第１章

ＥＤＲ電子化辞書

ＥＤＲ電子化辞書は、通常いわれる辞書、シソーラス、コーパスを含め、統合的な観点から開発が試みられた言語データである。対象とする言語は、日本語と英語であり、分野としては、日常一般の基本語と情報処理専門用語を扱う。ＥＤＲ電子化辞書は、日本語単語辞書、英語単語辞書、概念辞書、日英対訳辞書、英日対訳辞書、日本語共起辞書、英語共起辞書、日本語コーパス、英語コーパス、専門用語辞書（情報処理）からなる。

==========[ＥＤＲ電子化辞書の構造]==========

〈日本語単語辞書〉　：（→1.1.1節，２章）
　　〈日本語単語辞書レコード〉...
　　〈連接テーブル〉

〈英語単語辞書〉　：（→1.1.2節，３章）
　　〈英語単語辞書レコード〉...

〈概念辞書〉　：（→1.1.3節，４章）
　　〈概念見出し辞書〉
　　　　〈概念見出しレコード〉...
　　〈概念体系辞書〉
　　　　〈概念体系レコード〉...
　　〈概念記述辞書〉
　　　　〈概念記述レコード〉...

〈日英対訳辞書〉　：（→1.1.4節，５章）
　　〈日英対訳辞書レコード〉...

〈英日対訳辞書〉　：（→1.1.5節，６章）
　　〈英日対訳辞書レコード〉...

〈日本語共起辞書〉　：（→1.1.6節，７章）
　　〈日本語共起辞書レコード〉...
　　〈日本語動詞共起パターン副辞書〉
　　　　〈日本語動詞共起パターン副辞書レコード〉...

〈英語共起辞書〉　：（→1.1.7節，８章）
　　〈英語共起辞書レコード〉...

〈日本語コーパス〉　：（→1.1.8節，９章）
　　〈日本語コーパスレコード〉...

〈英語コーパス〉　：（→1.1.9節，１０章）
　　〈英語コーパスレコード〉...

（更新 ： 2001.06.08）

〈専門用語辞書(情報処理)〉　　：(→1.1.10節，１１章)
　　〈日本語専門用語単語辞書(情報処理)〉
　　〈英語専門用語単語辞書(情報処理)〉
　　〈日英専門用語対訳辞書(情報処理)〉
　　〈英日専門用語対訳辞書(情報処理)〉
　　〈専門用語概念見出し辞書(情報処理)〉
　　〈専門用語概念体系辞書(情報処理)〉
　　〈日本語専門用語共起データ(情報処理)〉
　　〈英語専門用語共起データ(情報処理)〉
==========

1.1 ＥＤＲ電子化辞書の構成

1.1.1 日本語単語辞書

日本語単語辞書は日本語単語辞書レコードを単語見出しの読み順(五十音順)に並べたものである。日本語単語辞書レコードは、レコード番号と、見出し情報、文法情報、意味情報、運用・その他情報、および管理情報から構成される。日本語単語辞書の基本的な役割は、日本語単語と概念の対応関係を記述し、この対応関係が成り立つときの文法的特性を与えることである。日常一般の基本語を対象とする。

1.1.2 英語単語辞書

英語単語辞書は英語単語辞書レコードを単語見出しのアルファベット順に並べたものである。英語単語辞書レコードは、レコード番号と、見出し情報、文法情報、意味情報、運用・その他情報、および管理情報から構成される。英語単語辞書の基本的な役割は、英語単語と概念の対応関係を記述し、この対応関係が成り立つときの文法的特性を与えることである。日常一般の基本語を対象とする。

1.1.3 概念辞書

概念辞書は、単語辞書，対訳辞書，共起辞書の各辞書から参照される概念を規定するための辞書である。各概念を言葉で説明する概念見出し辞書、また、概念間の関係を与える概念体系辞書、概念記述辞書からなる。概念体系辞書は概念間の上下関係を規定するもの。概念記述辞書はそれ以外の関係を規定するものである。

1.1.4 日英対訳辞書

日英対訳辞書は日英対訳辞書レコードをかな見出しの五十音順に並べたものである。日英対訳辞書レコードは、レコード番号、見出し情報、文法情報、意味情報、対訳情報、管理情報から構成される。日英対訳辞書の基本的な役割は、日本語単語見出しがその概念に対応する時の英語の対訳を与えることである。

