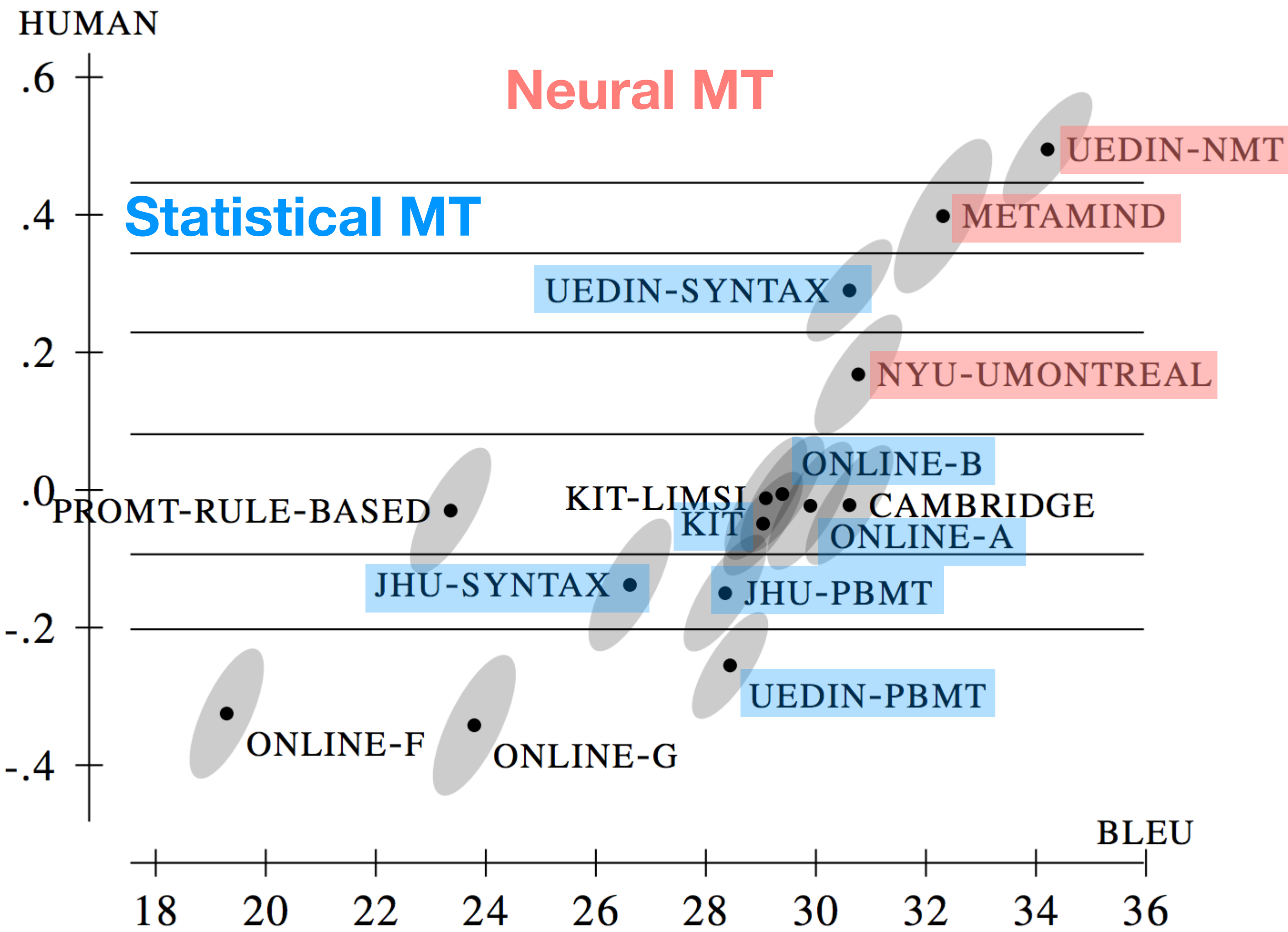
Current Challenges

Philipp Koehn

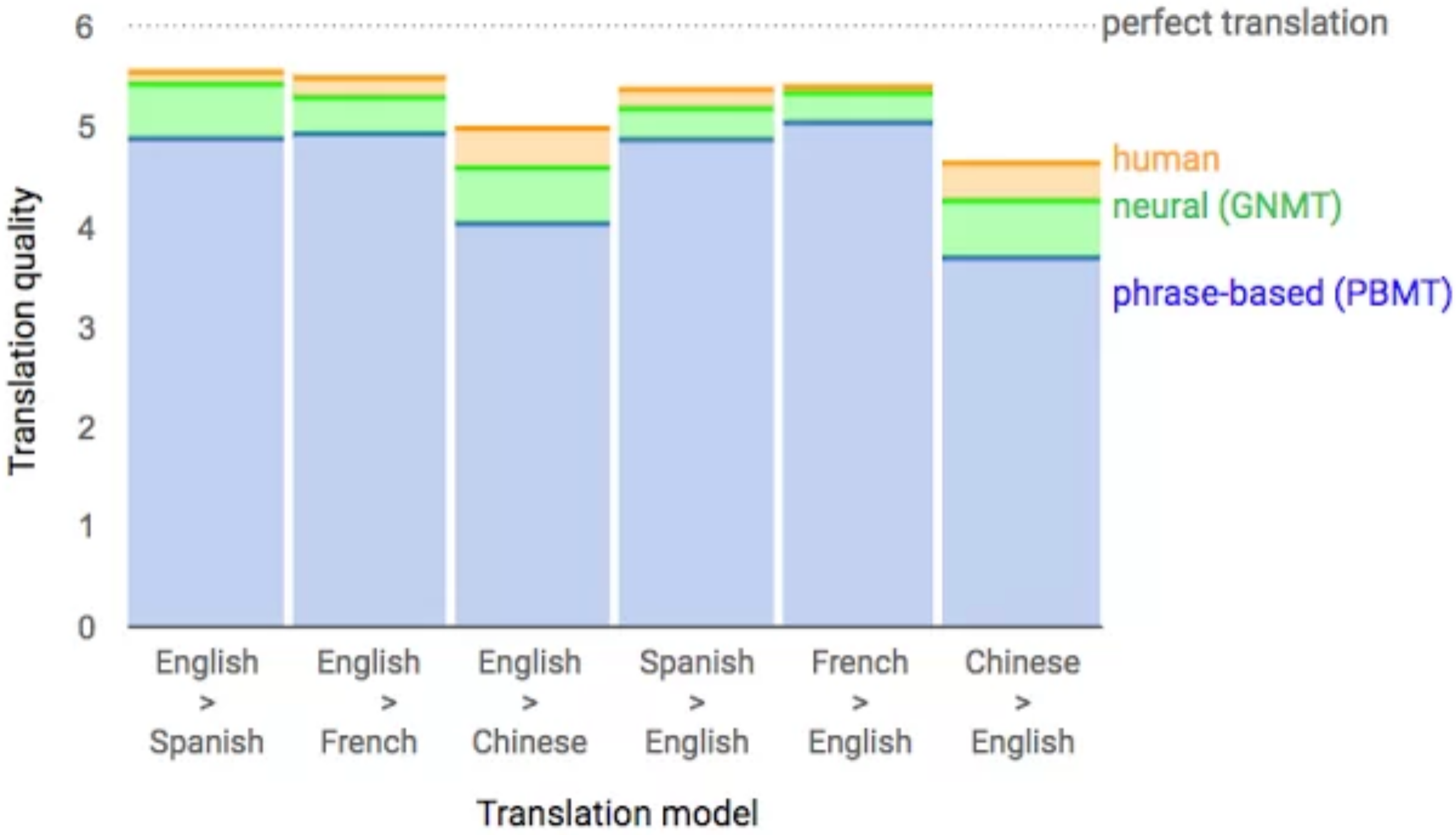
6 November 2018

WMT 2016



(in 2017 barely any statistical machine translation submissions)

2017: Google: "Near Human Quality"



2018: More Hype

Microsoft Research Achieves Human Parity For Chinese English Translation

