

200401047A

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と連携する日本語医学用語シソーラスの実用性  
に関する評価研究

平成16年度 総括・分担研究報告書

主任研究者 脊山 洋右

平成17（2005）年4月

## 目 次

I. 総括・分担研究報告	
UMLSと連携する日本語医学用語シソーラスの実用性に関する評価研究	.....1
脊山洋右、鈴木博道	
II. 分担研究報告	
1. 各種コードとの連携に関する評価研究	.....4
開原成允	
2. MeSH 翻訳管理システムの医学用語シソーラスへの適用 (MTMS : MeSH Translation Maintenance System) に関する評価研究	.....6
野添篤毅、豊玉速人	
3. UMLSとの機械的連携に関する研究	.....8
小野木雄三	
4. 持続的なシソーラス構築及びMeSH翻訳システムへの適応に関する評価研究	.. 10
篠原恒樹	
III. 研究成果の刊行に関する一覧表	
IV. 研究成果の刊行物・別刷	
参考資料	.....

厚生労働科学研究費補助金（医療技術評価総合研究事業）  
総括研究報告書

UMLS と連携する日本語医学用語シソーラスの実用性に関する評価研究

主任研究者 脊山 洋右 お茶の水女子大学  
分担研究者 鈴木 博道 (財) 国際医学情報センター

**研究要旨**

医学文献の検索においても、電子カルテシステムにおいても、E BMにおいても、また様々な医学情報の円滑な流通のためにはシソーラスの存在は欠かすことが出来ないものである。また、わが国で医学情報流通の問題を考える際には、更に日本語と英語との混在についても十分に配慮することが必須である。米国国立医学図書館 (National Library of Medicine--NLM) が主導する国際的なプロジェクトに参画し、UMLS (Unified Medical Language System) と連携しての日本語医学用語シソーラス開発を提唱し、実際に開発を試み、その方法論と成果とについての実用性を評価することが本研究の目的である。平成16年度には、「UMLSとの連携」と言う考え方についてより広範な理解を求め、本研究班の成果などを公開することを目的に、ワークショップを開催した。また、方法論について検証するためには、日本医学会医学用語辞典と SNOMED-CT との間のマッピングを試みた。また、現在、わが国で実際に文献検索で使われている医学中央雑誌刊行会の医学用語シソーラス、並びに Mesh(Medical Subject Headings)日本語訳との間の連携についても、NLM のMTMS (Mesh Translation Maintenance System) 実験に参画する形での実験を行った。

途中経過ではあるが、日本医学会医学用語辞典との間の連携については英語医学用語では見通しが立ち、日本語医学用語に関する一部課題が残っている。SNOMED-CT については技術的な追加処理検討が必要とされている。医学用語シソーラスと Mesh 日本語版に関しては特に問題なく進行した。

**分担研究者**

開原 成允 国際医療福祉大学  
野添 篤毅 愛知淑徳大学図書館情報学科  
小野木 雄三 東京大学大学院医学系研究科  
篠原 恒樹 医学中央雑誌刊行会  
鈴木 博道 (財) 国際医学情報センター

**A. 研究目的**

UMLS (Unified Medical Language System) と連携した日本語シソーラス開発の方法論に関する再評価、並びに、日本医学会医学用語辞典改訂作業との連携による研究成果シソーラス活用策の試行を行い、UMLS と連携したシソーラス開発の実用性を明らかにする。

また、米国 NLM (National Library of Medicine)とのより緊密な相互協力のもとで、シソーラスの拡充策を模索する。

**B. 研究方法**

UMLS のグローバルな展開、フランス・ドイツ・ロシア・イタリア・スペイン・オランダ・フィンランド・ポルトガルなど、他国語への対

応、を踏まえ日本語対応が取り残されることの無いよう、日本語シソーラスも拡充策を講じて行く。

具体的には、SNOMED-CTへのマッピング、日本医学会の医学用語取り込みにより、一層の拡充を行って行く。

同時平行で、NLMとの定期協議を行い、未だ対応付けされていない言語国との連携可能性を追求し、日本医学会の用語との一層のマッピングと改訂作業の連携、米国内でのUMLS活用状況調査、なども実施を心がけた。

**C. 研究結果**

SNOMED-CTとのマッピングを試みた結果から、概念を通じて日本語にマッピングで

きる可能性はあまり高くないことが明らかとなつた。

サンフランシスコで開かれた国際医療情報学会では、わが国におけるUML Sとの連携による日本語シソーラス開発について報告すると同時に、参加していたNLM関係者やUML S共同研究者等との情報交換、今後の方向性に関する意見交換も活発に行われ、更に、シソーラス・メンテナンス・システムなどについて協議した。

本研究班と日本医学会医学用語委員会とは綿密な連携を図ることで合意され、同学会が進めようとしている医学用語辞典の改訂作業でも協力することを予定している。

国内でのUML S連携に関する理解を深めてもらうため、11月24日にはワークショップ「医学用語・シソーラスを巡る国際的展開—UML Sそして日本語によるPubMed検索」を開催した。Mayo Clinicから招いたChute博士の講演も含め、本研究班の途中成果も紹介した。

#### D. 考察

年末に開催したワークショップ以降、本研究班HPへのアクセスも増し、UML Sとグローバルな視点からの連携について、認識はよりなされる様になってきている。更に、一層、普及啓蒙活動も力を入れることによって、結果的に途中成果としてのモデル・シソーラスも使われ、評価されることと期待している。

#### E. 結論

シソーラス開発手法とその途中成果に関する実用性検討は、日本医学会の場でもようやく動きが始めたところである。また、途中成果の

HP公開についても、ワークショップや関連報道などで多少の認知度はあがったものと思われる。

#### F. 健康危険情報

なし

#### G. 研究発表

##### 1. 論文発表

1) 脊山洋右, 開原成允, 野添篤毅, 小野木雄三, 篠原恒樹, 鈴木博道. UMLSと連携した日本語医学用語シソーラスの開発実験. 医療情報学, 24(suppl.), pp1201-2, 2004

##### 2. 学会発表:

1) Yuzo Onogi, Kazuhiko Ohe, Masaaki Tanaka, Atsutake Nozoe, Tetsuro Sasaki, Megumi Sato, Yuko Kikuchi, Tsuneki Shinohara, Hiromichi Suzuki, Shigekoto Kaihara, Yousuke Seyama: Mapping Japanese Medical Terms to UMLS Metathesaurus. in Proceedings of the IIth World Congress on Medical Informatics. Sep.7-11,2004, San Francisco, California, USA. (EDS) M. Fieschi et al. p406-410, 2004

#### H. 知的財産権の出願・登録状況

なし。

#### I. 特許取得

なし。

## 医学用語・シソーラスを巡る国際的展開 —UMLS そして日本語による PubMed 検索—

私は平成13年度以来、厚生労働省医療技術評価総合研究の研究費補助金を受け、日本語医学用語シソーラス開発手法の開発研究とその試行を行っておりまます。当該研究は米国国立医学図書館NLM(National Library of Medicine)との協力によりUMLS(Unified Medical Language System)との連携に基礎をおいております。この度、米国から医学用語やUMLSに造詣の深いMayo ClinicのDr. Chuteをお招きし、米国内での状況をご講演頂き、併せて、本研究班の成果の一端を広くご紹介する機会を持つことを企画しました。

