

末尾語にサ変接続名詞を持つ実践医療用語の語彙分類

相良かおる・小野正子（西南女学院大学 保健福祉学部）

山崎 誠（国立国語研究所）

今回、我々は「複合語では、末尾語が品詞および意味を決定する主要部となる」という規則に基づき、日本語シソーラスである『分類語彙表』を利用した実践医療用語の意味的な語彙分類を試みた。その結果、医療記録情報より抽出した用語の1,185語中、1,103語（93.1%）の末尾語が『分類語彙表』の見出し語と一致し、108グループに分類でき、その内81グループ（75%）は意味的に妥当な分類であった。

Classification of the medical and health-related terms with verbal noun endings

Kaoru Sagara / Masako Ono (Faculty of Health and Welfare, Seinan Jo Gakuin University)
Makoto Yamazaki (Research Dept., NINJAL)

We are now conducting research to build a thesaurus of medical and health-related terms.

Then we have classified, the medical and health-related terms extracted from EMR documents (75,000 sentences) using the Japanese thesaurus "Bunruigoihyou", based on the rule that the meaning of compound word is commonly determined by the last two Kanji characters of the term.

As a result, we found that the last two Kanji characters of 1,103 terms (93.1%) out of the 1,185 extracted terms, are listed in headwords in 108 different synonym groups of "Bunruigoihyou", and of which 81(75%) were available as a thesaurus of medical and health-related terms.

1. まえがき

一般病床（400床以上）の2016年度の電子カルテ導入率は85.1%（予定含む）、300床以上では79.6%となっている[1]。診療録（カルテ）の電子化は、紙媒体では困難だった看護記録等の異なる職種により作成される医療記録との統合を可能にする。そして、チーム医療においては、医師に限らず看護師や療法士等もカルテを作成する。従って、蓄積された電子カルテデータ（以下、医療記録情報という）には、多様な専門分野の用語が含まれる。またカルテは患者の初診から退院までの全体像が記述される。初期記録では、問診時の患者や家族との対話で得た情報や、診断病名、治療計画・方針が、経過記録では体温、呼吸数、心拍数や血圧、手術記録、そして、主観的な患者の訴えや、医師や看護師等が観察した所見等の叙述的経過が記録される。従って医療記録情報には、専門用語だけでなく、数値や記号、日常一般に使われる言葉や表現も含まれる。このように、多様な表現、同義語、誤字脱字、そして非文法的で断片化した表現が含まれる非定型データであることが障壁となって医療記録情報の利活用を困難にしている。

我々は、医療記録情報の自然言語処理を支援するために医療記録情報に含まれる用語（以下、実践医療用語という）を収集し、2008年に語分割

（分かち書き）用辞書 ComeJisyoV1（登録語数30,146語）を公開し、以後随時更新し、2013年11月には ComeJisyoV5-1（登録語数77,760語）を公開している[2]。

しかしながら、医療記録情報の語分割だけでは、非定型データのままであり、人工知能アルゴリズム、学習アルゴリズムの利用は困難である。非定型の医療記録情報を2次利用可能な状態に変換するためには、統制語の定められたシソーラスが有用である。

そこで我々は、実践医療用語シソーラスの作成・公開を目指し、現在、①シソーラスの作成・維持には、多大な時間と労力を要すること、②医療記録情報には、学術医学用語の他、看護学学術用語等の関連する分野の専門用語や日常語が含まれていることを勘案し、既存のシソーラスや類語辞典、概念の体系化された用語集等の活用可能性を調査している。

既に医療記録情報から抽出した用語739語について、医学およびその関連領域の学術用語を対象とした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）」のカテゴリー分類に着目し、照合を行い、医療行為や看護行為などの動作・行為を表す用語の語彙分類での利用が困難であることが分かっている[3]。また、医療記録情報と『現代日本語書き言葉均衡コーパス』[4]の比較において、①医療記録情報の方が

はるかに高密度に漢字連続が出現すること、②名詞の割合が高く、動詞の割合が低いことが分かっている[5]。実際、医療記録では、医療行為や看護・介護の行為は、サ変動詞の「する」が省略された体言止めで表現されている。

