

中国語の意味文法の構成とその処理系の作成†

楊 頤 明† 堂 下 修 司†

本論文では、中国語の意味解析のための文法の構成と、それを用いて統語・意味的曖昧性の解決や、意味表現のための格構造の抽出を行うシステムについて述べる。ここで定義する意味文法は特定のインプリメンテーションに独立した論理的な情報構造であり、意味拘束条件に基づく句構造・格構造対応規則と優先規則からなる。我々は中国語解析について特に重要である「連動構造」(連続する動詞句), 「是字句」(「である」), 「~的」(「~の」), 前置詞句について詳細化した規則を構成した。意味文法を解釈する系は、統語解析部で抽出された不完全な句構造(未解明の部分も含む)を入力とし、文法で規定された意味拘束条件に従って格構造へ変換し、その優先度を計算する。対象指向のプログラミング手法を導入して、階層的意味分類の上で整理された一般規則、および単語の個別規則をシステムの中に一様な形で組み込むようにした。本解析システムでは、入力文全体の解析を、各部分の独立な統語・意味解析によりボトム・アップ的に実現する。この処理方式では、意味解析で得られた部分解析結果の優先度により探索方向を導き、もっともらしい結果を先に得るようなヒューリスティックな解析を実現する。残りの可能性も保存しており、先の探索が失敗した場合、次の選択として提供することができる。本方式は現在一部汎用計算機上で実現されている。最初のテストとして、入力文20例の解析を行ったところ、第一位選択で70%, 第二位選択まですべての正しい格構造が抽出できた。この実験結果により、本論文で行った意味処理の有効性が確かめられた。

1. まえがき

中国語の非形態性(英語の語尾や関係詞、日本語の格助詞などのようなものが少ない)が原因で、形態素や統語情報の解析に基づいた計算機構文解析では、爆発的に生じる曖昧性の解消は重大で本質的な問題になる。その解決手法の一つとして、中国語の「特徴語」という数百個の常用語を利用して、入力文全体の構文解析の前に、前処理として統語構造を部分的に予測することにより探索空間を抑える方法を、すでに以前の論文で報告した¹⁾が、本論文では、統語解析の上でのいっそう深い処理-意味解析の導入について述べる。

構文解析に意味的知識の利用について、近年来日本語や英語などを対象にする研究が目立っている。その代表的なものは、数十人の研究者を集め、数年間にわたる機械翻訳プロジェクト(日本科学技術庁日英訳のMu-Project²⁾、ヨーロッパの多言語翻訳プロジェクトEUROTRA³⁾)の一部として、意味解析を大規模システムの上で実現することである。それに対して、中国語から他言語への翻訳には、上記のような大規模プロジェクトが結成されていないし、中国語解析はまだ統語解析の段階にすぎず、意味解析はほとんど導入されていない一方、従来の言語学では、計算機解析に利用

できる中国語意味文法はまだ提案されていない現状である。

中国語意味解析の問題の一例を図1に示す。図1の七つの例文は統語構造あるいは意味構造の上でそれぞれ異なるものであるが(2.1.1節を参照)、統語情報のみによる構文解析では、いずれも NP_V_NP_V_NP の系列に対応するものであるので、区別できないものである。このような曖昧性を意味解析の段階で解明しなければならない。

本論文では、意味拘束条件によって統語解析結果中に残った統語・意味的曖昧性を解析することを目的とし、それに必要な知識を意味処理系に使える文法としてまとめ、システムの上でこの意味文法の利用法と有効性を検討する。

意味解析に利用した知識は

- 1) 中国語の構文において、句構造の統語機能(ここで表層格と呼ぶ)に関する知識、すなわち句構造の並び順序および機能語を表層格の決定に利用する。
- 2) 単語の意味分類と構文の意味制限:動詞の格解析を中心にして、動詞と名詞などの意味分類、およびそれに基づいた格構造抽出に意味的拘束条件を利用する。
- 3) 意味構造の最尤さに関する知識:意味対応関係上の妥当性として、人間の常識を数値化したものを最尤結果の選択に利用する。特に典型情況に関する「肯定的」(正数)と「否定的」(負数)二種類の知識を強調的に利用する。

† The Construction of a Semantic Grammar for Chinese Language Processing and Its Implementation by YIMING YANG and SHUJI DOSHITA (Department of Information Science, Faculty of Engineering, Kyoto University).

†† 京都大学工学部情報工学教室

	N P	V	N P	V	N P	
例文 1.	小孩子 子供	笑 からかう	他 彼	是 です	一個大胖子. 一個, 天肥り	子供は彼を大肥りであるとからかう.
例文 2.	我 私	勧 める	他 彼	学 ぶ	医. 医学	私は彼に医学を学べと勧める.
例文 3.	他 彼	有 有る	一個妹妹 一人妹	喜歡 好む	美術. 美術	彼は美術の好きな妹がいる.
例文 4.	我 私	知道 知る	他 彼	抽 吸う	煙. たばこ	私は彼がたばこを吸うことを知っている.
例文 5.	他 彼	教 教える	我 私	打 打つ	球. ボール	彼は私にボールを打つことを教える.
例文 6.	你 君	唱 歌う	歌 歌	影響 影響する	別人. 他人	君が歌っていることが他人の邪魔になって いる.
例文 7.	他 彼	抽 吸う	煙 たばこ	喝 飲む	酒. 酒	彼はたばこを吸い酒を飲む.

N P : 名詞句 V : 動詞

図 1 連動構造 NP_V_NP_V_NP の七つの例文

Fig. 1 Seven sentences of serial verb construction NP_V_NP_V_NP.

