

# Отчёт БДЗ 2

Алеся Иванова

19 марта 2023 г.

Код был частично взят из [https://pytorch.org/tutorials/beginner/translation\\_transformer.html](https://pytorch.org/tutorials/beginner/translation_transformer.html)

## Эксперимент 1 (checkpoint)

Построила словарь из слов с минимальной частотой 10 в обучающей выборке. Размер немецкого словаря получился 15019, размер английского словаря получился 12501.

Обучила трансформер с параметрами:

$EMB\_SIZE = 512$

$NHEAD = 8$

$FFN\_HID\_DIM = 512$

$BATCH\_SIZE = 128$

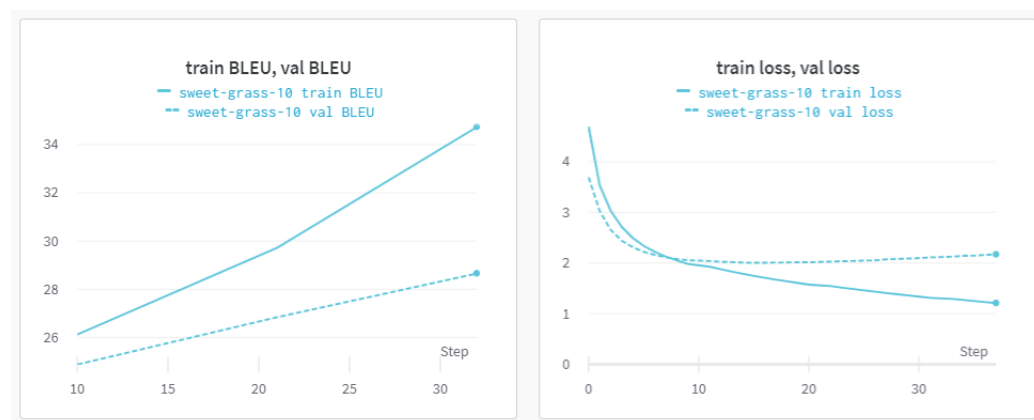
$NUM\_ENCODER\_LAYERS = 3$

$NUM\_DECODER\_LAYERS = 3$

где  $EMB\_SIZE$  - размерность эмбедингов,  $NHEAD$  - количество голов трансформера,  $FFN\_HID\_DIM$  - размерность Feed-Forward слоёв,  $BATCH\_SIZE$  - размер батча,  $NUM\_ENCODER\_LAYERS$  - количество слоёв энкодера,  $NUM\_DECODER\_LAYERS$  - количество слоёв декодера.

Оптимайзер Adam с параметрами  $lr = 0.0001$ ,  $betas = (0.9, 0.98)$ ,  $eps = 10^{-9}$ .

Раз в 10 эпох считую BLEU на 100 семплах из теста и 100 семплах из валидации, чтобы проверить, что модель получает адекватные значения BLEU. При этом, так как BLEU берётся по маленькому количеству сэмплов, то оно не отражает действительное качество. Действительное качество отражает посылка ответов в бота :)



На тесте лучший результат 27.73 получился после 20 эпох.

