

# 日本語を含む言語対における 制約付き翻訳タスクの近況

東北大学 情報科学研究科

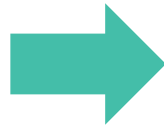
乾・坂口・徳久研究室 博士3年

阿部 香央莉

# 背景：特定の用語を適切に翻訳したい！

## 1. 専門用語の翻訳（科学論文，特許文書...）

*We have conducted a **shared task** focusing on terminology **consistency** from 2021.*

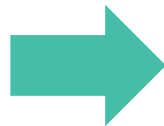


(System G) 2021年から用語の**一貫性**に重点を置いた**共有タスク**を実施しました。  
(System D) 2021年からの用語の**統一**に焦点を当てた**共有タスク**を実施しました。  
(System M) 2021年から用語の**一貫性**に焦点を当てた**共同作業**を行った。

🤔 「"shared task"は共有タスク、"consistency"は一貫性と訳したいなあ...」

## 2. 特定の固有名詞の翻訳（新商品の名前...）

ポケモン新作のパートナーには**ニャオハ**を選びました。

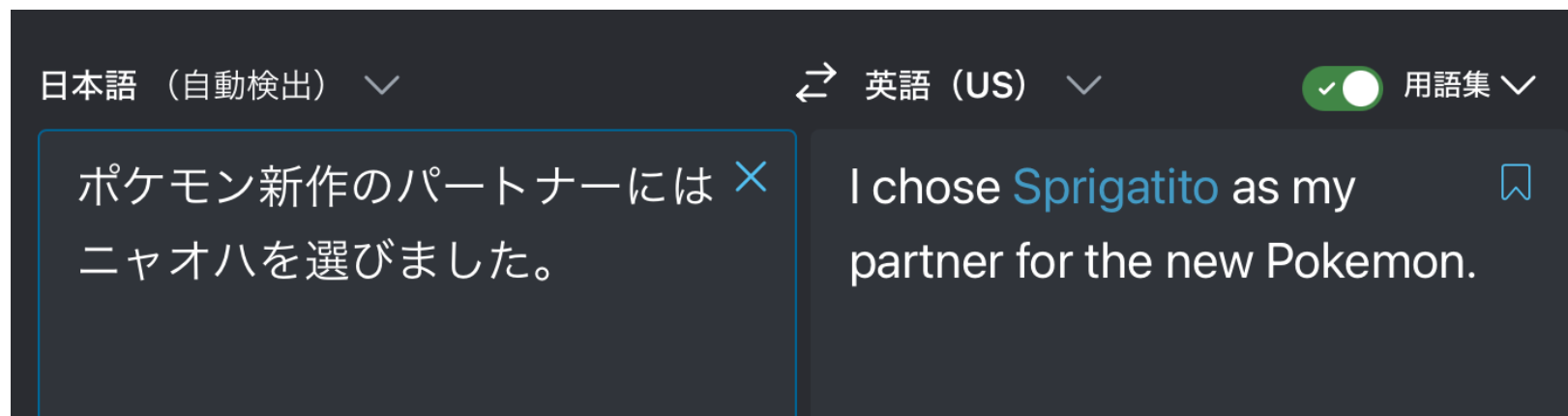


(System G) I chose **Nyaoha** as my partner for the new Pokemon game.  
(System D) I chose **Meowha** as my partner for the new Pokemon.  
(System M) I chose **Nyaoha** as my new Pokemon partner.

🤔 「"ニャオハ"は英語名だと"Sprigatito"だから、英語圏の人には伝わらないかも...」

# 背景：特定の用語を適切に翻訳したい！ → 実はできる

- System D, System Mには「用語集」機能あり



System D



本講演は、この内部で動いている（であろう）仕組みの紹介になると思います

System M

# 背景：研究分野における「用語エラー」への着目

- “用語（Terminology）”：専門用語・特定の固有名詞など
  - 2021年の機械翻訳シェアードタスクで提出されたシステムでも、「用語」に関するエラーが一定数出現 [Freitag+’21]

Error Categories	Errors (%)	Major (%)	Human MQM	All MT		Tohoku		OPPO		eTrans	
				MQM	vs H.	MQM	vs H.	MQM	vs H.	MQM	vs H.
Accuracy/Mistranslation	33.2	27.2	0.296	1.285	4.3	1.026	3.5	1.219	4.1	1.244	4.2
Style/Awkward	14.6	4.6	0.146	0.299	2.0	0.289	2.0	0.315	2.1	0.296	2.0
Fluency/Grammar	10.7	4.7	0.097	0.224	2.3	0.193	2.0	0.215	2.2	0.196	2.0
Accuracy/Omission	3.6	13.4	0.070	0.091	1.3	0.063	0.9	0.063	0.9	0.120	1.7
Accuracy/Addition	1.8	6.7	0.067	0.025	0.4	0.018	0.3	0.024	0.4	0.021	0.3
Terminology/Inappropriate	8.3	7.0	0.061	0.193	3.2	0.171	2.8	0.189	3.1	0.193	3.2
Fluency/Spelling	2.3	1.2	0.030	0.039	1.3	0.030	1.0	0.039	1.3	0.028	0.9
Accuracy/Untranslated tex	3.1	14.9	0.024	0.090	3.8	0.082	3.5	0.066	2.8	0.098	4.2
Fluency/Punctuation	20.3	0.2	0.014	0.039	2.8	0.067	4.9	0.013	1.0	0.011	0.8
Other	0.5	5.2	0.005	0.010	1.9	0.009	1.6	0.010	1.9	0.007	1.2
Fluency/Register	0.6	5.0	0.005	0.014	3.0	0.009	1.9	0.015	3.2	0.015	3.3
Terminology/Inconsistent	0.3	0.0	0.004	0.005	1.2	0.004	0.9	0.005	1.2	0.005	1.2
Non-translation	0.2	100.0	0.003	0.083	28.3	0.041	14.0	0.065	22.0	0.094	32.0
Fluency/Inconsistency	0.1	1.3	0.003	0.002	0.7	0.001	0.3	0.001	0.3	0.003	1.0
Fluency/Character enc.	0.1	3.7	0.002	0.001	0.7	0.002	1.0	0.001	0.6	0.000	0.2

