



[Winter Seminar]

Quality Estimation

Sugyeong Eo

CONTENTS

1. Recap
2. Task1
3. Task2 (sent)
4. Task2 (word)
5. Quality estimation_KO-EN
6. Summary

1. RECAP

- Aims at estimating **quality scores/categories** for given translations from an unknown MT system **without reference translation** at various granularity(sentence/word/phrase)
- Composed of 3 subtasks
 1. Sentence-level direct assessment task: to predict the quality of translation sentences
 2. Word- and Sentence-level post-editing effort: This task evaluates the application of QE for post-editing purposes
 3. Document-level task: to predict document-level quality scores as well as fine-grained annotations

1.1. SENTENCE-LEVEL DIRECT ASSESSMENT TASK

- Sentence-Level Direct Assessment shared task
- The datasets were collected by **translating sentences sampled from source language articles** and annotated with **Direct Assessment (DA) scores** by professional translators.
- DA scores:
 - rated with a score between 0 and 100
 - 10-30: completely incorrect
 - 70-90: closely preserves the semantics
 - 90-100: perfect translation
 - standardized using the z-score

index	original	translation	scores	mean	z_scores	z_mean	model_scores	
0	0	ඔහු ජෙෆර්සන් ගේ දිවිරුම් දීමට ප්‍රථමයෙන් වොම්...	He was the first to give Jefferson's oath of a...	[30, 35, 16, 14, 8, 1]	17.333333333333332	[-1.3041309086819917, -0.6301174512913427, -1....	-1.246499	-0.272417
1	1	මෙම සිතුවම් මගින් පුරාතන රාජධානියේ ජීවිතය දිවි...	These paintings represent the life and cultura...	[100, 70, 91, 100, 94, 82]	89.5	[1.5639505897683126, 0.7933273313367781, 1.707...	1.368105	-0.160187
2	2	ජ්ලේටෝ ඉදිරිපත් කරන පරිදි මරණාභියුක් වූ සොක්‍රටීස්...	As Plato presents the deathly Socrates, the gr...	[56, 40, 26, 94, 96, 98]	68.33333333333333	[-0.2388434949718787, -0.42676819663018256, -0....	0.403440	-0.262659
3	3	ආසියාවේ වර්ධනය වෙමින් පවතින සමාධිමත් බවද මෙයට ...	This was a cause of growing prosperity in Asia.	[62, 42, 32, 100, 89, 95]	70.0	[0.006992062038147374, -0.3454284947657185, -0...	0.472506	-0.275005
4	4	භූමිකාවන්ගේ මුල් බිම සම්බන්ධයෙන් විද්වතුන් අත...	The original floor of the mystics has also be...	[51, 37, 21, 92, 92, 92]	64.16666666666667	[-0.44370645914690043, -0.5487777494268786, -0...	0.260778	-0.278933
...	...	...	...	...	...	...	...	...
995	995	සොයාගැනීම් සඳහා නිමැවුම් වලට වඩා අවධානයක් හිමි...	Possessing more attention than buildings for d...	[41, 11, 11, 70, 88, 79]	50.0	[-0.853432387496944, -1.6061938736649113, -1.3...	-0.173618	-0.414427
996	996	ඇතැමුන් පවසන්නේ මෙයට හේතුවක් ලෙස ඇල්කොහොල, විෂ...	Some say this can be caused by alcohol, toxic...	[46, 32, 16, 93, 91, 89]	61.166666666666664	[-0.6485694233219221, -0.7521270040880388, -1....	0.174589	-0.218976

1.2. WORD- AND SENTENCE-LEVEL POST-EDITING EFFORT

[task2] en-zh

HTER score

MT

Post-edits

Source-tags

SRC

SRC-MT alignments

MT-Tags

0.818182

最后的征服者骑着他的剑继续前进。 → 총 토큰 9개

最后一个征服者骑上了马，拔出了剑。

OK BAD OK BAD BAD BAD BAD BAD BAD BAD

the last conquistador then rides on with his sword drawn .

