

SD式モデルによる物語文の 意味処理の試み

脇山 正博[†] 野崎 剛一^{††} 河口 英二[†]

[†] 九州工業大学工学部電気工学科 ^{††} 長崎大学総合情報処理センター

本論文では、自然言語概念の意味表現法として著者らが提案している”意味構造記述式モデル (Semantic-structure Description Form Model, SD式モデル)”を採用する。本論の目的はこのモデルを用いて自然言語の概念についての様々な意味処理を定式化することである。ここでは特に、このモデルの物語理解への応用について考察する。著者らはSD式モデルの実験システム (SDENV) を試作し、本モデルの有効性を確かめた。

Semantic processing of story data by the SD-form model approach

Masahiro Wakiyama[†], Koichi Nozaki^{††} and Eiji Kawaguchi[†]

[†] Kyushu Institute of Technology ^{††} Nagasaki University

In this paper we adopt a context-free language (SD-Forms) to deal with meaning representation of a natural language. The objective of this paper is to formalize most of the meaning processing of a natural language. The SD-forms are applied to the story understanding problem. We implemented experimental system (SDENV) to base on our model. The experiments using SDENV worked well and our model proved appropriate.

1. はじめに

物語文の研究としては物語文法という枠組みを用いた研究が見られるが¹⁾、これだけでは物語理解のモデルとしては十分とは言えない。一般に、物語の理解のためには、物語が成立する世界の定義とその世界における各種の知識が必要である。

自然言語の研究において従来から提案されてきた古典的ともいえる知識表現形式としては、フレーム²⁾や意味ネットワーク³⁾が知られている。その他にも意味表現に関する多くの試みがなされている^{4), 5)}が、認識、理解、学習するといった問題を一つの原理に基づいて扱っている研究報告は少ない。

本論文では、自然言語概念の意味表現法として“意味構造記述式モデル (Semantic-structure Description Form Model, SD式モデル)”を採用する⁶⁾。本論の目的はこのモデルを用いて自然言語の概念についての様々な意味処理を定式化することである。物語は一つの世界を構成するものであるため、本論では特にそのような物語データにおける意味処理の問題点をSD式モデルを通して議論する。

概念の意味処理には意味の差を定量的に扱うための枠組みが必要である。これに対してSD式モデルでは2種の詳述関係を基に任意の概念間の意味差の尺度を導入している。

本論文は以下、2章ではSD式モデルの概要、3章では物語データの形式やその特殊性、4章では知識システムにおける認識と理解、5章では学習について述べる。さらに、6章では物語の意味処理で最も問題である照応語の処理をSD式モデルを通して示す。7章ではSD式モデルによる意味処理実験を行なうためのシステム“SDENV”についてふれ、最後に本論のまとめと今後の課題について述べる。

2. SD式モデルの概要

SD式^{6), 7)}とは自然言語における個々の概念、陳述表現、感情表現、或はシステムに与える事実データ等を記述するための中間言語であり、その構文は曖昧さのない一つの文脈自由文法SDGで規定されている。

2.1 SD式生成文法の定義

SDGは構文的な曖昧さのない記号列を生成する一つの文脈自由文法であり、形式的には

$$SDG = \{ \Sigma_N, \Sigma_T, P, \Phi \}$$

と表わせる。ここに、

Φ は、開始記号であり Σ_N の一要素である。

Σ_N は、非終端記号の有限集合。

Σ_T は、終端記号の有限集合であり、概念ラベル、

規定子、結合子、機能項目等からなる。

P は、生成規則の有限集合であり、

$$S1: X \rightarrow [s(X), v(X)] \quad (4.0)$$

$$S2: X \rightarrow [s(X), v(X), o(X)] \quad (6.0)$$

$$S3: X \rightarrow [s(X), v(X), c(X)] \quad (6.0)$$

$$S4: X \rightarrow [s(X), v(X), i(X), o(X)] \quad (8.0)$$

$$S5: X \rightarrow [s(X), v(X), o(X), c(X)] \quad (8.0)$$

のような形式で定められている。S1-S5 は何れも陳述SD式を生成する規則の例である。() 内の数値は変数が右辺の形に置き換えられるときに加わる意味の詳述量として定義されている値である。

その単位を“semit”とする。

2.2 詳述関係

SD式で表記された二つの概念 D_1, D_2 について、構文構造に基づく詳述関係 $ELAB_{synt}(D_1, D_2)$ と知識データに基づく詳述関係 $ELAB_{know}(D_1, D_2)$ を統合したものを、あらためて“詳述関係”と定義し、 $ELAB(D_1, D_2)$ と表記する。