医学用語や医学シソーラス、医学にとらわれることなく広く専門用語管理やシソーラスにご興味有る方々、シソーラスなどを通じて医学医療情報の流通に関わり有る方々など、多くの方々のご参加をお待ちしております。

平成16年度厚生労働科学研究「UMLSと連携した日本語医学用語シソーラスの実用性に関する評価」研究班（主任研究者 お茶の水女子大学教授 脊山洋右）

### プログラム

14:30～14:45	あいさつと研究概要	脊山洋右(お茶の水女子大学)
14:45～15:45	特別講演(英語) "Medical terminology and UMLS—state of the art in USA" Christopher G Chute, M.D., Dr.P.H. Professor of Medical Informatics, Mayo Clinic	
16:00～16:20	「日本語医学用語の UMLS メタシソーラスへのマッピング」 小野木雄三(東京大学)	
16:20～16:40	「MTMS(MeSH Translation Maintenance System)について」 野添篤毅(愛知淑徳大学) 豊玉速人(NPO 医学中央雑誌刊行会)	
16:40～17:00	デモンストレーション 日本語による PubMed 検索 小野木雄三(東京大学)	

### 日時

日時： 2004年11月24日(水) 14:30～17:00

会場： 東京大学鉄門記念講堂（東京大学医学部新研究棟14階）

参加費： 無料

共催： 特定非営利活動法人 医学中央雑誌刊行会

財団法人 国際医学情報センター

### 参加申し込み方法

参加ご希望の方は、所属機関名・会社名、所属部署・役職、氏名(ふりがな)、電話番号 E-mail、をご記入の上 e-mail または faxにてお申し込み下さい。特に受講票は発行致しませんので、当日、そのまま会場にお越し下さい。

### - お問い合わせ先 -

(財)国際医学情報センター

事業推進室 UMLS ワークショップ担当

〒160-0016 東京都新宿区信濃町35

信濃町煉瓦館3階

TEL 03-5361-7089 FAX 03-5361-7110

E-mail [jigyo@imic.or.jp](mailto:jigyo@imic.or.jp)

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 と連携する日本語医学用語シソーラスの実用性に関する評価研究  
各種コードとの連携

分担研究者 関原 成允 国際医療福祉大学副学長

**研究要旨**

現在、日本における最も権威のある医学用語集としては、日本医学会の編纂した医学用語辞典があるが、この辞典には二つの大きな欠点があった。第一は、この医学用語辞典に採用されている英語が必ずしも英語圏での標準的な英語医学用語ではないこと、第二は、シソーラスの概念がないために、基幹となる用語も同義語も同列に扱われ、利用者にどの用語が基幹的な用語であるかがわからないことであった。

本研究では、この欠点を改善するために、UMLS の Concept Unique Identifier (CUI) を用いて、医学用語辞典の英語と UMLS の連携を図ると共に、日本の用語辞典にもシソーラスの概念を定着させることを目的とした。

研究方法としては、医学用語辞典の英語と UMLS とを正規化した上でマッチングをとったが、約 8 万語の中の 40% でマッチングがとれた。とれないものについては、今後個別の検討が必要であることが明らかになったが、この方法で目的が達成されるであろうことは、ほぼ確実となった。

キーワード：医学用語、UMLS、シソーラス、標準病名集

**A. 研究目的**

現在、日本における最も権威のある医学用語集としては、日本医学会の編纂した医学用語辞典がある。この辞典は、英語の医学用語を日本語に変換する際の標準的な日本語医学用語として使われてきたが、二つの大きな欠点があった。第一は、この医学用語辞典に採用されている英語が必ずしも英語圏での標準的な英語医学用語ではないこと、第二は、シソーラスの概念がないために、基幹となる用語も同義語も同列に扱われ、利用者にどの用語が基幹的な用語であるかがわからないことであった。

本研究では、この欠点を改善するために、UMLS の Concept Unique Identifier (CUI) を用いて、医学用語辞典の英語と UMLS の連携を図ると共に、英語のシソーラスを医学用語辞典に埋め込むことを研究する。また、第二は、対応する日本語については、医学会各分科会の用語辞典を参照しつつ、基幹となる日本語と同義語を区別することを研究する。

この方法が確立した後は、これを実際の医学用語辞典の改定に役立てることが本研究の目的

である。

**B. 研究方法**

日本医学会医学用語辞典・英和の電子版を用い、その英語を正規化し、UMLS の CUI の英語との間でマッチングをとる。マッチ（対応がとれた）した英語は、それが基幹的英語か同義語かを区別して医学用語辞典の英語に対応するコード番号を付す。これによって、まず、医学用語辞典の中の国際的にみて標準的な英語を区別する。

次に、マッチしなかった英語について、それが重要な基幹語であるか、または削除してもいいものかを判定し、現在の英語単語を整理する。

次にマッチした英語に対応する日本語について、1 対 1 の対応であれば、それを日本語の基幹語として採用する。もし、複数の日本語が対応している場合には、医学会各分科会の用語辞典を参照しつつその中の基幹語と同義語を判別し、その印を付す。また、病名については、厚生労働省と医療情報システム開発センターが開発したいわゆる「標準病名集」があり、広く

普及しはじめている。この病名集は日本医学会の監修を経たものであり、基幹語と同義語が整理されている。従って、病名についてはこの病名集との間の整合性をとることによって、日本医学会用語集を整理することを試みる。

### C. 研究結果

現在、日本医学会医学用語辞典には、83791の英語が収載されている。単純に対応をとると、この中の約40%がUMLSとの対応をとることができた。

しかし、対応がとれたものの中にも、UMLSのさまざまなレベルでの対応関係があるので、それを個別に精査することが必要である。これにより、英語の部分の整理が可能であることは確かであった。

一方の対応つかなかった英単語については、さまざまな原因が考えられた。第一は、UMLSは名詞のみであるが、日本医学会用語辞典は形容詞をも含むからこれは対応がとれない。また、単純なスペルミスがまだ残っているものもあった。また、薬剤やアイソトープなど表記上の問題が原因のものもあった。これらを今後一つ一つ検討していくことが必要であることが判明した。

複数の日本語訳を持つものの整理については、本年度は検討できず、来年度の検討課題である。

標準病名集との間の整合性についての検討も本年はまだ実現していないが、漢字の略字体の問題、カナ表記の必要な用語にカタカナをもちいるか、ひらがなを用いるかなど、多くの解決しなければならない問題が残されている。

### D. 考察

これまで、多くの医学用語集が作られたきたが、通常はさまざまな場所で使われている英語を無差別に収集して出発点とした用語集も多く、そこに収載されている英語が国際的に標準的な英単語であるか否かについてはほとんど検討されたことはなかった。今回UMLSを一つの基準

として、収載する英単語を整理するという試みはこれまでにないものである。

また、これまでの医学用語集にはシソーラスの概念がまったく存在せず、これも大きな混乱の原因となっていた。今回の研究によって日本にもシソーラスの概念が定着すれば非常に大きな意義がある。

### E. 結論

本年は、第一年目として、日本医学会医学用語辞典の見出し英単語とUMLSのCUIの間の対応を研究し、この対応をとることによって、日本語の辞書の英単語を国際標準の観点から整理できることが明らかになった。

