Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2023”

June 14–16, 2023

Text simplification as a controlled text style transfer task

Tikhonova Maria
HSE University, SberDevices
Moscow, Russia
m_tikhonova94@mail.ru

Fenogenova Alena
SberDevices
Moscow, Russia
alenush93@gmail.ru

Abstract

The task of text simplification is to reduce the complexity of the given piece of text while preserving its original meaning to improve readability and understanding. In this paper, we consider the simplification task as a subfield of the general text style transfer problem and apply methods of controllable text style to rewrite texts in a simpler manner preserving their meaning. Namely, we use a paraphrase model guided by another style-conditional language model. In our work, we perform a series of experiments and compare this approach with the standard fine-tuning of an autoregressive model.
Keywords: text simplification, natural language processing, machine learning, text style transfer
DOI: 10.28995/2075-7182-2023-22-507-516

Задача симплификации текста как задача управляемого переноса стиля

Тихонова Мария
НИУ ВШЭ, SberDevices
Москва, Россия
m_tikhonova94@mail.ru

Феногенова Алена
SberDevices
Москва, Россия
alenush93@gmail.ru

Аннотация

Задача автоматического упрощения текста состоит в том, чтобы уменьшить сложность подаваемого текста с целью улучшения удобочитаемости и понимания, но при этом сохраняя первоначальный смысл. В данной статье мы рассматриваем упрощение текста как задачу переноса стиля (style transfer). Мы исследуем методы управляемой генерации при переносе стиля текста для автоматической генерации упрощенных текстов. А именно, мы используем исходную модель перефразирования текста и дополнительный стилевой дискриминатор (GeDi-classifier), который контролирует выход и направляет генерацию модели в нужный стиль "упрощения"текста. В работе мы проводим серию экспериментов и сравниваем этот подход со стандартным дообучением авторегрессионной модели.
Ключевые слова: автоматическое упрощенние текстов, обработка естественного языка, текстовый стайл трансфер, перенос стиля, генеративные модели

1 Introduction

The goal of text simplification (or TS, in short) is to reduce the linguistic complexity of the given text fragment to improve its readability and to make it easier to understand. Text complexity depends on the presence of participial and adverbial constructions, complex grammatical structures, infrequent and ambiguous words, and subordinate sentences. Thanks to its numerous applications, the TS problem has received significant attention in Natural Language Processing (or NLP). For instance, it may simplify communication for non-native speakers and people with cognitive disorders such as aphasia or dyslexia. In addition, text simplification can improve language model performance on such NLP tasks as semantic role labeling, summarization, information extraction, machine translation, etc.

One standard approach to solving this task is to fine-tune a pre-trained language model on a large text corpus containing aligned complex and simplified sentences.

In this paper, we step aside from this paradigm and consider TS as a text style transfer task, regarding the "simplicity of the text" as a particular style. For this purpose, we use methods of controllable text generation. Namely, the GeDi algorithm proposed in (Krause et al., 2020) and further developed in (Dale et al., 2021). Following their methodology we use a paraphrase model (the main model) guided by another language model conditioned for the "simple" style (or GeDi-classifier). The choice of such an approach was motivated by its several advantages compared to standard fine-tuning of the pre-trained language model. First, it does not change the main language model. The trained GeDi-classifier can be used with different main models (for example, rewriter based on RuT5-Large, rewriter based on RuT5-XL, summarizer based on RuT5-Large, summarizer based on RuT5-Large, etc.), which gives more freedom for its usage. Thus, it simplifies the fine-tuning process as the classifier should only be trained once and then can be used in combination with various main models. Second, we can train several GeDi-classifiers with different target styles (sentiment, simplification, toxicity, etc.) and use them with any of the main language models we have. Thus, we only need to fine-tune M main models and N GeDi-classifiers instead of fine-tuning $N * M$ models for each combination.

In this work, we perform a series of experiments on the simplification dataset from the RuSimpleSentEval-2021 Shared Task (Sakhovskiy et al., 2021). We compare the controllable text style transfer approach with standard fine-tuning of autoregressive language models and show that GeDi-based approach of controllable text style transfer achieves quality comparable with standard fine-tuning.

The rest of the paper is structured as follows: first, in section 2 we overview the papers related to the field of TS and a paraphrase task, which can be regarded as its generalization, as well as the methods for controllable style generation. Next, in section 3 we discuss the controllable text style transfer approach we use. Then, section 4 describes the experimental setup. Section 5 presents evaluation results. Finally, section 6 concludes the paper.

2 Related Work

The task of text simplification is a popular generation task in NLP, useful in many applications: from pre-processing for machine translation to assistive technology for people with cognitive disorders. The systems of TS improve text readability and simplify text understanding while retaining its original information content as much as possible. The automation of this process is a complex problem which has been explored from many points of view. Several good extensive surveys cover the datasets and most of the classical methods for TS problem (Shardlow, 2014), (Al-Thanyyan and Azmi, 2021).

