

厚生労働科学研究費補助金

医療技術評価総合研究事業

UMLSと連携した日本語医学用語シソーラスの作成
に関する研究

平成15年度 総括・分担研究報告書

主任研究者 脊山 洋右

平成16(2004)年4月

目 次

I. 総括・分担研究報告	
UMLSと連携した日本語医学用語シソーラスの作成に関する研究	1
脊山 洋右、 鈴木 博道	
II. 分担研究報告	
1. UMLSと連携した日本語医学用語シソーラスの作成に関する研究	5
(UMLSのメタシソーラスを用いた日本語病名用語のICD10コーディングに関する研究)	
開原 成允、 波多野 賢二	
2. UMLSと連携した日本語医学用語シソーラスの作成に関する研究	7
篠原 恒樹、 野添 篤毅	
3. UMLSと連携した日本語医学用語シソーラスの作成に関する研究	9
(UMLSとの機械的連携に関する研究)	
小野木 雄三	
4. UMLSと連携した日本語医学用語シソーラスの作成に関する研究	12
佐藤 恵	
III. 研究成果の刊行に関する一覧表	
なし	
IV. 研究成果の刊行物・別刷	
なし	
参考資料	15

厚生労働科学研究費補助金（医療技術評価総合研究事業）

総括・分担研究報告書

UMLS と連携した日本語医学用語シソーラスの作成に関する研究

主任研究者 脊山 洋右 お茶の水女子大学教授

分担研究者 鈴木 博道 （財）国際医学情報センター

研究要旨

EBM の重要な基礎となるエビデンスの集積や検索利用のためには、医学用語シソーラスの整備が不可欠である。全く初めからシソーラスを構築することを避け、米国の UMLS(Unified Medical Language System)と連携し日本医学会医学用語辞典・医学中央雑誌シソーラスその他既存用語集やシソーラスの活用によって、より実現性の高い日本語医学用語シソーラス開発技法の確立を目指して調査検討し部分的なシソーラス開発を試みた。NLM(National Library of Medicine)との共同研究開発プロジェクトとして主任研究者・分担研究者・研究協力者の一体となった共同作業を中心に研究を進め、試作したシソーラスによって①日本語医学用語による Pubmed 検索、②学会投稿演題の自動分類、の小規模な実験を行った。

分担研究者

開原 成允

国際医療福祉大学副学長 教授

野添 篤毅

愛知淑徳大学 教授

小野木 雄三

東京大学大学院 助教授

佐藤 恵

(財)医療情報システム開発センター

主任研究員

篠原 恒樹

医学中央雑誌刊行会 理事長

鈴木 博道

(財)国際医学情報センター

事業推進室長

A. 研究目的

実用的な日本語医学用語シソーラスの開発技法について、UMLS を利用し、既存の医学用語集やシソーラスなどの医学用語をコンセプトに対するマッピングする方法を検討し、試み、評価の一助とすべくシソーラス使用実験を行うものである。結果として、日本語医学用語シソーラス開発手法のフィージビリティを把握できる。本研究は NLM との連携協力によって進めており、その成果は将来の国際的情報交換にも活用可能である。

B. 研究方法

研究初年度から NLM との連絡・情報交換、米国を中心とした UMLS の実用例の調

査、非英語圏における国際的なシソーラス開発・活用に関する事例の把握、を実施してきており、先行する実験・研究などの成果をいち早く反映させる様に努めてきている。また、連携の対象となる UMLS に関する詳細な中身の把握を行い、各種医学用語集などとのマッピングを行うことによって UMLS 自体の部分的評価も得られた。

シソーラス開発は、当該医学用語の英語表記をプログラムによって正規化処理した上でコンセプトに対応付ける手法を核とした、機械的処理が基本となっている。また、この結果を点検し、問題点を抽出し、問題点に関する判定を投票するための、Web システムを構築し、班員・協力者などで利用した。

シソーラス開発技法を評価するためには、開発手順や開発コスト、シソーラス開発上の問題点などを明らかにすることは、当然、必要不可欠である。同時に、成果となるシソーラスが実用に耐えるものでなくてはならない。それ故、研究 3 年目の平成 15 年度には、開発技法に係わる問題点の把握と、部分的に完成されたシソーラスを使った実用性テストも実施した。具体的な方法は、シソーラスのポキャブラリー・ファイルを活用・評価するために日本語医学用語から NLM による英語の文献データベース (Pubmed) 検索を試みた。また、シソーラスのコンセプトに基づく階層性を活用した学会演題自動分類の実験も行った。

なお、直接本研究の成果にはなりがたいものの、UMLS に日本語医学用語を付与する件についても、NLM と協議を進めてきている。

(倫理面への配慮)

