

# 特許請求項自動後編集のための対訳並列要素の自動抽出・二言語間対応付け

石丸司<sup>1</sup> 宇津呂武仁<sup>1</sup> 永田昌明<sup>2</sup>

<sup>1</sup>筑波大学 <sup>2</sup>NTT

## ⚠ 特許翻訳の特徴と翻訳の課題



### 現在の特許翻訳（専門家への外注）

- ・ 翻訳費用が数十～数百万円程度と高額
- ・ やり取りに時間がかかる

- ✔ 専門性が高く翻訳料が高額
- ✔ 弁理士や翻訳者委託の手続き等に時間がかかる
- ✔ 特許文は専門用語・語数が多く一文が非常に長い
- ✔ 文構造が複雑で翻訳が難しい

上記構造が含まれ、翻訳が最も難しい項目の一つ  
 ➤ 特許請求項

**[!]** 特許請求項とは…特許出願書において、発明の範囲を定義する文章のこと

機械翻訳（MT）では翻訳に誤りが生じやすい

- ・ 一文が非常に長い（300～500文字/1項）。
- ・ 文構造が複雑。

➤ 複雑な **並列構造** が複数含まれることが多い。

## 並列構造とは？

文中で品詞的に対等な関係にある語句の並び。  
 翻訳時にこの関係が崩れると、  
 特許の権利範囲が変わってしまうなどのリスクがある。

例.

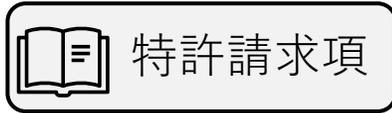
日：請求項 1 および請求項 2 の特徴を備えた機械。

英：A machine comprising the features of  
**claim 1 and claim 2.**

機械翻訳の誤り：訳抜け・繰り返し・並列関係のずれ等



LLMを用いて並列構造の翻訳誤りを検出・修正し、  
 より正確な特許請求項翻訳の実現を目指す。



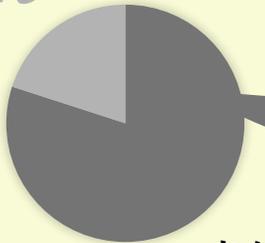
機械翻訳

並列構造が関係する割合

なし

翻訳誤りの割合

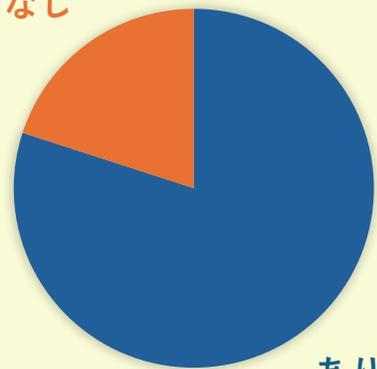
なし



あり

何かしらの翻訳誤りが生じる割合

約80%



あり

翻訳誤りの中で並列構造の不整合により生じている翻訳誤りの割合

約60%

つまり・・・

特許請求項を翻訳した場合、  
並列構造が原因で誤訳となる割合は、



80%(全体のエラー率)×60%(並列構造による割合)  
=約**48%**

並列構造の整合性を保持することでかなりの割合の翻訳誤りが解消される

## 原文（日本語特許請求項）：

請求項 3 に記載の動的経路変更型迷路アトラクションシステムであって、来場者が所持する端末装置と通信する通信手段と、前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段と、をさらに備えることを特徴とする動的経路変更型迷路アトラクションシステム。

## 正しい対訳特許請求項（英語）：

The dynamic route-changing maze attraction system according to claim 3, further comprising:  
a communication means for communicating with a terminal device possessed by a visitor; and a means for providing individualized information, services, or content through the communication means.

## 誤った機械翻訳文（英語）：

The system according to claim 3, further comprising a communication means to communicate with a terminal device, individualized information, services, or content, and a providing means.

## 原文（日本語特許請求項）：

請求項 3 に記載の動的経路変更型迷路アトラクションシステムであって、  
来場者が所持する端末装置と通信する通信手段と、  
前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段と、  
をさらに備えることを特徴とする動的経路変更型迷路アトラクションシステム。

## 正しい対訳特許請求項（英語）：

The dynamic route-changing maze attraction system according to claim 3, further comprising:  
a communication means for communicating with a terminal device possessed by a visitor; and a means for providing individualized information, services, or content through the communication means.

## 誤った機械翻訳文（英語）：

The system according to claim 3, further comprising a communication means to communicate with a terminal device, individualized information, services, or content, and a providing means.

## 原文（日本語特許請求項）：

請求項 3 に記載の動的経路変更型迷路アトラクションシステムであって、  
来場者が所持する端末装置と通信する通信手段と、  
前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段と、  
をさらに備えることを特徴とする動的経路変更型迷路アトラクションシステム。

## 正しい対訳特許請求項（英語）：

The dynamic route-changing maze attraction system according to claim 3, further comprising:

a communication means for communicating with a terminal device possessed by a visitor; and  
a means for providing individualized information, services, or content through the communication means.

