

## 概念依存表現からの日本語文の生成

重永 幸 松尾豊文 安在弘幸  
(徳山高専) (九州工業大学)

## 1. まえがき

人間が言語を理解したかどうかは、それに対する質問に答えられるかどうかによって推測できる。同様に計算機で言語理解ができれば、問い合せに対して応答ができるはずである。この場合計算機の内部にはそれらに対する情報(知識)をもっていなければならない。ここでの応答とは文章を生成する事である。

文章生成にはランダムに文を生成する試みと、情報を内部表現から自然言語に変換する試みとがある。前者は文法や文法理論をテストするために行われ、後者は文の「意味」のある種の内部表現を入力として、これを表層構造、すなわち適切な単語列に変換する事である。この種のシステムの例としてはSchankのMARGIE<sup>1)</sup>がある。これは英文を入力して、推論と言い換えを行うシステムである。このうち言い換えを出力する部分は文の生成としてGoldmanが扱っている。

筆者らも内部表現にSchankの概念依存(Conceptual Dependency:CD)表現を用い、ここから日本語文を生成する過程を計算機でシミュレートした。本システムにおいては、CD表現を、LISPのa-listに当る表現(BNFで定義した)に変換し、動詞を決定した後、構文情報を用いて文を生成する。この過程の中で特に主点を置いたところは、係り受けの取扱い、CD表現からの動詞の選択及び述語文節の合成である。

本稿では、2章でCD表現から日本語文を生成する原理について説明を行い、3章でこれに基づいて作成した我々のシステムについて述べる。そして4章で生成された日本語文とその生成過程の例を示す。

## 2. システムの概要

計算機上に何らかの形で、ここではCD表現を用いて蓄えられている知識から文を生成する過程を考える。但し、ここで扱っている知識は相互に何ら関係をもたず、断片的に蓄えられてい

るものとする。

本章では本システムで扱っている日本語の文と、知識を表現するのに用いるCD表現について説明する。次にシステムの設計方針とシステムで参照するために必要な辞書、動詞の定義特性について述べる。

## 2.1 日本語文

本システムで扱う日本語文を次のように規定する。

$\langle \text{文} \rangle ::= \langle \text{格要素の並び} \rangle \langle \text{述語文節} \rangle$

$\langle \text{格要素の並び} \rangle ::= [N_1 \cdot \cdot N_i \cdot \cdot N_n]$

( $N_i$ は格要素)

$\langle \text{格要素} \rangle ::= \langle \text{名詞} \rangle \langle \text{助詞} \rangle | \langle \text{節} \rangle \langle \text{格要素} \rangle$

$\langle \text{節} \rangle ::= \langle \text{形容詞節} \rangle | \langle \text{名詞節} \rangle | \langle \text{副詞節} \rangle$

$\langle \text{形容詞節} \rangle ::= \text{Cas}(が、は)(\text{Mdf}(i)(\text{文}))$

( $i$ は格要素の並びの中でこの節が占める位置を示す)

$\langle \text{名詞節} \rangle ::= \text{Cas}(が、は)(\langle \text{文} \rangle)$

$\langle \text{副詞節} \rangle ::= \text{Cas}(が、は)(\langle \text{文} \rangle)$

$\langle \text{述語文節} \rangle ::= \langle \text{語幹} \rangle [P_1 \cdot \cdot P_i \cdot \cdot P_l]$

( $P_i$ は付属語)

$\langle \text{語幹} \rangle ::= \langle \text{動詞の語幹} \rangle | \langle \text{形容詞の語幹} \rangle |$

$\langle \text{形容動詞の語幹} \rangle$

$P_i ::= \langle \text{可能} \rangle | \langle \text{否定} \rangle | \langle \text{時制} \rangle | \langle \text{疑問} \rangle |$

$\langle \text{叙述形式} \rangle$

なお、 $[X_1 \cdot \cdot X_i \cdot \cdot X_n]$ はその中から任意の要素 $X_i$ ( $1 \leq i \leq n$ )を取って並べる事を示す。但し、それによってできた並びを $X_{j(1)} \cdot \cdot X_{j(k)} \cdot \cdot X_{j(n)}$ ( $1 \leq j(k) \leq n$ )とすると $j(k-1) < j(k)$ でなければならない。

