

## **A Way for Using the Verb Conceptual Dictionary in an Integrated Narrative Generation System: the Use of Co-occurrence Information on Verb Concepts**

**Takashi Ogata**

*Faculty of Software and Information Science, Iwate Prefectural University  
152-52, Sugo Takizawa, Iwate 020-0693, Japan*

**Jumpei Ono**

*Graduate School of Software and Information Science, Iwate Prefectural University  
152-52, Sugo Takizawa, Iwate 020-0693, Japan  
E-mail: t-ogata@iwate-pu.ac.jp, g236m001@s.iwate-pu.ac.jp*

### **Abstract**

This study presents a method of selecting verb concepts based on analysis of co-occurrence information of verbs in modern Japanese novels collected from the Japanese digital library, “Aozora Bunko”. We plan to incorporate the proposed method into an integrated narrative generation system (INGS) that synthesizes various mechanisms (story generation, sentence generation, and so on) for narrative generation. We are currently developing such a system. In this study, we provide an overview of INGS and examine mechanisms related to the proposed method. In a verb conceptual dictionary that we have developed, as verb concepts of a variety of difficulty (comprehensibility) are mixed, the objective in this study is to adjust difficulty of verb concepts by using co-occurrence information on verb concepts. Our hypothesis is that the original concepts of co-occurrence and the resulting concepts have a proportional relationship. We investigate this hypothesis by incorporating proposed mechanism into an INGS.

*Keywords:* Co-occurrence information, Integrated narrative generation system, Verb conceptual dictionary, Verb concept

### **1. Introduction**

This paper mainly focuses on an event generation in the story generation mechanism and verb conceptual dictionary in an integrated narrative generation system (INGS)<sup>1-3</sup> (these details are mentioned in the section 2). An event composed of a verb concept and noun concepts. However, their concepts are selected randomly in an event generation in the INGS. Therefore, we previously proposed a method of the verb concept selection based on frequency information<sup>4</sup>. Here, we propose a method of the verb concept selection based on co-occurrence information.

### **2. The Architecture of an Integrated Narrative Generation System**

The architecture of our INGS is shown in Fig. 1. The system is composed of a concept generation mechanism used to generate a story structure and edit it (discourse), and a surface representation mechanism to represent a narrative. These mechanisms use dictionaries (stored concepts<sup>5</sup>, language notations<sup>6</sup>), story contents knowledge base (stored knowledge for generating structure of narrative)<sup>7</sup>, the state-event translation knowledge base (stored knowledge for connecting between each event and states in the conceptual structure)<sup>3</sup>.

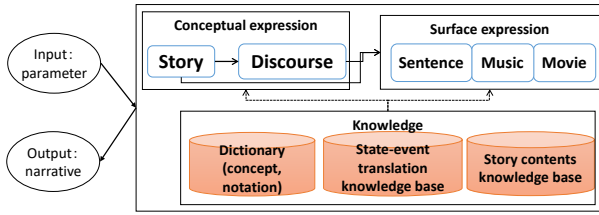


Fig. 1. Architecture of INGS

### 3. Overview of Conceptual Dictionaries

Conceptual dictionaries have a hierarchical structure from higher to lower concepts.

#### A) Hierarchy of Noun Concepts

The noun conceptual dictionary currently contains 115,765 terminal concepts and 5,809 intermediate concepts. Each intermediate concept has a list of hyponymy terminal concepts, the number of depths in the hierarchy, the super-ordinate concepts, and a list of the hyponymy intermediate concepts.

#### B) Hierarchy of Verb Concepts

The verb conceptual dictionary has 11,951 terminal concepts and 36 intermediate categories. Fig. 2. shows the description of the verb concept “bite.” Each terminal verb concept has the following three elements: 1) a sentence that includes the verb (“sentence-pattern” in Fig.2.), 2) the types of required noun cases (“case-frame”), and 3) defining the range which each noun concept in the case-frame(s) requires (“constraint”).

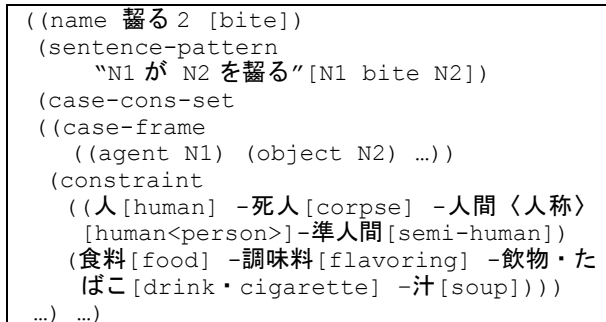


Fig. 2. Description of the verb concept “bite”

#### C) Generating an Event Concept in the Story Generation Mechanism

A story generation process expands a story structure by using story techniques. This process includes an event generation. An event is generated based on the description of the case-frame. A verb concept that is used in an event concept generation is decided based on

story techniques and story contents knowledge. The objective of this paper is to revise this selecting method.