意味文法では、複数個のアスペクトから階層関係の上で分類的に整理される規則を含みながら、各単語に依存する個別規則も大量に利用するため、対象指向の手法を導入して、性質の異なる規則を各オブジェクトの上で独立にモジュール化して、対象指向処理系で統一に扱う。また、ここで利用される知識は、つねに正しいという性質のものではなく、むしろ多くの場合正しいというヒューリスティックな性質をもち、完全なものではなく、むしろ典型的な現象あるいは局所的な意味対応関係しか扱わないものであるので、優先度処理を用いて意味解析規則の相対比較的な利用法を与える。

意味解析部の入力は、統語解析で入力中国語文の一部に対して文脈自由句構造文法に基づいて解析して得た句構造である（ただし、句構造のうち、完全に解明していない部分もありうる、2.1.1節を参照）。意味解析部では、意味拘束条件で句構造における統語・意味的曖昧性を解析しながら、適切な格構造を抽出し、それに優先度を与える。

上記のような統語・意味解析は、入力文の各部分に独立に行い、部分結果から全体の解析結果をボトム・アップ的に組み立てる。解析の制御方式では、ベスト・ファスト戦略を採用して、部分結果の優先度で全体の解析の探索方向をリードすることにより曖昧性を抑え、無駄の計算を避ける。残りの可能性も保存しており、先の探索が失敗した場合次の選択として計算して提供する。

以下では本解析方式の各部分について詳しく述べる。

2. 意味文法

2.1 格 解 析

意味解析の記述モデルに格構造を選んだ。しかも、格ラベル（深層）として、Mu-Project とほぼ同じ集合（現在格ラベル 40 種類）を採用する。その理由は

- 1) 機械翻訳での、格文法の考え方を探る傾向を重視する、
- 2) 格構造は中国語伝統言語学と一致する点が多い、
- 3) 異なる言語体系の間の共通性と可換性を検討する、
- 4) 本システムを、既存の翻訳システムと結び付け、多言語間翻訳の可能性を開くためである。

格解析は句構造から中間表現へ、中間表現から格構造への 2ステップで行われる。

中間表現は、表層格ラベルで標記する依存構造である。ここで、この中間表現を「表層的中間表現」と呼ぶ。

格構造は、深層格ラベルで表記する依存構造である。

表層格ラベルとして、中国語文の基本成分に基づいて規定した機能ラベル（6 個）、特殊単語（前置詞、方位詞、助詞「的」など、約 240 個あるうち現在そのうちの代表的なもの 30 個を扱っている）、および統語カテゴリの一部を用いる。

深層格ラベルとしては、動詞の格解析を中心に分類したものである。

表 1 表層格と深層格対応一覧表
Table 1 Correspondence between surface-case and deep-case.

表 層 格	深 層 格 (現在 40 個)
機能ラベル (6個)	主 語 主体 (Subject)
	第一目的語 対象 (Object), 受け手 (Recipient), 主体, 内容規定, 与え手 (Origin)
	第二目的語 対象, 内容規定
	後 方 述 語 目的 (Goal), 原因 (Cause), 内容規定, 修飾節 (Descriptive), 並列述語 (主述語)
	名詞状況語 時 (Time), 時間 (During), 場所 (Space)
	補 語 度合 (Degree)
特殊単語 (現在 30 個)	“把” (を) 対象
	“在” (に) 場所, 時, 条件 (Condition), 範囲規定 (Range), 観点 (Viewpoint)
	“的” (の) 関連語 (Associative), 修飾節, {前置詞の深層格}, {方位詞の深層格}, {形容詞の深層格}, {数量詞の深層格}
	“跟” (に, と) 相手 (Partner), 比較 (Comparison)
	“由” (から) 主体, 場所, 始点 (Space-from), 構成要素 (Component), 始状態 (Source), ...
	“了” (た) テンス (Tense)

統語カテゴリ	〈数量詞〉 数量分類子 (Quantity)
	〈動量詞〉 回数 (Times)
	〈形容詞〉 属性 (Attribute), 関係 (Relation), ...

〈 〉: 統語カテゴリ

() : 対応日本語あるいは対応英語ラベル

{ } : 集 合

… : 省 略

表 1 に表層格ラベル, 深層格ラベル, およびそれらの対応を (部分的に) 列挙した。図 2 に, 前に挙げた例文 3 の句構造解析結果, 表層的中間表現, 格構造を示す。この例で示したように, 句構造には「?」の部分, 中間表現では「後方述語」("NEXT-VP") のような未解明の部分があり, ゆえに, ここでの句構造も, 表層的中間表現も不完全なものであるが, 残った曖昧性を格構造の上で解明する。

句構造の表層格は, 句構造の並び順序と機能語 (例えば, 前置詞や助詞など) とに依存するので, 中間表現の生成規則を, 図 3 に示した中国語文のプロトタイプに基づいて, 句構造生成規則ごとに規定する。句構

造から中間表現への変換では, 意味情報を利用していないので, 統語的な変換である。

意味制限を中間表現から格構造への変換に導入する。意味拘束条件を単語の意味分類 (意味マーク) に基づいて書く。性質の異なる曖昧性に対して別々の意味マーク集合を用いるが, 以下では, 中国語意味解析の三つの典型問題を例として, 意味拘束条件の利用を示す。

2.1.1 連動構造 (serial verb constructions)

中国語文の連動構造は, 言語学者によって次のように定義した: 二つあるいはそれ以上の動詞句や埋め込み文を含むが, それらの関係を示す目印がない文を連動構造と呼ぶ⁴⁾。連動構造の典型パターンは

