

Verb Concept Selection using Co-occurrence Information of Verb Concepts: A Mechanism in an Integrated Narrative Generation System

Jumpei Ono^a, Takashi Ogata^b

^a Graduate School of Software and Information Science, Iwate Prefectural University, Japan

^b Faculty of Software and Information Science, Iwate Prefectural University, Japan

^a g236m001@s.iwate-pu.ac.jp, ^b t-ogata@iwate-pu.ac.jp

Abstract

An integrated narrative generation system, which we develop, generates an event concept by referring to noun and verb concepts than a conceptual dictionary. However, the noun and verb concepts were selected at random. Considering the choice of the noun concept, we used the frequency information and co-occurrence information. Furthermore, we implement a similar method considering the verb concept selection. The method of the verb concept selection based on frequency information assumed in this paper premises that the appearance frequency of all concepts is prepared for. When it is extreme that all the candidates of a certain verb concept are of frequency 0, the concept of selection based on the frequency is in good agreement with the conventional random selection without functioning. Therefore, we suggest a method using the co-occurrence information between the concepts. We demonstrate a calculation result of the frequency information using the co-occurrence information in the verb concept when the frequency information becomes 0, in particular in the calculation method of the frequency information.

Keywords: Conceptual dictionary, Verb concept, Integrated narrative generation system.

1. Introduction

The authors have developed a narrative generation system known as an integrated narrative generation system (INGS)⁽¹⁻³⁾. Fig.1 shows the overview of the INGS.

The INGS generates narrative events and sentences according to two types of macro level parts i.e., generation mechanism and knowledge mechanism. The former is

divided into three main elements that include the story generation mechanism, discourse mechanism, and surface generation mechanism. The main elements of the latter are conceptual dictionaries (for noun concepts and verb concepts chiefly), language notation (letter notation) dictionary, and state-event transformation knowledge base. This paper is mainly related to the story generation mechanism and verb conceptual dictionary. The most important element of a story structure generated by the story generation mechanism is an event that consists of the verb concept and the corresponding noun concept. However, the noun and verb concepts were selected at random. Furthermore, we implement a similar method for the verb concept selection. The method of the verb concept selection is based on frequency information.

2. The Architecture of an Integrated Narrative Generation System

The INGS integrates various mechanisms for narrative generation that we have been suggesting and developing⁽¹⁻³⁾. The entire overview is shown in Fig. 1. The system consists of a concept generation mechanism to generate a story and edit a structure of narration (discourse). In addition, it consists of a surface representation mechanism to represent a narrative by letter, sound, and picture. The INGS requires the conceptual dictionary knowledge (noun concept, verb concept, etc.) for constructing an event concept that refers to an occurrence in the narrative, as well as story content knowledge (causal relationship, script, etc.) for constituting the structure of the narrative. The conceptual dictionary⁽⁴⁾, language notation dictionary⁽⁵⁾, story content knowledge base (Store knowledge show each relationship between events such as causal relationship and script)⁽⁶⁾, etc. are