Алеся Иванова

bhw2\_answer-1

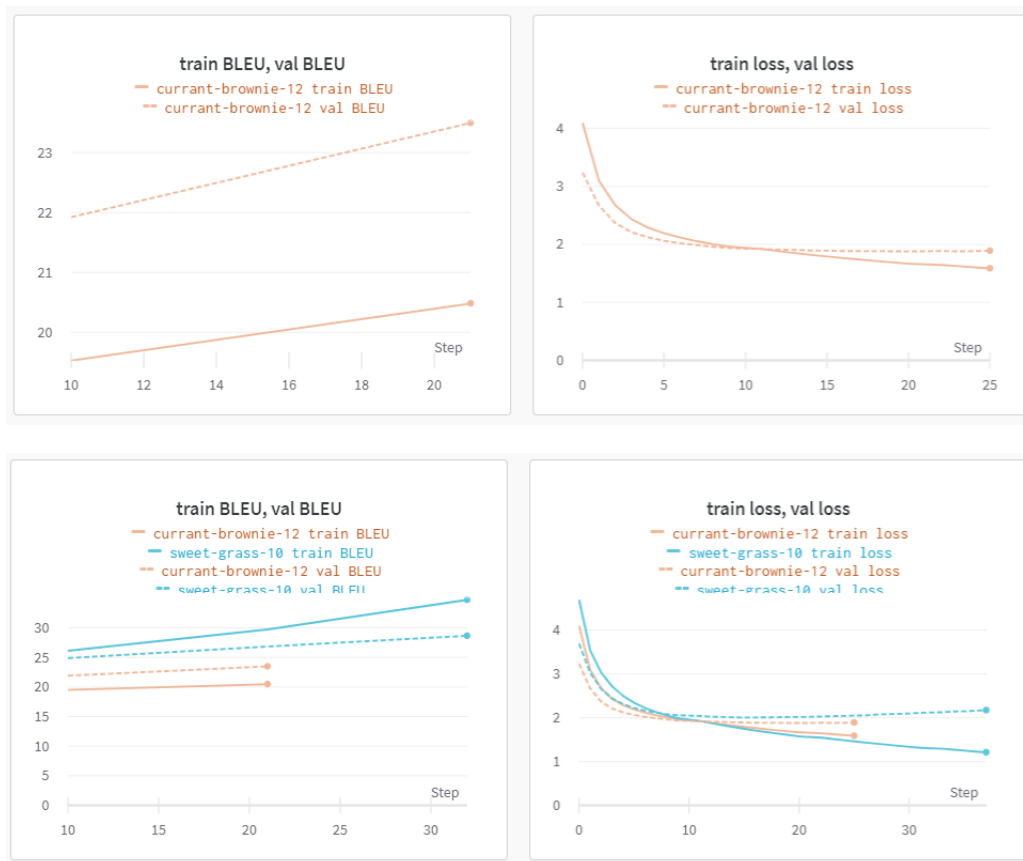
Я принял твоё решение.

Значение BLEU на публичном сплите 27.73.

Ты получаешь за это решение 5.865 баллов. 20:13

## Эксперимент 2

Решила изменить минимальную частоту для попадания слова в словарь, поставила частоту 50. Размер немецкого словаря получился 4285, размер английского словаря получился 4278. Модели должно быть проще обучаться, когда в словаре содержится меньше слов. Но при этом она не сможет выдавать перевод достаточно редких слов.



Модель медленнее переобучалась, но на тесте лучшее качество было 24.29. Стоит отметить, что в этом эксперименте значения loss моделей нельзя сравнивать между собой, так как размер словаря разный и функции потерь тоже разные.

