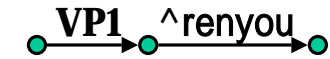
237

## 統語的制約と意味的制約の処理

### サブネットワークによる方法

#### 文型パターンの例

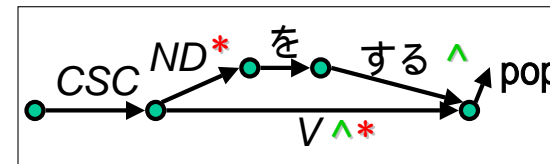
VP1 ^renyou



#### VPの定義

VP CSC ND\* を する ^ \* ... 意味的制約の係り先  
 VP CSC V^\* ^ ... 統語的制約の係り先

#### VPのネットワーク



#### エージェントの記憶域

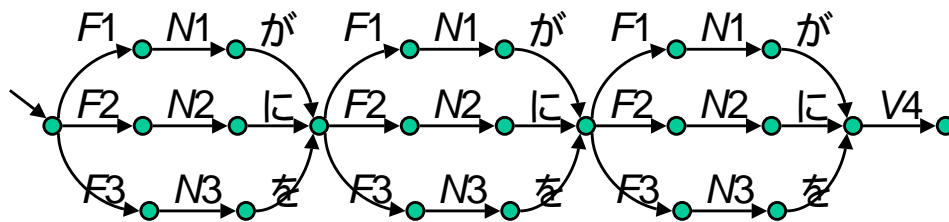
\* に対する形態素  
 ^ に対する形態素

238

## 順序任意記号の処理

### 複数のパスに展開

<例> { N1が, N2に, N3を } V4



#### 通過したパスで判定

F1, F2, F3のすべてが揃えば、適合と判定

239

## エージェントによる照合状態の管理

### エージェントが保持する主な情報

- ネットワーク上の位置
- push/popのためのスタック
- ネットワークの移動経路
- 変数・記憶記号のバインド値
- 制約条件の係り先
- フラグ

240

## エージェントの基本動作

Move	条件を満たすときアークを移動
Kill	条件を満たさないとき消滅
Memorize	変数等のバインド値を記憶
Sleep	同期をとるために休眠
Awake	同期をとるために覚醒
Breed	曖昧なときに子供を産む
Join	複数エージェントが合体

241

## (5) 照合速度向上の方法

### 3つの高速化の技術

- (1) 字面によるパターン絞り込み前処理
- (2) エージェントの合体機能(JOIN機能)
- (3) サブネットワーク照合履歴の再利用

242

## 字面によるパターン絞り込み前処理

文型パターン中の必須の字面で絞り込む

(文) 家に帰ってご飯を食べる。  
(パターン辞書)

P1: N1にV2(て | で)VP3

P2: N1なのでVP2と思う。

P3: N1にV2(て | で)N3をV4.kako.

### 効果が期待される理由

単語レベル文型パターン辞書(12万パターン)中の  
必須字面の種類数: 93,792 種類 殆どが頻度1

**頻度1回の字面: 88,303 種類(94%)**

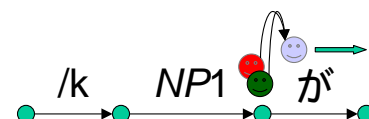
頻度2回の字面: 3,335 種類(4%)

## エージェントの合体(JOIN)機能

ATNでの状態遷移の非決定性(曖昧性)

- ・文型パターンのネットワークは**直線的**
- ・一端分岐したものが合流しやすい

例) 裏の門が



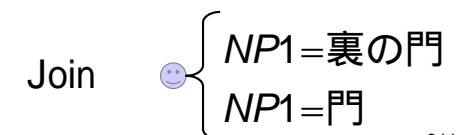
😊 NP1=裏の門

😄 /k=裏の, NP1=門

### Join機能

エージェントの保持する  
情報を統合

変数のバインド値を,  
複数通り記憶



244

## サブネットワーク照合履歴の再利用

- 入力表現と変数との照合により得られた情報を保存
  - 保存する情報
    - 変数の種類
    - その変数に適合する形態素の先頭の位置
    - その変数に適合する入力表現の文字列
- 同じ入力表現を別のパターンの変数と照合するとき、その情報を使用する
  - Pop時に当該エージェントの履歴に追加登録

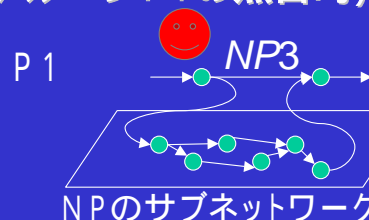
245

## 照合履歴の記録と再利用の方法

(入力文)

… 7 太郎 8 の 9 家 10 の 11 門 12 が …

(パターンP1の照合時)



(パターンP2の照合時)

