読影所見タグ付け基準

伊藤　薫 荒牧　英治 黒橋　禎夫

　　　　NAIST 京都大学

Ver. 5.3

2019/2/7

1 概要 3

1.1 ガイドライン中の記法について 3

1.2 例文について 3

1.3 改行について 3

2 タグ種類 4

2.1 病名・症状タグ (Diseases and Symptoms)，確実性属性(certainty) 4

2.1.1 certainty=”positive”の場合 4

2.1.2 certainty=”suspicious”の場合 5

2.1.3 certainty=”negative”の場合 5

2.1.4 TNM分類 6

2.2 臓器・部位タグ (Anatomical Entities) 6

2.3 特徴・尺度タグ (Features and Measurements) 7

2.4 変化タグ (Change) 8

2.5 保留タグ (Pending) 8

3 複数のタグ種類について共通のルール 9

3.1 並列 9

3.2 誤記について 9

3.3 　　「術後」を含む表現 10

3.4 タグ範囲について 11

4 タグ付け例 11

5 検討中の事項 12

5.1 変化タグへのcertainty属性導入 12

6 解決済みの事項 12

6.1 特徴・尺度タグへのcertainty属性導入 12

# 概要

　本ガイドラインでは，創薬標的を目標とした医療人工知能開発研究で用いられる学習データ作成のためのアノテーション基準について説明する．アノテーションの対象は病名・疾患名，部位名などの医学的な概念とする．具体的な作業内容は，電子カルテや読影所見など，実際の医療テキストに出現した表現のうち，上記の概念に相当する範囲に対してタグを付与することである．以下，本節では本ガイドライン全体に関する留意点，2節では各タグの説明，3節では複数のタグ種類に共通するルールについて，4節では実際のアノテーション例，5節では議論中の事項，6節では解決済みの事項について記述する．

## ガイドライン中の記法について

　本ガイドラインでは，アノテーション例を示す際にXMLタグと書式変更の2種類を使用する．実際の作業ファイルにはXML形式でタグを付与するが，可読性のため本ガイドライン内では書式変更を併用する．タグと書式の対応関係は2節で最初にタグ名（一部は属性名）が出てきた際に，そのタグを該当する書式で記述することで示す．ガイドライン内において，XMLタグは3節で説明対象となっていて，かつ，初回登場時にのみ付与し，他の場合は書式の変更で示す．また，仕様について議論が必要なもののうち，特定の事項に関するものはコメント機能で示す．

## 例文について

　本ガイドラインで使用している例文は疑似文章であるため，実際の医療知識に合致しない文面を含む．

## 改行について

　読影所見には改行が含まれており，作業用ファイルでも行を変えることで表現している．この際，ID列に記載した番号を用いて読影所見の同一性を示す．例えば，次の文を作業用ファイルで扱いたいとする．

1. ＸＸＸＸ年ＸＸ月ＸＸ日のＣＴ検査の結果と比較しました。原発性肺癌と考えます。

＃４Ｌ縦隔にリンパ節あり、短径は１．９ｃｍです。前回より増大しています。転移が疑われます。

腹水なし．脳転移を疑う所見なし。骨転移を疑う所見なし。

上の文章は3行から構成されている．これにID=3が割り当てられていた場合，作業用ファイルでは以下の図1のように表示される．赤色で囲んだ部分が分割された本文，青色で囲んだ部分が文章IDを示す．

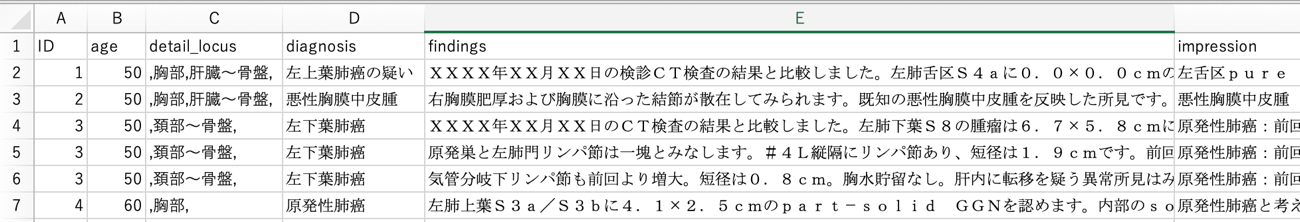


図 1　作業ファイル内における本文改行の表現

# タグ種類

## 病名・症状タグ (Diseases and Symptoms)，確実性属性(certainty)

* 対象

病名・症状タグは病名・症状を表す表現に付与する．その病名・症状が実際に見られた場合は，positive, 存在が疑われた場合はsuspicious, 存在が否定された場合はnegativeを確実性属性の値とする．

* XMLタグ

<d certainty={“positive” or “suspicious” or “negative”}>

※ <d>タグおよびcertainty属性から構成され，certaintyはpositive, suspicious, negativeいずれかの値を取る．

### certainty=”positive”の場合

* 手がかりとなる表現

「認めます」「見られます」

* 例

1. 右肺に<d certainty=“positive”>一様すりガラス影</d>が散在。
2. 感染の合併が見られます。
3. 左第1肋骨に骨折後の変化が認められます。

