

ニューラル自然言語処理の進展と 社会課題への取り組み

京都大学 大学院情報学研究科 教授

黒橋 禎夫

言語リソース

自己紹介：研究歴と言語リソース

1991～1993 日本語基礎解析
京大博士課程 JUMAN・KNP

1994～2000 京大テキストコーパス(科研基盤A)
京大助手・講師

2001～2005 格フレーム(科研学術創成)
東大助教授

2006～2022 検索エンジン基盤TSUBAKI(科研特定領域)
京大教授
日中・日英対訳コーパスASPEC(JST)
京大ウェブ文書リードコーパス(科研基盤B)
日本語BERTモデル(CREST)
診療テキストコーパス(PRISM)
対話コーパス(NEDO)

[博士取得者]

河原大輔(格フレーム,JGLUE)

清田陽司(**HOME'Sデータセット**)

荒牧英治(医療NLPリソース)

鍛冶伸裕(評価極性コーパス)

柴田知秀(日本語BERTモデル)

笹野遼平(名詞格フレーム)

中澤敏明(ASPEC,MTリソース)

村脇有吾(入力誤りデータセット)

Fabien Cromieres(NMTモデル)

原島純(**クックパッドデータセット**)

萩行正嗣(リードコーパス)

Chenhui Chu(ASPEC,V&Lリソース)

John Richardson(EBMTモデル)

Raj Dabre(MTリソース)

坂口智洋(時間情報アノテーション)

栗田修平(V&Lリソース)

Huang Yin Jou(対話コーパス)

Tolmachev Arseny(JUMAN++)

清丸寛一(対話コーパス)

黒橋・楮・村脇研究室

[Top](#)[メンバー](#)[研究紹介](#)[研究発表一覧](#)[NLPリソース](#)[内部ページ](#)

自然言語処理のためのリソース[↑]

ツール[↑]

- [日本語解析器 KWJA](#)
- [日本語形態素解析システムJUMAN](#)
- [日本語形態素解析システムJuman++](#)
- [日本語構文解析システムKNP](#)
 - [PyKNP \(JUMANとKNPのPythonバインディング\)](#)
- [解析結果汎用表示ツール TableDisplay](#)
- [KyotoEBMT](#)
- [Webページ変換ツール WWW2sf](#)
- [検索エンジン基盤 TSUBAKI](#)
 - [Webページ構造解析モジュール DetectBlocks](#)
 - [Webページ発信者候補抽出モジュール DetectSender](#)

データ・コーパス[↑]

- [京都大学テキストコーパス](#)
- [京都大学ウェブ文書リードコーパス](#)
- [京都大学格フレーム](#)
- [京都大学名詞格フレーム](#)
- [BERT日本語Pretrainedモデル](#)
- [基本料理知識ベース](#)

ニューラル自然言語処理の進展

2013年 : 単語のベクトル表現、word2vecの提案

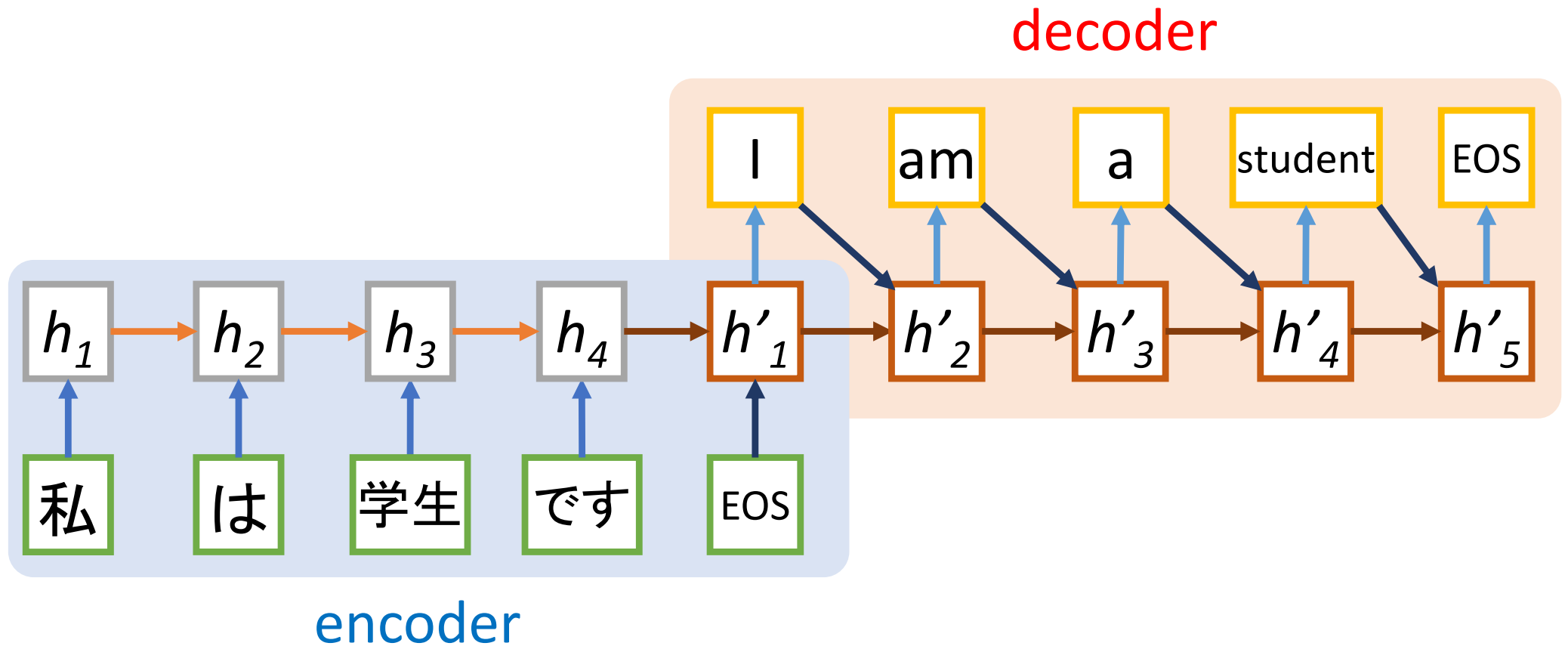
2014年 : ニューラル機械翻訳の提案

2015年頃 : NLPにおいてニューラルネットワークの
利用が本格化

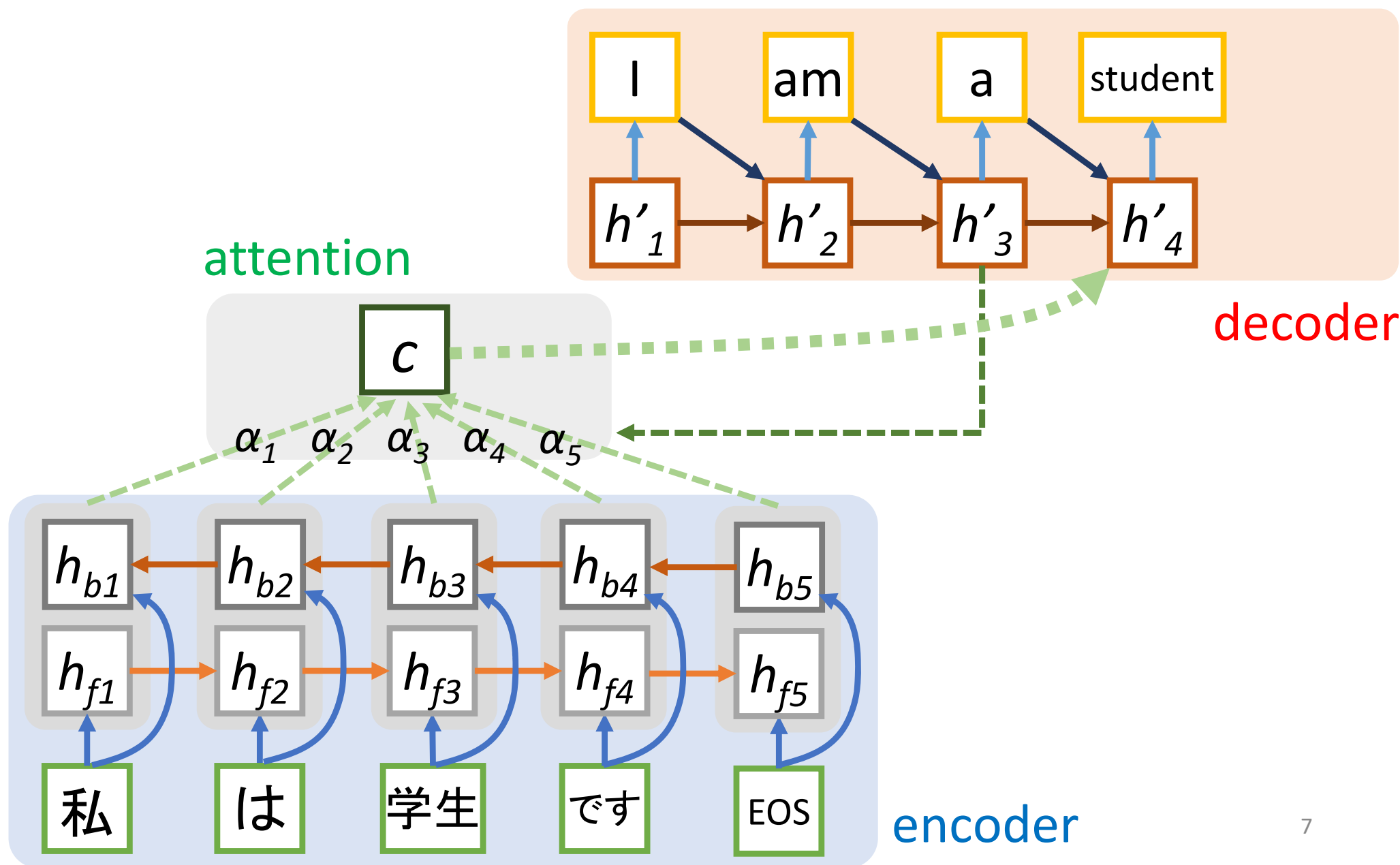
2017年 : Transformerの提案

2018年 : 汎用言語モデルBERTの提案

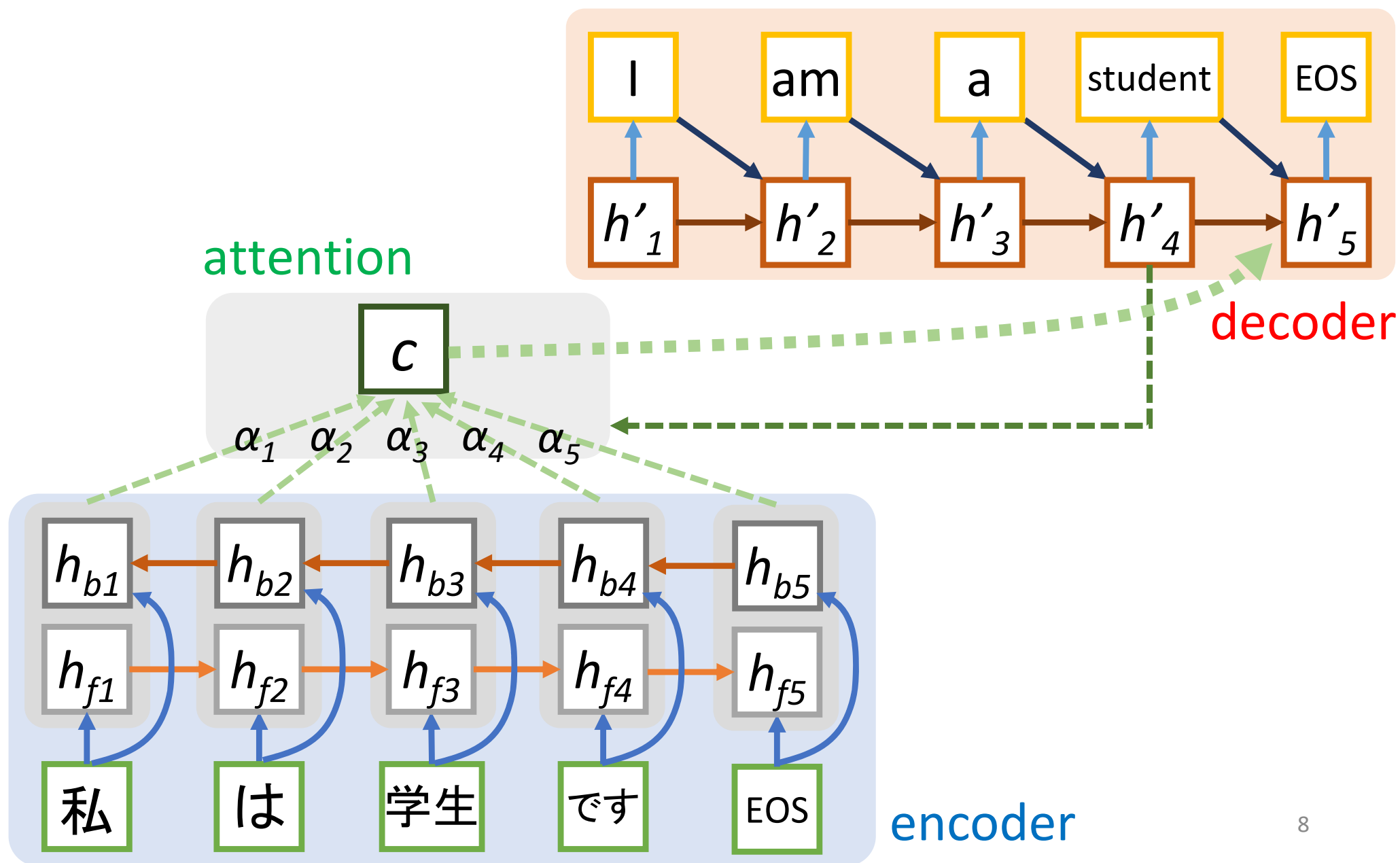
ニューラル翻訳 [Sutskever+, 2014]