### 4. Acquiring Co-occurrence Information

#### A) Used Texts and Method

We used “Aozora Bunk”, which includes 4980 texts of primarily Japanese novels published from 1872 to 1963, to acquire co-occurrence data of verbs. At first, co-occurrence information of all verbs in the texts was acquired.

The following (1) is a formula in KH Coder<sup>8</sup> to calculate the co-occurrence data. In the formula,  $S_a$  is equal to the total number of sentences that include verb  $a$  in the text;  $S_b$  is the total number of sentences that include verb  $b$  in the text;  $S_a \cap S_b$  is the total number of sentences including both  $a$  and  $b$  in the text; and  $S_{all}$  is the total number of sentences included in the text.

$$f = \frac{|S_a \cap S_b|}{|S_a|} / \frac{|S_b|}{S_{all}} \quad (1)$$

#### B) Result

The co-occurrence information was corresponded to 4866 terminal verb concepts in the verb conceptual dictionary in INGS. As an example, Table. 1 shows co-occurrence data of “bite”. Each verb concept connects to 143.44 verb concepts in average.

Table. 1. Co-occurrence data of “bite”

Co-occurrence information	Relating verb concepts
141.1954	打ち上げる 1[shoot up], 慕う 1[pout], etc.
70.5977	飽きる 1[loss intent], 謀る 1[deceive], etc.
	...<Omitted>...
1.035663	出来る 7[be able to]
1.001386	戻る 1[return]

### 5. Preliminary Questionable Research

We evaluated the effectiveness in the verb concepts selection based on the acquired co-occurrence information.

#### 5.1 Preparation

Four lists of verb concepts, which were extracted from the stories generated by using a partial mechanism in INGS, were used in the questionable research.

**A) A Used Story Generation Method in INGS**

The story structures were generated by using “story contents grammar”<sup>9</sup> and the mechanism that we have been developing for a part in INGS. A hierarchy of the story contents grammar has a layer of “functions”. The level is the most fundamental and a sequence of functions form a basic structure of a story. And, each function is extended to the lower level where one or more verb concepts are instantiated by selected from the candidates.

Fig. 3. shows used functions and the fixed sequence. Each of the functions is rewritten by the lower levels to reach to the level of one or more concrete verb concepts.

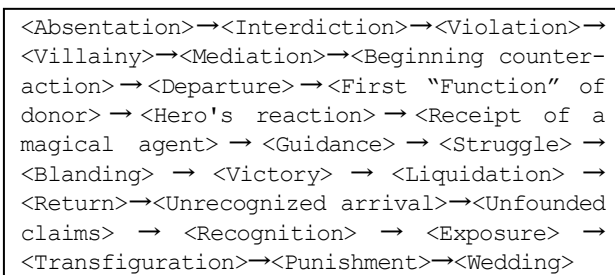


Fig. 3. A used sequence of functions

**B) Making Verb Concept Lists**

We acquired co-occurrence data using the above story generation mechanism by the following method. When the mechanism selects a concrete verb concept from generally many verb concepts, it utilizes co-occurrence data as shown in the next types.

**[Max]** The story generation mechanism generated 5 stories by repeatedly using a verb concept that has the strongest co-occurrence with the prior verb concept and finally decided 39 verb concepts in one story selected randomly from the above stories.

**[Middle]** The mechanism generated 5 stories by repeatedly using a verb concept that has the medium co-occurrence with the prior verb concept and finally decided 41 verb concepts in one story selected randomly from the above stories.

**[Min]** The mechanism generated 5 stories by repeatedly using a verb concept that has the weakest co-occurrence with the prior verb concept and finally decided 47 verb concepts in one story selected randomly from the above stories.

**[Random]** The mechanism generated 5 stories by repeatedly using a verb concept randomly for the prior verb concepts and finally decided 41 verb concepts in one story selected randomly from the above stories.

In addition, in the following cases, a verb concept is selected randomly: 1) the first verb concept in story generation, 2) when two or more candidates exist in the selection timing, 3) when no candidates with co-occurrence data in the selection timing.

**5.2 First Evaluation**

**A) Goal and Process**

The goal is to check the relationship between the degree of co-occurrence information and the degree of comprehensibility by the users. The subjects were five students. They read four kinds of verb concepts lists (max, middle, min, and random) to select an value from 1 (easy) to 5(difficult). We set a time limit of five minutes for each verb list.

**B) Result**

As shown in Table. 2 shows the result, we could not see clear differences among the subjects. In short, co-occurrence information of verb concepts does not influence the comprehensibility.