Written by Sue Gee

Wednesday, 21 March 2018

Researchers in Microsoft's labs in Beijing and in Redmond and Washington have developed an AI machine translation system that can translate with the same accuracy as a human from Chinese to English.

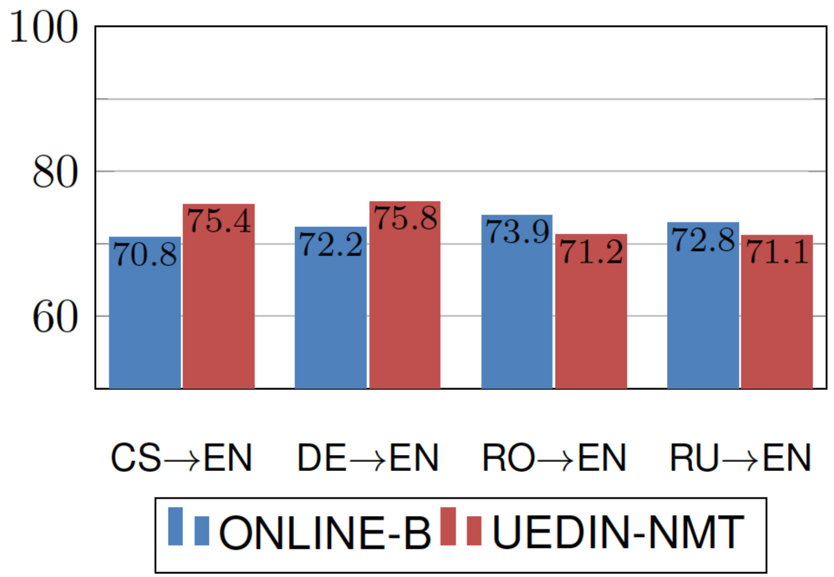
SDL Cracks Russian to English Neural Machine Translation

