



資料ダウンロード

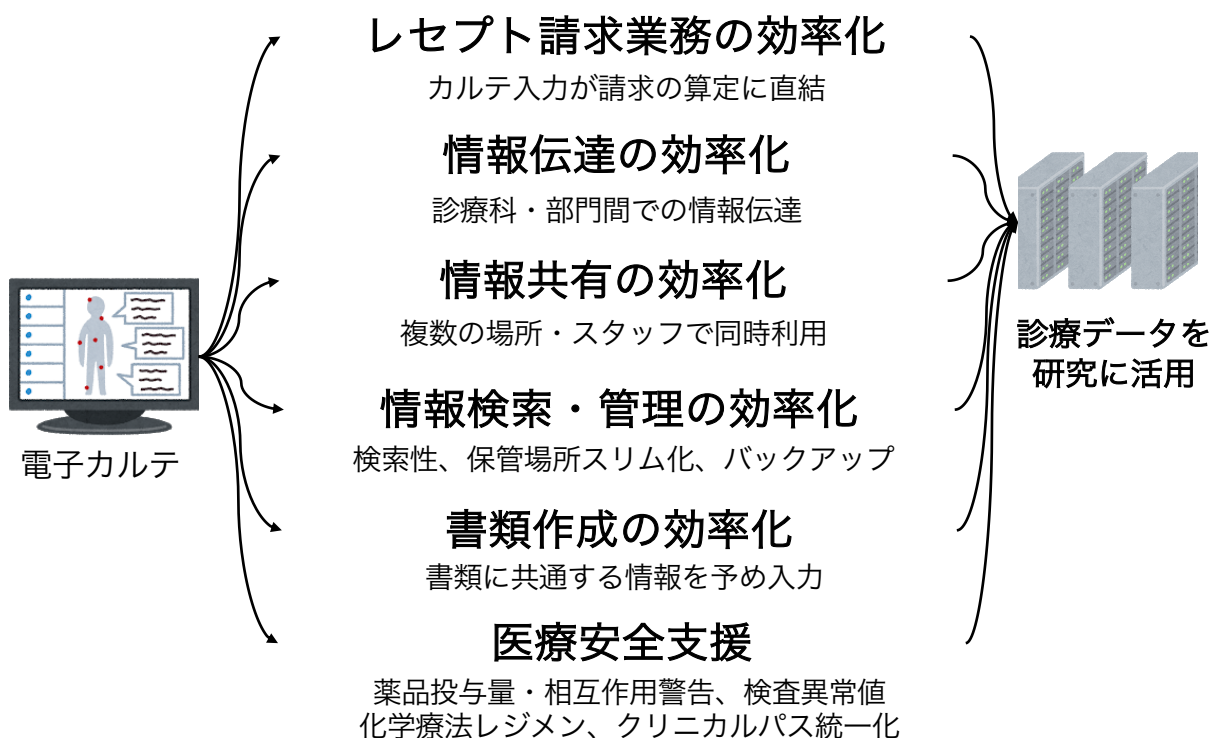
診療テキストの構造化に向けた 症例報告コーパスからの情報抽出

柴田 大作¹、◎ 河添 悦昌¹、篠原 恵美子¹、嶋本 公徳¹

1. 東京大学大学院医学系研究科 医療AI開発学講座

1

背景：電子カルテに期待される効率化と研究利用



2

背景：診療テキストの利用の難しさ

・診療テキスト

- ・ 医師の診療録や退院サマリ、看護記録、各種レポートなど。
- ・ 患者さんの様々な情報が記録されるも2次的な活用は限定的。

表記が雑多

収縮期は120mmHg、血圧110-120 mmHg、臥位153/91mmHg

同義語

転倒した、転んだ、転落した、ベッドから落ちた

事実性（あり、なし）

関節炎を**認めた**、筋力低下や錐体路徴候を**伴わない**、胸水**ある**も腹水**なく**

事実性（本人、家族）

家族歴：糖尿病なし、**母方**の糖尿病歴、**祖父**にSBMAの発症

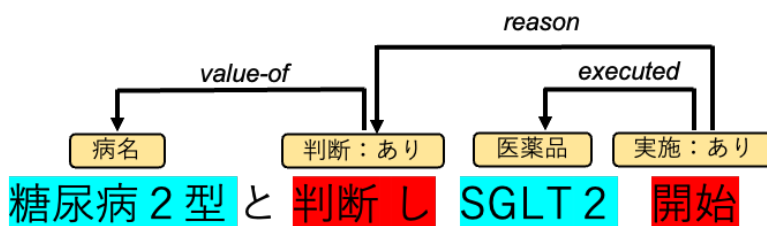
事実性（一般知識）

網膜症が合併**しやすい**ため、術後に**出血**が予想される

3

背景：医療のテキストコーパス

- ・ テキストからのNLP技術を開発するためのデータセット（テキスト＋アノテーション情報）
 - ・ AIの学習・評価用データセットとして利用
 - ・ 良いコーパスであればAIのベンチマークとなる
- ・ 言語学分野と医療分野のコーパスの違い
 - ・ 言語学：形態素、品詞、文節境界、係り受け構造などに興味
 - ・ 医療：医療の情報モデルに基づく情報に興味



