

エピソードネットワークを用いた物語のあらすじ生成^{†*}中澤俊哉^{††} 重永実^{†††}

物語のあらすじを生成するには、まず物語を深く理解し、文の因果関係を十分考慮することが必要である。そこで日本語物語理解システム JStory が生成するエピソードネットワークを使ってあらすじ生成を試みた。JStory ではさまざまな知識を用いて事象間のより深い関係を推論したり、入力文章の行間にいくつかの補助的な文を埋めていき、その結果をエピソードネットワークとして表現している。つまり、物語の展開(筋)や各文の因果関係をエピソードネットワークに変換し、それを物語の内部表現としている。そこでこのエピソードネットワークに基づいてキーセンテンスを抽出し、あらすじを生成する。その際、表層的にキーワードやキーセンテンスを抽出するのではなく、原因-結果、理由-行動に注目して、物語の展開(筋)やキーセンテンス間の因果関係を維持した飛躍のないあらすじ生成を試みている。目的達成に対し何度も失敗を繰り返した後ゴールを達成するような、話の展開の異なる物語についても、基本方針を変えることなく対応できる。かつ四つの昔話について人間によるあらすじ生成と比較してシステムの動作を検証している。この方法の特徴は、大局的な要約が可能であり、連鎖関係に基づいており、かつ、スクリプトや概念ユニットなどの知識構造に着目した局所的な要約も取り入れているので、あらすじに飛躍が生じにくい点にある。

1. はじめに

あらすじ(要約)生成の手法はこれまでにいくつか提案されている。単語の頻度、キーワードなどから文章内の重要箇所を求める表層的な手法¹⁾、スクリプト²⁾などの知識構造に基づいてトップダウン的に文章内の重要箇所を求める手法³⁾、全文章を意味表現に変換してから何らかの評価基準に基づいてボトムアップ的に文章内の重要箇所を求める手法⁴⁾、これらを融合した手法⁵⁾などさまざまなレベルがあり、目的や対象により使い分ける必要がある。特にその対象を物語とする場合には、まず物語を深く理解し、文の因果関係を十分考慮して、あらすじを構成する文章間の連鎖関係を維持することが必要であると考えられる。しかしこれまでにこのような意図をもって具体的にあらすじを生成したものはないと思われる。

一方、筆者らは「自然言語で記述した一つのまとまりのある文章を、できるだけ人間の情報処理過程に近い方法で計算機に理解させる」ということを目的として、物語理解システム JStory (Japanese Story Understanding System)⁶⁾を構築し、物語を理解した結果として生じる事象のエピソードネットワークの構築について報告し⁷⁾、さらに複数名詞やプラン-ゴール

におけるゴールに対する失敗への対応⁸⁾、省略の補填⁹⁾などの補強をしてきた。ここでは、人が物語を理解する過程を「読み手が既に持っている実世界に関する知識や常識や物語の展開についての予測などを用いて、物語内の事象を意味的に結び付ける過程である」と見なし、以下のことに主眼を置いている。

- 物語中で陽に述べていない事柄を様々な種類、様々なレベルの推論規則を用いて物語表現に組み込む、
- 物語中の事象間の関係を求める、
- 物語表現や各事象間の関係をネットワークとして記憶の中に残す。

つまり、物語理解を事象のネットワークを作成する過程であると思なしている。

そこで本稿では JStory が物語を理解した結果として生成する事象のネットワークに基づいて、ボトムアップ的にキーセンテンスを抽出する手法を採用した。それは前述のように、事象のネットワークは物語を理解し、文章内の各事象間の因果関係を表現したものであるから、このネットワークに基づいてキーセンテンスを抽出することにより、物語の展開(筋)を追跡し、それを構成するキーセンテンス間の連鎖関係(原因-結果、理由-行動)を維持したあらすじの生成ができると思われるからである¹⁰⁾。文献 6)でも同じ考えの下に種々試みているが、本稿ではその結果を勘案、補強し、かつ 6)では十分な対応ができなかった異なるタイプの物語にも対応できるように改善されており、物語の筋を維持し、文と文との間の因果関係を保持したあらすじを生成することができている。そして不十分な実験ではあるが、人間の指摘とかなり同

† Composition of Outline of a Story by Its Episode Network by TOSHIYA NAKAZAWA (Fuji Xerox Co.) and MINORU SHIGENAGA (Department of Computer Science, School of Computer and Cognitive Sciences, Chukyo University).

†† 富士ゼロックス(株)

††† 中京大学情報科学部情報科学科

* この研究は筆者らが山梨大学工学部計算機科学科に在籍中に行ったものである。

じようなキーセンテンスを抽出できていることを確かめている。

2. システムの概要

システムは構文解析部、意味解析部、分脈解析部、文生成部、およびあらすじ生成部から成る。構文解析部では、ローマ字で分かち書きした入力文から句構造を抽出し、構文木を作成する。意味解析部では、構文木を基にして係り受けなどの処理を行い、入力文の意味表現を作る。文脈解析部では、意味解析部でできた意味表現から、適切な知識構造に起動をかけ、事象（イベント）間のつながり（エピソードネットワーク（以下 EPN と略記））を作る。文生成部では、物語内部表現を日本語文に変換し出力する。あらすじ生成部では、キーセンテンスの抽出を行い、あらすじ文を生成する。

EPN を生成する構文・意味・文脈解析部の詳細は文献 7) を参照していただきたい。

現在、本システムでは「こぶとりじいさん」「舌切り雀」「大きなかぶ」「笠地蔵」の四つの物語（昔話）が処理可能になっている。

3. エピソードネットワーク

物語をどのように表現すれば良いのかという問題は、人が物語を理解した後、物語を記憶の中でどのように表現しているのかという問題が解明されていない現段階では、明確な答を得ることができない問題である。本研究では、構文・意味・文脈解析の結果、物語内で生じた事象に相当するイベントと、事象間を関連付けている連鎖に相当するリンクによって成り立つネットワークを物語の内部表現としている。そのネットワークを EPN, EPN 内のリンクをエピソードリンク（表 1 参照）、EPN 内のノードのうち入力文を表現したノードを入力イベント、システムが推論して新たに作成したノードをデータイベントと呼ぶことにする。入力イベントとデータイベントの形式を図 1 に示し、例として「笠地蔵」の「おじいさんはお地蔵様の雪を払ってあげました」という文に相当する入力イベントを図 2 に、同じく「笠地蔵」の「お地蔵様は雪を取りたい」というゴールに相当するデータイベントを図 3 に示す。

また、相異なる二つのタイプの物語の代表と

表 1 エピソードリンクの分類
Table 1 Categories of episode links.

