1ZA-4

単語連想による筋道を用いた物語プロット自動生成システム

加賀谷卓宏　乾伸雄　小谷善行
東京農工大学　工学部　情報コミュニケーション工学科

１．はじめに

コンピュータに自動的に物語を作らせる場合、使用する単語の知識が必要となる。例えば、「私は病院で病気を治した」などの文を生成するには、「病院は場所、病気を治す場所」などの知識が必要になってくる。そこで、連想という知識表現を用いてこれを解決する。

本稿では、物語の流れを大局的に見ることのできるプロット(設計図)という出力表現を用い、連想という知識表現が一文の正確さ、物語の道筋の出来を向上できるかを示すことを目的とする。

２．知識辞書

本稿は三つの知識辞書を使う。それらを処理し、物語プロットを生成する。

２．１ 連想概念辞書

この辞書は、慶応義塾大学石崎研究室によって、被験者を用いて連想実験をし、その結果を電子化した辞書[1]である。この辞書に記述されている情報の内の「動作概念」「環境概念」「関連語」の三つの情報を利用する。例として以下がある。

動作概念：刺激語「病院」連想語「入院する」
環境概念：刺激語「病院」連想語「部屋」
関連語　：刺激語「病院」連想語「医者」

主に、システムが使用できる単語の管理、一文の評価に用いる。

２．２ 単語カテゴリー辞書

この辞書は、EDR 概念体系辞書と EDR 単語辞書[2]からデータを抽出し作成した。単語がどのカテゴリーの下位概念にあたるかを記したものである。本稿で利用する単語のカテゴリーは

人物、場所、人物を除く生き物、小道具、
飲食物、乗り物店、書物、情報、武器

の 10 種類である。

実際のデータ構造は「単語,カテゴリー番号」となる。(例：「学校,2」「ご飯,5」「噂,9」「包丁,10」)

２．３ 動詞辞書

この辞書は、動詞ごとに必要な情報を手動で記し辞書を作成した。この辞書には、本稿で使用する動詞と、その動詞ごとの深層格、各格で使用できる単語のカテゴリーが記してある。

本稿で使う動詞は 52 語であり、深層格は、「agent, object, beneficiary, implement, goal, place」の 6 種類である。(EDR の基準による)

文を生成する際の動詞による格支配に用いる。

３．物語プロット自動生成システム

本システムは 2 章で説明した三つの辞書を四つのモジュールで段階的に処理する。四つのモジュールは繰り返し動作させることによって、物語プロットの一文一文を順々に生成していく。本稿の出力の構造は図 1 のようになる。

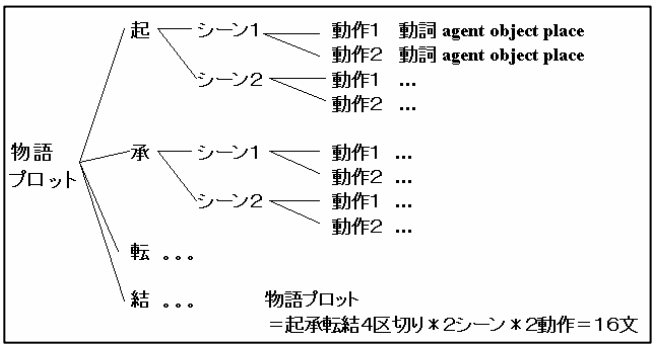


図 1　物語プロットの構造

３．１ 単語選択モジュール

単語選択モジュールでは、

- 基本単語の読み込み
- 連想概念辞書の刺激語から初期単語を選択
- 初期単語から連想概念辞書を参照し「環境概念」の単語を選択
- 前の文の単語から連想概念辞書を参照し「環境概念」「関連語」の単語を選択

のように、主に使用できる単語を増やす処理を行う。基本単語とは「主人公」といった物語を表現する上で最低限必要な単語のことである。

３．２ 単語分類モジュール

単語分類モジュールでは、単語選択モジュールにおいて選択された単語を、単語カテゴリー辞書を用いて、カテゴリーに分類する。

３．３ 単語間概念生成モジュール

単語間概念生成モジュールでは、単語に付属する設定を生成する。設定する項目は以下のとおりである。

- 主要人物の人数を設定
- 主要人物の現在地を設定
- 主要人物同士の交友関係を設定
- 主要人物の生死(病気)状態を設定

Automatic Plot Generator for Stories that use route by Word Association
Takuhiro Kagaya, Nobuo Inui, Yoshiyuki Kotani
Tokyo University of Agriculture and Technology