## 誤った機械翻訳文（英語）：

The system according to claim 3, further comprising a communication means to communicate with a terminal device, individualized information, services, or content, and a providing means.

## 原文（日本語特許請求項）：

請求項 3 に記載の動的経路変更型迷路アトラクションシステムであって、  
来場者が所持する端末装置と通信する通信手段と、  
前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段と、  
をさらに備えることを特徴とする動的経路変更型迷路アトラクションシステム。

## 正しい対訳特許請求項（英語）：

The dynamic route-changing maze attraction system according to claim 3, further comprising:

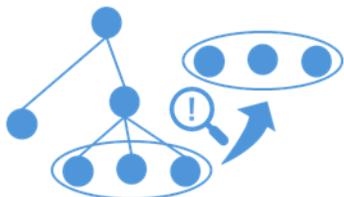
a communication means for communicating with a terminal device possessed by a visitor; and  
a means for providing individualized information, services, or content through the communication means.

## 誤った機械翻訳文（英語）：

The system according to claim 3, further comprising a communication means to communicate with a

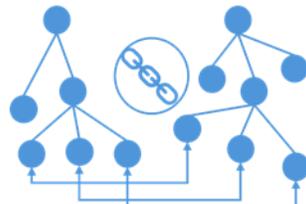
terminal device, individualized information, services, or content, and a providing means.

独立した3つのモジュールを統合し、自動後編集を行う



## ① 並列構造抽出 LLM

日本語原文・英語翻訳文から  
それぞれの並列構造を  
タグ付きで抽出



## ② 対応付け LLM

原文と翻訳文の  
並列構造を照合し  
不整合（ズレ）を検出



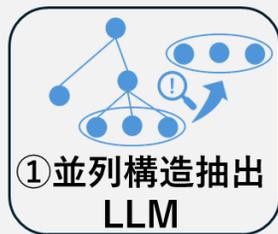
## ③ 誤り訂正 LLM

不整合の情報を基に  
翻訳文を原文と同様の  
並列構造を持つように修正

提案システムは、あくまで翻訳前後の並列構造整合性に着目したシステムであり、すべての翻訳誤りを修正できるものではない。

**原文：**

請求項 3 に記載の動的経路変更型迷路アトラクションシステムであって、来場者が所持する端末装置と通信する通信手段と、前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段と、をさらに備えることを特徴とする動的経路変更型迷路アトラクションシステム。



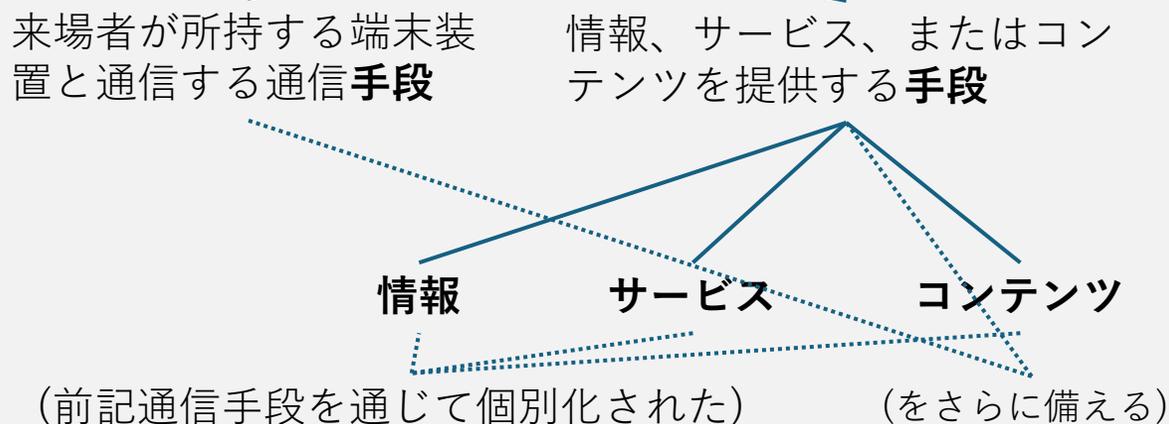
**機械翻訳文：**

The system according to claim 3, further comprising a communication means to communicate with a terminal device, individualized information, services, or content, and a providing means.

