

解説

辞書構築における諸問題†



吉田 将†

1. はじめに

日本語ワードプロセッサの爆発的な普及、機械翻訳システムの研究開発プロジェクトの進行、第5世代コンピュータにおける自然言語理解や機械翻訳の研究、国内主要メーカからの商用機械翻訳システムの登場、データベースの自然語インタフェースなどなど、最近自然言語処理に関する研究開発の話題は絶えない。

我が国における自然言語処理の研究は昭和30年代初めに開始され、その後の数十年の間に上述の話題のほとんどに関して、すでに研究が開始されている。にもかかわらず、ここ数年の研究開発、実用化のテンポの速さは著しい。その原因は、計算機環境の発達、とくに計算機システムの高速度・大容量化と低価格化、漢字処理装置の普及などをあげることができる。

我々がかな漢字変換の開発研究を始めたのが昭和30年代後半であり、昭和40年代初めには、今日の日本語ワードプロセッサの基本原則となっている、かな漢字変換の基礎研究はほとんど完了している¹⁾。同時に我々は、必要な機械辞書の作成に取り組んだ。10万語足らずの辞書の作成作業を始めたのが昭和40年初め、3年後には一応完成したのであるが、当時の計算機環境では漢字ディスプレイなど思いも寄らず、出力装置としては漢字テレと称する1字/秒程度のスピードで機械的にプリントする印刷装置があるのみで、作成した辞書の内容の全部を印刷するためには24時間運転しても1年はかかるという時代で、プリント出力を見て内容をチェックすることなど不可能なことであった。このために、数字コードのまま人間が見てチェックする必要があり、それを支援するための種々の正当性チェック方式を考えたものである。この辞書の作成・使用を通じて、我々は多くの貴重な経験をした。

自然言語処理に機械辞書（以後、辞書と呼ぶことが

多い）が必要であることを今さら説く必要はないが、辞書の重要性、作成・管理・運用上の問題点については案外気付かれていないのではないかと思ひ、関係者の間ではすでに常識になっていることではあるが、あえて解説することにした。ただし、辞書の作成については、大方の合意というものはほとんどないのが現状である。したがって、私個人の常々持っている見解を述べることになるので、異なる意見を持った関係者も少なくないことを断っておきたい²⁾。

従来、自然言語処理の研究の中心は形態素、構文、意味の各段階の解析、意味表現、理解、文脈処理の研究など、どちらかと言えば処理方式の研究に重点が置かれ、使用する辞書の規模は小さく、実用規模のシステムに関するものは少なかった。しかし、ワードプロセッサを始めとして³⁾、最近各所で開発されつつある機械翻訳システムなどでは、実用規模の処理が要求されており、したがって、そのための辞書が必要となり、その開発が重要な課題となっている⁴⁾。現在のところ、各システムごとに、おのおのの考えに従って、既存の辞書（多くは国語辞典、用語集などを機械可読にして⁵⁾、それから編集している）を再編集して情報の追加・削除をしながら辞書の開発を行っている。このため、機械翻訳システムで必要とされる解析用、変換用、生成用辞書が各々独立に製作され、収容語集も異なるといった事態も発生しているのが実状である。

しかし、小規模、実験段階のシステムでは明確でなかった問題が、実用システムにして始めて認識されるということも少なくないようである。特に、自然言語の本質であるあいまいさ、漠然さ、多様さに起因する問題は避けられない。

辞書は規模が大きく、内容も複雑で、その開発には多大の経費と労力を要する上に、利用者一人一人その内容に対する要求が異なる。それにもかかわらず、個人で開発することが困難であるという大変やっかいな問題をかかえている。

以下に、辞書開発に係わる問題点をあげてみよう。

† Problems in Dictionary Construction by Sho YOSHIDA (Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology).

