

教詞を含む日本語文の意味処理に関する一考察

田 中 穂 積 (電子技術総合研究所)

§1. はじめに.

中学の理科の教科書文を対象にした言語理解システムを実現するために、筆者等は、文を構文解析し意味構造を抽出するプログラムを作成している。これはEXPLUSとよばれ、その一部は電総研LISP上で動作している。(2)

中学の理科の教科書文の特徴の一つは、数詞を含む文*が比較的多数みられることである。このような理科の教科書文にあらわれる数詞には、定義の明確な物理単位が含まれているため、文中の数詞の意味をプログラムで機械的に確定することは、比較的容易のように思われる。しかし、我々人間にとって単純とも思える数詞の意味確定が、プログラムの容易に行なえるとは必ずしもいえない。数詞の意味を詳細に検討すれば、文中の数詞の意味が、文脈や常識を考慮して決まる場合があるからである。

§2では、中学の理科の教科書文にあらわれる物理単位について、それらの興味ある現象を取り上げて説明する。

§3では、数詞を含む文には、意味的な同一性が自明であるような多数の“いいまわし”が存在することを述べる。そして、意味的な同一性が自明であるような多数の“いいまわし”から、同一の意味構造を抽出することの重要性と問題点を指摘する。§4では、EXPLUSによる意味構造抽出の概略を説明する。

(詳細は、別稿にゆずる。)

最後に、同様な問題に対する言語学者の見解を批判的に検討し、数量詞に関してEXPLUSで未解決の問題を論じる。

§2. 物理単位と数詞の意味

物理単位の意味と、それを含む数詞の意味との間には、強い相関関係がある。相関の程度に応じて、物理単位をつぎの3通りに分類することができる。

1. …… 物理単位の意味により、数詞の意味が一意に決定される。
2. …… 物理単位の意味に多義性があり、そのため数詞の意味が一意に決定できない。(このような場合には、数詞をとりまく文環境(文脈, 常識)を用いて、数詞の意味が決定される。)
3. …… 物理単位の意味に多義性はないが、数詞をとりまく文環境により、数詞の意味が、物理単位のもとの意味から変質する。

1., 2., 3. の物理単位をそれぞれ A類, B類, C類とよぶことにする。主として、中学の理科の教科書で扱われる物理単位をこれらの3種の類に分類して、論じることとする。

* 以下では、数もしくは変数に物理単位が付いたものを数詞とよぶ。たとえば、「1 cm」, 「y Kg」などである。不定の限量数詞、たとえば「スベテ」などは扱わない。

§2.1 A類の物理単位

A類の物理単位は、数詞の意味を一意に決める。物理単位の多くは、このような性質をもっている。たとえば、「1 cd ノ ロウソク」という名詞句に含まれている数詞「1 cd」の意味は、物理単位 cd が光度の強さをあらわすことから、光度の強さをあらわす。中学の理科の教科書文にみられる物理単位のうち、A類の物理単位をまとめると、表Aのようになる。

表1 A類の物理単位

物理単位	意味	物理単位	意味
cm/s, m/s, km/h ...	速度	J, W・s, W・h kg重・m ...	仕事
cm/s ² , m/s ² ...	加速度	PS, HP	馬力
mg, g, kg ...	質量	cal/°C	熱容量
g/cm ³ ...	密度	cal/g・°C	比熱
kg重/cm ² , mmHg ...	圧力	°C, °F, K	温度
atm, bar	気圧, 圧力	rpm	回転数
mA, A, ...	電流	cd	光度
Ω, MΩ, ...	抵抗	lx	照度
ms, s, h sec.	時間	Hz	周波数 (振動数)

§2.2 B類の物理単位

B類の物理単位は、それ自身で、すでに多義性をもっている。このような物理単位は、そう多くない。B類の物理単位を説明するために、つぎの名詞句を考えよう。

- | | |
|--------------|----------------|
| B1. 45度ノ水 | B4. 壁ニ対シテ45度の棒 |
| B2. 45度ノウォッカ | B5. 温度ガ45度ノ棒 |
| B3. 45度ノ棒 | B6. 45度ノ失敗 |