$\text{Cas}(a_1, a_2)$ は助詞 $a_2$ を助詞 $a_1$ に変更する操作を表す。

$\text{Mdf}(i)$ は $[N_1 \cdot \cdot N_{i-1} N_i N_{i+1} \cdot \cdot N_n]$ を $[N_1 \cdot \cdot N_{i-1} N_{i+1} \cdot \cdot N_n]$ に変える操作を表す。

## 2.2 CD表現

CD表現の表現形式を図-1に与える。図式表現としてのCD表現は、計算機に入力される時には、記号列言語表現であるリスト表現で表される。



報であり、動詞を含む概念の種類によって次のように分ける。

- 1～4 カラム：一番外側の概念のとき用いる
- 5～8 カラム：先行概念で#がCONCEPT,ACTORのとき用いる
- 9～12:先行概念で#がOBJECTのとき用いる。

各カラムに対して接続する助詞は1:は、2:に、3:を、5:に、6:を、9:が、10:に、であり、その他のカラムの情報に対しては助詞は接続しない。

図-4に様相度合辞書の構造を示す。MENTS (mental state)やSIZEなどの様相に対して各値に適当と思われる形容詞あるいは形容動詞を対応させる。これらは文のなかで連体修飾関係を表す場合と述語文節を合成する場合に用いる。活用の種類としては連体形が「な」のものを1、「の」を2、「い」を3、その他を4とする。

## 2.4 設計方針

日本語文の生成において、本システムで重点を置いた事項は、係り受けの扱い、CD表現からの動詞の選択および述語文節の合成である。

一般に、日本語には次の性質がある<sup>2)</sup>。(i)同一文中の複数の係り受けは交差しない、(ii)文末にない文節はその右側のいずれかの文節に係る。この性質は、上述のCD表現のリスト表現への変換の際、また、計算機内での処理の際にも利用される。

生成の手順を次のように与える。

- STEP 1 時制概念を抽出する。
- STEP 2 動詞辞書を用いて動詞を選択する。
- STEP 3 構文情報を用いて概念を表す単語間に関係を与える。
- STEP 4 格要素並びを合成する。
- STEP 5 叙述情報により述語文節を合成し、STEP 4までに合成した部分と接続する。
- STEP 1から5までの手続きによって単文の生成を行う。STEP 3では連体修飾関係の処理を行う。

様相度合概念では、生成する形容詞あるいは形容動詞によって語尾変化が異なる。たとえば、活用1:幸せ+な+少女、活用2:灰色+の+空、活用3:悲し+い+少年、および活用4:いろいろ+していた+少年、となるため、どの活用を

するのかという情報は様相度合辞書に蓄えておく。但し活用4の場合には活用は時制によってもかわるので活用形自身が蓄えられる。(図-4参照)

統語上接続する助詞は表-2のようにする。

表-2 CD表現上での位置と助詞

基本格	位置	助詞
動作主体	ACTOR CONCEPT	は
動作・作用の及ぶ対象	OBJECT	を
動作・作用の相手または動作・位置の終点	TO	に
動作・作用・位置の起点	FROM	から

## 2.5 動詞の定義

本システムではCD表現から直接日本語文を生成する方法をとるが、両者の対応を考えた場合に、一番問題となるのはSTEP 2の動詞の選択をどう行うかである。筆者等は、あらかじめ動詞の定義特性をあいまいさがないように注意しながらCD表現で表し、入力したCD表現とそれとのマッチングをとって動詞を選択する方法をとった。たとえば「投げる」という動詞の定義特性は「XがYに力を作用してYの所有をXからZへ変える」と表される。これをCD表現で表すと図-5(a)のようになる。ここで判るように、概念には表層文として現れるものと現れないものがある。

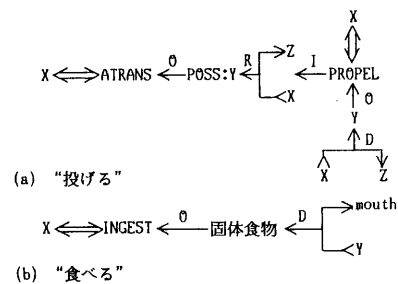


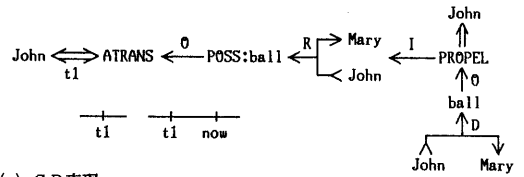
図-5 動詞の定義特性

## 3. システム構成

本章では、前述の設計方針に基づいて実際に作成したシステムについて、その詳細を述べる。3.1でCD表現の内部形式、3.2で生成アルゴリズムについて述べ、3.3で動詞の選択方法について説明する。