ニューラル翻訳 [Bahdanau+, 2014]

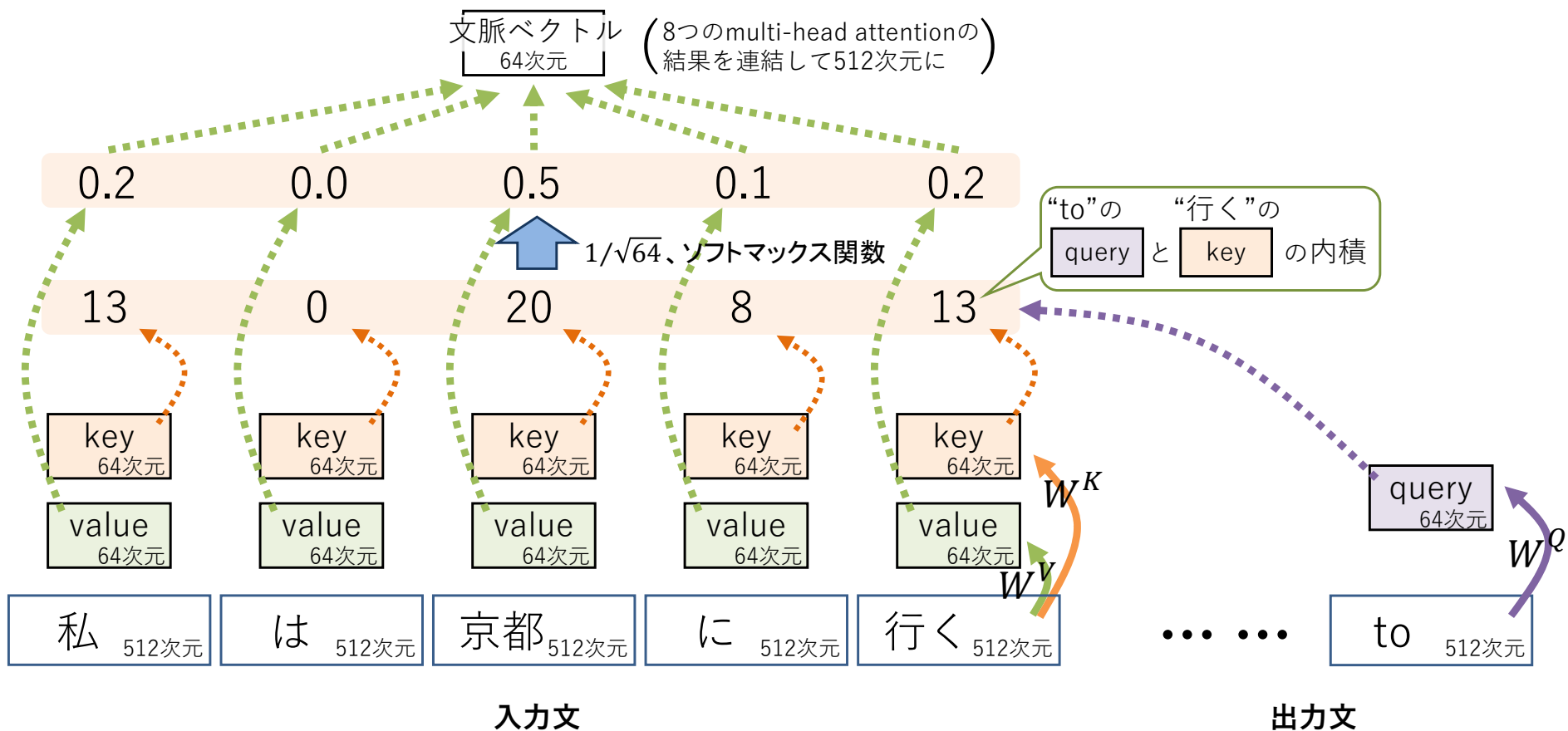


ニューラル翻訳 [Bahdanau+, 2014]



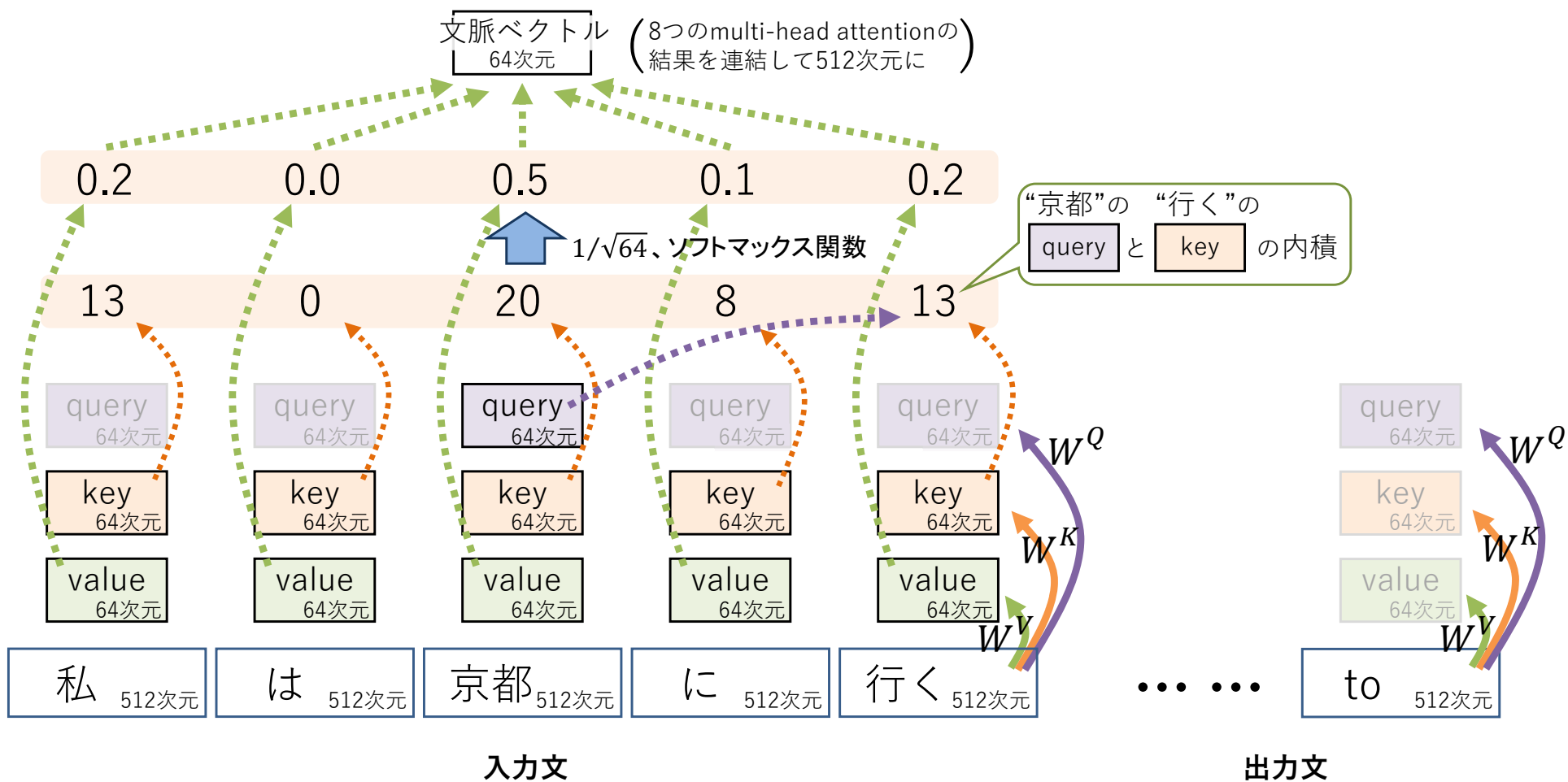
Transformer [Vaswani+, 2017]

Cross-Attention



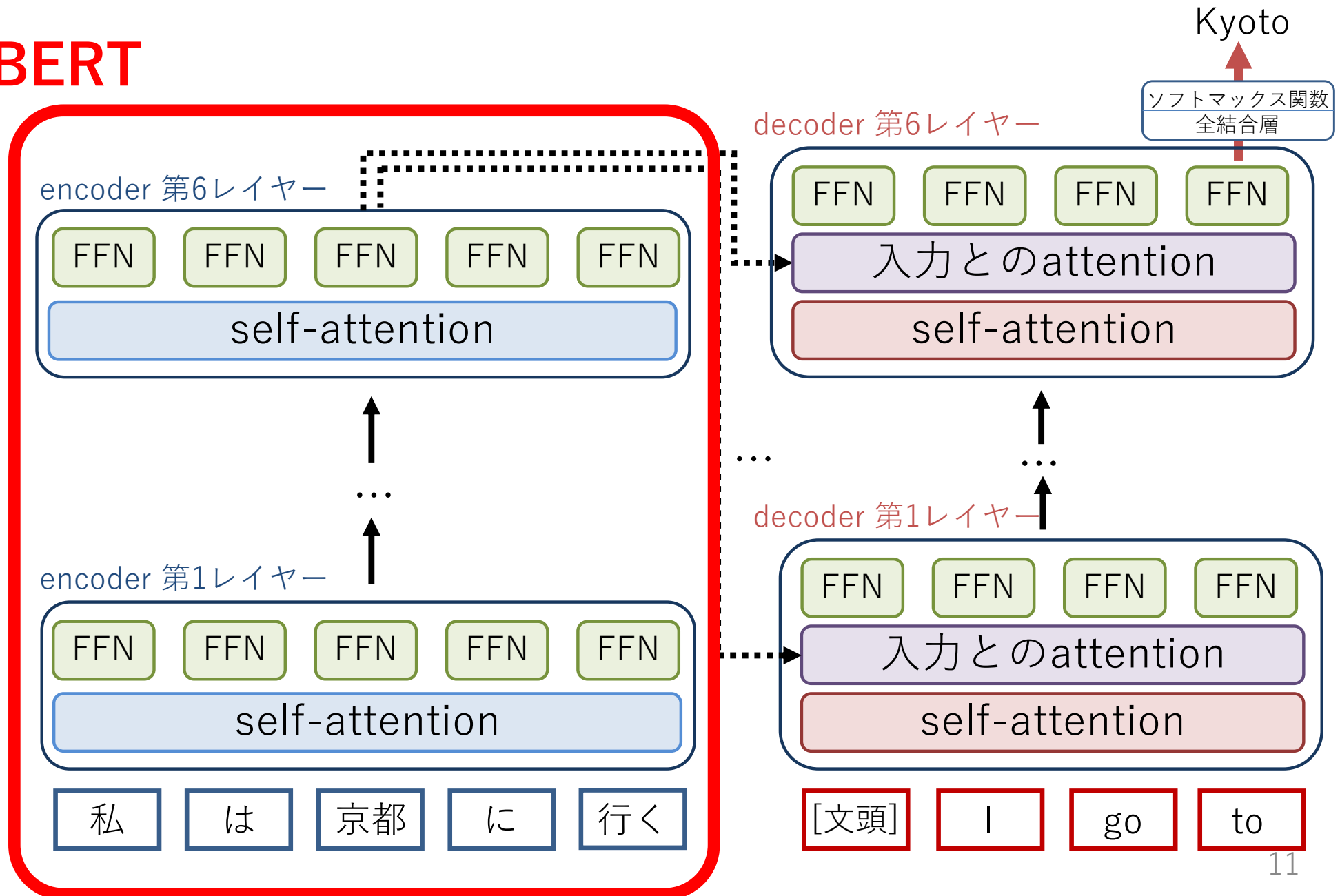
Transformer [Vaswani+, 2017]

