

国立国語研究所学術情報リポジトリ

A Quantitative Analysis of Words Included in Practical Medical Terms

| | |
|-------|---------------------------------------------------------------------------------------------------|
| メタデータ | 言語: jpn 出版者: 公開日: 2021-03-05 キーワード (Ja): キーワード (En): 作成者: 山崎, 誠 メールアドレス: 所属: |
| URL | https://doi.org/10.15084/00003156 |

実践医療用語を構成する語の計量的分析

山崎 誠（国立国語研究所研究系言語変化研究領域）[†]

A Quantitative Analysis of Words Included in Practical Medical Terms

Makoto Yamazaki (National Institute for Japanese Language and Linguistics)

要旨

本稿では、実践医療用語を語彙の量的な分布を通して概観するとともに、語構成的な特徴を明らかにする。利用するデータは、2020年5月に公開された ComeJisyoUtf8-2 に収録されている 114,957 語である。これを MeCab0.996 と UniDic-cwj-2.2.0 で解析し、『分類語彙表増補改訂版』の情報を付与したものを利用した。主な結果は、以下のとおり。(1) 1語あたりの短単位数は平均 2.67、最大値は 13 であった。(2) 中項目の意味分類では、全体では「量」「身体」「作用」「生命」「心」が上位 5 カテゴリーであるが、語頭・語中においては、「身体」がもっとも多く、語末では「生命」がいちばん多かった。(3) 分類項目では、全体では「病気・体調」がいちばん多かったが、語頭では、「膜・筋・神経・内臓」、語中では「性質」、語末では「病気・体調」が最多であった。(4) 国家試験（看護師、助産師、管理栄養士）や看護師及び管理栄養士養成校で使用されている教科書での出現状況においても意味分布に違いが見られた。(5) ComeJisyo を語の専門用語としての汎用性・一般性から類別し、特徴を見た結果、一般性が高くなると語が長くなるなどの違いが見られた。

キーワード：実践医療用語，ComeJisyo，語構成，短単位，『分類語彙表』

1. はじめに

実践医療用語とは、医師、看護師などが電子カルテを作成する際に使う用語で、医療・介護などの専門用語のほかに、医療現場での慣用的な用語や助詞が省略された臨時一語などが含まれている。実践医療用語を広範囲に収集したデータが ComeJisyo¹ という名で 2013 年より公開されている²（相良 2014）。

実践医療用語の実態を把握する研究が相良らの研究グループによって行われている（内山他 2018，相良 2014，相良他 2010，2012，2015，2016，2017，2019，東条他 2019 山崎他 2014，2019）。これらのうち、内山他（2018），相良他（2019），東条他（2019），山崎他（2019）は、実践医療用語の語構成に着目した研究であり、品詞や意味ラベルの分布による分析を行っている。これらの分析においては、ComeJisyo 全体を対象とするのではなく、そこから抽出した一部を対象としている。とくに、抽出においては、「方言や医療施設特有の語を含む合成語の排除」、『分類語彙表』に収録される語（一般的な語）を含む合成語データの抽出、「臨時一語」の認定要件（石井 2007）に該当しないもの 7,139 語を抽出（山崎他 2019: 162）のような手順が行われている。これらの手順によって抽出された語は、基本的な医療・介護用語である可能性が高く、医療現場で用いられる臨時一語などが

[†] yamazaki [AT] ninjal.ac.jp

¹ 読み方は「コメジショ」である。

² <https://ja.osdn.net/projects/comedic/> 左記サイトによると、ComeJisyo は、「医療記録の分かち書きを支援するための実践医療用語辞書」とされている。

含まれていないのではないかと推測される。そこで本稿では、語彙の選別をせず、ComeJisyo 全体を対象として分析した場合の語彙の実態を報告するものである。併せて、ComeJisyo を専門用語としての汎用性・一般性という観点から類別を行った際、それぞれの区分どのような特徴があるかを観察する。

2. 見出し語の類別

山崎他 (2019) では、以下の手順で ComeJisyo³から分析対象の見出し語を抽出している。以下、山崎他 (2019: 162) から再掲する。

(1) 方言や医療施設特有の語を含む合成語の排除 (汎用性の確保) :

