

The 8th Workshop on Asian Translation (WAT2021) 開催報告

中澤 敏明

東京大学

2021/12/09

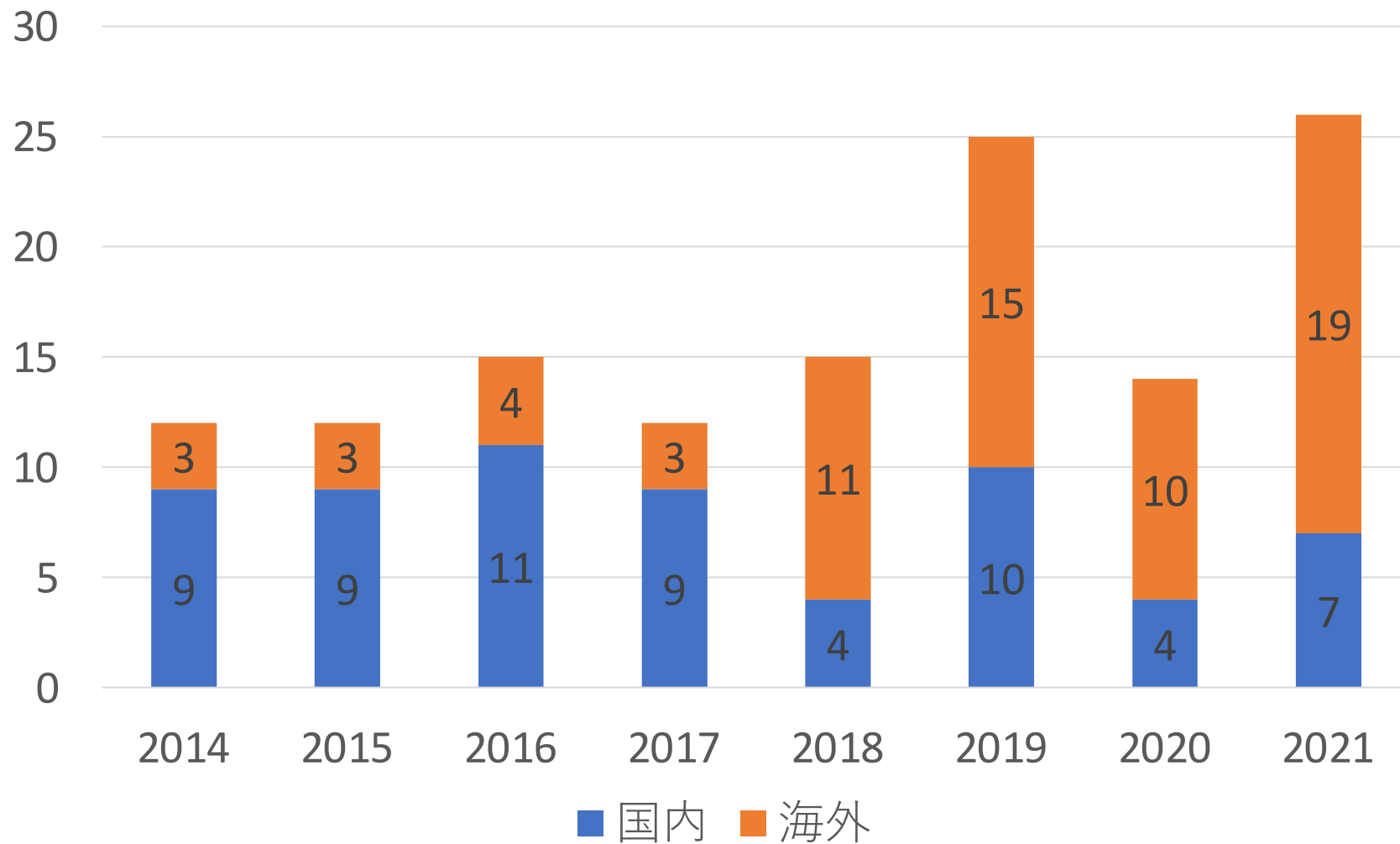
アジア翻訳ワークショップ (WAT)

