「言語処理の応用システム」特集号

技術資料

病名アノテーションが付与された医療テキスト・コーパスの構築

荒牧　英治[†]・若宮　翔子[†]・矢野　　憲[†]・永井　宥之[†,††]・
岡久　太郎[†,††]・伊藤　　薫[†]

高度な人工知能研究のためには，その材料となるデータが必須となる．医療，特に臨床に関わる分野において，人工知能研究の材料となるデータは主に自然言語文を含む電子カルテである．このようなデータを最大限に利用するには，自然言語処理による情報抽出が必須であり，同時に，情報抽出技術を開発するためのコーパスが必要となる．本コーパスの特徴は，45,000 テキストという我々の知る限りもっとも大規模なデータを構築した点と，単に用語のアノテーションや用語の標準化を行っただけでなく，当該の疾患が実際に患者に生じたかどうかという事実性をアノテーションした点の 2 点である．本稿では病名や症状のアノテーションを対象に，この医療コーパス開発についてその詳細を述べる．人工知能研究のための医療コーパス開発について病名や症状のアノテーションを中心にその詳細を述べる．本稿の構成は以下の通りである．まず，アノテーションの基準について，例を交えながら，概念の定義について述べる．次に，実際にアノテーターが作業した際の一致率などの指標を算出し，アノテーションのフィージビリティについて述べる．最後に，構築したコーパスを用いた病名抽出システムについて報告する．本稿のアノテーション仕様は，様々な医療テキストや医療表現をアノテーションする際の参考となるであろう．

キーワード：電子カルテ，医療テキスト，症例報告，病名，ICD-10，医療情報

Development of the Clinical Corpus with Disease Name Annotation

Eiji Aramaki [†], Shoko Wakamiya [†], Ken Yano [†], Hiroyuki Nagai [†,††],
Taro Okahisa [†,††] and Kaoru Ito [†]

Sufficient data is required for research on advanced AI. In the field of medicine, especially clinical medicine, information retrieval is necessary to utilize the data fully since the data—mainly clinical records—uses natural language. The corpus we developed in this study has the following strong points: (i) The corpus consists of 45,000 case reports, which is the largest to our knowledge, and (ii) not only did we standardized the terminology and the method for annotation, we also annotated “factness,” which notes whether or not a disease name is actually the state of the patient in a case

[†] 奈良先端科学技術大学院大学, Nara Institute of Science and Technology
[††] 京都大学大学院人間・環境学研究科, Kyoto University Graduate School of Human and Environmental Studies

report. This paper describes the methods to develop the medical corpus for AI research, focusing on the annotation of the disease or symptom name. First, we define the concepts contained in the annotation criteria using examples. Next, we discuss the feasibility of the annotation through giving some indexes such as agreement rate. Finally, we report the development of the disease-name extraction system based on the corpus. We believe this corpus is a good reference for future clinical annotation.

Key Words: *Electronic Health Record (EHR), Clinical Text, Case Reports, Disease Name, ICD-10, Medical Infomatics*

1 はじめに

医療現場で生成される多様なデータ（以下，**医療データ**と呼ぶ）の大部分は自然言語文であり，今後もその状況はただちに変わりそうにない．医療データの利活用としては，診療への応用，もしくは学術研究や政策への応用が挙げられるが，現在，盛んに医療データの利活用の重要性が叫ばれているのは，後者の二次利用である (科学技術振興機構研究開発戦略センター 2017)．二次利用されることが期待される医療データとしては，**健診データ**や**診療報酬データ**がある． 健診データは健康診断の際に作成されるデータであり，検査名と検査値から構成される．健診データは受診者が多く，組織で一括して収集されるため，大規模な医療データとしてよく用いられる．一方，診療報酬データは医療費の算定のために用いられるデータであり，医療行為がコード化されたものである．このデータは厚生労働省が収集し管理するため，同じく大規模な医療データとしてよく用いられる．両データは，数値やコードから構成される構造化されたデータのためコンピュータでの扱いは容易であるが，詳細な情報が含まれていないことが解析の限界となっていた．

そこで，より詳細な情報が含まれる**診療録**，**退院サマリ**，**症例報告**といったテキスト化された医療データの活用に注目が集まっている．診療録とは，病院において患者が受診した際や入院時の回診の際に記述されるテキストであり，詳細な患者情報が記述される．また，退院サマリとは，退院時に記述される情報であり，入院中の診療録の要約である．症例報告も退院サマリと同じく入院時の要約であるが，学会に報告されるものである．他にも，病院内にはテキスト化された医療データが存在しており，本稿ではこれらのテキスト化された医療データ全般を指し，**電子カルテ**と呼ぶ．

電子カルテは，自然言語文が中心となる非構造データであるため，扱いは困難であるが，詳細な情報が記述されており，その量は年々増加しつつある．この動きは，1999 年に医療データであっても，一定の基準を満たした電子媒体への保存であれば，記録として認められる，という法改正が行われて以降，特に急速に進展した．2008 年には，400 床以上の大規模病院で 14.2%，一般診療所で 14.7% であった電子化率は，2014 年には，400 床以上の大規模病院で 34.2% ，一

般診療所で 35.0%と倍以上に増加している[1]．このまま増加すれば，ほとんどの病院で電子カルテが用いられるであろう．

電子化の第一の目的は，病院の運営の効率化によるコスト削減であるが，副次的な利用法として，これまで膨大な労力をかけて行われてきた調査への応用が期待されている．例えば，医薬品の安全に関わる情報や疫学的情報の収集をより大規模かつ容易に実行可能にしたり，これまで不可能であった医療情報サービスも構築可能にすると期待されている．しかし，このような期待は高まるものの，具体的な成功事例は乏しい．これは，電子カルテに多く含まれる自然言語文の扱いが困難であることが原因で，電子カルテの情報を最大限に活用するには自然言語処理が必須となる．

本研究では病名のアノテーション基準を提案し，45,000 例もの症例報告を材料としてアノテーションを行う．このアノテーションでは，症例報告の対象患者の疾患や症状についての情報を整理することを目指し，単に病名のみをマークするだけでなく，症状が患者に発生しているかどうかの区別まで行う．海外では，医療分野における同様のコーパスは政府の協力のもと開発，公開がされているが，日本では公開された大規模コーパスは存在せず，コーパスの仕様についても十分な資料がなかった．本稿では，日本で初となる大規模な医療分野のコーパス開発の詳細について述べる．

本研究が提案するアノテーションは，症例報告のみならず，さまざまな医療テキストへ利用可能である汎用的なものである．また，これが実行可能なアノテーションであることを示すために，複数のアノテーター間における一致率やその問題点などの指標を示し，フィージビリティの検討を行った．最後に，病名アノテーションを利用して構築した病名抽出器についても紹介する．

本コーパスの特徴は，以下の 2 点である．

(1) 従来，小規模な模擬データが配布されるにとどまっていた利用可能な医療分野のコーパス (Morita, Kano, Ohkuma, Miyabe, and Aramaki 2013; Aramaki, Morita, Kano, and Ohkuma 2014, 2016) と比較し，約 45,000 テキストという大規模なデータを構築した点．

(2) 単に用語の範囲をアノテーションしただけでなく，用語で示された症状が実際に患者に生じたかどうかという**事実性**をアノテーションした点．

特に，症状の事実性を記述することは応用を考えると重要である．例えば，以下のような 2 つの応用システムを用いたシナリオを想定できる．

【医薬品副作用調査シナリオ】 ある医薬品 A と医薬品 B がどれくらい副作用を起こすかを比較したいとする．この場合，医薬品 A と医薬品 B で検索して得られたテキストセット A とテキストセット B をつくり，それぞれに出現する副作用と関連した病名の頻度を比較

[1] 厚生労働省医療施設調査より (http://www.mhlw.go.jp/toukei/list/79-1.html)

すればよい．だが，これを実際に行うと，「副作用による軽度の<P>咳嗽</P>は認めたが、<N>間質性肺炎</N>は認めなかった。」[2]といったように，想定はされるが実際には起こっていない副作用も記述される．よって，事実性を判定する必要が生じる．

【診断支援シナリオ】 診断を行う際には，ガイドラインに沿って症状の有無を調べ，合致する診断を下す．これはフローチャートになっており，例えば，意識消失，痙攣あり，嘔吐あり，発熱ありの際に考えられる症状には心筋炎，脳梗塞，脳炎など曖昧性があるが，ここで血液検査を行って炎症所見のない場合は心筋炎が除外される．このような場合，診断がガイドラインに沿っていることを明確にするために，事実性のない症状についても記述される（この例では「炎症所見なし」）．よって，診断支援のデータとして用いる場合には，事実性を判定する必要が生じる．

本研究の貢献は以下の通りである．

(1) 医療テキストへのアノテーションについての詳細な仕様を示した．
(2) 実際にアノテーションした結果について，一致率や問題点などのフィージビリティを議論した．
(3) 本研究で構築したコーパスを用いて病名抽出器を構築し，アノテーションの妥当性を検証した．

2 関連研究

海外でも電子カルテに自由記載された自然言語文からどのように有益な情報を抽出するかについて関心が高まっていた．このため，2006 年から NIH (National Institutes of Health) のサポートで開始された i2b2 NLP Challenge (Uzuner 2008) というワークショップにより，様々なコーパスのリリースが行われている．i2b2 NLP では，退院サマリを利用してコーパスが構築されており，これは，筆者らの知る限り，医療分野のコーパスを最初に公開したワークショップである．その後，日本語ワークショップの MedNLP(Morita et al. 2013; Aramaki et al. 2014, 2016) やヨーロッパでの CLEF e-Health (Kelly, Goeuriot, Suominen, Schreck, Leroy, Mowery, Velupillai, Chapman, Martinez, Zuccon, and Palotti 2014) など，i2b2 NLP の仕様を参考にしたコーパス開発が行われているが，いずれも小規模なものに留まっている．例えば，MedNLP で扱われたコーパスの一部は，言語資源協会 (GSK) により「GSK2012-D 模擬診療録テキスト・データ」として公開されている[3] が，わずか 20 件にすぎない．

