大規模電子化辞書開発における高機能辞書エディタ

6S－5

安達久博*，　三輪和弘，　中沢正幸，　鈴木美穂

株式会社　日本電子化辞書研究所

１．はじめに

現在，㈱日本電子化辞書研究所（以下，ＥＤＲと略す）では，単語辞書，概念辞書を核とする大規模な自然言語処理用の電子化辞書を構築中である．　また，言語データとして日英各２，０００万文の大規模テキストベースを構築している．　これ等の複数の電子化辞書は，互いに辞書情報の共有と参照を行なえる多種類のリンクにより有機的に結合されている．　従ってこの点に着目し，電子化辞書を大規模なハイパーテキストの構造としてとらえ，オブジェクト指向の概念を導入することにより辞書開発者（ユーザ）の自由な検索要求（情報の動的交換）に柔軟に対応できる編集システムとして実現した．[4]

本論文では，この大規模な辞書開発に利用される高機能辞書エディタの概要について報告する．

２．背景と目的

編集対象となるＥＤＲ電子化辞書と言語データベースを以下に示す．[1]

①ＥＤＲ電子化辞書は，単語辞書，概念辞書，共起辞書，対訳辞書の４種類の辞書から構成されている．　単語辞書は基本語と専門用語に分れ，概念辞書は概念体系と概念記述に分れている．　また，言語は日本語と英語を対象にしているため全部で１０個の辞書からなる．（図１参照）

②日英各２，０００万文からなる大規模なテキストベース（新聞，小説，百科辞典等）とそのＫＷＩＣ，更に文を選択し情報（形態素情報，構文木，意味表現）を付加したＥＤＲコーパスからなる．

これらのデータは，開発の初期段階では，それぞれ個別の作成支援ツールを利用していた．[2]　しかし，開発の途中段階においては，各辞書情報項目（ノード）間の整合性の考慮と実データ上での検証作業が発生し，これらのデータを統合化して編集作業の行なえる高機能辞書エディタが必要となる．　また，従来のエディタは個々のデータ属性に依存した操作体系をユーザーに強制し，必ずしも統一的なユーザインタフェースを実現していない．　本エディタの目的は，編集作業における統一的な環境を構築するにはデータの多様性とデータの結合性さらにはデータに対する操作性を吸収する枠組みを提供することである．　ＥＤＲ電子化辞書は，概念ＩＤにより単語辞書と概念辞書とがリンクされた構成になっている．　従ってリンクの概念を拡大することにより容易にハイパーテキストの概念を導入できる構造を内在しているデータといえる．本エディタでは，その様なシーケンシャルな階層構造をリンクする組織リンクとユーザーが自由に情報項目間を非シーケンシャルにリンクできる参照リンクに分けて仮想的に辞書項目を展開し，編集可能なユーザー環境を実現した．

実現手段としては，辞書情報を階層化したクラスで定義し，メッセージの送受信により“情報の動的交換”を行うオブジェクト指向の概念を導入した．

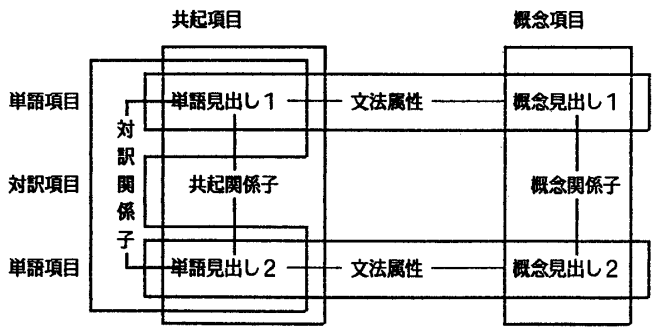
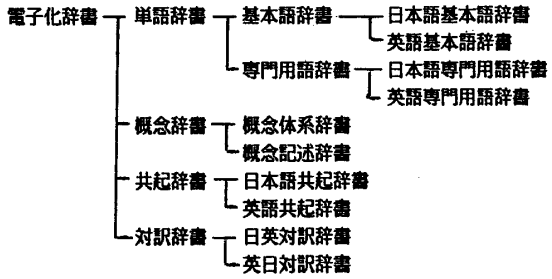


図１．単語項目と概念項目の関係



概念項目は，２つの概念の間に成立しうる関係を記述したものである．関係子としては，格，因果，意味関係がある．
共起項目は，２つの単語間の文法的な関係を記述したものである．文法関係としては，被覆，連接，上位／下位関係がある．
対訳項目は，２つの単語間の対応関係を示したものである．関係子としては，同義，類縁，上位／下位関係がある．

＊ 現在，㈱東芝　総合研究所

High Quality Editor for Developing Electronic Dictionary

Hisahiro ADACHI, Kazuhiro MIWA, Masayuki NAKAZAWA and Miho SUZUKI
Japan Electronic Dictionary Research Institute, Ltd.