- アジア言語を対象とした機械翻訳評価ワークショップ
 - 日中韓、インドネシア語、タイ語、マレーシア語、ヒンディー語、インド系諸語、ミャンマー語、ロシア語、クメール語、英語
 - 文書単位の翻訳、マルチモーダル翻訳、制限翻訳など
- 2014年より毎年開催
 - WAT2014 @ 東京、WAT2015 @ 京都
 - WAT2016 @ 大阪 (Coling2016併設)
 - WAT2017 @ 台湾 (IJCNLP2017併設)
 - WAT2018 @ 香港 (PACLIC2018併設)
 - WAT2019 @ 香港 (EMNLP2019併設)
 - WAT2020 @ 中国 (AAACL-IJCNLP2020併設、オンライン)
 - WAT2021 @ タイ (ACL-IJCNLP2021併設、オンライン)
 - WAT2022 @ 韓国 (COLING2022併設、10/12-17)



ドメイン	言語対	2014	2015	2016	2017	2018	2019	2020	2021
科学技術論文	Ja ⇔ En/Zh	→ (2020から文書単位)							
特許	Ja ⇔ Zh	→ (Zh→Jaのみ)							
	Ja ⇔ Ko	→ (Ko→Jaのみ)							
	Ja ⇔ En	→							
新聞	Id ⇔ En	→							
	Ja ⇔ En	→							
	Ja ⇔ Ru	→							
インド諸語	Hi ⇔ En	→							
	Hi ⇔ Ja	→							
	Ta ⇔ En	→							
	Odia ⇔ En	→							
	Indic Multilingual	→ (7言語) (10言語)							
混合	My ⇔ En	→							
	Km ⇔ En	→							
レシピ	Ja ⇔ En	→							
適時開示文書	Ja → En	→							
IT&Wikinews	Hi/Th/Ms/Id ⇔ En	→							
マルチモーダル	En → Hi	→							
	En → MI	→							
	Ja ⇔ En	→ (データ追加)							
ビジネス対話	Ja ⇔ En	→							
制限翻訳	Ja ⇔ En	→							

参加者一ム数



WAT2021参加チーム一覧

Team ID	Organization	Country
TMU	Tokyo Metropolitan University	Japan
NTT	NTT Corporation	Japan
NICT-2	NICT	Japan
NICT-5	NICT	Japan
NLPHut	Idiap Research Institute Switzerland, IIT BHU, BITS Pilani India, KIIT University India, Silicon Techlab pvt. Ltd India, University of Chicago	Switzerland, India, USA
TMEKU	Tokyo Metropolitan University, Ehime University, Kyoto University	Japan
*goodjob	Dalian University of Technology	China
YCC-MT1	University of Technology (Yatanarpon Cyber City)	Myanmar
YCC-MT2	University of Technology (Yatanarpon Cyber City)	Myanmar
NECTEC	National Electronics and Computer Technology Center (NECTEC)	Thailand
mcairt	CAIR	India
nictrb	NICT	Japan
sakura	Rakuten Institute of Technology Singapore, Rakuten Asia.	Singapore
IIT-H	International Institute of Information Technology	India
*gauvar	Amazon	Singapore
*JBJBJB	Individual participant	Korea
SRPOL	Samsung R&D Poland	Poland
NHK	NHK	Japan
CFILT	Computing for Indian Language Technology	India
iitp	IIT Patna	India
Volta	International Institute of Information Technology Hyderabad	India
coastal	University of Copenhagen	Denmark
CFILT-IITB	Indian Institute of Technology Bombay	India
CNLP-NITS-PP	NIT Silchar	India
Bering Lab	Bering Lab	South Korea
tpt_wat	Transperfect Translations	USA

Team ID	ASPEC + ParaNatCom EJ	ASPEC Restricted		ALT + UCSY		NICT-SAP			
		EJ	JE	En-My	My-En	En-Hi/Id/Ms/Th IT	Wikinews	Hi/Id/Ms/Th-En IT	Wikinews
TMU			✓						
NTT		✓	✓						
NICT-2						✓	✓	✓	✓
goodjob	✓								
YCC-MT1				✓					
YCC-MT2				✓					
NECTEC					✓				
nictrb		✓	✓						
sakura				✓	✓	✓	✓	✓	✓
NHK		✓	✓						
human eval	✓	✓	✓			✓		✓	