1.1.5 英日対訳辞書

英日対訳辞書は英日対訳辞書レコードを単語見出しのアルファベット順に並べたものである。英日対訳辞書レコードは、レコード番号、見出し情報、文法情報、意味情報、対訳情報、管理情報から構成される。英日対訳辞書の基本的な役割は、英語単語見出しがその概念に対応する時の日本語の対訳を与えることである。

1.1.6 日本語共起辞書

日本語共起辞書は、日本語コーパスに格納された実文の解析結果から係り受けを構成している部分を抽出し、句の表記の五十音順に並べたものである。日本語共起辞書レコードは、レコード番号、見出し情報、構文情報、意味情報、共起状況情報、および管理情報から構成される。日本語共起辞書の基本的な役割は、日本語コーパス中での共起状況の情報に基づいて妥当な自立語の組合せ方の実例を示すことである。日本語動詞共起パターン副辞書は、日本語の主要な動詞約 5,000 について、格フ

レーム情報を記述したものである。すなわち、各動詞の各概念について、共起しうる表層格の組、それぞれの表層格に対応する深層格(概念関係子)の種類、およびそれぞれの深層格におけるフィラーとなりうる概念の範囲を記述している。日本語動詞共起パターン副辞書の主要な役割は、意味解析において、動詞およびそれと共起する名詞に属する複数の概念から、適切な概念を選択するための手がかりを提供することである。

1.1.7 英語共起辞書

英語共起辞書は、英語コーパスに格納された実文の解析結果から係り受けを構成している部分を抽出し、句の表記の五十音順に並べたものである。英語共起辞書レコードは、レコード番号、見出し情報、構文情報、意味情報、共起状況情報、および管理情報から構成される。英語共起辞書の基本的な役割は、英語コーパス中での共起状況の情報に基づいて妥当な内容語の組合せ方の実例を示すことである。

1.1.8 日本語コーパス

日本語コーパスは、日本語コーパスレコードを文のEUC（Extended Unix Code）順に並べたものである。日本語コーパスレコードは、レコード番号、文情報、構成要素情報、形態素情報、構文情報、意味情報、および管理情報から構成される。日本語コーパスの基本的な役割は、大量の実際の文に対して文の構成要素の認定をおこない、それらの構成要素がどのようにまとまって文を形態的・構文的・意味的に構成するかを示すことである。

1.1.9 英語コーパス

英語コーパスは、英語コーパスレコードを文のアルファベット順に並べたものである。英語コーパスレコードは、レコード番号、文情報、構成要素情報、形態素情報、構文情報、意味情報、および管理情報から構成される。英語コーパスの基本的な役割は、大量の実際の文に対して構成要素情報の認定をおこない、構成要素情報がどのようにまとまって文を形態的・構文的・意味的に構成するかを示すことである。

1.1.10 専門用語辞書（情報処理）

専門用語辞書（情報処理）は、日本語および英語の情報処理分野の専門用語に関する辞書である。本辞書は、複数の辞書から成り立っている。日本語専門用語単語辞書（情報処理）、英語専門用語単語辞書（情報処理）、日英専門用語対訳辞書（情報処理）、英日専門用語対訳辞書（情報処理）、専門用語概念見出し辞書（情報処理）、専門用語概念体系辞書（情報処理）、日本語専門用語共起データ（情報処理）、英語専門用語共起データ（情報処理）から構成されている。

1.2 仕様の説明形式

各辞書の仕様の説明は、共通の説明形式を用いてなされる。したがって、各章は同じドキュメント構造をしている。各章がどのような論理的構成をしているかを同じ説明形式を用いて説明する。

==========[章の構造]==========
〈章番号〉
〈章のタイトル〉　　：辞書名
〈要点の紹介〉
〈辞書レコードの構造〉　　：辞書レコードの論理仕様の記述
(→1.2.1節)
〈辞書レコードの例示〉
.
.
.
〈詳細説明を行う節〉
.
.
.
〈諸表〉
〈表名一覧〉
〈表〉
.
.
.
〈物理仕様と標準サンプルデータ〉
〈物理仕様〉　　：拡張BNFによるCD-ROM版レコードの仕様。物理的な辞書レコードは１つの文字列である。
〈標準サンプルデータ〉　　：辞書の内容を把握するための標準となるサンプルデータ
〈サンプル辞書レコード〉...
==========