（骨折後の変化は単なる変化ではなく，「骨折後の変化」という表現が一体となって病状を表しているので，変化タグではなく疾患・症状タグを付与．）

1. 間質性肺炎の増悪傾向

（「の」で連結されているが，後半は変化を表すので別々にタグを付与．）

1. 右肺S2結節

（「右肺S2」は部位名だが，複合名詞を形成しているため一括でタグを付与．）

1. 一方右肺下葉の濃度上昇はわずかに減少している

（「の」で連結されているが，前半は部位を表すので別々にタグを付与．）

### certainty=”suspicious”の場合

* 手がかりとなる表現

「疑います」「否定はできません」「可能性があります」「鑑別にあがる」

* 例

1. <d certainty=“suspicious”>急性肺炎</d>の増悪傾向を疑いますが
2. NSIP感染の合併は否定できません。

### certainty=”negative”の場合

* 手がかりとなる表現

「認めません」「○○なし」「○○は消失」

* 例

1. <d certainty=“negative”>小リンパ節腫大</d>を指摘できません。
2. 左肺尖部の気胸嚢胞は消失しています。

（症状がなくなる変化を表す表現があった場合にもこのタグを付与．）

### TNM分類

タグ付け対象の文章にはTNM分類と呼ばれるがんの分類に関する記号が頻繁に出現する．これは本文中ではT, N, Mの前後に数字やアルファベットを伴うという形で改行前に出現することが多い．

TNM分類の記号にも以下の例のように病名・症状タグを付与する．

1. 原発性肺癌と考えます。ｃＴ１ｃ

縦隔・肺門に病的有意なリンパ節腫大を認めない。Ｎ０

胸水貯留なし。

TNM分類については予め自動で症状・病名タグとcertinty=”positive”属性を付与済みなので，必要に応じて修正作業を行う．また，一部の作業用ファイルではTNM分類記号後の改行を整形しているが，自動で行っているため不自然な箇所も含まれる．

## 臓器・部位タグ (Anatomical Entities)

* 対象

臓器・部位をはじめとした場所を表す表現に付与する．「辺縁」「内部」のように抽象的な表現も含む．

* XMLタグ

<a>

* 例

1. <a>右肺</a>に小リンパ節が散在。
2. 両側肺下葉優位に不均一な濃度上昇が認められます．
3. 左第3肋骨に骨折後の変化が認められます。
4. 両側肺下葉胸膜直下の結節
5. 内部空洞の辺縁に沿って充実部分が認められます

## 特徴・尺度タグ (Features and Measurements)

* 対象

病名・症状の特徴及び尺度・値・範囲・程度に関する修飾語句に付与する．程度表現（「軽度」など）も含めるが，変化を表す表現と連続している場合は変化タグに含め，特徴・尺度タグには含めない．

* XMLタグ

<f>

* 例

1. <f>境界明瞭・辺縁平滑</f>な結節影が認められます。
2. 縦隔内にリンパ節が散見されます
3. 不均一な濃度上昇
4. 局在性の均等影
5. 両肺にびまん性に気管支拡張影が出現
6. 径3cm以下のリンパ節腫大を散見する
7. 有意なサイズのリンパ節腫大を認めません。
8. Positive sizeのLNは認められません。
9. 少量の両側肺胸水あり。
10. 軽度の陰影を認めます。

原則として臓器・部位タグは上記のような（複合）名詞を単位として付与することを想定しているが，「○○術後」の「○○」に部位を表す表現が含まれる場合は，その表現にも付与する．詳細は3章を参照．

## 変化タグ (Change)

* 対象

増悪，減少など，症状の変化を表す表現に付与する．「骨折後の変化」に見られるように，「変化」という単語自体は必ずしも変化タグの付与対象とはならない．変化を表している場合でも病名・症状を表す複合名詞に組み込まれている場合はこのタグを使わず，病名・症状の一部とする．変化がないことを表す場合は付与しない．

* XMLタグ

<c>

* 例

1. 急性肺炎の疑い　前回より<c>少し増悪</c>
2. 右肺下葉の濃度上昇はわずかに増強

（「濃度上昇」は病名・症状の一部とする）

1. 左肺尖部のびまん性気胸腔は消失しています。
2. 全体的に中程度増強しています。

（「中程度増強」は連続しているので変化タグに含める）

1. 収縮性変化が強くなっています。

* 付与しない場合の例

1. 全体としては著名な変化は指摘できない

（変化なし，付与しない）

## 保留タグ (Pending)

* 対象

上記のタグをつけるべき専門用語と思しいが，自身では判断がつかず専門家の判断が必要な表現に付与する．アノテータ自身が必要だと思う状況で使用する．

* XMLタグ

<p>

* 例

1. 胸膜直下は<p>spare</p>されており
2. 縦隔内に短径4cm程度までのLNsが認められます。
3. Coronaryに網状影

# 複数のタグ種類について共通のルール

## 並列

　中黒(・)，スラッシュ(/)，ハイフン(-)，読点(，)のような記号や「もしくは」「および」「ならびに」のような表現をはさんで並列されている場合は，一括でタグを付与する．

