

意味および文脈情報を用いた日本語文の 解析—名詞句・単文の処理*

長尾 真** 辻井潤一** 田中一敏**

Abstract

We have constructed a powerful parser of Japanese sentences. This parser can resolve ambiguities of input sentences by utilizing various kinds of semantic relationships among words. Especially, the 'case' concept introduced by C. J. Fillmore¹⁾ plays an important role in this parser. Case is the relation which is established between an object and an event. In a sentence, an object is usually expressed by a noun phrase and an event by a verb.

The information as to what kinds of cases should be described and what kinds of nouns should be selected as case-elements of a verb of a sentence, is called 'case-frame' of the verb and is described in the verb dictionary.

A verb may have more than one case frames corresponding to different usages of the verb. A noun represents a set of objects which have certain same properties. We express properties by attribute-value pairs in the noun dictionary. The dictionary also contains the set-inclusion relationships among nouns. In this paper, we will explain how the parser utilizes these dictionaries in order to analyze a noun phrase and a simple sentence.

1. まえがき

従来自然言語文の構文解析では、文法的な解析(統語処理—syntactic analysis)のみで、意味を取入れた解析(semantic analysis)は殆んどなされなかったため、曖昧さの解消が大きな問題となり解析に限界があった。ところで英語においては最近になって、入力された自然言語の語る「世界」を限定することにより、あるいは意味論を詳しく展開することにより構文解析を行い、かなりの成果をおさめてきている^{1)~4)}。しかしながら日本語では、栗原等が意味的基本構文の考えを用いた分析を行っている^{5),6)}。他は、この種の研究はあまりなされておらず、意味、文脈を有機的に結びつける日本語の文法体系はまだまだ模索の段階である。そこで我々は、日本語の特徴をふまえ格文法の考え方

を導入し、意味分析だけでなく、前後の文が表現する文脈情報(context)をも用いた構文解析を行うことにした。

なお、本論文では主として意味辞書を用いた名詞句・単文の分析について述べ、別の論文(参考文献13))で文脈の情報を用いた分析について報告する。

2. 構文解析の方針

我々は、一応対象とする分野を中学校理科化学の分野に限定し、以下の方針のもとに文の分析を行った。

- (1) 統語処理・意味処理・文脈処理を同時に行う。
- (2) 単語間の意味的なつながりを記述した意味辞書を用いる。
- (3) 対象分野を中学校理科化学の分野に限定する。
- (4) プログラム言語：記号処理用言語 LISP ならびに統語処理・意味処理・文脈処理の結合を容易にする PLATON¹⁰⁾を用いる。
- (5) 我々は日本語の特徴が、主として動詞と格助

* Analysis of Japanese Sentences by Using Semantic and Contextual Information—Semantic Analysis, by Makoto NAGAO, Jun-ich TSUJII and Kazutoshi TANAKA (Faculty of Engineering, Kyoto University)

** 京大工学部電気工学第2学科

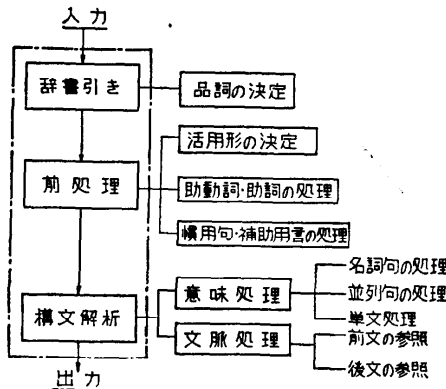


Fig. 1 System Construction

詞とによって規定されることに着目し, C. J. Fillmore の提唱する格文法の考え方を基礎におき¹¹⁾, なるべく多くの例文を調べることにより種々の heuristic な手法で解析を行う。

- (6) 文脈情報を名詞を中心に集約し, これを参照して省略語の補足・代名詞や指示連体詞の処理等を行う。

全体の構成を Fig. 1 に示す。なお入力文はローマ字で, 文節単位に分ち書きされている。

3. 名詞辞書の構造と名詞句の分析

3.1 名詞辞書

名詞句は, 格助詞「の」を介在して多くの名詞が連なり, 一つのまとまった意味を表わす。これを分析するためには, 各名詞の持つ意味内容がわからなければならない。例えば

「試験管の中の 100 g の質量の青色の液体の温度」において意味を考えない場合, 132 通りの係り受けが成立する。このことからそれぞれの名詞のもつ機能や意味を与える辞書を, 「解析する立場」から構成することが必要となる。このような立場から我々の作った名詞辞書は

- (1) 他の名詞との結合の仕方
- (2) 名詞間の上位—下位概念のリンク
- (3) 名詞がもつ属性ならびに属性値

の3つの側面から構成されている。

3.1.1 名詞のカテゴリ

一般に名詞は具体的な概念を指示すること (entity word) が多いが, ある種の名詞においては, 他の単語と関係してはじめて意味の定まることがある (relational word)。このような名詞に対して固有の意味記