そこで今回、右側主要部の規則、すなわち末尾の造語成分（以下、末尾語という）が語全体の品詞および意味を決定する主要部となること[6]を利用し、小児看護および助産領域の経過記録から抽出した動作・行為を表す名詞連続語 1,185 語を対象に、『分類語彙表』の分類番号によるカテゴリ分類の活用可能性を調べた。本稿ではその結果について述べる。

2. 関連研究

2.1 医療記録情報の言語的特徴

医療記録情報の多くは医療の現場で短時間のうちに作成するという性質から、総体的に短い文が多い。また、メモのような要点だけを記録するような文も多く見られる。樺島（1979）では、このような要約的な文章では、そこに含まれる名詞の比率が高いことが指摘されている[7]。そこで実際に調査したところ、助詞・助動詞や記号等を除いた場合、本稿で扱った医療記録情報の名詞の比率は 77.23% であり、この割合は樺島（1979）に挙げられている新聞の見出しや新聞記事よりも高い割合であることが分かっている[8]。

表 1. 医療記録情報と現代日本語書き言葉均衡コーパス (BCCWJ) における品詞の割合

品詞	医療記録情報(%)	BCCWJ(%)
名詞	39.76	29.53
代名詞	0.19	1.22
動詞	7.99	11.40
形容詞	1.27	1.28
形状詞	0.98	1.06
連体詞	0.25	0.80
副詞	1.12	1.47
接続詞	0.10	0.39
感動詞	0.06	0.13
助詞	16.12	25.33
助動詞	3.76	8.28
接頭辞	0.71	0.70
接尾辞	2.65	2.70
記号	3.95	0.41
空白	2.72	2.04
未知語	1.17	0.17
補助記号	17.21	13.02
その他	0.00	0.07
合計	100.00	100.00

本稿で対象とする医療記録情報の全品詞を「現代日本語書き言葉均衡コーパス」(以下、BCCWJ と略す) [4]と比較したのが表 1 である。医療記録情報は BCCWJ に比べて、名詞の比率が高いことと同時に、助詞の割合が低いことが観察される。名詞の割合が高く、助詞の割合が低いということは、林（1982）や石井（1993）で言う、「臨時一語」的複合語が多いのではないかと推測される[9] [10]。そして複合語を構成する造語成分の結合力において、和語と漢語を比較すると、漢語の結合力が圧倒的に強いことが分かっている[6][11]。

既に山崎・相良（2014）において、同じ条件で医療記録情報と BCCWJ から漢字連続を抽出し分析した結果、医療記録情報の方がはるかに高密度に漢字連続が出現することを確認している[8]。

また、表 1 から医療記録情報では動詞の割合が BCCWJ に比べて低いことも観察される。これは例えば、サ変動詞の「する」を省略して使うような状況が想定され、このことも名詞連続を増やす一因になっていると考えられるが、具体的な調査には至っていない。

2.2 専門用語の語彙分類

相良・小野他（2015）は、実践医療用語（739 語）について、学術用語を対象とした日本語シソーラス「医学用語シソーラス第 7 版」のカテゴリ分類を用いて分類し、未定義語やカテゴリ分類の異なるものに、医療行為、看護行為、処置名が多く含まれていたことを報告している[3]。

石井（2007）では、学術専門用語 23 分野 65,576 語を分割して求めた造語成分延べ 147,822 語、異なり 19,853 語について、1964 年に刊行された初版『分類語彙表』[12]を一般語のリストと考え、一致すれば「一般語」とみなすという調査を行い、以下のことを明らかにしている[6]。

- (1) 「一般語」の割合は、造語成分の異なり語数で 3 割、延べ語数で 7 割であること。
- (2) 「一般語」の造語成分は、「専門語」の造語成分に比べて、より多くの用語生成に寄与していること。
- (3) 造語成分の語種比率では、「一般」の造語成分には漢語が 68.3%、次いで和語 22.8%、「専門語」では、漢語 55.1%、次いで「外来語」41.5% であること。
- (4) 「専門語」は単独または用語の前要素に多い。一方、「一般語」は「専門用語」の造語成分において後要素（右側要素）に多く、造語の後要素として類概念を表すことが多い。

