

仮想電子辞書の設計

都司 達夫† 芳野 学† 上杉 和弘† 宝珍 輝尚†

†福井大学工学部 †ホロンシステム

あらまし 記憶メディアの大容量化と低廉化およびマルチメディア技術の進展にともなって、各種電子辞書の出版が盛んに行われている。手引きでは、辞書が潜在的に保有している豊富で多角的な情報を手軽に利用することは不可能であるのに対して、電子辞書の存在意義は手引き辞書の持つ構造の強制を解放し、ユーザの望む情報をユーザの望む配列で提示することであると考えられる。本稿では、母体とする複数の辞書をその論理構造にしたがって構成要素に分解し、項目結合可能性の制限の下にユーザが必要な項目のみを自由に組み合わせて新たに定義できる「仮想電子辞書」の提案とその設計について述べる。仮想電子辞書として、一種類の辞書内で有用な情報のみを再配列して得られる「部分辞書」と複数の辞書内の有用な情報を「結合項目」を介して結合して得られる「結合辞書」がある。さらに、これらの仮想電子辞書の定義構文を提案し、その実装方式を説明する。

Design of Virtual Electronic Dictionaries

Tatsuo TSUJI† Manabu YOSHINO† Kazuhiro UESUGI† Teruhisa HOCHIN†
†Fukui University †Horon System Inc.

Abstract Various kinds of electronic dictionary are being published and widely available. While it is not possible by hand to extract all the significant information underlying paper dictionaries, electronic dictionaries must provide users the facility to retrieve the information and show it in the user-specified arrangement. In this paper, we describe a design of *virtual electronic dictionary system*, in which one or more dictionaries are decomposed into their components according to their logical structures and some of them by users are corrected and combined logically into a virtual electronic dictionary under the *connectivity condition*. There are two kinds of virtual electronic dictionary, one of which, called a *simple dictionary*, can be constructed using a single base dictionary, and the other, called a *complex dictionary*, can be constructed more than one base dictionaries. The definition and retrieving primitives of virtual electronic dictionaries are introduced and their implementation scheme is discussed.

1. はじめに

記憶メディアの大容量化と低廉化およびマルチメディア技術の進展にともなって、各種電子辞書（たとえば、[9]）の出版は盛んに行われている。特に、画像や写真、音声を取り込んだ「マルチメディア辞書」の発展は著しい。本来、辞書にはきわめて豊富で多角的

な情報が内包されており[1]、これらの情報は「辞書の文法」にしたがって、構造的に記述されている。手で引く場合にはこの構造の強制に縛られており、辞書が潜在的に保有している豊富で多角的な情報を手軽に利用することは不可能である。電子辞書の存在意義はこのような構造の強制を解放し、ユーザの望む情報を

ユーザの望む配列で提示することができることであり、考えられる。これには、望む情報を特定するために、辞書の文法にしたがった構造化を行う必要がある。電子辞書を含めた電子化出版物の構造化には、JIS規格[8]が最近制定され、それにしたがった、電子辞書も出回り始めている。また、[2]では辞書のハイパーテキスト化の方式が提案されている。ところが、辞書の論理構造(スキーマ)を前面に押し出し、ユーザレベルで既存辞書からニーズにかなった、仮想的な辞書を定義し利用するための枠組みは見当たらない。

そこで、本報告では、上記のような目的に沿った「仮想電子辞書」の設計について述べる。仮想電子辞書として、一種類の辞書内で有用な情報のみを再配列して得られる「部分辞書」と複数の辞書内の有用な情報を「結合項目」を介して結合して得られる「結合辞書」を提案する。さらに、これらの仮想辞書に関連する諸概念を提案するとともに、仮想電子辞書の定義・操作言語を提案し、その実装方式を説明する。

