20000114 5A

タイトル：　構造化臨床医学用語集の構築に関する研究

英文タイトル：　Toward a Structuerd Healthcare Terminology

著者：　劉亜斌[1]　里村洋一[1]　佐々木哲明[2]

Ya-bin Liu　Yoichi SATOMURA　Tetsuaki SASAKI

木村通男[3]　廣瀬康行[4]　山﨑俊司[4]

Michio KIMURA　Yasuyuki HIROSE　Shunji YAMAZAKI

1　千葉大学医学部附属病院医療情報部　260-8677 千葉市中央区亥鼻１－８－１

Chiba University Hospital,　Inohana 1-8-1, Cyuou-ku, Chiba 260-8677

2　医療情報システム開発センター 107-0052　東京都港区赤坂２－３－４　ランディック赤坂１０Ｆ

MEDIS-DC, Akasaka 2-3-4, Minato-ku, Tokyo 107-0052

3　浜松医科大学附属病院医療情報部　431-31 浜松市半田町３６００

Hamamatsu University Hospital, Handacho 3600, Hamamatsu 431-31

4　琉球大学医学部附属病院医療情報部　903-0215 沖縄県西原町字上原２０７

University Hospital of the Ryukyus,　Uehara 207, Nishihara-cho, Okinawa 903-0215

抄録：
電子カルテの実用期を目の前にして構造化医学用語集の登場が待たれている。これまでも、分野ごとに各種の用語集（原資用語集）が作られているが、電子カルテで用いるには、これらを統合するための原則を必要とする。用語集には非曖昧性、非冗長性、一貫性など多くの要素が整っているべきとされるが、生き物である現実の用語集では困難な点が多い。理想と現実を調和させ、かつ統合的な管理を可能とするために、正規化用語集、基本用語集、用例索引及び用語分解（Parsing）と用語合成（cpmposing）からなる用語システムを提案した。これによって、意味ネットワークを直接管理することなく（それらは原資用語集の意味構造にゆだねる）用語集の編集を遂行し保守する事ができる。また標準用語に精通しない一般ユーザーの入力用語を標準用語に変換し、原資用語集の分類と照合することも可能となる。このモデルを検証するために、同じモデルによる、病名集のＩＣＤ１０－自動コーディングシステムとＳＮＯＭＥＤの日本語自動翻訳システムを試作、実行し、良好な結果を得た。

キーワード：標準化、構造化、日本語、医学用語、電子辞書

Abstract: Structured healthcare terminology is one of the most important bases of EPR system. Standard terminology should respond to several basic requirements, which are non-redundancy, non-ambiguity, consistency, etc. Actual terminologies, however, hardly satisfy these requirements. In order to harmonize the formal requirements and practical usage, authors proposed a system based on Atom dictionary that contain prior representations of basic concepts and synonyms with domain classification. It has also, normalized terminology, quotation index, and tools of parsing and composing. This model can provide an easier process for editing standard healthcare terminology without significant consideration on semantic network, which can be given by original reference terminologies. It will offer users better accessibility to standard terminology. This paper introduce the structure of the system and reports the results of two applications on this idea to automatic diagnostic representations to ICD-10 and Japanese interpretation of SNOMED.

(Keywords: Standardization, Structured healthcare terminology, Digital Dictionary, Japanese)

1　はじめに

電子カルテが技術的な意味で実用期を迎えていることは疑いない事実である。現に、少なくとも数カ所の大病院で運用に供されている。しかし、オーダーエントリーの運用が進んでいる病院でも、次のステップとして電子カルテの採用を確定している所は多くない。経済的要因や医師の抵抗など深刻な問題に決着が付いていないこともあるが、電子化の目的であるデータの共通利用に至る条件が十分でないことも理由の一つである。通信やセキュリティー面を含めて標準化活動が進行中であるが、中でも、用語の問題はどの標準化課題においても共通の重要事項で、電子化された標準医学用語辞書の登場が待たれている。

厚生省は標準病名集[1)]をはじめ、処置・手術、薬品、検査、材料などの電子化用語集を開発中である。一方、医学界の様々な専門領域で、学会が編集した用語集が作られており[2)]、時代を反映して電子化の動きもある。しかし、病名集開発を担当した筆者らの経験では、信頼性のある、また使いやすい用語集の編集のためには、様々な標準化の要件を検討し、統一したルールを設定することが必要である。とはいうものの、**用語は生物のように変化し、進化するものであり、**また論理的な一般ルールが常に当てはまるものでもない。

本論文では、電子化用語集に求められる基本要素を検討した上で、現実的な編集ルールと作成された用語集の機能を補う方法を提案した。ここでは、あくまでもコンピュータ上で専門用語辞書として扱うことを前提としており、印刷された冊子体の辞書を想定してはいない。

2．この論文で用いた用語

本文で使われる用語を次のように定義する。

1）用語：概念を１つの言語で表現したもの
2）用語集：用語の集合
3）意味：用語が指し示す概念
4）文脈：用語が使用される背景（専門領域や使用局面）

3．理想的な用語集の要件

ASTM (American Society for Testing and Materials) は、標準的な用語集の備えるべき要素の基準を１９９７年に公表している[3)]。これを参照しながら、現実に存在する、あるいは存在しうる用語集とは別に、もし理想的な用語集というものが存在し得ると仮定して、**その用語集が備えているべき機能や属性を検討してみる。**