一般に複合語等の複数の造語成分からなる用語では、右側主要部の規則が成り立つ。すなわち末尾の造語成分（以下、末尾語という）が語全体の品詞および意味を決定する主要部となることが分かっている[13]。

3. 研究方法

3.1 提案手法

関連研究より、①学術専門用語の造語成分の約7割が『分類語彙表』の見出し語と一致する一般語であり、その多くが末尾語であること、②日本語において複数の造語成分からなる用語の品詞および意味は末尾語によって決まることが分かっている。

本研究では、石井（2007）の調査に倣い[6]、実践医療用語の末尾語について、『分類語彙表』の見出し語と照合し、一致したものを「一般語」とみなし、未登録語を「専門語」とする。「一般語」については、分類番号を付加し分類した後、分類結果についてその妥当性について検証する。

なお、本調査では1964年に出版された初版ではなく、2004年に刊行された『分類語彙表増補改訂版』[14]（以下、『分類語彙表』と略す）を用いる。『分類語彙表』は、出版当初から日本語の語彙・意味の研究に活用され、現在、延べ95,811語、異なり79,516語を収録しており、類、部門、中項目、分類項目の4つのレベル（段落まで含めると5つのレベル）で語彙を分類している。

3.2 調査方法

相良・小野他（2015）の調査過程において、同じ用語であっても、医学用語辞典に記載の意味と医療記録や申し送りなどで使われる際の意味とにズレのある用語があることが分かっている。

そこで、実践医療用語の意味を基にして語彙分類するために、小児看護の臨床経験を持つ共著者および、助産師の資格・経験を持つ研究協力者の専門領域である小児・助産を扱う診療科（2施設分）の医療記録情報を対象とし、言語処理担当の共著者を含めた3名でデータの作成を行った。

以下にデータの作成方法を述べる。

- (1) 小児・助産領域の医療記録情報 75,000行から、正規表現を用い、Perlプログラムで連続するJIS第1水準、第2水準の漢字の末尾にひらがなの“す”、“し”、“さ”が付加している文字列を抽出。
- (2) 目視により“食事少し”などの接辞が副詞であるものを削除しサ変接続可能な複合語を抽出。
- (3) 日本語として適切ではないものを削除。
- (4) 末尾の“す”、“し”、“さ”を削除後、末尾語がサ変接続漢字である1,185語を抽出。
- (5) 漢字連続語の末尾語と『分類語彙表』の辞書系を照合し、一致するものに分類番号を付加。
- (6) なお、『分類語彙表』において、複数の分類番号が付加されている用語（多義語）については、最初に一致した分類番号（分類番号の小さいもの）を付加（表2）。

例：「吸収」には3つのコードが付加されて

いるが、本調査では“1.1533”を付加。

(7) 末尾語に付加された『分類語彙表』の分類番号により分類。

(8) 分類の妥当性について検討。

表2. 「吸収」の分類番号

分類番号	見出し語	部門	中項目	分類項目
1.1533	吸収	関係	作用	漏れ・吸入など
1.3393	吸収	活動	生活	口・鼻・目の動作
1.38	吸収	活動	事業	事業・業務

3.3 匿名化处理

本研究で用いる医療記録情報は、医療施設において以下の匿名化处理のなされた個人情報を含まないテキストデータである。

- (1) 数値は“9”に置換し、固有名詞および個人名は“X”に置換されている。
- (2) 利用者が意図的に改行を行った箇所および文中の“.”の直後で分割したものを1行の文字列とする。
- (3) 同文を削除。

従って、本研究で用いるテキストデータは、(1)の匿名化により、検査項目である“Co2”や“HbA1c”など数値を含む固有名詞は“Co9”、“HbA9c”となっている。また(2)の処理で“患児が「お腹がすいた。何か食べたい。」と言った。”というデータは、“患児が「お腹がすいた。”、“何か食べたい。”、“”と言った。”の3行に分割され、さらに(3)の同文削除により、文の入力順序の情報が失われるため、内容についての分析は勿論のこと、構文についての分析調査にも適さない文字列データである。