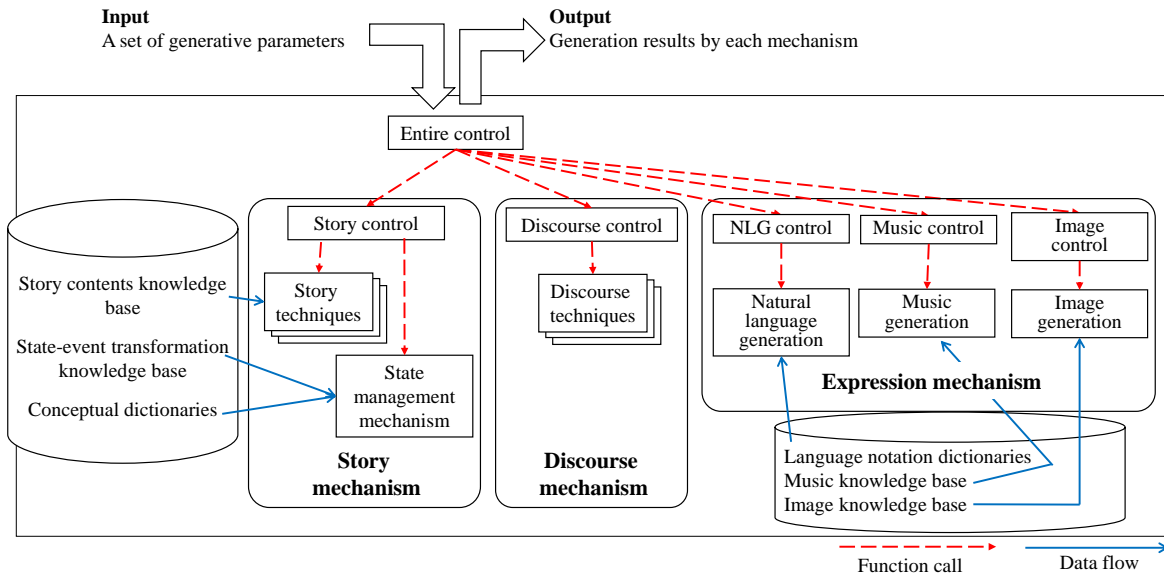


Fig. 1. Overview of the integrated narrative generation system

used to obtain the required information for each generation mechanism. Moreover, each event in the conceptual structure is connected states in front and behind using the state-event transformation knowledge base⁽³⁾.

3. Conceptual Dictionaries

An event is generated with the instantiation of conceptual materials stored in conceptual dictionaries for noun and verb concepts. Each dictionary has a hierarchical structure, from higher to lower concepts. The noun dictionary currently contains 115,765 terminal concepts and 5809 intermediate concepts. Each intermediate concept has (1) a list of hyponymy concepts, (2) the number of depth in the hierarchy, (3) the serial number of the super-ordinate concepts, and (4) the range of serial numbers of the hyponymy concepts. Fig. 2 shows the description of a noun concept.

On the other hand, the verb concepts hierarchy has 11,951 terminal concepts and 36 intermediate categories. Each terminal verb concept describes the following three elements: (1) a “sentence-pattern” for a sentence including the verb, (2) one or more “case-frame(s)” to show the types of required noun cases, and (3) one or more “constraint(s)” to define the range in the noun conceptual dictionary, in

```
(男 [Man]
(hierarchy
(depth 8)
(hype 人間<男女> [Human<Gender>])
(hypo 男 [接辞] [Man[Affix]])
(terminal 男 [Man] 父 [Father] 少年 [Boy]...))
```

Fig. 2. The description of a noun concept

which each noun concept in the above case-frame(s) requires. For example, Fig. 3 shows the description of the verb concept, “eat”. When the INGS materializes the framework of a case-frame, it selects concepts in the noun conceptual dictionary. The objective of the paper is to revise the selecting method.

A story generation process involves expanding or transforming a story structure. In particular, under some parameters (“macro-structure”, “length”, etc.) given in the first step, a story technique generates a new event or sub-structure including one or more new events using a variety of story content knowledge to integrate them into the original structure using various relations. According to a generated new event, new states are also generated. When an event is generated, the above story content knowledge gives the basic form based on the description of a case-frame shown in Fig. 3. Each constraint indicates the range in the noun conceptual dictionary to decide a noun concept. Essentially, each noun concept in an event is transformed into an instance using the attribute information if necessary. Finally, each generated event is transformed into a natural language sentence using specific word

```
((name 食べる 2 [eat])
(sentence-pattern "N1 が N2 を食べる" [N1 eat N2])
(case-cons-set
((case-frame ((agent N1) (object N2) ...))
(constraint
((人 [human] -死人 [corpse] -人間 <人称>
[human<person>]-準人間
[semi-human])
(食料 [food] -調味料 [flavoring] -飲物・たばこ
[drink・cigarette])))) ...))
```

