Title
Building Conceptual Dictionary for Providing Common Knowledge in the Integrated Narrative Generation System

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 34(34)

ISSN
1069-7977

Authors
Oishi, Kensuke
Kurisawa, Yasunari
Kamada, Mami
et al.

Publication Date
2012

Peer reviewed

Building Conceptual Dictionary for Providing Common Knowledge in the Integrated Narrative Generation System

Kensuke Oishi (g231g010@s.iwate-pu.ac.jp)
Graduate School of Software and Information Science, Iwate Prefectural University, 152-52 Sugo
Takizawa, Iwate 020-0193 Japan

Yasunari Kurisawa (g031g054@s.iwate-pu.ac.jp)
Faculty of Software and Information Science, Iwate Prefectural University

Mami Kamada (g031i301@s.iwate-pu.ac.jp)
Faculty of Software and Information Science, Iwate Prefectural University

Itaru Fukuda (g031h134@s.iwate-pu.ac.jp)
Faculty of Software and Information Science, Iwate Prefectural University

Taisuke Akimoto (g236i001@s.iwate-pu.ac.jp)
Graduate School of Software and Information Science, Iwate Prefectural University

Takashi Ogata (t-ogata@iwate-pu.ac.jp)
Faculty of Software and Information Science, Iwate Prefectural University

Abstract

We explain the current version of a conceptual dictionary containing two hierarchies of verb concepts and noun concepts to be functioned in our narrative generation system. It is used for operating naturalness or validity of generated events and realizing or adjusting the intentional defamiliarization. Namely, this dictionary is a mechanism to be able to flexibly adjust a variety of generation from realistic narratives to fantastical narratives as well as the foundation for a narrative event and the elements. In the current version, verb concept dictionary has originally defined 5338 case frames and modified 1158 constraints and noun concept dictionary contains 121573 concepts including 5808 intermediate concepts.

Keywords: Narrative generation system; conceptual dictionary; verb/noun concept hierarchy; case frame; constraint.

Introduction

This paper explains the development of a conceptual dictionary or hierarchical systems of concepts in the narrative generation system framework which is our main research theme. A basic unit for a narrative in the system is an event concept containing a verb concept and noun concepts and the information of these concepts is held in the conceptual dictionary. It is one of the central components in the system.

Narrative is the strongest method for organizing fragmentary knowledge human being has. We have been developing a narrative generation system as an intelligence tool for the creation of future literature & narrative (Ogata & Kanai, 2010). For digital art and entertainment such as computer game, perhaps narrative and story can become an important element in the same manner as traditional genres. Our research of narrative generation is done for the application to novel contents such as computer game and narrative generation based narrative or literature.

As a related work, Okada and Endo (1992) proposed a system to generate stories like Aesop fables. A story generated is a kind of simulation of the process that a main character or actor plots a sequence of planning actions. In contrast, our narrative generation system architecture is constructed as an organic fusion of diverse narrative knowledge and techniques including planning, discourse structure, story grammar, script, discourse relation, and so on. Although the Aesop system has a feature as an application of conceptual dictionary research, our goal of the narrative generation system is pursuing the mechanism of narrative generation itself. Our extreme purpose is not developing conceptual dictionary itself, but creating narrative generation system. Therefore, a basic policy here is to use existing dictionaries as possible to customize and expand them according to the architecture and mechanism.

As a narrative generation study, Oz project (Bates, 1992) attempted the development of an interactive drama with dialogue and actions in autonomous agents. This system mainly focuses on the interactive techniques for the user's narrative experiments. In contrast, our system contains a variety of narrative and linguistic knowledge for generating deep and conceptual narrative structures. On the other hand, BRUTUS (Bringsjord & Ferrucci, 2000) is an interactive narrative generation architecture which has an integrative feature including story grammar, planning, and so on. However, it deals with only a specialized narrative theme, "betrayal". Whereas, we are intend to develop a more general mechanism for various types of narratives.

The goal of this paper is the proposition of a conceptual dictionary in the narrative generation system architecture we have been developing. The conceptual dictionary has two components of verb concept dictionary and noun concept dictionary and each system has a hierarchical structure based on single inheritance. A main issue in the development of conceptual dictionary is currently defining constraints, which means the knowledge for deciding the range of value for each case in an event as a basic unit in a narrative. A constraint is described in a verb concept and prescribes the possible range of noun concepts. In this paper, we describe the whole structure and some detailed parts of the conceptual dictionary by especially putting a focus on the description and role of constraints. Although this paper uses existing studies of conceptual dictionaries as a reference, in the combination with the domain of narrative generation system, a variety of novel and difficult issues emerges. For example, realistic narratives and fantastical narratives respectively need different widths of conceptual constraint. This issue is directly treated in other papers (Zhang, Ono & Ogata, 2011, 2012). A characteristic of this study is an exploratory approach through the incremental system development, and it is hoped that models and theories progress through the repetition of design-implementation-experiment. The proposed conceptual dictionary in this paper provides a foundation for such incremental process toward more principled modeling.

