

実践医療用語の語構成要素

— 意味を基準とした分割 —

相良かおる¹⁾ 山崎誠²⁾ 麻子軒³⁾ 東条佳奈⁴⁾ 小野正子¹⁾ 内山清子⁵⁾

1) 西南女学院大学 2) 国立国語研究所 3) 大阪大学大学院文学研究科
4) 目白大学 5) 湘南工科大学

個人情報を含む医療記録に含まれる合成語の語構成は明らかになってはいない。一方、電子化された医療記録の自然言語処理を支援するため、医療記録から用語を収集し公開している分ち書き用辞書 ComeJisyo の登録語数は 111,664 語となり、言語資源としての利用が可能となった。

本研究では本辞書の登録語から一般的な語を含む合成語 7,194 語を選定し、これらの語構成を明らかにする。具体的には、合成語 7,194 語は意味的に適切な単位（語構成要素という）分割し、得られた語構成要素に意味的ラベルを付与する。

その結果、合成語は 16,451 要素（異なりでは 3,230 要素）に分割され、101 種類の意味ラベルによって分類された。合成語の語構成要素列において語頭に多く出現する意味ラベルは「身体部位 (37%)」次いで「状態 (32%)」、語末に多く出現する意味ラベルは「疾患 (56%)」次いで「症状 (9%)」であった。

Semantic analysis of medical compound words

Kaoru Sagara¹⁾ Makoto Yamazaki²⁾ Ma Tzu-Hsuan³⁾ Kana Tojo⁴⁾
Masako Ono¹⁾ Kiyoko Uchiyama⁵⁾

1) Seinan Jo Gakuin University 2) National Institute for Japanese Language and Linguistics
3) Osaka University 4) Mejiro University 5) Shonan Institute of Technology

The word structure of compound words in medical records containing sensitive personal information has not been clarified. On the other hand, Come-JisyoSjis-1 a dictionary for word segmentation of electronic medical records, has 111,664 words collected from medical records and can be used as a language resource.

In this study, we select 7,194 compound words including general words from words in Come-JisyoSjis-1 and clarify their word structure. Specifically, 7,194 compound words are divided into 16,451 semantically appropriate word components (the number of word types is 3,230). These 3,230 word components can be classified by 101 kinds of semantic labels.

As a result, the followings are found. The semantic label “body part (37%)” is the most frequently used as the first component of compound words, and then “state (32%)”. The semantic label “disease (56%)” and “symptom (9%)” are frequently appeared in the end of compound words.

1. はじめに

2001年、厚生労働省は、全国400床以上の病院の6割に電子カルテシステムを導入するという具体的な目標を掲げた。

医療記録文の分ち書きを支援するために2008年より無償で公開したComeJisyoV1(登録語数約3万語)は、随時更新を続け2013年11月にはComeJisyoV5-1(登録語数77,760語)を、2018年11月にはUTF版ComeJisyoUtf8-1(登録語数75,831語)を、2019年4月には登録語数111,664語のComeJisyoSjis-1を公開している^[1]。

ComeJisyoV1の作成に着手した2004年には、

標準化された日本語の医療用語は公開されておらず、医療記録に含まれる医療用語(以下、「実践医療用語」)の語構成の実態は不明であった。そのため語の単位認定の規則を定めず、看護師、管理栄養士、医師としての臨床経験を持つ者が、「ひとまとまりの語」としたものを登録している。従って本辞書には「急性びまん性管内増殖性糸球体腎炎」などの合成語が多く含まれている。

筆者らは本辞書を実践医療用語の言語資源と捉え、2018年度より、ComeJisyoSjis-1の合成語を対象に語構成の分析に着手している。

本研究の第一目標は、医療記録に含まれる合成

語を構成する意味的、統語的に妥当かつ実用的な「語構成要素」を収集し、これらに医療現場および領域で利用する際に有益な情報を付与した「実践医療用語分類語彙表」を作成することである。

第二の目標は、実践医療用語の語構成に関する言語学的な知見を得ることである。筆者らは先行研究並びに既存の言語資源を調べ、利用可能なものを活用し分析方針を定め、実践医療用語の分析と修正の過程の中で知見を得たいと考えている。

第三の目標は、得られた知見を含む成果物を医療実践、医療教育の領域で、出来得れば言語研究の領域においても利用可能な形で公開することである。

本稿では、このうち第二の目標に関して、ComeJisyoSjis-1に登録されている合成語 7,194 語を語構成要素に分割し意味ラベルを付与した結果について述べる。