**抽出：**



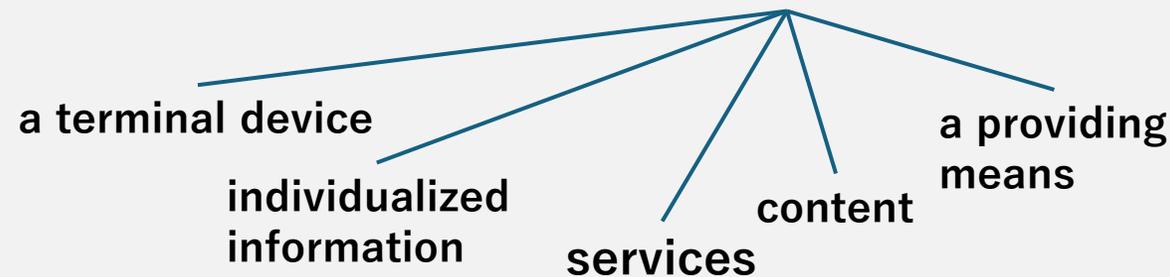
請求項 3 に…あって、**S** と、をさらに…システム。



**抽出：**



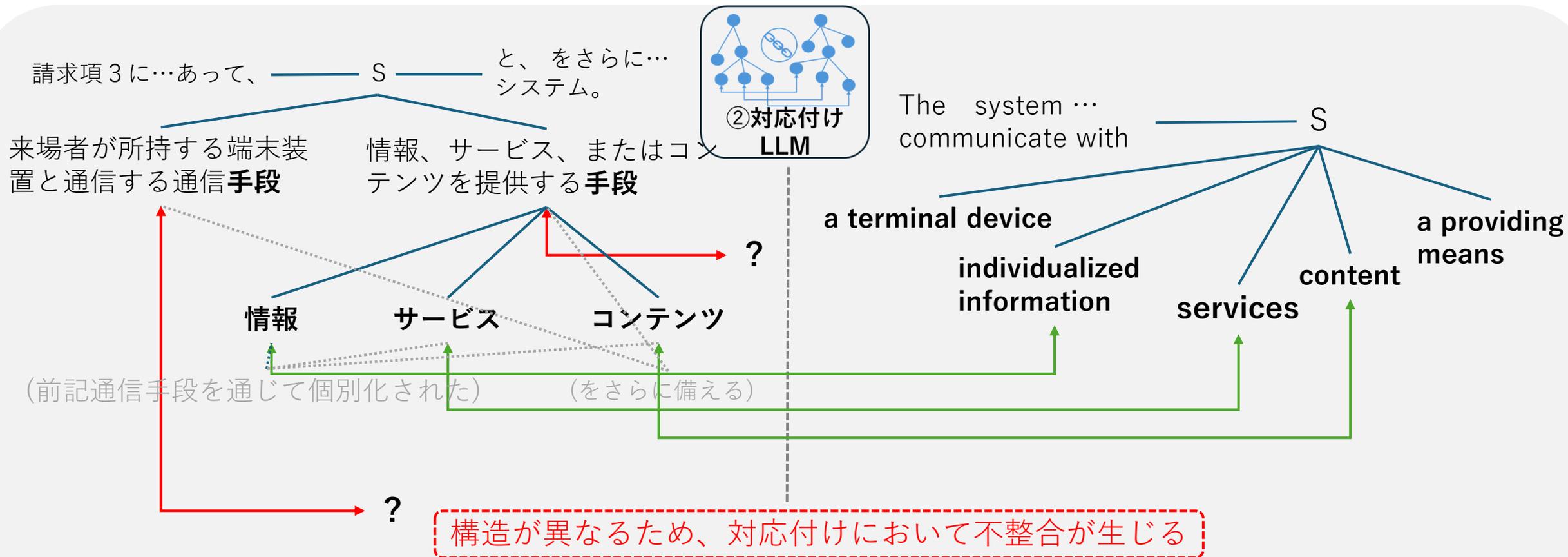
The system…communicate with **S**



原文の文構造とは異なり、一つの平坦な並列構造のみになっている



LLMによる原文・翻訳文からの並列構造抽出



**[!]誤り情報：**  
 「来場者が所持する端末装置と通信する通信手段」と「前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段」に該当する並列句が翻訳文の並列構造内にありません。

## 原文：

請求項 3 に記載の動的経路変更型迷路アトラクションシステムであって、 来場者が所持する端末装置と通信する通信手段と、 前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段と、 をさらに備えることを特徴とする動的経路変更型迷路アトラクションシステム。

## 機械翻訳文：

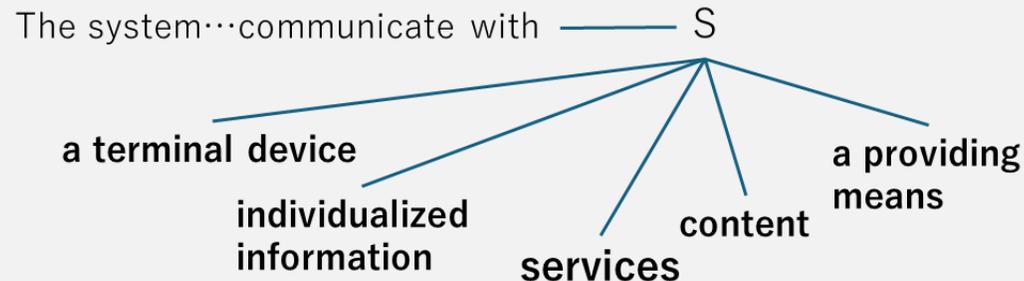
The system according to claim 3, further comprising a communication means to communicate with a terminal device, individualized information, services, or content, and a providing means.

