自 然 言 語 処 理 66 － 2
（ 1 9 8 8. 5. 2 7 ）

自然言語の意味処理用辞書の構成法

堺 和宏*，徳永健伸**，奥村学**，田中穂積**
*日本電気 **東京工業大学工学部

[概要]
本稿では，自然言語の意味処理の中核となる辞書の構成法について述べる．まず，辞書の記述内容は言語に依存する情報と言語に依存しない情報に分けられることを述べる．これまでの意味処理方式では，これらの情報を一つの辞書の中に混在させて記述していたが，本稿では，それぞれを「概念対応辞書」,「概念辞書」に独立して記述することにより意味処理用の辞書を構成することを提案する．次に，自然言語処理システム LangLAB 上で意味処理の実験をおこなうことにより，情報の混在した一つの意味辞書を利用するこれまでの方式の問題点をあげ，それらが「概念対応辞書」,「概念辞書」を利用する本方式によってどのように解決されるかを述べる．

A Costruction Method of Semantic Dictionary for Natural Language Processing

SAKAI Kazuhiro*, TOKUNAGA Takenobu**, OKUMURA Manabu** and TANAKA Hozumi**
*NEC Corporation
**Department of Computer Science, Tokyo Institute of Technology
(2-12-1 Ookayama Meguro-ku Tokyo 152 Japan)

Abstract
This paper presents a construction method of semantic dictionary for natural language processing. We claim that a semantic dictionary should be divided into two parts, a language dependent part and a language independent part. The former is called "Word-Concept Mapping Dictionary" and the latter is called "Concept Dictionary". Through an experiment on a natural language processing system, we will discuss some problems of usual semantic dictionaries, and will show how such problems are solved by our method.

1 はじめに

計算機による自然言語処理の歴史は約40年におよぶが，その技術の確立までには未だに多くの難題を残している．しかし，多くの研究者の努力と最近の計算機環境の改善によって，統語解析についてはかなりのことがわかってきている．また，知識表現に関する研究も精力的におこなわれ，「意味」の問題に取り組める状況が整ってきた．

自然言語の意味がどのようなものであるかは必ずしも明確になっていないが，本研究では自然言語の「概念的意味」を対象とする．我々は，世界に関する知識を概念の体系として持っている．自然言語の役割は，その概念体系を利用し相手に意味を伝えることであると考えてよい．つまり，ある文（章）の発話（または記述）があったときに，その発話（または記述）者がどういう概念を指していたかは重要な情報である．このような情報を以下では「概念的意味」と呼ぶ．

ところで，発話（または記述）に表われる文（章）は無限に生成可能なものであり，それぞれに対して前もって意味を与えることは不可能である．そこで，意味処理の方法としては，意味を与える最小単位を認定して，それらの意味から全体の意味を構成するという立場をとるべきであろう．辞書の役割は，その最小単位に対して意味を与えることであるから，意味処理の中核をなすものとして研究を進める必要がある．

本稿では，意味処理における辞書の役割と構造を明らかにし，その構成法を提案する．さらに，具体的な意味処理プログラムの作成を通じて，その有効性を実証する．

2章では，単語の「概念的意味」を『概念辞書』と『概念対応辞書』の2層に分けて記述し，これらに統語的情報を記述する『単語辞書』を加えて，全体として3層から成る辞書構造を提案する．『概念辞書』には，上位/下位関係に関する概念の分類体系を記述し，『概念対応辞書』には，単語と概念の対応を記述する．『概念辞書』の内容は言語に依存しない情報であり，『概念対応辞書』の内容は言語に依存する情報である．

3章では，2章で提案した辞書の有効性を実証する．まず，これまでおこなわれてきた多くの意味処理手法の問題点を考察し，2章で提案した辞書を用いた意味処理方式によって，その問題点がどのように解決されるかを具体例に基づいて述べる．

4章では，本稿のまとめと今後の課題について述べる．また，(株)日本電子化辞書研究所(EDR)で作成されている辞書の構成と本稿で提案する辞書構成との比較についても述べる．

2 自然言語の意味における辞書の役割と構成

2.1 言語の意味のとらえかた

言語現象である「文」あるいは「文章」は無限に作ることができ，それぞれに対して前もって意味を与えるのは不可能である．そこで，意味処理の方法としては，フレーゲの構成性原理に従い，単語に意味を与えて，単語の意味から全体の意味を構成する，という立場をとるべきであろう．その際に，単語の意味を記述することが辞書の役割となる．