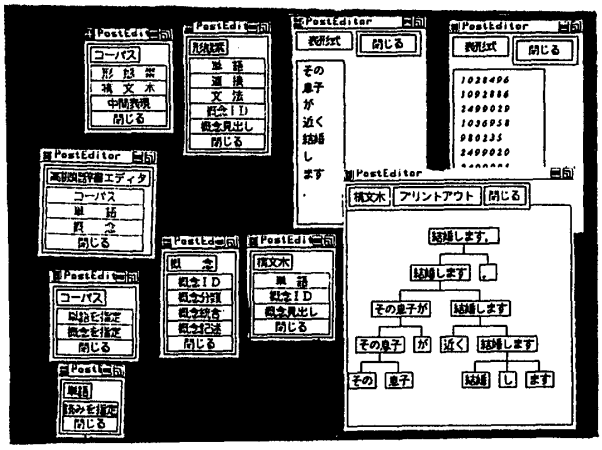
3．特徴

従来の電子化辞書がシーケンシャルな構造（例えば，冊子体の辞書と同様の階層構造）を持つ情報であり，そのためのエディタもいわゆるワークシート形式のデータ属性を反映したものであった． ＥＤＲでも，初期の辞書情報の作成においては，人手によるデータ作成が必要なため作成支援システムは従来方式を採用した．[2]　しかし，本格的に辞書を電子化する段階において，辞書情報を編集するユーザの必要性に応じて，様々な流れをもって情報をリンクし，非シーケンシャルな構造の情報に変えていく必要がある．（例えば，単語辞書の見出しにリンクされている概念見出しが概念辞書のどの位置を占め，他の概念見出しとの関係を調べ，リンクを辿ってまた単語辞書に戻って辞書情報の整合性を検証してみる等をナビゲートするものが今回提案する本エディタである．）

その特徴として，①情報と情報を直接リンクするオーサリング(authoring) とリンクを辿って情報を参照するブラウジング(browsing)の２つのモードを同時にユーザに提供していること．②従来，固定的なリンクであった組織リンク(organizational link)と参照リンク(referencial link)を同一視した，バイパスリンクを新たに設定したこと．これによりユーザにとって不必要な情報はバイパスして仮想的に必要な電子化辞書を構築できる．（図１の複数の辞書を仮想的に展開できる）③ある辞書情報を編集する場合，影響を受ける関係辞書項目を即座に表示し，同時に変更を行え，かつ検証できるシミュレーション機能を備えていること．（最終的な変更は，バッチ処理で行うが暫定的な変更の多い対話的環境においてはユーザ固有の差分ファイルに保管しておくことにより，データ修正の有効性の確認を対話的に行なえる）④これら一連の操作体系は，単語，概念コーパスのどの切口から情報にアクセスする場合でも統一されたユーザインタフェースを提供していること．

4．機能

本エディタの基本機能は，以下の内容である．
①辞書項目定義エディタ
②グラフィカルリンクエディタ
③辞書データの変更が辞書全体のどこに影響するかを示すシミュレーション機能
④現在の注目点に到達した過程を提示するモニタリング機能　これは，大規模なハイパーテキスト構造を構築する際の問題としてDisorientation Problem（迷よえる子羊）[3] を解決する手段である．
⑤データの復元と操作の再現性を保証したプレイバック機能を用意している．



5．おわりに

本論文では，自然言語処理用の大規模電子化辞書を大規模なハイパーテキストとしてとらえ辞書情報の持つ様々な形態，構造の編集機構にオブジェクト指向の概念を導入し編集データを統一的な枠組みの下で柔軟に，かつ容易に扱える高機能辞書エディタを提案した．本エディタは，基本的に辞書項目間のリンクの接続と切断の操作により編集し，そのリンクを辿ることにより情報の参照を行うという統一的なユーザインタフェースを実現し，電子化辞書を効率的に編集するエディタである．

今後は，ユーザが大規模なデータ間にはりめぐらしたリンクを自動的に最適化する機構を検討する必要がある．

謝辞

最後に，この研究の機会を与えて下さった横井所長，第五研究室天野室長，及び貴重なご意見を頂いた第一研究室内田室長に感謝します．

参考文献

[1] EDR 電子化辞書: EDR,TR-016　November 1989
[2] 辞書開発支援システム（第１版）: EDR,TR-015　Noovember 1989
[3] Conklin,J. "Hypertext:An Introduction and Survey" COMPUTER,vol.20:p17-41.　September 1987
[4] 知的ハイパーテキストに関する調査研究報告書：（財）機会システム振興協会　March 1990