Global Enterprises to Capitalize on Near Perfect Russian to English Machine Translation as SDL Sets New Industry Standard

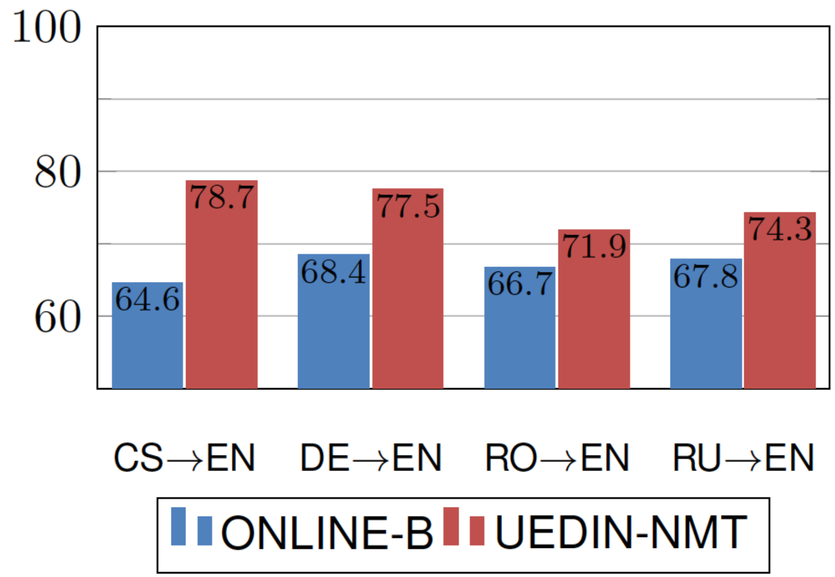
"90% of the system's output labelled as perfect by professional Russian-English translators"

Just Better Fluency?

Adequacy
+1%



Fluency
+13%



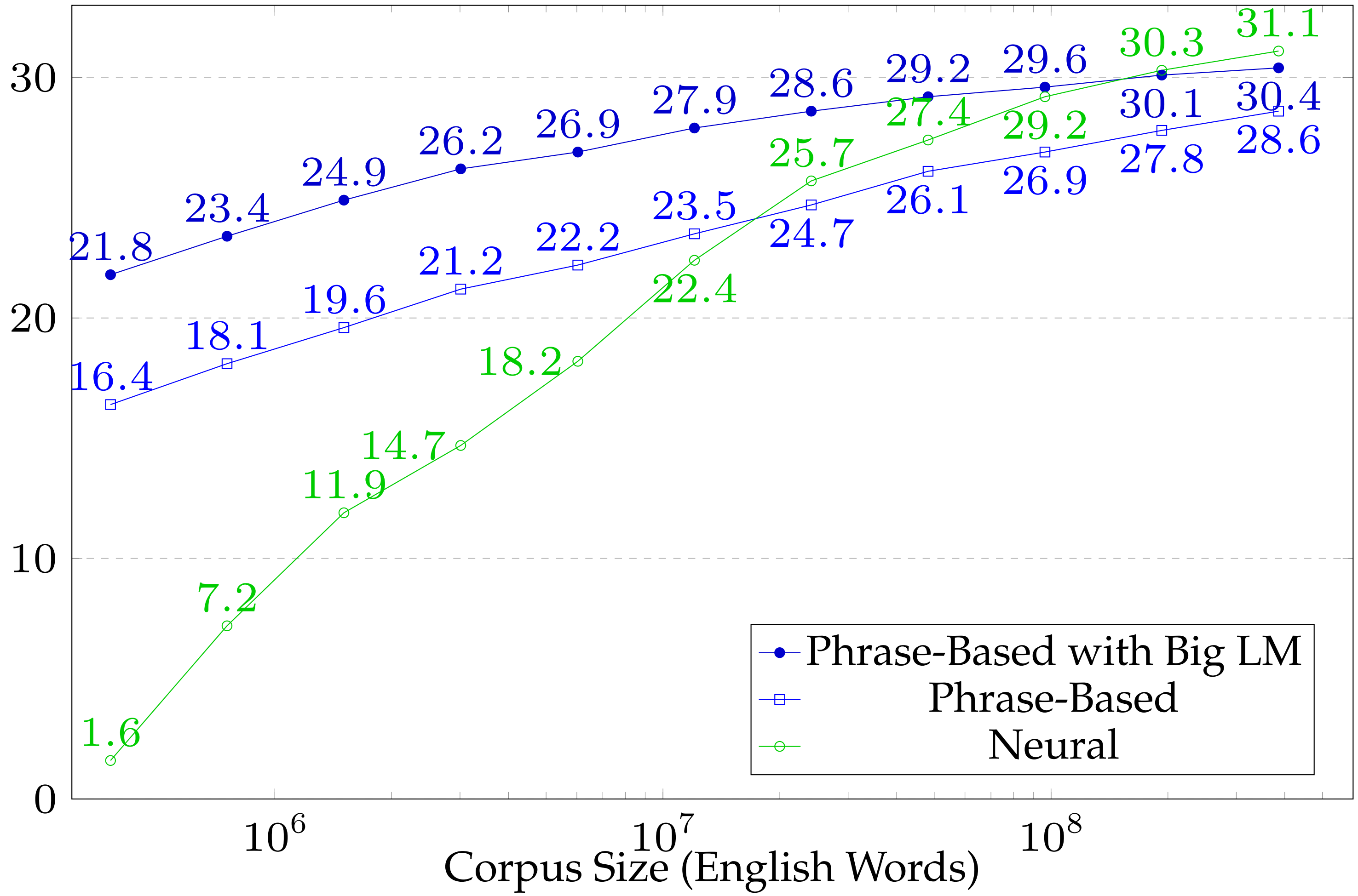
(from: Sennrich and Haddow, 2017)

