

文章解析のアルゴリズム化への試み (続編)

石綿敏雄 (茨城大学)

ここで文章というのは英語の *discourse* に当る概念である。*discourse* には最近「談話」という訳語が当てられることが少なくない。したがって文章と談話は同じ事柄を表わしているが、既に指摘されているように、談話は話しことばにかたより、文章は書きことばにかたよることは否めない。そこで訳語を当てずに英語の *discourse* をそのまま使う方がかえって危険が少ない。しかしここであえて文章ということばを使うについては、多少の理由が存する。それは当面の目標が物語の理解ということにあるからであり、書きことばを主要な対象と考えているからである。

この研究は文部省科学研究費特定研究言語のなかの「計算機による日本語談話行動の総合モデル化」(課題番号410202)のための報告である。(続編)とあるが一応のまとめのつもりである。

(続編)のつかない、前回の発表では、文章解析のケース・スタディのあと、問題点を整理し、小シーンと大シーンの関係を考え、小シーンに「言語表現」が対応すると考えた。今回は大シーンと物語の構造の関係を考え、「言語表現」と動詞の意味記述の問題を考え、最後に今回組み立てたアルゴリズムを試験的に運用してみたいと思う。

1. 物語の分類と構造

前回は小シーンと、小シーンを組み立てて成る大シーンの関係を考えた。今回はそれを含めて、物語の分類と構造上の差に注目し、それと大シーンとの連続についても考えてみたい。

前回で取りあげた「暴動」という大シーンでは、それに「警官隊が走る」「青年たちが走る」というような小シーンの束が対応し、一つ一つの小シーンには「警官隊が走る」「警官が青年を追いかける」などの言語表現が対応していることも考えた。これらの小シーンは暴動の大シーンに属することはもちろんであるが、そのうちのいくつかは全く他のシーンの一部にもなりうる。たとえば青年たちが走るというのは運動会のシーンでもあり、尾車に乗り遅れまいと努力するシーンでもありうる。このような関係をもう少し一般的に考えてみることにしよう。いま大シーン P, Q, R があるとし、 P は小シーン p, q, r を含むものとする。 Q は小シーン q, t, u を含み、 R は小シーン r, q, u を含むとする。このような包含関係、対応関係のなかには大シーンにとってオブリガトリーな小シーンと、オプショナルな小シーンが含まれるであろう。また順序のあるものと、そうでないものがあると考えられる。

ここでいう小シーンとは生活のなかの最も小さな単位のようなものとも考えられよう。たとえば字を書くとか、ビールを飲むとか、歩くとか、雨が降るとか。これらはいくつかの動作の組み合わせであることもあり、一つの動作であることもある。これらの一つ一つの動作(状態、……その他一つの述語であらわされるもの)は一つ一つの動詞が対応している。たとえば字を書くについては

筆をとる すみをする ます目を埋める

などがある。

一つの動作 act が一つの言語表現に対応することもあれば、いくつかの言語表現に対応することもある。前号で述べたように、一つの「意味」



は「男が女を愛する」「男が女を恋する」「男が女にほれる」……などいくつかの表現となつて言語として実体化するからである。同時にこれらの表現と意味と小シーンと大シーンの関係を上のように考えておく。

さてこれらの表現にはそれぞれ特有の統語基本構造がある。これらの成分と述語の意味関係を「格」と考えることができよう。つまり「格」あるいはフィルモアの所属格より更に深いところに「意味」の一つの世界を考えるのである。

小シーンは一つの文またはいくつかの文によって言語として実現する。大シーンはこれらのさらに大きなまとまりとなる。大シーンの例をあげれば、店で買物をする。郵便局で小包を送る。銀行に行く。レストランで食事する。汽車に乗るなど。短編小説のなかには一つの大シーンあるいはいくつかの大シーンで終りになるものがある。しかしやや長い小説では、異なる大シーンをいくつかならべたものを一つの単位とすることもあるだろう。「章」ということもあるが、外来語にそろえてブロックといつてみる。たとえば、漱石の「三四郎」の第一ブロックは漢字の「一」とあるのがこれである。「三四郎」の「一」ではまず車中の大シーンが現われ、「じいさん」が降りたあと「女」と名を屋まで行き、そこで下車して同じ宿にとまる、といういくつかの大シーンが続く。翌日また汽車に乗り、後に広田先生として再登場する「男」と話しをする大シーンが続く。こうして「一」が終り、一ブロックが終わるのである。つまり「一」では、三四郎の上京、女と同宿およびじいさんに見られる三四郎の性根、「男」との出会いなどの内容が盛り込まれている。

