

意味ラベルを用いた「-性」を含む病名の言い換え

麻子軒¹⁾ 相良かおる²⁾ 高崎智子²⁾ 東条佳奈¹⁾ 山崎誠³⁾

1) 大阪大学 2) 西南女学院大学 3) 国立国語研究所

医療用語は、専門家以外の人にとって難解なものが多い。その理由に、医療用語を作る語構成要素間の統語構造に関する情報がないことが考えられる。本研究は、医療用語のうち、特に頻度が高い「-性」を含む病名を対象に、それぞれの語構成要素に付与された意味ラベルを利用し、統語構造に関する情報を機械的に加えることによって、病名を分かりやすく言い換えることを試みた。その結果、〈部位〉を表す意味ラベル（例：身体部位）は「での」、〈原因〉を表す意味ラベル（例：病原体）は「による」、〈状態〉を表す意味ラベル（例：経過、症状）は「である」、〈主体〉を表す意味ラベル（例：患者属性）は「に生じた」を当該語構成要素の直後に加えることによって、機械的な言い換えが可能であることが明らかになった。

Rewording disease names that include expressions of nature or quality using semantic labels

Tzu-Hsuan Ma¹⁾ Kaoru Sagara²⁾ Satoko Takasaki²⁾
Kana Tojo¹⁾ Makoto Yamazaki³⁾

1) Osaka University 2) Seinan Jo Gakuin University
3) National Institute for Japanese Language and Linguistics

Medical terms are often difficult for non-professionals to understand probably due to a lack of information on the syntactic structure between word-building elements that create medical terms. In this study we attempted to reword disease names for easier comprehension using semantic labels assigned to respective word-building elements, especially disease names, which often include “nature” or “quality” expressions and mechanically adding information on syntactic structure. The results showed that it is possible to mechanically reword medical terms by adding “at/in” for semantic labels that express the “affected part” (example: physical site), “due to” for semantic labels that express “cause,” (example: pathogen), “is” for semantic labels that express “state” (example: course, symptom), and “occurred in” for semantic labels expressing the “subject” (example: patient attributes).

1. はじめに

電子カルテシステムが導入されて以来、医療現場では日々医療記録が蓄積されている。しかし、電子カルテに含まれる医療用語は、専門家以外の人にとって難解なものが多く、誤解も生じやすい。その理由の1つに、医療用語を作る語構成要素間の統語構造に関する情報がないことが考えられる。本研究は、医療用語のうち、特に頻度が高い「-性」を含む病名を対象に、それぞれの語構成要素に付与された意味ラベルを利用し、統語構造に関する情報を機械的に加えることによって、病名を分かりやすく言い換えることを目的とする。その成果は、外国人看護師に対する日本語教育や、患者に対する病状説明などに応用できると思われる。

医療用語を分かりやすく言い換える試みとして、国立国語研究所編(2009)などが挙げられる⁽¹⁾。例えば、「悪性腫瘍」は「細胞が異常に増えてかたまりになったもののうち、すぐに治療が必要なもの」への言い換えが提案されている。こ

の方法は当該医療用語に詳しい説明を加えることが可能だが、語ごとに言い換えの文章をそれぞれ作成する必要があり、大量の語を機械的に処理するには向いていない。医療用語が難解な理由には、用語そのものになじみがないことや統語構造が分かりにくいことなど、複数考えられるが、本研究では後者の問題点を解決するために、統語構造の情報を病名に含めた形で言い換えることを試みる。

統語構造を考えるにあたって、まず病名の語構成の実態を明らかにする必要がある。医療用語の語構成に関する先行研究に、内山他(2018)⁽²⁾、相良他(2019)⁽³⁾、東条他(2019)⁽⁴⁾、山崎他(2019)⁽⁵⁾などがある。上記の研究では、医療用語の「意味」を基準とした最小単位を「語構成要素」として規定している。例えば、「アルコール性小脳性運動失調症」という病名は、「アルコール性 | 小脳性 | 運動失調症」のように、3つの語構成要素に分割され、それぞれ「病因 | 状態 | 病名」という意味ラベルが付けられている。本発