以降、第 2 章では言語資源、すなわち本稿で対象とする合成語の選定について述べる。第 3 章の 3.1 節では予備実験について、3.2 節では本研究における「語構成要素」の定義、3.3 節では合成語から語構成要素列を求める方法、3.4 節では「語構成要素」に付与する意味ラベルについて述べる。第 4 章では、今回の分類手順について述べる。第 5 章では、分類した結果について述べる。5.1 節で語構成要素について、5.2 節で意味ラベルについて、5.3 節で意味的パターン（意味ラベル列）について述べる。第 6 章では、分類結果についての考察を述べ、最後に第 7 章でまとめる。

2. 言語資源

本研究で対象とする合成語(実践医療用語)は、以下の手順で選定した 7,194 語である。

(1) 汎用性の高い合成語: 31,162 語

2018 年 10 月当時、既に公開されていた ComeJisyoV5-1 (登録語数 77,760 語) と、公開予定 (2019 年 4 月公開済み) の ComeJisyoSjis-1 (登録語数 111,664 語) は、4 医療施設の医療記録より抽出した語を登録している。従って登録語には方言や医療施設特有の業界用語、略語、そして誤字も含まれている。そこで汎用性を確保するために、Web 上で公開されている辞書など研究用に収集した医療用語データと本辞書の登録語を照合し、一致した 31,162 語を抽出した。

(2) 合成語に含まれる一般的な用語: 231 語

- (1)により求めた合成語を MeCab 0.996^[2]と Unidic-cwj-2.2.0^[3]を用いて自動形態素解析を施し、以下の(a)~(d)に沿って語を抽出した。
- (a) 「名詞」または「形状詞」となる単一語
 - (b) 二語以上の語で「名詞」+「接尾辞」、「形状詞+接尾辞」、「接頭辞」+「名詞」、漢字 2 字の「名詞」+「名詞」、「記号」の

みの組み合わせであるもの

(c) 上記以外で文字数 3 文字以下のもの

(d) 自動形態素解析の結果、「感動詞」と誤解析された語を、目視で確認し妥当と認めたもの更にこれら(a)~(d)の規則に従ったものから「アルファベット」、「ひらがな」、「カタカナ」のみで表記された語を除いた 3,728 語について、2018 年 11 月公開予定の ComeJisyoUtf8-1 の登録語 75,831 語における出現頻度を求め、頻度の降順上位 25%の 768 語を抽出した。

専門外の筆者らが分析を行う上で、意味的な語の切れ目の判断を容易にすること、および、『分類語彙表一増補改訂版一』(以下、『分類語彙表』)^[4]の分類項目の活用可能性を調べるために 768 語と『分類語彙表』の見出し語を照合し、一致した 231 語を語構成要素候補とした。

(3) 一般的な語を含む合成語: 7,194 語

(1)の合成語 31,162 語の内、231 語を含む 7,194 語を本稿の対象合成語とした。

3. 語構成要素と意味ラベル

3.1 予備実験

語構成要素を合成語より切り出し、意味分類する方法について、以下の提案が出された。

① 『分類語彙表』より構成要素に付与し得る意味ラベルを洗い出す。

② 構成要素を以下の 6 タイプに分類した後に各構成要素に意味ラベルを付与する。

A: 一般語で医療分野では単一で一般語としては使われない

B: 一般語だが医療分野では合成語内でしか現れず、専門的な意味を持つ

C: 一般語だが、医療分野では単一で専門的な意味を持つ場合がある

D: 一般語だが医療分野では合成語になった際に専門的な意味を持つ場合がある

E: 医療分野でしか使われない

F: 一般語であり、一般的な意味でしか使われない

なお、ここでは『分類語彙表』に記載されている語を「一般語」という。

予備実験 1: 一般語と専門用語の判別

提案①②を検証するため、231 語について上記 6 分類を試み、分類項目の妥当性を調べた。

その結果、231 語中一般的な意味でしか使われない F と判断した語は 55 語であり、『分類語彙表』に掲載された語には「一般語」以外の語が含まれること、医療の知識を持たない筆者らにとって 6 分類が容易ではないことがわかった。また、分類項目が適切でない語が 59 語あり、残り 172 語についても複数の項目が付与され、ラベルを 1 つにする必要があることがわかった。そこで「分類

語彙表』に掲載の 231 語を一般的な語とし、これらに付与された分類項目を基に意味ラベルを付与する」という提案は見送ることとした。

予備実験 2：文法的分類と意味的分類

先行研究を調べ、語構成要素の分割方法として、①野村(1977)の複合語の類型⁵⁾と②石井(2007)の意味分類カテゴリー(以下、「意味ラベル」)⁶⁾の弁別法が候補となった。

