Author	Model Architecture	ISO 639-3	WER
SIG21: Clematide and Makarov (2021) Link	CLUZH models 1-7. LSTM-based neural transducer with pointer network-like monotonic hard attention trained with imitation learning. All models 1-7 are majority-vote ensembles with different number of models (5-30) and different inputs (characters or segments). Achieved good results in nld (14.7), ice (10), jpn (5.0), fra (7.5) and vie (2.0) but not better than SIG20.	medium (8.000 train pairs)	
		hye (arm_e)	6.4
		hun	1.0
		kat (geo)	0.0
		kor	16.2
		low (800 train pairs)	
		ell (gre)	20
		ady	22
		lav	49
		mlt_ltn	12
		cym (wel_sw)	10
SIG21: Lo and Nicolai (2021) Link	UBC-2 outperforms the baseline. They analysed the errors of the baseline and extend it by adding penalties for wrong vowels and wrong diacritics. Errors on vowels actually decreased. Best macro average (low-resource).	ady	22
		khm	28
		lav	49
		slv	47
SIG21: Gautam et al. (2021) Link	Dialpad-1: Majority-vote ensemble consisting of three different public models (weighted FST, joint-sequence model trained with EM and a neural seq2seq), two seq2seq variants (LSTM and transformer) and two baseline variations.	high (32.800 train pairs)	
		eng (eng_us)	37.43
SIG20: Peters and Martin (2020) Link	DeepSPIN-2,-3,-4: Transformer- or LSTM-based enc-dec seq2seq models with sparse attention. Add language embedding to enc or dec states instead of language token.	3.600 train pairs	
		jpn (jpn_hira)	4.89
		fra (fre)	5.11
		rum	9.78
		vie	0.89
SIG20: Yu et al. (2020) 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).	hin	5.11
		nld (dut)	13.56

Table 1: SOTA models