また、これを用いることにより、医学用語辞典にシソーラスの概念を導入することが可能になることも明らかになった。

しかし、英単語の整理は可能となつても、日本語医学用語の整理は、UMLSの利用によって行うことはできないから、この部分は日本の医学会の総力をあげて行う他はない。この部分の方法論については、本年度はまだ未検討であった。

### G. 研究発表

なし。

### H. 知的財産権の出願・登録状況

なし。

### I. 特許取得

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 と連携する日本語医学用語シソーラスの実用性に関する評価研究  
MeSH 翻訳管理システムの医学用語シソーラスへの適用  
(MTMS : MeSH Translation Maintenance System)

分担研究者 野添篤毅 愛知淑徳大学文学部図書館情報学科  
研究協力者 豊玉速人 特定非営利活動法人 医学中央雑誌刊行会

**研究要旨**

MTMS を使用した MeSH 翻訳語の提供に関する取り組みと経緯、及び今後の展開を検討した。

**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いる。2003 年までに、フランス、ドイツ、イタリア、オランダ、フィンランド、ポルトガル、ロシア、スペインの 8 カ国語に渡り、Concept レベルでの体系化ができており、スウェーデン、ポーランドと言った東欧ばかりでなく、ベトナムを始めとするアジア諸国、漢字文化の日本、韓国の対応も進行中である旨が発表されており、多言語対応のシソーラスとしても、幅広く展開されてきた。NLM は UMLS の多言語対応をより良く、取り込めるシステムとして MeSH Translation Maintenance System(以下、MTMS)を開発した。今回 NLM より NPO 医学中央雑誌刊行会に日本語医学用語の取り込みに限って提供された MTMS を使用して、日本語医学用語をいかに提供し、メンテナンスを行っていくかを試みた。

**B. 研究方法**

英語圏以外の国々でも Web による情報検索サービス、PubMed が幅広く利用されていることは言うまでもない。だが、自国の言語で PubMed 等の検索に利用できる医学用語辞書、あるいはシソーラスと言ったものを作成していくところは、先にも述べたとおり、僅か 8 ケ国

にすぎない。日本以外のアジア諸国では「医学教育は英語で行っている」というのが通常であることから、あまり、必要性がないという現状もあるかもしれない。東欧のチェコなども英語圏ではないが、MeSH の翻訳に取り組んでいるが、UMLS から見れば数パーセントの対応である。これは先にも述べた 8 ケ国にも言えるであるが、対応が完全に取れているとは言えないのが現状である。ところが、医学中央雑誌刊行会が提供している「医学用語シソーラス」は 99% の用語が MeSH とリンクが取れており、検索ばかりでなく、論文等で使用されている英語の医学用語も幅広く取り入れられており、なおかつ MeSH と Concept レベルでのリンクがあることは UMLS へ日本語医学用語の提供をある程度容易にすることことができた。

医学中央雑誌刊行会発行の「医学用語シソーラス」から、MeSH に対応する日本語の用語抽出し、MTMS にてデータベースを構築してもらい、登録された用語に対して、各々、更新、修正していく。それらの用語が NLM の Supervisor の手によって承認され、UMLS へ登録される。この手順を必ず踏むことで UMLS への日本語医学用語が登録されていく。

図 3 に示した画面から、用語を入力し、検索し、その用語に関してそれぞれ更新等の作業を行う。

図 1 UMLS のトップ画面

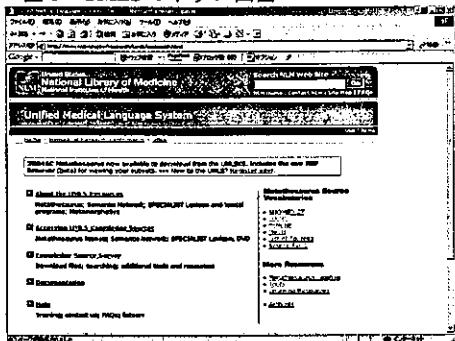


図 2 トップ画面

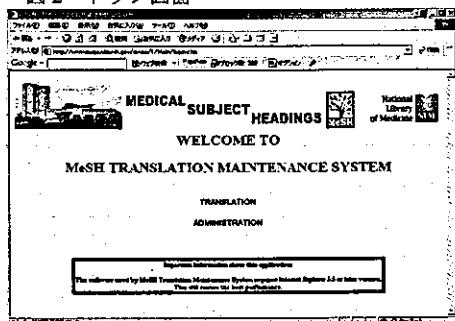
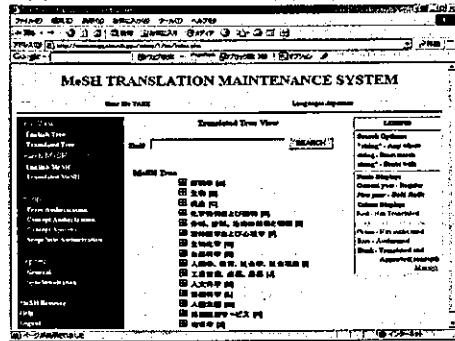


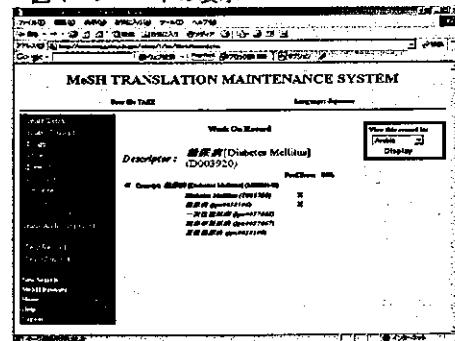
図 3 Translated MeSH のトップ画面



検索された用語は Concept ごとに Entry Term、Entry Term(Non Print)等に分けて表示される。

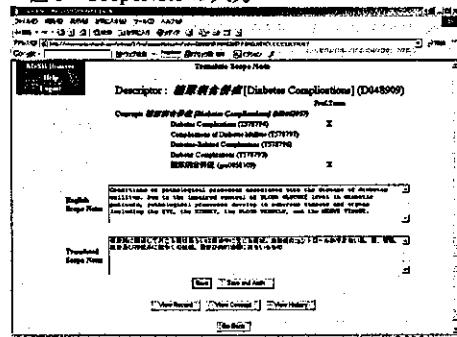
(図 4) 承認などの情報はフォント及び色によって表現される。

図 4 レコードの表示



新規登録の場合は対象の MeSH 用語を検索してレコードを呼び出し、必要項目を入力していく。当然のことながら日本語入力(自国語入力)が可能である。(図 5) それぞれのメニューがあり、呼び出された画面に対して更新、修正等を行っていく。また、登録された内容をカテゴリー構造で表示することも可能である。

図 5 ScopeNote の入力



### C. 研究結果

MTMS による用語のメンテナンスに関しては十分に行える。今回、約 2 万語の用語を登録し、MeSH にて変更のあったものに関しては更新作業を行ったが、問題なく実行することができた。

### D. 考察

今回の作業は UMLS の中でも MeSH に関してのみ、行うことが可能であることは他の言語に対しても、対応することが十分に可能であることが示唆された。

### E. 結論

今回の作業が他の言語への対応するテストケースとしては十分に適応でき得ることがわかった。この成果は情報検索の支援、索引作業の支援ばかりでなく、各種、英語圏以外のデータベースへのリンクも十分に可能であると思われる。

