Chinese→English

	Ave. %	Ave. z	System
1	78.8	0.140	NIUTRANS
	77.7	0.111	ONLINE-B
	77.9	0.109	UCAM
	78.0	0.108	UNISOUND
	77.5	0.099	TENCENT-ENSEM
	77.5	0.094	UNISOUND
	77.9	0.091	LI-MUZE
	77.0	0.089	NICT
	76.7	0.078	UMD
10	75.0	−0.005	ONLINE-Y
	74.5	−0.017	UEDIN
12	73.6	−0.061	ONLINE-A
13	65.9	−0.327	ONLINE-G
14	64.4	−0.377	ONLINE-F

English→Chinese

	Ave. %	Ave. z	System
1	80.7	0.219	TENCENT-ENSEM
	80.3	0.206	UNISOUND
	80.5	0.199	GTCOM-PRIMARY
	79.7	0.185	ALIBABA-ENS-RERANK
	79.2	0.173	ALIBABA-GENERAL-A
	79.5	0.166	ONLINE-B
	79.0	0.165	ALIBABA-GENERAL-B
8	78.1	0.094	UMD
	77.5	0.082	NICT
	77.1	0.069	ONLINE-Y
	75.5	0.037	ONLINE-A
12	70.7	−0.202	UEDIN
13	63.3	−0.419	ONLINE-F
	63.4	−0.435	ONLINE-G

Czech→English

	Ave. %	Ave. z	System
1	71.8	0.298	CUNI-TRANSFORMER
2	67.9	0.165	UEDIN
3	66.6	0.115	ONLINE-B
4	62.1	−0.023	ONLINE-A
5	57.5	−0.183	ONLINE-G

English→Czech

	Ave. %	Ave. z	System
1	67.2	0.594	CUNI-TRANSFORMER
2	60.6	0.384	UEDIN
3	52.1	0.101	ONLINE-B
4	46.0	−0.115	ONLINE-A
5	42.0	−0.246	ONLINE-G

German→English

	Ave. %	Ave. z	System
1	79.9	0.413	RWTH
	79.4	0.395	UCAM
	78.2	0.359	NTT
	77.3	0.346	ONLINE-B
	77.4	0.321	MLLP-UPV
	77.0	0.317	JHU
	76.9	0.315	UBIQUS-NMT
	76.7	0.310	ONLINE-Y
	75.7	0.268	ONLINE-A
	75.4	0.261	UEDIN
11	72.5	0.162	LMU-NMT
	72.2	0.149	NJUNMT-PRIVATE
13	65.2	−0.074	ONLINE-G
14	58.5	−0.296	ONLINE-F
15	45.4	−0.752	RWTH-UNSUPER
16	42.7	−0.835	LMU-UNSUP

English→German

	Ave. %	Ave. z	System
1	85.5	0.653	ONLINE-Z
2	82.2	0.561	ONLINE-B
	81.9	0.551	MICROSOFT-MARIAN
	81.6	0.539	MMT-PRODUCTION
	82.3	0.537	UCAM
	80.2	0.491	NTT
	79.3	0.454	KIT
8	77.7	0.396	ONLINE-Y
	76.7	0.377	JHU
	76.3	0.352	UEDIN
11	71.8	0.213	LMU-NMT
12	67.4	0.060	ONLINE-A
13	53.2	−0.385	ONLINE-F
	53.8	−0.416	ONLINE-G
15	36.7	−0.966	RWTH-UNSUPER
16	32.6	−1.122	LMU-UNSUP

Estonian→English

	Ave. %	Ave. z	System
1	73.3	0.326	TILDE-NC-NMT
2	71.1	0.238	NICT
	69.9	0.215	TILDE-C-NMT
	69.0	0.187	TILDE-C-NMT-2BT
	69.2	0.186	UEDIN
	68.7	0.171	TILDE-C-NMT-COMB
	67.1	0.117	ONLINE-B
	66.4	0.106	HY-NMT
	66.8	0.106	TALP-UPC
10	65.4	0.063	ONLINE-A
	64.0	0.007	CUNI-KOCMI
12	59.4	−0.117	NEUROTOLGE.EE
13	52.7	−0.341	ONLINE-G
14	34.6	−0.950	UNSUPTARTU