Table. 2. Evaluation result

	A	B	C	D	E	Average
[Max]	1.23	2.54	2.87	3.08	2.82	2.51
[Middle]	1.29	2.73	3.00	2.80	3.10	2.59
[Min]	1.26	2.50	3.02	2.85	2.65	2.46
[Random]	1.15	2.60	2.93	2.80	2.63	2.42

This result was similar to the case of noun concept selection based on co-occurrence information that we previously implemented<sup>10</sup>. Also, frequency and co-occurrence are clearly different methods to adjust verb concepts selection in INGS.

**5.3 Second Evaluation**

**A) Goal**

On the other hand, in the noun concepts selection using co-occurrence information, we confirmed another method. In the method, a noun concept is associated with other noun concepts that have the similar degree of

co-occurrence information<sup>11</sup>. We adapt this method to verb concepts selection.

**B) Process**

For the above stories of max, middle, and min, we calculated the absolute values of difference (*D*) between the value of a verb concept *X* and the value of the next verb concept *Y*. Fig. 4 shows the procedure. As the case of noun concepts selection, we consider if *D* is lower, it indicates the similar comprehensibility. However, it is necessary to be actually investigated the consideration. In this paper, we check whether lower *D* is resulted by higher degree of co-occurrence. It enables to adjust selecting verb concepts using co-occurrence of verb concepts.

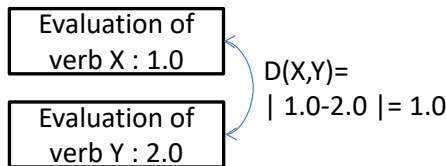


Fig. 4 Calculating *D*

**C) Result**

Table 3 shows the result of calculating *D*. In this Table 3, we did not use verb concepts that were selected randomly. The average of *D* in [Max] was smaller than others. If lower *D* is related to the similar degree of comprehensibility of the verb concepts, we may be able to use this result to adjust the degree of comprehensibility of verb concepts.

Table 3. The result of calculating *D*

Method	The number of the evaluated verb concepts	Average of <i>D</i>
[Max]	12	0.43
[Middle]	15	0.76
[Min]	14	0.96

**6. Conclusion**

This paper preliminarily proposed a method of a verb concept selection by using co-occurrence information for the conceptual event generation mechanism in INGS. The results indicate a next design direction of the verb concepts generation in INGS: to use verb concepts that have co-occurrence relationship for prior verb concept enables to select them by similar degree of comprehensibility. The proposed method may useful to adjust the verb concepts selection in INGS.

In the future, we plan to precisely improve the method to finally integrate methods of concepts selection in event generation in INGS using frequency and co-occurrence of words.

**References**

- Ogata, T. and Kanai, A.: An Introduction of Informatics of Narratology: *On Thought and Technology of Narrative Generation*, Gakubunsha, 2010. (In Japanese)
- Akimoto, T. and Ogata, T.: “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.
- Ogata, T.: “Development of the Integrated Narrative Generation System”, *Proc. of the 13th Forum on Information Technology II*, 323-330, 2014. (In Japanese)
- Ogata, T. and Ono, J.: “Controlling the use of Semantic Concepts in an Integrated Narrative Generation System: The use of the Verb Frequency Information”, *Proc. of 21st International Symposium on Artificial Life and Robotics*, 2016.
- Ogata, T.: “Building Conceptual Dictionaries for an Integrated Narrative Generation System”, *Journal of Robotics, Networking and Artificial Life*, 1(4), 270-284, 2015.
- Ogata, T. and Ono, J.: “Language Notation Dictionaries and the Use in an Integrated Narrative Generation System”, *IEICE Technical Report*, 115(69), 25-30, 2015. (In Japanese)
- Ogata, T.: “Computational and Cognitive Approaches to Narratology from the Perspective of Narrative Generation”, Ogata, T. & Akimoto, T, (Eds.), *Computational and Cognitive Approaches to Narratology*, IGI Global, 1-74, 2016.
- Higuchi, K.: “Quantitative Analysis of Textual Data: Differentiation and Coordination of Two Approaches”, *Sociological Theory and Methods*, 19(1), 101-115, 2004.
- Imabuchi, S. and Ogata, T.: “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.
- Ono, J. and Ogata, T.: “Noun Concepts Selection in Events Based on the Co-occurrence Relation in Verb Concepts and Noun Concepts: Mechanisms for the Story Generation in an Integrated Narrative Generation System”, *Proc. of the 14th Forum on Information Technology II*, 239-242, 2015. (In Japanese)
- Ono, J. and Ogata, T.: “Selecting Noun Concepts Based on Accounting Data: As a Mechanism in the Integrated Narrative Generation System”, *IEICE Technical Report*, 114(366), 49-54, 2014. (In Japanese)