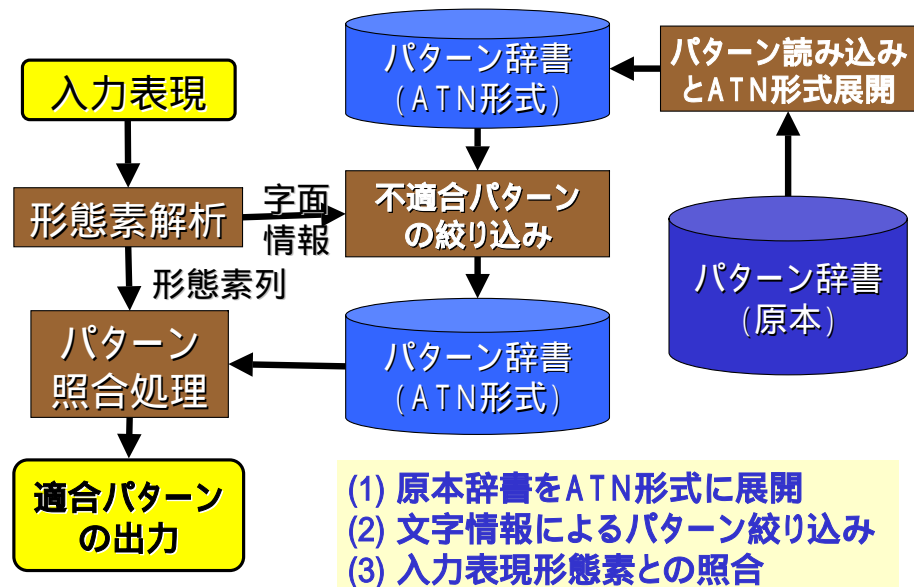


履歴テーブル

キー	値
NP・7	「太郎」 「太郎の家」 「太郎の家の門」
NP・9	「家」 「家の門」

246

## (6) パターンパーサの実装と評価



- 原本辞書をATN形式に展開
- 文字情報によるパターン絞り込み
- 入力表現形態素との照合

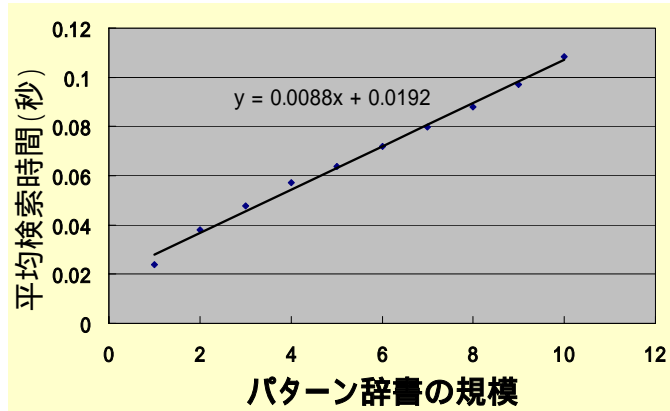
## 動作環境

- CPU: AMD Athlon64 2.4GHz
- メモリ: 2GB
- OS: SUSE-Linux (64bit)

248

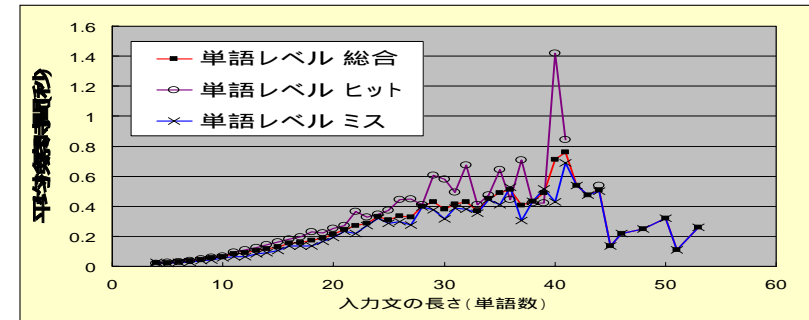
## 辞書の規模と検索時間

- 単語レベルの文型パターン辞書の場合 -

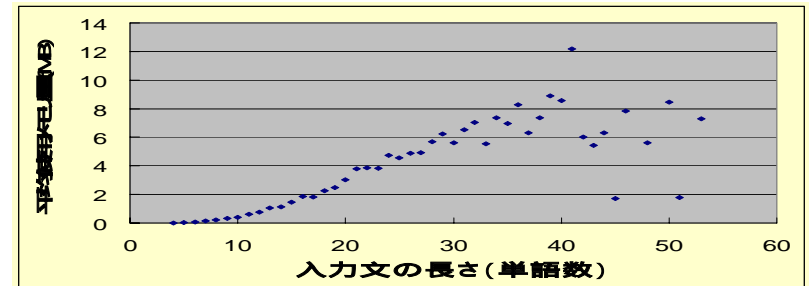


249

## 文の長さや検索時間



## 文の長さやメモリ使用量



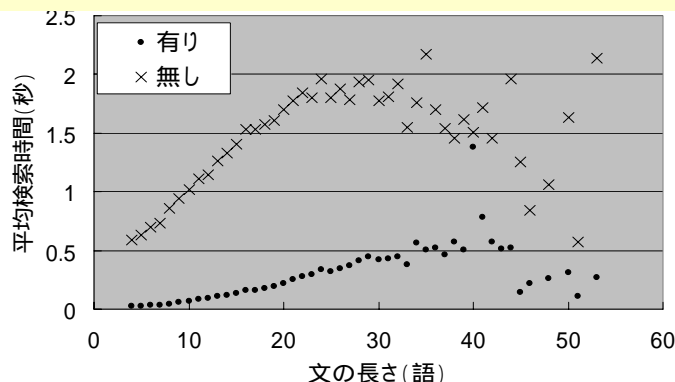
150

## (7) 高速化手法の評価

### 絞り込みの効果

絞り込みのない場合の評価

単語数20以上で急速にメモリ不足が増大する  
照合できた場合のみで評価



検索時間を6分の1以下に短縮

251

## 遷移状態のJOINの効果

「JOIN」しない場合、計算不能が増大  
照合可能な場合のみを評価

(1) 状態数の圧縮 = 平均 5分の1 に圧縮

(2) 平均検索時間 = **9分の1に圧縮**

(0.94秒/文 0.11秒/文)