Team ID	JPC						Multimodal						MS COCO EJ	
	EJ	JE	CJ	JC	KJ	JK	TX	En-Hi HI	MM	En-MI TX	HI	Flickr EJ		JE
TMU	✓	✓			✓	✓								
NLPHut							✓	✓		✓	✓			
TMEKU												✓		✓
sakura												✓	✓	
iitp									✓					
Volta							✓		✓					
CNLP-NITS-PP							✓		✓					
Bering Lab	✓	✓	✓	✓	✓	✓								
tpt_wat	✓	✓	✓	✓	✓	✓								
human eval	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓

Team ID	Indic21																			
	En-X									X-En										
	Bn	Kn	Ml	Mr	Or	Hi	Gu	Pa	Ta	Te	Bn	Kn	Ml	Mr	Or	Hi	Gu	Pa	Ta	Te
NICT-5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NLPHut	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
mcairt	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
sakura	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IIT-H	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
gauvar	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
JBKJB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SRPOL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CFILT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
coastal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CFILT-IITB										✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
human eval	✓	✓	✓	✓	✓					✓	✓	✓	✓	✓						

研究論文

- 5件の研究論文と、ACL-IJCNLP2021から2件の findings論文を採択
- 論文タイトル
 - BTS: Back TranScription for Speech-to-Text Post-Processor using Text-to-Speech-to-Text
 - Zero-pronoun Data Augmentation for Japanese-to-English Translation
 - Evaluation Scheme of Focal Translation for Japanese Partially Amended Statutes
 - Optimal Word Segmentation for Neural Machine Translation into Dravidian Languages
 - Itihasa: A large-scale corpus for Sanskrit to English translation

招待講演

- “Massively Multilingual Translation and Evaluation”
by Francisco Guzmán and Angela Fan (Facebook AI)
- “Understanding and Improving Context Usage in
Context-aware Translation”
by Graham Neubig (Carnegie Mellon University)

翻訳結果評価

- 自動評価
 - BLEU, RIBES, AM-FM
 - 自動評価サーバーに常時、翻訳結果を提出可
- 人手評価
 - 正確性評価
 - Direct Assessment
 - WMT(Conference on Machine Translation)の公式指標
- 評価費用はスポンサー費などによる (AAMT、NICT、サンフレア、川村インターナショナル)

#	Team	Task	Date/Time	DataID	BLEU								Method	Other Resources	
					human	kytea	meCab	Moses tokenizer	stanford-sentimentec-ctb	stanford-sentimentec-rtku	indic tokenizer	unuse			
1	srcb	en-ja	2018/09/16 15:52:17	2480	43.43	44.94	44.05	-	-	-	-	-	NMT	No	pc
2	NICT-5	en-ja	2018/09/03 16:49:57	2219	42.87	44.42	43.49	-	-	-	-	-	NMT	No	E
3	srcb	en-ja	2018/09/16 15:26:37	2479	42.49	44.11	43.20	-	-	-	-	-	NMT	No	pc

正確性評価 (Adequacy)

- 特許庁による特許文献機械翻訳の「内容の伝達レベルの評価」

評価値	評価基準
5	すべての重要情報が正確に伝達されている。(100%)
4	ほとんどの重要情報は正確に伝達されている。(80%~)
3	半分以上の重要情報は正確に伝達されている。(50%~)
2	いくつかの重要情報は正確に伝達されている。(20%~)
1	文意がわからない、もしくは正確に伝達されている重要情報はほとんどない。(~20%)

- 一対評価対象文中の200文を評価
- 文ごとに、2人が評価

Direct Assessment



Sentence: 1

Is the English text (SRC) a good caption for the highlighted area of the image? : Yes No

SRC Text:

Four baseball players on field.

Indicate to what extent each of these candidate translations expresses the meaning of the English source text (independently of the other candidate).