すなわち、

$$ELAB(D_1, D_2) = \min \{ ELAB_{synt}(D_1, D_2), ELAB_{know}(D_1, D_2) \}$$

である。 D_1, D_2 間に詳述関係が成立することを、 $D_1 \supset D_2$ と表わす。

このとき D_1 は D_2 の先祖、 D_2 は D_1 の子孫と呼ぶ。

詳述関係は半順序関係であるのでこれを用いて概念を体系化できる。詳述関係は一般には多段階に連鎖する。

D_1, D_2 についてそれぞれに最も近い共通の先祖 D^* を D_1, D_2 の最近共通先祖と呼ぶ。そして

$$DIFF(D_1, D_2, D^*) = ELAB(D^*, D_1) + ELAB(D^*, D_2)$$

となる値を D_1, D_2 の意味差の尺度と定義する。

2.3 事実データの階層

自然言語概念に関する知識を内蔵し、これに基づいて外部入力に対する意味処理を行うシステムを「知識システム」と呼ぶ。

SD式モデルでは、一つの知識システム内の事実データを次の2種に大別する。

- (1) 対象世界を規定する世界データ
- (2) 対象世界で有効となる知識データ

図1はSD式モデルにおけるこれらの階層を示す。

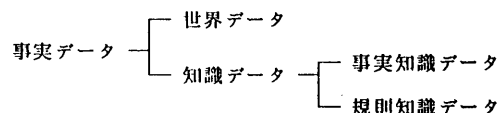


図1 事実データの階層

Fig.1 Hierarchy of fact data

知識システム内の事実データは詳述関係“ $\supset$ ”で体系化することができる。

以下に簡単な例を示す。

<例1>

世界データ

W1:[s(桃太郎),v(退治/過去),o(赤鬼)]

事実知識データ

F1:[s(人間/SOME),v(退治),o(鬼)]

F2:(人間)incl(桃太郎)

F3:(悪者)incl(鬼/悪)

規則知識データ

R1:(assu((s(桃太郎),v(発見/過去),o(X)))

andx((悪者)incl(X))))

caus([s(桃太郎),v(退治/過去),o(X)])

これらの体系化については4. 1節に示している。

2. 4 知識システムにおける推論

SD式モデルでは D_{root} が事実データであるとき、

$$D \supseteq D_{root}$$

なる全ての D を真とする。すなわち、

$$true(D)$$

である。さらに、 $true(D_1)$ なる概念 D_1 を用いて

$$D \supseteq D_1$$

なる D も真とする。また、事実を基にした帰納的推論の結果や、システム固有の信念に基づく $true(D)$ を定義することも可能である。

SD式モデルでは $true$ の他にも不確実な推論を行うための *plausible*, *possible*, *uncertain* 等の関数も定めている。

3. 物語データの形式や特殊性

世界データの特別な場合として“物語データ”を考えることができる。人間にとって物語を理解するということは、筋の展開に論理的な矛盾がなく全体の意味内容を把握できることである。この場合、人間が持っている一般常識や物語に関連した個別知識が大きな役割を果たす。

自然言語での物語文の中には、“呼掛け”、“感嘆”、“独言”等も含まれるが、大半は“順序付けられた陳述文”の集合と考えてよい。本論では、自然言語の物語文が既にSD式に変換済みであり、各物語素片は

物語世界名:文番号:物語素片

なる形式のデータであると仮定する。このうち、“物語素片”部分が自然言語の各文の意味構造を陳述SD式として表現している部分である。例えば、付録に示す事実データは「桃太郎」のあらすじを示す。

SD式で表わされた物語データが一般の世界データと異なる点は次のようなことである。

- 1: 文番号で示されるデータの順序には重要な意味がある。
- 2: 照応表現がある。
- 3: 文脈を利用した省略が生じていることがある。

以上のことから物語データの特長性として以下のことが結論される。

【特徴1】まず、1: に関連し、物語データがシステムの外部から一文毎に（物語素片毎に）入力された場合、各文は、システム内部に蓄積されている知識データと、既に入力された物語素片の集合に照らしてみても

a) 無矛盾であること

b) 解釈して認識・理解ができること

という条件を満たすものでなければならない。

【特徴2】物語によく現れる照応語は物語の聞き手（または読者）の想像力を刺激し、物語の各部分を強く印象づける効果がある。しかし、その実体（照応語が意味しているもの）は聞き手（すなわち知識システム）にとって同定可能な関係を保持していなければならない。