本研究では，その単語の意味として「概念的意味」を取り上げる．「概念的意味」とは，単語が表わす概念によって与えられる意味をいう．例えば，「開ける」という動詞は，『空間をつくる』，『閉めてあったものを開放状態にする』，『営業を開始する』という概念を指す [8]が，このとき「開ける」は3つの「概念的意味」を持つといえる．

「概念的意味」を表わすのに，単語の指す概念を羅列するだけでは十分ではない．例えば，「十時に店を開けた」という文の場合，「開ける」の指し得るいくつかの概念のうちから『営業を開始する』という概念をダイナミックに決定しなければならないが，このような決定をするためには，「開ける」と共起する「店」が『商売をする場所』を指し，「十時」が『時間』を指すという情報を利用しているはずである．すなわち，「概念的意味」には，単語の指す概念とあわせてその概念が指されるための条件，特に共起する単語の統語的あるいは意味的な条件を記述しておく必要がある．

また，単語の指す概念は，知識として体系的に記述する必要があり [6]，以上をまとめると，意味処理用の辞書で記述するべきことは以下の3種類となる．

1. 単語が表わす概念
2. 単語の表わす概念を特定するための（統語的・意味的）共起条件
3. 知識として体系化された概念

2.2 辞書の構成

本研究では，「概念の分類体系」の普遍性を尊重し，辞書の構成を図1のような3層構造とする．単語辞書は統語的な情報を記述するところで，意味情報は概念辞書と概念対応辞書の2層で表現される．

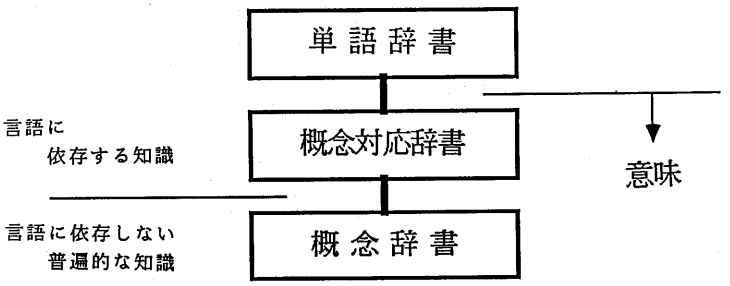

図 1: 辞書の構成

概念辞書には，上位/下位関係に基づく概念の分類体系を記述する．これは，田中 [6]や荻野 [2]らのシソーラスと基本的に同様のものであるが，シソーラスが「言葉」の上位/下位関係を記述するものであるのに対し，本研究では言語に依存しない「概念」の上位/下位関係を記述するという立場をとる．

概念対応辞書には，単語と概念体系の結び付きを記述し，それにより，単語の「概念的意味」を記述する．一般に，一つの単語は複数の概念を指すことが多いが，どの概念を指すかは他の単語との共起条件に依存する．その場合，共起条件は，言語に依存する条件と，言語には依存しない条件に分けることができる．例えば，

I took a plane.

という英文を「私は飛行機に乗った」と訳す場合に，動詞"take"に関する次のような条件は言語に依存する．

1. "take"という動詞に対し主語と目的語がそろったら，"take"は，「乗り物に乗る」という概念を指すことができ，主語に位置する名詞句の指す概念，目的語に位置する名詞句の指す概念は，それぞれ，「乗り物に乗る」という概念の agent, object になる．
2. "I"は，「私」という概念を指す．
3. "a plane"は，「飛行機」という概念を指す．

一方，次の条件は言語に依存しない．

4. 「私」という概念は「人間」という概念の下位概念である．
5. 「飛行機」という概念は「乗り物」という概念の下位概念である．
6. 「乗り物に乗る」という概念の agent は「人間」で，object は「乗り物」である．

上記の条件のうち，言語に依存しない条件 4,5,6 は「概念辞書」に記述する．一方，「概念対応辞書」では，言語に依存する条件 1,2,3 を記述することになる．

3 LangLAB システムにおける意味処理の枠組

本章では，東工大田中研究室で開発された自然言語処理システム LangLAB と知識表形式 DCKR について簡単に説明し，それらを利用した今までの意味処理方式の問題点を指摘する．その上で，2.2 節で提案した辞書構造を組みこんだ新しい意味処理方式を提案し，それによって旧方式の問題点がどのように解決されるかを具体例に基づいて述べる．

3.1 自然言語処理システム LangLAB

LangLAB システムは東工大田中研究室で開発された自然言語処理システムで次の特徴をもつ [9].