Table 1 Categories of Nouns and their Examples

カテゴリ	記号	例	語数
物質名詞	ME	水素, 試験管	28
属性名詞	MZ	色, 温度, 質量, 状態	
属性値名詞	MZV	青色, 気体, 100°C	41
サ変動詞	MS	化合, 実験, 測定	29
機能名詞	MFUNC	前, 中, 上	35
形式名詞	MK	こと, もの, とき, ため	9
形容詞的名詞	MADJ	全体, 別, 一定	24
量関係の名詞	MQ	比, 和, 差	6

述を与える必要がある。

例えば, サ変動詞の語幹が名詞化 (nominalization) した名詞—サ変名詞—は動詞と類似の働きを示す。即ち「質量の測定」は「質量を測定する」と同じ意味である。また「前」, 「中」, 「上」等の名詞は英語の前置詞に相当する役割を果たし, 前にくる名詞に応じて時間・場所等の意味を表わす。このような relational word としての名詞は MF というフラッグのもとにそれぞれの機能を表わす記号をもち, その機能に応じた意味の記述がなされている。その記号と名詞の例を Table 1 に, 辞書例を Fig. 2 に示す。

3.1.2 上位—下位概念

上位—下位概念とは, 物質—化合物—食塩などの名

(ONDO M ((MF MZ) (ZST ONDO)* (SP RYOU ZOKUSEI)))
(OOKISA M ((MF MZ) (ZST TAISEKI MENSEKI)* (SP RYOU ZOKUSEI)))

(AOIRO M ((MF MZV NREF) (Z-ADJ (IRO BLUE)**))

(KAGAKUHENKA M ((MF MS NREF) (SP ACTIVE) (MCASE SUBJ)**))

(SOKUTEI MC (MF MS NREF) (SP ACTIVE) (MCASE OBJ**))

(MAE M ((MF MFUNC NREF)
(FTRANS (ACTIVE JIKAN BEFORE)

(T BASHO IN-FRONT-OF))****

(TAME M ((MF MK)(KTRANS
(BUN TAME GOAL CAUSE)
(BUSSHITSU TAME CAUSE)
(ACTIVE TAME GOAL CAUSE)))

(ITTEIRYOU M ((MF MZV MADJ NREF) (Z-ADJ (RYOU
CONSTANT)

(SHITSURYOU CONSTANT) (TAISEKI CONSTANT)))

(WA M ((MF MQ MFUNC NREF) (FTRANS ((MQCHECK)
WA SUM) (T WA SUM)))

* ZST: 属性名詞に対して, それを示す標準属性名が示される。

** Z-ADJ: 属性値名詞に対して, それが形容する属性と, その値が示される。

***MCASE: 「M+ の +MS」の構造をとった時, 前の名詞が入り易い格を記す。この例では, 「酸素の化学変化」, 「銅の化学変化」等のように, 前の名詞が「化学変化する」の主格に入り易いことを示す。一般に MCASE には, 複数個の格が記述される。

****FTRANS: 「前にくる名詞の条件」, 「機能名詞の表わす意味」, 「前名詞とのリンクの種類」の3字組の集合が記述される。

Fig. 2 Typical Examples of Noun Descriptions

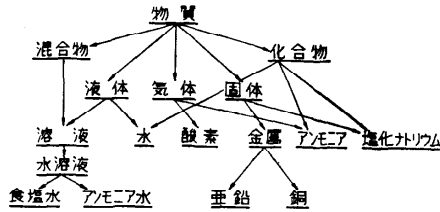


Fig. 3 Super-sub Relationships among Nouns

詞間の包含関係を示すもので、トリーではなくFig. 3の様な構造になる。この上位一下位の名詞は次項で述べる属性-属性値対による意味の特殊化により上位の名詞から下位の名詞へ移行する。例えば、化合物-酸化物ではその化合物を構成する要素を酸素に限定したとき、酸化物というのである。この上位一下位概念の記述は、動詞の格構造の記述と関連づけられ、動詞文の処理の際に大きな役割を果たす。

3.1.3 属性-属性値対

我々は、名詞が内包する意味に対して属性ならびに属性値という考え方で記述を行った。この属性の一覧をTable 2に、またFig. 4に「物質」・「液体」の辞書内容を示す。「物質」の辞書内容において(体積)、(温度)等の様に属性値が記されていないものはその属性に対して自由な値をとり得ることを示している。一方「液体」の辞書内容としては上位概念である「物質」に関して特殊化された部分のみが記述されている。即ち「液体」の属性「状態」は液体であり、また

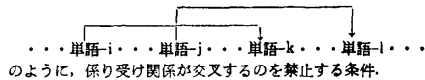
Table 2 Table of Attributes

属性	記号	属性	記号
体積	TAISEKI	酸-アルカリ	S-A
面積	MENSEKI	化学式	CEF
長さ	NAGASA	粘性	NENSEI
温度	ONDO	厚み	ATSUMI
密度	MITSUDO	形	KATACHI
質量	SHITSURYOU	かたさ	KATASA
質量	IRO	濃度	NOUDO
状態	STATE	値	VAL
におい	NIOI	部分	P
名前	NAME	全体	W
成分	SEIBUN		