【特徴3】記述の省略も聞き手にある種の推論を要求し物語を印象づける効果がある。しかしこの場合も欠落した情報はシステム内の知識と既に入力されている物語素片から復元できる範囲になければならない。

本論では6章でこれらに対してSD式モデルでの対処を示している。以下、4章と5章ではSD式モデルに基づいて、物語の知的処理の前提となる認識、理解、学習等の形式化を行なう。

4. SD式モデルによる知的処理

4. 1 コンパクト概念体系

知識システム内の世界データや、知識データは一つ一つの概念として孤立して蓄えられているのではなく、全体を有機的に関連づけて蓄積しておく。この場合、4. 2節以下に示す各種の知的処理の効率を高めるための工夫と、知識の蓄積に要する記憶量を出来るだけ少なくすることが重要である。本論では、システムに具体的に与えられた事実データとしての概念集合 Ω_0 を基にして創り出される一つの概念体系を以下のように定義する。

概念体系 $\{\Omega/\Omega_0, \Psi\}$ とは、概念集合 Ω と詳述量の集合 Ψ からなる意味ネットワークである。ここで概念集合 Ω は変数 X と与えられた事実データの集合 Ω_0 を含むものである。概念は意味ネットワークにおける節点であり、詳述関係は意味ネットワークの枝によって表される。与えられた事実データ Ω_0 のコンパクト概念体系（略して、コンパクト体系）とは以下の条件を満足する概念体系 $\{\Omega/\Omega_0, \Psi\}$ のことである。

- (1): Ω_0 は Ω の部分集合である。
- (2): Ω_0 における任意の二つの節点 $D1$ と $D2$ に対して最近共通先祖 D^* が存在し、 D^* は Ω に属する。
- (3): Ψ は一段の詳述関係からなる。
 Ω において、すべての節点間の詳述量は Ψ の詳述量を累積することによって与えられる。
- (4): もし差集合 $\Omega - \Omega_0$ に属する任意の節点を除け

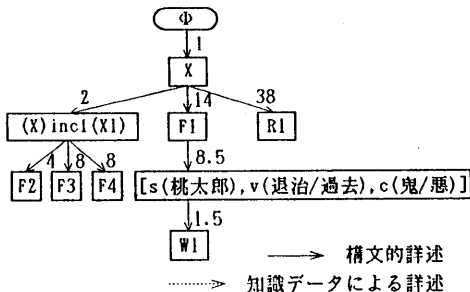
ば (2) または (3) が満たされなくなる。
 (5) : もし Ψ に属する任意の枝を除くと, (3) が満た
 れなくなる。

以上のようなコンパクト概念体系を用いると知識シ
 ステムにおける認識動作, 理解動作, さらに学習動作等
 の形式化が容易となる。

概念体系における総事実記憶量(TM)と実知識量(NK)と
 は, それぞれ Ω_0 の意味量の全体, 及び Ψ の総詳述量で
 ある⁶⁾。

以下, 図2に<例1>の事実データにさらに事実知識
 データとして

F4:(鬼/悪)incl(赤鬼)
 がある時のコンパクト概念体系図を示す。



TM = 100 (semit) NK = 85 (semit)

図2 コンパクト概念体系図

Fig.2 A conceptual hierarchy

4. 2 知識システムにおける無矛盾性

知識システムは無矛盾でなければならない。

すなわち, 世界データと知識データとして与えられるデ
 ータを Ω_0 とし, それから得られるコンパクト概念体系
 $\{\Omega/\Omega_0, \Psi\}$ に関して矛盾が生じてはならない。本モデ
 ルでの矛盾とは以下のいずれかが成立することである。

- 或るSD式 D とその否定 ($\text{nega}(D)$) が同時に
 true となる。
- 或る概念から出発し, 半順序関係により別の知
 識を辿るうちに自分自身に帰着する。