3．１　使用目的（対象範囲）

言葉というものは、同じ表現が同じ意味を持つと先験的に信じられているところがあるが、実際には、その言葉が使われる目的や適応される世界によって明らかに異なる。たとえば、「心」は医学領域で使われれば「心臓」の同意語であるが、文学や宗教では「精神」や「魂」と同義に使われる。同じ医学でも専門領域によって意味するところが異なる場合も少なくない。従って、以下に述べる曖昧性や冗長性についても、その用語集が、どの領域で使用されるかによって左右される。そこで、その用語集が医学一般を対象とするか、臨床医学に限定するか、あるいは外科の領域に絞るか、保険請求の記載に限定するかなど、用語集使用目的を明確に示すことが必要である。

３．２　非曖昧性

特定の一語が使用者によって複数の意味に解される場合がある。このような用語の意味の曖昧さは、略語の場合によく見られる。「ＤＭ」は「Diabetes Mellitus」と「Dermatomyositis」どちらの場合にも用いられる。このような用語は収録すべきでない。漢字を用いた日本語表現では少ないが、仮名を用いる場合には、意味の重複に充分注意する必要がある。

３．３　非冗長性

一つの概念の表現として、複数の異なった用語が対応する場合がしばしばある。印刷された用語集は一般に重複を許している。なぜならば、実用的な意味で同義語は極めて重要であり、その収録がむしろ用語集の価値を高めているからである。しかしながら、コンピュータで使われる用語集の場合には、冗長性は他の機能（同義語辞書、基本概念辞書など）にゆだねて、用語集本体からは排除すべきである。表現の多様性を許せば、用語の数が無限大に膨張すること、データ収録後の検索が困難になることなどがその理由である。

３．４　網羅性

網羅性は対象範囲を明確にすることと不可分である。たとえば、病名集の場合、保険診療で許されている対象疾患に限定するか、あるいは、健康相談や、交通外傷、予防接種など、保険適応をを受けない診療行為の理由も含むのか否かによって、網羅すべき病名（診療事由）の範囲が異なる。しかし、一旦、対象範囲が明確にされれば、その範囲で用語集は必要な全ての用語網羅していなければならない。そうでなければ、利用の局面で表現できない事象に遭遇することとなる。また、網羅性を維持するためにも適正な期間毎の改訂を必要とする。

３．５　緻密性

用語集の目的によって緻密性の程度も異なる。すなわち、収録される用語の種類と数が異なる。たとえば臨床目的の用語集では「腫瘍」について、「悪性腫瘍」「良性腫瘍」、「原発腫瘍」、「転移腫瘍」などの概念と、それらの発生部位が最低限の緻密性として必要とされるが、病理学の領域で使用される用語集としては、組織型やＴＮＭ分類のような進行の程度に関する情報も必要とされる。緻密性が十分でなければ表現力を保証できない。

３．６　一貫性

以上に述べた要件は、用語集の全体を通じて適用されなければならない。用語集の編集に当たっては、しばしば各用語の実際の適用を想定して、個別に判断が行われ、結果として用語集全体を通じてのルールが守られないことがある。

３．７　非文脈依存性

用語は、３．１の項で述べたように、用途によって意味するところが異なる場合がある、一つの用語集の中で、文脈に依存した用語をそのまま採用すると混乱が生じる。たとえば、「ＭＳ」は、「多発性硬化症」「僧帽弁狭窄症」「メニエール症候群」のいずれをも意味するが、一般には文脈によって使い分けられている。循環器疾患のみを対象とした用語集では「僧帽弁狭窄症」として採用可能であるが、一般的な医学用語集では標準用語として採用できない。従って、このような用語は用語集の中に収録すべきではない。強いて収録する場合には、一意の表現に対応させるためのそれぞれの文脈を設定する必要がある。

４．正規化用語集の作成

さて、上記のような特性を全て満たした用語集は、論理的には完璧であって、このような用語集に採用された用語のみを使って作成された文書は、全てのユーザーに紛れなく理解される。しかしながら、このような人工的な用語集は現実には作成が難しい。なぜならば、後に述べる基本語（ＡＴＯＭ）を含めて、既存の用語は過去の概念の変遷を引きずり、概念は表現の多様性を備え、それぞれの表現は新しい概念への援用を可能性として保持している。標準の用語といえども、大多数はそのような混沌の中から拾い集められるのであって、人工的に合成される訳ではないからである。そこで、筆者らは、実在する各種の用語集に以下のような操作を加えて正規化を行った。