Challenges

- Lack of training data
- Domain mismatch
- Rare words
- Sentence length
- Word alignment
- Beam search
- Noise

lack of training data

Amount of Training Data



English-Spanish systems trained on 0.4 million to 385.7 million words

Translation Examples

Source	A Republican strategy to counter the re-election of Obama
$\frac{1}{1024}$	Un órgano de coordinación para el anuncio de libre determinación
$\frac{1}{512}$	Lista de una estrategia para luchar contra la elección de hojas de Ohio
$\frac{1}{256}$	Explosión realiza una estrategia divisiva de luchar contra las elecciones de autor
$\frac{1}{128}$	Una estrategia republicana para la eliminación de la reelección de Obama
$\frac{1}{64}$	Estrategia siria para contrarrestar la reelección del Obama .
$\frac{1}{32}+$	Una estrategia republicana para contrarrestar la reelección de Obama

domain mismatch

Domain Mismatch

System ↓	Law	Medical	IT	Koran	Subtitles
All Data	30.5 32.8	45.1 42.2	35.3 44.7	17.9 17.9	26.4 20.8
Law	31.1 34.4	12.1 18.2	3.5 6.9	1.3 2.2	2.8 6.0
Medical	3.9 10.2	39.4 43.5	2.0 8.5	0.6 2.0	1.4 5.8
IT	1.9 3.7	6.5 5.3	42.1 39.8	1.8 1.6	3.9 4.7
Koran	0.4 1.8	0.0 2.1	0.0 2.3	15.9 18.8	1.0 5.5
Subtitles	7.0 9.9	9.3 17.8	9.2 13.6	9.0 8.4	25.9 22.1

Translation Examples

Source	Schaue um dich herum.
Ref.	Look around you.
All	NMT: Look around you. SMT: Look around you.
Law	NMT: Sughum gravecorn. SMT: In order to implement dich Schaue .
Medical	NMT: EMEA / MB / 049 / 01-EN-Final Work progamme for 2002 SMT: Schaue by dich around .
IT	NMT: Switches to paused. SMT: To Schaue by itself . \t \t
Koran	NMT: Take heed of your own souls. SMT: And you see.
Subtitles	NMT: Look around you. SMT: Look around you .