†† 九州工業大学情報工学部創設準備室

(1) 辞書の目的・種類：どのような目的・方式で使用するのか。それによって、内容が全く異なるものとなる。目的を共通グループ化することは可能と思われるが、処理方式については、各々異なるのが普通であるから、同じ辞書を共通に使用することは、ごく限られた応用以外、困難と思われる。

(2) 辞書の規模：収容語彙の数と種類、その選択基準、分野・対象領域による違い、分野の設定基準などの問題がある。また、収容語彙の追加・削除の基準も問題である。

(3) 辞書のデータ項目、内容：収容すべき語彙を定めたとしても、その語彙に関してどのような項目についてその内容を収容するのか、収容すべき内容はどのように表現するのか（主として意味表現の問題）、誰がその内容を作成・収集するのか（知識獲得の問題）などの問題がある。

(4) 辞書のデータ構造：辞書に収容すべき項目、内容が与えられたとして、これをいかなるデータ構造で表現し、蓄積するかが重要である。その構造は辞書の使用法（検索方法、利用形態など）により最適なものを採用する必要がある。辞書の検索効率に大きな影響をもつことになる。また、辞書を通常のデータベースと見なして、汎用の DBMS を利用するか、使用環境に応じた専用 DBMS にするかという問題もある。

(5) 辞書内容の共通化：辞書項目、内容を定めたとしても、その利用方法、利用者によりその要求は異なる。一人として同じ内容で同意することはないであろう。どのようにすれば、内容を共通化して蓄積することができるであろうか。共通化の対象となるものにはどのような項目があるのであろうか。もし、共通化できないとすれば、どうすれば良いのか。

(6) 辞書の構築、追加・修正ツール：辞書の構築は大規模なものであり、たとえ蓄積すべき内容が与えられても、それをインプットするためのツールを用意しなければならない。このツールには追加・修正、チェック、校正、テストなどの機能をもっていることが要求される。

(7) 辞書の管理・運用：誰が作るか、どういう体制で辞書の内容を定めるか、データの収集・作成は誰が行うのか。どのようにして維持・管理・更新し、利用者は誰か。評価は誰が行うのか。経費は誰が負担するのか。

以下、各項目について検討する。

2. 辞書の目的・種類

辞書は目的によって種々存在する⁶⁾。これを言語処理のレベルに従って分類すると、

1. 形態素レベル辞書
 - かな漢字変換用辞書
 - 形態素解析用辞書
2. 構文レベル辞書
 - 文解析用辞書（係り受け、格構造）
 - 意味カテゴリ
3. 意味レベル辞書
 - 格構造
 - 概念依存構造生成
 - セマンティックネットワーク生成
 - シソーラス
 - 多義構造
4. 多言語間辞書
 - 対訳辞書
 - 翻訳辞書（トランスファ辞書）
 - パーソナルな翻訳、文書作成支援型辞書
5. その他
 - 専門用語集など

2.1 形態素レベル辞書

形態素レベル辞書には、代表的なものとしてかな漢字変換用辞書がある。その内容は、すでに、多くの場で公表されている。学術研究用として公開されている辞書もある¹⁾ので参照されたい（検索システム付^{7),8)}）。

形態素レベル辞書では以下のような問題がある。

1. 蓄積語彙の種類と数—汎用、専門（対象）分野ごとに検討の要あり。

2. 見出し語の形式
 - かな見出しか漢字見出しか
（表記法の選択）
 - 活用語の扱い
 - 複数項目見出し

3. 付属語の扱い

4. データ構造（汎用 DBMS にのせるか（汎用機）、特殊（専用）DBMS とするか（かな漢字変換用など））

5. 記憶構造、検索方式

利用方法（アルゴリズム）によって異なる。一般に、形態素解析用辞書を文解析に使用する場合には、

① ランダムにアクセスして使う。

② 使用（検索）頻度も多い。

したがって、検索能率は重要である。

6. 未知語対策

存在を前提としなければ不可。

7. 音韻情報のゆれ、その他、内容・表記のゆれと標準化の問題。

見出し語の形式については、以下のような問題がある。

見出し ID		かな見出し		格	見出し切れ目	文節切れ目	文法・意味情報													漢字	表記					
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
A: 見出し語句 ID 部	B: 校正ずみ表示部	C: かな見出し部	D: 格助詞部	E: 古語・外来語表示部	F: 見出し語句切れ目表示部	G: 文節切れ目表示部	H: 合自立語の表示部	I: 小文字数表示部	J: 同形意義語句優先順位部	K: 重要語表示部	L: 自・他動詞種類表示部	M: サ変動詞表示部	N: 活用の種類表示部	O: 活用の行表示部	P: 形容詞の種類表示部	Q: 形容動詞の種類表示部	R1: 副詞の種類表示部	R2: 副詞に後続可能文字表示部	S: 接続・感動・連体詞部	T: 見出し語句の種類表示部	U: 語句構成表示部	V: 普通名詞部	W: 代表漢字数部	X: 固有名詞部	Y: 助数詞, 代名詞部	Z: 漢字表記部