４．１用字法の統一（正規化）

大文字、小文字の別はあるもののアルファベットで統一されている欧米に比べて、我が国は用字法に特別な注意が必要である。漢字、かな、カナ（１バイト、

２バイト）、アルファベット（１バイト、２バイト）、ギリシャ文字、ローマ数字など、同じ概念を表記する手法が多様で、一般にはこれを様々組み合わせてに使ってきた。しかし、組み合わせを自由に許せば用語の表現が膨大なものとなる。使い方に一定の制限を与えて、冗長性を排除しなければならない。
１バイト文字と２バイト文字の混在を許すのは、徒に表現の多様性を助長する。１バイト文字でなければ表現できない用語は存在しないから、少なくとも、２バイト文字による統一がなされるべきである。また、６３００あまりの漢字（ＪＩＳ第２水準）の中には、同義異字が数多く含まれている。（図１）。同じ意味の文字のどちらを使用するのか、明確な編集方針を決めなければならない。
生物学では、動植物の名称をカタカナで記述する事になっているが（例：人間＝ヒト、蛙＝カエル）、医学では必ずしも厳密に守られているわけではない。このルールに従えば、Streptococcus の訳語としては、生物名としての「レンサキュウキン」が妥当だと言えるが、一般的には「連鎖球菌」「れんさ球菌」などと表記される。日本医学会の医学用語辞典でもこの表現を採用している。
Cyst の概念を「のう腫」とするか「嚢腫」とするかも同じ例である。このように漢字とカタカナ、ひらがなの使い分けを統一する事は正規化の基本である。同様に、漢数字、アラビア数字、ローマ数字の使い分けにも統一したルールが必要である。たとえば、「十二指腸」を「１２指腸」と表記するのを許せば、数字を内蔵するあらゆる表現に影響がある。
学術領域によって、表現が異なるケースも多い。著名なものとして、「精巣」と「睾丸」、「副甲状腺」と「上皮小体」などがある。「摘除」と「剔除」のように、同義であるかどうかの判断も分かれるケースもある。どちらが正しいかどうかの決定はさておいて、少なくとも、一つの用語集の中では表現の統一だけは保たなければならない。これは、用語集の使用目的によっても異なるので、臨床医学用語集ではどちらかを優先語として扱う必要がある。
外国語のカタカナ表記は極めてやっかいな問題を提起する。特に英語圏以外の固有名詞の表記が混乱している。例を挙げれば Buerger 氏病のカタカナ表現には以下のようなものがある。

バージャー氏病
ビュルガー氏病
バーガー氏病
ビュルゲル氏病

固有名詞については、原語の表記をそのまま２バイト文字のアルファベットに移して使う方法もあるが議論の有るところである。カタカナ表記を使うならば、少なくとも一つの固有名詞について一つのカタカナ表記を決定すべきであろう。上記の例では、日本医学会医学用語辞典が「バージャー」としている。

日本医学会のような権威ある機関の決定に従うのがよいが、同辞典でも複数の表記を列記してある例もあり。筆者ら編集者が独断で優先順をゆだねざるを得ない場合も多い。
ここで注意して置かねばならないのは、その用語が独立して使われる場合においても、他の用語表現の一部として使われる場合も、ルールが一様に適応されるべき事である。
以上のような考察の結果次のような表記ルール作り、これを正規化用語集の編集に適用する事とした。

1）２バイト文字のみを使用する
2）漢字は日本医学会用語辞典（以下、医学用語辞典）に採用されている文字を使用する
3）漢字、ひらがな、カタカナの混在は、医学用語辞典に採用された表現を使用する
4）医学用語辞典に複数の表記がある場合は、医学用語辞典の先頭に記載されている表現を採用する
5）数の表現には原則としてアラビア数字を用いる。但し、「十二指腸」のように数値としての意味が薄く、漢字表記が定着しているものは漢数字を用いる。
6）外国語の固有名詞をカタカナ表記する場合は、できる限る英語の発音に忠実なものを選ぶ。（医学用語辞典にあるものは、これを採用する）

４．２複合語の表記

専門用語の多くは、複数の概念の組み合わせで表現される。その用語を構成する各々の構成要素である概念の表現が統一されていたとしても、組み合わせの方法が正規化されていなければ、表現の多様化を招く。
表現の組み合わせは、語順と括弧、句読点、ハイフォンなどの記号によって実現されるから、これらの使い方のルールを決定する必要がある。筆者らは、括弧や句読点を一切用いない表記を選択した。また、アルファベットによる略語や化学式などの場合を除いてハイフォンやその他の記号による表現は排除した。

４．３同義語、慣用語

正規化用語集においては冗長性を排除しなければならないが、日常使われる用語の冗長性自身は必要なものである。それは、ユーザーによけいなストレスを与えない点からも、また、表記法に関する専門家たちの果てしのない議論に巻き込まれないためにも重要である。用語集の非冗長性を補うために、用語集本

体とは別に、同義語や慣用語、あるいは古い過去の記録を参照する場合に備えて、廃語をも含めた辞書を用意しなければならない。これを同義語辞書と呼ぶこととする。この辞書によって、「バセドウ病」や「グレーブス病」を「甲状腺中毒症」の同義と認識し、用語集の中からそれを選び出す事ができる。

４．４　分類法と意味関連

用語集はその使用目的に従って、何らかの分類法に関わりをもつものである。たとえば、現行の健康保険における報酬請求を目的とするならば、料金表（診療報酬点数表）との関連が明らかでなければならず、疾病統計に用いるものならば、ＩＣＤの分類に関連付けられているはずである。しかし、特定の狭い目的に絞られた用語集の場合を除いて、収録された用語のすべてがそれぞれの分類に対応しているとは限らない。

用語集には、その目的に応じて、独自の意味関連構造を持つことが許されている。印刷物の用語集ではアルファベットや五十音順にならべられており、意味構造を持たない場合が多いが、ＳＮＯＭＥＤ（Systematized Nomenclature of MEDicine）は、その名のとおり，独自の分野別（部位・臓器、病理形態、医療行為、薬剤などの１１軸）の分類にしたがって収録しているばかりでなく、それぞれの分野の内部でも、独自の分類概念にしたがった配列を行っている。