公開予定の ComeJisyoSjis-1 の見出し語 111,664 語と研究用に収集した医療用語データ 52,974 語を照合し、一致する語 31,162 語 (以下、「合成語データ」という) を抽出

(2) 『分類語彙表』に収録されている語 (一般的な語) を含む合成語データの抽出 :

① 「合成語データ」を MeCab0.996 と UniDic-cwj-2.2.0 により短単位に分割

② 「臨時一語」の認定要件 (石井 2007) に該当しないもの 7,327 語を抽出

③ 更に、アルファベット、ひらがな、カタカナのみからなる語を削除した 3,728 語を抽出

④ 既に 2018 年 11 月公開の ComeJisyoUtf8-1 (75,831 語) における出現頻度を求め、頻度の多いもの上位 25% となる 768 語を抽出

⑤ 『分類語彙表』の項目と照合し、一致した 231 語を抽出

⑥ 合成語データ (31,162 語) より⑤の 231 語を含む合成語データ 7,139 語を抽出

(再掲ここまで)

以上を図示すると、図 1 のようになる。図 1 の「B.汎用」は、上記 (1) に対応し、「C.一般」は、上記 (2) に対応する。また、ComeJisyo 全体から「B.汎用」を除いた部分を「A.非汎用」と位置付ける。山崎他 (2019) では、「C.一般」の一部を扱った。実践医療用語の実践という特徴を表すと推測される「A.非汎用」の部分については、まだ詳細な分析はされていない。これらを踏まえて、以下のようなリサーチ・クエスチョンを設ける。

RQ1 : ComeJisyo 全体の語彙の統計的情報はどのようになっているか。

RQ2 : ComeJiSyo を専門用語としての汎用性・一般性という観点で類別したとき、それぞれの区分にどのような特徴が現れるか。

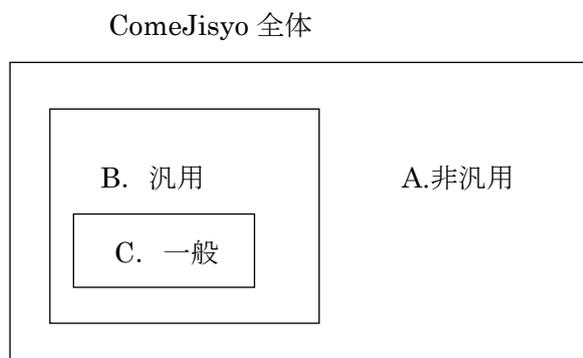


図 1 ComeJisyo の語彙の類別

³ 山崎他 (2019) で使っているデータは ComeJisyoSjis-1 であり、今回使用した ComeJisyoUtf8-2 と比べると収録語数が 3293 語少ない。

3. データ

本稿で使用したデータは、ComeJisyoUtf8-2 に収録されている 114,957 語である。そのほか、出現情報として該当する語に、「看」「助」「栄」「教」が付与されている。これらは、ComeJiSyο の概要によると、以下のような情報となっている。

「看」「助」「栄」は、看護師国家試験、助産師国家試験、そして管理栄養士国家試験 5 年分（2013 年～2018 年）の問題文に出現した用語であることを示しています。「教」は、看護師及び管理栄養士養成校で使われている教科書 57 冊の索引に使われている語であることを示しています。

4. 短単位への分割と分類語彙表番号の付与

上記 114,957 語を MeCab0.996 と UniDic-cwj-2.2.0 で解析し、『分類語彙表番号－UniDic 語彙素番号対応表』(wslp2unidic) を用いて、分類語彙表番号を付与した⁴。wslp2unidic は、分類語彙表番号から UniDic の語彙素 ID への対応表であるが、そこから、UniDic の語彙素から分類語彙表番号を逆引きできるようにして、分類語彙表番号を付与した。なお、処理の都合上、UniDic の品詞が補助記号となっている短単位には、0.0001 という分類語彙表番号を、対応表に現れなかったものには、9.9999 という番号を与えた。

5. 全体の傾向

まず、114,957 語について、短単位数、品詞、意味分類の 3 つの観点から概要を述べる。

5. 1 語の長さ（1 語あたりの短単位数）

語の長さは 1 語あたりの短単位数で計測する。図 2 に結果を示す。語の長さの平均は 2.67 であり、中央値は 2 である。最大値は「在宅/人工/呼吸/器/使用/特定/疾患/患者/緊急/時/一時/入院/事業⁵」の 13 であった。