(NP)_V_(NP)_V_(NP)

である (図 1 の例文を参照)。ただし, () は, 「ない場合もある」ことを示す。図 1 と図 2 ではその例を挙げた。

本解析方式では, 統語解析段階で連動構造 V_NP_V_NP を抽出して, 中間表現では, その系列を述語—目的語—後方述語に対応させる (図 2 を参照)。

連動構造を, 格構造の上で 6 タイプに大別する。それらを図 4 に示す。格構造決定に, 動詞の意味属性に基づいた分類, すなわち「使役型」「感情型」「所有型」「知覚型」「特殊型 1」「特殊型 2」6 種類を利用する。格解析規則は以下に示す:

- 1) 前方の動詞が「使役型」ならば, タイプ 1 の (1), すなわち S2 は S1 の「目的」である。
動詞の例: 叫=させる, 讓=させる, 委託=頼む,
請=要請する, 引誘=誘惑する, 要求,
禁止, etc.
- 2) 前方の動詞が「感情型」ならば, タイプ 1 の (2), すなわち S2 は S1 の「原因」である。
例: 喜歡=かかる, 表揚=ほめる, 怪=せめる,
討厭=嫌がる, 恨=恨む, 笑=からかう, 感謝,
etc.
- 3) 前方動詞が「所有型」ならば, タイプ 2, すなわち, S2 の S1 の最後の名詞の修飾節である。
例: 有=有る, 偷=盗む, 買=買う, 找=捜す, 作=作る, etc.
- 4) 前方動詞が「知覚型」ならば, タイプ 3.
例: 看見=見える, 知道=知る, 相信=信じる, 断定=判断する, 主張=主張する, etc.
- 5) 前方動詞が特別型 1, すなわち「行動」あるいは「抽象事物」を第二目的語とする型ならば,

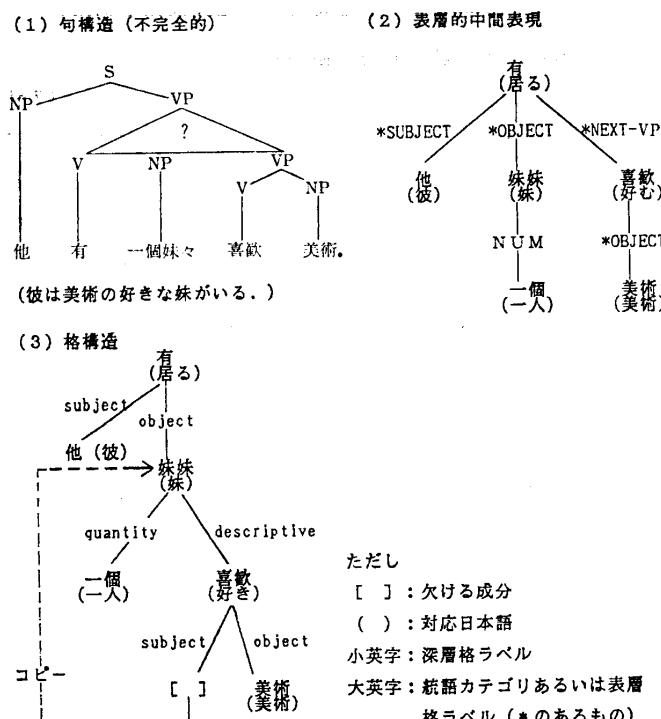


図 2 句構造-中間表現-格構造の対応の一例
Fig. 2 Examples of corresponding expressions of phrase-structure, intermediate expression, and case-structure.

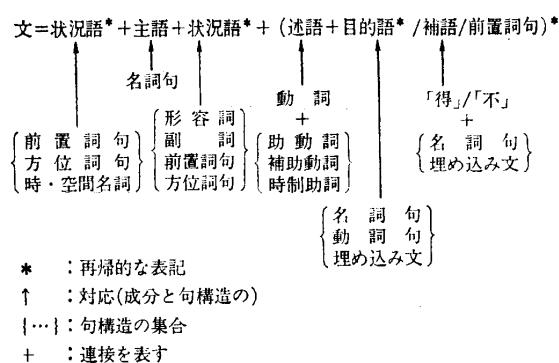


図 3 中国語文の基本成分と構造
この式 “文=…” では、文構造の最上層部は、正規表現で表されている。
Fig. 3 Prototype structure and elements of Chinese sentence.

タイプ 4.

- 例: 教=教える, 問=問う, 指導=指導する, etc.
6) 後方動詞が特別型 2, すなわち「行動」あるいは「抽象事物」を主語とする型とするならば,
タイプ 5.
例: 是=である, 解決=解決する, 使=make, etc.
7) その他の場合(一般型として), タイプ 6, すなわち S1 と S2 は共通な主語をもつ並列文

である。

2.1.2 前置詞句

日本語の格助詞と対照的に、中国語の前置詞は機能語のように働き、時間格や場所格から動作主や対象格まで多様な成分表現に用いられる。

典型的な中国語前置詞句は

前置詞+名詞句+方位詞

の統語構造をもち、これが動詞を修飾する場合、深層格を、前置詞、名詞意味マーク、方位詞意味マーク、動詞 4 要素間の意味対応関係により決める。このために利用した意味分類では、名詞意味マークとして Mu-Project と同一の集合、51 種類の意味マークを採用するが、方位詞の意味マークとしては、その部分集合の〈時刻〉、〈時期〉、〈場所〉、〈抽象〉、〈条件〉 5 種類のマークを利用する。