物質 ← ((ATR (体積) (温度) (密度) (質量) (色) (形) (状態) (酸-アルカリ……)))
 液体 ← ((SP 物質) (ATR (状態 液体) (形 NIL) (粘性)))

Fig. 4 Example of Attribute-Value pairs

* 係り受け関係に対する制限条件で



属性「形状」は持ち得ないことが示されている。

ある名詞の属性-属性値対は上位概念の名詞を次々とたどることにより全て出すことができるが、その際下位の属性値ほど優先される様になっている。

これにより形容詞-名詞、名詞-名詞結合の意味的なチェックを行うことができる。

3.2 名詞句の処理

3.2.1 名詞二語間の処理

名詞句の意味を解析するには、名詞二語の間の解析が基礎となる。現在、化学の分野に見られる15種類の係り受けが処理可能である。構成としては、名詞のカテゴリによりまず大きく分類し、その後意味辞書を用いて係り受けのチェックを行うものである。

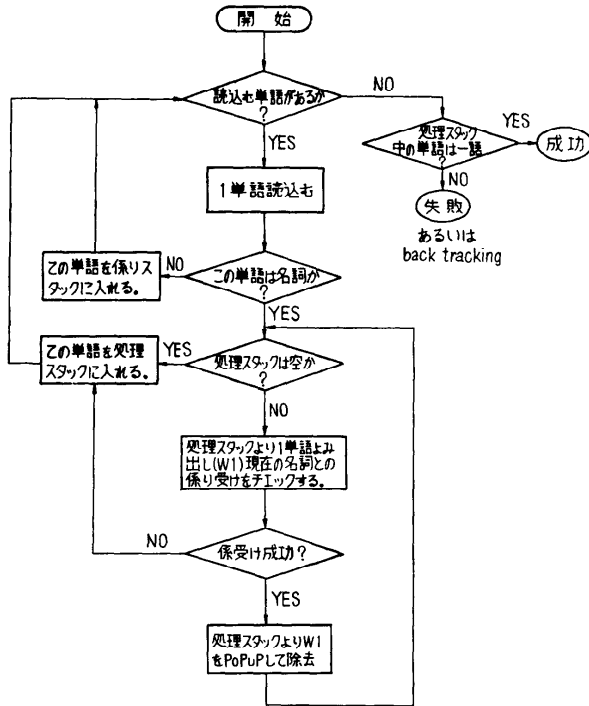
この15種類の係り受け関係は用例の頻度ならびに二語の結合の強さを考慮に入れた処理順位に従って処理が行われる。Table 3に名詞二語の用例とその頻度を示す。

3.2.2 名詞句の処理

名詞句の処理は前項で述べた名詞二語の処理並びに非交叉条件*を用いて行う。Fig. 5(次頁参照)に処理のアルゴリズムを示す。このアルゴリズムはPLATONで書かれ、バックトラッキングの機能を用いて、相当複雑な名詞句でも処理できる。また、Fig. 6(次頁参照)の様に同一の内容を表わしている名詞句は、表現

Table 3 Examples of Noun-noun Combination and their Frequency

処理順位	分類	例数	例
1	MZV-MZ (MZ-MZV)	10	100gの質量 気体の状態(の酸素)
2	MZV-M (M-MZV)	74	青色の固体 銅 50g
3	MADJ-M	55	もとの物質 余分の体積
4	M-MZ	79	亜鉛の質量 液体の色
5	M-MFUNC	118	試験管の中 反応の前
6	MZ-M	13	(100gの) 質量の銅 (高い) 温度の液体
7	M-MS	58	酸化銅の還元 水素の発生
8	TIME-M	11	(反応の) 前の物質 (~した) 時の水
9	PLACE-M	33	(ビーカーの) 中の水
10	M-MK	17	化合のため
11	ELEMENT-M	13	ナトリウムの化合物 食塩の水溶液
12	M-ELEMENT	3	酸化銅の酸素
13	SB-SP (例示)	3	ナトリウム(など)の金属
14	NAME-M	28	質量保存の法則
15	W-P (全体一部分)	10	アルコールランプのしん
16	その他	42	1cm ² あたりの力 化合物の調べ方



* 係りスタック…形容詞, 形容動詞, 連体詞などを入れておくスタック
そして次に名詞が入ってきたとき, これらの単語と名詞のチェックがなされる。

Fig. 5 Algorithm of Analyzing a Noun phrase

が種々の形をとっていても同一の処理結果になる。

3.2.3 並列句の処理

名詞句の中の並列要素は次の2つに分類できる。

- (1) 単純並列句: 名詞2語が単純に並列となるもの。

【例】 銅とイオウ, Na や Cu

- (2) 複合並列句: 並列要素が名詞句中で係り受け関係をもつもの (タイプ(a)~(d))