リンク名	記号	説明
planfor	p	あるゴールに対して使うプランを示す
subgoal	s	あるプランを達成するためのサブゴールを示す
instan	in	プランから行為イベントに向かい、そのプランの具体化したものを示す
motive	m	ゴールとそのゴールに動機を与えたイベントをつなぐ
reason	R	精神的な事象、または状態は行為の理由となり得る
initiate	it	ある客観的事象、または状態は偶然に精神的な事象を引き起こし得る
result	r	ある行為はある状態を引き起こす
enable	e	ある状態はある行為を可能にする
disable	da	ある状態はある行為を不可能にする
achieve	a	あるイベントがゴールを達成させる
fail	f	あるイベントがゴールを失敗させる
sum	sm	いくつかの複数の事象から合意される

- ・入力イベントの形式
(<イベント番号><述語の見出し語>
<スロット 1><スロット 2>…)
- ・データイベントの形式
(<データ番号>DATA
<スロット 1><スロット 2>…)

図 1 入力とデータイベントの形式
Fig. 1 Forms of input and data events.

```

(13
harawu : イベント番号
(item zl) : 述語の終止形
(syntax-inf end p a s ageru) : 述語辞書番号
(kaku) : 法情報
      : 格構造
(age oziisax-3) (dog te-40)
(frl ozizoosama-32) (tol *g00116)
(tai yuki-19 (pas-ownship ozizoosama-32)))
(imi : CD 表現
(do (actor oziisan-3)
    (obj yuki-19 (pas-ownship ozizoosama-32))
    (result
      (change
        (obj yuki-19 (pas-ownship ozizoosama-32))
        (from phys-cont (part ozizoosama-32))
        (to pbunri (part ozizoosama-32))))
      (mode)
      (con con-30)))
(loc
(scenario (totyuu-30 (part miti-29))))
(time (scenario nil) : 場所情報
(rel-time (after (event 12 con-28))
          (before (event 14 con-31))))
(infer : エピソードリンク
(backward (reason (link (event 12 con-28))))
(forward (achieve (link (data 15 con-26)))
          (initiate (link (data 26 con-49)))))
(category Act) : イベントの分類
(mod-conj sokode) : 接続詞情報
(mod-adv (sadv yasaki)) : 副詞情報
(con con-30) : 節番号
(real +) : 現実/非現実
(character (oziisan-3)) : 登場人物
(prop (yuki-19 te-40)) : 登場物
(relation) : 属性情報
(loc yuki-19 *g00116) (state yuki-19 pbunri))
(conceptual) : 概念ユニット情報
($pos-effect (initiate 1 main))))
    
```

図 2 “おじいさんはお地蔵様の雪を払ってあげました”
に対する入力イベント
Fig. 2 The input event for “OZIISAN (An old man)
took off YUKI (snow) from OZIZOOSAMA.”

```

(15                                     : データ番号
data                                     :
(goal)                                  : データ分類
(iml)                                    : CD 表現
(a-bunri ozizoosama-32 yuki-19 (con con-26))
(con con-26)                             : 節番号
(infer)                                  : エピソードリンク
(backward
(motive (link (event 11 con-25)))
(achieve (link (event 13 con-30))))))
    
```

図3 “お地蔵様は雪を取りたい”に対するデータイベント
Fig. 3 The data event for “OZIZOOSAMA wants to take off snow.”

①「こぶとりじいさん」
(1) 物語文章

昔々、ある所に良いおじいさんと悪いおじいさんがいました。良いおじいさんにも悪いおじいさんにも顔に悪いこぶがありました。ある日、良いおじいさんは山へ仕事に行きました。芝刈をしていると、雨が降ってきました。おじいさんは急いで大きな木の陰に入りました。夜になって、雨がやみました。その時、たくさんの鬼が現れました。おじいさんは鬼に見つかってしまいました。もしもおじいさんが餌をしなければ、鬼はおじいさんを食べるとおじいさんに言いました。しかし、おじいさんの餌はたいへん上手だったので、鬼はおじいさんの顔からこぶを取りました。鬼が悪いおじいさんの顔からこぶを取ったことを聞いた悪いおじいさんは、次の日、山に行きました。悪いおじいさんは鬼と一緒に餌をしました。おじいさんの餌はたいへん下手でした。鬼は、悪いおじいさんの顔から取ったこぶを悪いおじいさんの顔につけてしまいました。

(2) データイベントの内容

D-1 -MANは	鬼を取りたい	D-2 +MAN	鬼を取りたい
D-3 +MANは	雨を避ける	D-4 +MANが	仕事場へ行く
D-5 +MANが	芝刈をする		
D-6 +MANが	芝刈をする		
D-7 +MANは	芝刈をする		
D-8 +MANの	雨を避ける	D-9 +MANが	木に近づく
D-10 +MANは	自力で移動	D-11 +MANは	雨を待つ
D-12 +MANと	鬼と近づく	D-13 +MANは	鬼を見る
D-14 +MANが	鬼を怖がる		
D-15 +MANは	鬼を食べたい		
D-16 +MANは	鬼を食べたい		
D-17 鬼は	餌を取る	D-18 鬼の	餌のプラン
D-19 鬼は	餌を取る	D-20 鬼は	喜ぶ
D-21 鬼は	餌を取る		
D-22 +MANは	鬼を避ける		
D-23 -MANは	鬼を代行させたい		
D-24 -MANは	山へ行く	D-25 -MANは	自力で移動
D-26 鬼は	何をする	D-27 鬼の	餌のプラン
D-28 鬼は	怒る	D-29 鬼は	餌を付ける

* -MAN : 悪いおじいさん +MAN : 良いおじいさん

図4 「こぶとりじいさん」の入力文章とデータイベントおよびそれらのノード番号
Fig. 4 Input sentences, data events and their nodes in the EPN for “KOBUTORI ZIISAN (Two old men with a wen).”