The interest and the development of TS systems for the Russian language rapidly increased with the RuSimpleSentEval Shared Task (Sakhovskiy et al., 2021), for which the authors presented the dataset and baselines. In addition, other Russian datasets exist for TS, among which are ruBTS (Galeev et al., 2021) and the aligned parallel TS dataset from language learner data (Dmitrieva et al., 2022).

The TS task can be considered the sub-task of the paraphrase task due to the similarity of the task definition and criteria of the generated text: the format should be changed while preserving the original text content. For the Russian language, several paraphrase models in the open source are commonly used, for example, paraphrased library (Fenogenova, 2021), or models by David Dale [1]. These models work on the sentence level. In addition, there exist a model from Sber [2] that rewrites extensive texts, which can contain many sentences.

For the evaluation of paraphrase tasks, the standard natural language generation (NLG) metrics are commonly used. There are surface-based metrics such as variations of BLEU, ROUGE, CHRF+; and BERT-base metrics such as LABSE (Feng et al., 2020) and BertScore (Zhang et al., 2019). For instance, their combinations are presented in the GEM benchmark (Gehrmann et al., 2021). Besides, for the TS task, special metrics such as SARI (Xu et al., 2015), included in the EASSE [3] package and Lens (Maddela

[1] https://huggingface.co/cointegrated/rut5-base-paraphraser
[2] https://sbercloud.ru/ru/datahub/rugpt3family/demo-rewrite
[3] https://github.com/feralvam/easse

et al., 2022), were proposed.

The controllable text style transfer approach has received considerable attention in recent years. One of the pioneers in this field was (Keskar et al., 2019), where authors use conditioned controlled codes for guided text generation.

GeDi (Krause et al., 2020) uses a small external language model classifier (or simply GeDi-classifier) to guide the generation of the main language model, re-weighting next token probabilities and, thus, increasing the probabilities of words in the given style. ParaGeDi (Dale et al., 2021) adopts this idea to the paraphrasing task by applying the GeDi approach in combination not with the standard language model but with the paraphraser fine-tuned to rephrase the original text preserving its original meaning.

In (Liu et al., 2021) the authors proposed DExperts. Their approach uses two extra language models conditioned towards and against the desired style (or topic), which are used to re-weight the probabilities of the next tokens predicted by the main language model.

(Yang and Klein, 2021) explores the usage of text classifiers for controllable text generation with FUDGE. This idea is further developed in (Sitdikov et al., 2022), where authors proposed CAIF sampling, which is a method for controllable text generation based on re-weighting logits with a free-form classifier.

Thus, while most methods for controllable text style transfer concentrate on controllable text generation in a given style, we focus on the task of paraphrasing the original text in a given style, preserving the meaning and applying the ideas from the ParaGeDi method for text simplification, regarding the simplicity of the text as a specific style. It should also be noted that while the work ParaGeDi uses GPT-2 language models, we use RuT5-Large based models. In other words, both components are derived from the same pre-trained language model version. Such an approach avoids problems with the difference in the vocabulary in the process of fine-tuning.

In addition, we perform our research for the Russian Language, which distinguishes our work from the papers mentioned above, which concentrate on English.

3 Method

Besides the standard approach of fine-tuning a pre-trained language model used as a baseline for the style-transfer experiments, we consider several versions of controlled text generation models based on the GeDi algorithm proposed in (Krause et al., 2020). In it a language model performs text generation guided by another language model conditioned for the specific topic or style or topic. More precisely, in our work, we adopt the extension of this method presented in (Dale et al., 2021), where the authors enable the model not only to generate but to paraphrase the input text. Below, a brief description of the method is given.

3.1 GeDi

In the original GeDi algorithm, the whole model consists of two parts. The first component is a generation autoregressive model. The second model is an autoregressive discrimination model, trained on sentences labeled with a specific style or topic, which we will further refer to as **GeDi-classifier**. Thus, in the process of training GeDi-classifier learns the word distributions conditioned on a particular label. At each generation step, the distribution of the next token predicted by the main language model P_{LM} is adjusted using the Bayes rule and an additional class-conditional language model P_D:

$$P(x_t|x_{<t}, c) \propto P_{LM}(x_t|x_{<t})P_D(c|x_t, x_{<t})$$

Here, x_t is the current token, $x_{<t}$ is the prefix of the text, and c is the desired style (e.g. simplicity or sentiment) — one of C classes. The first term in the formula is predicted by the main language model P_{LM}. The second term is calculated using GeDi-classifier P_{DC} via the Bayes rule. As a result the tokens which are more likely to appear in a text of the chosen style receive a higher probability:

$$P_D(c|x_t, x_{<t}) \propto P(c)P_{DC}(x, x_{<t}|c)$$

In the original paper, GeDi was successfully used for guided text generation with GPT-2 language model making the generation of the less toxic texts.

