

## 概念モデルに基づく

## 概念記述の表現方法の一考察

3B-5

三輪 和弘、岸本 行生、塚田 康博、小寺 次夫

(株)日本電子化辞書研究所

まえがき

電子化辞書は日本語や英語といった言葉に依存した単語辞書とその言葉の間にある事実を知識ベースの形で表現した概念辞書から構成される。人間は言葉を理解するとき単に言葉に関する情報だけで理解しているわけではない。言葉が表している具体物や経験などから得られるイメージや事柄に関する知識(常識)を用いて理解していると考えられる。概念辞書はこの知識のうちの言語的な情報を計算機が利用できるような形式で保持する辞書である。本辞書は自然言語処理のために必須の知識ベースである。単語辞書は文法的情報と語がもつすべての概念に細分化される。概念記述は各概念について常識として成り立つ事実を記述する。事実は概念見出し間に成り立つ関係を与えることで言葉間の概念として記述される。概念辞書の中で概念体系は単語辞書で設定された概念見出し間を上位下位関係にしたがって構造化される。この上位下位関係は概念の持つ属性の継承関係を保証するものであり、概念記述の抽象化のために用いられる。本報告はこれらの辞書を包含しさらに種々の分野で用いられる常識ベースを構築するためのモデルの提案である。また記述データを検証するための支援システムについても述べる。

## 1. 概念モデル

概念を表現するための概念モデルは、人工知能・認知科学の分野で必要性が指摘されていたにも関わらず、従来きわめて小規模なものが試作されただけで、現実の知識処理に適用された例はほとんどない。人間の持っている知識を幾つものパラダイムに分けて計算機上に表現しようという試みが人工知能や認知科学の分野で盛んに行われている。人の持つ知識には共通な知識と個人によって少々異なる知識があると考えられている。モデルはこれらの差異をも克服し記述できるものでなければならない。

人の持つ概念を以下のように定義する。

## ● Base Concept

ある概念を成り立たせている基本的な概念であり、その概念自信が内包している本来の概念及び他の概念との関係によって成り立つ概念とする。

## ● Case Concept

他の概念との関係つまり一般の文中において成り立つ概念との関係によって表される概念の集合とする。

人の持つ言語に対する概念はBase Conceptを核として場面・情況に合わせてCase Conceptに定義されている語の組み合わせによって文章を構成し理解しているものと考えられる<sup>[2]</sup>つまりCase ConceptはBase Conceptを個人の経験的な規則によって変化させた物であり、その人の考え方や思考が作り出しているものである。図1にこの関係を現す。

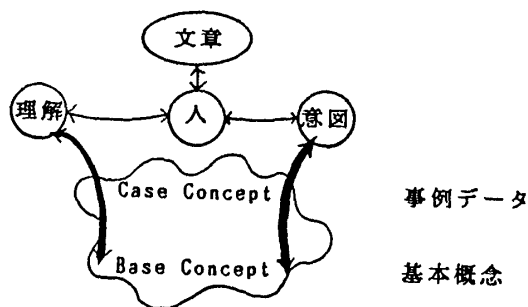


図1 概念と文章理解の関係

Base concept とCase Concept を表現するに当たって以下の制限を加える。

計算機上で明確に知識を表現するためには例外概念、否定概念をBase Conceptの中に記述するものとする。これは例外事項や知らない知識を頻繁に持つ人間の常識を処理するためには絶対に必要である。Conceptは項として記述され、項間の関係は“,”、“;”によって記述される。<sup>[3]</sup>

A conceptual discription method based on the conceptual model

Kazuhiro MIWA, Yukio Kishimoto, Yasuhiro TUKADA, Tuguo KODERA, Kazuyuki MIYATA

Japan Electronic Dictionary Reseach Institute, Ltd. Sharp System Product Corp.

## 表現形式

```

defconcept ConceptName
  is_a      SuperConcept;
  part_of  SubConcept;
  constraint
  base:
    func(Value);
    brelation1(Concept1, Concept2);
    irregal (IConcept);
    not(NConcept)
    ...
  case:
    crelation1(Concept1, Concept2);
    ...
    crelationN(ConceptN1, ConceptN2).