Narrative Generation and Conceptual Dictionary

In the macro design, the architecture of narrative generation system consists of conceptual generation phase and surface representation phase by natural language, visual media, and music. The former is divided into story as the narrative content to be narrated and discourse as the structure that a story is narrated. According to this framework, we have been developing various mechanisms and modules independently of each other. Currently, we are advancing the developing of an integrated narrative generation system in which a variety of modules are organically blended into a whole. The goal of this system is to execute the generation of narrative structures in some levels such as story, discourse, and surface representations in a unified method. Story and discourse are represented in formally same a tree structure description form. Each leaf node in the structure is corresponding to an event or state and each intermediate node is corresponding to a relation combining the child nodes including events, states, and the sub-structure. An event as a unit of conceptual representation is the most fundamental element of story and discourse. It is described in the form of case frame which is linked to a verb concept and some noun concepts in the conceptual dictionary. When story or discourse generation mechanism generates an event concept, the conceptual dictionary provides or constraints the semantic elements. A case frame for a verb concept is created and a constraint in the verb concept combining with noun concepts decides the semantic range of each noun concept in the case frame.

The proposed conceptual dictionary is actually positioned in a pilot version of integrated narrative generation system (Akimoto & Ogata, 2011b). It is implemented by Common Lisp with about 800 functions. The main macro modules are control mechanism, conceptual generation module, and surface representation module. The control mechanism calls each module according to a set of parameters as the goal of generation to automatically execute a generation process from story to expression. The conceptual generation module contains next modules to generate or transform narrative conceptual structure: a story generation mechanism including a single event generation mechanism using rhetorical techniques (Zhang, Ono & Ogata, 2011, 2012), a state-event transformation mechanism (Onodera & Ogata, 2012), an events sequence generation mechanism by story grammar based on Propp's narratology (Propp, 1969; Imabuchi & Ogata, 2011), and a discourse mechanism containing 13 kinds of narrative discourse techniques redefined Genette's narrative discourse theory (Genette, 1972; Akimoto & Ogata, 2011a). The surface representation mechanism consists of a sentence generation mechanism, an animated movie generation mechanism (Ogata, 2008), and a cyclical mutual transformation mechanism between conceptual narrative and music (Ogata, Akimoto & Seito, 2011).

Our narrative generation project is a longitudinal and exploratory study, and we have been employing both top-down approach relevant to the system's macro design and bottom-up one programming various modules in parallel. Although the above system is a kind of "bricolage" in the current state, the conceptual dictionary plays an important role in an organized integration of some modules developed independently. The standardizing of an event form and the reference of a common conceptual dictionary by various modules enables the combination as a whole narrative generation system. This is the most fundamental role of the conceptual dictionary and advanced topics of research for upgrading narrative generation emerge on the ground.

The Structure of Conceptual Dictionary

As described above, the basic unit of a narrative is an event or an event concept described as a conceptual representation. A story and a discourse are represented as each tree structure (Figure 1). Each terminal node is corresponding to an event and each intermediate node is a relation for binding some the lower nodes. Each event is represented as a case frame with a verb concept and some noun concepts. We prepare next eight cases: agent (an subject in an action), counter-agent (an object in an action and a living thing), object (an object in an action and a no living thing), instrument (an tool used in an action), location (a place of an action), time (a pair of starting time and ending time of an action), from (a starting place of an action), and to (an ending place of an action). In addition, we prepare seven kinds of optional cases for treating such concepts which are not able to describe using above cases (Table 2).

The conceptual dictionary performs as a background for each event in narrative generation. It is divided into verb concept hierarchy and noun concept hierarchy. Both have a hierarchical structure based on is-a relations by the mechanism of single inheritance. A verb concept defines one or more case frames which become the templates of the event concept(s). In addition, it has the description of constraint condition (simply constraint) for limiting the range of each case's value. A constraint is defined by one or more noun concepts in noun concept hierarchy and means the range that an element in each case contained in a verb concept can refer inside the noun conceptual hierarchy. The current version of verb concept hierarchy has 5338 case frames and 4881 constraints. On the other hand, current noun concept hierarchy contains 121573 noun concepts including 5808 intermediate concepts.