4. 結果

表3. 末尾語の照合結果

	延べ語数	異なり語数
一般語	1,103 (93.1%)	245 (87%)
専門語	82 (7%)	36 (13%)
計	1,185 (100%)	281 (100%)

調査データの末尾語 1,185 の内、1,103 語 (93.1%) が「一般語」で、「専門語」はわずか 82 語 (7%) であった。異なり語数においても、末尾語 281 の内、245 語 (87%) が「一般語」で、「専門語」は 36 語 (13%) と、末尾語の 9 割が「一般語」であった（表3）。

「一般語」について分類番号でまとめると、1,185 語が 108 グループに分類された（表4）。

表 4. 末尾語の分類結果

要検討	通番	分類番号	分類項目名	延べ語数	異なり語数
	1	1.111	関係	1	1
	2	1.1112	因果	1	1
	3	1.112	相対	3	1
※1	4	1.12	存在	2	2
	5	1.121	出没	19	2
	6	1.1211	発生・復活	3	2
※2	7	1.122	成立	9	4
※3	8	1.124	保存	6	2
	9	1.125	消滅	13	1
	10	1.1251	除去	2	2
	11	1.13	様相・情勢	2	2
	12	1.1302	趣・調子	5	1
	13	1.1341	弛緩・粗密・繁簡	2	1
	14	1.1342	調節	8	2
※4	15	1.15	作用・変化	94	7
	16	1.1501	変換・交換	1	1
※5	17	1.1502	開始	63	2
※6	18	1.1503	終了・中止・停止	59	5
※7	19	1.1504	連続・反復	55	3
	20	1.151	動き	3	2
	21	1.1513	固定・傾き・転倒など	1	1
	22	1.152	進行・過程・経由	10	3
	23	1.1521	移動・発着	2	1
	24	1.1522	走り・飛び・流れなど	1	1
	25	1.1524	通過・普及など	4	2
※8	26	1.1527	往復	26	5
※9	27	1.1531	出・出し	16	4
※10	28	1.1532	入り・入れ	15	7
※11	29	1.1533	漏れ・吸入など	9	4
※12	30	1.154	上がり・下がり	28	3
	31	1.155	合体・出会い・集合など	2	2
	32	1.1553	開閉・封	5	3
	33	1.156	接近・接触・隔離	6	4
	34	1.1561	当たり・打ちなど	1	1
	35	1.1563	防止・妨害・回避	5	2
	36	1.157	成形・変形	2	2
※13	37	1.158	増減・補充	59	12
	38	1.1581	伸縮	4	1
	39	1.1583	進歩・衰退	7	3
	40	1.1584	限定・優劣	1	1
	41	1.165	順序	3	2
	42	1.166	新旧・遅速	5	3
	43	1.178	ふち・そば・まわり・沿い	1	1
※14	44	1.1902	数	1	1

要検討	通番	分類番号	分類項目名	延べ語数	異なり語数
	45	1.2416	職人	1	1
	46	1.3	心	1	1
	47	1.3001	感覚	1	1
	48	1.3002	感動・興奮	3	2
	49	1.3003	飢渴・酔い・疲労・睡眠など	3	2
※15	50	1.3013	安心・焦燥・満足	7	3
	51	1.3021	敬意・感謝・信頼など	1	1
	52	1.3042	欲望・期待・失望	8	2
	53	1.305	学習・習慣・記憶	5	2
※16	54	1.3061	思考・意見・疑い	19	2
	55	1.3062	注意・認知・了解	64	8
	56	1.3064	測定・計算	8	2
	57	1.3065	研究・試験・調査・検査など	19	5
	58	1.3066	判断・推測・評価	15	2
	59	1.3067	決心・解決・決定・迷い	7	3
	60	1.3071	論理・証明・偽り・誤り・訂正など	2	1
	61	1.3075	説・論・主義	1	1
	62	1.3081	方法	11	1
	63	1.3084	計画・案	1	1
※17	64	1.3091	見る	17	4
	65	1.3092	見せる	13	3
	66	1.3093	聞き・味見	8	2
	67	1.3122	通信	2	1
	68	1.3133	会議・論議	2	1
	69	1.3135	批評・弁解	1	1
	70	1.3136	説明	6	1
	71	1.3151	書き	5	2
	72	1.333	生活・起臥	4	3
※18	73	1.3331	食生活	20	5
	74	1.3333	住生活	6	1
※19	75	1.3334	保健・衛生	1	1
	76	1.3371	旅・行楽	1	1
	77	1.3393	口・鼻・目の動作	7	1
※20	78	1.343	行為・活動	50	3
	79	1.347	成功・失敗	2	2
	80	1.3511	出欠	5	2
	81	1.352	応接・送迎	5	2
	82	1.3522	仲介	2	1
	83	1.353	約束	4	1
	84	1.3532	賛否	4	2
	85	1.3543	争い	1	1
※21	86	1.3613	捕縛・釈放	1	1
	87	1.362	運営	7	1
	88	1.364	教育・養成	3	1
※22	89	1.365	救護・救援	2	2