2. 仮想辞書の設計

市販電子辞書における相互検索機能を使えば、複数辞書にまたがった場合の検索操作性は向上するものの、個々の辞書を意識して、切り替えるといった操作が必要であり、煩わしい。ここでは、個々の辞書の実体とは別個にユーザが定義する仮想辞書を設計する。仮想辞書は説明項目の実体を持たずに、辞書実体のどの部分を使用するかを定義情報のみ有する。ここでの仮想辞書の考え方は、データベースシステムにおけるビュー機能に相当している。各実体辞書の文法構造を表現する辞書スキーマ群に対するビュースキーマとして仮想辞書のスキーマをユーザが定義できる。仮想辞書に対する検索要求は辞書実体に対する3で述べる検索ライブラリ中の検索関数呼び出しにより、実行される。以降では特に「仮想辞書」と断らない限り、単に辞書といった場合は実体辞書を指すこととする。なお、本研究で用いた辞書の元データは三省堂(株)から提供を受けた、新クラウン和英辞典[5]、ニューセンチュリー英和辞典[6]であり、Websterの英語ソーラス[7]を参考にした。

2.1 結合項目

「仮想辞書の自動生成」を行う場合には、項目同士の結合操作がポイントとなる。例えば、英和辞書から、「反意語仮想辞書」を作成する場合、英和辞書の見出

し説明中に現れる「反意語」が結合項目となり、(同じ)英和辞書の見出しと結合してその反意語の詳しい情報を与える仮想辞書が定義される。なお、辞書の見出しは必ず、結合項目となり得る。ここでは、種々の辞書の辞書構造と各項目のデータの分類を掲げて、仮想辞書生成時の結合項目について考察する。

(1) 和英辞書

和英辞書の論理構造を図1に示す。[]付きの項目は辞書の部分構造を決定するための項目であり抽象項目と呼ぶこととする。また、[]なしの項目は辞書に記載されているデータであり、具体項目と呼ぶ。

それぞれの具体項目のデータは、結合可能性の観点から表1のように分類される。「見出し」以外の項目には記号(1byte)も含まれている。ここでは、「英訳の例文」「用例の例文」「関連語の例文」をまとめて「例文」としている。また、用例や例文は(和単語列、英単語列)の対として与えられるが、ここでは例えば、用例の場合、和単語列成分は「用例和」、英単語列成分は「用例英」として別個の項目としている。

なお、以降の結合可能性分類表における右端の欄の“-”, “O”, “△”, “×”は、“結合は無意味”, “結合可能”, “データに処理を施すことで結合可能”, “結合は困難”をそれぞれ表している。

見出しは例えば国語辞書など、仮名和単語や片仮名和単語を見出しとする辞書とそのまま結合可能である。“△”のついたものは、適当な構文解析や整形によって、有意な単語に分解することにより、他の結合項目と結合可能である。「漢字訳」は、見出しとほとんど1対1の関係にある。しかし、たとえば「いたみ」のように「痛み」と「傷み」のような同音異義語があるが、和単語を結合項目とする他の辞書と結合可能である。「英訳」は多くの場合単一英単語であり、英単語を見出しとする他の辞書と結合可能である。英単語列の場合には、英語辞書の成句と結合したり、名詞+名詞や形容詞+名詞のような複合語の場合には最後の名詞が有意であり、結合項目となり得る。「関連語和」の場合には、見出しを含む複合語である場合が多く、たとえば、国語辞書の参考語と結合可能であり、より詳しく、その日本語の意味を知ることができる。また、用例や例文はいずれも、単語列として他の説明項目と全く一致することはほとんどない。ただし、構文解析によりその中の名詞や動詞について他と結合することは可能である。

[和英辞書] → 見出し [見出し項目]
 [見出し項目] → 漢字訳 [英訳項目]
 [英訳項目] → 英訳 [対訳項目]
 [対訳項目] → [例文] [用例] [関連語]
 [例文] → 例文和 例文英
 [用例] → 用例和 用例英 [例文]
 [関連語] → 関連語和 関連語英 [例文]

図1 和英辞書の構造

項目	分類レベル	文字種	結合可能性
見出し	辞書	(片)仮名和単語	○
漢字訳	見出し項目	和単語	○
英訳	英訳項目	[英単語(列)]の列	△
用例和	英訳	和単語(列)	×
用例英	英訳	英単語(列)	×
関連語和	漢字訳	和単語	△
関連語英	漢字訳	英単語(列)	△
例文和	英訳, 用例, 関連語	和単語(列)	×
例文英	英訳, 用例, 関連語	英単語(列)	×

