

# 項の復元は必要か？

## 日英機械翻訳における省略された項の扱いの分析

野末 慎之介<sup>1</sup>, 松林 優一郎<sup>1,2</sup>, 藤井 諒<sup>3,1</sup>, 岸波 洋介<sup>3</sup>, 森下 睦<sup>3,1</sup>, 坂口 慶祐<sup>1,2</sup>

<sup>1</sup>東北大学 <sup>2</sup>理化学研究所 <sup>3</sup>フューチャー株式会社

# 背景：日英翻訳では省略項の読解・復元が性能に影響を及ぼす？

- 日本語では主語や目的語などの述語の項がしばしば省略される
- 省略された項の理解と、その項を表出させて（復元して）翻訳することが重要  
[1, 2, 3, 4]

ある男性Aさんは三年前に休暇でバンコクに行き、そこで美しいタイ人女性の、Bさんに一目ぼれ。  
ドイツに (φが) (φを) 連れ帰って 甘い結婚生活を始めた。



「誰が、誰を、連れ帰った」のか、  
表出させて（復元して）翻訳する必要

Three years ago, a man named A traveled to Bangkok on vacation, where he fell in love with a beautiful Thai woman named B.  
He brought her back to Germany, and they began a blissful married life together.

# 背景：日英翻訳では省略項の読解・復元が性能に影響を及ぼす？

- 項の認識や復元が翻訳に影響しないことが報告されている [5,6,7]
- 項の復元がなくとも翻訳が成り立つような例も考えられる
  - 2の英訳例では、受動態を使用することで明示的な項を復元せずに翻訳できている

委員会が新計画を審査。協議の末、その計画を (φが) 承認した。

1. The committee reviewed a new plan. After deliberation, **it approved** the plan.
2. The committee reviewed a new plan. After deliberation, the plan **was approved**.

