

原 著

医療用語のシソーラス作成にむけた予備的調査

相良 かおる* 小野 正子** 上野 恵子***

<要 旨>

電子カルテシステムの導入により、医療施設では電子医療記録文書データが日々蓄積される。

我々は、専門用語に加え、略語や隠語などの多様な表現で記載され、また誤字脱字も含まれる電子医療記録文の自然言語処理を支援するために医療用語を収集し、分ち書き用辞書ComeJisyoを無償公開し、現在は、医療用語シソーラスの作成に向け予備的調査を行っている。

今回、ComeJisyoの登録語739語を対象データとし、医学およびその関連領域の日本語シソーラスである「医学用語シソーラス第7版（登録語約59万語）」の 카테고리分類に着目し、照合を行った。

その結果、739語中、シソーラスに登録されているものが354語あり、その内、我々研究者が付加した分類コードと一致したものは228語、一致しなかったものは126語だった。また未登録語が385語あり、「医学用語シソーラス第7版」の 카테고리分類では分類できないものも111語見つかった。

本稿では、これらの照合結果について述べる。

キーワード：医療用語、医療記録、シソーラス、自然言語処理、語彙分類

1. はじめに

近年、電子カルテシステムを導入している医療施設では電子医療記録文書（以下、医療記録情報という）が日々蓄積され、施設内での情報共有が容易になってきた。また日々蓄積される大量の医療記録情報は、従来の患者診療用途（一次利用）に加え、統計資料、臨床研究や疫学研究、教育訓練、そして診療情報管理などの2次利用が可能となり要望されるようになってきた。

しかしながら、医療記録情報に使われる医療用語の標準化がなされていないこともあり、医療記録情報には、専門用語に加え、略語や隠語、多様な表現、また誤字脱字も含まれるため、これらをコンピュータで処理（以下、自然言語処理という）し、2次利用することは容易ではない。

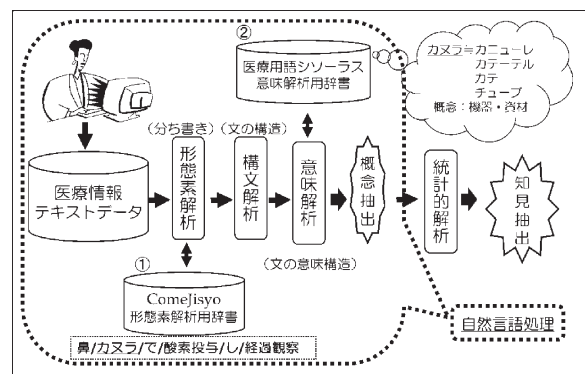


図1. 医療記録情報のテキストマイニングにおける自然言語処理と辞書

そこで我々は、医療記録情報の自然言語処理を支援することを目的に、2004年より看護実践用語の収集を開始し、用語の分析^{1) 2) 3)}と看護用語の標準化に関する調査研究^{4) 5) 6)}を行った。これらの結果を基に、2008年に医療記録文書の語分割（分ち書き）用辞書ComeJisyoV1（登録語数30,146語）の無償公開を開始、

* 西南女学院大学保健福祉学部栄養学科

** 西南女学院大学保健福祉学部看護学科

*** 西南女学院大学短期大学部生活創造学科 非常勤講師

以後随時更新を続け2013年11月からはComeJisyoV5-1 (登録語数77,760語)を公開中である⁷⁾⁸⁾⁹⁾¹⁰⁾¹¹⁾(図1①)。

テキスト形式の医療記録情報から、ある情報を探し出す(以下、情報検索という)ためには、文字で表記された「ことば」を用いる。しかしながら、情報を文字で表記した「ことば」と「情報の意味」は1対1に対応している訳ではない。同義語、関連語、反義語、広義語、狭義語、連語などの意味的な関係をたどる高度な情報検索には、語義で整理した意味解析用辞書、すなわちシソーラスが必要になる。また、大量のテキストデータから新たな知見を発見するテキストマイニングの中でも意味的にまたは概念的に整理された辞書が使われる(図1②)。

本研究の目的は、看護実践用語を含む医療記録情報のための意味解析用の辞書(シソーラス)の作成であり、今回、ComeJisyoV5-1の登録語77,760語の中で4種類の電子医療記録に出現する739語について、『医学用語シソーラス第7版』¹²⁾の第2層分類カテゴリーで分類を行い、登録用語の照合を行った。

2. 用語の定義

本稿で用いる用語の定義を以下に示す。

医療記録情報：医療施設で蓄積されるテキスト形式の電子医療記録文書データ。

自然言語：我々人間が日頃使っている日本語や英語などの言葉。なお、コンピュータが理解できるプログラム言語を人工言語という。

自然言語処理 (natural language processing: NLP)：自然言語をコンピュータで処理する一連の技術(図1)。

テキストマイニング：大量のテキストデータを自然言語処理し構造化した後、統計的手法を用いて特徴あるデータを検出して提示する技術(図1)。

情報検索 (information retrieval: IR)：収集蓄積した文書に索引付けをし、索引と検索語を照合し検索語を含む文書を探し出すこと。

全文検索：文書全体を調べて、検索語が出現した文書を全て探し出すこと。

ディスクリプタ：同義語、類義語の中で索引語、検索語として利用する語を意味する。「統制語」または「優先語」ともいう。(図2)