rare words

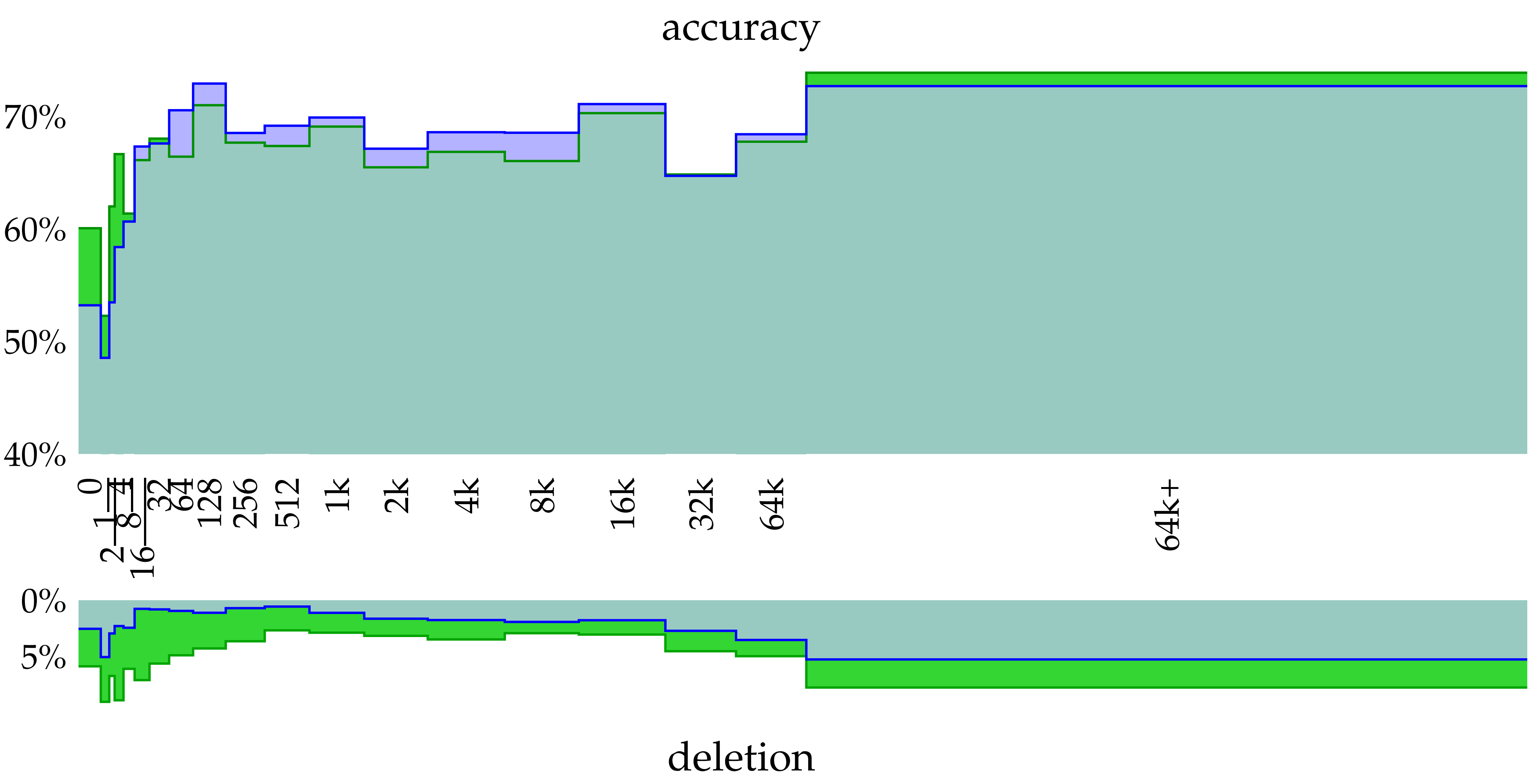
Rare Words

- More frequent in training $\rightarrow$ more likely to get right in test
- Let's measure this
- One problem
  - frequency measured for input words
  - translation correctness measured for output words

Translation Accuracy for Input Words

- Generate word alignment between input and output words
- Look up count of input word in training
- Link to output word via word alignment
- Check if it is also in the reference translation
- A lot of tedious special cases
  - one-to-many alignment, only some output words in reference
  - input word not aligned to any target word
  - many-to-one alignment
  - output word occurs multiple time in output or reference sentence