0-1 1-0 2-1 2-2 3-2 4-3 5-3 6-5 7-4 8-6 9-7 10-8

OK OK OK BAD OK OK BAD BAD OK BAD OK BAD OK BAD OK BAD OK

 : MT token

 : GAP token

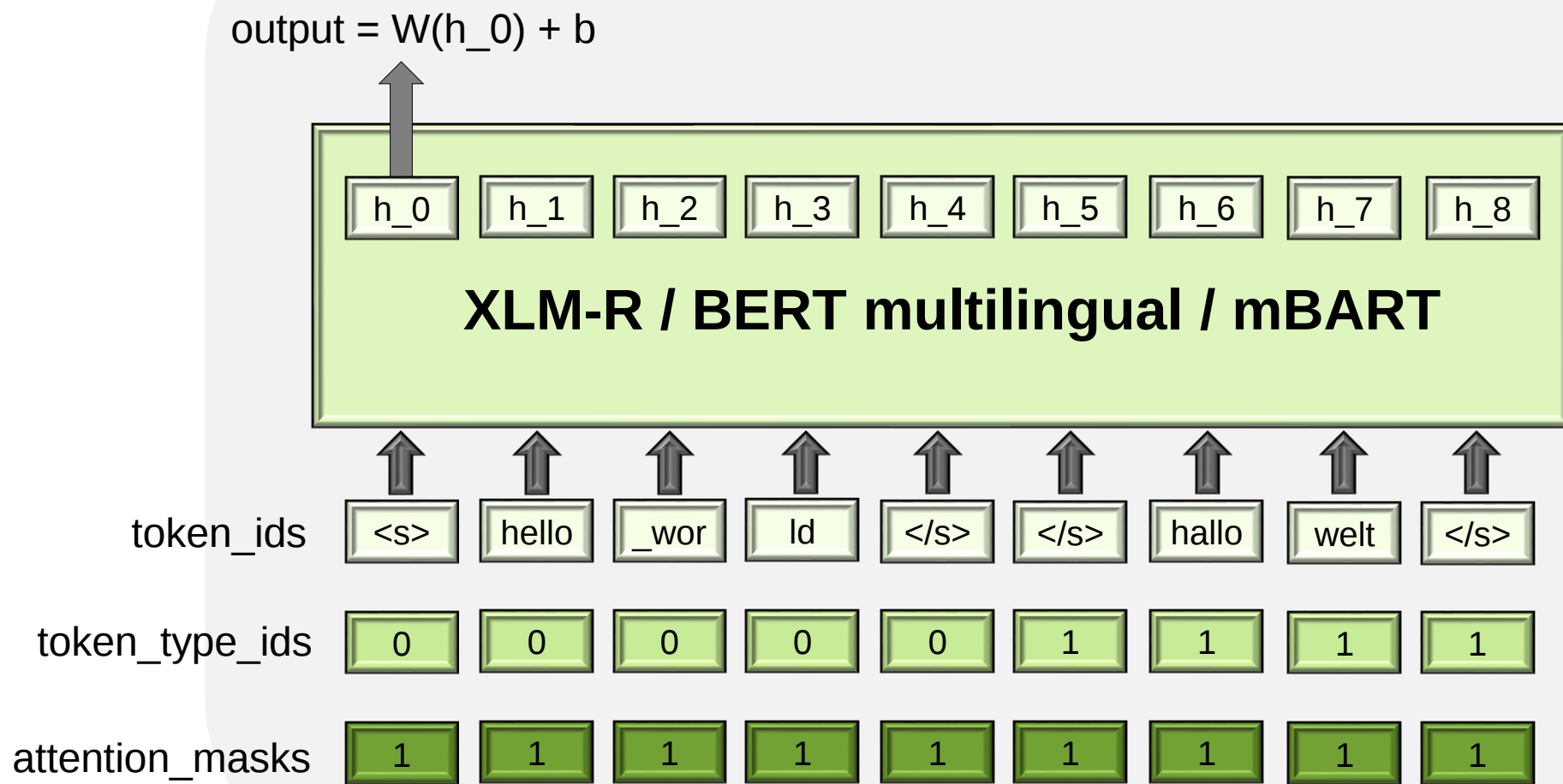
of MT Tags(N) = # of MT tokens + # of GAP tokens(MT tokens +1) = 2N+1

