

実践医療用語_語構成要素語彙試案表 Ver.2.0の構築

東条佳奈(大阪大学), 黒田航(杏林大学), 相良かおる・高崎智子(西南女学院大学),
西嶋佑太郎(医師), 麻子軒(関西大学), 山崎誠(国立国語研究所)

1. はじめに

- 電子カルテシステムの普及により、医療記録データの二次利用の需要が高まっているが、医療用語については自然言語処理に必要な語分割の処理のための情報が不足している。
- 医療記録に現れる実践医療用語には合成語が多く、医療関係者にとって意味のある最小単位を認定することも難しい。
例) 腹腔鏡補助下低位前方切除D2郭清施行
→語構成や語種構成の実態の調査が必要

『実践医療用語_語構成要素語彙試案表 Ver.2.0』作成の流れ

- 医療記録文の自然言語処理支援のためのMeCab用辞書 ComeJisyo (ComeJisyoSjis-1) の登録語より『分類語彙表 増補改訂版』に収録の語を含む合成語7,194語を抽出。
(手順の詳細は相良他2019,2020,相良2021を参照)
- 語構成要素とそれぞれに意味ラベルを付与した『実践医療用語_語構成要素語彙試案表 Ver.1.0』(試案表Ver.1.0)を2021年3月に公開。
- 試案表Ver.1.0の合成語や意味ラベルを見直し、新規の情報を追加したアップデート版であるVer.2.0を2022年3月に公開。
→本発表では試案表Ver2.0の紹介を行う。

2. 本試案表における用語について

- (相良2021を踏襲)
- 実践医療用語**: 医療施設で使われる医療記録に含まれる「学術上の専門用語」と「それ以外の専門用語」のこと。
 - 語構成要素**: 合成語を構成する要素で、合成語を医療の観点から意味的にまたは統語的に分割可能なすべての部分文字列のこと。
分割できない合成語は元の合成語を語構成要素とする。
例) 合成語: 脳幹多発性硬化症
語構成要素: 脳幹, 多発性, 硬化症, 多発性硬化症
 - 意味ラベル**: 医療の観点での意味を表すラベル。用語の意味を表すものと医療者の使用面での心的な捉え方を表すものが含まれる。
 - 語構成要素が多義の場合は意味ラベルを複数付与する。
例) 語構成要素「流動食」の場合
意味ラベルは《医療行為, 食品, 治療食》

データ概要

- 「対象語」シート: 対象合成語7,087語と読み・語構造・文字長情報
- 「語構成要素」シート: 語構成要素6,633種と読み・意味ラベル・出現頻度・語頭出現頻度・語末出現頻度
- 「意味ラベル」シート: 意味ラベル41種・意味ラベルに該当の語構成要素の個数・意味ラベルの属性(具体/抽象)

3. 『実践医療用語_語構成要素語彙試案表 Ver1.0』からの主な変更点

① 不要な対象合成語の削除

	Ver.2	Ver.1
対象合成語	7,087	7,192
語構成要素	6,633	6,380

- 異体字による重複や現在では使われない語などを削除。
- 語構成要素の分割位置の見直しにより語構成要素数は増加。

② 合成語の語構造情報の付与

悪性脳腫瘍→悪性[脳[腫瘍]]
脳淡蒼球内オイルプロカイン注入療法
→[脳[淡蒼球内]][[オイル[プロカイン]][注入[療法]]]

③ 意味ラベルの見直し・統合

80種類の意味ラベルを41種類へ

- 意味ラベルの用語の変更 例) 《体内物質》《体外物質》→《生体物質》《化学物質》
- 類義の意味ラベルの統合 例) 《感覚》《動き》《機能》→《身体機能, 生理》
- 上位概念となる意味ラベルを想定し、語構成要素が少ない意味ラベルを統合した。
例) 「白色」《色彩》→《状態》、「任意」《意向》→「任意入院」《医療行為》
- 命名が困難なもの・医療用語特有のものではない一般的な語構成要素に《#未定》ラベルを付与した。

