On the Impact of Various Types of Noise on Neural Machine Translation

Huda Khayrallah & Philipp Koehn

{huda, phi}@jhu.edu

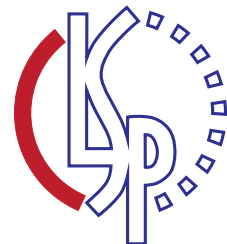This talk was presented at WNMT at ACL2018

It is based on this paper:

http://aclweb.org/anthology/W18-2705

bib:

http://aclweb.org/anthology/W18-2705.bib

On the Impact of
Various Types of Noise on
Neural Machine Translation

Huda Khayrallah & Philipp Koehn

{huda, phi}@jhu.edu

More data is better!
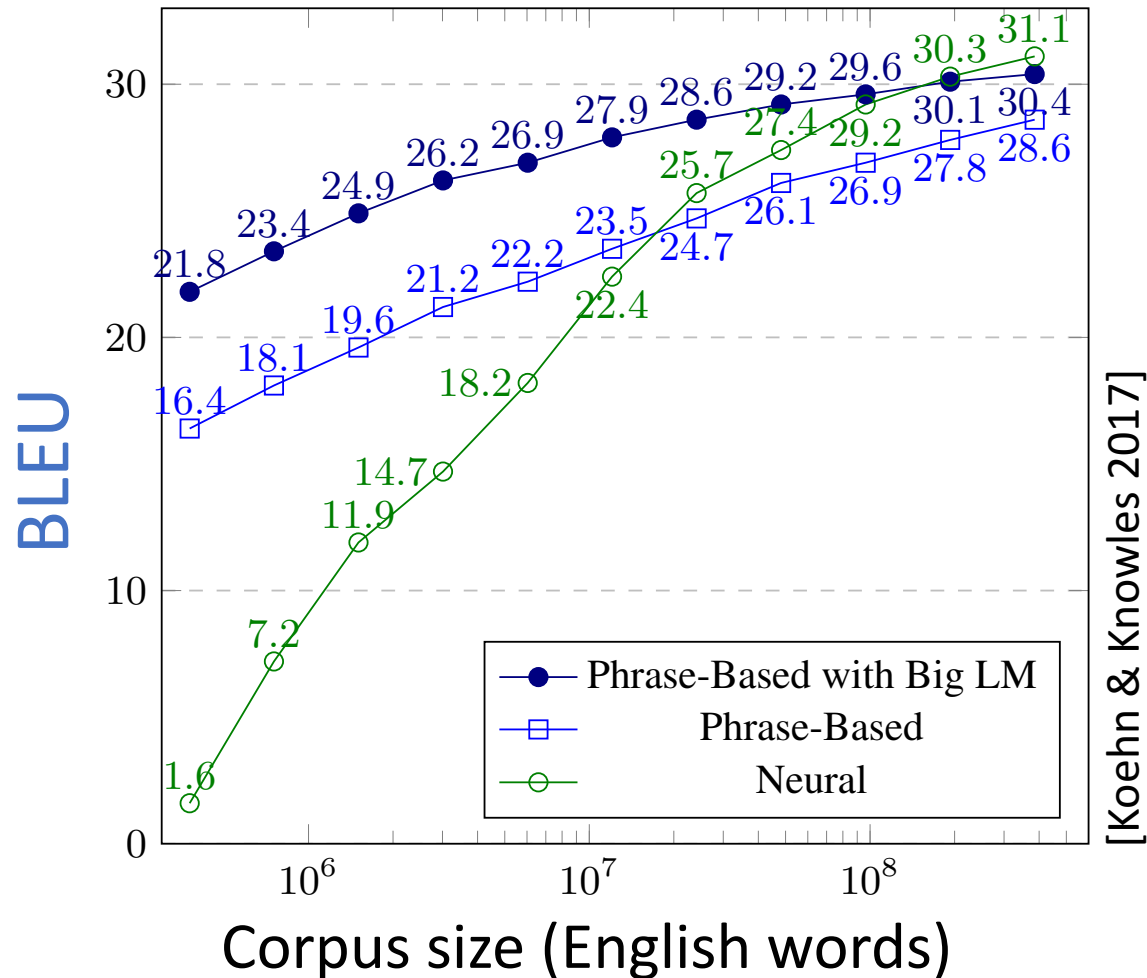
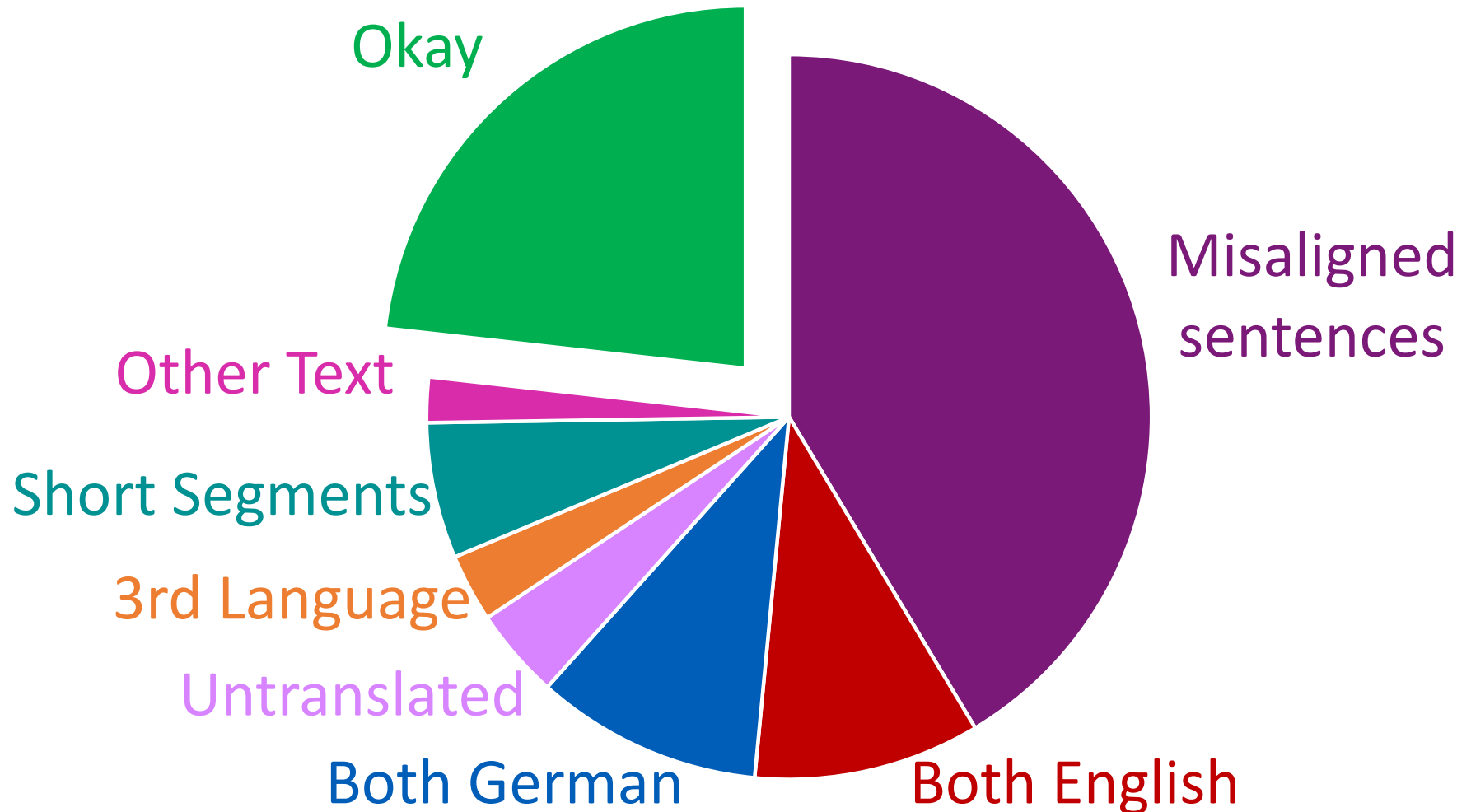|	NMT	SMT
WMT17	27.2	24.0
+ noisy corpus	17.3 (-9.9)	25.2 (+1.2)