表1 和英辞書の結合項目分類表

(2) 英和辞書

英和辞書の論理構造を図2に示し、結合項目分類表を表2に示した。発音、重要度、UC フラグ、語形変化については、他の辞書項目と結合して意味のある仮想辞書の定義は考えにくいと判断してここでは、結合可能性を「-」とした。品詞種別については、国語辞書などと組合わせて品詞別の仮想辞書を定義することができるので、「○」とした。「類義語」や「反意語」は説明の形式に統一性はないが、そのなかの英単語を取り出すことにより、英語見出しや他の英語辞書の類義語と結合し得る。さらに、「解説」や「文法情報」については説明が不定型であり、有意な結合は行えないとして、「×」とした。

[英和辞書] → 見出し [見出し項目]
 [見出し項目] → 重要度 発音 [品詞項目]
 解説 文法情報
 [品詞項目] → 品詞 語形変化 UC フラグ [語義項目]
 他の品詞形 反意語 類義語 解説
 文法情報 [成句項目]

[語義項目] → UC フラグ 語義 [用例] 解説
 文法情報 [下位語義]
 [用例] → 用例英 用例和
 [下位語義] → 語義 [用例] 解説 文法情報
 [成句項目] → 重要度 成句 意味 [用例]

図2 英和辞書の構造

項目	分類レベル	文字種	結合可能性
見出し	辞書	英単語	○
重要度	見出し項目, 成句項目	数値	-
発音	見出し項目	発音記号	-
品詞	品詞項目	品詞種別数値	○
語形変化	品詞項目	英単語列	-
UC フラグ	品詞項目, 語義項目	可算, 非可算フラグ	-
語義	語義項目, 下位語義	和単語列	△
用例英	用例	英単語列	×
用例和	用例	和単語列	×
反意語	品詞項目	英単語列	△
類義語	品詞項目	英単語列	△
他の品詞形	品詞項目	英単語	△
解説	全レベル	不定	×
文法情報	成句レベルを 除く全レベル	不定	×
成句	成句項目	英単語列	○
意味	成句項目	和単語列	×

表2 英和辞書の結合項目分類表

(3) 英語シソーラス

英語シソーラス辞書の構造を図3に示し、結合分類表を表3に示す。

[英語シソーラス辞書] → 見出し [見出し項目]
 [見出し項目] → 品詞 [品詞項目]
 [品詞項目] → 語義 用例 [語義項目]
 [語義項目] → [類義語項目] [関連語項目]
 [成句項目] [反対語項目] [反意語項目]
 [類義語項目] → 類義語
 [関連語項目] → 関連語
 [成句項目] → 成句
 [反対語項目] → 反対語
 [反意語項目] → 反意語

図3 英語シソーラス辞書の構造

項目	分類レベル	文字種	結合可能性
見出し	辞書	英単語	○
品詞	見出し項目	品詞種別数値	○
語義	品詞項目	英単語列	△
用例	品詞項目	英単語列集合	×
類義語	類義語項目	英単語列	△
関連語	関連語項目	英単語列	△
成句	成句項目	英単語列	△
反対語	反対語項目	英単語列	△
反意語	反意語項目	英単語列	△

表3 英語シソーラス辞書の結合項目分類表

2.2 仮想辞書定義

2.1 で述べた結合項目にしたがって、複数の辞書を結合した場合には、複数の辞書が単に結合項目を介して合わさっただけであり、実用にはならない。仮想辞書を実用的に使用するためには、ユーザ自身が仮想辞書を自分のニーズにしたがって、きめ細かく定義できる必要がある。このために、ここでは、本辞書システムにおける、仮想辞書定義の定義要素を以下に述べる。以降の説明においては例として、英和辞書と英語シソーラス辞書を使って、和訳付き英語シソーラスの仮想辞書を作成する場合を取り上げる。

(1) 辞書構造の定義

複数の辞書を結合しても、それらの辞書の対応する項目がすべて必要であるわけではない。各辞書の項目を選択的に取り込み、それを可能な配置により再編成することにより、仮想辞書の構造を定義できる。和訳付き英語シソーラス辞書の場合、例えば、

和訳付き英語シソーラス辞書(見出し、
見出し項目(品詞項目(品詞、
類義語項目(類義語、語義))))

