

国立国語研究所学術情報リポジトリ

実践医療用語_語構成要素語彙試案表 Ver.3 の公開にむけて

メタデータ	言語: 出版者: 公開日: 2023-11-24 キーワード (Ja): キーワード (En): 作成者: 相良, かおる, 黒田, 航, 東条, 佳奈, 西嶋, 佑太郎, 麻, 子軒, 山崎, 誠 メールアドレス: 所属:
URL	https://doi.org/10.15084/0002000139

実践医療用語_語構成要素語彙試案表 Ver.3 の公開にむけて

相良かおる（奈良先端科学技術大学院大学）

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）

Word Component Database for “Hands-On Medical Terms” (Version 3) Specification

Kaoru Sagara (Nara Institute of Science and Technology)

Kow Kuroda (Kyorin University)

Kana Tojo (Osaka University)

Yutaro Nishijima (Kyoto University)

Tzu-Hsuan Ma (Kansai University)

Makoto Yamazaki (National Institute for Japanese Language and Linguistics)

要旨

医療用語の合成語の語構造および語構成要素とその意味を明らかにすることを目的に、合成語 7,087 語を分析し『実践医療用語_語構成要素語彙試案表 Ver.2』を作成する過程で、(1)医療用語の選定方法、(2)分割単位の曖昧性、(3)語構造の記述方法、(4)意味ラベルの命名と付与方法に課題が見つかった。そこでこれらの課題を検討し、改良版の試案表 Ver.3 の作成に着手した。具体的には、(1)合成語の追加は今後の課題とし試案表 Ver.3 では行わない。(2)語分割が複数考えられる場合は、一つに絞るか分割せずに一つの語構成要素とする。(3)並列および不連続な語構造も記述できる複層化形態素解析 (Multi-Layered Morphological analysis : MLMA) を提案し、試案表 Ver.3 では、不連続な構造は原則として記述しないが、一部の並列構造を不連続な構造として記述できる場合は記述する。(4) 語構成要素の意味ラベルの集合と合成語の意味との関係がわかるように意味ラベルを合成語にも付与する。また意味ラベルが「#未定」の語構成要素の見直しを行うこととした。

1. はじめに

医療記録データには、複数の語が連結された医療用語の合成語（以下、「医療合成語」という）が多く存在する。一方、施設内で蓄積される少ないデータで自然言語処理を行うためには、医療合成語の語構成やそれらの構成要素の意味の利用は有用である。しかし、医療記録は個人情報が含まれるため非公開であり、言語学的な調査があまり行われてこなかった。

そこで、我々は、(1)医療合成語の語構造を明らかにすること、(2)医療合成語を構成する語構成要素を抽出し、意味ラベルを付与すること、(3)医療合成語の日本語学または言語学的知見を得ること、(4)得られた知見を含む成果物を医療実践、医療教育の領域で、出来得れば言語学研究の領域においても利用可能な形で公開することを目的とし、『実践医療用語辞書 ComeJisyoSjis-1』より抽出した医療合成語 7,192 語を対象に 2020 年に『実践医療用語_語構成要素語彙試案表 Ver.1』（以下、「試案表 Ver.1」という）を作成した。次いで 2021 年には、見直しの必要があった意味ラベルの妥当性（相良 2021）を踏まえて、医療合成語

や意味ラベルを見直し、『実践医療用語__語構成要素語彙試案表 Ver.2』¹（以下、「試案表 ver.2」）を公開した（東条ら 2022）。

医療合成語の語構造の解析の過程で、医療合成語には①同じ見出し語であっても数詞の表記法にゆれがあること、算用数字は、順序尺度が多く名義尺度が少ないこと、ローマ数字は名義尺度が多いこと（東条ら 2021）②「～性」を含む病名が多く、「先天性」「多発性」「急性」などが高頻度に用いられ、「急性細菌性髄膜炎」のように他の「～性」との共起も見られること（東条ら 2020）、③「急性」と「～性」の共起については、「急性」は語頭に多く出現し、医療現場では「急性」無しの同義の病名が多く使われていること、「急性」とその他の「～性」の語順を変えた同義語が存在すること（相良ら 2022）などの知見を得ている。また、語構造と意味に関しては、④《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ている（麻ら 2020）。

しかしながら、これらの研究結果は、日本語学的な成果というよりは、言語処理をする上で有益な成果といえる。また、医療現場では「急性」無しの病名が使われている（相良ら 2022）ということから、対象としている医療合成語は、「実践医療用語」と言えるのかという疑問も浮上している。