4

背景：希少・難治性疾患を対象とした症例報告コーパス^{1,2} iCorpus: Corpus of clinical case reports of intractable diseases

- ・ 希少・難治性疾患を対象
 - ・ 幅広い診療科や疾患領域をカバー（77の臨床系学会誌）
- ・ 診療テキストから情報抽出するため開発
 - ・ 病名、症状・所見、治療理由などはテキストにのみ記載される
- ・ 汎用性・網羅性・一貫性を意識したアノテーション基準
 - ・ 複数の事実性（状態有無、判断の種類、実施有無など）を区別
 - ・ 時間関係、解剖部位関係、因果関係なども
 - ・ 異なる医療用語集への正規化を見据えて粒度を細かく
- ・ 機械学習モデルの学習・評価に利用可
 - ・ NLPタスク(固有表現抽出、関係抽出)としても定式化
- ・ 一般公開し研究利用可能
 - ・ 症例報告テキストのライセンス問題をクリア済み
<https://ai-health.m.u-tokyo.ac.jp/home/research/corpus>

1. 篠原 恵美子ら、医療テキストに対する網羅的な所見アノテーションのためのアノテーション基準の構築、第25回日本医療情報学春季学術大会、5
2. 河添 悦昌、篠原 恵美子、希少・難治性疾患を対象とした症例報告テキストコーパスの構築、第41回医療情報学連合大会、

76 歳 男性

主訴：開眼困難、言葉が出にくい

既往歴：高血圧、気管支喘息にて近医通院中。

家族歴：特記事項なし。

現病歴：2009 年 3 月頃より、両眼の開眼が困難な様子を家人に指摘されるようになった。同年 4 月下旬頃より発語が聞き取りづらくなったことを家人に指摘され、同年 7 月、当科に入院した。

一般理学所見：身長 155.4cm、体重 57.4kg（2 カ月間で約 6kg の減少）、体温 36.8 度、血圧 110/60mmHg、脈拍 75 分整、SpO2 93%、聴診上両肺で wheezing を聴取した。

神経学的所見：意識は清明であるが、病識は欠如し、多幸的であった。HDS-R は 1630、MMSE は 1730 であり、とくに計算、数字の逆唱、語想起、遅延再生での減点がめだつた。本例は失語をとめない、誤答はその要因によらず減点とした。Frontal Assessment Battery は 418 であった。新しい質問に対し、以前の質問の返答をくりかえす滞続言語をみとめた。努力性で失構音をともなう電文体の非流暢性発話や自発語の減少をみとめ、喚語困難をとまっていた。一方、二語文の理解と日常会話の聴理解は概ね保たれていた。以上より Broca 失語と判断した⁷⁾。また、書字は助詞が乏しく、電文調であり、“リハビリ”を“リハサル”と書くなどの字性錯書がめだつ一方で、“小豆”を“こまめ”と読むなどの類音的錯読をみとめた。強制把握、強制泣き笑いなど前頭葉徴候をみとめた。瞳孔は左右同大、対光反射は正常で、眼球運動制限はみとめなかった。開眼を命ずると、眼裂は両側で数十秒かけてゆっくりと 4~5mm まで開き、呼びかけや驚きにより、9~10mm までの開眼を確認できた (Fig. 1)。自動・随意運動乖離をみとめることから、開眼失行と診断した。随意的な開眼は困難であり、手もちいて眼瞼を挙上することが習慣となっていた。閉眼は敏速であった。開眼に先行して皺眉筋や眼輪筋の収縮はみられなかった。両側軟口蓋の挙上不良、催吐反射陰性、構音障害、嚥下障害をみとめた。挺舌は可能で、舌の軽度萎縮をみとめたが、線維束性収縮はみとめなかった。全身に軽度の筋萎縮がみられ、四肢で線維束性収縮をみとめたが、筋力低下はみとめなかった。口尖らし反射は陽性で、下顎反射をふくめ、四肢腱反射は亢進し、病的反射も四肢で陽性であった。感覚系、自律神経系は異常をみとめなかった。