[2] <P>で示した病名は事実性のあるもの，<N>で示した病名は事実性のないものを表す．詳しくは 4.2 節．
[3] http://www.gsk.or.jp/catalog/gsk2012-d/

また，本研究でも利用する医学オントロジーである ICD-10[4]と，自然言語文に含まれる疾患名や症状名を対応させるための辞書は，これまでにも何度か作成が試みられている (Fabry, Baud, Ruch, Le Beux, and Lovis 2003; Yamada, Aramaki, Imai, and Ohe 2010; Bouchet, Bodenreider, and Kohler 1998)．しかし，これらの研究は病名辞書を記述することに終始しており，電子カルテ内の文脈がもつ情報が失われている．本研究では，コーパス中に出現する病名に対してアノテーションすることで，文脈も含んだ形で利用可能なリソースの構築を目指す．

本研究は，大規模な日本語の医療コーパスを構築する初めての試みである．これまでにも医療コーパスを用いた研究はあったが，その仕様の詳細については明らかにされておらず，全体像がつかめないのが実情であった．本稿では，今後同様のアノテーションにおいて指針となるように，アノテーションの仕様を実際にコーパスが構築可能な程度の粒度で紹介する．

3　コーパス構築の全体像

3.1　材料

本研究で扱う電子カルテは**症例報告**といわれるテキストである（表 1）．症例報告とは，学会に提出される患者情報の要約である．症例報告には，患者の診断名・転帰，入院時の症状および所見，治療後の経過などが簡潔に記述され，医師の教育や，類似した症例の参考のために参照される．このため，症例報告は患者に相対して記述する診療録よりも高い可読性で記述される傾向にある．症例報告は学会への報告であるため，記録先は異なるが，入院時の患者の要約であるという点で退院サマリ（図 1）と類似しており，症例報告のアノテーションの枠組みを退院サマリに転移することが可能である．

本コーパスの材料となったテキストは日本内科学会に報告された 44,761 の症例報告[5]である．これは，2004 年以降に報告された全症例であり，11,866 施設，26,235 人の医師からなる．また，これらがカバーする診療科は内科全領域である．症例報告が対象としている分野と，各分野における報告数を表 2 に示す．なお，日本内科学会の会員は学会ホームページ[6] を通じて本コーパスのデータを検索，閲覧可能である．また，本コーパスの研究利用についてはウェブサイト[7]にて，利用に関しての最新情報を参照可能なようにしている．

[4] ICD とは，世界保健機関 (WHO) による，疾病及び関連保健問題の国際統計分類 (International Statistical Classification of Diseases and Related Health Problems) の略称であり，世界中で集計された死亡や疾病のデータに基づく分類体系である．ICD にはいくつかのバリエーションがあり，日本では ICD-10 が用いられている．詳しくは 3.4 節を参照．

[5] この段階ではコーパスから除外する複数症例についての報告（3.2 節参照）も含まれているため，完成後のコーパスの症例報告件数とは一致しないことに注意．

[6] http://www.naika.or.jp/

[7] http://mednlp.jp/

表 1 症例報告「頚部硬膜外血腫を合併した特発性血小板減少性紫斑病」の例

症例は８１歳、男性。２０２０年２月起床直後発症の<P>後頚部痛</P>と進行性の<P>四肢麻痺</P>にて救急搬送。口腔粘膜に<P>出血傾向</P>を認め、Ｈｂ　１１．７ｇ／ｄｌ、ＷＢＣ　５１３０／μｌ（分類正常）、血小板は０．９万／μｌと著明に減少、D-ｄｉｍｅｒ　１０．８μｇ／ｄｌ、ＰＴ・ＡＰＴＴ・Ｆｂｇは正常範囲であった。ＭＲＩにてＣ３～胸椎上部の背側硬膜外に<P>血腫</P>を示唆する高信号を認め、これによる頚髄の圧迫が<P>麻痺</P>の原因と考えた。以上より、臨床的に<P>ＩＴＰ</P>を疑ったが、重篤な<P>出血</P>と<P>全身状態不良</P>のため、診断確定を待たずに血小板輸血とデキサメサゾン大量投与（４０ｍｇ，５日間）を直ちに開始した。翌日には血小板数２．７万／μｌと増加し、<P>出血傾向</P>と<P>麻痺</P>の改善を認めた。第４病日に行った骨髄検査では明らかな異常は認めず、小型巨核球がやや増加しており、ステロイド投与後のためＰＡＩｇＧは５４．８ｎｇ／１０Ｅ７と微増であったが、他に<P>血小板減少</P>の原因となる基礎疾患もないことより<P>ＩＴＰ</P>と診断した。その後、血小板数は１５万以上となり、<N>硬膜外血腫</N>についても第１５病日のＭＲＩで消失を確認し、後遺症を残すことなく治癒した。<SKIP/>ＩＴＰは時に重篤な<N>深部出血</N>をきたすが、<P>頸部硬膜外血腫</P>を合併した例は非常に稀である。このような症例に対しては、ＩＴＰのみならず脊髄の<P>圧迫障害</P>を回避するためにも、発症早期のデキサメサゾン大量療法が有用と考えられた。

3.2 除外データ：複数症例を扱った記録

症例報告の中には，ある症状について，１つの病院で観察された複数の患者のことをまとめて報告したものや特定の病気に対し特定の治療法が有効であるかどうかを複数の事例から考察したものが含まれている．このような報告においては，患者１人１人についての記述が少なく，記述されている情報がどの患者に当てはまるものであるかを判断することが難しい場合が多々ある．そのため，特定の患者１名に関する報告のみをアノテーションの対象とすることとした．

3.3 アノテーションの流れ

電子カルテには専門用語が多く含まれており，正確なデータ整理のためには医学知識が必須となる．例えば，「DIC あり」という表現に含まれる「DIC」が「播種性血管内凝固」という疾患を指すといった判断は，医師や看護師，医療事務員といった医療関係者（以降，**医療従事者**）以外の者（以降，**非医療従事者**）にとっては容易ではないため，本来はすべての作業を医療従事者によって行うのが理想的である．しかし，それはコスト的にも人材的にも非常に困難である．最も人数が多い医療従事者は看護師であるが，慢性的に人手不足であり，また，雇用単価も高い（時給換算で 2,000 円〜2,300 円）．そもそも，看護師の本来の職務はアノテーションの作業と大きく乖離しており，熱意を持ってアノテーションに従事可能な人材の確保は容易でない．

そこで，本研究では，データ整理の作業に熟練し，かつ，事務作業とも親和性の高い，医療

病歴総括

患者ＩＤ：I000000002　**氏名**：　**性別**：女　**入院回数**：1
生年月日：　年　月　日　**入院時年齢**：　68年　0か月
住所：
電話番号：
診療科：内科　**病棟**：　**病室番号**：
主治医：　**担当医(研修医)**：
入院日時：　年　月　日　時　分
退院日時：　年　月　日　時　分　**入院日数**：　13日
入院経路：　**紹介の有無**：　**病院名**：　**医師名**：
退院経路：　**紹介の有無**：　**病院名**：　**医師名**：
死亡退院：　年　月　日　時　分　**剖検の有無**：無　**検死の有無**：無
転科：　年　月　日　**転科先**：　**(主治医**：　)
併科受診：
転帰：軽快

最終診断：
＃1 E13　その他の明示された糖尿病

退院時所見：
入院経過の概要

入院目的）TKA術前、血糖コントロール目的

既往歴）
・10年前（58歳）からDM、平成16年（63歳）右大腿骨頸部内側骨折→当院整形外科にて人工骨頭置換術
・10年前頃（58歳）から健康診断でDM指摘されていた。　近くの診療所でフォローしていた。最近はベイスン（0.2mg）2T　1日2回朝･夕。
食事療法はできておらず、ジュース・お菓子の摂取が多かった。
今回整形外科でTKAを勧められた。
4月28日当科受診時ＨｂＡ１ｃ6.6%あり、術前コントロール目的にて5月25日入院となった。

身長142.2cm　体重57,3kg　BMI26.4

治療）
・diet：DM1,200kcal＋Nacl6g食
・drug：ベイスン（0.2mg）2T　1日2回朝･夕にて３PBSでも、FBS＜100と大変良好になった。
・6月6日当科退院となった。

問題点）
＃　網膜症
＃　高血圧
＃　変形性膝関節症

図 1　退院サマリの例（GSK2012-D 模擬診療録テキスト・データから抜粋）

表 2　症例報告の対象分野と各分野における報告数

分野	報告数
消化器	7,553
循環器	6,982
内分泌・代謝	5,973
呼吸器	5,171
血液	4,316
神経	4,176
アレルギー・膠原病	3,614
感染症	3,527
腎臓	3,201
一般	248

事務員の経験者を中心に雇用し，コーパスの構築を行う．ただし，医療事務員は看護師や検査技師などの他の医療従事者と比較し，そもそも人数が多くない．そこで，アノテーションのプロセスを医療事務経験者でなくとも従事可能な部分と，医療事務経験者が必要な部分という以下の 2 つに分けた．

(1) a. **病名タグ付け**
医学知識を用いず，非医療従事者が病名だと判断したもの全てにタグを付与する．

b. **病名コーディング**
上記 (1a) のプロセスでタグ付けされた表現のうち，頻度の高いものから順に医療従事者が医学知識を用いて，病名であるか否かを判断し，病名である場合は後述する ICD コードを付与する．

本稿では，病名の範囲の同定 (1a) を**タグ付け**，病名の分類 (1b) を**コーディング**と呼び，(1a) タグ付けと (1b) コーディングの両方を含む作業を**アノテーション**と呼ぶことにする．