要検討	通番	分類番号	分類項目名	延べ語数	異なり語数
	90	1.366	請求・依頼	4	2
	91	1.37	取得	11	2
	92	1.3701	所有	2	1
	93	1.3711	需給	1	1
	94	1.377	授受	1	1
	95	1.381	農業・林業	2	1
※23	96	1.3811	牧畜・漁業・鉱業	9	3
	97	1.383	運輸	1	1
※24	98	1.3831	医療	94	14
	99	1.3842	炊事・調理	1	1
※25	100	1.385	技術・設備・修理	2	2
	101	1.3852	扱い・操作・使用	20	2
	102	1.513	水・乾湿	1	1
※26	103	1.517	熱	10	2
	104	1.5607	体液・分泌物	1	1
	105	1.571	生理	9	4
※27	106	1.572	障害・けが	1	1
	107	1.5721	病気・体調	6	4
	108	3.1913	速度	6	1
	109	専門語		82	36
		合計		1,185	281

表 5. 延べ語数の多い上位 20 グループ

分類番号	分類項目名	延べ語数			異なり語数		
		語数	累積	累積割合	語数	累積	累積割合
1.3831	医療	94	94	8%	14	14	5%
1.15	作用・変化	94	188	16%	7	28	10%
1.3062	注意・認知・了解	64	252	21%	8	35	12%
1.1502	開始	63	315	27%	2	43	15%
1.158	増減・補充	59	374	32%	12	45	16%
1.1503	終了・中止・停止	59	433	37%	5	57	20%
1.1504	連続・反復	55	488	41%	3	62	22%
1.343	行為・活動	50	538	45%	3	65	23%
1.154	上がり・下がり	28	566	48%	3	68	24%
1.1527	往復	26	592	50%	5	71	25%
1.3331	食生活	20	612	52%	5	76	27%
1.3852	扱い・操作・使用	20	632	53%	2	81	29%
1.3065	研究・試験・調査・検査な	19	651	55%	5	83	30%
1.121	出没	19	670	57%	2	88	31%
1.3061	思考・意見・疑い	19	689	58%	2	90	32%
1.3091	見る	17	706	60%	4	92	33%
1.1531	出・出し	16	722	61%	4	96	34%
1.1532	入り・入れ	15	737	62%	7	100	36%
1.3066	判断・推測・評価	15	752	63%	2	107	38%
1.3092	見せる	13	765	65%	3	109	39%

表 6. 延べ語数毎のグループ数 (昇順上位 10 位)

延べ語数	グループ数	累積	累積割合
1	27	27	25%
2	16	42	39%
3	7	50	46%
4	6	56	52%
5	8	64	59%
6	6	70	65%
7	5	75	69%
8	4	79	73%
9	4	83	77%
10	2	85	79%

表 5 は、延べ語数の多いもの上位 20 位をまとめたものである。108 グループの内、上位 20 グループ内に延べ語数で 65%、異なり語数で 39% の用語が含まれていた。

一方、延べ語数で 1 語 1 グループは 27 あり、グループ全体の 25%、2 語 1 グループは 16 あり全体の 14% を占め、全グループの 4 割が延べ語数 1 語または 2 語から成っていた (表 6)。