一つのブロックはいくつかの大シーンから成り立つ。「上京」というブロックは「汽車にのる」「車中の出来事」「目的地東京下車」などの大シーンからなりたつ。これを P, Q, R であらわせば、上京 R は

$$R \rightarrow P - (Q) - R$$

で表わされる。 Q はあつてもなくてもよい。「三四郎」の場合はいきなり Q で始まっているので、そこに一つの文学的技法があるとみることもできよう。恋愛という大シーンは、たとえば出会い E 、デート D 、愛 L の三大シーンから成ると考えれば

$$L \rightarrow E - D - L$$

平安貴族の恋の形式は、べつ見 P 風聞 H 、和歌の贈または答 W 、結婚 M などの要素からなる。

$$L \rightarrow P / H - W - W - W \dots W - M$$

和歌の贈答の間の事件を X で代表させれば、また初期値を I とすれば

$$L \rightarrow I - P / H - X - W - X - W \dots W - X - M$$

のようになる。初期値は主人公、場所の紹介であり、はじめの文の主語になる。

「伊勢物語」「大和物語」のような歌物語は

$$U = R \rightarrow L - L - L \dots - L$$

のようになろう。

いま読んでいる作品がどのジャンルに属するかについては、このような物語のジャンル分けとその構造上の区別をあらかじめ検討しておくことにより可能となろう。たとえば「軍記物語」については (= B=R)

$BR = I - B - X - B - X - B \dots$

となる。ここで B は合筆描写をあらわす。

紀行文学 T=R は紀行のブロック T を交えて次のようになろう。

$T=R \rightarrow I - T - X - T - X \dots$

もし

$T \rightarrow X - T - H$

で、H が俳句を表わすとすれば、この紀行文は俳諧紀行文（たとえば「奥の細道」のような）である。

さて、いま対象を文学に限定し、作品テキストを T とすれば

$T = \{ T=R, U=R, B=R, \dots \}$

となる。和歌物語を生成しようと思うときは、U=R を選択し、上述の生成のための書き換え規則を適用すればよいことになる。もしいま読んでいるテキストがいかなるカテゴリーに属するかを知らなければ、通常の構文解析と同様に上記の句構造規則と照合すればよい。文学作品には順序の入れ替えや必要な場所の省略という「変形」もある。ブロックや大シーン、小シーン生成にはこの種の「変形規則」が働いているとみることができよう。

次に以上の形式化の事例をあげてみる。

「むかし、おなかわたらひしける人の子ども^①、井のもとにいでて遊びけるを^②、大人になりければ、男も女も恥ぢかはしてありけれど^③、男はこの女をこそ得めと思ふ^④、女はこの男をと思ひつゝ^⑤親のおはすれども聞かでないありける^⑥。さて、このとなりの男のもとより、かくなむ^⑦

筒井つの井筒にかけしまろがたけ過ぎにけらしも妹見ざるまに^⑧
女、返し^⑨

くらべにしふりわけ髪も肩過ぎぬ君ならずして誰かあぐべき^⑩
など言ひ言ひて^⑪つひに本意のごとくあひにけり^⑫-----」（伊勢物語 23段）

以上の例で、①は時と人物の紹介、②は、④のおなかとあわせて場所の紹介であり、上の形式化では I を形成する。③は、特殊ケースであるが P/H に当たる。④~⑥は X、⑦も X（⑦は歌の前部とみてもよい）。8 は W、⑨は⑦と同じ。⑩は W、⑪は X または W の後部分。⑫は M。かくして

$I - P/H - X - W - X - W - M$

となり、L を形成することになる。この種の L の連続であればいま読んでいるテキストは

$U=R$

であることになる。

2. 文章の内容の累積

上述の方法はテキスト構成を句構造的に表現する方法であり、内容のときはテキストの種類判定に役立つと考えられる。これに好してテキストが文の集積であると考へた場合、

$$T = S_1 + S_2 + S_3 + \dots + S_m$$

と表わされる。

いま S_i の意味を SM_i , S_m までの文内容の累積を TM_m で表わせば

$$TM_m = \sum_{i=1}^m SM_i$$

i が 1 のときは S_1 の意味が TM_1 の意味であり, i が 2 のときは S_1 と S_2 の意味の累積がそれぞれテキストを読んだときの TM_2 の内容である。