中国語の前置詞は 60 程度ある(呂叔湘が 61 個⁵⁾、Li, C. が 58 個⁴⁾を集めた)が、我々は数百例を調べた上、その中から常用されかつ多くの曖昧性を引き起こす 28 個(現在)について、深層格判断規則を規定した。より詳細な規則は各動詞の個別規則として、また、一般化した規則は各前置詞ごとにまとめられる。

2.1.3 「～的」

「的」は中国語の特有な助詞であり、「～+的+名詞」の形で修飾部分(「～」で示す)を名詞に繋げる。「的」を「～」の表層格とするが、その深層格を「的」前後の統語・意味関係で決める。すなわち、

- 1) 名詞+「的」: 「～的」の深層格を「関連語」(associative word)とする。
- 2) 数量詞+「的」: 「数量子」(quantity)とする。
- 3) 形容詞+「的」: 形容詞の意味属性、例えば「属性」、「関係」などを深層格とする。
- 4) 前置詞句+「的」: 前置詞句の深層格とする。ただし、「的」の後は動名詞である場合、その動名詞の意味属性を前置詞句深層格の決定条件に入れる(2.1.2 項を参照)。
- 5) 動詞句/埋め込み文+「的」: 「修飾節」とする。ただし、「的」の前の部分(修飾節)に後の名詞による「補間」を行う。補間とは、修飾される名詞を修飾節に欠けた成分としてあてはめる。補間の種類を図 5 に例示する。補間では、名詞に意味マークと修飾節の述語動詞の意味制限(主語、

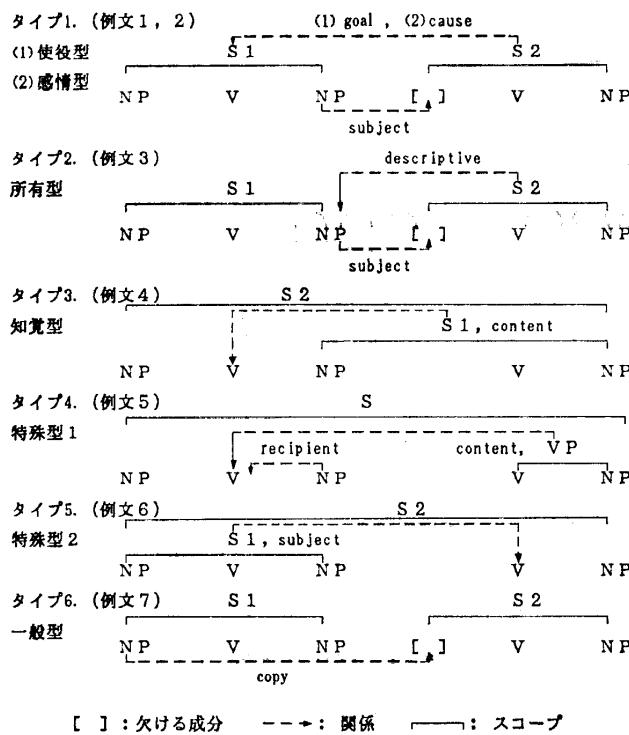


図4 連動構造 NP_V_NP_V_NP の6タイプ
Fig. 4 The six types of serial verb construction.

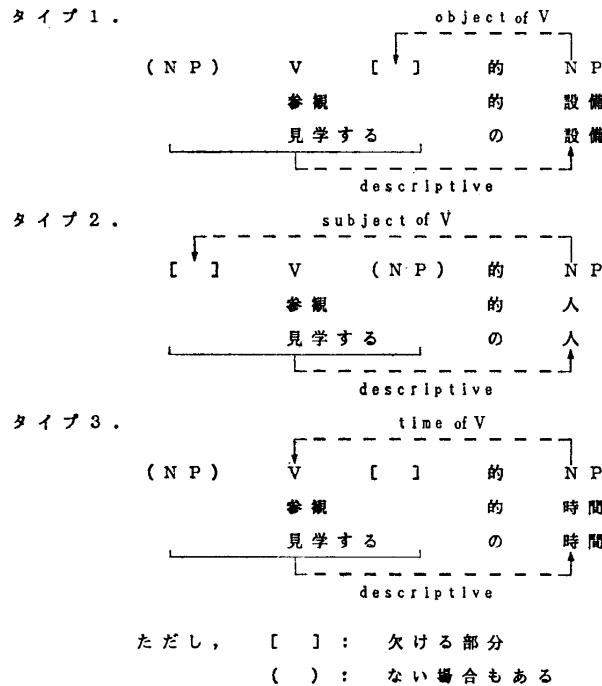


図5 「的」による修飾節補間の三つのタイプ
Fig. 5 The three types of implicit noun determination in descriptive clause.

目的語などに対しての)を検査する。補間規則は各動詞ごとに規定される。

2.2 優先規則

意味拘束条件として利用された知識は、第一に、100% 正しいものではなく、むしろ多くの場合正しいという常識的なもの。

第二に、完全ではないものの、すなわち、全構造よりも部分構造の意味対応に関する拘束条件、あるいは、意味関係のすべてよりも、典型的な現象しか扱わないものである。そのような規則を使うと、誤りや一つ以上の結果を導くことがありうる。そこで、我々は優先度を導入して、各格構造のもっともらしさの計算で最尤選択を求める。優先規則では、