省略された項の正確な読解や復元は  
適切な翻訳をする上で必須なのか？

# 本研究の概要

大規模言語モデルの日英翻訳における省略項の同定・復元の影響と重要性の再検討  
同一データセットでゼロ照応解析と日英翻訳を実施 → 解析成否で比較

タスク実施

日本語  
項省略事例  
サブセット

~~~~~  
~~。~~~Aさん。  
-(*φ*ガ) 倒れて----  
-----。

文間ゼロ照応事例



GPT-4o

ゼロ照応  
解析

解析成功

Aさん



解析失敗

Bさん



英訳

比較・分析

翻訳性能

ゼロ照応解析成功

ゼロ照応解析失敗

解析失敗事例では  
性能が下がる？

省略項の復元

項の復元率  
は違う？

項を復元しない時  
どう訳す？

# 本研究の詳細な観点と結果のまとめ

- 翻訳の文意伝達性
  - 解析成功事例で翻訳成功率**94%**、解析失敗事例で**98%**と、解析に失敗しても文意伝達は可能
  - 解析失敗事例では翻訳エラーの原因が全て省略に関係
- 項の復元傾向
  - 解析成功事例で**60.6%**、解析失敗事例で**37.5%**と、解析に失敗すると項の復元を避ける傾向
  - 主節の述語は項が復元される傾向が強いものの、**30%以上**は復元されず
  - 補完された項は翻訳失敗事例の**3件**（解析成功**1件**, 解析失敗**2件**）を除き、全て文脈的に妥当
  - 復元された項の代名詞率は解析成功事例の方が若干高い
- 英語の構文構造
  - 使用された構文構造に、解析の成否による傾向の違いはなし
- 省略されたままの項は翻訳から推論可能か？
  - 省略の成否に限らず、省略されたまま項を英訳文のみから推測可能

**NLP2026の予稿 (C9-20)  
をぜひご覧ください！**

# 本研究における事例の定義

- 本研究で扱う事例：文間ゼロ照応事例
  - 文間ゼロ照応：省略された項が前方文脈に存在する事例
  - 対象述語, 対象の格助詞 ( $c \in \{\text{が}, \text{を}, \text{に}\}$ ) , 対象文, 先行文脈 (最大3文) , 正解の先行詞クラスター
    - 対象文：対象述語が含まれる文
  - NAIST Text Corpus 1.5[8] (日本語新聞記事コーパス) から抽出
- 実際の例

|             |                                        |
|-------------|----------------------------------------|
| 対象述語        | 受賞し (原形：受賞する)                          |
| 対象の格        | が (ガ格)                                 |
| 先行文脈 (最大3文) | 雄大な北アルプスを望む長野県梓川村は、梓川の清流ではぐくまれたリンゴの産地。 |
| 対象文         | 1984年には天皇賞を <u>受賞し</u> 、その味には定評がある。    |
| 正解先行詞クラスター  | 長野県梓川村, 産地, リンゴ                        |

# ゼロ照応解析：手法概要



- ゼロ照応解析タスクで省略項の理解を評価
  - 入力の記事に基づき、指定された述語の指定された格助詞（格）の項を同定

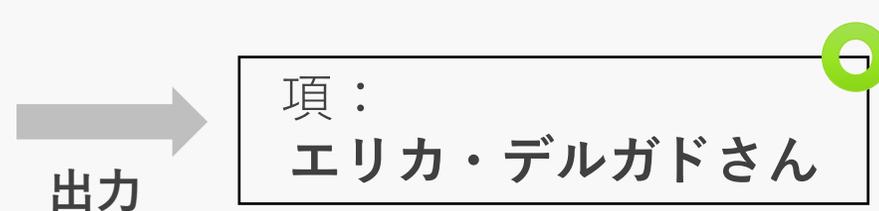
## 文章

乗客の九歳の少女一人が救助されただけで、... 救助されたのはエリカ・デルカドさん。  
... 沼地で、気を失って倒れているのを発見された。

## 正解先行詞クラスタ

少女, エリカ・デルガドさん

{文章 (先行文脈 + 対象文)}  
{タスク指示}  
対象述語：倒れて  
対象の格：が (主語)



照合・正誤判定

# ゼロ照応解析：評価の詳細



## ● 設定

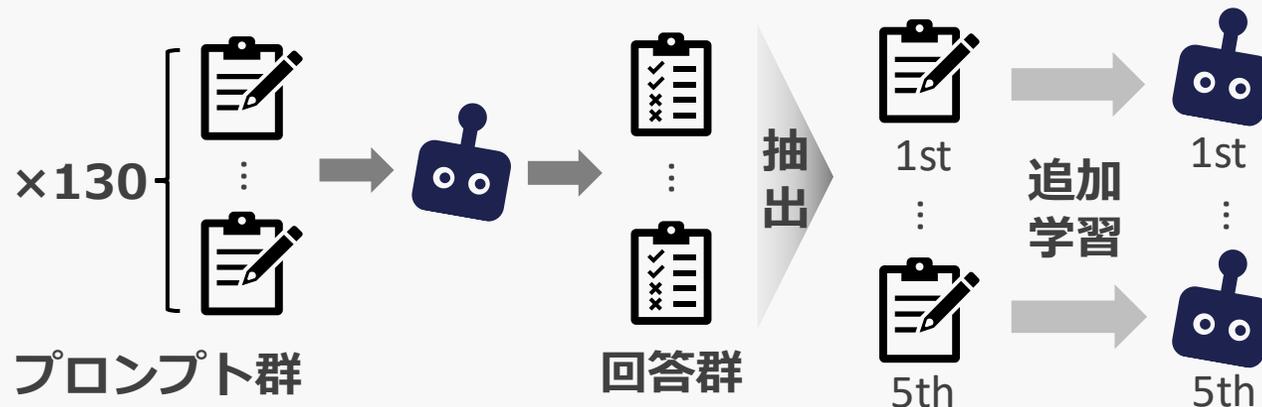
- gpt-4o-2024-08-06 (temperature: 0)
- 評価指標

- 成否 (2値) : 5つのモデルの**過半数 (3つ以上)** で正解するか否か

## ● タスクの理解度や出力の揺れを抑制

1. プロンプト大量作成 → Top 5抽出 → 少量追加学習

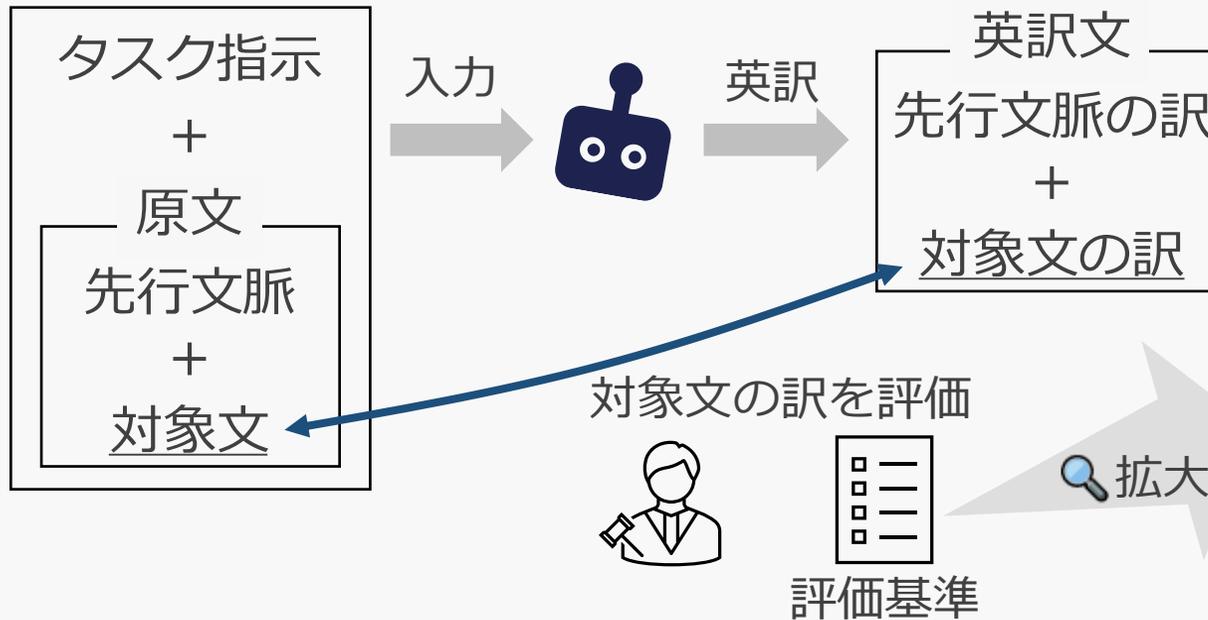
2. 過半数 (3つ以上) のモデルが正解で成功



# 日英翻訳：文意伝達性の評価



- GPT-4oを用いて日英翻訳を行い、文意伝達性指標<sub>[9,10]</sub>に基づいて人手評価
  - モデル：gpt-4o-2024-08-06 (temperature: 0)
  - 評価指標：成否



## 評価基準（文意伝達性<sub>[9,10]</sub>）

- 7: 参照文と文意が同一と考えて差し支えない
- 6: 参照文と文意に僅かな違いがあるが、ほぼ誤解の心配はない
- 5: 参照文と文意に若干の齟齬はあるが、大きな誤解を招くほどではない
- 4: 参照文と矛盾とまではいかないが重要な情報の誤りや過不足があり文意の重大な誤解が起こり得る