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 と連携する日本語医学用語シソーラスの実用性に関する評価研究  
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 クリニカル  
バイオインフォマティクス研究ユニット・臨床情報工学

#### A. 研究目的

医学中央雑誌刊行会による医学用語シソーラスと MeSH との対応に関し、昨年度までに概念間対応の確認作業を完了した。本年は UMLS に含まれる世界最大の登録語数を誇る SNOMED-CT と日本語医学用語とのマッピングを取るための手法および問題点を調査した。

#### B. 研究方法

UMLS の中で MeSH 概念と SNOMED-CT 概念との対応付けを利用して、機械的な意味対応を取る際の問題点を調査した。材料として UMLS2004AB(2004 年の第 2 版)を使用し、その中に含まれる MeSH および SNOMED-CT を利用した。

#### C. 研究結果

MeSH の概念数が約 2 万件であるのに比べて、SNOMED-CT の概念数は約 30 万件である。これらの間には明らかに多対 1 の関係があり、概念としての粒度の違いが存在する。実際、機械的に MeSH に対応した日本語をそのまま SNOMED-CT に当てはめることにより、約 12000 件の SNOMED-CT 概念が日本語で記述された。すなわち MeSH と SNOMED-CT とが共通しない部分が 8000 件も存在するということである。粒度の違いだけではなく、共通する概念が 60% しか存在しないことも明らかになった。

#### D. 考察

UMLS は世界各国の医療に関する多数の統制用語を集積し、共通概念に関して同一の概念

コードを振ったメタシソーラスであるが、統制用語は使用目的に応じて異なるものが歴史的に発展してきていることを反映して、UMLS の共通概念であっても個々の統制用語集内の意味に違いがあり、特に概念粒度に大きな隔たりが存在する。文献検索の標準語彙は MeSH であり、臨床の語彙は SNOMED-CT であるが、これらを相互に交換可能とするためには、機械的一致に加えた何らかの補完が必要である。

例えば、粒度の荒い MeSH から SNOMED-CT への投射には条件を付加し、逆に粒度の細かい SNOMED-CT の概念は、MeSH 概念の組み合わせで記述できる可能性がある。これは同時に SNOMED-CT 概念を日本語に対応させる方法でもある。今後はこうした概念の分解により日本語との機械的対応を拡大させることが課題である。

#### E. 結論

SNOMED-CT の概念を、UMLS を利用して機械的に日本語で記述する試みを行った。対応する概念は 30 万件中 1 万 2000 に過ぎず、今後は概念要素の分解などの手法によって日本語との対応を行う必要があることが明らかとなった。

#### G. 研究発表

1. 論文発表：
  - 1) 小野木雄三, UMLS を利用した日本語医学知識サポートシステム. 医療情報学, 24(suppl.), pp386-387, 2004
  - 2) 脊山洋右, 開原成允, 野添篤毅, 小野木雄三, 篠原恒樹, 鈴木博道. UMLS と連携

した日本語医学用語シソーラスの開発実  
験. 医療情報学, 24(suppl.), pp1201-2, 2004

2. 学会発表 :

- 1) Yuzo Onogi, Kazuhiko Ohe, Masaaki Tanaka, Atsutake Nozoe, Tetsuro Sasaki, Megumi Sato, Yuko Kikuchi, Tsuneki Shinohara, Hiromichi Suzuki, Shigekoto Kaihara, Yousuke Seyama: Mapping Japanese Medical Terms to UMLS Metathesaurus. in Proceedings of the 11th World Congress on Medical Informatics. Sep.7-11,2004, San Francisco, California, USA. (EDS) M. Fieschi et al. p406-410, 2004

H. 知的財産の出願・登録状況

(予定を含む)

1. 特許取得 : なし
2. 実用新案登録 : なし
3. その他 : 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 と連携する日本語医学用語シソーラスの実用性に関する評価研究  
持続的なシソーラス構築及び MeSH 翻訳システムへの適応

分担研究者 篠原恒樹 特定非営利活動法人

**研究要旨**

UMLS と連携した医学用語シソーラス作成から、UMLS への日本語医学用語の持続的な提供と今後の展開について検討した。

**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)はフランス、ドイツ、イタリア、オランダ、フィンランド、ポルトガル、ロシア、スペインの 8 カ国に渡る言語に対応している(2003 年)。2004 年以降も多言語に対応していく旨が既に発表されている。その中にはベトナムを始めとするアジア諸国、漢字文化の日本、韓国への対応も計画に入っていることが含まれている。この計画は UMLS が多言語対応のシソーラスとして、幅広く、全世界で利用されていることを示している。英語圏以外の国々でも Web による医学文献検索サービス、PubMed、MEDLINE が NLM から提供されている Medical Subject of Headings(以下、MeSH)、シソーラス(UMLS に含まれる)を介して幅広く利用されていることからも、自国の言語で医学文献の検索が可能になることは医学研究の発展に大いに寄与するばかりでなく、他の言語のデータベース(医学関連分野)とのリンクが UMLS を介して可能になることを示す。その為には如何に持続的に日本語医学用語を提供できる環境を構築するかが重要課題になることが考えられる。私ども医学中央雑誌刊行会は Web による医学文献検索を提供していく上で経験的にこれらの諸問題を熟知している。医学文献検索に用いられる医学用語は多岐にわたり、変化し、様々な形態をとる。これらの問題に取り組むことは医学関

連分野の用語による障壁を低くするばかりでなく取り除くことも可能であると思われる為、持続的な日本語医学用語の提供形態を検討した。

**B. 研究方法**

医学用語集(辞典含む)として、日本国内で使用されている医学用語と UMLS のマッチングを行い、マッチングした用語の Supervising、マッチしない用語をどのように処理すれば持続的な提供形態をとることが可能であるかを検討した。

**C. 研究結果**

日本の代表的な辞書と提供されている日本医学会用語集(第 2 版)には 78,316 語が収録されている。この用語集に含まれている約 40%が UMLS の用語とマッチした。うち、MeSH の用語とマッチしているものは 25%。残りの 60%はなんらかの理由でマッチしていない。医学用語集は辞書である為、名詞以外の形容詞等も含まれている為、このような数字になる(約 25%は形容詞等と思われる)。これらは他の用語集あるいは英語以外の MeSH 等からあいまい検索等の手法をとることによってある程度は解決できるかもしれないが、検索に使用される、されないは関係なく典拠されていることから、これらの用語類を他の手法で UMLS に提供できる形態を構築することは一時的には可能であっても、持続的に提供することは難しいと考えられる。その理由として、UMLS は Concept ごとに言語の違い、同義語類等、それぞれにコードが付与さ

れ、分類されている。PubMed 等で提供されている MeSH もディスクリプターのもとに同義語があつめられて構成されているが、UMLS の中では Concept で分類されている。この Concept レベルでのリンクが取れていれば、日本語医学用語がそれぞれの Concept に分類される。このことに着目し、他の医学用語集やシソーラスを検証してみたが、日本医学会用語集と同じ問題が常に立ちはだかる為、持続的な提供は容易ではない。その中、NPO 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は現在、版を重ね、第 5 版が発行されているが、MeSH とのリンクが 99%、略全ての用語についてとれている。一部、とれていらないのは米国特有の保健機構等であり、疾病、医薬品名等は MeSH のカテゴリコードとも一致している。一部、MeSH の Concept(Descriptor とは異なる)に振り分けなければならないものもあったが、このことは MeSH、UMLS の改定に準じてメンテナンスを行うことで十分に解決できる。

