

## СТАТЬЯ

УДК 004.8

**МЕТОД ОЦЕНКИ КАЧЕСТВА МАШИННОГО ПЕРЕВОДА  
НА ОСНОВЕ ФРАЗОВОГО ВЫРАВНИВАНИЯ****Козина А.В., Черепков Е.А., Белов Ю.С.***Калужский филиал ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)», Калуга, e-mail: fn1-kf@mail.ru*

В настоящее время одним из наиболее эффективных подходов в области разработки систем машинного перевода (МП) является подход на основе выравнивания фраз. Методы на основе такого подхода используют непрерывные последовательности слов в качестве единиц перевода. Выравнивание потенциально может улучшить качество систем машинного перевода, поскольку учитывает общие отношения, порядок, контекст между словами. В данной статье предлагается подход к оценке качества машинного перевода с использованием фразового выравнивания. Вводятся функции выравнивания в пределах фразы и за её границами, а также функция покрытия, функция корреляции между частотами исходных и целевых фраз. На основе предложенных функций строится итоговая модель оценки машинного перевода. Проводится тестирование разработанной модели. Определяется наиболее подходящая размерность N-грамм – N последовательных слов для каждой фразы выравнивания. Проводится сравнение данного подхода с традиционными метриками оценки качества перевода, такими как BLEU, GTM, TER и METEOR. Результаты показывают, что предлагаемый подход имеет наилучшую производительность, а также для набора данных WMT 2013 обеспечивает повышение точности не менее 13,92% по сравнению с традиционными метриками.

**Ключевые слова:** оценка качества машинного перевода, выравнивание в машинном переводе, выравнивание на основе слов, выравнивание на основе фраз, BLEU, GTM, TER, METEOR

**QUALITY ASSESSMENT APPROACH OF MACHINE TRANSLATION  
BASED ON PHRASES ALIGNMENT****Kozina A.V., Cherepov E.A., Belov Yu.S.***Bauman Moscow State Technical University (Kaluga Branch), Kaluga, e-mail: fn1-kf@mail.ru*

Currently, one of the most effective approaches in the field of developing machine translation (MT) systems is an approach based on phrase alignment. Methods based on this approach use continuous sequences of words as translation units. Alignment can potentially improve the quality of machine translation systems, since it takes into account general relationships, order, context between words. This article proposes an approach to assessing the quality of machine translation using phrase alignment. Alignment functions are introduced within the phrase and beyond its boundaries, as well as a coverage function, a correlation function between the frequencies of the source and target phrases. Based on the proposed functions, the final model for evaluating machine translation is built. The developed model is being tested. The most suitable dimension of N-grams is determined – N consecutive words for each alignment phrase. This approach is compared with traditional translation quality assessment metrics, such as: BLEU, GTM, TER and METEOR. The results show that the proposed approach has the best performance, and for the WMT 2013 dataset it provides an increase in accuracy of at least 13.92% compared to traditional metrics.

**Keywords:** evaluation of the quality of machine translation, alignment in machine translation, word-based alignment, phrase-based alignment, BLEU, GTM, TER, METEOR

Этап выравнивания является важной частью машинного перевода, который заключается в определении соответствия между словами/фразами имеющими одинаковое значение на одном языке со словами/фразами на другом языке. Результатом выравнивания является параллельный текст, который используется для инициализации систем машинного перевода [1]. Улучшение выравнивания позволяет увеличить точность создаваемого системой перевода и кроме того способно внести значительный вклад при оценке качества такого перевода. Целью исследования является разработка подхода к оценке качества машинного перевода с использованием фразового выравнивания.

*Механизм выравнивания на основе слов*

Рассмотрим механизм выравнивания на примере [2]. Одним из способов визуализации задачи выравнивания является матрица (рис. 1).

Здесь соответствующие выравнивания между словами представлены точками в матрице выравнивания. Выравнивания слов не обязательно должны иметь однозначное соответствие. Слова могут иметь несколько точек выравнивания или не иметь их вовсе. Так, на рис. 1 английское слово *assumes* соответствует трем немецким словам *geht davon aus*. Однако не всегда легко установить корректное выравнивание слов. Основная проблема заключается в том, что

некоторые слова не имеют четкого эквивалента на другом языке.

	michael	geht	davon	aus	,	dass	er	im	haus	bleibt
michael	■									
assumes		■	■	■						
that						■				
he							■			
will										■
stay										■
in								■		
the								■		
house									■	

Рис. 1. Пример матрицы выравнивания слов: Слова на английском языке (строки) выровнены по словами на немецком языке (столбцы), как указано точками в матрице

Для учета выравнивания по нескольким словам можно применить модель выравнивания в двух направлениях: от источника к цели и от цели к источнику. Полученные в результате выравнивания слова могут быть объединены, например, путем объединения точек пересечения или объединения всех точек выравнивания. Этот про-

цесс обычно называют симметризацией выравнивания [2].

