

小説群からの動詞をベースとした事象連鎖の獲得と利用

Acquiring and Using Event Sequences Based on Verbs from Novels

荒井達也*¹
Tatsuya Arai

小方孝*¹
Takashi Ogata

*¹ 岩手県立大学ソフトウェア情報学部

Faculty of Software and Information Science, Iwate Prefectural University

An Integrated Narrative Generation System (INGS), which have been developing by the authors, needs to use scripts as semantic event sequences for the story generation. Further, the authors aim at automatic acquisition and composition of generation knowledge including scripts. For acquiring scripts automatically, this paper proposes a method to acquire two verb sequences using the simplest bigram and insert necessary elements into the case structure of each of the acquired verbs using a semi-automatic script composition tool presented by the authors.

1. はじめに

筆者らが研究・開発を行っている統合物語生成システム (Integrated Narrative Generation System: INGS)[Ogata 16]において、ストーリー生成を行う際に利用する知識として、意味的なまとまりを持つ事象の連なりである事象連鎖を利用する。これまで、筆者らはスクリプト型事象連鎖の獲得を行ってきたが、本稿では、N-gram[天野 07]を利用して、上のスクリプト型事象連鎖と異なる既存の小説をベースとした事象連鎖の獲得、獲得した事象連鎖を INGS におけるストーリー生成の際に利用する方法を示す。今回は試験的に、2つの事象からなる事象連鎖を bigram から獲得する。事象連鎖獲得の手続きは、動詞 bigram の抽出と動詞 bigram の格構造化の2つに分かれる。以降、2節では、INGS における事象連鎖の説明を行い、続く3節と4節で動詞 bigram の作成と動詞 bigram の格構造化を行う。最後に、5節で動詞 bigram に格構造を挿入した事象連鎖を用いたストーリー生成を行う。

2. INGS における事象連鎖

INGS はストーリー生成を行うために様々な知識ベースを持ち、中でも本稿は、動詞概念辞書と事象連鎖知識と特に関連する動詞概念辞書は現在、動詞概念を 11,951 個格納している。多くの動詞概念は、同一の表記を持つが異なる意味及び格構造を持つため、例えば「食べる 1」、「食べる 2」のように、番号により意味を区別している。同一表記による重複を除いた場合、5,000 個の単純な動詞によって辞書は構成される。

INGS における事象連鎖は、意味的なまとまりをもつ事象の連なりであり、それぞれの事象は動詞概念辞書中の動詞概念によって記述される。これまで筆者らは、3 から 10 個の事象を含むスクリプト型事象連鎖[荒井 17]の自動獲得を試みたが、ここで獲得する事象連鎖は 2 個の事象のみから成る。

3. 動詞 bigram の作成

まず、青空文庫中の小説群から動詞 bigram の作成を行う。そのために、小説に形態素解析を行い、動詞を抽出する。青空文庫中の 13,331 個の小説群[青空文庫 98]に対し、小説ごとに形態素解析を行い、動詞の抽出を行った。抽出した動詞群から作成した bigram を動詞 bigram と呼ぶ。総計 3,041,450 種類の

動詞 bigram が作成された。

作成した動詞 bigram は、動詞概念辞書に未格納の動詞を含む。INGS において使用可能とするために、作成した動詞 bigram に対し、動詞概念辞書中の単純な動詞 5,000 個からなる動詞 bigram の抽出を行う。本稿では、動詞 bigram からの事象連鎖の獲得と利用を試験的に行うために、動詞概念辞書中の単純な動詞 5,000 個に対し、『日本語彙体系』[池原 99]中の他動詞を用い、動詞の選別を行った。動詞概念辞書中の他動詞 1,174 個に基づいて動詞 bigram の抽出を行い、130,810 種の動詞 bigram を抽出した。表 1 に、出現頻度が高い順に 10 個、出現頻度が低い順に 10 個の動詞 bigram をそれぞれ示す。

表 1 抽出した動詞 bigram の例

| 動詞 bigram 上位 10 個 | 出現頻度 | 動詞 bigram 下位 10 個 | 出現頻度 |
|----------------------|------|----------------------|------|
| (見る 見る) | 7199 | (間引く 殺す) | 1 |
| (言う 言う) | 3983 | (払い戻す くれる) | 1 |
| (見る 思う) | 3464 | (炒める 言う) | 1 |
| (知る 知る) | 3214 | (使い果たす 構う) | 1 |
| (思う 見る) | 2434 | (暈す 見せる) | 1 |
| (思う 思う) | 2429 | (割り付ける 引く) | 1 |
| (考える 見る) | 1995 | (競り合う 申す) | 1 |
| (しまう 思う) | 1892 | (絞り込む 図る) | 1 |
| (書く 書く) | 1778 | (貸し出す くれる) | 1 |
| (見る 言う) | 1732 | (繰り広げる 持つ) | 1 |