Fig. 3. The description of the verb concept,

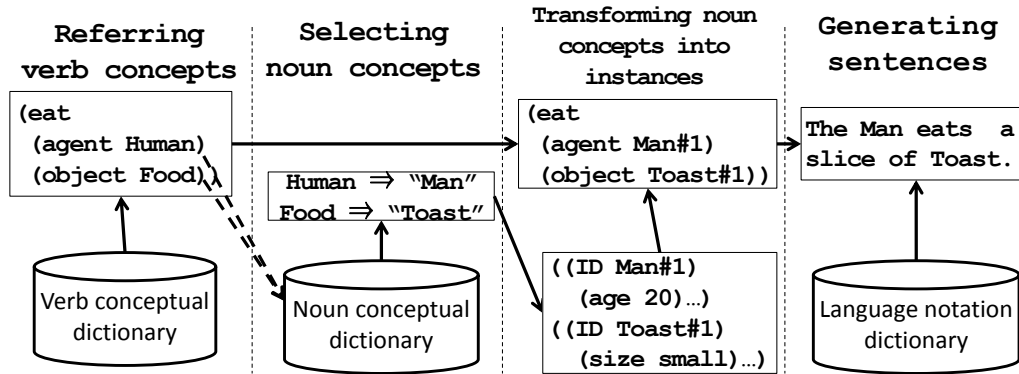


Fig. 4. The process of generating an event

representation. Fig. 4 shows the process of generating an event.

4. Acquiring the Frequency Information and the Co-occurrence Information

We have analyzed the word frequency in 4980 texts (mainly novels) in Aozora Bunko from 1872 through 1963 to use it for verb concept selection in story generation. The image of the process is shown in Fig. 5. KH Coder⁽⁷⁾ is used in the analysis in the first step to calculate the word frequency. In addition, the mechanism links the acquired frequency information to the co-occurrence information

4.1 Acquiring the Frequency Information

Each terminal concept in the last place of the hierarchy supports the vocabulary item in the language notation dictionary. We let the terminal concept in the verb conceptual dictionary of the INGS reflect the result of the frequency investigation into the verb vocabulary from the target texts.

According to the conditions, the mechanism analyzes the word appearance frequency in the target texts by using morphological analysis. Next, the proposed mechanism relates the above verbs with frequency to the terminal verb concepts in the verb conceptual dictionary by simple matching. Although each description of the verb concepts is theoretically a type of label and the linguistic notation is proposed using the language notation dictionary, a general notation is actually used for each of the description.

This study investigates the appearance frequency of the verb vocabulary from the text. Fig.5 shows (1) the process of acquiring the frequency information (2) the result of the verb concept dictionary terminal concepts. A verb concept is stored in the verb concept dictionary in the INGS hierarchically.

However, for the verb concept, a meaning perceives a different concept using a number with the same sound as the notation in the verb conceptual dictionary. For example, “eat 1(食べる 1)” and “eat 2(食べる 2)”.

Therefore, at first, in the case of acquisition of the frequency information, we compile a redundant concept. For example, we compile “eat 1(食べる 1)” and “eat 2(食べる 2)” to “eat(食べる)”. Finally, considering each verb concept, we return the number that we eliminated. The value of the frequency information of the verb concept and co-occurrence information then just applies a value before returning a number to each verb concept. For example, in the case of 10, the frequency information “eat” makes “eat 1” and “eat 2” from the verb concept and so does each frequency information with 10.

As a result, we have acquired word frequency for 4,885 verbs in the target texts and linked the frequency to 4,886 terminal verb concepts.

4.2 Acquiring the Co-occurrence Information between Verb Concepts

The acquisition method of the co-occurrence information is approximately same as an acquisition method of the frequency information. Because information to obtain the difference in a handling of (1) process of Fig.5 is co-occurrence information, the information to relate to a verb concept in (2) is the point that is a verb concept and

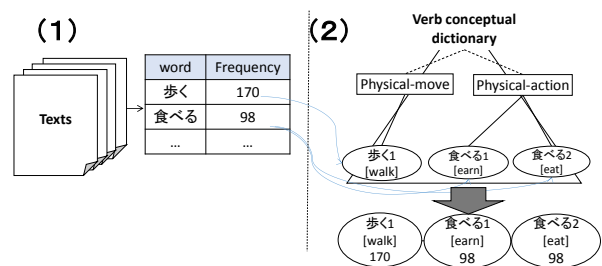


Fig. 5. Acquiring the frequency information

the group of the value indicating the strength of the collocation having a collocation as well as numerical value.

We show the calculating formula of the co-occurrence information in Fig.6. Where, “ S_a ” is the total number of sentences that “A” targeted for a search and is included in the text of the sample; the “ S_b ” is the total number of sentences that “B” targeted and is included in the text of the sample; “ $S_a \cap S_b$ ” is the total number of sentences including neither of “A” and “B” in the text of the sample; “ S_{all} ” is the total number of sentences included in the text of the sample.