【例】 (a) 銅の質量とイオウの質量

(b) 銅とイオウの質量

(c) 気体の体積と質量

(d) 硫化銅の銅とイオウの質量の比

これらを解析するために次の様なアルゴリズムを作成した。

① 並列助詞「と」を見つけたら, それに呼応する「と」があるかどうか調べる。あれば最初の「と」の前の名詞 (M1) と2つの「と」にはさまれた名詞句が並列となる。この名詞句を処理して④へ。

② 呼応する「と」がない場合, あるいは他の並列

(a) 表現の多様性

- 100gの{x}固体
- 100gの質量の{x}固体
- 質量100gの{x}固体
- {x}100gの質量の固体
- {x}質量100gの固体
- {x}100gの固体
- {x}固体100g

ただし

$$\{x\} = \begin{cases} \text{白色の} \\ \text{白い} \\ \text{色の白い} \\ \text{白い色の} \end{cases}$$

(b) 解析結果

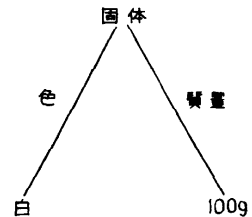


Fig. 6 Variety of structures and the Result of parsing

助詞, コンマの場合は, それらのデリミターの前の名詞と同一の語を探す。

②' なければ同一のカテゴリの名詞をさがす。

②" なければ共通の上位概念をもつ名詞をさがす。

見つかった名詞を M2 とすると,

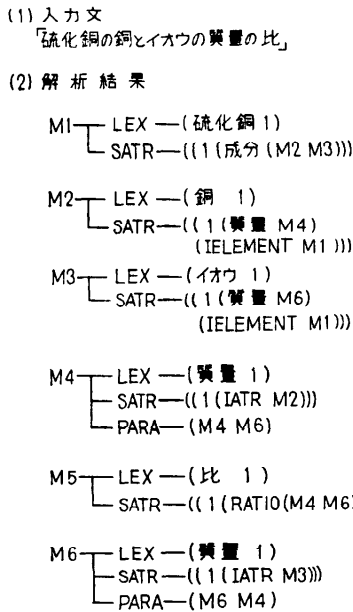
③ デリミターから M2 までの名詞句を処理する。

④ デリミターより前の名詞句 (NP1) を処理する。

⑤ M2 に係り得る名詞を NP1 の最初の方からさがして処理する。

⑥ M2 の後の名詞列と, M1 及び M2 との係り受けの処理を行い, 双方が係る場合は新しく名詞を作り出して, それを新しい並列要素とする。

一例として例(d)の「硫化銅の銅とイオウの質量の比」について述べる。アルゴリズムの②'により銅とイオウが並列になる。次に③の処理は名詞が1つだけ(イオウ)なので終了し, ④へくる。④では「硫化銅の銅」を M-ELEMENT (銅は硫化銅の成分である) の関係にあると解析し, また⑤によりイオウに対して



(注) 処理後、並列になっているものは二つの質量であり、銅とイオウではない。

Fig. 7 An Example of Analysis of Coordinate Conjunctions

「硫化銅のイオウ」を処理する。これも M-ELEMENT と判断される。次に⑥に従って、「銅の質量」、「イオウの質量」が解析され、新たにイオウのための名詞「質量」が発生される。結局「質量(—銅の—)と質量(—イオウの—)の比」を解析することになる。解析結果を Fig. 7 に示す。

4. 格構造と動詞辞書

名詞句の解析において名詞の意味辞書を用いたのと同様に、動詞を中心とした構文を解析する際には、動詞の意味辞書が大きな役割を果たす。特に動詞の連体形による埋込み文の処理においては、名詞と動詞の意味的なつながりの考慮なしでは解析が不可能であることが多い。また次の様な構造の文は、日本語においては頻繁にみられる。

(例) A班がピーカーに熱した硫酸銅水溶液を入れる。

ここで、「A班が」、「ピーカーに」の連用修飾句は統語的な観点からは「熱した」、「入れる」のいずれにも係り得る。これを決定するためには、各名詞と動詞

CF-1: ((ACT 人間) (OBJ 物質) (IN 液体))
 CF-2: ((ACT 人間) (OBJ 固体) (INST NDOUGU))
 CF-3: ((ACT 液体) (OBJ 物質))

Fig. 8 Case-Frames of the Verb *TOKASU*

が意味的に係り受け可能かどうかを調べる必要がある。更に「硫酸銅水溶液」は埋込み文の構成要素となっているが、いかなる関係でこの埋込み文に入るかを決定する際にも意味処理が必要となる。

動詞辞書は格文法の考え方に従って整理されており、その動詞の持つ格構造が記述されている。格構造とは、その動詞が要求する格と、その格に入り得る名詞の対の集合である。Fig. 8 に動詞「溶かす」の格構造を示す。この格構造は、名詞の意味辞書と組合せて用いられ、例えば「溶かす」の行為者格としては、「人間」あるいは「液体」の下位概念である名詞が優先して選出される。