なお，作業の効率化を測るため，予め非医療従事者によってタグ付けがなされた少数のデータを教師データとして，機械学習によって自動で全てのデータにタグ付けしたデータ（以降，**自動タグ付けデータ**）を作成し，(1) の両プロセスにこのデータを利用している．具体的には，(1a) の作業は，自動タグ付けデータをアノテーターが修正（タグの追加・削除）する形で行う．また，(1b) の作業においては，(1a) のプロセスが完了するよりも前に，自動タグ付けデータにおける頻度に基づいてコーディングを行い，(1) の作業が完了次第，自動タグ付けデータからは収集することができなかった表現に対して，(1b) のコーディング作業を行う．

3.4 ICD-10

ICD (WHO 1992) とは疾病及び関連保健問題の国際統計分類 (International Statistical Classification of Diseases and Related Health Problems) の略であり，世界中で集計された死亡や疾病のデータの体系的な記録，分析，解釈および比較を行うため，世界保健機関憲章に基づき，世界保健機関 (WHO) が作成したオントロジー的な性質を持つ分類体系である．ICD には各国の事情を反映したバリエーションがあり，例えば米国では ICD-9-CM, オーストラリアでは ICD-9-AM, 日本では ICD-10 が用いられている．

ICD-10 は，アルファベット 1 桁と数字 2〜4 桁の組み合わせによって表記され，この表記は**コード**（または **ICD コード**）と呼ばれる．それぞれのコードには，体系的に分類された疾病や死因の概念が対応づけられている．また，実際の電子カルテ内に現れる疾病や病名などを表す表現に対して，コードを割り当てる行為は**コーディング**（**ICD コーディング**）と呼ばれる．一般の計算機科学で用いられる，プログラムを構築する意味でのコーディングとは異なるので注意されたい．コードは表 3 のような階層構造を持つ．コードの最初のアルファベット（**軸**と呼ばれることもある）は，感染症や新生物（がん）などの全身症 (A–E)，循環器や消化器系疾患な

表 3 ICD-10 の概要

コード（3 桁目まで）	分類見出し
A00–B99	感染症および寄生虫症
C00–D48	新生物
D50–D89	血液および造血器の疾患ならびに免疫機構の障害
E00–E90	内分泌，栄養および代謝疾患
F00–F99	精神および行動の障害
G00–G99	神経系の疾患
H00–H59	眼および付属器の疾患
H60–H95	耳および乳様突起の疾患
I00–I99	循環器系の疾患
J00–J99	呼吸器系の疾患
K00–K93	消化器系の疾患
L00–L99	皮膚および皮下組織の疾患
M00–M99	筋骨格系および結合組織の疾患
N00–N99	尿路性器系の疾患
O00–O99	妊娠，分娩および産じょく
P00–P96	周産期に発生した病態
Q00–Q99	先天奇形，変形および染色体異常
R00–R99	症状，徴候および異常臨床所見・異常検査所見で他に分類されないもの
S00–T98	損傷，中毒およびその他の外因の影響
V00–Y98	傷病および死亡の外因
Z00–Z99	健康状態に影響をおよぼす要因および保健サービスの利用
U00–U99	特殊目的用コード

ど (F–N)，奇形や新生児疾患 (O–Q)，症状や兆候 (R)，障害 (S–T)，傷病 (V–Y) などの分類記号となっている．さらに，次桁からの数字で詳細な部位などが示される．例えば，「右上肺葉がん」は C341 に分類されるが，最初の 3 桁の C34 が「気管支および肺の悪性腫瘍」を示し，最後の 1 が上肺を示している．ただし，コードに左右の区別はなく，右肺に生じた疾患であるということは C341 というコードから判別することはできない．

なお，本コーパスの規模としては，56 万の病名の出現 (TOKEN) を収載し，異なり病名としては 19,000 種類の病名 (TYPE) をカバーしている．これを ICD-10 コードの種類にまとめると 2,260 コードとなる．なお，ICD-10 は全体として膨大な体系であり，実際には頻出しない病名や日本では存在しない病名も含めると，数万規模の体系となっている．実際に日本で一般的に用いられている ICD-10 コードの数は約 4,900 であり[8]，本コーパスはこのうち約半分（2,260 コード）をカバーしていることになる．

[8] https://www.medis.or.jp

4 病名タグ付け

4.1 病名タグ付けの指針

病名タグ付けは，症例報告に含まれる病名にタグを付与する作業である．しかし，3.3 節で述べた通り医療従事者の確保は容易ではないことから，この作業は特に医療やその他の学術的知識を持たない作業者にも可能なように基準を設定した．また，実際に作業を行ったのは，2017 年 7 月現在で医療従事者・非医療従事者を共に含む 10 名である．以下では，タグ付けの指針や基準の詳細，および作業結果の一致率について述べる．

病名タグ付けは，以下の 3 つの指針に基づいて行った．

(I) 非医療従事者がタグ付け対象の言語的単位を判断しやすい基準を設ける

症例報告においては，同一の症状・疾患の記述であっても，医師によって様々な表現の仕方が用いられる．特に，複数の形態素によって特定の症状・疾患を表している場合，非医療従事者がタグ付けの範囲を決定することが困難である場合が多い．そこで，タグ付けに際しては医療的知識を用いることなく，言語的な情報からタグ付けの有無や範囲の決定を行うことができる基準を設ける．

(II) 病名コードを付与できる可能性を持つ表現を最大限抽出するための基準を設ける

本コーパスは，ICD コードに基づいて，複数の表現がなされている病名を標準化することを目指している．しかし，タグ付けを行う非医療従事者は ICD コードに関する知識を有していないため，データ中に見られる症状のどれが ICD コードを持っているのかを判断をすることが難しい．そこで，ICD コードが付与される可能性を持つ表現を最大限抽出するための基準を設ける．

(III) 症例報告の患者に関する情報（事実性）を整理する

1 節で述べた通り，本研究は症例報告の対象患者の疾患や症状の情報について整理することを目的としている．しかし，実際の症例報告には，患者による罹患が実際に確認された病名だけではなく，存在が否定された病名や施設名等に含まれる病名，さらに特定の症例報告のレベルではなく一般論のレベルで登場する病名も多い．例えば，表 1 の症例報告では，実際の患者に生じた症状ではなく，医学的な知識として「ＩＴＰは時に重篤な深部出血をきたす」といった記述がなされている．そこで，病名アノテーションにおいては，患者に実際に確認された病名と，そうではない病名を区別し，それぞれに対応するタグを設ける．

以下では，上記の指針に基づき行った実際のタグ付け作業の詳細を述べる．

4.2 タグの種類

指針 (III) に則り，本コーパスで用いるタグは次の 3 種とする．

(2) a. **陽性タグ（P タグ：<P> ... </P>）**
患者に関する症状，疾患名で実際に罹患が認められたもの，あるいは疑われたものに対して P タグを付与する．

b. **陰性タグ（N タグ：<N> ... </N>）**
患者に関する症状，疾患名で罹患が否定されたものに対して N タグを付与する．

c. **スキップタグ（SKIP タグ：<SKIP/>...）**
各データ末尾の一般論には P タグ，N タグを付さず，直前に SKIP タグを付すことで，以降のテキストはタグ付け対象外とすることを明示する．P タグ，N タグと異なり，終了タグは付与しない．なお，一般論であっても，データの末尾以外にある場合には SKIP タグを付与せず，N タグを付与する．

症例報告において，一般論はテキスト末尾に考察として記述される場合が多く，考察部分はそれ以前の具体的な症例報告とは性質が異なる．そこで，テキスト末尾に登場する一般論は SKIP タグによって区別し，本コーパスからは除外する．なお，テキスト末尾以外に登場する一般論に関しては，N タグを付与することで患者が実際に罹患した病気と区別する．これらのタグの種類の決定方法については，4.5 節にて例とともに詳細を述べる．

先行研究においては，N タグの内容を区別するものも存在する．例えば，(Aramaki, Miura, Tonoike, Ohkuma, Mashuichi, and Ohe 2009) では，「疑い」，「必要」，「可能性」，「否定」，「延期」，「予定」，「希望」，「勧める」，「方針」といった細かい区別を用いていた．また，NTCIR の MedNLP2 タスクでは，「否定」，「疑い」，「家族歴」を区別していた．しかし，場合によってはこの区別が困難な場合がある．さらに，今回対象とするコーパスが大規模であることから，できるだけタグの仕様を単純化するのが好ましい．想定される応用例においても，N タグを除外して検索したい要求はあっても，N タグの内容を区別して扱うケースは多くないと考えている．以下に想定される応用例を列挙する．

- 作用調査：P タグのみを抽出
- 副診断支援（類似した症状の患者を検索）：P タグのみを抽出
- 診断支援（次に行うべき検査や見るべき所見を予測）：P タグと N タグの両方を抽出
- 病名表現の統計調査：P タグと N タグの両方を抽出
- 入力支援（サジェスト）：P タグと N タグを区別せず抽出

4.3 病名タグ付けの原則

病名タグ付けは，4.1 節に挙げた指針を反映し，下記の原則に基づいて行った．

(3) a. 名詞で表現される病名に対してのみタグを付与する

b. P タグを付与するのは患者が罹患したと医師が判断した病名（陽性所見）とする

原則 (3a) は，指針 (I, II) を反映している．症例報告において，当該の報告を記述した医師によって病名の表現の仕方や表記法が異なるため，そのような表記の揺れに対して，非医療従事者であっても言語的な特徴からタグ付け範囲の決定をすることができる基準を設ける必要がある．そこで，タグ付け対象を名詞（サ変動詞の語幹も名詞に含めることとする）に限定することによって，統一的なタグ付けを目指した．

原則 (3b) は，指針 (III) を踏まえている．症例報告には，実際に患者が罹患していると医師が判断していない病名が多く含まれている．この点を明示化するために，患者が罹患したと医師が判断したと考えられる病名には P タグを付与することで，それ以外の病名と区別した．

次節では各原則が実際のタグ付けにおいてどのように実現されるかを述べる．

4.4 タグ付けの単位

原則 (3a) に示したように本コーパスでは，タグ付け対象を名詞に限定することで，非医療従事者によるタグ付け単位の判断の統一化を図っている．これは，症例報告において以下のような表記の揺れが散見されるためである．