The number of verb concepts that could obtain the co-occurrence information is the same as the number of verb concepts that were able to get frequency information.

5. Calculating the Frequency Information based on the Co-occurrence Information

In this section, we calculate the value of the estimate using the co-occurrence information between the verb concepts using the value calculated in section 4. When the verb concept is 0 in the frequency information of the verb concept, we used the value calculated in section 3.

5.1 A Method of Calculating the Frequency Information

Considering the verb concept of appearance frequency 0, we find the co-occurrence information of the verb concept based on a supposition “Appearance frequency of concept “B” having a collocation and the appearance frequency of concept “A” are proportional to verb concept “A””. In addition, we suggest a method to set the frequency information to a verb concept of frequency 0 using the frequency information in the verb concept.

The procedure is as follows: (1) The system extracts a verb concept having verb concept “A” and a collocation in order of the strength of the collocation using the KH Coder. (2) The system calculates the mean of the frequency information of each verb concept that the system extracted, and assumes this frequency information of verb concept

$$f = \frac{|S_a \cap S_b|}{|S_a|} \bigg/ \frac{|S_b|}{S_{all}}$$

Fig. 6. The calculating formula of the co-occurrence information

“A”.

For example, when “lay in stock” co-occur “over (frequency information 20,711)”, “to give (frequency information 66,190)”, and “to come back (frequency information 49,435)”, as for the frequency information of “to lay in stock” is 45,445.33 that is average of frequency information of “over”, “to give” and “to come back”.

In addition, we cannot estimate the frequency information when the frequency information of a certain verb concept and all verb concepts in the collocation is 0 only by this method. Therefore, we devised the frequency information acquisition technique by using the repetition of the suggestion technique. For example, in Fig 7, “(1)” shows a state of frequency provided using a method without the co-occurrence information, and “(2)” shows the frequency information setting by the method that uses the co-occurrence information. In concept “E”, the frequency remains 0 for frequency 0 then all of co-occurrence concepts. The case that sets the frequency information using the co-occurrence information more here is in “(3)”. The frequency information of concept “E” in “(3)” is set to 1 by calculating the average of concept “A” and “F”.

5.2 The result of Calculating the Frequency Information

We select 30 novel works of “Riichi Yokomitsu” among “Aozora Bunko” so that verb concepts of 0 frequency information sharply increased as a calculation object of the frequency information. We obtained the frequency information of 2,364 verb concepts. The total number of the verb concept of 0 frequency is 2,521. As a result of calculating the frequency information using the above-mentioned method, 3,672 verb concepts have frequency information, and 1,308 verb concepts have frequency information in comparison with the initial calculation of the frequency information. Fig.8 shows its transition by a graph.

We then show a generation result using the frequency information. We also show a generation result using the

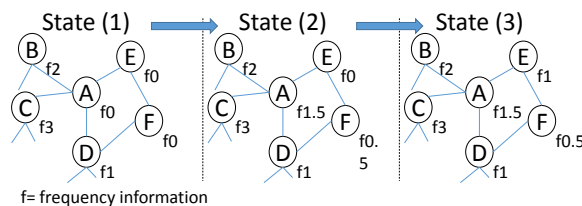
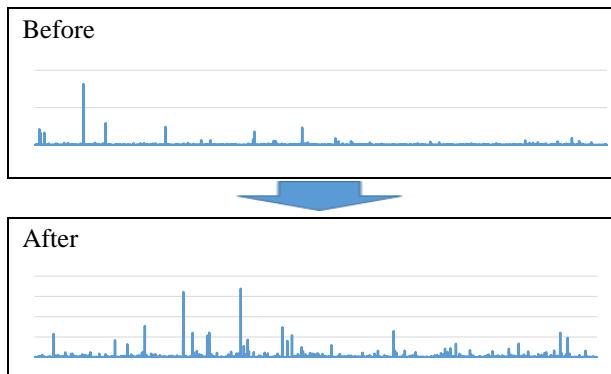


Fig. 7. A method of calculating the frequency information

frequency information before calculation using the co-occurrence information shown in Fig.9, and we present the generation result using the frequency information after having calculated in Fig.10.

We plan an estimate of the frequency information from the study of “Riichi Yokomitsu” and other authors used the condition by “Aozora Bunko” This experiment is intended to examine confirmation of the validity of the method and other methods from the tendency confirmation of every plural writer.