Self-Attention



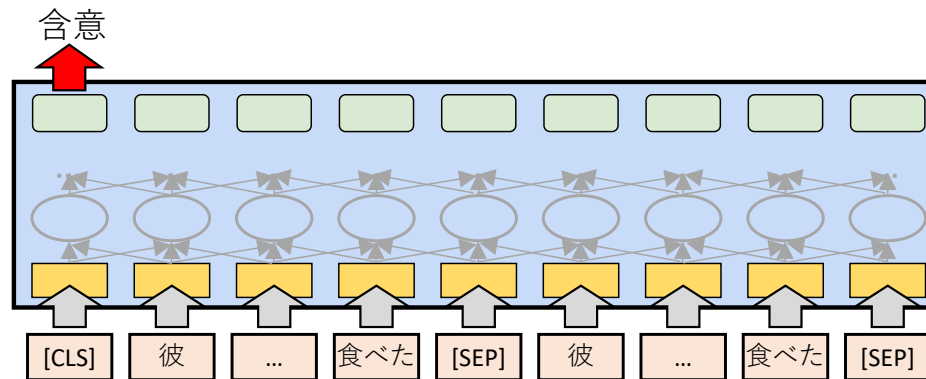
Transformer [Vaswani+, 2017]

BERT

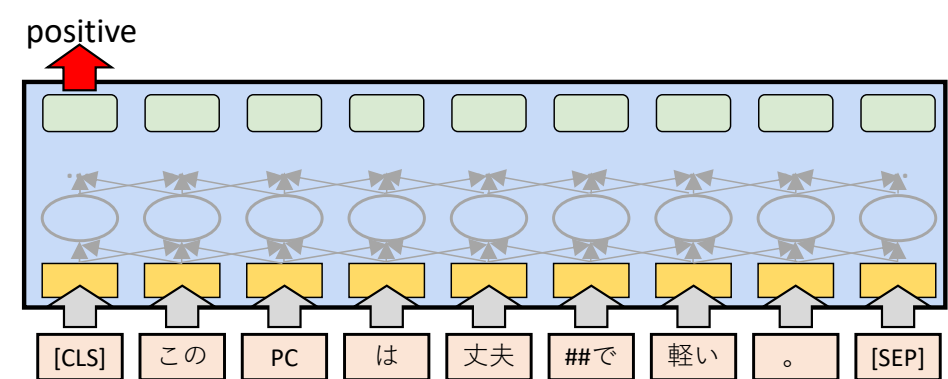


様々なタスクでFine-Tuning

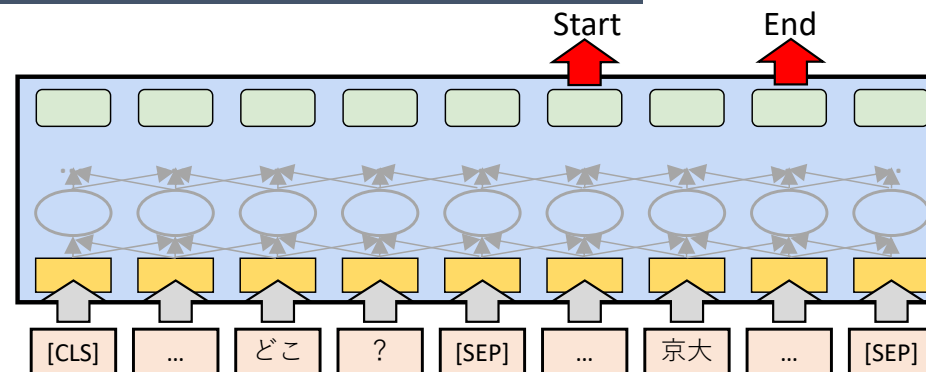
文ペア分類問題（例: 含意関係認識）



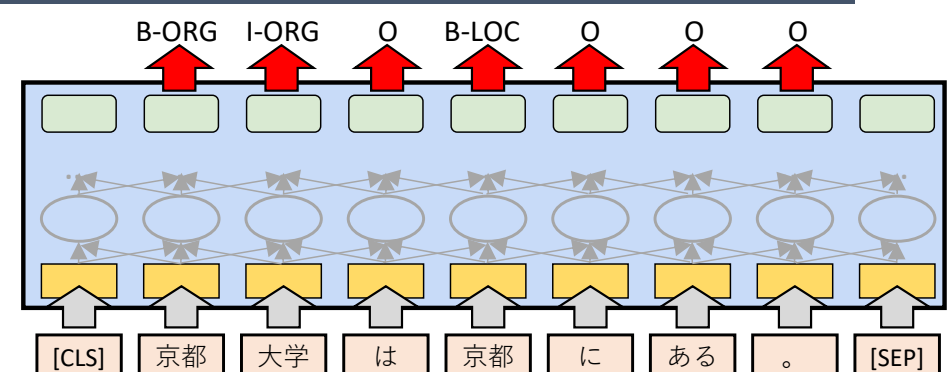
1文分類問題（例: 極性判定）



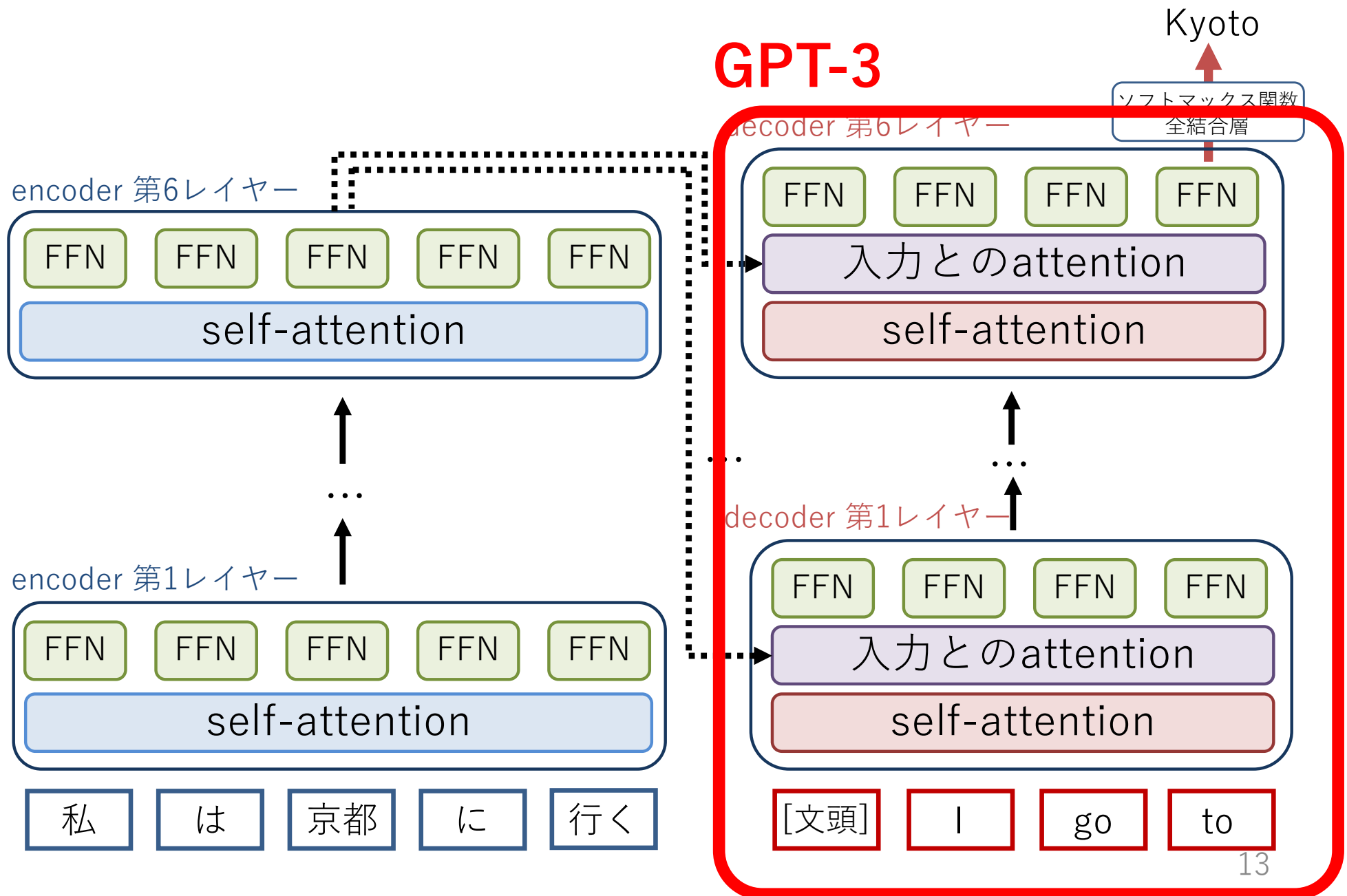
スパン抽出（例: 質問応答）



系列ラベリング（例: 固有表現解析）



Transformer [Vaswani+, 2017]



GPT-3

Answer the following question.

Question: I was hungry, but I didn't eat anything. Why?

Answer: 私はお腹がすいたけど、何も食べなかった。なぜでしょう？



There could be a number of reasons why someone might not eat even though they're hungry.

Maybe they're trying to lose weight, or maybe they can't afford to buy food. Or, there could be a medical reason why they can't eat, such as an allergy or a condition that makes it difficult to digest food.

おそらく減量中、または、食べ物を買うお金がない。あるいはアレルギーや食べ物の消化が難しいなどの医学的理由があるのかも知れない。

<https://beta.openai.com/docs/quickstart/start-with-an-instruction>

<https://arxiv.org/abs/2005.14165>

最近の大規模汎用言語モデル

model	organization	date	#params	d _{model}	#layers	#heads	#tokens
GPT-2	OpenAI	19.02	1.5B	1,600	48	25	?
GPT-3	OpenAI	20.05	175B	12,288	96	96	300B
Gopher	DeepMind	21.12	280B	16,384	80	128	300B
LaMDA	Google	22.01	137B	8,192	64	128	168B
MT-NLG 530B	MS/NVIDIA	22.01	530B	20,480	105	128	270B
Chinchilla	DeepMind	22.03	70B	8,192	80	64	1.4T
PaLM	Google	22.04	540B	18,432	118	48	780B
OPT	Meta	22.05	175B	12,288	96	96	180B
BLOOM	BigScience	22.07	176B	14,336	70	112	350B
HyperCLOVA 39B	LINE	21	39B	8,192	48	64	300B?
HyperCLOVA 82B	LINE	22?	82B	10,240	64	80	300B?