表では語の分割規則は上記の基準を踏襲する。用語規定の詳細は2節にて述べる。

2. 方法

2.1 データ

2014年に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 ComeJisyo⁽⁶⁾が言語資源として公開されているが、登録語の単位認定と語分割がなされていない。筆者らは、2018年より ComeJisyoSjis-1 に収録された医療用語の合成語 7,194語を対象に、その語構成解析に着手し、医療の観点から意味的に妥当な単位（以下、「語構成要素」）に分割し、これらに意味ラベルを付与している⁽³⁾。

現在、臨床看護および臨床医の経験を持つ共同研究者らで見直しを行っている。現時点での解析で得られた語構成要素は5,787種類、意味ラベルは93種類となっている。なお、語分割の単位規定は2.2節、意味ラベルの規定は2.3節で詳述する。

本研究では、ComeJisyoSjis-1に収録された医療用語の合成語 7,194語より、下記の例のような、「-性」の語構成要素を含み、かつ、語末の語構成要素の意味ラベルが「病名」である合成語 803語を分析対象とする。

合成語：アルコール性小脳性運動失調症
要素列：アルコール性 | 小脳性 | 運動失調症
意味ラベル：病因 | 状態 | 病名

2.2 語分割の単位規定

本研究で扱ったデータには、医療従事者が医療現場の観点から作成した、意味により分割した「短い単位」の語構成要素および、統語的なまとまりを考えた「長い単位」の語構成要素の両方があるが、本研究の目的は統語的な情報の付与による言い換えのため、実際に分析に使用するのは「長い単位」である。以下、合成語およびその長い単位（長）と短い単位（短）の例を挙げる。

合成語：アルコール性多発ニューロパチー
要素列（短）：アルコール性 | 多発 | ニューロパチー
意味ラベル列（短）：病因 | 数量 | 病名
要素列（長）：アルコール性 | 多発ニューロパチー
意味ラベル列（長）：病因 | 病名

2.3 意味ラベルの規定

意味ラベルは、意味カテゴリーによってなされた分類である。本研究では、石井(2007)の分類⁽⁷⁾を参考にし、共同研究者らで医療用語の性質に合わせて追加・修正したものである。今回の病

名を表す合成語の語構成要素に付与された意味ラベルを以下の表1に示す。

表1 意味ラベル分類

意味ラベル	例（下線部分）
位置	多発性 <u>後部</u> 色素上皮症
患者属性	<u>若年性</u> 脱毛症
感覚	先天性 <u>聴覚</u> 無知覚症
形状	壊疽性 <u>丘疹状</u> 結核疹
経過	<u>亜急性</u> 肺炎
行為	<u>性行為</u> 感染症
症状	心因性 <u>妄想</u> 精神病
状態	<u>拡大性</u> 頭蓋骨折
身体部位	<u>下肢</u> 悪性腫瘍
数量	<u>多発</u> 梗塞性認知症
体外物質	<u>食塩</u> 欠乏性脱水症
体内物質	<u>ヘキソキナーゼ</u> 欠乏性貧血
動き	多巣性 <u>運動</u> ニューロパチー
病因	<u>アルコール性</u> 肝硬変
病原体	<u>グラム陰性菌</u> 感染症
病態	<u>ケトン血性</u> 嘔吐症
病名	腎悪性 <u>腫瘍</u>

なお、文脈によって複数の意味ラベルが考えられる語もある。これらの語には、「病因・病態」のように、「・」を用いて複数の意味ラベルを列挙している。また、「テント下」のように位置と身体部位の両方の意味を持つ場合は、「/」を用いて「位置/身体部位」のように列挙する。今回の分析で用いた803語は、このような多義の語構成要素の場合の病名を除外している。

3. 言い換えのルール

上述のように、本研究の目的は、病名を中心とする難解な医療用語に統語構造に関する情報を付与することによって、機械的に分かりやすく言い換えることである。異なる意味ラベルであっても、統語構造上同じ役割になることもありうるため、意味ラベルと統語関係との関係を整理すれば、ある程度機械的な言い換えが可能になる。次節より、語構成要素に与えられた意味ラベルによって、機械的に言い換えられる規則を記す。なお、意味ラベル「病名」は、今回の調査対象では語末にしか出現しないため、言い換えるのルールが不要である。以下の本文では、山括弧〈 〉は統語的な役割、鉤括弧「 」は意味ラベルを表す。