以上には、どれも「45度」という数詞が含まれている。これは、この数詞に含まれている物理単位「度」のもつ多義性による。

まず、文B1では「45度」は温度をあらわす。(理科の教科書では、温度をあらわす物理単位として°C, °F, Kを用いる(表A)。しかし、日常の会話

では、B1のような使われ方をすることがある。) そして、B1の数詞「45度」が温度を意味することは、数詞に後続する語、すなわち「水」の意味的性質によるものと考えられる。

B2にあらわれる数詞「45度」は、筆者の常識によれば、ウォッカに含まれるアルコールの濃度をあらわす。この場合、注意しなければならないことは、B2の数詞の意味が、後続語「ウォッカ」の他に、数詞の数値的な大きさに依存して決まるということである。筆者は、ウォッカのアルコール濃度が約45度であることを知っている。そのような人にとっては、「45度」が温度をあらわすとは考えにくい。それを知らない人は、「45度」に、アルコールの濃度とは別の意味を手えるかもしれない。逆に、ウォッカのアルコール濃度が、約45度であることを知っている人が、「3度ノウォッカ」という名詞句を手えられれば、そこに含まれている数詞「3度」の意味が、アルコールの濃度をあらわすとは考えにくくなるかもしれない。このように、文脈や常識を用いて数詞の意味が決定される場合がある。

つぎのB3は曖昧である。しかし、つぎのことは分かる。B3の数詞「45度」は、アルコールの濃度や回数を意味するのではなく、温度か角度を意味している。これは、後続語の「棒」の意味的性質による。

B4の数詞「45度」は角度をあらわす。B4では、数詞に先行する語句、すなわち「壁ニ対シテ」によって、数詞の意味が決まる。そして、B1、B2と同様にB4では数詞の意味を直接指す語(角度)がないにもかかわらず、(間接的に)数詞の意味が決定できる。

これに対しB5では、数詞に先行する語「温度」によって、直接、数詞の意味が決定される。B6は、後続語「失敗」の意味的性質から、この数詞は、回数をあらわしていることが分かる。

このように、B1からB6までの数詞「45度」の意味は、数詞に先行、もしくは後続する語(句)の意味に応じて、直接ないしは間接的に決定される。B2にみたように、数詞の意味決定に、常識が必要になることもある。これらは、数詞の意味を決めるプログラムで考えておかなければならない問題である。

「度」の他にB類の物理単位として「秒(″)」、「分(′)」がある。

- | | | |
|------|-------------------------|-----------|
| B7. | 北緯 135度(°) 32分(′) 5秒(″) | (角度) |
| B8. | 10時 25分(′) 30秒(″) | (時刻) |
| B9. | 10時間(時) 25分(′) 30秒(″) | (時間) |
| B10. | 10分の1 | (割合) |

§2.3 C類の物理単位

C類の物理単位は、B類と異なり、それ自身に意味的多義性はないが、数詞をとりまく文環境により、数詞の意味が、物理単位のもとの意味から微妙に変質する。つぎの名詞句を考えてみよう。

- | | | |
|-----|---------------|------------|
| C1. | 2000 cc / 水 | (体積) |
| C2. | 2000 cc / ビン | (内容積) |
| C3. | 2000 cc / 自動車 | (排気量) |

C1からC3には、いずれも「2000cc」という数詞が含まれている。この数詞中の物理単位「cc」は体積をあらわしており、B類の物理単位「度」のような意味的多義性はない。しかし、C2、C3では、数詞の意味が体積から内容積、排気量にそれぞれ変質している。この理由を考えてみる。