```

## 2. 概念辞書

概念辞書では言語を計算機が理解するために必要な常識に関する情報が形式化され記述される。人が知っている知識を計算機が利用できる形式で記述することである。こうした常識を用いられれば本当の意味での自然言語の意味解析が可能になる。しかしこの概念記述には常識自体が無限であり、また人の知識は不明確であるという問題がある。つまり量的・質的な問題がある。これらの問題を解決するために以下の方法が考えられる。

- ・ 概念記述をするための洗練された関係子の設定
- ・ 概念体系による圧縮
- ・ 基本概念と複合概念の分類

本研究はこの部分にあたる

- ・ 概念の文章データからの自動学習、自己組織化

## 3. 概念記述

概念辞書は単語辞書で設定された概念見出しについてその意味を記述するものである。つまり概念見出しが代表する事物や事象について成り立つ事実関係を記述するものである。概念は共起する言葉間の関係によって他の言葉との相異つまり意味が識別される。概念を記述する上で関係子が非常に重要である。現時点の概念記述の仕様案では、約40種の関係子が仮設定されており、それらは次の3つのクラスに分類される。

- ◎ 意味関係 : 上位下位、全体部分、属性関係
- ◎ 格関係 : 格文法にもとづく関係
- ◎ 事象関係 : 因果関係や事象の時間的關係

上記の関係子のうち、概念の上位下位関係によって関係付けられた構造を概念体系と呼ぶ。現在はこの仮関係子を用いて概念記述データを作成している<sup>[1]</sup>。

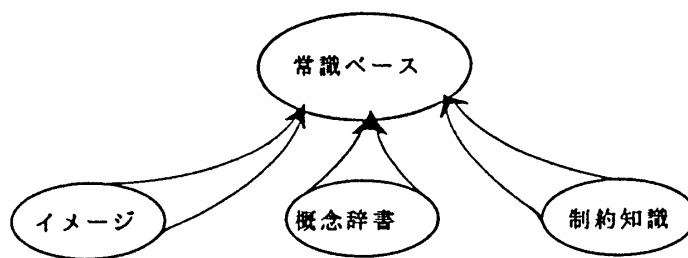


図2 常識ベースと辞書データの関係

## 4. 概念記述データの検証方法

概念記述の検証支援ツールも開発している。システムは関係付けられた概念を二項関係で表わし、表層レベルで重複チェック、矛盾チェックツール及び単言語辞書参照ツールからできている。さらに概念データの検証方法は実データに対応した評価方法が現在のところ確実であると思われるのでK W I Cによる共起関係との対応も考えている。図3に本ツールの構成を示す。

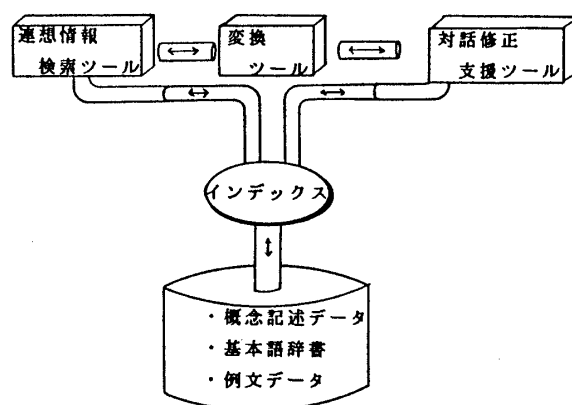


図3 検証システム支援ツールの構成

## まとめ

電子化辞書における概念辞書は常識ベースの一部であり単に自然言語処理のみでなく今後知識ベースシステムの基礎知識となることを述べてきた。また電子化辞書を常識ベースに拡張するための枠組みを示した。今後この枠組みの元で常識ベースのプロトタイプを試作していきたい。

## Reference

- [1] 岸本 他, 概念モデルに基づく概念記述の実験と評価, 情報処理学会第37回全国大会, 1988
- [2] R.C.Shank, Conceptual information processing, North-Holland, 1978
- [3] 井佐原 均, 自然言語処理と推論, 人工知能学会研究会資料, 1988