表 7 では意味のおおまかな分布を比較した。ここでは分類番号の上位 3 桁の「中項目」と呼ばれる分類を比べる。医療記録情報が現れている 20 の中項目数の総数に占める割合において、医療記録情報の上位 5 位中、4 位までが医療記録情報の割合の方が大きく、この意味分類が中心的であると考えられる。

表 7. 意味の分布 (中項目, 異なり)

番号	中項目	分類語彙表	医療記録情報
1.15	作用	8.2%	< 33.5%
1.30	心	14.2%	< 20.0%
1.38	事業	4.8%	< 9.8%
1.12	存在	1.6%	< 6.1%
1.33	生活	7.9%	> 4.9%

5. 考察

『分類語彙表』の見出し語との照合において、末尾語がサ変接続の漢字連続 1,185 語の内、1,103 語 (93.1%) が一致し (表 3)、108 グループに分類することが出来たことから、漢字連続の実践医療用語については、末尾語と『分類語彙表』の見出し語の照合による語彙分類法が利用可能であることが示唆された。

一方、『分類語彙表』の分類の中には、表 4 の通番 17 分類項目名「開始」と通番 18 「終了・中止・停止」、通番 27 「出・出し」と通番 28 「入り・入れ」のように対義語が別になっている場合と、通番 30 「上がり・下がり」や通番 37 「増減・補充」など、類義語としてまとめられている場合

がある。表 5 に示すように本調査データでは、これらの対義語の出現頻度が高いこともあり、実践医療用語の語彙分類においては、対義語を別分類とするか、類義語としてまとめるかを検討し、どちらかに統一すべきだと考えている。

実践医療用語として使われる場面や意味から細分類が必要と思われるものが見つかった。例えば、表 4 の通番 64 「見る」に分類された末尾語は「観察」「散見」「所見」「発見」であったが、「観察」は過程を、「散見」は頻度・程度を、「所見」「発見」は結果を表すことから、我々は別の分類とすべきだと考えている。その他に細分類が必要と考えられるものに通番 29 「漏れ・吸入など」に分類された「吸引」「吸収」「吸入」「点滴」があった。

表 5 より、分類項目名「医療」に分類された末尾語は延べ語数 94 語、異なり語数 14 語と多くなっている。異なり語 14 語は、「加療」「浣腸」「解熱」「検尿」「採血」「再診」「止血」「受診」「診察」「接種」「穿刺」「投薬」「投与」「輸血」と多様な医療用語が含まれ、実践医療用語の意味分類としては、これらを一まとめに分類するのは不適切であり、新たな分類項目を定め細分類すべきだと考えている。

表 6 より、1 語 1 グループのものが 27 語あるが、その中の「打撲 (分類項目名: 当たり, 打ちなど)」は、実践医療用語では、「打撲傷 (分類項目: 障害・けが)」と同義であり、分類項目名「障害・けが」に分類すべきものも含まれる。

『分類語彙表』に収録されている「末尾語」の中には、「一般語」の意味と異なる意味を持つ語がある。相良・小野他 (2015) において、Mesh カテゴリー分類に準拠した「医学用語シソーラス第 7 版」において「E7 機器と資材用品」に分類される「カテーテル留置」は、「機器や資材用品」ではなく、「治療の手段」としての意味を持つことを述べているが [3]、『分類語彙表』において「留置」は、分類項目名「捕縛・釈放」(表 4 通番 86) に分類された。実践医療用語における「留置 (indwelling)」は「体内の臓器や管腔に挿入・内在された状態」を意味し、「人を一定の場所に拘束する」という意味では用いない。また「鎮静」は分類項目「安心・焦燥・満足」となっているが (表 4 通番 50)、「落ち着いて静かな状態」という意味としてではなく、ほとんどが「鎮静剤等の投与により鎮め落ち着かせること」の意味で用いる。従って、これらについては『分類語彙表』の分類項目名をそのまま利用することは適切ではない。

さらに、末尾語が語全体の品詞および意味を決定する主要部とならない例も見つかった。「寛解

導入」の末尾語「導入」の分類項目名は「入り・入れ」(表4通番28)であるが、「寛解導入」は、「寛解導入療法(induction chemotherapy)」の略語として使われることから、省略された「療法」が主要部であり、分類項目名「方法」(表4通番62)となる。