これまで取り上げてきた<数詞>+<ノ>+<名詞>のような名詞句の場合<数詞>は、普通、<名詞>の数量をあらわしている。たとえばこれは、B1、B2、B4、B5、B6、C1を見れば理解できるだろう。ところがC2とC3では事情がやや異なる。すなわち数詞「2000cc」は、それぞれ最後の名詞「ビン」と「自動車」の体積を意味しない。C2では「ビン」の内容積を、またC3では、自動車の（エンジンの）排気量をあらわしている。

以上の観察から、C2、C3の数詞「2000cc」は、<名詞>の内生素性的な“もの”の存在を仮定して、その“もの”の体積が「2000cc」であると解釈すべきことが分かる。C2の場合の「2000cc」は、「ビン」そのものの体積を意味するのではなく、「ビン」に入る内容物の最大体積をあらわしている。「ビン」という語とともに、内容物という“もの”の存在が仮定されているのである*。C3も同様に、「自動車」という語とともに、エンジンという“もの”の存在が仮定されている。C3の場合には、エンジンから更に、シリンダが仮定され、その内容積の総和が「2000cc」とであると解釈できよう。

このような場合には、実際に、何の体積であるかをはっきり示すために、(内)容積や排気量などという特別の語が用いられる。C類の物理単位は比較的多い。(詳細な検討を進めれば、A類の物理単位のうちのいくつかを、C類に所属させなければならなくなるかもしれない。) なお、ccと cm^3 はともに体積をあらわし、置換可能のはずである。C1、C2については、ccを cm^3 で置換しても意味的な変化はない。ところが、C3では奇妙な文になる。(これは、我が国の慣習によるものなのだろうか。) 体積に関係する物理単位の他に、C類に所属する物理単位を以下に列挙する。

— m^2 , cm^2 ... など、面積をあらわす物理単位

- C4. 20 cm^2 / 円 (面積)
- C5. 表面積が 20 cm^2 / 球 (表面積)
- C6. 断面積が 20 cm^2 / 棒 (断面積)

— m , cm , mm ... など長さをあらわす物理単位

- C7. 20 cm / 細イ糸 (長さ)
- C8. 20 cm / 円 (?)
- C9. 半径 20 cm / 円 (半径)
- C10. 焦点距離 2 cm / 凸レンズ (焦点距離)
- C11. 時速 10 Km (時速)
- C12. 光源カラ 10 cm / 衝立 (距離)

この種のものは、その他に思いつくまま列挙しても、縦、横、高さ、巾、波長、

* B2でも同様な解釈が可能である。B2では、アルコールの存在が仮定されているからである。

波高, など多数ある。C12では数詞「10 cm」が, 2つの基準点(光源と衡立の位置)間の距離の意味を強くあらわしている。

— mV, V, ... などの物理単位

C13. 1.5 V ノ 乾電池 (起電力)

C14. AB 間ノ 電位差 10V (電位差)

— Y(年), d(日), h(時間)

C15. 10h ノ 労働 (時間)

C16. 今カラ 10h 前 (相対時点)

C17. 1978 年 (絶対時点)

§3. 数詞を含む文のいいまわし

数詞を含む文には, 意味的な同一性が自明な多数の言いまわし(paraphrase)が存在する。このような文の意味処理を行なうプログラムは, それら多数の言いまわしから, 同一の意味構造を抽出すべきである。「1 cd ノ 電球」という名詞句が, 少なくとも何通りの言いまわしをもつか列挙してみると, 図1のように12通りある。ここで

()内は省略可能を, また { }内からは, いずれか1つが選択されるものとする。図1はさらに, 「光度」の代りに「光度ノ強サ」をも許すとなれば, 23通りの言いまわしが存在することになる。このようなことを考えると, 「1 cd ノ 電球ノ 前 30 cm ノ 所ニ 半径 1 cm ノ 穴ノ 空イタ 衡立ヲ 置く」のように, 数詞をいくつか含む文において, 意味的な同一性が自明な