In a verb concept, if noun concepts inside the range of constraint are used, a natural and possible event like "A knight fight an enemy" is formed. But, if noun concepts outside the range are used, an unnatural and impossible event like "A knight fight a windmill" is created. We show another example. In an event concept (eat (agent old-person) (object rice-ball)) ("an old-person eats a rice-ball"), when (eat (agent N1) (object N2)) is defined as the case frame corresponding to "eat" and the noun concept "food" is defined as the constraint of N2, the

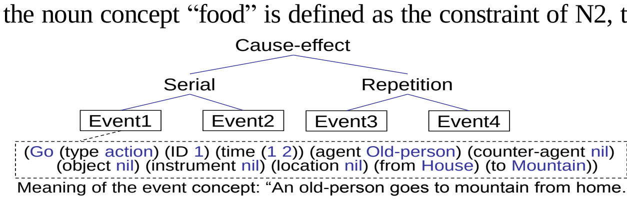

Figure 1: An example of narrative conceptual structure

noun concept to inserted into N2 is the subordinate concept of "food" in the noun concept hierarchy. Therefore, we can confirm that this event concept is constructed inside the constraint.

By the way, fantastical events or physically impossible ones appear in narratives normally. One of concerns for us in our narrative generation project is to be able to flexibly generate from more realistic narratives or events sequences to more fantastical ones. Although the objective of this paper does not discuss the topic, it has been treated in the study of advertising scenario generation system as an application of the narrative generation system in our group (for details, Zhang, Ono & Ogata, 2011, 2012). One of main objectives of this research is to adjust the semantic range in a single event using the concept and techniques of "defamiliarization". Its idea was acquired from a rhetorical analysis of actual television commercial films. The analysis defined three types of standard rhetoric, which are for generating events corresponding to ordinary events, and nine types of irregular rhetoric, which are for generating events corresponding to extraordinary events. We call twelve types of rhetoric "product introduction rhetoric (or technique)". In advertising narratives, events with a product are represented through the rhetoric. The former types of rhetoric draw the processes of "manufacturing", "purchasing", and "usage" of a product. In contrast, the latter types of rhetoric perform a kind of deviance by techniques of defamiliarization for "the actor's action", "the state of product", "the background or place", "the actor himself (herself)", and so on. Defamiliarization means a literary technique for changing a familiar object into unfamiliar one to reinforce the impression. For example, the impression of a familiar product is reinforced by the application of defamiliarization techniques to the objects and agents. This idea will be able to generalize to narratives other than the advertising narrative.

We show the overview of the processing flow. First, user selects a product name from a products list prepared by the developer and the number (1-3) of a type of standard rhetoric. Based on the information, the system generates an ordinary event by acquiring a verb concept and some noun concepts within the standard rhetoric for the product. These concepts are prepared according to each product. Next, when the user designates an irregular rhetoric, the system rewrites the ordinary event to an irregular or extraordinary event by applying the corresponding rhetorical technique. Conceptual dictionary is used in the application of irregular rhetoric. Specifically, the system replaces an original concept in the target event with another concept in the different category by changing the reference region in the noun concept hierarchy. For example, to generate an event with a product "car" according to "the defamiliarization of location" rhetoric, first, the system generates an event like "(drive (agent woman) (instrument car) (location plateau))" using a noun concept "car" inside the constraint. Next, the system replaces a noun concept corresponding to the location with another concept outside the constraint using the defamiliarization processing to generate an event at the strange place like "(drive (agent woman) (instrument car) (location seabed))". The condition of this processing is that the constraint in regular rhetoric is limited in the actual range. In this research, the conceptual dictionary basically defines the standard semantic range for each noun concept in an event from the viewpoint of actual possibility or physical possibility principally. On the other hand, extraordinary or irregular events are constructed by several types of defamiliarization rhetoric based on a kind of constraint relaxation. Although a weak point of the current version is that the defamiliarization techniques are randomly applied, we succeeded at the experimental implementation of a simple framework of defamiliarization processing using the conceptual dictionary.

In addition, a story generation system by McIntyre and Lapata (2009) generates stories by using a knowledge base about compositions of an event sentence and chains of events extracted from a narrative corpus based on co-occurrence of words. Stories are generated by a kind of tree search of possible stories. Here, the system has several scoring criteria for pruning low scored branches that are strength of connection of words in an event and words between adjacent events, interestingness, and coherence.

Existing Conceptual Dictionaries

Table 1 compares existing Japanese conceptual dictionaries. "Goi-Taikei, A Japanese Lexicon" (Ikehara et al., 1999) is the largest scale Japanese lexical and conceptual dictionary for Japanese-to-English machine translation, which hierarchically organizes semantic attributes based on single inheritance. We mainly referred to it as a starting point to construct the overall structures of two hierarchies. The lexicon is basically a lexical dictionary, and because the definition of constraints and the granularity of noun concepts' categories are too rough to directly use in the narrative generation system, we considerably refined their organizations and uniquely added case frames for verb concepts. Especially, for the purpose of complementing the shortage of intermediate concepts in the noun conceptual hierarchy to a large degree, we referred to "Japanese WordNet" (Bond et al., 2009). Although we used only Japanese lexicon and Japanese conceptual dictionaries as a reference, narrative has an aspect which transcends linguistic difference and we also need to them.