このような意味関連構造は、単一の階層構造である必要はない。意味の世界は複雑な関係のネットワークを構成しており、いわゆる「分類」は、そのネットの一部をある目的に合わせて抽出したものにすぎない。この意味ネットワークを一種の知識の表現だとみなして、体系的に整理しようとしたのが、ＮＬＭ(National Library of Medicine)の手になるＵＭＬＳ(Unified Medical Language System)である。このような試みは偉大な仕事であるが、大変な労力を要する、おそらく完結に何十年とかかるであろう。

実用的でしかも適応対象がある程度広い用語集を求めるならば、むしろ標準用語集の内部には、独自の意味構造を構築せず、ＩＣＤ[4]やＳＮＯＭＥＤ[5]、薬品の薬効分類など既存の分類との関連を、それらに対応した用語との関係で記録することで、十分に可能である。なによりも、新しい独自の意味関連体系を構築し、これを一般に納得させるのは至難のことであるからである。

上記の理由により、筆者らが作成した正規化用語集には意味構造が定義されていない。

５．用語システムの構成

これまでの記述で、相互に矛盾する要件を挙げてきた。冗長性を排除すべきとしながら、冗長性は実用面で重要であると指摘した。意味構造の記述が大切であるとしながら、実際には構造を持たない用語集が望ましいとした。これは、理論的にあるべき姿と、実現しうる、かつ実用になる用語集との間に乖離があるためである。実際の用語集を開発する場合は、何らかの形でこの矛盾を解消しなければならない。
筆者らは、その解決法として基本語辞書および用例索引の整備とこれを用いた自動命名法を提案する。

５．１　基本語辞書

基本語とは表記された概念を分解して、それぞれの概念が失われない最小の単位としたものである。たとえば、「右上葉肺扁平上皮癌」は、「右」「上葉」「肺」「扁平上皮癌」の４つの基本語に分解される。ここで、「扁平」と「上皮」、「癌」に分解しないのは、これらを切り離すと、「扁平上皮癌」の概念が失われてしまい、臨床医学用語集としての素質を失うからである。このように、基本語といえども用語集の適用範囲から逃れることは出来ない。もし、この用語集が解剖学の領域においても使われると仮定すれば、扁平上皮癌は「扁平上皮」と「癌」の二つに切り離すことが必要とされるであろう。
基本語辞書は、その細分化が行き届くほど、適応範囲が拡大すると同時に実用性が低下する。この欠点を補うために、基本語には属性と用例の情報を付け加える必要がある。まず、用語の属性として、品詞と分野分類が重要である、医学用語の大半は名詞であるが、用語を分解すると、形容詞、動詞、接続詞などが出現する。分野は、身体の位置や部分を現すもの、病因となるもの、医療行為を表現するもの、生態機能を現すもの、固有名詞などに分類できるが、固有名詞を除いて、ＳＮＯＭＥＤの持つ１１の軸はこの分野（ドメイン）をよく分類している。これを利用するのは、用語選択の便宜ばかりでなく有利である。筆者らは、この１１軸に「固有名詞」を加えて１２分野とした。
もう一つの重要な属性は、同義語である、材料となった元の用語集（これを原資用語集と呼ぶこととする）から標準用語に移す際に排除した冗長性をこれによって補償する。同義語には、原資用語集には採用されていないが一般に使われている慣用語、既に使われなくなっているが、古い文献などに出現する廃語、表記の簡略化のために使われる略語などがある。同義語は、ユーザーの自由な表現と、原資用語集の正規表現を結ぶ重要な橋渡しである。

５．２　用例索引　(Index)

基本用語と正規化用語集を関係づけるための情報を集めたものを「用例索引」と呼ぶこととする。たとえば、
原資用語集に「肝切除術（拡大葉切除）」があるとすると、これは正規化されて「肝切除術拡大葉切除」となるが、これを基本語辞書を参照して分解し、「肝」、「切除術」、「拡大」、「葉切除」に分ける。これらの基本語を優先語に置換して、「肝臓」、「拡大」、「葉切除術」の３語に整理し、この３つのキーワードのいずれからも、元の正規化用語を検索できるように索引を作成するのである。このようにすれば、参照用語の「肝切除術（拡大葉切除）」はもちろんのこと、「拡大肝臓葉切除術」や「肝臓拡大葉切除」「肝臓葉切除（拡大）」のような多様な表現のいずれを使っても、原資用語集の表現に到達する事ができる。

５．３　自動命名法

上記の用例索引と基本語辞書を利用して、ユーザーの自由表現を正規表現に翻訳するシステムである。入力された表現を基本語に分解し、それぞれを同義の正規語に置きかえるとともに、同じ物があればこれを選択し、ない場合には類似表現を提案する。ここでは類似性を評価する論理が重要な鍵であるが、単に構成する基本用語だけではなく、文字数や、語順、領域分類などを利用して、類似度を算出し、優先度順位をつけるプログラムを作成し利用した。詳細は別に報告する。
図２に自動命名法と基本語辞書、用例索引及び各種の原資用語集の関係を示した。

６．　システムの検証

これまで述べてきた構成のシステムが、実用的に働くか否かを，実際的な応用プログラムによって試験した。

６．１　病名自動コーディング

筆者らは１９９４年以来厚生省の依頼を受けて、ＩＣＤ１０[4]対応の標準病名集[6][7]の開発に携わってきた。その結果、約１８０００の病名を収録し、これらに、３で述べた正規化を行った上、ここから病名表現に使われた基本概念１０４７４語、同時に同義語訳を定義したもの８２８語を得た。