3.1 〈部位〉を表す意味ラベル

症状が生じた〈部位〉を表す意味ラベルが現れた場合、その直後に「での」を付けることによって、統語関係を明示することができる。今回の調査結果で該当した意味ラベルは「位置」「身体部

位」である。本研究で用いた解析データでは、医療従事者による実用的な観点から、体の一部を特定できる場合は「身体部位」、特定できない場合は「位置」と、それぞれ異なる意味ラベルを付与しているが、いずれも統語的な役割において〈部位〉を表すため、構造的に言い換えルールは同様である。

- ・多発性後部色素上皮症
→ 多発性**後部**での色素上皮症（位置）
- ・門脈周囲性肝硬変
→ **門脈周囲**での肝硬変（位置）
- ・下肢悪性腫瘍
→ **下肢**での悪性腫瘍（身体部位）
- ・頸椎多発性脱臼
→ **頸椎**での多発性脱臼（身体部位）

なお、上記の例は、〈部位〉を表す意味ラベルがどのように言い換えられるかに焦点を当てて提示したもので、最終的な言い換えの結果ではない。例えば、「多発性後部色素上皮症」は最終的な言い換えの結果は「多発性である後部での色素上皮症」になるが、ここでは「位置」という意味ラベルを表す語構成要素「後部」の言い換えルールを明示するために「多発性後部での色素上皮症」という形で示した。

3.2 〈原因〉を表す意味ラベル

症状が生じた〈原因〉を表す意味ラベルの場合、その語構成要素の直後に「による」を付けることによって言い換えられる。下記に示す、意味ラベル「行為」「病因」「病原体」を含んだ病名がその例である。

- ・性行為感染症
→ **性行為**による感染症（行為）
- ・アルコール性肝不全
→ **アルコール**による肝不全（病因）
- ・感染性下痢症
→ **感染性**による下痢症（病因）
- ・グラム陰性菌感染症
→ **グラム陰性菌**による感染症（病原体）
- ・先天性ウイルス感染症
→ 先天性**ウイルス**による感染症（病原体）

この項目の意味ラベルで、特に「病因」に関しては「疾病を引き起こす原因」という解釈ができることから、意味カテゴリーよりも統語的な役割に近い分類が存在することは否めない。ただ、「病因」は特に疾病に関わる「原因」なので、意味カテゴリーとしての側面もあると思われる。また、

本研究で用いた解析データの意味ラベルは医療従事者の実用的な観点で付与されたものであり、医療現場では「原因」が「疾病」を引き起こすものかどうかを分ける必要がある事情を考慮し、あえて変更せずに採用した。

3.3 〈状態〉を表す意味ラベル

症状が生じた〈状態〉を表す意味ラベルの場合、その直後に「である」を挿入することで言い換えを行う。具体的に、「形状」「経過」「症状」「状態」「病態」の意味ラベルが該当する。

- ・壊疽性丘疹状結核疹
→ 壊疽性**丘疹状**である結核疹（形状）
- ・冠状加齢性白内障
→ **冠状**である加齢性白内障（形状）
- ・亜急性膝炎
→ **亜急性**である膝炎（経過）
- ・一過性関節症
→ **一過性**である関節症（経過）
- ・心因性妄想精神病
→ 心因性**妄想**である精神病（症状）
- ・拡大性頭蓋骨折
→ **拡大性**である頭蓋骨折（状態）
- ・びまん性脱毛症
→ **びまん性**である脱毛症（状態）
- ・多発性外傷
→ **多発性**である外傷（病態）
- ・ケトン血性嘔吐症
→ **ケトン血性**である嘔吐症（病態）

上記の例から、「経過」や「病態」などは分類上、「状態」の一種であるという見方もできる。「状態」は意味カテゴリーにおいて「経過」や「病態」の上位に位置すると考えられるためである。意味ラベルの階層関係について、今後まだ見直しが必要である。