JOHNS HOPKINS
U N I V E R S I T Y

Noisy Corpus

5%		10%		20%		50%		100%	
27.4	24.2	26.6	24.2	24.7	24.4	20.9	24.8	17.3	25.2
+0.2	+0.2	-0.6	+0.2	-2.5	+0.4	-6.3	+0.8	-9.9	+1.2

NMT **SMT**

Manual Analysis

Noise Types

- Misaligned Sentences
- Misordered words
- Wrong Language
- Untranslated Sentences
- Short Segments

Misaligned Sentences

Misaligned Sentences

Die Koalas sind süß   The koalas are cute

Die Kängurus springen   The kangaroos jump

Der Koala ist weich   The koala is soft

Das Känguru ist schnell   The kangaroo is fast

Misaligned Sentences

Die Koalas sind süß The kangaroos jump

Die Kängurus springen The koala is soft

Der Koala ist weich The kangaroo is fast

Das Känguru ist schnell The koalas are cute

Misaligned Sentences

5%		10%		20%		50%		100%	
26.5	24.0	26.5	24.0	26.3	23.9	26.1	23.9	25.3	23.4
-0.7	-0.0	-0.7	-0.0	-0.9	-0.1	-1.1	-0.1	-1.9	-0.6

NMT   **SMT**

Misordered Words

Misordered Words (source)

Die Koalas sind süß   The koalas are cute

Die Kängurus springen   The kangaroos jump

Der Koala ist weich   The koala is soft

Das Känguru ist schnell   The kangaroo is fast

JOHNS HOPKINS
UNIVERSITY

Misordered Words (source)

Koalas Die sind süß   The koalas are cute

Kängurus springen Die   The kangaroos jump

ist Der weich Koala   The koala is soft

schnell Känguru ist Das   The kangaroo is fast

Misordered Words (source)

	5%		10%		20%		50%		100%	
	26.9	24.0	26.6	23.6	26.4	23.9	26.6	23.6	25.5	23.7
	-0.3	-0.0	-0.6	-0.4	-0.8	-0.1	-0.6	-0.4	-1.7	-0.3

NMT   **SMT**

JOHNS HOPKINS
UNIVERSITY

Misordered Words (target)

Die Koalas sind süß   The koalas are cute

Die Kängurus springen   The kangaroos jump

Der Koala ist weich   The koala is soft

Das Känguru ist schnell   The kangaroo is fast

JOHNS HOPKINS
UNIVERSITY

Misordered Words (target)