そこで、筆者らが興味を持つ語を含む対象合成語を①または②の方法で語構成要素に分割し検討した結果、野村(1977)の文法的分類ではなく、石井(2007)の意味カテゴリーによる分類を採用することとした。

予備実験 3：合成語 1,878 語の分割

語構成要素への分割に際し、①接尾辞を切り離すか否か、②重要な意味を担う「高」「低」「無」「半」などは分割するか否かなどが問題となった。そこで、各自が興味を持つ合成語 1,010 語を語構成要素に分割し、分析者全員で分割方法および意味ラベルについて見直し、意味ラベル表を作成し、分析結果を修正した。これら 1,010 語を分割して得られた語構成要素で最も頻度の高い「先天性」についての考察および石井(2007)の意味カテゴリーを用いた分類については東条他(2019)にまとめられている⁷⁾。

次に対象合成語より乱数を用いてランダムに抽出した 1,000 語について分担して語構成要素に分割し意味ラベルを付与した。分割結果の詳細は山崎他(2019)にまとめられている⁸⁾。

なお、語構成要素への分割手順は、第 4 章の手順 1 から手順 3 に従う。

3.2 語構成要素の定義

当初、本研究の語構成要素を「医療用語の意味を基準とした最小単位を『語構成要素』とする」と定義したが、予備実験 3 の結果を受け、「最小の単位」とは限らずに、現段階における「語構成要素」の定義を以下のように改訂した。

「医療の観点から意味的に分割可能な語をすべて語構成要素とする。」

以下に例を示す。

合成語：先天性中枢神経異常

語構成要素：

[先天性][中枢神経][異常][中枢][神経]
[中枢神経異常]

合成語：気管支肺胞洗浄異常

語構成要素：

[気管支肺胞洗浄][異常][気管支]
[肺胞][洗浄]

合成語：性染色体異常

語構成要素：[性染色体][異常][染色体]

3.3 分割規則

予備実験 3 の結果より、合成語を語構成要素に切り分ける規則を以下のように定めた。

- (a) 接辞「前(ぜん)」「後(ご)」「上(じょう)」「下(げ・か)」「両(りょう)」「片(かた・へん)」などは切り離さない。
- (b) 「非」「不」「無」「反」「半」などの接頭辞は、係る範囲までを 1 要素とする。
- (c) 「腎癌」「肝癌」は、「[身体部位][疾患]」となるが、「腎|癌」「肝|癌」と 2 要素に分割するのは不自然である。そこで「胃癌」「大腸癌」「幹細胞癌」などの「[身体部位]・癌」を 1 要素とする。

3.4 意味ラベル

語構成要素に付与する意味ラベルは、石井(2007)の 22 種の意味カテゴリーを参考に必要なものを追加することとし、接尾辞「性」の付く語構成要素の意味ラベルはすべて[状態]とする。また、一つの要素に複数の意味ラベルを付与する場合、同位であれば「・」で[医療行為・状態]のように列挙し、階層構造を持つ場合は上位のラベルを左側にして「>」を用いて[行為>サービス]のように列挙する。

なお、予備実験 3 終了時点で蓄積された意味ラベルは 86 種、これらの内複数のラベルが列挙されたのは 21 種である。

例：「充血 [疾患・状態]
「破裂 [動き・状態]

以下に、参考にした石井(2007)の意味カテゴリーを示す。

表 1 石井(2007:180) 意味カテゴリー

自然物	動植物	物品	食品
道具	薬品	力	人間
機械	衣料	部分	家具
資材	地類	容器	建物
空間	形状	数量	動き
状態	時間		

4. 方法

対象合成語を語構成要素に分割し、意味ラベルを付与する作業手順を以下に示す。

手順1 231 語をキーとして合成語を昇順に並べ分担者に割り当てる。分担者は同じキーを含む合成語を手順 2 に従い処理し、処理した分類データファイル(以下、分類データ)を共有のフォルダに提出する。

[キー] [対象合成語]
吸引 羊水大量吸引症候群
吸引 有機溶剤吸引

拒絶 慢性拒絶反応
 拒絶 拒絶反応
 強迫 強迫神経症
 強迫 強迫観念人格障害

手順2 分類データの合成語を語構成要素のパターン（以下、語構成要素列）に分割し、各語構成要素に意味ラベルを付与し、語構成要素列に対応した意味ラベルのパターン（以下、意味ラベル列）を求める。