3.4 〈主体〉を表す意味ラベル

今回の調査で症状が生じた〈主体〉を表す意味ラベルについて、該当したのは「患者属性」のみである。言い換えの規則は、その直後に「に生じた」を付加することである。

- ・老人性骨萎縮症
→ **老人**に生じた骨萎縮症（患者属性）
- ・若年性脱毛症
→ **若年**に生じた脱毛症（患者属性）

3.5 〈その他〉を表す意味ラベル

以上分析したもののほかに、統語関係を「の」

でしか示すことができない意味ラベルがある。例えば、「感覚」「数量」「体外物質」「体内物質」「動き」が該当する。

- ・先天性聴覚無知覚症
→ 先天性**聴覚**の無知覚症 (感覚)
- ・多発性梗塞性認知症
→ **多発**の梗塞性認知症 (数量)
- ・食塩欠乏性脱水症
→ **食塩**の欠乏性脱水症 (体外物質)
- ・アミノ酸欠乏性貧血
→ **アミノ酸**の欠乏性貧血 (体内物質)
- ・多巣性運動ニューロパチー
→ 多巣性**運動**のニューロパチー (動き)

なお、「の」はほとんどの統語関係を表せるので、頻繁に用いると言い換えの意義がなくなる。本研究では、「による」などのように、他の選択肢がある場合はそちらの表現を優先にする方針を採用する。

4. 「-性」の言い換え

ここで注意が必要なのは、「-性」に前接する要素が体言の場合、「性」を取って言い換えたほうが自然だということである。東条他 (2020) は、水野 (1987) の分類⁽⁸⁾を用いて、医療用語でも「-性」は前接する語基の品詞性によって異なる特徴をもっている現象があると指摘した⁽⁹⁾。

水野 (1987) は語基の文法的性格によってその機能を分析した。具体的に、①格助詞「が」を伴って文の要素となる「体言類」、②「する」を伴ってサ変動詞となる「用言類」、③「な」を伴って連体修飾成分となる「相言類」、④そのままで連用修飾成分となる「副言類」、⑤上に挙げた四つの類のどれにも当てはまらず、必ず接辞などと結合して用いられる「結合類」の五つの分類がある。

以下は、すでに3節で挙げたものも含めて、「-性」に前接する語基が「体言類」に当たる例とその言い換えを示す。

- ・門脈周囲性肝硬変
→ **門脈周囲**での肝硬変 (位置)
- ・アルコール性肝不全
→ **アルコール**による肝不全 (病因)
- ・ウイルス性中毒疹
→ **ウイルス**による中毒疹 (病因)
- ・老人性骨萎縮症
→ **老人**に生じた骨萎縮症 (患者属性)
- ・若年性脱毛症
→ **若年**に生じた脱毛症 (患者属性)

上記の例では、例えば、「アルコール」や「老人」などはいずれも体言であるため、言い換える結果、「アルコール性による〇〇」「老人性に生じた〇〇」よりも、「アルコールによる〇〇」「老人に生じた〇〇」のほうが自然に感じられる。機械的に言い換えを行う際に、「-性」の前に「体言類」の語基が現れた場合、「性」を削除する特別な処理が必要である。

次に、「-性」に前接する語基が「用言類」に当たる例を確認しよう。

- ・感染性下痢症
→ **感染性**による下痢症 (病因)
- ・麻痺性眼瞼下垂症
→ **麻痺性**である眼瞼下垂症 (状態)
- ・妄想性醜形恐怖症
→ **妄想性**である醜形恐怖症 (状態)
- ・多発性外傷
→ **多発性**である外傷 (病態)

上記の例では、「感染する」「麻痺する」という表現が存在することから、「感染」「麻痺」が「用言類」であることが分かる。この場合、「感染症による〇〇」「麻痺性である〇〇」というように、特別な処理が必要なく、機械的な言い換えが可能である。

「相言類」と「副言類」は今回の調査対象にはなかった。「結合類」は「-性」と前接する語基と分割できないので、言い換える際は、必ず下記の例のように「性」が付く形で現れる。

- ・一過性関節症
→ **一過性**である関節症 (経過)
- ・先天性冠状動脈瘤
→ **先天性**である冠状動脈瘤 (状態)
- ・悪性脊髄腫瘍
→ **悪性**である脊髄腫瘍 (病態)