(4) a. 腎機能低下が見られた
b. 腎機能が低下していた
c. 腎機能が高度に障害されていた

「腎機能低下」は ICD コード (N289) を有する症状である．しかし，症例報告には (4) に例示したように，様々な表現の仕方で，同一の事象が記述されている．このような表記揺れを過不足ない単位で医療知識の無いアノテーターが全て採取することは極めて困難である．また，仮に全ての表記揺れを統一的な単位でタグ付けすることができたとしても，そのような表記揺れは病名コーディングの段階で，頻度の少なさからコーディング対象外となってしまうことが予想される．

以降，4.4.1 節から 4.4.7 節では具体的にどのような表現をタグ付け単位として認定したかを詳細に述べる．

4.4.1 複合名詞

複合名詞は，1 つの名詞として扱い，まとめてタグ付け対象とする．

(5) 複合名詞の例
a. <P>軽度網膜血管炎</P>
b. <P>AFP 高値</P>
c. <P>腎機能異常</P>
d. <P>全身倦怠感著明</P>

e. <P>両側肺門リンパ節腫大</P>

f. <P>ｃｌａｓｓ Ⅴ腺癌</P>

また，病名そのものではなくても，患者の症状を表す特定の語彙を含んだものについてもタグ付け対象とする．

(6)　**患者の症状を伴う特定の語彙を含む複合名詞の例**

a. <P>血痰程度</P> [〜程度]

b. <P>皮膚病変痂皮傾向</P> [〜傾向]

c. <P>壊疽部</P> [〜部]

d. <P>腫瘤辺縁</P> [〜辺縁]

e. <P>項部硬直陽性</P> [〜陽性（陰性）]

4.4.2 英語表記・略号

英語表記やアルファベットによる略記も名詞として扱い，タグ付け対象とする．

(7)　**英語表記・略号の例**

a. <P>ｃａｒｃｉｎｏｉｄ ｔｕｍｏｒ</P> [英語表記]

b. <P>ＡＩＰ</P> [英語略記]

4.4.3 修飾句

本コーパスでは，以下のような修飾句を形成する表現についてはタグ付け対象外とする．

(8)　**タグ付け対象としない修飾句の例**

a. 急性肝炎様に<P>ＡＩＨ</P>を発症した

b. ポリープ状の<P>腫瘍</P>を認め

c. 肉芽腫性の<P>炎症</P>

ただし，以下のように，「に」や「の」等の助詞が介入せず，複合名詞となっている場合は，原則 4.4.1 に示したように，合わせてタグを付与する．

(9)　**助詞の介入しない修飾句を伴う複合名詞の例**

a. <P>急性肝炎様症状</P>

b. <P>ポリープ状腫瘍</P>

c. <P>肉芽腫性炎症</P>

4.4.4 動詞

本コーパスでは，以下のような動詞が表す症状はタグ付け対象外とする．

(10)　**タグ付け対象としない動詞の例**

a. 両足が痛み，右膝が腫れてきた

b. 皮膚はいずれも硬くなった

c. 夜はなかなか寝つけない，とのこと

ただし，サ変動詞の語幹については，それ単体で病名を表す名詞と認定できるものに限ってタグを付与する．

(11) 病名を表す名詞と認定できるサ変動詞の語幹の例

a. <P>狭窄</P>していると考えられた

b. <P>腎機能低下</P>する

4.4.5 セパレーション

中黒（・），スラッシュ（/），ハイフン（-）や，読点（,）をはさむ場合は，その前後が独立した名詞の場合，それぞれを別の名詞として個別にタグを付与する．一方，前部ないし後部がもう一方と連結して複合名詞を作る場合は，当該記号を挟んで 1 つのタグを付与する．

(12) 分割された前後が独立した名詞として認定できる例

<P>血痰</P>・<P>下血</P>も出現した

(13) 前部または後部がもう一方と連結する例

a. <P>ウイルス性，細菌性肺炎</P>疑いあり [ウイルス性肺炎＋細菌性肺炎]

b. <N>腸蠕動音低下，亢進</N>ともになし． [腸蠕動音低下＋腸蠕動音亢進]

c. <P>下咽頭、気管、甲状腺浸潤</P>、<P>頚部リンパ節転移</P>と診断され [下咽頭湿潤＋気管湿潤＋甲状腺湿潤]

「および」「ならびに」といった表現は中黒やスラッシュと同様に扱う．

(14) 「および」「ならびに」が使用されている例

a. <P>体幹および四肢運動失調</P>などを認めた。[体幹運動失調＋四肢運動失調]

b. <P>結節性ならびに小浸潤性陰影</P>を認めた。[結節性陰影＋小湿潤性陰影]

4.4.6 丸括弧

丸括弧で囲まれた言い換えの表現がある場合は，個別にタグを付与する．

(15) 丸括弧による言い換え表現の例

<P>胃食道逆流症</P>（<P>ＧＥＲＤ</P>）

4.4.7 特殊記号を含む名詞

症例報告においては，患者の症状を表すために医師の間で慣習的に使用されている記号が現れることがある．そのような記号は，複合名詞を構成する一部として解釈し，まとめてタグ付けする．

(16) 「上昇／低下」の意味で「↑／↓」が使用されている例

a. <P>ＩｇＡ２８００↑</P>
（IgA 高値は慢性肝疾患や感染症等を示す血液検査所見）
b. <P>皮膚ツルゴール↓</P>
（皮膚ツルゴール低下は脱水症、特に低張性脱水症）

(17) 「陽性／陰性」という意味で「(+) ／ (−)」が使用されている例
a. <P>項部硬直 (+)</P>
b. 検尿も<N>蛋白 (−)</N>、<P>潜血 (+)</P>であった

4.5 タグの種類決定

症例報告に登場する病名には，医師が実際に患者に認めたものだけでなく，検査の結果否定された病名や一般論において登場する病名，患者の症状とは関係のない複合名詞の一部として登場するものが多く存在する．本コーパスでは，そのような表現に対応するために，4.2 節に挙げた 3 種のタグ（P タグ，N タグ，SKIP タグ）をアノテーションに用いる．本節では，具体的にどのような表現に各タグを付与するかを述べる．

4.5.1 患者の症状以外の病名を含む複合名詞

患者の症状としてではなく，施設名，手術名，検査名といった複合名詞の一部に含まれた病名にはタグを付与しない．

(18) 患者の症状を表さない病名を含む複合名詞の例
a. かかりつけの■■■内科リウマチ科クリニックより当院紹介受診となった．
b. 左下肢静脈瘤手術を行った．
c. 細菌培養、膠原病検査を行い１週間経過観察

4.5.2 まだ発症していない病名

実際には，まだ発症していないが，今後患者に発症が予想されている病名には N タグを付与する．

(19) 発症が予想されている病名の例
a. <N>気道閉塞</N>の危険が高く
b. <N>肺塞栓症</N>の発症が危惧された

4.5.3 家族歴

家族歴（患者の家族が罹患したことのある症状）は，患者が実際に罹患していない病名であるため，N タグを付与する．

(20) 家族歴の例

a. 母が<N>高脂血症</N>

b. 長兄、祖母にも<N>大動脈疾患</N>の既往歴があり

なお，患者本人の既往歴（過去に患っていた症状）に関しては，患者の罹患した病名であるため，P タグを付与する．

(21) **患者の既往歴の例**

XX 歳時に<P>胃潰瘍</P>のため胃亜全摘術既往歴がある。

4.5.4 罹患が疑われている病名

罹患が疑われている病名に関しては，同一文内で陰性であることが示されている場合に限り N タグを付与し，それ以外は P タグを付与する．

(22) **同一文内で罹患が否定されている例**

a. <N>脳血管障害</N>を疑いＣＴ・ＭＲＩを施行したが異常所見を認めなかった。

b. <N>抗酸菌感染症</N>を疑い喀痰抗酸菌検査を施行したが陰性で、外来での厳重な経過観察とした。

(23) **同一文内で罹患が否定されていない例**

a. 臨床経過より、アミオダロンによる<P>薬剤性肺障害</P>を疑った。

b. 喘息既往や血液検査から<P>好酸球性肉芽腫性多発血管炎</P>を疑った。

また，疑いを表す述部の例としては以下のものが挙げられる．

(24) **疑いを表す述部を伴う例**

a. バルサルバ負荷によるＴＭＦの変化、肺静脈血流と流入血流左室内伝播速度を観察したところ、いずれも<P>左室拡張障害</P>を示唆する結果であった。

b. モニター心電図上、心拍数２００回／ｍｉｎ程度の<P>ｗｉｄｅ ＱＲＳ ｔａｃｈｙｃａｒｄｉａ</P>を認めており、<P>心室頻拍</P>と考え、マグネゾール１Ａを静注後に<N>頻拍</N>は停止した。

c. ■■月■■日縦隔腫瘍摘出し<P>ｇａｎｇｌｉｏｎｅｕｒｏｍａ</P>と診断。良性で全身病態に関連少ないと判断し、<P>ＧＢＳ</P>を想定し、免疫吸着を行った。

d. <P>脳卒中</P>が疑われる<P>神経症状</P>を認めたため脳ＭＲＩを施行、右小脳・右脳幹に比較的最近に発症したと思われる<P>脳梗塞所見</P>を認めた。

e. <P>ヘパリン起因性血小板減少症</P>の可能性を考え抗ヘパリン-ＰＦ４複合体抗体を測定したところ陽性であった。

f. <P>悪性腫瘍</P>も否定できないため、診断目的でエコー下肝生検を施行。

4.5.5 治癒表現

治癒を表す表現が伴っている病名については，症状や疾患が完全に消失したことが明示されている場合のみ N タグを付与する．

(25) 完全に消失したことが分かる治癒表現を伴う例

a. <N>リンパ腫所見</N>は消失しており

b. <N>血尿</N>は陰性化し

なお，症状・疾患が完全に消失したことが明示されていない場合は P タグを付与する．

(26) 症状・疾患が完全に消失したことが判断できない例

a. <P>腫瘤</P>はほぼ消失していた

b. <P>発疹</P>も消退傾向となり

c. <P>心不全症状</P>の軽快を認めた

d. 腸管壁の<P>肥厚</P>は改善した

「寛解」「完全寛解」「奏効」「完全奏効」という用語は，癌の徴候が消失しただけであり，完全な治癒を表すわけではないため，これらを伴う病名には P タグを付与する．

4.5.6 一般論に見られる病名

症例報告に見られる病名には，個別のケースについて言及しているのではなく，これまでに得られている知見の一般的記述の中に登場するものが多く存在する．また，そのような一般論は，本文末尾に記述されることが極めて多い．本研究におけるタグ付け作業は，3.3 節で述べたように，自動タグ付けデータを修正する形で実施したため，本文末尾の一般論に含まれる病名に予め付与されたタグを 1 つ 1 つ人手で削除する作業は非効率的である．そのため，タグ付け作業では SKIP タグを設け，本文末尾の一般論の直前にこれを付すことで，それ以降のテキストに付されたタグを無視するようにした．