Count vs. Accuracy

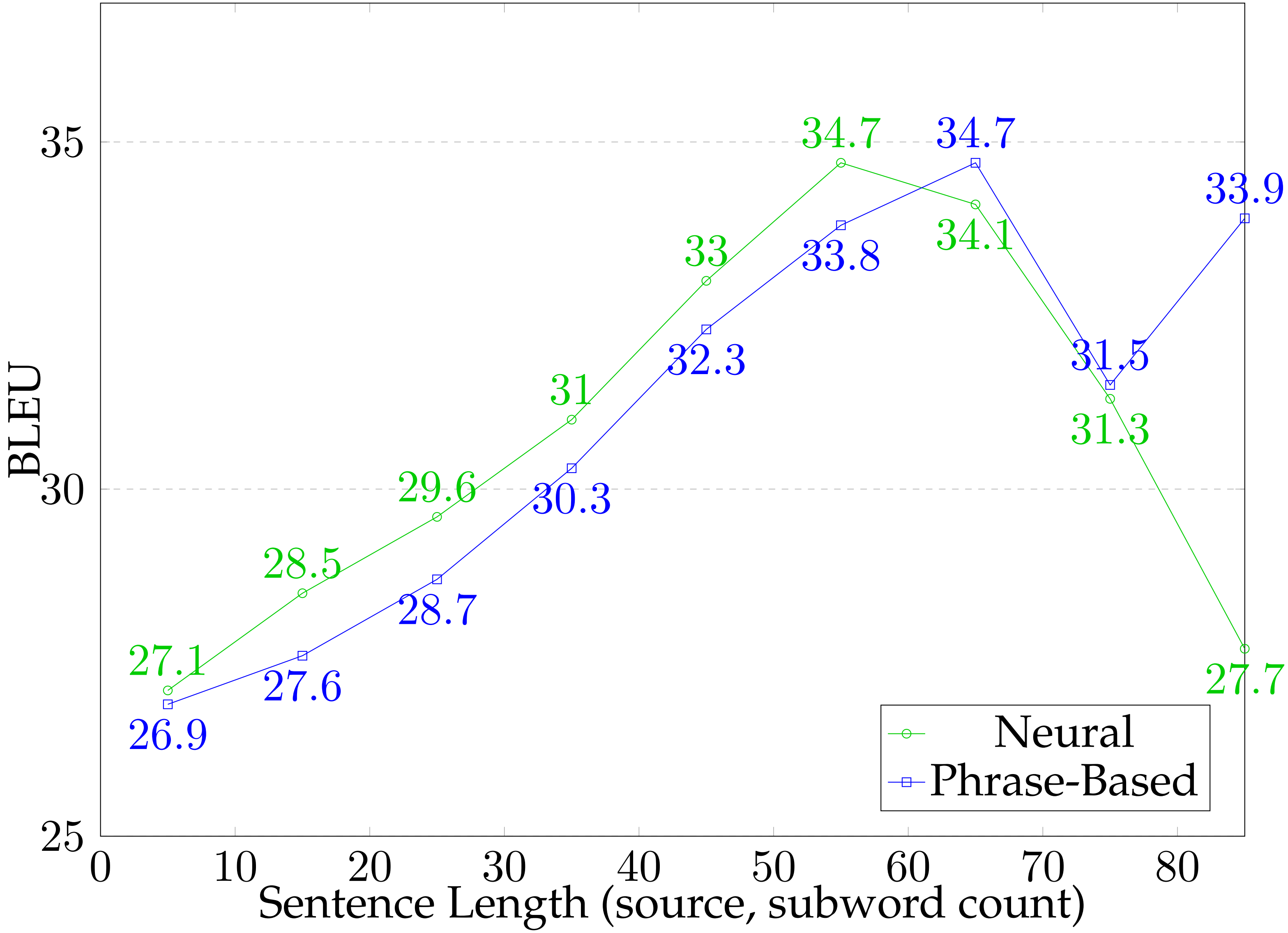


sentence length

Sentence Length

- For longer sentences, harder to keep track of coverage
- Training sentence length limit: longer sentences not seen in training