- ...

# 結果：日英翻訳の文意伝達性

- ゼロ照応解析の成否により日英翻訳の文意伝達性に大きな差はない

|          | 翻訳成功 | 全事例数 |
|----------|------|------|
| ゼロ照応解析成功 | 94   | 100  |
| ゼロ照応解析失敗 | 98   | 100  |

省略項の誤訳が原因の  
翻訳エラー

- 解析成功：1/6 件
- 解析失敗：2/2 件

省略された項が同定できるか否かに関わらず  
概ね誤解の生じない日英翻訳が可能

# 日英翻訳：項の復元の評価



対象述語の対象の格が復元されているかを判定

ある男性Aさんは三年前に休暇でバンコクに行き、そこで美しいタイ人女性の、Bさんに一目ぼれ。  
ドイツに（φが）連れ帰って甘い結婚生活を始めた。

対象述語

↓  
主語の「Aさん」が省略されている

Three years ago, a man named A traveled to Bangkok on vacation, where he fell in love with a beautiful Thai woman named B.  
**He** brought her back to Germany, and they began a blissful married life together.

③ 項が文脈的に妥当か判定

① 対象述語（連れ帰って）に対応する述語を特定

② 対象の格（今回は主語）に該当する項を特定

## 結果：項の復元率

- ゼロ照応解析成功事例では復元率が高いものの…

|          | 復元事例数 | 全事例数 | 復元率 (%) |
|----------|-------|------|---------|
| ゼロ照応解析成功 | 60    | 99   | 60.6*   |
| ゼロ照応解析失敗 | 33    | 88   | 37.5*   |
| 全体       | 93    | 187  | 49.7    |

\*: 1%水準で有意差  
(カイ二乗検定)

省略項が同定できる事例でも  
40%程度の事例で復元されずに翻訳

## 結果：項の復元率

- ゼロ照応解析失敗事例では復元率が下がる

|          | 復元事例数 | 全事例数 | 復元率 (%) |
|----------|-------|------|---------|
| ゼロ照応解析成功 | 60    | 99   | 60.6*   |
| ゼロ照応解析失敗 | 33    | 88   | 37.5*   |
| 全体       | 93    | 187  | 49.7    |

\*: 1%水準で有意差  
(カイ二乗検定)

省略項の同定が困難な事例では  
項の復元を避けるも翻訳できている

# 結果：項を復元せずにどのように翻訳しているか？

- 項が復元されない事例では述語の名詞化や受動態などを使用
  - 言語学で非明示的な項（Implicit Argument, PRO等）を許容する構文として議論 [11,12,13]

|          |         | ゼロ照応解析    |           |    |
|----------|---------|-----------|-----------|----|
|          |         | 成功        | 失敗        | 全体 |
| 日英<br>翻訳 | 名詞・動名詞  | 12 (31.6) | 14 (25.5) | 26 |
|          | 受動態     | 11 (28.9) | 15 (27.3) | 26 |
|          | 不定詞     | 5 (13.2)  | 11 (20.0) | 16 |
|          | 原文と対応なし | 5 (13.2)  | 8 (14.5)  | 13 |
|          | 分詞      | 3 (7.9)   | 4 (7.3)   | 7  |
|          | その他     | 2 (5.3)   | 3 (5.4)   | 5  |
|          | 全体      | 38        | 55        | 93 |

大きな傾向の  
差はなし

# 本研究の意義と今後の展望

- 省略された項の同定・復元が翻訳に必須ではないことを最先端のLLMにおいて検証（現在地の特定）
- 今後は「理想の翻訳における省略の扱い」を探求
  - 意味は読み取れる。では、読みやすい文章にはなっている？
  - 人間の書く英語においてどれくらい明示しない項を含むのか、その条件は？
  - 読みやすく、意味の伝わる理想の文では項がどれくらい明示されるのか？

## 参考文献 / References

1. Taku Kudo, Hiroshi Ichikawa, and Hideto Kazawa. 2014. [A joint inference of deep case analysis and zero subject generation for Japanese-to-English statistical machine translation](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 557–562.