# 本講演の概要

## 1. 制約付き翻訳シェアードタスクの設立

- アノテーション付きデータセットの作成
- 簡易的な自動評価指標の策定

## 2. 制約付き翻訳モデルの紹介

- 2種類の手法：「緩い」「厳しい」制約付き翻訳モデル
- 上記モデルの応用（実際にタスクに提出されたシステム）

## 3. タスクの結果 & 今後の課題

- 現状の高性能なモデルとは？
- 評価（スコアリング）の信頼性
- エラー分析

# 本講演の概要

## 1. 制約付き翻訳シェアードタスクの設立

- アノテーション付きデータセットの作成
- 簡易的な自動評価指標の策定

## 2. 制約付き翻訳モデルの紹介

- 2種類の手法：「緩い」「厳しい」制約付き翻訳モデル
- 上記モデルの応用（実際にタスクに提出されたシステム）

## 3. タスクの結果 & 今後の課題

- 現状の高性能なモデルとは？
- 評価（スコアリング）の信頼性
- エラー分析

# 「制約付き翻訳（の共有）タスク」の設立

- ASPEC科学論文ドメインデータ [Nakazawa+'16] を元に再設計

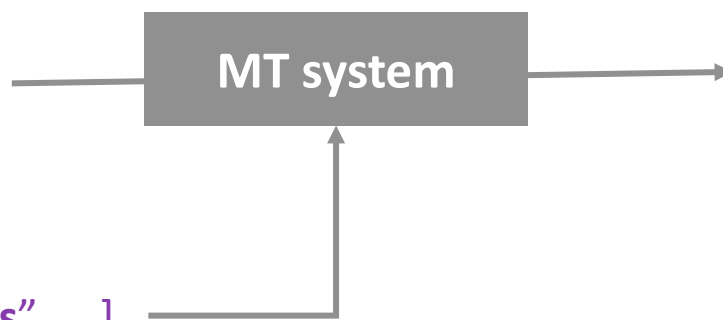
専門用語が多く含まれる

- 「与えられた**制約語彙**に従い，文を翻訳する」
  - 制約語彙リスト：対象言語側の用語集
  - ^に含まれてない同義の表現を許容しない

...シフトする帰還回路であり ...物体の**磁気特性**の変化...で検出することができる。

## 制約語彙リスト

["feedback circuit", "magnetic features", ...]



... can detect changes in the **magnetic characteristics** of the object in the center ...



... can detect change of **magnetic features** of an object present ...

# なぜ「共有タスク」を作るのか？

---

- 共有タスク = **統一的な評価基盤** の作成
  - 様々なモデルを, 同じデータセット・同じ条件下で比較可能
    - ? 「現時点で最も良いモデルは何か？」
    - ? 「最も良いモデルでもまだ解けていない課題とは？」
- 新規参入, 技術共有の促進
  - データセットが整備されていると, 参入しやすい
  - シェアードタスク参加者には論文投稿の義務あり
    - ? 「どのような工夫が効果的だった or でなかったか？」



# 共有タスク設立に向けて必要な準備

## ASPEC (科学論文翻訳データ)

1. ...物体の**磁気特性**の変化  
...で検出することができる。  
2. ...

1. .... can detect change of  
**magnetic features** of an  
object present ....  
2. ...



## 制約語彙リスト

1. [..., **magnetic features**, ...]  
2. [...]

...シフトする帰還回路であり ...物  
体の**磁気特性**の変化...で検出する  
ことができる。

### 制約語彙リスト

[“feedback circuit”, “**magnetic features**”, ...]

MT system



(System A) ... can detect changes in  
the **magnetic characteristics** of the  
object in the center ...

Score: 34



(System B) ... can detect change  
of **magnetic features** of an object  
present ....

Score: 92

1. 元データに対応する「制約語彙リスト」作成
2. 評価指標（スコア付けの基準）の策定

# データセットの作成：制約語彙のアノテーション

- 評価のために，“制約語彙”の正解を文単位でアノテーションする必要



## 1. 制約語彙（用語）の抽出

...物体の**磁気特性**の変化...  
で検出することができる。  
...

... can detect change of  
**magnetic features** of an  
object present ....  
...



## 2. 抽出された用語同士の対応付け

### 制約語彙リスト

[  
 (“帰還回路”,  
 “**磁気特性**”),  
 (...),  
 ]

[  
 (“feedback circuit”,  
 “**magnetic features**”),  
 (...),  
 ]

[2021年] 全工程，人手作業（英日・日英）

• **高コスト**

# データセットの作成：制約語彙のアノテーション

- 評価のために，“制約語彙”の正解を文単位でアノテーションする必要



## 1. 制約語彙（用語）の抽出

...物体の**磁気特性**の変化...  
で検出することができる。  
...

... can detect change of  
**magnetic features** of an  
object present ...