システムに矛盾が発見されたときどのようにして矛盾を
 解消するかは大きな問題であるが, システム自身に自己
 修正能力は無く, 外部から指示せざるを得ない。

4. 3 外部入力に対する認識動作の定義

外部入力として概念 D_{in} が与えられたとする。知識
 システムにおいて, この D_{in} を認識するとは, 知識シ
 ステム内のコンパクト概念体系 $\{\Omega/\Omega_0, \Psi\}$ において
 D_{in} に最も近い先祖 D_{rso} を見いだすことである。す
 なわち,

$$ELAB(D_{rso}, D_{in}) = \min_{D \in \Omega} ELAB(D, D_{in})$$

である。このことを $D_{rso}(D_{in})$ と表記する。

例えば, 例1の事実データがあるものとする。
 このとき次のような外部入力

$$D_{in} = [s(\text{桃太郎}), v(\text{発見/過去}), o(\text{鬼/悪})]$$

が与えられたとするとシステムは

$$D_{rso}(D_{in}) = [s(\text{人間/SOME}), v(\text{退治}), o(\text{鬼})]$$

と認識する。

4. 4 外部入力に対する理解動作の定義

外部入力概念 D_{in} の理解動作とは, 知識システム内
 のコンパクト概念体系 $\{\Omega/\Omega_0, \Psi\}$ において $D \in \Omega$
 なる任意の D に対して $D_{in} \supset D$ を確認することであ
 る。

例えば, <例1>の事実データがあり, さらに事実知
 識データとして F4 : (鬼/悪)incl(赤鬼)があるものと
 する。外部入力

$$D_{in} = [s(\text{桃太郎}), v(\text{発見/過去}), o(\text{鬼/悪})]$$

が与えられたとすると

$$\text{true}([s(\text{桃太郎}), v(\text{退治/過去}), o(\text{鬼/悪})])$$

となり, 外部入力 D_{in} をシステムは理解できる。

4. 5 外部入力に対する解釈の定義

D_{in} を外部入力とすると, 以下のような関係

$$\text{INTP}(D_{in}, D)$$

を満たす D をその解釈と定義する。

この D を $D = D_{\text{INTP}}(D_{in})$ と表わす。

1. $\text{INTP}(D_{in}, D)$

但し, $\text{ELAB}(D_{in}, D)$ or $\text{ELAB}(D, D_{in})$ 。

2. $\text{INTP}(D_{in}, D)$

但し, 以下の (2-a), (2-b) を満たす。

(2-a): D_{in} と D は同じタイプのSD式である。

(2-b): D_{in} と D の互に対応する部分 D_{in}' , D'
 について

$$\text{INTP}(D_{in}', D')$$

<例2>

知識データ

(動物)incl(雉)

(人間)incl(桃太郎)

外部入力

$$D_{in} = [s(\text{雉}), v(\text{随伴}), o(\text{人間})]$$

この場合, 以下は解釈の例となる。

- $D_{\text{INTP}}(D_{in}) = [s(\text{雉}), v(\text{随伴}), o(\text{桃太郎})]$
- $D_{\text{INTP}}(D_{in}) = [s(\text{雉}), v(\text{随伴/過去}), o(\text{桃太郎})]$
- $D_{\text{INTP}}(D_{in}) = [s(\text{動物}), v(\text{随伴}), o(\text{桃太郎})]$
- $D_{\text{INTP}}(D_{in}) = [s(\text{動物}), v(\text{随伴}), o(\text{人間})]$

5. 学習の定義

学習とは, 認識や理解の能力を高めるための新たな知
 識データの獲得や, 既存知識の再編成動作である。一般
 に知識データの記憶量が増えたと応答の質 (正しい応答)
 は向上するが, システムの処理速度は低下する。従って

実知識量をなるべく減らさずに、知識データの記憶量の削減と処理速度の向上が学習のねらいである。

5. 1 外部からの新たな知識の獲得

外部入力 D_{in} は 事実データとの関係で以下の4つに分類できる。

- (1) $Dr_{oo}(D_{in})=X$ (全く未知の概念)
- (2) $ELAB(Dr_{oot}, D_{in})$ (既知の概念枠内での新しい概念であることを認識)
- (3) $true(D_{in})$ または $true(D_{oooo}(D_{in}))$ (理解するが新しい情報はない)