以上見てきたように、「-性」を含む語構成要素の性質が特殊であり、特に「-性」に前接する語基が「体言類」の場合は単純に言い換えを行えないことがうかがえる。

5. 量的分布から見た語順の傾向

本節では、今回調査した病名の言い換えパターンの量的分布を確認する。その結果から、病名を構成する語構成要素の語順における傾向が観察できる。次の表2に、頻度が多い10位までの統語構造のパターンを示す。

表 2 統語構造のパターン (10 位まで)

順位	言い換えのパターン	頻度
1	「病因」による「病名」	308
2	「病態」である「病名」	106
3	「状態」である「病名」	86
4	「身体部位」での「状態」である「病名」	74
5	「身体部位」での「病態」である「病名」	48
6	「経過」である「病名」	28
7	「身体部位」での「病名」	17
8	「病因」による「身体部位」での「病名」	16
9	「病因」による「病態」である「病名」	11
10	「病因」による「状態」である「病名」	9

最も頻度が高いのは、【「病因」による「病名」】というパターンである。病名にとって最も重要な情報は、「なぜ」それを引き起こしたかであるということがうかがえる。なお、「病因」を表す語構成要素は、ほぼ「〇〇性」である。具体例を以下に挙げる。

- ・アルコール性肝硬変
→ **アルコール**による**肝硬変**
- ・遺伝性腎疾患
→ **遺伝**による**腎疾患**
- ・先天性直腸狭窄症
→ **先天性**による**直腸狭窄症**

次に多いのは、【「病態」である「病名」】(2位)と【「状態」である「病名」】(3位)というパターンである。病名は、その属性を細分化するために、「どのような」という情報も入れることが重要だと思われる。「病態」「状態」を表す語構成要素も、「〇〇性」が多かった。以下がその例である。

- ・悪性組織球症
→ **悪性**である**組織球症**
- ・多発性腎結石
→ **多発性**である**腎結石**
- ・局所性脳挫傷
→ **局所性**である**脳挫傷**
- ・複雑性尿路感染症
→ **複雑性**である**尿路感染症**

続いて頻度が高いのは、先に挙げたパターンに「身体部位」を加えた、【「身体部位」での「状態」である「病名」】(4位)と【「身体部位」での「病態」である「病名」】(5位)というパターンである。「どのような」という情報のほか、「どこに」という情報も病名において重要であることが分

かる。注目すべきなのは、このパターンでは「〇〇での良性腫瘍」「〇〇での悪性腫瘍」が大半を占めていることである。下に挙げるのはこのパターンの例である。

- ・リンパ節良性腫瘍
→ **リンパ節**での**良性**である**腫瘍**
- ・末梢神経悪性腫瘍
→ **末梢神経**での**悪性**である**腫瘍**
- ・頸髄中心性損傷
→ **頸髄**での**中心性**である**損傷**
- ・脊髄多発性神経鞘腫
→ **脊髄**での**多発性**である**神経鞘腫**

6位になっているのは、【「経過」である「病名」】というパターンである。このパターンも【「病態」である「病名」】と【「状態」である「病名」】と類似しており、病名の「どのような」という属性を表す情報だと言える。下の例を参照されたい。

- ・一過性関節症
→ **一過性**である**関節症**
- ・急性薬物中毒
→ **急性**である**薬物中毒**

7位以降のパターンは基本的にこれまでのパターンの組み合わせである。語順の傾向としては、「病因」が語頭に現れるパターンが多く、その次に「身体部位」か「病態」が続く現象が観察される。7位～10位のパターンの例をそれぞれ1例挙げる。

- ・尿路性器トリコモナス症
→ **尿路性器**での**トリコモナス症**
- ・先天性脊柱後側弯症
→ **先天性**による**脊柱**での**後側弯症**
- ・中毒性多結節性甲状腺腫
→ **中毒性**による**多結節性**である**甲状腺腫**
- ・遺伝性小脳性運動失調症
→ **遺伝**による**小脳性**である**運動失調症**

4位と5位のパターンから分かるように、「身体部位」が「状態」または「病態」より先に現れる傾向がある。そのため、10位までのパターンに出現した語構成要素は、おおよそ「病因」→「身体部位」→「状態」または「病態」→「病名」という順番になっている。語構成要素が多い病名は必然的に出現頻度が低いので、表2には表示していないが、下記の例がこれまでの結論を裏付けている。