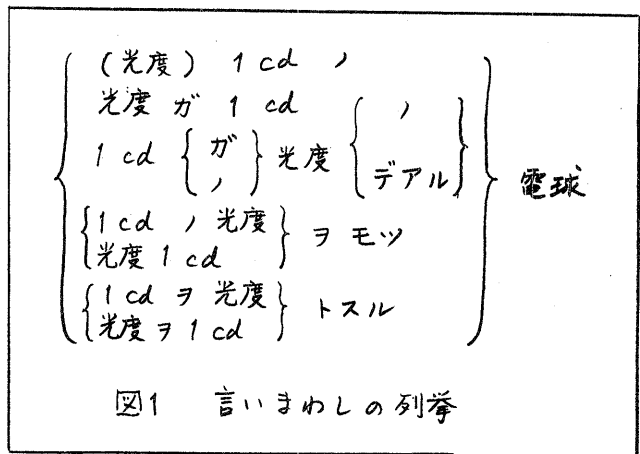


図1 言いまわしの列挙

言いまわしの数は, 軽く1万通りを越える。

これとは別に, 言語学者は, 数詞を含む文に, つぎのような興味ある現象⁽³⁾⁽⁴⁾を指摘している。これは, 数量詞後置変形とよばれている。

D1. 2000cc ノ 水ヲ 買ウ

D2. 水ヲ 2000cc 買ウ

D3. 2000cc ノ ビンヲ 買ウ

D4. * ビン ヲ 2000cc 買ウ

D5. 2000cc ノ 自動車ヲ 買ウ

D6. * 自動車 ヲ 2000cc 買ウ

D7. 二本 ノ ビンヲ 買ウ

D8. ビン ヲ 二本 買ウ

D9. 二台 ノ 自動車ヲ 買ウ

D10. 自動車 ヲ 二台 買ウ

D11. * 二本 ノ 水ヲ 買ウ

D12. * 水 ヲ 二本 買ウ

D1-D2, D7-D8, D9-D10 は, 数詞の後置が可能なる例である。これに対し D3-D4, D5-D6 では数詞を後置することができない。数詞を含む文の意味

処理を行なうプログラムは、これらを正しく処理できなくてはならない。

§4. 意味抽出プログラム EXPLUS⁽²⁾の概要

これまでの説明から、人間にとって自明ともいえる数詞の意味確定にも（仔細に検討すると）さまざまな問題が含まれていることが理解できたのではないかと思われる。§4では、我々の意味抽出プログラム EXPLUS が、どのようにしてそれらの問題を解決しているか、その概略を述べる。

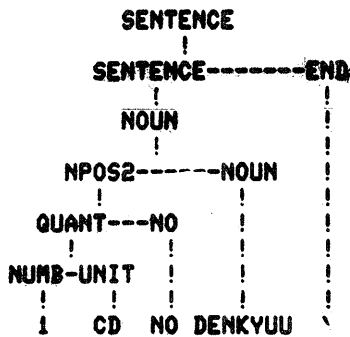
EXPLUS は、我々の開発した拡張 LINGOL とよばれるパーザ⁽¹⁾上に作成されている。拡張 LINGOL は、パーズング・システムのベースとして、Pratt の作成した LINGOL を用い、これに幾つかの機能拡張を施したものである。機能拡張の主なものは、

- (i) 分かち書きを自動化するプログラムが組込まれている。これは形態素の分析（活用、音便処理など）にも応用することができる。
- (ii) パーズング過程で、文法規則から得られる予測情報を利用したり、制御することができる。これにより、文法規則数を増大させずに、文法の精密化が可能となる。また、 unnecessary 部分パーズング木の生成を阻止し、パーズングを高速化することができる。（パーズング速度 = 40 msec / 語が実現されている。）予測情報として、必要なら意味的な情報をも、パーズング過程に反映させることができる。
- (iii) 新しいユーティリティ・プログラムが付加されている。これにより、自然言語処理の基礎となる文法や辞書の作成を会話的に行なうことができる。