#### D. 考察

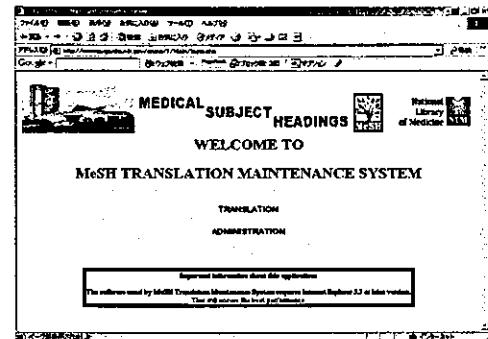
日本国内で提供されている医学用語集の中から持続的に日本語医学用語を提供できる形態を構築することができるか否かは UMLS と Concept レベルでのリンクがいかに取れているかであることがわかった。新規に用語が追加される際もこの点を踏まえて作成すれば良く、シソーラス等の用語集を作成する際にも同じことが言える。辞書、辞典の類はシソーラスという検索に使用される用語の集合とは異なるため、全ての用語に関して Concept レベルでのリンクをとることを前提としているが、用途が違うので一時的には可能であっても、持続的にはシソーラスのような形態をもちつつ、辞書、辞典的な要素を残すと言う、ある意味、矛盾した要素を盛り込まねばならなくなり、持続的な提供は困難とも言える。

また、現在 NLM は MeSH の翻訳語登録及びメンテナンスシステムとして MeSH Maintenance Translation System(以下、MTMS)が

提供されている。(図 1)

MTMS に登録されたデータは用語毎に更新、修正が可能である。これらの用語は NLM の Supervisor の手によって承認され、UMLS へ登録されていく。現在、NPO 医学中央雑誌刊行会では、日本語の医学用語の提供を行っている。

図 1:MTMS、トップ画面



#### E. 結論

今回の作業が日本語ばかりでなく他の言語とのリンクへのテストケースとして十分になり得、持続的な提供形態をとることが可能であることを示すことができた。

# Mapping Japanese Medical Terms to UMLS Metathesaurus

Yuzo Onogi<sup>a</sup>, Kazuhiko Ohe<sup>b</sup>, Masaaki Tanaka<sup>c</sup>, Atsutake Nozoe<sup>d</sup>, Tetsuro Sasaki<sup>e</sup>, Megumi Sato<sup>e</sup>, Yuko Kikuchi<sup>f</sup>, Tsuneki Shinohara<sup>g</sup>, Hiromichi Suzuki<sup>h</sup>, Shigekoto Kaihara<sup>e</sup>, Yousuke Seyama<sup>i</sup>,

<sup>a</sup> Clinical Bioinformatics Research Unit, Graduate School of Medicine, the University of Tokyo, Japan

<sup>b</sup> Department of Medical Informatics, Graduate School of Medicine, the University of Tokyo, Japan

<sup>c</sup> Kawasaki University of Medical Welfare, Japan

<sup>d</sup> School of Library and Information Science, Aichi Shukutoku University, Japan

<sup>e</sup> The Medical Information System Development Center, Japan

<sup>f</sup> Department of Health Service Management, International University of Health and Welfare, Japan

<sup>g</sup> Japan Medical Abstracts Society, Japan

<sup>h</sup> International Medical Information Center, Japan

<sup>i</sup> Faculty of Human Life and Environmental Sciences, Ochanomizu University, Japan

## Abstract

This paper introduces and reports the results for a project to map Japanese medical terms to the UMLS Metathesaurus. The "Thesaurus for Medical and Health related Terms version 5" published in 2003 by the Japan Medical Abstracts Society and UMLS version 2002AC provided by NLM were used in this study. The goal was to judge the validity of the correlation between the Japanese and English terms that belong to the same MeSH concept. Fifteen medicine, nursing, and library science professionals, excluding JAMAS, used a custom designed Web interface to perform this task. About 10% of the concepts were judged as invalid, and the reasoning behind these failures were analyzed. Experience from this project can be used to estimate the manpower required to revise the Japanese thesaurus after future revisions to UMLS or MeSH.

## Keywords

UMLS, MeSH, Japanese Medical Terms

## Introduction

In order to accurately exchange information, it is imperative that concepts and terms have universal meanings. This is the motivation behind the development and maintenance of standard health terminologies and controlled vocabularies. The Unified Medical Language System (UMLS) project was initiated in 1986 by the U.S. National Library of Medicine to integrate these controlled vocabularies [1]. The UMLS Knowledge Sources, which now provides valuable information about biomedical concepts, incorporates tools to accomplish searches on the root terms used to name the concepts [2,3].

However, information exchange is much more difficult in a multilingual environment [4,5]. Although the UMLS currently contributes to cross-language retrieval by supporting terms in

15 languages, Japanese vocabulary is excluded. There is a great need for Japanese professionals to enter queries in Japanese and retrieve documents, even if they are written in English. Currently, professionals typically use MEDLINE via Ovid technologies, Inc. or PubMed services provided by NLM. A set of Japanese biomedical terms that correspond to or closely relate to MeSH or UMLS concepts is needed in order to search articles in Medline using Japanese keywords.

Only a few controlled vocabularies have been developed and maintained in Japan. The primary interest of the Japanese medical society has instead been focused on nomenclature alone, without semantic structure, in order to coordinate with the English representation. This might have originated from difficulty inputting kanji characters using the keyboard, which delayed the establishment of health information databases. And Japanese terms are written using more than 2000 Kanji characters, 83 Hiragana characters, and 96 Katakana characters. Most terms of foreign origin are written using Katakana, which is a phonetic alphabet. However, when there are variations in pronunciation for a foreign term, a number of Katakana variations exist for a single term. This characteristic is prominent in medical terms that originate from personal names. Hiragana is also phonetic, however it is primarily used to detail the pronunciation of the kanji character(s), when they appear in nouns. This feature of Japanese writing may be another reason why controlled vocabularies for Japanese medical and health-related terms have lagged.

In the information age, it is increasingly important for medical information specialists to collect, accumulate, and analyze health related information written in Japanese. This advancement is particularly important for electronic health records. This application has led to increasing research interest in developing controlled vocabularies in Japanese [6-8]. Now we do have a set of standard Japanese disease names

corresponding to ICD10, but there are no vocabularies in regard to Japanese clinical terms.

In 2003, the Japan Medical Abstracts Society published the "Thesaurus for Medical and Health Related Terms, version 5." This Japanese thesaurus corresponded to MeSH, but were not checked in public, might not correspond to MeSH correctly, nor cover enough clinical terms used in practice.

Therefore, a team of health-related professionals checked the validity of the Japanese terms in this thesaurus to determine if they properly correspond to their English counterparts based on both the indexers' and the health professionals' points of view. Terms that were judged poorly were discussed and either removed or replaced by other more appropriate Japanese medical terms. The purpose of this article is to describe the validation project and report its initial results.

## Materials and Methods

### Materials

The following resources were used to research the health-related terms:

1. UMLS version 2002AC, provided by NLM.
2. The "Thesaurus for Medical and Health Related Terms version 5," referred to as the TMH, was used as a collection of Japanese medical terms. The TMH was published in 2003 by the Japan Medical Abstracts Society (JAMAS).
3. The Database and WWW server for this system used Apache version 1.3.27 with php 4.2.3 module and PostgreSQL 7.2.3 DBMS, running on Red Hat Linux 7.3.

