

# 名詞／動詞概念辞書の開発と物語生成システムにおける利用 Development of Noun/Verb Conceptual Dictionaries and the Use in the Narrative Generation System

栗澤 康成<sup>\*1</sup>  
Yasunari Kurisawa

大石 顕祐<sup>\*2</sup>  
Kensuke Oishi

小方 孝<sup>\*3</sup>  
Takashi Ogata

<sup>\*1</sup> 岩手県立大学大学院  
Graduate School of Iwate Prefectural University

<sup>\*2</sup> 富山県警察  
Toyama Prefectural Police

<sup>\*3</sup> 岩手県立大学  
Iwate Prefectural University

In our narrative generation system, conceptual dictionaries for noun concepts and verb concepts are central mechanisms for manipulate case information in an event which is the basic element in narrative conceptual structure. In this paper, we describe the concept and current version. Noun concept dictionary has 5808 intermediate node (concepts) and 115765 terminal concepts, and verb concept dictionary has 12174 verb concepts and 15234 pairs of case frame and constraint. We describe some issues to be solved and show the future plan including the development of noun attribute frames.

## 1. まえがき

筆者らが構想・開発している物語生成システム(構想の詳細は[小方 2010])における概念構造生成の部分(物語生成及び物語言説の生成部)のために動詞概念及び名詞概念の概念辞書の開発を進めて来た[Oishi 2011] [大石 2012a] [Oishi 2012b]. 本稿では一旦構想レベルに立ち返って概念辞書開発の構想を示した後開発の現状を整理し、現状での問題点及び次の作業の目標を設定する。

## 2. 物語生成システムにおける概念辞書の構想

概念辞書としては以下の三種類が必要である。

最初に、動詞概念辞書は動詞概念を階層的に分類した体系であり、各動詞概念にはそれが取る深層格の種類及びその深層格に入る概念の範囲(制約条件)が定義される。これらは特に物語内容生成において事象概念を構成する際の最も基礎的な知識となる。また、動詞概念ごとの基本的な文の形態を表す文型パターンが格納される。これは文章表現の生成に利用する。

名詞概念辞書は、物語に登場する人物、物、場所、時間等のインスタンスを構成するための基礎的な知識となる階層的体系であり、各名詞概念にはそれが持つ基本的な属性情報が付与される。動詞概念辞書における制約条件は名詞概念辞書の階層構造における範囲指定により定義される。

修飾概念辞書は、動詞概念や名詞概念に対する修飾的な役割を持つ概念(主に形容詞概念と形容動詞概念)を階層的に体系化する。林檎は「赤い」のように、名詞概念の属性情報に付与する情報として利用される。またこの修飾的な属性情報は、物語言説に描写表現を挿入する際に参照される。

図 2 はこれらを含む概念辞書と物語の概念表現機構との関係を示す。物語生成システムにおける物語の概念表現は、物語の最小単位である事象概念を物語関係で結び付けた物語木(物語を木構造で表したものの)の形式となっている。また、ひとつの事象概念の前後には、物語内の登場人物や場所等の状態を表す状態の情報が付加される。物語内容機構が事象概念を生成する際、概念辞書は動詞概念体系から格フレームと文型パターンの情報を提供し、名詞概念体系からは制約条件に沿った名詞概念を提供する。一方名詞概念や修飾概念を参照する際、

概念辞書は名詞概念に記述された概念フレームによって概念の持つ意味情報や、修飾語によって示された形容的状態情報を状態情報に付与する役割を果す。

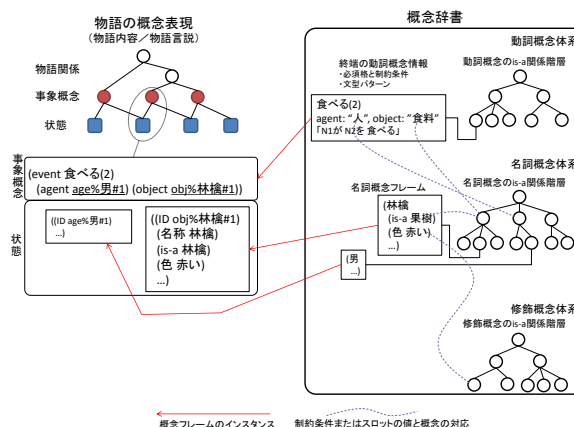


図1 概念辞書と概念表現機構との関係

また、動詞概念辞書には常識的な意味で可能な事象を生成するのに適切な制約条件が設定されるが、その制約条件の範囲を意図的に拡大させることで、いわば異化的な事象の生成を可能とする。