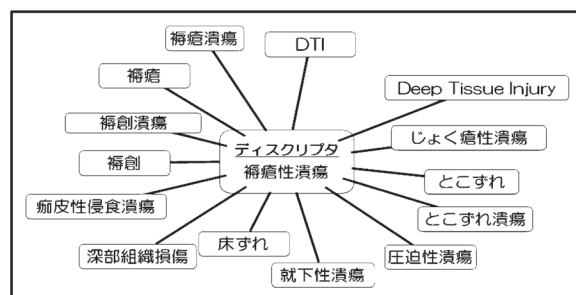


図2. ディスクリプタと同義語・類義語

索引語 (indexing term)：対象となる学術論文などの文書の内容を特徴付ける単語。なお、文書中から索引語を抽出する処理を索引付け (indexing) という。

ライフサイエンス：医学、歯学、薬学、看護学、介護、ヘルスケア、生物学など広範囲にわたる分野をいう。

看護実践用語：看護の現場で行われる実践を記述するために用いられる用語であり、口頭伝達もしくは文書による伝達の双方で用いられている用語全てを含む。

3. シソーラス

シソーラス (thesaurus) とはもともと、辞書や百科事典などの知識の宝庫を意味するギリシア語である。イギリスのPeter Mark Rogetが著した*Thesaurus of English words and phrases* (1852)により広まったとされている。このRogetのシソーラスを参考に、日本語のシソーラス(同義語や類義語を分類、配列した辞書)が複数構築されている。日本語においては、類義語辞典と呼ばれることが多く、日本語の主なシソーラス・類語辞典として、『角川類語新辞典(1981)』、『EDR概念体系辞書(1993)』、NTT『日本語語彙体系(1997)』、『類語大辞典(2002)』、『日本語大シソーラス(2003)』、『類語例解辞典(2003)』、国立国語研究所『分類語彙表(2004)』、『三省堂類語新辞典(2005)』がある¹³⁾。

医学およびその関連領域の日本語シソーラスとしては、医学中央雑誌刊行会『医学用語シソーラス』がある。この『医学用語シソーラス』は、米国国立医学図書館(National Library of Medicine, NLM)の医学、歯学、薬学、看護学、介護、ヘルスケア、生物学など広範囲にわたる分野(以下、ライフサイエンスという)の学術文献データベースMEDLINE (MEDLARS Online)のシソーラスであるMeSH(3.2節参照のこと)に準拠している¹⁴⁾¹⁵⁾¹⁶⁾。

同義語がまとめられディスクリプタとして正式名称

が付加され、ディスクリプタ間の階層構造や関連が整理された機械可読な統制されたシソーラス（統制シソーラスという）は、情報検索において、検索網羅性を高め、検索効率を向上させる。

また、大量のテキストデータを自然言語処理し、概念を抽出し、統計的解析を行うことで新たな知見を得る場合がある。この技術をテキストマイニングと言い、テキストマイニングにおいてもシソーラスが用いられる（図1）。

3.1. 本研究の意義

学術論文は、専門家が読むことを想定し、正式な学術用語が使われ、誤字脱字などの少ない信頼性の高いテキストデータである。学術文献データベースの提供者は、これらの論文内容から主要な概念を抽出し、索引語を統制シソーラスのディスクリプタから見つけて付加した後、データベースに収録する。そして文献検索の際には、検索者が入力した検索キーワードと統制シソーラスを照合しディスクリプタに変換後、文献に付加された索引語との照合を行う。

一方、本研究で扱う医療記録情報は、入力者、入力時間、診療科名などの識別子は付加されるものの、索引語は付加されないため、検索技術としては全文検索が使われる。電子カルテシステムの多くは、改竄予防の措置が取られているため、入力後の加筆修正は容易ではなく、誤字脱字が含まれたままである。そして忙しい中、限られたスペースで立ったまま短時間に入力するなどの入力環境にも起因し、一般的な日本語文に比べ、1文が短く、主語が省略される傾向があり、さらに格助詞の省略による造語も多く、略語や隠語、方言など多様な表現が含まれている。そのため、医療記録情報から「褥瘡」に関する情報を取得したい場合、検索キーワードには「褥瘡」だけでなく、「床ずれ」、「とこずれ」、「褥創」など、想定される同義語を指定しなければならない。

図3は、CiNii（NII学術情報ナビゲータ）で“テキストマイニング”と、“医療”および“看護”のAND検索をした結果である。近年、看護の分野でテキストマイニングツールを使った研究が行われるようになってきたことが分かる。医学書院の「看護研究」2013年8月号、10月号で「看護研究におけるテキストマイニング」の特集号を出版しており、2013年は、“テキストマイニング”と“看護”のAND検索結果は、記事検索で28件、タイトル検索で17件と多くなっている。そして、タイトル検索17件の内、学生の看護実習記録を

含め看護記録やアンケート調査の自由記述文を対象としたものが9件となっている。



図3. 医療と看護領域でのテキストマイニングを使った研究

したがって、看護学生の実習記録など看護領域の経験・知識が十分でない記録者による記述を含めた自由記述文においては、誤字を含めた自然言語処理が必要であると我々は考えている¹⁷⁾。