(27) 本文末尾に登場する一般論の例

a. ... その後徐々にＡＬＴは低下し、ウィルス量も低下したが、<P>肝不全</P>からの回復には至っておらず、現在も治療中である。<SKIP/>近年ではステロイドフリーの化学療法や、他の免疫抑制剤でも、免疫抑制の回復期におこるＨＢＶの再活性化が数多く報告されている。また劇症化症例においては肝炎発症後のラミブジン投与では効果に乏しく、予後も不良である。化学療法に伴うＨＢＶの再活性化、劇症肝炎はいかなる化学療法においても引き起こす可能性があり、ＨＢｅ抗原やキャリアの肝予備能に関わらず、ラミブジンの予防投与をすることが必要と考える。

b. ... 薬物的除細動は合部調律が続き、その後洞調律となり、薬物療法を行って整脈を維持出来た。<SKIP/>長年に渡り心房細動が続き、左心房の拡大がかなり拡大している症例でも心筋シンチ検査を行い、その所見から、除細動が可能かの判定に役立て

る事が出来ると思われた。

なお，本文末尾以外に登場する一般論については，患者に実際に見られた症例とは区別して，N タグを付与することとした．

(28) **本文末尾以外に登場する一般論の例**

a. <N>平滑筋肉腫</N>は、進行又は再発症例では有効な治療法が確立されておらず、その予後も不良であるのが現状である。今回我々はＧｅｍｃｉｔａｂｉｎｅ／Ｄｏｃｅｔａｘｅｌの化学療法にて予後の改善が得られた後<P>腹膜原発平滑筋肉腫肝転移</P>の１例を経験したので報告する。...

4.6 作業者間一致率

本節でこれまで述べてきた，病名タグ付けの基準の妥当性を調査するため，実際に作業に従事した 10 名のうち 2 名（非医療従事者）がランダムに抽出された 100 件の症例報告を対象に行ったタグ付け結果について，一致率の評価を行った．評価は一方の作業者の作業結果を正解，もう一方の結果をシステムの出力結果とみなし，再現率 (recall)，精度 (precision)，F 値を算出した．正解か否かについては，タグ種類とタグ範囲が共に一致した場合に正解，それ以外は全て不正解とした．結果を表 4 に示す．

まず，病名に関するタグ（P タグ，N タグ）付与の不一致の原因について述べる．P タグについては，「菲薄化」「心室細動」などの医学用語に関する知識の差が原因とみられる不一致がみられた．一方，N タグについては，P タグに比べ一致率が低くなる傾向がみられた．また，作業者間で SKIP タグの範囲が異なる場合，その範囲の P タグと N タグは一方で作業対象となるが，もう一方では SKIP タグの範囲内として処理されることになり，必ず不一致が生じる．このため，P タグ，N タグの一致率は SKIP タグの一致率にも影響される．

SKIP タグについては一致率が顕著に低い値となった．不一致の場合，同一の症例報告に対し，両方の作業者が SKIP タグを付与しているものの，その範囲が異なっている例が散見された．そのような例を，原因とともに (29), (30) に示す．

表 4 タグ付与についての作業者間一致率

タグ種類	適合率 (%)	再現率 (%)	F1
<P>	93.4 (1,379/1476)	81.1 (1,379/1,699)	86.8
<N>	80.0 (128/160)	57.7 (128/222)	67.0
<SKIP/>	20.4 (19/93)	34.0 (19/56)	25.5

表 5　タグ範囲が一致した場合の作業者間タグ種類一致率

適合率 (%)	再現率 (%)	F1
98.2 (1,378/1,403)	99.0 (1,378/1,392)	98.6

(29)　**一般論や考察部分とも，患者に関しての言及とも捉えることが可能で判断が分かれる場合**

a. 考察：不明熱にて発症し、確定診断まで難渋した症例を経験した。<SKIP/>Ａｂｉｏｔｒｏｐｈｉａ　ｄｅｆｅｃｔｉｖａは...

b. <SKIP/>考察：不明熱にて発症し、確定診断まで難渋した症例を経験した。Ａｂｉｏｔｒｏｐｈｉａ　ｄｅｆｅｃｔｉｖａは...

(30)　**位置がほぼ一致しているが，開始時点の見出しの扱いに相違が見られる場合**

a. 【考察】<SKIP/>透析患者における悪性リンパ腫の報告は少ない。...

b. <SKIP/>【考察】透析患者における悪性リンパ腫の報告は少ない。...

このように，何をもって一般論・考察の開始とするかについては判断が難しいと考えられ，今後ガイドラインの検討を要する．また，見出しの取扱いについてもガイドラインの修正が必要である．

両作業者が同一の範囲に P タグもしくは N タグを付与した場合についても同様に，一方の作業者の結果を正解，もう一方の作業者の結果を出力とみなし評価した（表 5）．なお，SKIP タグについては P タグや N タグと同一範囲に付与された例は見られなかった．結果として，F 値が 98.6 と高い一致率を示した．

5　病名コーディング

病名コーディングとは，前節で述べた病名タグ付けの作業によって症例報告の文章中から抽出された症状や疾患を表す表現に対して，ICD コードと標準病名を属性として付与する作業である．なお，病名コーディングは，P タグが付与された表現のみを対象とする．この作業は，先の病名タグ付けとは異なり，医療知識を有する医療従事者が行った．

本研究では，コーパスの作成にあたり 3 名の医療従事者（以下，「**作業者**」と呼ぶ）が病名コーディングの作業を担当した．コーディングにあたっては，3 名の作業者が意見を交換しつつ，P タグが付与された表現に対して可能な限り ICD コードを付与することを目指し，コーディングを実施した．本節では，病名コーディング作業の基本的な手順，およびコーディング作業にあたっての留意点について述べる．

5.1 コーディング作業の手順

病名コーディングの作業には万病辞書を用いた．万病辞書とは，奈良先端科学技術大学院大学が開発しているデータベースであり，ICD コードと，それに対応する標準病名が記載されている[9]．コーディング作業は，P タグが付与された表現と，万病辞書に記載された項目との一致を利用して行った．

具体的なコーディング作業の流れは次の通りである．P タグが付与された表現について，まず，全文一致検索による自動コーディング処理を行い，そこでコーディング処理がされなかったものについては，作業者が人手によるコーディングを行った．詳細な手順を以下に述べる．

5.1.1 自動コーディング

P タグが付与された全ての表現を万病辞書で全文一致検索する．万病辞書の記載に全文一致する項目が見つかった場合，その項目に対応する ICD コードを付与する．

5.1.2 人手によるコーディング

人手によるコーディングは，(I) 全文一致検索に基づくコーディングと，(II) 部分一致検索に基づくコーディングの 2 段階に分かれる．それぞれの手順を以下に述べる．

(I) 全文一致検索に基づくコーディング

先の自動コーディング処理による全文一致検索の取りこぼしを補うため，P タグが付与された表現のうち，略語や英語を伴う表現をパラフレーズしながら万病辞書で検索を行う．全文一致する項目が見つかった場合，その ICD コードを付与する．

例えば，(31) では “Wegener” という英語表記が用いられている．この語は，「ウェゲナー」あるいは「ウェジナー」と読まれることもあるため，「ウェゲナー」もしくは「ウェジナー」としても検索を行う．

(31) Wegener （ウェゲナー，ウェジナー）肉芽腫（M313：多発血管炎性肉芽腫症）

(II) 部分一致検索に基づくコーディング

P タグが付与された表現について，万病辞書と全文一致する項目がみられない場合，部分的に一致する項目を検索したうえで，対応する ICD コードを付与する．部分検索は，P タグを付与された表現が最も多くの修飾語を含む状態から始め，万病辞書内の適切な項目に一致するまで，段階的に修飾語を省略しながら実施する．万病辞書と部分的に一致する項目が見つかった場合，その段階で検索を終了する．

例えば，「LQT2 型 QT 延長症候群」という表現は，万病辞書内に全文一致する項目が存在しない．そこで「LQT2 型」という修飾語を省略し，「QT 延長症候群」で万病辞書を検索するこ

[9] http://www.mednlp.jp/dic-ja.html

表 6　修飾語を伴う病名の例「重症熱性血小板減少症候群（A938：重症熱性血小板症）」

年齢	病期・頻度	重症度	部位	性質・原因	病変・疾患名
—	—	重症	—	熱性	血小板減少症候群

とで，対応する ICD コードである「I490：QT 延長症候群」を見つけることができる．なお，「LQT2 型」は遺伝子の型別を表すと解釈できるため，病名は「遺伝性 QT 延長症候群」とする．