## 履歴機能の効果

履歴機能を使用しない場合、計算不能が増大  
照合可能な場合のみを評価

平均検索時間 = **38分の1に短縮**

(1.58秒/文 0.04秒/文)

# 高速化効果のまとめ

## < 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

200

## (8) パターン検索の例

(パターンパーサ)

多くのパターンがヒット  
意味的に不一致あり

変数を意味的に制約



## 意味制約の例

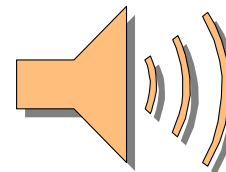
用言の意味属性で制約した結果、32件に絞り込まれた



## 【補足説明19】 パーサー照合能力の限界は？

**自己パターンへの適合率は100%になる筈？**

各パターンが作成元の例文に適合する割合



現状	単語レベル	99.02%
	句レベル	94.85%
	節レベル	92.07%

現状では、なぜ100%にならないか？

- ・ 変数への適合範囲の判定に構文情報が必要

どうしたら解決するか？

- ・ 単文パターン辞書、句表現パターン辞書、複合語パターン辞書が揃えば解決する



## 6.2 パターン意味検索プログラム

### (1) 機能条件と設計条件

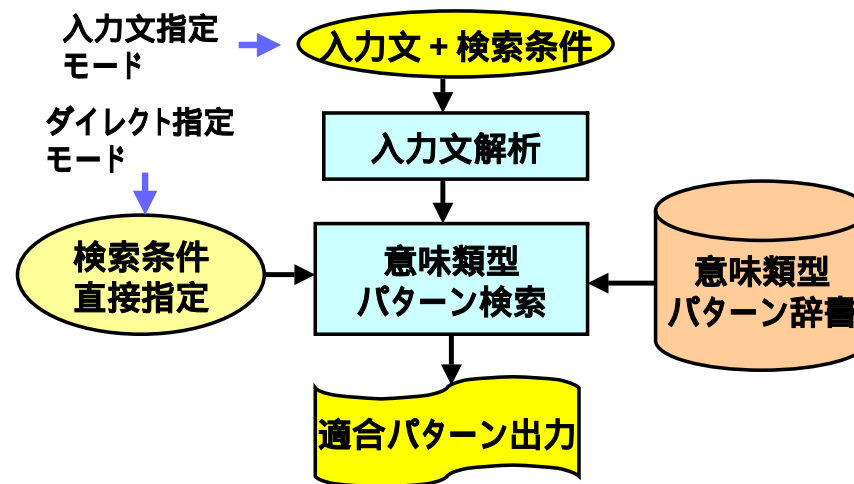
#### 機能条件

- ・ 入力文と統語的、意味的分類コードの一致するパターンの検索
- ・ 指定した統語的、意味的分類コードのパターンの検索

#### 設計条件

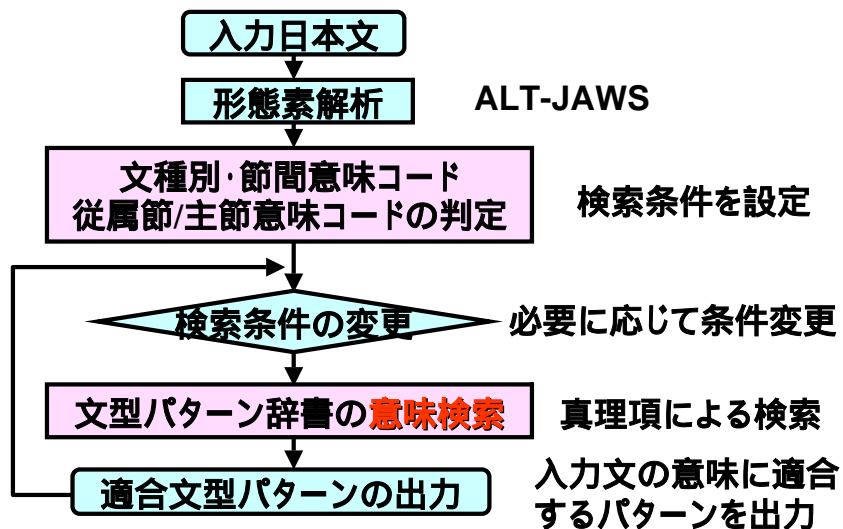
- ・ 入力文の持つ統語的、意味的分類コードは自動的に判定すること
- ・ 入力文の統語的、意味的分類コードに対して任意の制約条件を追加削除できること