2. QUALITY ESTIMATION-TASK 1

English-German

	Team	Codalab username	Run	Pearson r ▼	MAE	RMSE
1	NiuTrans	NiuTrans		0.562	0.558	0.676
2	• TransQuest	tharindu		0.554	0.613	0.740
3	• Bergamot-LATTE	ssun32		0.544	0.451	0.616
4	IST and Unbabel	Joao1996		0.523	0.470	0.635
5	Papago	zaemyung		0.498	0.454	0.637
6	TMUOU	TMUOU		0.482	0.455	0.625
7	Bergamot	Lisalan		0.476	0.483	0.636
8	XC	xc2019		0.465	0.739	0.861
9	nc	zhanghuimeng		0.392	0.715	0.964
10	IST and Unbabel	ddk		0.267	0.525	0.683
11	Bergamot-LATTE	mfomicheva		0.259	0.819	0.940
12	WL Research	mhkane_225		0.253	0.527	0.683
13	nc	DexinWang		0.246	0.503	0.680
14	nc	Hancheng_Deng		0.171	0.490	0.726
15	Organisers	fblain	BASELINE	0.146	0.679	0.967
16	FVCRC	leizhou		0.111	0.805	1.063

2. QUALITY ESTIMATION-TASK 1



2. QUALITY ESTIMATION-TASK 1

English-German

	Team	Codalab username	Run	Pearson r ▼	MAE	RMSE
1	NiuTrans	NiuTrans		0.562	0.558	0.676
2	• TransQuest	tharindu		0.554	0.613	0.740
3	• Bergamot-LATTE	ssun32		0.544	0.451	0.616
4	IST and Unbabel	Joao1996		0.523	0.470	0.635
5	Papago	zaemyung		0.498	0.454	0.637
6	TMUOU	TMUOU		0.482	0.455	0.625
7	Bergamot	Lisalan		0.476	0.483	0.636
8	XC	xc2019		0.465	0.739	0.861
9	nc	zhanghuimeng		0.392	0.715	0.964
10	IST and Unbabel	ddk		0.267	0.525	0.683
11	Bergamot-LATTE	mfomicheva		0.259	0.819	0.940
12	WL Research	mhkane_225		0.253	0.527	0.683
13	nc	DexinWang		0.246	0.503	0.680
14	nc	Hancheng_Deng		0.171	0.490	0.726
15	Organisers	fblain	BASELINE	0.146	0.679	0.967
16	FVCRC	leizhou		0.111	0.805	1.063

TASK1	Pearson	MAE	RMSE
XLM-Roberta			
mBERT			
mBART			

2. QUALITY ESTIMATION-TASK 1

translation	scores	mean	z_scores	z_mean
He was the first to give Jefferson's oath of a...	[30, 35, 16, 14, 8, 1]	17.333333333333332	[-1.3041309086819917, -0.6301174512913427, -1....	-1.246499
These paintings represent the life and cultura...	[100, 70, 91, 100, 94, 82]	89.5	[1.5639505897683126, 0.7933273313367781, 1.707...	1.368105
As Plato presents the deathly Socrates, the gr...	[56, 40, 26, 94, 96, 98]	68.33333333333333	[-0.2388434949718787, -0.42676819663018256, -0...	0.403440
This was a cause of growing prosperity in Asia.	[62, 42, 32, 100, 89, 95]	70.0	[0.006992062038147374, -0.3454284947657185, -0...	0.472506
The original floor of the mystics has also be...	[51, 37, 21, 92, 92, 92]	64.16666666666667	[-0.44370645914690043, -0.5487777494268786, -0...	0.260778
...	...	...	...	...
Possessing more attention than buildings for d...	[41, 11, 11, 70, 88, 79]	50.0	[-0.853432387496944, -1.6061938736649113, -1.3...	-0.173618
Some say this can be caused by alcohol, toxici...	[46, 32, 16, 93, 91, 89]	61.166666666666664	[-0.6485694233219221, -0.7521270040880388, -1....	0.174589

