

汎電子化辞書：概念レベルの構造

横井俊夫 田中裕一 仲尾由雄 荻野孝野 野口喜洋
日本電子化辞書研究所

1 はじめに

汎電子化辞書[1]における意味記述に関しては、基本構造の記述のレベルの軸に沿い、意味の表現形式の選択に従って種々のレベルを考慮することができる。従来用いられてきた意味表現形式は、意味というものをどう考えるのか、あるいは意味のどの側面を扱うのか、表現能力の幅と前提とする推論機構(意味処理機構)の能力等に関し、それぞれに特徴を持っていて、言葉の意味を、ある種、近似的に捉えている。

これらの表現形式の中で、深層格による意味表現は、言語処理においても、また言語学においても、共通的な手法としてしばしば用いられてきた。そこでわれわれは、これを拡大・改良した形式を意味の表現に関する標準的なレベルとして採用したいと考える。このレベルを概念レベルと称する。概念レベルは次のような理由から、標準的なものとして位置付けてよいと考えられる。

- (1) 概念レベルの情報は、必ずしも深い意味処理の結果得られるものではないが、一般性は高い。そして、このレベルを対象とする意味処理機能だけでも、新たな有用な言語処理機能を実現することができる。
- (2) 表層レベルからより深い意味処理レベルへと意味表現を解釈・変換する際に、概念レベルは中間ステップとしての役割を果たす。また、より深い深層情報に関する辞書の開発の素材として有効に利用できる。
- (3) 表層に近い意味表現ということから、表層レベルの辞書を利用して概念レベルの辞書を大規模に開発することができる。

2 概念辞書の構造

2.1 基本構造

概念辞書の役割は概念に関する基礎的な知識を提供することである。この知識は、表層上異なった表現を持つ要素(語、句、文、文章、…)同士がほぼ等しい意味を持っているか否かの判定、あるいはその等しさの度合いの計算などに用いられる。

(1) 概念

概念は実体概念と関係概念とに分けられる。表層レベルの構成単位(語、文、文章、文書)に対し、それぞれ実体概念と関係概念が定義でき、構成単位間のコンテキスト関係に対応した概念の合成関係が定義できる。

	実体概念	関係概念
語概念	名詞、動詞等の概念語によって想起されるもの	助詞等の関係語によって想起される関係
文概念	文の事象表現部分によって想起されるもの	接続詞等によって想起される事象間の関係
文章概念	文列が示す事象列が想起させるもの	文書構造等が想起させる事象列間の関係

(2) 意味関係

概念は、各種の意味的な関係によって定義付けがなされる。これらの関係の成立の多くは局所的な整合性のみが保証されるものである。また、文概念に文概念が含まれるなど、循環的である。

合成関係: 概念と、その概念の下位の構成要素の概念によって合成された合成概念との間の関係。

階層関係(属種関係): 構成単位内の概念に定義される上位-下位関係(語概念、文概念)。

全体-部分関係: 全体概念と部分概念の対応関係(語概念)。

同値関係: 概念間の(ほぼ)同義の関係。

対位関係: 異言語間の概念間の同値、階層関係等。

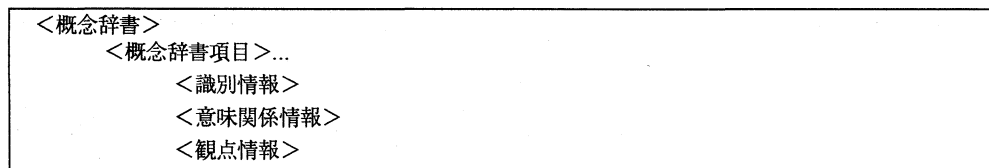
(3) 意味表現と概念表現

ある構成単位の意味表現とは、極力、忠実に詳細化した合成概念である。省略した骨格部分のみの合成概念を用いる時は概念表現と呼ぶ。

(4) 概念レベルの役割

概念レベルの役割は、概念間の関係付けの情報の内容、形式を規定することにより定まる。1つの語に対応する語義を異なった概念にいか分割するか。また逆に、1つの概念にどの程度の揺らぎを認めるか。このような規準が概念レベルの各サブ辞書の基本的な機能を規定することになる。

(5) 基本構造



2.2 語概念辞書

(1) 概念(語義)設定の規準

工学的には、概念の幅はある程度広く取り、語から概念への分割を少なくするというのが、応用を考慮した場合に重要だと考えられる。さらに、語義の立て方の規準を作るに当たって考慮すべきこととしては、generic / specificの対応関係、また、そこからのズレ、1語の語義をなす概念間の包含関係、派生語の概念と元の語の概念との対応関係、(動詞の場合など)格のフィラーの間の包含関係などがある。

(2) 階層関係(概念体系)の観点と構成

概念体系の組み立てに関しては議論の分かれるところである。連想も含め、“似た”語を集めた体系なのか、百科事典的、博物学的な体系なのか、あるいは用言との意味共起を属性とする意味素性の体系なのか、といったことが問題となる。これらは、必ずしも全く異なる体系を形作るわけではなく、主要部分は基本的に同一か対応関係のつくものになると考えられる。