図-1 形態素辞書の項目の例

イ. 表記法 日本語は片かな、平かな、漢字かな混り、ローマ字など多様な表記法がある。さらに、送りがなの不統一、漢字の不統一、原文のまま、または片かな表記による原語の表記など、また、原語の片かな表記における読みの不統一など、同一の語に対する表記法が多様であり、いずれも許容されている。

ロ. 活用語 活用語の見出しは、通常の国語辞典などでは終止形が採用されている。しかし、機械処理では活用した形で現れるわけであるから、終止形に限る理由はない。語幹（不変化部分）と変化部分とに分けて、不変化部分のみを見出しとする考えもある。ただし、不変化部分のない語（する、あるなど）はすべての活用形を見出しにするなどの特別な措置が必要である。また、辞書内での標準表記をどう選択するかも問題である。

ハ. 複数項目見出し 形態素辞書の一例を図-1に示す。どの項目を見出しとして使用できるようにするか、その場合の検索方式・効率をどう維持するかなどの問題がある。また、形態素辞書は、日本語の場合などはべた書き文の解析に使用されることが多い。この場合は辞書の記憶構造に特別な工夫を必要とする。大規模辞書を多数回繰り返し検索する必要があるため、検索時間の増大を防止する方式を工夫する必要がある⁹⁾。

2.2 構文用辞書

このレベルの辞書の内容は品詞によって異なる。このレベルの辞書の目的は、与えられた語の列（形態素解析によって得られた結果の語列）から、各語が文中でどのような役割で存在しているかを求めるために必要な情報を与えるものである。すなわち、ある語が文中で主語として存在するか述語として存在しているのか、または目的語として存在しているのか。また、このような文法的な関係に加えて、意味的な関係すなわち、ある語が他のある語に対して主体として存在しているのか、場所を表す語として存在しているのかな

どを求めることが必要である。

このような構文解析は、用いる文法、意味的關係などによって辞書が持つべき内容、形式も異なる。いずれにしても、これらの情報は品詞により全く異なるので、品詞別に主な問題をあげてみよう。

品詞の種類として、どのようなものを立てるかは言語により、また同一言語においてもその言語の捉え方（文法）、利用の方式により異なる。

機械辞書を作成するという点からは自立語、特に動詞と名詞が問題となる。これらは、①数が多い、②語彙が不定である、③意味・用法が複雑・多様で多義性もある、④文法の違い、解析・利用方法の違いにより内容も異なる、などのため一度作成を始めると変更が困難である。他の品詞については語彙数が少ないので、辞書作成上の問題よりはその語の捉え方、利用方法などの問題（文法の問題）に研究・開発の焦点がある。

(1) 名詞辞書

まず、語彙の問題がある。

収容語彙(数)の多い辞書での問題として、以下のようなものがある。

イ. 汎用の語彙か専門語彙か

ロ. 汎用、専門それぞれにおける使用頻度の大小とその評価

ハ. 専門語彙が属する分野の分類法と各語のそれへの登録基準

ニ. 複合語の登録基準

ホ. 語彙（特に専門複合語の）収集方法、専門用語としての認定・情報付加

ヘ. 専門語彙の整理

次に、名詞の意味・概念と辞書に収容すべき内容については次のような問題がある。

イ. 名詞の意味分類（意味マーカ）

ロ. 名詞の意味分類は概念の分類。しかし、その分類の捉え方は多岐⁹⁾

ハ. 名詞意味分類（意味マーカ）のレベルはどうし

て決めるか。

ニ. 他の名詞(語)との意味的關係(同義, 反義, 上位-下位などの)をどう捉えるか。

ホ. シソーラスの自動構成 上記の意味的關係の付与から構成可能であるべきであるが, 逆に自動構成した結果を見て語の定義の妥当性の判定に利用できる。

ヘ. データ構造 特に他の語との関係をどう表現するか。

ト. 継承属性をどう利用して表現の簡潔性を求めるか。

(2) 動詞辞書

語彙量としてはサ変動詞の問題が大きい。種類が多いことと, 比較的自由に複合・合成されるためである。サ変動詞は文法的には動詞であるが, 複合名詞と本質的には同じ構造であると考えてよい。