TASK1	Pearson	MAE	RMSE
XLM-Roberta			
mBERT			
mBART			

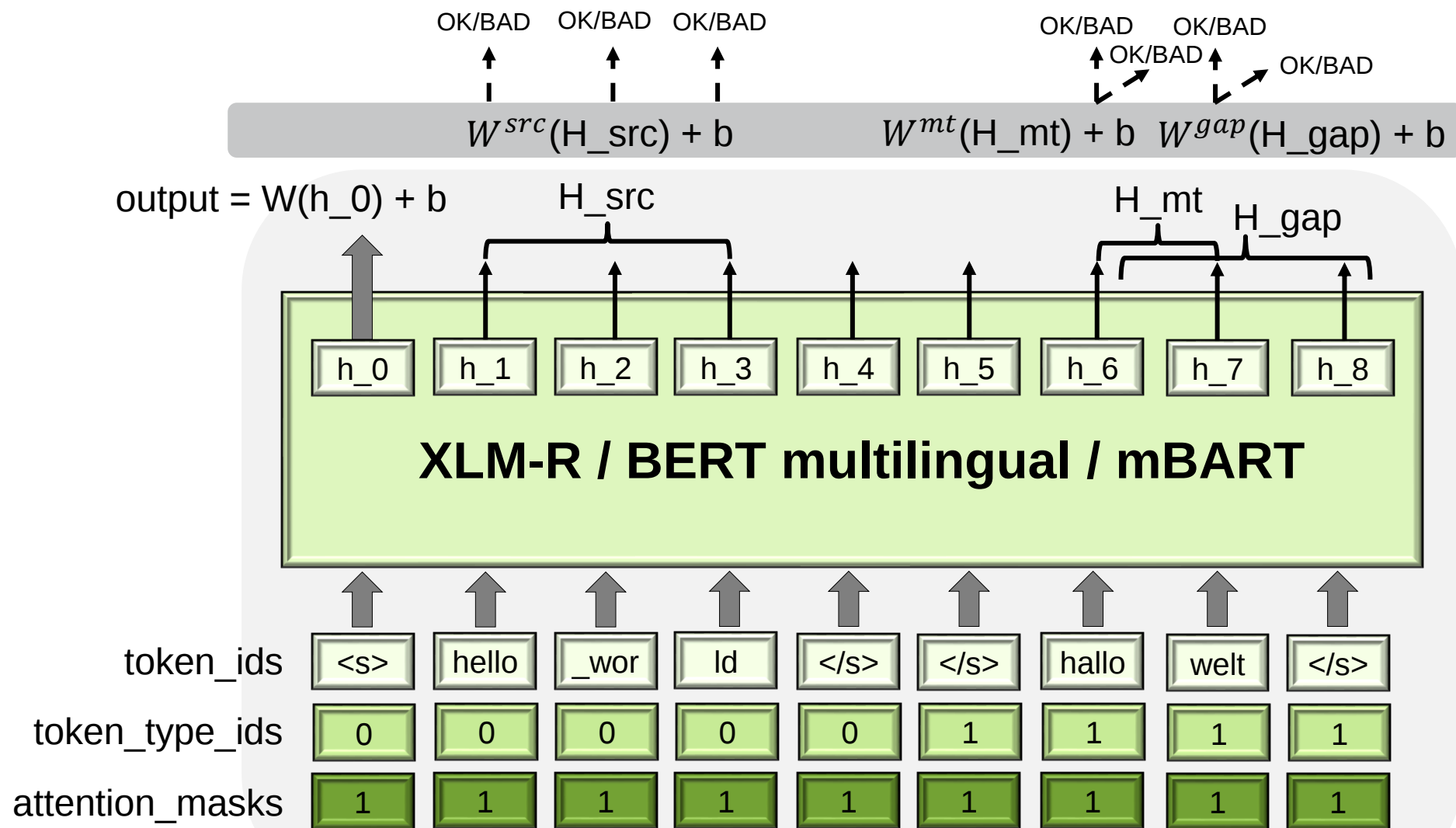
TASK1	Pearson	MAE	RMSE
3 scores			
1 scores			