Concepts and strings that were related to MeSH were used from UMLS. There were 20,603 concepts (descriptors) and 37,864 MeSH entries used. The concepts numbered approximately half of the MeSH tree entries because some concepts have multiple entries. For example, the concept "Diabetes Mellitus" has two entries: MeSH tree numbers C19.246 and C18.452.297, and "WAGR Syndrome" has 18 MeSH numbers. However, many concepts only correspond to one MeSH tree number.

The TMH thesaurus contains 84,617 Japanese biomedical terms, which are categorized in 19,179 MeSH concepts, based on the MeSH 2001 version. Japanese indexers from the Japan Medical Abstracts Society extracted these terms from Japanese articles that were published in Japan. The indexers have also extracted additional terms that relate to each concept, resulting in a total of 95,071 terms.

### Methods

Phase one involves checking the correlation between the Japanese terms and the UMLS strings for every UMLS concept within MeSH. More specifically, a web interface was used to repeatedly judge the correlation between a set of

Japanese terms and a set of English strings for the same MeSH (UMLS) concept. During this process, concepts that had been changed from 2001 to 2002 were excluded. The MeSH terms that appeared more than once in the MeSH tree were judged multiple times. If there were multiple judgments for one set of Japanese terms or UMLS concept, the final result will be NO if there exists only one NO judgment.

The goal of Phase two was to correct the sets that were judged as NO. On a case by case basis, members discussed the reasons for the mismatch, and members made a final decision by voting via web interface.

### Web interface for Phase one judgment

The WWW was used as the medium to verify whether the Japanese terms properly corresponded to the UMLS concepts. Fifteen people with a background in Japanese health related terms participated in this project. Each member individually selected their preferred areas from the top MeSH nodes and checked the validity of the Japanese terms that fell under the chosen nodes.

Figure 1 shows the judging interface. There are two areas in the window. The primary Japanese term and its synonyms are shown on the left side, and corresponding UMLS concept, its definition, and synonyms are shown on the right side. In the upper left region, the top left area is the judging interface. This interface includes a text area for adding comments and OK and NO buttons. Below this, Japanese synonyms and their sources and a table detailing Japanese reference terms that relate to this concept and their sources are displayed. In the right area, the UMLS concept definition, the list of English strings for the concept, and a list of all MeSH numbers for the concept are shown.

Each member judged whether the Japanese terms listed in the synonym table corresponded to the English terms and definitions listed on the right. The member then clicked either the OK or NO button, and the reasons for the decision were detailed in the text box. This process is continued until all of the Japanese terms assigned to a member were completed.

Figure 2 shows an interface in which members can explore all of the MeSH numbers, even those assigned to other members. In this figure, the left window shows only a list of Japanese primary term due to space limitation. But in reality, a list of MeSH numbers and corresponding English primary strings are shown at the left side. When a user clicks on a MeSH number, the information relating to the node appears in the middle and the right windows. The contents of these windows are the same as those in Figure 1.

Haversian System = 骨単位 (コツタノイ)  
(A10.165.265.507, C0824691, S0046764, 02606)

この日本語で正しく対応していると判断される場合はOKを、  
そうでない場合は理由を下に記入の上、NOのボタンを押してください。

OK (正しい)  
 NO (正しくない)

統制語: Havers系 / ハバース系

辞書	code	典拠
骨単位 (コツタノイ)	0.000	1) MeSH 1997 医学件名標目表 英和版 2) 解剖学辞典
オステオン (オステオノン)	1.001	1) 解剖学辞典

以下参考情報

辞書	code	典拠
Havers系 (Haversケイ)	2.001	1) MeSH 1997 医学件名標目表 英和版

CUI-Source and Definition

C0824691-Mesh2002\_06\_01

A circular structural unit of bone tissue. It consists of a central hole, the Haversian canal through which blood vessels run, surrounded by concentric rings, called lamellae.

STR in the same CUI	TS	STT
Osteon	P	PF
osteon	P	VC
Osteon, NOS	P	VO
Osteons	P	VP
Haversian System	S	PF
Osteonum	S	PF
haversian system	S	VC
Haversian system	S	VC
System, Haversian	S	VW

MeSH entries for same CUI

TRACE
4: Haversian System
3: Bone and Bones
2: Connective Tissue
1: Tissues

Fig 1 - User interface for verification of correspondence between a set of terms within a Japanese medical concept and English terms in the same UMLS concept. On the left, Japanese medical terms and their synonyms and sources are displayed. On the right, the corresponding English terms for the same UMLS concept are displayed. The judging process is completed by pushing the "OK" or "NO" button in the upper left corner. When judged as "NO," the reasons for the judgment are required in the text box.

汗  
初乳  
胃液  
胃酸  
腸分泌物  
母乳  
粘液  
子宮頸管粘液  
肺液  
唾液  
皮脂  
精液  
恥垢  
喀痰  
汗  
涙液  
毒液  
胎脂  
体液  
眼涙水  
腹水  
血液  
膽汁  
血漿  
体液区画  
体内水分  
気管支肺胞洗浄液  
體液  
細胞外液  
血管外肺水分

**CUI-Source and Definition**  
**C0015352-Mesh2002\_06\_01**  
Interstitial space between cells, occupied by fluid as well as amorphous and fibrous substances.

STR in the same CUI	TS	STT
Extracellular Space	P	PF
extracellular space	P	VC
Extracellular space	P	VC
Extracellular space, NOS	P	VO
Extracellular Spaces	P	VP
Space, Extracellular	P	VW
Spaces, Extracellular	P	VWP
Intercellular Space	S	PF
Intercellular space, NOS	S	VO
Intercellular Spaces	S	VP
Space, Intercellular	S	VW
Spaces, Intercellular	S	VWP

**MeSH entries for same CUI**

TRACE
3: Extracellular Space
2: Cellular Structures
1: Cells

TRACE
3: Extracellular Space
2: Body Fluids
1: Fluids and Secretions

以下参考情報

辞書	code	典拠
細胞外液 (サイボウガイエキ)	0.000	1) 医学用語シーラス 第3版 2) MeSH 1997 医学件名標目表 英和版
間質液 (カンシヅエキ)	1.001	1) MeSH 1997 医学件名標目表 英和版

**Extracellular Space**  
**(A12.207.395,C0015352,S0039883, 08855)**

この項目はチェック済です by yseyama.  
不適  
→細胞外腔、細胞外スペースの方が適当(yseyama)

判定をやり直す場合は、下に再度入力してください

この日本語で正しく対応していると判断する場合はOKを、  
そうでない場合は理由を下に記入の上、NOのボタンを押してください。

細胞外腔、細胞外スペースの方が適当  
(yseyama)

OK (正しい)  
 NO (正しくない)

統制語: 細胞外空間

辞書	code	典拠
細胞外液 (サイボウガイエキ)	0.000	1) 医学用語シーラス 第3版 2) MeSH 1997 医学件名標目表 英和版
間質液 (カンシヅエキ)	1.001	1) MeSH 1997 医学件名標目表 英和版

以下参考情報

辞書	code	典拠
----	------	----

Fig 2 - The left window shows the list of MeSH numbers with their corresponding English primary string and Japanese primary term (here shows only Japanese primary term due to space limitation). When the user selects a MeSH number, the relevant information appears in the middle and right windows. These contents are the same as those in Figure 1.

