Has NMT really bridged the gap between MT and human translation? What problems remain?

Six Challenges for Neural Machine Translation

- domain mismatch
- amount of training data
- rare words
- long sentences
- beam search
- NMT systems are much less interpretable

System ↓	Law		Medical		IT		Koran		Subtitles	
All Data	30.5	32.8	45.1	42.2	35.3	44.7	17.9	17.9	26.4	20.8
Law	31.1	34.4	12.1	18.2	3.5	6.9	1.3	2.2	2.8	6.0
Medical	3.9	10.2	39.4	43.5	2.0	8.5	0.6	2.0	1.4	5.8
IT	1.9	3.7	6.5	5.3	42.1	39.8	1.8	1.6	3.9	4.7
Koran	0.4	1.8	0.0	2.1	0.0	2.3	15.9	18.8	1.0	5.5
Subtitles	7.0	9.9	9.3	17.8	9.2	13.6	9.0	8.4	25.9	22.1

Towards Robust Neural Machine Translation

Input	tamen *bupa* kunnan zuochu weiqi AI.
Output	They are not afraid of difficulties to make Go AI.
Input	tamen *buwei* kunnan zuochu weiqi AI.
Output	They are not afraid to make Go AI.

Source	zhongguo dianzi yinhang yewu guanli xingui jiangyu sanyue yiri qi shixing
Reference	china's new management rules for e-banking operations to take effect on march 1
MLE	china's electronic bank rules to be implemented on march 1
$AST_{lexical}$	new rules for business administration of china 's electronic banking industry will come into effect on march 1 .
$AST_{feature}$	new rules for business management of china 's electronic banking industry to come into effect on march 1
Perturbed Source	*zhongfang* dianzi yinhang yewu guanli xingui jiangyu sanyue yiri qi shixing
MLE	china to implement new regulations on business management
$AST_{lexical}$	the new regulations for the business administrations of the chinese electronics bank will come into effect on march 1 .
$AST_{feature}$	new rules for business management of china's electronic banking industry to come into effect on march 1

Achieving Human Parity on Automatic Chinese to English News Translation

Перевод новостей с китайского на английский.

Проблемы оценки качества перевода: метрики основаны на эталонах, эталонные переводы часто низкого качества, нужно проверять на людях.

Подход к переводу: Dual Learning, Deliberation Networks (2-pass decoding), Agreement Regularization, фильтрация исходных данных для уменьшения шума.

Оценка качества перевода на основе статистической различимости оценок «толокеров» (How accurately does the above candidate text convey the semantics of the source text?). Базовый уровень 68.5%, получилось 69%.