### [!] 誤り情報：

「来場者が所持する端末装置と通信する通信手段」と「前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段」に該当する並列句が翻訳文の並列構造内にありません。



③ 誤り訂正



LLMによる修正



**The dynamic route-changing maze attraction system according to claim 3, further comprising:  
 a communication means for communicating with a terminal device possessed by a visitor;  
 and a means for providing individualized information, services, or content through the communication means.**

## 原文：

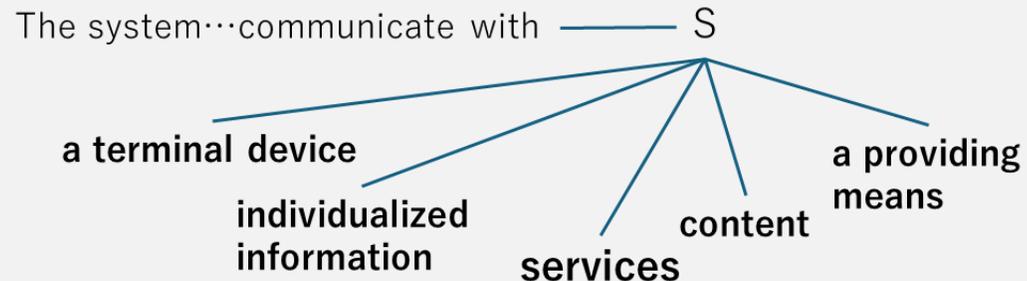
請求項 3 に記載の動的経路変更型迷路アトラクションシステムであって、来場者が所持する端末装置と通信する通信手段と、前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段と、をさらに備えることを特徴とする動的経路変更型迷路アトラクションシステム。

## 機械翻訳文：

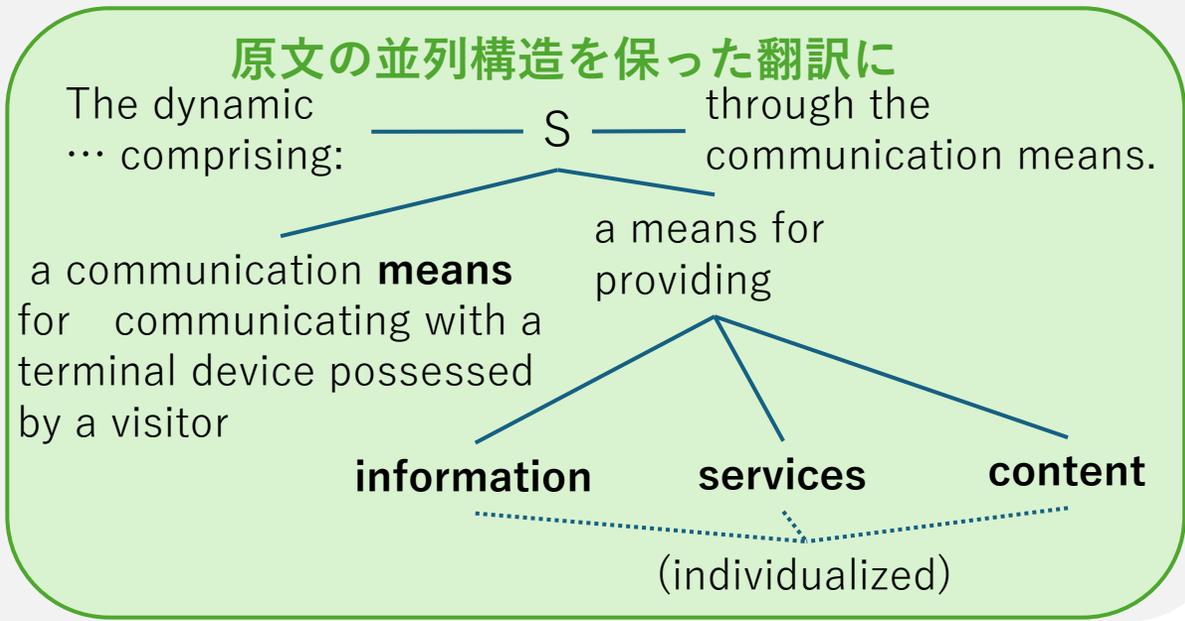
The system according to claim 3, further comprising a communication means to communicate with a terminal device, individualized information, services, or content, and a providing means.