A Verb Concept Hierarchy

Figure 2 shows the overall structure of verb concept hierarchy. Although the Japanese lexicon contains about 12000 verb concepts under 36 categories, the current version of our verb concept hierarchy contains 4260 verb concepts in 6 categories to represent physical state changes ("physical transfer", "possessive transfer", "change of attribute", "change of body", "body motion", and "generation"). A verb concept defines a sentence pattern ("A Japanese Lexicon" contains about 12000 sentence patterns), which is the template for a sentence that takes a verb as the predicate, one or more sets of case frames, the constraints, and the superordinate concept (by an "is-a" relation). The number of case frames is 5338 for 4260 verb concepts.

In Figure 3, we show the form by an example of a verb concept's description by Common Lisp. The information relating to a verb concept is stored into a variable. In this example, next information is stored into the variable: (1) A corresponding sentence pattern; ("N1 eat N2"), (2) Two case frames; "(eat(2) (agent N1) (counter-agent N2))" and "(eat(2) (agent N1) (object N2))" ("(2)" means the second meaning of "eat" covering some meanings), (3) Constraints for each case, and (4) the information that the superordinate concept is a verb concept category ("body motion").

Defining Case Frames

In the description of a case frame, we decided adequate cases corresponding to respectively a noun concept based on the

Table 1: Existing conceptual dictionaries

	EDR Electronic Dictionary (National Institute of Information and Communications Technology, 2001)	Japanese WordNet (Bond et al., 2009)	Goi-Taikei, A Japanese Lexicon (Ikehara et al., 1999)
Number of concepts	About 410,000	57,238	About 3,000
Number of words	About 270,000	93,834	About 300,000
Structure of concept hierarchy	Multiple inheritance	Multiple inheritance	Single inheritance
Feature	Bilingual dictionary, collocation dictionary, corpus, etc. are recorded.	The words are divided into group by synonym (synset).	About 12000 sentence patterns corresponding to the meaning of verb concept are recorded.

sentence pattern and the constraints. We show an example of the definition based on a sentence pattern, "N1 eat N2". The constraints of N1 and N2 are respectively "(human, animal)" and "(food, life)". This means that N1 takes the subordinate concepts of "human" or "animal" and N2 takes the subordinate concepts of "food" or "life". In the N2, we distinguish "a living thing" from "a nonliving thing" and define two types of case frames, "(eat(2) (agent N1) (object N2))" and "(eat(2) (agent N1) (counter-agent N2))". Here, the constraints of N1 in two frames are "human" and "animal". However, in the N2, the constraints of two frames are respectively "food" and "life". As mentioned above, a verb concept sometimes holds some different case frames.

Currently, case frames for 174 verb concepts are not defined because it is difficult to find an adequate case for a noun

