

# 医療言語処理と リアルワールドデータが拓く共創

医療分野のテキスト + 自然言語処理 = 膨大なリアルワールドデータ

奈良先端科学技術大学院大学 (NAIST)

矢田 竣太郎

けいはんなR&Dフェア2022

# Medical text NLP MedNLP

## 医療分野のテキスト + 自然言語処理 = 医療言語処理

- 自然言語処理 (NLP):  
機械翻訳に代表されるような, 人間の言語をコンピュータで解釈する技術
- 近年, わずかな訓練事例だけから回答を生成できるGPT-3など話題に
- 電子カルテなど医療テキストを処理するNLPを, 以降, 医療言語処理と呼ぶ

rinna社の日本語版GPT-3  
(<https://prtimes.jp/main/html/rd/p/000000025.000070041.html>)

GPTの入力: 宮本武蔵は、

GPTの出力: 生涯人生を修行に費やし心を鍛え精進した人でした。そして宮本武蔵の弟子たちの記録も数多く残されています。

GPTの入力: 最悪だよはネガティブ。いいねはポジティブ。素晴らしいはポジティブ。良くはないは

GPTの出力: ネガティブ。

GPTの入力: Helloは、こんにちはという意味です。Good morningは、おはようという意味です。Thank youは、

GPTの出力: ありがとうという意味です。感謝の気持ちを表す簡単なフレーズになります。



「夕暮れの神社の夏祭りを描いた水彩画」

rinna社の日本語版stable diffusion  
(<https://prtimes.jp/main/html/rd/p/000000035.000070041.html>)

Vision & Language  
画像処理 + 言語処理

夕暮れの神社の夏祭りを描いた水彩画

# 野望と現実

## NLPの応用

## FDA認可AI機器

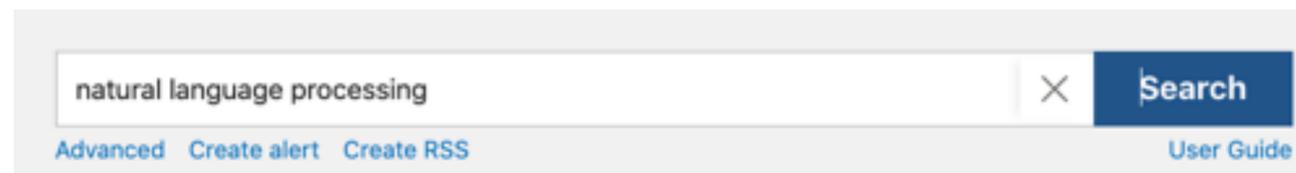
目的	入力	出力	手法
診断支援・類似症例検索	文書集合	文書	文書類似度
見落とし防止	文書	文	重要文判定 文類似度
副作用シグナル抽出	文書集合	用語	文書分類 文分類
カルテ入力支援	文集合	文	文分類
カルテ自動生成	画像, 検査値	文書	生成
悩み不安調査	文書	文	重要文抽出
個人情報匿名化	文書	個人情報	NER

Device	Company	Panel (Lead)	Primary Product Code
Precise Position	Philips Healthcare (Suzhou) Co., Ltd.	Radiology	JAK
Qmenta Care Platform Family	Mint Labs, Inc., D/B/A. QMENTA	Radiology	LLZ
LINQ II Insertable Cardiac Monitor, ZELDA AI ECG Classification System	Medtronic, Inc.	Cardiovascular	MXD
IDx-DR	Digital Diagnostics Inc.	Ophthalmic	PIB
Cognoa Asd Diagnosis Aid	Cognoa, Inc.	Neurology	QPF
CINA CHEST	Avicenna.AI	Radiology	QAS
HYPER AiR	Shanghai United Imaging Healthcare Co.,Ltd.	Radiology	KPS
Cartesion Prime (PCD-1000A/3) V10.8	Canon Medical Systems Corporation	Radiology	KPS
MEDO-Thyroid	MEDO DX Pte. Ltd.	Radiology	QIH
Preview Shoulder	Genesis Software Innovations	Radiology	QIH
Automatic Anatomy Recognition (AAP)	Quantitative Radiology Solutions	Radiology	QKR

<https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-ai/ml-enabled-medical-devices>

荒牧英治, 自然言語処理の医療への応用, 先進医療 NAVIGATOR 医療とAI最前線, 2022.