### [!] 誤り情報：

「来場者が所持する端末装置と通信する通信手段」と「前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段」に該当する並列句が翻訳文の並列構造内にありません。



LLMによる修正



✓ LLMによる翻訳文の並列構造修正

### 提案システム全体における新規性

- ・ 翻訳前後の並列構造整合性に着目した翻訳誤り検出・修正  
 関連研究：翻訳文全体における誤り個所の検出・修正  
 ➤ 特許請求項特有の文構造に特化しているわけではない

### 各モジュールにおける新規性

#### 1. 並列構造抽出

- ・ TANLフレームワークを用いた一段階のデコーダモデル (LLM) による並列構造抽出  
 関連研究：エンコーダモデルを用いた2段階のシステムで抽出(Wang et al., 2023)  
 ➤ 2段階のシステムが独立していることによる相互依存関係の欠如・訓練の煩雑さ

#### 2. 並列構造対応付け

- ・ デコーダモデル(LLM)を用いるスパン対応付け  
 関連研究：エンコーダモデルを用いたスパン対応付け・LLMによる単語対応付け  
 (Chousa et al., 2019) (Miao et al., 2025)  
 ➤ 単語対応ではエンコーダモデルを超える精度をLLMが達成、スパン対応に応用

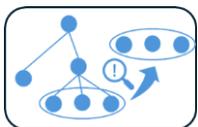
#### 3. 翻訳後編集

- ・ 並列構造整合性に基づく誤り情報による翻訳後編集  
 関連研究：翻訳文の訳抜け・繰り返しによる誤り情報を基に後編集  
 ➤ 並列構造に着目することでより特許請求項の翻訳後編集に特化

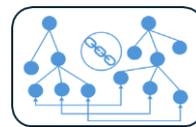
独立した3つのモジュールを統合し、自動後編集を行う



今回の発表では、上記2モジュールの実装のみを行った実験結果を示す。



## ① 並列構造抽出



## ② 並列構造対応付け

評価データ：日英対訳特許請求項100文（JaParaPat\_2020年）  
 ※機械翻訳（MT）の誤りを含むデータでは評価していない

評価指標：各文に含まれる並列句単位で**F1スコア**を計算

評価対象モデル：

- ・ クローズドLLM（GPT-5 API呼び出し）
- ・ 訓練したオープンソースLLM（自前訓練）

・ 非LLM（BERT）ベース先行手法（英語のみ）

・ 非LLM（BERT）ベース単語対応モデル

LLM訓練データ：  
 Treebank形式のデータからルールベース  
 で抽出し作成

LLM訓練データ：  
 対訳特許請求項から手作業で作成

**■概要：**

- ・ NLPタスクを「拡張自然言語間の翻訳」として統一するフレームワーク
- ・ 構造情報をタグとして自然言語内に埋め込む

**■基本アイデア**

入力：

Barack Obama was born in Hawaii.

出力：

[Barack Obama | person] was born in [Hawaii | location].

➤すべてのタスクをtext-to-textに統一

**■特徴・利点**

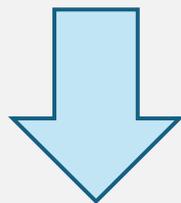
- ・ タスクへの依存性が低い
- ・ LLMの生成能力と高い依存性
- ・ モデル設計が単純で再現性が高い
- ・ 多言語に拡張が容易

**■本研究との関係**

- ・ 並列構造をタグとして表現
- ・ 日英間の対応付けをタグとして表現

**原文：**

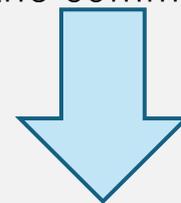
請求項 3 に記載の動的経路変更型迷路アトラクションシステムであって、来場者が所持する端末装置と通信する通信手段と、前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段と、をさらに備えることを特徴とする動的経路変更型迷路アトラクションシステム。

**並列構造抽出後：**

請求項 3 に記載の動的経路変更型迷路アトラクションシステムであって、**<1\_NP>**来場者が所持する端末装置と通信する通信手段**</1>**と、**<1\_NP>**前記通信手段を通じて個別化された**<2\_NP>**情報**</2>**、**<2\_NP>**サービス**</2>**、または**<2\_NP>**コンテンツ**</2>**を提供する手段**</1>**と、をさらに備えることを特徴とする動的経路変更型迷路アトラクションシステム。

**対訳文：**

The dynamic route-changing maze attraction system according to claim 3, further comprising: a communication means for communicating with a terminal device possessed by a visitor; and a means for providing individualized information, services, or content through the communication means.