また、ライフサイエンス分野での欧文文献データベース用のシソーラスにはMeSH（3.2節）が、日本語の文献データベース用のシソーラスとしては『医学用語シソーラス』（3.3節）があるものの、医療記録情報に含まれる多職種間で利用可能な略語や隠語、業界用語を対象とした医療用語シソーラスは見当たらない。医療記録情報に含まれる医療用語を対象とした実践医療用語シソーラスは、医療記録情報を自然言語処理する際に、また、専門用語の概念を適切に学ぶ辞書としても有用である。

そこで我々は、医療記録情報に含まれる用語を対象とした実践医療用語シソーラスの作成を目指し、研究を進めている。

シソーラスの作成・維持には、多くの経費が必要である。そこで近年自動的にシソーラスを作成する手法が提案されている^{18) 19) 20) 21) 22)}。また、シソーラス作成の基盤となる単語間の意味的な関係の近さを数値化する研究も報告されている^{23) 24) 25)}。

しかし、これらは大量のテキストデータを必要とするため、個人情報が含まれ門外不出の医療記録情報からこれらの手法を用いてシソーラスを作成することは難しい。

そこで、既存のシソーラスや類語辞典、そして概念が体系化され公開されている用語集の活用を考え、調査したところ、看護師国家試験に出現する看護領域の用語の一部についてはMeSH（概要は次節を参照のこと）のカテゴリーで分類することが困難であることが

分かっている²⁶⁾。

次に筆者は、世界的規模での看護実践用語の体系化を目指し開発されたICNP[®] (International Classification for Nursing Practice / 看護実践国際分類) のカテゴリー分類の妥当性を調査した。具体的には、ICNP[®] ベータ 2 日本語訳と NANDA 看護診断ラベル第 3 版の診断ラベル、そして NIC 看護介入分類ラベルとの照合を行った。併せて、看護記録には看護実践用語だけでなく、患者の嗜好や日常生活に関する記述など一般的な日本語も多く含まれることから、一般的な日本語を対象とした類義語辞書である国立国語研究所で開発された「分類語彙表」の妥当性についての調査も行った⁶⁾。

広義の看護学術用語を対象としたシソーラスに CINAHL (Cumulative Index to Nursing and Allied Health Literature) シソーラスがある。CINAHL シソーラスのツリー構造は MeSH カテゴリーの階層構造に看護に用いられる分類体系 (P カテゴリー) を加えたものであり、看護学術用語が多く登録されているものの、看護実践用語に含まれる①量や質などの程度、②疾病・症状の程度、③時間的な要素 (服薬前、術前、術後、早い、遅い、時々など) の記述、④身体の上、下、左、右、部分、全体などの部分の表記について該当するカテゴリーがないことが分かっている^{6) 30)}。なお、医療施設で蓄積される医療記録情報の記載者は看護師だけではないことから医療記録情報の自然言語処理には看護実践用語だけでなく医療従事者の用いる用語を対象とした医療用語シソーラスが必要となる。

そこで、MeSH カテゴリー分類に準拠しつつ、日本国内のライフサイエンス領域の論文に含まれる用語を対象としている「医学用語シソーラス」に着目し、今回、そのディスクリプタと ComeJisyo V5-1 登録語の内、4 種の文書に出現した用語 739 語との照合を行った。

次節では、MeSH および「医学用語シソーラス」の概要について述べる。

3.2 MeSH (Medical Subject Headings)

3.2.1 概要

MeSH は、NLM が、1960 年に開発したライフサイエンス分野のデータベース MEDLARS (Medical Literature Analysis and Retrieval System) の 件名標目 (索引語) の用語集である。現在 MeSH は、MEDLARS のオンライン検索データベースである MEDLINE に論文を登録する際の索引付けのための索引語集や検索者のための検索語集として使用される

他、NLM が作成するデータベースの索引や目録作成などに使用されている²⁸⁾。なお、インターネットを介して無料で MEDLINE を検索できるサービスとして PubMed がある。

毎年何十万件もの論文が生産されるライフサイエンス分野では、日々新しい用語が生まれており、それらに対応して、MeSH は毎年改訂され、2013 年版には 26,000 語以上の用語が収録されている。

MEDLINE に収録される文献は多くの専門家に参照されることを意識して記述され、記述内容の正確性がより厳密に問われるため、略語や隠語、そして誤字脱字の出現頻度は極めて低く、MeSH に登録されている用語は学術的な専門用語がほとんどである。

3.2.2 カテゴリーと階層構造 (Tree Structure)

MeSH は 16 のカテゴリー (第 1 層) に分かれ (図 4)、各カテゴリーはさらにサブカテゴリーに分かれている。サブカテゴリーは一般的な用語から特定の専門的な用語へと最大 12 階層まで細分されている。

A	Anatomy (解剖)
B	Organisms (生物)
C	Diseases (疾患)
D	Chemicals & Drugs (化学物質と薬物)
E	Analytical, Diagnostic & Therapeutic Techniques & Equipment (分析・診断・治療の技術・機器)
F	Psychiatry & Psychology (精神医学・心理学)
G	Phenomena and Processes (現象と過程)
H	Disciplines and Occupations (学問分野と過程)
I	Anthropology, Education, Sociology & Social Phenomena (人類学・教育・社会学・社会現象)
J	Technology, Industry, Agriculture (工業技術・産業・農業)
K	Humanities (人文科学)
L	Information Science (情報科学)
M	Persons (人間)
N	Health Care (保健医療)
V	Publication Characteristics (出版特性)
Z	Geographicals (地理)

