

7S-4

概念体系の開発思想

高井貞治, 岸本行生, 三池誠司, 松川智義, 横田英司
 (株)日本電子化辞書研究所

1. はじめに

(株)日本電子化辞書研究所(EDR)では, 自然言語の意味処理を可能にする, 大規模な概念辞書の研究開発を行っている。概念辞書は概念記述と概念体系によって構成される。概念記述は概念間に成立し得る結び付きを表し, 概念体系は概念の階層関係を表す。昨年度までに, 例文に基づいた概念記述データの収集を行うとともに, 概念体系の試作を行った。本発表では, 概念体系の目的, 及び体系試作の手順と方法論について報告する。

2. 概念体系の目的

EDR概念辞書の中心をなす概念記述は, ある概念が他の概念とどのような関係をもって結び付き得るかという情報を記述している。概念記述の素データとなる概念関係の情報は, EDRでは大量の例文の解析によって抽出している。しかし, いかに大量の文を解析しても, 大規模なEDR単語辞書に含まれるすべての概念の関係が収集できるとは限らない。また概念間のあらゆる関係を個別に記述すれば, 概念辞書の規模が膨大となり, 利用する上でも効率的ではない。したがって, 概念辞書としては, 概念間の関係の記述量をコンパクトにすると同時に, 例文からは漏れていた概念関係の情報を補う方策が必要である。

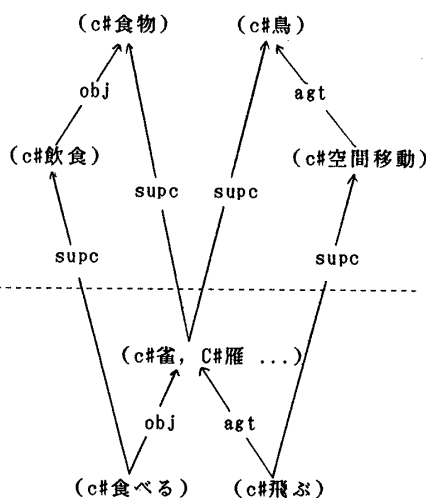
本概念体系は, 他の概念との関係の観点からみて等しいふるまいをする概念をグループ化し, そのグループ毎に上位概念を設定することによって, 概念を体系化したものである。図1に上位概念と下位概念の関係の模式を示す。概念体系は, このような上位-下位関係に基づいて概念の属性継承を行い, 概念間の関係の演繹的な推論を可能とするものである。

概念体系による概念間関係の推論によって, 以下の2点が可能となる。

- ①推論可能な個々の概念記述を削除することにより, 概念記述のデータ量を取扱い可能な量にする

- ②あらゆる概念を体系の中に位置づけることにより, 例文から得られなかった概念関係に情報を与える

<上位概念>



<個別概念(下位概念)>

註. (C#X)は, 単語Xによって表される概念の一つを表す。また関係子は以下のとおり;
 agt: 動作主, obj: 対象,
 supc: 上位-下位関係

図1. 概念関係の属性

3. 概念体系の開発手順

概念体系の開発は, 概念記述データの開発と並行して行う。図2に開発手順の概略を示す。

- ①上位概念の候補となる分類項目を決定する^[2]。
- ②一方で, 概念間の関係を示す関係子セットを決定する^[1]。
- ③単語辞書に含まれる概念を, 決定した分類項目によって分類する。分類作業の過程で, 必要に応じて分類項目を調整, 改良する。
- ④一方で, 分類項目を代表概念と見なして, その間の関係記述を行う^[3]。

PHILOSOPHY AND METHODOLOGY OF THE HIERARCHICAL CLASSIFICATION OF CONCEPTS
 Sadaharu TAKAI, Yukio KISHIMOTO, Seiji MIIKE,
 Tomoyoshi MATSUKAWA, Eiji YOKOTA, Shin-ya AMANO
 Japan Electronic Dictionary Research Institute, Ltd.