Алеся Иванова

bhw2\_answer\_exp2\_20

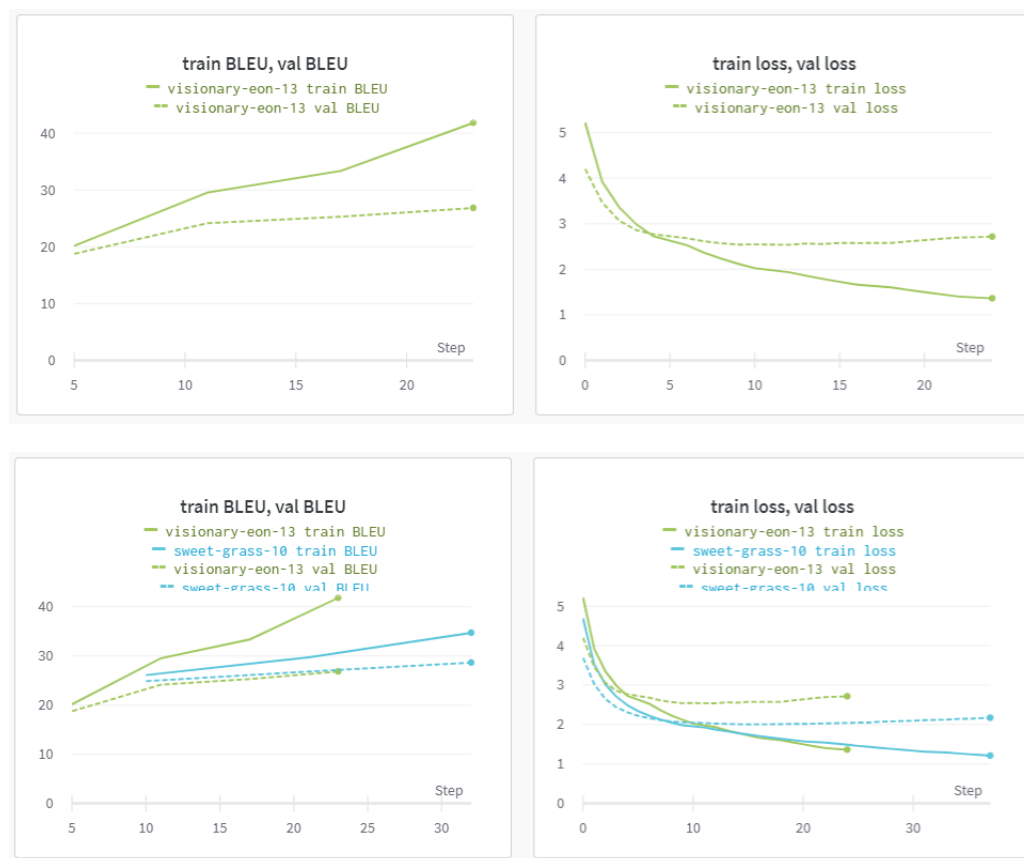
Я принял твоё решение.

Значение BLEU на публичном сплите 24.29.

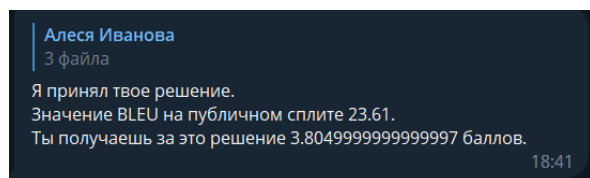
Ты получаешь за это решение 4.145 баллов. 19:36

## Эксперимент 3

Решила попробовать построить словарь из всех слов обучающей выборки. Размер немецкого словаря получился 123554, размер английского словаря получился 56326.