図 4. MeSH2010 の第 1 層カテゴリー²⁸⁾
注: カテゴリーの日本語訳は文献²⁸⁾ p.7 を転載

3.3 医学用語シソーラス

ここでは、本研究で用いた「医学用語シソーラス」について文献^{12) 14) 15) 16)} を基に概要を述べる。

3.3.1 概要

日本国内の医学および関連領域の文献データベース 医中誌 Web を提供している医学中央雑誌刊行会は、

文献データに付加する索引語を同義関係、階層関係、関連関係によって関連付けた「医学用語シソーラス」を構築している。1983年に第1版が発行され、最新版は2011年発行の第7版である。

「医学用語シソーラス」は4年に1度改訂され、MeSHの新設語はフリーキーワードとして登録し、改訂の際にシソーラス用語に昇格する。

2014年8月27日現在、収録されている索引・検索に使用するディスクリプタは28,205語、これ以外に医中誌フリーワードが約4万語あり、索引に用いられるディスクリプタは合わせて約7万語となっている。ディスクリプタに対する同義語が約59万語登録され、医中誌で管理する用語数は、約66万語となっている(表1)。

表1. 医学用語シソーラス登録用語数

(2014年8月27日現在)		
	ディスクリプタ数	同義語数
シソーラス語	28,205	409,711
フリーワード	44,541	182,321
計	72,746	592,032

2011年発行の第7版はMeSH2010に準拠し、ディスクリプタ25,588語のうち、25,002語(98%)を収録し、MeSHのディスクリプタのうち、アメリカの医療制度や地名など、国内の文献の索引において必要性の低いものを除き、国際一般名(INN)のついた医薬品、生薬、漢方薬、介護保険制度など日本特有の制度に基づく用

語、MeSHでは比較的登録の少ない看護に関する用語(例:「清拭」)など、国内に必要な語3,203語が追加登録されている¹⁴⁾。

3.3.2 同義関係(ディスクリプタと同義語)

医学用語シソーラス第7版における同義関係は、厳密な同義語、異表記語だけでなく、下位語や類義語も含まれる。同義関係として登録される用語には、以下のものがある。

- ① 漢字、カナ表記などの異表記(例:蛋白質とタンパク質)
- ② 英文と和訳(例:Neoplasmsと腫瘍、新生物)
- ③ 英文とカタカナ表記(例:Bacteroidesとバクテロイデス)
- ④ フルスベルと略語(例:Amyotrophic Lateral SclerosisとALS)
- ⑤ 別名(例:上皮小体と副甲状腺)
- ⑥ 医薬品の一般名、商品名、治験番号(開発コード)、化学名
- ⑦ 下位語(例:感染性関節炎と細菌性関節炎)
- ⑧ 類義語(例:外来手術と日帰り手術)

3.3.3 分類と階層関係(カテゴリーコード)

第7版のカテゴリーの第2層階層まではMeSH2010と共通であり(図5)¹²⁾、最も深い階層は13階層となっている。

A 解剖学	C06 消化器疾患	E 分析、診断、治療の技術と機器	H 学問分野と専門分野
A01 身体	C07 口頭疾患	E01 診断	H01 自然科学
A02 筋骨格系	C08 気道疾患	E02 治療	H02 保健医療業務
A03 消化器系	C09 耳鼻咽喉疾患	E03 麻酔法と無痛法	
A04 呼吸器系	C10 神経系疾患	E04 外科手術	I 人類学、教育、社会学や社会現象
A05 泌尿生殖器系	C11 眼疾患	E05 調査研究法	I01 社会科学
A06 内分泌系	C12 泌尿生殖器疾患—男性	E06 歯科学	I02 教育
A07 心臓血管系	C13 女性生殖器疾患と妊娠合併症	E07 機器と資材用品	I03 人間活動
A08 神経系	C14 心臓血管疾患		
A09 感覚器	C15 血液疾患とリンパ疾患	F 精神医学および心理学	J 工業技術、産業、農業
A10 組織	C16 先天性、遺伝性、および新生児疾患と奇形	F01 行動と行動メカニズム	J01 技術、産業、農業
A11 細胞	C17 皮膚疾患と結合組織疾患	F02 心理現象と過程	J02 食品と飲料
A12 体液と分泌物	C18 栄養疾患と代謝性疾患	F03 精神障害	
A13 動物の身体構造	C19 内分泌系疾患	F04 行動訓練と活動	K 人文科学
A14 口頭系	C20 免疫系疾患		K01 人文科学
A15 血液免疫系	C21 環境起因障害	G 現象と過程	
A16 胚構造	C22 動物の疾病	G01 物理的現象	L 情報科学
A17 外皮系	C23 病理学的状態、症状と徴候	G02 化学的現象	L01 情報科学
A18 植物の構造		G03 代謝現象	
A19 真菌構造	D 化学物質および薬物	G04 細胞生理学的現象	M 人間集団
A20 細菌構造	D01 無機化学物質	G05 遺伝学的現象	M01 人間
A21 ウイルス構造	D02 有機化学物質	G06 微生物学的現象	
	D03 複素環式化合物	G07 生理学的過程	N 保健医療サービス
B 生物	D04 多環式炭化水素	G08 生殖と泌尿の生理学的現象	N01 人口集団特性
B01 真核生物	D05 高分子物質	G09 循環と呼吸の生理学的現象	N02 保健医療の施設、人的要因、サービス
B02 古細菌	D06 ホルモン、代用物質、拮抗剤	G10 消化器系生理学現象と口腔生理学的現象	N03 保健医療経済と保健医療組織
B03 細菌	D08 酵素と補酵素	G11 筋骨格生理学的現象と神経生理学的現象	N04 保健医療サービス管理
B04 ウイルス	D09 炭水化物	G12 免疫学的現象	N05 保健医療の品質、アクセス、評価
B05 生物生活環形態	D10 脂質	G13 外皮系生理学的現象	N06 環境と公衆衛生
	D12 Amino Acids, Peptides, and Proteins	G14 眼の生理学的現象	
C 疾患	D13 Nucleic Acids, Nucleotides, and Nucleosides	G15 植物生理学的現象	Z 地理的位置
C01 細菌性疾患と真菌性疾患	D20 複合混合物類	G16 生物学的現象	Z01 地理的位置
C02 ウイルス性疾患	D23 生物学的要因	G17 数学的概念	
C03 寄生虫疾患	D25 医用材料と歯科材料		
C04 腫瘍	D26 薬物		
C05 筋骨格系疾患	D27 化学作用と効能		