(4) 既知の知識データと矛盾するもの。

従って、(1),(2) の場合のみが知識獲得の対象となる。

(1) の場合は、そのまま事実データとして登録する。

(2) の場合は、知識獲得を行い新しい知識を獲得後コンパクト概念体系を更新しなければならない。

5. 2 自己学習

知識システムにおける自己学習とは、既にシステムに蓄えられている事実データを再編成して事実データの記憶量を減らすことと、処理速度の向上を計ることである。

以下は総記憶量を削減する自己学習の例である。

(1) 事実データの一般化

<例3>

いま、 $W1, W2, W3$ を以下の事実データであるとする。

$W1 : [s(\text{赤鬼}), v(\text{である}), c(\text{悪者})]$

$W2 : [s(\text{白鬼}), v(\text{である}), c(\text{悪者})]$

$W3 : [s(\text{黒鬼}), v(\text{である}), c(\text{悪者})]$

総記憶量 = 39.0

この場合これらを次の $W11, W12$ に置き換える。

$W11 : [s(\text{LABEL1/ANY}), v(\text{である}), c(\text{悪者})]$

$W12 : (\text{LABEL1})\text{csof}([s(\text{赤鬼}), s(\text{白鬼}), s(\text{黒鬼})])$

総記憶量 = 30.0

但し、 LABEL1 はシステム自身がつける概念ラベルである。

(2) 事実データの抽象化

<例4>

事実データ ($W1, \dots, W6$) を事実データ ($W10$) と規則データ ($R1$) に置き換える。

$W1 : [s(\text{犬}), v(\text{随伴}), o(\text{桃太郎})]$

$W2 : [s(\text{桃太郎}), v(\text{与える/過去}),$

$i(\text{犬}), o(\text{きびだんご})]$

$W3 : [s(\text{猿}), v(\text{随伴}), o(\text{桃太郎})]$

$W4 : [s(\text{桃太郎}), v(\text{与える/過去}),$

$i(\text{猿}), o(\text{きびだんご})]$

$W5 : [s(\text{雉}), v(\text{随伴}), o(\text{桃太郎})]$

$W6 : [s(\text{桃太郎}), v(\text{与える/過去}),$

$i(\text{雉}), o(\text{きびだんご})]$

↓

総記憶量 = 102.0

$W10 : (\text{LABEL2})\text{csof}([s(\text{犬}), s(\text{猿}), s(\text{雉})])$

$R1 : [\text{assu}([s(\text{LABEL2/ANY}), v(\text{随伴}), o(\text{桃太郎})])]$

$\text{caus}([s(\text{桃太郎}), v(\text{与える/過去}),$

$i(\text{LABEL2/ANY}), o(\text{きびだんご})])]$

総記憶量 = 58.0

(3) 既存の事実データを忘れる学習

次の例は事実データの記憶量を削減する効果は大きい、システムの動作が多少変わる例である。

<例5>

($W1, \dots, W4$) を ($W4, W11$) に置き換える

$W1 : [s(\text{鬼/赤}), v(\text{である}), c(\text{悪者})]$

$W2 : [s(\text{鬼/白}), v(\text{である}), c(\text{悪者})]$

$W3 : [s(\text{鬼/黒}), v(\text{である}), c(\text{悪者})]$

$W4 : [s(\text{鬼/桃}), v(\text{である}), c(\text{善者})]$

↓

総記憶量 = 68.0

$W4 : [s(\text{鬼/桃}), v(\text{である}), c(\text{善者})]$

$W11 : [s(\text{鬼/MOST}), v(\text{である}), c(\text{悪者})]$

総記憶量 = 34.0

自己学習の目的には以上の他に、多段推論における処理速度の向上を目指す場合がある。以下に簡単な例を示す。

(4) コンパクト概念体系のネットワーク上の特定概念間に直結の詳述枝を付ける。

<例6>

$ELAB([s(\text{桃太郎}), v(\text{退治}), o(\text{悪者})],$

$[s(\text{桃太郎}), v(\text{強く成る})])=1.0$

$ELAB([s(\text{住民}), v(\text{平和})],$

$[s(\text{桃太郎}), v(\text{退治}), o(\text{悪者})])=1.0$

↓

$ELAB([s(\text{住民}), v(\text{平和})], [s(\text{桃太郎}), v(\text{強く成る})])$

=2.0

これをコンパクト概念体系の Π に追加する。しかし、このような追加をするともはやその体系はコンパクト体系でなくなる。

(5) 推論の帰結を事実データとする

<例7>

$K1 : [(s(\text{犬})\text{incl}(\text{ボチ}))]$

$R1 : (\text{assu}((s(\text{犬})\text{incl}(X))\text{caus}([s(X), v(\text{吠える})]))]$

$K1, R1$ から推論した帰結は

$K2 : [s(\text{ボチ}), v(\text{吠える})]$

であるが、これを事実データとして登録する。

6. 物語を対象とする意味処理

3章で述べたように、物語データは世界データの種類ではあるが、物語に固有な性質を持つものである。本論では知識システムが行なう意味処理として、物語に関する外部からの質問に対し、意味的に正しい応答が出来るような処理を目標とする。