(32) a.　「LQT2 型QT 延長症候群」で検索 → 一致なし

b.　「QT 延長症候群」で検索 → I490：QT 延長症候群

また，タグ付けの段階で欠損したと考えられる情報を補完することで対応する ICD コードの特定が可能になる場合は，情報を補完してコーディングを行った．例えば，(33) は病名の一部が欠けた状態で抽出された表現である．この場合，まず「グレン症候群」で検索を行い，部分一致した項目である「シェーグレン症候群」の ICD コードを付与した．

(33)　グレン症候群（M350：シェーグレン症候群）

修飾語の有無がコーディングに影響する疾患名や病名も存在するため，部分一致の検索を行う際には特に留意する必要がある（5.2.3 節参照）．なお，部分一致検索にあたって修飾語を省略する際は，表 6 のように修飾語が表す意味のカテゴリごとに区分を設けた．

5.2　コーディング作業の指針

本節では，コーディング作業における具体的な指針について，実例を挙げながら述べる．

5.2.1　表記のゆれ

P タグが付与された表現の中には，同じ疾患や症状を表すものであっても異なる表記で出現するものがある．例えば，「R91：胸部異常陰影」という ICD コードに対応する表現は，(34) のようにさまざまな表記で現れる．このような場合は，表記の異なりを捨象してすべて同一の ICD コードを付与した．

(34)　{スリガラス／すりガラス／スリガラス状／すりガラス状／スリガラス様}　陰影

5.2.2　ウェブ上の情報の利用

P タグが付与された表現の中には，略語で表記されたものや希少疾患，作業者に馴染みが薄い病名など，一見するとコーディングが難しいものがある．(35) に例を示す．

(35) a.　電撃性紫斑病（D692：紫斑病）

b.　ペットボトル症候群（E872：ケトアシドーシス）

このような表現についても可能な限り ICD コードを付与するために，ウェブサイトの情報を参考にすることもあった．なお，ウェブサイトの情報の利用頻度にはそれぞれの作業者の間で個

人差があった.

なお，インターネット上の情報を参照するにあたり，情報の信頼性についても考慮した．特に，公的機関のウェブサイトや，日本内科学会症例検索システム「症例くん」[10]，または検索上位のウェブサイトを参考に，最も妥当であると考えられる ICD コードを付与した.

5.2.3 コーディングの精度

同じ病名を含む表現であっても，修飾語の有無によって対応する ICD コードが異なることがある．そのような場合は最も妥当な分類に対応する ICD コードを付与するように留意した．例えば，(36) に挙げたような修飾語は，コーディングに影響する．(37) はその具体例である．また，(38) のように，癌に関わる表現は，「術後」や「再発」などの語の有無によって対応する ICD コードが異なることがある.

(36) 二次性，心因性，1 型・2 型，原発性，中枢性，家族性，遺伝性，細菌性，完全性，再発性，本態性，疾患部位を表す表現など

(37) a. H814：心血管性めまい
b. H811：体位性めまい
c. T752：低音性めまい
d. F456：心因性めまい

(38) a. C169：早期胃癌
b. Z080：早期胃癌術後
c. C761：乳癌術後胸壁再発
d. C798：乳癌術後胸壁転移

5.2.4 複数コーディング

ある 1 つの P タグが付与された表現に対して，複数の ICD コードや病名が対応すると考えられる場合，最大 2 つまで ICD コードおよび病名を付与した.

(39) a. 嘔気嘔吐（R11：嘔気・嘔吐症）
b. 脂肪肝合併 2 型糖尿（K760：脂肪肝・E11：2 型糖尿病）
c. 呼吸障害（J060：呼吸困難・J969：呼吸不全）
d. 全身関節痛（M2550：多発性関節症・M2559：関節痛）

(39a, 39b) のように，2 つの語が並列されて 1 つの表現を構成している場合は，それぞれの単語について万病辞書の検索を行ったうえで，対応する ICD コードおよび病名を 2 つまで付与する．また，(39c) のように，病状の程度を表す表現の意味が曖昧で，病状の重症度が特定できな

[10] http://www.naika.or.jp/meeting/endaikensaku/

い場合は，対応すると考えられる ICD コードを最大 2 つまで付与する．(39d) のように，作業者の間でコーディングについて意見が分かれ，協議の結果 1 つの ICD コードに断定できなかった場合，最大 2 つまで ICD コードを付与する．

なお，P タグが付与された 1 つの表現に対して，3 つ以上の ICD コードが対応すると考えられる場合は，その表現をコーディングの対象外とする．(40) はコーディング対象外とされた表現の例である．

(40) a. 血管炎
　　b. 血栓
　　c. 肉芽腫

作業者間での意見の相違と，その解決方法については，5.4.2 節にて改めて詳述する．

5.2.5 医学的知識の利用

コーディングを行う際には，医学的知識を利用して表現の意味を解釈しなければならない場合がある．そのような場合は，解釈を補った上でコーディングを行った．

ある検査の異常な結果を表す表現には，その検査異常に対応する ICD コードを付与する．

(41) a. 動脈血ガス（R798：血液ガス値異常）
　　b. 膵性胸水（R848：胸水検査異常）
　　c. 陰性 T 波（R943：心電図異常）

ある検査で陽性の反応が出たことを示す表現の場合も，その検査結果に対応する ICD コードを付与する．

(42) 胸水結核菌陽性（R845：胸水結核菌陽性）

表現中の医学用語を解釈することで対応する ICD コードが同定できる場合は，その ICD コードを付与する．例えば (43) では「副腎外」という語が用いられているが，副腎外という人体の部位は存在しないため，「異所性」を含む病名を付与する．

(43) 副腎外褐色細胞腫（D447：異所性褐色細胞腫）

病状の程度を含む表現は，病状の程度を解釈し，それに応じて対応する ICD コードを付与する．

(44) a. 腎機能障害（N289：腎機能低下）
　　b. 腎障害（N289：腎障害）
　　c. 急性腎障害（N289：腎障害）
　　d. 腎病変（N289：腎疾患）
　　e. 腎機能異常（R944：腎機能検査異常）
　　f. 軽度腎機能障害（N289：腎機能低下）

疾患を特定できる臨床所見を表す表現には，対応する疾患の ICD コードを付与する．

(45) a. 骨破壊（M0690：関節リウマチ）
b. ニボー像（K567：イレウス）

症状を表す表現であっても，対応する ICD コードが存在する場合は，そのコードを付与する．

(46) a. 脱水（E86：脱水症）
b. めまい（R42：めまい）

5.2.6 コーディング対象外の表現

以下に挙げる表現は，コーディングの対象外とした．

(47) **患者に生じている疾患，症状を表さない表現**
a. 急性期
b. ステージ II
c. 予後不良

(48) **細胞変性などを表す表現**
a. ○○変化
b. ○○浸潤
c. 形態異常

(49) **病変部位を 1 つに特定できない表現**
a. 多発潰瘍
b. 嚢胞
c. 基礎疾患

(50) **意味が漠然としており，対応する ICD コードが推測できない表現**
a. 拡張
b. 貯留
c. 多発病変
d. 偶発症

(51) **部分一致検索で対応項目が見つからず，ICD コードを推測できない表現**
a. 症
b. 壁運動障害
c. 嚢胞状

(52) **検査所見を表す表現（「R91: 胸部異常陰影」に対応する表現を除く）**
a. ○○状陰影
b. 狭窄所見
c. 壁肥厚

d. low density area

e. 異常集積

5.3 病名コーディングの具体的作業

本節では，コーディング作業の具体的な内容について述べる．特に，作業者間でのコーディング作業の分担およびコーディングの結果について詳しく述べる．

5.3.1 コーディング作業の分担

病名のコーディングは，コーディング対象データの総数 13,207 件のうち，自動コーディング処理がされた 4,603 件を除いた 8,604 件のコーディングを 3 名の作業者が担当した．この 8,604 件のうち，出現頻度が 5,055 回から 30 回までの，1,071 件の表現については，3 名全員がコーディングを行った．残りの 7,533 件については，3 名がそれぞれ分担して個別にコーディングを行った．ただし，判断に迷う場合は 3 名で協議しながらコーディングを実施した．

5.3.2 コーディングの統一作業

上述の通り，高頻度の表現（出現頻度 5,055 回から 30 回の 1,071 件）のコーディングは 3 名の作業者全員が行ったものであるため，3 名の間でコーディングの判断が分かれる部分もあった．この 1,071 件のコーディングについては 3 名で協議を行い，最終的なコーディングを決定した．最終的なコーディングは，以下の基準にしたがって決定した．

(53) a. 1 つの P タグが付与された表現に対し，3 名全員がコーディングの対象外と判断した場合は，その表現をコーディング対象外とする．

b. 1 つの P タグが付与された表現に対し，3 名全員が同一の ICD コードを付与した場合は，その ICD コードを採用する．

c. 作業者の間でコーディングについて見解の相違があった場合は，協議を行い，統一した ICD コードを付与する．

最終的なコーディング決定に至る手順は次の通りである．(I) 作業者各自が行ったコーディングの結果を照らし合わせた後，(II) 作業者間で意見の対立がある部分について協議し，(III) 最終的なコーディングを決定した．

(I) 作業者各自が行ったコーディング結果の照合

まず，それぞれの作業者が個別に行ったコーディングの結果を照合した．この時点で，(53) の基準に従い 3 名の作業者間でコーディング結果が一致した場合，コーディングを確定した．結果を表 7 に示す．

また，作業者間のコーディングの一致率 (%) を以下の式によって求めた．

$$\frac{\text{作業者間のコーディングの一致数}}{\text{コーディングされた表現数}} \times 100$$

表 7 コーディングの照合結果

コーディング結果	件数
コーディング対象外に確定した表現	349/1,071
コーディング対象に確定した表現	498/1,071
コーディングが一致せず協議が必要な表現	224/1,071

表 8 コーディング作業の一致率

コーディング	作業一致率 (%)
h_1, h_2, h_3 間一致	68.9 (498/722)
h_1, h_2 間一致	4.1 (30/722)
h_1, h_3 間一致	4.8 (35/722)
h_2, h_3 間一致	7.8 (57/722)