図5. 「医学用語シソーラス第7版」第2層カテゴリー¹²⁾

注: 文献¹²⁾ p.2-3 カテゴリー一覽より作成

4. 研究方法

4.1 対象データ

我々は、ComeJisyo²⁹⁾の登録語を対象としたシソーラスの作成を目指している。

現在公開中のComeJisyoV5-1の登録語は、看護学教科書の索引語、看護師国家試験問題（2002-2007年）に含まれる用語、看護領域の文書より抽出した用語、Web上で公開されている用語辞書、栄養管理・栄養指導分野で使われる用語、倫理的配慮がなされた3医療施設の医療記録情報に含まれる用語、言語資源協会において研究用に公開されている模擬診療録テキストデータに含まれる用語、併せて77,760語である。

今回の調査では、複数の医療施設で使われていると考えられる用語について検討するためにComeJisyoV5-1登録語のうち、3医療施設の医療記録情報および模擬診療録の4文書全てに出現する739語を対象データとした。

4.2 調査方法

Step 1. 人手による分類

臨床看護の経験者2名を含む研究者3名で、対象データ739語に「医学用語シソーラス第7版」のカテゴリー第2層（図5）の分類コード（以下、第2層分類コードという）を付加した。

第2層分類コードの付加に際しては、経過記録を記載する際の語義・概念を意識し、冊子版の「医学シソーラス第7版」を参考にし、1語に1つの分類コードに絞るのではなく、該当する第2層分類コード全てを付加した。語義の不明なものについては、冊子版および電子辞書の以下に示す辞書を用いた。

- (1) 医学書院 医学大辞典 第2版 2009
- (2) 治療薬マニュアル2010
- (3) ステッドマン医学大辞典 改訂6版
- (4) ステッドマン医学略語辞典
- (5) 医学英語実用語法辞典
- (6) リーダーズ英和辞典 第2版
- (7) リーダーズ・プラス
- (8) New Oxford American Dictionary Second Edition
- (9) 看護大事典 第2版 2010
- (10) カルテを読むための医学用語・略語ミニ辞典 第3版 2011
- (11) 看護学学習辞典 第2版 Gakken
- (12) 広辞苑 第6版

Step 2. 機械的分類

研究者により分類コードを付加した後、表1に示す「医学用語シソーラス第7版」の電子データ（2014年8月27日時点）を用い、プログラム言語PerlおよびExcel 2013を用いて機械的にディスクリプタおよび同義語（登録語592,032語）との照合を行った。

5. 結果

5.1 照合結果

対象データ739語を「医学用語シソーラス第7版」の登録語と照合した結果は、表2に示すとおりである。ディスクリプタと一致した語は156語、同義語と一致した語198語あり、354語（約48%）が「医学用語」シソーラスの登録語と一致していた（これらを以下、一致語という）、そして医学用語シソーラス第7版に登録されていない語（以下、未登録語とする）は385語（約52%）であった。

表2. 対象データと「医療用語シソーラス」登録語との照合

	語数
ディスクリプタと一致	156
同義語と一致	198
未登録	385
計	739

一致語354語（約48%）を「医学用語シソーラス第7版」の第1層、第2層分類コードと照合した結果は、表3のとおりである。分類コードと一致した語は228語、一致しなかった語は126語であった。

表3. 一致語との「医学用語シソーラス」の分類コード照合結果

	語数
分類コード一致	228
不一致	126
	354

分類コードのカテゴリーと一致しなかった126語の内訳は表4のとおりである。第1層カテゴリーが一致しなかった語は36語、第2層カテゴリーが一致しなかった語は90語であった。第2層カテゴリーとの照合で一致しなかった項目の内、語数の多いカテゴリーとして『D 化学物質および薬物』44語、『C 疾患』18語、『E 分析, 診断, 治療の技術と機器』15語があった。