合成語：滑車神経損傷
 構成要素列：滑車 | 神経 | 損傷
 意味ラベル：滑車 [物品]
 神経 [身体部位]
 損傷 [疾患]
 意味ラベル列：物品 | 身体部位 | 疾患

合成語：切断四肢再接合術
 構成要素列：切断 | 四肢 | 再接合術
 意味ラベル：切断 [医療行為・作用]
 四肢 [身体部位]
 再接合術 [医療行為]
 意味ラベル列：医療行為・作用 | 身体部位 | 医療行為

手順3 分類データを一つのファイルに統合し、初めに①語構成要素列を見直し（例：「滑車 | 神経 | 損傷」→「滑車神経 | 損傷」）、②語構成要素に付与された意味ラベルを見直す。次いで③語構成要素と意味ラベルの対応表を作成し（表 2）、④作成した対応表を基に意味ラベル列を更新する。

なお、予備実験 3 の結果作成した対応表に収録された語構成要素数は 1,754 種、意味ラベルは 86 種である。

表 2 語構成要素と意味ラベルの対応表

【語構成要素】	【意味ラベル】
突発性	状態
湿 疹	症状
湿疹状	形状
湿疹性	状態
湿疹様	症状
再破裂	動き・状態

手順4 合成語 7,194 語の内、予備実験 3 の 1,010 語とランダムに抽出した 1,000 語（重複 132 語）の 1,878 語を除いた、5,316 語について、手順 1 に従って分担し、予備実験 3 終了時に作成した対応表（語構成要素 1,754 要素、意味ラベル 86 種）を基に手順 2 に従って語構成要素に分割後、意味ラベルを付与する。

手順5 提出された分類データを統合し、手順 3 と同様に、①語構成要素列を見直し、②意味ラベル対応表を見直し、③意味ラベル列

の更新を行う。（対応表の語構成要素 3,453 要素、意味ラベル 176 種に更新）

手順6 予備実験 3 の分類データと手順 5 の分類データを統合し、全対象合成語 7,194 語の語構成要素に付与された意味ラベルを見直した上で①意味ラベル対応表を更新（語構成要素 4,106 要素、意味ラベル 184 種）し、②意味ラベル列の更新を行う。次に③語構成要素列の語末の語構成要素をキーとして語構成語を昇順に並べ替え、確認用データを作成し、分担して確認作業を行う。

【語末】	【語構成要素列】
発疹	突発性 発疹
発疹	ウイルス性 発疹
発疹	ボストン発疹
発疹	湿疹様発疹
発疹	風疹様発疹

手順7 確認データを統合し、語構成要素の定義、意味ラベルの設定および表記法について見直し、手順 3 と同様に修正し合成語 7,194 語の語構成要素列、意味ラベル列、意味ラベル対応表を作成する。

5. 結果

5.1 語構成要素

合成語 7,194 語を分割して得られた語構成要素の総数（延べ）は 16,451 要素、種類（異なり）は 3,230 要素、合成語 1 語あたりの語構成要素数の平均は、2.6 要素、中央値は 2 要素であった。なお、Unidic-cwj-2.2.0 により自動形態素解析を施した結果は 26,782 短単位、1 合成語当りの平均および中央値は共に 4 短単位であった。

第 4 章の手順 6 での語構成要素数 4,106 要素に比べ、876 要素減少しているのは、①「手 | 関節」「食道 | 静脈」のように接続する語構成要素に付与された意味ラベルが[身体部位][身体部位]となる場合、「手関節」「食道静脈」と 1 要素にまとめたこと、②「膵臓 | 癌」などを「肝臓」に合わせて 1 要素にまとめたこと、また③「胸管 | 内頸静脈（胸管静脈と内頸静脈）」を「胸管内頸静脈」と 1 要素にしたこと、そして「内側 | 足底 | 神経 | 損傷」を「内側足底神経 | 損傷」とし、「内側足底神経[身体部位]」を 1 要素にしたことなどに依る。

表 3 は合成語 1 語あたりの語構成要素数をまとめたものである。前述の通り接続する意味ラベル[身体部位]を 1 要素にまとめたこと、接辞を含む「非糖尿病性」などを 1 語構成要素としたことから、合成語の約 7 割が、語構成要素数 1 または 2 要素となった。

表 3 合成語 1 語あたりの語構成要素数

要素数	合成語数	
1	735	10%
2	4,141	58%
3	1,913	27%
4	342	5%
5	51	1%
6	12	0%
総計	7,194	

表 4 は、頻度 50 以上の語構成要素 25 要素の一覧である。これら 25 要素中 15 要素の意味ラベルは「状態」であり、8 要素は「身体部位」であった。

表 4 語構成要素 (頻度 50 以上)