表 9 最終的に決定したコーディングの結果

コーディング結果	件数
コーディング対象外に確定した表現	408/1,071
コーディング対象に確定した表現	659/1,071
コーディングを保留した表現	4/1,071

結果を表 8 に示す．表 8 では便宜上それぞれの作業者を h_1, h_2, h_3 と表記する．表 8 から，3 名の作業者全員のコーディングを対照すると，7 割程度の一致がみられたことがわかる．

(II) 作業者間での協議

3 名の間でコーディングについて見解の相違がみられた場合，協議を行ったうえで最終的なコーディングを決定した（5.4.2 節参照）．

(III) 最終的に決定したコーディング結果

作業者間での協議の結果，最終的に決定したコーディングの件数を表 9 に示す．

5.4 コーディング作業の問題点

本節では，実際に病名コーディング作業を行った結果，明らかになった問題点について述べる．具体的な問題点としては，ICD コードそのものが抱える問題（5.4.1 節）と，3 名の作業者間での意見の対立によって生じた問題（5.4.2 節）の 2 点がある．以下，それぞれについて述べる．

5.4.1 ICD コードそのものが抱える問題

病名コーディングの作業を行う中で，ICD コードの分類や表記が抱える問題が示唆された．この問題は，ICD コードそのものが原因であるため，コーディング作業者の努力だけでは解決することができなかった．今後の課題といえる．

まず，ICD コードが，どのような基準によって分類されているか不明確なため，P タグを付与された表現に対応する ICD コードの検索や同定が困難になることがあった．例えば，(54) は，類似した病名に対応するが異なる分類をもつ ICD コードの例である．

(54) a. N40：前立腺症
b. N429：前立腺障害

また，さまざまな部位に発生すると考えられる疾患ではあるが，特定の部位についての ICD コードしか存在しないため，推測によるコーディングをしなければならない場合があった．

(55) a. 伸展不良（Q344：軟産道伸展不良）
b. 側副血行（Q258：主要大動脈肺動脈側副血行路）

ICD コードに対応する病名の表記にゆれがあり，検索が困難になることがあった．

(56) a. 「癌」と「がん」
b. 「嚢胞」と「のう胞」

5.4.2 作業者間での意見の対立によって生じた問題

コーディング作業は，3 名の作業者が行ったものであるため，意見の相違が生じることもあった．この問題に対しては，作業者間で協議を行うことによって解決を試みた．本節では，作業者間の意見の相違から生じた問題と，その解決法について述べる．

具体的な問題としては，(I) P タグが付与された表現のコーディング可否についての問題と，(II) P タグが付与された表現にどの ICD コードを付与するかという 2 つの問題が生じた．

(I) コーディング可否の問題

ある P タグが付与された表現をコーディングの対象とみなすかどうかという基準について，3 名の作業者の間で意見が分かれることがあった．本コーパスの目的からすると，P タグが付与された表現には可能な限りコーディングをすることが望ましい．しかし，作業者の主観的な推測によって，本来コーディングの対象外である表現にまでコーディングをするようなことは避けなければならない．

例えば，作業者の中には，(57) に挙げた表現には対応する ICD コードの存在が推測できるため，コーディング可能であるという意見があった（カッコ内はコーディング候補）．しかし，これらの表現は，ある疾患や症状の発生を推測する手掛かりになりうる表現ではあるものの，患者に生じている疾患や症状そのものを示す表現ではない．そこで，このような場合はコーディングの対象外と判断した

(57) a. 転移（C80：転移性腫瘍）
b. 単核球（B270：EB ウイルス伝染単核症，B279：伝染性単核症）
c. 止血困難（R58：出血）

画像所見を表す表現に対しても，可能な限りコーディングを施した．ただし，コーディング

の対象は，実際に患者にその症状が生じていることが明らかな表現に限った．例えば，(58) に挙げた「腫瘤影」や「腫瘤陰影」という表現は，患者の身体に腫瘤が存在することを示唆するため，「R229：腫瘤」の ICD コードが対応するという意見があった．しかし，「腫瘤影」や「腫瘤陰影」という表現は，実際に患者の身体に腫瘤が存在がしているかどうかにかかわらず，腫瘤を疑わせる影が認められるだけの場合でも用いることができる．これに対して，「腫瘤形成」や「腫瘤病変」といった表現は，医師が患者を診察した結果，腫瘤が存在することを確認した場合に用いられる．したがって，最終的に「腫瘤影」と「腫瘤陰影」は，実際に患者の身体に腫瘤が存在していることが明らかではない表現と判断して，コーディングの対象外とした．

(58) a. 腫瘤影（コーディング対象外）
b. 腫瘤陰影（コーディング対象外）
c. 腫瘤像（R229：腫瘤）
d. 腫瘤形成（R229：腫瘤）
e. 腫瘤病変（R229：腫瘤）

なお，異常陰影を表す表現も画像所見であるが，「R91：胸部異常陰影」のみ ICD コードが存在するため，胸部の陰影を表すと解釈できる表現についてはコーディングを行った．

(59) スリガラス陰影（R91：胸部異常陰影）

(II) 対応する ICD コードの問題

P タグが付与された表現の中には，作業者によって，どの ICD コードを付与するか見解が分かれたものがあった．そのような場合，3 名の作業者で協議したのち，統一した ICD コードを付与するか，統一できない場合は 2 つまで ICD コードを付与した．(60) は多数意見を採用した例である．なお，以下では 3 名の作業者をそれぞれ h_1, h_2, h_3 と表記し，それぞれが行ったコーディング結果をカッコ内に例示する．

(60) 心窩部不快感（R198：心窩部不快）
h_1, h_2, h_3：（R198：心窩部不快，R198：心窩部不快，R908：胸部不快）

なお，ICD コード付与の判断にあたっては，必ずしも多数意見を採用するのではなく，特に重要と思われるものであれば少数意見も採用した．(61) はその例である．

(61) 虚血性小腸炎（K559：虚血性腸炎）
h_1, h_2, h_3：（K529：小腸炎，K529：小腸炎，K559：虚血性全腸炎）

また，1 つの表現に対応する ICD コードを 1 つに断定できない場合は，5.2.4 節に示した基準に依拠して，2 つまで ICD コードを付与した．

(62) 呼吸障害（J060：呼吸困難・J969：呼吸不全）
h_1, h_2, h_3：（R060：呼吸困難，J969：呼吸不全，R060：呼吸困難）

6 応用システム：病名抽出器の構築

ここまで，医療テキストコーパスの構築方法について述べてきたが，本節ではこのコーパスを用いた応用システムの可能性について議論する．本コーパスの最も素朴な応用は，このコーパスを学習データとして類似症例検索システムや診断支援システムといった，様々な高次の応用システムで利用可能な病名抽出器を作ることである．以下では，タグ付けされていない医療テキストから自動で病名を抽出する病名抽出器の概要について述べる．なお，前節までは疾患名・症状名などを区別してきたが，本節では 4 節でタグ付け対象とされていたものを単に**病名**と呼ぶことにする．

6.1 病名抽出器の処理

前節までに説明した医療テキストコーパスを教師データとして用いて，病名を自動で抽出する病名抽出器を開発した．以下では，本病名抽出器の処理方法について述べる．提案する病名抽出器は以下の 2 つの処理を同時に行う．

(i) 事象認識 (ER): 医療テキストにおける病名および疾患名を識別する．これは，一般的な固有表現認識タスクと類似した処理である．以降，この処理を **ER** (Entity Recognition) とも呼ぶ．

(ii) 陽性／陰性 (P/N) 分類：テキスト中の P タグと N タグの区別を行うタスクである．以降，この処理を **P/N 分類**と呼ぶ．

病名抽出器の 2 つの処理は，4 節で説明したアノテーションによる病名タグと対応している．タグの学習にあたっては，文字単位で事象認識と事象の事実性判別を同時に系列ラベリングの問題として解いた．一般的に，医療テキストには長く複雑な複合名詞（例えば，「傍大動脈リンパ節郭清」など）や，ひらがなのみからなる医療用語（例えば，「びまん」など）が多く出現することにより，形態素解析の誤りがしばしば発生する．そのため，病名抽出器では，単語単位ではなく，より頑健な文字ベースでの解析 (Asahara and Matsumoto 2003) を採用した．図 2 に文字ベースでの系列ラベリングと単語ベースでの系列ラベリングの違いを示す．

事実性の判定に関しては，これまでにも多くの先行研究があるが (北川，小町，荒牧，岡崎，石川 2015; 松田，吉田，松本，北 2016)，本研究では一般的かつ実装が簡易な手法でコーパスの規模と精度の関係を調査するため，事象認識のラベルとして事実性を表現した．通常，事実性判定（P/N 分類）は事象認識の後に適用される．しかし，P/N 分類に必要な情報は ER で必要な情報と重複する部分も多い．例えば，「〜が認められる」「〜が認められない」は，ともに病名出現の大きな手がかりであるとともに，P/N 分類の手がかりにもなる．このため，ER と P/N 分類の 2 タスクを 1 つに融合する方式を採用した．

6.2 実験

6.1 節では，医療テキストコーパスの応用システムとして，病名抽出器の処理方法について述べた．以下では，実際に病名抽出器を用いて症例報告に自動タグ付けを行い，精度について評価を行った．まず，教師データにはタグ付けにより病名に対して P タグ，N タグが付与された 500 件の症例からなるコーパスを用いた．なお，本タスクは病名抽出を目的としているため，コーパスの SKIP タグを削除し，本来 SKIP タグがあった箇所以降も通常の P タグ，N タグの基準によってタグ付けを行った．前処理として，初めにテキストデータを文に分解し，一般の固有表現抽出の手法にしたがって，文字単位で開始 (B), 内側 (I), 外側 (O) の IOB2 ラベルを付与した （図 2）．系列ラベリングの学習には CRF[11] を用いた．表 10 に CRF で用いた文字ベースの特徴テンプレートを示す．特徴として表層文字と文字種（漢字，ひらがな，カタカナ，英数字）のみを用いた．ウィンドウサイズは前方 2 文字，後方 5 文字に設定した．後方のウィンドウサイズを大きく設定したのは，P/N 分類の手がかりとなる，否定に関わる述語が病名の後方に現れるためである．

評価は 500 件の症例報告を用いた 10 分割交差検証で行い，結果は ER と P/N 分類を個別に評価した．表 11 に ER の結果，表 12 に P タグ，N タグのそれぞれの抽出性能を示す．評価

入力テキスト
腫瘤は肝細胞癌ではなく肝の孤立性形質細胞腫と診断された.

