

### French–English

625–836 judgments per system

System	C?	≥others
GOOGLE ●	no	.76
DCU ★	yes	.66
LIMSI ●	no	.65
JHU ★	yes	.62
UEDIN ★	yes	.61
UKA	yes	.61
LIUM-SYSTRAN	no	.60
RBMT5	no	.59
CMU-STATXFER ★	yes	.58
RBMT1	no	.56
USAAR	no	.55
RBMT3	no	.54
RWTH ★	yes	.52
COLUMBIA	yes	.50
RBMT4	no	.47
GENEVA	no	.34

### English–French

422–517 judgments per system

System	C?	≥others
LIUM-SYSTRAN ●	no	.73
GOOGLE ●	no	.68
UKA ●★	yes	.66
SYSTRAN ●	no	.65
RBMT3 ●	no	.65
DCU ●★	yes	.65
LIMSI ●	no	.64
UEDIN ★	yes	.60
RBMT4	no	.59
RWTH	yes	.58
RBMT5	no	.57
RBMT1	no	.54
USAAR	no	.48
GENEVA	no	.38

### Hungarian–English

865–988 judgments per system

System	C?	≥others
MORPHO ●	no	.75
UMD ★	yes	.66
UEDIN	yes	.45

### German–English

651–867 judgments per system

System	C?	≥others
RBMT5	no	.66
USAAR ●	no	.65
GOOGLE ●	no	.65
RBMT2 ●	no	.64
RBMT3	no	.64
RBMT4	no	.62
STUTTGART ●★	yes	.61
SYSTRAN ●	no	.60
UEDIN ★	yes	.59
UKA ★	yes	.58
UMD ★	yes	.56
RBMT1	no	.54
LIU ★	yes	.50
RWTH	yes	.50
GENEVA	no	.33
JHU-TROMBLE	yes	.13

### English–German

977–1226 judgments per system

System	C?	≥others
RBMT2 ●	no	.66
RBMT3 ●	no	.64
RBMT5 ●	no	.64
USAAR	no	.58
RBMT4	no	.58
RBMT1	no	.57
GOOGLE	no	.54
UKA ★	yes	.54
UEDIN ★	yes	.51
LIU ★	yes	.49
RWTH ★	yes	.48
STUTTGART	yes	.43

### Czech–English

1257–1263 judgments per system

System	C?	≥others
GOOGLE ●	no	.75
UEDIN ★	yes	.57
CU-BOJAR ★	yes	.51

### Spanish–English

613–801 judgments per system

System	C?	≥others
GOOGLE ●	no	.70
TALP-UPC ●★	yes	.59
UEDIN ★	yes	.56
RBMT1 ●	no	.55
RBMT3 ●	no	.55
RBMT5 ●	no	.55
RBMT4 ●	no	.53
RWTH ★	yes	.51
USAAR	no	.51
NICT	yes	.37

### English–Spanish

632–746 judgments per system

System	C?	≥others
RBMT3 ●	no	.66
UEDIN ●★	yes	.66
GOOGLE ●	no	.65
RBMT5 ●	no	.64
RBMT4	no	.61
NUS ★	yes	.59
TALP-UPC	yes	.58
RWTH	yes	.51
RBMT1	no	.25
USAAR	no	.48

### English–Czech

4626–4784 judgments per system

System	C?	≥others
PCTrans ●	no	.67
EUROTRANXP ●	no	.67
GOOGLE	no	.66
CU-BOJAR ★	yes	.61
UEDIN	yes	.53
CU-TECTOMT	yes	.48

Systems 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.

C? indicates constrained condition, meaning only using the supplied training data and possibly standard monolingual linguistic tools (but no additional corpora).

● indicates a **win** in the category, meaning that no other system is statistically significantly better at  $p\text{-level} \leq 0.1$  in pairwise comparison.

★ indicates a **constrained win**, no other constrained system is statistically better.

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