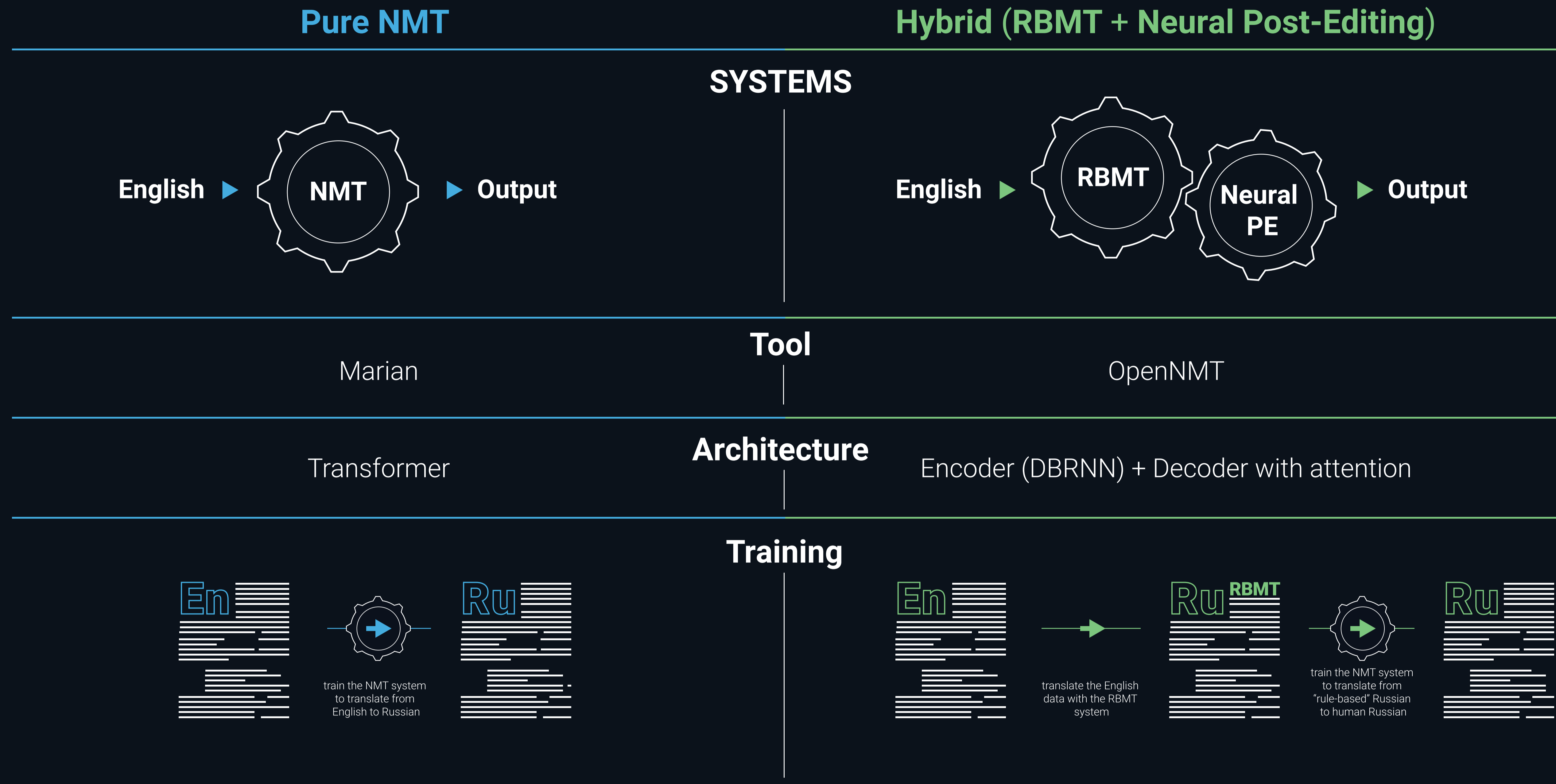




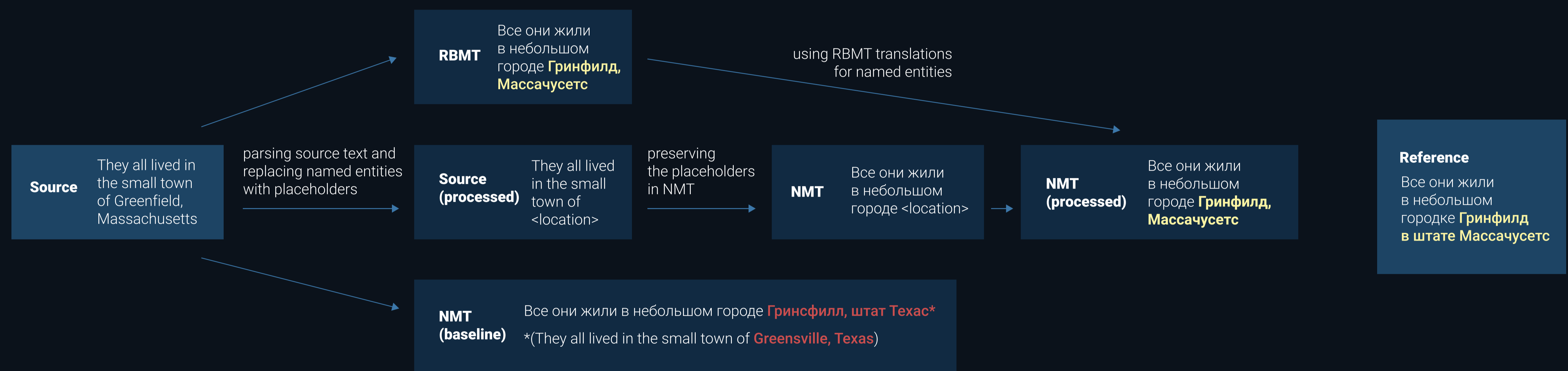
PROMT Systems for WMT 2018 Shared Translation Task

Alexander Molchanov, Statistical and Neural MT team lead, PROMT LLC



Features

Named Entities processing (Locations, Organizations, Persons)



DATA

Corpus	#sent	#tokens EN	#tokens RU
MultiUN	6.0	140.8	129
ParaCrawl	1.5	28.4	24.3
Yandex corpus	0.6	16.8	15.4
Private data	0.6	15.6	15
CommonCrawl	0.4	10.3	9.5
NewsCommentary	0.3	6.2	5.9
TED Talks	0.1	2.4	2.1
Total	9.5	220.5	201.2

• BPE 85K • synthetic data (back-translations of news from statmt.org), 10M • **Hybrid system only** copied monolingual data (news from statmt.org) in source and target, 10M
 • synthetic data with unknown words, 19M

RESULTS

System	newstest2017	newstest2018
RBMT	22.9	18.1
NMT	31.0	27.4
NMT+names	30.9	27.3
Hybrid	29.5	25.3



PROMT.COM