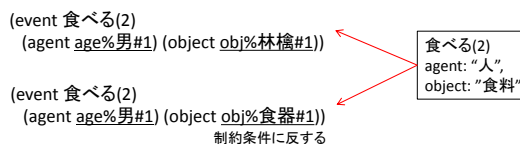


図2 制約条件による常識的事象生成と異化的事象生成の例

さらに必要な機構は、各名詞概念が保持する属性情報(名詞概念やその他の定数)の集合であり、これを名詞属性フレームと呼ぶ。動詞概念とその格要素から成る事象が動詞概念のためのフレームだとすれば、これはその名詞概念版に当る。

名詞属性フレームを構成する属性情報に関して、[熊谷 2010]は小説の分析を通じて人物に関する属性スロットの素案を示した(表 1)。今後はこれを拡張して行く予定である。

連絡先: 栗澤康成, 岩手県立大学大学院ソフトウェア情報学研究科, 滝沢村滝沢字菓子 152-52, g231k013@s.iwate-pu.ac.jp

表1 人物のフレームの属性情報 ([熊谷 2010]より引用)

ID		変更後
time		
location		
短期的情報		
所持		
体力		
外的状態	体勢	その他
心理的特性	態度	信念
社会的特性		
内的関係: 血縁関係など		
長期的情報		
is-a		
名前		
性別		
身体情報		
住所		
経験: 人物の過去の情報		
社会的特性		
所属集団		
分属集団		
外的関係: 友人関係など		
個人的特性		
年齢		
職業		
経済状況		
性格		
行動パターン		
心理的特性	信念	

以上のような構想に基づいて名詞概念辞書及び動詞概念辞書の開発を進めており、以下の二つの節では、それぞれの開発の現状を要約する。

### 3. 名詞概念体系の概要

#### 3.1 名詞概念体系の構造

図3は名詞概念体系の構造である。中間概念 5808 個、終端概念 115765 個を格納している。終端概念は具体的な個々の名詞概念に対応し、中間概念はそれらを束ねる分類項目かその分類項目をさらに束ねる項目である。全体は単一継承の体系を成している。中間概念のうち、比較的上位概念は『日本語語彙大系』[池原 1990]の属性情報を参考に構築し、比較的下位の体系には『日本語 WordNet』[Bond 2009]を参考に追加ないし修正を施した。終端概念には『日本語語彙大系』の登録語由来の情報を使用している。例えば、“獣”は、『日本語語彙大系』の属性情報に由来し、その下位概念“獣[海獣]”は新たに追加した中間概念である。また、“獣[鯨]”の下位概念“鯨”、“白長須鯨”、“せみ鯨”は、『日本語語彙大系』の語彙情報由来で、元々は“獣”と直接関連付けられていた。

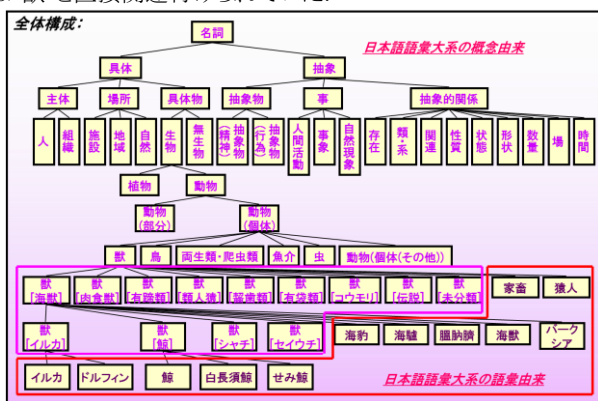


図3 名詞概念体系の構造

なお、この名詞概念体系と連結される言語辞書の開発も行っている。現在の言語辞書は、一つ概念に対し漢字・平仮名・片仮名・ローマ字による語彙情報を保持する形式となっている。これは文生成機構によって主に利用される。文生成機構は、概念辞書を利用して生成された物語の事象概念を入力とし、それを自然言語文に変換する機能である。

