

## Machine translation

For the task of machine translation, [NLLB-200](#) from Facebook was used. The model of my own choice was [OPUS](#) from Helsinki-NLP due to being lightweight and fast when compared to NLLB. I chose to translate English to Russian and vice versa, it made collecting my own dataset easier. [FLORES](#) devset consisting of 1012 sentences was used to translate and compare as ground truth. As for my own dataset, I randomly picked 25 sentences from [DEMETR](#) dataset's human translation of news, and another 25 from official fiction translations. [BLEU](#) scores were evaluated. Both datasets were also translated and evaluated on Google and Yandex translators. All evaluations have been double checked with independent code.

|                      | NLLB-200           | OPUS               | Yandex             | Google             |
|----------------------|--------------------|--------------------|--------------------|--------------------|
| FLORES (eng>rus)     | 15.8165936968724   | 16.923267918690055 | 30.09429889037876  | 29.66732122901695  |
| FLORES (rus>eng)     | 22.14250071345737  | 31.314224813827344 | 46.401882441470555 | 17.395797375642235 |
| my dataset (eng>rus) | 27.021693845694813 | 23.56370192487632  | 53.007145129171796 | 40.1577332834242   |
| my dataset (rus>eng) | 29.644501864987678 | 11.705589020696324 | 30.96815816279626  | 46.04495152573003  |

Table 1: BLEU scores for 4 translation models

## Observations

- Some sentences in FLORES are actually more than a single sentence, some are separated by a full stop comma, some by semicolons. They are a rare occurrence.
- All dataset-model combinations had sentences with a BLEU score of 100, meaning they were identical to ground truth.
- There was no clear and consistent trend among models with respect to datasets.
- Google translator performed much worse translating FLORES from russian to english, however it had the best accuracy on my own dataset translating in the same direction. I presume it is due to small size of the dataset.
- Given that accuracies on FLORES are a better representation of performance due to size, Yandex was all around the best translation model.
- As can be seen from tables below, sentences yielding the minimum BLEU score are either wrong in FLORES (are not atomic) or yield translations that still make sense. So none of the models failed miserably.
- Fiction sentences are harder to get a high score on, even if the translation is fine, ground truth tends to be more creative. I believe even human translation would be far from a BLEU score of 100 with the majority of fiction. For instance, 'I wondered what kind of interior life might give rise to such an exterior.' has a ground truth translation of 'Меня интересовало, какой внутренний мир мог породить такое внешнее проявление.', and NLLB-200 translated it as 'Я удивился, какая внутренняя жизнь может привести к такому внешнему.'. And Google - 'Я задавался вопросом, какая внутренняя жизнь могла породить такую внешнюю.'. Then Yandex - 'Я задавался вопросом, какая внутренняя жизнь могла бы породить такую внешность.'. One could argue that all of these translations are good enough and convey the same meaning, maybe except Yandex, but all of them had scores lower than 10.
- Given 2 languages, accuracies of translating one way or the other may differ greatly. I theorize that it is due to language features and the way they are utilized, one may use it right, but still deviate from a native speaker. We can observe that the average BLEU scores for FLORES were lower translating from english to russian in 3 models (except Google).
- NLLB-200 performed well on my dataset. Both directions of translation were higher than scores on FLORES. Google as well, performed well translating in both directions. The highest score overall was on Yandex translating my dataset from english to russian, it was greater than 50.

| NLLB-200    | BLEU | Translation  | Ground truth  |
|-------------|------|--|---|
| F(en>ru)min | 0.79 | В конце Средневековья западная Европа начала развивать свой собственный стиль.         | Собственные стили начали развиваться в Западной Европе к концу Средневековья. Одним из самых больших достижений того времени, благодаря крестовым походам, стало использование пуговиц для застегивания одежды. |
| F(en>ru)max | 100  | В конце 2017 года Симинофф появился на торговом телеканале QVC.                        | В конце 2017 года Симинофф появился на торговом телеканале QVC.   |
| F(ru>en)min | 1.45 | It's very hard to detect, even for someone who knows it exists.                        | In fact, it is not easy to find at all even if one knew it existed. Once inside the cave, it is a total isolation.  |
| F(ru>en)max | 100  | Suspected cases of H5N1 in Croatia and Denmark remain unconfirmed.                     | Suspected cases of H5N1 in Croatia and Denmark remain unconfirmed.  |
| M(en>ru)min | 4.03 | Я удивился, какая внутренняя жизнь может привести к такому внешнему.                   | Меня интересовало, какой внутренний мир мог породить такое внешнее проявление.  |
| M(en>ru)max | 100  | Все животные равны, но некоторые животные равнее других.                               | Все животные равны, но некоторые животные равнее других.  |
| M(ru>en)min | 7.67 | The common truth is that a bachelor with a great fortune should be looking for a wife. | It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife.   |
| M(ru>en)max | 100  | It was the best of times, it was the worst of times...                                 | It was the best of times, it was the worst of times...  |