Die Koalas sind süß   koalas cute are The

Die Kängurus springen   kangaroos The jump

Der Koala ist weich   is The soft koala

Das Känguru ist schnell   fast The is kangaroo

Misordered Words (target)

5%		10%		20%		50%		100%	
27.0	24.0	26.8	24.0	26.4	23.4	26.7	23.2	26.1	22.9
-0.2	-0.0	-0.4	-0.0	-0.8	-0.6	-0.5	-0.8	-1.1	-1.1

NMT **SMT**

JOHNS HOPKINS
UNIVERSITY

Wrong Language

Wrong Language (French source)

Die Koalas sind süß   The koalas are cute

Die Kängurus springen   The kangaroos jump

Der Koala ist weich   The koala is soft

Das Känguru ist schnell   The kangaroo is fast

Wrong Language (French source)

Les koalas sont mignons	The koalas are cute
Les kangourous sautent	The kangaroos jump
Le koala est doux	The koala is soft
Le kangourou est rapide	The kangaroo is fast

JOHNS HOPKINS
UNIVERSITY

Wrong Language (French source)

5%		10%		20%		50%		100%	
26.9	24.0	26.8	23.9	26.8	23.9	26.8	23.9	26.8	23.8
-0.3	-0.0	-0.4	-0.1	-0.4	-0.1	-0.4	-0.1	-0.4	-0.2

NMT **SMT**

JOHNS HOPKINS
UNIVERSITY

Wrong Language (French target)

Die Koalas sind süß   The koalas are cute

Die Kängurus springen   The kangaroos jump

Der Koala ist weich   The koala is soft

Das Känguru ist schnell   The kangaroo is fast

Wrong Language (French target)

Die Koalas sind süß   Les koalas sont mignons

Die Kängurus springen   Les kangourous sautent

Der Koala ist weich   Le koala est doux

Das Känguru ist schnell   Le kangourou est rapide

JOHNS HOPKINS UNIVERSITY

Wrong Language (French target)

5%		10%		20%		50%		100%	
26.7	24.0	26.6	23.9	26.7	23.8	26.2	23.5	25.0	23.4
-0.5	-0.0	-0.6	-0.1	-0.5	-0.2	-1.0	-0.5	-2.2	-0.6

NMT **SMT**

JOHNS HOPKINS
UNIVERSITY

Untranslated

Untranslated (English Source)

Die Koalas sind süß	The koalas are cute
Die Kängurus springen	The kangaroos jump
Der Koala ist weich	The koala is soft
Das Känguru ist schnell	The kangaroo is fast

Untranslated (English source)

The koalas are cute   The koalas are cute

The kangaroos jump   The kangaroos jump

The koala is soft   The koala is soft

The kangaroo is fast   The kangaroo is fast

Untranslated (English source)

5%		10%		20%		50%		100%	
27.2	23.9	27.0	23.9	26.7	23.6	26.8	23.7	26.9	23.5
-0.0	-0.1	-0.2	-0.1	-0.5	-0.4	-0.4	-0.3	-0.3	-0.5

NMT **SMT**

JOHNS HOPKINS
UNIVERSITY

Khayrallah & Koehn

28

Untranslated (German target)

Die Koalas sind süß   The koalas are cute

Die Kängurus springen   The kangaroos jump

Der Koala ist weich   The koala is soft

Das Känguru ist schnell   The kangaroo is fast

Untranslated (German target)

Die Koalas sind süß

Die Kängurus springen

Der Koala ist weich

Das Känguru ist schnell

Die Koalas sind süß

Die Kängurus springen

Der Koala ist weich

Das Känguru ist schnell

Untranslated (German target)

	5%	10%	20%	50%	100%
	17.6 23.8	11.2 23.9	5.6 23.8	3.2 23.4	3.2 21.1
	-0.2	-0.1	-0.2	-0.6	-2.9
	-9.8	-16.0	-21.6	-24.0	-24.0