**並列構造抽出後：**

The dynamic route-changing maze attraction system according to claim 3, further comprising: **<1\_NP>** a communication means for communicating with a terminal device possessed by a visitor **</1>**; and **<1\_NP>** a means for providing individualized **<2\_NP>** information **</2>**, **<2\_NP>** services **</2>**, or **<2\_NP>** content **</2>** through the communication means **</1>**.

LLMを用いてTANLを適用した並列構造抽出



LLMによる原文・翻訳文からの並列構造抽出

**原文：**

請求項 3 に記載の動的経路変更型迷路アトラクションシステムであって、**<1\_NP>来場者が所持する端末装置と通信する通信手段</1>**と、前記通信手段を通じて個別化された情報、サービス、またはコンテンツを提供する手段と、をさらに備えることを特徴とする動的経路変更型迷路アトラクションシステム。



**対訳文：**

The dynamic route-changing maze attraction system according to claim 3, further comprising: **<1\_NP>a communication means for communicating with a terminal device possessed by a visitor</1>**; and a means for providing individualized information, services, or content through the communication means.

**LLMを用いてTANLを適用した並列構造対応付け**

- 
- 
- 
- 
- 

- 
- 
- 
- 
- 

**原文：**

請求項 3 に記載の動的経路変更型迷路アトラクションシステムであって、来場者が所持する端末装置と通信する通信手段と、前記通信手段を通じて個別化された情報、サービス、または**<2\_NP>コンテンツ</2>**を提供する手段と、をさらに備えることを特徴とする動的経路変更型迷路アトラクションシステム。



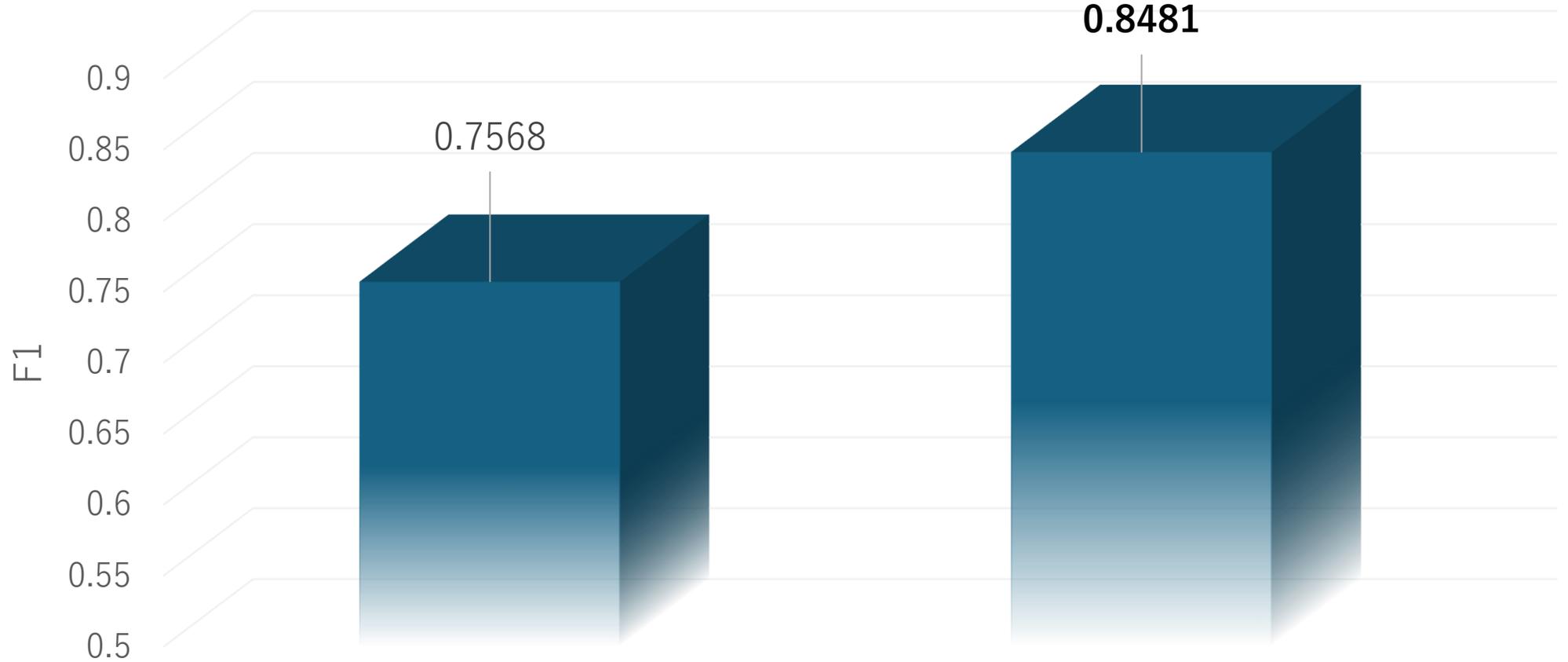
**対訳文：**

The dynamic route-changing maze attraction system according to claim 3, further comprising: a communication means for communicating with a terminal device possessed by a visitor; and a means for providing individualized information, services, or **<2\_NP> content </2>** through the communication means.