して「こぶとりじいさん」と「大きなかぶ」とを取り上げ、「こぶとりじいさん」の全文とデータイベントを図4に、そのEPNを図5に示す。また「大きなかぶ」の全文とデータイベントを図8に、そのEPNを図9に示す。図中 E-n と記述しているノードは入力イベント、D-n と記述しているノードはデータイベントである。また、E-*. * と記述しているノード

は埋め込み文を持つイベント、もしくは埋め込み文をイベントにしたものを表している。しかし「～て」接続型の埋め込み文に関しては、前文と後文が時間的に同時に起こったものであれば E-n, E-n+1 とし、同時でなければ E-n, m, E-n.m+1 としている。点線で囲んだ部分はスクリプト、もしくは概念ユニット⁷⁾で、一まとまりの連鎖であると認識したノード群である。EPN 内には登場していないイベントもあるが、それらはリンクによる連鎖がないので省略したが、実際には、時間・空間・語と語の関係(概念関係)などで関連がある。

4. あらすじ生成

一般に物語は、登場人物などの行動の記述が始まるまでの設定部とそれ以後の展開部とに分けられ、本研究では物語理解は事象の連鎖関係の把握であり、EPN が理解の成果であると考えている。したがって、

あらすじ=物語設定部+キーセンテンスの連鎖と定義し、「抽出したキーセンテンス間においてもなんらかの連鎖関係を維持していなければ、あらすじには成り得ない」と考えている。したがって本稿におけるあらすじ生成は重要な連鎖関係を維持しながら、キーセンテンスに相当するノードをネットワーク内の全ノードの中から抽出する作業となり、その出力形式としては、通常、抽出したノードの文章の連鎖となる。

4.1 キーセンテンスの抽出

キーセンテンス抽出の方針を次のように定めた。

- (1) キーセンテンス抽出過程において、人の援助は必要としない。
- (2) キーセンテンス間でなんらかの連鎖関係を維持していなければならない。
- (3) キーセンテンスの数は、基となる文章の全節数の半数以下にする。一つの節が EPN の一つのノードに相当するため、対象となる単位は節である。ここで述べるキーセンテンス抽出の結果が元の文章の全節数の半数以下にならない場合、物語の形式に依存したキーセンテンス抽出を行う(4.3 節参照)。
- (4) キーセンテンスと成り得るイベントは入力イベントのみに限定する。キーセンテンス抽出過程ではデータイベントもキーセンテンス候補に含めている。しかしデータイベントはシステムが推論し

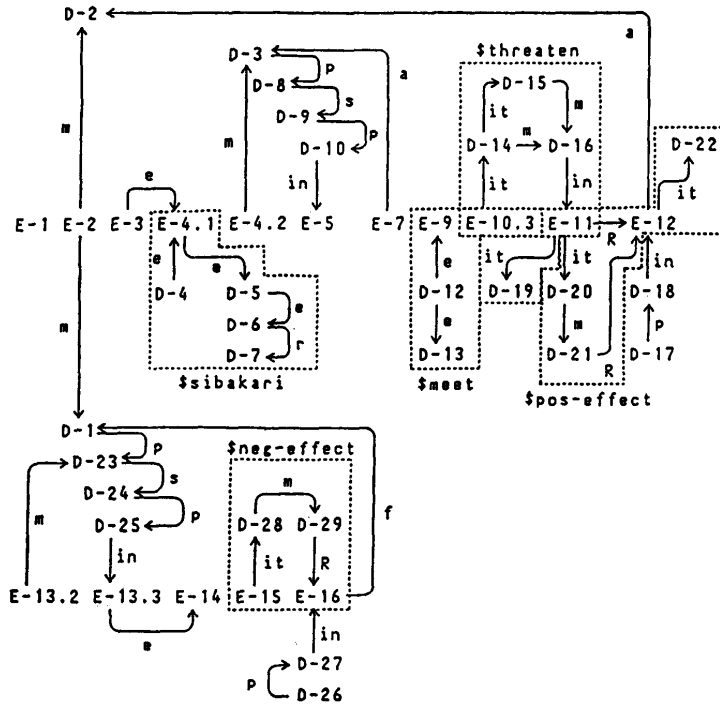


図5 「こぶとりじいさん」のエピソードネットワーク
Fig. 5 An episode network for "KOBUTORI_ZIISAN (Two old men with a wen)."

たイベントであり、システムが推論したことなら人にも推論可能であり、補完した文が本文より重要な役目を果たすことは通常考えられない。したがって最終的なキーセンテンスからはデータイベントを省略する。

以上の考えの下に、以下のようにしてキーセンテンスを抽出する。

第1段階：キーセンテンス抽出の基となるノードを抽出する。

出入りするリンクの本数が多いノードは、それだけ物語の内容と強く結びついていると考えられ、重要なノードであると判断される。

何本以上のリンクを持つノードを重要なノードであるとするのはシステムが決める。決定方法を以下に示す。

(四捨五入 (一つのノードに連鎖するリンクの本数の平均)) + 1

この本数以上のリンクを持つノードをキーセンテンス抽出の基となるノードとする。

第2段階：第1段階で抽出したノード間で連鎖関係があるものだけを残す。

(1) 抽出したノードの前向き連鎖だけをたどり、

抽出したノード内で連鎖関係があるものがないものでグループ分けする。

(2) グループ分けしたもののうち、設定部と連鎖関係があるグループを基となるノードとする。本研究ではあらずじを“設定部+キーセンテンス”にとらえているためである。もしも設定部と連鎖関係があるグループが存在しない場合は、グループ内で最大のリンク数を持つグループを基となるノードとする。