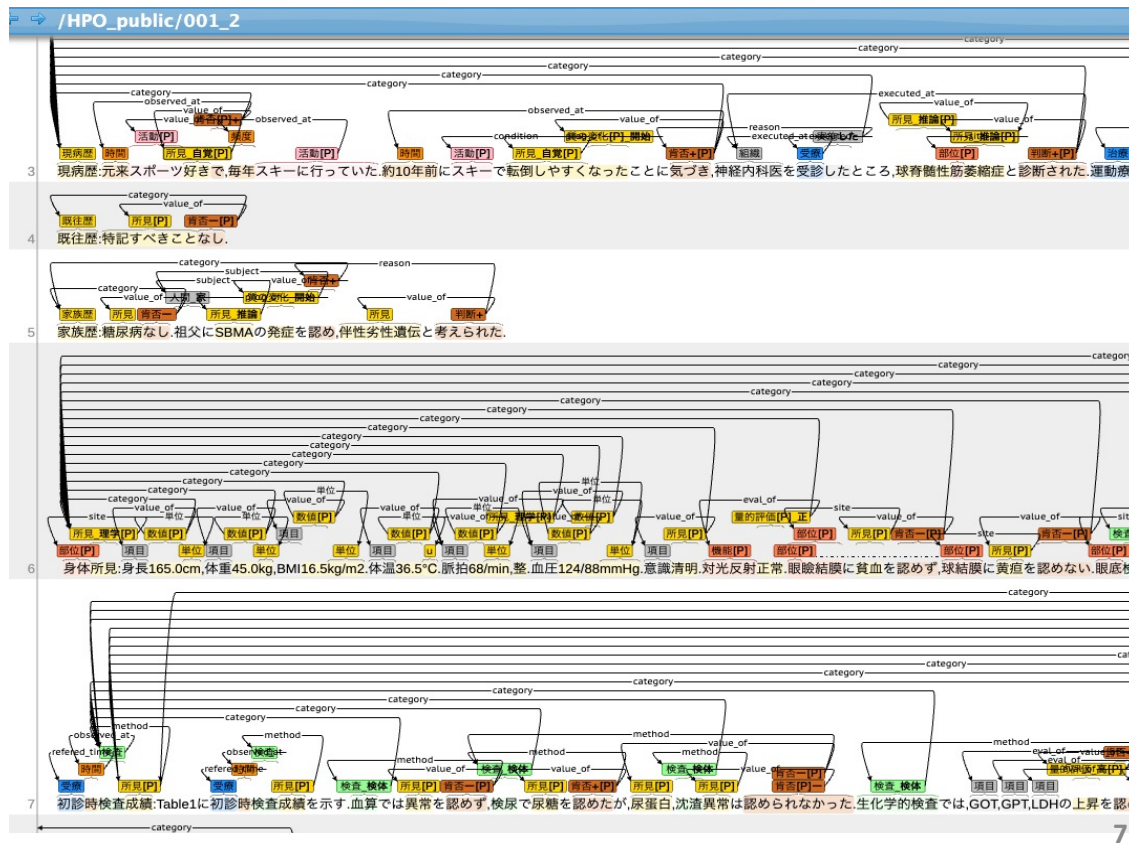
検査所見：一般血液検査では CK が 339I/UL と軽度高値である以外は異常所見なく、抗アセチルコリンレセプター (ACh-R) 抗体、および抗筋特異的チロシンキナーゼ (MuSK) 抗体は陰性で、甲状腺機能は正常であった。吸気呼気位で比較した胸部レントゲンで横隔膜の動きが乏しく、血液ガス分析は pH 7.437、PCO2 38.9mmHg、PO2 75.2mmHg、HCO325.6mmol/l と軽度の低酸素血症をみとめ、呼吸機能検査は %VC が 71.3% と低下していた。FEV1.0% は指示理解が困難のため評価不能であった。嚥下造影検査 (VF) では明らかな誤嚥はみとめなかった。頭部 MRI では、とくに前頭葉穹隆部を中心に両側前頭葉優位で両側側頭葉におよぶ脳萎縮と、T2強調画像および FLAIR 画像で両側前頭葉優位に大脳半球深部白質の広範な高信号変化をみとめた (Fig. 2A, B)。99mTcECD SPECT では、両側前頭葉前部、側頭葉前部、海馬傍回前部に左右対称性のいちじるしい血流低下をみとめた (Fig. 3A, B)。針筋電図は舌と右上腕二頭筋でおこない、いずれも神経原性変化をみとめた。エドロフォニウム試験は陰性で、低頻度反復刺激誘発筋電図でも waning はみとめなかった。

入院後経過：自発性の低下がより顕在化した一方で、排便前に必ず病棟を周回するなど常同行動もみられるようになった。常同行動に対し、選択的セロトニン再取り込み阻害薬 (selective-serotonin reuptake inhibitor ; SSRI) のフルボキサミンマレイン酸塩を 25mg 日から開始し、50mg 日まで漸増したところ、常同行動の改善のみならず、開眼失行にも軽度改善がみられた。しかし Broca 失語の進行をみとめ、入院時には日記をつけていたが、1 カ月間で筆談もできなくなった。一方で四肢筋力の低下はみとめなかった。入院初期の VF では明らかな誤嚥はみとめなかったものの、その後嚥下障害が顕在化したため、入院約 2 カ月後に胃瘻を造設し、療養型病院へ転院した。