1) 部分結果の優先度を各意味解析規則で与える。

2) 全結果の優先度は各部分の総和である。

ただし、各意味解析規則の優先度を経験的に決める。

優先度を最適格構造の選択を利用するだけではなく、句構造解析で得られた複数個の可能結果の比較にも利用する。その典型例は「是字句」の意味解析である。

「是字句」とは「A是B」(AはBである)の型の文である。「是」も動詞の一種であるが、「A是B」は中国語のもっとも一般的な文の一種で、その意味処理上の特殊性によって、ここで、特別な文の一種とする。解析中に問題になるのは、「是字句」が埋め込み文になったとき、その文がどこからどこまでになっているかの判定である。そこで、AとBの意味対応関係によって複数個の可能性に優先度を与える必要がある。

A是Bの文に意味制限を導入するため、名詞の意味分類を利用した。51個の名詞意味マークからその部分集合(14個)を選んで、「A是B」の典型意味パターンを作った。意味マークを図6に、AとBの意味対応関係に基づく優先規則を表2に示す。この優先規則では、意味対応関係を20対程度しか利用していない。部分的な規則であるが、典型パターンを強調する方針で単純化したもので、「肯定的」と「否定的」な意味パターンに、それぞれ「1」と「-1」の値を与え、その他の非典型的対応に対して優先度をゼロにする。

例えば、「前一個字母是T的可能性比較大.」(前

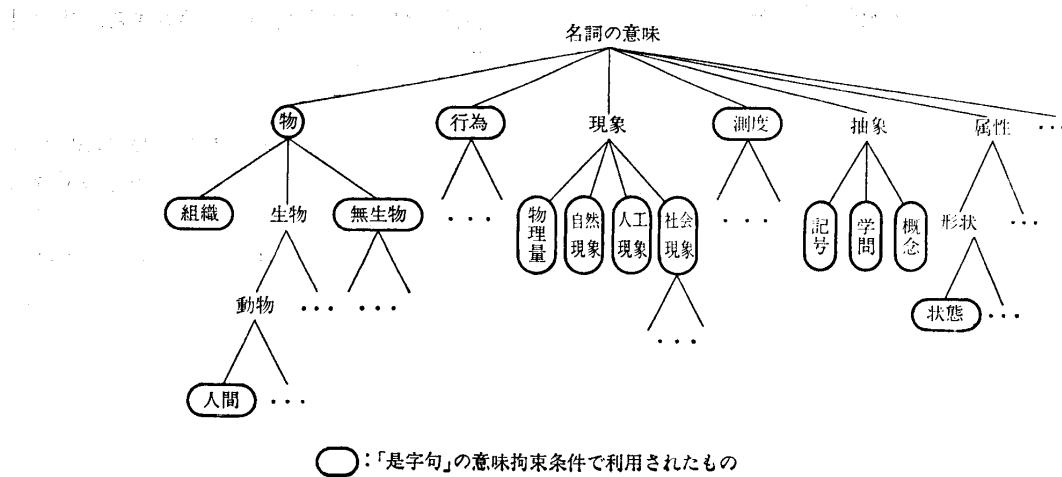


図 6 「是字句」解釈に用いた名詞意味分類と階層関係

Fig. 6 The semantic classification of nouns used in the analysis of sentence "A is B".

表 2 "A 是 B" の意味対応関係による優先度

Table 2 Priorities based on semantic correspondence in sentence "A is B".

A	測度	物理量	...	記号	學問概念	行為	状態	物	人間	組織	無生物	<VP>	*	**
測度	1	1		1								-1		
物理量		1						-1	/	/	/	-1		
...		1				1		-1	/	/	/	1		
記号			1									-1		
学問概念				1				-1	/	/	/	-1		
行為					1			-1	/	/	/	1		
状態						1	-1	/	/	/	/	1		
物							1					-1		
人間								1	1	-1	/			
組織									1	-1	/			
無生物										1	/			
<VP>											<CLAUSE>	1		
												*	1	1

ただし、表において、対応関係に互換性のある部分は省略した

<> : 統語カテゴリ
... : 人工現象、社会現象、自然現象
* : その他の意味マーク
** : *に属する下位の意味マーク
空白 : 優先度が零であるもの
/ : 上位の意味マークに従うもの

のシンボルが T である可能性は比較的大きい。」の解析には、「A 是 B」のスコープに 3通りの可能性がある(図 7 の a, b と c)。それぞれの意味対応関係は〈記号〉対〈記号〉、〈記号〉対〈属性〉、〈記号〉対文である。上記の優先規則によって a の方がしか選ばれない。

3. 意味処理系

3.1 対象指向の処理系

3.1.1 処理方式について

意味文法では、単語に依存する個別的規則を大量に利用する一方、多方面から分類的に整理された一般規則も必要とする。このような大量かつ性質の異なる知識の利用には、単に辞書(個別規則)中心のやり方あるいは、単に導出規則(例えば、CFG)で取り扱う方法は、いずれも困難を伴い、また不十分である。むしろ、複数個のアスペクトから任意のレベルでモジュール化された知識を統合的に利用でき、かつ各モジュール間の通信や切り換えに対し柔軟なアクセス・メカニズムを提供した処理系が適切である。対象指向の考え方は Smalltalk⁶⁾ や LOOPS⁷⁾ によってプログラミング言語に導入されたものであるが、我々はそれが自然言語処理に適するものとして、筆者らによって、英文解析に用いられた PAL

言語(Purely Application Language)とその処理系⁸⁾を基本にして、中国語の意味解析システムを作成した。以下において、対象指向処理型の一般的な記述は省略して、中国語意味解析に関する部分のみについて述べる。

3.1.2 意味解析規則の記述

意味解析規則は、句構造—表層的中間表現の対応規

則と、表層的中間表現—格構造の対応規則二つの部分からなる。

本システムでは、表層的中間表現を PAL 言語で記述する。PAL は内包論理の記述言語であり、すべて例文 B.