1. Prolog インタープリタに組み込みの機能をそのまま解析に利用するので，パーザを作成する必要がない．
2. ボトムアップ，縦型探索で構文処理をおこなう．

```
open ::                                          % 意味規則 1
   [subj  $  isa:event;
             isa:thingOpen          => object]
     ::                                          % 意味規則 2
   [subj  $  isa:instrument         => instrument;
             isa:wind               => reason]
   [obj   $  isa:thingOpen          => object]
     ::                                          % 意味規則 3
   [subj  $  isa:human              => agent;
   [obj   $  isa:thingOpen;
             isa:event              => object]
   (with  $  isa:instrument
             where
             obj!caller isa:thingOpen => tool)
   (with   $   isa:animal           => partner).
```

図 2: SRL/O による“open”の意味記述

3. 純粋なボトムアップではなくトップダウンの予測も利用することにより，左外置の文法規則を効率的に扱うことができる．

4. 辞書が TRIE 構造化されているので，効率的な辞書引きと柔軟な熟語の扱いができる．

5. ユーザーは，XGS [1] 形式の文法と DCG [2] 形式の辞書を記述すればよく，あとは LangLAB に組み込みのトランスレータによって実行可能な Prolog プログラムに変換される．

なお，詳細は [9]を参照されたい．

3.2 知識表現形式 DCKR

DCKR [3] は，東工大田中研究室で開発された知識表現形式で [5]，次の特徴を持つ．

1. オブジェクトを構成する各スロットを，述語 sem をヘッドとする 1 つのホーン節で表現する．

2. 1 つのオブジェクトは，第一引数が等しい sem 述語をヘッドとするホーン節の集合で表わされる．

3. 知識はすべて Prolog プログラムで記述されるため，知識に関する推論が Prolog のインタープリタに組み込みの機能で代用できる．

4. ユーザーは SRL/O というフレーム形式の高水準言語で知識を記述すればよく，あとはトランスレータによって DCKR の sem 述語に変換される．

3.3 意味処理の枠組

3.3.1 これまでの方法とその問題点

SRL/O による辞書記述　意味解析用の辞書記述には，フレームによる表現が有効である．図 2に，英語の動詞“open”の意味記述を SRL/O によって記述した例を示す [1]．これは，“open”に対して三つの意味規則を与えており，それぞれの規則は次の形式のスロットを複数個もっている．

[<統語的制約条件> $ <意味的制約条件> => <深層格>]

例えば，意味規則 3 の第一スロットは，

[1] eXtraposition Grammar with Slash categry
[2] Definite Clause Grammar
[3] Definite Clause Knowledge Representation

```
1. we opened the door .

No. 1    time : 367 msec
9~open#82
        tense:  past
        agent:  we#81
        object: door#84
                    det:    the
```

```
2. we opened the debate.

No. 1    time : 366 msec
9~open#87
        tense:  past
        agent:  we#86
        object: debate#89
                    det:    the
```

図 3: 旧方式の意味解析結果

- 統語的に SUBJ[4]の位置にある単語が意味的に*human[5]を表わすならば，その単語は“open”と深層的に *agent*[6]という関係をもつ

ということを表わしている．

図 2 の辞書を利用して，次の 2 つの英文を意味解析した結果は，それぞれ図 3 のようになる．

1. We opened the door.
2. We opened the debate.

問題点　問題は 2 つある．第一に，意味解析のレベルが浅い．

文 1 と文 2 において“open”の表わす概念は異なるものである．文 1 の“open” は*「しまっているものを開放状態にする」，文 2 の“open”は*「イベントを始める」，をそれぞれ指している．しかし，図 3 を見ると，その違いは“door”と“debate”の部分のみであり，“open”の表わす概念の違いは反映されていない．つまり，意味解析結果の情報は，

「“we”,“open”,“door”,“debate”という表層的な単語が深層的にどういう関係にあるか」

ということだけで

「それらの表層的な単語が概念的にはどういうものを表わしているのか」

という情報は含まれていない．

第二に，意味記述のなかに言語に依存する部分と依存しない部分が混在している．2.2 節で示した辞書の枠組でいうなら，概念対応辞書と概念辞書の記述が混在している．図 2の意味記述には，統語的制約条件と意味的制約条件の両方を記述していたが，意味的制約条件が言語に依存せず多言語間で共用が可能であることを考えれば，これらは分けて記述するべきである．