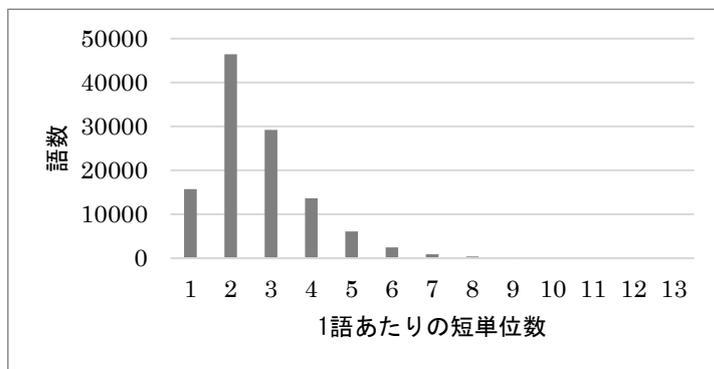


図 2 1 語あたりの短単位数の分布

5. 2 使用語彙の頻度

ComeJisyo 全体（114,957 語）を形態素解析した結果、延べ 306847 語、異なり 21559 語が得られた。表 1 に頻度上位 20 語を示す。最も多く使われたのは接尾辞の「性」である。2 番目の unk はいわゆる未知語であり、UniDic に登録されていない語であることを示す。内訳はカタカナ語や記号が多いようであるが詳細は分析していない。

⁴ wslp2unidic は、分類語彙表番号から UniDic の語彙素 ID への対応表であるが、そこから、UniDic の語彙素から分類語彙表番号を逆引きできる対応表を作成し、分類語彙表番号を付与した。

⁵ /は短単位の境界を示す。この語には「教」という出現情報が付与されていて、看護師及び管理栄養士養成校で使われている教科書 57 冊の索引に使われている語であることが分かる。

表1 頻度表（上位20語）

| Rank | 語彙素 | 語彙素読み | 品詞 | 語種 | 頻度 |
|------|-----|-------|---------------|-----|------|
| 1 | 性 | セイ | 接尾辞-名詞的-一般 | 漢 | 9739 |
| 2 | unk | unk | 名詞-普通名詞-一般 | unk | 9341 |
| 3 | 症 | ショウ | 接尾辞-名詞的-一般 | 漢 | 4846 |
| 4 | 炎 | エン | 接尾辞-名詞的-一般 | 漢 | 2293 |
| 5 | 的 | テキ | 接尾辞-形状詞的 | 漢 | 2154 |
| 6 | 部 | ブ | 名詞-普通名詞-助数詞可能 | 漢 | 1692 |
| 7 | 術 | ジュツ | 名詞-普通名詞-一般 | 漢 | 1509 |
| 8 | 神経 | シンケイ | 名詞-普通名詞-一般 | 漢 | 1509 |
| 9 | 法 | ハウ | 名詞-普通名詞-一般 | 漢 | 1221 |
| 10 | 動脈 | ドウミヤク | 名詞-普通名詞-一般 | 漢 | 1190 |
| 11 | 群 | グン | 名詞-普通名詞-一般 | 漢 | 1056 |
| 12 | 後 | ゴ | 接尾辞-名詞的-副詞可能 | 漢 | 1025 |
| 13 | 関節 | カンセツ | 名詞-普通名詞-一般 | 漢 | 1010 |
| 14 | 障害 | ショウガイ | 名詞-普通名詞-サ変可能 | 漢 | 1002 |
| 15 | 症候 | ショウコウ | 名詞-普通名詞-一般 | 漢 | 988 |
| 16 | 型 | ガタ | 接尾辞-名詞的-一般 | 和 | 980 |
| 17 | 体 | タイ | 名詞-普通名詞-助数詞可能 | 漢 | 944 |
| 18 | 細胞 | サイボウ | 名詞-普通名詞-一般 | 漢 | 936 |
| 19 | 骨折 | コッセツ | 名詞-普通名詞-サ変可能 | 漢 | 906 |
| 20 | 骨 | コツ | 接尾辞-名詞的-一般 | 漢 | 900 |