4.1 必須格と自由格

動詞は種々の格成分を要求するが、その格は大きく2つに分けることができる。それを必須格・自由格と呼ぶ。必須格とは、それがなければ、動詞の意味が確定しなくなる格をいう。例えば、動詞「熱する」は①熱する主体、②熱される客体、③熱する際に用いる道具、の3つを必要とする。

ところで「ナフタリンをガスバーナーで熱する。」においては、①の主体は文中に現われていない。しかしこれは前後の文脈、あるいは文が述べられた状況により省略されているだけでインプリシットには表わされているのである。(特に化学の文章では主体を表わす人間は殆どの場合省略される。これは実験などの操作を行う人間がその場面で常に設定されているからである。)

自由格とは、いつ(時間)、どこで(場所)、何故(原因・理由)等の外的状況を示す格であり、自由に動詞に係ることができる。従って、辞書中には必須格成分のみが記述されている。

Table 4 (次頁参照) に格助詞と格の関係を示す。

4.2 表層格構造と深層格構造

構文解析の究極の目的は、意味が同じであれば異なる表層の表現(入力文)であっても同一の深層の表現形式にすることであるが、入力文を直接に深層表現に変換することは不可能に近く、またアルゴリズム、プログラムは複雑となる。そこで動詞辞書には、解析の際に利用される表層格構造と、それに対応する深層格構造とを対にして記述し (Fig. 9 次頁参照)、1つの

Table 4 Post-Positions in Japanese and Case Relationships

格助詞	格
が	ACT, SUBJ
の	NMOD (ACT, SUBJ)
を	OBJ
に	RESULT, IN, IOBJ, TO, PLACE, CAUSE, TIME
へ	TO
と	FACT, RESULT, TAISHO
から	FROM, SOURCE, CAUSE, METHOD, PLACE, TIME
より	FROM, SOURCE
で	INST, SOURCE, CAUSE, METHOD, PLACE, TIME
まで	TO

は
副助詞

SSUB
ALLCASE

- * 「は」, 副助詞は前処理でそれぞれ SSUB, ALLCASE に変換される。
- * SSUB, ALLCASE とともに上記の様々な格をとり得る。
- * 「まで」は現在格助詞の扱いをする。

単文の解析が終わった時点で深層格構造におとすことにした。

例えば, Fig. 9 に示した「混合する」の辞書内容により

- (1) 物質Aに物質Bを混合する。
- (2) 物質Aを物質Bと混合する。
- (3) 物質Aと物質Bを混合する。

を同一の深層格構造に変換することができる。

- CF-1: (((ACT 人間) (OBJ 物質) (IOBJ 物質))
(D-CASE ((ACT 人間) (OBJ 物質 PL))
(ACT ACT) (OBJ (PARA (* OBJ) (* IOBJ))))))
- CF-2: (((ACT 人間) (OBJ 物質) (TAISHO 物質))
(D-CASE ((ACT 人間) (OBJ 物質 PL))
(ACT ACT) (OBJ (PARA (* OBJ) (* TAISHO))))))
- CF-3: (((ACT 人間) (OBJ 物質 PL)))

Fig. 9 Case-Frames and Deep-Case Structures of the Verb *KONGOUSURU*

5. 単文の分析アルゴリズム

動詞文の処理は単文の処理が基礎となる。複文・重文などの処理はこの単文処理のプログラムを帰納的に使用することにより可能となるので, 本章では単文処理における種々の考え方, アルゴリズムについて述べる。処理の構成としては,

- (1) 文中から動詞を見出す。
- (2) 動詞より前の部分を取り出し, それを格助詞の存在によってセグメンテーション (segmentation) を行う。ただし「の」の所では切らない。
- (3) セグメンテーションされた格助詞より前の部分を切り出し, その構造を名詞句処理にわたす。
- (4) (3)をすべての名詞句に対して行い, それが

終了すると, 動詞に近い名詞句の文節から順次, 動詞との係り受け関係を調べる。

- (5) 動詞の格構造を参照し, まだ埋められずに残っている必須格を埋めるために前後の文脈を参照する。

この(5)は論文(13)を参照していただくことにし, 本章では(4)について述べることにする。

5.1 係り受けと多重処理

名詞と動詞との係り受けの可否は主として①名詞句に付いた格助詞が示し得る格の集合, ②名詞句の中心名詞, ③動詞の格構造, の3つの情報を用いて行われる。

動詞の格構造は比較的上位の名詞との関係という形で記述されている。従って「食塩を水に溶かす。」という文が, 「溶かす」の格構造——((ACT 人間) (OBJ 物質) (IN 液体))——に合致するかどうかは, 名詞の上位一下位のリンクを上位の方にたどり, 食塩の上位概念として物質があること, 水の上位概念が液体であることにより調べられる。この様に, 格助詞による統語的な情報のほかに, 動詞の格構造及び名詞の上位一下位のリンクによって示される意味的な情報をその都度参照することにより名詞—動詞間の係り受け関係が決定される。