2. QUALITY ESTIMATION-TASK 1

translation	scores	mean	z_scores	z_mean
He was the first to give Jefferson's oath of a...	[30, 35, 16, 14, 8, 1]	17.333333333333332	[-1.3041309086819917, -0.6301174512913427, -1....	-1.246499
These paintings represent the life and cultura...	[100, 70, 91, 100, 94, 82]	89.5	[1.5639505897683126, 0.7933273313367781, 1.707...	1.368105
As Plato presents the deathly Socrates, the gr...	[56, 40, 26, 94, 96, 98]	68.33333333333333	[-0.2388434949718787, -0.42676819663018256, -0...	0.403440
This was a cause of growing prosperity in Asia.	[62, 42, 32, 100, 89, 95]	70.0	[0.006992062038147374, -0.3454284947657185, -0...	0.472506
The original floor of the mystics has also be...	[51, 37, 21, 92, 92, 92]	64.16666666666667	[-0.44370645914690043, -0.5487777494268786, -0...	0.260778
...	...	...	...	...
Possessing more attention than buildings for d...	[41, 11, 11, 70, 88, 79]	50.0	[-0.853432387496944, -1.6061938736649113, -1.3...	-0.173618
Some say this can be caused by alcohol, toxic...	[46, 32, 16, 93, 91, 89]	61.166666666666664	[-0.6485694233219221, -0.7521270040880388, -1....	0.174589

TASK1	Pearson	MAE	RMSE
2000개			
1000개			
500개			
200개			

3. QUALITY ESTIMATION-TASK2 (SENT)

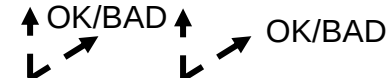


3. QUALITY ESTIMATION-TASK2 (SENT)

OK/BAD OK/BAD OK/BAD



OK/BAD OK/BAD



$W^{src}(H_{src}) + b$

$W^{mt}(H_{mt}) + b$ $W^{gap}(H_{gap}) + b$

English-German

	Team	Codalab username	Run	Pearson r ▼	MAE	RMSE
1	• HW-TSC	yuriak		0.758	0.099	0.133
2	Bering Lab	djlee_beringlab		0.723	0.107	0.140
3	NiuTrans	NiuTrans		0.649	0.123	0.154
4	IST and Unbabel	Joao1996		0.633	0.137	0.178
5	NJUNLP	cuiqu		0.618	0.129	0.160
6	NICT Kyoto	Raphael_NICT		0.615	0.151	0.197
7	Bergamot	Lisalan		0.613	0.130	0.160
8	IST and Unbabel	ddk		0.531	0.138	0.180
9	TransQuest	tharindu		0.499	0.149	0.184
10	nc	zhanghuimeng		0.494	0.163	0.198
11	nc	erickrf		0.442	0.146	0.183
12	nc	DexinWang		0.402	0.155	0.196
13	Organisers	fblain	BASELINE	0.392	0.150	0.190