このような物語の意味処理に関して特に問題点として次の2点を挙げる事ができる。

a) 或る物語データが与えられたとき、どのような知

識データが前提となっているか。

b) 照応語の処理として、システムにどのような機能を持たせればよいのか。

a) に関しては、全ての分野にわたり広く知識を用意しておけば良いが、現実には不可能である。従って物語素片データが逐一入力される度に、システム自身がそれを認識理解するためにまだ不足している情報を外部に質問出来るようになっていれば好都合である。以下、6.1節ではSD式モデルにおけるそのような仕組みを検討する。

b) に挙げた照応語処理の機能は知識システムに不可欠である。SD式モデルでは構文的な条件と知識に基づいて照応語の対象概念を同定するが、いくつかの具体的な方法を6.2節で述べる。

6.1 物語の意味処理の枠組み

物語データの一つの素片 D_{in} がシステムに与えられるとき、それまでのコンパクト概念体系 $\{\Omega/\Omega_a, \Psi\}$ に照らして D_{in} が

$plausible(D_{in})$

であるとき、 D_{in} にはあまり新しい情報はなく、システムはこれを当然のこととして“受け流す”。しかしそうでない場合は

$plausible(D_{intp}(D_{in}))$

となるような解釈を仮定する。そしてこの解釈が正しいか否かを外部に質問する。

<例8>

知識データ

(人間)incl(桃太郎)

(assu([s(X), v(随伴), o(人間)]))

caus((動物)incl(X))

入力物語素片

$D_{in}=[s(雉), v(随伴), o(桃太郎)]$

この場合、(動物)incl(雉)を仮定すれば

$D_{intp}(D_{in})=[s(雉), v(随伴), o(人間)]$

なる解釈に対して

$plausible(D_{intp}(D_{in}))$

となる。従ってシステムは外部に対して次のような確認のための質問を行なう。

fact((動物)incl(雉))?

この例では未知ラベル“雉”に対する既知概念との関係“incl”についての質問であるが、一般には、概念ラベル、修飾子形式、結合子形式、陳述式、間の結合関係

incl, ptof, indu, csof, caus

に関する質問を行なうことが可能である。

6.2 照応処理の枠組み

談話や物語における照応解析は意味処理の重要な課題である。照応問題とは、一般に、人称代名詞(彼, 彼女, he, she), 指示代名詞(これ, それ, あれ, this, it, that)等の照応語が意味するものの実体を物語デー

タの中で探索することである。

本モデルでは照応解析のレベルを

(1) 構文情報だけによる解析

(2) 構文情報と意味情報による解析

の2段階に分け、さらに照応ラベルの実体が、

(A) 一つ概念ラベルである場合

(B) 一つ以上の物語事実データ(すなわち、部分物語)である場合に分けて扱うことにする。

構文情報のみによる解析結果の妥当性の判定は、同定されたラベルを照応語と置き換えてみればよい。その結果、「意味的に妥当なSD式」となっていればそのラベルを受け入れ、「不当なSD式」であれば棄却して別の照応を求める。

<例9>

物語データ

momotaro: 4:([s(妻/老), v(洗濯/過去)])pseq

([s(桃/(SOME)para(大きい)), v(流来/過去)])

momotaro: 5:[s(彼女), v(拾う/過去), o(それ)]pseq

([s(彼女), v(帰宅/過去)])

知識データ

K1: [s(人間), v(拾う), o(物事)]

K2: (人間)incl(彼女)

K3: (物事)incl(桃)

この例における照応問題は、照応語「彼女」、「それ」の実体を同定することである。

文番号5の照応語「彼女」は、「妻/老」と同定し、照応語「それ」は、「桃/(SOME)para(大きい)」と同定できればよい。(7.2.1参照)

7. SDENVの試作と実験

7.1 SDENVの試作

SD式による意味処理方式が有効であるか否かは、実際のSD式データを用いて具体的な処理を試してみなければならない。そのため、本研究ではSD式処理システムとしてSDENV(SD-Formula ENVironment)を試作した。このシステムはPROLOGを用いて記述したものであり、プログラム量は約240KBである。SDENVはSD式の処理に関する200種以上の関数/述語を備えている。SDENVは全体としては、多様な関数/述語群と、システムに内蔵された事実データからなり、豊富な意味処理実験が可能である。特に、SD式を表す「変数」の扱いが可能であるため、知識や規則の一般的な記述、SD式同士のマッチングに於ける部分SD式の同定が容易である。