4. 動詞 bigram の格構造化

3 節で抽出した動詞 bigram を構成する 2 つの動詞に対し、INGS で利用可能とするために、動詞の意味を決定し、動詞 bigram に格構造を挿入する作業を行う。以下、作業手続き並びに動詞 bigram を格構造化した際に利用したツールについて説明する。

4.1 動詞 bigram を含む文の抽出

3 節で獲得した動詞 bigram を事象連鎖とするために、動詞 bigram 中の動詞の意味を決定する必要がある。ここでは、動詞 bigram を構成する 2 つの動詞が実際の小説中でどのように使われているかを知るために、動詞 bigram の抽出元となった文を抽出する。今回は、筆者が比較的動詞の格構造化が容易であると判断した他動詞から構成される 20 種類の動詞 bigram を選択した。動詞 bigram の抽出元となった文を小説から抽出する際、今回は、1 つの小説から 20 種類の動詞 bigram の格構造

連絡先: 荒井達也, 岩手県立大学ソフトウェア情報学部, 岩手県滝沢市菓子 152-52, g0311011@s.iwate-pu.ac.jp

化を行う。今回は、動詞 bigram の種類を 1 番多く含む宮本百合子著『道標』と、『道標』に含まれない動詞 bigram を含む 2 番目に動詞 bigram の種類が多い横光利一著『旅愁』の 2 作品を用いた。表 2 に、選択した動詞 bigram とその出現頻度、これらの動詞 bigram を含む実際の文の対応を示す。

表 2 動詞 bigram を含む小説中の文

| 小説名 | 動詞 bigram | 出現頻度 | 小説中の文 |
|------|-----------|------|--|
| 『道標』 | (考える 見る) | 1995 | いろいろの場合を考えて見るように云った。 |
| | (見る 知る) | 1446 | 素子は、伸子が見たのを知りながら、吉之助の上においている視線を動かさなかった。 |
| | (言う 思う) | 1302 | という風に彼女は言った。思えば、そうなはずだった。 |
| | (知る 思う) | 1097 | あなたがたの生活を知りたいと思っています」と云おうとした。 |
| | (思う 考える) | 960 | そう思って考えてみると、 |
| | (見る 書く) | 832 | 服装のその記者を見た。いつか、どこかで比田礼二という名のひとが小市民というものについて書いている文章をよんだ記憶があった。 |
| | (見る 感じる) | 695 | 明け暮れの中で見て感じているソヴェト青年の二十歳の人生の内容からみると、 |
| | (読む 書く) | 358 | 伸子の読み書き能力は、非常に劣っていた。 |
| | (買う くれる) | 319 | 泰造が饅頭別を買ってくれたものでもあった。 |
| | (描く 見る) | 245 | 化学工場の内部を描いた画を見ていたが、何か考えが湧いたらしく、 |
| | (見る 描く) | 237 | しているのを見て、げんこをにぎっているゴリキイ。それらはみんな一九〇〇年頃の漫画であった。「プロジェクトル」のゴリキイ特輯号のために新しく描かれた漫画で |
| | (知る 書く) | 229 | 力をはたらかすことさえ知らないのでしょうか」書いて自分の脇で、紅絹(もみ)の針さしを床におとしてしまったのにも心づかないで、 |
| | (食べる 飲む) | 150 | 同量にたっぷり食べ、また、たっぷり飲んだことだったろう。 |
| | (着る 見る) | 144 | 「あなた、ワンピース着て見る気はない？」 |
| | (向ける 見る) | 131 | 笑顔を伸子に向けて、リン博士は、「あなた、孫逸仙大学の女学生たちを見ましたか？」ときいた。 |
| 『旅愁』 | (見る 笑う) | 816 | 千鶴子は二人を見ると、にっこりと笑い懐しそくに近よって来た。 |
| | (言う 笑う) | 783 | だなんて、そう仰言ったわ。」と千鶴子は多少からかい気味に笑った。 |
| | (見る 読む) | 297 | 見ると、カソリックの若い僧侶が椅子にかけて聖書を一心に読み初めた。 |
| | (殺す 思う) | 218 | この女性をどうとう責め殺した。」矢代はまたそう思うと、一層涙がとまらなそうになり、 |
| | (見る 作る) | 198 | 大きな千鶴子の眼を見ながら、矢代は笑顔を軽く作って下へ沈んでいった。 |