### (2) 意味検索プログラムの構成



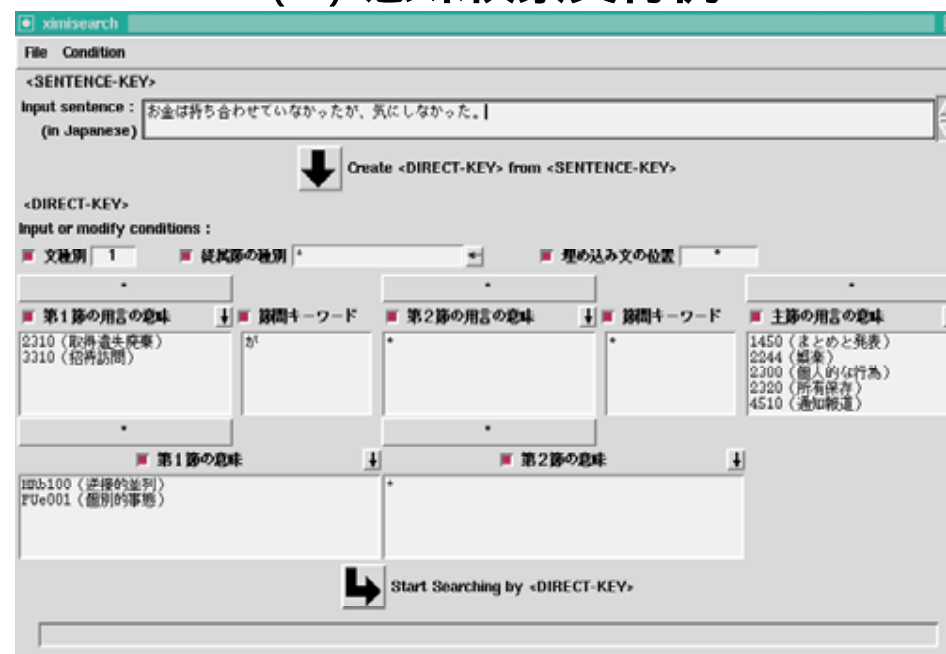
258

### 意味検索プログラムの構成



259

### (3) 意味検索実行例



# 検索結果

Search Result

File View Operation

検索結果: 日本語パターン数 23 / 英語パターン数 23 / 事例文数 12

AC007991-00  
祖父は来年喜寿を迎えるが、なおかくしゃくとしている  
My grandfather will be 77 next year, but he is still strong and healthy.  
W: /y\$1^(/tefkN1は)#2[/tefkTIME3]/tekn4を/ef迎えるが、\$1#5[/uADV6]!かくしゃくと/efしている。  
W: N1 will be NUM(N4) #2[ADV(TIME3)], but N1 be #5[ADV6] strong and healthy.

AC023295-00  
申しわけありませんが、いま即答はできません  
I'm sorry, but I can't give you a ready answer.  
W: /y\$1^(/tefkN1は)ありませんが、</utkN1は>/tefkいま</tekn2に>N3は/efできません。  
W: <I|N1> be sorry, but <I|N1^pron> can not give <you|N2^obj> N3.

AD000210-00  
私は彼に会ったが、彼は私に知らぬ顔をした  
I met him, but he did not acknowledge me.  
W: /y\$1^(/tefkN1は/tefkN2に/efV3.kakoが、/y\$1^(/tefkN4は/tefkN5に/tefk知らぬ顔を/efした。  
W: N1 V3.past N2^obj, but N4 did not acknowledge N5^obj.

AF001162-00  
お金は有るが、今ここには持っていない。  
I have money, but I don't have it with me here now.  
W: /y\$1^(/tefkN1は/efV2#52(.genzai|.kako)が、</utkN3は>#4[/tefkTIME5]#6[/tekn7には]/efV8.teiru.hitei#51(.genza  
i|.kako)。  
W: <I|N3> V2#52(^present|^past) N1, but <I|N3> do not V8#51(^present|^past) N1 with N3^obj #6[ADV(N7)  
] #4[ADV(TIME5)].