のようにその構造を入れ子で指定する。

(2) 母体となる辞書群の指定

結合の対象となる実体辞書を1つ以上指定する。1つの場合には、結果として得られる辞書を仮想部分辞書といい、2つ以上の時は仮想結合辞書という。和訳付き英語シソーラスの仮想辞書の場合には、母体辞書群は(英和辞書、英語シソーラス辞書)である。

(3) 実体辞書との対応付けの定義

(1) で定義した仮想辞書の具体項目が実体辞書のどの項目に関連づけられるかを定義する。実体辞書の論理構造(スキーマ)にしたがって、見出しから始まって辞書項目に至るパスを実体辞書の項目として指定す

る。たとえば、図2の英和辞書スキーマにおいて、“見出し項目.品詞項目.成句項目.成句”とすれば、以下の(5)の選択条件を満足するすべての成句を指示することになる。また、“見出し項目.品詞項目.成句項目”とすれば、選択条件を満足するすべての成句について、その「重要度」、「成句(の綴り)」、「意味」、「用例」を指示する。和訳付き英語シソーラス辞書の場合、例えば、

見出し = ET.見出し、
品詞 = ET.見出し項目.品詞、
類義語 = ET.見出し項目.品詞項目.語義項目.
類義語、

語義 = EJ.見出し項目.品詞項目.語義
となる。ここで、ETは英語シソーラス辞書、EJは英和辞書を表す。

(4) 結合項目の指定

一般に母体辞書群を $\{D_1, D_2, D_3, \dots, D_n\}$ として、各 D_i ($1 \leq i \leq n$) 中からひとつづつ選んだ項目の集合を $\{D_1.E_1, D_2.E_2, D_3.E_3, \dots, D_n.E_n\}$ とする。 $D_i.k.E_i.k$ のある並び、 $D_{i1}.E_{i1}, D_{i2}.E_{i2}, D_{i3}.E_{i3}, \dots, D_{in}.E_{in}$ について、すべての隣接する辞書項目 $D_{im}.E_{im}$ と $D_{im+1}.E_{im+1}$ の間に、相互結合可能性表において○または△の結合可能性が存在するとき、母体辞書群 $\{D_1, D_2, D_3, \dots, D_n\}$ は $\{D_1.E_1, D_2.E_2, D_3.E_3, \dots, D_n.E_n\}$ を介して結合可能であるという。この時、 $\langle D_{i1}.E_{i1}, D_{i2}.E_{i2}, D_{i3}.E_{i3}, \dots, D_{in}.E_{in} \rangle$ を結合ベクトルという。

和訳付き英語シソーラスの仮想辞書の場合には、結合ベクトルは $\langle ET.見出し, EJ.見出し \rangle$ となる。

(5) 選択条件の指定

和訳付き英語シソーラスの仮想辞書の場合、基本的には、英和辞書も英語シソーラス辞書も品詞ごとに分けられているが、互いに異なる品詞を結合しても、有意な情報とはいえないし、組み合わせの数も膨大になる。このような場合には、「選択条件」として、“ET.見出し項目.品詞項目.品詞 = EJ.見出し項目.品詞項目.品詞” というようにして結合の条件を指定する。他に、動詞ばかりの仮想辞書を作成する場合には、“ET.見出し項目.品詞項目.品詞=動詞 and EJ.見出し項目.品詞項目.品詞=動詞” というような指定になる。

(6) 項目間の関連指定

結合項目および選択条件を指定しても選択項目の関連のさせ方にあいまいさが残る。ET.見出し項目.品詞

項目、類義語はその品詞における類義語集合であり、その和訳をどのように関連づけ、配列するかについて、次の2つの方法が考えられる。

(a) ET の類義語集合 ET. 見出し項目. 品詞項目. 類義語に対して、EJ の語義集合

{ m | p ∈ ET. 見出し項目. 品詞項目. 類義語, m は同じ品詞の p の語義集合 }

(b) p ∈ ET. 見出し項目. 品詞項目. 類義語 に対して、同じ品詞の p の語義集合

(a) の場合には、要素同士の関連はユーザ自身の解釈に任されるが、類義語の一覧が分かりやすい。また、(b) の場合には、関連付けはシステムが行ってくれるが、類義語の一覧は得られない。母体辞書における構造をそのまま反映するという立場から、ここでは (b) を既定とする。(a) の指定については、2.3(2) を参照されたい。