語構成要素	意味ラベル	頻度	割合
1	先天性	状態	1,040 6%
2	多発性	状態	281 2%
3	急性	状態	243 1%
4	慢性	状態	204 1%
5	新生児	人間	165 1%
6	外傷性	状態	147 1%
7	結核性	状態	139 1%
8	遺伝性	状態	111 1%
9	一過性	状態	106 1%
10	後天性	状態	100 1%
11	耳介	身体部位	95 1%
12	中毒性	状態	93 1%
13	転移性	状態	81 0%
14	大腿骨	身体部位	70 0%
15	顔面	身体部位	66 0%
16	妊娠性	状態	66 0%
17	皮膚	身体部位	65 0%
18	特発性	状態	61 0%
19	卵管	身体部位	56 0%
20	胃	身体部位	53 0%
21	感染性	状態	53 0%
22	再発性	状態	52 0%
23	術後	時間	51 0%
24	脊髄	身体部位	51 0%
25	角膜	身体部位	50 0%
種類	小計	3,499	21%
3,230 種	総計	16,451	100%

表 5 は語頭に頻度 50 以上で出現する語構成要素の一覧である。意味ラベル[状態]が付与された「～性」の要素が 7 要素中 6 要素を占めた。頻度 50 以上の語構成要素数が 7 要素であること、全体に占める割合が 11%であることから、語頭には多様な要素が出現すると推測される。

一方、表 6 は語末に頻度 50 以上出現する語構成要素である。意味ラベル[疾患]が付与された要素が 21 要素中 15 要素であった。

表 5 語頭にくる語構成要素 (頻度 50 以上)

語構成要素	意味ラベル	頻度	割合
1	先天性	状態	370 5%
2	多発性	状態	108 2%
3	急性	状態	80 1%
4	慢性	状態	68 1%
5	結核性	状態	62 1%
6	新生児	人間	56 1%
7	外傷性	状態	52 1%
	小計	796	11%
	総計	7,194	100%

表 6 語末にくる語構成要素 (頻度 50 以上)

語構成要素	意味ラベル	頻度	割合
1	腫瘍	疾患	434 6%
2	損傷	疾患	407 6%
3	出血	症状	181 3%
4	手術	医療行為	174 2%
5	障害	障害	146 2%
6	中毒	疾患	137 2%
7	挫傷	疾患	135 2%
8	麻痺	疾患・状態	133 2%
9	狭窄症	疾患	106 1%
10	貧血	症状	99 1%
11	感染症	疾患	94 1%
12	脱臼	疾患	90 1%
13	皮膚炎	疾患	79 1%
14	熱傷	疾患	74 1%
15	捻挫	疾患	66 1%
16	狭窄	状態	66 1%
17	結核	疾患	66 1%
18	後遺症	疾患	64 1%
19	髄膜炎	疾患	56 1%
20	破裂	動き・状態	55 1%
21	腎炎	疾患	52 1%
種類	小計	2,714	38%
1,765 種	総計	7,194	100%

5.2 意味ラベル

語構成要素 3,230 要素に付加した意味ラベルは第 4 章の手順 6 での意味ラベル 184 種から 83 種減少し、101 種類であった。これは[時間]と[時期]を[時間]に、[薬品]と[薬剤]を[薬品]など大きな概念にまとめたこと、また[身体部位]を解剖学的な名称だけでなく身体における部分や位置を表すラベルとし、4 種あった意味ラベル[身体部位]、[身体部位・位置]、[身体部位・空間]、[身体部位・組織]を[身体部位]と[身体部位・患部]の 2 種類にまとめるなど、複数の意味ラベルを持つ意味ラベルを減らしたことに依る。

表 7 は頻度 50 以上 (中央値は 8 語) の意味ラベル 17 種の一覧である。これら 17 種で全体の 94%を占め、「身体部位 (37%)」「状態 (35%)」で全体の 7 割以上を占めていた。なお表 7 右列

の“※”印は、石井 (2007) が立てた 22 種の意味分類にある意味ラベルであり、全 101 ラベル中 22 ラベルが一致し、残り 79 ラベル (78%) は新たに設定したものである。

表 7 意味ラベル (頻度 50 以上)

意味ラベル	頻度	割合
1 身体部位	6,096	37%
2 状態※	5,766	35%
3 疾患	648	4%
4 時間※	528	3%
5 物質	490	3%
6 人間※	309	2%
7 種類	237	1%
8 動植物	217	1%
9 空間※	208	1%
10 行為	194	1%
11 生理	164	1%
12 医療行為	134	1%
13 形状	134	1%
14 症状	130	1%
15 薬品※	106	1%
16 動き	98	1%
17 方法	58	0%
小計	15,517	94%
総計	16,451	100%

