

=====

『医療縮約表現_語構成要素語彙試案表 Ver.1.0』

◇ 概要 ◇

『医療縮約表現_語構成要素語彙試案表 Ver.1.0』（以下、「本試案表」という。）は、2024年に言語資源協会より配布を開始した電子データです。

本試案表には、医療記録に含まれる医療縮約表現 810 語を構成する語構成要素 1,342 要素、およびそれぞれの要素に付与した 50 種類の意味ラベルが収録されています。

本試案表における「語構成要素」は、「医療縮約表現を構成する要素で、言語学や日本語学の形態認定の単位ではなく、医療の観点から有意味性を反映する単位に分割した文字列を語構成要素」としています。例えば、「呼吸状態変調」の語構成要素は、「呼吸状態」と「変調」の 2 要素となります。なお、「医療の観点から有意味性を反映する単位」には、臨床医学では「扁平上皮癌」となるものが、解剖学では、「扁平上皮 | 癌」となるように専門分野により分割単位が異なることがわかっています。そこで、本試案表では、「|」により短い単位で分割したものに加えて、複数の分割位置がわかるように複層化形態論解析 MLMA 記法による記述も併記しています。例えば、「前腕 | 皮弁 | 再建」を MLMA 記法により「<[前腕]><{皮弁}><再建}>」と記述しています。MLMA 記法で記述することで、医療縮約表現を構成する 5 種類の語構成要素「前腕」「皮弁」「再建」「前腕皮弁」「皮弁再建」の抽出が可能になります。なお、MLMA による記述から、機械的に語構成要素を抽出する Perl 版スクリプトを公開しています。MLMA の詳細および Perl 版スクリプトについては、参考文献をご覧ください。

各語構成要素には、医療の観点からみた 50 種類の意味ラベルを付与しています。

例えば前述の「前腕皮弁再建」の 5 要素では、「前腕」には「身体部位」という意味ラベルを付与し、複数の意味を持つ「皮弁」と「前腕皮弁」には「患部、身体部位、部分」と 3 種類の意味ラベルを 50 音順に付与し、「再建」と「皮弁再建」には「治療行為」という意味ラベルを付与しています。

「再建」のように一般語として使われる語の場合も、医療の観点に基づいた意味ラベルを付与しています。

また、「FU（エフユー／フォローアップ）」のように異なる読み方をするものがある場合は、これらのヨミガナを併記しています。

加えて、本試案表では、「右側主要部の法則」を利用した意味の推測や、医療記録からの合成語の抽出に役立つように各語構成要素の語頭と語末に出現する頻度（医療縮約表現 810 語中）を付与しています。

◇ ファイル構成 ◇

本データの拡張子はxlsx (Excel ファイル) です。ファイルには次の6シートが入っています。

はじめに	: 各シートの説明および用語の説明
対象縮約表現	: 医療縮約表現 810 語の一覧
縮約表現 MLMA 記述	: 医療縮約表現を MLMA により記述したもの
全語構成要素 (MLMA 含む)	: 医療縮約表現に含まれる語構成要素の一覧
意味ラベル (個ラベル)	: 意味ラベル 50 種類の一覧
意味ラベル (行為まとめ)	: 4 種類の行為ラベル*の組み合わせ 15 通りを一つの意味ラベルとしてまとめた 60 種類の一覧

以下、「はじめに」以外の5シートの各列について簡単に説明します。

1. 対象縮約表現: 分析対象の医療縮約表現 810 語の一覧

ID	: 1~810
医療縮約表現	:
ヨミガナ	: 複数の読みがある場合は「/」で、併記 (例: エスステージイチビー/エスステージワンビー)
タイプ	: 縮約表現を I~III の3タイプに分類
要素数	: 医療縮約表現を語構成要素に分割した場合の語構成要素数
文字長	: 医療縮約表現の文字長
語構成要素列	: 語構成要素の分割位置を「 」で表記した医療縮約表現 例 「左 腎 尿管 全摘」→語構成要素は4要素
意味ラベル列	: 語構成要素に付与した意味ラベルを「 」でつなげた医療縮約表現の意味ラベル列 (例 「位置, 部分 身体部位 身体部位 治療行為」)
語頭要素	: 語構成要素に分割した際の先頭にくる語構成要素
語頭要素の意味ラベル	: 語頭要素の意味ラベル
語末要素	: 語構成要素に分割した際の最後にくる語構成要素
語末要素の意味ラベル	: 語末要素の意味ラベル