動詞の構造・意味をどう捉え表現するかには多くの議論がある。それに従って, 辞書に記載すべき内容も異なる。

一般的には, 以下のような情報の扱いが中心課題となる。

- 文型パターン (特に英語の場合)
- そのパターンに表れる主語や目的語の性質 (意味マーカ, 文法的性質など)
- 格支配のパターン (特に日本語の場合)
- アスペクト素性 (継続, 瞬時など)

上記の情報をもった汎用の日本語基本動詞辞書が動詞 850 語について作成されている¹⁰⁾。

動詞に関する基本的な項目としてはその格構造がある。すなわち, 動詞がどのような格をとるか, ということで, この情報はその動詞の“意味”をある程度表現したことになる。

動詞の格構造とは, たとえば,

“走る”=〈人〉が〈場所〉ヲ走る

のように“何が”“どこを”“走る”のか, すなわち“走る”という動作が成立する(考えられる)ためには走る主体(agent)が存在し, それはある“場所”を走るのだからその場所(place)が存在するはずであると考えると, “走る”にはこのような agent や place の存在が不可分であるからこれを“走る”の定義内容の一部として記述しようとするものである。しかし, この格構造にも多くの問題がある¹¹⁾。

イ. 格とは何か。格の捉え方に種々ある。

ロ. 格の種類と数(20~100種類と人により異なる)。

ハ. そもそも格は数えあげ可能なのか。

ニ. 格の構造は? 深層格と表層格の問題。

ホ. 必須格と任意格。

ヘ. 格の共起関係と共起名詞の問題。共起名詞は決められるか。

ト. 共起名詞(の意味マーカ)は固定できるか。

チ. 共起名詞の意味マーカを用いることにどのような意味があるか。

以上の他にも, 動詞をどう捉え, どういう項目を辞書に収容すべきかについては多くの立場・考え方があり, 現状では機械辞書のための項目が整理されているとはいえない。ほとんど共通の合意が得られたものはないと言っても過言ではない²⁾。

2.3 意味解析用辞書

意味解析のためには, 前述の構文解析の結果の情報とともに語と語との間の意味的關係を決定すること, 及び各語の意味を求めることが必要である。各語の意味を知るとは, その語に関する次のような情報を知ることである。

イ. 意味マーカ(意味マーカ間の関係)

ロ. その語の表す概念

ハ. その概念の他の概念との意味的關係(上位-下位, 同類・反義などいわゆるシソーラス情報)

ニ. その概念が文中の他の概念とどのような依存関係(文の概念依存構造)にあるか。

通常の国語辞典などでは, 見出し語に対する意味は, 簡単な定義文などで与えられている。機械辞書では, これを上記のような諸関係を用いてどの程度まで与えてやればよいのか, どの範囲を推論にまかせるのか, などの問題がある。

語の意味とその理解, 文の意味とその理解とはいかなることであるかは議論のあるところである。利用目的によって異なるといえる。上述の意味の捉え方も広く合意されているわけではない。

2.4 多言語間辞書

多言語間辞書といっても事実上は2言語間辞書に限って論じるのが現段階では適当であろう。その理由は, 2言語間の対応関係に限っても単なる訳語の置き換えのみでは不可であり, その対応関係は複雑であるからである²⁾。2言語間辞書の代表的なものは翻訳用辞書(トランスファ辞書)である。2言語間辞書においては以下のような問題がある。

(1) 対応関係は語対語に限られない。原語に対する訳語が句, 文であったり, 存在しなかったりする。

(2) 同一の原語に対して訳がいくつか存在する。したがって、その選択情報を与えておく必要がある。この選択情報は広い意味での文脈によって与えられるものであるから、①選択に必要な情報は何か、②それを文脈中からどうやって見出すか、を与えておくことが重要となる。