表 8 語頭にくる意味ラベル (頻度 50 以上)

意味ラベル	頻度	割合
1 身体部位	2,692	37%
2 状態	2,298	32%
3 疾患	418	6%
4 時間	218	3%
5 物質	216	3%
6 人間	116	2%
7 動植物	103	1%
8 行為	102	1%
9 症状	93	1%
10 種類	87	1%
11 生理	79	1%
12 医療行為	78	1%
13 空間	76	1%
14 形状	54	1%
15 薬品	51	1%
16 動き	50	1%
小計	6,731	94%
総計	7,194	100%

表 8 は語頭に出現する頻度 50 以上の意味ラベル 16 種の一覧である。これらで全体の 94% を占めていた。

表 9 は語末に頻度 50 以上で出現する意味ラベル 12 種である。これらで全体の 93% を占めていた。

表 9 語末にくる意味ラベル (頻度 50 以上)

意味ラベル	頻度	割合
1 疾患	3,999	56%
2 症状	644	9%
3 医療行為	555	8%
4 状態	513	7%
5 疾患・状態	210	3%
6 障害	166	2%
7 動き・状態	136	2%
8 身体部位	128	2%
9 行為	118	2%
10 生理	98	1%
11 動き	80	1%
12 症状・疾患	52	1%
小計	6,699	93%
総計	7,194	100%

5.3 意味ラベル列

合成語 7,194 語の意味ラベル列は 1,086 種類あった。

表 10 は、頻度が 50 以上の 19 種類の意味ラベル列をまとめたものである。これらで全体の 58% を占めており、「[身体部位][疾患]」と「[状態][疾患]」で 29% を占めていた。

表 10 意味ラベル列 (頻度 50 以上)

意味ラベル列	頻度	割合
1 身体部位 疾患	1,346	19%
2 状態 疾患	694	10%
3 状態 身体部位 疾患	297	4%
4 疾患	221	3%
5 身体部位 状態 疾患	193	3%
6 状態 症状	153	2%
7 状態 状態 疾患	143	2%
8 状態	137	2%
9 物質 疾患	127	2%
10 身体部位 症状	121	2%
11 身体部位 医療行為	115	2%
12 疾患 医療行為	99	1%
13 身体部位 疾患・状態	96	1%
14 動植物 疾患	85	1%
15 状態 身体部位 状態	84	1%
16 身体部位 状態	81	1%
17 身体部位	76	1%
18 身体部位 動き・状態	70	1%
19 症状	62	1%
小計	4,200	58%
総計	7,194	100%

6. 考察

本章では、第 1 章で示した第二の目標について、対象合成語 7,194 語を分析して得られた語構成要素の定義と分割規則、得られた語構成要素 3,230 要素について考察する。

6.1 語構成要素の定義と分割規則

当初、語構成要素を「意味を基準とした最小単位」と定め、3.1節の予備実験3では、類義語「大腿骨頸部内側骨折」と「大腿骨内側頸部骨折」を、「大腿骨 | 頸部 | 内側 | 骨折 ([身体部位][身体部位][位置][疾患])」と「大腿骨 | 内側 | 頸部 | 骨折 ([身体部位][位置][身体部位][疾患])」というように異なる意味ラベル列に分類した。

並行して接辞「前」「後」「左」「右」「上」「下」などを切り離すか否かの判断のために、第2章(2)の3,728語の内、これらの接辞を含む308語を調べ、「前額骨」に対し「後額骨」はない、「後天性」「先天性」に対し「前天性」はない、「上皮腫」「中皮腫」に対し「下皮腫」はないなど、「前・後」「左・右」「上・下」「内・外」で、対になっていないものがあることを確認した。また、[身体部位]の「前庭部」に対して「後庭部」はないが、一般的な「庭(にわ)」の意味では「後庭部」があることを確認した。加えて予備実験3の結果を検討し、分割規則は、生成規則にも成り得ることも踏まえ、本研究では、[位置]を示す接辞は切り離さず「右脚」「前腕部」を[身体部位]の1要素とすることにした。

意味ラベル[身体部位]を人体における位置・空間の意味を含めた広い概念としたことで、「大腿骨頸部内側骨折」と「大腿骨内側頸部骨折」は、「大腿骨 | 頸部内側 | 骨折」と「大腿骨内側 | 頸部 | 骨折」の3要素に分割され、意味ラベル列は共に[身体部位][身体部位][疾患]となった。