#### Механизм выравнивания на основе фраз

Как уже было указано ранее, иногда одно слово на исходном языке переводится как два или более слова на целевом языке. Процесс симметризации не всегда помогает исправить подобную ситуацию [3]. Один из вариантов решения данной проблемы – использовать более крупные единицы перевода, например фразы. Процесс выравнивания на основе фраз включает в себя два шага:

- создание выравнивания на основе слов между каждой парой предложений параллельного корпуса;
- извлечение фразовых пар, которые соответствуют полученному выравниванию слов.

Для рассматриваемого примера пошаговый процесс извлечения выровненных фраз отражен в таблице.

Оценка выравнивания для каждой пары фраз  $(\hat{s}, \hat{t})$  строится поверх оценок выравнивания слов следующим образом:

$$\beta_s(\hat{s}, \hat{t}) = - \sum_{s_i \in \hat{s}} \sum_{t_j \in \hat{t}} \ln a_s(i, j), \quad (1)$$

$$\beta_t(\hat{s}, \hat{t}) = - \sum_{s_i \in \hat{s}} \sum_{t_j \in \hat{t}} \ln a_t(i, j). \quad (2)$$

#### Процесс извлечения выровненных фраз

Этап	Фраза на английском языке	Фраза на немецком языке
1	michael	michael
2	michael assumes	michael geht davon aus; michael geht davon aus,
3	michael assumes that	michael geht davon aus, dass
4	michael assumes that he	michael geht davon aus, dass er
5	michael assumes that he will stay in the house	michael geht davon aus, dass er im haus bleibt
6	assumes	geht davon aus; geht davon aus,
7	assumes that	geht davon aus, dass
8	assumes that he	geht davon aus, dass er
9	assumes that he will stay in the house	geht davon aus, dass er im haus bleibt
10	that	dass ; , dass
11	that he	dass er ; , dass er
12	that he will stay in the house	dass er im haus bleibt ; , dass er im haus bleibt
13	he	er
14	he will stay in the house	er im haus bleibt
15	will stay	bleibt
16	will stay in the house	im haus bleibt
17	in the	im
18	in the house	im haus
19	house	haus

*Функция внешнего выравнивания*

Выравнивание подразумевает не только то, что слова в пределах исходной фразы выровнены со словами в пределах целевой фразы, но также что оставшаяся часть исходного предложения, выровнена с подобной частью целевого предложения. Предположим, что набор фраз в исходном предложении –  $S$ , в целевом –  $T$ . Функция внешнего выравнивания будет выглядеть следующим образом:

$$f_{os}(\hat{s}, \hat{s}', \hat{t}, \hat{t}') = -\ln \sum_{\hat{s}_i \in \hat{s}_{out}, \hat{s}', \hat{t}'} \phi_{os}(\hat{s}, \hat{s}', \hat{t}, \hat{t}') = -\ln \sum_{\hat{s}_i \in \hat{s}_{out}, \hat{s}', \hat{t}'} \frac{\sum_{\hat{t}_j \in \hat{t}_{out}, \hat{s}_i, \hat{t}'} \beta_t(\hat{s}_i, \hat{t}_j)}{\sum_{\hat{t}_j \in \hat{t}', \hat{s}_i, \hat{t}'} \beta_t(\hat{s}_i, \hat{t}_j)}, \quad (3)$$

$$f_{or}(\hat{s}, \hat{s}', \hat{t}, \hat{t}') = -\ln \sum_{\hat{t}_j \in \hat{t}_{out}, \hat{s}', \hat{t}'} \phi_{or}(\hat{s}, \hat{s}', \hat{t}, \hat{t}') = -\ln \sum_{\hat{t}_j \in \hat{t}_{out}, \hat{s}', \hat{t}'} \frac{\sum_{\hat{s}_i \in \hat{s}_{out}, \hat{s}_i, \hat{t}'} \beta_s(\hat{s}_i, \hat{t}_j)}{\sum_{\hat{s}_i \in \hat{s}', \hat{t}_j, \hat{t}'} \beta_s(\hat{s}_i, \hat{t}_j)}, \quad (4)$$

$s_{out}$  и  $t_{out}$  – набор фраз за пределами соответствия между  $\hat{s}'$  и  $\hat{t}'$ .

