

『実践医療用語_語構成要素語彙試案表 Ver.2.0』

◇ 概要 ◇

『実践医療用語_語構成要素語彙試案表 Ver.2.0』(以下、「本試案表」という)は、2022年に言語資源協会より無料で配布を開始した電子データです。

本試案表は、2021年に公開した『実践医療用語_語構成要素語彙試案表 Ver.1.0』の合成語7,192語を見直し、異体字による重複、誤字、現在では使われない語などを除いた7,087語を対象合成語とし、これらを構成する語構成要素6,633要素、およびそれぞれの要素に付与した41種類の意味ラベルが収録されています。

実践医療用語には、読み方が一般的でないものもあることから、本試案表では対象合成語と全ての語構成要素に読みを付与しています。加えて、対象合成語には、語構成要素の境界を示した語構造の情報を以下の例のように付与しています。

対象合成語：悪性脳腫瘍

語構造：悪性[脳[腫瘍]]

語構成要素：悪性，脳，腫瘍，脳腫瘍

本試案表における「語構成要素」は、「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」と定義します。従って、例えば合成語「オリーブ橋小脳萎縮症」では、「オリーブ」、「橋(キョウ)」、「小脳」、「オリーブ橋小脳」、「萎縮症」の5つの語構成要素から構成されるものとします。

これらの語構成要素には、医療の観点からみた意味ラベルを付与しています。Ver.2.0では、Ver.1.0で用いた80種類の意味ラベルを41種類にまとめました。命名が困難なもの、および「岩波国語辞典第五版タグ付きコーパス2004」に立項されている医療用語特有のものではない一般的な語構成要素には「#未定」のラベルを付与しています。

また、語構成要素には、結合する語構成要素によって意味が変わるものと、そもそも複数の意味を持つものなどがあります。本試案表では、複数の意味ラベルを「,」で列挙しています。例えば語構成要素「治療食」の意味ラベルは「医療行為, 食品, 治療食」となっています。「治療食」には、豆腐やカボチャなどの一般の食品が含まれ、また医療行為の一つとも考えられます。そこで「医療行為」と「食品」と「治療食」の意味ラベルを「,」により併

記しています。

最後に、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、各語構成要素が対象合成語の語頭、語末に出現する頻度（合成語 7,087 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（エクセルファイル）です。ファイルには次のシートが入っています。

- はじめに : 用語の定義及び各シートの解説
- 対象語 : 対象語一覧
- 語構成要素 : 語構成要素一覧
- 意味ラベル : 意味ラベル一覧

以下、「はじめに」以外を中心に簡単に説明します。

1. 「対象語」シート

1.1 レコード数

レコードの総数は、7,087 件です。

1.2 項目の種類と説明

レコードを構成する項目は、次の 5 項目です。

- ID : 通し番号
- 対象語 : 収録した語の見出し
- フリガナ : 対象語の読み
- 語構造 : 語構成要素の境界を示した情報
- 文字長 : 対象語の文字列の長さ

2. 「語構成要素」シート

2.1 レコード数

レコードの総数は、6,633 件です。

2.2 項目の種類と説明

レコードを構成する項目は、次の 8 項目です。

ID : 通し番号
語構成要素 : 対象語を構成する要素
フリガナ : 語構成要素の読み
意味ラベル : 語構成要素の意味を表すラベル
出現頻度 : 対象語における語構成要素の出現頻度
語頭 : 当該語構成要素が語頭に出現した頻度
語末 : 当該語構成要素が語末に出現した頻度
対象語 ID : 語構成要素が対象語と一致する場合の対象語 ID

3. 「意味ラベル」シート

3.1 レコード数

レコードの総数は、41 件です。

3.2 項目の種類と説明

レコードを構成する項目は、次の 5 項目です。

ID : 通し番号
意味ラベル : 語構成要素の意味を表すラベル
フリガナ : 意味ラベルの読み
語構成要素数 : 該当意味ラベルに分類された語構成要素の数
属性 : 「日本語語彙大系」を参考に「具体」と「抽象」の属性を付与

◇ ライセンス情報と免責事項 ◇

本試案表を研究・教育等に利用する場合は出典情報として 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/gsk2020-g>

複製および改変して利用することも可能ですが、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◇

[1] 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－意味を基準とした分割－. じんもんこん 2019 論文集. 2019, p.57-64.

- [2] 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―. 現代日本語研究. 2019, No.11, p.40-58.
- [3] 麻子軒, 相良かおる, 高崎智子, 東条佳奈, 山崎誠. 意味ラベルを用いた「一性」を含
む病名の言い換え. 人文科学とコンピュータシンポジウム論文集. 2020, p.283-288.
- [4] 相良かおる, 小野正子, 高崎智子, 東条佳奈, 麻子軒, 山崎誠. 実践医療用語の語構成
と意味―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. 2020, p.289-296.
- [5] 黒田航, 相良かおる. 医療用語の is-a オントロジー構築の FCA を使った効率化. 言語
処理学会 第 28 回年次大会 発表論文集. 2022, p.705-710.

◇ 開発者 ◇

研究代表者： 相良かおる（西南女学院大学）

共同研究者： 黒田航（杏林大学）

高崎智子（西南女学院大学）

東条佳奈（大阪大学）

西嶋佑太郎（医師）

麻子軒（関西学院大学）

山崎 誠（国立国語研究所）

◇ 更新履歴 ◇

2021年3月30日 Ver.1.0 公開

2022年3月31日 Ver.2.0 公開

◇ 謝辞 ◇

『実践医療用語_語構成要素語彙試案表』は, JSPS 科研費 JP18H03499 および 21H03777 の
助成を受けています。

以上