Axis of X: Frequency, Axis of Y: Verb concepts
Fig.8. The transition of the frequency information

親方[chief]が隠れ[hide]に出かける[go-out]。蛇[snake]が名馬 [fine-horse]を皇女 [princess]に与える [give]。皇女 [princess]が眠る [sleep]。細民[the-poor]が皇女[princess]を鞘 [sheath]に詰める [pack]。細民[the-poor]が鞘[sheath]を投げ [throw]。イワン[Ivan]が町中[downtown]に来る[come]。王様[king]が助け[help]をを求める[demand]。王様[king]がイワン [Ivan]を送り出す[see-off]。イワン [Ivan]が一戦 [a-battle]に備える [prepare]。イワン [Ivan]が出国する [depart]。贈与者[donner]がイワン[Ivan]に来訪[visit]を命じる [order]。イワン[Ivan]が従う [be-engaged]。イワン [Ivan]がビーフ [beef]を食べる [eat]。イワン [Ivan]が友 [friend]と会う [meet]。友 [friend]がイワン [Ivan]に譲渡する [transfer]。骨無し [spineless]がイワン [Ivan]を連れる [take]。イワン [Ivan]が蛇 [snake]と戦う [fight]。蛇 [snake]が鞘 [sheath]を落とす [drop]。イワン [Ivan]が鞘 [sheath]を拾う [pick-up]。蛇 [snake]がイワン [Ivan]に敗れる [be-defeated]。イワン [Ivan]が皇女 [princess]を発見する [discover]。イワン [Ivan]が飛び出す [rush-out]。イワン [Ivan]が皇女 [princess]を捕らえる [catch]。イワン [Ivan]が皇女 [princess]を連れる [take]。イワン [Ivan]が脱出する [escape]。イワン [Ivan]が到着する [arrive]。ニセ主人公 [false-hero]が言い張る [insist]。ニセ主人公 [false-hero]が要求する [demand]。イワン [Ivan]が王様 [king]と会う [meet]。イワン [Ivan]が持参する [bring]。イワン [Ivan]が皇女 [princess]に鞘 [sheath]を返す [return]。皇女 [princess]が王様 [king]に推理 [Reasoning]を伝える [convey]。王様 [king]が悪徳 [Vice]を知る [know]。王様 [king]が追撃 [Pursuit]を気付く [notice]。ニセ主人公 [false-hero]が露見する [come-to-light]。イワン [Ivan]が昇格する [be-promoted]。王様 [king]が蛇 [snake]を容赦する [pardon]。王様 [king]がニセ主人公 [false-hero]を容赦する [pardon]。イワン [Ivan]が皇女 [princess]と結婚する [marry]。

Fig. 9. The result (before calculation)