NMT   SMT

JOHNS HOPKINS
UNIVERSITY
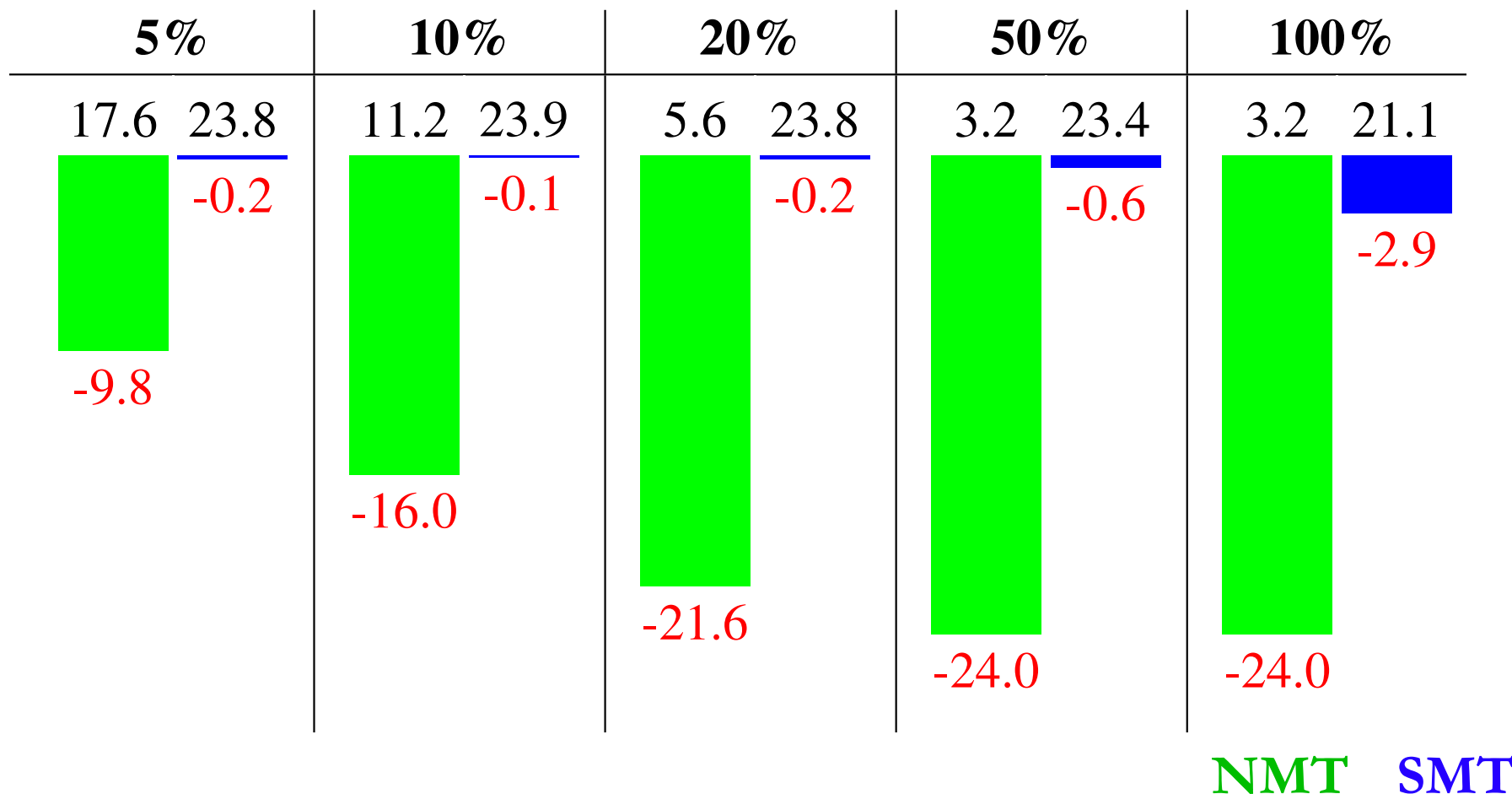
Short Segments

Short Segments

Die

süß

Känguru

schnell

The

cute

Kangaroo

fast

Short Segments

≤ 2 words

	5%		10%		20%	
	27.1	24.1	26.5	23.9	26.7	23.8
	-0.1	+0.1	-0.7	-0.1	-0.5	-0.2

3-5 words

	5%		10%		20%		50%	
	27.8	24.2	27.6	24.5	28.0	24.5	26.6	24.2
	+0.6	+0.2	+0.4	+0.5	+0.8	+0.5	-0.6	+0.2

	5%		10%		20%		50%		100%	
MISALIGNED SENTENCES	26.5 -0.7	24.0 -0.0	26.5 -0.7	24.0 -0.0	26.3 -0.9	23.9 -0.1	26.1 -1.1	23.9 -0.1	25.3 -1.9	23.4 -0.6
MISORDERED WORDS (SOURCE)	26.9 -0.3	24.0 -0.0	26.6 -0.6	23.6 -0.4	26.4 -0.8	23.9 -0.1	26.6 -0.6	23.6 -0.4	25.5 -1.7	23.7 -0.3
MISORDERED WORDS (TARGET)	27.0 -0.2	24.0 -0.0	26.8 -0.4	24.0 -0.0	26.4 -0.8	23.4 -0.6	26.7 -0.5	23.2 -0.8	26.1 -1.1	22.9 -1.1
WRONG LANGUAGE (FRENCH SOURCE)	26.9 -0.3	24.0 -0.0	26.8 -0.4	23.9 -0.1	26.8 -0.4	23.9 -0.1	26.8 -0.4	23.9 -0.1	26.8 -0.4	23.8 -0.2
WRONG LANGUAGE (FRENCH TARGET)	26.7 -0.5	24.0 -0.0	26.6 -0.6	23.9 -0.1	26.7 -0.5	23.8 -0.2	26.2 -1.0	23.5 -0.5	25.0 -2.2	23.4 -0.6
UNTRANSLATED (ENGLISH SOURCE)	27.2 -0.0	23.9 -0.1	27.0 -0.2	23.9 -0.1	26.7 -0.5	23.6 -0.4	26.8 -0.4	23.7 -0.3	26.9 -0.3	23.5 -0.5
UNTRANSLATED (GERMAN TARGET)	17.6 -9.8	23.8 -0.2	11.2 -16.0	23.9 -0.1	5.6 -21.6	23.8 -0.2	3.2 -24.0	23.4 -0.6	3.2 -24.0	21.1 -2.9
SHORT SEGMENTS (max 2)	27.1 -0.1	24.1 +0.1	26.5 -0.7	23.9 -0.1	26.7 -0.5	23.8 -0.2				
SHORT SEGMENTS (max 5)	27.8 +0.6	24.2 +0.2	27.6 +0.4	24.5 +0.5	28.0 +0.8	24.5 +0.5	26.6 -0.6	24.2 +0.2		
RAW CRAWL DATA	27.4 +0.2	24.2 +0.2	26.6 -0.6	24.2 +0.2	24.7 -2.5	24.4 +0.4	20.9 -6.3	24.8 +0.8	17.3 -9.9	25.2 +1.2