第3段階：基となるノードを補う、あるいはノード間を補間するノードを抽出。

(1) 基となるノードから in リンク (表1参照)で連鎖している入力イベントノードを候補に加える。

(2) 基となるノードから a, f, m, r, R リンク (表1参照)によって連鎖しているノードを候補に加

える。

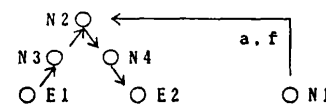
(3) 候補となっているノードのうち a, f リンクを持つノードについて、そのノードから前向きにリンクをたどり、最初に到達する入力イベントノードを候補に加える (図6参照)。

(4) スクリプト、概念ユニット内のシーンが候補になっておればスクリプトや概念ユニット内の要約を利用する。

例えば、概念ユニット \$POS-EFFECT⁷⁾ のシーンがキーセンテンスの候補に含まれているとする。

\$POS-EFFECT の内容は以下のようにになっている。

シーン1：行為者のなにかの状態が通常よりも良



N1 は基となるノード
N1 を前向きにたどり最初に到達する入力イベントノード E2 を候補に加える

図6 第2段階の(3)の対象となる連鎖
Fig. 6 The link used in the process 3 in the 2nd step.

い. もしくは上手だ. ある
いは, 行為者が相手にポジ
ティブなことをする.

シーン 2 : 相手は好印象を持つ.

シーン 3 : 相手は次のシーン 4 をしよ
うと思う.

シーン 4 : 相手は行為者にポジティブ
なことをする.

そして, \$ POS-EFFECT の LO-
CAL-SUMMARY スロットには (1,
4) が記述してある. これは, \$ POS-
EFFECT の局所的な要約としては
シーン 1 とシーン 4 を文章化すれば良
いということを表している. それ以外
のシーンは文章化しなくても推論可能
であるということである. そこで,
\$ POS-EFFECT 内の局所的要約は
「行為者が相手にポジティブなことを
した. (したがって) 相手は行為者に
ポジティブなことをした」となる. こ
のように, LOCAL-SUMMARY スロ
ット内のシーンを文章化すれば, その
概念ユニット内の因果関係を保つこと
ができる.

(5) キーセンズが物語設定文を含んで
おれば, 物語設定文をキーセンテ
ンスから除く.

(6) 展開部の最初の文がキーセンテ
ンス候補になっていないが, その文が言
及している場所, 時間を抽出したキー
センテンスで使用している場合には,
その文もキーセンテンスに加える.

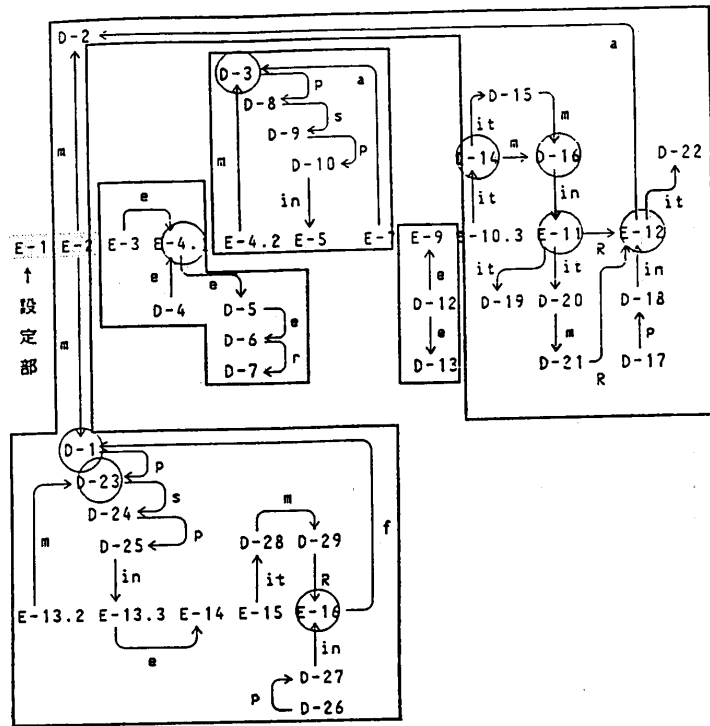
以上の結果から得られるノードをキーセ
ンテンスとし, ノードを出現時間順に並べ
直し, 設定部を加え (文献 7) 参照) 連続
文生成をしてあらすじを作り出す.

4.2 キーセンテンスの抽出例

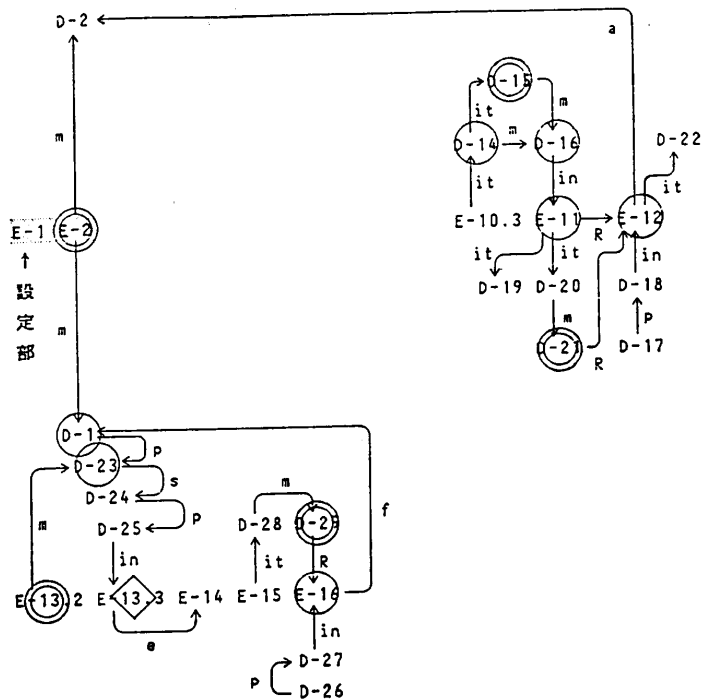
キーセンテンス抽出の具体的な例を「こ
ぶとりじいさん」を用いて説明する.