3.3.2 新しい枠組

概念辞書と概念対応辞書を組み込んだ新しい意味処理の枠組について述べる．

概念対応辞書は MRL で，概念辞書（知識ベース）は SRL/O で記述する．MRL とは，概念対応辞書の記述用に新しく設計した表現形式であり，図 4に示すフレーム形式をもつ．<単語>を一つのユニットとし，表わす概念と対応条件条件（表層関係と深層関係の対応）の組を複数個記述する．単語がどのような概念を表わすかが中心的な記述であり，対応条件部分はなくてもよい．

3.3.1 節の“open”に相当する部分を MRL と SRL/O を用いて記述すると図 5(a)（概念対応辞書），(b)（概念辞書）のようになる．

概念辞書には，単語“open”の表わすことのできる概念が記述されている．記述の都合上，*open11 などのラベルを用いているが，その意味は以下の通りである．

[4] 大文字は表層格を表わす．
[5] 左肩に*のついたものは概念を表わす．
[6] 斜体字は深層格を表わす．

```
<単語>
      ::  <対応する概念 1>
              [<表層関係 a>:<深層関係 a>]
                  ..............
      ::  <対応する概念 2>
              [<表層関係 a>:<深層関係 a>]
      ..........................

      ::  <対応する概念 n>
              [<表層関係 a>:<深層関係 a>]
                  ..............
```

図 4: MRL の記述形式

```
open
  :: open11              %対応規則 1
        [subj : agent]
        [obj  : object]
        [with : tool]
  :: open12              %対応規則 2
        [subj : cause]
        [obj  : object]
  :: open1               %対応規則 3
        [subj : object]
  :: open2               %対応規則 4
        [subj : agent]
        [obj  : object]
  :: open2               %対応規則 5
        [subj : object]
  :: open11              %対応規則 6
        [subj : tool]
        [obj  : object].
```

(a) 概念対応辞書(MRL)

```
open11 ::
      [isa:humanact]
      [isa:open1].
open12 ::
      [cause:wind]
      [isa:open1].
open1 ::
      [object:thingOpen].
open2 ::
      [object:event]
      [isa:humanact].
humanact ::
      [agent:human]
      [isa:action].
action ::
      [partner:animal].
```

(b) 概念辞書(SRL/O)

図 5: “open”の辞書記述

```
1. we opened the door .

No. 1    time : 434 msec
open#42
      prototype:      open11
      tense:  past
      agent:  we#41
                prototype:      c_we
      object: door#44
                prototype:      c_door
                det:    the
```

```
2. we opened the debate .

No. 1    time : 466 msec
open#47
      prototype:      open2
      tense:  past
      agent:  we#46
                prototype:      c_we
      object: debate#49
                prototype:      c_debate
                det:    the
```

図 6: 新方式の意味解析結果)

*humanact 人間の行動を表わす概念. *agent*となる概念は*human (人間という概念) の下位概念でなければならない.

*open1 閉じているものが開くという概念. *object*となる概念は*thingOpen (開くものという概念) の下位概念でなければならない.

*open2 人間の行為を表わす概念. *humanactを上位概念としてもつことによりagentに関する条件を継承する. *object*となる概念は*event (行為という概念) の下位概念でなければならない.

*open11 人が閉じているものを開くという概念. *humanactと*open1を上位概念としてもつので, *agent,object*に関する条件を継承する. *tool*となる概念は*instrument (道具という概念) の下位概念でなければならない.

*open12 (人以外の) 何かの原因で閉じているものが開くという概念. *open1を上位概念としてもつので *object* に関する条件を継承する. *cause*となる条件は*wind (風という概念) の下位概念でなければならない.

以上の概念は, 言語に依存しない知識であり, 英語以外の自然言語処理や自然言語処理以外の知識処理システムとも共用できるものである.

一方, 概念対応辞書には, 単語"open"がどういう概念を表わすことができるかという情報が記述されている. 例えば, 図 5(a)の対応規則 1 は,

- "open"は*open11 に対応し, そのときには, SUBJ の位置にくる語の対応する概念が*open11 の *agent*, OBJ の位置にくる語の対応する概念が*open11 の *object*, WITH の位置にくる語の対応する概念が*open11 の *tool*, となる

ということを表わしている. このとき, 対応関係のみが記述され, 概念間の制約条件は一切記述する必要がない.

図 5の辞書による

1. We opened the door.
2. We opened the debate.

の解析結果を図 6 に示す. 図 3 の解析結果と比べて, "open"の表わす概念が正しく抽出されている.