６．２　自動コーディングのプロセス

上記の１０４７４基本語の組み合わせから、任意の病名集に収録されている病名を自動的に標準病名集に変換するシステムを構築した。ＡＵＴＸ－１０と名

づけられたこのシステムは、８ヶ所の病院で実際にその病名集のＩＣＤ１０対応版編集に用いられた。このシステムでは、元の病名集がＩＣＤ９に準拠したコードが与えられていることを前提としている。ＩＣＤ９が付与されていない場合にも利用できるが、変換確率は低下する。システムは変換効率向上のために、ＩＣＤ９とＩＣＤ１０の対応表を備えており、約６０％のコードを自動的に変換できるからである。

図３にしたがって処理過程を述べると、先ず、入力された病名表現と直接に一致するＩＣＤ１０対応病名を辞書から検索する。一致病名が辞書に存在すれば、文句なく変換できる。次に、文字変換による正規化と、さらに句読点・括弧などの処理を行い、再び、対応するＩＣＤ１０対応病名を検索する。

このプロセスで適当な対応が発見されなかった場合は、パーサーによって、入力病名を、基本語に分解する。分解した基本語の組み合わせを用例索引から発見し、ＩＣＤ１０のコードを得る。基本語の一部に同義語がある場合は、これに置きかえて、再び用例索引を参照する。これでも該当するものが発見できなければ、一部の基本語を一時的に削除して類似病名の検索を進める。更に、対照する二つの病名について、それぞれを構成する同じ基本語の数による類似判定を行う、すなわち、同じ基本がをより多く含んでいる病名を類似性が高いと判定する。また、分類上の兄弟となる病名も類似病名とする。

該当する病名もしくは類似病名が複数ある場合には、優先順位をつけて、候補をリストし、オペレータに選択を要求する。

６．３　病名集の変換結果

ＡＵＴＯＸ－１０を用いた自動コーディングシステムによって、千葉大学医学部附属病院で１０数年来使用されてきたＩＣＤ９対応病名集を処理した。

結果を以下に示す。

対象病名数	９１６２	
単一の確定候補	６６８７	（７３％）
複数の候補の提示	１１９２	（１３％）
候補なし	１２０９	（１３％）
その他　（エラーなど）	７４	（１％）

千葉大学の病名集は、かなりの部分、標準病名集と起源を共有していることから、比較的良好な成績を示したと思われるが、同じシステムを琉球大学の病名集変換に用いて、同様の変換効率を得ている[8)]。　基本用語集による自動命名法の効果を示したといえよう。

7．　ＳＮＯＭＥＤの翻訳への応用

ＳＮＯＭＥＤはその前身であるＳＮＯＰ(Systematized Nomenclature of Pathology)を含めると、既に４０年に及ぶ歴史のある構造化医学用語集である。今日では、１１軸に分けて約１５万語を収録し、コンピュータで利用できる、最も実用的な辞書として、世界中で使われている。筆者らは、ＳＮＯＭＥＤを翻訳する試みをこの１０年来続けてきたが、翻訳作業の効率化のためには、コンピュータによる支援が必要であると感じてきた。

そこで、この基本語辞書と自動命名法を利用して、ＳＮＯＭＥＤの翻訳に応用することを試みた。まず、ＳＮＯＭＥＤから、汎用される単語を頻度別に抽出し、頻度の高い単語に日本語訳を与えた（表１）。こうして作成した基本語を用いて、ＳＮＯＭＥＤの英語表現を日本語に半自動的に翻訳した。　詳細は別に報告するが、おおむね良好な結果を得て、実用的な翻訳に利用する準備が進んでいる。

8．考察

医学学術研究や診療録の記述に共通基盤を持ちたいという要求は、近代医学が始まって以来のものである。医学用語の辞書作成や標準化は１００年以上にわたって営々と続けられてきた。様々な医学辞書、用語集が出版され、時代に合わせて改訂が行われている。しかし、医療の情報がコンピュータを利用して処理される時代になって、様相は大きく変わってきた。ある程度の基礎医学知識を持った人を対象とした出版物で済ませられる状態から，そのような知識のないコンピュータにも利用が可能なように整理され，しかも関連情報を盛り込んだ辞書が必要になってきたのである。

このような需要への対応は、必ずしも新しいことではない。ＳＮＯＭＥＤの前身であるＳＮＯＰ（Systematized Nomenclature of Pathology）は１９６０年代に編集されている。その後のＳＮＯＭＥＤ、ＵＭＬＳ，Ｒｅａｄ、などの歴史も企画の開始から数えれば１０年を越える。これらは、それぞれ独自の目標を掲げながら、その時代の情報処理の進歩に沿って変わってきている。

Rossi Mori[9]　は、用語集の発展を３世代に分類している。すなわち、従来の辞書形式（意味構造を持たず、アルファベットやアイウエオ順に並べたもの）を第１世代とし、ＳＮＯＭＥＤinternational やＩＣＤ１０、ＬＯＩＮＣなどの構造化されてはいるが、特定の階層構造や、カテゴリー分類に限られているものを第２世代、包括的な意味ネットワークを表現し、これをコンピュータで自由に操作できるようなモデルを第３世代と呼んでいる。ＵＭＬＳ[10]　はこうした第３世代への挑戦の一つと位置づけられるが、この数年、第２世代の用語集の中にも、第３世代を目指して改訂を加えつつあるものもある。ＳＮＯＭＥＤ—ＲＴ