TASK2	Pearson	MAE	RMSE
XLM-R			
mBERT			
mBERT			

attention_mas

3. QUALITY ESTIMATION-TASK2 (SENT)

OK/BAD OK/BAD OK/BAD

OK/BAD OK/BAD

OK/BAD OK/BAD OK/BAD

English-German

	Team	Codalab username	Run	Pearson r ▼	MAE	RMSE
1	• HW-TSC	yuriak		0.758	0.099	0.133
2	Bering Lab	djlee_beringlab		0.723	0.107	0.140
3	NiuTrans	NiuTrans		0.649	0.123	0.154
4	IST and Unbabel	Joao1996		0.633	0.137	0.178
5	NJUNLP	cuiqu		0.618	0.129	0.160
6	NICT Kyoto	Raphael_NICT		0.615	0.151	0.197
7	Bergamot	Lisalan		0.613	0.130	0.160
8	IST and Unbabel	ddk		0.531	0.138	0.180
9	TransQuest	tharindu		0.499	0.149	0.184
10	nc	zhanghuimeng		0.494	0.163	0.198
11	nc	erickrf		0.442	0.146	0.183
12	nc	DexinWang		0.402	0.155	0.196
13	Organisers	fblain	BASELINE	0.392	0.150	0.190

TASK2	Pearson	MAE	RMSE
XLM-R			
mBERT			
mBERT			

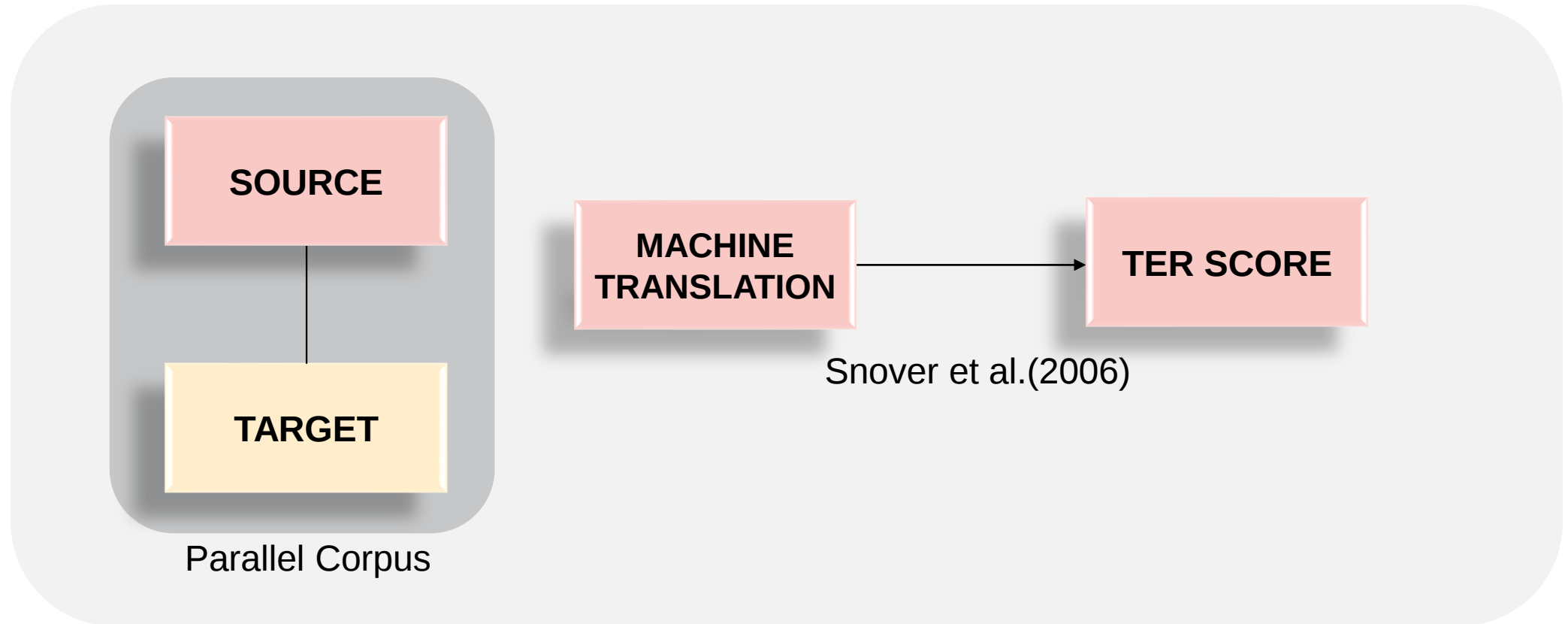
lr = 2e-5

TASK2	Pearson	MAE	RMSE	Grad Acc
XLM-R				
XLM-R				
mBERT				
mBERT				
mBART				
mBART				