開眼失行を呈した筋萎縮性側索硬化症/前頭側頭葉変性症の1例

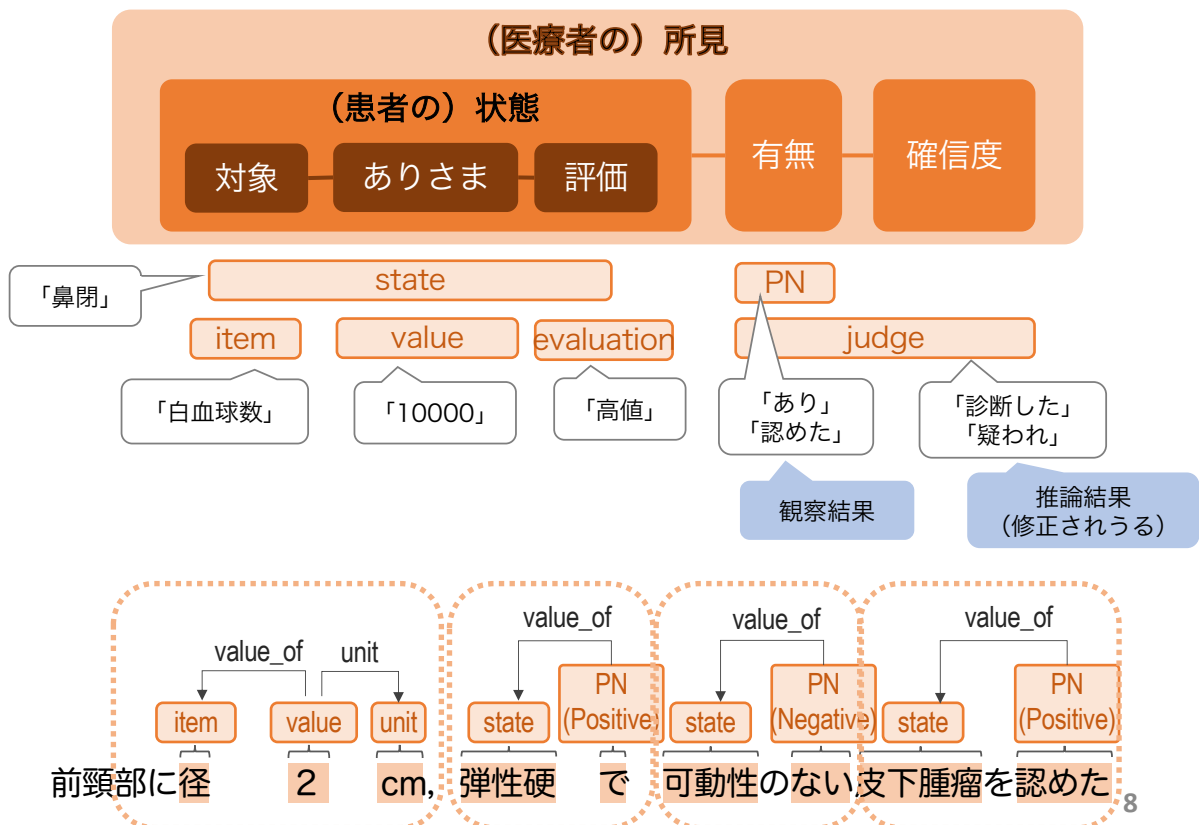
https://www.neurology-jp.org/Journal/public_pdf/050090645.pdf