(3) 逆に原文における文、句が訳文における語に対応することもある。すなわち、原文のどの語群を訳文のいかなる語句に対応させるかを決め、さらにそれをいかなる文脈情報によって選択するかその条件を与える必要がある。

(4) 上記のように対応関係は、語対語の単なる置き換えではないので、①その対応関係を分類し、②それらの対応関係に付随している選択情報を整理・分類し、③さらに選択情報間の衝突、欠如が生じた場合の選択ルール適用順序などを与えることが必要である。

(5) これらの項目・内容すなわち辞書は、見出し語(句)ごとに作成する必要がある。

3. データ項目

辞書に収容すべき項目は、その使用目的・方法によって異なる。一般の国語辞典における項目の一部を図-2に示す¹³⁾。機械辞書用には、これらに機械辞書特有の情報を追加する必要がある。人間が使うための辞書には不要であるが、機械辞書には必要であるという項目をあげることは必ずしも容易ではない。むしろ困難である。一般的には、人間にとっては常識であるため記述を省略してある部分かどの程度存在するかを明らかにしておく必要がある。すなわち、人間の高度な判断力の下で使われることを頼りに作成してある辞書

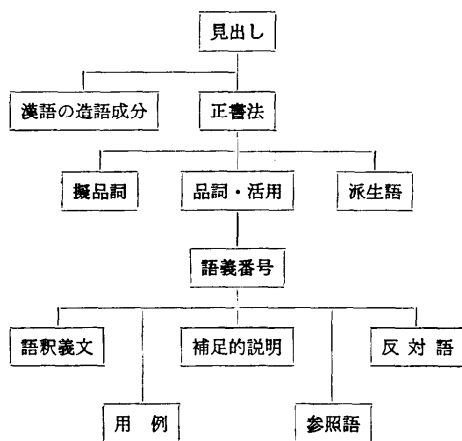


図-2 国語辞書における項目とその関係 (一部)

を機械処理用を使用することによるあいまいさを除去し、明確に表示・記述することが必要となる部分はどこであるかを明らかにしておく必要がある。特に意味に関する情報、すなわち①定義文の内容と記述の仕方、②多義の場合の意味選択基準、③例文による定義からの意味の類推・抽出、④他の語(上位語、下位語、同類語、関連語など)との関連において意味が記述されているので、それら関連した語との関係の明示、などの明確な表示を機械辞書には要求したい。

人間のための汎用の辞書に記載されている意味定義文を機械可読にし、記述形式を整理し、今後の利用に供するための作業は有用であろう。この立場から汎用辞書における見出し語とその定義文などの例を見てみよう。

辞書に記載されている、これらの情報の“項目”を整理してみると、図-2のようになる。

機械辞書においては、これらの項目ごとに、整理してその内容が記憶されていることが要求される。しかしながら、人間用国語辞典には簡潔さが重要であるため、一定の形でこれらの項目とその内容が記載されているわけではない。また、人間が見て始めて理解できるような記述法(略記など)が採用されているのみならず、その記述形式が一定していないことがある。

このようなことは、利用(検索)面からは不都合であるので、記述形式にあいまいさのない形に変換する必要がある。記述形式のあいまいさ(および、不統一)は項目の指定方法においても、また定義文の内部においても存在する。

一つの辞書の内部での不統一、あいまいさのみではなく、辞書の違いによる項目、内容、定義方法などの違いも大きい。

4. データ構造と記憶構造

辞書のデータ構造、記憶構造についても、その使用目的、使用方法によって全く異なるものにならなければならない。

辞書の構造を決定する際に考慮すべきことをあげてみる。

(1) 語彙数: その大小によって記憶装置の割当などで記憶構造に影響する。その結果、検索能率に大きな影響を与える。もちろん解析・生成における未知語の遭遇率にも影響する。

(2) 語彙の種類: 一般分野と特定対象分野に分けて考える必要がある。特定(専門)分野をどのように

して分類するか、分野を決めたとしてもある語がその分野に属する語かどうかの判定基準をどうするかの問題がある。語彙の種類は一般分野については新聞・雑誌などの語彙調査、特定（専門）分野については、その分野の文献などから収集し、出現頻度、出現範囲の分布、使用者の分布、時間経過などを考慮して整理する必要がある。