2. 縮約表現 MLMA 記述: 「対象縮約表現」の先頭5列、語構成要素の一覧

ID	:
医療縮約表現	:
タイプ	:

要素数	:
語構成要素列	: 語構成要素の分割位置を「 」で表記した医療縮約表現
MLMA 要素数	: MLMA 記述から抽出した語構成要素数
MLMA 記述※	: 重複および非連続な語構成要素を MLMA 記法で記述したもの 例 「<{[左]><腎}><尿管}><全摘>」 → 語構成要素は、6 要素
D 列-F 列	: 語構成要素数の差異（「要素数」 - 「MLMA 要素数」） 例 「左腎尿管全的」の場合は、4 要素 - 6 要素で「-2」

3. 全語構成要素(MLMA 含む): 全語構成要素 1,342 要素の一覧

No.	: 1 ~ 1342
語構成要素	:
ヨミガナ	: 複数の読みがある場合は「/」で、併記
意味ラベル	: 語構成要素に付与した意味ラベル
出現回数	: 医療縮約表現 810 語における出現回数
MLMA 抽出	: MLMA 記述により抽出した語構成要素に「1」を付与
語頭出現頻度	: 語末に出現する語構成要素数
語末出現頻度	: 語頭に出現する語構成要素数

4. 意味ラベル (個ラベル): 意味ラベル 50 種類の一覧

ID	: 1 ~ 50
意味ラベル	: 語構成要素に付与した意味ラベル
付与した語構成要素数	: 語構成要素 1,264 要素における付与数
付与した語構成要素数 (含む MLMA 記述)	: MLMA 記述により抽出した語構成要素を含めた全 1,342 要素における付与数

5. 意味ラベル (行為まとめ): 意味ラベル 60 種類の一覧

ID	: 1 ~ 60
意味ラベル	: 付与した 4 種類の意味ラベル「治療行為」「患者行為」「支援行為」「非医療行為」の組み合わせ (14 種類 / 全 15 種類) を一つの意味ラベルとした場合の一覧
付与した語構成要素数	: 語構成要素 1,264 要素における付与数
付与した語構成要素数 (含む MLMA 記述)	: 全語構成要素 1,342 要素での付与数

◇ ライセンス情報と免責事項 ◇

本試案表をご利用の際は、出典情報として GSK の言語資源カタログに記載の下記 URL を明記して下さい。

<https://www.gsk.or.jp/catalog/gsk2024-b/>

複製および改変してご利用頂いて構いませんが、再配布は禁止します。

本試案表の利用によって生じた損失や損害について、言語資源協会および本データの開発者は一切の責任を負いかねます。あらかじめご了承ください。

◇参考文献◇

医療縮約表現の抽出方法などの分析過程：

医療縮約表現の分析と課題，相良かおる,黒田航,東条佳奈,西嶋佑太郎,麻子軒,山崎誠，言語処理学会 第 30 回年次大会 発表論文集，160-164 2024.

MLMA 記述：

要素の重複と不連続性を扱える抽出型の語構成要素解析:並列分散型形態素解析の提案，黒田航,相良かおる,東条佳奈,麻子軒,西嶋佑太郎,山崎誠，言語処理学会第 29 回年次大会，772-777 2023.

◇ 研究代表者：相良 かおる（奈良先端科学技術大学院大学）

共同研究者： 黒田 航（杏林大学）
東条 佳奈（大阪大学）
西嶋 佑太郎（京都大学）
麻 子軒（関西大学）
山崎 誠（国立国語研究所）

◇ 更新履歴 ◇

2024 年 7 月 5 日 Ver.1.0 公開

◇ 謝辞 ◇

『医療縮約表現_語構成要素語彙試案表』は、科学研究費補助金 (JP21H03777) の助成を受けています。

以上