AF028819-00

# 6.3 第6章のまとめ

## 1. パターンパーサの開発

入力表現の構造に適合するパターンを検索

<基本アルゴリズム>

・ATNによる照合 + サブネットワークによる変数照合

<高速化の手法>

- (1) 字面によるパターン絞り込み前処理
- (2) 遷移状態の合体機能(JOIN機能)
- (3) サブネットワーク照合履歴の再利用

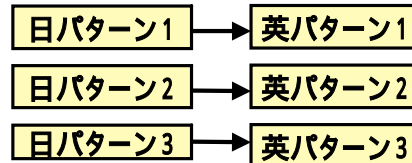
## 2. パターン意味検索プログラムの開発

統語的、意味的分類コードによる検索

## 【補足説明20】 日英機械翻訳の2方式

### 1. パターン翻訳方式

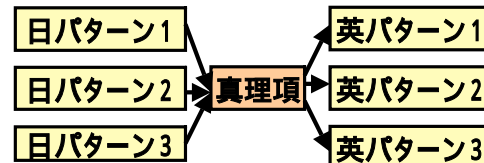
・入力日本語に適合したパターンを使用する



日英パターン間で変数の対応関係が明確

### 2. 意味的等価変換方式

・真理項を介して英語パターンを選択する



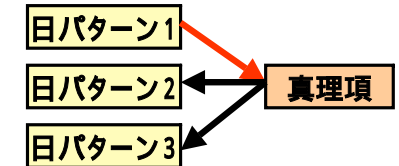
日英パターン間で変数の対応関係が明確でない

いずれも複数の対訳候補が得られるが、後者の方が構造的なバリエーションが多い

## 【補足説明21】 日本文書き換えの2方式

### 1. 真理項を媒介する方法

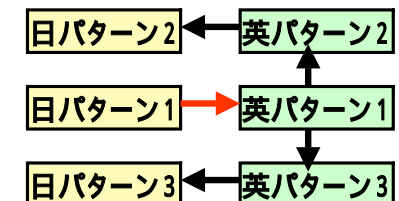
・同一概念を持つパターンのグループ内で変換



意味的近似度が低い

### 2. 英語パターンを媒介する方法

・同一または包含関係の英語パターンを持つ日本語パターンで変換



意味的近似度が高い

# 第7章 適合パターンの絞り込みと翻訳への適用例

## 第7章の目次

- 7.1 パターン検索結果絞り込みの例
- 7.2 パターン翻訳実験の例
- 7.3 訳し分け候補生成機能の評価
- 7.4 第7章のまとめ

## 7.1 パターン検索結果の絞り込み

### (1) 実験の目的と実験方法

#### パターンパーサ

- ・ 統語構造に着目したパターン検索プログラム
- ・ 意味的に不適切なパターンにも適合 = 適合情報も出力 **自動翻訳向き**

#### 意味検索プログラム

- = 適合パターンのみ出力 **翻訳支援向き**

#### < 目的 >

パターンパーサの適合文型が絞り込めるか

## 検索の条件

対象: **単語レベル文型パターン**

### (1) パターンパーサの検索条件

**変数の意味的制約条件を無視**

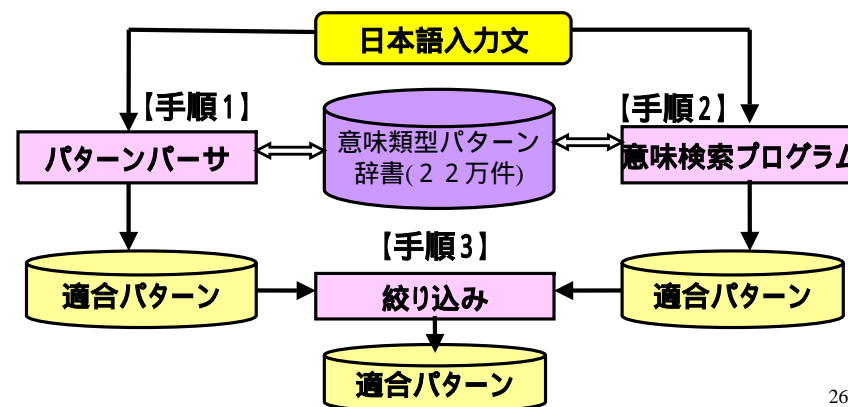
構造的に適合したパターンのみを抽出

### (2) 意味検索プログラムの検索条件

- < 条件1 > 「節間意味分類」
- < 条件2 > 「節間意味分類」+ 「主節意味分類」
- < 条件3 > 「節間意味分類」+ 「主節意味分類」+ 「従属節意味分類」による検索

## 絞り込み実験の手順

- 【手順1】 パターンパーサによる検索
- 【手順2】 意味検索プログラムによる検索
- 【手順3】 両者による絞り込み



## (2) 絞り込み実験の結果

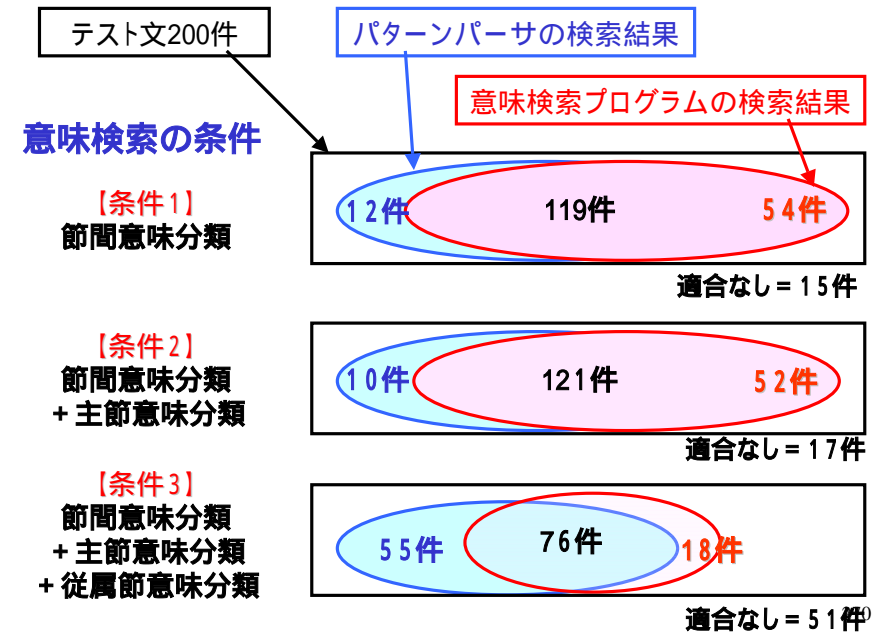
適合パターンから見た入力文の分類

分類	説明	意味検索の条件		
		条件1	条件2	条件3
A類	パーサのみ適合	10	14	57
B類	意味検索のみ適合	54	50	22
C類	両者適合	121	117	74
D類	どちらも不適合	15	19	47
-	合計	200	200	200

入力文全体



## テスト文の分類



## 平均適合パターン数と適合率の比較

評価パラメータ	パーサでの検索	意味による検索		
		条件1	条件2	条件3
適合率R1	64.0%	87.5%	83.5%	48.0%
平均適合文型数N	61件	9741件	546件	36件

適合率: R1 (%)

文型パターンに適合した入力文の割合 (%)

平均適合文型数: N (件)

$$N = \frac{\text{入力文に適合したパターンの総数}}{\text{1つ以上のパターンに適合した入力文}}$$

## 絞り込み効果

分類		文種別1 (重文)		文種別3 (複文)	
		N	P1	N	P1
パーサ照合		58.8件	24.2%	2.6件	67.6%
意味検索	条件1	9233件	12.0%	7019件	2.7%
	条件2	69件	24.2%	504件	2.7%
	条件3	6件	53.2%	37件	6.4%
絞り込み結果	条件1	57.4件	14.7%	1.9件	48.0%
	条件2	1.1件	78.6%	1.4件	55.6%
	条件3	1.0件	87.5%	0.25件	0%