(7) 見出し指定と見出しフィルタ

見出しの指定を行う。見出しとして複数の辞書項目を指定してもよい。見出しには既定として、高速アクセスのための索引が付与される。通常、結合条件の選択演算により、仮想辞書のレコード集合は選択的に定義されるが、見出し集合に働くフィルタによっても選択することができる。見出しフィルタとして、正規表現が利用できる。また、逆引き辞書[4]のように、見出しの読みを逆順にして、並べ替えることもできる。

2.3 仮想辞書の定義

仮想辞書には、仮想部分辞書と仮想結合辞書がある。いずれの場合にも「母体辞書の指定」と「選択項目の指定」は必須である。

(1) 仮想部分辞書

仮想部分辞書では辞書の結合は行われず、単一辞書について、ユーザの必要とする項目のみを再編成する。数ある辞書項目をすべて検索対象とはせずに、使用目的に沿って選んだ辞書項目のみをよりコンパクトにまとめた辞書を単一辞書から抽出することが目的である。たとえば、英和辞書の中から、成句のみを抽出して成句辞書を作成したい場合には、

DEFINE 成句辞書(見出し、見出し項目(品詞、成句))

FROM EJ AS

ASSIGN 見出し = EJ. 見出し、

品詞 = EJ. 見出し項目. 品詞項目. 品詞、

成句 = EJ. 見出し項目. 品詞項目.

語義項目. 成句項目

WITH ENTRIES 見出し

となる。DEFINE 句では、仮想辞書名の下に辞書項目とその構造が定義される。実体辞書では「品詞項目」の下に3つの項目があり、「語義項目」には5つの項目があったが、成句のみを選ぶことにより、これら2つの抽象項目はなくなり、構造は浅くなっている。FROM 句は母体辞書の指定であり、AS 以降は母体辞書による定義内容である。ASSIGN 句は仮想辞書の各辞書項目が実体辞書のどの項目に関連づけられるかを定義する。WITH ENTRIES 句は見出しを仮想辞書の辞書項目で指定する。

見出しに対する索引は、今の場合、実体辞書の索引をそのまま使用できないこともないが、実体辞書のすべての見出しが成句項目を持っているわけではない。したがって、仮想辞書の見出しに対してフィルター検索をかけた場合、無駄な探索を行ってしまうので、仮想辞書用に新たに別途作成する。各項目間は規定として前節(6)(b)のように関連させて、品詞ごとに成句項目に関連させるが、(6)(a)のように関連させたいときには、グルーピングを行う。

次に、英作文を支援する仮想辞書の構成を考える。英文作成において、実際にもっとも役に立つ情報は用例であろう[3]。例えば次のように、和英辞書から例文、用例、関連語を集めて、用例対訳辞書Aを定義する。

DEFINE 用例対訳辞書A(見出し、

用例項目(用例和、用例英))

FROM JE AS

ASSIGN

見出し = JE. 見出し、

用例和 =

<JE. 見出し項目. 英訳項目. 対訳項目. 例文. 例文和、

JE. 見出し項目. 英訳項目. 対訳項目. 用例. 用例和、

JE. 見出し項目. 英訳項目. 対訳項目. 用例. 例文. 例文和、

JE. 見出し項目. 英訳項目. 関連語. 関連語和、

JE. 見出し項目. 英訳項目. 関連語. 例文. 例文和>、

用例英 =

<JE. 見出し項目. 英訳項目. 対訳項目. 例文. 例文英、

..... (用例和 と同様)

>

WITH ENTRIES 見出し

ASSIGN 句の「用例和」は和英辞書の5つの和文の

辞書項目を1つにまとめた新たな単一項目である。「用例英」も同様の新たな単一項目である。このような操作を定義レベルのグルーピング(GROUPING)という。このときの問題は、「用例和」と「用例英」の対応のさせ方である。一般に2つのグルーピングA < a1, a2, ..., an > および B < b1, b2, ..., bn > があり、仮想辞書の DEFINE 句において(A, B) と並び、母体辞書において(ai, bi) (i = 1, 2, ..., n) と並んでいるとき、(A, B)の項目配置は既定では(a1, b1) (a2, b2) ... (an, bn) と並ぶ。なお、ai と bi の関連のさせ方は既定では前節(6)(b)である。