また格構造に記述された必須格成分のチェックが終わっても係り受けが成立しないときには自由格成分とみて, 再びチェックが行われる。即ち格助詞が自由格を取り得るものであるかどうかを調べ, 取り得る場合には名詞が時間・場所・理由等を表わす名詞であるか否かを調べる。例を Fig. 10 (次頁参照) に示す。

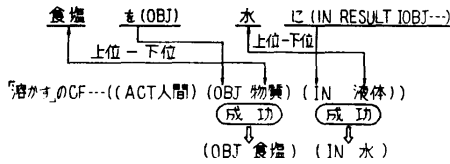
しかしながら, 上で述べた情報は次の様な点で必ずしも一意的な決定を与えるとは限らない。

- (1) 各動詞には複数個の用例がある。即ち, それらに対応する複数個の表層格構造が存在する。
- (2) 「格助詞」がなく「副助詞」のみが付いた名詞句については, 格を明確に規定できず, 様々な格をとり得る。また格助詞でも普通複数個の深層格に対応する。
- (3) 埋込み文によって形容されている名詞は埋込み文の格要素になることが多いが, 統語的にはどの格で入るかについての情報が無い。

そこで我々は, 全ての可能な係り受け関係が解析結果として得られる様にプログラムを構成した。

また, 非交叉条件を考慮して, 係り受けのチェックは動詞に近い「名詞+格助詞」から行われる。

(1) 必須格の処理例 「食塩も水に溶かす」



(2) 自由格の処理例 「熱したときに水素が発生した」

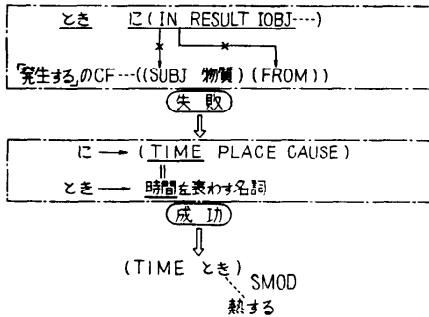


Fig. 10 Determination of Case-Relations between Nouns and Verbs

Table 5 Classification of Embedded Sentences

分類	用例数	用例
連体形+形式名詞	94	(a) 電気分解したとき、水の量はへる (b) 銅を熱したときの色の变化
連体形+名詞句	112	(c) 熱した銅の質量を測る (d) 熱した質量 100 g の銅 (e) 測定した銅の質量 (f) 試験管に入れた液体の化合物 (g) 試験管に入れた液体の体積 (h) ビーカーに水を入れた試験管を入れる (i) 固体が液体になる温度 (j) 体積と質量を測定した結果
名詞+の+連体形+名詞句	18	(k) 色のついた液体 (l) 水のもつエネルギー (m) 水蒸気のできる変化 (n) 青色のとけた金属
連体形+格助詞+接続助詞	20	(o) 水素を発生するので (p) 水を分解するには (q) 粒子が分離するだけ

5.2 連体形の処理

名詞を形容する埋込みの動詞文では、動詞とその動詞文が形容する名詞との関係を調べる必要がある。例文 250 程度を調べた結果を Table 5 に示す。この分類に応じて、4 つの処理アルゴリズムを作成した。以下それらを簡単に述べる。

(1) 連体形+形式名詞+(名詞句) (例: Fig. 11 (a)(b))

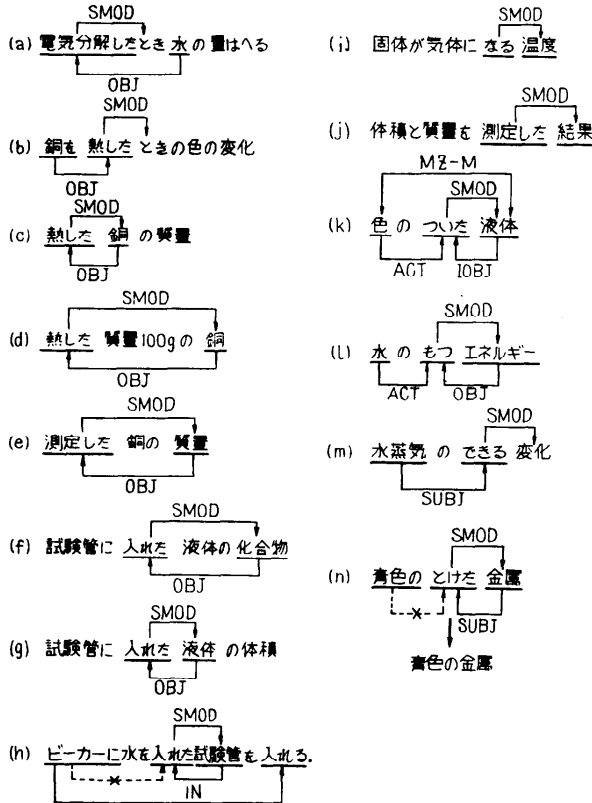


Fig. 11 Examples of Analysis of Embedded Sentences

この形をとる時には

- (i) 埋込み動詞文は直後の形式名詞を形容する。
- (ii) 形式名詞の後に名詞句が続く場合には、名詞句内の名詞が動詞文の格要素になることが多い。

従って、処理手順としては、動詞文を解析した後、それを形式名詞に文形容 (SMOD) の関係でかけ、形式名詞より後の名詞が動詞文中の空白格要素になり得るかどうかを順次チェックする。