医学用語を対象としており、プライバシーに係わる様な情報は扱っていない。

C. 研究結果

UMLS について、米国内並びに英語圏の国を中心に NLM との共同研究プロジェクトとして様々な実用化実験が行われてきているものの他にも、文献・研究報告・診療データ管理などの応用研究は多数行われており、利用価値はそれなりに評価されてきている。しかし、一歩英語圏から出ると、Mesh(Medical Subject Headings)を介した UMLS との連携に、ほぼ留まっているのが現実となっている。

一方では、今回本研究班で実際に試したことと同様に、各国語 (ドイツ語・フランス語・ロシア語・スペイン語・フィンランド語他) の各種医学用語集・シソーラス (ICD10、MedDRA、SNOMED、ICPC) など UMLS に対するマッピングが多数行われてきており、UMLS が国際的な医学用語シソーラスのスタンダードになりつつあることが窺えた。

構築したモデル・シソーラスは全 50,000 語、コンセプトとして 20,000 を超えるものであり、分担して行った点検作業の結果この内 9% 程には何らかの問題が指摘された。モデル・シソーラスの対象語は UMLS の性格上からも Mesh 由来の用語が圧倒的である。

平成 15 年度に力を入れて取り組んだモデル・シソーラスの実用性テストについては、12 月 19 日 (金) に拡大班会議の場で班員・協力者、更に研究班関係者以外の医学用語シソーラスに興味ある人達にも披露・紹介し、意見などを聴取した。また、

日本語モデル・シソーラスによる Pubmed 検索は、3月上旬 NLM に打合せに赴いた際、デモンストレーションを行った。実用性テストについては、今後更に継続的に広く紹介して行くことで、適切な評価が可能となる。

D. 考察

シソーラスの開発技法に関するモデル、教科書は、現在存在していないが、かつては、UNESCO 他国際機関によるシソーラス開発ガイドラインが存在した。基本的手法は、論題中のキーワードをカード化し、類似語・同義語を集め、各ボキャブラリ間の関連づけを定義して行く、といった手順であって、現代のコンピュータ時代には大凡マッチしないものである。今回採用した技法は、人手を掛けることは最小限の点検などに押さえて極力機械的手段によるものであり、問題点の発生率、実用性テストの中間状況からも、概ね有効性が得られたと考えている。但し、今回開発を試みたシソーラスは、いわば機能するシソーラスであり、表示を目指したリードターム決定などには関知しない構成に留めているものである。リードタームの問題は今後の課題と考えている。

構築したモデル・シソーラスの実用性テストについて、未だ、評価に十分な検討まで進んでいない。日本語医学用語による Pubmed 検索を試した人達からは好印象は得られている。また、NLM でのデモンストレーションでは、日本語で Pubmed 検索をそして逆に英語で日本語データベース（医中誌 DB）検索を、との趣旨で紹介したところ、非常に好評であった。更に、非英語

圏の各国で UMLS 連携の各国母国語シソーラスを構築してくれば、NLM として好ましいことである、という別の観点からの評価も受けた。NLM の負担は無く、UMLS という研究プロジェクトの立派な成果でもある。

学会演題自動分類は、発表演題の Web 登録が普及してきている現在、セッション分けやプログラム編成を旧来通りの手作業で行うことは、演題数の多い学会では無理であろうとの趣旨から、シソーラスのカテゴリーや階層性を活用した試みである。学会の対象分野によって利用する階層を工夫することは必須であるものの、期待通りの結果は得られた。

なお、UMLS に日本語医学用語を搭載する件は、NLM 側の Mesh 翻訳支援システムとの連携をも対象に、更に協議を続けることになっている。

E. 結論

コンピュータ処理を中心として UMLS との連携並びにコンセプトの利用、日本医学会医学用語辞典・医中誌シソーラス・生化学用語辞典・看護学事典・MedDRA/J・MEDIS 病名ファイルその他とのマッピング、を前提とした日本語医学用語シソーラス開発技法は、未だ十分な根拠があるとは言えないものの、以下の理由から有効であった。シソーラス開発の効率、出来上がったシソーラスの実用性、グローバルな情報交換を可能とする、と言う 3 つの理由である。

今後、特に実用性に関するより充実した検討が望まれる。

F. 健康危険情報

特になし。

G. 研究発表

シソーラスの開発という共同作業について、進行に応じ、各年度毎にその一部を研究班員・研究協力者・各員の所属機関で分担実施してきている。それ故、全体の成果をまとめて発表する以外の研究発表は未だ実施できていない。

3年間の研究全体の概要を簡単にまとめ、研究発表に代わる参考資料として添付した。

H. 知的財産権の出願・登録状況

特になし。

UMLS と連携した日本語医学用語シソーラスの作成に関する研究
(UMLS のメタシソーラスを用いた日本語病名用語の ICD10 コーディングに関する研究)