表4. 「医学用語シソーラス第7版」と異なるカテゴリーに分類された126語の内訳

	語数
第1層カテゴリーの相違	36
第2層カテゴリーの相違	90
A 解剖学	9
C 疾患	18
D 化学物質および薬物	44
E 分析, 診断, 治療の技術と機器	15
G 現象と過程	2
N 保健医療サービス	2
計	126

「医学用語シソーラス第7版」未登録語385語について分類コードを付加した結果は表5に示したとおりである。未登録語数の多かったのは、『E 分析, 診断, 治療の技術と機器』95語、『C 疾患』63語、『A 解剖学』38語、『G 現象と過程』30語であった。分類コードが付加出来なかったものが111語あった。

表5. 医学用語シソーラス第7版未登録語385語

カテゴリー	語数
A 解剖学	38
C 疾患	63
D 化学物質および薬物	17
E 分析, 診断, 治療の技術と機器	95
F 精神医学および心理学	5
G 現象と過程	30
H 学問分野と専門分野	2
I 人類学, 教育, 社会学, 社会現象	2
M 人間集団	3
N 保健医療サービス	19
該当なし	111
計	385

6. 考察

表2において、ディスクリプタと一致した対象データ156語に比較して、同義語と一致したものは198語と多くなっている。これは、医療記録情報においては、「病気」と「やまい」などの異表記が多いことと、正式名称ではないものが使われているためである（図2参照のこと）。

掲載語数約66万語の「医学用語シソーラス第7版」と対象データ739語を機械的照合した結果、一致語354語（156語+198語）に比べて、未登録語は385語（約

52%）とやや多くなっていた。これは未登録語の中には、「腹部膨満感」「冷感」「熱感」「鼻閉感」「搔痒感」など患者の主訴を表す感覚的な言葉が多く含まれており、これらが学術用語として文献に出現することが少ないため、未登録となっていると考えられる。

又、第1層カテゴリーが異なるもの36語（表4）の中には、患者の体位を表す「仰臥位」「座位」「側臥位」「立位」「良肢位」がある。「医学用語シソーラス第7版」では“G11：筋骨格生理学的現象と神経生理学的現象”の分類コードが付加されているが、本論文の研究者は、“G11”に加え、“E2：治療”を付加している。これは、体位が患者の状態を改善する治療の一環として使われているためである。また、「尿量減少」については、「医学用語シソーラス第7版」では、“C12：泌尿生殖器—男性”と“C13：女性生殖器疾患と妊娠合併症”、“C23：病理学的状態、症状、徴候”の分類コードが付加されているが、本論文の研究者は、それらに加え、“E1：診断”を、そして水分の出納バランスが崩れたときにも生理的な現象として「尿量減少」が現れるため“G8：生殖と泌尿の生理学的現象”を付加している。

第2層分類コードが異なるもの90語（表4）のうち、「利尿剤」「降圧剤」「造影剤」「鎮痛剤」などの薬剤について、「医学用語シソーラス第7版」では、これらに“D27：化学作用と効用”が付加されているが、本論文の研究者は“D26：薬物”を付加している。

『C 疾患』における不一致語18語の中には、「出血傾向」がある。「医学用語シソーラス第7版」では“C15：血液疾患とリンパ疾患”の分類コードを付加している。しかし「出血傾向」から考えられる疾患は他分野（DIC）にも存在するため、本論文の研究者の分類では、多角的な視点でとらえて「C23病理学的状態、症状、徴候」を加えた。このように「C23病理学的状態、症状、徴候」の分類コードを付加したケースとして、「シバリング」「便潜血」「起立性低血圧」「皮膚障害」「呼吸抑制」などの語がある。

『E 分析, 診断, 治療の技術と機器』における不一致語15語の例としては、「カテーテル留置」がある。「医学用語シソーラス第7版」では、「E7機器と資材用品」の分類コードが付加されている。しかし、臨床ではカテーテル留置は機材であると同時に治療の手段として活用する場面が多いため、本論文の研究者は「E2 治療」の分類コードを追加している。その他同様なケースに「ギプス固定」「良肢位」などがある。

このように「医学用語シソーラス第7版」との照合

により、学術文献データベースで用いることを想定している「医学用語シソーラス第7版」の登録語と医療記録を記載する上での語義には相違があることが分かる。

表5の未登録語385語では、『E 分析, 診断, 治療の技術と機器』95語の中には「酸素投与」「持続点滴」「脈拍数」「換気量」「血ガス」「血糖測定」「人工肛門」などが含まれている。

『C 疾患』63語の中には、「顔面浮腫」「気分不快」「倦怠感」「食欲低下」「搔痒感」「背部痛」「四肢冷感」「末梢冷感」「咽頭痛」など患者の状態を表す語があった。

『A 解剖学』38語の中には、身体の特出出来ない部位、例えば「上腹部」「下腹部」「左季肋部」「肘部」「体部」がある。本論文の研究者は、これらに“A01: 身体”の分類コードを付加しているが、医療記録を記載する上で、もう少し細かく身体の部位を特定できる分類コードで分類する必要があると考えている。

『G 現象と過程』30語には、生理的な働きを表す語「肝機能」「甲状腺機能」「腸蠕動」「睫毛反射」や身体の動きを表す語「可動域」「挙上」「外旋」「外転」「前屈」「前傾」「背屈」などが含まれている。