・遺伝性赤芽球増殖性ポルフィリン症
 → **遺伝**による**赤芽球**での**増殖性**である
 ポルフィリン症

5節の結果をまとめると、病名は「どこに」「なぜ」「どのような」「症状」がある、というパターンで現れることが多いことが分かった。なお、最も語構成要素が多い語は「急性後部多発性斑状色素上皮症」で、言い換えの結果は「急性である後部での多発性である斑状である色素上皮症」である。

6. まとめと今後の課題

本研究では、「-性」を含む病名を対象に、統語関係の情報を付加する試みを行った。その結果、意味ラベルを利用すれば、病名をある程度機械的に分かりやすく言い換えられることが分かった。

特に、医療従事者などの専門家ではない人にとって、医療用語は「それぞれの語構成要素の意味が分かっても、それらが組み合わさった語の意味の理解は難しい」という存在である。統語関係を示す情報を付加することで、それぞれの語構成要素の間の関係が明らかになり、語全体の意味を理解するのに役に立つ。これは、外国人看護師に対する日本語教育や、患者に対する病状説明に活かすことができる。ただし、今回の調査を経て、以下の課題がまだ残っている。

1点目は意味ラベルの再整理である。例えば「病態」は「状態」の一種と捉えることも可能なので、このように階層関係が存在する意味ラベルの見直しが必要である。

2点目は意味ラベルに「統語レベル」と「意味レベル」のものが混在している問題である。例えば「ウイルス性」は「病因」に分類されたが、言い換えの際は「性」を取った「ウイルス」で行ったため、実際の意味ラベルは「病原体」のはずである。すなわち、「病因」は統語レベル、「病原体」は意味レベル、それぞれ異なる次元に属する分類であり、記述する際は別の枠組を設定する必要があるように思われる。

3点目は、文脈によって解釈が変わる語への処理である。例えば、今回の調査で暫定的に複数のラベルが与えられた語に関して、今後どのような言い換え規則を適用するかを決めることが重要である。

4点目は細かい処理が必要な語への対応である。今回、「-性」の前接要素が体言の場合は「性」を削除したほうが自然である例があったが、このような処理が必要な語はほかにもあると思われるので、さらに用例を精査し、これらを見つけ出す作業も必要であろう。

5点目は、今回提案した言い換えの方法が平面的な記述になっており、3つ以上の語構成要素を含む合成語の係り受け関係が示されていないという問題である。例えば、「先天性聴覚無知覚症」は本研究の言い換えの結果、「先天性による聴覚の無知覚症」になるが、統語構造上、「先天性」に係るのはその後続く「聴覚無知覚症」全体であるため、「[先天性による [聴覚の無知覚症]]」のように、係り受けの情報を明示する工夫が今後必要であると考えられる。

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)の助成を受けています。

参考文献

- [1] 国立国語研究所「病院の言葉」委員会編. 病院の言葉を分かりやすく一工夫の提案一, 勁草書房, 2009
- [2] 内山清子, 岡照晃, 東条佳奈, 小野正子, 山崎誠, 相良かおる. 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, 2018, p.463-467.
- [3] 相良かおる, 麻子軒, 東条佳奈, 小野正子, 内山清子. 実践医療用語の語構成要素一意味を基準とした分割. じんもんこん 2019 論文集, 2019, p.57-64.
- [4] 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 試案一「先天性」を例に一. 現代日本語研究, 2019, No.11, p.40-58.
- [5] 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 発表論文集, 2019, p.161-168.
- [6] 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, 2014, Vol.24, No.2, p.204-209.
- [7] 石井正彦. 現代日本語の複合語形成論. ひつじ書房, 2007.
- [8] 水野義道. 漢語系接辞の機能. 日本語学, 1987, Vol.6, No.2, p.60-69.
- [9] 東条佳奈, 麻子軒, 相良かおる, 高崎智子, 山崎誠. 病名における「-性」の分析—一般書籍との比較から—. 言語資源活用ワークショップ 2020 発表論文集, 2020.