#### 3.2 記述方法

名詞概念体系は、各中間概念の通し番号のリストと、中間概念ごとに階層情報を保持する変数群として Common Lisp で記述される。図4に記述形式と記述例を示す。中間概念の通し番号は、名詞概念体系の中間概念のみを深さ優先探索で辿る時と同様の順番で割り振られている。中間概念ごとに作成した変数は、下位概念の集合の情報と階層の深さの情報を保持している。また、通し番号によって上位概念と下位概念の範囲が規定される。

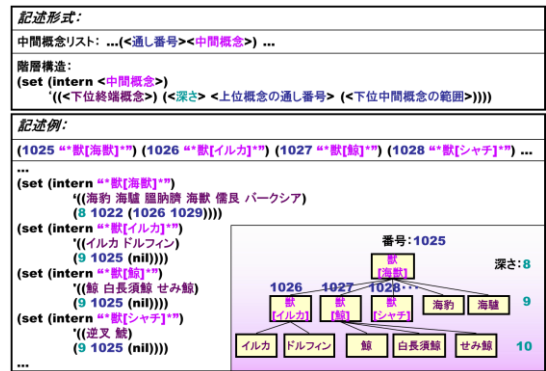


図4 名詞概念体系の記述形式と記述例

#### 3.3 構築過程:登録語の概念化及び中間概念の追加

次のような手順で名詞概念体系の構築を行った。まず、『日本語語彙大系』には「一般名詞意味属性体系」の 2710 個の属性(本研究の概念辞書における中間概念に相当)に対して、延べ 145360 個の語彙が登録されている。これらの語彙には「狸」と「タヌキ」のような同音同義語が含まれる。この登録語彙から同音動義語を取り除き、115765 個の概念を本辞書に収載した。

次に、詳細な制約条件の設定を可能にする目的で『日本語 WordNet』を参考に中間概念の追加(分類の詳細化)を行った。図5はその具体例である。ここで、制約条件とは後述の動詞概念体系に登録された個々の動詞概念が持つ、その格要素が取り得る値の範囲であり、名詞概念体系の部分と対応する。同音同義語を纏め概念として登録した段階では多くの終端概念が同じ一つの中間概念と直結していた。例えば、“海豹”と“狸”は共に“獣”の下位概念で、それ以上の細かい分類はされていなかった。一方『日本語 WordNet』には水棲動物の分類として「aquatic mammal」という概念が登録されている。そこでこの概念を参考に、“獣”の下に“\*獣[海獣]\*”という中間概念を追加し、“海豹”と“海驢”をその下位に移動した。



図5 中間概念の追加方法

図6は、中間概念の追加作業の前後で、各階層の中間概念における下位概念の数がどのように変化したかを示す。作業前は、5 階層から 12 階層における下位概念の数は平均 60 個~40 個であったのに対して、作業後には平均で 40 個~10 個以下になっており、実際により詳細な分類が実現されている。

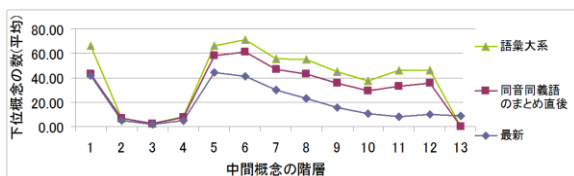


図 6 各階層の中間概念における下位概念の数

3.4 現状の問題点

図 7 に示す, 同義概念の重複, 対義概念の同一範疇化, is-a 関係以外の要素の存在, 妥当性の疑わしい分類が現時点での問題点として挙げられる. 例えば, 名詞概念体系は is-a 関係に基づく体系であるが, 一部 part-of 関係も混じっている. この例は“顔面”の全体部分関係を示しているが, その他特に場所のような空間的概念の場合 part-of 関係をより体系的に整理することが必要になって来ると思われる.

