ＬＴＢマスタ辞書の構造と内容

田中裕一　　海野　敏
（ＩＣＯＴ）　（東京大学大学院）

１．はじめに

汎用日本語処理系ＬＴＢは、ＩＣＯＴが談話理解実験システムＤＵＡＬＳを開発する過程で蓄積してきた日本語文処理のためのさまざまな技術を、ひとつの独立したソフトウェアとしてまとめあげた汎用のシステムである。ＬＴＢは、形態素解析部ＬＡＸ、構文解析部ＳＡＸ、文生成部、マスタ辞書、および意味表現言語ＣＩＬから構成されている。本発表では、ＬＴＢの一部を構成するマスタ辞書の第１版について、その構造と内容を報告する。

２．マスタ辞書の特徴

ＬＴＢの研究開発は、構文解析と文生成を共通の辞書と文法を用いて行うという基本的なコンセプトのもとに進められている。このコンセプトに基づき、ＬＴＢマスタ辞書には、品詞、活用型など、語（語句）の構文的、文法的情報に加えて、シソーラスコード、制約など、若干の意味的情報が記載されている。

マスタ辞書のレコードは品詞ごとに独自の構造をもっており、記載項目も品詞によって異なっている。ＬＴＢでは、語は名詞、動詞、形容詞、副詞、連体詞、助詞、助動詞の７つの品詞に分類されており、マスタ辞書には、このうち助詞、助動詞を除く５つの品詞についての情報が、それぞれの辞書フォーマットに従って記述されている。

マスタ辞書は、逐次型推論マシンＰＳＩ上に構築されている。現在までに、マスタ辞書に蓄積されている語の総数は、およそ３６００語である。その内訳は、名詞が約２２００語、動詞が約１１００語、形容詞が約１００語、副詞が約２００語、および連体詞が約３０語である。

Structure and Contents of the LTB Master Dictionary
Yuichi Tanaka (ICOT)
Bin Umino (Tokyo University)

３．辞書フォーマットの概要

ＬＴＢマスタ辞書は、見出し語ファイルと語義ファイルから構成されている。これらはそれぞれ、見出し語レコードと語義レコードから構成されている。ひとつの語についての情報は、ひとつの見出し語レコードと、その見出し語レコードから参照されている（一般には）ひとつ以上の語義レコードによって記述されている。

見出し語レコード、語義レコードは、いずれも、属性ラベルと属性値の対が、リストに並んだかたちをしている。品詞によっては、属性値が、さらに属性ラベルと属性値の対のリストになっている場合がある。また、いくつかの属性値に関しては既定値が定められており、記載が省略されているときには、この既定値が仮定される。

４．見出し語レコードの記載内容

見出し語レコードの構造を、図１に示す。見出し語レコードの各項目の記載内容は、以下の通りである。

ＩＤ　：見出し語レコードを参照するための一連番号
品詞　：名詞、動詞、形容詞、副詞、連体詞のいずれか
活用型：ＬＴＢの文法に準拠した活用型
表層表現：表層の記述形式。活用語の場合、語幹と語尾を '-' でつなげたもの
読み　：読みのカタカナ表記
分解　：分解式のリスト
語義指標：見出し語レコードが参照している語義レコードのＩＤのリスト

分解は、当該語が慣用句で、さらに細かい語に分解できる場合にのみ記入される項目である。

5．語義レコードの記載内容

次に、語義レコードの構造と記載内容を、動詞を例にあげて説明する。動詞の語義レコードの構造を図2に示す。語義レコードの各項目の記載内容は、以下の通りである。

ＩＤ　：語義レコードを参照するための一連番号
深層格：使用する深層格のラベル名のリスト
能動態表層格：深層格のラベル名と表層格の対のリスト
受動態表層格：深層格のラベル名と表層格の対のリスト
意味構造：形式的意味記述による意味構造
制約　：意味構造選択の規則、または格要素のカテゴリ検査の規則
シソーラスコード：ＬＴＢシソーラスに準拠した分類番号
補語　：当該語に対する補語の情報
態　　：直接受動、間接受動、使役、授受表現のあり／なし
相　　：状態、準状態、瞬間、継続のいずれか
自他　：絶対自動、絶対他動、相対自動、相対他動、両用動詞のいずれか
可能動詞化：可能動詞化の可／不可
意志性：意志性のあり／なし
派生名詞：派生名詞のリスト

動詞が直接受動態をとらない場合、受動態表層格の項は存在しないし、動詞が補助動詞ではない場合、補語の項は存在しない。また、派生名詞の項には、Nullリストが既定値に定められている。

6．おわりに

ＩＣＯＴでは、現在、ＬＴＢマスタ辞書の記載内容と記載形式についてさまざまな角度から検討を重ね、各品詞ごとに辞書フォーマットの改善を逐次行っている。意味的情報に関しては、特に研究課題が多く、ＤＵＡＬＳ第３版の完成へ向けて、目下実験的に記述を進めている。今後も、ＬＴＢマスタ辞書の完成度を高めるため、いっそう研究を深める予定である。

参考文献

田中裕一，et al：ＬＴＢマスター辞書の意味記述の構想，ソフトウェア科学会，論理と自然言語研究会ワークショップ，1988．
瀧塚孝志，et al：ＬＴＢマスター辞書の構成，ソフトウェア科学会，論理と自然言語研究会ワークショップ，1988．

```
<見出し語レコード>::=
    〔ＩＤ／          <見出し語レコードＩＤ>，
     品詞／           <品詞>，
     活用型／         <活用型>，
     表層表現／       <表層表現>，
     読み／           <読みリスト>，
     分解／           <分解式リスト>，
     語義指標／       <語義レコードＩＤリスト>〕
```

図１．見出し語レコードの構造

```
<動詞語義レコード>::=
    〔ＩＤ／          <語義レコードＩＤ>，
     深層格／         <深層格リスト>，
     能動態表層格／   <格リスト>，
     受動態表層格／   <格リスト>，
     意味構造／       <意味構造>，
     制約／           <制約>，
     シソーラスコード／<シソーラスコード>，
     補語／
       〔品詞／       <補語品詞>〕，
     態／
       〔直接受動／   <ありなし>，
        間接受動／    <ありなし>，
        使役／        <ありなし>，
        授受表現／    <ありなし>〕，
     相／             <相>，
     自他／           <自他>，
     可能動詞化／     <可能動詞>，
     意志性／         <ありなし>，
     派生名詞／       <派生名詞リスト>〕
```

図２．動詞語義レコードの構造