- ⑤分類項目間の関係記述をもとにして，上位概念を設定し，体系化する．また，分類データによってその体系を検証する^[4]．
- ⑥並行して，大規模なテキストの解析によって概念記述データを得る．
- ⑦実例による個々の概念記述データと，体系から演繹される概念関係を，個々の概念の分類データを介して照合し，体系を検証，改良する．
- ⑧これらの過程の繰り返しによって，互いに矛盾のない概念体系と概念記述を得る．

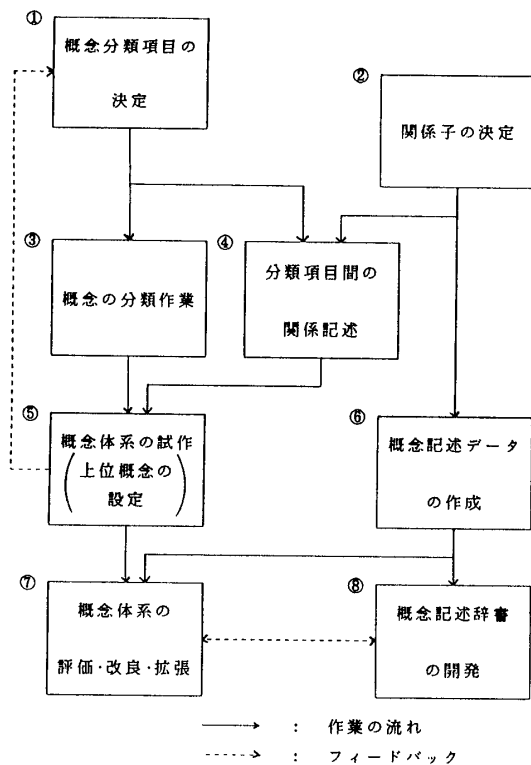


図2. 概念辞書の開発手順

単語辞書に含まれる概念の分類作業(③)は基本的に人手によって処理し，分類項目間の関係記述(④)は，内省によって行った．概念記述データの作成は，計算機のサポートによって大規模な例文の構文・意味解析データを作成し，それを人が確認して完成するという手順で行っている．また，概念体系の構築は，分類項目間の関係記述データから関係のパターンを抽出するツールを作成し，計算機のサポートによって効率的に行えるようにした^[4]．

4. 概念体系の内容

前述したように，本概念体系は，概念間の関係の要因となる概念の属性を上位概念として設定し，その観点からの属性継承を目的とした体系である．概念はそれぞれが様々な関係の中心となるので，属性の束として表すことができる．したがって，本体系は，属性の多重継承を許す比較的浅い階層構造となる．

属性継承の健全性については，実際の例文データから収集した，個別の概念記述からフィードバックを行うことによって，その検証を行う．一方，体系項目間の概念記述に演繹することのできない，例外事項や個別の事実・事象に関わる概念記述データは，この過程で抽出され，体系項目間の概念記述に吸収されることなく，概念辞書にそのまま保存される．

5. まとめ

以上，概念体系の開発方法について述べた．本概念体系は，概念記述データ量の圧縮という単純な目的に沿ったものなので，構造も単純であり，拡張に対する柔軟性が高いといえる．その反面，辞書としての完成度や推論の過程が，トップダウン的な体系と，ボトムアップ的な個々の概念の分類・記述との相互依存性の上に成り立っているため，質の安定化の方法が大きな課題となる．

今後は，質の安定化の方法の策定と，推論の有効性の検証を併せて進めていきたい．

【謝辞】

本研究の機会を与えていただいた，EDRの横井所長，千葉第2研究室長，及び有益なご意見や多大なご協力をいただいた，内田第1研究室長をはじめとするEDRの研究員の皆様には感謝いたします．

【参考文献】

- [1] 日本電子化辞書研究所：『概念辞書（第2版）』TR-012，1989．
- [2] 岸本他：情報処理学会第41回全国大会「概念分類項目の設定」
- [3] 横田他：情報処理学会第41回全国大会「概念分類項目間の関係記述」
- [4] 松川他：情報処理学会第41回全国大会「2項関係抽象化支援システムの基本機能」
- [5] 三池他：情報処理学会第41回全国大会「概念分類項目間の関係記述に基づく概念体系の試作」