Sentence Length



word alignment

Word Alignment

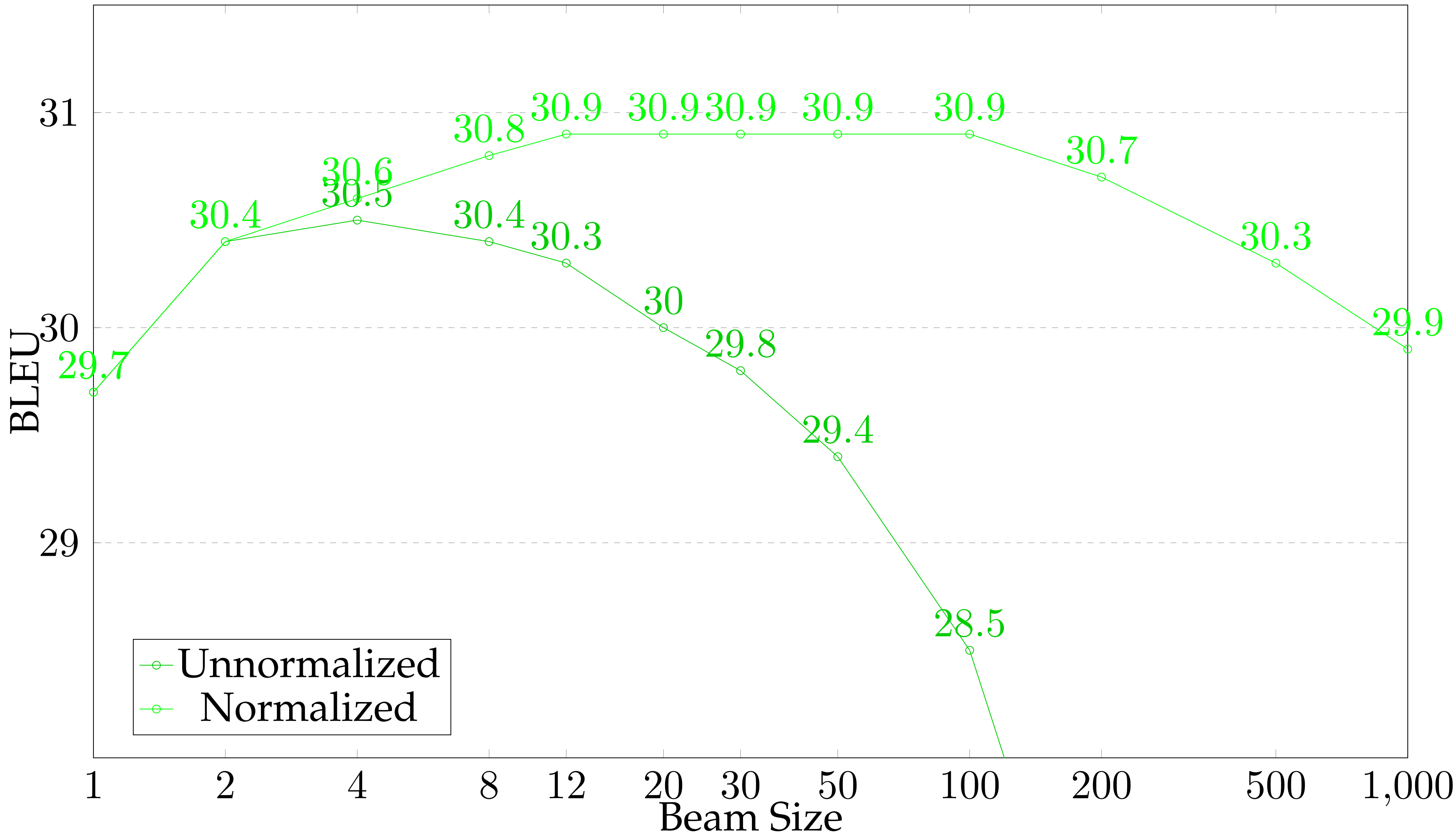
	relations	between	Obama	and	Netanyahu	have	been	strained	for	years	.
die	56		16								
Beziehungen	89										
zwischen		72	26								
Obama			96								
und				79							
Netanjahu					98						
sind						42	11	38			
seit								22	54	10	
Jahren										98	
angespannt								84			
.						11	14	23			49

Word Alignment?

	das	Verhältnis	zwischen	Obama	und	Netanyahu	ist	seit	Jahren	gespannt	.
the		47							17		
relationship			81								
between				72							
Obama					87						
and						93					
Netanyahu							95				
has								38	16		26
been								21	14		54
stretched											77
for									38	33	12
years										90	
.	11								19	32	17

beam search

Beam Search



31
30
29
BLEU
29.7
30.4
30.6
30.5
30.8
30.4
30.9
30.3
30.9
30
30.9
29.8
30.9
29.4
30.9
28.5
30.7
30.3
29.9
Unnormalized
Normalized
1
2
4
8
12
20
30
50
100
200
500
1,000
Beam Size

noisy data

Noise in Training Data

- Crawled parallel data from the web (very noisy)

	SMT	NMT
WMT17	24.0	27.2
+ Paracrawl	25.2 (+1.2)	17.3 (-9.9)

(German-English, 90m words each of WMT17 and Crawl data)

	5%		10%		20%		50%		100%	
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	-0.9	+0.2	-2.5	+0.4	-6.3	+0.8	-9.9	+1.2

- Corpus cleaning methods [Xu and Koehn, EMNLP 2017] give improvements

Types of Noise

- Misaligned sentences
- Disfluent language (from MT, bad translations)
- Wrong language data (e.g., French in German–English corpus)
- Untranslated sentences
- Short segments (e.g., dictionaries)
- Mismatched domain

Mismatched Sentences

- Artificial created by randomly shuffling sentence order
- Added to existing parallel corpus in different amounts

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

- Bigger impact on NMT (green, left) than SMT (blue, right)