4.2 スクリプト半自動構築ツールを利用した動詞 bigram への格構造の挿入

スクリプト半自動構築ツール[荒井 16]は、INGS における事象連鎖知識の 1 つであるスクリプト知識の構築を補助するツールである。ユーザはスクリプトを入力し、スクリプトを構成する事象列中の動詞概念の意味を決定することで事象列に格構造を挿入することができる。この機能を利用し、作成した動詞 bigram を構成する 2 つの動詞の格構造を決定し、事象列に格構造を挿入する。

4.1 節で抽出した 20 個の動詞 bigram を含む小説中の文を実際にユーザに提示し、スクリプト半自動構築ツールを用いて、動詞 bigram を構成する 2 つの動詞の格構造を決定してもらった。例として、小説中の文「同量にたっぷり食べ、また、たっぷり飲んだことだったろう。」をユーザに提示し、動詞 bigram「食べる 飲む」を構成する 2 つの動詞「食べる」と「飲む」の意味をユーザが決定する手順を図 1 に示す。動詞 bigram に格構造を挿入した事象列を図 2 に示す。今回、図 2 で示した格構造において、動詞概念「食べる 2」と「飲む 2」の動作主である agent 格は、図 1 の動詞 bigram を含む小説中の文から分かるように、どちらも同じであるから「age1」を agent 格に挿入している。動詞 bigram を含む小説中の文において、動詞の動作主が異なる場合は、agent 格中の動作主「age1」をユーザが手入力により動作主「age2」に変更する必要がある。

「同量にたっぷり食べ、また、たっぷり飲んだことだったろう。」中の動詞 bigram「食べる 飲む」の意味を決定してください。

「食べる」の意味を決定してください。
1: コメンテーターが原色版で食べる
2: 客がしらすぼしを食べる
番号を選択> 2

「飲む」の意味を決定してください。
1: 渦が下枝を飲む
2: 原人がアブサンを飲む
<<中略>>
番号を選択> 2

図 1 ユーザによる格構造の決定

```
(event 食べる 2 (agent (&v age1)) (counter-agent (&v age2)) (location (&v loc1)))
(event 飲む 2 (agent (&v age1)) (location (&v loc1)) (object (&v obj1)))
```

図 2 格構造を挿入した動詞 bigram

5. ストーリー木としての動詞 bigram の利用

格構造を挿入した動詞 bigram を INGS にて実際に利用する方法を示す。INGS では、ストーリーを事象と事象どうしの関係からなるストーリー木で表しており、ストーリー生成を行うことは、ストーリー木の拡張を行うことと等しい。

ストーリー木の拡張に、ストーリー技法を利用する。ストーリー技法は、ストーリー木の任意の事象を入力とし、ある部分木を出力する。技法毎に生成される部分木の構造は決まっており、その具体的な内容は対応する知識ベースに格納されている。図 3 では、ストーリー木 S を展開する事象 E2 に、「結果」技法により生成された事象 E2、E3 からなる部分木が、置換されることで、ストーリー木が拡張されている。拡張されたストーリー木は S' となる。