5.3 品詞・語種

表2は、品詞の割合である。名詞と接辞（接尾辞・接頭辞）で約96%を占める。

表2 品詞の頻度と割合

| 品詞 | 頻度 | 割合(%) |
|------|--------|--------|
| 名詞 | 243175 | 79.25 |
| 代名詞 | 38 | 0.01 |
| 動詞 | 661 | 0.22 |
| 形容詞 | 349 | 0.11 |
| 形状詞 | 1933 | 0.63 |
| 副詞 | 402 | 0.13 |
| 連体詞 | 10 | 0.00 |
| 接続詞 | 30 | 0.01 |
| 感動詞 | 231 | 0.08 |
| 接頭辞 | 8559 | 2.79 |
| 接尾辞 | 43089 | 14.04 |
| 助詞 | 527 | 0.17 |
| 助動詞 | 218 | 0.07 |
| 記号 | 6157 | 2.01 |
| 補助記号 | 1464 | 0.48 |
| 空白 | 4 | 0.00 |
| 合計 | 306847 | 100.00 |

表 3 は、語種の割合である。漢語と外来語とで全体の約 85% を占める。和語も約 6% 弱見られる。

表 3 語種の頻度と割合

| 語種 | 頻度 | 割合(%) |
|------|--------|--------|
| 和語 | 17592 | 5.73 |
| 漢語 | 231668 | 75.50 |
| 外来語 | 31451 | 10.25 |
| 混種語 | 920 | 0.30 |
| 記号 | 10109 | 3.29 |
| 固有名詞 | 5552 | 1.81 |
| unk | 9555 | 3.11 |
| 合計 | 306847 | 100.00 |

5. 4 意味分類

表 4~7 は、『分類語彙表増補改訂版』(以下、『分類語彙表』) の分類番号に基づく集計である。表 4 は品詞に当たる「類」、表 5 は「部門」、表 6 は「中項目」、表 7 は「分類項目」による集計である。

一つの語彙素に対して複数の分類語彙表番号が対応する場合があるため、意味分類の集計における合計は 411568 である。平均して 1 語に対して 1.34 個の意味分類が対応していることになる。なお、「記」は記号を、「未」は未知語に対する、それぞれ臨時的な分類である。「語頭」「語末」は短単位数が 2 以上の語を対象として、「語中」は短単位数が 3 以上の語の場合を対象としてそれぞれにカウントした。

表 4 「類」の割合 (全体および語中の位置別)

| 類 | 全体 | 語頭 | 語中 | 語末 |
|----|--------|--------|--------|--------|
| 体 | 75.87 | 70.20 | 78.89 | 78.06 |
| 用 | 0.51 | 0.82 | 0.72 | 0.29 |
| 相 | 4.35 | 7.16 | 4.52 | 1.89 |
| 他 | 0.17 | 0.37 | 0.15 | 0.05 |
| 記 | 0.36 | 0.02 | 1.03 | 0.07 |
| 未 | 18.75 | 21.43 | 14.69 | 19.64 |
| 合計 | 100.00 | 100.00 | 100.00 | 100.00 |

表 5 「部門」の割合 (全体および語中の位置別)

| 部門 | 全体 | 語頭 | 語中 | 語末 |
|--------------|--------|--------|--------|--------|
| 抽象的關係 | 34.52 | 31.30 | 40.60 | 32.17 |
| 人間活動の主体 | 3.63 | 3.11 | 4.28 | 3.56 |
| 人間活動—精神および行為 | 16.54 | 15.49 | 14.60 | 19.26 |
| 生産物および用具 | 4.51 | 4.11 | 3.65 | 5.61 |
| 自然物および自然現象 | 21.54 | 24.16 | 20.99 | 19.64 |
| 記 | 0.36 | 0.02 | 1.03 | 0.07 |
| * | 0.17 | 0.37 | 0.15 | 0.05 |
| 未 | 18.75 | 21.43 | 14.69 | 19.64 |
| 合計 | 100.00 | 100.00 | 100.00 | 100.00 |

表 4 から全体の約 75% が体の類 (名詞) であることが分かる。未知語となった 18.75% の中にもかなりの名詞が含まれていると推測されることから、全体の 9 割近くは体の類で