CAND1 Text:

क्षेत्र में बेसबॉल खिलाड़ी

CAND1 Score: worst



best

CAND2 Text:

क्षेत्र में चार बेसबॉल खिलाड़ी ।

CAND2 Score: worst



best

評価結果（の一部）

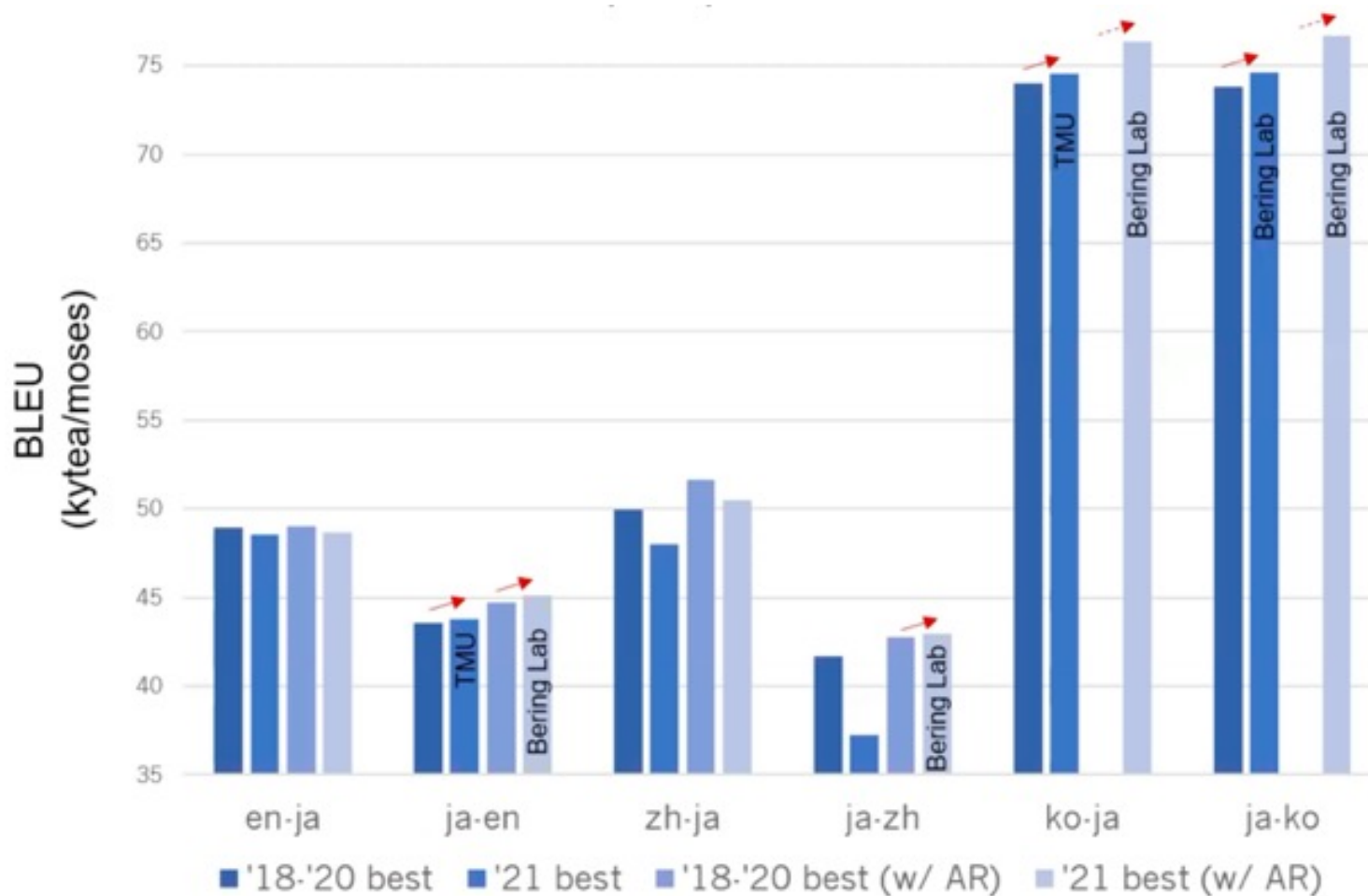
詳細はoverview paperや各チームのsystem description paperを参照してください