EXPLUS は、拡張 LINGOL に埋め込まれた日本語文法を用いて日本語文を構文解析し、構文解析木を得る。ただし、その後の意味処理過程が仮定されているので、構文解析木は、深層構造ないしは意味構造を忠実に反映していなくてもよいとする。つぎに、この構文解析木をベースにして、意味構造が抽出される。

格文法の考え方をを用いて文の意味構造を抽出する場合、動詞などに含まれる格フレームを中心にして意味抽出作業が進行する。したがって、格フレームのような意味抽出の鍵となる情報をできるだけ早い機会に取り出して利用可能な状態にしておくことが望まれる。日本語では、動詞は文末に位置することが多い。そこで文末の動詞を中心に意味構造を抽出する方法が考えられる。しかし助動詞が複雑に相互承接した文や、体言止めの文などでは、文末に動詞を認識することができない。このような不都合さは、あらかじめ、文法規則を用いて文をパーズングし、構文解析木を抽出しておくことにより避けることができる。構文解析木から、文中の動詞や名詞句の所在を確認することができるからである。

一方、後述するように、構文解析木の形は、どの語とどの語を、どのように関連させ、その結果をどのように統合化して、意味構造に写像してゆくかを定める。換言すれば、構文解析木を、意味解釈を行なうプログラム木として見なすことができる。これを、図2を用いて説明する。（図中、本質的でないとと思われる部分は……として省略されている。）



a. 構文解析木

Grammar:

```
[SENTENCE (SENTENCE END) (...) <S-expression1>]
[SENTENCE NOUN (...) <S-expression2>]
[NOUN (NPOS2 NOUN) (...) <S-expression3>]
[NPOS2 (QUANT NO) (...) <S-expression4>]
[QUANT (NUMB UNIT) (...) <S-expression5>]
.....
```

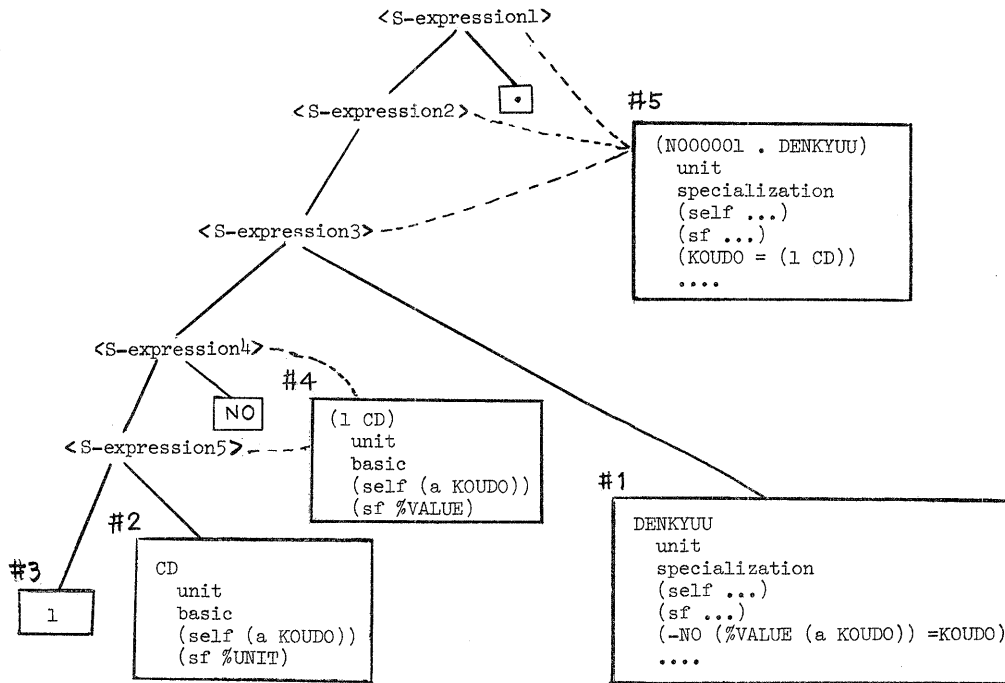
Dictionary:

```
[DENKYOU NOUN (...)]
'(DENKYOU
  unit
  specialization
  (self ...)
  (sf ...))
(-NO (%VALUE (a KOUDO)) =KOUDO)
.....
```

```
[CD UNIT (...)]
```

```
'(CD
  unit
  basic
  (self (a KOUDO))
  (sf %UNIT))
.....
```

b. 文法規則と辞書項目



c. プログラム木

図2. EXPLUSによる意味構造抽出
(入力文 = "1 cd) 電球")