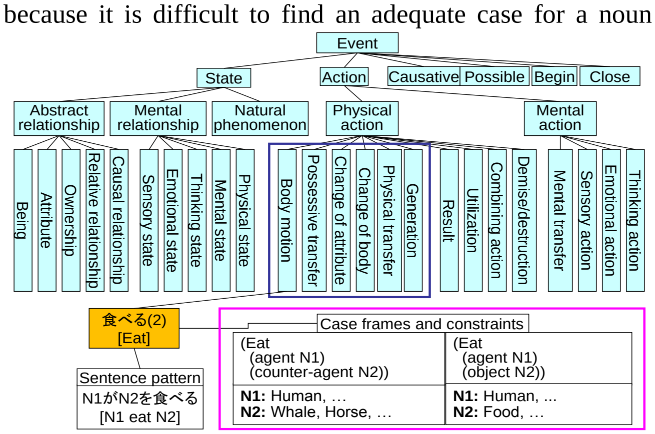

Figure 2: The overall structure of verb concept hierarchy

```
Description format
(set (intern <verb concept>) '((name (<verb concept's name>))
    (sentence-pattern <sentence-pattern>)
    (case-cons-set
        ((case-frame ((agent [<noun term> nil]) (counter-agent [<noun term> nil])(location
                      [<noun term> nil]) (object  [<noun term> nil])(instrument [<noun term>
                      nil]) (from [<noun term> nil]) (to [<noun term> nil])))
         (constraint ((<noun concepts>)))))
    (is-a (<superordinate concept>))))

An example
(set (intern "食べる[eat](2)") '((name (食べる[eat] (2)))
    (sentence-pattern "N1が N2を 食べる"[N1 eat N2])
    (case-cons-set
      ((case-frame ((agent N1) (counter-agentN2) (location nil) (object nil) (instrument nil)
                    (from nil) (to nil)))
       (constraint (("人<human>" "-死人<dead person>" "-人間〈人称<personal pronoun>〉"
            "-準人間<demi-human>") ("*獣[鯨<whale>]*" "*獣[馬<horse>]*" "*獣[牛<cattle>]*"
            "*獣[豚<pig>]*" "*獣[山羊<goat>]*" "*獣[羊<sheep>]*" "*獣[鹿<deer>]*"
            "*獣[猪<wild boar>]*" "*獣[兎<rabbit>]*" "*鳥<bird>[家禽<poultry)]*"
            "*鳥[猟鳥<game bird>]*" "魚<fish>" "-*魚[伝説<legendary>]*"
            "たこ･いか･えび･かに<octopus･squid･prawn･crab>"))))
      ((case-frame ((agent N1) (counter-agent nil) (location nil) (object N2)(instrument nil)
                    (from nil) (to nil)))
       (constraint (("人<human>" "-死人<dead person>" "-人間〈人称<personal pronoun>〉"
            "-準人間<demi-human>") ("食料<food>" "-調味料<seasoning>"
            "-飲物･たばこ<drink･cigarette>")))))
      (is-a (身体動作<Body motion>))))
```

Figure 3: The description of a verb concept

concept within the current cases. Such noun concepts are categorized into following seven types: "adverbial concept", "possessive case", "event", "purpose of action", "experiencer", "target of comparison", and "idiom". We show an example of "purpose". In a verb concept, "go(3)", the sentence pattern is "N1 go to N3 from N2 for N4" and the constraint of N4 is "(abstraction)" which shows the purpose of "go". For instance, "a man goes to a mountain from a house for mowing". If we interpret it as "to", we can not distinguish N4 from N3 ("(place, field)"). Table 2 set straight above seven cases to be expanded and the relevant studies. Okada (1991) and Takeuchi (2011) show other case frames. The former uses ten types of cases. And, the latter uses 71 types of cases to represent the argument structure of verbs which defines the relationship between a verb and nouns in a sentence. Because the description of case frames by the reference may become too long and complex, we tentatively describe seven types of case frames with "optional" sign.

In addition, for example, FrameNet (Fillmore & Baker, 2010) and Japanese FrameNet (Ohara, 2008) define the semantics of a word by a semantic frame, which means the structured knowledge about a typical scene, constructed with the frame elements. And, VerbNet (Kipper-Schuler, Dang, & Palmer, 2000; Kipper-Schuler, 2005) is a hierarchical verb lexicon in which each class is described by semantic predicates, thematic roles, and basic syntactic frames. Their resources and above Takeuchi (2011) are advanced semantic knowledge to deal with the complex narrative knowledge unit. In the modular approach of our narrative generation system architecture, a common and a large amount of conceptual dictionary treats the semantic knowledge which has a comparatively simple structure and other mechanisms handle more complex semantic processing. For example, Onodera and Ogata (2012) resolve an action into two states and the relations to generate sequences of states which form a basis of a story. The above resources are more directly related to the part of such complex semantic processing.

Editing Constraints

The result of a previous evaluation and consideration shows that if we adopt "actual naturalness" as the criterion, the definition of constraints become too difficult or ambiguous (Oishi & Ogata, 2011). So, as a basic policy, we define constraints in the range of physical possibility (in our daily life in common sense). We think this criterion is comparatively clear though it may actually contain difficult problems. In addition, constraints contained in "A Japanese Lexicon" have originally the purpose and role of conflict resolution for Japanese-English translation. For the context of narrative generation, they are too wide to adequately limit the semantic range for each noun concept to be generated. For example, in the case of "eat(2)", the subordinate concepts of a constraint "food" corresponding to the N2 contain inadequate concepts such as "drink". And, the subordinate hierarchy of "life" also contains "the animal on a legend". One of primary purposes of the extension of hierarchy is to avoid such ambiguity and inaccuracy.