key word	CD表現またはそれに対応した意味
ACTOR	行為主
ACT	根元的行為
OBJECT	← <sup>0</sup> 対象
DIRECT-F TO	← <sup>D</sup> →
RECEIP-F TO	← <sup>R</sup> →
INSTR	← <sup>I</sup> →
TYPE PERFORM ATTRIB VARY	概念単位の種類 ↔ ↔ ↔
CONCEPT	概念
RELT	関係
ATRB	様相
VAL	値
TO(VALUE) FROM(VALUE)	← ( ) → ( )
MODE ? C /	法 疑問 可能 否定
TIME TIMERELT BASETIME	事象発生時点 時間関係 基底時間との関係
( )	概念単位
CONJTYPE r R E ROTAT	結果因果関係の種類 ↑↑ <sub>r</sub> ↑↑ <sub>R</sub> ↑↑ <sub>E</sub> ↑↑↓
X	未知の行為主
Y, Z	場所を表す未知概念

図-6 CD表現と属性



(a) CD表現

```
(( TIMERELT t1 BASETIME t1-now ) ( ACTOR John TYPE
perform ACT atans-p OBJECT ball RECEIP-F John TO
Mary INSTR ( ACTOR John TYPE perform ACT propel
OBJECT ball DIRECT-F John TO Mary ) TIME t1 ))
```

(b) 入力形式

図-8 “ジョンはメアリーにボールを投げた”

表-3 時制情報表

NEW TENSE	BASE TENSE	TENSE	NEW TENSE	BASE TENSE	TENSE	NEW TENSE	BASE TENSE	TENSE
左	PAST	PAST	左	PRESENT	PAST	左	FUTURE	PAST
=	PAST	PAST	=	PRESENT	PRESENT	=	FUTURE	PAST
右	PAST	FUTURE	右	PRESENT	FUTURE	右	FUTURE	FUTURE

ドに置きかえられ、さらに図-7に示すBNFに従う入力形式に書きかえられる。キーワードを属性とし、直後に実際の概念が続く。これを属性値と呼び、入力形式は属性と属性値の並びで表

される(これはLISPのa-listに当る)。たとえば「John threw a ball to Mary」という意味をCD表現で表すと図-8(a)のようになり、それは図-8(b)の入力形式に変換される。本システムではこの作業は人間が行っている。

### 3.2 日本語文生成アルゴリズム

本システムは、あらかじめ準備された辞書類と入力であるCD表現、そして時制処理ルーチンと文生成ルーチンよりなる。以下でそれぞれのルーチンについて説明する。

#### 3.2.1 時制処理ルーチン

時制処理ルーチンでは、入力したCD表現から時制概念を抽出し、各事象発生点に時制を与える。この時制情報を得るために基本時点(BASETIME)、基本時制(BASETENSE)、新時点(NEWTIME)および新時制(NEWTENSE)の4つの変数を導入する。その処理アルゴリズムを次に示す。

```
<概念依存表現> ::= ( (<時制概念>) (<概念情報>) )
<時制概念> ::= TIMERELT <timerelt> BASETIME <basetime>
<概念情報> ::= <様相概念> | <様相変化概念> | <事象概念>
<様相概念> ::= <様相概念> | <関係概念>
<様相概念> ::= CONCEPT <埋込み概念> TYPE attrib ATRB <atrb>
VAL <val> {MODE ?} {TIME <time>}
<関係概念> ::= CONCEPT <埋込み概念> TYPE attrib RELT <relt>
OBJECT <parameter> {MODE ?} {TIM <time>}
<様相変化概念> ::= CONCEPT <埋込み概念> TYPE vary ATRB <atrb>
FROM (VALUE <key> <val>) TO (VALUE <val>)
{MODE ?}
<事象概念> ::= ACTOR <埋込み概念> TYPE perform ACT <act>
OBJECT <埋込み概念> {DIRECT-F <埋込み概念> TO
<埋込み概念>} {RECEIP-F <埋込み概念> TO <埋込み概念>}
{INSTR (<事象概念>)} {MODE <mode>} {TIME <time>}
<埋込み概念> ::= <parameter> | <parameter> (<概念情報>)
<act> ::= INGEST | EXPEL | GRASP | MOVE | PROPEL | PTRANS | ATRANS-0 | ATRANS-P
<relt> ::= POSS | PART | LOC-N | LOC-B | LOC-0 | LOC-R | LOC-L
<timerelt> : 時間関係 <val> : 様相概念の値 (+10~-10の数字)
<basetime> : 基底時点 <key> ::= < | >
<atrb> : 様相概念 <mode> ::= C | ? | / (または組み合わせ)
<time> ::= t1 | t2 | t3
<parameter> ::= 英単語 | #
```