最も単純なA類の物理単位を含む文「1cdノ電球」から、どのようにして（EXPLUSによって）意味構造が抽出されるか見てみよう。「1cdノ電球」は最初構文解析され、図2-aのような木が作られる。図2-aを作り出す時に用いた文法規則と辞書項目の一部が図2-bに示されている。各文法規則と辞書項目のCADDR部（<Sem>部）は、意味解釈を行なうプログラムであり、その具体的な記述はユーザに開放されている。構文解析木を構成する文法規則にはこのような意味解釈を行なうプログラムが付随しているから、図2-aの構文解析木には図2-cのようなプログラム木が対応していることになる。プログラム木の最上部のプログラム <S-expression1> を評価することにより意味解釈が開始される。

図2-cの各ノードでは、直下のノードのプログラムを評価して、評価結果を知ることができる。分枝があるときに、右直下のノードのプログラムの評価を左直下のノードのプログラムの評価に優先させれば、右直下のノードを先にたどって情報を持ち込むことができる。このようにして、意味抽出の鍵となる語のもつ情報をできるだけ早い機会に取り出すことができる。ノードの各プログラムには必ず一つの文法規則が対応しているので、それを見ながら、どちらのノードの情報を先に取り出すかを定めることができる。さらにプログラム木の構造は、各プログラムの局所変数のscopeと直接関係する。これを利用して下位のノードのプログラムから、上位ノードのプログラムの変数（情報）を参照することができる。具体的なプログラムの動作を、ざっと追跡してみる。

図2-cの#1, #2のフレームは図2-bの辞書項目の<sem>部に対応している。各ノードの評価が進み<S-expression5>の評価が行なわれていくものとしよう。このプログラムでは、#2と#3のフレーム（unit）を干渉させる。そして、#4のようなunitを作り出す。EXPLUSでは、これをユニット 干渉とよんでいる。#2と#3の干渉により、ユニット名がCD → (1CD)に、意味素性スロット of (semantic Feature) に書かれた意味素性が%UNIT → %VALUEに変化していることがわかる（#4）。#4は、<S-expression3>まで持たせられて、右のノードの#1と2度目の干渉を行なう。

各unitは、複数個のスロットから構成されている。はじめのスロットはユニット名である。（例：#1のDENKYUU.）第2, 第3のスロットは、重要ではないので説明しない。第4のスロット selfには、上位概念が記述されている。#4では、(1CD)が光度である(a KOU DO)ことが記述されている。第4のスロット selfには、意味素性が記述される。#4では、(1CD)は、値(%VALUE)であるとされている。第5スロット以下は、そのスロットを満たすことのできるunit（これをフィラーとよぶ）の意味記述を行なう。たとえば#5のスロット

(-NO (%VALUE (a KOU DO)) = KOU DO)

は、フィラーとなりうるunitには、格助詞「ノ」が付随し、フィラーの意味的性質が%VALUE & (a KOU DO) でなければならない、ということを書述べている。これは、このスロットを満たし得る条件（premise）とみることができ。最後の=KOU DOは、このスロットが満たされた時に起動されるアクション（action）名である。見方を変えれば、第5スロット以下には、premise — actionの対、すなわち、プログラム・システムで用いられる