(早稲田大学 河原大輔教授提供)

ChatGPT



私はお腹がすいたけど、何も食べなかった。なぜでしょう？

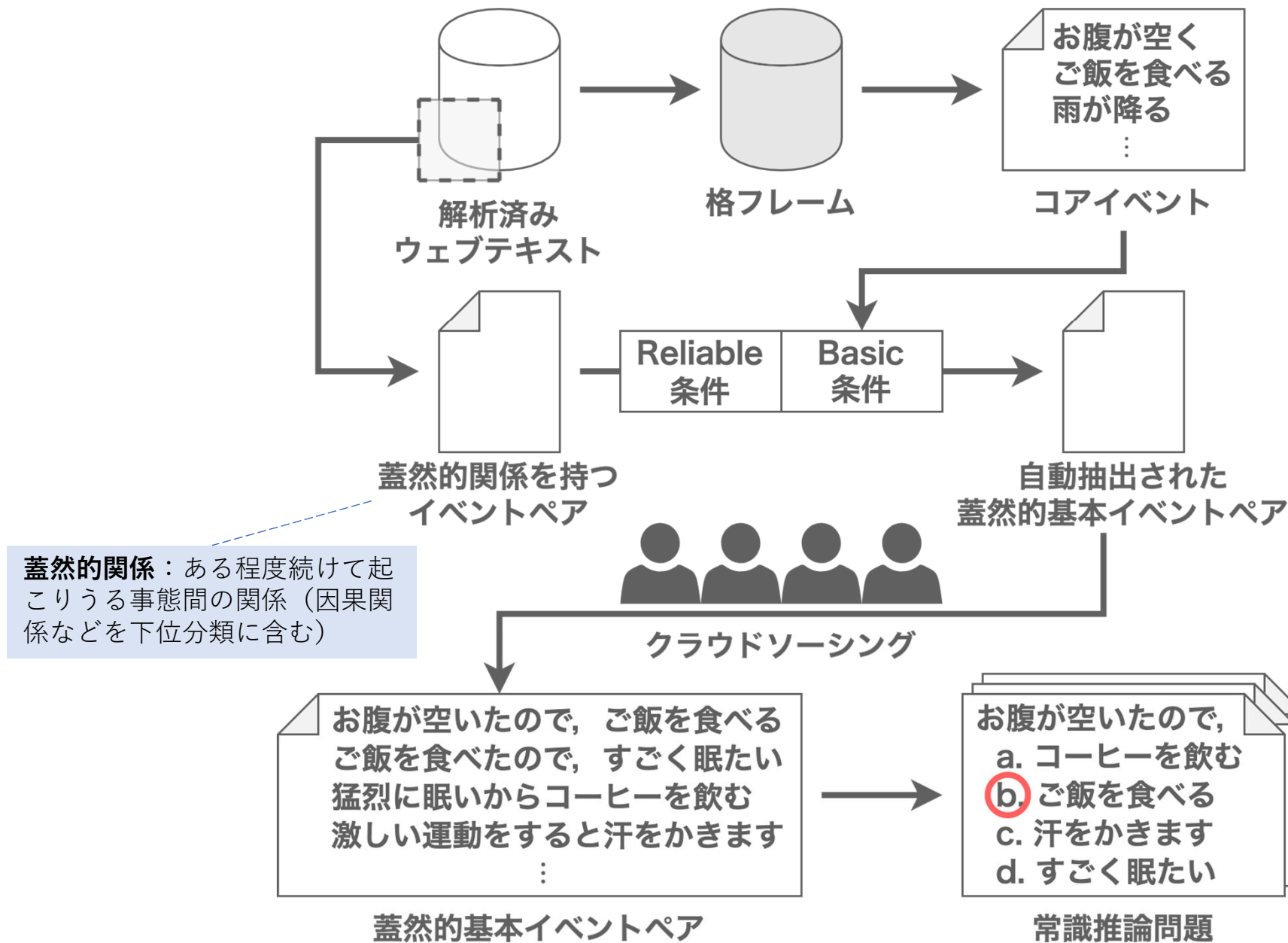


お腹がすいたのに食べなかった理由は、いくつか考えられます。ひょっとしたら、周りに食べるものがなかったか、まだ食べる時間じゃないと思っているかもしれません。また、ストレスや不安があると、食欲が減退することがあるので、そういった状況があるかもしれません。お腹がすいたら、できるだけ早めに何かを食べるようにしましょう。



常識推論データセットの構築と利用

(日本漢字能力検定協会との共同研究 2017.7-2022.3)



蓋然的関係：ある程度続けて起こりうる事態間の関係（因果関係などを下位分類に含む）

常識推論データセットの構築と利用

- 日本語**7億文**のウェブコーパスから17万組の蓋然的な基本イベントペアを自動抽出
- クラウドソーシングによって蓋然的関係を持つと判断された**10万組**のイベントペアを利用し、**10万問**の常識推論問題を生成

霧が晴れると、

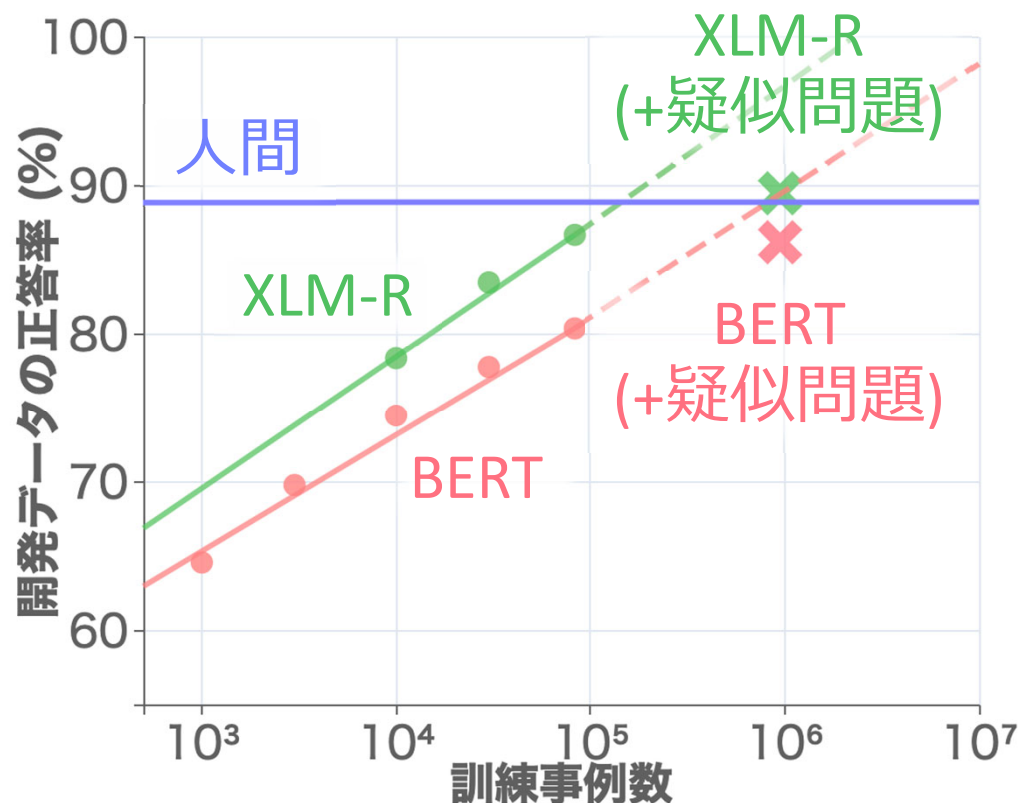
- ☒ a. 景色が素晴らしい
- b. 川の音がすごい
- c. 雪遊びも楽しそうだ
- d. 写真写りがいい

仕事辞めたら

- a. 仕事でいっぱい
- ☒ b. 生活は大変だ
- c. 勉強がはかどるはずだ
- d. ますます犯罪が増える

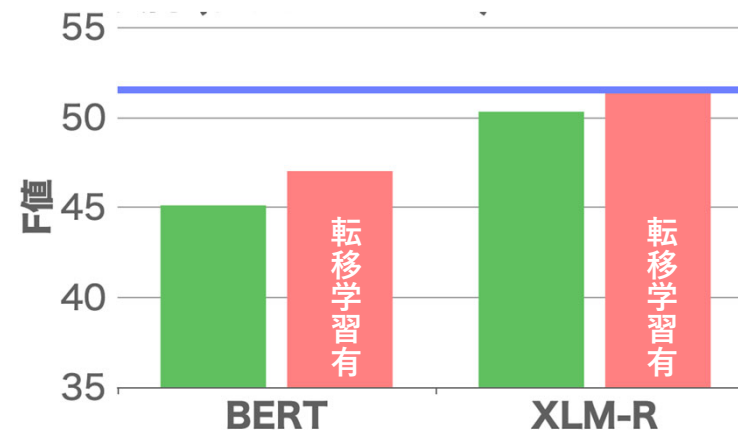
- さらに、日本語**33億文**のウェブコーパスから、クラウドソーシングのチェックを省略して**86万問**の疑似問題を生成

常識推論問題の 解答結果

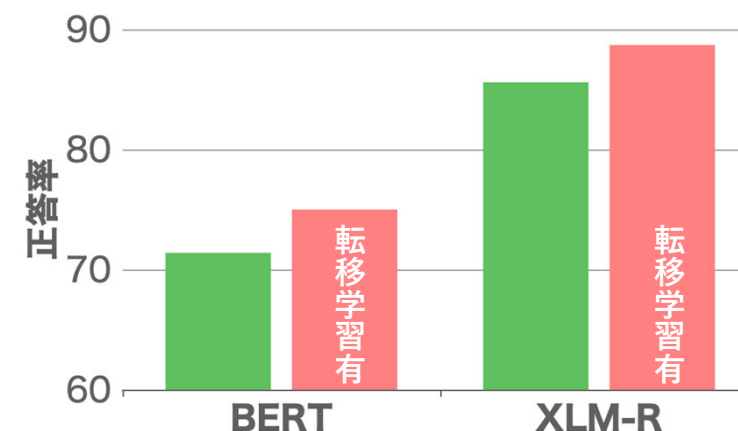


蓋然的関係知識の 転移学習

談話関係解析 (京大ウェブ文書リードコーパス)



日本語WSC (照応解析の難問)



K. Omura, D. Kawahara and S. Kurohashi: A Method for Building a Commonsense Inference Dataset based on Basic Events, EMNLP 2020 (2020.11).

K. Omura, S. Kurohashi: Improving Commonsense Contingent Reasoning by Pseudo-data and its Application to the Related Tasks, COLING 2022, Korea (2022.10). **Outstanding Paper Award**

小学生用の作文学習ゲーム： 「いちまるとはじめよう！ことばむすび」

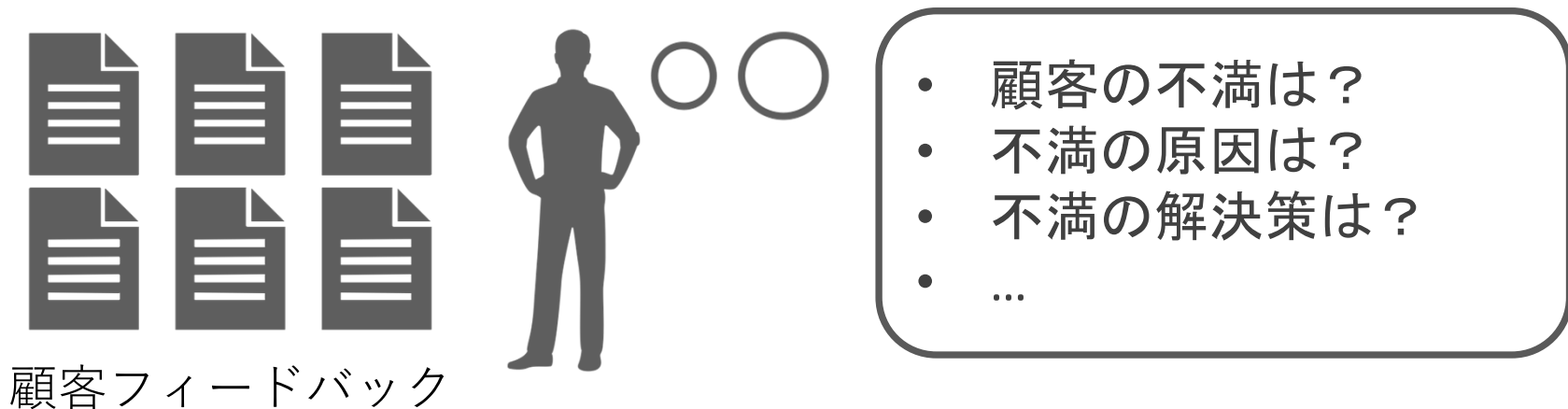
- 語と語、文と文を格助詞や接続助詞でつないでいくことで、言葉の使い方(共起関係)や文同士の関係(蓋然的関係)を学べるゲーム



