hyperparameter	system						
hidden layer size	1024						
embedding size	512						
encoder depth	1	2				1	
encoder recurrence transition depth	1	2					
decoder depth	1	2				1	
dec. recurrence transition depth (base)	2	4				2	
dec. recurrence transition depth (high)	1	2				1	
tie decoder embeddings		yes					
layer normalization		yes					
lexical model					yes		yes
hidden dropout		0.2			0.5		
embedding dropout		0.2			0.5		
source word dropout		0.1			0.3		
target word dropout					0.3		
label smoothing		0.1				0.2	
maximum sentence length	200						
minibatch size (# tokens)	4000			1000			
learning rate	0.0001					0.0005	
optimizer	adam						
early stopping patience	10						
validation interval:							
IWSLT 100k / 200k / 400k	50		100		400		
IWSLT 2 / 800k / KO-EN 2.3M	1000		2000		8000		
beam size	5						

Table 5: Configurations of NMT systems reported in Table 2. Empty fields indicate that hyperparameter was unchanged compared to previous systems.