そこで本稿では、試案表 Ver.3 を日本語学的な分析にも利用可能な言語資源とするための問題点と解決方法について述べる。具体的には、先ず本研究で対象とする医療合成語と各試案表の概要を述べ、次いで、日本語学的知見を得る上での問題点と、試案表 Ver.3 に反映できる解決方法について述べる。

2. 医療合成語の概要

2.1 選定方法

以下に本医療合成語の選定方法を示す（相良ら 2019 再掲）。

手順 1. 医療合成語候補 31,162 語の抽出

『実践医療用語辞書 ComeJisyoSjis-1』の見出し語の抽出に用いた医療記録は、それぞれ 1 施設分の医師経過記録、看護経過記録、多職種共有経過記録であり代表性はない。そこで、対象合成語のバイアスを減らすために Web 上で公開されている辞書など研究用に収集した医療用語データと本辞書の見出し語を照合し一致した 31,162 語を医療合成語候補とした。なお、これらには、英数字のみからなる語も含まれている。

手順 2. 医療合成語に使われる一般的な用語 231 語の抽出

医療の知識がなくても語の境界の判定が容易になるように、一般的な日本語を含む医療合成語を対象にすることとし、医療合成語候補 31,162 語を MeCab 0.996² と Unidic-cwj-2.2.0³ を用いて自動形態素解析を施し、以下の(a)～(d)に沿って語を抽出した。

- (a) 「名詞」または「形状詞」となる単一語
- (b) 二語以上の語で「名詞」+「接尾辞」、「形状詞+接尾辞」、「接頭辞」+「名詞」、

¹ <https://www.gsk.or.jp/catalog/gsk2020-g/>

² MeCab : <https://taku910.github.io/mecab/>

³ 現代書き言葉 UniDic : https://clrd.ninjal.ac.jp/unidic/back_number.html#unidic_bccwj

漢字2字の「名詞」+「名詞」、「記号」のみの組み合わせであるもの

(c) 上記以外で文字数3文字以下のもの

(d) 自動形態素解析の結果、「感動詞」と誤解析された語を、目視で確認し妥当と認められたもの

次に(a)~(d)の規則に従ったものから「英数字」、「ひらがな」、「カタカナ」のみで表記された語を除いた3,728語について、2018年11月公開のComeJisyoUtf8-1の登録語75,831語における出現頻度を求め、頻度の降順上位25%の768語を抽出した。

医療合成語の語構成要素に意味ラベルを付与する上で、『分類語彙表一増補改訂版一』（以下、『分類語彙表』⁴）の分類項目の利用可能性を考え、768語と『分類語彙表』の見出し語を照合し、一致した231語を語構成要素候補とした。なお、ComeJisyoUtf8-2の見出し語114,957語に含まれる語を『分類語彙表』の分類番号で集計（意味分類）した結果については、山崎（2020）に詳述されている。

手順3. 一般的な語を含む医療合成語：7,192語

医療合成語候補31,162語の内、語構成要素に一般的な語231語を含む7,192語を試案表Ver.1の医療合成語とした。

2.2 出現頻度

表1と表2は、医療合成語候補31,162語から英数字のみの語および人名などの固有名詞を除いた30,881語（以下、「全医療合成語」という）および一般語を含む試案表Ver.1の医療合成語7,192語についてComeJisyoUtf8-2r1（見出し語数114,957語）⁵に記載の出処に関する情報をまとめたものである。表1は、5種類の文書における頻度を示したものであり、試案表Ver.1の医療合成語7,192語において文書頻度が0の語が57.36%占めている。

表1 文書頻度 (0~5) ⁶

	計	5	4	3	2	1	0	未登録語
全医療合成語	30,881	583	1,401	5,897	3,553	4,320	15,104	23
	100%	1.89%	4.54%	19.10%	11.51%	13.99%	48.91%	0.07%
試案表 Ver.1	7,192	143	251	1,025	688	958	4,125	2
	100%	1.99%	3.49%	14.25%	9.57%	13.32%	57.36%	0.03%

※未登録語：ComeJisyoUtf8-2r1の見出し語にない語

表2は、電子カルテシステム導入の際に推奨されるMEDIS標準マスター⁷の内、病名マスター、手術・処置マスター、症状所見マスター、看護実践用語マスターの4分野と照合した

⁴ 国立国語研究所：分類語彙表 増補改訂版，大日本図書，2004.