Table 2: Min and max scores and corresponding sentences in NLLB-200

| OPUS        | BLEU | Translation   | Ground truth  |
|-------------|------|---|---|
| F(en>ru)min | 0.79 | К концу средневековья Западная Европа начала развивать свой собственный стиль.  | Собственные стили начали развиваться в Западной Европе к концу Средневековья. Одним из самых больших достижений того времени, благодаря крестовым походам, стало использование пуговиц для застегивания одежды. |
| F(en>ru)max | 100  | Однако водитель получил серьезные травмы головы.  | Однако водитель получил серьезные травмы головы.  |
| F(ru>en)min | 1.87 | A Polish skier with visual impairment, Macea Kresel, and an athlete leading Anna Ogarginska finished 13th in a supergiant, South Korean skier Park Jong-sook finished 24th in a supergiant among men sitting. | Poland's men's visually impaired skier Maciej Krezel and guide Anna Ogarczyńska finished thirteenth in the Super-G. South Korea's Jong Seok Park finished twenty-fourth in the men's sitting Super-G.           |
| F(ru>en)max | 100  | The Internet combines elements of both mass and interpersonal communication.  | The Internet combines elements of both mass and interpersonal communication.  |
| M(en>ru)min | 2.36 | В США число инфицированных составляет 4 миллиона человек, в Бразилии — 2,3 миллиона, а в Индии — 1,3 миллиона.  | По количеству заражений лидируют США (4 млн); потом идут Бразилия (2,3 млн) и Индия (1,3 млн).  |
| M(en>ru)max | 100  | Это была любовь с первого взгляда.  | Это была любовь с первого взгляда.  |
| M(ru>en)min | 7.67 | Vladimir Putin and Vladimir Zelinski &apos;s phone call was initiated by the Ukrainian side and confirmed by the Kremlin press service.   | The telephone conversation between Vladimir Putin and Volodymyr Zelensky took place on Ukrainian initiative, the Kremlin aide states.   |
| M(ru>en)max | 100  | It was the best of times, it was the worst of times...  | It was the best of times, it was the worst of times...  |

Table 3: Min and max scores and corresponding sentences in OPUS

| Google      | BLEU  | Translation   | Ground truth  |
|-------------|-------|---|---|
| F(en>ru)min | 1.12  | Другие номинации включают «Лучший фильм», «Режиссер», «Оперативная работа», «Дизайн костюмов», «Монтаж фильма», «Оригинальная музыка», «Художник-постановщик», «Звуковой монтаж», «Сведение звука» и «Оригинальный сценарий». | Среди других номинаций: "Лучшая картина "Режиссёр "Операторская работа "Дизайн костюмов "Монтаж фильма "Оригинальная музыка "Художник-постановщик "Звуковой монтаж "Звук"и "Оригинальный сценарий". |
| F(en>ru)max | 100   | В конце 2017 года Симинофф появился на торговом телеканале QVC.   | В конце 2017 года Симинофф появился на торговом телеканале QVC.   |
| F(ru>en)min | 2.86  | Harmony in the workplace is extremely important, and the results of group work are valued above individual achievements.  | Workplace harmony is crucial, emphasizing group effort rather than praising individual accomplishments.   |
| F(ru>en)max | 100   | Suspected cases of H5N1 in Croatia and Denmark remain unconfirmed.  | Suspected cases of H5N1 in Croatia and Denmark remain unconfirmed.  |
| M(en>ru)min | 4.46  | Я задавался вопросом, какая внутренняя жизнь могла породить такую внешнюю.  | Меня интересовало, какой внутренний мир мог породить такое внешнее проявление.  |
| M(en>ru)max | 100   | Об этом в интервью ERR рассказал генеральный секретарь Центристской партии Михаил Корб.   | Об этом в интервью ERR рассказал генеральный секретарь Центристской партии Михаил Корб.   |
| M(ru>en)min | 10.15 | After all, who can love a monster?  | For who could ever learn to love a beast?   |
| M(ru>en)max | 100   | This was shown by a survey conducted by the sociological research company Vilmarus, commissioned by the newspaper Lietuvos rytas.   | This was shown by a survey conducted by the sociological research company Vilmarus, commissioned by the newspaper Lietuvos rytas.   |

Table 4: Min and max scores and corresponding sentences in Google

| Yandex      | BLEU  | Translation  | Ground truth  |
|-------------|-------|--|---|
| F(en>ru)min | 1.24  | Другие номинации включают лучшую картину, режиссуру, операторскую работу, дизайн костюмов, монтаж фильма, оригинальную партитуру, дизайн постановки, монтаж звука, сведение звука и оригинальный сценарий. | Среди других номинаций: "Лучшая картина "Режиссёр "Операторская работа "Дизайн костюмов "Монтаж фильма "Оригинальная музыка "Художник-постановщик "Звуковой монтаж "Звук"и "Оригинальный сценарий". |
| F(en>ru)max | 100   | 64-летний водитель грузовика в аварии не пострадал.  | 64-летний водитель грузовика в аварии не пострадал.   |
| F(ru>en)min | 2.36  | Plants produce oxygen that a person breathes and absorb carbon dioxide that a person exhales.  | Plants make oxygen which humans breathe, and they take in carbon-dioxide which humans exhale (that is, breathe out).  |
| F(ru>en)max | 100   | Rolando Mendoza fired his M16 rifle at the tourists.   | Rolando Mendoza fired his M16 rifle at the tourists.  |
| M(en>ru)min | 4.07  | Я задавался вопросом, какая внутренняя жизнь могла бы породить такую внешность.  | Меня интересовало, какой внутренний мир мог породить такое внешнее проявление.  |
| M(en>ru)max | 100   | В Сиэтле полиция применила светошумовые гранаты и слезоточивый газ и арестовала 45 человек, после того как демонстрации в поддержку протестов в Портленде переросли в насилие.                             | В Сиэтле полиция применила светошумовые гранаты и слезоточивый газ и арестовала 45 человек, после того как демонстрации в поддержку протестов в Портленде переросли в насилие.                      |
| M(ru>en)min | 10.15 | After all, who can love a monster?   | For who could ever learn to love a beast?   |
| M(ru>en)max | 100   | It was the best of times, it was the worst of times...   | It was the best of times, it was the worst of times...  |

Table 5: Min and max scores and corresponding sentences in Yandex