背景：brat¹を利用しアノテーションを実施

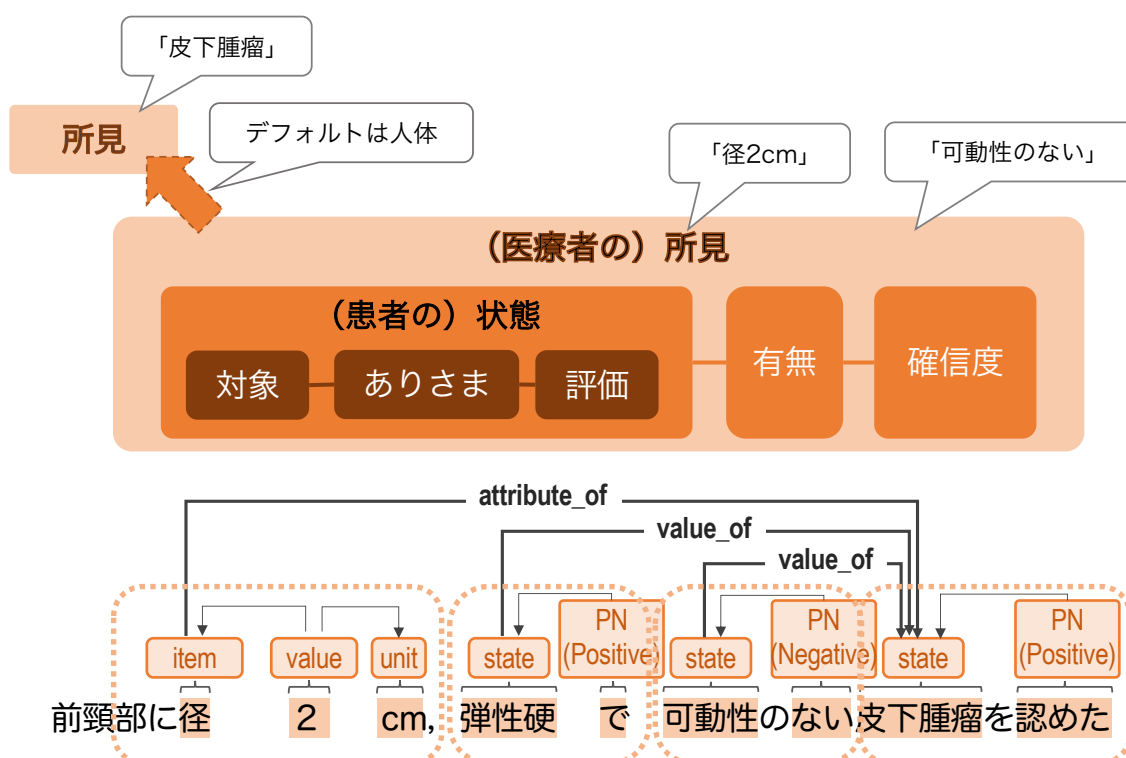


1. Pontus Stenetorp, et.al., brat: a Web-based Tool for NLP-Assisted Text Annotation. In Proceedings of the Demonstrations Session at EACL 2012.

背景：状態と所見の考え方



背景：何についての状態・所見か



9

背景：固有表現タグの定義（一部）

タグが有する情報				タグ付け基準の概略	タグ付けされるテキスト例
	名称	属性	本人属性		
1	age	-	o	年齢を示す表現.	63 歳, 56 歳, 1 歳 6 か月, 高校生
2	sex	-	o	性別を示す表現.	男性, 女性, 男児, 女児
3	smoking	Current Stop	o	喫煙に関する表現.	喫煙, タバコ, 喫煙歴, 禁煙
4	drinking	Current Stop	o	飲酒に関する表現.	飲酒, アルコール, 飲酒歴, 禁酒
5	state	-	o	患者の状態全般を示す表現. いわゆる, 病名, 症状 (患者の訴え), 所見 (観察結果) などを含む.	吐き気, 萎縮症, 糖尿病, 口渇
6	body	-	o	人体部位. 特定の部位を示す表現.	頭, 胃, 肝, 手足, 眼瞼結膜
7	tissue	-	o	人体組織. 人体各所で繰り返し出現するもの.	筋, 筋肉, 粘膜, 細胞, 繊維
8	item	-	-	患者の状態を表すために参照される項目.	血糖, 血糖値, HbA1c, 食欲
9	clinical_test	-	-	臨床検査に関する表現. item との違いは計測法を含むか否か.	神経学的検査, 徒手筋力検査
10	PN	Positive Negative None	o	患者の状態が, ある (Positive) , ない (Negative) , わからない (None) ことを示す表現.	で, 認め, 示す, 認める なし, 認めず, ではなく, なく 不明であった, 詳細不明
11	judge	Positive Suspicious Future Negative None	o	医療者により, 患者の状態がある (Positive) , あることが疑われる (Suspicious) , 将来あるかもしれない (Future) , ない (Negative) , 不明 (None) であることが判断されたことを示す表現.	診断された, 考えられた, 疑われた, 可能性も考え 否定的, 明らかではなかった 確定診断に至らなかった
12	quantity_evaluation	High Normal Low	o	数値への評価. 高い (High) , 正常 (Normal) , 低い (Low) .	上昇, 異常高値, 増加 正常, 基準値, 保たれて 低下, 減少, 減弱
13	quantity_progress	Increase NoChange Decrease	o	数値の変化. 上昇 (Increase) , 変化なし (NoChange) , 低下 (Decrease) .	
14	quality_evaluation	Mild Moderate Severe	o	数値以外の状態の程度を質的に示す表現. 主に疾患の重症度を示す. 軽度 (Mild) , 中等度 (Moderate) , 重度・高度 (Severe) .	軽度, 軽い, わずか, やや 中等度, 中度, 中等症 強い, 著名, 著しい, 重度
15	quality_progress	Start Worsen NoChange Improve Recover	o	数値以外の状態の時間的な変化を質的に示す表現. 出現 (Start) , 悪化 (Worsen) , 持続 (NoChange) , 改善 (Improve) , 軽快 (Recover) .	出現した, なった, きたした 悪化, 増悪, 進行, 顕在化 持続, 保たれて, 変わらず 改善, 軽快, 回復 落ち着き, 復帰, 軽快, 回復
16	value	-	o	検査値など, 身体や検体を測定し得られる数値.	7.5, 20, 1, 5, 165.0
17	unit	-	-	数値との組で表される単位.	m g / 日, 行, cm, kg/m2
18	time	-	-	時間軸上における特定位置の時点や区間を示す表現.	約 10 年前, その後, 直後
19	time_span	-	-	時間軸上の位置を問わず時間幅を示す表現.	1 日, 長時間, 2 カ月間