これらにおいても、患者の全体像を適切に表す別の分類コードを設け、分類した方が良く考えている。

第1層のカテゴリー分類の『該当なし』111語の中には、程度を表す「頻回」「高さ」「深さ」、疾病・症状の程度を表す「著明」「著変」、時間的な要素を表す「就寝前」「急性期」「退院時」「労作時」「安定性」「可動性」など末尾語に「性」がつくもの、「自覚的」「肉眼的」など「的」がつく語などがある。

前述の3.1節でCINAHLでは分類出来ない看護実践用語として、①量や質などの程度、②疾病・症状の程度、③時間的な要素（服薬前、術前、術後、早い、遅い、時々など）の記述、④身体の上、下、左右、部分、全体などの部分の表記があると述べたが、「医学用語シソーラス第7版」においてもこれらの看護実践用語を適切に分類するコードがないことが分かった。

7. まとめと今後の課題

今回、ComJisyoの登録語の内、3医療施設と模擬診療録に出現する739語を対象に、「医学用語シソーラス第7版」と照合し、臨床看護の経験者2名を含む研究者3名による人的分類との比較を行った。

その結果、「医学用語シソーラス第7版」に登録さ

れている一致語が354語あり、うち、本論文の研究者が付加した分類コードと「医学用語シソーラス第7版」の分類コードが一致したものが228語あることから、医療用語を対象としたシソーラスを作成する上で「医学用語シソーラス第7版」の分類コードの活用が可能であると考えられる。一方、「換気量」「脈拍数」など、数量や程度を表現した語や、「腹部膨満感」「鼻閉感」「搔痒感」など患者の主観の状態を表す語、そして「下腹部」「仰臥位」などの患者の身体的部位や体位を表す語については、新たな分類方法・分類コードの設定が必要だと考えている。

また、「医学用語シソーラス第7版」の分類コードと本論文の研究者が付与した分類コードが異なる126語について調べたところ、学術用語での意味と医療記録を記載する際の意味に相違があることが示唆された。

学術文献データベースにおける索引付けでは、分類コードの定義や付加する範囲についてのマニュアルを基に、論文の主題を見つけるのに十分な知識を持つ専門家が、索引付けのトレーニングを受け、索引語を付加している。

一方今回の調査では、情報科学、小児看護学、助産学を専門とする3名の研究者で、冊子体の「医学用語シソーラス第7版」の分類コードを参考にし、専門領域以外の用語を含む対象データ739語に分類コードを付加した。専門外の語義の分からない用語については複数の辞書や辞典を調べた。その結果、分類コードが一致したものの中には「医学用語シソーラス第7版」の分類コードや辞書・辞典の語義の影響を受けたものもあると考えられ、4.2節の人手による分類に記載の「経過記録を記載する際の語義・概念を意識し分類する」という条件が徹底されていない。そこで、今後は他領域の専門家により検証し、分類の精度を高めるとともに、研究者の専門領域である小児看護学、助産学の医療記録で使われる用語を対象に、再度、「医学用語シソーラス第7版」との照合、および分類コード付けを行う予定である。