分担研究者 開原 成允 国際医療福祉大学副学長 教授

研究協力者 波多野 賢二 東京大学医学部附属病院中央医療情報部

研究要旨

UMLS のメタシソーラスを用いた日本語病名用語の ICD10 コーディングの可能性を検討するため、比較的シンプルな方法で日本語病名用語と ICD10 コードのマッチングを試みた。

A. 研究目的

WHO が刊行する国際疾病分類第 10 版 (ICD10) は現在日本において公的に利用される主要な病名コードであり、医療機関にとって病名用語に適切な ICD10 コードを割り当てることが重要な課題である。一方 UMLS のメタシソーラスには ICD10 分類項目のすべてが概念として登録され、それぞれのレコードには固有の概念コードである CUI (Unique Concept Identifier) が割り振られている。よって、日本語の病名用語に対応する CUI を見つけられれば、それを仲立ちに病名用語に適切な ICD10 コードを与えることが出来る。

このような ICD10 コーディング手法を活用することにより、ICD10 コードの未知の病名用語に対して CUI を介して自動的に ICD10 コードを割り振ることができる可能性がある。このようなシステムが実現すれば、これまでのコードブック片手の手作業に頼っていた ICD10 コーディング作業の効率性と正確性が向上し、ICD10 コードベースの疾病分類による包括診療報酬などが始まりつつある医療現場への貢献が大いに期待できる。

今回の研究では、このような UMLS のメタシソーラスを用いた日本語病名用語の ICD10 コー

ディングの可能性を検討するため、比較的シンプルな方法で日本語病名用語と ICD10 コードのマッチングを試みた。

B. 研究方法

対象は日本医学会医学用語辞典および医学中央雑誌刊行会医学用語シソーラスに収載された日本語医学用語である。それらの日本語医学用語の英語訳フィールドと UMLS メタシソーラスの MRCON テーブル STR (文字列) フィールドや MRXNW.ENG テーブルの NWD (正規化用語) フィールドを対応させることにより、日本語医学用語に対応する CUI を見つけることが出来る。さらに、UMLS の MRSO テーブルを通じて、CUI に対応する ICD10 コードを見つえられる。

このような操作によって、日本語医学用語に対応する ICD10 コードをほぼ自動的に割り当てることが可能である。ただし、ICD10 コードは病名 (疾患名) に与えられるコードであるため、すべての CUI に ICD10 コードが付与されているわけではない。そのため、対象となる CUI は ICD10 コードが対応している (MRSO で ICD10 由来の) CUI のみとなる。

C. 研究結果

MRSO テーブルで、ICD10 由来の CUI は延べ 13505 レコードである。複数の ICD10 コードが同一の CUI に対応する重複を除くとユニークな CUI は 11564 個であった。

日本医学会医学用語辞典、および医学中央雑誌刊行会医学用語シソーラスに収載された医学用語のうち、その英訳語が UMLS メタシソーラスの CUI と対応付けられた用語は 117252 語で、そのうち MRSO テーブルで ICD10 由来の概念であった用語（すなわち ICD10 コードが付与できる用語）が 25405 語であった。その中で、ICD10 準拠標準病名マスター・厚生労働省が刊行する ICD10 日本語訳などにより ICD10 コードが既知の病名用語は 24640 語、上記の資料に一致する用語がなく、今回新たに ICD10 コードと対応付けられた病名用語は、785 語であった。
(表)

D. 考察

今回のマッチング作業で、表にあるようないくつかの重要な新しい病名用語見だし、正確な ICD10 コードを付与することが出来た。ただし、今回新たに ICD10 コードと対応付けられた 785 語の病名用語のうち多くは、ICD10 コードが既知の用語の倒置形・かな漢字または英語日本語などの表記の違いなど、同一概念の表記上のバリエーションであった。すなわち、見出された ICD10 コードに対応する新しい概念は数としては少数であった。現状のマッチ率では、入力された病名用語に自動的に ICD10 コーディングを行うシステムを実用化するのは困難といわざるを得ない。

しかし、今回取ったようなシンプルなものマッチングの手法では、UMLS の概念データベース体系の一部のみ利用したに過ぎず、概念間の関係を記述した MRREL テーブルのデータを用いて当該概念の上位・下位概念を検索するなどの工

夫をすることによって、さらにマッチ率を向上させることも可能である。また、UMLS のもうひとつの概念データベースである Semantic Network との連携なども行い、実用的なマッチ率が得られる自動コーディングシステムの可能性を探る予定である。

表：ICD10 コードと対応付けられた用語（一部）

病名用語	ICD10 コード
胎児性アルコール症候群	P04.3 母体のアルコール使用<飲酒>により影響を受けた胎児および新生児
肺ヒアリン膜症	P22.9 新生児の呼吸窮<促>迫, 詳細不明
小唇症	Q18.7 小(口)唇(症)
滑石肺症	J62.0 タルク粉じん<塵>によるじん<塵>肺(症)