### (3) 検索結果の例

#### 【手順1】 パターンパーサの検索結果(上位5件)

<入力文> 「私は友人を頼って上京した。」 適合パターン数 = 2532件

- × [1] /y\$1/tcfkN1を/cfV2(て)で\$1^(ytckN3は)/cf(V4.kako|ND4をした),  
N3 V(V4|ND4).past when N3 V2.past N1.  
<例> 話を聞いて彼は逆上した。 He went wild when he heard that.
- × [2] /y\$1/tcfkN1を/cfV2(て)で\$1^(ytckN3は)/cf(V4.kako|ND4をした),  
N3 be.past V(V4|ND4).past to V2 N1.  
<例> それを聞いて私は安心した。 I was relieved to hear it.
- × [3] /y\$1/tcfkN1を/cfV2(て)で\$1^(ytckN3は)/cf(V4.kako|ND4をした),  
V2^grn N1 V(V4|ND4).past.  
<例> 死体を見て彼は茫然自失した。 Seeing the dead body freaked him out.
- × [4] /y\$1/tcfkN1を/cfV2(て)で\$1^(ytckN3は)/cfV4.kako,  
With N(V2) of N3^poss N1 N3 V4.past.  
<例> 手を振って彼女は立ち去った。 With a wave of her hand she went away.
- × [5] /y\$1/tcfkNP1を/cfV2(て)で\$1^(ytckN3は)!VP4.kako,  
N3 V2.past NP1 to VP4.  
<例> 盲目という悪条件を克服して彼は偉大な学者になった。  
He overcame the handicap of blindness to become a great scholar.

#### 【手順2】 意味検索プログラムの検索結果(5件)

<入力文> 「私は友人を頼って上京した。」 適合パターン数 = 23件

- [1] /y</tkN1は>/cfいささかの/k知辺を/cf頼って/ytckN2へ/cf(上っ)のぼっ|上ぼっ)た。  
<l|N1> went to N2, looking to a slight acquaintance for assistance.  
<例> いささかの知辺を頼って都へ上った。  
I went to town, looking to a slight acquaintance for assistance.
- [2] /y\$1^(tcfkN1は)/tcfkN2を/cf頼って\$1/ycf上京した。  
N1 came to Tokyo from the country counting on N1^poss N2's help.  
<例> わたしはおばを頼って上京した。  
I came to Tokyo from the country counting on my aunt's help.
- × [3] /y</tkN1は>/tcfkN2/tck一人の/k知人を/c頼って/ycf上京した。  
<l|N1> came to town, looking to an AJ(N2) friend for assistance.  
<例> たった一人の知人を頼って上京した。  
I came to town, looking to an only friend for assistance.
- [4] /y</tkN1は>/tcfkN2を/cf頼って</ycfN3は>/cf(V4.kako|ND4をした),  
<l|N3> N(V4|ND4) where <l|N1> could rely on <my|N1^pron^poss> N2.  
<例> 親類を頼って上京した。 I went to Tokyo where I could rely on my relatives.
- × [5] /y</tkN1は>/tcfk雲の/k中を/tcfkN2に/cf頼って/ycfV3.kako,  
<We|N1> V3.past through the clouds with the help of N2.  
<例> 雲の中を計器に頼って飛んだ。  
We flew through the clouds with the help of the instruments.

#### 【手順3】 絞り込み後のパターン(上位5件)

<入力文> 「私は友人を頼って上京した。」 適合パターン数 = 16件

- [1] /y\$1^(tcfkN1は)/tcfkN2を/cf頼って\$1/ycf上京した。  
N1 came to Tokyo from the country counting on N1^poss N2's help.  
<例> わたしはおばを頼って上京した。  
I came to Tokyo from the country counting on my aunt's help.
- [2] /y</tkN1は>/tcfkNP2を/cf頼って/ycf上京した。  
<l|N1> came to town, looking to NP2 for assistance.  
<例> たった一人の知人を頼って上京した。  
I came to town, looking to an only friend for assistance.
- [3] /y</tkN1は>/tcfkN2を/cf頼って</ycfN3は>/cf(V4.kako|ND4をした),  
<l|N3> N(V4|ND4) where <l|N1> could rely on <my|N1^pron^poss> N2.  
<例> 親類を頼って上京した。 I went to Tokyo where I could rely on my relatives.
- [4] /y</tkN1は>/tcfkNP2を/cf頼って/ycf上京した。  
<l|N1> came to town, looking to NP2 for assistance.  
<例> たった一人の知人を頼って上京した。  
I came to town, looking to an only friend for assistance.
- × [5] /y</tkN1は>/tcfkN2を/cfV3(て)で/ycf(V4.kako|ND4をした),  
<l|N1> N(V4|ND4) after having V3.past <my|N1^pron^poss> N2' consent by persuasion.  
<例> 両親を説得して上京した。  
I came to Tokyo after having obtained my parents' consent by persuasion.