前一個字母是 T 的可能性比較大.
 a
 b
 c

- a : 前のシンボルが T である (可能性は比較的大きい).
 〈記号〉 〈記号〉
 - b : 前のシンボルが T の可能性である (ことは、比較的大きい).
 〈記号〉 〈属性〉
 - c : 前のシンボルは、T の可能性が比較的大きいということである.
 〈記号〉 文
- a. b. c : 「A 是 B」スコープの三つの可能性
 [] : スコープ
 < > : 意味マーク
 () : 説文

図 7 「A 是 B」のスコープ判断の曖昧性
 Fig. 7 The ambiguity in scope of sentence "A is B".

統語規則: $S \rightarrow NP VP$

PAL 生成規則:

導出規則の右辺 1 番目 (NP) は “* SUBJECT” という表層格ラベルで右辺 2 番目 (VP) に作用する。

PAL 式:

例文：“我是学生。” (“私は学生です。”)

$PAL = (*SUBJECT(我)) (*OBJECT(学生))(是)$
 ↑
 NP の PAL VP の PAL

図 8 PAL 式生成の一例

Fig. 8 An example of PAL generation.

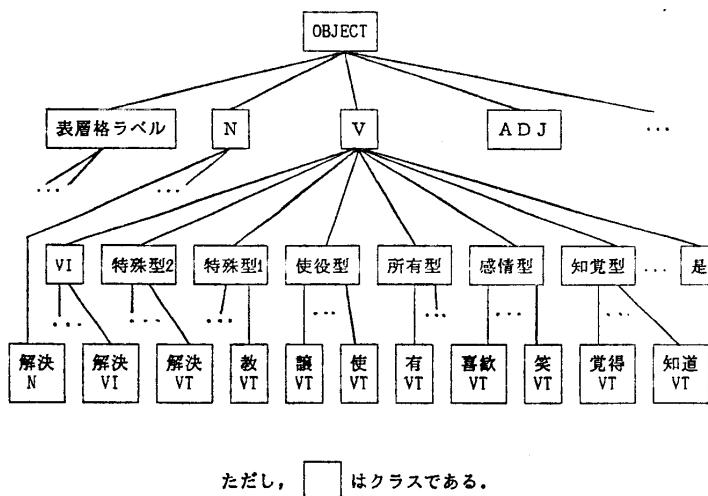


図 9 意味文法クラスの階層関係 (一部)

Fig. 9 The hierarchy of semantic classes.

の表現が作用子 (述語) と引数との二つの要素で作用的形式で (回帰的に) 書かれる。PAL 式は句構造から変換して得たものであり、句構造の各部分に対応する意味文法規則モジュールへのポインタ (単語あるいは表層格ラベル) を与え、意味評価の順序を決める。PAL の生成規則は各統語規則 (CFG の導出規則) に対して規定する。その例を図 8 に示す。

PAL から格構造を生成する規則をオブジェクト群を用いて記述する。

オブジェクト群をクラスの階層関係として構成する。クラスは、共通な性質をもつものであるが、本意味文法の記述に用いたクラスの種類と階層関係 (の一部) を図 9 に示す。文法の記述では、

1) 単語の各統語カテゴリのクラスを最下位クラスとし、その単語の意味属性と意味解析の個別規則を与える。

2) 上位クラスでは、意味解析の一般的な規則をモジュール (クラス) に分け、階層的に記述する。最上位クラスでは、最も一般化された (default) 規則を与える。

3) 各クラスでは、意味処理上に共通性をもつものの手続きをまとめる。同じ品詞に分類される語の多くは同じ手続きをもつので、クラスは各品詞、およびそれらの意味に基づいて分類されるものが多い。

4) クラス内部の意味解析規則には各表層格ラベルを付けて分類される。すなわち、表層格ラベルは各規則の呼び出し語である。

我々は、表 1 の表層格に従ってオブジェクトを作成した。例として、動詞「笑」(笑う) とその上位クラス「感情動詞」のオブジェクトを図 10 で挙げる。

3.1.3 意味処理系の制御方式

PAL 式は、統語解析で句構造規則とともに、統語規則に対応する PAL 生成規則を起動することによって生成される。

PAL 式を、対象指向なインタプリタによって評価すると、各語の項目にオブジェクトを代入することにより意味解析規則を呼び出し、格構造を生成しながら、その優先度を計算する。自然言語の処理系として、この制御方式では次の特

徴がある:

- 個別規則を優先する: インタプリタは「オブジェクトの実例 (instance)」—クラス (class)—上位クラス

動詞「笑. VT」のクラス:

上位クラス=他動詞		準上位クラス: SENSE 1=感情型	
格 解 析 規 則	表 層 格	規 则	
	後続述語	上位クラスを SENSE 1 で入れ換えて格解析を行ってから、その結果をもって、上位クラスを復元する。	
	主 語	深層格 <— 主体 (SUBJECT), 優先度 <— 意味マークは“人間”でなければ -1	
	前 置 詞	“対”, “跟”	深層格 <— 相手 1 (PARTNER) 優先度 <— 名詞意味マークは“人間”でなければ -1

上位クラス「感情型」のクラス

上位クラス=他動詞		準上位クラス: SENSE 1=一般型	
格 解 析 規 則	表 層 格	規 则	
	後方述語	深層格 <— 主述語の原因 (CAUSE) 補 間: 主述語の目的語で原因節の主語として入れこむ 優先度: 補間の意味妥当性によって計算する。	

図 10 動詞「笑」のクラスとその準上位クラス
Fig. 10 Examples of verb class.