<https://aclanthology.org/events/wat-2021/>

特許翻訳

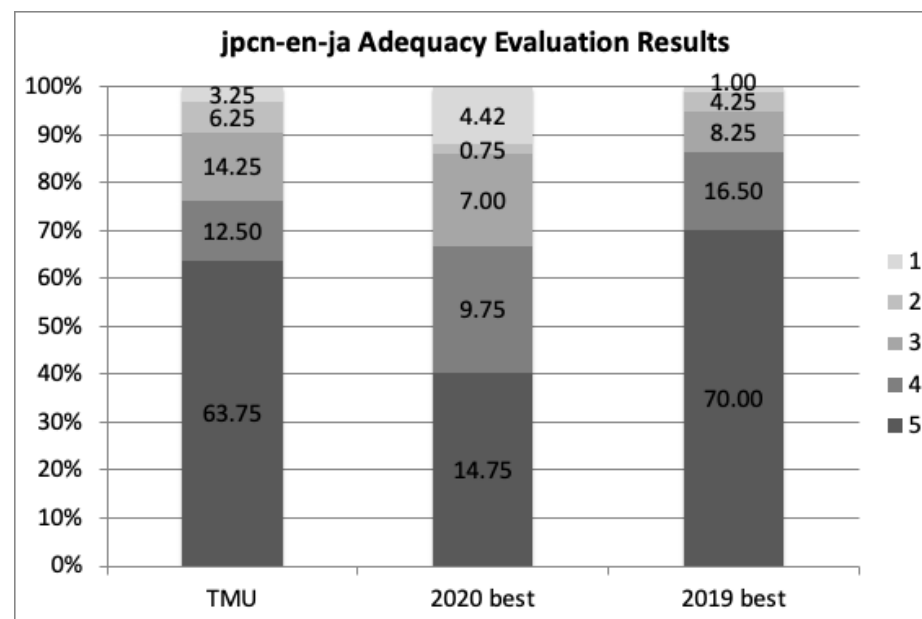
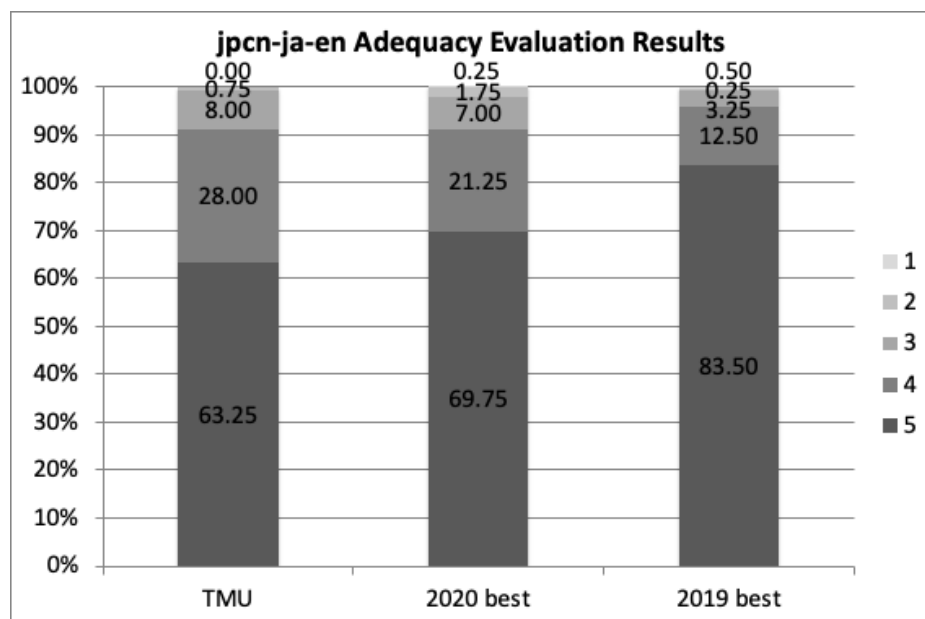
- 特許庁より提供された特許対訳コーパスJPCを利用
 - 日英・日中・日韓の双方向の翻訳タスク
 - 訓練データ25万文
- 日英・日韓には3チーム、日中には2チームが自動評価サーバーに翻訳結果を提出
 - 人手評価用の翻訳結果を提出したのは1チームのみ
- 予算の都合上、実際に人手評価を行ったのは日英、英日タスク 1チーム分のみ