単語ベース系列ラベリング

腫瘤	は	肝	細胞	癌	で	は	なく	孤立	性	形質	細胞腫	と	診断	さ	れ	た
O	O	B-N	I-N	I-N	O	O	O	B-P	I-P	I-P	I-P	O	O	O	O	O

文字ベース系列ラベリング

腫	瘤	は	肝	細	胞	癌	で	は	な	く	孤	立	性	形	質	細	胞	腫	と	診	断	さ	れ	た
O	O	O	B-N	I-N	I-N	I-N	O	O	O	O	B-P	I-P	I-P	I-P	I-P	I-P	I-P	I-P	O	O	O	O	O	O

タグ付き出力テキスト
腫瘤は<N>肝細胞癌</N>ではなく肝の<P>孤立性形質細胞腫</P>と診断された.

図 2 提案する病名抽出器による医療テキストからの病名抽出の例，ならびに，単語ベースの系列ラベリングと文字ベースの系列ラベリング（本病名抽出器で適用）の比較

表 10 文字ベース CRF における特徴テンプレート

特徴	N-gram	ウィンドウサイズ（文字）
表層文字	1, 2	−2, −1, 0, 1, 2, 3, 4, 5
文字タイプ	1, 2	−2, −1, 0, 1, 2, 3, 4, 5

[11] http://taku910.github.io/crfpp/

表 11　病名抽出器による ER の精度

	適合率 (%)	再現率 (%)	F1
単語ベース CRF	92.0	79.9	85.5
文字ベース CRF	91.1	82.6	86.6

表 12　P/N 分類の精度

(a) P タグ抽出性能

	適合率 (%)	再現率 (%)	F1
単語ベース CRF	85.4	79.4	82.3
文字ベース CRF	85.3	81.5	83.3

(b) N タグ抽出性能

	適合率 (%)	再現率 (%)	F1
単語ベース CRF	67.6	33.3	44.4
文字ベース CRF	73.5	44.7	55.4

は CoNLL2000[12]で提供されたツールを用いた．比較のために単語ベース CRF での結果も並記する．

結果としては，ER, P/N 分類いずれにおいても，若干であるが文字ベースによる手法が単語ベースによる手法の性能を上回った．ER については 85.0 以上の高い精度であり，P タグについても 80.0 以上の高い精度での抽出に成功している．一方，N タグについては，文字ベースでも 55.0 と低い精度であった（いずれも評価指標は F 値）．この原因の 1 つとしては，P タグに比べて N タグの出現頻度が低いことが考えられる．また，もう 1 つの原因として，陰性であると判断するためには，設定よりも大きな文脈を要する場合があり，CRF ではこれが困難であることが挙げられる．

本医療テキストコーパスを利用し，いかに N タグを高い精度で捉えられるかが今後の課題の 1 つである．なお，本システムはウェブサイト[13] にて配布している．

7　おわりに

本稿では，自然言語処理による電子カルテからの情報抽出に必須となる，病名がアノテーションされたコーパスの開発について述べた．また，実際にアノテーターが作業した際の問題を，一致率を含め議論を行った．さらに，コーパスを用いて構築した病名抽出器を題材に，コーパスの応用可能性について議論した．本稿のアノテーション仕様が，今後の医療分野におけるコーパス開発の一助となることを祈念する．

[12] https://www.clips.uantwerpen.be/conll2000/
[13] http://sociocom.jp/parser.html

謝　辞

本研究の一部は国立研究開発法人日本医療研究開発機構（AMED）の臨床研究等 ICT 基盤構築研究事業：総合診療医の診療支援及び診療業務効率化の支援基盤構築に関する研究（課題番号：16930323）の支援によって行われた．また，本論文の内容の一部は，言語処理学会第 23 回年次大会で発表したものである (荒牧，岡久，矢野，若宮，伊藤 2017).

参考文献

Aramaki, E., Miura, Y., Tonoike, M., Ohkuma, T., Mashuichi, H., and Ohe, K. (2009). "Text2table: Medical Text Summarization System Based on Named Entity Recognition and Modality Identification." In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pp. 185–192. Association for Computational Linguistics.

Aramaki, E., Morita, M., Kano, Y., and Ohkuma, T. (2014). "Overview of the NTCIR-11 MedNLP-2 Task." In *Proceedings of the 11th NTCIR Conference*, pp. 147–154.

Aramaki, E., Morita, M., Kano, Y., and Ohkuma, T. (2016). "Overview of the NTCIR-12 MedNLPDoc Task." In *Proceedeings of the 12th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 71–75.

荒牧英治，岡久太郎，矢野憲，若宮翔子，伊藤薫 (2017). 大規模医療コーパス開発に向けて. 言語処理学会第 23 回年次大会発表論文集, pp. 1200–1203. 言語処理学会.

Asahara, M. and Matsumoto, Y. (2003). "Japanese Named Entitiy Extraction with Redundant Morphological Analysis." In *Proceedings of HLT-NAACL 2003*, pp. 8–15.

Bouchet, C., Bodenreider, O., and Kohler, F. (1998). "Integration of the Analytical and Alphabetical ICD10 in a Coding Help System. Proposal of a Theoretical Model for the ICD Representation." *Medinfo 1998*, **9** (1), pp. 176–179.

Fabry, P., Baud, R., Ruch, P., Le Beux, P., and Lovis, C. (2003). "A Frame-based Representation of ICD-10." *Studies in Health Technology and Informatics*, **95**, pp. 433–438.

科学技術振興機構研究開発戦略センター (2017). 研究開発の俯瞰報告書：ライフサイエンス・臨床医学分野 （2017 年）. 技術資料, 国立研究開発法人科学技術振興機構.

Kelly, L., Goeuriot, L., Suominen, H., Schreck, T., Leroy, G., Mowery, D. L., Velupillai, S., Chapman, W. W., Martinez, D., Zuccon, G., and Palotti, J. (2014). "Overview of the ShARe/CLEF eHealth Evaluation Lab 2014." In *Information Access Evaluation: Multilinguality, Multimodality, and Interaction*, Vol. 8685, pp. 172–191. Springer, Heidelberg; New York; Dordrecht; London.

北川善彬，小町守，荒牧英治，岡崎直観，石川博 (2015). インフルエンザ流行検出のための事実性解析. 言語処理学会第 21 回年次大会発表論文集, pp. 218–221.

松田紘伸，吉田稔，松本和幸，北研二 (2016). Twitter を用いた病気の事実性解析及び知識ベース構築. In *Proceedings of the 30th Annual Conference of the Japanese Society for Artificial Intelligence*, pp. 1–4.

Morita, M., Kano, Y., Ohkuma, T., Miyabe, M., and Aramaki, E. (2013). "Overview of the NTCIR-10 MedNLP Task." In *Proceedings of the 10th NTCIR Conference*, pp. 696–701.

Uzuner, O. (2008). "Second i2b2 Workshop on Natural Language Processing Challenges for Clinical Records." In *AMIA Annual Symposium Proceedings*, pp. 1252–1253.

WHO (1992). *ICD10: International Statistical Classification of Diseases and Related Health Problems*. World Health Organization. `http://www.who.int/classifications/icd/en/`.

Yamada, E., Aramaki, E., Imai, T., and Ohe, K. (2010). "Internal Structure of a Disease Name and Its Application for ICD Coding." *Studies in Health Technology and Informatics*, **160** (2), pp. 1010–1014.

略歴

荒牧　英治：2000 年京都大学総合人間学部卒業．2005 年東京大学大学院情報理工系研究科博士課程修了．博士（情報理工学）．以降，東京大学医学部附属病院特任助教を経て，奈良先端科学技術大学院大学特任准教授．医療情報学，自然言語処理の研究に従事．

若宮　翔子：2013 年兵庫県立大学大学院環境人間学研究科博士後期課程修了．博士（環境人間学）．以降，京都産業大学コンピュータ理工学部研究員を経て，2015 年より奈良先端科学技術大学院大学博士研究員．ソーシャル・コンピューティングに関する研究に従事．情報処理学会会員．

矢野　　憲：2009 年広島大学大学院工学研究科情報工学専攻博士後期課程修了．博士（工学）．以降，大阪大学臨床医工学融合研究教育センター技術補佐員，福岡大学工学部ポストドクター，国際電気通信基礎技術研究所研究技術員を経て，2016 年より奈良先端科学技術大学院大学博士研究員．機械学習，自然言語処理に関する研究に従事．情報処理学会，電子情報通信学会，人工知能学会各会員．

永井　宥之：2017 年京都大学大学院人間・環境学研究科修士課程修了．修士（人間・環境学）．現在，同大学院博士後期課程在学中．専門は日本語学，認知言語学．日本認知言語学会会員．

岡久 太郎：2016 年京都大学大学院人間・環境学研究科修士課程修了．修士（人間・環境学）．現在，同大学院博士後期課程在学中．専門は認知言語学．コミュニケーション研究，マルチモーダル研究．日本語用論学会，日本社会言語科学会各会員．

伊藤 薫：2012 年京都大学大学院人間・環境学研究科修士課程修了．修士（人間・環境学）．現在，奈良先端科学技術大学院大学研究員．自然言語処理に関する研究に従事．専門は認知言語学，談話・テクスト言語学．言語処理学会，日本認知言語学会，日本語用論学会各会員．

（2017 年 5 月 22 日 受付）
（2017 年 8 月 7 日 再受付）
（2017 年 9 月 21 日 採録）