KWJA：汎用言語モデルに基づく日本語解析器

(Kyoto Waseda Japanese Analyzer)

- NLP タスクは end-to-end 学習で解くのが主流になった
- しかしテキスト分析などゴールが不明確で試行錯誤を要するタスクでは end-to-end 学習の適用が困難

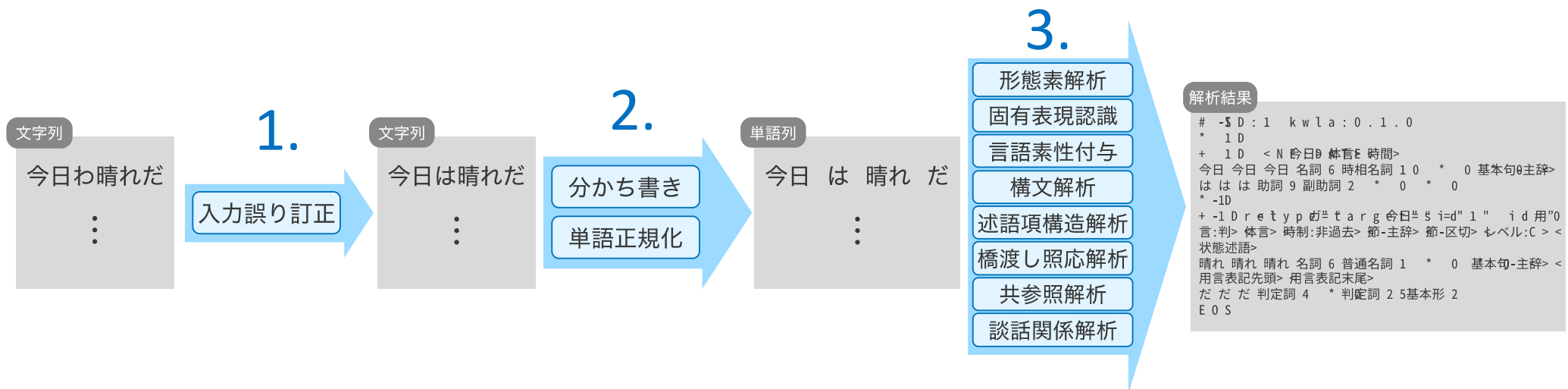


- テキスト分析では**人間が分かる粒度でテキストを整理**するのが有効で、そこでは依然として**言語解析が有用**

KWJA：汎用言語モデルに基づく日本語解析器

(Kyoto Waseda Japanese Analyzer)

- 汎用言語モデルベースで**あらゆる言語解析を実行**
- 汎用言語モデルの能力を積極的に活用することで、**人手によるコード・ルールの記述を最小化**（性能改善は汎用言語モデルの改善で達成）
- データセットさえ整備すれば**新たなタスクにも容易に適用可能**



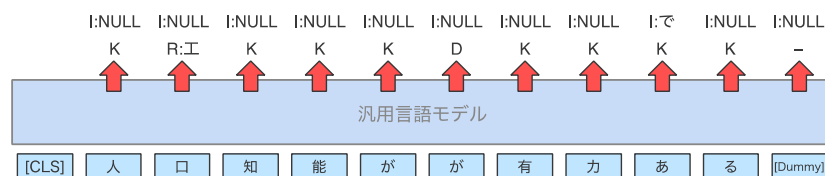
KWJA：汎用言語モデルに基づく日本語解析器

(Kyoto Waseda Japanese Analyzer)

すべての解析タスクを**系列ラベリング**または**単語選択**で解く

1. 入力誤り訂正モジュール

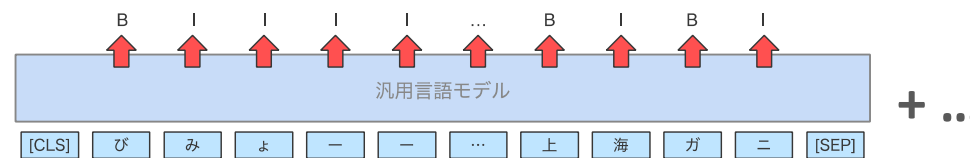
入力誤り訂正（系列ラベリング）



▶ 各文字に施す編集操作（削除等）を予測

2. 文字レベル解析モジュール

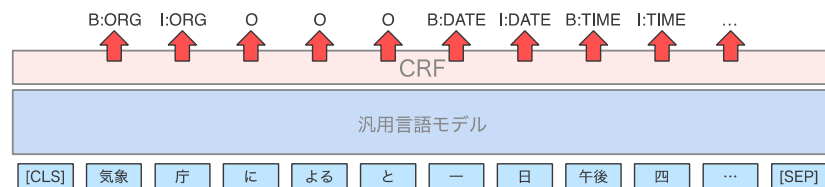
分かち書き（系列ラベリング）



▶ 各文字が単語の先頭か続きか予測

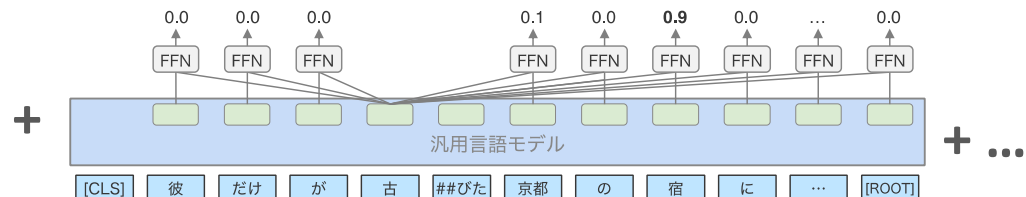
3. 単語レベル解析モジュール

固有表現抽出（系列ラベリング）



▶ 各単語が固有表現（人名等）か予測

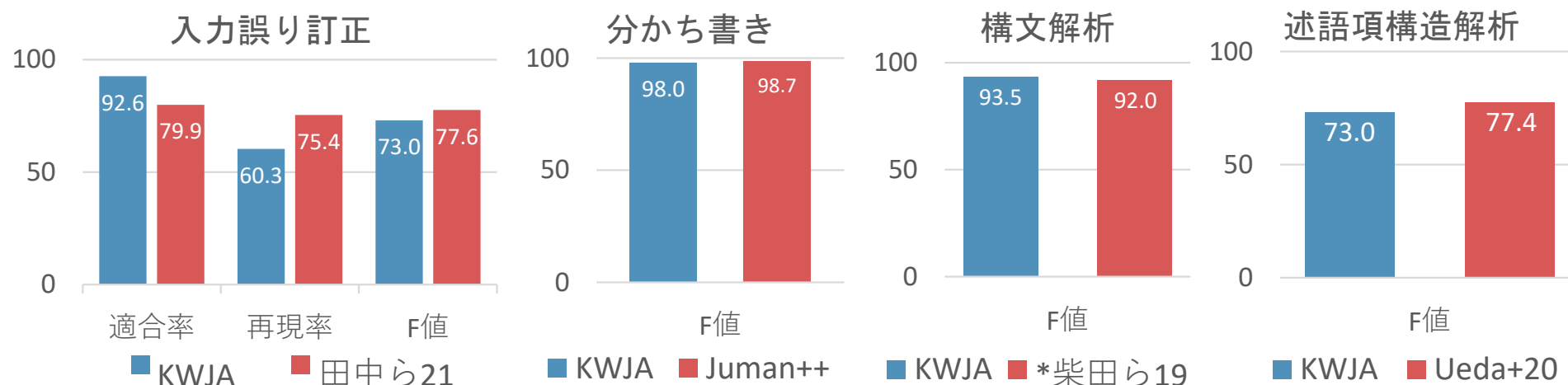
構文解析（単語選択）



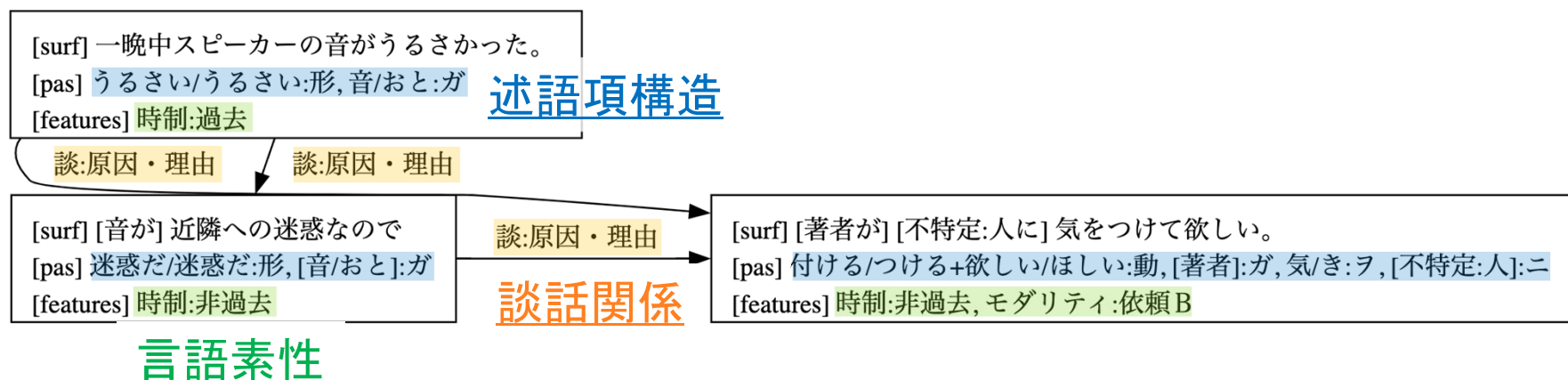
▶ 各単語のかかり先の単語を予測

解析結果

- 多くの解析タスクで既存モデルに匹敵する精度を達成



- 解析例



応用：因果関係グラフ [清丸ら20]

- 顧客フィードバックを分析し、不満の原因・結果・解決策を提示



応用：因果関係グラフ [清丸ら20]

- 顧客フィードバックを分析し、不満の原因・結果・解決策を提示



社会課題への取り組み

- 医療テキストの構造化
- 行政対話ボット
- 暗黙知表出インタビュー

PRISM:医療テキスト構造化のための言語・知識処理基盤の構築

黒橋 禎夫(京大) 荒牧英治(NAIST) 荒瀬由紀(阪大) 戸次大介(お茶の水女子大)

(医薬基盤研、国立がん研究センターとの共同研究 2018～)

目標: 医療分野における臨床テキスト、患者テキストの解析・構造化のための言語・知識処理基盤を構築し、**特発性肺線維症 (IPF)** および**肺がん**の創薬標的予測に資するテキスト構造化を実現する

読影所見の解析

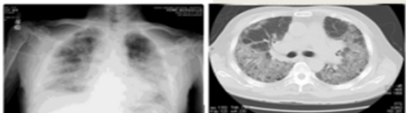
CT-① 胸部単純 CT 胃がん術後

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、**蜂巢肺**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

positive, negative

医薬基盤研・伊藤先生



用語・関係抽出

胸水貯留を認めないが、**大動脈の石灰化**を認めます

J90 NEG POS I700

事実性判定

構造化情報の抽出

網状影:s:下|下葉|両側肺|背側|胸膜
すりガラス影:s:下|下葉|両側肺|背側|胸膜
胸膜下線状影:s:下|下葉|両側肺|背側|胸膜
牽引性気管支拡張:s:下|下葉|両側肺|背側|胸膜
間質性肺炎:s:
多形状結節:s:右肺|胸膜|葉間
石灰化結節:s:上葉|右肺
炎症後変化:s:上葉|右肺
リンパ節腫:s:
脂肪沈着:s:左室壁
心内膜下梗塞後変化:s:左室壁
石灰化:s:冠動脈
肥厚:s:右乳腺A領域|右腋窩の皮膚

間質性肺炎:s:
すりガラス影:s:下葉|両側肺
網状影:s:下葉|両側肺
牽引性気管支拡張:s:下葉|両側肺
NSIP:s:
炎症後変化:s:上葉|両側肺
石灰化:s:上葉|両側肺
間質性肺炎:p:
...