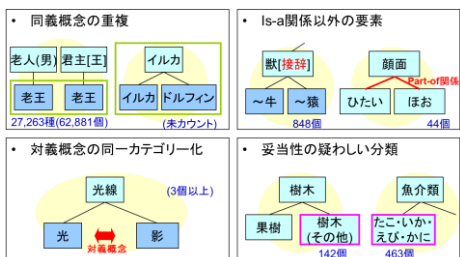


図 7 名詞概念体系における現状の問題点

4. 動詞概念体系の概要

4.1 動詞概念体系の構造

図 8 に動詞概念体系の構造を示す. 36 個の中間概念に対して終端概念 12174 個が定義されている. 各終端概念には一つの文型パターン(その動詞概念を述語とする文型の雛形)が定義され, ひとつ以上の格フレームと制約条件の対 15234 組が関連付けられている. 例えば, 動詞概念“食べる(2)”は, “身体変化”の下位概念で, 文型パターン「N1 が N2 を食べる」と格フレーム“(食べる(2) (agent N1)(counter-agent N2))”及び“(食べる(2) (agent N1)(object N2))”, 各制約条件の項に対応する制約条件, 上位概念が“身体動作”であるという情報を記述している. 動詞概念体系の階層構造と文型パターンは『日本語語彙大系』に基づいて構成し, 格フレームは新たに作成し, 制約条件は『日本語語彙大系』において定義されているものを本研究の目的に沿う形に大幅に変更した(主に詳細化).

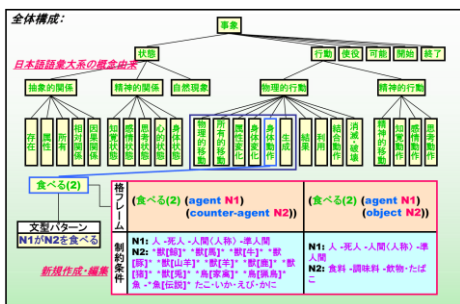


図 8 動詞概念体系の構造

4.2 記述形式

図 9 に動詞概念体系の記述形式と具体例を掲げる. 各終端概念は, 階層情報, 文型パターン, 格フレーム-制約条件対の

情報を保持する変数であり, Common Lisp で記述されている. 各動詞概念の情報は, 同名の変数に格納されている.

```

記述形式:
(set (intern <動詞概念>) '(name (<動詞概念名>))
(sentence-pattern <文型パターン>)
(case-cons-set
((case-frame ((agent [<名詞項> nil]) (counter-agent [<名詞項> nil])
(location [<名詞項> nil]) (object [<名詞項> nil])
(instrument [<名詞項> nil]) (from [<名詞項> nil]) (to [<名詞項> nil]))
(constraint (<制約条件(名詞概念の配列)>))))(is-a (<上位概念>))))

記述例:
(set (intern "食べる(2)")
(name (食べる(2)))
(sentence-pattern "N1がN2を食べる")
(case-cons-set
((case-frame ((agent N1) (counter-agent N2) (location nil)
(object nil) (instrument nil) (from nil) (to nil)))
(constraint ((("人" "死人" "人間(人形)" "準人間")
("鳥[鶏]" "鳥[鷹]" "鳥[馬]" "鳥[牛]" "鳥[豚]" "鳥[山羊]" "
"鳥[羊]" "鳥[鹿]" "鳥[猪]" "鳥[蛙]" "鳥[兎]"
"鳥[家畜]" "鳥[猛鳥]" "鳥[魚]" "鳥[魚伝説]"
"たこいか" "えび" "かに"))))
((case-frame ((agent N1) (counter-agent nil) (location nil)
(object N2) (instrument nil) (from nil) (to nil)))
(constraint ((("人" "死人" "人間(人形)" "準人間")
("食料" "調味料" "飲料" "たばこ")))) (is-a (身体動作))))))
    
```

図9 動詞概念体系の記述形式

4.3 格フレームの作成方法

動詞概念体系は, 各動詞概念に対応する①格フレームの作成及び②制約条件の編集の順で作成した. ここでは①を, 次節で②を説明する.

最初に前処理として『日本語語彙大系』の「用言意味属性体系」を構成する属性と, 用言のうち動詞を述語とするものを抽出し, Lisp コードに変換した. 具体的にはデータ抽出プログラムを作成し, 動詞概念の階層的分類と, 動詞概念とその文型・制約条件を抽出した. 次に, 階層的分類については手作業で, 文型・制約条件については自動変換プログラムを作成して, Lisp コードに変換した.