これらの設定はランダムで設定され、一文一文を評価する際のパラメーターとして利用される。また、前の文の動作に応じて、これらの設定が更新される。例えば、前の文が「動詞:行く agent 主人公,goal 病院」であれば主人公の現在地が病院に更新される。

3．4 プロット生成モジュールと評価関数

プロット生成モジュールでは、得た単語と設定を基に一文を生成する。文は図2のように動詞と深層格の組で表現される。文は一度に 52 個生成され(52 個は使用可能な動詞の語数である。が、各動詞ごとに文が生成されるわけではない),評価関数によって採点を行い、一番得点の高いものがその段階での一文に決定される。

文の生成方法は、以下の順である。

(1) 動詞辞書から動詞の決定
(2) 主要人物から agent の単語を決定
(3) object, beneficiary, implement, goal のうち必要な格の単語を決定
(4)place は agent の現在地が代入

評価関数は、以下の点を採点する。

- 前の文と評価中の文に連想概念辞書と同じ単語の組があるか
- 評価中の文の動詞と単語に連想概念辞書と同じ単語の組があるか
- 単語間概念生成モジュールの設定が使用されているか
- 動詞の連動が使用されているか
- 起承転結に応じた動詞か
- 途中プロットの動詞と評価中の文に連想概念辞書と同じ単語の組があるか。

途中プロットとは、予め承・転・結の最初の文を生成しておいた文のことである。

前の文との連想概念辞書による採点が、繰り返されることにより結果として道筋になる。16 文生成し一つの物語プロットとなる。

実際の実行結果では図 2 のような物語プロットが出力される。

```
区切り「起」　シーン1　動作1
動作:拾う agent:主人公 object:石 place:学校
区切り「起」　シーン1　動作2
動作:行く agent:主人公 implement:徒歩 goal:山 place:学校
                    <中略>
区切り「結」　シーン1　動作1
動作:壊す agent:主人公 object:石 place:道
区切り「結」　シーン1　動作2
動作:行く agent:主人公 implement:徒歩 goal:海 place:道
```

図2 出力された物語プロットの一つ

4．評価

評価は、出力された物語プロットをアンケートによって評価してもらった。物語プロットは図2のような深層格による表現ではなく、格に応じた助詞を補った文体とした。

評価項目は、次のとおりである。

(1)一文だけ見た場合の適合性・創造性(1~5 点)
(2)前の文との流れの適合性・創造性(1~5 点)
(3)全体を見て物語が思い浮かぶか(1~5 点)
(4)部分的に見て良い流れはあるか(文番号)

(1)と(2)は各文について回答してもらった。

適合性とは、回答者から見て、常識的に考えて動詞と単語、単語と単語の組み合わせが正しいか。創造性とは物語的に考えて面白いと思う単語の使われ方がされているか、とした。

アンケートには、システムが出力した四つの物語プロットの他に自作で作った物語プロットを一つ、ランダムで作られた物語プロットを一つ、用意した。ランダムで作られたプロットでは、評価関数による採点と連想による使用単語増加が起こらず、初期単語を 100 としたものである。(動詞辞書による格支配は考慮する)

評価結果は、表1のとおりであった。

表1 物語プロットの評価の各平均点

	一文		前文			
	適合	創造	適合	創造	全体	優良部
物語1	**3.43**	2.79	3.06	2.90	2.30	35.4%
物語2	3.09	**3.19**	2.95	**3.19**	2.63	36.6%
物語3	2.74	3.01	2.64	2.83	2.40	32.8%
物語4	3.19	3.14	**3.09**	3.08	**3.00**	—
自作	4.14	2.88	3.48	3.22	4.22	61.1%
ランダム	2.86	3.03	2.47	2.99	2.17	30.0%

5．考察

本システムが出力したプロットの各最高値はランダムを上回る結果となった。自作と比べると、下回っている物語が多いが、近い値や上回っている数値もある。創造性に関しては、システムによる曖昧さが斬新とも見て取られ、高い評価が得られた。適合性の部分は、動詞辞書による影響が強いため、ランダムに対して下回る結果となるものもあった。一文、前文という尺度になると自作に近い評価が得られることもあるが、全体や、連続した部分となると、自作と比べかなり低い評価となった。これは、道筋を前文からの連想の連続と考え、評価関数を設計したのが影響している。

6．おわりに

連想辞書という知識を導入し、一文の正確さや、物語の道筋を向上できることが解った。

＜参考文献＞

[1]岡本　潤 石崎　俊:「概念辞書の構築と概念空間の定量化　連想実験による概念空間の抽出」、2001、情報処理学会研究報告 自然言語処理 No.130 - 011
[2]EDR 電子化辞書：
http://www.jsa.co.jp/EDR/J_index.html