はないかと思われる。「体」の類は語頭より、語中、さらに語中より語末での割合が高い。それとは逆に、「相」の類は語末に向かって割合が低くなっている。また、表5から「抽象的關係」がもっとも多く、ついで「自然物および自然現象」、「人間活動—精神および行為」と続く。語中の位置では、「抽象的關係」は語中に多く、「人間活動—精神および行為」は語末に、「自然物および自然現象」は語頭に多いことがそれぞれ見て取れる。

表6 「中項目」の頻度と割合（「全体」の上位10、全体および語中の位置別）

| 順位 | 中項目 | 全体 | 語頭 | 語中 | 語末 |
|----|-----|-------|-------|-------|-------|
| 1 | 未 | 18.75 | 21.43 | 14.69 | 19.64 |
| 2 | 量 | 8.62 | 7.82 | 9.69 | 8.38 |
| 3 | 身体 | 8.42 | 12.77 | 10.67 | 2.76 |
| 4 | 作用 | 7.64 | 6.63 | 7.93 | 8.28 |
| 5 | 生命 | 7.49 | 4.68 | 4.93 | 11.97 |
| 6 | 心 | 6.44 | 5.87 | 5.40 | 7.84 |
| 7 | 様相 | 6.07 | 3.26 | 9.97 | 5.21 |
| 8 | 類 | 3.11 | 3.08 | 3.90 | 2.46 |
| 9 | 事業 | 3.05 | 2.70 | 2.78 | 3.59 |
| 10 | 時間 | 2.73 | 3.37 | 2.39 | 2.50 |

表6は、中項目の頻度（上位10）である。中項目は分類番号のピリオド以下2桁で集計している。したがって、「類」の区別はなされていない。1位は「未」で未知語であるので、これを除くといちばん多いのが「量」に関するもの、次いで「身体」「作用」と続く。上位には心身に関する意味領域が多い。語中の位置別で見ると、「身体」は語頭および語中に多く、語末には少ない。「生命」は語末に多く、語頭、語中は語末の半分程度の分布である。また、「様相」は語中に多い。

表7 「分類項目」の頻度と割合（「全体」の上位20、全体および語中の位置別）

| Rank | 分類項目 | 全体 | 語頭 | 語中 | 語末 |
|------|-----------|-------|-------|-------|-------|
| 1 | 未 | 18.75 | 21.43 | 14.69 | 19.64 |
| 2 | 病気・体調 | 4.64 | 1.78 | 2.81 | 8.56 |
| 3 | 膜・筋・神経・内臓 | 3.71 | 5.08 | 5.43 | 1.08 |
| 4 | 性質 | 3.04 | 1.06 | 6.84 | 1.53 |
| 5 | 助数接辞 | 2.41 | 1.10 | 3.16 | 2.89 |
| 6 | 医療 | 1.47 | 1.30 | 1.31 | 1.74 |
| 7 | 生理 | 1.27 | 1.60 | 1.17 | 1.05 |
| 8 | 生物 | 1.26 | 1.31 | 1.37 | 1.12 |
| 9 | 異同・類似 | 1.05 | 0.54 | 1.86 | 0.80 |
| 10 | 作用・変化 | 1.02 | 0.78 | 1.00 | 1.24 |
| 11 | 頭・目鼻・顔 | 1.02 | 1.97 | 0.99 | 0.23 |
| 12 | 感覚 | 1.01 | 0.77 | 1.22 | 1.03 |
| 13 | 障害・けが | 1.01 | 0.36 | 0.49 | 1.98 |
| 14 | 一般・全体・部分 | 0.98 | 1.18 | 1.26 | 0.57 |
| 15 | 骨・歯・爪・角・甲 | 0.89 | 1.23 | 1.29 | 0.26 |
| 16 | 順位記号（甲乙丙） | 0.87 | 1.14 | 1.16 | 0.38 |
| 17 | 体液・分泌物 | 0.84 | 1.24 | 0.96 | 0.40 |
| 18 | 様相・情勢 | 0.83 | 0.21 | 1.30 | 0.95 |
| 19 | 元素 | 0.75 | 0.87 | 0.87 | 0.54 |

| | | | | | |
|----|------|------|------|------|------|
| 20 | 手足・指 | 0.72 | 1.19 | 0.86 | 0.19 |
|----|------|------|------|------|------|