attention_mas

3. QUALITY ESTIMATION-TASK2 (SENT)

- Parallel corpus : WMT2020 dataset



3. QUALITY ESTIMATION-TASK2 (SENT)

- Parallel corpus : WMT2020 dataset

$lr = 2e-5$

TASK2	Pearson	MAE	RMSE	Grad Acc
XLM-R				
XLM-R				
mBERT				
mBERT				
mBART				
mBART				

Data: 2M Pretrain: 1 epoch

TASK1	Pearson	MAE	RMSE
XLM-R (+)			

3. QUALITY ESTIMATION-TASK2 (SENT)

- Data Filtering

TASK2	Epoch	Pearson	MAE	RMSE
XLM-R				
above mean				
below mean				
≤ 0.8				
≤ 0.9				

4. QUALITY ESTIMATION-TASK2 (WORD)

[SRC sentence] José Ortega y Gasset visited Husserl at Freiburg in 1934.

<s>	_José	_Ort	ega	_y	_G	asset	_Visite	d	_Hus	ser	l	_at	...
-----	-------	------	-----	----	----	-------	---------	---	------	-----	---	-----	-----

Token_ids

0	17151	12426	2765	113	527	110896	36997	71	20387	2189	141	99	...
---	-------	-------	------	-----	-----	--------	-------	----	-------	------	-----	----	-----

src_labels

													...
--	--	--	--	--	--	--	--	--	--	--	--	--	-----

[MT sentence] 1934 besuchte José Ortega y Gasset Husserl in Freiburg.

<s>	_1934	_besucht	e	_José	_Ort	Ega	_y	_G	Asset	_Hus	Ser	l	...
-----	-------	----------	---	-------	------	-----	----	----	-------	------	-----	---	-----

Token_ids

2	58020	138438	13	17151	12426	2765	113	527	110896	20307	2189	141	...
---	-------	--------	----	-------	-------	------	-----	-----	--------	-------	------	-----	-----