（Reference Terminology）もその一つで、ＳＮＯＭＥＤの用語に意味関係や利用のための付加情報を付加する作業を行っている。
用語問題を理論的に追求すると、個々の用語が医学用語世界の全体でどのような位置づけにするか、言い換えれば、その用語が表現する概念（Concept）を如何に定義するか、が問題となる。Cimino[11] やEvans[12] はこの課題に知識工学的なアプローチをしている。一つ一つの用語の意味関連情報を、個々に検討してあらゆる用途に使えるように定義するのは、実務的に困難である。むろん、ＵＭＬＳはこれに挑戦するプロジェクトであるが、国家的な規模で莫大な投資をしなければ遂行できない。
意味関連を重視した医学用語集の例として、MEDIS の病名集改定作業[10] がある。この病名マスターの改訂作業は単に病名集を改訂するのではなく、実用的な関連情報を含めた体系化を目指して以下のような原則を提示している。一病名概念には一つのコードを割り当て、このコードに対応する代表的表現（リードタームと称する）と、それと同一概念を表現する同義語を収録すること、病名を入力するシステムで必要となる補助情報を揃えること、修飾語を付加する規則を明確にすること、などである。
この方法は、実際の臨床現場での病名入力に際して、臨床医が伝えたい概念を違和感なく表現できることに重点が置かれており、臨床医にとっては使いやすいと思われる。一方で、伝達される情報がコードの組み合わせ、表記の組み合わせ、入力された原文など多様であり、情報の利用者は処理の必要に応じてどれを選択するかを問われる。また、多様な使用局面での必要な付帯情報を網羅することは、かなり困難な仕事であって、ＵＭＬＳに比べられるほどのことはなくとも完成に至るまでに相当な期間を要するであろう。
筆者らがここで提案しているのは、既存の用語集の保持している情報を集積する事によって、半自動的に個々の用語の持つ概念を定義する事である。この方法では、資源とした元の用語集の情報によって制約を受けるが、比較的容易に遂行できること、新たに、資源として利用できる用語集が登場したときに、これを取り込むことによって、精度と適応範囲の拡大を図ることができること、これを繰り返すことによって、次第に理想的な用語集を構築できる可能性があること、などの利点がある。一方、複数の用語集の間に用語の使い方の矛盾がある場合には、これをそのまま内包してしまうという危険がつきまとう。矛盾がありそうな用語については、最終的には、人間の目で訂正・修正を行わねばならない。筆者らの経験では、このような用語集間の矛盾よりは、個々の用語集の内部の誤りの方が遙かに多かったことを指摘しておく。

９．謝辞

本研究の一部は厚生科学研究費補助金（情報技術評価総合研究）の支援を受けて遂行された。

１０．　引用文献

１）社会保険診療報酬支払基金. 診療科別標準傷病名集. 社会保険業務研究協会；1996.

２）社会保険診療報酬支払基金. 診療科別標準傷病名集. 社会保険業務研究協会; 1996.

３）日本医学会用語管理委員会. 医学用語辞典. 南山堂; 1991.

４）American Society for Testing and Materials. Standard Guide for Construction of a Clinical Nomenclature for Support of Electronic Health Records. ASTM E-1284-97.

５）厚生大臣官房統計情報部. 疾病、障害および死因統計分類提要－ＩＣＤ１０準拠. 厚生統計協会; 1995.

６）College of American Pathologists. SNOMED international. CAP; 1993.

７）里村洋一,佐々木哲明　ＩＣＤ－１０に準拠した標準病名集、 日本医事新報 1998; 3876:23-27,.

８）里村洋一,山﨑俊司,佐々木哲明　ＩＣＤ１０対応病名集と自動コーディングシステム　第１５回医療情報学連合大会論文集 1995; 957-960.

９）山﨑俊司, 廣瀬康行, 比嘉直樹, 劉亜斌, 里村洋一, 菅田厚.　病名マスターについての一考察－病名集変換作業を通して－　第19回医療情報学連合大会論文集 1999; 758-759.

１０）　大江和彦　電子的診療情報交換のための実用的な病名概念マスターの在り方　第２０回医療情報学連合大会論文集　2000;　2-A-7-3

１１）　Rossi Mori A, Consorti F, Gareazzi E. Standards to Support Development of Terminological Systems for Healthcare Telematics. *Meth Inform Med* 1998; 37:551-563.

１２）　National library of medicine. Unified Medical Language System 11th edition. NLM; 2000.

１３）　Cimino JJ, Clayton PD, HripcsaK G, Jhonson SB. Knowledge-based approaches to the maintenance of large controlled medical terminology. *JAMIA* 1994 ; 1(1):35-50.

１４）　Evans DA, Cimino JJ, Hersh WR, Huff SM, Bell DS. Toward a Medical-concept Representation Language. *JAMIA* 1994; 1-3:207-217.