### <参考> その他の絞り込みの方法

#### (1) 変数の意味的制約条件による方法

文種別	用言意味属性を使用せず		用言意味属性を使用			
			汎化なし		第1段に汎化	
	N	P1	N	P1	N	P1
文種別1(重文)	46.4件	25.2%	5.9件	19.4%	10.0件	35.9%
文種別3(複文)	8.4件	46.2%	0.5件	41.7%	2.3件	56.6%

#### (2) 多変量解析による方法

使用するパラメータ  
パターン適合率、 字面適合率、 変数適合率、  
用言意味属性距離、 名詞意味属性距離、 など

## 7.2 パターン翻訳実験の例

### (1) 実験の狙いと方法

#### 実験の狙い

- ・ 複数翻訳結果の生成
- ・ 意識型翻訳機能の検証

#### 実験の方法

- ・ 絞り込みパターンを使用した翻訳
- ・ 従来システムとの比較

277

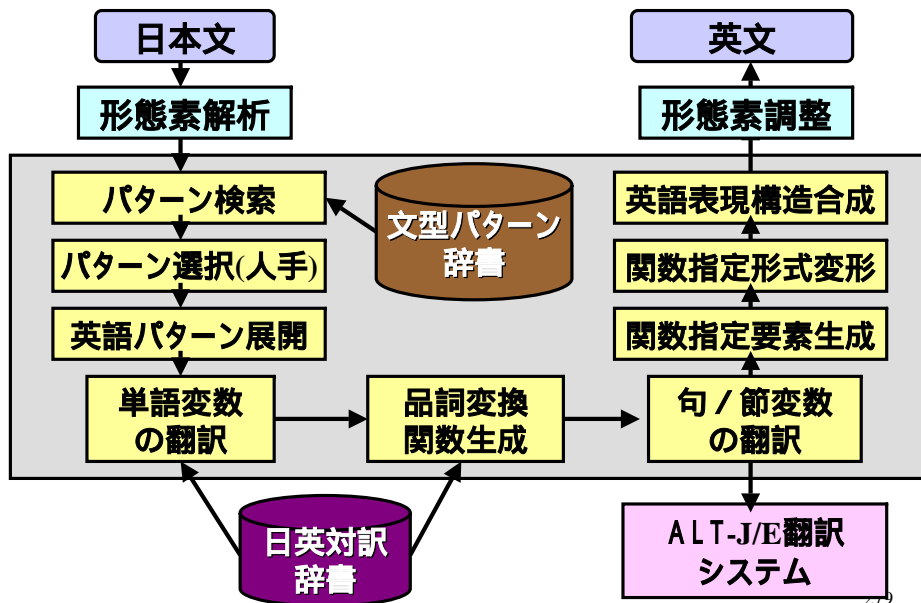
### (2) パターン翻訳実験システムの試作

- 1) 使用するパターンの指定(手動)
- 2) 変数適合部分の英訳
  - ・ 単語 対訳辞書を検索
  - ・ 句/節 既存の翻訳システムを使用
- 3) 文法属性の変換処理
- 4) 関数、記号で指定された語形変形など
- 5) 英語形態素調整(未着手)

注目点: **どれくらい意識ができるか**

278

### パターン翻訳実験システムの構成



279

### 日英パターン翻訳の例

英語文型パターンは人手で選択、 各変数の局所訳を代入、 N-gramで訳文選択

Search Result

File Data

検索結果: 日本語パターン数 [延べ32/異なり32] 表示範囲: 100%

順位: パターンID (適合率)

解析された部分  
適合パターン  
原文

1: AC032920-00 (100.0%)  
5年かけて言語知識ベースを開発した。  
VP1 NP3 V4

日文型: /VP1(て/で)/N2は/NP3を/of V4, kaka.  
英文型: <I|N2> @have^present VP1^ed V4^ing NP3.  
日原文: 1か月かけてこの計画を練り上げた。  
英原文: I have spent a whole month working out this plan.

1: AC032920-00 Translation Result  
英語パターン  
VP1 NP3 V4  
<I|N2> @have^present VP1^ed V4^ing NP3.  
翻訳結果  
I have spent five years developing a language knowledge basis .

1: AF032920-00  
VP1 NP3 V4

ITM Reboot Translate

### (3) 実験結果と翻訳例

入力文	春になると木の枝がのびる。
出力文	The branch of a tree grows in spring.
【参考】 使用したパターンとその例文 <日> /ytcfkTIME1に/(成る なる)と/fNP2が/cfV3. 春になるといろいろな花が咲く。 <英> NP2 V3 in N1. Many kinds of flowers bloom in spring.	
入力文	金沢文庫は鎌倉時代以来の歴史を持つ町です。
出力文	Kanazawabunko be town with history since the kamakura era.
【参考】 使用したパターンとその例文 <日> /ycfN1は/ycNP2を/cf(持つ もつ)/N3.#da. 「取る」は多くの意味を持つ言葉だ。 <英> N1 be N3 with NP2. "Toru" is a word with many meanings.	

281

## 7.3 意味的等価変換方式の効果

### (1) 評価の目的と方法

#### 評価の目的

統語的に異なる訳文候補の生成  
 多彩な訳文候補の生成

#### 評価対象文

- (1) 文種別1(重文)と文種別3(複文)、各20文
  - (2) 適合文型パターンがいずれもゼロでないこと
  - (3) 適合文型パターン数がいずれも極度に大きくないこと  
 平均適合文型数Nの10倍以下
- 理由: 文型パターン辞書は未縮退化

### (2) 評価結果

#### 適合パターン数の比較

評価パラメータ		パーサ の検索	意味による検索		
			条件1	条件2	条件3
文種別1 (重文)	平均適合パターン数N	46.4 件	691 件	243 件	12.2 件
	正解適合パターン数Nt	11.7 件	45.6 件	31.3 件	5.6 件
	適合パターン正解率P1	25.2 %	12.0 %	24.2 %	53.2 %
	英文統語構造異なり数	9.7 件	45.2 件	31.3 件	5.5 件
文種別3 (複文)	平均適合パターン数N	8 件	715 件	753 件	48.5 件
	正解適合パターン数Nt	4.3 件	45.7 件	4.6 件	1.6 件
	適合パターン正解率P1	53.0 %	2.6 %	2.7 %	6.4 %
	英文統語構造異なり数	2.3 件	4.4 件	4.6 件	1.6 件

283

### パターン翻訳方式との比較

区別	パターン検索 の方法	文種別	適合正解 パターン数	統語構造 の異なり
パターン 翻訳方式	パターン パーサ	重文	11.7	9.7
		複文	4.3	2.3
意味的等価 変換方式	意味検索 プログラム	重文	45.6	45.2
		複文	45.7	4.4

#### 評価の条件

意味検索で「節間意味分類」を使用した場合

284

## 翻訳結果の例(1 / 2)

【例1】 **そういう善行の記事を読むと心が温まる。**  
It get warm heart to read articles about good deeds like that .