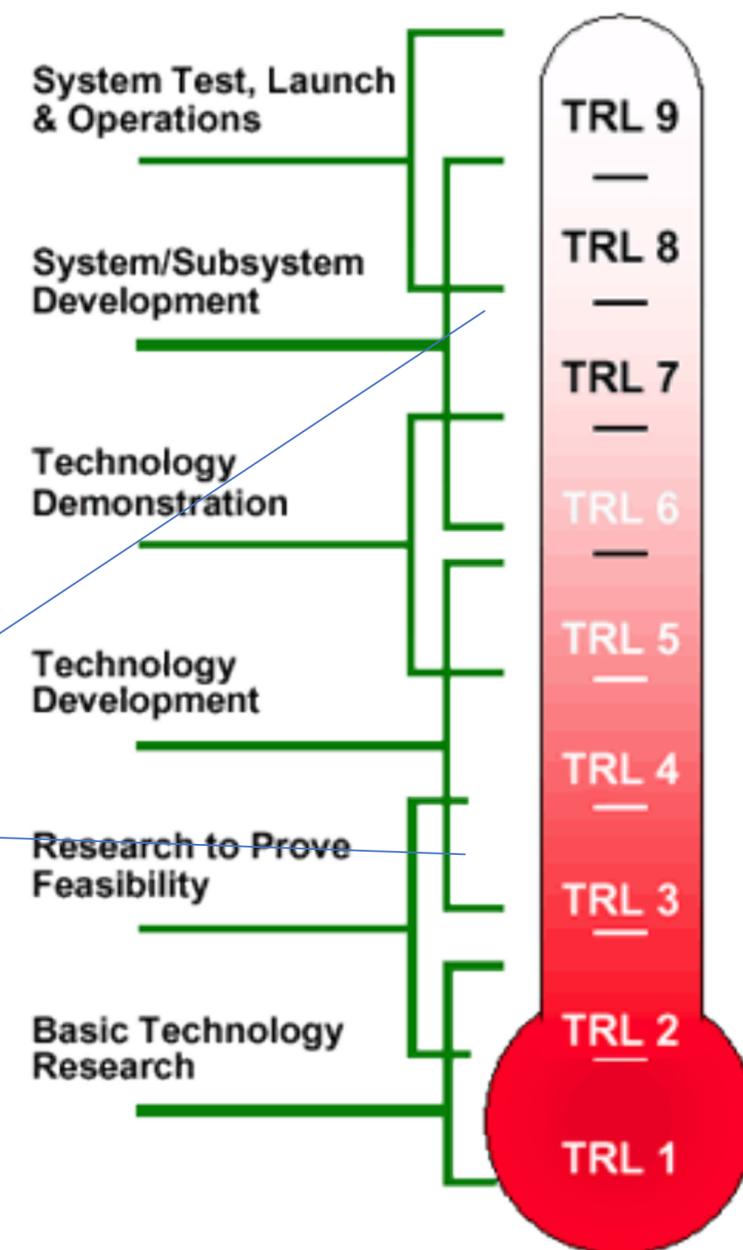
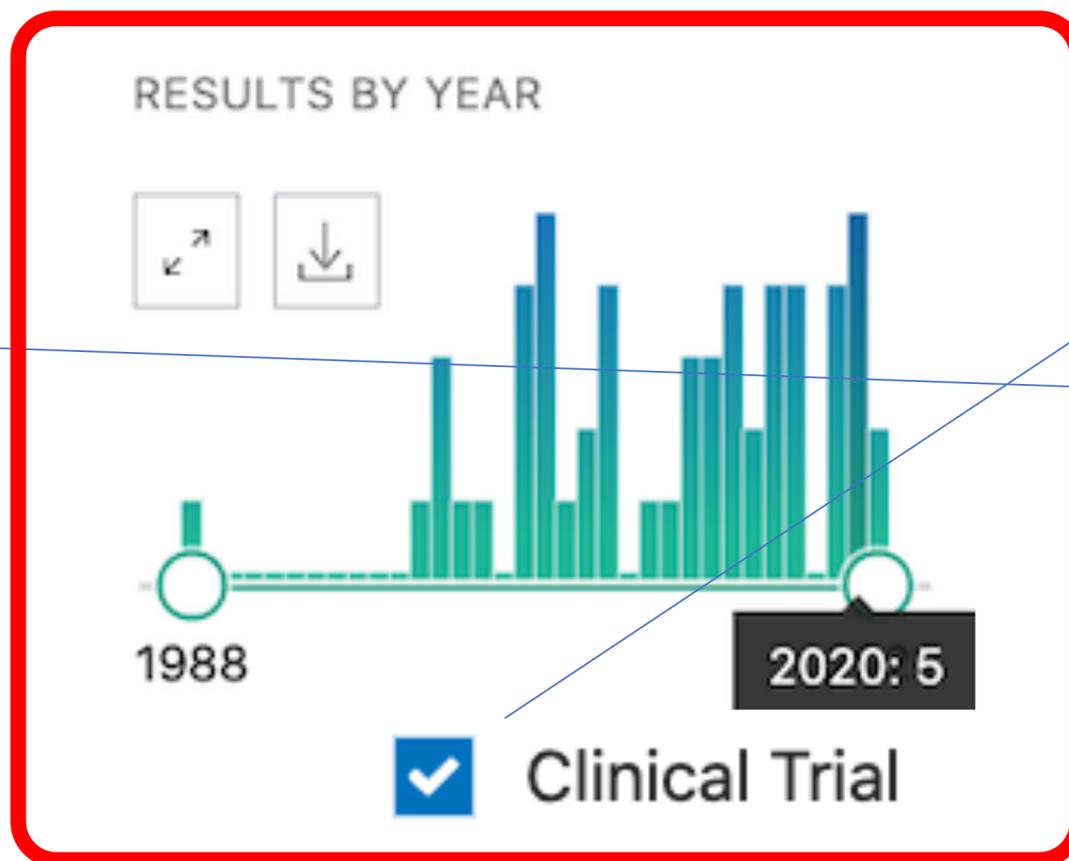
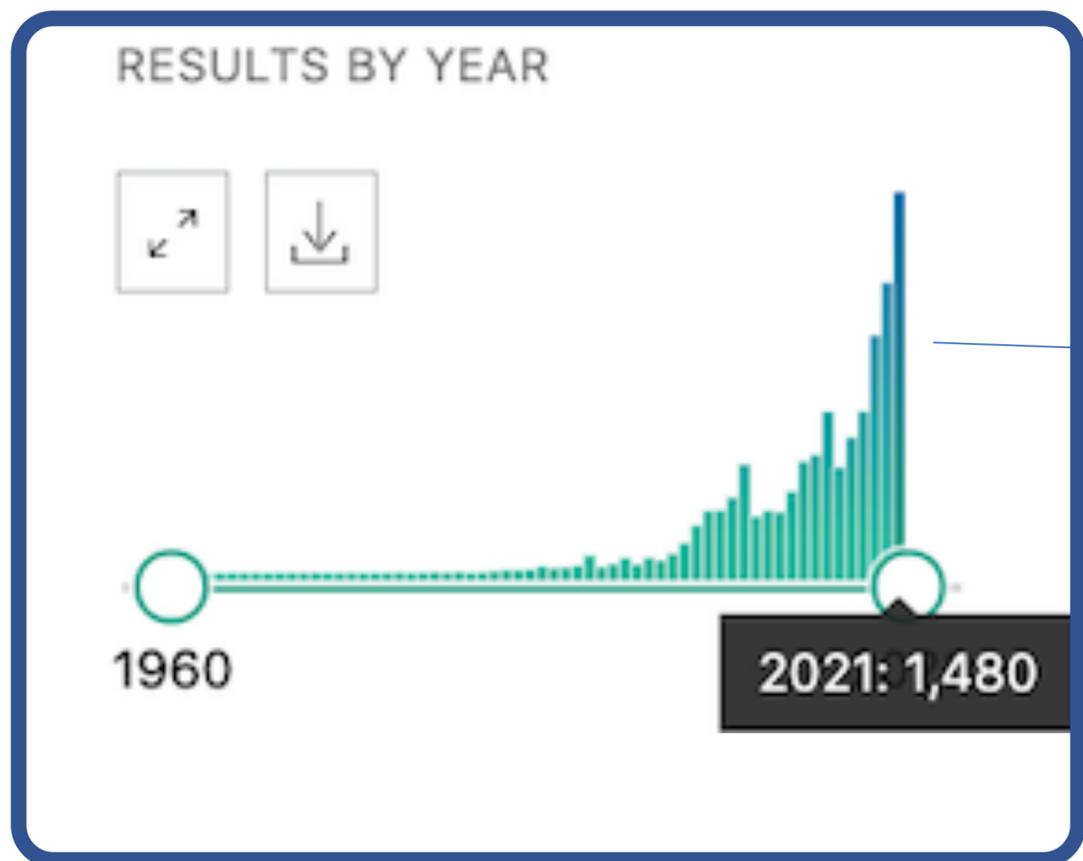
# 医療言語処理は まだまだ基礎研究



できそうから、できたが求められる時代へ

NLP論文は増加傾向(1000報以上)

NLPで臨床試験を行っている論文は希少



# Natural Language Processing: from Bedside to Everywhere

Eiji Aramaki, Shoko Wakamiya, Shuntaro Yada, Yuta Nakamura

2022 IMIA and Georg Thieme Verlag KG

英語中心

Reference	Objectives	Study design	Target language	Corpus	NLP task	Method	Performance [95% confidence interval]
Wissel et al., 2020 [24]*	Detection of surgical candidates in epilepsy patients	Prospective study	English	Progress notes	Document classification	SVM	ROC-AUC 0.79 [0.62–0.96]
Seol et al., 2021 [25]*	Prevention of acute exacerbation of pediatric asthma with decision support tool	RCT	English	Progress notes	NER	Not applicable	Odds ratio 0.82 [0.374–1.96]
Zheng et al., 2016 [26]	Detection of patients with diabetes prior to structured coding	Prospective study	English	Progress notes, discharge summaries, and other clinical notes	Document classification	NER + Normalization + Random forest	ROC-AUC 0.929
Wang et al., 2015 [27]	Detection of chronic heart failure patients prior to structured coding	Prospective study	English	Progress notes, discharge summaries, and other clinical notes	Document classification	NER + Normalization + Random forest	ROC-AUC 0.919
Tian et al., 2017 [28]	Detection of deep vein thrombosis and pulmonary embolism	Retrospective study	English	Radiology reports	Document classification	Rule-based	Sensitivity 0. [0.88–0.97] 0.96 [0.95–
Bucher et al., 2020 [29]*	Surveillance of surgical site infections	Retrospective study	English	Operative reports, progress notes, nursing notes, radiology reports, and discharge summaries	Binary classification	Existing tool	ROC-AUC 0.912
Shi et al., 2016 [31]	Detection of cerebral infarction patients, pneumonia patients, and coronary artery disease patients	Retrospective study	Chinese	Progress notes and discharge summaries	Document classification	CNN	F1-score 0.934–0.966
Waston et al., 2011 [32]	Identification of psychosocial re-admission risk factor of heart failure patients	Retrospective study	English	EHR notes	Multivariate analysis	Rule-based feature extraction + Logistic regression	(Not applicable)
Wang et al., 2017 [33]	Detection of arterial fibrillation patients with high risk of cerebral infarction	Retrospective study	English	Clinical notes	Information extraction	Rule-based	Positive Predictive Value 0.92–1.00
Buchen et al., 2017 [34]	Prediction of developing coronary artery disease	Retrospective study	English	Clinical notes	Document classification	NER-based feature extraction + SVM	F1-score 0.774
Chase et al., 2017 [35]	Detection of multiple sclerosis patients	Retrospective study	English	EHR notes	Document classification	NER + EL + Naive Bayes classification	ROC-AUC 0.94 [0.93–0.95]
Forferko-Shodrach et al., 2019 [38]	Information extraction from clinical letters for epilepsy patients	Retrospective study	English	Clinic letters	Information extraction	Existing tool	F1-score 0.911
Selby et al., 2018 [39]	Surveillance of postoperative deep vein thrombosis and pulmonary embolism	Retrospective study	English	Radiology reports	Document classification	(Not applicable)	Sensitivity 0.851–0.900, Specificity 0.717–0.801
Tibbo et al., 2019 [41]	Surveillance of periparturient femur fractures	Retrospective study	English	Operative reports and progress notes	Document classification	NER + EL	Sensitivity 1.000, Specificity 0.998
Zhang et al., 2021 [91]*	Automatic summarization of doctor-patient conversation	Retrospective study	English	Outpatient transcription	Summarization	BART	ROUGE-L 0.3412