*Функция покрытия*

Кроме того, полезно учесть, какую часть переводимого текста покрывает рассматриваемая фраза. Пусть  $\hat{s}$  и  $\hat{t}$  представляют исходную и целевую фразы,  $S$  и  $T$  представляют исходное и целевое предложения. Отношения  $\frac{|\hat{s}|_{words}}{|S|_{words}}$  и  $\frac{|\hat{t}|_{words}}{|T|_{words}}$  отражают покрытие текста каждой выровненной фразой (1 – предложение охвачено полностью, 0 – отсутствие покрытия). Соответствующие функции выглядят следующим образом:

$$f_{sc}(\hat{s}, \hat{t}) = \varphi_{sc}(\hat{s}, \hat{t}) = \frac{|\hat{s}|_{words}}{|S|_{words}}, \quad (5)$$

$$f_{tc}(\hat{s}, \hat{t}) = \varphi_{tc}(\hat{s}, \hat{t}) = \frac{|\hat{t}|_{words}}{|T|_{words}}. \quad (6)$$

*Корреляция между частотами исходных и целевых фраз*

С целью сохранить количество предложений в корпусе, в которых встречается конкретная выровненная фраза, вводится дополнительная функция. Пусть  $C(\hat{s})$  – число вхождений фразы  $\hat{s}$  в корпус. Аналогичным образом,  $C(\hat{t})$  – число вхождений фразы  $\hat{t}$  в корпус. Функция частот будет выглядеть следующим образом:

$$f_{freq}(\hat{s}, \hat{t}) = \varphi_f(\hat{s}, \hat{t}) = \frac{(C(\hat{s}) - C(\hat{t}))^2}{(C(\hat{s}) + C(\hat{t}) + 1)^2}. \quad (7)$$

*Функция итоговой оценки*

Учитывая пару фраз  $(\hat{s}, \hat{t})$  и выровненное соответствие  $(\hat{s}', \hat{t}')$ , определяется функция  $\varphi_f(\hat{s}, \hat{s}', \hat{t}, \hat{t}')$  для обеспечения единого представления функций, описанных ранее:

$$\varphi_f(\hat{s}, \hat{s}', \hat{t}, \hat{t}') = \begin{cases} f(\hat{s}, \hat{t}) & \text{если } (\hat{s}', \hat{t}') \text{ не существует} \\ f(\hat{s}, \hat{s}', \hat{t}, \hat{t}') & \text{если } (\hat{s}', \hat{t}') \text{ существует} \end{cases} \quad (8)$$

Используем метод аппроксимации логарифмической модели для оценки единиц машинного перевода [4]. Общая функция оценки  $E(\hat{s}, \hat{t}, \theta)$  для пары фраз  $(\hat{s}, \hat{t})$  будет выглядеть следующим образом:

$$E(\hat{s}, \hat{t}, \theta) = \ln \sum_{\hat{s}, \hat{t}} e^{\sum_{i, \hat{s}, \hat{t}} \Phi_i(\hat{s}, \hat{s}; \hat{t}, \hat{t})}. \quad (9)$$

### Результаты работы метода

Для оценки предложенного подхода использовался набор данных WMT 2013. Данные сопровождаются широким спектром языковых пар. В данной работе рассматривались две пары параллельных корпусов: французско-английский и испанско-английский. Параллельные корпуса сопровождаются ручными и автоматическими оценками перевода [5]. Для оценки эффективности дан-

ного подхода результаты сравнивались с традиционными метриками оценки МП, такими как BLEU [6], GTM, TER и METEOR [7].

Для определения корреляции с человеческими суждениями использовался коэффициент корреляции баллов Спирмана.

$$\rho = 1 - \frac{6 \sum d_{ij}^2}{n(n^2 - 1)}, \quad (10)$$

где  $d_{ij}$  – евклидово расстояние, вычисляемое как разница между данной метрикой оценки и оценкой суждения человека,  $n$  – число переводов.

В предлагаемом подходе необходимо было выбрать наиболее подходящую размерность  $N$ -грамм.  $N$ -грамма состоит из  $N$  последовательных слов для каждой фразы выравнивания. Рис. 2, 3 демонстрируют влияние выбора различных значений  $N$  на точность предложенного подхода.