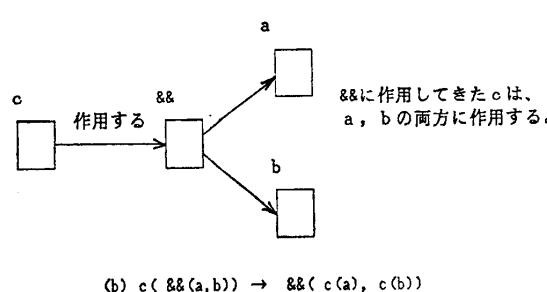
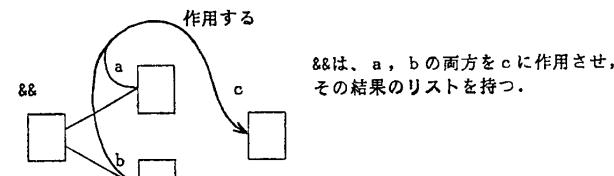


図 11 「&&」変換子の 2 種類の操作
Fig. 11 The two operations of the transformer “&&”.

(super class) — 上位クラス — …」のように、下位クラスから上位クラスへ順に指定された意味処理の規則が成功的に適用されるまでたどることによって下位クラスの規則を優先する。

2) 非決定的探索機能: 上位クラスを変数化して、解析中変数を代入することによって、下位クラスに複数個の上位クラスに対応させることができ、選択的あるいは並行的な処理を実現する。例として、動名詞の上位クラスを通常は「名詞」クラスに設定するが、前置詞が掛かってくる(2.1.3 節を参照)ときには、動名詞の上位クラスを一時的に動詞に対応させることによって、前置詞句の深層格を決める。

3) 中間結果の管理: 非決定的処理に導かれた複数個の中間結果を同様に扱うために、特別な変換子「&&」を用いる。「&&」の二種類の操作を図 11 に示す。さらに、中間結果 (instance) の中味を変更することに、コピーの手法を用いて互いの副作用を防ぐことができる。

3.2 優先度を利用した統語・意味解析の制御

今まで意味解析を(統語解析に)独立な部分として述べてきたが、次に統語解析と意味解析との関連について、優先度の利用を述べる。

中国語解析システムの統語・意味解析部の制御方式を図 12 に示す。

統語意味解析の入力は前処理系¹⁾の出力とする。前処理では、特殊語により入力文をブロック(「断片」と呼ぶ)化して、入力文の句構造を部分的に指定または予測した断片構造を、独立した最下位部分結果(単語あるいは指定された句構造)の集合に分解して、断片のもつトップ・ダウン的な予測情報や優先度(前処理で得られた情報)をボトム・アップ的に変形して各部分結果に付ける。統語意味解析はこのような部分結果の集合の上でボトム・アップ的に行われる。解析の各ステップを次に示す。

- 部分結果の集合から、優先度のいちばん高いものを選び出す。
- その部分木に対し句構造解析を一步進め、上位部分木を新しく作る。
- 新部分木に意味解析を行い、意味構造と優先度を与える、新しい部分結果として集合に入れてから、a へ。
- 以上各ステップを最終結果に到達するまで繰り返す。

この方式は、統語・意味解析のフィードバック・

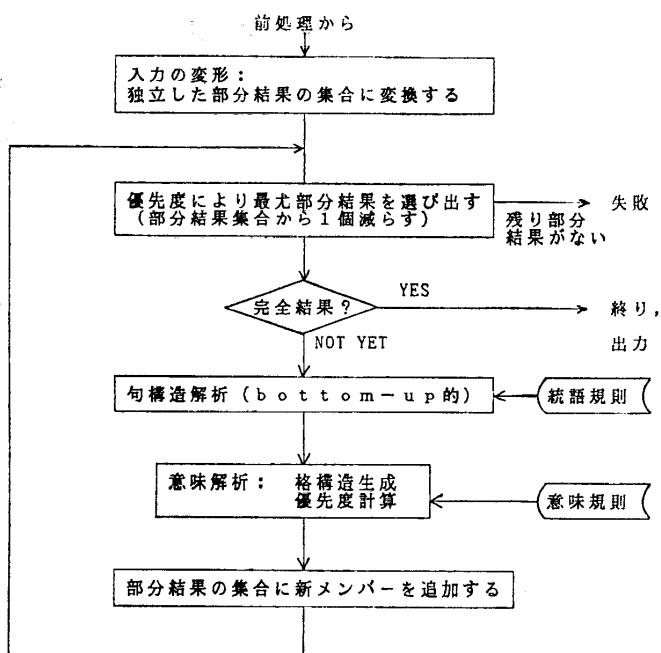


図 12 統語・意味解析の制御方式

Fig. 12 The control graph of syntactic-semantic analysis.

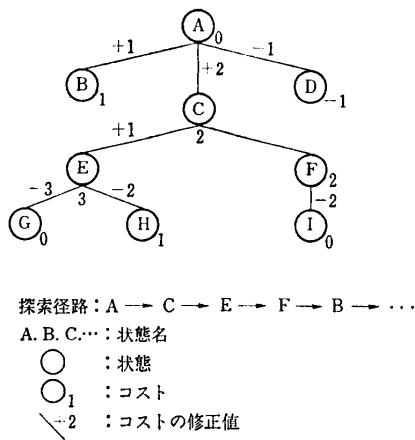


図 13 ベスト・ファーストの探索順

Fig. 13 Best first searching algorithm.