7.2 SDENVによる実験

7.2.1 照応処理

自然言語で書かれた文に関する質問応答システムを考

えるとき、以下のような問題点に気づく。まず、一般の知識や概念世界に関する質問応答では、

- 1) 同義語、反義語、類義語
- 2) 概念の階層関係
- 3) 部分全体関係

などが問題である。さらに物語文に関しては、

- 4) 照応問題
- 5) 物語個別知識

が問題となる。その他、具体的な自然言語文を処理する際には、修飾による意味の限定、能動態・受動態などの言い替え、さらに否定の意味範囲なども問題となる。このような問題の解決には曖昧さのない具体的な意味表現が必要となる。

質問応答処理には、特に 4) の照応問題の解決が前提となることが多い。

S D E N V では照応問題を以下の例のように処理している。

まず、前方照応の入力形式は以下のようにしている。

anaphora(_oorigin, _oobj, _iorigin, _iobj).

先行物語事実 _oorigin 中のラベル _oobj と、
後行物語事実 _iorigin 中のラベル _iobj とが
anaphora なる関係である。

<前方照応コマンドの例>

<例 9> の事実データがあるものと仮定する。

anaphora(_oorigin, _oobj, [momotaro:5], それ)

_oorigin = [momotaro:4]

_oobj = 桃/(SOME)para(大きい)

照応処理として後行物語事実(文番号 n) が入力されると先行物語事実の文番号 n-1, n-2, ... に対して構文的処理と意味処理を行う。構文的処理では、照応文の前後の機能項目 s, i, o 等のラベルの候補を調べる。その同定が終わると意味処理、たとえば「拾う」というような述語項目の知識データを調べその主語や目的語項目等に対してそれらを満足する妥当な解釈が存在するか調べる。

7.2.2 物語に関する質問応答処理

S D E N V 上では物語データに対する質問応答処理も可能である⁹⁾。質問文は以下のように2つのタイプに分けて処理している。

- (1) 概念の status (true, plausible, possible, false, uncertain) に関する質問

この場合の質問文の入力形式は以下のようにしている。

status(_worldname, _answer, _statement).

但し, _worldname : 物語世界名

_answer : 応答

_statement : 入力陳述

<例>

以下の質問について付録の事実データがあるものと仮定する。

- a) status 型質問

```
status(momotaro, _answer,  
[s(犬/some), v(到着/(過去)para(場所/鬼が島))]).  
_answer = true  
status(momotaro, _answer,  
[s(桃), v(浮く/(媒体/水))]).  
_answer = plausible
```

- (2) 5W1H 型質問

これは 何時(when), 何処(when), 誰(who),
何(what), 何故(why), 如何(how) に関する質問である。

この場合の質問文の入力形式は以下のようにしている。

query(_worldname, _answer, _statement).

但し, _worldname : 物語世界名

_ans : 応答

_statement : 入力陳述

<例>

以下の質問について付録の事実データがあるものと仮定する。

- b) 何時(when) 型質問

```
query(momotaro, _ans, [s(桃太郎),  
v(成る/過去/何時), c(若者/強)]).  
_ans = [s(桃太郎), v(成る/過去/ある日),  
c(若者/強)]
```

- c) 何処(when) 型質問

```
query(momotaro, _ans, [s(桃太郎),  
v(出征/何処)]).  
_ans = [s(桃太郎), v(出征/(過去)para  
(目的/鬼退治)para(場所/鬼が島))]
```

- d) 誰(who) 型質問

```
query(momotaro, _ans, [s(誰),  
v(洗濯/過去)]).  
_ans = [s(妻/老), v(洗濯/過去)]
```

- e) 何(what) 型質問

```
query(momotaro, _ans, [s(桃太郎),  
v(与える/過去), i(犬/当該), o(何)]).  
_ans = [s(桃太郎), v(与える/過去),  
i(当該/犬), o(きびだんご)]
```

物語データに関する質問を行うと、照応処理を解決し対応する物語事実を同定する。対応する物語事実が見つからなかった場合には知識データを参照し推論機能を用いて応答する。

8. おわりに

PROLOG で試作した SD 式処理システム「S D E N V」は多様な能力を持つものであり、物語 SD 式データを用いた意味処理は一般知識や推論規則データを増やせば、S D E N V 上でかなりきめ細かく行える見通しがついた。この S D E N V を用いて今後さらに高度な意味