⁵ <http://comedic.osdn.jp/> ※現在OSDN側の都合でダウンロードが出来ない場合がある。

⁶ 実際の医療記録データ3種類と看護師教育で使われる看護記録データ、そして言語資源協会より公開されている模擬診療録テキスト・データの5種類における文書頻度(0~5)

⁷ 一般財団法人医療情報システム開発センター(以下、MEDIS-DC)が厚生労働省から委託を受け、診療情報の「用語・コード」の標準化として作成したもので、以下の9分野の標準マスターがある。①病名マスター(ICD10対応標準病名マスター)、②手術・処置マスター、③臨床検査マスター、④医薬品HOTコードマスター、⑤医療機器マスター、⑥看護実践用語標準マスター、⑦症状所見マスター、⑧歯科分野マスター、⑨画像検査。なお、本稿では①②⑥⑦の4分野を扱っている。

結果の頻度を示したものである。試案表 Ver.1 の医療合成語の 85.55%が病名マスターに登録されている病名となっている。なお、試案表 Ver.2 の医療合成語 7,087 語における病名マスターの割合は 86.24%、文書頻度が 0 の割合は 57.99%である。

表 2 MEDIS 標準マスターの登録語数

	計	病名	手術・処置	症状所見	看護実践用語	マスター登録無	未登録語
全医療合成語	30,881	20,499	995	38	185	9,141	23
	100%	66.38%	3.22%	0.12%	0.60%	29.60%	0.07%
試案表 Ver.1	7,192	6,153	425	7	41	564	2
	100%	85.55%	5.91%	0.10%	0.57%	7.84%	0.03%

※未登録語：ComeJisyoUtf8-2r1 の見出し語にない語

表 3 は、全医療合成語の文書頻度と MEDIS 標準マスターの内、病名マスター、手術・処置マスター、症状所見マスター、看護実践用語マスターの 4 分野と照合した結果のクロス集計表である。全体の 66.43% (20,499 語) が病名マスターに登録の病名であり、その内 52.75% (10,814 語) が文書頻度 0 (全体に占める割合は 35.04%) である。

表 3 文書頻度と MEDIS 標準マスターのクロス集計

	計	病名	手術・処置	症状所見	看護実践用語	マスター登録無	
5	583	236	19	6	28	294	
4	1,401	653	41	6	52	649	
3	5,897	3,595	141	13	64	2,084	
2	3,553	2,255	115	5	17	1,161	
1	4,320	2,946	160	5	11	1,198	
0	15,104	10,814	519	3	13	3,755	
計	30,858	20,499	995	38	185	9,141	

3. 試案表の概要

表 4 と表 5 は、各試案表の概要をまとめたものである。試案表 Ver.1 と Ver.2 の変更点は、① 不要な対象医療合成語の削除、② 合成語の語構造情報の付与、③ 意味ラベルの見直し・統合である (東条ら 2022)。①については、異体字による重複と誤字および現在では使われていない語 105 語を削除した。②の語構造情報は、医療合成語に語の境界位置を [] で示したものである。

医療合成語：感染後甲状腺機能低下症

語構造：感染後 [甲状腺 [機能低下症]]

語構成要素：感染後、甲状腺、機能低下症、甲状腺機能低下症

語構成要素の認定は、長単位や短単位といった言語学や日本語学の観点の形態認定の単位とするのではなく、「医療の観点からみた有意味性を反映する単位 (以下、「医療の観点からみた単位」という)」とし、その判断は、医師、看護師と議論を交えて行っている。

以下に医療の観点からみた有意味性を反映する単位の例を示す。なお、医療従事者が分割

マスター登録の病名となっており、実際の医療記録に出現する合成語の殆どが病名であるとは考えにくく、偏りがあること、合成語の選定方法に問題があることは否めない。

4.2 語の分割の曖昧性

言語学的な分析では、品詞などを考慮した機能的な単位で複合語を分割するのが一般的である。しかし、意味の理解なくして医療合成語を品詞単位で分割することは困難である。そして意味の解釈には曖昧性がある。同じ医学でも専門領域によって意味の解釈がことなる例として、「右上葉肺扁平上皮癌」を「概念が失われない最小の単位」に分割する際、臨床医学の領域では、「右 | 上 | 葉 | 肺 | 扁平上皮癌」となるが、解剖学の領域では「右 | 上 | 葉 | 肺 | 扁平上皮 | 癌」となることが指摘されている（劉 2000）。本研究でも「乳汁分泌抑制」の分割において、使用頻度の高い助産師は「乳汁分泌 | 抑制」と分割し、医師と看護師は「乳汁 | 分泌抑制」と分割した。なお、試案表では、医学大辞典の部分一致検索の結果⁸を根拠に、「乳汁 | 分泌抑制」としている。このように、語の意味の解釈には曖昧性があることから、意味による語分割において斉一性を保証することは困難である。