Reference	Objectives	Study design	Target language	Corpus	NLP task	Method	Performance [95% confidence interval]
Valtchinov et al., 2020 [52]*	Identification of implantable device posing MRI safety risks	Retrospective study	English	Radiology reports and other clinical notes	NER	Existing tool	Accuracy 0.83–0.91
Trivedi et al., 2018 [53]	Determination to use or not to use contrast enhancement for MRI imaging	Retrospective study	English	Free-text MRI protocols and indications	Document classification	Existing tool	Accuracy 0.832
Chillakuru et al., 2021 [54]*	Determination to use or not to use contrast enhancement for MRI imaging	Retrospective study	English	Free-text MRI protocols and indications	Document classification	TF-IDF vectorization + GBDT, Word embedding + shallow NN	Accuracy 0.8338–0.8543
Kim et al., 2020 [71]*	Information extraction from prostate cancer	Retrospective study	English	Pathology report	NER	BERT	Accuracy 0.9795–0.9839
[72]*	Automated diagnosis of multiple psychiatric diseases	Retrospective study	English and Chinese	Discharge summaries	Multilabel document classification	CNN, Random forest	Weighted F1-score 0.972 (document), Accuracy 0.930
Dai et al., 2020 [72]*	Automated diagnosis of multiple psychiatric diseases	Retrospective study	English and Chinese	Discharge summaries	Multilabel document classification	RoBERTa	Micro F1-score 0.584
Moon et al., 2022 [80]*	Recognition of gynecological surgical history	Retrospective study	English	Clinical notes	Multiclass classification	NER + Rule-based approach	Weighted F1-score 0.76
Sterckx et al., 2020 [81]*	Birth risk estimation	Retrospective study	English and Dutch	Clinical notes and structured data	Regression	Random forest	Accuracy 0.88
Wang et al., 2021 [82]*	Information extraction from cancer surgery	Retrospective study	English	Radiology reports	Information extraction	Rule-based	Accuracy 0.88
Gaskin et al., 2016 [85]	Surveillance of postoperative complications of cataract surgery	Retrospective study	English	EHR notes and structured data	Binary classification	Random forest	ROC-AUC 0.62–0.84
Moen et al., 2018 [87]	Synonym replacement for better readability	Retrospective study	Finnish	Progress notes and nursing notes	Synonym detection	Word embedding + cosine similarity	Top-1 accuracy 0.3464
Gopinath et al., 2020 [88]*	Auto-completion of progress notes	Retrospective study	English	Progress notes and structured data	Recommendation	Rule-based entity type detection + Shallow NN or Bayes statistics	Keystroke reduction 67%
[89]*	Information extraction from nursing notes	Retrospective study	Finnish	Nursing notes	Sentence classification	LSTM	Accuracy 0.71
Krishna et al., 2021 [90]*	Automatic summarization of doctor-patient conversation	Retrospective study	English	Outpatient transcription	Summarization	BERT-LSTM classifier + T5	ROUGE-L 0.3838
Zhang et al., 2021 [91]*	Automatic summarization of doctor-patient conversation	Retrospective study	English	Outpatient transcription	Summarization	BART	ROUGE-L 0.3412

後ろ向き中心

分類中心

新しい手法が少ない  
古典主義

英語カルテを材料にすこし古い手法を用いて  
文書分類による後ろ向き解析

# 何が医療言語処理に向いているのか？

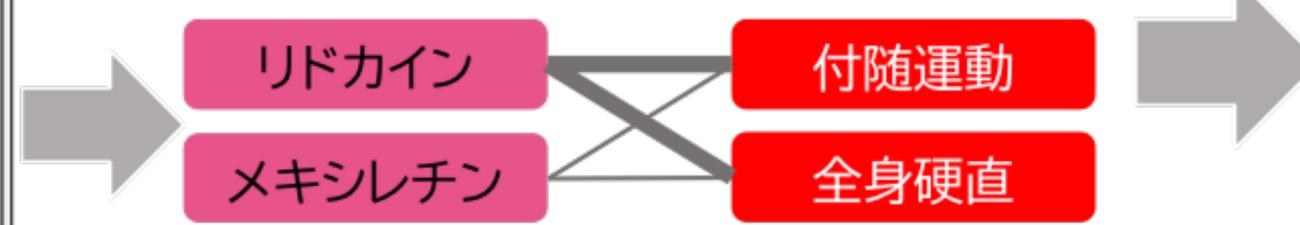
## 1つの解がフェノタイプ

- 膨大なデータに対して、後ろ向き解析で行える
- 1症例ベース単位のエラーに大きく影響されない
- 創薬, ドラッグリポジショニングなどビッグゴールにつながる

時(DATE)その後 Mk(-)リドカインは  
C休薬、減量し Mv 78mg/h の  
時、Mk(+)メキシレチンを  
Mv(-) 200mg/日 から  
Mv(+) 450mg/日に C増量 した  
時(DATE)初日に、  
D(+)全身硬直、不随運動が C出現 した  
(Tk リドカイン血中濃度は  
Tv 5.7μg/mL)。

リアルワールドデータ

ステップ1: 医療言語処理による情報抽出  
固有表現抽出技術で医薬品とフェノタイプ抽出



医薬品 × フェノタイプ

	痛み	便秘	下痢	口内炎	吐き気	味覚障害	発熱	しびれ	倦怠感	浮腫	頭痛	むくみ	痺み
乳剤ドセタキセル	107	31	49	40	28	36	39	36	30	32	9	31	11
乳剤FEC (100) (イメンド)	26	52	20	41	42	12	19	7	12	9	13	1	11
乳剤TC (ドセタキセル+エンドキサン)	34	29	22	20	19	13	14	9	18	10	11	7	7
乳剤P+T+M+HER+DOC (2回目以降)	12	9	40	16	3	6	8	4	6	13	2	1	6
乳剤AC (イメンド)	14	20	4	16	12	9	6	1	11	2	4	5	3
乳剤FEC (100) (APR)	11	25	9	15	13	4	11	2	8	1	8	0	3
乳剤Per+Tra+DOC (2回目以降)	12	3	34	6	5	6	1	8	5	9	4	8	4
乳剤AC (APR)	15	22	11	15	10	14	1	8	2	0	4	2	3
乳剤ddAC (イメンド)	8	19	3	3	9	5	4	1	7	2	2	0	1
乳剤EC90 (APR)	8	5	4	3	0	4	2	0	2	1	1	1	1
乳剤ddAC (APR)	7	2	4	2	1	5	1	2	1	0	5	2	2