創薬

臨床テキストのアノテーションコーパス構築

医療エンティティ・[属性]

- ・病名・症状(D) [有・無・疑]
- ・臓器・部位(A)
- ・特徴・尺度(F)
- ・変化(C)
- ・検査(T-TEST, T-KEY, T-VAL) [予定・実施・中止]
- ・薬品(M-KEY, M-VAL) [予定・実施・中止]
- ・治療(R) [予定・実施・中止]
- ・クリニカルコンテキスト(CC) [予定・実施・中止]
- ・時間表現(TIMEX3) [DATE・TIME・DURATION・SET・AGE・MED]

エンティティ間の関係

- ・changeSbj
- ・changeRef
- ・featureSbj
- ・subRegion
- ・keyValue
- ・timeOn
- ・timeBefore
- ・timeAfter
- ・timeBegin
- ・timeEnd

汎用的な アノテーション仕様

- ・診療録, 読影所見, 症例報告など, 幅広い臨床テキストに
- ・病変検出, 副作用抽出, 診断経過推定など, 様々な応用に
- ・乱立する仕様を統合

医療エンティティ 属性

病変・症状エンティティ<d>は具体的な病変や症状を指す表現に付与する。病変の存在を示す異常な見え(「すりガラス状」, 「網状影」など), 異常な聴診音(「粗い雑音」, 「coarse crackle」など)などの所見にもタグを付与する。

その病変・症状が実際に認められたかどうかの程度(確実性: certainty)を属性として付与し, 以下の値をとる。

positive: その病変・症状が患者に実際に認められた場合

suspicious: 患者がその病変・症状に該当するのではないかと疑われている場合(鑑別疾患²³として提案されているなど)

negative: 患者についてその病変・症状の存在が否定された場合

general: 患者に関係ない一般的な病変・症状に関する記述や, 将来生じる可能性のある病変・症状を表現している場合

XMLタグ

```
<d certainty="{positive" or "suspicious" or "negative" or "general"}">...</d>
```

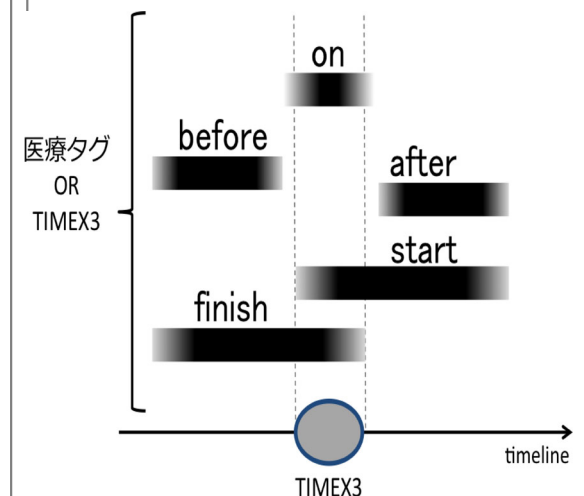
2.3.1.1 エンティティ範囲の目安

原則として「の」などの助詞等を含まない, ひとまとまりの複合名詞が単位の目安である。複合名詞として表現される病変や症状の一部には, 特徴・変化的な表現(→2.3.3, 2.3.4節)や部位名(→2.3.2節)を含む場合もあるが, 全体として病変・症状タグを与える。また, 「○○の変化」のような表現の場合, 助詞「の」も含めて病変・症状エンティティとみなすことがある(→2.3.4節)。他のタグの方が適切な場合など, 適切なタグの選択にあたっては, 当該タグの説明も参照されたい。

例

(2) 一方<a>右肺下葉の<d certainty="positive">濃度上昇</d>は<c>わずかに減少</c>している。

エンティティ間関係



臨床テキストのアノテーションコーパス構築

医療エンティティ・[属性]

- 病名・症状(D) [有・無・疑]
- 臓器・部位(A)
- 特徴・尺度(F)
- 変化(C)
- 検査(T-TEST, T-KEY, T-VAL)[予定・実施・中止]
- 薬品(M-KEY, M-VAL)[予定・実施・中止]
- 治療(R)[予定・実施・中止]
- クリニカルコンテキスト(CC)[予定・実施・中止]
- 時間表現(TIMEX3)[DATE・TIME・DURATION・SET・AGE・MED]

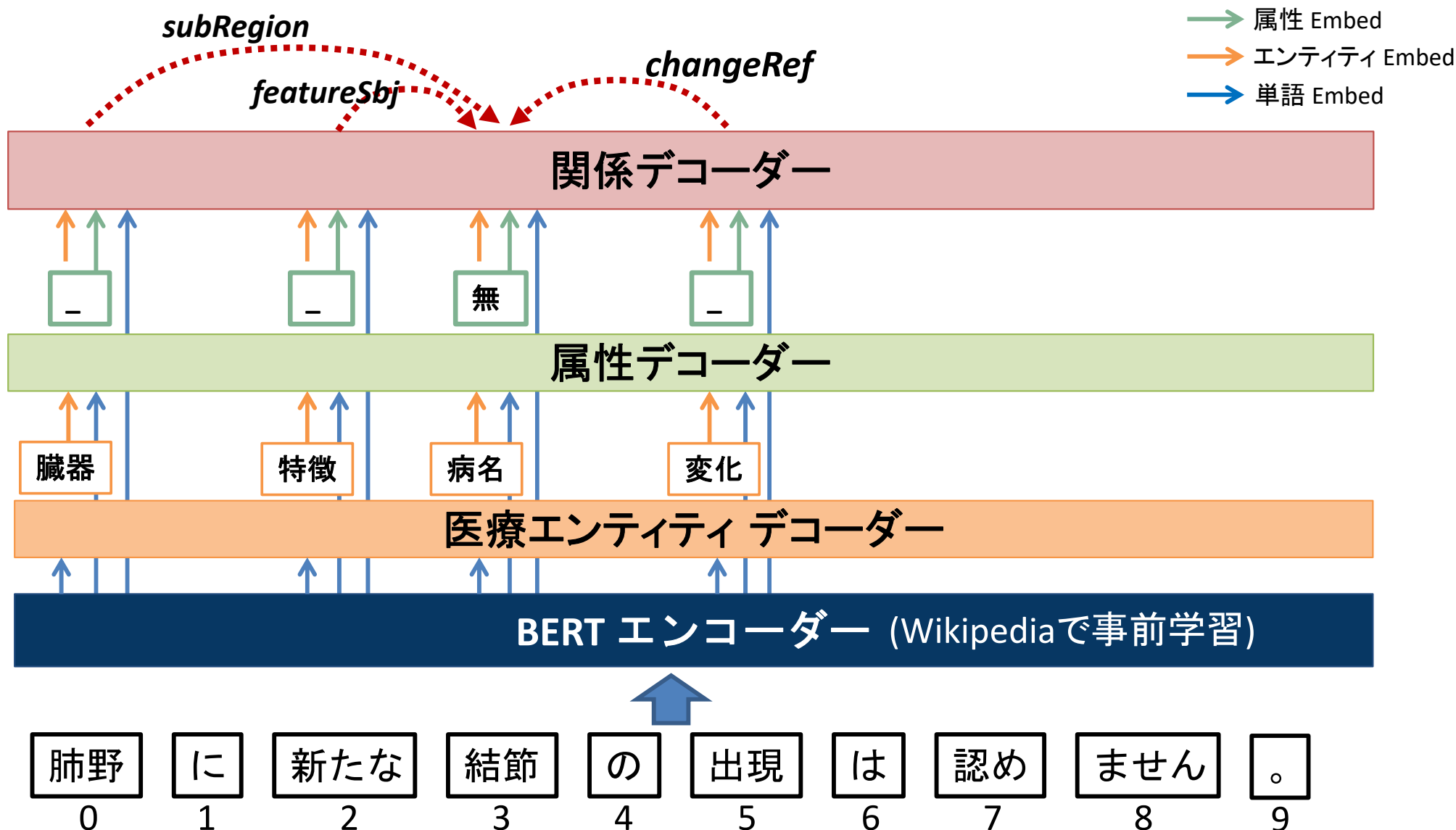
エンティティ間の関係

- changeSbj
- changeRef
- featureSbj
- subRegion
- keyValue
- timeOn
- timeBefore
- timeAfter
- timeBegin
- timeEnd

	肺がん読影所見	IPF読影所見	IPF診療録	
文書数	2,699	603	461	1,500
1文書の平均文数	12.5	13.3	40	28.9
1文書の平均単語数	160.4	155.2	408.8	338.6
患者数	2,699	603	461	81
主要な医療エンティティタグ数				
病名・症状 (D)	36,216	5,029	7,359	16,058
臓器・部位 (A)	18,400	2,946	1,930	3,139
検査・問診 (T-Test)	2,037	568	2,829	6,380
時間表現 (TIMEX3)	3,754	1,641	6,321	13,596
主要なエンティティ間関係数				
subRegion	26,147	4,813	2,123	2,899
timeOn	41,512	8,155	31,428	51,977

医療エンティティ・属性・関係認識 Jointモデル

[Cheng+ LREC2022; Cheng+ EMNLP2020 Findings; Wan+ EMNLP 2022]



弱教師学習に基づく症例報告の構造的要約

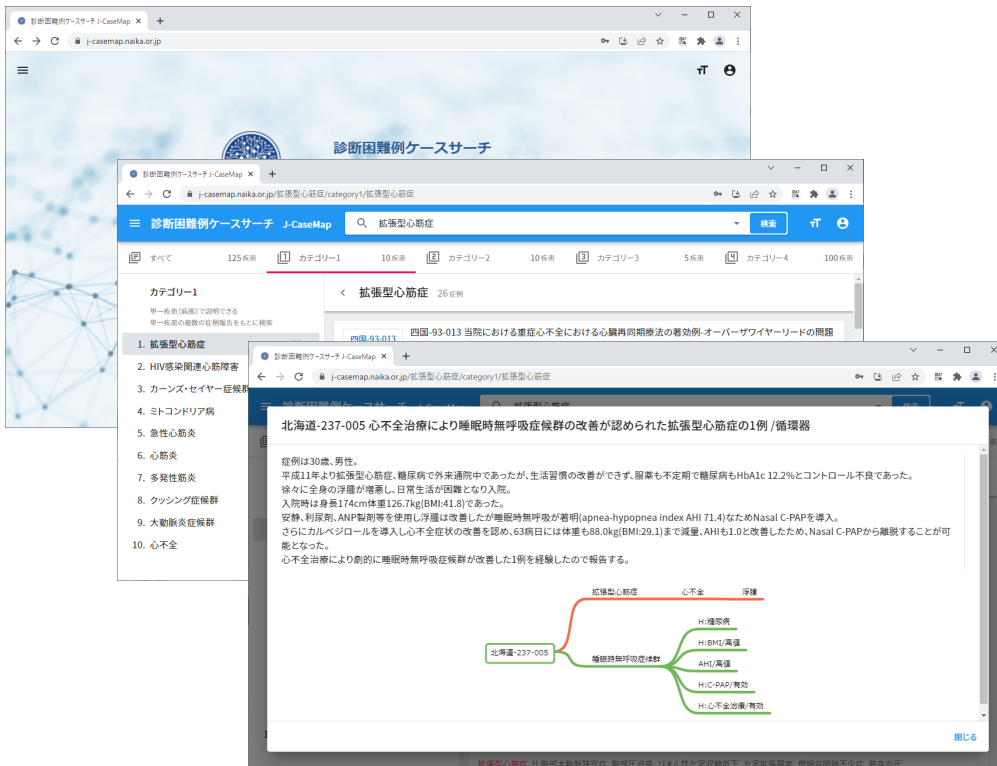
第26回日本医療情報学会春季学術大会(2022.7.2)

尾崎立一 清丸寛一 Cheng FEI 黒橋禎夫(京都大学) 永井 良三(自治医科大学) 佐藤 寿彦(株式会社プレジジョン)

研究の概要

- 専門化が進んだ医学分野では過去の症例の参照が重要
- 永井は症例報告の効率的な検索のため、1.5万件規模の症例報告に対して**グラフ構造要約**(下図)を手作業で整備
- 本研究では作業コストの低減と一貫性向上のため、**症例報告からグラフ構造要約を自動生成**する手法を提案
 - グラフ構造要約の各要素を症例報告中の言及箇所に対応づけることで**情報抽出の弱教師データを構築**し、情報抽出モデルを訓練・適用することでグラフ構造要約を生成
- 提案手法により**F値69.8**でグラフ構造要約中の関係を抽出可能