Existing filtering methods

- BiCleaner [Miquel Espla-Gomis and M Forcada 2009]
- Zipporah [Xu & Koehn 2017]

Existing filtering methods
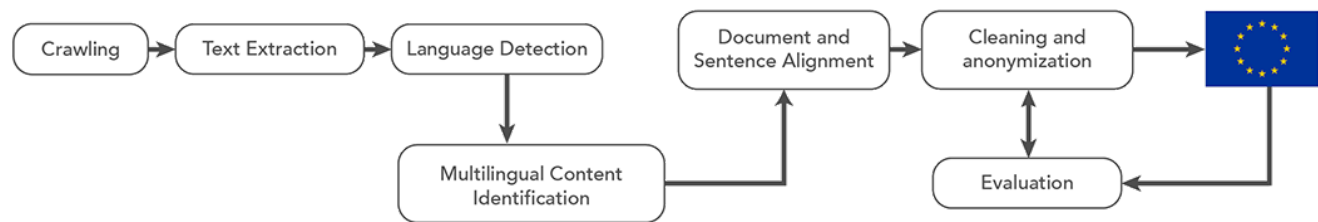
- BiCleaner [Miquel Espla-Gomis and M Forcada 2009]
- Zipporah [Xu & Koehn 2017]
- WMT shared task

Questions?

	5%		10%		20%		50%		100%	
MISALIGNED SENTENCES	26.5	24.0	26.5	24.0	26.3	23.9	26.1	23.9	25.3	23.4
	-0.7	-0.0	-0.7	-0.0	-0.9	-0.1	-1.1	-0.1	-1.9	-0.6
MISORDERED WORDS (SOURCE)	26.9	24.0	26.6	23.6	26.4	23.9	26.6	23.6	25.5	23.7
	-0.3	-0.0	-0.6	-0.4	-0.8	-0.1	-0.6	-0.4	-1.7	-0.3
MISORDERED WORDS (TARGET)	27.0	24.0	26.8	24.0	26.4	23.4	26.7	23.2	26.1	22.9
	-0.2	-0.0	-0.4	-0.0	-0.8	-0.6	-0.5	-0.8	-1.1	-1.1
WRONG LANGUAGE (FRENCH SOURCE)	26.9	24.0	26.8	23.9	26.8	23.9	26.8	23.9	26.8	23.8
	-0.3	-0.0	-0.4	-0.1	-0.4	-0.1	-0.4	-0.1	-0.4	-0.2
WRONG LANGUAGE (FRENCH TARGET)	26.7	24.0	26.6	23.9	26.7	23.8	26.2	23.5	25.0	23.4
	-0.5	-0.0	-0.6	-0.1	-0.5	-0.2	-1.0	-0.5	-2.2	-0.6
UNTRANSLATED (ENGLISH SOURCE)	27.2	23.9	27.0	23.9	26.7	23.6	26.8	23.7	26.9	23.5
	-0.0	-0.1	-0.2	-0.1	-0.5	-0.4	-0.4	-0.3	-0.3	-0.5
UNTRANSLATED (GERMAN TARGET)	17.6	23.8	11.2	23.9	5.6	23.8	3.2	23.4	3.2	21.1
	-9.8	-0.2	-16.0	-0.1	-21.6	-0.2	-24.0	-0.6	-24.0	-2.9
SHORT SEGMENTS (max 2)	27.1	24.1	26.5	23.9	26.7	23.8				
	-0.1	+0.1	-0.7	-0.1	-0.5	-0.2				
SHORT SEGMENTS (max 5)	27.8	24.2	27.6	24.5	28.0	24.5	26.6	24.2		
	+0.6	+0.2	+0.4	+0.5	+0.8	+0.5	-0.6	+0.2		
RAW CRAWL DATA	27.4	24.2	26.6	24.2	24.7	24.4	20.9	24.8	17.3	25.2
	+0.2	+0.2	-0.6	+0.2	-2.5	+0.4	-6.3	+0.8	-9.9	+1.2

De➔En translation

- **MARIANNMT**
  - RNN encoder-decoder
  - With attention
  - BPE
  - Dropout
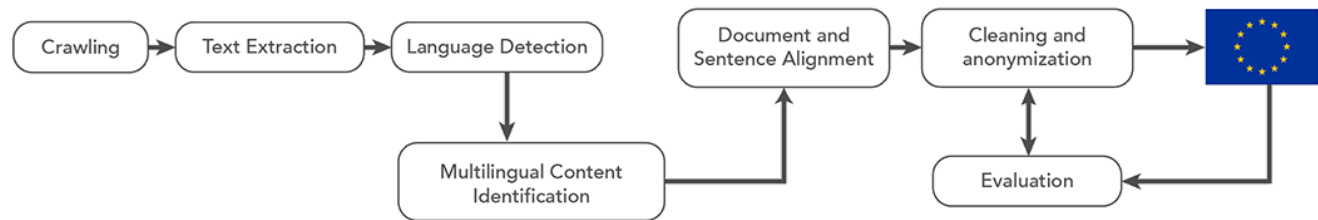 - (Shallow!)
  - …

- Moses
  - 5gram LM
  - hierarchical lexicalized reordering
  - Operation sequence model
  - MBR decoding
  - Cube pruning
  - …

Let's go get more data!