E. 結論

日本医学会医学用語辞典および医学中央雑誌刊行会医学用語シソーラスに収載された病名用語に自動的に ICD10 コードを付与するシステムの実用化は困難であり、これを実現するためには、UMLS のもうひとつの概念データベースである Semantic Network との連携なども同時に行っていく必要がある。

F. 健康危険情報

特になし

G. 研究発表

なし

H. 知的財産権の出願・登録状況

特になし

厚生労働科学研究費補助金（医療技術評価総合研究事業）
分担研究報告書

UMLS と連携した日本医学用語シソーラスの作成に関する研究

分担研究者 篠原 恒樹 特定非営利活動法人医学中央雑誌刊行会理事長
分担研究者 野添 篤毅 愛知淑徳大学図書館情報学科教授

研究要旨

UMLS と連携した医学用語シソーラス作成にあたり、日本国内における医学文献検索に実績のある「医学用語シソーラス 第5版」（医学中央雑誌刊行会発行）を用い、本研究班班員、小野木雄三考案による「登録語彙修正投票システム」を用い、投票による修正投票を行った。

A. 研究目的

米国国立医学図書館(National Library of Medicine:以下 NLM)が医学文献検索に使用しているメタシソーラス、UMLS(Unified Medical Language System)は多種の医学用語を包括しており、多言語にも対応が進んで来ているが、日本語の対応がまだ行われていない。そこで、日本医学会用語集を始めとする医学用語とのリンクを作成することは医学研究者にとって大いに役立ち、かつ医学研究の発展に大きく貢献できるものと思われる。UMLSの中にはPubMed等、NLMで提供されている医学文献データベースに索引語として付与されているMeSH(Medical Subject Headings)をも当然のことながら含まれており、これらのシソーラス用語との日本語対応を如何にして、作成できるかを試みた。

B. 研究方法

医学用語集(辞書含む)として、日本国内で使用されている用語と UMLS 2000(2000

年発行)を用い、マッチングを行った。以前、日本医学会用語集(83,791語収録)とのマッチングを行った際には、UMLSの34.8%の用語がマッチした。今回は、UMLSの中で、特に根幹となるMeSHのConceptとなる用語との日本語対応が取れないかとの観点から、種々試みた。日本国内において、MeSHのConceptとの連携が取れている用語集は医学中央雑誌刊行会で発行している「医学用語シソーラス」しかなく、一部、MeSHの翻訳語集が紀伊國屋書店から発行されているが、2000年発行のMeSHを最後に未だアップデートされていない。医学中央雑誌刊行会発行の「医学用語シソーラス」(以下、医学用語シソーラス)は版を重ね、現在、2003年発行の第5版が発行されている。そこに収録されているディスクリプタは日本特有の保健機構等を除き、略99%の用語がMeSHとのConceptレベルでのリンクが取れている。分野は異なるが、MedDRA/Jなどでは、約10%程度である。UMLSに含まれる、ICD-10の訳語はすでに対応している、

というのが現状である。また、MeSH にある用語で「医学用語シソーラス」に登録されていないものは、米国特有の保健機構等であり、疾病、医薬品名等はすべてリンクがあり、MeSH のカテゴリーコードとも略一致している。カテゴリーによるマッチングも可能と言える。これはマッチングが正確に行える一つの要素と言える。今回は UMLS の用語のうち、MeSH と「医学用語シソーラス」とのマッチングを行い、日本語シソーラスとしての適語、不適語、また、他候補（日本語としての）の選択を小野木雄三考案による〔登録語彙修正システム〕を用いて行った。また、「医学用語シソーラス」に含まれない、MeSH の用語に関しては、医学中央雑誌刊行会で昨年、リンクを作成し、医学文献の検索に十分対応できるよう、作成した。この為、MeSH の用語と「医学用語シソーラス」はすべて、Concept レベルでのリンクが取れ、今回の投票システム作成の前準備がすべて整ったと言える。投票による、選定内容は、(I) 同義語追加提案、(1)、(2)、(II) 不適語削除からなっており、医学文献の検索等を十分に考慮し、それぞれの用語について、投票による選定を行った。(今回は特に倫理面での配慮はしていない。というのは一部研究者のみの閲覧によるものである為)

C. 研究結果

同義語による追加提案の中、(I)同義語提案(1)の約 600 語について、投票による用語選定を行った。また、(II)「不適語」削除 103 個についても同様に行った。一部、変換ミスによるものが含まれていたが、作業には特に支障はなかった。

D. 考察

提案を認める(OK)が多数を占めたが、却下するものも数%あった。倒置形と呼ばれる MeSH 特有の用語形態も「医学用語シソーラス」は対応しているので、特に、選定に問題はなく提案された追加も略、受け入れられるものと考えられた。

