

**French-English**

589–716 judgments per combo

System	$\geq$ others
RWTH-COMBO ●	0.77
CMU-HYP-COMBO ●	0.77
DCU-COMBO ●	0.72
LIUM ★	0.71
CMU-HEA-COMBO ●	0.70
UPV-COMBO ●	0.68
NRC	0.66
CAMBRIDGE	0.66
UEDIN ★	0.65
LIMSI ★	0.65
JHU-COMBO	0.65
RALI	0.65
LIUM-COMBO	0.64
BBN-COMBO	0.64
RWTH	0.55

**German-English**

743–835 judgments per combo

System	$\geq$ others
BBN-COMBO ●	0.77
RWTH-COMBO ●	0.75
CMU-HEA-COMBO	0.73
KIT ★	0.72
UMD ★	0.68
JHU-COMBO	0.67
UEDIN ★	0.66
FBK	0.66
CMU-HYP-COMBO	0.65
UPV-COMBO	0.64
RWTH	0.62
KOC-COMBO	0.59

**Spanish-English**

1385–1535 judgments per combo

System	$\geq$ others
UEDIN ★	0.69
CMU-HEA-COMBO ●	0.66
UPV-COMBO ●	0.66
BBN-COMBO	0.62
JHU-COMBO	0.55
UPC	0.51

**English-French**

740–829 judgments per combo

System	$\geq$ others
RWTH-COMBO ●	0.75
CMU-HEA-COMBO ●	0.74
UEDIN	0.70
KOC-COMBO ●	0.68
UPV-COMBO	0.66
RALI ★	0.66
LIMSI	0.66
RWTH	0.63
CAMBRIDGE	0.63

**English-German**

1340–1469 judgments per combo

System	$\geq$ others
RWTH-COMBO ●	0.65
DFKI ★	0.62
UEDIN ★	0.62
KIT ★	0.60
CMU-HEA-COMBO ●	0.59
KOC-COMBO	0.59
FBK ★	0.56
UPV-COMBO	0.55

**English-Spanish**

516–673 judgments per combo

System	$\geq$ others
CMU-HEA-COMBO ●	0.68
KOC-COMBO	0.62
UEDIN ★	0.61
UPV-COMBO	0.60
RWTH-COMBO	0.59
DFKI ★	0.55
JHU	0.55
UPV	0.55
CAMBRIDGE ★	0.54
UPV-NNLM ★	0.54

**Czech-English**

766–843 judgments per combo

System	$\geq$ others
CMU-HEA-COMBO ●	0.71
ONLINEB ★	0.7
BBN-COMBO ●	0.70
RWTH-COMBO ●	0.65
UPV-COMBO ●	0.63
JHU-COMBO	0.62
UEDIN	0.61

**English-Czech**

1405–1496 judgments per combo

System	$\geq$ others
DCU-COMBO ●	0.75
ONLINEB ★	0.70
RWTH-COMBO	0.70
CMU-HEA-COMBO	0.69
UPV-COMBO	0.68
CU-BOJAR	0.66
KOC-COMBO	0.66
PC-TRANS	0.62
UEDIN	0.62

System combinations are listed in the order of how often their translations were ranked higher than or equal to any other system. Ties are broken by direct comparison. We show the best individual systems alongside the system combinations, since the goal of combination is to produce better quality translation than the component systems.

- indicates a **win** for the system combination meaning that no other system or system combination is statistically significantly better at  $p\text{-level} \leq 0.1$  in pairwise comparison.
- ★ indicates an **individual system** that none of the system combinations beat by a statistically significant margin at  $p\text{-level} \leq 0.1$ .

For all pairwise comparisons between systems, please check the appendix.

Note: ONLINEA and ONLINEB were not included among the systems being combined in the system combination shared tasks, except in the Czech-English and English-Czech conditions, where ONLINEB was included.