Existing filtering methods

De➔En translation

	NMT	SMT
WMT17	27.2	24.0

Insitution	NMT 10m								NMT 100m								SMT 10m								SMT 100m						
	dev (r	avg	newst	iwslt2	Acqui	EMEA	Globa	KDE	dev (r	avg	newst	iwslt2	Acqui	EMEA	Globa	KDE	dev (r	avg	newst	iwslt2	Acqui	EMEA	Globa	KDE	dev (r	avg	newst	iwslt2	Acqui	EMEA	Globa
AFRL	8.91	13.83	11.22	6.11	15.53	23.84	8.90	17.39	30.09	30.18	36.99	26.32	26.48	35.10	27.99	28.17	21.52	21.90	26.15	18.86	18.30	25.96	20.03	22.12	24.56	25.24	29.93	22.22	21.47	29.68	22.62
AFRL	27.67	27.09	33.41	23.34	25.57	29.89	25.36	24.99	30.34	30.28	37.44	25.96	26.62	35.22	28.07	28.36	22.54	23.36	27.69	20.03	20.59	26.77	21.05	24.01	24.49	25.32	29.91	22.27	21.47	29.88	22.74
AFRL	15.99	19.82	19.68	10.94	23.85	26.94	14.81	22.68	30.08	30.14	37.40	26.08	26.36	34.79	28.10	28.13	21.68	22.48	26.48	19.42	20.23	25.50	20.45	22.80	24.51	25.22	29.79	22.32	21.51	29.67	22.57
AFRL	8.59	13.49	10.93	5.57	15.34	23.73	8.54	16.85	17.84	21.05	23.27	16.83	22.91	26.23	19.00	18.05	21.52	21.93	26.18	18.94	18.32	25.97	20.05	22.14	22.21	22.89	27.06	20.02	20.93	25.77	21.16
AFRL	23.45	25.10	29.24	21.45	24.18	29.03	22.68	24.00	29.40	29.62	36.24	25.13	26.18	35.00	27.44	27.71	21.62	22.43	26.63	19.46	19.70	25.69	20.28	22.84	24.33	25.02	29.45	22.18	21.29	29.52	22.35
Alibaba	28.38	27.60	35.00	25.20	24.08	29.81	25.80	25.72	31.94	31.92	39.52	27.06	28.43	36.72	29.11	30.66	22.28	24.11	29.05	**22.16**	20.57	26.67	22.00	24.19	25.64	26.42	31.16	22.94	22.45	31.18	23.99
Alibaba	28.06	27.60	35.18	25.64	24.24	29.40	25.60	25.54	31.86	31.93	39.67	27.27	28.43	36.44	29.09	30.65	22.26	24.11	28.95	22.13	20.53	26.84	22.02	24.18	25.63	26.44	31.12	23.00	**22.51**	31.17	**24.00**
ARC	16.32	19.77	20.33	11.40	21.13	27.43	14.67	23.68	31.38	31.34	39.01	26.60	27.82	35.89	28.32	30.37	20.82	22.68	26.91	18.91	19.34	27.24	20.38	23.32	25.50	26.13	30.80	22.70	22.35	30.86	23.49
ARC	26.09	25.79	31.32	21.20	22.94	30.17	23.39	25.73	31.46	31.33	38.99	26.57	27.63	35.96	28.20	30.63	20.49	22.39	26.30	18.83	18.71	26.50	20.20	23.78	25.28	26.12	30.58	22.82	22.27	30.88	23.43
ARC	25.01	23.97	30.42	20.20	21.49	28.77	22.89	20.02	31.35	31.32	39.01	26.48	27.58	35.81	28.30	30.72	21.48	21.95	26.02	18.15	18.49	25.84	20.01	23.20	25.30	26.03	30.73	22.68	22.06	30.94	23.43
University of Tartu	25.80	25.17	31.62	21.78	21.83	28.15	24.03	23.63	30.63	30.60	38.24	26.16	27.52	35.76	28.14	27.77	20.84	22.31	26.81	19.46	18.71	24.78	20.63	23.49	24.89	25.70	30.35	22.26	21.87	30.53	23.15
DCU	3.50	3.34	3.43	1.95	1.10	4.79	3.10	5.69	4.27	4.78	4.46	2.36	1.41	7.68	4.57	8.19	5.22	6.56	6.74	3.29	2.39	10.15	6.23	10.57	12.17	13.22	14.68	10.00	9.04	16.46	12.30
DCU	4.56	3.42	4.68	1.81	1.71	5.88	3.22	3.23	7.40	8.61	7.98	5.13	5.22	12.81	8.02	12.47	9.81	12.86	14.13	9.11	9.65	16.45	11.49	16.33	17.52	18.57	21.03	14.81	14.96	22.12	17.18
DCU	5.71	5.01	5.98	2.92	3.18	8.23	4.81	4.91	17.68	18.39	21.17	12.77	14.83	24.19	17.04	20.34	12.35	15.26	17.23	11.94	11.30	18.85	13.80	18.41	20.11	21.09	24.58	17.57	17.00	24.65	19.44
DCU	6.36	6.27	7.51	3.65	5.06	10.96	5.25	5.20	17.88	18.60	20.64	13.15	15.42	24.43	17.81	20.12	12.62	15.67	17.65	12.32	11.83	19.24	14.15	18.82	20.19	21.19	24.59	17.89	17.06	24.64	19.54