第 1 段階

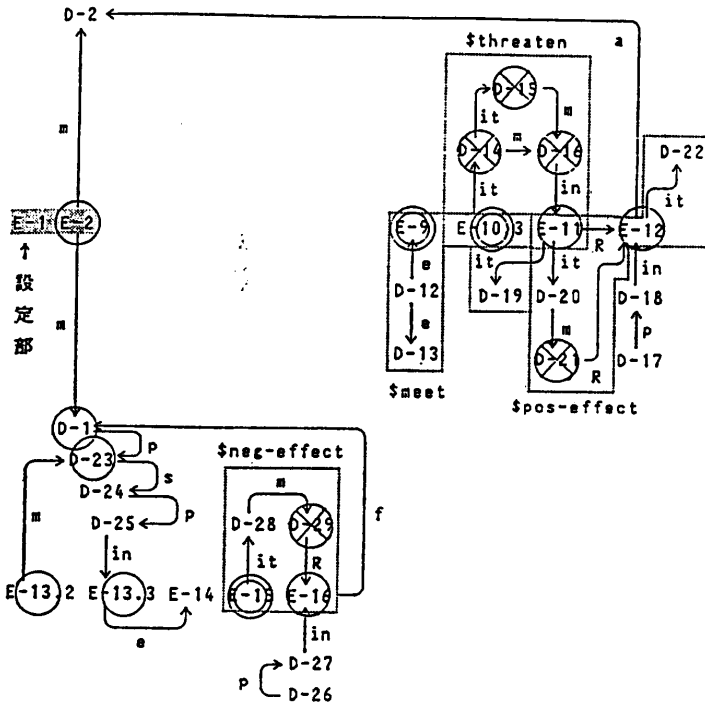
リンク数 3 のノードを抽出する. 抽出
ノードのリンク数は上述の方法で 3 本以
上となる. 図 7 (a) の○印ノードがリン
ク数 3 以上のノードである.



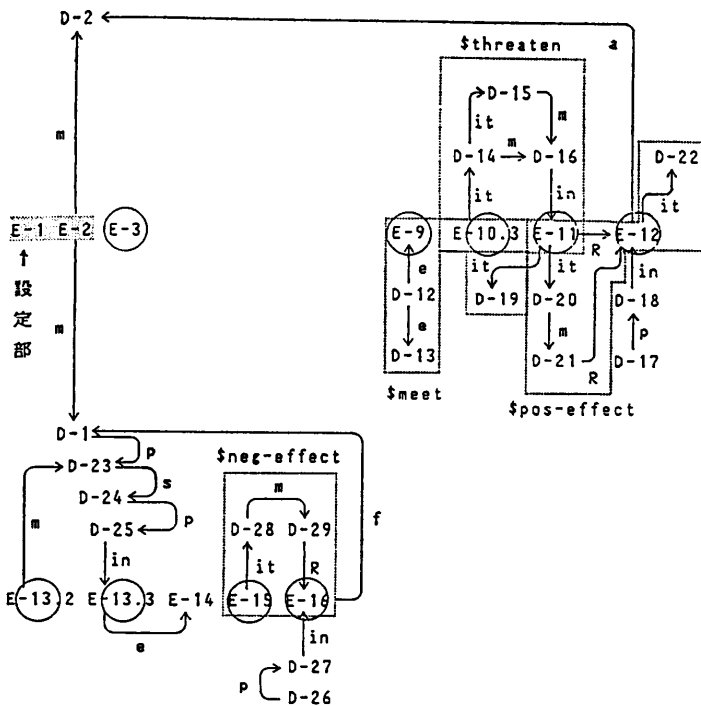
(a) 第 1, 第 2 段階終了時の EPN
(a) The EPN composed after the 1st and 2nd steps.



(b) 第 3 段階 (3) 終了時の EPN
(b) The EPN composed after the process 3.



(c) 第3段階(4)終了時の EPN
(c) The EPN composed after the process 4 in the 3rd step.



(d) 最終的なキーセンテンスを表すノード
(d) The nodes composing the final key sentences.

図7 “こぶとりじいさん” に対するエピソードネットワーク (Fig. 5) からのキーセンテンス抽出過程

Fig. 7 Some processes extracting key sentences from the episode network for “KOBUTORI ZII SAN” (Fig. 5).

第2段階

- (1) 第1段階で抽出したノードの連鎖関係に基づいてグループ分けを行う (図7(a)の太線で囲んだものが一つのグループ)。図では4グループ存在している。
- (2) 設定部と連鎖関係のあるグループが基になるノードを含むグループとなる。

第3段階

- (1) 図7(b)の○印ノードから in リンク (表1参照) で連鎖していて、まだ○印のついていないノードはないので、この処理の該当ノードはないことになる。
- (2) 図7(b)の○印ノードと a, f, m, r, R のうちのいずれかのリンクで連鎖しているノードを抽出する。図7(b)の◎印ノードがそのノードである。
- (3) 図7(b)の○印, および◎印ノード内で, a, f リンクを持つノードを前向きにたどり, 最初に到達する入力イベントを抽出する。図7(b)の◇印ノード (E-13.3) がそのノードである。
- (4) 図7(b)の○印, ◎印および◇印のノードが概念ユニット, あるいはスクリプトに含まれていれば, その中で要約を試みる。対象となる概念ユニットは \$MEET, \$THREATEN, \$POS-EFFECT, \$NEG-EFFECT である。E-9 は連鎖関係がないために最初は候補ではなかったが, \$MEET 内の局所的な要約を試みる時に, 候補ノードに加わる。この処理の結果, 抽出するノードは図7(c)の◎印ノードであり, 候補から削除するノードは ⊗印ノードである。
- (5) 候補からデータイベントノードと設定部であるノードを削除する。
- (6) 展開部の最初の文, つまり「ある日, 良いおじさんは山へ仕事に行きました」はキーセンテンスの候補になっていないが, 「山」をキーセンテンス