診断困難例ケースサーチ J-CaseMap (症例報告数1.5万件のグラフ要約)



ケーススタディ

症例報告

70歳 男性 帯状疱疹ウイルスによる出血性脳炎の70歳男性例。
【症例】70歳、男性【主訴】なし(意識障害)【現病歴】既往に胃全摘と2回の帯状疱疹がある。
当院搬入時、意識障害に加え、右側胸部の帯状疱疹と髄膜刺激徴候がみられた。各種精査にて、ビタミンB12欠乏症とリンパ球優位の髄液細胞増加を認め、くわえて、血清および髄液中のVZV-DNAが陽性であった。頭部MRIでは、大脳半球の皮質・灰白質境界部の類円形病変と、小脳半球および中脳被蓋部に微小出血を伴う多発性の高信号病変を認め、VZV脳炎と診断した。MRAでは異常所見は無かった。抗ウイルス剤に加え、ステロイドおよびメコバロミン投与を行ったところ、臨床症状は改善し、頭部MRIでも病巣の縮小を認めた。

人手の要約

△ビタミンB12欠乏症
△H:胃全摘
△H:メコバロミン/有効
—帯状疱疹
—帯状疱疹@側胸部
△VZV-DNA/陽性
—脳炎
—意識障害
—髄膜刺激徴候
✖髄液=(✖)細胞数/(✖)上昇
—髄液=リンパ球/上昇
—髄液=VZV-DNA/陽性
—MRI=類円形病変@大脳半球皮髄境界
✖MRI=(✖)T2高信号@ (✖)小脳半球
—MRI=微小出血@中脳被蓋部
✖MRI=(✖)T2高信号@ (✖)中脳被蓋部
—H:抗ウイルス剤/有効
—H:ステロイド/有効

自動要約

—帯状疱疹
—帯状疱疹@側胸部
△H:胃全摘
—脳炎
—意識障害
—髄膜刺激徴候
△ビタミンB12欠乏症
—髄液=リンパ球/高値
—髄液=VZV-DNA/陽性
—MRI=類円形病変@大脳半球
—MRI=微小出血@ (✖)小脳半球
—MRI=微小出血@中脳被蓋部
—H:抗ウイルス剤/有効
—H:ステロイド/有効
△H:メコバロミン/有効

凡例：正解 / 足りない予測 / 誤った予測 / 親子関係のズレ

弱教師学習に基づく症例報告の構造的要約

第26回日本医療情報学会春季学術大会

尾崎立一 清丸寛一 Cheng FEI 黒橋禎夫(京都大学) 永井 良三(自治医科大学), 佐藤 寿彦(株式会社プレジジョン)

Step 1: グラフ構造要約の関係三つ組への分解

拡張型心筋症
- 浮腫
睡眠時無呼吸症候群
- 糖尿病
- AHI/高値

グラフ構造要約



- (拡張型心筋症, 親子, 浮腫)
- (睡眠時無呼吸症候群, 親子, 糖尿病)
- (睡眠時無呼吸症候群, 親子, AHI)
- (AHI, モダリティ, 高値)

関係三つ組

症例は30歳男性。拡張性心筋症、糖尿病で外来通院中。全身の浮腫が増悪し入院。睡眠時無呼吸が著明（AHI 71.4）なため…

症例報告

Step 3: 情報抽出 モデルの学習



症例は30歳男性。拡張性心筋症、糖尿病で外来通院中。全身の浮腫が増悪し入院。睡眠時無呼吸が著明（AHI 71.4）なため…

アノテーション付き
症例報告

Step 2: 症例報告テキストへのマッピング

関係	親子	検査	場所	補足	モダリティ	全体
三つ組数	703,257	56,470	50,226	36,588	121,344	967,885
F値	65.8	66.3	65.9	76.0	75.8	69.8

対話ボットをベースとした行政と市民の 新たなコミュニケーションチャネルの構築

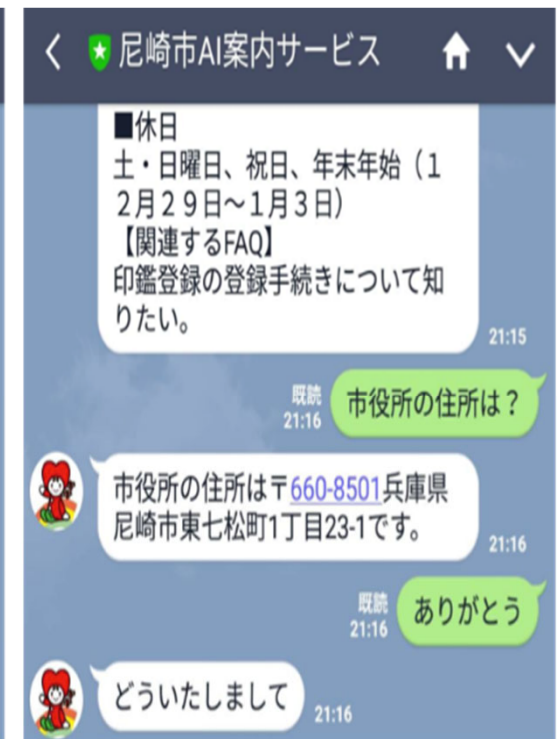
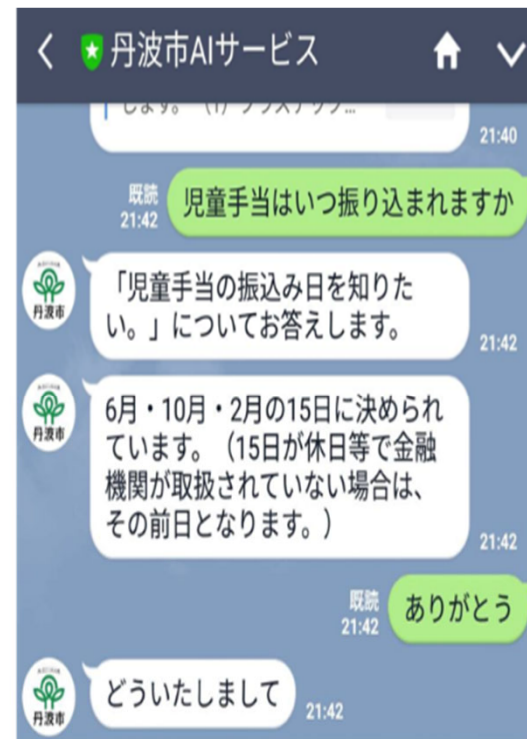
- 兵庫県、尼崎市、丹波市、京都大学大学院情報学研究科、国立情報学研究所ロバストインテリジェンス・ソーシャルテクノロジー研究センター（CRIS: Center for Robust Intelligence and Social Technology）、LINE株式会社の6者は、2018年4月に**LINEを活用した社会課題解決手法の研究に関する連携協定**を締結
- 京都大学情報学研究科 黒橋研究室を中心にCRISの研究プロジェクトとして実施



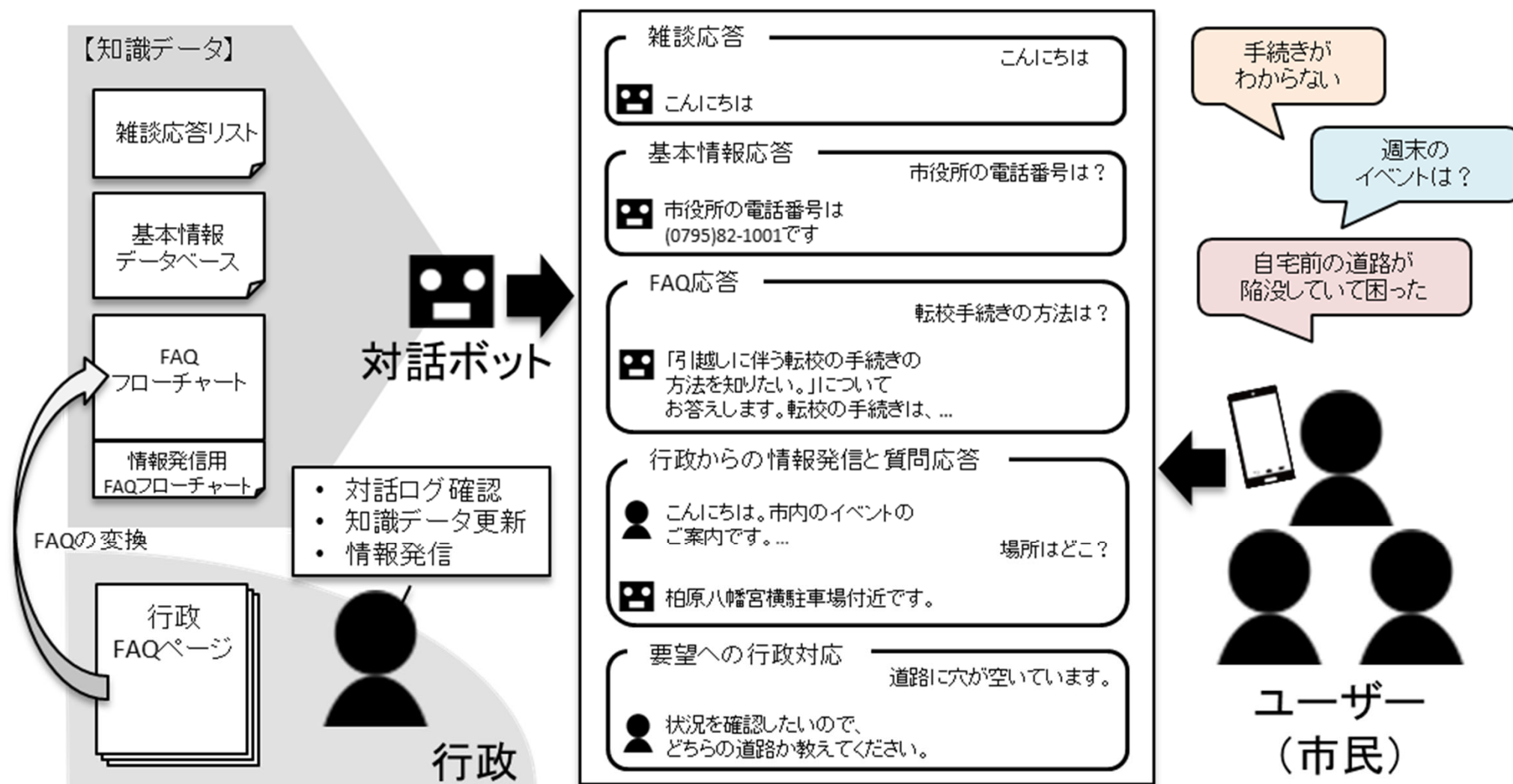
実施内容

- 行政サービスに関する対話システム「尼崎市AI案内サービス」「丹波市AIサービス」をLINE上で運用する実証実験
(実施期間:2018年6月～2021年3月)

- 目的:
 - 社会課題の解決に資する人工知能(AI)の開発
 - 対話分析による住民サービス向上
 - 地域活性化施策の策定支援



対話システムの機能の概要



FAQ検索の高度化

クエリ: ハチの巣の撤去はお願いできますか？

クエリとQの類似性

検索エンジンTSUBAKI
[Shinzato+ 08]

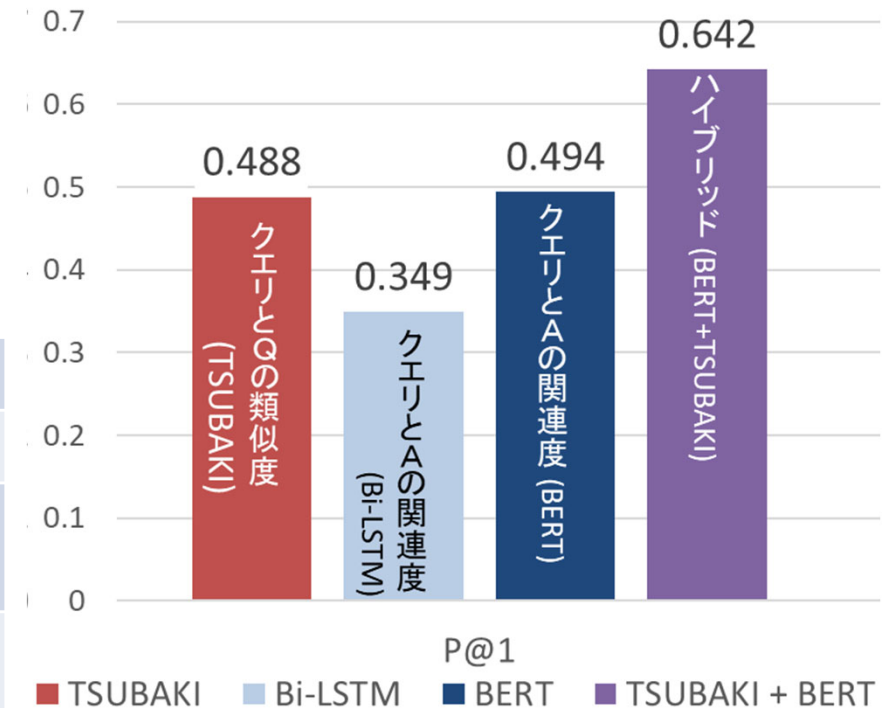
+

クエリとAの関連性

多数自治体のQAペアを収集,
BERTで学習(※クエリとQの大
規模学習データは存在しない)

FAQ

Q: 大型ごみの出し方を教えて	A: 生活環境センターに電話して...
Q: 粗大ごみの捨て方は	A: センターに電話してください。
Q: ハチの巣を駆除してほしい	A: 一般家庭の住宅や敷地内にハチの巣が...
Q: 郵送で税の証明書を取得したい	A: 証明書の郵送は...
...	...