表7は、分類項目の上位20個である。『分類語彙表』には895個の分類項目があるが、そのうち535個の分類項目が出現している。これは全体の約6割であり、意味領域の偏りが示唆される。「未」を除いていちばん多かったのが「病気・体調」である。これは医療用語のデータであるため当然のことであろう。語中の位置に関する特徴を以下にまとめた。身体分は語頭に、症状は語末に、性質は語中に多いことが分かる。

- (1) 語頭の割合が語末の割合の2倍以上あるもの：
「膜・筋・神経・内臓」「頭・目鼻・顔」「一般・全体・部分」「骨・歯・爪・角・甲」「体液・分泌物」「手足・指」
- (2) 語末の割合が語頭の割合の2倍以上あるもの：
「病気・体調」「助数接辞」「障害・けが」「様相・情勢」
- (3) 語中の割合が語頭および語末のどちらの割合より2倍以上あるもの：
「性質」「異同・類似」

5.5 見出し語の属性

ここでは、見出し語の属性として、試験や教科書における出現情報を利用する。表8は、教科書（看護師及び管理栄養士養成校で使われている57冊）、看護師国家試験、助産師国家試験、管理栄養士国家試験における出現状況別に意味分類を集計したものである。全体に共通するのは、「序数接辞」「医療」「生理」の3項目だけであり、それぞれの分野の特徴が出ていることが分かる。

表8 出現情報による「分類項目」（それぞれの上位10）

| Rank | 教科書 | 看護師 | 助産師 | 管理栄養士 |
|------|-----------|------------|-----------|-------|
| 1 | 未 | 未 | 未 | 未 |
| 2 | 病気・体調 | 助数接辞 | 生理 | 生物 |
| 3 | 膜・筋・神経・内臓 | 一般・全体・部分 | 助数接辞 | 助数接辞 |
| 4 | 様相-性質 | 医療 | 膜・筋・神経・内臓 | 元素 |
| 5 | 助数接辞 | 病気・体調 | 医療 | 病気・体調 |
| 6 | 医療 | 時間的前後 | 数記号（一二三） | 量 |
| 7 | 生物 | 様相・情勢 | 病気・体調 | 性質 |
| 8 | 生理 | 感覚 | 順位記号（甲乙丙） | 生理 |
| 9 | 順位記号（甲乙丙） | 左右・前後・たてよこ | 生 | 医療 |
| 10 | 感覚 | 生理 | 異同・類似 | 数 |

6. 語の類別による違い

「2. 見出し語の類別」で述べた ComeJisyo の語彙的類別「A.非汎用」「B.汎用」「C.一般」により語彙の量的な違いが見られるのかどうかを調べる。この類別は、概ねこの順に語の一般性が高くなると推測されるので、以降、本稿では一般性による違いと位置付ける。表9は1語あたりの短単位数を「A.非汎用」「B.汎用」「C.一般」および全体で比べたものである。平均値、中央値、分散、標準偏差を示した。

表9 1語あたりの短単位数（語の一般性別）

| 類別 | 平均値 | 中央値 | 分散 | 標準偏差 |
|---------------|------|-----|------|------|
| 全体（114957語） | 2.67 | 2 | 1.64 | 1.28 |
| A.非汎用（83828語） | 2.54 | 2 | 1.38 | 1.18 |
| B.汎用（31129語） | 3.03 | 3 | 2.16 | 1.47 |
| C.一般（7139語） | 3.72 | 4 | 2.13 | 1.46 |

表9からA, B, Cの順に1語あたりの短単位の平均値が大きくなり、中央値も上がっている。すなわち、専門用語としての汎用性・一般性が高くなると語が長くなることが分かる。