Keratinocytes 角化細胞/ケラチン細胞/表皮細胞 A11.436.397	06746	NO	表皮細胞(は上位語 (yseyama)	角化細胞/ ケラチノサイト/ ケラチン細胞/ 表皮細胞/ Keratinocyte/ Keratinocytes/ 角質細胞/
Nucleolus Organizer Region NOR/核小体形成体部/核小体形成域 A11.284.170.279.190.160.650 Q05.275.658	06790	NO	AgNOR(は不適当 (yseyama)	核小体形成域/ AgNOR/ 核小体形成体/ 核仁形成体域/ NOR (Nucleolus Organizer Region)/ Nucleolar Organizer/ Nucleolus Organizer Region/ 核小体形成体部/ AgNORs/ Nucleolar Organizer Regions/ 核小体形成領域/ 仁形成体部/
Dilatation, Pathologic 病的拡張症/異常拡張症 C23.300.325	06802	NO	拡張症-病理的(hliu)	病的拡張症/ 拡張症-病的/ Dilatation, Pathologic/ Ectasia/ Pathologic Dilatation/ Pathologic Dilation/ 異常拡張症/

Fig 3 - Members are able to browse comments issued by other members in real time. The first column shows a MeSH descriptor and the corresponding Japanese term, the 3rd column shows current judgment, the 4th column shows the member's comments, and the 5th column shows corresponding Japanese terms for the concept.

#### Web interface for browsing comments given

Figure 3 shows the web interface used to browse the member's comments, which can be examined in real-time. Comments are displayed in the 4th column with the member's name in parentheses. The results of decisions are shown in the 3rd column. Some concepts that were judged as OK were submitted with comments. These concepts will be treated the same as those judged NO in the correction phase.

#### Web interface for discussions and voting – Phase two judgment

Members are able to submit further opinions for concepts judged as NO or as OK that already contain comments. This additional interface is similar to Figure 1. After a thorough discussion, the changes are proposed. For example, "The word A should be changed to word B in the Japanese synonyms for concept C" will be proposed in Japanese. The final decision is then made by a vote.

#### Results

Phase one judgment was started in January 2003 and was finished in September. Phase two checking is currently in progress. This section presents the results thus far.

#### Analysis of Phase one judgment

1503 concepts (7.3%) were judged as NO and 1865 (9.1%) were submitted with comments. As previously stated, concepts that included comments were treated the same as those judged as NO., and then we analyzed whole cases being issued any comments. The comments were categorized into the following six groups:

1. A concept contains English terms, and sometimes no Japanese term. (44.0%)

2. There is no Japanese class or group name that corresponds to the plural English form. (11.8%)
3. Other Japanese terms are more suitable and should be included. (33.6%)
4. Some Japanese terms are not appropriate and should be excluded. (5.5%)
5. Some terms are questionable and need further discussion. (2.1%)
6. No change is needed. (2.8%)

Category 1 means that some English terms appear directly in Japanese articles. Category 2 is a feature of Japanese language that there is no explicit plural form of the Japanese noun. Categories 3 through 5 were the intended target of the Phase two process, which number 770 (41.3%). This process, which involves discussion and voting, will be completed within a month.

#### Estimation of workload for revision of Japanese thesaurus

Fifteen members of professionals relating to Japanese biomedical vocabulary participated in the Phase one work, requiring a total of 124 days. Each member worked approximately a week on the project. The members averaged 167 concepts a day, while one member was able to check a maximum 1920 concepts in only one day.

847 descriptors were added and 47 descriptors were deleted between the 2001 and 2002 MeSH versions, resulting in a total of 894 concepts that needed to be updated in the Japanese thesaurus. From 2002 to 2003, 1250 concepts were added and 20 concepts were deleted, necessitating 1270 new updates.

From these previous experiences, it is estimated that six man-days are required to check 1000 concepts during the Phase one process.

## Discussions

### Information retrieval via Japanese medical terms

When the Phase two process are completed, professionals will be able to use a Japanese medical thesaurus that corresponds to the UMLS Metathesaurus. This means that Japanese keywords can be used to search for articles in MEDLINE. This will be useful not only for health professionals, but also the Japanese population at large.

However, current UMLS Metathesaurus consists of seven bit ASCII characters. It should be noted that UMLS should have any multi byte character sets, such as UNICODE.

### Comparison by concept

In this project, Japanese terms and corresponding English terms were compared for a concept in the UMLS. However, when the UMLS is revised, some of the English terms in a concept will be changed and others will not. When this occurs, a considerable amount of work must be performed on the corresponding set of Japanese terms, because all of the Japanese terms within the concept will need to be checked. Therefore, just as the UMLS introduced the CUI-LUI-SUI scheme, the Japanese terms must be managed efficiently while updating the identification system. An option is to introduce an LUI for Japanese terms that correspond to the English strings that have the same LUI and the same mental meaning in UMLS. Table 1 shows one such example. This table displays the relationship between the English and Japanese terms for the concept "Eye". A kanji character in upper two rows and the former character in lower three means "eye". The latter kanji character in lower three rows means "ball". As is evident from this example, introducing an LUI on Japanese side may dramatically reduce the workload during the UMLS concept updates.

Table 1 –Sample relationships between English and Japanese

UMLS LUI	English	Japanese
L0015392	Eye, eyes	眼
L0876007	Oculus	眼
L0015418	Eyeball	眼球
L1811070	Bulbus oculi	眼球
L0847402	Globe	眼球

### A working style for collaborating on the Web

This project required the development of web interfaces to view the relationships between English and Japanese terms within a UMLS concept, make judgments and submit

comments, discuss the issues, and vote on a decision. These interfaces were extremely useful and efficient for the participating members.

Some members, however, reported that it was difficult to understand the judgment policy and the remarks form when they first started. During this study, the judgment policy was added to the web site. Periodic meetings were necessary even though the judgment process was conducted remotely over the Web.

### Acknowledgments

This research entitled "Development of thesaurus on Japanese medical terms corresponding to UMLS Metathesaurus" (13070101) was supported by research funds from Japan's Ministry of Health, Labor and Welfare.

The authors thank the following people for participating in the study: H. Toyotama, S. Sakuma, M. Hamada, Y. Uchijima, K. Hatano, K. Kashiwagi, and Y. Liu, and three anonymous reviewers for their comments and suggestions,

### References

- [1] Lindberg D, Humphreys B, McCray A. The Unified Medical Language System. Methods Inf Med 1993; 32(4): 281-91.
- [2] Campbell KE, Oliver DE, et al. Representing Thoughts, Words, and Things in the UMLS. JAMIA. 1998;5:421-31.
- [3] Cimino JJ. Auditing the Unified Medical Language System with Semantic Methods. JAMIA. 1998;5:42-51.
- [4] Bodenreider O, McCray AT. From French vocabulary to the Unified Medical Language System. Medinfo. 1998; 9 Pt 1:670-4.
- [5] Hersh WR, Donohoe LC. SAPHIRE International: a tool for cross-language information retrieval. Proceedings/AMIA Annual Symposium. 1998; 673-7.
- [6] Tanaka M. A Mapping Approach of Japanese Medical Vocabulary to UMLS. Proceedings/Japan Congress of Medical Informatics 2000; 920-1
- [7] Liu Y, Satomura Y. Building a Controlled Health Vocabulary in Japanese. Method Inform Med 2001; 40(4): 307-14.
- [8] Hatano K, Hamada A, et.al. Developing of Clinical Disease Name Classification Based on ICD 10-based Code Set of Disease Names. Proceedings/Japan Congress of Medical Informatics 2001; 739-41

### Address for correspondence

Yuzo Onogi, MD, PhD, Clinical Bioinformatics Research Unit, Graduate School of Medicine, the University of Tokyo. 7-3-1 Bunkyo Hongo Tokyo, 113-8655 Japan.