ルールが書かれている。さて、〈S-expression 3〉では、#4を、#1のフィルターとして干渉させる。その結果、#4は#1のロット(-NO (%VALUE (a KOUDO)) =KOUDO)を満たすことが分かる。=KOUDOというactionは書き換え規則の名称になっており、現ロットを(KOUDO = (10 CD))に書き換える(#5)。このようにして、最終的に#4が〈S-expression 1〉の値として返され、それとともに意味構造の抽出が終了する。C類の物理単位を含む数詞の意味確定は、#2のsellロットに(a KOUDO)というプロトタイプ記述があるのを利用して、#1と#4の干渉で、数詞の意味が光度であるとか一意に決定することができた。B類、C類の数詞の意味を決定するためには、数詞の前後の語の意味に応じて、新しいロットを付加したり、削除したり、変形する複雑なunit変形が繰り返される。5番目以下のロット記述では、premiseとして、意味素性と、小文字のaをもとに幾つかのプロトタイプ記述(たとえば(a KOUDO))との、任意の論理的組合せが許されている。特にpremise中の(@REQUIRE <S-expression>)という記述によって、任意の<S-expression>を実行して、さらに詳細な意味処理ができる。これを用いれば、#1で、(1 cd)が、電球の光度として常識的な大きさであるか否かをチェックすることもできる。

Winogradの用語に従えば、(@REQUIRE <S-expression>)はTo-fill methodに、またactionは、when-filled methodに対応している。(5)

EXPLUSでは、これらのmethodとプロダクション・システムの考え方を利用して、意味抽出過程を制御し、図1の言い廻しから、同一の意味構造を抽出している。

§5 数量詞後置変形

文D1から文D2を派生する変形は、言語学者により数量詞後置変形とよばれている。数量詞後置変形については、神尾、井上が詳しく論じている。⁽³⁾⁽⁴⁾前者は、この変形を認める立場に、また後者はそれを批判する立場に立っている。彼等は、D1からD12以上に複雑な多数の文をも対象にしているが、本稿では、この範囲の文に限って検討する。神尾は、D5とD1の対応関係をつぎのように説明する。D5には図3の、D1には図4のような構造を対応させる。

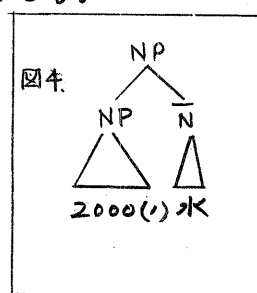
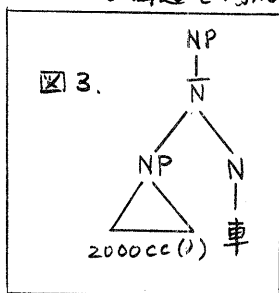
Nに埋め込まれた数詞は形容詞などと同じ修飾語句であり、「二台」は限定詞でNPに直接支配されている。

そして、移動できる数詞はNPに直接支配される、限定詞としての数詞であるとする。

この考え方の要点は、Nという技巧的なノードを設けて、それに支配される数詞は後置することができない、と

することである。しかしながらこれは、意味構造抽出の観点から検討すると、あまり有用な考え方とはいえない。木構造の差違を前提にした議論だからである。図3、4の2つの「2000cc」という数量詞の片方がなぜ形容詞などと同様な修飾語句なのかの説明されていないからである。しかも、数量詞の意味が、それに先行もしくは後続する語の意味の影響を(直接的ないしは間接的に)受ける(S