次に, これらの情報に対して格フレームの作成を行った. 抽出情報は, 文型パターン・制約条件の情報とセットになっている. 例えば, 動詞「食べる」の場合, 「用言意味属性体系」に登録されている対応する文型パターンは「N1 が N2 で食べる」, 「N1 が N2 を食べる」の二つであり, それぞれに制約条件が定義されている. そこで, 文型パターン-制約条件の対ごとに格フレームを定義した. 図 10 に沿って具体例で説明する. 「N1 が N2 で食べる」の場合, N2 の制約条件が“仕事”であることから, N2 に対応する深層格は行為に用いられる道具を表す深層格: instrument であると定義した. また, 「N1 が N2 を食べる」の場合, N2 の制約条件は(“生物”, “食料”)である. この場合, N2 が“生物”の場合と“食料”の場合で, 被行為者: counter-agent, 行為対象: object がそれぞれ当て嵌まる. このような場合, 一つの文型パターンに対して二つの格フレームが作成される.

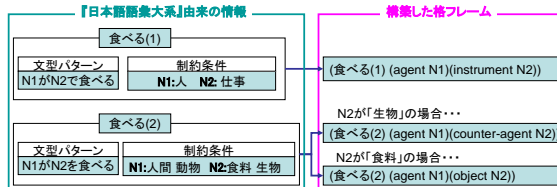


図10 格フレームの作成方法

4.4 制約条件

『日本語語彙大系』における制約条件は粗過ぎると判断された. 例えば, “食べる(2)”の場合, この制約条件をそのまま使用すると, 「たばこを食べる」のような事象が生成される. そこで制約条件設定の作業を新たに行った. しかし, 制約条件の範囲をどのように決めるのかは大変難しい問題である. 本研究でのゆるやかな目標は, ある事象を, その格要素をランダムに設定して

も、常識的に・日常的な感覚において妥当な範囲の名詞概念が含まれた事象が生成される、というものである。しかし、常識的・日常的な制約は論理的に決まるものではない。本研究では、人間の感覚に基づいて、人手によりすべての動詞概念における制約概念を設定し、その後評価を通じてそれを修正して行くものとした。

制約条件の基本的な編集方法は、①より下位の概念による定義と、②ある概念からの特定範囲の除外の 2 種類である。それぞれの例を図 11 に示す。

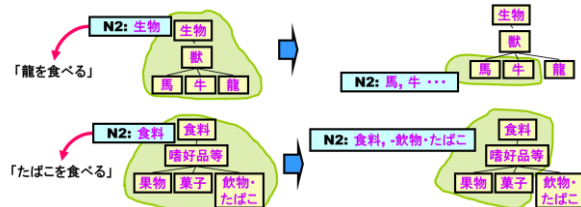


図11 制約条件のふたつの編集方針

この方法で、予備的に 300 個の格フレームの制約条件を編集したが、この方法をすべての項で行うと作業時間が膨大になるため、より組織的な方法として制約条件のセットを作成し、対応する項にセットを流用する方法を考案し、残りのすべての動詞概念の制約条件の設定を行った。例えば“N1 が N2 に引越す”の N2 の制約条件を、「居住可能な場所」を意味する一つのセットと見なし、複数の動詞概念にこのセットを適用する。

動詞概念体系では「物理的・論理的に可能な事象」を生成する制約条件の設定し、評価実験を行った。評価実験は、動詞概念体系から抽出した 150 個の格フレームを用いて生成した事象概念 3000 個に対し、可能か不可能かでの評価を行った。その結果、123 個の格フレームで目標に達する評価が得られた。

しかし、制約条件の設定だけでは解決できない問題も残っており、その一つが「N1 が N2 を N3 から N4 に/へ 流す」のように項が多いパターンである。このような格フレームを持つ動詞概念は、各制約条件自体にはなんら問題が無い場合(N2: “血”, ”涙” N4: “外傷のできる部位”, “目(器官)”)でも、「腕から涙を流す」のような不可能な事象が生成されやすい。

このようなパターンを定義する方法として、①制約条件に(N2: “血” N4: “外傷のできる部位”), (N2: “涙” N4: “目(器官)”)という二通りのパターンを定義する、②名詞概念“涙”の属性として、発生する部位: “目(器官)”を定義するという 2 通りの案が考えられる。しかし、前者を採用する場合は動詞概念体系の、後者を採用する場合は名詞概念体系の形式を変更する必要がある、今後の課題となっている。