- ②「大きなかぶ」
 (1) 物語文章
 おじいさんがかぶの種をまきました。
 やがてそれは大きなかぶになりました。
 おじいさんはかぶを抜こうとしました。
 ところが、かぶは抜けませんでした。
 おじいさんはおばさんをお呼んできました。
 おばあさんがおじいさんを引っ張って、おじいさんがかぶを引っ張りました。
 それでもかぶは抜けませんでした。
 おばあさんは種をお呼んできました。
 種がおばあさんを引っ張って、おばあさんがおじいさんを引っ張って、おじいさんがかぶを引っ張りました。
 それでもかぶは抜けませんでした。
 種は犬をお呼んできました。
 犬が種を引っ張って、種がおばあさんを引っ張って、おばあさんがおじいさんを引っ張って、おじいさんがかぶを引っ張りました。やがて、かぶが抜けました。
- (2) データイベントの内容
 D-1 Zはかぶの種を持っている
 D-2 Zはかぶを抜きたい
 D-3 ZはBの協力を欲す
 D-4 Zはどこかに行きたい
 D-5 Zはかぶを抜きたい
 D-6 Zはどこかに行きたい
 D-7 Zは歩く
 D-8 Bはかぶに力をかける
 D-9 BはMの協力を欲す
 D-10 Zはかぶを抜きたい
 D-11 Bはどこかに行きたい
 D-12 Bは歩く
 D-13 Mはかぶに力をかける
 D-14 MはIの協力を欲す
 D-15 Iはかぶを抜きたい
 D-16 Mはどこかに行きたい
 D-17 Mは歩く
 D-18 Iはかぶに力をかける
 * Z:おじいさん B:おばあさん M:種 I:犬

図8 「大きなかぶ」の入力文章とデータイベントおよびそれらのノード番号

Fig. 8 Input sentences, data events and their nodes in the EPN for "OOKINA KABU (A large turnip)."

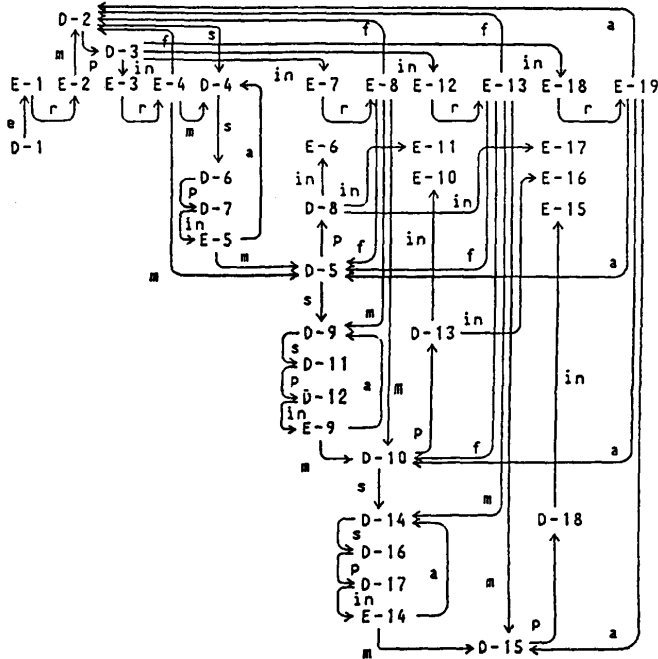


図9 「大きなかぶ」のエピソードネットワーク

Fig. 9 An episode network for "OOKINA KABU (A large turnip)."

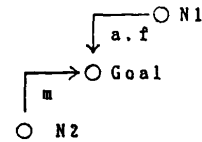


図10 (3)の処理対象となる連鎖
 Fig. 10 The link used in the process 3.

内で用いているので、この文もキーセンテンスに加える。以上の結果、キーセンテンスとなるのは図7(d)の○印ノードであり、連鎖関係も同図に示すように維持できている。全節数 21 に対し、キーセンテンス数9なのであらすじ生成の方針を満たしている。

物語設定部と抽出したキーセンテンスを連続文生成した結果を以下に示す。これが「こぶとりじいさん」のあらすじとなる（「舌切り雀」「笠地蔵」もこのキーセンテンス抽出法であらすじを生成できている）。

◇「こぶとりじいさん」のあらすじ:

顔に醜いこぶがある良いおじいさんと悪いおじいさんが、昔々、ある所にいました。

ある日、良いおじいさんは仕事に山へ行きました。良いおじいさんは鬼に見つかってしまいました。

もしも良いおじいさんが踊りをしなければ、鬼は良いおじいさんを食べると良いおじいさんに言いました。良いおじいさんの踊りは上手だったので、鬼は良いおじいさんの醜いこぶを取ってあげました。

そのことを聞いた悪いおじいさんは、次の日、山へ行きました。

悪いおじいさんの踊りは下手だったので、鬼は良いおじいさんの醜いこぶを悪いおじいさんの顔に付けてしまいました。

4.3 特別なキーセンテンス抽出

4.1 節で示したキーセンテンス抽出法では、抽出したキーセンテンスが基の文章の全節数の半分を越えてしまうことがある。これは「物語の形式によりあらすじ生成法を換えなくてはならない」ことを示している。例えば、全入力およびデータイベント文を図8に、その EPN を図9に示す「大きなかぶ」に、4.1 節のキーセンテンス抽出法を試みると、すべての入力イベントノードを抽出してしまう。「大きなかぶ」の文章が既にあらすじに近い（よけいな記述が少なく、短い）ということや、ほとんどのノードが a, f, m, r

リンクで連鎖していて、しかも同じリンクが繰り返し出現することが原因である。「大きなかぶ」のように何度か失敗を繰り返した後、ゴールを達成するような筋の物語用のキーセンテンス抽出法を述べる。

- (1) 4.1 節のキーセンテンス抽出法で抽出したノードが入力文章の全節数の半分以下でなければ、第2段階(1)の処理まで戻る。
- (2) ある基となるノードに複数のfリンクと一つのaリンクがあり、複数のfリンクの方が一つのaリンクよりも出現順序が早ければ、最初のfリンクよりも出現順序が前のノードと、複数のfリンクで直接連鎖している入力イベントノードと、一つのaリンクで直接連鎖している入力イベントノード以外は基となるノードの中から削除する。
- (3) 図10に示す連鎖があり、図中のノードN1が基となるノードに含まれていたなら、図中のノードN2を候補に加える。