## 2. 抽出された用語同士の対応付け

### 制約語彙リスト

[  
 (“帰還回路”,  
 “**磁気特性**”),  
 (...),  
 ]

[  
 (“feedback circuit”,  
 “**magnetic features**”),  
 (...),  
 ]



## 3. 対応付（3）の確認

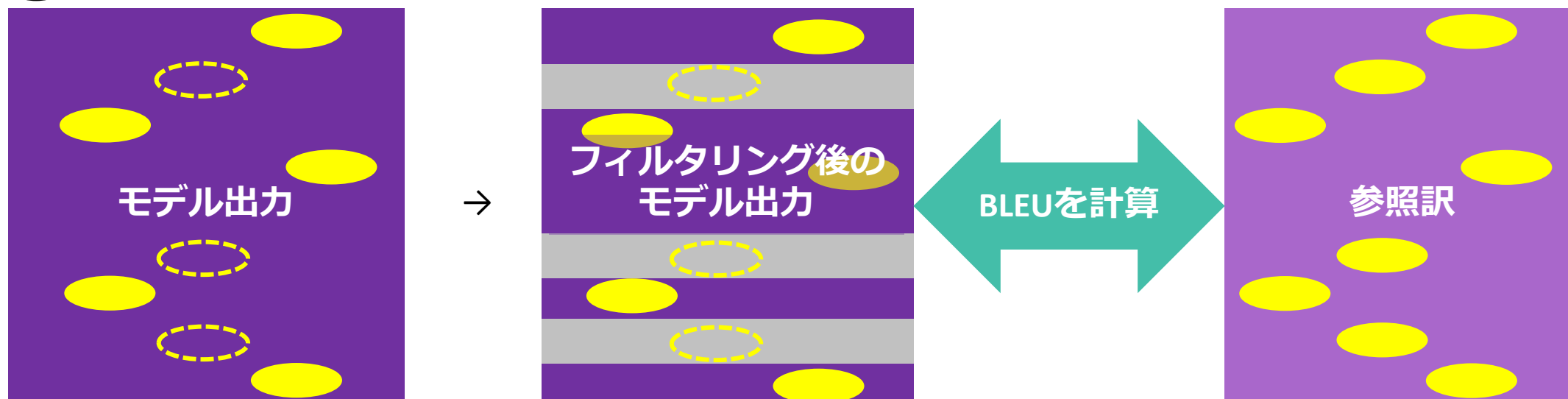
[2022年] 自動抽出 → 人手対応付け・確認 (中日・日中)

- 依然として高コスト（自動対応付けはうまくいかず...）

# 自動評価指標の策定：翻訳の質 & 用語一貫性

- 「翻訳の質」と「用語の一貫性」の両方を考慮
  - 制約語彙が含まれていない例をフィルタリング (exact match)
  - 参照訳とのoverlapを計測 (BLEU)
    - BLEUにおける, BP (出力が短いと減点) を用語一貫性のスコアとして利用

● : 制約語彙



# 人手評価指標：翻訳の質

- 自動翻訳だけでは、完全には翻訳の質を測りきれない → 人手評価
- **Source-based Direct Assessment (src-based DA)** [Federmann+'18]
  - 評価者（バイリンガル）に、元文とシステム出力文を提示して点数付与

The screenshot shows a web interface for evaluating translations. At the top, there are navigation links for 'Appraise' and 'Dashboard', and a user identifier 'zhoeng2701'. Below this, a header bar indicates '0/10 blocks, 10 items left in block', the task ID 'AppenEvalFY1827 #3672: Segment #640', and the language pair 'Chinese (中文) → English'. The main content area displays the source text in Chinese: '而安特卫普为全球最大的钻石交易中心之一，当地工匠的钻石切割技术名满天下，所出售的钻石经过严格鉴定，深受内地女士的欢迎。' Below the source text is the candidate translation in English: 'Antwerp is one of the world's largest diamond trading centers, local artisans diamond cutting technology name world, the sale of diamonds after rigorous identification, by the mainland ladies welcome.' A slider is positioned below the candidate translation, with a label: 'How accurately does the above candidate text convey the original semantics of the source text? Slider ranges from Not at all (left) to Perfectly (right)'. The slider is currently set to a low value. At the bottom, there are 'Reset' and 'Submit' buttons.

シーケンスバー（100段階）で評価

# 2021-2022年における共有タスクの結果

年	提出システム	En-Ja		Ja-En	
		自動評価 (Exact+BLEU)	人手評価 (src-based DA)	自動評価 (Exact+BLEU)	人手評価 (src-based DA)
2021	<b>Hard&amp;Soft const (ensemble) (NTT)</b>	<b>57.2</b>	<b>77.5</b>	<b>44.1</b>	<b>75.6</b>
	Soft const (NHK)	33.9	74.1	37.5	73.9
	Hard&Soft const (NICT)	28.8	73.6	31.8	72.1
	Hard const (TMU)	-	-	22.6	50.2
	(human ref.)	-	73.4	-	74.1
2022	Soft const+Post-edit (ensemble) (TMU)	52.7	76.4	40.8	74.1
	Soft const+Post-edit (TMU)	50.5	76.6	38.1	72.0
	Soft const (TMU)	37.6	74.9	23.0	73.3
	(human ref.)	-	76.6	-	74.7