処理実験（物語の要約，話題関連知識の制御，人間の心理言語行動規則等々）を行う予定である。一方，システムの高速化は今後の大きな課題である。

参 考 文 献

- 1) Rumelhart, D.E. : Notes on a Schema for Stories, in Bobrow, D.G. and Collins, A. (eds), Representation and Understanding : Studies, in Cognitive Science, Academic Press (1975).
- 2) Minsky, M. : "A Framework for Representing Knowledge." in The Psychology of Computer Vision, edited by Winston, P.H., McGraw-Hill Book Company (1975).
- 3) Quillian, M.R. : "Semantic memory, in Semantic Information Processing", edited by Minsky, M.L., pp.227-270, MIT Press (1968).
- 4) Montague, Richard : "Proper Treatment of Quantification" in Ordinary English, in Formal philosophy, edited by Thomason, R.H., pp.247-270. Yale University Press (1974).
- 5) Schank, R.C. and Abelson, R.P. : "Scripts, Plans, Goals and Understanding", Lawrence Erlbaum (1977).
- 6) Eiji Kawaguchi, Masahiro Wakiyama and Koichi Nozaki, "A Semantic Structure Description Model of General Concepts in a Natural Language World" Proceeding of Pacific Rim International Conference on Artificial Intelligence '90, pp.298-303, (1990).
- 7) 野崎, 脇山, 河口 "自然言語概念の意味構造モデルと会話文データベースへの応用", 第81回自然言語処理研究会本大会別稿
- 8) 脇山, 鎌田, 河口 "物語の意味記述データを基にした質問応答システム", 情報処理学会39回(平成元年度後期)全国大会N0.2G-2, pp.686-687

付 録

<桃太郎物語データ>

STORY(momotaro: 1:[s(夫婦/(some)para(老)),v(住む/(過去)para(場所/或)para(時/ある時))]).
 STORY(momotaro: 2:[s(夫/老),v(行く/(過去)para(習慣)para(場所/山)para(目的/柴刈))]).
 STORY(momotaro: 3:[s(妻/老),v(行く/(過去)para(習慣)para(場所/川)para(目的/洗濯))]).
 STORY(momotaro: 4:[s(妻/老),v(洗濯/過去)]pseq([s(桃/(SOME)para(大きい)),v(流来/過去)]).
 STORY(momotaro: 5:[s(彼女),v(拾う/過去),o(それ)]pseq([s(彼女),v(帰宅/過去)]).
 STORY(momotaro: 6:[s(夫/老),v(割る/(過去)para(手段/包丁)),o(桃/前述)]pseq([s(嬰兒/(或)para(可愛い)para(男)),v(出現/(過去)para(起点/桃/当該))]).
 STORY(momotaro: 7:[s(彼等),v(命名/過去),o(彼),c(桃太郎)]).
 STORY(momotaro: 8:[s(彼),v(成る/(過去)para(時/ある日)),c(若者/強)]).
 STORY(momotaro: 9:[s(彼),v(出征/(過去)para(目的/鬼退治)para(場所/鬼ガ島))].
 STORY(momotaro: 10:[s(犬/some),v(随伴/(過去)para(場所/村)),o(桃太郎)]).
 STORY(momotaro: 11:[s(桃太郎),v(与える/過去),i(犬/当該),o(きびだんご)]).
 STORY(momotaro: 12:[s(猿/some),v(随伴/(過去)para(場所/山)),o(桃太郎)]).
 STORY(momotaro: 13:[s(桃太郎),v(与える/過去),i(猿/当該),o(きびだんご)]).
 STORY(momotaro: 14:[s(雉/some),v(随伴/(過去)para(場所/野原)),o(桃太郎)]).
 STORY(momotaro: 15:[s(桃太郎),v(与える/過去),i(雉/当該),o(きびだんご)]).
 STORY(momotaro: 16:[s(桃太郎),v(到着/(過去)para(場所/鬼ガ島))].
 STORY(momotaro: 17:[s(彼等),v(撲滅/過去),o(鬼/複数)]).
 STORY(momotaro: 18:(上述)caus([s(住民/当該),v(暮らす/(過去)para(平穩))]).

<事実知識データ>

F1:(夫婦)cs of([夫,妻])
 F2:(仕事)incl([柴刈,洗濯])
 F3:(人間)incl([夫婦,彼女,彼等])
 F4:(勝利)incl([撲滅])
 F5:(鬼)incl([赤鬼,白鬼,黒鬼])
 F6:(物事)incl([桃,きびだんご])
 F7:(場所)incl([鬼ガ島])
 F8:(時)incl([ある日])

<規則知識データ>

R1:(assu([s(X),v(浮く/(媒体/水))]))
 caus([s(X),v(流来)])
 R2:(assu([s(X),v(到着/場所/Y)])
 andx([s(Z),v(随伴),o(X)]))
 caus([s(Z),v(到着/場所/Y)])