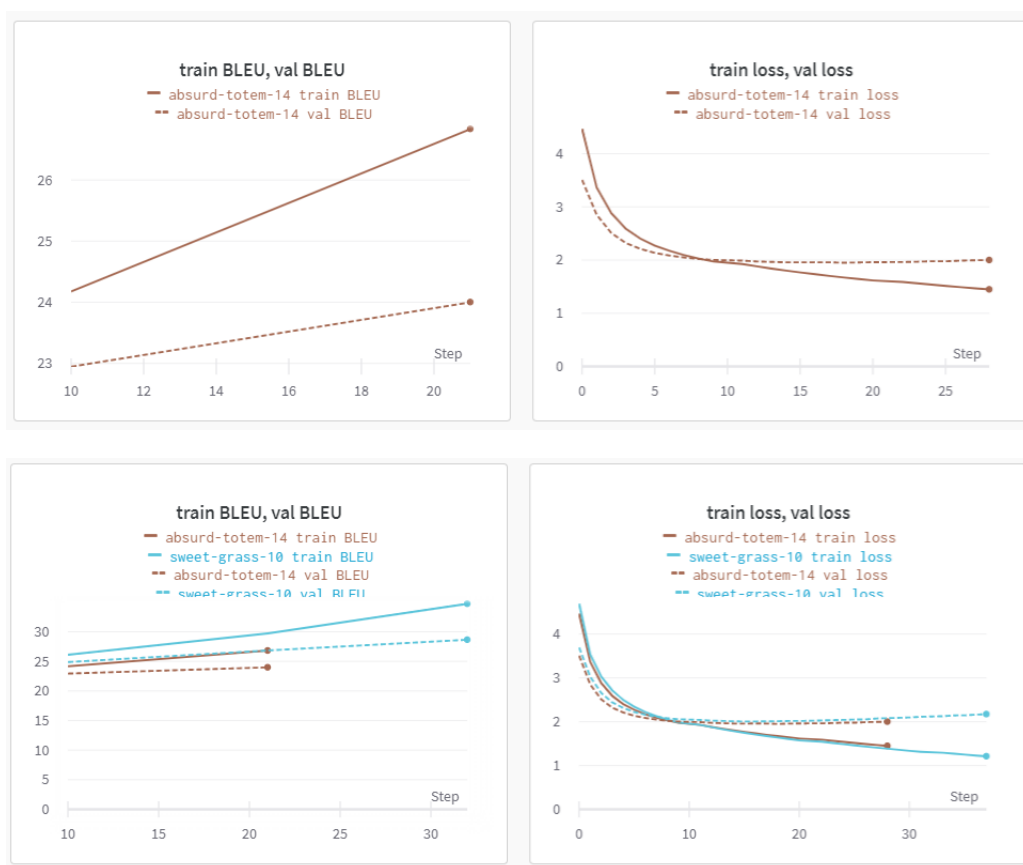


Модель достаточно быстро переобучилась, лучшее качество на тесте было 23.61.

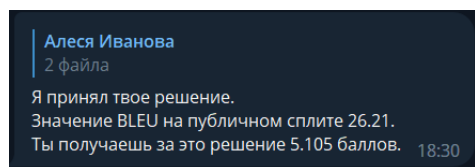


## Эксперимент 4

Решила попробовать другие размеры словаря, попробовала поставить пороговую частоту 20. Размер немецкого словаря получился 8875, размер английского словаря получился 8082.

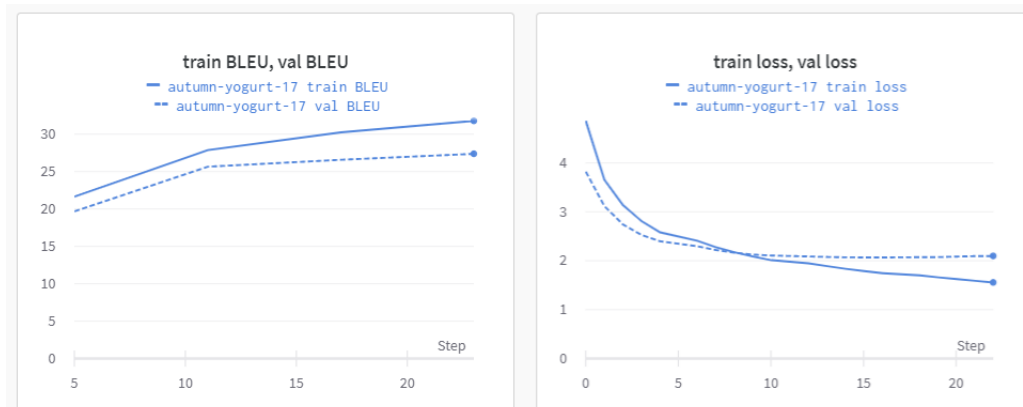


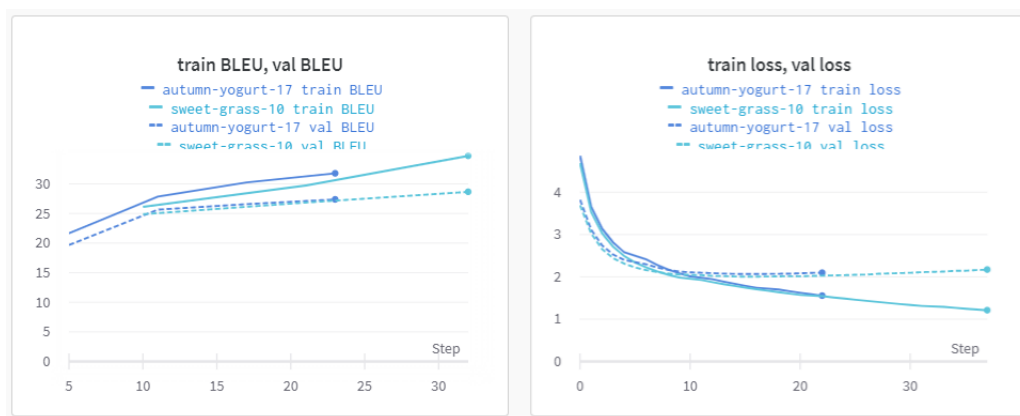
Лучшее качество на тесте было 26.21.