# 本講演の概要

## 1. 制約付き翻訳シェアードタスクの設立

- アノテーション付きデータセットの作成
- 簡易的な自動評価指標の策定

## 2. 制約付き翻訳モデルの紹介

- 2種類の手法：「緩い」「厳しい」制約付き翻訳モデル
- 上記モデルの応用（実際にタスクに提出されたシステム）

## 3. タスクの結果 & 今後の課題

- 現状の高性能なモデルとは？
- 評価（スコアリング）の信頼性
- エラー分析

# 制約付き翻訳のためのモデル

---

- 統計的機械翻訳 : 「特定のフレーズを入れて翻訳」が指定可能
- ニューラル機械翻訳 : // が簡単にはできない

- ニューラル機械翻訳における「制約付き翻訳」モデルの試み

## Soft constrained : 「緩い」制約モデル [Song+'19][Chen+'20]

- モデル学習時のデータに制約を課す (データ拡張)
- 必ずしも制約語彙を出力できるとは限らない

## Hard constrained : 「厳しい」制約モデル [Arthur+'16][Hokamp&Liu'17][Post&Vilar'18]

- デコーディング時に強制的に制約語彙を出力させる
- 必ず制約語彙を出力できる



# 「緩い」制約：Lexical-Constraint-Aware NMT (LeCA) [Chen+'20]

• モデル学習時に，以下の形式でデータを学習

- input: 元文 + <sep> + 制約語彙1 + <sep> + 制約語彙2 ...
- output: 翻訳文

→ 学習を通じて，入りに付加された制約語彙を含むよう仕向ける

- 明示的なルール・制約のための機構を組み込んではいない  
→ **必ず制約語彙含むとは限らない**

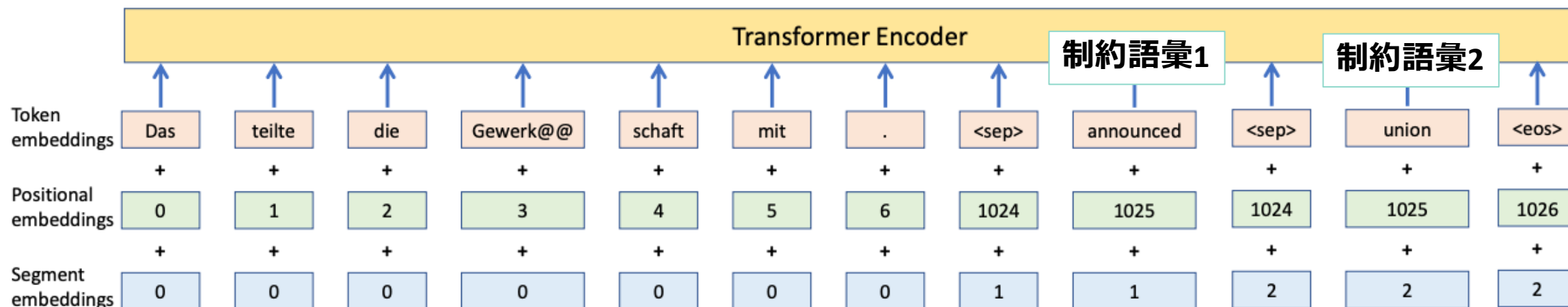


Figure from [Chen+'20]

# 「厳しい」 制約 : Dynamic Beam Allocation (DBA) [Post&Vilar'18]

- デコーディング時, 出力探索中に制約語彙に相当するトークン列を強制的に出力させる
  - 強制的な出力により, モデルの**翻訳の質が下がる**ことがある
  - (翻訳の質を求めて探索幅を多くすると) **推論速度が遅くなる**

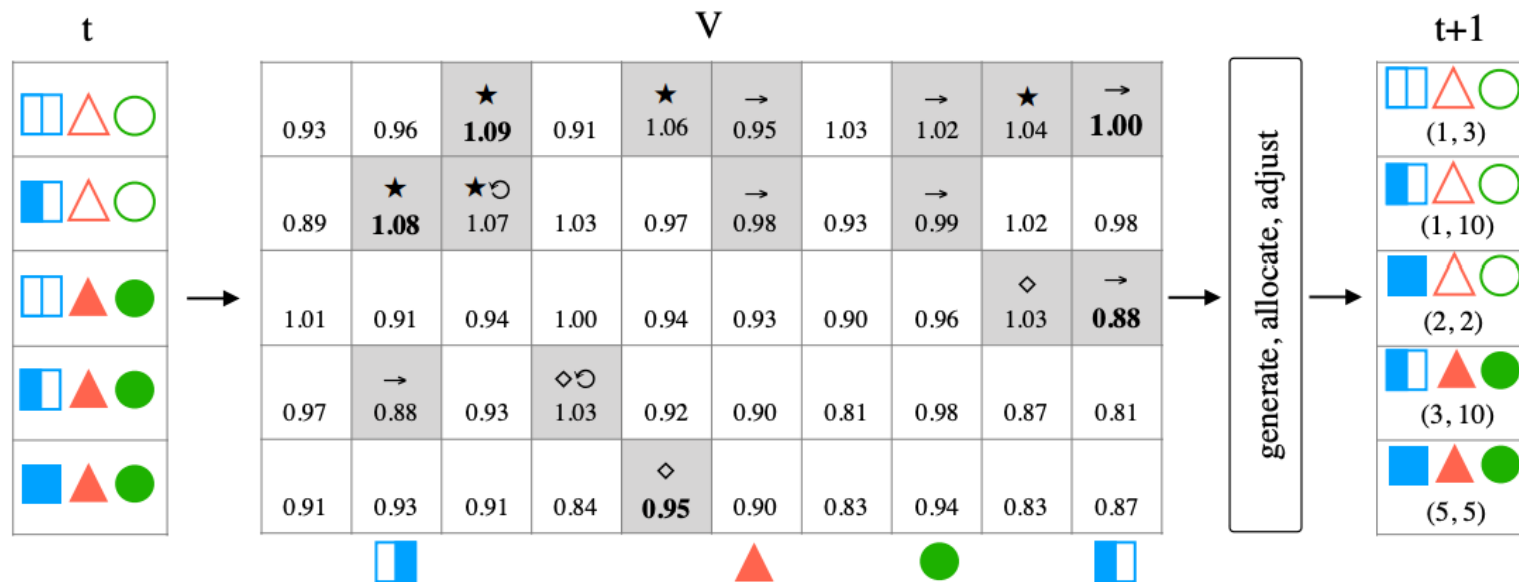


Figure from [Post&Villar'18]