１０．参照サイト

１）SNOMED http://www.snomed.org
２）UMLS　http://www.NLM.nih.gov/research/umls/umlsmain.html
３）Read code　**http://www.coding.nhsia.nhs.uk**
４）MEDIS **http://www.medis.or.jp/**

靭 (9078) ⇨ 靭 (E805)

鈎 (8A62) ⇨ 鉤 (E7EA)

腦 (E449) ⇨ 脳 (9450)

頸 (E8F2) ⇨ 頚 (8C7A)

髄 (E992) ⇨ 髄 (9091)

図１　ＪＩＳコードがそれぞれに与えられている同義文字（医療で汎用される文字の例）、日本医学会では右側の文字を採用しており、本研究でもこれに準じる。

英語	頻度	日本語	カナ	頻度	領域	品詞	優先度
INFARCT	40	梗塞	コウソク	49	M	1—名詞	a—優先語
ISOLATED	40	孤立性	コリツセイ	5	T	2—形容詞	a—優先語
ISOLATED	40	隔離の	カクリの	2	P	2—形容詞	a—優先語
ISOLATED	40	単独の	タンドクの	1	G	2—形容詞	a—優先語
MAGNET	40	マグネット	マグネット	7	P	1—名詞	b—同義語
MAGNET	40	磁石	ジシャク	3	P	1—名詞	a—優先語
MAINTENANCE	40	維持	イジ	24	P	1—名詞	a—優先語
MAINTENANCE	40	メインテナンス	メインテナンス	3	P	1—名詞	b—同義語
MAKERS	40	生産者	セイサンシャ	0	J	1—名詞	a—優先語
NYSTAGMUS	40	眼振	ガンシン	141	F	1—名詞	a—優先語
OXIDE	40	酸化物	サンカブツ	9	C	1—名詞	a—優先語
PERITONEUM	40	腹膜	フクマク	18	T	1—名詞	a—優先語
PLEURA	40	胸膜	キョウマク	34	T	1—名詞	a—優先語
SMEAR	40	スミア	スミア	26	P	1—名詞	a—優先語
SMEAR	40	塗抹標本	トマツヒョウホン	9	P	1—名詞	b—同義語
SOUND	40	音	オト	206	A	1—名詞	a—優先語
SOUND	40	音響	オンキョウ	16	A	1—名詞	b—同義語
SPINOUS	40	棘状の	キョクジョウの	2	T	2—形容詞	a—優先語
SYNTHETIC	40	合成の	ゴウセイの	2	C	2—形容詞	a—優先語
SYNTHETIC	40	合成薬品	ゴウセイヤクヒン	1	C	1—名詞	a—優先語