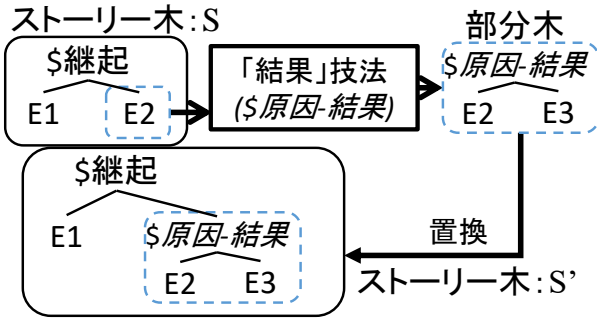


図3 拡張前後のストーリー木の記述

| | |
|-----|--|
| 拡張前 | <pre> (\$継起 (event 食べる 2 (type action) (ID 32) (time nil) (agent age%どら息子#1) (counter-agent age%コーチン#1) (location loc%砂防ダム#1)) (event 飲む 2 (type action) (ID 33) (time nil) (agent age%どら息子#1) (counter-agent nil) (location loc%砂防ダム#1) (object obj%斗酒#1))) </pre> |
| 拡張後 | <pre> (\$継起 (\$原因-結果 (event 空く 1 (type action) (ID 309) (time nil) (agent age%どら息子#1) (location loc%砂防ダム#1)) (\$原因-結果 (event 食べる 2 (type action) (ID 32) (time nil) (agent age%どら息子#1) (counter-agent age%コーチン#1) (location loc%砂防ダム#1)) (event 成る 5 (type action) (ID 304) (time nil) (agent age%どら息子#1) (counter-agent nil) (location loc%シネマテーク#1)))) (\$原因-結果 (event 飲む 2 (type action) (ID 33) (time nil) (agent age%どら息子#1) (counter-agent nil) (location loc%砂防ダム#1) (object obj%斗酒#1)) (\$原因-結果 (event 酩酊する 1 (type action) (ID 307) (time nil) (agent age%どら息子#1) (counter-agent nil) (location loc%貧民窟#1)) (event 放浪する 1 (type action) (ID 308) (time nil) (agent age%どら息子#1) (location loc%貧民窟#1))) (event 寝る 3 (type action) (ID 311) (time nil) (agent age%どら息子#1) (counter-agent nil) (location loc%貧民窟#1))))) </pre> |

図4 拡張前後のストーリー木の記述

4 節で格構造化した動詞 bigram「食べる 2 飲む 2」に対し、\$継起の関係を用いてストーリー木を生成した。図4にストーリー木の拡張例を示す。図4中の太字部分は動詞 bigram を構成する2つの事象部である。図4の上に、拡張前のストーリー木を、拡張後のストーリー木を図4の下に示す。図4の拡張前のストーリー木は「どら息子がコーチンを食べる。どら息子が斗酒を飲む。」という流れになり、拡張後のストーリー木は、「どら息

子がお腹が空く。どら息子がコーチンを食べる。どら息子が満腹に成る。どら息子が斗酒を飲む。どら息子が酩酊する。どら息子が貧民窟で放浪する。どら息子が寝る。」というより具体化された流れになる。

6. おわりに

本稿では、事象連鎖の最小の単位である2つの事象からなる事象連鎖の獲得を小説中の動詞 bigram から行い、獲得した動詞 bigram 中の20個の動詞 bigram に対し、試験的に、動詞 bigram を含む小説中の文を参考に、格構造化を行った。さらに、格構造化した動詞 bigram からストーリー生成を行った。具体的には以下のことを行った—

- ① 2つの他動詞で構成される動詞 bigram を抽出した。
- ② 20個の動詞 bigram を含む文を2つの小説から抽出した。
- ③ 獲得した事象連鎖をストーリー木としストーリーを生成した。

①については、今後、自動詞、他動詞を考慮しない動詞 bigram からの幅広い事象連鎖の獲得を行う。次に②について、抽出した文において、動詞の否定形が使われていることや、名詞的に動詞が使われていることがあった。また、事象連鎖とは言えない単なる動詞の繋がりも多く存在した。今後、動詞 bigram の格構造化を行う際、ユーザに複数の動詞 bigram を含む小説中の文を提示する等の対策が考えられる。また、本研究の目的は文の正確な分析ではないので、得られた結果をもとに自由にスクリプト化することも考えられる。最後の③において、獲得した事象連鎖からストーリー生成を行ったが、今後、ストーリー木以外への事象連鎖の利用を考案する。

今後は本稿の方法を進め、扱えるデータ量を増やしつつ自動化の程度を高め、実際の物語生成に自由・柔軟に使用できるようにする。

参考文献

- [青空文庫 98] 青空文庫, <http://www.aozora.gr.jp> (1998). (最終アクセス日 2016年11月)
- [天野 07] 天野真家, 石崎俊, 宇津呂武仁, 成田真澄, 福本淳一: IT Text 自然言語処理, オーム社, p.12 (2007).
- [荒井 16] 荒井達也, 小野淳平, 小方孝: 物語生成のための事象連鎖知識の半自動生成—統合物語生成システムにおける利用—, 2016年度人工知能学会全国大会(第30回)論文集 (2016).
- [荒井 17] 荒井達也, 小野淳平, 小方孝: 統合物語生成システムにおけるスクリプト型事象連鎖の獲得と利用, 言語処理学会第23回年次大会 (2017). (印刷中)
- [池原 99] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦: 日本語語彙大系 CD-ROM 版, 岩波書店 (1999).
- [Ogata 16] Ogata, T.: Computational and Cognitive Approaches to Narratology from the Perspective of Narrative Generation in Ogata, T., and Akimoto, T. (Eds.), Computational and Cognitive Approaches to Narratology, IGI Global, pp.1-74 (2016).