Ver1.0より削除したラベル46種

意向, 動き, 音, 化学現象, 課題, 感覚, 関係, 基準, 軌跡, 基礎, 規則, 機能, 教育, 距離, 傾向, 行為, 作用, 時間, 色彩, 社会, 手技, 主体, 順序, 数量, 成育, 性質, 制度, 増減, 属性, 体外物質, 体内物質, 知的産物, 調節, 程度, 認識, 熱, 能力, 波動, 光, 文法用語, 変化, 方向, 保健衛生, 保留, 様相, 例示

Ver2.0のラベル41種

維持行為, 位置, 医薬品, 医療行為, 衛生物品, 化学物質, 患者属性, 患部, 機器, 経過, 形状, 検査, サービス, 施設, 指標, 種類, 症状, 状態, 食品, 身体機能, 身体部位, 精神, 生体物質, 生理, 組織, 治療食, 動植物, 排泄物, 場所, ヒト, 費用, 病因, 病原体, 病態, 病名, 物品, 部分, 法規, 方法, 予防行為, #未定

4. 意味ラベルからみる語構成要素の概観

語構成要素数上位順の意味ラベル(延べ)

順位	意味ラベル	語構成要素数	順位	意味ラベル	要素数	順位	意味ラベル	要素数	順位	意味ラベル	要素数
1	病名	1,918	11	病因	145	21	組織	46	31	施設	11
2	状態	1,669	12	#未定	141	22	患者属性	45		物品	11
3	身体部位	1,639	13	病原体	123	23	機器	39	33	サービス	10
4	病態	1,202	14	身体機能	96	24	指標	28	34	衛生物品	8
5	医療行為	319	15	生体物質	89	25	患部	26		法規	8
6	症状	249	16	種類	83	26	検査	23	36	食品	6
7	化学物質	241	17	生理	77	27	方法	22	37	予防行為	5
8	部分	239	18	医薬品	72	28	場所	20	38	治療食	4
9	位置	235	19	形状	60	29	精神	19		排泄物	4
10	経過	176	20	ヒト	54	30	動植物	17	40	維持行為	2
	総計	9,183								費用	2

- 上位4種の意味ラベルが全体の約7割
- 対象語のうち73.6%が病名であり(《病名》or《病態》ラベルが付与)、病名にはその症状が発生している患部が同時に含まれることが多いため《身体部位》ラベルも多い。

5. おわりに

- 本試案表を利活用することで、医療用語の語彙的な特徴の分析のほか、将来的に新たな医療記録データの解析に応用していくことが可能と思われる。
- 意味ラベル間の関係の整理や、未定ラベルの検討など、精緻化に向けた作業を引き続き行う。
- 語構成要素がすべて専門用語からなる合成語も分析対象に含める必要がある。

意味ラベルの変更と課題

- 《状態》ラベルが付与された語構成要素のほとんどは《状態》の下位にあたる《病態》ラベルが併記された。
- Ver.1.0で意味ラベルが《状態》のみ付与していた語構成要素は915あったが、Ver.2.0では218へ
→意味ラベルを複数付与したことで語構成要素が判別しやすくなった
例) 「サルコイドーシス性」《状態, 病態》
「コクシジオイデス性」
《状態, 病態, 病因, 病原体》
- 複数の意味ラベルの併記には、上下関係または包含関係と解釈されるものが混在している。
- 例) 「細菌性食中毒」
「細菌性」《状態, 病態, 病因, 病原体》
→, は>で読み替え可

謝辞

本研究はJSPS科研費JP18H03499およびJP21H03777の助成を受けています。

文献

- 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」『人文科学とコンピュータシンポジウム論文集』pp.57-64.
相良かおる・小野正子・高崎智子・東条佳奈・麻子軒・山崎誠(2020)
「実践医療用語の語構成と意味一語構成要素語彙試案表の作成にむけて」『人文科学とコンピュータシンポジウム論文集』pp.289-296.
相良かおる(2021)「実践医療用語における語構成要素の意味ラベルについて」『言語処理学会第27回年次大会発表論文集』pp.559-562.

関連URL

実践医療用語辞書 ComeJisyo <https://ja.osdn.net/projects/comedic/>
実践医療用語_語構成要素語彙試案表Ver.2.0
<https://www.gsk.or.jp/catalog/gsk2020-g>