## 誤記について

本文に誤記と思われる表現が見られた場合は，タグにcorrection属性を加え，正しいと思われる表記をcorrection属性の値として記入する．本文に「すりが荒らす影」という表現があり，これを「すりガラス影」の誤表記だと判断した場合のタグ付け例を以下に示す．

1. <d certainty="positive", correction="すりガラス影">すりが荒らす影</d>

## 「術後」を含む表現

読影所見では「術後」を含む表現が頻出するが，場合により付与すべきタグの種類が異なる．

* 「○○術後」

「○○術後」のように複合名詞の最後に「術後」が出現し，全体として「○○という手術の後」という意味になる場合は原則としてタグを付与しない．但し，以下のように臓器・部位タグに相当する表現の一部に「術後」が含まれる場合は臓器・部位タグに相当する表現のみにタグを付与する．

1. 胸部左肺上葉切除術後

* 「（△△）術後○○」

「術後」が複合名詞の最初または中間に出現し，全体として本タグ付け作業の対象となる意味を指す場合（病名・症状タグの場合が多い）は，まとめて全体の意味に相当するタグを付与する．

1. 術後左肺胸水

* 「△△術後　○○」（スペースが含まれる場合）

「△△術後　○○」のようにスペースが含まれる場合は以下の例のように「△△術後」と「○○」について別々に判断し，タグを付与する．「△△」部分に臓器・部位名がある場合はタグ付けを行う．

1. 胃術後　局所再発

## タグ範囲について

タグ範囲については原則として，病名・症状タグ及び臓器・部位タグの場合に複合語程度，特徴・尺度タグの場合は病名・症状タグの修飾部，変化タグの場合には動詞句を想定しているが，全体としてまとめてタグ付けした方がよいと判断した場合は，以下の例文に見られる「細菌・ウイルス・ニューモシスチスのような感染症の合併」のようにまとめて付与する．文脈により様々な場合があるので，作業者がまとめるかどうか迷った場合はまとめてタグ付けを行う方針とする．

1. 間質性肺炎の<c certainty=“suspicious”>増悪</c>や細菌・ウイルス・ニューモシスチスのような感染症の合併が疑われます。

# タグ付け例

* 例1

CT-①　胸部単純CT　胃がん術後

両側胸膜直下の肺底部に右側優位に網状影や牽引性気管支拡張像を認め、蜂巣肺は認めません。Possible UIP patternを疑います。右肺に斑状すりガラス影が散在。一連の間質性肺炎に伴う変化で矛盾ありません。有意なリンパ節腫大を指摘できません。胸水貯留を認めません。大動脈の石灰化を認めます。甲状腺両葉に小LDAを認めます。両腎嚢胞あり。

診断　間質性肺炎の疑い

* 例2

CT-②　胸部単純CT

右側肺底部に右側優位に網状影や牽引性気管支拡張像を認めます。胸膜直下はspareされており、NSIP　patternを疑います。右肺に斑状すりガラス影が散在。一連の間質性肺炎に伴う変化で矛盾ありません。前回より少し陰影増強を認めます。ｃＴ１ｃ

縦隔内に小リンパ節を散見しますが、有意なサイズのリンパ節腫大を指摘できません。Ｎ０

胸水貯留を認めません。大動脈の石灰化を認めます。甲状腺両葉に小LDAを認めます。前回と著変なし。

診断　間質性肺炎の疑い　前回より少し増悪

# 検討中の事項

## 変化タグへのcertainty属性導入

変化タグについてもcertainty属性を導入すべき例があるが，確実性スコープの判断がアノテータにとって負担となる可能性があり，導入検討中である．

　具体的には，アノテータに以下の区別を要求することになる．書式変更では表現しにくいため，XMLタグでそれぞれの例を示す．上の例では急性肺炎自体は疑いの対象ではなく，増悪傾向であることが疑いの対象となっているが，下の例では急性肺炎であることが疑われており，増悪傾向であることが疑われているという情報は付与されない．

1. <d>急性肺炎</d>の<c certainty=“suspicious”>増悪傾向</c>を疑いますが
2. <d certainty=“suspicious”>急性肺炎</d>の<c>増悪傾向</c>を疑いますが

※現在，2018/12月99件発注分id=53の事例では「<c certainty=”negative”>経時的変化</c>に乏しい。」とタグ付けしている．

# 解決済みの事項

## 特徴・尺度タグへのcertainty属性導入

　特徴・尺度タグについても以下の例文のような違いを表現するためにcertainty属性を導入すべき例があるが，特徴・尺度により修飾される症状・病名タグに付与するcertainty属性による表示で足りると考え，導入は不要だと判断した．

1. <f certainty=”negative”>有意なサイズ</f>の<d>リンパ節腫大<d>を指摘できません。
2. <f>有意なサイズ</f>の<d certainty=”negative”>リンパ節腫大<d>を指摘できません。