一方、意味が捨象される統計的自然言語処理や機械学習による語分割では、大量のデータが必要であり、秘匿性の高い医療記録データを大量に集めることは困難である。また、語分割された結果が有意味か否かの判断は人に委ねられるため、意味の解釈の曖昧性の問題に直面することになる。

本研究では、医療合成語を医療の観点から有意味性を反映する単位に分割し、語構成要素を抽出している。

医療合成語：一過性甲状腺機能低下症

語 構 造：一過性 [甲状腺 [機能低下症]]

語構成要素：一過性、甲状腺、甲状腺機能低下症、機能低下症

しかし「機能低下症」は、「機能低下」に有意味性があると判断すると「機能低下 [症]」となり、「低下症」に有意味性があると判断すると「機能 [低下症]」となる。このように医療の分野により判断が分かれ、語末が「症」や「障害」など前要素に意味を添える接尾語的な機能を持つ語の場合は、分割せずに一つの語構成要素としている。

その結果、病気を表す語末語「病」および「症」の前要素の相違または特徴を調べる際には、試案表 Ver.2 をそのまま利用することはできず、語構造の見直しが必要となる。

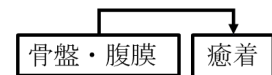
4.3 語構造の記述方法

試案表 Ver.2 では、「・」で区切って列挙した語構成要素の後要素との関係は、以下に示すように2種類あり、医療の知識がないと、どちらのタイプかを判断するのは困難である。

医療合成語：骨盤腹膜癒着

語 構 造：[骨盤・腹膜] 癒着

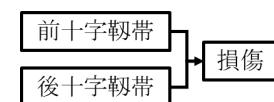
意 味：骨盤と腹膜が癒着している状態



医療合成語：前後十字靭帯損傷

語 構 造：[[前・後] 十字靭帯] 損傷

意 味：前十字靭帯と後十字靭帯が損傷している状態



⁸ 南山堂医学大辞典第20版の検索結果では、「分泌抑制」の「乳汁」以外の前要素として「成長ホルモン」「胃酸」「酸」がある。

4.4 意味ラベル

試案表 Ver.1 では、一般的な日本語に意味ラベルを付与し、その数は 80 種類となった。しかし、「～性」となる語構成要素をはじめ、患者の状態全てに意味ラベル《状態》を付与した結果、意味ラベル《状態》が全語構成要素に占める割合は、22.24%となった。そこで Ver.2 では、医療の観点による意味ラベルの付与に注力し、意味ラベル《状態》を細分化し、一方で国語辞典⁹に見出し語のある語構成要素については《#未定》の意味ラベルを付与した。その結果、「サルコイドーシス性」は《病態》で、「マイコプラズマ性」は《病因》かつ《病原体》というように「～性」の前要素に関する知識がなくとも意味判断ができるようになった（東条ら 2022）。一方、一般的な日本語に《#未定》の意味ラベルを付与したことで、《#未定》が付与された語構成要素 141 語の全体に占める割合は、2.13%となっている。

複数の意味を持つ語構成要素には、複数の意味ラベルを無秩序に「,」で区切って列挙している。表 6 は、語構成要素に付与した意味ラベル数をまとめたものである。Ver.2 では、意味ラベル《状態》を細分化したことにより、複数の意味ラベルを付与した語構成要素が多くなっている。

表 6 語構成要素に付与した意味ラベル数

	1	2	3	4	計
Ver.1	5,603	745	32	0	6,380
Ver.2	4,289	2,171	138	35	6,633

多義の語構成要素の中には、連結する語構成要素により意味が一意に決まるものがある。

語構成要素：ウイルス
意味ラベル：病因,病原体

語構成要素：ウイルス性
意味ラベル：状態,病態,病因,病原体

「ウイルス」に付与した《病原体》と《病因》に一般的な日本語の意味での関係はない。そして後ろに「肝炎」がつく「ウイルス肝炎」での「ウイルス」の意味は《病因》となり、「肝炎」が前につく「肝炎ウイルス」での「ウイルス」は《病原体》となる。なお、「ウイルス肝炎」は「ウイルス性肝炎」の「性」が省略されたものと考えられる。