# 提出システム1: 「緩い」 & 「厳しい」 制約の結合 [Chousa&Morishita'21]

LeCA: 「緩い」  
制約モデル

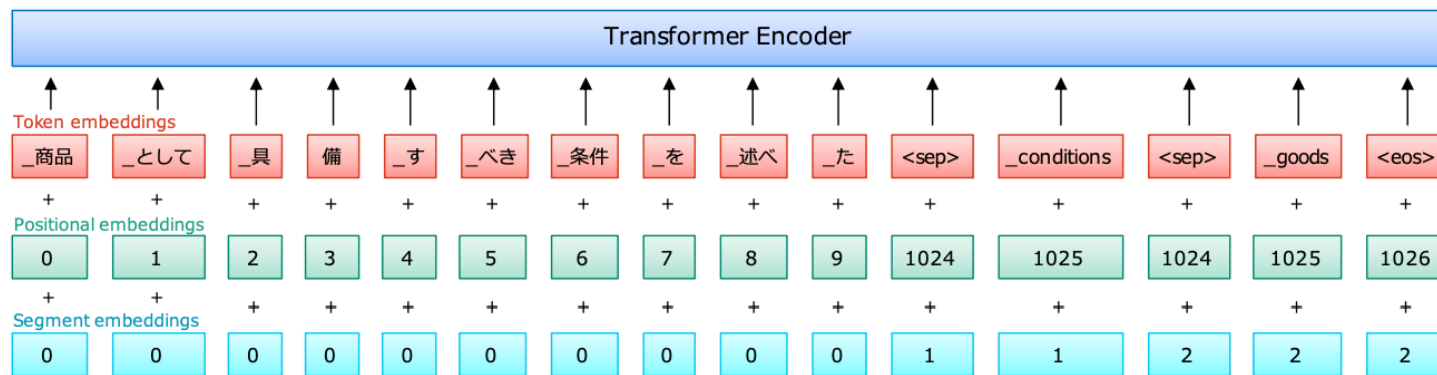
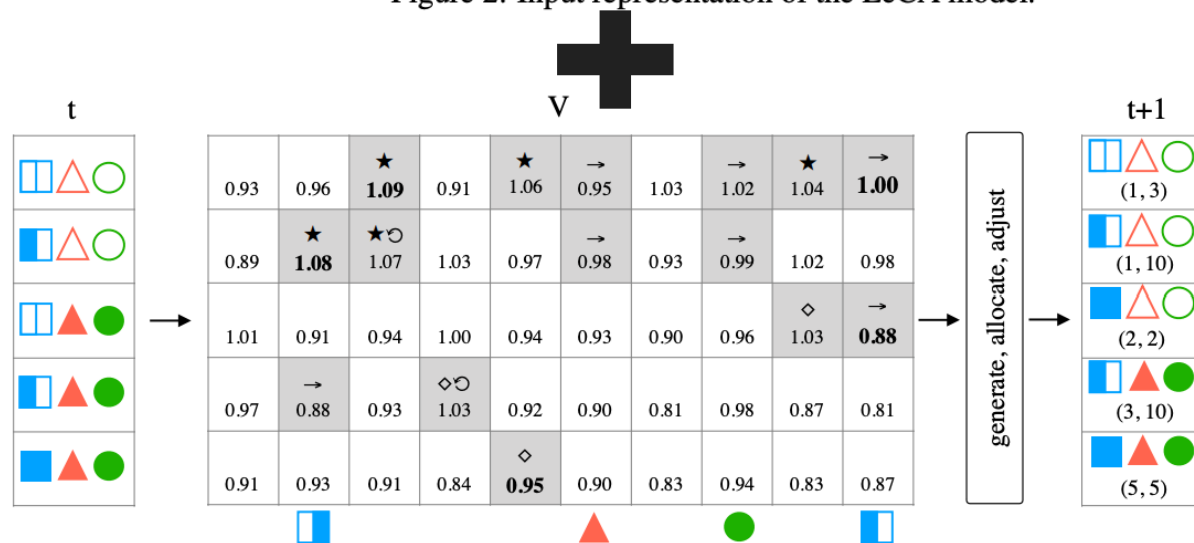


Figure 2: Input representation of the LeCA model.