The method is divided into two parts. First is by hyponymy concepts and "furniture" is substituted with "chair" and "bed". Second is by the partial exclusion of hyponymy concepts and "dead person" is excluded from "human" using a minus sign such as "-dead person". Oishi and Ogata (2011) have modified 300 constraints by the methods. For noun concepts defined as constraints of all terms in these case frames, we investigated

Table 2: Towards the expansion of deep case definition

problem	example of sentence pattern	Okada (1991)	Takeuchi (2011)	New case
1) adverbial concept	N1が N2を N3に **N4** 縮める (N1 shorten N2 by N4 to N3)	OC: complement of attribute	numeral / correspond to adverb	**Adverb**
2) possessive case	N1が **N2**の N3に/へ 潜る (N1 go under N2)			**Possessive**
3) event	N1が N2を **N3**で 破る (N1 beat N2 at/in N3)		situation	**Situation**
4) purpose of action	N1が N2から/より N3へ/に/まで **N4**に 出かける (N1 go to N3 from N2 for N4)		purpose	**Purpose**
5) experiencer	N1が **N2**に 疲れる (N1 get tired of N2)		experiencer	**Experiencer**
6) target of comparison	N1が N2を **N3**の 倍にする (N1 make N2 twice as much as N3)	OS: souse		**Souse**
7) idiom	N1は **N2(足)**が N3から 遠のく (N1 go to N3 less frequently)		usage	**Ideom**

the subordinate concepts and edit them according to the methods. For instance, for the constraints of N3 in "N1 return to N3 from N2", we extended the original constraints "(location place building)" to "(lodging housing area -area(scope) -area(human_activity) -land -world foot_of_mountain mountain_pass valley ground island/cape shore -bank farm site point_of_compass edge distance house(body)[housing])". However, this work spent a lot of time.

To improve the work efficiently and organically, we prepared a lot of "sets" that describe a group of constraints to be used commonly in similar plural case frames. For instance, a common set can be applied to the case frames consisting of a place a person can sit down and the set can be used to the opposite concepts such as "stand up" too. Specifically, to a set of noun concepts used as constraints, we give a name that the set means as a whole to apply it to the constraints for other case frames. For example, we give a name, "inhabitable location" to a set of constraints, "(lodging housing area -area(scope) -area(human_activity) -land -world foot_of_mountain mountain_pass valley ground island/cape shore -bank farm site point_of_compass edge distance house(body)[housing])". The set of noun concepts is commonly used for the constraint of N2 in other events like "N1 move to/into N2", and so on.

In this time, we extracted 26 kinds of sets from the modified constraints in 300 case frames and applied the sets to other 5037 case frames. As a result, we could apply the sets to one or more elements in 4581 case frames. The number of cases that are adequate sets to all elements was 1158. For example, in "N1 take out N2 from N3", we could apply a set, "a person of such age who can take care by herself or himself" to N1 and "transportable object" to N2.

A Noun Concept Hierarchy

Figure 4 is the overall structure of noun conceptual hierarchy. The current version contains 5808 intermediate concepts and 115765 terminal concepts. Many of the higher level concepts are based on the information registered in "general semantic attributes system (hierarchy) for nouns" in "A Japanese Lexicon". On the other hand, the hierarchy of lower intermediate concepts is basically organized referring to "Japanese WordNet". The terminal concepts are based on the words in the Japanese lexicon.

Next Figure 5 is the format for describing the noun concept hierarchy and the actual example by Common Lisp. The upper list means the serial number showing registered intermediate concepts, and variables into which the information for each intermediate concept is stored. An intermediate concept has a list of hyponymy concepts, a number of the depth in hierarchy, the serial number of the superordinate concepts, and the range of serial numbers of the hyponymy concepts. In the example, "(seal sea-lion fur-seal sea-animal dugong bakushia)" is defined as the subordinate concepts of "sea-animal".

The method of constructing the noun conceptual hierarchy is as follows. The referred Japanese lexicon contains 141870 words in 2710 attributes in "general semantic attributes system (hierarchy) for nouns". Here, each attribute is corresponding to a concept in our conceptual dictionary. These words have homonyms like "狸" ("tanuki" in Kanji) and "タヌキ" ("tanuki" in Katakana), which means commonly "raccoon dog" in Japanese. In the narrative generation system, we divide the generation process into a conceptual level to make a narrative content and a surface representation level to make the variations by language and other representation media. The former uses a conceptual dictionary and the latter uses a language dictionary. To realize the mechanism, after we integrated two or more homonymous words into a single concept, we viewed all words as concepts to store 115765 concepts into the noun concept hierarchy.