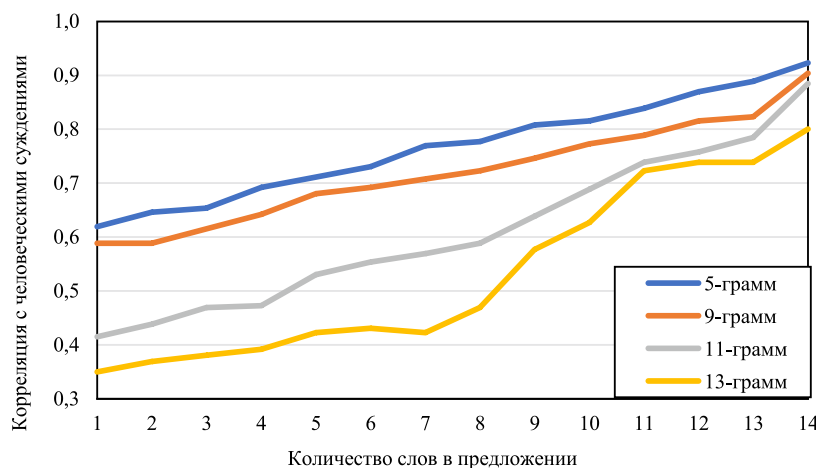


Рис. 2. Влияние разной длины  $N$ -грамм на корреляцию с суждениями человека с использованием французско-английских данных WMT 2013

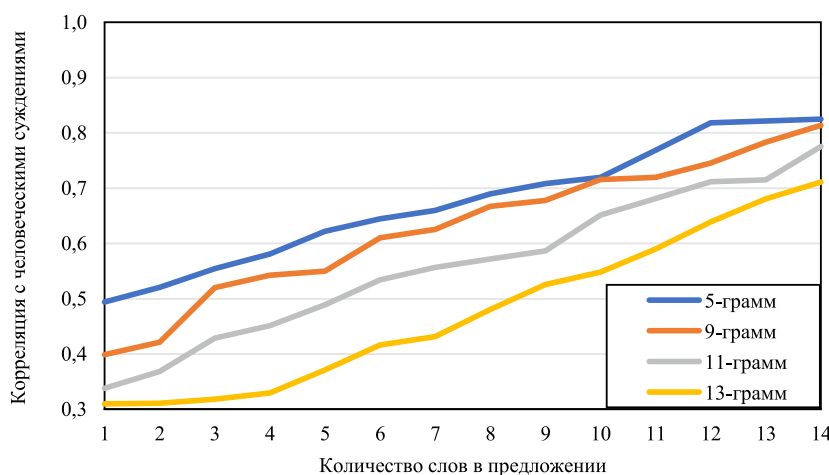


Рис. 3. Влияние разной длины  $N$ -грамм на корреляцию с суждениями человека с использованием испанско-английских данных WMT 2013