謝辞

本研究は西南女学院大学共同研究費の助成を得て行われている。

参考文献

- 1) 相良かおる：看護記録に含まれる文書の統語構造, 日本医療情報学会 第5回看護情報研究会論文集, pp. 85-88, 2004
- 2) 相良かおる, 小野正子, 鈴木隆弘, 嶋田元, 小作浩美：看護記録文の計量的用語調査, 人文科学とコンピュータシンポジウム, p.103-110, 2010
- 3) 小木曾智信, 相良かおる：医療分野で使われる複合語の語種構成, 第29回社会言語科学会研究大会発表論文集, p.158-161, 2012
- 4) 相良かおる, 小作浩美, 小暮潔：標準看護実践用語の特徴, 第6回看護情報研究会論文集, P.73-75, 2005
- 5) Kaoru Sagara, Akinori Abe, Hiromi itoh Ozaku, Noriaki Kuwahara, and Kiyoshi Kogure: Features of Standardized Nursing Terminology Sets in Japan, In Proceedings of the 9th on Nursing Informatics (NI2006), p.471-475, 2006
- 6) 相良かおる, 小作浩美, 小暮潔, 納谷太, 桑教則彰：看護文書の意味解析用辞書の構築におけるICNP[®]と「分類語彙表」の活用可能性, 医療情報学 第24巻 第6号, p.657-665, 2005
- 7) 相良かおる, 浅原正幸, 小野正子, 小作浩美：形態素解析器MeCab用看護用語ユーザ辞書の作成と公開, 第28回医療情報学連合大会論文集, p.938-939, 2008
- 8) 相良かおる, 浅原正幸, 小野正子, 外山健二：形態素エンジンMeCab用辞書 ComeJisyoV2および看護教育支援用かな漢字変換辞書の作成と公開, 第29回医療情報学連合大会論文集, p.983-984, 2009
- 9) 相良かおる, 小野正子, 小木曾智信, 小作浩美：電子医療記録の分ち書き用ユーザ辞書ComeJisyo の紹介と単語生起コスト, 言語処理学会 第18回年次大会 発表論文集, p. 621-624, 2012
- 10) 相良かおる, 小野正子, 小作浩美, 鈴木隆弘, 高崎光浩, 嶋田元：分ち書き用辞書ComeJisyoの評価, 医療情報学 第32巻 第6号, p.301-307, 2012
- 11) 相良かおる, 小野正子：実践医療用語辞書ComeJisyoの紹介, 第33回医療情報学連合大会論文集, p.828-830, 2013
- 12) 医学中央雑誌刊行会：医学用語シソーラス第7版, 2011
- 13) 言語処理学会 編：言語処理事典, p.92-93, 2009
- 14) 浜田雅美, 医学用語シソーラス第7版 改訂のポイント, 医学図書館 58 (1), p.61-64, 2011
- 15) 浜田雅美, 医学用語シソーラス：MeSHとの関連性, 情報の科学と技術, 63巻5号, p.193-200, 2013
- 16) 浜田雅美, 佐久間せつ子, 三沢一成：「医学用語シソーラス」と索引作業, 情報管理 Vol.51 No.9, p.542-652, 2008
- 17) 相良かおる：ComeJisyoの紹介と医療情報に含まれる誤字調査, 情報知識学会誌第22回年次大会予稿 Vol.24 No.2, p.204-209, 2014
- 18) 浦本直彦：コーパスに基づくシソーラス：統計情報を用いた既存のシソーラスへの未知語の配置, 情報処理学会論文誌 37(12), 2182-2189, 1996
- 19) 佐々木稔, 新納浩幸：単語クラスターリングの語義判別問題への応用, 情報処理学会研究報告, 自然言語処理研究会報告 2003(23), 145-152, 2003
- 20) 有田一平, 菊池英明, 白井克彦：検索語の共起情報を利用した単語クラスターリングとWeb検索への応用, 電子情報通信学会技術研究報告, NLC, 言語理解とコミュニケーション, 107(158), 115-120, 2007
- 21) 川前徳章, 青木輝勝, 安田浩：統計的モデルを用いた単語クラスターリング, 情報処理学会研究報告, 自然言語処理研究会報告 2001(69), 55-60, 2001
- 22) 山本英子, 神崎享子, 井佐原均：出現状況の包含関係による語彙の階層構造の構築, 情報処理学会論文誌 47(6), 1872-1883, 2006
- 23) 笠原要, 稲子希望, 加藤恒昭：テキストデータを用いた類義語の自動作成, 人工知能学会論文誌, AI 18, 221-232, 2003
- 24) 土田正明, デ・サーガステイン, 鳥澤健太郎, 村田真樹, 風間淳一, 黒田航, 大和田勇人：単語分布類似度を用いた類推による単語間の意味の関係獲得法, 情報処理学会論文誌 52(4), 1761-1776, 2011
- 25) 原一夫, 鈴木郁美, 新保仁, 松本裕治：文法的・意味的共起を利用した単語類似度の計算, 人工知能学会論文誌 28(4), 379-390, 2013
- 26) 穴井めぐみ, 相良かおる, 小野正子, 青山和子, 小田正枝：過去11年間の看護国試問題の形態素解析による看護師国家試験問題の頻出語の分析, 西南女学院大学紀要 Vol.8, p.24-34, 2004
- 27) MeSH Tree Structures - 2010 : http://www.nlm.nih.gov/mesh/2010/mesh_trees/trees.html, (2014年9月14日参照)
- 28) シソーラス研究会：MeSH入門, 情報科学技術協会, 2013
- 29) ComeJisyo : <http://sourceforge.jp/projects/comedic/> (2014年9月14日参照)
- 30) Hyun S. , Park H. A. , 輪湖史子：INR SELECTION ICNPとNANDAおよびHHCC,オマハ・システムのクロスマッピング：統合看護言語の開発を目指して, インターナショナルナーシング・レビュー 26(1), 7-16, 2003

Preliminary Investigation for the Creation of a Clinical Terms Thesaurus

Kaoru Sagara*, Masako Ono**, Keiko Ueno***

<Abstract>

Data from electronic medical record (EMR) documents are accumulated daily through the operation of EMR systems in medical facilities.

In order to support the natural language processing of EMR documents that include technical terms, abbreviations, jargon, and typographical errors, we have created a dictionary called ComeJisyo, which can parse sentences in EMRs for terms using the Japanese morphological analyzer MeCab, and we have been offering it to the public free of charge.

We are now conducting preliminary research to build a thesaurus of medical and health-related terms. In this paper, we describe the results of comparing terms that are commonly used in medical facilities and are registered in ComeJisyo V5-1 with those in an existing medical thesaurus.

Out of 77,760 terms registered in ComeJisyo V5-1, we extracted 739 terms that appear commonly in four types of EMR documents. We arbitrarily classified the extracted terms according to the category classification system of the Thesaurus of Medical and Health-related Terms, 7th Edition (in Japanese) and compared their classifications with those in the existing thesaurus. Of the 739 extracted terms, 354 were registered in the thesaurus, of which 228 were classified into the same category as that in the thesaurus. There were 385 extracted terms not registered in the thesaurus, of which 111 could not be classified into the same category as the thesaurus.

Keywords: medical terms, medical records, thesaurus, natural language processing,
vocabulary classification

* Department of Nutritional Sciences, Faculty of Health and Welfare, Seinan Jo Gakuin University

** Department of Nursing, Faculty of Health and Welfare, Seinan Jo Gakuin University

*** Part-time Instructor in the Department of Life Studies, Seinan Jo Gakuin University Junior College