In addition, to be able to set more detailed and elaborated constraints, we added intermediate concepts and refined the classification referring to "Japanese WordNet". In the step that we integrated homonymous words into a single concept to register them as concepts, many terminal concepts were directly combined with the intermediate concepts. For example, "seal" and "raccoon-dog" are respectively the subordinate concepts of "beast", and more detailed categories do not exist. If we want to define "aquatic mammal" as a constraint, we have to list all the terminal concepts. On the other hand, in "Japanese WordNet", the concept of "aquatic mammal" exists as a subordinate category of "aquatic mammal". Therefore, we added an

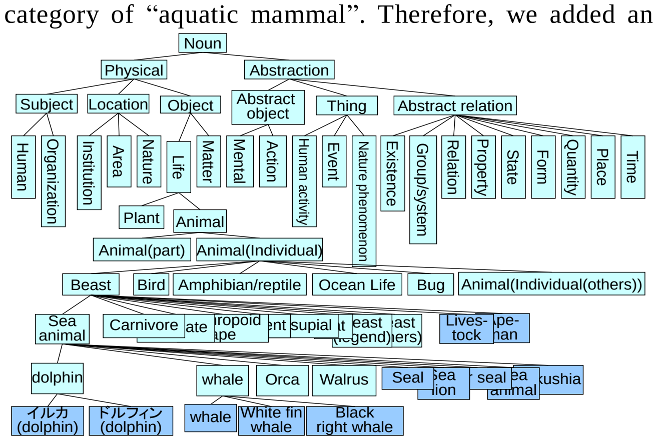

Figure 4: The overall structure of the noun concept hierarchy

Description Format

```
(set (intern [noun concept]) '(([subordinate concept(terminal)])
      ([depth]  [number of superordinate concept] ([range of subordinate concept]))))
```

Examples

Intermediate Concepts List

```
…(1025 "*獣[海獣]*<sea animal>") (1026 "*獣[イルカ]*<dolphin>") (1027 "*獣
[鯨]*<whale>") (1028 "*獣[シャチ]*<orca>") …
```

Hierarchy Structure

```
…
(set (intern "*獣[海獣]*<sea animal>")
          '((海豹<seal> 海驢<sea lion> 膃肭臍<fur seal>
                    海獣<sea animal> 儒艮<dugong> バークシア<bakushia>)
          (8 1022 (1026 1029))))
(set (intern "*獣[イルカ]*<dolphin>")
          '((イルカ<dolphin> ドルフィン<dolphin>)
          (9 1025 (nil))))
(set (intern "*獣[鯨]*<whale>")
          '((鯨 白長須鯨<white fin whale>
                    せみ鯨<black right whale>)
          (9 1025 (nil))))
(set (intern "*獣[シャチ]*<orca>")
          '((逆叉<orca> 鯱<orca>)
          (9 1025 (nil))))
```

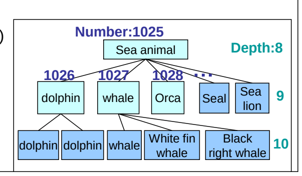

Figure 5: The format and an example of noun concept hierarchy

intermediate concept, “sea-animal”, under “beast”, and moved “seal” and “sea-lion” to the subordinate category.

A Tentative Evaluation

For the concepts that constraint sets are applied, we attempted a confirmation of the validity of the constraints and the classification of noun concepts regarding to the constraints. The criterion is whether physically possible events can be generated by each constraint. For example, we determine that such event as “A salesgirl divorces her husband” is possible. In contrast, such event as “A spokesman is sailing on the paper” is impossible.

We prepared an experimental event generation program which selects a noun concept in the noun concept hierarchy for each case in the case frame at random according to the constraint. And, we generated 20 events by 100 case frames selected at random from 1158 case frames that all constraints were set. Three evaluators checked the generated 2000 events according to the criterion of possible/impossible. As a result, if 14/20 or more events were “physically possible”, we decided that the constraints in the case frame were comparatively adequate. When the results in the evaluators were different, we employed the result of majority. The success was 70 and the failure was 30. For 12 in the 30 failures, we could modify them by applying another set of constraint. For example, in the case of “N1 collect N2 from N3”, we may replace the set of “N2:object” from “goods, status, and notion to be able to send” to “goods to be able to send”. For the 15, we modified the constraints themselves. For example, in the case of “N1 sail N2”, we changed the constraint of “N2:location” to ”(river waterway lake sea)” from ”(river waterway lake sea sky)” which is corresponding to a set, “the place to be able to pass with ship and aircraft”. Other two ones are the description errors in case frames. Remaining 1 problem is the case that the verb concept itself is impossible like “N1 cast a spell on N2”. By applying the above criterion of evaluation, the difference of feelings by evaluators and the ambiguity of evaluation diminished more than previous criterion, natural/unnatural. This brings the simplicity of system development. However, there is also the difficulty that possible but ordinarily unnatural events can not be eliminated. Because it is difficult that the conceptual dictionary absorbs this sort of contextual knowledge, the solving needs to be given in the relation with the above-mentioned defamiliarization rhetoric. We are considering a basic framework that the conceptual dictionary gives comparatively simple and standard conceptual knowledge and a variety of narrative knowledge such as the defamiliarization operates and adjust more complex and advanced literary or artistic rhetorical techniques.