ステップ2: Drug Repositioning探索  
大量に収集する過程で含まれるバイアスを統制し解析

- 医療言語処理(MedNLP)の現状
- MedNLPの手法
  - 用語抽出から関係付与まで
- MedNLPの新しい材料
- おわりに

# まずは結果から： 1施設からこのような情報が即座にとれる

## 乳がん入院患者(n=2000)の統計

	Anticancer drug	痛み	便秘	下痢	口内炎	吐き気	味覚障害	発熱
乳腺ドセタキセル	ドセタキセル	107	31	49	40	28	36	
乳腺FEC (100) (イメンド)	エピルビシン+フルオロウラシル+シクロホスファミド	26	52	20	41	42	12	
乳腺TC (ドセタキセル+エンドキサン)	ドセタキセル+シクロホスファミド	34	29	22	20	19	13	
乳腺 <sup>+</sup> ルツズマブ+HER+DOC (2回目以降)	ペルツズマブ+トラスツズマブ+ドセタキセル	12	9	40	16	3	6	
乳腺AC (イメンド)	ドキソルビシン+シクロホスファミド	14	20	4	16	12	9	
乳腺FEC (100) (APR)	エピルビシン+フルオロウラシル+シクロホスファミド	11	25	9	15	13	4	
乳腺Per+Tra+DOC (2回目以降)	ペルツズマブ+トラスツズマブ+ドセタキセル	12	3	34	6	5	6	
乳腺AC (APR)	ドキソルビシン+シクロホスファミド	15	22	11	15	10	14	
乳腺ddAC(イメンド)	ドキソルビシン+シクロホスファミド	8	19	3	3	9	5	
乳腺EC90 (APR)	エピルビシン+シクロホスファミド	8	5	4	3	0	4	
乳腺ddAC(APR)	ドキソルビシン+シクロホスファミド	7	2	4	2	1	5	

# 副作用関係の自動抽出の実際

固有表現抽出  
Named Entity Recognition

メキシレチン を 200mg/日 から 450mg/日 に 増量 した 初日 に 全身硬直, 付随運動 が出現し,

薬品名

投与量

投与量

行為

時間

症状

行為

事実性判定  
Negation

固有表現正規化  
Named Entity  
Normalizaiton

メキシレチン  
一般名: Mexiletine Hydrochloride  
ATCコード: C01BB02  
KEGGID: D00639

ICD: R252  
MedDEA: 痙攣  
HPO: 痙攣

ICD: R298  
MedDEA: 項部硬直

ICD: M2569  
MedDEA: 関節硬直  
HPO: 関節拘縮

関係抽出  
Relation Extraction

Drug-ADE relation  
因果関係

# 大規模な用語正規化辞書

	C	D	E	F	G	H	I	J	K	L	
1	出現形	出現形よみ	標準病名	標準病名よみ	症状フラグ	診断名フラグ	ICDコード	信頼度LEVE	頻度レベ	共起症状 (標準病名)	共起診断
87	逆行性胆管炎	ギャッコウセイタンカンエン	逆行性胆管炎	ギャッコウセイタンカンエン	0	0	K830	S	55-60%		
88	外傷性くも膜下出血	ガイショウセイクモマクカシュクケツ	外傷性くも膜下出血	ガイショウセイクモマクカシュクケツ	0	1	S066	S	55-60%	出血, 疼痛, 意識障害, 圧痛, 発熱, めまい	出血
89	臍管拡張	スイカンカクチョウ	臍管拡張	スイカンカクチョウ	0	0	K868	S	55-60%		
90	全身痛	ゼンシンツウ	全身痛	ゼンシンツウ	1	0	R529	S	55-60%	疼痛, 発熱	疼痛, 発熱
91	乳頭腫	ニュウトウシュ	乳頭腫	ニュウトウシュ	0	0	D369	S	55-60%		
92	妄想性障害	モウソウセイショウガイ	妄想性障害	モウソウセイショウガイ	0	0	F220	S	55-60%		
93	高ビリルビン血症	コウビリルビンケツショウ	高ビリルビン血症	コウビリルビンケツショウ	1	1	R798	S	55-60%	黄疸, 発熱, 疼痛, 浮腫, 出血	発熱
94	腹圧性尿失禁	フクアツセイニョウシツキン	腹圧性尿失禁	フクアツセイニョウシツキン	0	0	N393	S	55-60%		
95	顎変形症	ガクヘンケイショウ	顎変形症	ガクヘンケイショウ	0	0	K079	S	55-60%		
96	排便障害	ハイベンショウガイ	排便障害	ハイベンショウガイ	1	0	R198	S	55-60%	疼痛, 嚥下障害, 発熱, 圧痛, 排尿障害, 食欲不振	疼痛
97	鎖骨骨折	サコツコッセツ	鎖骨骨折	サコツコッセツ	0	0	S4200	S	55-60%		
98	副甲状腺機能低下症	フクコウジョウセンキノウテイカショウ	副甲状腺機能低下症	フクコウジョウセンキノウテイカショウ	0	1	E209	S	55-60%	浮腫, 発熱, 発赤, 嘔吐症, 疼痛, 嚥下障害	浮腫, 低力
99	便通異常	ベンツウイジョウ	便通異常	ベンツウイジョウ	0	0	K590	S	55-60%		
00	残胃癌	ザンイガン	残胃癌	ザンイガン	0	1	C169	S	55-60%	発熱, 疼痛, 嘔気, 腹痛症, 食欲不振, 腫瘍, リンパ節腫大	胃癌, 発熱
01	肛門管癌	コウモンカンガン	肛門管癌	コウモンカンガン	0	1	C211	S	55-60%	疼痛, 腫瘍, リンパ節腫大, 発熱, 嘔気, 出血	疼痛, 腫瘍
02	爪囲炎	ソウイエン	爪囲炎	ソウイエン	0	0	L030	S	55-60%		
03	重症急性膵炎	ジュウショウキウセイスイエン	重症急性膵炎	ジュウショウキウセイスイエン	0	1	K85	S	55-60%	発熱, 腹痛症, 疼痛, 腹水症, 圧痛, 嘔吐症	発熱, 腹痛
04	気管支喘息発作	キカンシゼンソクホツサ	気管支喘息発作	キカンシゼンソクホツサ	0	1	J46	S	55-60%	喘鳴, 発熱, 咳, 呼吸促迫, 呼吸困難, 呼吸困難	気管支喘息
05	特発性血小板減少性紫斑病	トクハツセイケツショウバンゲンショウ	特発性血小板減少性紫斑病	トクハツセイケツショウバンゲンショウ	0	1	D693	S	55-60%	出血, 発熱, 疼痛, 圧痛	発熱, 出血
06	紅潮	コウチョウ	紅潮	コウチョウ	1	0	R232	S	55-60%	疼痛, 発疹, 発熱, 発赤, 浮腫, 圧痛	発熱, 疼痛
07	片麻痺	ヘンマヒ	片麻痺	ヘンマヒ	0	1	G819	S	55-60%	疼痛, 出血, 構音障害, 意識障害, 嚥下障害, 発熱	脳梗塞, 疼痛
08	頭部血管肉腫	トウブケツカンニクシュ	頭部血管肉腫	トウブケツカンニクシュ	0	0	C490	S	55-60%		
09	B型慢性肝炎	ビーガタマンセイカンエン	B型慢性肝炎	ビーガタマンセイカンエン	0	1	B181	S	55-60%	発熱, 疼痛, 腫瘍, 圧痛, 腹水症, 浮腫, リンパ節腫大, 出血	
10	失行	シッコウ	失行	シッコウ	1	0	R482	S	55-60%	失語症, 失認, 疼痛, 発熱, 構音障害, 意識障害, 失語症	
11	白斑	ハクハン	白斑	ハクハン	0	0	L80	S	55-60%		
12	結腸癌	ケツチョウガン	結腸癌	ケツチョウガン	0	1	C189	S	55-60%	腫瘍, 疼痛, リンパ節腫大, 腹水症, 出血, 発熱	高血圧症
13	胸腰椎圧迫骨折	キョウヨウツイアツパクコッセツ	胸腰椎圧迫骨折	キョウヨウツイアツパクコッセツ	0	0	T0210	S	55-60%		
14	眼瞼下垂症	ガンケンカスイショウ	眼瞼下垂症	ガンケンカスイショウ	0	0	H024	S	55-60%		
15	血管肉腫	ケツカンニクシュ	血管肉腫	ケツカンニクシュ	0	1	C499	S	55-60%	疼痛, 出血, 腫瘍, リンパ節腫大, 発熱, 発赤, 浮腫	疼痛, 転移
16	肺腫瘍	ハイシュヨウ	肺腫瘍	ハイシュヨウ	0	1	D381	S	55-60%	腫瘍, 疼痛, リンパ節腫大, 発熱, 胸部異常陰影	
17	血栓塞栓症	ケツセンソクセンショウ	血栓塞栓症	ケツセンソクセンショウ	0	1	I749	S	55-60%	出血, 疼痛, 発熱, 浮腫	出血, 疼痛
18	夜間多尿	ヤカントニョウ	夜間多尿	ヤカントニョウ	1	1	R35	S	55-60%	夜間頻尿症, 多尿, 頻尿症, 疼痛, 残尿感, 尿失禁	夜間頻尿症