(2) 連体形+名詞句 (例: Fig. 11 (c)~(j))

この形式をとる用例が最も多く、しかも係り受け関係、処理プログラムともに複雑となる。特に名詞句内に複数個の名詞がある場合には、どの名詞が埋込み文により形容されているか、また、その名詞が埋込み文中のいかなる格要素に相当するかを決定しなければならない。

処理手順としては、動詞の後の名詞句を処理

した後その名詞句内の各名詞を順々に動詞の直前において動詞文を解析する。ここでは格助詞の情報がないために主として動詞と名詞の意味辞書により係り受けが調べられる。

また名詞の内容を埋込みの文が述べているだけのもの、即ち名詞句中の名詞が動詞の格要素にならない場合には、単に文形容の関係で、名詞句の最初の名詞にかける。

(3) 名詞+の+連体形+名詞句(例: Fig. 11(k) ~ (n))

この構造をもつ埋込み文は次の3つに分けられる。

(i) 「属性名詞+の+連体形+その属性をもつ名詞」の形をもつもの。属性名詞は埋込み文の表層の主格または行為者格となる。従って、「の」を伴う名詞が属性名詞であれば、その属性を持つ名詞を後続の名詞句から探し出す。また「の」を「が」に変えて動詞文の処理を行う。例えば「体積の増加した気体」。

(ii) 「の」の前の名詞は属性名詞ではないが埋込み文の表層の主格または行為者格となるもの。例えば「水蒸気のできる変化」。

(iii) 名詞句の中に埋込み文があり「の」の前の名詞は埋込み文とは直接関係していないもの。例えば「青色のとけた金属」。この場合は、「連体形+名詞句」の処理を(2)の場合に従って行い、その後で、「名詞+の+名詞句」を名詞句として係り受けを決定する。

この3種類の係り受け関係を決定するための手順を Fig. 12 に示す。

(4) 連体形+格助詞あるいは接続助詞
これは格助詞・接続助詞との接続(結合)において

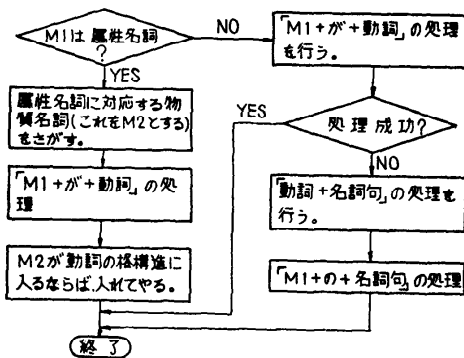


Fig. 12 Analysis of the construction 'Noun + Postposition 'No'+ Verb + Noun-phrase'

連体形をとっているだけなので、解析は連用形や終止形と同じ処理を行う。

5.3 受身文・使役文の処理

化学の文においては、行為者である人間は殆どの場合省略され、物質等に着目して事柄が述べられるため、しばしば受身・使役の形がとられる。処理としては、格助詞の表わす格を変換規則により読みかえて、動詞の格構造との係り受けを調べるものである。使役文に関する変換規則・処理例を Table 6, Fig. 13 に示す。変換規則による係り受けのチェックが失敗した場合には、表に示す様に再びもとに戻して処理を行う。

Table 6 Rewriting Rules of Cases in a Causative Sentence

格助詞	変換後	変換後の処理が失敗	例
(1) を に 連体 CASE* ALLCASE**	ACT IOBJ OBJ OBJ	OBJ IOBJ 以外の *に* の格 OBJ 以外のすべての格 OBJ 以外のすべての格	Fig. 13 (a) (a) (d)
(2) を に 連体 CASE ALLCASE	SUBJ ACT SUBJ SUBJ	OBJ *に* の示す格 SUBJ 以外のすべての格 SUBJ 以外のすべての格	Fig. 13 (b) (b) (c)

* 連体 CASE; 埋込み文によって形容されている名詞につけられるダミーの格助詞

** ALLCASE; 動助詞のみを伴った名詞句につけられるダミーの格助詞

*** (1) ...動詞の格構造に IOBJ がある場合の変換規則

(2) ...動詞の格構造に IOBJ がない場合の変換規則

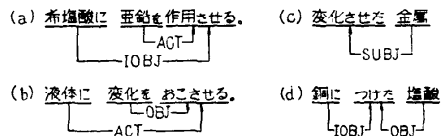


Fig. 13 Examples of Analysis of Causative Sentences

5.4 格構造の選択

1つの動詞文(単文)の解析が終了すると、複数個の処理結果が生じるが、これを Fig. 14 の様な評価関数により優先順位を決定し、優先度の高い処理結果を採用して次の文へと処理を進める。

$$f(CFN, C1, C2, C3) = \frac{6 \times C1 + 2 \times C3 + C2}{CFN} + \frac{C2}{2}$$