語構成要素：離開
意味ラベル：状態

語構成要素：嚥下困難
意味ラベル：状態,症状

語構成要素：嚥下性
意味ラベル：状態,病態

語構成要素：呼吸困難
意味ラベル：症状,状態,病態

一方、《症状》《病態》《状態》には、意味的な重なりがある。ものごとのありさまには《状

⁹ 「岩波国語辞典第五版タグ付きコーパス 2004」の見出し語と照合

態》を付与し、「脱水」など病気により患者に現れた状態には《症状》を、「栄養不良」などの身体機能の状態には《病態》を付与している。

語構成要素に付与した意味ラベルは、前後の語構成要素により変わる場合がある。今後、多様な医療合成語を分析する中で、語構成要素に付与される意味ラベルの種類は増える可能性があり、語構成要素に付与した意味ラベルの組み合わせが変わる場合がある。

現在は複数の意味を持つ場合、無秩序に「,」で区切って意味ラベルを列挙しており、見かけ上意味ラベル間に包含関係があるか否の判断はできない。

4.5 医療合成語の多様性

試案表の医療合成語の8割以上が病名となっている(表2)。しかし医療合成語には、病名だけでなく「残存機能」「皮膚脆弱」「局所圧迫」など、複数の一般的な日本語からなる合成語も多く、また、「冠動脈バイパス術」のような専門用語からなる合成語も多い。従って代表性を高めるために、医療分野での多様な文書に出現する医療合成語を追加する必要がある。

現在、非医療従事者が日本語学的な研究・調査を目的に医療施設で蓄積される実際の医療記録データを入手するのは困難である。しかしながら、公開されている標準マスター、例えば診療行為マスター¹⁰や診療ガイドライン¹¹、厚生労働省が公開している国家試験問題文などは、非医療従事者であっても入手可能である。

5. 対処方法と試案表 Ver.3

本章では、(1)実践医療用語の選定方法、(2)語の分割単位の曖昧性、(3)語構造の記述、(4)意味分類についての対処方法を述べる。

5.1 実践医療用語の選定方法

本研究に着手した当初は、全医療合成語 30,881 語の内、一般的な日本語を含む約7千の医療合成語の語構造分析と語構成要素の意味分類を終え最終版の試案表を公開した後に、残りの医療合成語の語構造分析と語構成要素の意味分類に着手する予定であった。

しかし今回の結果から、試案表 Ver.3 を公開した後に、新たに入手可能なデータから分析対象とする医療合成語を選定し分析を行う予定である。

5.2 分割単位の曖昧性

試案表 Ver.3 において、「乳汁分泌 | 抑制」と「乳汁 | 分泌抑制」や「機能 | 低下症」「機能低下 | 症」のように分割方法が複数あった場合、医学事典や用語集などの情報を基に「乳汁 | 分泌抑制」にするなど、どちらか一つにするか、「機能低下症」のように語末が「症」や「障害」などの場合は分割せずに一つの語構成要素とする。

5.3 語構造の記述

分割単位の変更は、当事者が行う必要があるが、語構成要素の抽出方法については、**並列分散形態素解析**(parallel Distributed Morphological Analysis : PDMA)を提案し、加えてPDMAより抽出精度は低いものの、記述が簡便な**複層化形態素解析**(Multi-Layered Morphological analysis : MLMA)を提案している。また、MLMAで語構造を記述したデータから語構成要素を自動抽出するPerlスクリプトを公開している(黒田ら 2023)。

語構造の記述において、並列構造、不連続な構造の記述は、MLMAの記法を用いること

¹⁰ <https://www.ssk.or.jp/seikyushiharai/tensuhyo/kihonmasta/index.html>

¹¹ <https://minds.jcqh.or.jp/>

で記述することができ、また MLMA の記述から自動的に語構成要素を抽出することができる（黒田ら 2023）。

試案表 Ver.3 では MLMA を用い、不連続な構造は原則として記述しないが、並列構造を不連続な構造として記述できる場合は記述する。

5.4 意味ラベル

現在、試案表 Ver.3 の作成に向けて試案表 Ver.2 の語構造を見直し、抽出した 12,336 の語構成要素に意味ラベルを付与している。その際、「#未定」としている一般的な意味を持つ語構成要素については、「行為」「状態」「物」などの大まかな分類方法を決め、該当する語構成要素に意味ラベルを付与する予定である。意味ラベルの列挙方法については、今回は定めない。