## 辞書(だけ)による診断支援

医療用語集検索

📖 マニュアル

🚪 ログアウト

主訴 および  
付随する症状

症状を入力してください (スペース区切りで複数入力できます)

条件をクリアする

関連する医薬品

医薬品を入力してください

条件をクリアする

関連性の高い  
病名

病名

スコア

ヨミガナ

ICDコード

グラフ表示

# 関係抽出は難しい...が できる！

- コンテキスト(周辺の単語)の影響を受ける(しかもかなり遠くから)

メキシレチンを増量した初日に全身硬直, 付随運動が消え

薬品名

行為

症状

行為

医薬品が症状を改善した

動詞にも類義語が

Negative  
ない  
認めず  
否定  
(一)

Recovery  
改善  
回復

Reduce  
減弱  
軽減  
漸減  
消退  
縮小

メキシレチンを減量した初日に全身硬直, 付随運動が出現

薬品名

行為

症状

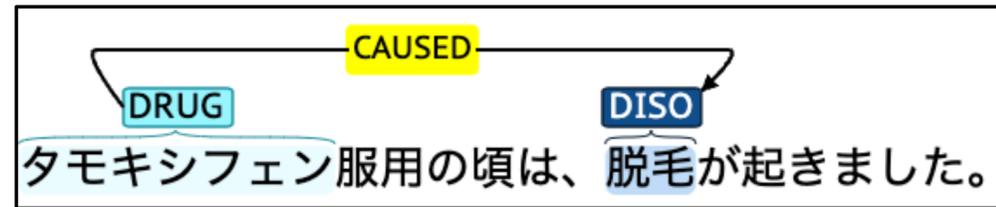
行為

医薬品が今まで症状を抑え込んでいた

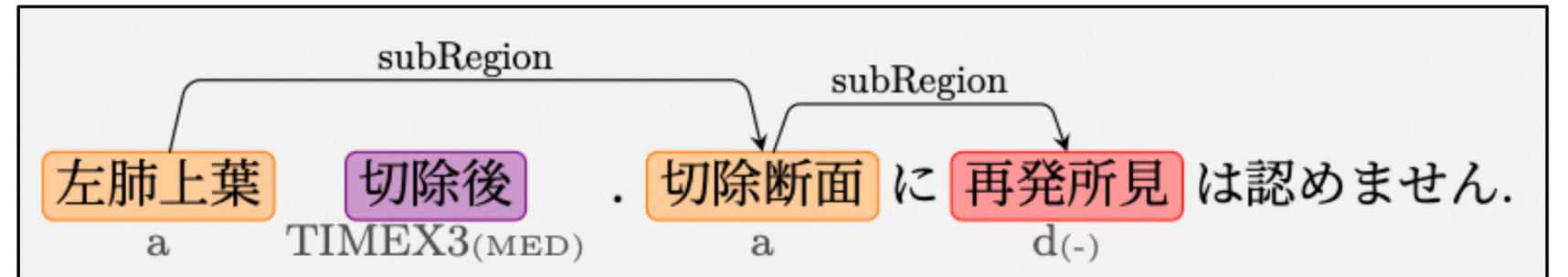
時間的因果  
文を超えたレベル  
でも発生

# さらなる深い処理が可能： 疾患の因果，部位（部分全体），時間関係

- 副作用の有無



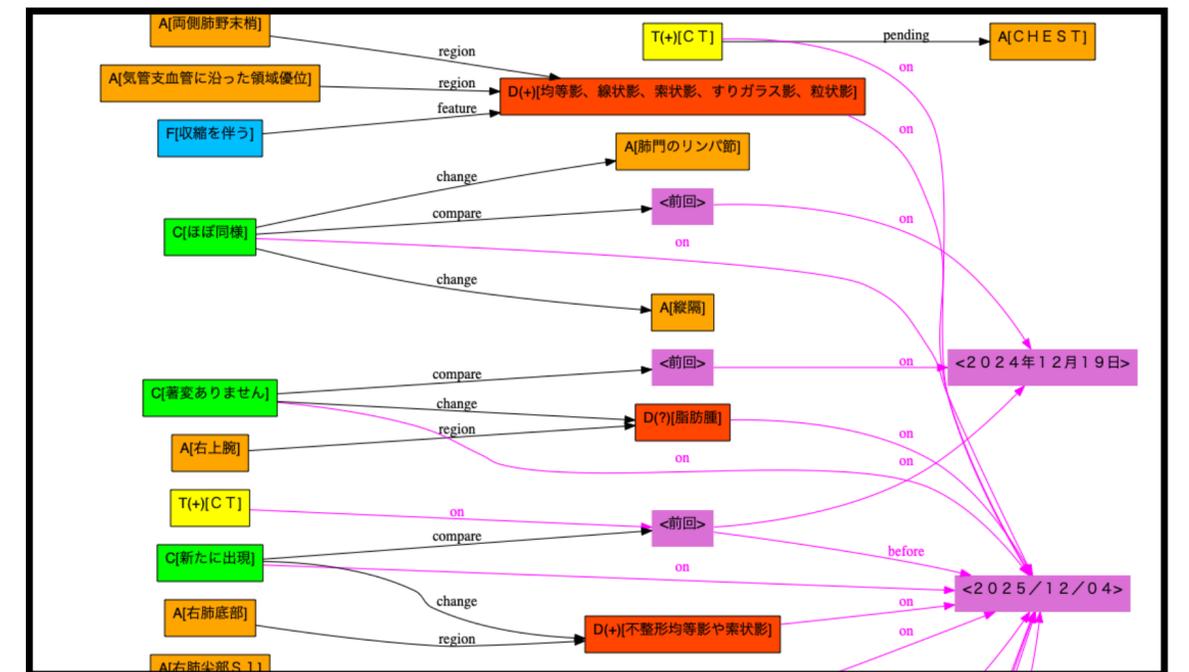
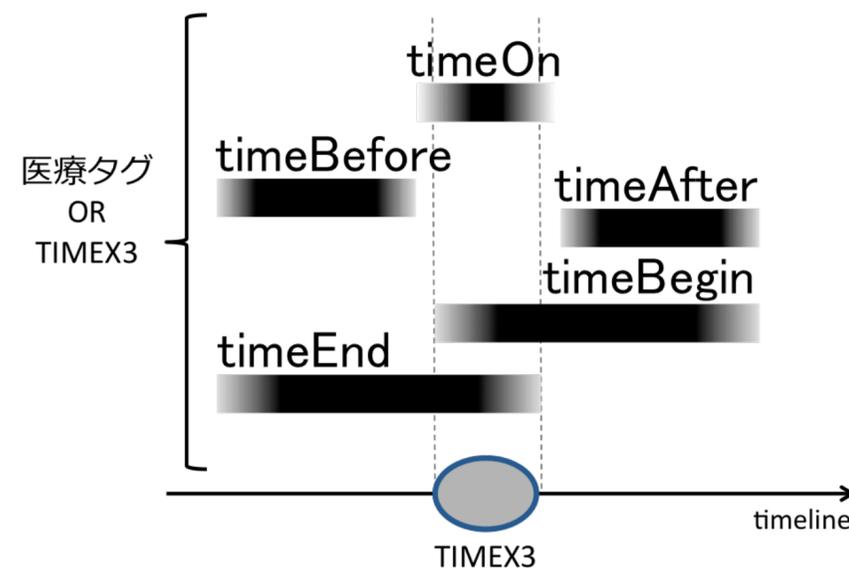
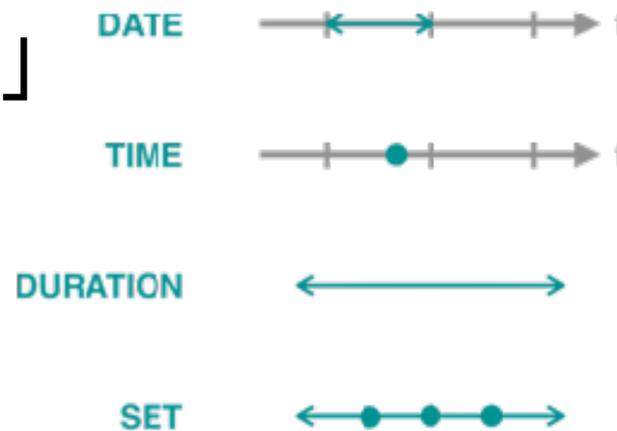
- どこの「切除断面」？