ところで、用例は英和辞書にも記載されており、その仮想辞書、“用例対訳辞書B(見出し、用例項目(用例英、用例和))”が用例対訳辞書Aと同様にして定義できる。

(2) 仮想結合辞書

結合項目の指定は関係データベースにおける結合演算に相当する。和訳付き英語シソーラス辞書の定義をまとめると、

```

DEFINE 和訳付き英語シソーラス辞書(英単語、
      品詞項目(品詞、類義語項目(類義語一覧、
      類義語語義(類義語、語義))))
FROM ET, EJ AS
ASSIGN 英単語 = ET.見出し、
      品詞 = ET.見出し項目.品詞、
      類義語一覧 =
      <ET.見出し項目.品詞項目.語義項目、
      類義語項目.類義語>、
      類義語 =
      ET.見出し項目.品詞項目.語義項目、
      類義語項目.類義語、
      語義 = EJ.見出し項目.品詞項目.語義
WITH ENTRIES 英単語、品詞項目.品詞
CONNECTED BY ET.見出し、EJ.見出し
WHERE EJ.見出し項目.品詞項目.品詞
      = ET.見出し項目.品詞項目.品詞
  
```

となる。結合項目の指定には、結合ベクトルを指定する CONNECTED BY 句を導入した。また、見出しは WITH ENTRIES 句により、ET の見出しと品詞の対が指定されており、この対をキーとして新たな索引が生成される。この例では、前節(6)の関連のさせ方について(a)と(b)の両方を併用している。すなわち、類義語一覧と個別

和訳の両方を記載している。類義語一覧の定義はグルーピングされ、ET.見出し項目.品詞項目.類義語項目.類義語 中の英単語列の並びが、並び中の英単語列の個別識別性が消失する。

仮想結合辞書は、複数の辞書を結合して得られるが、必ず、FROM 句において複数の辞書と、CONNECTED BY 句において結合項目を指定する必要がある。

2.4 仮想辞書の検索

(1) 構造検索

実体辞書および仮想辞書の構造定義に沿った検索であり、以下の FIND 文により行う。仮想辞書の DEFINE 文の構造定義中の項目に具体値を代入して、条件付けする。例えば、和訳付き英語シソーラス辞書の場合、

```

FIND 和訳付き英語シソーラス辞書(*take)、
      品詞項目(“動詞”、類義語項目)
  
```

とすれば、“take”で終わる動詞の英単語の(類義語、語義)がリストアップされる。類義語項目のように対象項目のままにしておくとして下位の項目がすべて表示される。

(2) 非構造検索

非構造検索とは、見出しからはじまる辞書の構造に沿わない検索であり、複数の見出しに対する説明項目を“縦断的”に検索することが可能である。たとえば、2.3 で述べた用例対訳辞書は便利であるが、調べたい言葉の見出しに関するものだけしか知ることができない。たとえば、「考える」という言葉の用例を他のすべての見出しの分も含めて検索できれば、手引きでは調べ尽くせない有用な例文を数多く得ることができる。

```

FIND 用例対訳辞書A(#、
      用例項目(*考え*、用例英))
  
```

FIND 文中の # は対応する位置の辞書項目(この場合、見出し)の個別識別性を無視して、それと同位の項目のみを、単に並置することを表す。このようなグルーピングを検索レベルのグルーピングと呼ぶ。なお、同位の項目内の構造は保たれる。この場合には、“用例対訳辞書A.用例項目.用例和と用例対訳辞書A.用例項目.用例和”の2つの項目を関連付けて、用例和に“考え”を含むような和英対訳表を作る。もちろん、用例英の検索も可能であり、

```

FIND 用例対訳辞書A(#、
      用例項目(用例和、"*consider*"))
  