6. おわりに

我々は、医療合成語の語構造と意味を分析し、その結果を『実践医療用語—語構成要素語彙試案表』として公開している。本稿では、試案表を日本語学または言語学での調査研究で利用する上での問題を明らかにした。試案表 Ver.2 を改善した試案表 Ver.3 においても、対象医療合成語の代表性の問題、語分割の単位の曖昧性、意味ラベルが確定されていないなどの問題は解決できておらず、日本語学または言語学で試案表を変更することなく利用するのは難しい。

しかし、意味が捨象される機械学習の結果を解釈するのは人であることから、医療従事者が実践医療用語をどのように解釈し利用しているかを知る上で、「医療の観点からみた有意味性」を重視した医療合成語の分析は有用である。

以下は、医療合成語の語構造の分析過程で興味を持ったこと、日本語学的な問いをまとめたものである。これらの問いが、何の役に立つのかは分からない。しかし、以下の問いに面白さを感じ、試案表を改変し、または新たな言語資源の作成に着手し、医療用語の語彙研究に取り組む人が増えることを期待している。

- ① 語構成要素の語順を換えた場合に意味がことなる医療合成語（例：「ウイルス肝炎」と「肝炎ウイルス」、「腎嚢胞」と「嚢胞腎」）の語構成要素の特徴
- ② 語構成要素の語順を換えても意味が同じである医療合成語（例：「境界型糖尿病」と「糖尿病境界型」）の語形成の特徴。なぜ語順の交換が生まれるのか
- ③ 語構成要素の並列構造の意味の分析（例：「[前・後] 十字靭帯」、「[脾・腎] 静脈吻合術」、「[絞扼性・癒着性] 腸閉塞症」、「[歯状核・赤核・淡蒼球・ルイ体] 萎縮症」）
- ④ 語の重複のある医療合成語の分析（例：[乳汁<分泌] 抑制>、先天性 [中枢<神経] 異常>）
- ⑤ 医療の観点からみた一つの語構成要素を更に有意味な短い単位に分割した要素の意味素性の集まりと語構成要素の意味との関係（例：「[気管支肺胞洗浄] 異常」：「気管支肺胞洗浄」は検査法で、気管支から肺胞までを洗浄した排泄液に異常があるという意味）

謝 辞

本研究は JSPS 科研費 JP18H03499 ならびに JP21H03777 の助成を受けている。

文 献

- 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.
- 東条佳奈、黒田航、相良かおる、高崎智子、西嶋佑太郎、麻子軒、山崎誠（2022）「実践医療用語_語構成要素語彙試案表 Ver.2 の構築」『言語資源ワークショップ2022 発表論文集』 pp.2-6.
- 東条佳奈、相良かおる、西嶋佑太郎、麻子軒、山崎誠（2021）「医療用語に含まれる序数詞について」『じんもんこん2021 論文集』 pp.194-199.
- 東条佳奈、麻子軒、相良かおる、高崎智子、山崎誠（2020）「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』 pp.357-364.
- 相良かおる、高崎智子、東条佳奈、西嶋佑太郎、山崎誠（2022）「『急性』を含む病名の語構成」『言語資源ワークショップ2022 発表論文集』 pp. 43-51.
- 麻子軒、相良かおる、高崎智子、東条佳奈、山崎誠（2020）「意味ラベルを用いた『-性』を含む病名の言い換え」『じんもんこん2020 論文集』 pp.283-288.
- 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.
- 山崎誠（2020）「実践医療用語を構成する語の計量的分析」『言語資源活用ワークショップ2020 発表論文集』 pp. 164-173.
- 劉亜斌、里村洋一、佐々木哲明、木村通男、廣瀬康行、山崎俊司（2000）「構造化臨床医学用語集の構築に関する研究」『医療情報学』20 卷6号 pp.513-522.
- 黒田航、相良かおる、東条佳奈、麻子軒、西嶋佑太郎、山崎誠（2023）「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」『言語処理学会第29回年次大会発表論文集』 pp. 772-777.