## Эксперимент 5

Сделала пороговую частоту 5. Размер немецкого словаря получился 25379, размер английского словаря получился 19057.





Получила качество на тесте 27.72, то есть примерно как с пороговой частотой 10, но при этом модель быстрее переобучалась, поэтому решила оставить пороговую частоту 10.

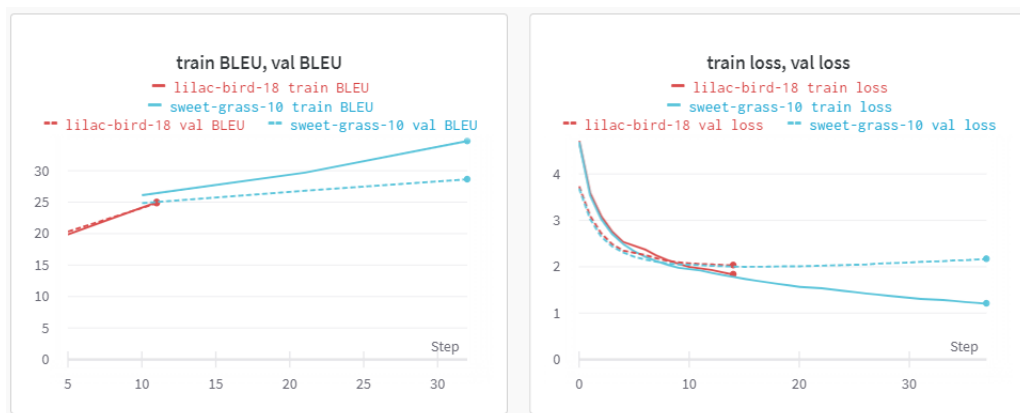
Алеся Иванова  
bhw2\_answer\_exp5\_15

Я принял твоё решение.  
Значение BLEU на публичном сплите 27.72.  
Ты получаешь за это решение 5.859999999999999 баллов.

23:05

## Эксперимент 6

Решила попробовать изменить параметры трансформера, увеличили количество голов вдвое, то есть с 8 до 16. Была гипотеза, что так модель сможет больше зависимостей найти между словами, но качество не улучшилось.



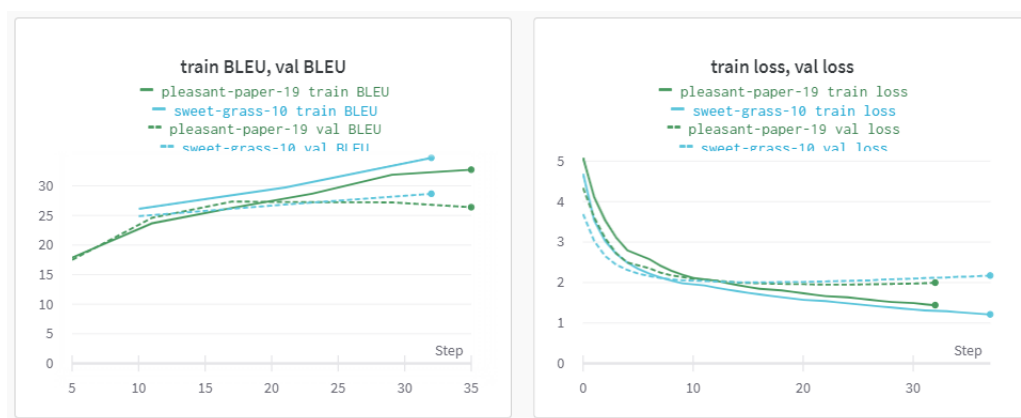
Алеся Иванова  
bhw2\_answer\_exp5\_10 (1)

Я принял твоё решение.  
Значение BLEU на публичном сплите 25.58.  
Ты получаешь за это решение 4.789999999999999 баллов.