* 我々にとって役立ったとすれば、言語学者によるそのような説明である。そのような説明は、数量詞後置変形規則をも説明することになると思われる。



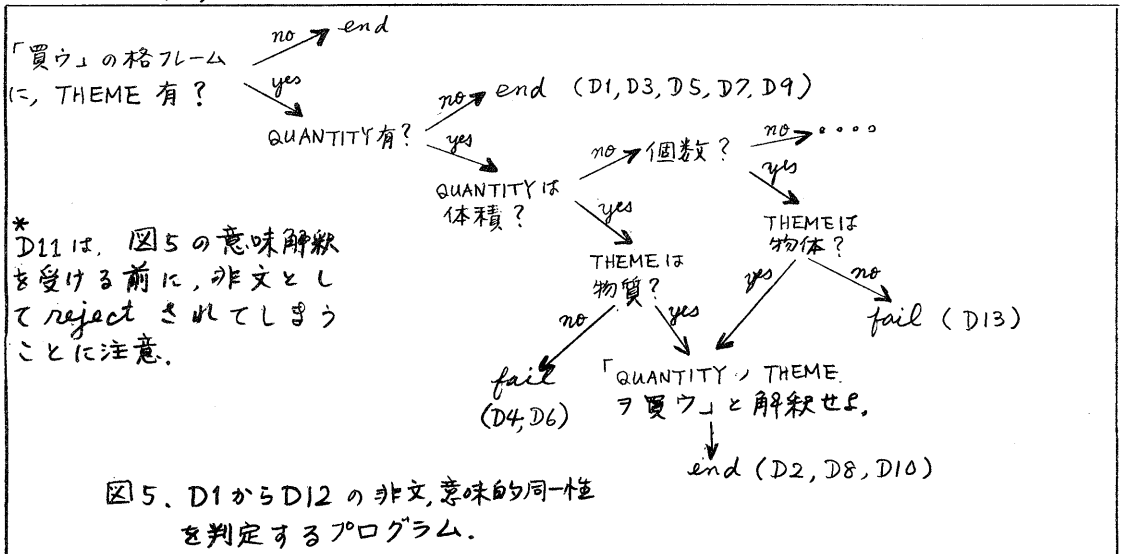
2) ことが論じられていない。したがって、つぎの事実を説明できない。

E1. 2000ccノ水ヲ増ヤス

E2. 水ヲ2000cc増ヤス

E1, E2は正しい文であるが、意味が異なる。E1はE2のパラフレーズではない。この事実は、D1, D2と良い対照をなしているとともに、数量詞後置変形が、動詞の影響をも受けることを示唆しているように思われる。

EXPLUSでは、神尾のような、図3や図4の2つの構造を考えない。全く同一の構文解析木を作り、数詞に先行する名詞や、後続する名詞や動詞の意味を考慮しながら意味を解釈し、意味構造を抽出する。たとえばD1からD12までの文は、「買う」という動詞に埋め込まれた意味解釈プログラムによって非文や意味的同一性の判定を正しく行なうことができる。EXPLUSでは、文D1からD12の二重下線と破線を施した名詞のそれぞれが、THEMEとQUANTITYであることを認識してから、非文の発見や、意味的同一性を判定するために、図5の意味解釈プログラムを動作させる。



なを、図5の手続きは、文生成用の意味解釈プログラムとしても利用可能であろう。

§6 おわりに EXPLUSでは、本稿で考察したような数(量)詞に関する問題はほとんど解決できると考えている。ただし、量(数)詞に係る範囲の問題が残されている。例えば、F1. 100円ノ野菜ヲ買う。 F2. 野菜ヲ100円買う。 F3. 100頁ノ本ヲ読む。 F4. 本ヲ100頁読む。 不定の数詞や、等位接続と数詞の問題(AトBヲ10コ買う)など未着手の問題も多い。

参考文献

(1) 田中, 佐藤, 元吉: "自然言語処理のためのプログラミングシステム—拡張LINGOLについて", 信学論誌, 60-D, No.12, 1977
 (2) 田中: "日本語の意味構造を抽出するシステムEXPLUSについて", 信学論誌 掲載予定
 (3) 井上: "日本語の文法規則", 大修館, 1978
 (4) 神尾: "数量詞のシンタックス", 「言語」6: 8, 大修館, 1977.
 (5) Winitz: "Frame Representations and The Declarative-Procedural Controversy" in Bobrow & Collins (Eds.): "Representation and Understanding", Academic Press, 1975. (訳: 監訳: "人工知能の基礎", 近代科学社, 1978, 8月出版予定)