そこで当初の定義「医療用語の意味を基準とした最小単位を『語構成要素』とする」を3.2節の定義に改訂した。

更に手順7の結果を見直し、連続する意味ラベル[身体部位][身体部位]を一つの[身体部位]にまとめることとした。従って「大腿骨頸部内側骨折」と「大腿骨内側頸部骨折」は、「大腿骨頸部内側 | 骨折」「大腿骨内側頸部 | 骨折」と2要素に分割され、統語的にもまとまりのある単位に分割することが可能となった。また「十二指腸空腸吻合術(十二指腸と空腸を繋ぎ合わせる手術)」は「十二指腸空腸 | 吻合術 ([身体部位][医療行為])」となり、「十二指腸と空腸を」という意味での「十二指腸空腸」が1要素となり、「端々 | 吻合術 (端と端を繋ぎ合わせる手術)」との整合性を保つことが可能となった。

「胸管静脈・内頸静脈」を略した表現である「胸管内頸静脈」も、[身体部位]の1要素となることから、「胸管内頸静脈吻合術 (胸管静脈と内頸静脈を繋ぎ合わせる手術)」は「胸管内頸静脈 | 吻合術」と分割され、意味的にも統語的にも妥当な要素に切り出すことが可能となった。

このように意味ラベル[身体部位]を付与する範囲を広げ、また連続する[身体部位]を一つにまとめることで、得られる語構成要素は統語的にも妥当なものとなった。

手順7の見直しの中で、「要素A+要素B」の意味が「要素A+要素B」とは別の意味Cになる合成語「気管支肺胞洗浄異常」が見つかった。

(a) 3.3節の分割規則による分割

構成要素列：気管支 | 肺胞 | 洗浄 | 異常
意味ラベル：気管支[身体部位] | 肺胞[身体部位] | 洗浄[行為] | 異常[状態]

(b) 医療の観点による分割

構成要素列：気管支肺胞洗浄 | 異常
意味ラベル：気管支肺胞洗浄[検査法] | 異常[状態]

『南山堂医学大辞典』第20版によれば、「気管支肺胞洗浄」とは「呼吸器疾患の診断・治療のため、気管支鏡により気管支肺胞領域を洗浄し、細胞成分・液性成分を採取する手技」を意味する。

しかしながら専門外の分析者がこのような医療領域特有の意味を持つ語を判別することは困難である。今回、1人当たりが担当する対象合成語は、分割作用と確認作業を併せて約2,500語であり、筆者らの内4名は、医療用語に馴染みがない。従って、今回分割した語構成要素列の中には(a)のケースが含まれている可能性があり、今後医療従事者による確認を行う予定である。

同様に、医療の観点による意味的な分割位置を複数持つ「乳汁分泌抑制」が見つかった。

(c) 3.3節の分割規則による分割

構成要素列：乳汁 | 分泌 | 抑制
意味ラベル：乳汁[分泌物] | 分泌[生理] | 抑制[行為]

(d) 医療の観点による分割 (筆者案)

構成要素列：乳汁 | 分泌抑制
意味ラベル：乳汁[分泌物] | 分泌抑制[生理]

(e) 医療の観点による分割 (助産師)

構成要素列：乳汁分泌 | 抑制
意味ラベル：乳汁分泌[生理] | 抑制[行為]

(d)の分割の根拠は、「分泌抑制」を『南山堂医学大辞典』第20版で部分一致検索した以下の結果に依る。

成長ホルモン	分泌抑制	ホルモン
	胃酸	分泌抑制
	酸	分泌抑制
		因子
		薬

一方、助産師に尋ねたところ(e)の回答を得た。同様に複数の分割が考えられるものに「焼身自殺未遂」がある。

(f) 3.3節の分割規則による分割

構成要素列：焼身 | 自殺 | 未遂
意味ラベル：焼身[行為] | 自殺[行為] |
未遂[状態]

(g) 医療の観点による分割①

構成要素列：焼身自殺 | 未遂
意味ラベル：焼身自殺[行為] | 未遂[状態]

(h) 医療の観点による分割②

構成要素列：焼身 | 自殺未遂
意味ラベル：焼身[行為] | 自殺未遂[行為]

語構成要素の定義を検討し、意味的に分割可能な全ての「語構成要素」に重複は認めないこととしたことから (3.2 節)、「構成要素列」では(e)と(g)により分割している。なお、意味ラベル対応表 (表 2) には、手順 7 の見直し前のデータ (a)(c)(f)の短く分割した語構成要素と、(h)「分泌抑制」、(g)「自殺未遂」の意味ラベルも収録しており、5,350 種の語構成要素を収集している。