自動評価結果(BLEU)



- Bering Lab：追加の1300万文からなる対訳コーパス
- TMU：fine-tuningされた日本語のBARTモデルを利用

人手評価結果



- 過去のWATでのベストなシステムと比べるとやや低い精度
- BARTという新たなNMTの枠組みを利用しても、過去のシステムに引けを取らない精度であることが示された

日英・英日制限翻訳

- 入力文とともに出力文で必ず使わなければならない訳語のリストが与えられる
 - システムはこれらの訳語を必ず含むように翻訳を生成
 - 与えられるのは訳語のリストのみ、入力文のどの語に対応するのかは与えられない
- ASPEC の日英データ (`dev/devtest/test`) を対象とし、10人のバイリンガルに依頼して専門用語(制限用語)を抽出

日英・英日制限翻訳

- **NMTは訳語統一が不得意**
 - 専門用語や固有名詞を特定の用語に常に正しく翻訳することが求められるような文書の種類も多い
 - この問題に対して、現在の精度を知ることや解決方法を模索するために提案されたタスク
- **自動評価はBLEUスコアと一貫性スコア**
 - 一貫性スコアは全ての制限用語が出力できた文の割合
 - 最終ランキングは全ての制限用語が出力できた文のみで計算したBLEUスコアにより決定
- **人手評価はDirect AssessmentおよびContrastive Assessment**

日英・英日制限翻訳 評価結果

En-Ja Team	final	Human Eval.	
		src-based DA	src-based CA
NTT	57.2	77.5	79.7
NHK	33.9	74.1	77.2
NICTRB	28.8	73.6	77.1
(human ref.)	—	73.4	76.4
Ja-En Team	final	Human Eval.	
		src-based DA	src-based CA
NTT	44.1	75.6	74.4
NHK	37.5	73.9	73.5
NICTRB	31.8	72.1	71.8
TMU	22.6	50.2	48.3
(human ref.)	—	74.1	72.9

日英・英日制限翻訳

- 全てのシステムが採用した手法
 - 入力文に制限用語を付加して入力
 - 翻訳時には制限用語を可能な限り出力するように探索
- システムごとの違いは訓練時の制限用語の準備方法（訓練データには制限用語が付加されていない）
 - 固有表現抽出技術により抽出された固有表現と、ベースラインとなるNMTにおいて誤訳となった表現を制限用語として用いる手法
 - LeCAと呼ばれる手法

日英マルチモーダル翻訳

- **Flicker30k (英日2チーム、日英1チーム)**
 - 訓練データサイズが増加、新たな翻訳技術が提案
→全ての翻訳結果が昨年度の結果よりも改善
 - 英日では画像情報を用いたシステムの方が高精度であったが、日英ではテキスト情報のみを用いたシステムの方が高精度
 - 用いたデータセットが英語から日本語へ、画像情報を参照しながら翻訳されたものであることが関係している？
 - 単語の範囲と画像とのソフトアライメントを考慮したシステムを構築したチームが良い精度を達成
→テキストと画像のグラウンディングが重要
- **Ambiguous MS COCO (英日1チーム)**
 - テキスト情報だけでなく画像情報を適切に用いることで翻訳精度を改善させることができることが示唆

今後の展望

- アジアの翻訳研究の活性化、データ整備等を目的として2014年に始めたアジア翻訳ワークショップ（WAT）は一定程度の成果を得ている
 - 言語対、参加者数の増加
- 実用上問題となる訳抜け、専門用語や低頻度語の翻訳、訳語の一貫性、文脈レベルの翻訳などを評価できるタスク設計が必要
 - 制限翻訳タスクはこのうち一貫性に注目
 - 特定の現象に絞ったようなタスク設計が重要と考えており、引き続きデータの収集や評価指標の定義などを行う
- **WAT2022はCOLING2022併設ワークショップとして、2022年10月12-17日に韓国で開催予定**
 - スポンサー募集！