DBA: 「厳しい」  
制約モデル



事前にLeCAである程度  
制約語彙が探索中に出  
現するモデルに  
→ DBAによる翻訳の質  
低下を抑える

# 提出システム2: 「緩い」制約 + 「後編集」 [Kondo&Komachi'22]

## Encoder (LeCA)

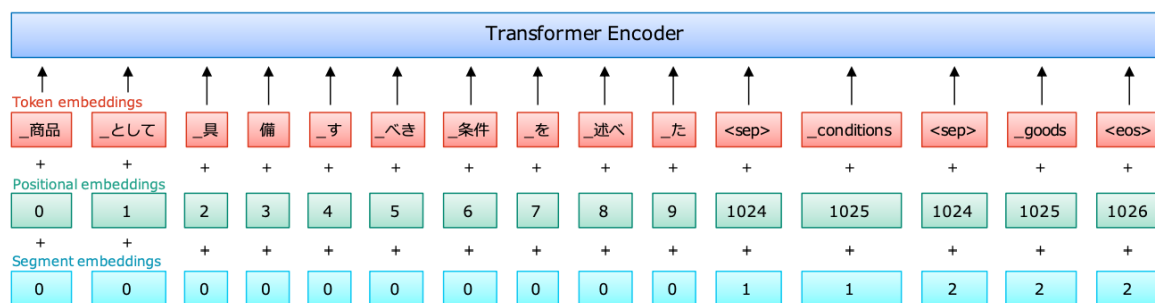
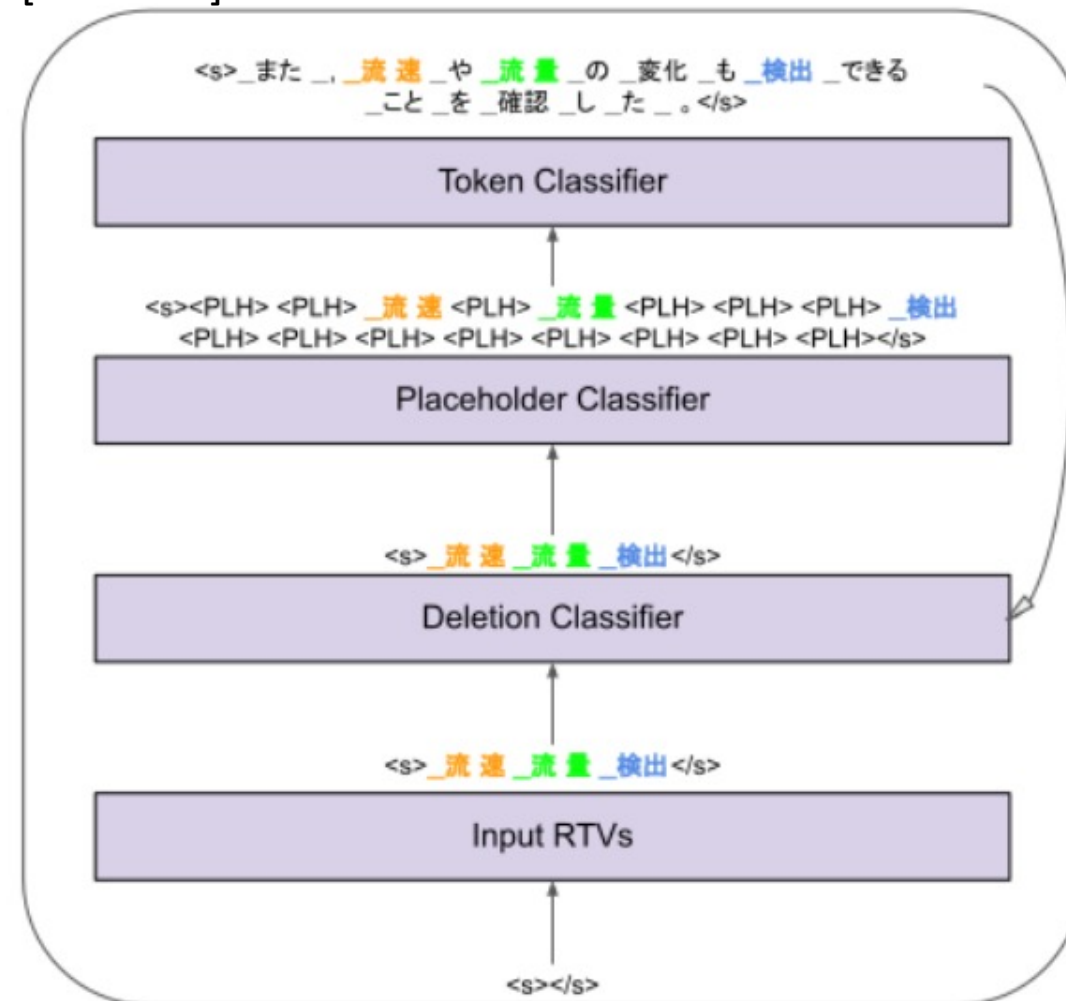


Figure 2: Input representation of the LeCA model.

LeCA（「緩い」制約モデル）は  
制約語彙を含まない可能性あり  
→ 制約語彙を必ず含むように  
「後編集」モデルを導入

## Decoder (Multi-Source Levenshtein Transformer: MSLevT) [Wan+'20]



# 本講演の概要

## 1. 制約付き翻訳シェアードタスクの設立

- アノテーション付きデータセットの作成
- 簡易的な自動評価指標の策定

## 2. 制約付き翻訳モデルの紹介

- 2種類の手法：「緩い」「厳しい」制約付き翻訳モデル
- 上記モデルの応用（実際にタスクに提出されたシステム）

## 3. タスクの結果 & 今後の課題

- 現状の高性能なモデルとは？
- 評価（スコアリング）の信頼性
- エラー分析

# 2021-2022年における共有タスクの結果（再掲）

年	提出システム	En-Ja		Ja-En	
		自動評価 (Exact+BLEU)	人手評価 (src-based DA)	自動評価 (Exact+BLEU)	人手評価 (src-based DA)
2021	<b>Hard&amp;Soft const (ensemble) (NTT)</b>	<b>57.2</b>	<b>77.5</b>	<b>44.1</b>	<b>75.6</b>
	Soft const (NHK)	33.9	74.1	37.5	73.9
	Hard&Soft const (NICT)	28.8	73.6	31.8	72.1
	Hard const (TMU)	-	-	22.6	50.2
	(human ref.)	-	73.4	-	74.1
2022	Soft const+Post-edit (ensemble) (TMU)	52.7	76.4	40.8	74.1
	Soft const+Post-edit (TMU)	50.5	76.6	38.1	72.0
	Soft const (TMU)	37.6	74.9	23.0	73.3
	(human ref.)	-	76.6	-	74.7

# Q. 本タスクにおいて高性能なモデルは？

year	System	En-Ja		Ja-En	
		Auto Eval. (Exact+BLEU)	Human Eval. (src-based DA)	Auto Eval. (Exact+BLEU)	Human Eval. (src-based DA)
2021	<b>Hard&amp;Soft const (ensemble) (NTT)</b>	<b>57.2</b>	<b>77.5</b>	<b>44.1</b>	<b>75.6</b>
	Soft const (NHK)	33.9	74.1	37.5	73.9
	Hard&Soft const (NICT)	28.8	73.6	31.8	72.1
	Hard const (TMU)	-	-	22.6	50.2
	(human ref.)	-	73.4	-	74.1