6.2 意味ラベル

今回新たに設定した 79 種類の意味ラベルは、表 1 の石井 (2007) の意味カテゴリーに倣い、『分類語彙表』や市販の類語辞典などを参考に設定しており、一般的な馴染みやすいものとなっている。例えば「焼身自殺」の意味ラベルは[行為]としている。一方、MeSH (Medical Subject Headings) では、「suicide attempt (自殺未遂)」のカテゴリーは“Psychiatry and Psychology(精神医学および心理学)”であり、国際医療用語集 SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms) におけるカテゴリーは“event(事象)”である。

学術医療用語を対象としたシソーラスのカテゴリーをそのまま実践医療用語の意味分類に利用できないことがわかっており、現在、実践医療用語を対象としたシソーラスは見当たらない^[9]。

医療現場で利用する上で、今回得られた語構成要素を、医療現場での利用を考慮した、医療実践の焦点、判断、時間・頻度、位置、手段、行為、該当者というような流れに沿った枠組みを定め、意味分類する必要があると考えている。

6.3 実践医療用語特有の接辞

対象合成語を語構成要素に分割する過程で、以下に示す実践医療用語特有の接尾辞約 70 種類が得られた。

- ～症：血症、硬化症、指症、虫症、失調症
- ～炎：神経炎、軟骨炎、包炎、膜炎、骨炎
- ～瘍：腫瘍、潰瘍、膿瘍、腺腫瘍
- ～傷：挫傷、損傷、挫滅傷 など

6.4 語構成要素の統語構造

本語構成要素の定義では、「十二指腸空腸」と

「胸管内頸静脈」は、共に意味ラベル[身体部位]の 1 要素となるが、その結合パターンは異なる。

そこで、野村 (1987 : 138) ^[10]の表記法を参考に語構成要素の結合パターンを「十二指腸空腸」は「□□□□+□□」、「胸管内頸静脈」は「(□□・□□)+□□」のように表記している。

7. まとめ

本稿では、今まで解明されてこなかった機密性の高い医療記録に含まれる一般語を含む合成語 7,194 語の語構成を明らかにした。

今後は、第一の目標である「実践医療用語分類語彙表」の作成・公開に向け、語構成要素列の見直しに着手し、「気管支肺胞洗浄[検査法]」などの専門用語を洗い出し、語構成要素と意味ラベルの対応表を更新した上で「語構成要素語彙表試案」を作成する。次いで、一般用語を含まない合成語 23,968 語 (31,162 語-7,194 語) の語構成の分析に着手する予定である。

また、実践医療用語の言語学的な分析は着手したばかりであり、第二の目標については継続して行い、得られた知見は公開する予定である。

謝辞

本研究は、科学研究費補助金「語形成および意味的情報を付加した実践医療用語辞書の構築」(JP18H03499)の助成を受けています。

参考文献

- [1] 相良かおる, 小野正子: 実践医療用語辞書 ComeJisyoSjis-1 の作成, 言語処理学会第 25 回年次大会発表論文集, p.1491-1494, 2019.
- [2] MeCab: <https://taku910.github.io/mecab/> (参照 2019-10-16)
- [3] UniDic: <https://unidic.ninjal.ac.jp/> (参照 2019-10-16)
- [4] 国立国語研究所: 分類語彙表 増補改訂版, 大日本図書, 2004.
- [5] 野村雅昭: 造語法, 宮島達男他, 岩波講座日本語 9 『語彙と意味』, p247-282, 1977.
- [6] 石井正彦: 現代日本語の複合語形成論, ひつじ書房, p.180-197, 2007.
- [7] 東条佳奈, 相良かおる, 小野正子, 山崎誠: 実践医療用語における構成要素の意味分類試案-「先天性」を例に一, 現代日本語研究, 11, p.40-58, 大阪大学大学院文学研究科日本語学講座現代日本語学研究室, 2019.
- [8] 山崎誠, 相良かおる, 小野正子, 東条佳奈, 麻子軒: 実践医療用語の語構成要素への分割と意味ラベル付与の試み, 言語資源活用ワークショップ 2019, 国立国語研究所, 2019.
- [9] 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むけた予備調査, 西南女学院大学紀要, Vol.19, 2015.
- [10] 野村雅昭: 複合漢語の構造, 水谷静夫他, 朝倉日本語新講座1 『文字・表記と語構成』, 朝倉書店, p.130-144, 1987.