この方法の具体例を示す。対象となる物語は「大きなかぶ」であり、EPNは図9である。

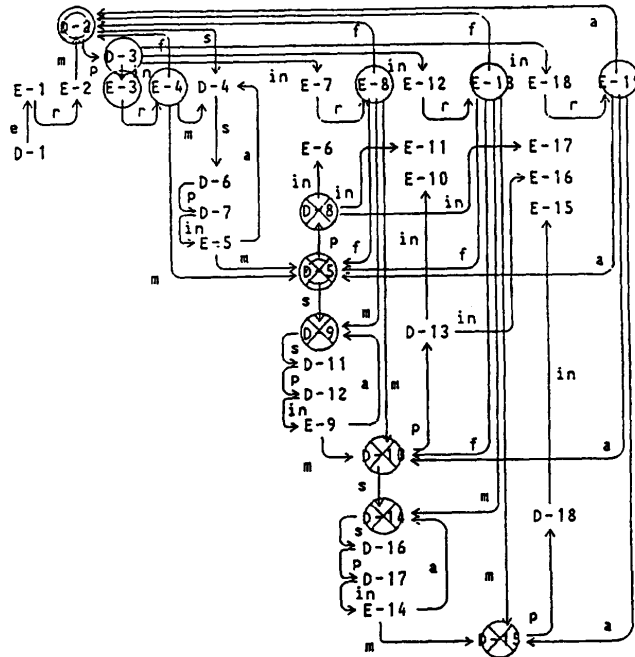
- (1) 第2段階(1)の処理まで戻ったときの基となるノードは図11(a)の○印ノードである(⊗, ⊙印を含む)。
- (2) 図11(a)内の⊙ノードが複数のfリンクと一つのaリンクを持つノードである。この処理で削除するノードは図11(a)の⊗印ノードである。
- (3) 図11(b)の太線で示したリンクが図10の連鎖に相当するものであり、図11(b)の⊙印ノードがこの処理で候補に加わるノードである。

最終的にキーセンテンスとなるのは図11(c)の○印ノードであり、連鎖関係は図に示すとおりである。

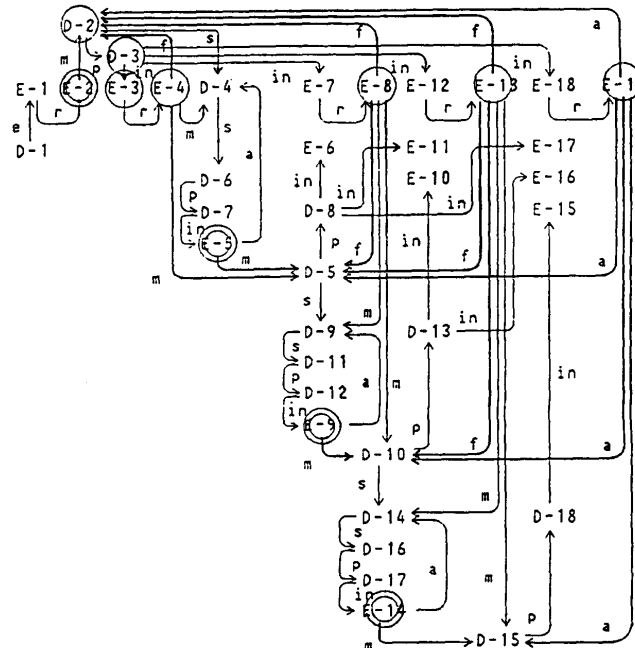
物語設定部と抽出したキーセンテンスを連続文生成した結果を以下に示す。これが「大きなかぶ」のあらすじとなる。

◇「大きなかぶ」のあらすじ:

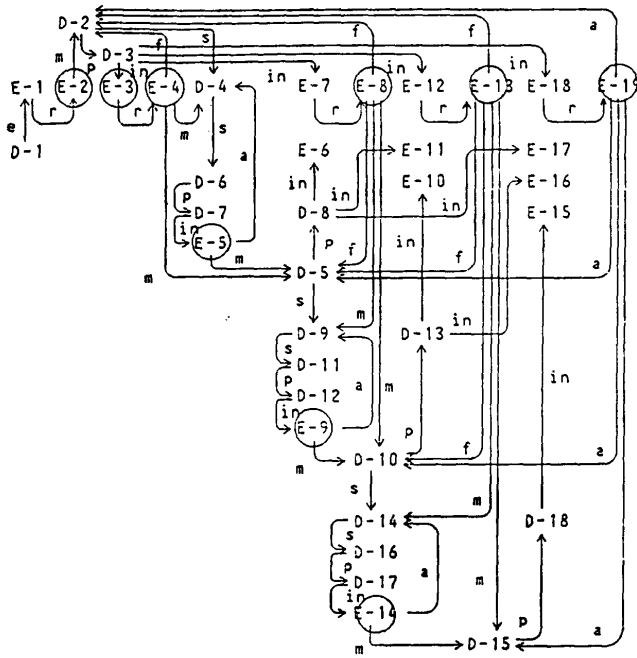
おじいさんがかぶの種を地面にまきました。
それは大きなかぶになりました。
それを抜こうとしましたがそれは抜けませんでした。



(a) 第2段階(2)の終了時のEPN
(a) The EPN composed after the process 2 in the 2nd step.



(b) 第2段階(3)の終了時のEPN
(b) The EPN composed after the process 3 in the 2nd step.



(c) 最終的なキーセンテンスとなるノード

(c) The nodes composing the final key sentences.

図 11 “大きなかぶ” に対するエピソードネットワーク (Fig. 10) からのキーセンテンス抽出過程

Fig. 11 Some processes extracting key sentences from the episode network for “OOKINA KABU” (Fig. 10).

