

ISO396-3	BS20						DeepSPIN20		IMS20		BS21	CL21	UBC21		DP21
	LSTM		transformer		pair n-gram							CL	UBC-1	UBC-2	
	WER	PER	WER	PER	WER	PER	WER	PER	WER	PER	WER	WER	WER	WER	
ady ^B	28.00	6.53	28.44	6.49	32.00	7.56	24.67 ³		25.00	5.79	22.00	22.00 ²³	25.00	22.00	
bul ^A	31.11	5.94	34.00	7.89	41.33	9.05	-		22.22	4.85	18.30	18.80 ⁶			
cym (wel) ^B											10.00	10.00 ¹	13.00	12.00	
ell (gre) ^B	18.89	3.30	18.89	3.06	21.78	4.05	-		18.67	2.97	21.00	20.00 ¹³	22.00	22.00	
eng(.us)											41.94				37.43
fra (fre) ^A	6.22	1.32	6.89	1.72	13.56	3.12	5.11 ³		6.89	1.60	8.50	7.50 ⁴⁵⁶			
hbs ^A											32.10	35.3 ⁷			
hin	6.67	1.47	9.56	2.40	12.67	4.05	-		5.11	1.20					
hun ^A	5.33	1.18	5.33	1.28	6.67	1.51	-		5.11	1.12	1.80	1.00 ⁶⁷			
hye (arm) ^A	14.67	3.49	14.22	3.29	18.00	3.90	-		12.67	2.94	7.00	6.40 ⁷			
ice ^B	10.00	2.36	10.22	2.21	17.56	3.62	-		9.33	2.04	12.00	10.00 ¹³	13.00	11.00	
ita ^B											19.00	31.00 ³	20.00	22.00	
jpn(.hira) ^A	7.56	1.79	7.33	1.86	9.56	2.07	4.89 ⁴		5.33	1.26	5.20	5.00 ⁷			
kat (geo) ^A	26.44	5.14	28.00	5.43	37.78	6.48	-		24.89	4.57	0.00	0.00 ⁴⁵⁶⁷			
khm ^B											34.00	32.00 ¹³	31.00	28.00	
kor ^A	46.89	16.78	43.78	17.50	52.22	15.88	24.00 ¹³		26.22	4.38	16.30	16.20 ⁴			

lav ^B											55.00	49.00 ²³	58.00	49.00
lit	19.11	3.55	20.67	3.65	23.11	4.43	-		20.00	3.63				
mlt(Jtn) ^B											19.00	12.00 ¹	19.00	18.00
nld (dut) ^A	16.44	2.94	15.78	2.89	23.78	3.97	-		13.56	2.36	14.70	14.70 ⁷		
rum ^B	10.67	2.53	12.00	3.62	11.56	3.55	9.78 ³		10.22	2.23	10.00	12.00 ³	14.00	10.00
slv ^B											49.00	50.00 ¹	56.00	47.00
vie ^A	4.67	1.52	7.56	2.27	8.44	1.79	0.89 ²		1.56	0.48	2.50	2.00 ⁵⁷		
macro	16.84	3.99	17.51	4.30	22.00	4.92	14.15	2.92	13.81	2.76				
macro low											25.10		27.10	24.10
macro medium											10.60			

Superscript numbers denote model numbers. Numbers in bold denote best model for one language (only WER). Languages in bold are in the 100LC corpus. ^A: medium resource languages in 2021. ^B: low resource languages in 2021

BS20: [Link](#)

DeepSPIN: Transformer- or LSTM-based enc-dec seq2seq models with sparse attention. Add language embedding to enc and dec states instead of language token. They did not report any results on the languages separately. Neither are the PER scores available. [Link](#)

IMS: Self training ensemble of one n-gram-based FST and 3 seq2seq (vanilla with attention, hard monotonic attention with pointer, hybrid of hard monotonic attention and tagging model) [Link](#)

BS21: similar to CL21 [Link](#)

CL21: LSTM-based neural transducer with pointer network-like monotonic hard attention trained with imitation learning 7 different, but still very similar ensembles of one model. [Link](#)

UBC21: UBC-2 baseline variant with vowel error punishment. UBC-1 baseline variant with syllable prediction. [Link](#)

DP21: Majority-vote ensemble consisting of 7 different models [Link](#)