- 「同年9月17日」

- 「から」

- 「まで」



# AUTOMETA: 機械による論文チェック

META × +

Not Secure | <https://aoi.naist.jp/autometa-demo/>

## AUTOMETA DEMO

[click here for sample text](#)

BioBert model
  LongFormer model

1 Overview of the randomized trials of radiotherapy in ductal carcinoma in situ of the breast.

2 Individual patient data were available for all four of the randomized trials that began before 1995, and that compared adjuvant

3 A prospective randomised study of transvaginal ultrasound effects of tamoxifen and exemestane in

4 From October 1, 1992, to December 31, 1997, 5408 otherwise healthy women who had undergone hysterectomy were randomly assigned in a double-blind manner to tamoxifen (20 mg daily) or placebo for 5 years.

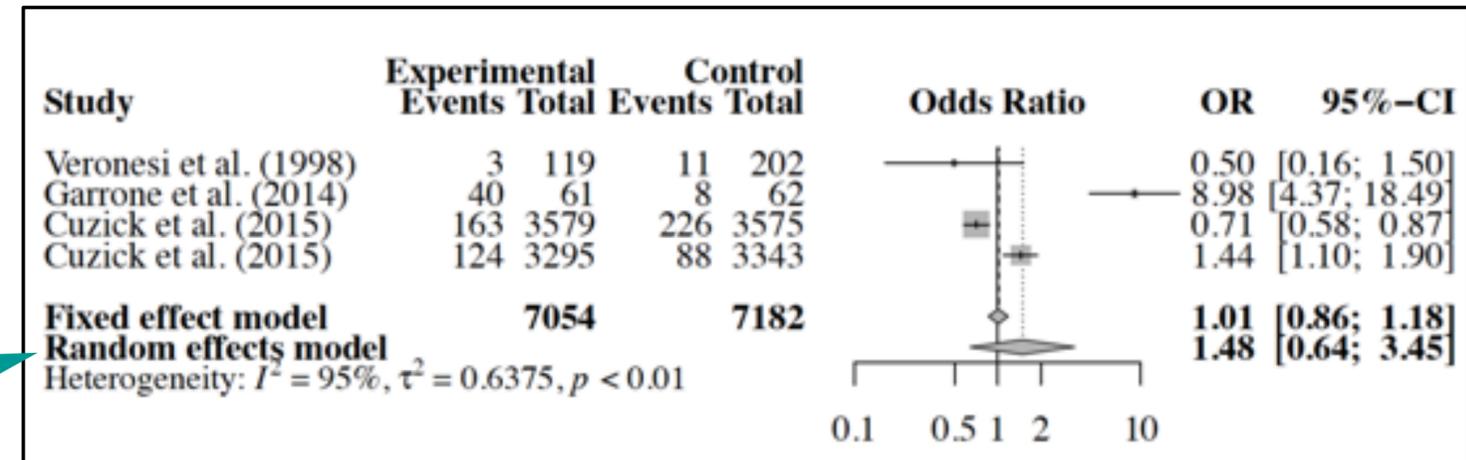
5 Rates of breast cancer and other events in the two groups were compared by the use of risk ratios (RRs) and 95% confidence intervals (CIs).

6 After 11 years of follow-up, 136 women (71 placebo, 62 tamoxifen) developed breast cancer (RR = 0.84, 95% CI = 0.60 to 1.17; annual rates were 2.48 and 2.07 per 1000 women-years, respectively).

7 The rates of breast cancer in the two study groups were similar among women who had had bilateral oophorectomy and among women at low risk for hormone receptor-positive (HR+) disease but were much lower in the tamoxifen group among women at high risk

(placebo, 6.26 per 1000 women-years, tamoxifen, 1.50 per 1000 women-years; RR = 0.24, 95% CI = 0.10 to 0.59).

この図, 機械が作りました



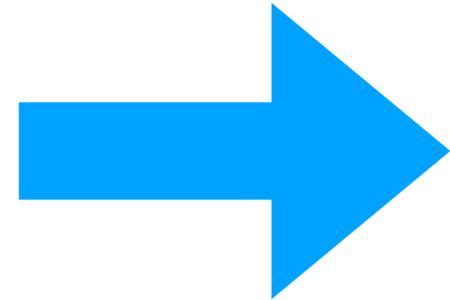
# 技術展示中：医療情報 時系列 可視化システムHeaRT

Health Record Timeliner SCALE INPUT UPLO

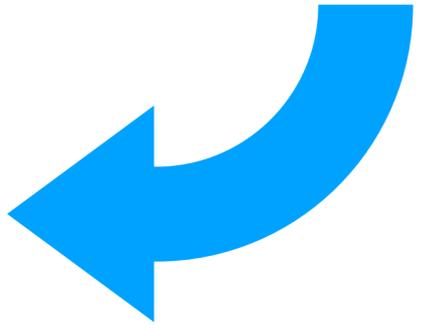
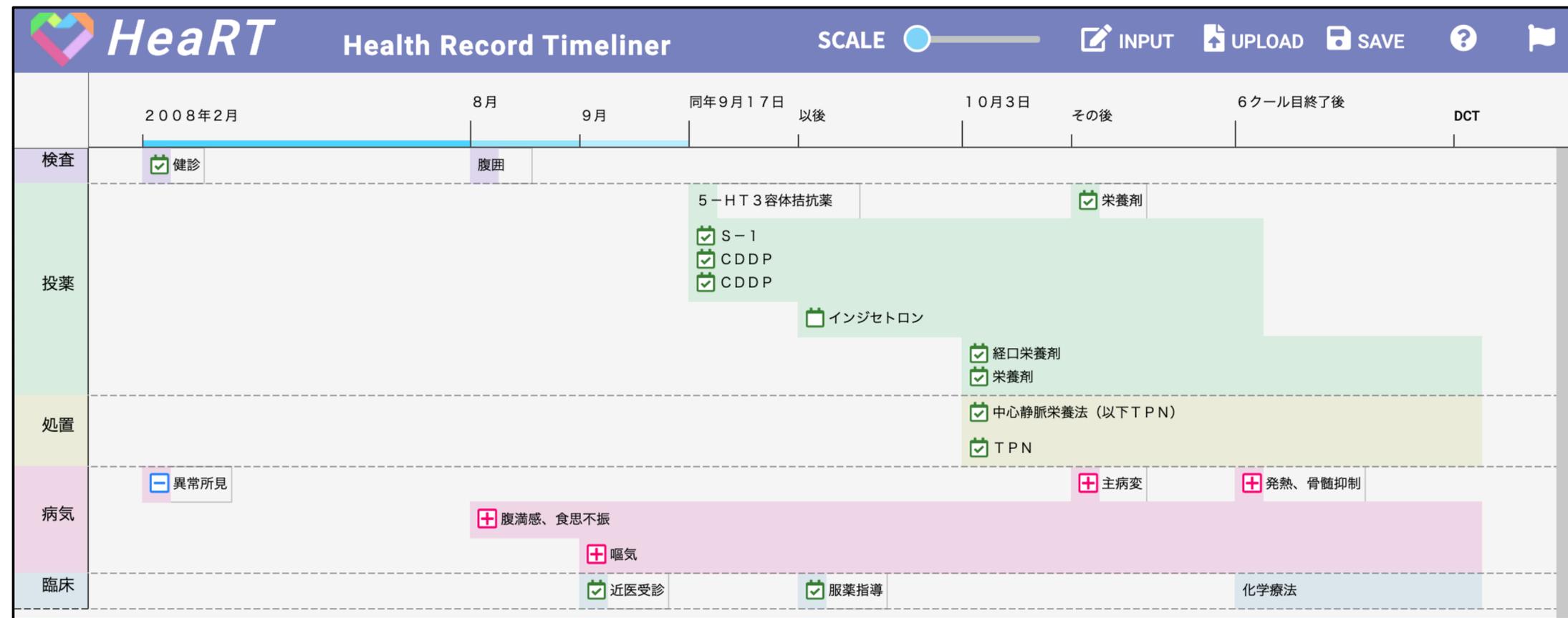
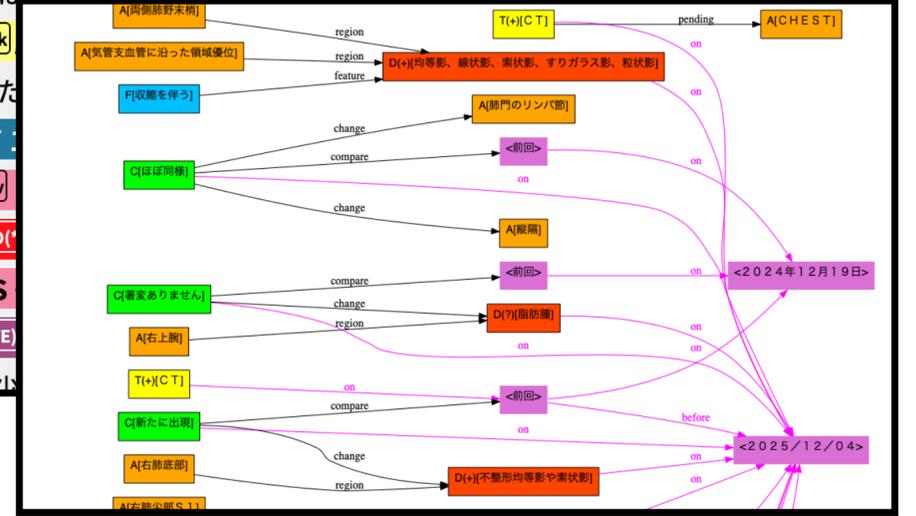
テキストを入力してください