そこで、おじいさんはおばあさんを選んで来ました。

しかし、大きなかぶは抜けませんでした。

そこで、おばあさんは孫を選んで来ました。

しかし、大きなかぶは抜けませんでした。

そこで、孫は犬を選んで来ました。

大きなかぶは抜けました。

5. 検 討

人のあらすじ生成と比較してシステムの評価をするため、大学生6人(質問(3)のみ8人)を被験者として、次の実験を行った。すなわち被験者にまず、配布した四つの物語を読んでもらい、(1)自由なあらすじの生成、(2)大切な文および部分の選択、(3)物語中で用いられている文を使用してのあらすじ生成を依頼した。その結果、(1)では意外にもほとんどの人が物語文章中の文をそのままつなぎ合わせたものをあらすじとしており、(3)による文章とほとんど同じであった。かつ文章の長さもほとんどのものが原文の半分を越えていた(本システムでは半分以下に抑えてい

る)。元の文章が子ども向きに簡略化されていることによるものと考えられる。(2)の結果から選ばれた文(節)に含まれている重要事項を因子と呼び、5人以上の人が選んだ因子を必須因子と名付け、(3)において6~8人の人があらすじに取り入れた因子を重要因子と名付けた。そして、これらの必須および重要因子を本システムが生成したあらすじが含み、極めて重要でない因子を含んでいないかどうかを調べたところ、どの物語に対しても、20~30個の因子のうち1,2の因子を除いてこの条件を満たしていた。したがって、裏付けとしては不十分ではあるが、本論文で取り上げた四つの昔話に対しては、ほぼ人間と同じように重要な文を取り上げてあらすじとしていえるであろう。満たされない主なものは、表層のみによるつながりしかない文が重要因子を含んでいる場合である。例えば「こぶとりじいさん」では、「その時、たくさんの鬼が現れました」という文は表層上のみによるつながりしかないため、EPN上には出てこない。本システムではこのような前後と密な関係のない文中の小設定は無視されるからである。しかし、この文をあらすじの中に取り入れている人が多くいた。今の場

合後続の文から容易に推論可能なので、この文がどうしても必要だとは考えられないが、一般的には、文中の小設定をどのような場合にどの程度考慮に入れるべきかは、今後検討する必要があるだろう。

なお、「特別なキーセンテンスの抽出法」を適用する必要があるかどうかは、構築されたEPNのノードの中に類似のa, f, m, r, Rリンクによる連鎖が何度も繰り返し続いている部分があるかどうかにより決まり、繰り返しがあつ場合には、その部分に局所的に適用することにより、あらすじ内の冗長性を軽減することができるであろう。

6. おわりに

これまで提案されてきたあらすじ(要約)処理でも確かに文中の重要箇所(キーセンテンス)を抽出することはできる。しかし、本研究で対象としているのは物語であり、物語のあらすじであれば話の内容を深く理解し、話の展開、キーセンテンス間の連鎖関係をも考慮にいれたキーセンテンス抽出を行わなければならない。そこで、EPNのリンクの種類に着目し、原因-

結果、理由-行動といった連鎖関係を維持したあらすじ生成を行った。その結果、対象が昔話にやや限定されているが、ここで取り上げた四つの昔話に対しては、人とほとんど同じキーセンテンスを抽出することができた。

この方法の利点は、大局的な要約が可能であり、連鎖関係に基づいており、かつ、スクリプトや概念ユニットなどの知識構造に着目した局所的な要約も取り入れているので、あらすじに飛躍が生じにくい点があり、十分目的を達し得ていると思われる。

しかしまだ文生成の能力が不十分で、生成されたあらすじがキーセンテンスの連鎖に近い場合が多い。今後の改善を必要とする。

参 考 文 献

- 1) 喜多壮太郎：説明文を要約するシステム，情報処理学会研究会報告，63-NL-6，pp. 41-48 (1987).
- 2) Schank, R. C. and Abelson, R.: *Scripts, Plans, Goals and Understanding*, Lawrence Erlbaum Associates (1977).
- 3) Lehnert, W.G.: Plot Units and Narrative Summarization, *Cognitive Science*, Vol. 5, pp. 293-331 (1981).
- 4) 内海功朗，重永 実：英語文章の大意生成，情報処理学会研究会報告，54-NL-8，pp. 57-64 (1987).
- 5) 安原 宏，小松英二，日比 孝，加藤安彦：要約支援システム COGITO，情報処理，Vol. 30, No. 10, pp. 1258-1267 (1989).
- 6) 野崎広志，重永 実：エピソードネットワークを用いた物語理解，情報処理学会知識工学と人工知能研究会資料，51-4，pp. 25-32 (1987).
- 7) 野崎広志，中澤俊哉，重永 実：物語理解におけるエピソードネットワークの構築，情報処理学会

論文誌，Vol. 30, No. 9, pp. 1103-1110 (1989).

- 8) 中澤俊哉，重永 実：「昔話」における物語理解，第36回情報処理学会全国大会論文集，5 T-1 (1988).
- 9) 木村岳男，重永 実：日本語物語理解システム JStory における省略された必須格の推定，1990年電子情報通信学会春季全国大会論文集，D-84 (1990).
- 10) 中澤俊哉，重永 実：エピソードネットワークを用いた物語のあらすじ生成，情報処理学会知識工学と人工知能研究会資料，90-69 (1990).

(平成2年4月24日受付)

(平成3年6月13日採録)

中澤 俊哉 (正会員)



1965年生。1988年山梨大学工学部計算機科学科卒業。同年同大学院工学研究科修士課程入学，物語理解の研究に従事。1990年修士課程修了。同年富士ゼロックス(株)入社，システム技術開発センタ勤務。

重永 実 (正会員)



1924年生。1947年京都大学工学部電気工学科卒業。金沢工業専門学校，金沢大学工学部を経て，1957年山梨大学工学部電気工学科，1971年計算機科学科，1990年中京大学情報科学部情報科学科勤務。工学博士。音声認識，物語理解，聴覚障害児用学習システムの開発等に興味を持つ。電子情報通信学会，日本音響学会等の会員。