表10 品詞の割合（語の一般性別）

| 品詞 | 全体 | A.非汎用 | B.汎用 | C.一般 |
|------|--------|--------|--------|--------|
| 名詞 | 79.25 | 80.25 | 76.99 | 76.14 |
| 代名詞 | 0.01 | 0.02 | 0.01 | 0.00 |
| 動詞 | 0.22 | 0.29 | 0.05 | 0.01 |
| 形容詞 | 0.11 | 0.13 | 0.07 | 0.03 |
| 形状詞 | 0.63 | 0.70 | 0.47 | 0.41 |
| 副詞 | 0.13 | 0.15 | 0.08 | 0.03 |
| 連体詞 | 0.00 | 0.00 | 0.00 | 0.00 |
| 接続詞 | 0.01 | 0.01 | 0.01 | 0.00 |
| 感動詞 | 0.08 | 0.09 | 0.04 | 0.01 |
| 接頭辞 | 2.79 | 2.76 | 2.85 | 3.48 |
| 接尾辞 | 14.04 | 12.04 | 18.57 | 19.75 |
| 助詞 | 0.17 | 0.21 | 0.08 | 0.03 |
| 助動詞 | 0.07 | 0.09 | 0.02 | 0.00 |
| 記号 | 2.01 | 2.57 | 0.73 | 0.10 |
| 補助記号 | 0.48 | 0.67 | 0.04 | 0.00 |
| 空白 | 0.00 | 0.00 | 0.00 | 0.00 |
| 合計 | 100.00 | 100.00 | 100.00 | 100.00 |

表11 語種の割合（語の一般性別）

| 語種 | 全体 | A.非汎用 | B.汎用 | C.一般 |
|------|--------|--------|--------|--------|
| 和語 | 5.73 | 6.18 | 4.73 | 3.36 |
| 漢語 | 75.50 | 73.07 | 80.99 | 92.99 |
| 外来語 | 10.25 | 10.78 | 9.05 | 2.50 |
| 混種語 | 0.30 | 0.33 | 0.24 | 0.21 |
| 記号 | 3.29 | 4.29 | 1.04 | 0.11 |
| 固有名詞 | 1.81 | 1.91 | 1.59 | 0.53 |
| unk | 3.11 | 3.45 | 2.36 | 0.30 |
| 合計 | 100.00 | 100.00 | 100.00 | 100.00 |

表10から、「A.非汎用」は「B.汎用」「C.一般」に比べて名詞の割合が高く、接尾辞、接頭辞の割合が低いことが分かる。表11の語種の割合からは、「A.非汎用」は和語、外来語、記号の割合が高いことが分かった。一方、unk（未知語）の割合も高いので、実際には誤差を含んだ値と考えるべきである。その意味では「C.一般」はunkが少なく、誤解析がなければ精度が高い結果になっている可能性がある。

表12は全体とそれぞれの上位20語であるが、非汎用には上位に「的」（接尾語）、「法」「術」

「後（ご，接尾辞）」「時（時）」のような一字漢語が見られる。

表 12 上位 20 語（語の一般性別）

| 順位 | 全体 | | A.非汎用 | | B. 汎用 | | C.一般 | |
|----|-----|------|-------|------|-------|------|------|------|
| | 語彙素 | 頻度 | 語彙素 | 頻度 | 語彙素 | 頻度 | 語彙素 | 頻度 |
| 1 | 性 | 9739 | unk | 7115 | 性 | 5750 | 性 | 2322 |
| 2 | unk | 9341 | 性 | 3989 | 症 | 3391 | 症 | 915 |
| 3 | 症 | 4846 | 的 | 1942 | unk | 2226 | 腫瘍 | 526 |
| 4 | 炎 | 2293 | 症 | 1455 | 炎 | 1921 | 損傷 | 453 |
| 5 | 的 | 2154 | 法 | 1093 | 部 | 835 | 炎 | 424 |
| 6 | 部 | 1692 | 術 | 932 | 骨折 | 740 | 先天 | 399 |
| 7 | 術 | 1509 | 部 | 857 | 神経 | 722 | 神経 | 287 |
| 8 | 神経 | 1509 | 細胞 | 795 | 骨 | 623 | 出血 | 256 |
| 9 | 法 | 1221 | 神経 | 787 | 関節 | 622 | 麻痺 | 236 |
| 10 | 動脈 | 1190 | 後 | 784 | 術 | 577 | 中毒 | 231 |
| 11 | 群 | 1056 | 切除 | 771 | 腫瘍 | 557 | 手術 | 226 |
| 12 | 後 | 1025 | ・ | 752 | 症候 | 556 | 多発 | 226 |
| 13 | 関節 | 1010 | 時 | 741 | 障害 | 537 | 部 | 226 |
| 14 | 障害 | 1002 | 型 | 721 | 群 | 534 | 動脈 | 213 |
| 15 | 症候 | 988 | 体 | 720 | 腫 | 523 | 感染 | 204 |
| 16 | 型 | 980 | 動脈 | 681 | 動脈 | 509 | 狭窄 | 198 |
| 17 | 体 | 944 | 量 | 674 | 損傷 | 464 | 関節 | 194 |
| 18 | 細胞 | 936 | 二 | 655 | 先天 | 459 | 骨 | 180 |
| 19 | 骨折 | 906 | 化 | 617 | 急性 | 443 | 結核 | 174 |
| 20 | 骨 | 900 | 薬 | 589 | 病 | 437 | 挫傷 | 170 |