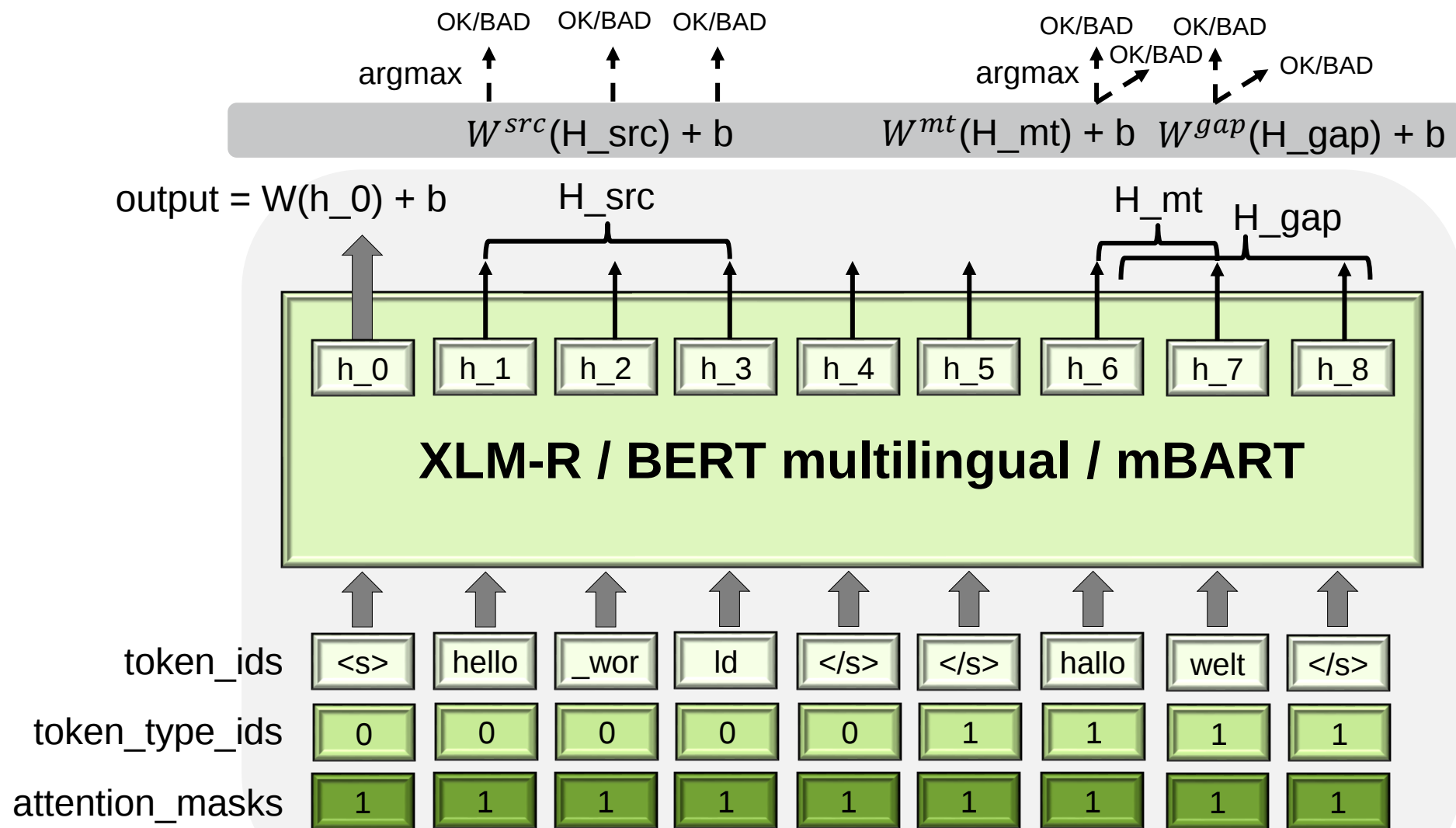
mt_labels

													...
--	--	--	--	--	--	--	--	--	--	--	--	--	-----

gap_labels

													...
--	--	--	--	--	--	--	--	--	--	--	--	--	-----

4. QUALITY ESTIMATION-TASK2 (WORD)



4. QUALITY ESTIMATION-TASK2 (WORD)



output = W'

TASK2-WORD	Target F1-BAD	Target F1-OK	Source F1-BAD	Source F1-OK
XLM-Roberta				

token_i

token_type_i

attention_mas

5. QUALITY ESTIMATION_KO-EN

- Parallel corpus: AI HUB, OpenSubtitles [2.1M]
- Data Augmentation

Machine Translation	Target	TER
Then you should go home quickly, it's not a problem to have a company dinner now.	Then you should quickly go home, the company meal is not important right now.	0.8571428571428571
The win-win fund is like handing over what the government should do to companies.	The win-win fund is like unloading the government's job to businesses.	0.6363636363636364
The win-win fund is like handing over what the government should do to companies.	I asked him already, but I think I should ask the accounting team again.	0.35714285714285715
People who make me spend money call me "salary."	People who cost me money usually call me "Umm umm umm"	0.5454545454545454
I work hard on home care, but I guess that's not enough.	I tried to take care of it home but guess that wasn't enough.	0.7692307692307693

6. SUMMARY

- TASK1
 - Ensemble
 - Data augmentation
- TASK2 – SENT
 - Data augmentation using parallel corpus
 - MT/SRC switching
 - Corpus filtering
- TASK2 – WORD
- KO-EN Quality Estimation
 - Pseudo quality estimation dataset
 - Corpus filtering

ANY QUESTIONS?

감사합니다.