# UMLSと連携した日本語医学用語シソーラスの開発実験

○脊山 洋右<sup>1)</sup> 開原 成允<sup>2)</sup> 野添 篤毅<sup>3)</sup> 小野木 雄三<sup>4)</sup> 篠原 恒樹<sup>5)</sup> 鈴木 博道<sup>6)</sup>  
 お茶の水女子大学 生活科学部<sup>1)</sup>  
 国際医療福祉大学<sup>2)</sup>  
 愛知淑徳大学<sup>3)</sup>  
 東京大学大学院 医学研究科<sup>4)</sup>  
 医学中央雑誌刊行会<sup>5)</sup>  
 国際医学情報センター<sup>6)</sup>

## Developmental experience of thesaurus of Japanese medical terms corresponding to UMLS metathesaurus

Yousuke Seyama<sup>1)</sup> Shigekoto Kaihara<sup>2)</sup> Atsutake Nozoe<sup>3)</sup> Yuzo Onogi<sup>4)</sup> Tsuneki Shinohara<sup>5)</sup>  
 Hiromichi Suzuki<sup>6)</sup>

Faculty of Human Life and Environmental Sciences, Ochanomizu University<sup>1)</sup>  
 International University of Health and Welfare<sup>2)</sup>  
 Aichi Shukutoku University<sup>3)</sup>  
 Graduate School of Medicinem, University of Tokyo<sup>4)</sup>  
 Japan Medical Abstract Society<sup>5)</sup>  
 International Medical Information Center<sup>6)</sup>

**Abstract:** We introduce the experience of developing Japanese medical thesaurus, according to the results of mapping Japanese medical terms to the UMLS Metathesaurus. We used a custom designed Web interface to evaluate the results. And we are now using the sample thesaurus to retrieve Pubmed from Japanese medical terms as a trial.

**Keywords:** UMLS, thesaurus, medical terminology

### 1. 目的

類似の情報を整理したり、適切な情報を検索するためには、カテゴリー性を持ったシソーラスの存在が有益であることに異論を挟む人はいないであろう。しかしシソーラスに関して大きな問題となるのは、シソーラスを作成し定期的に更新するための手間と実用性と考えられる。最初にシソーラスを構築するには巨大な手間がかかり、一度出来上がったシソーラスも更新されない場合には役に立たないことになりかねない。そこで、米国国立図書館(NLM?National Library of Medicine)が開発したUMLS(Unified Medical Language System)と連携しこれを活用した日本語医学用語シソーラスの開発技法を試み、評価と実用性の検討を行った

### 2. 方法

UMLSは87万のコンセプトを中心として、全世界200以上のボキャブラリー・ファイルを結びついているリレーショナル・データベースであり、UMLSメタシソーラスはMeSH、ICD-10、CPT、SNOMED、MedDRAなどを情報源としてドイツ語・フランス語・スペイン語その他、多言語の医学用語にもリンクがとれており、210万語からなるメタシソーラスはボキャブラリー・ファイルとして捕らえることも出来、これを介して各国語への変換も不可能ではないものである。UMLSをフルに活用し、日本医学会医学用語辞典・医学中央雑誌シソーラス、看護学事典、その他の日英対訳の形となっているシソーラス・用語集の、英語文字列の正規化処理その他の自動的処理を中心とした手法で、極力機械的手段を講じた日本語医学用語シソーラスの構築手段を試みた。

また、部分的ではあるものの構築した日本語医学用語シソーラスのWeb上での参照システム、そして、機械的手段でのマッチング結果を評価し最適訳語を投票するWebシステムも開発し、マッチング結果の検討を行った。

更に、シソーラスは単なるボキャブラリー・ファイルではなくに情報検索などの過程で機能してこそ役立ち、その存在意義が明らかとなるものである、との観点から、機械的手段で構築された日本語医学用語シソーラスの試用実験も行っている。シソーラスの用語集としての性格を活用した、日本語によるPubmed検索システムがその一つである。シソーラスの階層性を活用した試用として、日本生化学学会の発表演題を材料としてシソーラスのカテゴリー分類を利用した、演題自動分類も試みている。学会プログラム編成の自動化可能性を探る試みでもある。

なお、これらのテストには、UMLS2002を利用した。

### 3. 結果

機械的なマッチング作業は、日本医学会医学用語数の英語文字列をプログラムで正規化処理し、UMLSの正規化文字列索引と照合することで行った。当初、UMLS2001を材料としたトライアルでは、マッチング率が直接マッチングの11.6%に対して正規化処理後には34.8%に大幅に改善し、一定の効果が期待された。この段階で、英単語のスペルミスと思しきものが目についたため、正規化した英単語36,590語を複数の英語辞書と照合、いずれの辞書でも単語として照合不可能であった6,285語を取り出し、簡易スペルミス・チェックプログラムにかけたところ45%の2,832語がスペル訂正された。

以上を参考として、日本医学会医学用語集と医学中央雑誌刊行会の医学用語シソーラス第5版とを利用、MeSH関連語に重点を置いたより効率性を期待したマッチングを試みた。そして機械的マッチングの成果に関する評価を行った。対象となったコンセプトは20,743でこれに含まれる用語数は179,856語、それぞれUMLS全体の2%、10%に相当するサンプル数と言える。機械的マッチングで不備が指摘されたものは7.4%、何らかのコメントがつけられたものが、10.2%となった。

この作業に使い、日本語医学用語がUMLSに妥当な対応付けされたか否かを参照し、点検作業を分担して行い、問題点などの指摘と修整提案を掲示し、修整提案に対して投票できるようなWeb上のシステムを構築し

た。20,743コンセプト・コードに対して19,180は何らかの形で自動的にリンクされ、医学用語辞典8万語中では40から50%がマッピングされる結果となった。未だ、品詞を考慮した処理などで本開発手法の改善を図ることなども部分的に行ってみた。

日本語医学用語によるPubmed検索システムは、シソーラスそのものが形として出来ることも勿論大切であるが、シソーラスが有効に機能し情報検索などが容易になることこそ必要、との観点から、サンプル的に出来たシソーラスを試用するものであり、使用テストを続けている。

#### 4. 結論

シソーラス開発方法として、本技法は有用である。但し、本技法のみでシソーラスを完成させられるものは無いが、機能的に役立つシソーラスの部分的完成が期待出来る。また、更に標準病名マスターであるとかSNOMEDを追加して同じ作業を行うことでより完成度を高めることが期待出来る。

なおかつ、UMLSのコンセプト・コードを介することにより多言語間の情報交換にも利用可能である。

#### 5. その他

本研究は厚生労働省医療技術評価総合研究事業の研究補助を受けたものである。

---