背景：関係タグの定義（一部）

	関係の名称	関係の概略	関係の例
1	LR	source の側性は target である。	両側(source)の三角筋(target)。
2	subject	source は target についての情報である。	母(target)が糖尿病(source)。
3	attribute_of	source は target の属性である。	縦(source)-5cm、横(source)-2cm の腫瘍(target)。
4	site	source が target の部位である。	四肢(target)の筋力(source)低下。
5	referred_site	相対位置 source は target を基準とする。	鎖骨部(target)の周囲(source)。
6	part_of	source は target の一部である。	右示指(target)が腫れ、特に第二関節(source)で顕著。
7	region_contain	source の領域に target が含まれる。	頸部(source)に腫瘍(target)を認め。
8	func_of	source は target の機能である。	横隔膜(target)の動き(source)が乏しく。
9	eval_of	値/状態の程度 source は target についての評価情報。	軽度(source)の筋萎縮(target)を認めた。
10	prog_of	値/状態の変化 source は target についての評価情報。	CK(target)の上昇(source)。
11	value_of	source が target の値である。	身長(target)は 170 (source) cm であり。
12	unit	source は target の単位である。	身長は 170 (target) cm (source) であった。
13	after	治療 source の結果、target となった。	増量した(source)ところ不随意運動が消失した(target)。
14	start_from	source は時間 target から始まった。	2002 年(target)より筋力低下が進行した(source)。
15	end_at	source は時間 target に終わった。	2002 年(target)まで感覚障害あり(source)。
16	observed_at	source (状態)が観察された場所/時間は target である。	2002 年(target)に前医で振戦を認めた(source)。
17	referred_time	source の基準時間は target である。	開始(target)4 日目(source)に頭痛が出現。
18	condition	source は条件 targetのもとで起こる。	最大収縮時(target)に単位電位の減少がみられた(source)。
19	reason	source と判断された根拠は target である。	蜂窩織炎(target)から菌血症(source)を疑われ。
20	causal	source が原因となって target が起こった。	潰瘍性病変(source)による狭窄(target)を認めた。