## 5. 物語生成システムにおける概念辞書の利用法

構築した概念辞書は、物語生成システムにおける物語内容生成の機構で主に使用される。また、動詞概念体系の文型パターンは文生成機構で参照される。物語内容生成における主な利用方法は、①制約条件の範囲内で事象概念を生成する「自然な事象の生成」[Imabuchi 2011]、②概念辞書以外の知識を用いて生成した事象概念が制約条件の範囲内の概念で構成されているかをチェックする「事象における概念の意味チェック」[小野寺 2011]、③意図的に概念辞書での定義から逸脱した事象概念を生成する「異化的な事象の生成」[Zhang 2011]の 3 種類である。「自然な事象の生成」と「事象における概念の意味チェック」は概念辞書の基本的な利用方法であり、定義された制約条件の範囲に従った処理を行う。それに対して、「異化的な事象の生成」では、異化対象の名詞概念が挿入されている項に

定義されている制約条件を参照し、その範囲外の名詞概念を取得することで異化的な事象を生成する。

## 6. あとがき

物語生成システムにおいて主に事象概念の構成要素に情報を提供するために使用される、動詞概念・名詞概念・修飾概念のための概念辞書の開発を進めて来たが、本稿では、その構想を再整理し、開発の現状を要約した。現在実際に開発されているのは名詞概念辞書と動詞概念辞書である。名詞概念辞書は現在、中間概念 5808 個と終端概念 115765 個で構成されており、動詞概念辞書においては動詞概念 12174 個に対し、格フレームと制約条件の対 15234 組が設定されている。今後は、現状の説明中に記述した各種問題点の解決が直近の作業となるが、同時に、修飾概念辞書の開発及び暫定的な処理を行っている名詞属性フレームの開発に着手する必要がある。前者は物語言語説における描写の生成等に貢献し、後者は登場人物や場所等物語中に出現する諸概念に意味的厚みを付与する。

## 参考文献

- [Bond 2009] Bond, F., Isahara, H., Fujita, S., Uchimoto, K., Kuribayashi, T. & Kanzaki, K.: Enhancing the Japanese WordNet, The 7th Workshop on Asian Language Resources, ACL-IJCNLP 2009, 1-8, 2009.
- [池原 1990] 池原悟・宮崎正広・白井諭・横尾昭男・中岩浩巳・小倉健太郎・大山芳史・林良彦: 日本語語彙体系 CD-ROM 版, 岩波書店, 1990.
- [Imabuchi 2011] Imabuchi, S. & Ogata, T.: A Story Generation System based on Propp Combined with a Conceptual Dictionary, 7th International Conference on Natural Language Processing and Knowledge Engineering, 359-362, 2011.
- [熊谷 2010] 熊谷友子: 物語生成システムへの適用を目的とした人物情報の分析, 岩手県立大学ソフトウェア情報学部卒業論文, 2010.
- [小方 2010] 小方孝, 金井明人: 物語論の情報学序説—物語生成の思想と技術を巡って—, 学文社, 2010.
- [小野寺 2011] 小野寺康, 小方孝: 物語生成システムにおける物語内容の最下層を構成するための状態—事象相互変換機構, 人工知能学会第二種研究会ことば工学研究会(第 39 回)資料, 23-28, 2011.
- [Oishi 2011] Oishi, K. & Ogata, T.: Towards the Development of Conceptual Dictionary for Narrative Generation System, Proceedings of the 7th International Conference on Natural Language Processing and Knowledge Engineering, 351~358, 2011.
- [大石 2012a] 大石顕祐・小方孝: 物語生成システムのための名詞/動詞概念辞書の構築と応用, 信学技報 (IEICE Technical Report), 25-30, 2012.
- [Oishi 2012b] Oishi, K. and Ogata, T.: The Development of Conceptual Dictionary for Narrative Generation System The Structure and Functions, 4th IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, 168-170, 2012.
- [Zhang 2011] Zhang, Y., Ono, J. and Ogata, T.: An Advertising Rhetorical Mechanism for Single Event Combined with Conceptual Dictionary in Narrative Generation System, 7th International Conference on Natural Language Processing and Knowledge Engineering, 340-343, 2011.