```

用例英に“consider”を含む和英対訳表を作る。

この例の場合には、用例対訳辞書Aのみが検索の対象となっているが、用例対訳辞書Bの用例も併せて検索したい。そこで、これらの2つの辞書を“縦列結合”して、あらたな仮想用例辞書を定義し、それに対して検索をかけることにする。辞書を縦列結合してあらたな仮想辞書を作り出す構文は以下の通りである。

```
CREATE 用例対訳辞書(用例和、用例英)
FROM 用例対訳辞書A、用例対訳辞書B AS
ASSIGN 用例和 =
    [用例対訳辞書A. 用例項目. 用例和 | 用例対
    訳辞書B. 用例項目. 用例和]、
    用例英 =
    [用例対訳辞書A. 用例項目. 用例英 | 用例対
    訳辞書B. 用例項目. 用例英]
```

[| … |]は縦列結合を表す。縦列結合の場合には、結合項目の指定は不要であること、用例対訳辞書には見出しがないこと、仮想辞書同士あるいは仮想辞書と実体辞書の結合も可能であることなどに注意されたい。このように新たに定義された仮想辞書に対する先ほどの検索は

```
FIND 用例対訳辞書(*考え*, 用例英)
```

とすることができる。

2.5 仮想辞書の実体化

仮想辞書は実体辞書に対するビューであり、その検索は次の理由から一般的にオーバーヘッドが大きい。

- (a) 実体辞書にあった抽象項目が仮想辞書定義において、消失した場合でも、実際の検索は実体辞書を用いるために、抽象項目を介して行う必要がある。
- (b) 仮想結合辞書の結合処理コストが非常に高い。検索時に毎回結合処理を行うことは実用的でない。

仮想辞書の定義と実体辞書の辞書項目検索ライブラリルーチンを用いて、実体辞書と同様にその物理データ構造を二次記憶中に作り出す操作を「仮想辞書の実体化」という。これにより、上記のオーバーヘッドは回避できる。

なお、2.3(1)で述べたように、仮想辞書の見出しに対する索引だけは、仮想辞書定義時にすでに作成されており、実体化しなくても使用できる。また、(b)に対しては、結合によって連結される相手辞書のレコードの参照情報が結合索引として、仮想結合辞書定義時

に作成される。これらの対策により、項目データも含めて完全に実体化しなくても、見出し語指定検索の場合には、実用に耐えうるパフォーマンスは得られる。ただし、非構造検索を行う場合には、項目の全文検索が必要であり、それ用に最適化した辞書データの内部データ構造を自動生成することが必要となる。実体化は MATERIALIZE 文で行う。

```
MATERIALIZE 用例対訳辞書
```

3. 母体辞書の構造化

ここでは、我々が行った、辞書の構造分析(構文解析)による構造化の概要について記す。新クラウン和英辞典、ニューセンチュリー英和辞典の元データは組版用の制御コードとして JAM コードと呼ばれる独自のコードが使用されている。

3.1 構造化のための前処理

元データ中の JAM コードは、例えば、“文字の書体指定”や“行の境で分離しない文字列の指定”などの辞書のユーザイメージを保つための組版指定であり、辞書自身の構造とは直接関係ないものや、“見出しの始めの位置・段落の最初の位置の指定”など、辞書の構文要素であり、構造分析に本質的に関わるコードが混在している。後者の JAM コードは不可欠であり、新 JIS 第一水準中の適当なコードを用いて置き換えた。また、前者のコードは辞書のイメージに近い検索結果の表示にとって必要なもののみ、やはり新 JIS 第一水準中のコードで代替し、それ以外のものは除去した。このような前処理によって、

- (1) 構造分析が行い易くなる。
- (2) 本来の辞書のイメージにより近い表示が可能。
- (3) オリジナル辞書データの圧縮が行える。

といった利点が得られる。

3.2 構造化

(1) 構造化ファイルの構成

3.1の前処理で得た辞書データファイルを読み込み、辞書の文法に沿った構造分析を行い、次の3種類の構造化ファイルを生成した。各辞書の前処理後の辞書イメージと辞書に記載されている使用説明(凡例)に基づき、構造分析において使用する辞書の構造を決定した。辞書の構造分析の結果は、見出しごとに文法項目の階層構造にまとめ上げてファイルに格納した。