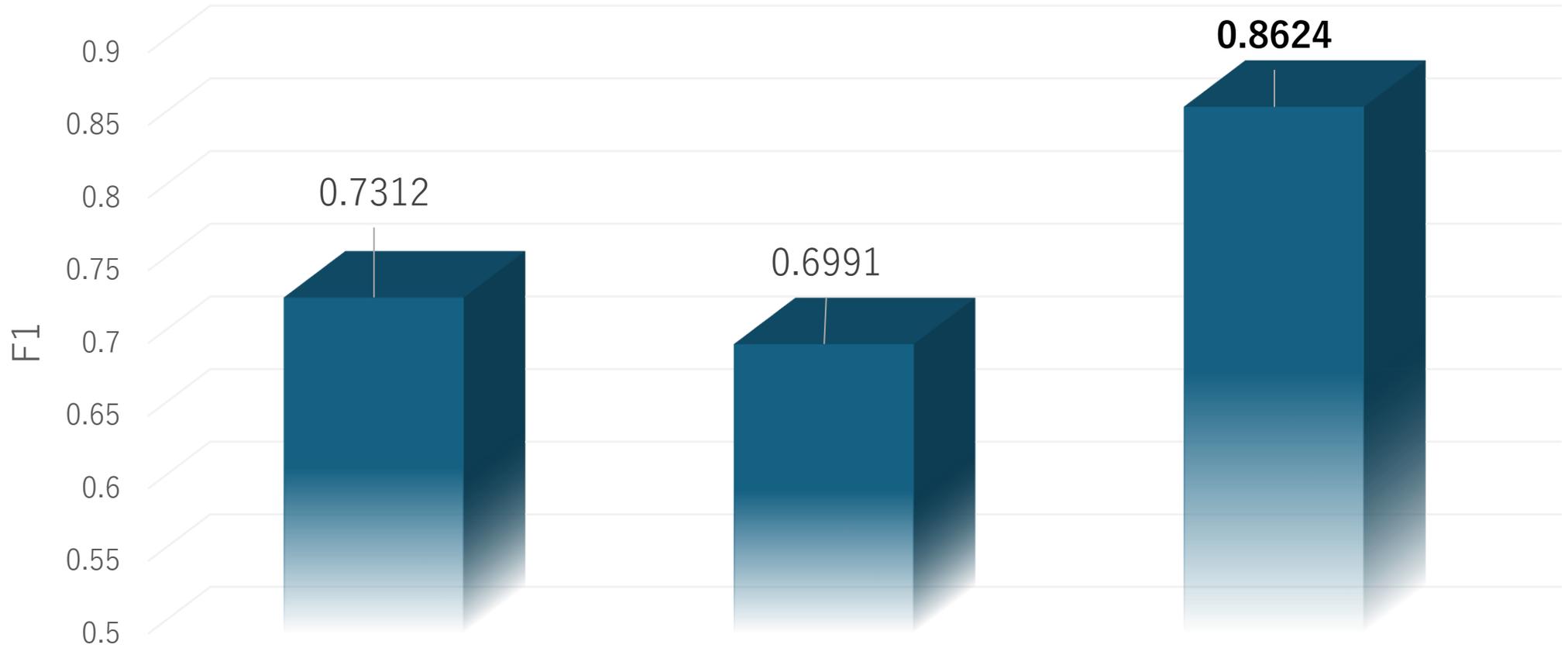
LLMによる原文・対訳文の並列構造対応付け

日本語抽出精度(F値)



提案手法

英語抽出精度(F値)



非LLM(BERT)  
ベース  
先行手法\_2023  
(CoRec)