E. 結論

今回携わった UMLS の MeSH という分野に関しては、「医学用語シソーラス」に一部追加等の作業を加えれば、日本医学用語シソーラスとしての役割は十分に果たしうられると思われる。

F. 健康危険情報

特になし。

G. 研究発表

なし。

H. 知的財産権の出願・登録状況

特になし。

厚生労働科学研究費補助金（医療技術評価総合研究事業）
分担研究報告書

UMLS と連携した日本語医学用語シソーラスの作成に関する研究
（UMLS との機械的連携に関する研究）

分担研究者 小野木 雄三 東京大学大学院医学系研究科特認助教授

研究要旨

日本語の医学用語と英語の医学用語との意味上の対応関係を確保することは極めて重要である。当研究班では日本語医学用語シソーラスの内容を点検し、UMLS と精度良く対応するようにした。

A. 研究目的

医学文献をはじめとして、国内外の情報を相互に利用する機会は増加の一途を辿っている。しかるに、日本語の医学用語と海外、特に米国の医学用語とが正しく対応しているか否かは極めて重要な問題である。ここで単なる英和/和英辞書的な対応関係を対象にすると、語彙の同義性と多義性の両方があるために結果が定まらない。まずは同義語のみを対象とし、医学概念とそれに属する同義語を持つシソーラスを元に、日本語の医学シソーラスと米国の代表的なシソーラスである UMLS との対応関係を確認することに意義がある。

そこで日本語の医学用語シソーラスと UMLS とを概念レベルで比較し、日本語医学用語の語彙を再点検することにより、日本語医学用語シソーラスの精度を向上させることを目的とする。またその応用として、日本語医学用語から PubMed (NLM の医学文献データベース) を直接検索できるようにする。

B. 研究方法

材料として日本語医学用語シソーラスと UMLS を使用する。日本語医学用語シソーラスは医学中央雑誌刊行会によって作成され、国内で出版された文献の内容を分類するためのキーワードを同義語 (シソーラス) として概念ごとにまとめたデータベースである。UMLS は米国 NLM によって作成され、海外の様々な医学用語集 (シソーラス) を同じ概念ごとにまとめなおしたメタシソーラスであるが、いまだに日本語は含まれていない。

日本語語彙と UMLS 概念との対応確認には、文献から得られた語彙が主体であることを考え、UMLS の大部分を構成する MeSH を対象とした。MeSH との対応が取れば、UMLS のメタシソーラスとしての性質から他の統制用語との対応も取れることになる。

語彙対応関係のチェックを行うために、インターネット上に確認作業のウェブインターフェイスを構築した。当研究班の班員はこれにアクセスし、日本語シソーラスの

各語彙が適切であるか否かの判定を概念ごとに行い、不適切と考える場合には削除提案を、より適切な語がある場合には提案を、疑問点がある場合にはその旨をコメントとして残し、他の班員がそれを閲覧できるようにした。その後、判定結果を集計し、不適切語彙および提案語彙それぞれに対し、インターネットを介して班員全員で投票作業を行うことにより、多数決で各語彙の判定を決定した。

C. 研究結果

作業対象である日本語シソーラスの全語彙数は 50322、概念数は 20743 であった。判定作業により、不適切および提案語が提示された概念数は 1865 件であり、これは全体の 9%にあたる。これから導出された削除提案語数は 154 語、新規提案語数は 689 語、また語彙の一部を交換しても同じ意味になるもの（例えば「蛋白」、「たんぱく」、「タンパク」）を機械的に生成したものが 1958 語であった。以上を投票で多数決をとり、最終的 UMLS と極めて良好に対応する、日本語医学用語シソーラスを構築した。

この応用として日本語からの PubMed 医学文献検索を行えるようにした。その仕組みを以下に示す。まず検索ウェブページに日本語語彙を入力する。システムはこの語彙を含む語彙を日本語シソーラスの中から探し出し、一致した語彙が含まれる概念を集約し、その概念を特定するリードタームのリストを提示する。ユーザーはそのリストから検索対象を選択する。するとシステムは選択された概念に対応する MeSH 語彙（英語）を使って PubMed 検索を行い、対応する文献の一覧を提示する。なお、この

日本語 PubMed 医学文献検索ページは当研究班の成果としてインターネットに公開されている。

D. 考察

語彙の判定作業をインターネット上で行う手法は極めて有効であった。従来は班員が物理的に 1 箇所に集まって作業をしなくてはならなかったため、作業がなかなか進まなかったが、この方法なら各班員は自宅からでも研究室からでも都合の良い時に好きな分量だけ、随時作業を行うことができるからである。このような作業形態は、今後どんどん導入していくべきと考える。