English→Estonian

	Ave. %	Ave. z	System
1	64.9	0.549	TILDE-NC-NMT
2	62.1	0.453	NICT
	61.6	0.427	TILDE-C-NMT
	61.2	0.418	TILDE-C-NMT-2BT
5	58.6	0.340	AALTO
	58.6	0.329	HY-NMT
	57.5	0.295	UEDIN
8	55.5	0.216	CUNI-KOCMI
	54.6	0.181	TALP-UPC
10	52.1	0.097	ONLINE-B
11	45.7	−0.132	NEUROTOLGE.EE
12	43.8	−0.195	ONLINE-A
13	37.6	−0.406	ONLINE-G
14	34.3	−0.520	PARFDA

Finnish→English

	Ave. %	Ave. z	System
1	75.2	0.153	NICT
	74.4	0.128	HY-NMT
	74.0	0.103	UEDIN
	72.7	0.083	CUNI-KOCMI
	72.9	0.078	ONLINE-B
	71.9	0.047	TALP-UPC
	71.5	0.045	ONLINE-A
8	66.1	−0.134	ONLINE-G
9	58.9	−0.404	JUCBNMT

English→Finnish

	Ave. %	Ave. z	System
1	64.7	0.521	NICT
	63.1	0.466	HY-NMT
3	59.2	0.324	UEDIN
	58.3	0.271	AALTO
	57.9	0.258	HY-NMT-2STEP
	57.4	0.238	TALP-UPC
	55.9	0.184	CUNI-KOCMI
	56.6	0.183	ONLINE-B
9	45.9	−0.212	ONLINE-A
	45.3	−0.233	ONLINE-G
11	42.7	−0.334	HY-SMT
	41.5	−0.369	HY-AH

Russian→English

	Ave. %	Ave. z	System
1	81.0	0.215	ALIBABA
	80.3	0.192	ONLINE-B
	79.6	0.170	ONLINE-G
4	77.5	0.110	UEDIN
5	76.2	0.034	ONLINE-A
6	74.1	−0.014	AFRL-SYSCOMB
	73.7	−0.027	JHU
8	64.2	−0.398	ONLINE-F

English→Russian

	Ave. %	Ave. z	System
1	72.0	0.352	ALIBABA-ENSEM
	71.4	0.324	ONLINE-G
3	66.8	0.159	ONLINE-B
	66.0	0.144	UEDIN
	64.9	0.115	PROMT-HYB-MARIAN
6	63.9	0.066	PROMT-HYB-OPENNMT
7	62.2	−0.004	ONLINE-A
8	59.1	−0.075	PROMT-RULE-BASED
9	44.5	−0.580	ONLINE-F

Turkish→English

	Ave. %	Ave. z	System
1	70.2	0.101	ONLINE-G
	69.3	0.077	ONLINE-A
	68.1	0.030	ALIBABA-ENSEMBLE
	68.0	0.027	ONLINE-B
	67.0	−0.008	UEDIN
	66.0	−0.040	NICT

English→Turkish

	Ave. %	Ave. z	System
1	66.3	0.277	ONLINE-B
	63.6	0.222	UEDIN
	63.5	0.216	ALIBABA-ENSEM-A
	62.0	0.128	NICT
	60.1	0.111	ALIBABA-ENSEM-B
	60.1	0.058	ONLINE-G
7	55.0	−0.060	RWTH
8	49.6	−0.254	ONLINE-A

Table 8: Official results of WMT18 News Translation Task. Systems ordered by standardized mean DA score, though systems within a cluster are considered tied. Lines between systems indicate clusters according to Wilcoxon rank-sum test at p-level $p < 0.05$. Systems with gray background indicate use of resources that fall outside the constraints provided for the shared task.