ループにより、意味解析で得られる部分結果の優先度を利用して次の探索方向を制御する（ベスト・ファースト探索）。図 13 では、このような探索順を例示する。優先度の利用により、解析を正しい方向へしづらし、最尤結果に先に届くことを保障する。残りの可能性も保存されているので、必要な場合（例えば、後のステップで失敗した場合）次の選択がシステムによって自動的に提供される。

4. 実験

本意味処理システムは汎用計算機上の UTI-LISP⁹⁾

によりプログラム化され、現在、部分的に稼動している。

意味文法およびシステムの有効性を確かめるため、種類の異なる、かつ解析上曖昧性のある典型例文 700 個^{4), 5), 10)}を考察した上、そのうちから 20 個を選んで、システムの上でテストした。それらの 20 個の典型例文において、連動構造が 9 個、前置詞句の例が 7 個、「～的」の例が 12 個、「是字句」の例が 8 個あり、意味拘束条件を利用した解析の結果、第一位選択で 70%，第二位選択まですべての正解が得られた。システムによる解析結果の例を図 14 に示す。前に述べた例文 3 (図 2)について、連動構造の 7 種の可能性から、入力文の正しい格構造を一番目の出力として得ることができた。

5. 結論

本論文では、中国語意味解析に必要な知識を、意味処理系向きに規則化すること、意味解析を対

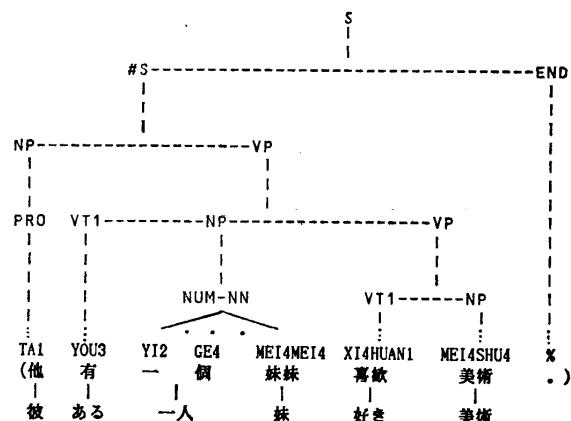


図 14 システムによる意味解析結果の一例

Fig. 14 An example of the result obtained by the semantic analysis system.

象指向処理系の上でシステム化すること、および優先度処理を用いてヒューリスティック探索を実現することについて述べた。意味文法では、曖昧性解消に有効となる意味拘束条件を、明確に定義した概念の有限集合（表層格、深層格、動詞や名詞などの意味マーク）の上で規則化した。なお、もっともらしさの比較に関する知識を優先度として各規則に組み込んだ。処理系としては、対象指向系の採用により、意味分類の上で階層的にモジュール化された意味文法に、柔軟なアクセス・メカニズムを提供したことが特徴である。優先度処理は、ベスト・ファスト戦略を用い、ボトム・アップ・パーザの上で、意味解析の結果で統語解析の探索方向を制御するように、探索空間の爆発的に増大することを避けた。この研究は人間のよく使うヒューリスティックな知識を、中国語意味解析システムに利用することを中心にして展開されるものであり、本意味解析方式は、曖昧性の多い自然言語、特に中国語の処理に適切かつ有効な解決方法を与えた。

本統語・意味解析システムと、すでに作成してあるわかつ書きシステム、および特徴語で局所的に統語解析を行う前処理システムとの結合を完成し、中国語解析の全体的システムを充実することが今後の残された研究課題である。

謝辞 末筆ながら、本研究に多くのご教唆をいただいた京都大学工学部の西田豊明助手、ご討論いただいた京都大学工学部の辻井潤一助教授、中村順一助手ならびに堂下研究室の諸氏に感謝いたします。

参考文献

- 1) 楊 頤明, 堂下修司, 西田豊明: 中国語解析システムにおけるヒューリスティックな知識の利用, 情報処理学会論文誌, Vol. 25, No. 6, pp. 1044-1054 (1984).
- 2) 長尾 真: 科技庁機械翻訳プロジェクトの概要, 情報処理学会自然言語研究会資料, 38-2 (1983).
- 3) King, M.: Design Characteristics of a Machine Translation System, Proceedings of the Seventh IJCAI, Vancouver, B.C., Canada, Vol. 1, pp. 43-46 (Aug. 1981).
- 4) Li, C. N. and Tohomponson, S. A.: *MANDARIN CHINESE—A Functional Reference Grammar*, University of California Press, Berkeley, CA (1981).
- 5) 呂 叔湘: 現代漢語八百詞, 商務印書館, 北京 (1980).
- 6) Goldberg, A. and Robson, D.: *Smalltalk-80—The Language and Its Implementation*, Xerox Palo Alto Research Center, Addison-Wesley Publishing Company, Reading, MA (1983).
- 7) Bobrow, D. G. and Stefik, M.: The LOOPS Manual, Working Paper, Memo KB-VLSI-81-13, Xerox Corporation, Pasadena, CA (1983).
- 8) 西田豊明, 川村 正, 堂下修司: LSI の動作記述からの知識獲得について, 情報処理学会知識工学と人工知能研究会資料, 38-6 (1985).
- 9) Chikayama, T.: UTILISP Manual, Technical Report, METR 81-6, University of Tokyo, Tokyo (1981).
- 10) 朱 德熙: 語法講義, 商務印書館, 北京(1982).
(昭和 60 年 3 月 20 日受付)
(昭和 60 年 9 月 19 日採録)