1:41

## Эксперимент 7

Решила увеличить вдвое количество слоёв энкодера и декодера, то есть с 3 до 6.



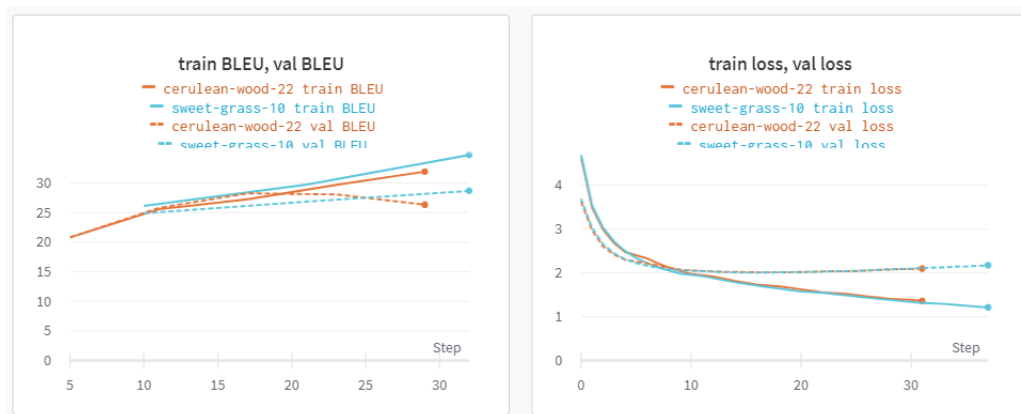
Модель обучалась дольше, но качество на тесте получилось чуть лучше - 27.82 (после 25 эпох).

Алеся Иванова  
dbhw2\_answer\_exp7\_25

Я принял твоё решение.  
Значение BLEU на публичном сплите 27.82.  
Ты получаешь за это решение 5.91 баллов. 22:07

## Эксперимент 8

Решила увеличить размерность Feed-Forward слоёв с 512 до 1024.



Качество модели не улучшилось.

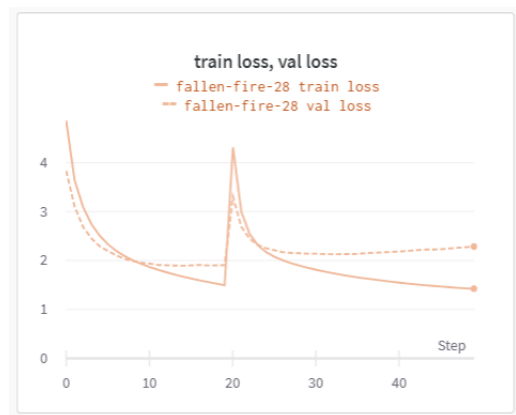
Алеся Иванова  
bhw2\_answer\_exp8\_20

Я принял твоё решение.  
Значение BLEU на публичном сплите 27.13.  
Ты получаешь за это решение 5.564999999999995 баллов. 21:40

## Эксперимент 9

Решила применить Back-translation, то есть сначала обучить модель для перевода с английского на немецкий, затем с её помощью сгенерировать новые пары строк на немецком и английском (переводя английские предложения из обучающей выборки на немецкий с помощью такой модели). То есть по сути получили аугментации текста. И после этого на выборке из изначальных и сгенерированных тестов учила модель для перевода с немецкого на английский.

Архитектуру обеих моделей взяла из эксперимента 1, так как получилось чуть лучше только в эксперименте 7, но там модели обучаются дольше. Количество настоящих и сгенерированных предложений взяла в пропорции 1:1. Вдохновлено статьей <https://arxiv.org/abs/1907.06616>.



На графике сначала loss первой модели, потом второй. Лучшее качество на тесте получилось 28.25 после 20 эпох.