**最高性能モデル**：「緩い」 & 「厳しい」 制約モデル

- 制約語彙を100%出力可能（「厳しい制約」モデルのため）
- 非常に高い質で翻訳を出力可能（自動評価，人手評価）

# Q. 本タスクにおいて高性能なモデルは？

year	System	En-Ja		Ja-En	
		Auto Eval. (Exact+BLEU)	Human Eval. (src-based DA)	Auto Eval. (Exact+BLEU)	Human Eval. (src-based DA)
2021	<b>Hard const+Soft const+Filtering (NTT)</b>	<b>57.2</b>	<b>77.5</b>	<b>44.1</b>	<b>75.6</b>

## 次点の高性能モデル: 「緩い」制約+ 「後編集」モデル

- 制約語彙を100%出力可能（「後編集」の補完により）
- （恐らく）一般的な「厳しい」制約モデルより早い (0.115 秒/文)

	(human ref.)	-	73.4	-	74.1
2022	<b>Soft const+Post-edit (ensemble) (TMU)</b>	<b>52.7</b>	<b>76.4</b>	<b>40.8</b>	<b>74.1</b>
	Soft const+Post-edit (TMU)	50.5	76.6	38.1	72.0
	Soft const (TMU)	37.6	74.9	23.0	73.3
	(human ref.)	-	76.6	-	74.7



# 分析：新しく考案した自動評価の妥当性

Q: 自動評価の結果が適切か？ → 人手評価との相関高ければ適切？

- スピアマンの順位相関の結果 (w/ 12システム)

自動評価	相関値
↔ BLEU	0.795
↔ Exact match	0.498
↔ <b>Ours (Exact+BLEU)</b>	<b>0.836</b>

- 考案した自動評価 (Exact+BLEU) は人手評価との相関高  
→ 「用語の一貫性」 (Exact match) を考慮することの重要性

# 分析：最高性能モデルにおけるエラー事例とは？

- 最高性能モデルの、低い人手評価スコアの例を確認
- 4つの要因に起因するエラー
  1. 制約語彙に含まれていない「用語」の欠落（データアノテーションの質）
  2. **1に含まれない誤訳（翻訳モデルの質）**
  3. **アノデータのドメイン知識の不足（人手評価の質）**
  4. ASPEC参照訳の問題（元のデータの質）

多かったのは、2&3

# 分析：アノテーション（制約語彙リスト）の質に起因するエラー

- 最高性能モデルの、低い人手評価スコアの例を確認
- 4つの要因に起因するエラー
  1. 制約語彙に含まれていない「用語」の欠落（データアノテーションの質）
  2. 1に含まれない誤訳（翻訳モデルの質）
  3. アノテータのドメイン知識の不足（人手評価の質）
  4. ASPEC参照訳の問題（元のデータの質）

Src: This paper presents security card renewal construction of **Hokkoku Newspaper Company**.

Ref: **北国新聞社**のセキュリティカード更新工事を紹介した。

Out: **北新聞社**のセキュリティカード更新工事について紹介した。

Score: 37

RV: [] (なし)

# 分析：翻訳モデルの質に起因するエラー

- 最高性能モデルの、低い人手評価スコアの例を確認
- 4つの要因に起因するエラー
  1. 制約語彙に含まれていない「用語」の欠落（データアノテーションの質）
  2. **1に含まれない誤訳（翻訳モデルの質）**
  3. アノテータのドメイン知識の不足（人手評価の質）
  4. ASPEC参照訳の問題（元のデータの質）

Src: "Ubiquitous" is the meaning of "it is omnipresent like God".

Ref: ユビキタス ( u b i q u i t o u s ) は「**神のように遍在する**」という意味である。

Out: 「ユビキタス ( u b i q u i t o u s ) 」とは、「**G o dのように不思議である**」という意味である。

RV: ["ユビキタス ( u b i q u i t o u s ) "]

Score: 33

# 分析：人手評価の質に起因するエラー

- 最高性能モデルの、低い人手評価スコアの例を確認
- 4つの要因に起因するエラー
  1. 制約語彙に含まれていない「用語」の欠落（データアノテーションの質）
  2. 1に含まれない誤訳（翻訳モデルの質）
  3. **アノテータのドメイン知識の不足（人手評価の質）**
  4. ASPEC参照訳の問題（元のデータの質）

Src: *Zizania latibolia* Turcz was cultivated in a laboratory, and changes of water quality in a rooting zone were measured.

Ref: マコモを実験室で栽培し、根圏における水質変化を測定した。

Out: マコモを実験室で栽培し、根圏の水質変化を測定した。

Score: 25

RV: ["マコモ", "根圏"]

# 分析：人手評価の質に起因するエラー

- 最高性能モデルの、低い人手評価スコアの例を確認
- 4つの要因に起因するエラー
  1. 制約語彙に含まれていない「用語」の欠落（データアノテーションの質）
  2. 1に含まれない誤訳（翻訳モデルの質）
  3. **アノータタのドメイン知識の不足（人手評価の質）**
  4. ASPEC参照訳の問題（元のデータの質）

Src: *Zizania latibolia* Turcz was cultivated in a laboratory, and changes of water quality in a rooting zone were measured.