4 おわりに

本稿では, 自然言語処理用の辞書の構造を以下の 3 層構造とすることを提案した.

- 単語の統語情報と概念に依存しない意味を記述する単語辞書
- 単語と概念の対応を記述する概念対応辞書
- 普遍的な概念体系を記述する概念辞書

単語の意味は概念対応辞書と概念辞書によって表現されるが，概念対応辞書，概念辞書にはそれぞれ言語に依存する情報，言語に依存しない情報を記述する．

意味の記述を2層に分けることにより，言語に依存する制約と概念的な制約を別々に扱うことが可能になるが，その有効性を示すために東工大田中研究室で開発された自然言語解析システム LangLAB 上で，小規模な実験をおこなった．その結果，これまでの方式の問題点が解決できることを具体例に基づいて示し，さらに辞書の記述も簡潔になることを示した．

自然言語処理用の辞書構成に関する最近の研究としては，(株) 日本電子化辞書研究所(EDR)によるものがある．

EDR では辞書を単語辞書と概念辞書に大別し，単語辞書には，(1) 概念の表層表現としての単語の見出し，(2) その見出しで表わされる概念，(3) 単語がある概念を表わすときの文法的特性，を記述し，概念辞書には，概念間の上位/下位関係(概念体系辞書)とそれ以外の関係等(概念記述辞書)を記述する．

EDR の辞書構成は単語辞書と概念辞書を区別しており，また概念辞書の中核に上位/下位関係を用いているという点では，我々の立場と同じであるが，単語と概念の対応記述においては異なっている．

単語と概念の対応について，本稿では『概念対応辞書』というひとつの独立した辞書を設けた．この辞書では単語と概念の関係とその場合の文法的特性を記述する．EDR の辞書でも「単語辞書」のなかに同様の記述があるが，EDR の「単語辞書」では単語の文法的特性(名詞，動詞など)を記述するだけなのに対し，本研究の『概念対応辞書』では単語の文法的特性に加えて単語の文法的な共起関係まで記述する．単語の文法的な共起関係とは，例えば，英語ならば，動詞に対する「主語」・「目的語」のような文法機能であり，日本語ならば，「が」・「を」のような表層的な格になる（3.3.2 節で英語の概念対応辞書の例を示した)．

ところで，本研究では概念辞書の記述として上位/下位関係に基づく概念の分類体系を利用するとしているが，実際に概念の分類体系を作成することは非常に難しい．辞書を開発するためには，まず言語データを調べ，それらをもとに人間が辞書を作成するという方法をとらざるを得ないが，この作業を支援するために次のようなツールが必要になるだろう．

1. 膨大な言語情報をさまざまな角度から検索するためのツール
2. 辞書内容の整合性を保ちながら，概念の追加や削除などの操作を支援するツール

我々は，これらのツールについても現在研究を進めている．1 については，三省堂新明解国語辞典を構造化し，さまざまな角度から高速に検索可能な機械辞書を開発している [3]．2 については，特に上位/下位関係に基づく分類に関して考察し，PSI 上にプロトタイプシステムをインプリメント中である [4,7]．

参考文献

[1] 奥村学, 田中穂積. LangLAB における高水準辞書記述言語 SRL/O. In 情報処理学会第 *33*回全国大会, pages 1431–1432, 1986.

[2] 荻野綱男. シソーラスについて. In ソフトウェア文書のための日本語処理の研究 - 5 , pages 1–61, 情報処理振興事業協会, 1983.

[3] 今津英世. 機械辞書の基礎的研究. 1988. 東京工業大学修士論文.

[4] 堺　和宏. 自然言語の意味処理のための辞書に関する研究. 1988. 東京工業大学修士論文.

[5] 田中穂積, 小山晴生, 奥村学. 知識表現形式 DCKR とその応用. コンピュータソフトウェア, 3(4):12–20, October 1986.

[6] 田中穂積, 仁科喜久子. 上位下位関係シソーラス ISAMAP1 の作成. In 情報処理学会自然言語処理研究会, pages 25–44, November 1987.

[7] 望月泰之. 上位/下位関係シソーラス作成支援ツール. 1988. 東京工業大学卒業論文.

[8] 計算機用日本語基本動詞辞書 *IPAL(Basic Verbs)* - 解説編 - 情報処理振興事業協会.

[9] T. TOKUNAGA, M. IWAYAMA, H. TANAKA, and T. KAMIWAKI. LangLAB:A Natural Language Analysis System. In *COLING88*, 1988. to appear.