(3) 検索アルゴリズム：辞書検索のアルゴリズムは辞書の使用方法すなわち、文の解析・生成などの方式に左右される。たとえば、日本語のようにべた書きされる文の形態素解析（特にかな文の形態素解析）では辞書の検索回数は、単純に計算すれば、文中の文字数を n とすると $O(n^2)$ となり、検索時間は記憶構造により決定的な影響を受ける。その他、人間がインタラクティブに使用するのか、パッチ的に使用するのか、などの違いによる検索の仕方の違いにより記憶構造、データ構造に影響を与える。

(4) 辞書の利用レベル：辞書には文の解析・生成に直接使用する辞書と、その辞書を生成するための源辞書あるいは源々辞書などの利用レベルの違いが存在し、データ構造、記憶構造に影響する。フロントエンドの辞書（解析・生成に直接使用する辞書）では検索能率が問題となるが、後方の辞書は、フロントエンド辞書の再編集や追加・修正・削除などのメンテナンスの容易さに重点を置いた構造にすべきである。

(5) 辞書内容の記述方式：国語辞典などでは、ある項目に対する説明は他の項目に対する説明とは、相互に関連がある場合でも、独立に記述されているのが普通である。意味解析などにおいては、上位一下位、類義、反義などの関係をたどって、推論を行う必要が生じる。この場合、各見出し語間で独立に内容が定義されていると、これらの意味的關係をたどることが非能率で事実上困難となる。したがって、これらの関係を辞書中に明示しておくことが有効であり、必要であると考え¹⁴⁾。このことは、大規模で複雑なデータベースにおいてはしばしば起こることで、知識を宣言的に記述し蓄積するか手続き的に記述して蓄積するかという問題でもある。ただし、辞書の場合は、自然語のもつあいまいさや漠然さゆえに生ずる語と語の同定の困難さがあり、通常の国語辞典のような宣言的な知識の蓄積では、文字どおりの良さを発揮できないと考える。このことは、辞書のデータ構造をフレーム形式にすることへの示唆を与えている¹⁵⁾。ただし、手続き的な記述を含むフレーム構造にすることの危険性も指

摘されている¹⁾。

5. 辞書内容の共通化

すでに述べたように、辞書の使用目的・使用方法によって、その記載内容に対する要求は異なる。したがって、どのような項目と内容をもった辞書を作るべきかについては何年議論しても大多数の同意が得られるとは考えにくい。一方において、辞書は膨大なものであり、一人で作成することは困難であるから、どうしても共同利用することを考えなければならない。解決の途はあまりないと思う。次のような方針で源辞書を作成する以外に方法はないのではないだろうか。①多くの人の合意が得られている項目一主として文法的な情報についてはそれを採用する。②多くの合意は得られていないが、違いが細部についての場合は、そのうちの一つ（または、いくつ）を選んでしまう。③多くの合意は得られなくとも、ある方式を定めれば有効であると考えられる情報は可能な限り記載する。

機械辞書は自然言語の研究者にとって必須のものである。必要な場合は源辞書から必要な部分を抜き出し、再編集し、情報の追加・変更・削除を行って自分の目的に合った辞書を作る以外に方法はない。上記のようにして作成した源辞書が存在することは、この作業を相当に軽減するものと信ずる。細部における内容の変更は極限された語彙の中で行われることが期待できると思う。

6. 辞書の構築と管理・運用

辞書構築上の問題点のうち、実際の構築作業にあたる際の問題点をあげておこう。

(1) 誰が（どういうグループ）辞書項目と内容を決めるのか。

(2) 誰がデータを収集するのか。

(3) 与えられた語がどの項目に属し、その内容をどう記述するのか。個人によるゆれ、同一個人でも日時の違いによるゆれが生じるがどう対処するか。

(4) 誰が登録された内容をチェックするのか。

(5) どのような体制で辞書の維持、管理、運用を行うか。

辞書に対する要求は使用者一人一人によって異なるにもかかわらず、個人で製作することが困難なものである。辞書が、自然言語研究に不可欠なものである以上源辞書は必要不可欠であると考え、源辞書の存在を疑問視する意見もあるが、それ以外に方法はないの