以上全108グループを見直したところ、細分類または併合の検討を要するものが27グループあった。

「専門語」すなわち『分類語彙表』に未収録の異なり語36語については、「亢進」「増悪」「著減」「帰院」など、『分類語彙表』にある分類項目名で分類できるものと、「血培」「自壊」「腹満」「体動」など、新たに分類項目名を設定すべきものが混在していた。

6. まとめ

今回、医療記録情報に含まれる末尾語にサ変接続名詞を持つ漢字連続1,185語を対象に、『分類語彙表』を用いた語彙分類を行った結果、1,103語(93.1%)が108グループに分類でき、そのうち、細分類、併合、意味的に再分類を要するものが27グループ(25%)あった。

今回の調査で約7割が適切に分類されたことから、医学およびその関連領域の学術用語を対象とした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)」のカテゴリー分類では語彙分類が困難であった、医療行為や看護行為などの動作・行為を表す用語の語彙分類に本手法が有用であることが示唆された。

また、「固定・傾き・転倒」など括りが大きく、細分類を要する分類項目については、分類項目名を「固定」「傾き」「転倒」等と『分類語彙表』の分類名に準拠して分割することで、『分類語彙表』との対応が可能となる。

一方、意味的に再分類を要する一因として、『分類語彙表』が日常用語のシソーラスであり、専門語への対応が十分になされていないこと、また、医療記録情報において略語が使われている場合に、適切な意味の対応付けが行えなかったことなどが考えられる。

『分類語彙表』の分類方式においても、Mesh(Medical Subject Headings)のカテゴリー分類においても、一度決められた分類方式が大幅に変更されることはない。我々も決定した分類方式を安易に変更すべきではないと考えている。

そこで、分類項目名「医療」に括られた末尾語、および『分類語彙表』の見出し語に一致しなかった『専門語』の分類基準および項目名については、『分類語彙表』の分類方法に倣いつつ、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

ソーラスであるMeshカテゴリー分類名、および看護学術用語を対象としたCINAHL(Cumulative Index to Nursing and Allied Health Literature)の分類体系等を参考に、慎重に検討を重ね分類項目名を決定する予定である。

謝辞

本研究は2015年度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) 九州医事研究会の資料ブログ、
<<https://qmir.wordpress.com/2016/07/22/ehr-2016/>>
(参照2016年10月25日)
- 2) ComeJisyo,
<<http://sourceforge.jp/projects/comedic/>>
(参照2016年9月19日)
- 3) 相良かおる, 小野正子, 上野恵子: 医療用語のシソーラス作成に向けた予備的調査, 西南女学院大学紀要 Vol.19, p.109-118, 2015
- 4) 『現代日本語書き言葉均衡コーパス(BCCWJ)』,
<http://pj.ninjal.ac.jp/corpus_center/bccwj/>
(参照2016年9月19日)
- 5) 山崎誠, 相良かおる: 医療経過記録における名詞連続の計量的特徴, 第7回コーパス日本語学ワークショップ予稿集, pp.303-308, 2015
- 6) 石井正彦: 現代日本語の複合語形成論, ひつじ書房, 2007
- 7) 樺島忠夫: 日本語のスタイルブック, 大修館書店, 1979
- 8) 山崎誠・相良かおる: 医療経過記録における漢字連続複合語の計量的分析, 人文科学とコンピュータシンポジウム論文集, pp.221-226, 2014
- 9) 林四郎: 臨時一語の構造, 『国語学』131, pp.15-26, 1982
- 10) 石井正彦: 臨時一語と文章の凝縮, 『国語学』173, pp.104-91, 1993
- 11) 沖森卓也, 木村義之, 田中牧郎, 陣力衛, 前田直子: 図解 日本の語彙, 三省堂, p.16-25, 2011
- 12) 林大: 分類語彙表, 国立国語研究所, 秀英出版, 1964
- 13) 鈴木良太 編: 言語化学の百科事典, 丸善, 2006
- 14) 国立国語研究所 編: 分類語彙表 増補改訂版, 大日本図書, 2004