<語概念辞書>
<語概念辞書項目>...
<識別情報>
<語概念表記>: 代表語義ラベル / symset / 概念カテゴリラベル
<同一化語義リスト>: symset
<概念説明>: 概念としての説明
<意味関係情報>
<合成関係>
<文概念リスト>: この語概念を含む文概念
<階層関係>
<上位概念リスト>
<下位概念リスト>
<全体部分関係>
<全体概念リスト>
<部分概念リスト>
<対位 (対訳) 関係>
<言語 1:対位概念リスト>: 同義、上位、下位
:
<観点情報>

2.3 文概念辞書

<文概念辞書>
<文概念辞書項目>...
<識別情報>
<文概念表記>: 文義ラベル、基本命題ラベル
<同一化語義リスト>
<意味関係情報>
<合成関係>
<文合成概念>: 語概念による合成概念表現
<文章概念リスト>: この文概念を含む文章概念リスト
<同義関係>
<階層関係>
<上位概念リスト>
<下位概念リスト>
<対位 (対訳) 関係>
<言語 1:対位概念リスト>
:

(1) 深層格 (関係概念) の種類

EDRでは深層格に対応する17種の関係、深層格に対応しない10種の関係を導入している。これまでなされてきた多くの議論に新たな一石を投じるつもりはないが、応用を考慮した上で必要十分な柔軟な関係を設定する必要があるだろう。

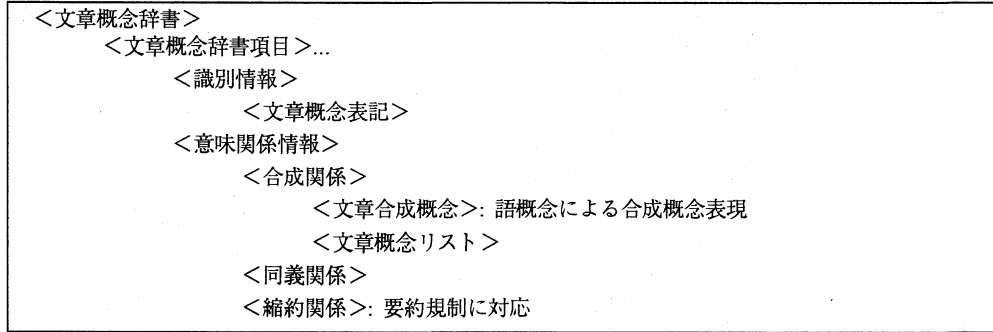
(2) 文合成概念表現と表層文表現との対応

語の関係概念 (深層格) や実体概念の選択の揺れへの対応を考える必要がある。

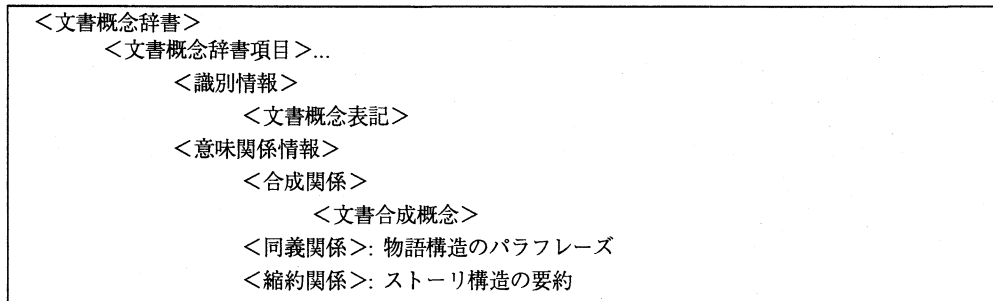
(3) 文概念の階層関係

語概念と同様、文概念にも階層関係を導入することができる。

2.4 文章概念辞書



2.5 文書概念辞書



3 EDR 電子化辞書における実現

EDR 電子化辞書 [2] の概念辞書においては、概念体系辞書に約 6000 の中間概念項目を設定し、日本語、英語に由来する計約 40 万概念を分類している。上記の記述では概念が由来する言語についてはほとんど述べなかったが、実際の作成過程では言語の違いは重要なファクタとなった。その結果、概念体系辞書は、日本語体系 (6000 項目) と英語体系 (600 項目) との 2 本立てとなっている。

4 おわりに

汎電子化辞書の概念レベルの構造および EDR 電子化辞書による実現状況を述べた。EDR においては仕様の的には概念体系辞書がほぼ要求を満たしたが、内容的にはまだ中途の段階にある。また、その上位の辞書については、データを収集しただけの段階である。今後は、概念体系辞書と概念記述辞書の充実、および単語辞書との間の整合性の向上を図りたい。

参考文献

- [1] 横井、安原、村木、原田、丸山：“汎電子化辞書：言語知識のアーキテクチャ”，言語処理学会全国大会 B3-2 (1995).
- [2] 日本電子化辞書研究所：“EDR 電子化辞書マニュアル”，<http://www.ijnet.or.jp/edr> (1995).