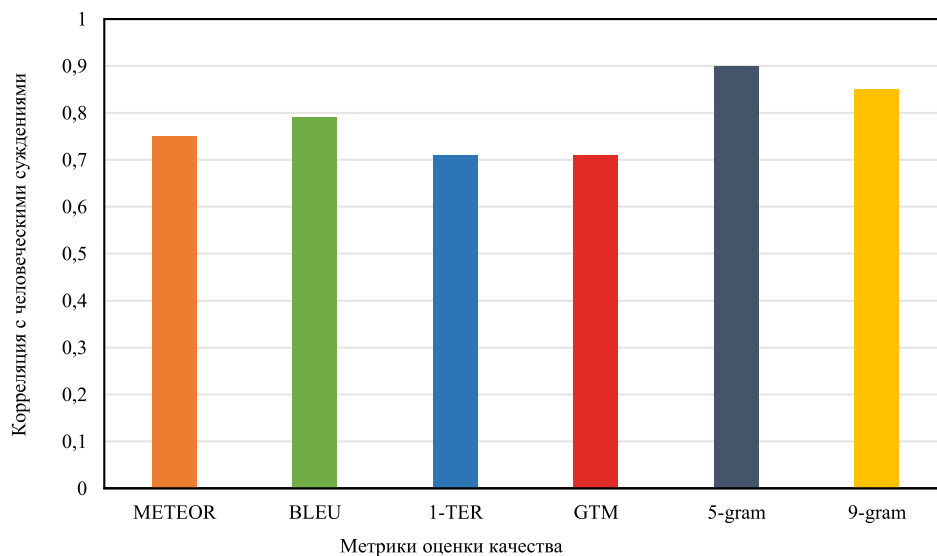


Рис. 4. Сравнение с использованием французско-английских данных WMT 2013

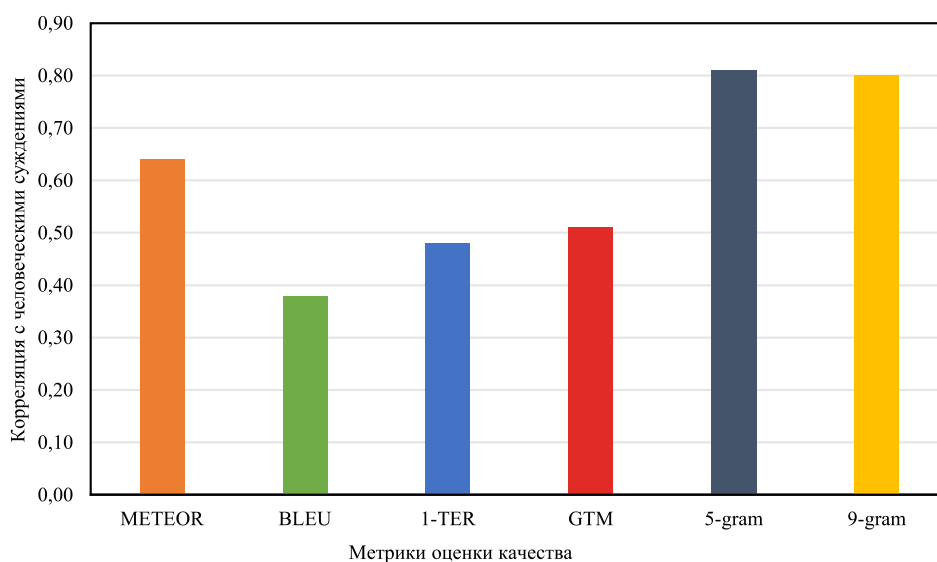


Рис. 5. Сравнение с использованием испанско-английских данных WMT 2013

Было обнаружено, что уменьшение значения  $N$  приводит к устойчивому улучшению корреляции с человеческими суждениями. В связи с этим было решено проводить дальнейшее тестирование с  $N$ -граммами в размере 5 и 9 слов.

На рис. 4, 5 представлена сводная информация о результатах корреляции предлагаемого подхода в сравнении с другими метриками.

Результаты показали, что предлагаемый подход имеет наилучшую производитель-

ность на тестируемом наборе данных. Предложенный подход обеспечивает повышение точности не менее чем на 3,85%.

#### Заключение

Таким образом, разработанный в данной работе подход показывает лучшие результаты в сравнении с традиционными метриками машинного перевода. Данный подход имеет потенциал, может быть в дальнейшем улучшен, например, введением дополнительных вероятностных функций

с учетом различных границ выравнивания, а также с включением человеческих оценок в процесс обучения системы.

#### Список литературы

1. Козина А.В., Черепков Е.А., Белов Ю.С. Обзор подходов к машинному переводу // Международный студенческий научный вестник. 2018. № 6. [Электронный ресурс]. URL: <https://www.eduherald.ru/ru/article/view?id=19408> (дата обращения: 25.12.2019).

2. Koehn P. Statistical Machine Translation, 1st ed. New York, NY, USA: Cambridge University Press, 2010. P. 488.

3. Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Barry Haddow, Matthias Huck, Chris Hokamp, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Carolina Scarton, Lucia Specia, and Marco Turchi. Findings of the 2015 workshop on statistical machine translation. In Proceedings of the Tenth Workshop on Statistical Machine Translation. Lisbon, Portugal // Association for Computational Linguistics. 2015. vol. Proceedings of the Tenth Workshop on Statistical Machine Translation. P. 1–46. [Electronic resource].

URL: <https://www.aclweb.org/anthology/W15-3001> (date of access: 25.12.2019).

4. Phillips B. Cunei: Open-Source Machine Translation with Relevance-Based Models of Each Translation Instance // Machine Translation. 2011. vol. 25. no. 2. P. 161–177. DOI: 10.1007/s10590-011-9109-6.

5. Козина А.В., Черепков Е.А., Белов Ю.С. Автоматические метрики оценки качества машинного перевода // Системный администратор. 2019. № 11 (204).

6. Chen B., Cherry C. A systematic comparison of smoothing techniques for sentence-level BLEU // Association for Computational Linguistics. 2014, vol. Proceedings of the Ninth Workshop on Statistical Machine Translation. P. 362–367. [Electronic resource]. URL: <https://www.aclweb.org/anthology/W14-3346> (date of access: 25.12.2019).

7. Liu C., Dahlmeier D., H.T. Ng. Better evaluation metrics lead to better machine translation. // Association for Computational Linguistics. 2011, vol. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. P. 375–384. [Electronic resource]. URL: <https://www.aclweb.org/anthology/D11-1035> (date of access: 25.12.2019).