日本語と英語の医学用語が概念レベルで適切に対応すると、例えば文献検索ではより高精度な結果を得ることができるようになる。またシソーラスとして適切に構成された日本語医学用語の存在価値も極めて大きいものと考えられる。今後電子カルテをはじめとして医療および保健衛生に関する文書が流通するに従い、医学用語の共通理解が極めて重要になる。すなわち医療用語の標準化としての側面から、ある医学概念を表記する様々な同義語を定義したことの意義は大きい。

応用として日本語キーワードから英語文献の検索を行ったが、逆に英語キーワードから日本語文献の検索を行うことが可能になり、国際化に大きく貢献すると考えられる。このことは同時に、国内の利用者の利益にとどまらず、海外の人々にとっても大きな福音となることが期待される。海外には英語を主とする研究者の他にも、多国籍・多言語の人々が医学・保健衛生に関する情報を求めている現状があるからである。

E. 結論

UMLS と精度良く対応する日本語医学用語シソーラスを構築した。

F. 健康危険情報

特になし。

G. 研究発表

1. 論文発表

2. 学会発表

Yuzo Onogi 他, Mapping Japanese Medical Terms to UMLS Metathesaurus. MedInfo 2004 受理

H. 知的財産権の出願・登録状況

特になし。

厚生労働科学研究費補助金（医療技術評価総合研究事業）
分担研究報告書

UMLS と連携した日本語医学用語シソーラスの作成に関する研究

分担研究者 佐藤 恵 （財）医療情報システム開発センター主任研究員

研究要旨

標準病名マスターに日本医学会用語集、日本内科学会用語集などの病名用語を連結、英語表記を付した上で、UMLS へのマッピングを試みた。学会横断的な用語標準化の一つの契機およびその手段としては有望である。

A. 研究目的

ICD-10 対応電子カルテ用標準病名マスター（標準病名マスター）は、厚生省（1994 年当時）が保健医療福祉情報システムの開発基盤を整備するために、各種用語・コード体系の標準化を積極的に推進することの一環として（財）医療情報システム開発センター（以下 MEDIS-DC）に新医療情報通信網基盤整備開発事業（厚生省健康政策局委託事業：1994 年より 3 年計画）を委託し、MEDIS-DC がその一つとして、1998 年に ver.1.00 を公開したマスターである。

病名表現の標準化を促進するために、Ver.2.00 への更新時に「Lead Term」をもつマスターとして再構築された。

2000 年にはレセプト電算処理マスターの傷病名マスターと統合され、更に厚生労働省保険局医療課長通知によりいわゆるレセプト病名として認知されることにより、医療施設における利用が急速に広まりつつある。

標準病名マスターはこのように政策的にも市場からも国内で「標準」の位置を確立しつつあり、今後は MedDRA や SNOMED などの用語集との連携も視野に入れる必要性も想像される。

本研究の目的は、米国国立医学図書館の長期プロジェクトである UMLS との連結を試行することにより、他の用語集との連携の可能性お

よびその手法を探ることである。特に今回は病名用語を対象とした。

B. 研究方法

1. 利用したファイル

- ICD-10 対応電子カルテ用標準病名マスター ver.2.22
- 日本医学会用語集
- 日本内科学会用語集
- UMLS 2002AC
- 医学中央雑誌シソーラス 同義語テーブル

2. 研究方法

(1) 標準病名マスターと日本医学会、日本内科学会用語集との連結

標準病名マスターを構成する 3 つのテーブルのうち、病名基本テーブルに収載された「表明表記」と日本医学会、日本内科学会用語集の用語とを連結する。「病名表記」とは標準病名マスターにおいて病態毎に設定された「Lead Term」である。

本作業を通じ、用語集と共通して収載される病名用語に英語表現を付与した。

(2) UMLS 2002AC を用いての作業用テーブルの作成

今回は UMLS の内容から icd_cui_mesh を連結し、医学中央雑誌シソーラスの同義語テ

ーブルから日本語を付与したテーブル (UMLS 作業テーブル)を用いた。

(3)標準病名マスターと UMLS 作業テーブルとの連結

標準病名マスターの「病名表記」と生成した UMLS 作業テーブルの文字列とを日本語表現、英語表現のそれぞれをキーに連結を行った。

C. 研究結果

D. 考察

今回試行に用いたファイル中、「病名表記」、日本医学会用語集、日本内科学会用語集の用語表現については一応の正規化がされていたものの、UMLS 作業テーブルの用語は、英語表現の正規化が不十分なものであった。

その結果比較以前の問題も含んだ結果となり、今回は数字での結果提示を断念した。連結が成立した組み合わせのうち、今後の作業の方向性を示す例として以下を示す。日本語表現、英語表現をキーとした連結例である。

①病名表記「帯状疱疹」を日本語表現で連結した結果

ソーステーブル	英語表現 1	英語表現 2	英語表現 3
UMLS 作業テーブル	Herpes Zoster		
日本医学会	zoster	shingles	
日本内科学会	zoster	cingulum	cingule