表１　英語対応した基本語の例。２列目の頻度はＳＮＯＭＥＤ中の出現頻度、５列目の頻度はＭＥＩＤ辞書に現れた翻訳語の頻度。領域はＳＮＯＭＥＤの軸に準じてある。

“ISOLATED”のように日本語が複数ある場合、領域毎に優先度を選んである。

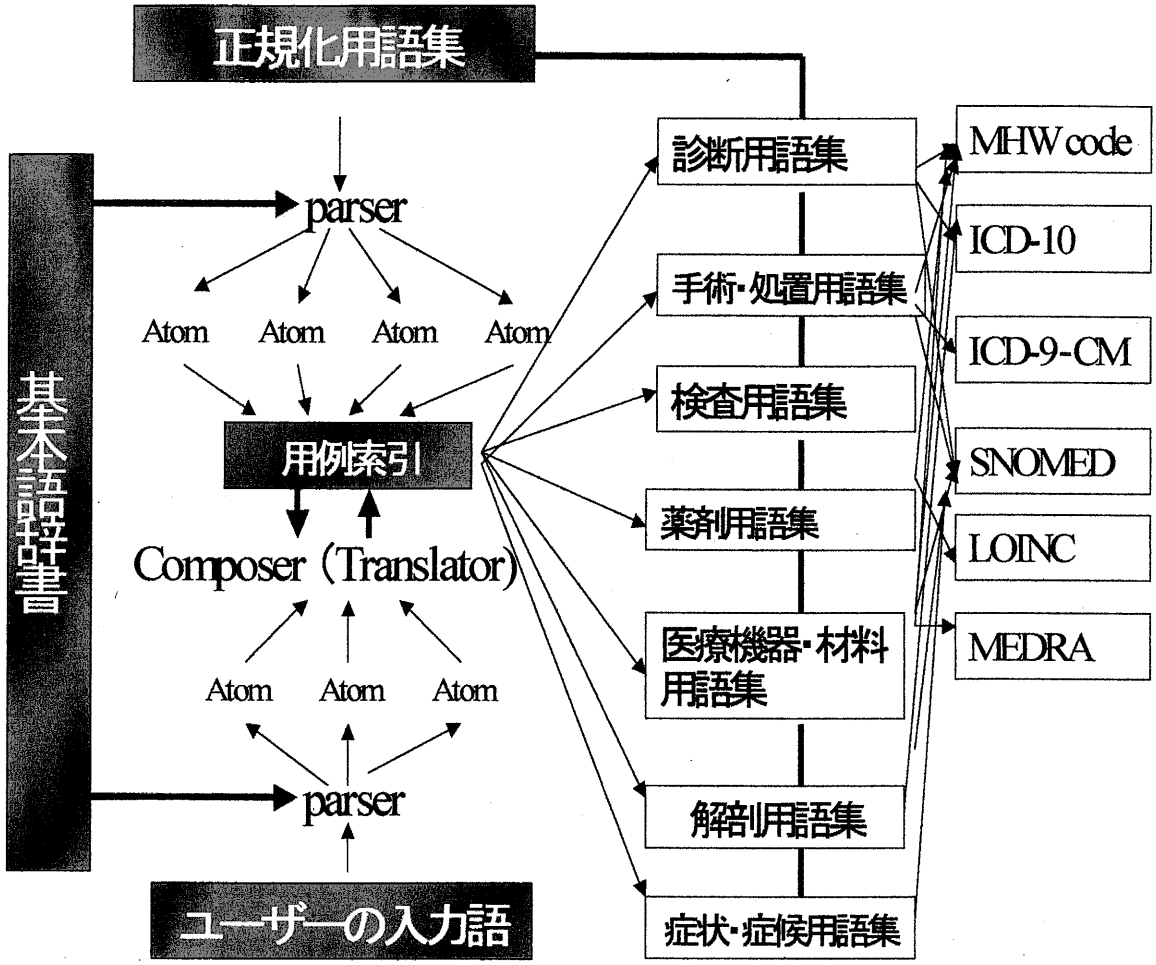

図２　用語体系の概念図　　目的別用語集（病名集、手術・処置用語集など）から標準用語を得て、これを基本用語に分解、基本用語の組み合わせから索引ファイルを作成する。
ユーザーが入力した（または、既存のテキストから取り出した）用語を、基本用語に分解し、基本用語の同義語や領域情報を利用して、標準用語に自動翻訳する。さらに標準用語の索引ファイルを利用して、目的別の用語集とそれに関連する分類体系を参照する事ができる。こうして、ユーザーが標準用語を意識せずとも、入力文の標準化が達成できる。

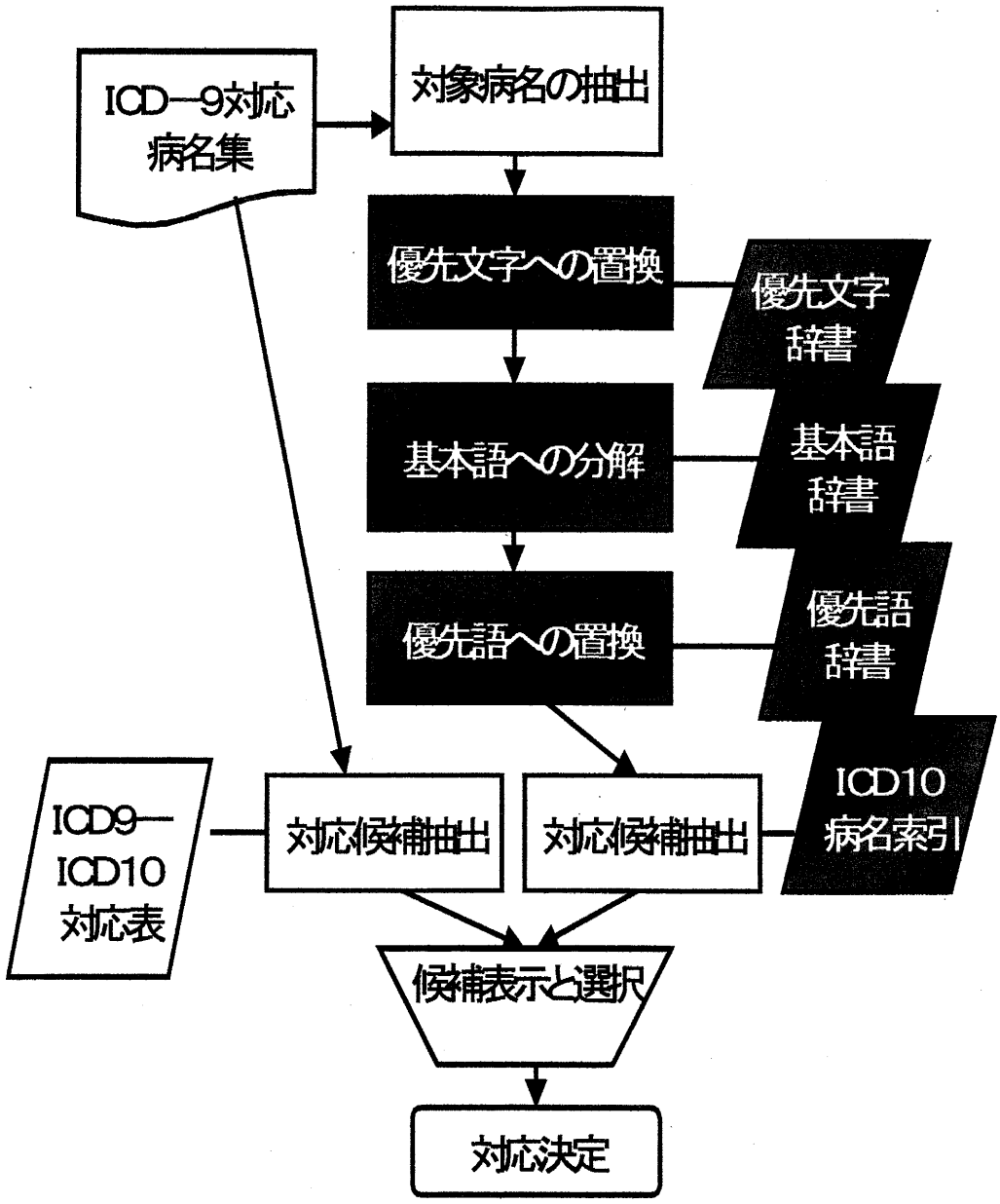

図３　既存の病名集からＩＣＤ１０対応標準病名集への変換過程
文字の正規化に続いて、基本用語に分解した後、同義語との入れ替えなどを行い、病名インデックスを参照する。こうして、オリジナル病名をＩＣＤ１０対応病名と対応させる。白抜き文字のブロックは用語集構造に関係する部分。