

# WMT 2017 - Biomedical task

June 2, 2017

## 1 Results for Automatic Evaluation

BLEU scores were calculated using the multi-eval tool and tokenization as provided in Moses.

\* indicates the primary run as informed by the participants.

### 1.1 Scielo dataset

Runs	pt/en	es/en	en/pt	en/es
baseline	36.35	31.50	30.52	27.31
UHH run1	43.84	37.14	39.14	36.08
UHH run2	43.93	37.47	39.38	35.93
UHH run3	43.88*	37.49*	39.21*	36.23*

### 1.2 EDP dataset

Runs	fr/en	en/fr
baseline	17.47	12.32
Hunter run1	15.10*	17.50*
Hunter run2	15.18	17.21
kyoto run1	25.21*	25.52
kyoto run2	-	27.04*
UHH run1	22.64	22.43
UHH run2	22.37	22.25
UHH run3	23.41*	22.79*

### 1.3 Cochrane dataset

<b>Cochrane</b>	<b>cs</b>	<b>de</b>	<b>fr</b>	<b>pl</b>	<b>es</b>	<b>ro</b>
Hunter run1	-	24.72*	30.75*	17.16*	-	14.74*
Hunter run2	-	-	30.76	-	-	-
lilt run1	-	34.91*	-	-	-	-
lilt run2	-	33.97	-	-	-	-
LMU	-	36.44*	-	-	-	-
PJIIT run1	19.96*	25.13*	-	18.86	-	24.91*
PJIIT run2	-	-	-	12.45	-	-
PJIIT run3	-	-	-	18.88*	-	-
uedin-nmt run1	28.54*	37.11*	-	29.04*	-	41.18*
uedin-nmt run2	-	-	-	27.69	-	38.89
UHH run1	-	22.03	32.46	-	48.99	-
UHH run2	-	22.37	32.59	-	48.45	-
UHH run3	-	22.63*	33.16*	-	48.70*	-

### 1.4 NHS dataset

<b>NHS</b>	<b>cs</b>	<b>de</b>	<b>fr</b>	<b>pl</b>	<b>es</b>	<b>ro</b>
Hunter	-	20.45*	22.99*	14.09*	-	10.56*
lilt run1	-	27.57*	-	-	-	-
lilt run2	-	26.79	-	-	-	-
LMU	-	29.46*	-	-	-	-
PJIIT run1	15.93*	21.88*	-	14.32	-	18.10*
PJIIT run2	-	-	-	10.75	-	-
PJIIT run3	-	-	-	14.34*	-	-
uedin-nmt run1	22.79*	33.06*	-	23.15*	-	29.32*
uedin-nmt run2	-	-	-	19.87	-	27.32
UHH run1	-	18.71	31.79	-	40.97	-
UHH run2	-	19.80	31.89	-	41.20	-
UHH run3	-	19.66*	33.36*	-	41.22*	-

## 2 Results for Manual Validation

(coming later)