Conclusions

We explained the current version of a conceptual dictionary containing two hierarchies of verb concepts and noun concepts to be functioned in our narrative generation system. It is used for operating naturalness or validity of generated events and realizing or adjusting the intentional defamiliarization. Namely, this dictionary is a mechanism to be able to flexibly adjust a variety of generation from realistic narratives to fantastical narratives as well as the foundation for a narrative event and the elements. In the current version, verb concept dictionary has originally defined 5338 case frames and modified 1158 constraints and noun concept dictionary contains 121573 concepts including 5808 intermediate concepts.

References

Akimoto, T., & Ogata, T. (2011a). Computational model of narrative discourse theory and reception theory in narratology and its implementation. *Proc. of the 13th Japanese Society for Language Sciences* (pp. 155-156).

Akimoto, T., & Ogata, T. (2011b). A consideration of the elements for narrative generation and a trial of integrated narrative generation system. *Proc. of the 7th NLPKE* (pp. 369-377).

Bates, J. (1991). Virtual reality, art, and entertainment. *The Journal of Teleoperators and Virtual Environments*, 1(1), 133-138.

Bond, F., Isahara, H., Fujita, S., Uchimoto, K., Kuribayashi, T., & Kanzaki, K. (2009). *Proc. of the 7th Workshop on Asian Language Resources, ACL-IJCNLP* (pp. 1-8).

Bringsjord, S. & Ferrucci, D. A. (2000). *Artificial Intelligence and Literary Creativity: Inside the Mind of BRUTUS*, a Storytelling Machine. New Jersey: Lawrence Erlbaum.

Genette, G. (1972). *Discours du récit. essai de méthode, Figures III*, Seuil: Paris. (Transl. Lewin, J. E. (1980). *Narrative discourse: an essay in method*. NY: Cornell University Press.)

Ikehara, S., Miyazaki, M., Shirai, S., Yokoo, A., Nakaiwa, H., Ogura, K., Ooyama, Y., & Hayashi, Y. (1999). *Goi-Taikei–a Japanese Lexicon CDROM*. Tokyo: Iwanami Shoten. (in Japanese)

Imabuchi, S., & Ogata, T. (2011). A story generation system based on Propp combined with a conceptual dictionary. *Proc. of the 7th NLPKE* (pp. 359-362).

McIntyre, N. & Lapata, M. (2009). Learning to tell tales: A data-driven approach to story generation. *Proc. of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP* (pp. 217-225).

National Institute of Information and Communications Technology (2001). *EDR Electronic dictionary technical guide*. http://www2.nict.go.jp/r/r312/EDR/J_index.html (Last access: 1.11.2012).

Ogata, T. (2008). Towards the movie construction in narrative generation system. In A. Kanai & Y. Niwa (Eds.), *Theory and practice of film editing*. Tokyo: Hosei University Press.

Ogata, T., Akimoto, T., & Seito, A. (2011). A new version of circular mutual transformation system between music and narrative. *Proc. of the 25th JSAI* (1H2-OS1-13in). (in Japanese)

Ogata, T., & Kanai, A. (2010). *An introduction of informatics of narratology: on thought and technology of narrative generation*. Tokyo: Gakubunsha. (in Japanese)

Oishi, K., & Ogata, T. (2011). Towards the development of conceptual dictionary for narrative generation system. *Proc. of the 7th NLPKE* (pp. 351-358).

Okada, N. (1991). *Representation and strage of word concepts*. Tokyo: The Institute of Electronics, Information and Communication Engineers. (in Japanese)

Okada, N., & Endo, T. (1992). Story generation based on dynamics of the mind. *Computational Intelligence*, 8(1), 123-160.

Onodera, K. & Ogata, T. (2012). Sequence generation based on mutual relationship between state and action: As a mechanism in narrative generation system. *Proc. of the 4th DIGITEL* (pp. 159-161).

Propp, V. (Пропп, В. Я.) (1969). *Морфологиясказки*, Иэ, 2е. Москва: Наука. (Transl. Scott, L. (1968). *Morphology of the Folktale*. Austin, TX: University of Texas Press.)

Takeuchi, K. (2011). Construction of thesaurus of predicate-argument structure for Japanese verbs. *Proc. of the 25th JSA I* (3H2-OS3-5). (in Japanese)

Zhang, Y., Ono, J., & Ogata, T. (2011). An advertising rhetorical mechanism for single event combined with conceptual dictionary in narrative generation system. *Proc. of the 7th NLPKE* (pp. 340-343).

Zhang, Y., Ono J., & Ogata, T. (2012). Single event and scenario generation based on advertising rhetorical techniques using the conceptual dictionary in narrative generation system. *Proc. of the 4th DIGITEL* (pp. 162-164).