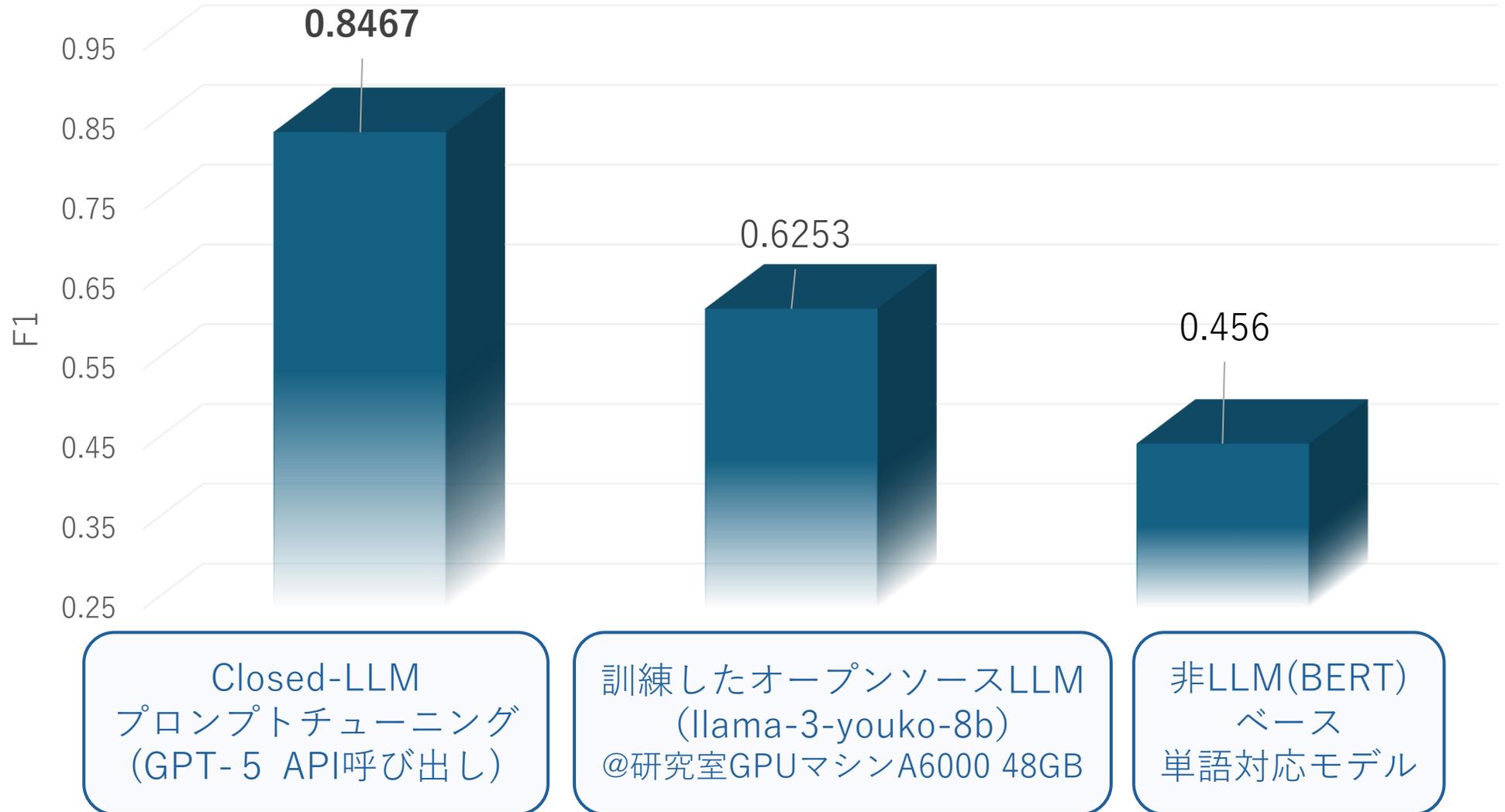
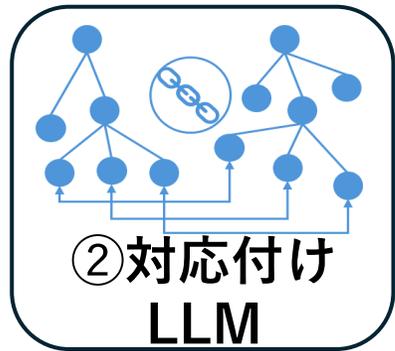
ベースライン

Closed-LLM  
プロンプトチューニング  
(GPT-5 API呼び出し)

提案手法

訓練したオープンソースLLM  
(gemma-3-4b-it)  
@研究室GPUマシンA6000 48GB

日英対応付け精度(並列構造抽出から対応付け)



提案手法

### 1. 並列構造抽出

**結果:** 訓練したオープンソースLLMが比較手法の中で最も高いスコアを記録。

**結論:** 特許文特有の複雑な構文解析において、提案手法の有効性が実証された。

### 2. 並列構造対応付け

**結果:** GPT-5による対応付け手法が提案手法を抑え最高精度を達成。

**考察:** 強力なLLMの文脈理解能力が、二言語間の構造的な一致を捉えるのに寄与した。

### 3. 今後の課題

**現状:** 現時点の訓練したオープンソースLLMの対応付け精度は、実用化レベルには至っていない。

**考察:** 精度を上げるには、疑似データ作成による訓練データ拡張、またはLLMのプロンプトエンジニアリングの最適化が必要。

## 今後の展望

独立した各モジュールを統合し、実用的なパイプラインを確立する

 並列構造抽出部(Extract) 

 対応付け部(Alignment) 

 翻訳後編集(Post-edit)



 並列構造の二言語間整合性に着目した特許請求項翻訳自動後編集