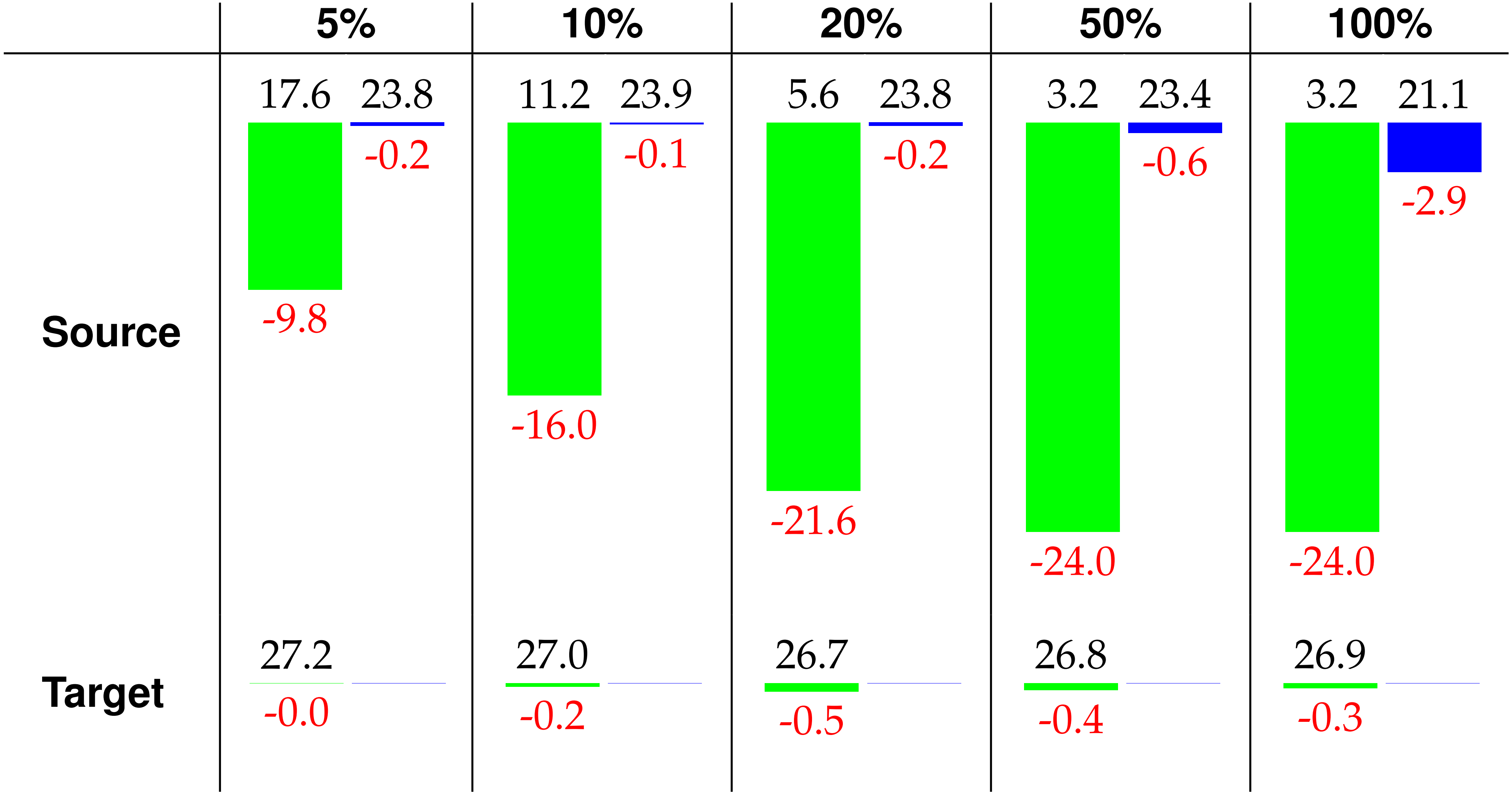
Misordered Words

- Artificial created by randomly shuffling words in each sentence

	5%		10%		20%		50%		100%	
Source		24.0 -0.0		23.6 -0.4		23.9 -0.1	26.6 -0.6	23.6 -0.4	25.5 -1.7	23.7 -0.3
Target		24.0 -0.0		24.0 -0.0		23.4 -0.6	26.7 -0.5	23.2 -0.8	26.1 -1.1	22.9 -1.1

- Similar impact on NMT than SMT, worse for source reshuffle

Untranslated Sentences



Wrong Language

	5%		10%		20%		50%		100%	
fr 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
fr 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

- Surprisingly robust, maybe due to domain mismatch of French data

Short Sentences

	5%		10%		20%		50%	
1-2 words	27.1	24.1	26.5	23.9	26.7	23.8		
	-0.1	+0.1	-0.7	-0.1	-0.5	-0.2		
1-5 words	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

- No harm done

questions?