行政現場からの反応

- 窓口では応じづらいニーズへの対応に一定の効果が見られた
 - 「24時間対応が可能になり、市民サービスの向上が図れた」
 - 「(市民から見て)どこに問い合わせたらよいか分からない内容や、AI案内サービスだから聞けること(人と上手くいっていない、いじめ等)などの質問の窓口になっている」

その後

- 各市のFAQの整備、行政ウェブページからのFAQ自動作成
- 尼崎市では2021年4月1日より「LINE AIサービス」を、丹波市では2022年4月より「丹波市AIチャットボット」をそれぞれリニューアルしてサービス継続

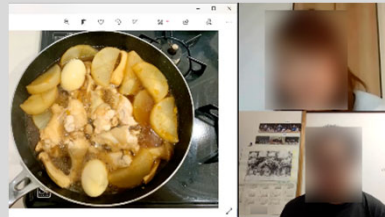
暗黙知表出インタビューのモデル化

目標: インタビュー対話の書き言葉変換と知識構造・マルチモーダル情報の自動解析の実現により, 熟練技能者から暗黙知を引き出すインタビュー支援システムを開発する

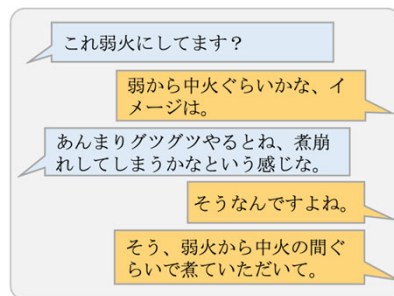
・ 料理技能者インタビューコーパスの構築 [Okahisa+ LREC2022]

コーパスの特徴

- ・ 10-15分の対話300件 (約7万発話)
- ・ マルチモーダル情報付き
- ・ 技能者は調理工程の写真を用意
- ・ インタビュアーはコツを引き出す

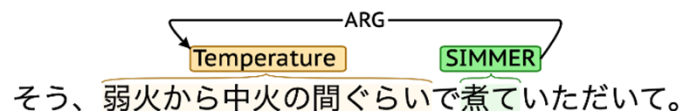


・ インタビュー対話の書き言葉変換

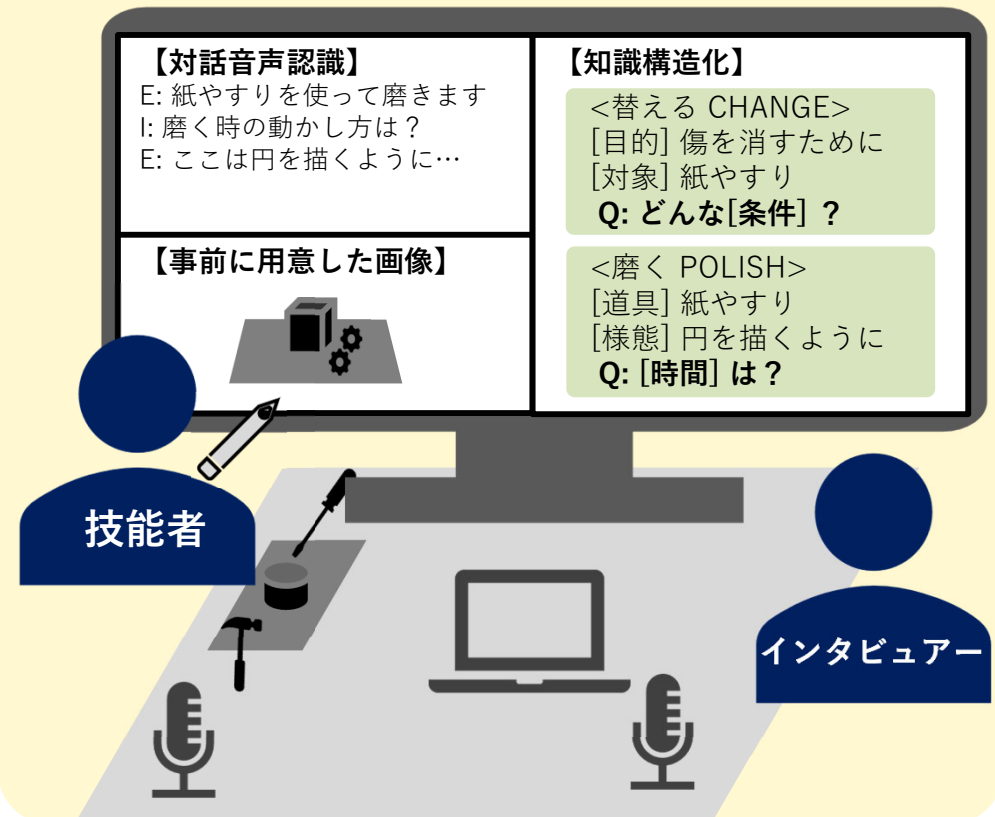


あんまりグツグツ煮込むと煮崩れてしまうので、弱火から中火の間ぐらいで煮ていきます。

・ 意味フレームに基づく知識構造の自動解析



インタビュー支援システム



これから

- データ駆動型社会／研究
- オープンサイエンス
- NLPは(かなり)昔からデータ駆動＋オープン
⇒ 学术界全体、産業界
- 知識基盤(分野横断による新たな価値、総合知)
- 汎用言語モデル・基盤モデルの発展

AI等の活用を推進する研究データエコシステム構築事業

研究データ基盤高度化チーム

NII Research Data Cloudを
7つの側面から機能拡張

NII

リーダ機関

研究データ基盤の機能実装

活用

コード付帯機能

データ・プログラム・解析環境の
パッケージ化と流通機能を提供し、
研究成果の再現性を飛躍的に向上

信頼

データプロビナンス機能

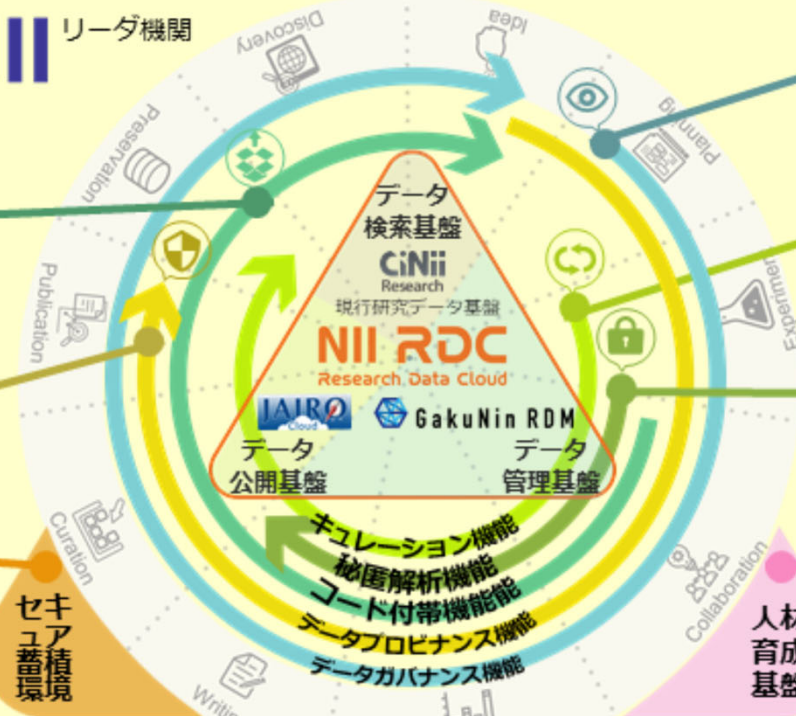
データの来歴情報の管理から利用
状況を把握でき、データ公開への
インセンティブモデルを提供

蓄積

セキュア蓄積環境

専用HWと高度な暗号化技術による
超鉄壁ストレージを提供し、
データの共有と保護を両立

セキュア蓄積環境



データガバナンス機能

管理

計画に基づきデータ管理等を機械
的に支援し、DMPをプロジェクト
管理に不可欠な仕組みへと変革

キュレーション機能

流通

専門的なキュレーションを実践
できるエコシステムを構築し、
データ再利用の促進に寄与

秘匿解析機能

保護

秘密計算技術で機微な情報も安心し
て解析できる環境の提供で、新しい
データ駆動型研究の世界を開拓

人材育成基盤

育成

RDMに必要なスキルを学ぶ環境
を提供し、全ての研究者を新しい
科学の実践者へと育成

プラットフォーム連携チーム

融合・活用開拓チーム

ルール・ガイドライン整備チーム

人材育成チーム

基盤の活用に係る環境整備



理化学研究所

リーダ機関

- ・ 機関内サービス等とNII RDC
の連携機能の整理と設計
- ・ 計測機器等からの大量データ
を効果的に管理するための要件
整理と機能開発
- ・ 管理対象となるメタデータの
設計と実証
- ・ 関連する高度化機能との仕様
調整と共同開発



東京大学
THE UNIVERSITY OF TOKYO

リーダ機関

- ・ 異なる分野間でのデータ活用
やデータ連携に発展する取り
組みを精査
- ・ 異なる分野間でのデータ活用
やデータ連携に関する具体的
なユースケースを創出
- ・ ユースケースをまとめたツール
キットの作成とそれを用いた
広報活動



名古屋大学
NAGOYA UNIVERSITY

リーダ機関

- ・ 研究データの活用に適した機
械可読データの統一的な記述
ルール設計
- ・ 研究データの公開に必要な要
項や作業フローの整備
- ・ 研究データを適切に取扱うた
めの指針のまとめ
- ・ 学内整備のための事例形成



大阪大学
OSAKA UNIVERSITY

リーダ機関

- ・ 人材育成を主とした研究デー
タ管理体制の構築を推し進め
る学内組織構築の事例形成
- ・ 研究データ管理人材に求めら
れる標準スキルに関する検討
- ・ 研究データ管理人材育成のた
めのカリキュラムの作成、オ
ンライン学習コースの整備

中核機関群の代表からなる運営委員会が全体を統括し研究データエコシステムの全国展開に向けて共同実施機関を随時拡大⁴¹

自己紹介の続き

2020.4～ 文部科学省科学官

(2022.4～研究データ基盤の将来構想に関わる連絡会 主査)

2021.4～ 京都大学図書館機構副機構長

(APCを一括とする転換契約などを担当)

2019.4～ デジタルアーカイブ学会 理事

(2021.4～ 産業とコンテンツ・データ部会長)

2023.4～ 国立情報学研究所(NII)所長

ありがとうございました
今後ともよろしくお願いいたします