親 [parent]が病 [disease]で死ぬ [die]。候 [lord]があれこれ [this-and-that]をイワン [Ivan]に説く [preach]。候 [lord]が「イワン [Ivan]が楼閣 [multistoried-building]を出る [go-out]」ことをイワン [Ivan]に禁止する [forbid]。イワン [Ivan]が楼閣 [multistoried-building]を出る [go-out]。イワン [Ivan]が禁物 [taboo]を破る [break]。皇女 [princess]が翼 [wing]に怪我をする [hurt]。皇女 [princess]は姿 [appearance]を消す [deaden]。王様 [king]がイワン [Ivan]を連れる [take]。イワン [Ivan]が冒険 [adventure]を決意する [determine]。イワン [Ivan]が自国 [home-country]から他国 [other-country]へ出国する [depart]。贈与者 [donor]が牢 [Prison]に居る [be-in]。贈与者 [donor]が「イワン [Ivan]が贈与者 [donor]を牢 [Prison]から解放する [release]」ことを頼む [ask]。イワン [Ivan]は頼み [request]を聞く [grant]。イワン [Ivan]は贈与者 [donor]を開放する [open]。イワン [Ivan]は呪具 [magical-tool]を飲む [drink]。イワン [Ivan]が能力 [ability]を得る [get]。贈与者 [donor]がイワン [Ivan]から遠く [far-away]へ逃げる [escape]。贈与者 [donor]が血痕 [bloodstain]を残す [leave]。イワン [Ivan]が血痕 [bloodstain]を辿る [follow]。イワン [Ivan]が他国 [other-country]から敵地 [enemy's-territory]へ移動する [move]。イワン [Ivan]が敵対者 [adversary]とカード [card]で争う [compete]。皇女 [princess]がイワン [Ivan]に声を掛ける [speak-to]。敵対者 [adversary]がイワン [Ivan]にカード [card]で負ける [be-defeated]。イワン [Ivan]が皇女 [princess]を発見する [discover]。イワン [Ivan]が貴金属 [precious-metal]を皇女 [princess]に見せる [show]。イワン [Ivan]が皇女 [princess]をおびき出す [lure]。イワン [Ivan]が皇女 [princess]を連れる [take]。イワン [Ivan]が敵地 [enemy's-territory]から脱出する [escape]。イワン [Ivan]が敵地 [enemy's-territory]から自国 [home-country]に到着する [arrive]。ニセ主人公 [false-hero]が「ニセ主人公 [false-hero]が王様 [king]を助ける [help]」ことを言い張る [insist]。ニセ主人公 [false-hero]が報酬 [reward]を要求する [demand]。王様 [king]がイワン [Ivan]と会う [meet]。皇女 [princess]がイワン [Ivan]と会う [meet]。イワン [Ivan]が顔 [face]に傷 [injury]を持つ [have]。皇女 [princess]が「イワン [Ivan]に傷が付く [get-hurt]」ことを思い出す [remember]。皇女 [princess]が真実を知る [know]。皇女 [princess]が王様 [king]に真実を伝える [notice]。王様 [king]が真実を知る [know]。王様 [king]が「ニセ主人公 [false-hero]が「ニセ主人公 [false-hero]が王様 [king]を助ける [help]」と嘘を言う [lie-to]」ことに気づく [notice]。ニセ主人公 [false-hero]の嘘偽り [lie]が露見する [come-to-light]。イワン [Ivan]が衣服 [clothes]を着る [put-on]。イワン [Ivan]が輝く [get-the-honor-of]。王様 [king]が敵対者 [adversary]を容赦する [pardon]。王様 [king]がニセ主人公 [false-hero]を容赦する [pardon]。王様 [king]がイワン [Ivan]に国を与える [give]。

Fig. 10. The result (after calculation)

6. Conclusions

In this paper, we implement a similar method for verb concept selection. The method of the verb concept selection is based on frequency information. The method of the verb concept selection based on frequency information assumed in this paper premises that the appearance frequency of all concepts is prepared for. Therefore, we suggest a method using the co-occurrence information between the concepts. We demonstrate a calculation result of the frequency information using the co-occurrence information in the verb

concept when the frequency information becomes 0, in particular in the calculation method of the frequency information. We plan an estimate of the frequency information from the study of “Riichi Yokomitsu” and other authors used the condition by “Aozora Bunko” This experiment is intended to examine confirmation of the validity of the method and other methods from the tendency confirmation of every plural writer.

References

- (1) T. Ogata and A. Kanai : An Introduction of Informatics of Narratology: on Thought and Technology of Narrative Generation, Gakubunsha, 2010. (In Japanese)
- (2) T. Akimoto and T. Ogata : “An Information Design of Narratology: The Use of Three Literary Theories in a Narrative Generation System”, The International Journal of Visual Design, 7(3), 31-61, 2014.
- (3) T. Ogata : “Development of the Integrated Narrative Generation System”, Proc. of the 13th forum on information technology II, 323-330, 2014. (In Japanese)
- (4) T. Ogata : “Building Conceptual Dictionaries for an Integrated Narrative Generation System”, Journal of Robotics, Networking and Artificial Life, 1(4), 270-284, 2015.
- (5) T. Ogata and J. Ono : “Language Notation Dictionaries and the Use in an Integrated Narrative Generation System”, IEICE technical report, 115(69), 25-30, 2015. (In Japanese)
- (6) S. Imabuchi and T. Ogata : “A Story Generation System Based on Propp Theory: As a Mechanism in an Integrated Narrative Generation System”, Advances in Natural Language Processing, Springer, 7614, 312-321, 2012.
- (7) K. Higuchi : “Quantitative Analysis of Textual Data: Differentiation and Coordination of Two Approaches”, Sociological Theory and Methods 19(1) , 101-115, 2004.