11

本研究の目的

- これまで開発してきた症例報告コーパスにおける、固有表現抽出（NER）と関係抽出（RE）のベースラインとなる精度を提示する。

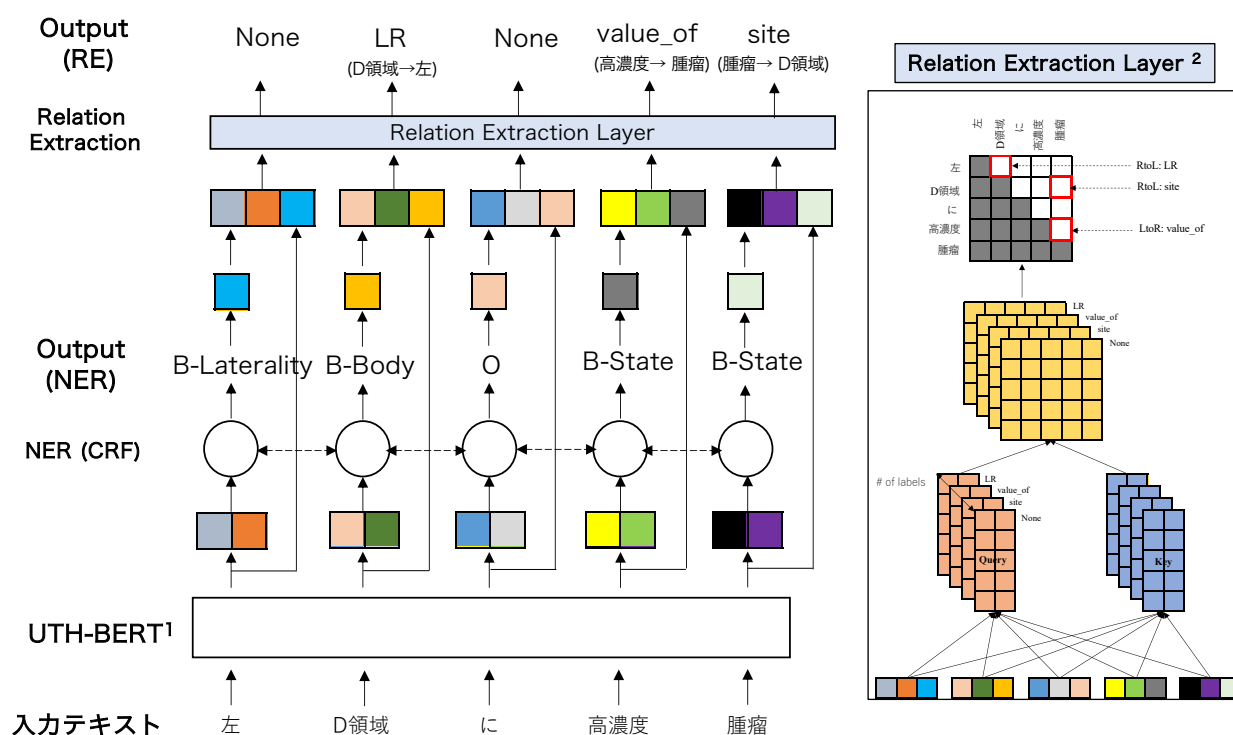
12

方法：実験に用いたコーパスの内訳

文書数 (=症例数)		183
1 文書あたり平均	文字数 (S.D)	1,915 (696)
	単語数 (S.D)	972 (330)
	固有表現数 (S.D)	394 (129)
	関係数 (S.D)	387 (127)
固有表現ラベルの種類		70種 (B-, I-, O : 141クラス)
関係ラベルの種類		35種 (+None : 36クラス)

13

方法：実験に用いた機械学習モデル(Joint NER-RE)



1. Kawazoe Y, et al. A clinical specific BERT developed using a huge Japanese clinical text corpus. PLoS One. 2021 Nov 9;16(11):e0259763.
2. Youmi Ma, et al., Named Entity Recognition and Relation Extraction using Enhanced Table Filling by Contextualized Representations, arXiv preprint arXiv: 2010.07522, 2020.

14

方法：NERとREの評価

- **【評価指標】** NER、REともに、5分割交差検定によるMacro-F1とMicro-F1の平均値。
- **【交差検定の分割】** 文書単位（183）で分割。学習データ（64%）、検証データ（16%）、テストデータ（20%）とした。
- **【NERの評価】** CoNLL-2000の評価方法を採用。1つの固有表現のB-タグとI-タグをすべて正解したものをTrue Positiveとした。
- **【REの評価】** 2つの固有表現間の35種の関係を正解したものをTrue Positiveとした。Noneを何らかの関係としたものをFalse Positive、何らかの関係をNoneとしたものをFalse Negativeとした。

15

結果：固有表現全タグ・全関係の平均値

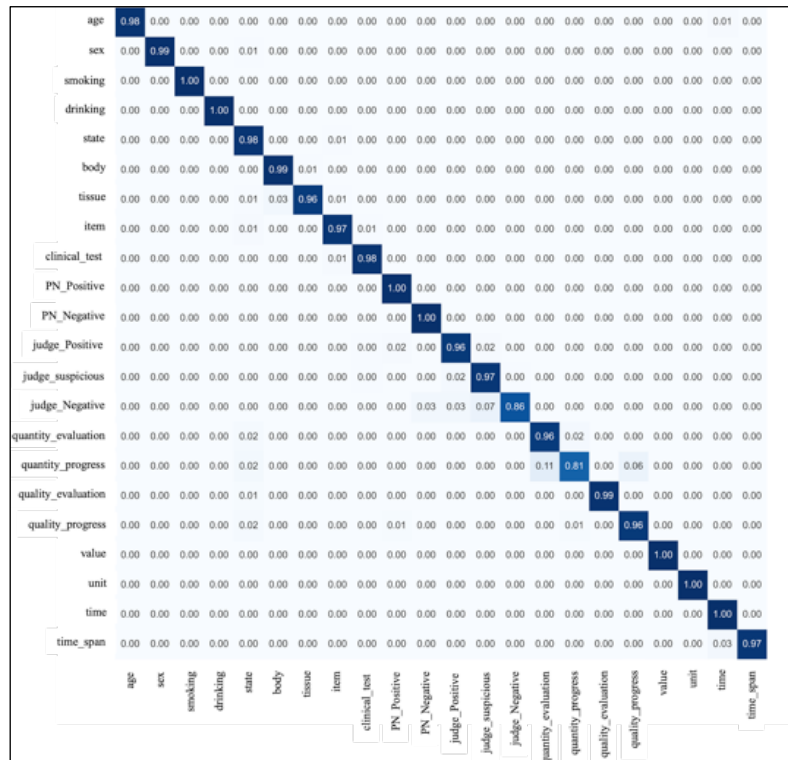
	固有表現抽出		関係抽出	
	Joint	Pipeline	Joint	Pipeline
Micro-F1	0.926	0.931	0.825	0.826
Macro-F1	0.753	0.779	0.640	0.643