図-7 システムで処理するCD表現

### 3.1 CD表現の内部形式

CD表現を本システムへ入力するためには、既に述べたように計算機で処理しやすい入力形式、すなわちリスト表現に変換しなければならない。CD表現で用いられる記号は図-6に示すキーワー

STEP 1 基底時点(NOW)より基本時点に基本時制を与える。

STEP 2 時制情報を求めるべき時制表示点を新時点とする。

STEP 3 新時点の時制(新時制)を時間関係より求める。

STEP 4 この新時制と基本時制の接続により時制情報表(表-3)から時制を求め、新たにこの時制を基本時制とする。

STEP 5 さらに時制を求めるべき時制表示点があればSTEP 3へ返る。

### 3.2.2 概念処理ルーチン

時制概念の処理が終了すると概念情報の処理に移る。図-7で表されるCD表現のうち特に問題となるのは埋め込み概念の処理である。埋め込み概念が<parameter>であれば名詞句の並びとして処理するが、<parameter>(概念情報)の場合、括弧内の概念情報はその前の<parameter>に係るとして係り受けの処理をする。処理は<概念情報>の種類によって次の三つに分けられる。

#### I <概念情報>::=<様相概念>

##### (i) <様相概念>::=<関係概念>

本システムで処理可能な関係概念には次のものがある。括弧内は生成される句を示す。

POSS(XのY) PART(XのY)

LOC-N(Xの近くのY)

LOC-B(Xの下のY)

LOC-O(Xの上のY)

LOC-R(Xの右側のY)

LOC-L(Xの左側のY)

##### (ii) <様相概念>::=<様相度合概念>

品詞によって次のような語尾変化を伴う。

品詞1:語幹+な(幸せな啓子)

品詞2:語幹+の(普通の人)

品詞3:語幹+い(赤い靴)

品詞4:大部分がサ行変格活用をする動詞「する」を接続しているので時制を考慮する。(いらいらしていた少年)

#### II <概念情報>::=<様相変化概念>

様相変化概念から生成される文の述語文節は、形容詞または形容動詞の連用形に、「なる」を接続した形になる。そのため時制を考慮しなけ

ればならない。

#### III <概念情報>::=<事象概念>

この場合、事象概念のACTORまたはCONCEPTの<parameter>に係る部分なのか、OBJECTの<parameter>に係る部分なのかによって処理が分けられる。

##### (i) ACTOR、CONCEPTのとき

たとえば・・・CONCEPT Keiko (ACTOR # TYPE perform ACT ingest OBJECT apple)・・・というCD表現から生成される文は「#はりんごを食べる」→「りんごを食べる」→「りんごを食べる啓子」という変形を受ける。

##### (ii) OBJECTのとき

・・・CONCEPT cake (ACTOR Keiko TYPE perform ACT atrans-o OBJECT # RECEIP-F Keiko TO Tom)・・・の処理手順は次のようになる。

1. 括弧の概念情報を処理する。「啓子はトムに#を与える」
2. Mdf(#)を行う。「啓子はトムに与える」
3. Cas(が、は)を行う。「啓子がトムに与える」
4. cakeを接続する。「啓子がトムに与えるケーキ」

CD表現において叙述情報としては否定、可能、疑問及び時制を扱っている。

### 3.3 動詞の選択

動詞の選択アルゴリズムを次のCD表現(リスト表現)を例にとって説明する。

```
((TIMERELT t1-t2 BASETIME t2-now)(ACTOR John (CONCEPT # TYPE attrib ATRB ments VAL-7 TIME t1) TYPE perform ACT propel OBJECT fist (CONCEPT # TYPE attrib RELT part OBJECT John) DIRECT-F John TO Tom INSTR (ACTOR John TYPE perform ACT move OBJECT fist DIRECT-F John TO Tom) TIME t2))
```