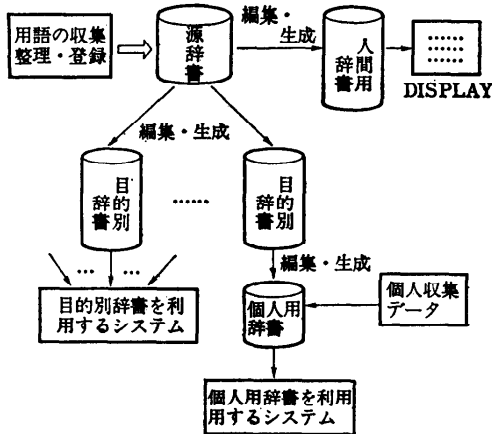


図-3 源辞書と目的別辞書

ではなかろうか。各人の要求に合わない部分は、各人が修正・変更する以外に方法はないと考える。源辞書から各人の要求に合った辞書が編集可能であるためには、源辞書の内容が明確に標準化されており、そのデータベース管理システムが汎用性の高いものであり、必要な項目内容の検索が容易にできるものでなければならない。

各人が使用するパーソナルな電子辞書に対する関心も高まりつつある。OA システムの一機能として、電子辞書が組み込まれるようになる日も近いようである。図-3 に源辞書から目的別辞書（フロントエンド辞書）、パーソナル電子辞書、人間用辞書を生成するシステムの構成を示す。

7. おわりに

物事を良くする秘訣は、それを誰でも自由に使えるようにすることである。ソフトウェアでもそうであるが、より多くの人々の使用に供して磨きをかけたものが良い製品となるように、辞書も多くの人に使われてこそ良いものに磨かれるに相違ない。

源辞書は国家的な体制で構築し、維持・管理・運用

しなければならないものであり、その国の国語政策にも深く係わる事業である。国民全体が自由に使用できることが最も重要である。

参考文献

- 1) 稲永, 吉田: 日本語処理のための機械辞書, 情報処理, Vol. 23, No. 2, pp. 140-146 (1982).
- 2) 日本情報処理開発協会: 自然言語のコンピュータ解析用辞書に関する調査研究報告書, 60-S 002 (1986).
- 3) 吉田: 機械辞書について, 日本語学, 7月号, 明治書院, pp. 26-35 (1984).
- 4) 辻井: 辞書の構成と機械翻訳, 情報処理, Vol. 26, No. 10, pp. 1174-1183 (1985).
- 5) 横山: 国語辞典データベース化の準備, 電総研報, Vol. 41, No. 11 (1977).
- 6) 荒木: 辞書, 長尾 真編, 日本語情報処理, 第7章, 電子通信学会 (1984).
- 7) 吉田, 日高, 稲永, 田中, 吉村: 公用データベース日本語単語辞書の使用について, 九大大型計算機センター広報, Vol. 16, No. 4, pp. 335-361 (1983).
- 8) 日高, 稲永, 吉田: 拡張 B-tree と日本語単語辞書への応用, 電子通信学会論文誌, Vol. J 67-D, No. 4, pp. 399-404 (1984).
- 9) Yoshida, S.: Conceptual Taxonomy for Natural Language Processing, JARECT, Ohm-sha & North-Holland (1982).
- 10) 情報処理振興事業協会: 計算機用日本語基本動詞辞書 IPAL の概要, Data Management (1986).
- 11) 村木, 青山, 六条, 村田: 辞書における格情報の記述, 情報処理学会, 自然言語処理研究会資料, 46-3 (1984).
- 12) 長尾, 辻井: 機械翻訳における訳語選択と構造変換過程, 情報処理, Vol. 26, No. 11, pp. 1261-1270 (1985).
- 13) 鶴丸: 単語の釈義文を利用した単語間の階層関係の抽出について, 情報処理学会, 自然言語処理研究会資料, 45-4 (1984).
- 14) Nicoletta Calzolari: Detecting Patterns in A Lexical Data Base, Proc. COLING 84, pp. 170-173 (1984).

(昭和 61 年 7 月 4 日受付)