JHU	25.21	25.34	31.44	23.13	22.82	26.31	24.02	24.32	29.56	30.16	36.75	24.24	27.57	35.36	27.70	29.32	20.24	22.61	26.31	20.21	19.88	24.69	20.28	24.30	24.85	25.77	30.19	22.61	22.12	29.92	23.35
JHU	25.20	25.41	31.31	23.14	22.51	26.56	24.38	24.53	29.13	29.75	36.43	23.25	27.26	35.06	27.32	29.20	20.18	22.57	26.26	20.36	19.33	24.76	20.37	24.32	24.81	25.84	30.35	22.63	22.13	30.12	23.26
LMU	25.47	22.07	31.16	21.98	16.83	23.98	23.82	14.67	28.92	29.02	36.21	25.69	24.45	33.21	27.48	27.09	21.47	23.07	27.57	20.80	17.67	26.57	21.48	24.35	24.64	25.45	29.96	22.34	21.40	29.92	23.08
LMU	26.17	23.65	31.89	22.40	18.51	27.01	24.60	17.46	29.14	29.47	36.99	25.48	25.19	33.46	27.52	28.17	21.73	23.29	28.03	20.61	17.97	26.95	21.45	24.73	24.86	25.57	30.14	22.42	21.47	30.08	23.09
LMU	26.22	23.65	31.79	22.09	18.43	27.14	24.53	17.94	29.33	29.54	36.71	25.48	25.25	34.15	27.67	27.95	21.71	23.26	28.03	20.57	17.96	26.96	21.46	24.58	24.86	25.51	30.00	22.31	21.25	30.18	23.19
LMU	23.03	22.96	28.79	21.06	16.01	26.98	23.30	21.64	30.82	30.54	37.78	25.95	25.77	35.61	28.48	29.62	19.62	21.46	25.35	19.67	15.30	25.32	20.03	23.08	25.09	25.61	30.34	22.37	20.98	30.44	23.27
MAJE	26.51	24.81	32.36	24.80	22.56	24.61	24.33	20.21	31.61	31.20	38.73	26.91	27.89	35.57	28.93	29.16	20.86	22.53	27.20	21.28	19.06	24.48	21.18	21.97	25.36	26.07	30.68	22.91	22.39	30.56	23.73
Microsoft	28.49	**28.62**	35.67	25.10	23.69	**32.72**	**26.72**	27.81	**32.41**	**32.06**	**39.85**	**27.43**	28.31	**36.70**	**29.26**	**30.79**	22.86	24.45	29.54	21.59	19.72	28.70	22.47	24.70	**25.80**	**26.50**	**31.35**	**23.17**	22.29	**31.45**	23.88
NICT	26.48	25.94	32.93	23.67	21.67	27.60	25.13	24.65	29.60	30.04	37.28	25.83	26.11	34.13	27.62	29.25	21.47	23.46	27.79	20.94	19.27	25.89	21.38	**25.51**	25.23	25.98	30.79	22.76	21.98	30.39	23.43
NRC	23.38	23.07	27.86	19.33	19.03	26.45	22.03	23.72	31.67	31.56	38.86	27.10	28.15	35.96	28.87	30.41	19.61	20.97	25.13	17.86	16.59	24.21	19.97	22.07	25.65	26.23	31.12	22.84	22.37	31.11	23.75
NRC	20.83	20.69	24.97	17.19	16.57	23.79	19.75	21.85	31.39	31.20	38.42	26.80	27.94	35.71	28.00	30.32	17.98	19.79	23.49	16.63	15.49	23.09	18.65	21.39	25.45	26.01	30.63	22.72	22.06	30.74	23.70
NRC	24.49	25.34	30.32	21.47	22.57	31.71	23.08	22.89	32.10	31.69	39.39	27.09	28.31	36.30	28.94	30.12	19.66	22.13	25.96	18.64	18.78	27.94	20.05	21.38	25.61	26.25	31.11	22.84	22.19	31.20	23.67
NRC	27.61	27.41	33.93	24.37	23.20	29.75	25.44	27.75	31.97	31.88	39.59	26.95	28.35	36.59	29.09	30.70	22.03	23.89	28.72	21.34	19.66	26.35	22.06	25.21	25.76	26.40	31.35	22.80	22.36	31.11	23.84
NRC	26.62	26.50	32.72	23.89	22.22	28.55	24.83	26.81	31.53	31.76	39.30	27.13	27.91	36.28	29.01	30.92	21.34	23.52	27.97	21.12	19.26	26.00	21.79	24.99	25.63	26.37	31.04	23.16	22.46	30.83	23.93
Prompsit	25.64	25.57	31.09	22.40	21.79	30.04	23.25	24.85	31.81	31.69	39.36	26.97	28.08	36.57	28.60	30.55	19.98	22.75	26.04	19.94	19.11	27.03	20.05	24.31	25.78	26.41	31.19	22.82	22.47	31.29	23.77
Prompsit	20.75	19.45	26.50	20.20	18.90	17.38	19.53	14.19	31.96	31.83	39.46	27.27	**28.36**	36.58	28.95	30.38	19.10	21.27	25.40	19.49	16.86	23.23	19.32	23.31	25.62	26.27	31.12	22.76	22.48	31.00	23.64
Prompsit	20.76	19.32	26.08	20.00	18.78	17.30	19.76	14.00	31.95	31.71	39.75	26.73	28.30	36.40	29.07	30.02	18.62	20.09	24.95	19.42	15.91	19.70	18.59	21.95	25.63	26.16	31.02	22.86	22.22	30.86	23.52