7. まとめと今後の課題

本稿では、ComeJisyo 全体の語彙的特徴を頻度、品詞、語種、意味の面からその一端を明らかにした (RQ1)。また、専門用語としての汎用性、一般性の観点から ComeJisyo を 3 つに区分し、それぞれの区分において語彙的な特徴があるかを調べた (RQ2)。

結果としては、以下のことが明らかになった。

- ・ 基本的情報として、1 語あたりの短単位数は平均 2.67、最大値は 13 である。
- ・ 品詞では、名詞と接辞（接尾辞・接頭辞）で全体の約 96% を占める。
- ・ 語種では、漢語と外来語とで全体の約 85% を占める。
- ・ 意味分類では、約 78% が体の類、部門では、「抽象的關係」がもっとも多く、ついで「自然物および自然現象」、「人間活動—精神および行為」で、これらで全体の 71% を占める。
- ・ 中項目の意味分類では、全体では「量」「身体」「作用」「生命」「心」が上位 5 個であるが、語頭・語中においては、「身体」がもっとも多く、語末では「生命」が最も多かった。
- ・ 分類項目では、全体では「病気・体調」がいちばん多かったが、語頭では、「膜・筋・神

経・内臓」, 語中では「性質」, 語末では「病気・体調」が最多であった。

・試験や教科書の出現状況別に見た場合, 意味分布に違いが見られた。

・専門用語としての汎用性・一般性という観点から実践医療用語を類別して特徴を見た結果, 汎用性・一般性が高くなると語が長くなるなどの違いが見られた。

今後の課題としては, 未知語を精査して結果の精度を高めること, および, 語構成における各要素の文法的関係などを視野に入れた分析が挙げられる。

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) の助成を受けたものである。

文 献

内山清子, 岡照晃, 東条佳奈, 小野正子, 山崎誠, 相良かおる (2018) 「実践医療用語の語構成要素抽出の試み」『言語資源活用ワークショップ 2018 発表論文集』 pp.463-467
doi/10.15084/00001680

相良かおる, 小野正子, 鈴木隆弘, 嶋田元, 小作浩美 (2010) 「看護記録文の計量的用語調査」『じんもんこん 2010 論文集』 pp.103-110.

相良かおる, 小野正子, 小作浩美, 鈴木隆弘, 高崎光浩, 嶋田元 (2012) 「分かち書き用辞書 ComeJisyo の評価」『医療情報学』 32(6), pp.301-307.

相良かおる (2014) 「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』 24(2), pp.204-209.

相良かおる, 小野正子, 上野恵子 (2015) 「経過記録に含まれる誤字・誤変換と同義語・類義語の紹介」『日本医療情報学会看護学術大会論文集』 JAMI-NS 16, pp.110-113.

相良かおる, 小野正子, 山崎誠 (2016) 「末尾語にサ変接続名詞を持つ実践医療用語の語彙分類」『じんもんこん 2016 論文集』 pp.183-190.

相良かおる, 山崎誠, 中島直樹, 山下貴範, 小野正子 (2017) 「医療記録における語彙の月別出現分布」『じんもんこん 2017 論文集』 pp.103-110.

相良かおる, 山崎誠, 麻子軒, 東条佳奈, 小野正子, 内山清子 (2019) 「実践医療用語の語構成要素—意味を基準とした分割」『じんもんこん 2019 論文集』 pp.57-64.

東条佳奈, 相良かおる, 小野正子, 山崎誠 (2019) 「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—」『現代日本語研究』, 11, pp.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.

山崎誠, 相良かおる (2014) 「医療経過記録における漢字連続複合語の計量的分析」『じんもんこん 2014 論文集』 pp.221-226.

山崎誠, 相良かおる, 小野正子, 東条佳奈, 麻子軒 (2019) 「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 発表論文集』 pp.161-168. doi/10.15084/00002565