Ref: マコモを実験室で栽培し、根圏における水質変化を測定した。

Out: マコモを実験室で栽培し、根圏の水質変化を測定した。

RV: ["マコモ", "根圏"]

Score: 25

現状、評価者は Src ⇔ **Outのみ**で評価 (src-based DA)

- 制約語彙を見せて評価？
- 使用する人手評価の枠組みを変更？

# 分析：元の翻訳データの質に起因するエラー

- 最高性能モデルの、低い人手評価スコアの例を確認
- 4つの要因に起因するエラー
  1. 制約語彙に含まれていない「用語」の欠落（データアノテーションの質）
  2. 1に含まれない誤訳（翻訳モデルの質）
  3. アノテータのドメイン知識の不足（人手評価の質）
  4. **ASPEC参照訳の問題（元のデータの質）**

Src: Japan is behind **other countries** in taking measures against misconduct.

Ref: 日本では不正行為への対応が遅れている。

Out: 日本は不正行為への対応が遅れている。 **Score: 58**

“other countries（「他の国より」）”が  
Ref, Outともに欠けている

RV: [“不正行為”]

# まとめ：制約付き翻訳タスクの現状

背景：適切に用語の翻訳がしたい！

## 1. 制約付き翻訳シェアードタスクについて（統一的な評価基盤作成）

- データセット・自動評価の枠組み整備

## 2. 現在の制約付き翻訳モデルの紹介

- 最高性能モデル：「緩い」 & 「厳しい」制約モデル
- 次点モデル：「緩い」制約 & 「後編集」モデル

## 3. 今後の課題

Q. 現状のモデルでも解けていない問題は？ → 用語以外の翻訳の質 & 推論速度

- 現状のタスク設定では、制約語彙は100%出力可能

Q. 評価の信頼性？

- 自動評価は人手評価と高相関
- 高い専門知識を必要とする例の人手評価が怪しい → 人手評価の見直し



# References

---

- [Freitag+’21] Markus Freitag, George Foster, David Grangier, Viresh Ratnakar, Qijun Tan, and Wolfgang Macherey. 2021. Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation. *Transactions of the Association for Computational Linguistics*, 9:1460–1474.
- [Nakazawa+’16] Toshiaki Nakazawa, Manabu Yaguchi, Kiyotaka Uchimoto, Masao Utiyama, Eiichiro Sumita, Sadao Kurohashi, and Hitoshi Isahara. 2016. ASPEC: Asian Scientific Paper Excerpt Corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 2204–2208, Portorož, Slovenia. European Language Resources Association (ELRA).
- [Federman+’18] Christian Federmann. 2018. Appraise Evaluation Framework for Machine Translation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 86–88, Santa Fe, New Mexico. Association for Computational Linguistics.
- [Song+’19] Kai Song, Yue Zhang, Heng Yu, Weihua Luo, Kun Wang, and Min Zhang. 2019. Code-Switching for Enhancing NMT with Pre-Specified Translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 449–459, Minneapolis, Minnesota. Association for Computational Linguistics.
- [Chen+’20] Guanhua Chen, Yun Chen, Yong Wang, Victor O.K. Li. 2020. Lexical-Constraint-Aware Neural Machine Translation via Data Augmentation. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*, pages 3587–3593. International Joint Conferences on Artificial Intelligence Organization.
- [Arthur+’16] Philip Arthur, Graham Neubig, and Satoshi Nakamura. 2016. Incorporating Discrete Translation Lexicons into Neural Machine Translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1557–1567, Austin, Texas. Association for Computational Linguistics.
- [Hokamp&Liu’17] Chris Hokamp and Qun Liu. 2017. Lexically Constrained Decoding for Sequence Generation Using Grid Beam Search. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1535–1546, Vancouver, Canada. Association for Computational Linguistics.
- [Post&Vilar’18] Matt Post and David Vilar. 2018. Fast Lexically Constrained Decoding with Dynamic Beam Allocation for Neural Machine Translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1314–1324, New Orleans, Louisiana. Association for Computational Linguistics.
- [Wan+’20] David Wan, Chris Kedzie, Faisal Ladhak, Marine Carpuat, and Kathleen McKeown. 2020. Incorporating terminology constraints in automatic post-editing. In *Proceedings of the Fifth Conference on Machine Translation*, pages 1193–1204, Online. Association for Computational Linguistics.

# References: Restricted Translation Taskに提出されたシステム論文

---

- 2021
  - NTT [Chousa&Morishita'21] Katsuki Chousa and Makoto Morishita. 2021. Input Augmentation Improves Constrained Beam Search for Neural Machine Translation: NTT at WAT 2021. In *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, pages 53–61, Online. Association for Computational Linguistics.
  - NHK [Mino+'21] Hideya Mino, Kazutaka Kinugawa, Hitoshi Ito, Isao Goto, Ichiro Yamada, and Takenobu Tokunaga. 2021. NHK's Lexically-Constrained Neural Machine Translation at WAT 2021. In *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, pages 46–52, Online. Association for Computational Linguistics.
  - NICT [Li+'21] Zuchao Li, Masao Utiyama, Eiichiro Sumita, and Hai Zhao. 2021. NICT's Neural Machine Translation Systems for the WAT21 Restricted Translation Task. In *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, pages 62–67, Online. Association for Computational Linguistics.
  - TMU [Kondo+'21] Seiichiro Kondo, Aomi Koyama, Tomoshige Kiyuna, Tosho Hirasawa, and Mamoru Komachi. 2021. Machine Translation with Pre-specified Target-side Words Using a Semi-autoregressive Model. In *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, pages 68–73, Online. Association for Computational Linguistics.
- 2022
  - TMU [Kondo&Komachi'22] Seiichiro Kondo and Mamoru Komachi. 2022. TMU NMT System with Automatic Post-Editing by Multi-Source Levenshtein Transformer for the Restricted Translation Task of WAT 2022. In *Proceedings of the 9th Workshop on Asian Translation*, pages 51–58, Gyeongju, Republic of Korea. International Conference on Computational Linguistics.