Prompsit	25.95	26.05	31.58	20.81	22.15	31.24	23.73	26.79	31.53	31.67	39.16	26.73	28.23	36.35	28.68	30.84	20.80	22.94	27.03	19.03	19.02	27.35	20.57	24.62	25.40	26.30	30.97	22.81	22.47	31.06	23.64
RWTH	28.29	26.65	34.77	25.00	24.40	27.70	25.92	22.11	30.92	31.05	38.59	26.94	27.46	35.44	29.00	28.85	22.36	23.85	28.62	21.79	21.01	26.79	22.03	22.83	25.32	25.91	30.67	22.89	21.98	30.18	23.47
RWTH	**29.44**	28.00	**36.04**	25.22	25.24	31.06	26.72	23.71	31.20	31.20	38.79	26.69	27.67	36.09	28.69	29.29	**23.23**	24.53	29.57	21.84	21.35	27.96	**22.67**	23.76	25.68	26.18	30.79	23.16	22.20	30.95	23.38
RWTH	29.31	28.01	35.95	**25.23**	**25.30**	31.09	26.62	23.86	31.46	31.29	39.22	26.54	27.43	36.29	28.70	29.56	23.14	**24.58**	**29.59**	21.76	**21.45**	28.12	22.63	**23.93**	25.57	26.21	30.79	23.11	22.06	30.90	23.64
Speechmatics	27.32	27.88	33.97	24.60	24.74	30.93	24.99	28.04	30.55	31.00	37.77	26.52	27.95	35.44	28.25	30.06	21.54	23.78	28.17	21.03	19.74	27.55	21.49	24.69	25.18	25.84	30.27	22.64	22.00	30.54	23.29
Speechmatics	27.30	27.97	34.72	24.07	24.42	31.40	24.95	**28.25**	30.32	30.84	37.36	26.47	27.83	35.12	28.17	30.08	21.55	23.88	28.23	20.53	19.58	**28.28**	21.41	25.26	25.08	25.85	30.35	22.46	21.86	30.72	23.34
Speechmatics	18.00	17.99	21.83	15.64	13.14	21.27	17.65	18.43	28.55	30.04	35.17	25.79	26.92	35.12	27.42	29.82	16.93	18.10	20.43	15.08	13.56	22.20	16.30	21.02	24.80	25.80	30.28	22.49	21.94	30.61	23.24
Systran	24.12	24.30	29.60	21.31	19.04	28.27	22.96	24.60	28.65	29.91	36.31	25.13	26.24	35.11	26.85	29.79	19.58	21.83	25.38	19.37	16.70	25.68	19.93	23.90	24.56	25.44	30.00	22.27	21.50	30.07	22.66
Tilde	26.22	26.56	32.40	22.06	22.08	31.33	24.38	27.13	31.40	31.22	38.57	26.78	27.54	36.57	28.22	29.62	21.02	23.03	27.27	19.79	18.32	27.69	21.02	24.08	25.18	25.97	30.65	22.77	21.94	30.89	23.35
Tilde	22.59	23.55	27.99	19.49	17.85	28.53	22.33	25.12	31.26	31.24	38.61	26.39	27.28	36.26	28.58	30.32	19.04	21.45	25.00	18.23	16.57	25.60	19.73	23.59	25.24	26.19	30.82	22.80	22.14	31.14	23.63
Tilde	21.62	22.60	26.59	18.85	16.80	27.38	21.76	24.22	30.81	30.75	38.02	25.76	26.66	35.74	27.89	30.44	18.33	20.99	24.34	17.39	16.24	25.08	19.39	23.50	25.03	25.91	30.56	22.46	21.96	30.77	23.25
UTFPR	10.33	11.44	13.20	7.77	10.43	17.53	9.93	9.80	8.61	11.88	10.47	6.83	11.69	18.21	10.11	13.94	16.11	17.55	20.51	14.67	14.04	20.99	16.09	18.97	19.45	20.67	23.68	18.16	17.10	23.90	18.85
UTFPR	22.19	21.75	27.21	18.53	18.60	24.86	19.21	22.09	19.25	22.23	25.00	16.67	19.15	28.83	19.70	24.05	19.75	20.81	25.09	18.58	16.16	23.74	19.12	22.18	21.74	22.35	26.48	20.17	17.36	26.04	20.53
UTFPR	5.57	6.57	6.54	2.93	4.19	11.52	5.89	8.34	4.16	6.15	4.66	2.13	3.50	12.32	5.00	9.29	11.69	13.22	14.91	9.90	10.57	16.83	12.12	14.98	15.74	16.96	18.66	14.43	13.91	20.40	15.15
Vicomtech	25.67	26.35	32.33	22.57	22.55	28.96	24.28	27.39	29.79	30.40	37.08	26.35	26.81	34.54	27.74	29.89	21.07	23.25	27.48	20.42	19.33	26.51	21.20	24.55	24.93	25.91	30.47	22.47	22.16	30.30	23.43
Vicomtech	25.31	25.64	31.25	21.81	20.67	29.09	23.48	27.56	21.14	24.91	27.23	22.44	23.15	26.92	22.94	26.76	21.27	23.29	27.52	19.80	19.33	26.84	21.12	25.14	25.13	25.80	30.17	22.39	22.12	30.03	23.36

On the Impact of
Various Types of Noise on
Neural Machine Translation

Huda Khayrallah & Philipp Koehn

{huda, phi}@jhu.edu