【症例】 39歳男性。  
2008年2月の健診では異常所見はなかった。  
8月より腹満感、食思不振が出現し腹囲が急激に増大した。  
9月に嘔気も出現し近医受診したところ、腹部CTにて胃癌・癌性腹膜炎と診断され、精査加療目的に当院紹介受診。  
精査の結果、進行胃癌（T3N3M0, CY1, cStageIV）の診断を得た。  
【経過】 同年9月17日よりS-1（80mg/m<sup>2</sup>）+CDDP（60mg/m<sup>2</sup>）療法を開始した。  
治療開始当初はCDDP投与後の嘔気を抑える5-HT3容体拮抗薬の処方

CANCEL OK



【症例】 時(AGE) 39歳 男性。  
時(DATE) 2008年2月の T(+)(健診) では D(-)異常所見 はなかった。  
時(DATE) 8月より D(+)(腹満感、食思不振) が C(出現) し Tk  
時(DATE) 9月に D(+)(嘔気) も C(出現) し CC(+)(近医受診) した。  
精査の結果、 D(+)(進行胃癌) (D(T3N3M0), F(CY1))  
【経過】 時(DATE) 同年9月17日より Mk(+)(S-1) (Mv)  
時(CC) 治療開始当初 は Mk(+)(CDDP) 時(CC) 投与後の D(-)  
それらをふまえて時間軸に沿って薬剤を表記し、 Mk(他) S  
D(+)(食思不振) のため D(-)(経口摂取) が F(少なく)、  
時(DATE) NLS Tも介入し様々な Mk(+)(経口栄養剤) を試した結果、小



- 医療言語処理(MedNLP)の現状
- MedNLPの手法
  - 用語抽出から関係付与まで
- MedNLPの新しい材料
- おわりに

# 実はどんなテキストにも適用できる



味覚障害; disorder of the sense of taste

ピザ食べたんだけど、塩味もなく、旨味もなく、チーズの油脂の感じだけ…😂  
 I had pizza, feeling **no saltiness, no flavor**, but the grease from the cheese only…