②病名表記「完全無歯症」を英語表現「anodontia」で連結した結果

ソーステーブル	日本語表現 1	日本語表現 2	日本語表現 3
UMLS 作業テーブル	無歯症		
日本医学会	完全 無歯症		
日本内科学会	無歯症		

日本語表現をキーにした場合も、英語表現をキーにした場合も、それぞれソースとなった用語集の収載内容の標準化が不十分であるという問題が浮かび上がっている。

また日本語表現をキーとした場合、付与されている英語表現が「適切な」英語表現であるかどうかの検証にもつながると考えられる。

E. 結論

本試行には日本医学会と日本内科学会の用語集のみを用いたが、筆者は標準病名マスターの更新作業を通じ、他の学会用語集とも接する機会が多く、学会を跨っての表現の標準化の必要性を強く考えているものである。今回 UMLS との比較検討を行った結果、正規化を欠いて十分な結果が得られなかったものの、学会横断的な用語標準化の一つの契機およびその手段としては有望であると判断した。

これはひいては国内において単独で十分であるという評価を受けつつある標準病名マスターと MedDRA 等の連携を進める上で、必要な作業と考えられる。

今後は元となるファイルの用語表現の正規化を十分に行い、また他の学会用語集などを集めるなどして、より実質的な作業に取り組むことが必要である。

また範囲を病名表現のみでなく修飾語表現その他にも広げ、その作業を通じて同

時並行的に医学用語の標準化の手段を提供
することも望まれる。

参考資料

「UMLS と連携した日本語医学用語シソーラスの作成に関する研究」3年間のまとめ

1. 日本語医学用語シソーラスの作成の目指すもの

日本語の医学用語は日本医学会の100近い分科会でそれぞれ固有の表現が使われてきたが、標準化を目指して日本医学会医学用語管理委員会により「日本医学会医学用語辞典英和」が平成3(1991)年に刊行され、平成15(2003)年には文部科学省により「学術用語集医学編」が刊行されるなど、医学用語の統一が図られてきた。

「日本医学会医学用語辞典英和」は7万9千語余りの用語を記載し、「学術用語集医学編」が1万9千語余りを収録しているものの、いずれも単に用語を五十音順に並べたものといっても過言ではない。個々の用語が持つ概念を定義し、上位語、下位語といった関係を明らかにして、同義語を示すという作業は為されておらず、シソーラスとしての体系をなしていないのが実情である。

コンピュータ技術の飛躍的な発達により、文献検索もWEB上で瞬時に行える時代になったが、そのためには日本語医学用語シソーラスの構築が欠かせない。米国の国立医学図書館(National Library of Medicine, NLM)がMeSHという英語による医学用語シソーラスを開発したお陰でMEDLINEという文献検索システムが稼動するようになって、現在では11,000,000件の論文が瞬時にして検索できるようになった。NLMではMeSHを骨格として数多くのシソーラスを統合したUMLS(Unified Medical Language System)というメタシソーラスを構築してきた。

日本語医学用語シソーラスができると、日本語で書かれた論文のデータベースでの検索が可能になるだけでなく、WEBを通じてMEDLINEを日本語で検索することができるようになる。また、このシソーラスがUMLSに組み込まれると、日本語の論文もMEDLINEの対象となり、情報交換の範囲が格段に広がることが期待される。

また、文献データの蓄積はEBMの骨格をなすが、シソーラスが存在しないと文献の内容が同じものを意味しているか否かがはっきりせず、Evidenceを形成する場合の大きな障害となる。標準病名用語集の整備はこの観点に沿ったものである。

NLMが開発したUMLSを背景として、日本医学会が作成した医学用語辞典及び医学中央雑誌刊行会のシソーラスを連携させることによって短期間に概念の整理、用語の標準化を図るのがこのプロジェクトの目的であり、MEDLINEの日本語による検索、学会演題の自動分類によるプログラム作成支援などの応用の道も探ることを目指した。

2. NLMのUMLSプロジェクト

NLMはメリーランド州ベセスダにあり、隣接するリスターヒル国立生物医学情報センター(Lister Hill National Center for Biomedical Communications)を中心に1968年

以来 30 年以上にわたって Health Information Resource Database の作成に取り組んでいる。

我々は 2002 年 2 月に NLM を訪ね、Library Operations の Associate Director である Betsy Humphreys 及び MeSH Section の責任者である Stuart Nelson らと本プロジェクトの研究方針および協力方法を協議した。UMLS の英語以外の言語によるレコード採用の状況を把握したことと、MeSH の翻訳ではなく日本語のシソーラス構築の必要性を先方に理解させ、相互協力の体制が整ったことは大きな成果であった。特に UMLS への日本語レコード追加のプロセスが協議されたことは意義深いものがある。NLM の Director である Donald A. B. Lindberg との会談でも、我々の研究の意義と社会への貢献が話し合われ相互理解が深まった。