いま SM の意味表現として dependency grammar の actant の状態, 格文法でいえばフィルモアの K の内容の状態として考えてみる。(SM の表わし方はもっとほかにあると考えられるがここではこのように考えておく)。actant はその動詞が現われることによってその内容を変える。 A, B を actant とすると

A が B を倒した (S_i)

という文が理われると, S_{i-1} まででは B は立っていたのであり, A はこれに (決定的には) $\#$ を加えていなかったのである。つまり

	A	B
TM_{i-1}	# 追加していない	立っていた (+ vertical)
TM_i	# 追加した	倒れている (- vertical)

さらに

A が B を C に売った (S_j) では

	A	C	B
TM_{j-1}	B を所有	B を所有せよ	A の所有
TM_j	B を所有せよ (代り: 金を得)	B を所有 (代り: 金を払)	C の所有

のようになる。動詞についている完了の「た」、現在の「る」、否定の「ない」、推量の「だろう」などは、この actant の内容の表現にかなりの影響を与える。ヨーロッパ語の過去完了、大過去などもこれを明示する。

さて、ここで、さきの伊勢物語23段の内容累積を再現してみよう。

- ① 出てくる actant ぶかし, 「おなかわたらひ」した人 こども
- ② " # (→ #筒)
- ③ " 男, 女 (← こども)
- ④ " " "
- ⑤ " " "
- ⑥ " 親 (= 「おなかわたらひ」した人)
- ⑦ " 男 (女へ) となり
- ⑧ " #筒 たけ まろ (男) いも (女)
- ⑨ " 女 返事
- ⑩ " ふりかけ 髪 肩 君 (男) 誰
- ⑪ " 一 (男, 女)
- ⑫ " 本意

以上のうち「ぶかし」は時間設定, 「おなかわ」「#筒」「となり」は場所を示し, 「たけ」「ふりかけがみ」「返事」「本意」など(誰)も小道具として出てくるが、スペースの関係で省略する。このように、時間、場所、もの... などに分けて全体をくわしく記述すれば、正確な読みが出来よう。

さて、残る 男, 女, 親についてだけ、以下に名文終了後のメモリーの状態を

示すことにする。

	男	女	親(田舎渡世)
S1	子ども	子ども	
S2	井戸のとこで遊ぶ(女といふ)	井戸のとこで遊ぶ(男といふ)	
S3	大人になる, 女に対し取っかいと思う	大人になる, 男に対し取っかいと思う	
S4	女を奪たいと思う	男の恋の対象(-)	
S5	女の恋の対象(-)	男を夫としたいと思う	
S6	親のいう人と結婚したくない	親のいう人と結婚したくない	両方の親が「ほかへ結婚相手を探せ」
S7	女へ恋の歌を贈る	男から母を receive (-)	
S8	恋歌(女とあんなうちに成人)	男の恋の対象(-)	
S9	女から返事をもらう(-)	女が「男へ返事	
S10	女の恋の対象(-)	恋歌(ほかの男と結婚せず)	
S11	互にいう	互にいう	
S12	女と結婚	男と結婚	

内容の累積といってもいままでのことを消してしまうのではなくすべておぼえていなければならぬ。書きかえることも行われるが、新しい情報がはいると前のを消すということが出来るようにするためには、男(女とせば)のなかに構造化する必要があろう。しかし男に限ってこれを行えば、

	男	女の動作	女に対する気持
S1	子ども		なごよし
S2	〃	のど 女と遊ぶ	〃
S3	大人		はずかしい
S4	〃		女を奪いたいと思う、
S5	〃	(女は男を奪いたい)	〃
S6	〃	親のいう人と結婚せず	〃
S7	〃	女に歌を贈る	〃
S8	〃		〃
S9	〃	女からの返事がある	〃
S10	〃		〃
S11	〃	互にいう	〃
S12	〃	女と結婚	

以上のうちで、毎令、気持などは前のメモリーのセットの状態が続くと考えてよからう。男の動作、女の動作、女への気持が互いにどのような関連をもつかは SCRIPT の運用とかサブルーチン化することもできようが、大筋は詩語的に表現される。