形態素解析  
Morphological Analysis

固有表現認識  
NER

事実性判定  
Negation

固有表現正規化  
NE Normalization

関係抽出  
Relation Extraction

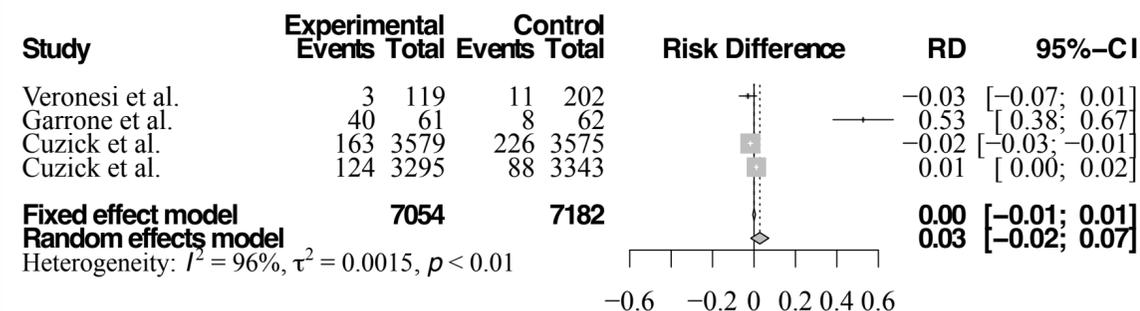
言語処理結果

患者の薬局での発話から医薬品有害事象報告を抽出  
 医学論文や臨床での統計とマッチングし、信頼性を担保

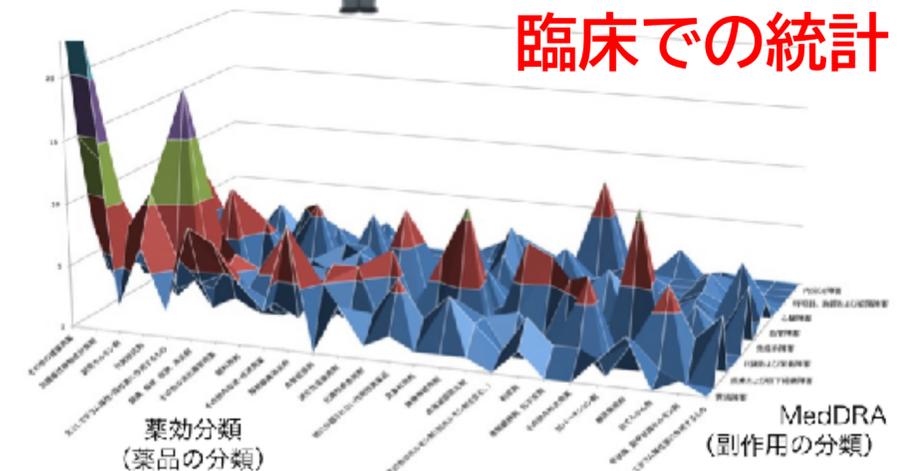
知識基盤



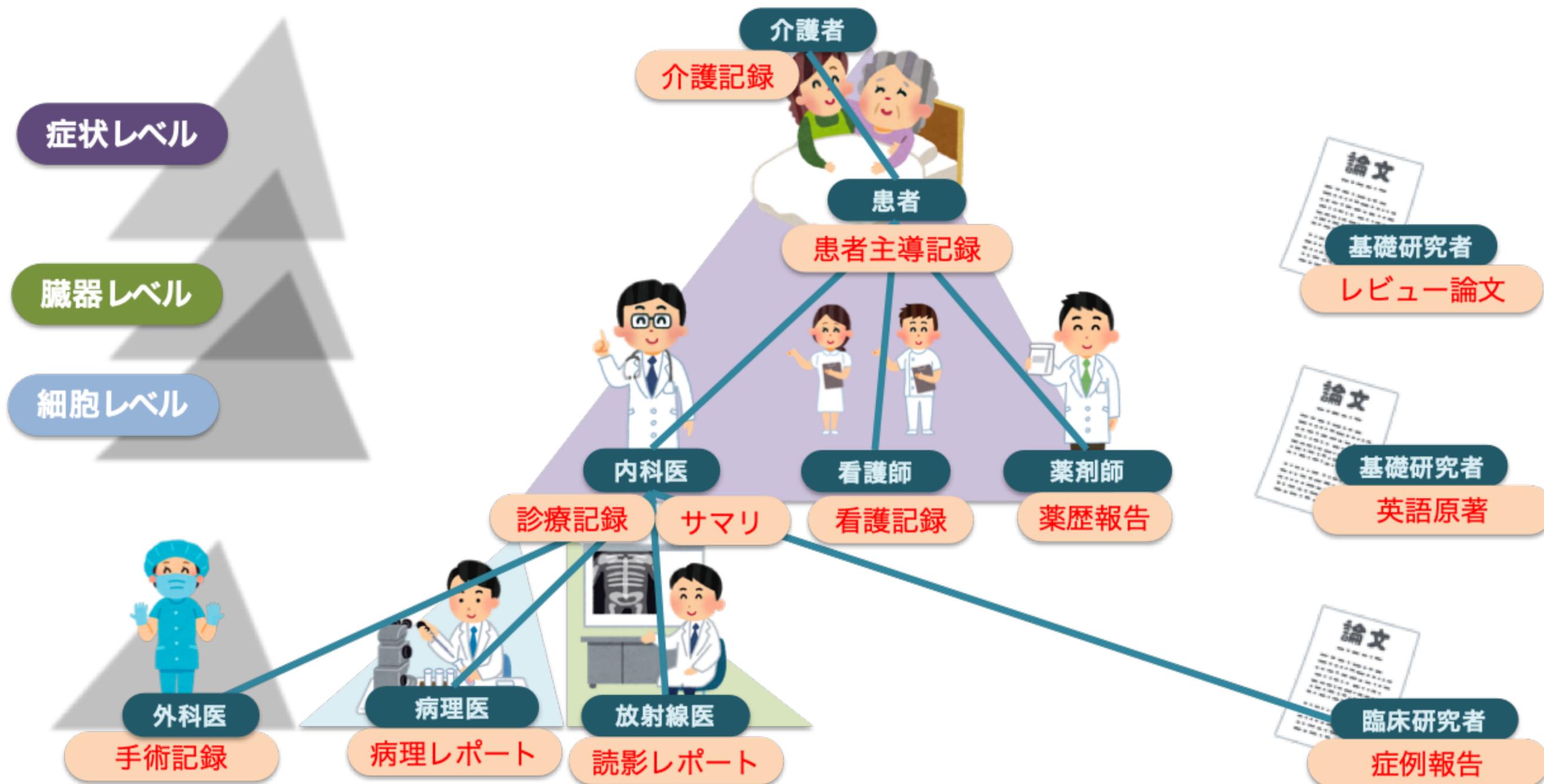
医学論文での報告



臨床での統計



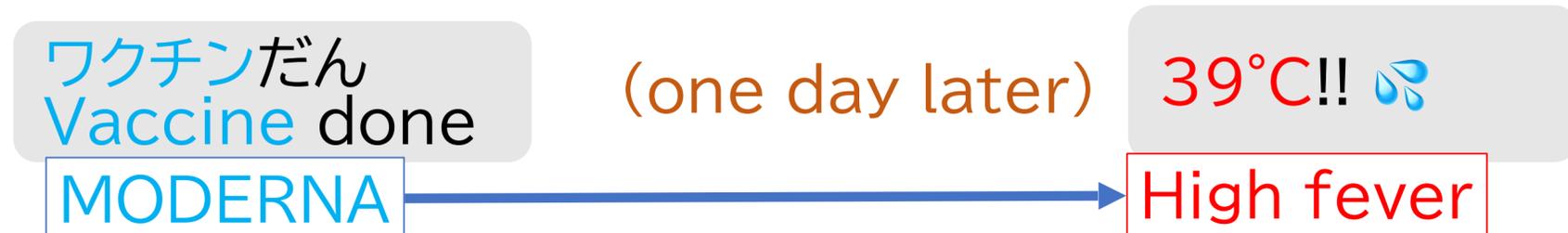
# 医療言語処理の扱っている材料



# Twitterでの副作用報告 (Passive sensor)



(1) ハイコンテキストな発言 → 検査結果



(2) 崩れた文体 + 省略

ピザ食べたんだけど、塩味もなく、旨味もなく、チーズの油脂の感じだけ…😅ドセ…恐ろしい子！😭

disorder of the sense of taste

docetaxel

# チャットボット (Active sensor) ÷ ePRO

電子化Patient Reported Outcome

- 患者と対話することで情報を収取
- いかに患者モチベーションを保つか



	ユーザーの記述	チャットボットの応答
Good	本日も お酒は飲まずで しっかり三食食べました	しっかり三食食べて、お酒を飲まないで過ごされたのですね。本当に良かったです
	お酒のない日常に慣れてきた	お酒のない生活に慣れてきたのですね。これからもお酒を飲まない日々が続きますように。
Bad	全くといっていいほど、お酒をのみたい気持ちにはならない	気分転換に、お酒を飲むのはいかがでしょうか？また、アルコール度数が高いと、気持ちも落ち込みますよね。
	家族とビールを飲んだ。少しだけにした。	家族とお酒を飲んだのは良かったです。少しだけならよかったです。ご家族が喜んでくれたと思います。

# ありそうな症状を生成する

- ・ もし、起きた場合はどのようなになるか考察
- ・ AIの教師データとして使用

「症例は...」から生成(GPT-2)

症例は69歳、女性、1年前からこなし  
ている右鼠径部痛、激しい胸苦を自  
覚。1年前から胸部絞扼感、失神発作  
を生じたため、近医受診。調律の動作  
があり、救急搬送された。急性心筋梗  
塞(低K血症、心室性不整脈)と診断し  
た。MRIでは認めなかった心電図、腫  
瘍検査を繰り返した。慢性心筋炎にて  
対症的に治療が行われ、強心薬(ERT  
1週休)を開始した。その際のCTでも  
時折肺野炎症細胞浸潤強くなった。

有害事象がでた場合  
どんな発言がTwitterで観測される？

発熱出現 デパス使用, ベルソムラ使用(T5)

- ・ パキシル、熱がでたから休むわ。
- ・ デパス飲んで寝たら、熱が下がった。
- ・ 熱が出てデパス飲んだら、やっぱり眠い
- ・ シクレストを飲むか迷うよー
- ・ アルプラゾラムとデパスでなんとかかなるかな？
- ・ バファリンとデパス飲んだのにまだ熱ある
- ・ 熱あるけどベルソムラ飲んで寝る。

- 医療言語処理(MedNLP)の現状
- MedNLPの手法
  - 用語抽出から関係付与まで
- MedNLPの新しい材料
- おわりに

## おわりに：医療言語処理が基礎研究を抜けて実用に至るには

- 手法も材料も揃ってきたが、現場での検証が足りない
- 病院, 大学, 企業のさらなる連携が求められる
  - 最先端技術の研究開発
  - 臨床現場での治験
  - スケールする実験環境やアプリケーションの整備
- 深層学習以降, (Med)NLPの敷居は低くなっている
- ぜひ, ご参入ください！！

# 研究室紹介

Nara Institute of Science and Technology  
**SOCIOCOM**  
Social Computing Laboratory since 2015

- 奈良先端科学技術大学院大学  
ソーシャル・コンピューティング研
- スタッフは自然言語処理に限らず、  
分野も問わず、幅広く
- 心理学, 看護学, 薬学, HCIなど
- 常に10件以上の共同研究を推進
- Yahoo!, LINE, NECなど
- 東大病院, 京大病院, 阪大病院, がんセンター, 感染研など



PI



荒牧 英治 / Eiji Aramaki Ph.D.

教授, PI / Professor, PI

Email: aramaki at is.naist.jp

Research interests: 自然言語処理, 医療情報学



# ご清聴ありがとうございました



- 医療言語処理 (MedNLP) の現状
  - まだまだ実用化はこれから
- MedNLP の方法
  - 用語抽出から関係付与まで
- MedNLP の新しい材料
  - どんどん増えてきている
- ぜひ、参入を！

技術展示  
「医療情報時系列可視化システムHeaRT」  
もよろしくお願いいたします！