1.2.1 論理仕様の記述形式

論理仕様の記述のために、直観的な理解し易さに重点を置いた記述形式（記述法）を定める。この記述形式は形式的な厳密さは欠くが、誤解の生じない伝達が可能であるように配慮されている。

レコード（項目）は、階層化された多数のフィールドからなるとする。フィールドとそのサブフィールドの関係は、タグ（フィールドの指定，名前付けを指示する）のインデンテーションで表す。フィールドの役割や格納される情報内容の記述が、タグの右側に書かれる。これらの記法は、SGMLから一部借用したものである。ただし、SGMLそのものでないことに注意されたい。

記述形式の基本的な文法（便法）は以下のとおりである。

(1)　〈aaa〉
　　〈bbb〉
　　〈ccc〉
－－－－－－－－－－－－－－－
フィールド〈aaa〉はサブフィールド〈bbb〉と〈ccc〉からなる。

(2)　〈aaa〉...
－－－－－－－－－－－－－－－
フィールド〈aaa〉が任意個連続して設定される。〈aaa〉がサブフィールドを持つ時は、それを含めて設定がなされる。

(3)　〈aaa〉。。。
－－－－－－－－－－－－－－－
フィールド〈aaa〉の下に設定されているサブフィールドが組で任意個連続して設定される。

(4)　〈aaa〉
　　.
　　.
　　.
－－－－－－－－－－－－－－－
フィールド〈aaa〉が任意個連続して設定される。サブフィールドを持たない時の便法である。

(5)　〈aaa〉　　:bbb
－－－－－－－－－－－－－－－
フィールド〈aaa〉の役割や格納される情報内容についての自然言語文bbbによる説明。'(→'と')'によって参照すべき情報（章，節，表等）へのリンク先が示される。

(6)　〈aaa〉　　bbb
－－－－－－－－－－－－－－－
フィールド〈aaa〉に格納される情報内容がbbbであることを例示する。(2)，(3)の場合については、別途便法を設ける。

(7)　〈aaa〉　　*〈数字〉
－－－－－－－－－－－－－－－
フィールド〈aaa〉に格納される情報内容の拡張BNF（次頁1.2.2参照）による記述が番号〈数字〉の脚注にあることを示す。脚注内容の記述についても、上記の便法を流用する。拡張BNFにおけるメタ変数とフィールドのタグとが、同じ表記法となることに注意されたい。この２つは全く別のものである。

1.2.2 物理仕様版電子化辞書の利用にあったての注意事項

＊　拡張BNFについて

　メタ変数を"<>"で表し、左辺と右辺との境界を"::="、選択の境界を"｜"のメタ記号で表す基本的なBNFに対し、次の拡張を行って使用しているので注意されたい。

(1) "。。。"は直前のメタ変数もしくはメタ変数をグループ化したものの一回以上の繰り返しを意味する。即ち、

　　<AAA>　::= <BBB>。。。

　　<AAA>　::= <BBB>｜<BBB><AAA>

　の二つのメタ文は等価である。

(2) "()"は繰り返しのセパレータを表すメタ記号である。

(3) "{}"は、複数の隣接するメタ変数などをグループ化する際に使用するメタ記号である。グループ化したものは一つのメタ変数と同様に扱うことができる。

　上述のメタ記号がデータとして現れる場合には、シングルクオート（'）で囲んでエスケープさせてある。

＊　辞書ファイルには Copyright文などのコメントも書かれている。
辞書データは各レコードのレコード番号のキーワードにより抽出すること。

＊　レコード長が４Ｋを超えるレコードは　CR を継続マークとして、複数行から構成されている。

＊　１つの値が記号やスペースを含む場合は " " で囲むことにより、１つの値を表現している。

　　例

　　"180-degree"
　　"take off"
　　"/"
　　" α "
　　"A.D."
　　"a science called archaeology"

　例外として、英語単語辞書の発音フィールドは１つの値しか持たず、そのほとんどが記号を含むために、 " " で囲むことはしていない。

＊　値を持たないフィールドは " "（連続するダブルクオート）で表している。

＊　ASCIIで表現できる文字は、ASCII文字を使用している。

　　例

　　RGBモデル
　　"アークランプという，２点に電流を流す灯"