CFN = 動詞が要求する必須格の個数

C1 = 文中から見出された必須格要素の個数

C2 = 文中から得られた自由格要素の個数

C3 = 論文(参考文献 13)に述べる文脈の参照によって補足された必須格要素の個数

Fig. 14 Evaluation Function for Selection of Case Frames

また最初選択された解析結果が、更に処理を進めた段階で不都合であると判断された場合には、その解析結果は放棄されて、次に高い評価値をもつ解析結果が採用されることになる。

6. むすび

格文法の導入、名詞・動詞の意味記述により能力の高い解析プログラムが構成できた。統語処理は PLATON で約 110 の規則、意味処理は LISP で約 1,350 行のプログラムで実現されている。名詞辞書・動詞辞書は各約 500 語、200 語である。

本システムの特徴を整理すると、

- (1) PLATON により統語処理と意味処理とが緊密に結びつけられ、しかも規則の変更が容易である。
- (2) 単文処理は統語的、意味的に可能な複数個の係り受け関係を出すことができる。
- (3) 名詞句処理は単一の処理結果を出すのが、バックトラッキング機能により、その処理結果が不適格と判断された時には、再度解析を途中から再開することができ、別の係り受け関係を得ることができる。
- (4) 並列句、長い名詞句、埋込み文等の処理は、統語的情報が少なく、意味記述を用いることにより初めて可能になった。
- (5) 従来の A・I アプローチよりもはるかに多い単語数を対象にしている。

本システムによる解析の成功例、失敗例および実行時間等は別論文にゆずるが、本論文で述べた単文処理、名詞句処理の問題点を次に挙げる。

- (1) 複数個の単文の解析結果から最適なものを選び出す操作、名詞句の解析結果を不適切であると判断する操作は、より広い文脈との意味的整合を考慮して行う必要がある。このためには、文章によって表現された対象世界の状態記述が必要であろう。
- (2) 否定形、仮定形の文を解析するためには、意味記述と矛盾する係り受け関係も許容する必要がある。
- (3) 意味処理は直接 LISP で書かれているため、システムの解読が困難になり、変更・修正がむずかしくなっている。意味情報の記述、参照を直接的に表現できるプログラミングシステムを

構成し、これと PLATON との結合を行う必要があろう。

- (4) 対象分野を広げた場合、意味辞書の構成をどの程度変更しなければならないかを検討する必要がある。

本システムは、質問応答システム (Q. A. S) の入力文処理に使用する予定であるが、Q. A. S の持つ演繹・推論の能力は特に (1) の問題点の解決に有効であろうと考えている。本研究の一部は文部省科学研究費補助金によった。

参考文献

- 1) Woods, W. A.: 'Procedural Semantics for a Question Answering Machine,' (FJCC, 1968), pp. 457-471
- 2) Winograd, T.: 'Procedures as a Representation for Data in a Computer Program for Understanding Natural Language,' (MAC TR-84, 1971)
- 3) Schank, R. C.: 'Margie: Memory, Analysis, Response, Generation, and Inference on English,' (Proc. 3rd IJCAI, 1973), pp. 255-261
- 4) Wilks, Y.: 'Understanding without Proofs,' (Proc. 3rd IJCAI, 1973), pp. 270-277
- 5) 栗原俊彦:「自然言語の機械処理」, 情報処理, vol. 14, No. 4, pp. 267-281, (1973).
- 6) 吉田将:「二文節間の係り受けを基礎とした構文分析」, 信学会, Vol. 55-D, No. 4, pp. 238-244, (1972).
- 7) 田中一敏:「意味情報・文脈情報を用いた日本語の構文分析」, (京大修士論文, 1975)
- 8) 長尾真, 辻井潤一, 田中一敏:「名詞の意味辞書とそれを用いた名詞句処理」, 情報処理学会第15回大会, (1974).
- 9) 長尾, 辻井, 田中:「名詞句に関する文脈処理」(信学会全大, 1974), pp. 258-259
- 10) 長尾, 辻井:「自然言語処理のためのプログラミング言語 PLATON」, 情報処理, Vol. 15, No. 9, pp. 654-661, (1974).
- 11) Fillmore, C.: 'The Case for Case,' in Universals in Linguistic Theory (Ed: Bach and Harms, 1968)
- 12) 村木, 堀江:「動詞の結合価・その1, 基本動詞の結合価」, (LDP 13, 国立国語研究所, 1974)
- 13) 長尾, 辻井, 田中:「意味および文脈情報を用いた日本語文の解析一文脈を考慮した処理」, 情報処理, Vol. 17, No. 1, pp. 19-28(1976).

(昭和50年4月28日受付)

(昭和50年7月4日再受付)