<理想訳> It warms my heart to read articles about good deeds like that.

<従来方式>

- ALT訳 When the article of such good conduct is read, a heart is warmed.  
A社訳 The mind gets warm if the article on such beneficence is read.  
B社訳 If the report of such good behavior is read, the heart will get warm.

【例2】 **彼はアメリカへ留学しようと決心した。**  
He resolved to go to america to study .

<理想訳> He resolved to go to America to study.

<従来方式>

- ALT訳 He decided to go to the U.S. to study.  
A社訳 He determined that went to study to the United States by him.  
B社訳 He decided to study in the United States.

## 翻訳結果の例(1 / 2)

【例3】 **彼はユーモアの分らない人だ。**  
That have no sense of humor .

<理想訳> He has no sense of humor.

<従来方式>

- ALT訳 He is people who do not understand humor.  
A社訳 He is a person who doesn't understand the humour.  
B社訳 He is the person who does not understand humor.

【例4】 **うそをつくのは悪い習慣だ**  
Jogging invention is bad habit .

<理想訳> Lying is a vice.

<従来方式>

- ALT訳 It is a bad habit to stab a bullfinch.  
A社訳 It is a bad practice to tell a lie.  
B社訳 A bad custom tells a lie.

## 7.4 第7章のまとめ

パターンパーサと意味検索プログラムの利用

### 1. 適合パターンの意味的な**絞り込み**

**意味的正解率向上効果**

- ・変数意味属性による方法・・・節間意味コードが効果的
- ・意味分類コードによる方法・・・汎化した方法が効果的  
改良の余地が大きい

### 2. パターン翻訳の効果

従来方式と比べて**意識の能力**が期待できる

### 3. パターン翻訳と意味的等価変換方式の比較

**訳し分け**のためのパターン候補が得られやすい

## 第8章 研究成果と今後の展望

### (1) 研究の到達点

重文複文の文型パターン辞書の開発とその意味類型化

意味類型パターン辞書 = 「**表現意味辞書**」

非線形な表現の意味を掘り取るための  
基本的な仕組みを実現

<学術的貢献>

- ・ **言語モデル**の提案  
非線形な表現の切り出しと汎化方法  
の基本原則の明確化
- ・ **パターン記述言語**の提案      汎用性が高い



## (2) 生産物

#	分類	項目	説明
1	報告書類	研究終了報告書	研究開発結果全体のまとめ
2		技術検討資料集	研究期間中の主な検討資料
3	パターン記述言語仕様書	パターン記述言語仕様書 <下記の資料を含む> (1)「節間意味分類体系表」(222分類) (2)「日本語単語意味属性体系」と「単語意味辞書」 ・用言(372分類、約6000語)、・名詞(370分類、約6万語)	
4	意味類型パターン辞書	意味類型パターン辞書	重文複文パターン(約22.7万件) 統語・意味分類コード付き
5		意味類型基本パターン辞書	上記辞書の圧縮版 (重要パターン約5万件)
6	検索プログラム	パターンパーサ・プログラム	要素照合型
7		パターン意味検索プログラム	真理項による検索
8	その他	重文複文対訳データのタグ付きコーパス(15万文)など	

## (3) 残された課題

= 重文・複文の意味類型辞書に関する課題 =

- (1) 変数の意味的制約条件の適正化  
動詞、名詞などの意味属性体系の見直し  
コスト大 パターン辞書縮退後に実施
- (2) 意味分類コード体系の適正化  
方式実験により問題点を把握
- (3) 適合パターン絞り込み方式  
多変量解析の方法が良さそう  
変数の意味制約条件見直し後に再検討
- (4) パターン照合方式との整合性  
被覆率向上の可能性

= 表現意味辞書全体の整備の課題 =

非線形言語モデルからの要請

各レベルの非線形な表現構造をパターン化すること

日本語表現の構成要素

- ・重文複文の表現構造……本研究で実現
- ・単文の表現構造……「日本語語彙大系」で一部実現  
非線形な主体的表現(助詞、助動詞類)の意味が対象外
- ・句と複合語の表現構造……未開発

< 今後の急務の課題 >

単文文型パターン辞書の再開発

## (4) 応用研究への期待

直接的応用

意識型の機械翻訳  
日本語言い換え技術

新研究分野への応用

キーワード型の情報検索  
質問応答型の意味検索  
波形分析型の音声認識  
意味追跡型の音声認識

## 意味類型パターン辞書のご利用について

研究利用は無料(原則)です。

4月からホームページ「**鳥バンク**」を開設し、

- (1) パーン辞書内容の閲覧
- (2) パターン辞書の簡易な検索試験

ができるようにします。また、

- (3) パターン辞書ファイルなどの入手方法
- もご案内します。

「鳥バンク」へは、下記のページからアクセスできます。  
覗いてみてください。

<http://unicorn.ike.tottori-u.ac.jp>