下線(1)で「いらいらしていたジョン」という名詞句が生成され、下線(2)で「ジョンのこぶし」が生成される。次に下線(3)を見るが、これはINSTRの属性値になっているので文を生成せず道具格のスタックに退避しておく。一番外側の概念情報は事象概念なので動詞の選択を行う。属性の並びを調べて<PΔ61PΔΔIM0214ΔΔΔ>とい

うキーを作る。このキーによって動詞辞書を引き、同一の構造を持つ動詞「殴る」を得る。

#### 4. 生成実験と検討

実験にはMELCOM COSMO 700を用い、システムの処理手続きはPASCALでコーディングした。時制処理ルーチンが約200ステップ、文生成ルーチンが1300ステップである。処理手続きの論理的なまとまりごととにできるだけ一つのプログラム単位を設定して全体のプログラムの構成を明確にするとともに、作成・管理および拡張が容易になるように心掛けた。その結果、副プログラム27本、前処理プログラム1本、主プログラム1本となった。

入力となるCD表現は様相概念、様相変化概念、事象概念および埋め込み概念を用いたものなど、図-7で表されるほとんどのものについて文の生成実験を行った。その何れの場合についても、意味的に不自然でない日本語文が生成された。(表-4参照)

本システムでは埋め込み概念についてLIFOスタックで処理をするので、CD表現の入力形式に忠実な日本語文が生成される。現在のところ、生成された日本語に対しては文法上の検討を行っていない。たとえば、表-4の例文15は「私が投げた黒いボールをつかむことができなかつたトムをあなたは殴ったか」の方が、より自然な日本語文である。意味上では表現されていても、表層文では表さない方が良いものがある(たとえば前述の「トム」)。また、目的語が長い場合主語の位置も問題となる。今後は生成の過程でこれらの解析もできるようにしたい。

#### 5. あとがき

本稿では知識の表現形式にSchankのCD理論を用い、そこから表層言語として日本語文を生成する方法と実験について述べた。この方法に基づいて我々が作成したシステムでは、CD表現から生成される単文については日本語として不自然さが無いことが確認された。また、埋め込み概念(CD表現で矢印↑で表される概念)についての係り受け処理も適当だと思われる。

表-4 日本語文の生成例

1. ケーキは私のだ
2. 私の家は公園の近くだ
3. あきは啓子に本を貸した
4. ジョンはメアリーにボールを投げた
5. トムは壁の近くにテーブルを運んだ
6. トムはあなたの大きいりんごを食べた
7. 嬉しかった啓子は笑った
8. メアリーは右を向くだろう
9. いらいらしていたジョンはトムを殴った
10. あきは汗をかいた
11. 幸せな啓子はメアリーに白い大きい帽子を与えた
12. あなたの大きいりんごを食べることができなかった啓子は悲しかった
13. あきは啓子がトムに与えたケーキを食べた
14. あなたは私がトムに投げた黒いボールをつかむことができなかったトムを殴ったか
15. 幸せなトムは家の近くの大きい公園に行くことができたか

しかし、本システムでは、CD表現のうち因果関係を表す概念について、また、接続詞を含む複文及び重文の生成について扱っていない。

言語理解システムを考える場合、文の生成だけでなく、知識ベースを構築する際の知識の表現形式、さらに構築の方法など多くの問題が考えられる。知識の表現形式として、CD表現が適当かどうかについては、知識の世界を狭いものに限定すれば有用だと思われる。

謝辞 本研究には九州大学吉田将教授に種々の御助言を頂きました。また、梨木隆志、本多寛人、青木清児、川上由起子の各氏の修論や卒論に深く負っています。ここに、各氏に深く感謝の意を表します。

#### 参考文献

- 1) Schank, R.C.: "Conceptual Information Processing", North-Holland, 1975
- 2) 田中、佐藤他: "自然言語処理技術と言語理論", 電総研205, 1980
- 3) 梨木隆志: "概念依存表現の表層構造化としての日本語文合成に関する研究", 九工大修論, 1977
- 4) 本多寛人: "自然言語処理—概念依存表現からの日本語文合成", 九工大卒論, 1978
- 5) 青木清児: "自然言語処理—概念依存表現からの日本語文合成", 九工大卒論, 1979
- 6) 川上由紀子: "日本語文の自動生成", 九工大卒論, 1982