2. Ryuichiro Kimura, Shohei Iida, Hongyi Cui, Po-Hsuan Hung, Takehito Utsuro, and Masaaki Nagata. 2019. [Selecting Informative Context Sentence by Forced Back-Translation](#). In *Proceedings of Machine Translation Summit XVII: Research Track*, pages 162–171.
3. Masaaki Nagata and Makoto Morishita. 2020. [A Test Set for Discourse Translation from Japanese to English](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3704–3709.
4. Longyue Wang, Siyou Liu, Mingzhou Xu, Linfeng Song, Shuming Shi, and Zhaopeng Tu. 2023. [A Survey on Zero Pronoun Translation](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3325–3339.
5. Hirotoshi Taira, Katsuhito Sudoh, and Masaaki Nagata. 2012. [Zero Pronoun Resolution can Improve the Quality of J-E Translation](#). In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 111–118.
6. Sho Shimazu, Sho Takase, Toshiaki Nakazawa, and Naoaki Okazaki. 2020. [Evaluation Dataset for Zero Pronoun in Japanese to English Translation](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3630–3634.
7. Ryokan Ri, Toshiaki Nakazawa, and Yoshimasa Tsuruoka. 2021. [Zero-pronoun Data Augmentation for Japanese-to-English Translation](#). In *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, pages 117–123.

## 参考文献 / References

8. Ryu Iida, Mamoru Komachi, Naoya Inoue, Kentaro Inui, and Yuji Matsumoto. NAIST Text Corpus: Annotating Predicate-Argument and Coreference Relations in Japanese, pp. 1177–1196. Springer Netherlands, Dordrecht, 2017.
9. Katsuhito Sudoh, Kosuke Takahashi, and Satoshi Nakamura. 2021. [Is This Translation Error Critical?: Classification-Based Human and Automatic Machine Translation Evaluation Focusing on Critical Errors](#). In *Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval)*, pages 46–55.
10. 樽本 空宙, 畠垣 光希, 宮田 莉奈, 梶原 智之, 二宮 崇, ChatGPT の日本語生成能力の評価, 自然言語処理, 2024, 31 巻, 2 号, p. 349-373.
11. Roeper, T. (1987). Implicit Arguments and the Head-Complement Relation. *Linguistic Inquiry*, 18(2), 267–310. <http://www.jstor.org/stable/4178538>
12. Williams, E. (1985). PRO and Subject of NP. *Natural Language & Linguistic Theory*, 3(3), 297–315. <http://www.jstor.org/stable/4047531>
13. Rajesh Bhatt and Roumyana Pancheva. (2007). Chapter 34. Implicit Arguments.