(a) 索引ファイル: B*木を使用した索引付き順次編成ファイルである。キーとして辞書の見出しが登録さ

れている。データとして文法項目ファイルにおける文法構造（構造分析結果）の先頭のオフセットを持つ。

(b) 文法項目ファイル： 各見出しについて、分析結果の文法構造が格納されている。文法構造には各辞書項目についての説明が書かれているデータファイルの部分の先頭からのオフセットが格納されている。

(c) データファイル： 見出しを除く辞書のデータ本体が格納されている。各文法項目は NULL 文字で終端される。

3.3 検索ライブラリ

3.2 で説明した実体辞書の内部データ構造をトラバースすることにより、任意の辞書項目を取り出すための関数群である。

4 仮想辞書の実装の方針

4.1 ユーザインターフェース

3. で述べた言語ベースのインターフェースでは仮想辞書はやはり使いにくい。仮想辞書の定義や操作には GUI ベースの分かりやすいインターフェースを提供することが必要である。特に本システムでは辞書スキーマを前面に押し出す必要がある。すなわち、仮想辞書の定義にあたっては、母体辞書のスキーマをグラフィカルに提示し、定義に必要な項目をその中から切り出し、可能な範囲で望みどおり組み合わせる機能を提供することにより、新辞書のスキーマ構築を支援する必要がある。

4.2 仮想辞書操作の実現

(1) 仮想辞書の定義

母体辞書の指定がすべて実体辞書または実体化された仮想辞書の場合には、仮想辞書定義文 (FIND 文) はそのまま内部表現に変換され、蓄えられる。母体辞書が仮想辞書を含むときは、仮想辞書の辞書項目は対応する実体辞書の辞書項目に変換されたものが埋め込まれた形での内部表現に変換される。

(2) 仮想辞書に対する検索構文

検索構文 (FIND 文) は(1)で蓄えられた仮想辞書の内部データ構造を参照しながら対応する実体辞書の検索構文の内部データ構造に変換される。変換された検索構文はインタープリタにより、解釈実行される。このとき、適宜 3.3 の検索ライブラリ関数により、検索が行われ、所定の配列で提示される。

(3) 仮想辞書の実体化

実体化 (MATERIALIZE 文) は(1)で蓄えられている仮想辞

書定義の内部データ構造を参照しながら、すべての定義項目について実体辞書を検索して、3.2 の三種類の構造化ファイルを自動生成する。

5. おわりに

仮想電子辞書の設計について述べた。本設計に基づき現在実装中である。既存の実体辞書の問題点として、記述が不統一でスキーマに沿った構造化ができない項目説明があったり、項目として切り出す時の終端子として決めたマークが必ずしも一定でなかったり、スキーマとしてこちらが想定していない項目が出現したりする例外が出現し、構造化できないものがあった。可能な限り手作業で対応したが、なお十分ではない。これらのいくつかは、言葉が持つ多様性を十分に説明するためには、ある程度不可避であると思われる。電子辞書の規格化がこのあたりも含めて十分に検討され、標準化されたならば、逆に辞書の編集段階から、規格準拠を考慮でき、統一のとれた電子辞書を構築できる。本研究における仮想電子辞書の場合には、このような標準化された実体辞書が母体辞書として使用できるような環境が望まれる。

謝辞 和英辞書および英和辞書の元データは三省堂(株)より提供していただいた。謝意を表する。

文献

- [1] 吉田 将：辞書構築における諸問題、情報処理、27, 8, pp.933-939, 1986.
- [2] 内藤、山下、松山、柵木：オンライン辞書のハイパーテキスト化手法、情報処理学会論文誌、34, 2, pp.320-329, 1993.
- [3] 武田明子、古郡延治：例文をもとにした英文書作成支援システム、情報処理学会論文誌、35, 1, pp.53-61, 1994.
- [4] 岩波書店辞典編集部：逆引き広辞苑、岩波書店。
- [5] 山田和男 編：新クラウン和英辞典、三省堂。
- [6] 木原、福村、芦川 編：ニューセンチュリー英和辞典、三省堂。
- [7] NeXT Computer Inc.: Chap.5 in "NeXT Applications", 1990.
- [8] 日本規格協会：日本語電子出版検索データ構造、JIS X 4081, 1996.
- [9] <http://www.m-w.com/cgi-bin/netdict>