16

結果：固有表現タグ22種, Pipeline

Normalized confusion matrix

	固有表現名	F値	固有表現数
1	age	0.958	637
2	sex	0.994	183
3	smoking	0.981	43
4	drinking	0.989	31
5	state	0.921	12,875
6	body	0.940	6,815
7	tissue	0.872	854
8	item	0.922	5,285
9	clinical_test	0.921	2,944
10	PN_Positive	0.958	5,005
11	PN_Negative	0.964	1,760
12	judge_Positive	0.902	478
13	judge_Suspicious	0.894	275
14	judge_Negative	0.710	34
15	quantity_evaluation	0.937	910
16	quantity_progress	0.808	342
17	quality_evaluation	0.971	578
18	quality_progress	0.931	1,555
19	value	0.960	4,417
20	unit	0.963	3,683
21	time	0.973	3,885
22	time_span	0.868	202

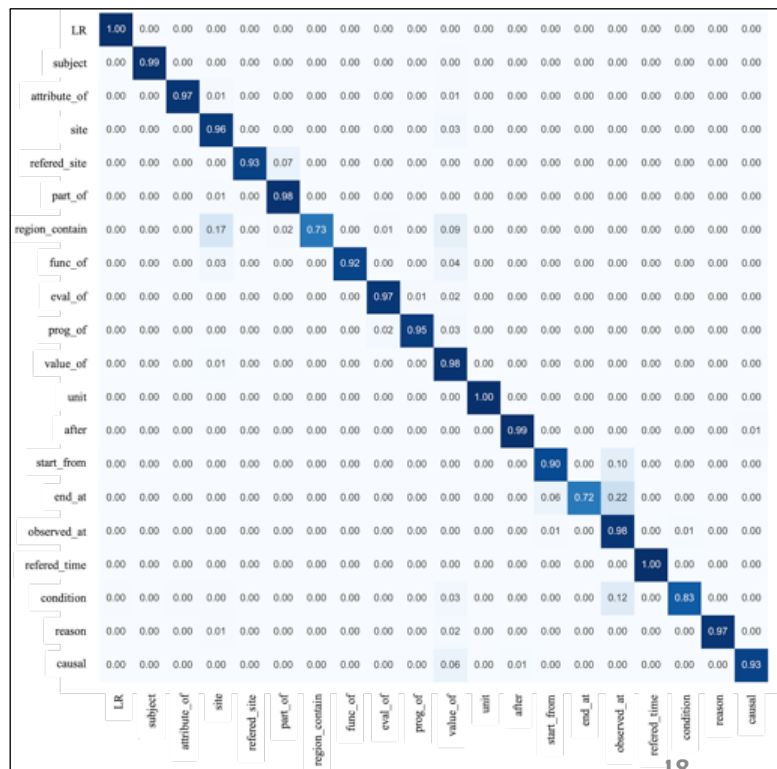


17

結果：関係タグ20種, Pipeline

Normalized confusion matrix

	関係名	F値	関係数
1	LR	0.892	1,177
2	subject	0.801	693
3	attribute_of	0.757	1,118
4	site	0.805	5,566
5	referred_site	0.738	273
6	part_of	0.701	2,214
7	region_contain	0.501	366
8	func_of	0.789	309
9	eval_of	0.885	1,701
10	prog_of	0.856	1,848
11	value_of	0.879	20,107
12	unit	0.970	3,963
13	after	0.377	450
14	start_from	0.745	218
15	end_at	0.519	27
16	observed_at	0.767	1,602
17	referred_time	0.916	2,698
18	condition	0.483	561
19	reason	0.634	2,534
20	causal	0.593	461



18

考 察

先行研究との比較

- ADEコーパス¹ (2972文書)
 - 固有表現:3, 関係:2, NER 91.3, RE 83.74
- 放射線レポート² (1000文書)
 - 固有表現: 9, 関係:10, NER 95.65, RE 86.5
- 診療テキスト² (156文書)
 - 固有表現: 9, 関係:10, NER 85.49, RE 71.04

本コーパスはNERタグ・関係数が先行研究と比較して多く、抽出タスクの難易度はより高いと思われるが、NER/REともに先行研究と同等の抽出精度を示した。ベースラインの精度としては十分と考えられた。

1. Gurulingappa H, et.al., Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. J Biomed Inform. 2012 Oct;45(5):885-92.
2. Fei Cheng, et.al., JaMIE: A Pipeline Japanese Medical Information Extraction System. arXiv:2111.04261.

19

結 語

- 開発してきた症例報告コーパス (iCorpus) における、70種の固有表現と35種の関係の抽出精度を提示した。
- BERTをベースとするモデルにより、Micro-F1で最大0.931 (NER) と0.826 (RE) の精度が得られた。
- 今後の課題として、実際の医療現場で作成されるテキスト (退院サマリなど) を対象とした、抽出精度の評価があげられる。

20