UMLS の Metathesaurus に採用された英語以外の用語はドイツ語が最も多く、次いでスペイン語、ポルトガル語、ロシア語などヨーロッパ言語であるが日本語は入っていない。MEDLINE のアクセス数で見ると、日本が 8% を占めアメリカ以外の国では最多であるにもかかわらず、その言語が検索対象となっていないことは情報交換を通じた研究成果の共有の上で負の要因といえよう。

3. 日本語医学用語の UMLS へのマッピング

日本医学会の「日本医学会医学用語辞典英和」、日本生化学会の「生化学用語辞典」、メヂカルフレンド社の「看護学大辞典」、文部科学省の「学術用語集医学編」に記載された日本語医学用語と UMLS の MeSH との対応を調べたところ、機械的なマッピングからは UMLS の 3%、医学会用語辞典の 40% にとどまることが明らかになった。形容詞の扱い方の違いなどもマッピング率の低さの一因であるが、マッピングされた用語についても、文字列からの一致が概念としても妥当な対応になっていないケースも明らかになり、個々の対応の点検と評価の必要性が浮かび上がってきた。

そこで、平成 15 年度は連携できた用語とできなかった用語を一語ずつ検討し、機械的な問題点を把握して、問題点の解決を図った。この際、概念と上位語、下位語の関係、同義語としての認定など、シソーラスとして必要な検討を加え、UMLS との対応を検討した。

指摘された問題点は整理して、その妥当性を検討した上で班員による合議に相当する、WEB 上での投票というプロセスを経てシソーラスの純化を図った。これまでに構築された日本語医学用語のシソーラスの規模約 50,000 語からなり、コンセプトとしては 20,000 を超えるものである。このうち、約 9% に何らかの問題点が指摘されたが、班員による投票というプロセスで、多くのものが系統的に整理された。

4. 日本語による MEDLINE の検索

MEDLINE は英語で検索するのが常道であるが、このプロジェクトで作成している日本

語医学用語シソーラスを経由すると日本語での検索が可能になる。例えば、「腎%」という日本語でシソーラスを表示させると 67 件の候補リストが表示されるので、そのうちで「高血圧一腎血管性」をクリックすると「Hypertension, Renovascular」として検索される。

このことは日本側だけでできるプロセスだが、NLM の UMLS が日本語シソーラスを認知して取り込んでくれると、双方向の検索が可能になる。英語の世界から日本語の文献が検索対象になるわけで、日本語で書かれた論文が世界から WEB を介して読まれる可能性が出てきたといえよう。

プロトタイプとしての日本語による検索ウィンドウはできたので、用語の規模の拡大と、日本語文献のデータベース構築が次の課題といえよう。

5. 学会演題の自動分類

日本生化学会は年に 1 回、会員の研究成果を発表する場として大会を行っている。2,000 題以上の演題が寄せられ、これを 3 日間で 30 近い会場で講演するが、そのプログラム作成に本研究で開発した医学用語シソーラスが役立つのではないかと考え、平成 14 年に京都で開催された第 75 回大会の演題を用いてシミュレーションを行った。

各演題には発表者によって「発表希望分類」と英語の「キーワード」が付けられているので、これらを対象にシソーラスをあてはめて分類し、数十題のグループに分けることを試みた。「キーワード」は分類のためのものではなく、抄録号の索引のために過去 5~6 年にわたって集積されてきたもので、1,720 語からなっているが、そのものはシソーラスの概念を入れた階層構造をなしていない。

このような限られた情報に基づく試みであったが、ほぼ希望サイズのグループを作ることができ、マンパワーで数日かけて行った作業を瞬時にして提示することができた。

勿論、発表者によって「キーワード」に対する考え方も異なるので、最後は人間の目で見た調整が必要なことは言うまでも無い。しかし WEB で受け付けた 2,000 題以上の演題が締め切った直後にグループ化されて示されるということは、その後の作業に有益なことである。

6. 非英語圏の文献検索

UMLS ではドイツ語をはじめ 13 カ国語の用語を組み入れているが、英語に比べると 2% に過ぎない。ましてや、いずれも訳語としての取扱いであって、このプロジェクトのように独自のシソーラスを構築して双方向にリンクするという考えには至っていない。翻訳から得られるのはあくまでも 1 対 1 の置き換えであって、奥行きが無い検索に過ぎない。

非英語国としてチェコ、オーストリア、イタリアのシソーラスへの取組状況を視察したが、まだ翻訳の域を出ず、それぞれの国語での文献データベースを別の世界として育

てている感が強い。

将来的にはいずれの言語も自由に交流できて、ことばの壁を意識することなく検索可能になれば、科学的な知識の共有から学問の急速な発展が実現できるものと期待される。

EBM の原点は各国言語の自由な交流である。