3.2 ParaGeDi

In our work, we adopt the approach of ParaGeDi, where the authors enable GeDi to preserve the meaning of the input text. For this, they replace the language model with a paraphraser. Thus, ParaGeDi models the following probability:

$$P(y_t|y_{<t}, x, c) \propto P_{LM}(y_t|y_{<t}, x)P(c|y_t, y_{<t}, x) \approx P_{LM}(y_t|y_{<t}, x)P_D(c|y_t, y_{<t})$$

where x is the original text, y is the generated text of length T, and c is the desired style.

The last transition in the equation above is an approximation which allows us to decouple the paraphraser from the GeDi-classifier model. As a result, the paraphraser and the GeDi-classifier can be trained independently in such a formulation.

As for the training process, ParaGeDi loss $\mathcal{L}_{ParaGeDi}$ consists of two components: the generative loss $\mathcal{L}_G$ used in language model training and the discriminative loss $\mathcal{L}_D$ which further pushes different classes away from one another.

$$\mathcal{L}_G = -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i}\sum_{t=1}^{T_i}\log P(y_t^{(i)}|y_{<t}^{(i)}, c^{(i)})$$

$$\mathcal{L}_D = -\frac{1}{N}\sum_{i=1}^{N}\log P(c^{(i)}|y_{1:T_i}^{(i)})$$

$$\mathcal{L}_{ParaGeDi} = \lambda\mathcal{L}_D + (1-\lambda)\mathcal{L}_G$$

where $\lambda \in [0, 1]$ is the weight of the discriminative loss.

Besides, to improve the preservation of the original content and to increase the style transfer accuracy, the following heuristics are used:

First, the conditional language model probability is raised to the power $w > 1$, which biases the discriminator towards the correct class in the process of generation:

$$P(y_t|y_{<t}, x, c) \propto P_{LM}(y_t|y_{<t}, x)P_{CC}(c|y_t, y_{<t})^w$$

Second, probabilities are smoothed by adding a small $\alpha > 0$ to all probabilities from the conditional language model:

$$P_\alpha(c|x_t, x_{<t}) = \frac{\alpha + P(c)P_{CC}(x, x_{<t}|c)}{\sum_{c' \in C}(\alpha + P(c')P_{CC}(x, x_{<t}|c'))}$$

Such a heuristic discourages the generation of tokens with low probability conditional on all classes.

Third, for class-conditional corrections, asymmetric lower and upper bounds (l and u) are used :

$$P_{\alpha,l,u}(c|x_t, x_{<t}) = \max(l, \min(u, P_\alpha(c|x_t, x_{<t}))).$$

This discourages the insertion of new tokens, as opposed to prohibiting existing tokens.

4 Experiments

4.1 Data

We perform a series of experiments on the dataset RuSimpleSentEval-2021 Shared Task (Sakhovskiy et al., 2021). This simplification dataset contains parallel pairs of sentences: complex – their corresponding simplified versions. Below, a sample from the dataset is presented.

Example from the dataset:

Source sentence:

“Климат Казани – умеренно континентальный , сильные морозы и палящая жара редки и не характерны для города”

Simplified paraphrases:

1. *“В Казани редко бывают и сильные морозы, и жаркая летняя погода”*
2. *“В Казани зимой не слишком холодно, а летом не слишком жарко”*
3. *“В Казани зимой не очень холодно, а летней жары почти не бывает”*

The organizers of the RuSimpleSentEval-2021 shared task constructed the dataset using automatic translation and post-processing WikiLarge corpus (Zhang and Lapata, 2017). The resulting dataset was split into the train, dev and two test sets (public and private). And while the train set was not filtered or verified, the organizers validated the dev, public and private test sets via crowd-sourcing using Yandex.Toloka [4] and filtered them. In this work, we evaluate the results on official public and private test sets. We additionally filtered the train part, which contains inappropriate examples due to its original automatic construction. For its cleaning, we used the following procedure: exclude examples with less than two lemmas in the intersection between the lemmatized source and target sentences (lemmatization was done via pymorphy2 [5] tagger (Korobov, 2015)); discard examples where the source sentence is a sub-string of the target one and the length is bigger than of the source one. Besides training and validation, we also use extra dev set filtered by the organizer.

4.2 Models

We conduct experiments and compare the results of the following models:

- **Golden testset**. We evaluate the golden references (first answer) from the fixed RuSimpleSentEval-2021 test sets (public/private);
- **Paraphraser**. We use a paraphrase model [6] trained on 7000 examples from different sources of various domains: 1) text level: literature domain, prose; back translation (with ru-en translation model [7]) of the texts from different domains filtered with Bertscore Rouge-L); 2) sentence level: Russian version of Tapaco corpus (Scherrer, 2020) and filtered ParaphraserPlus (Gudkov et al., 2020) corpus.
- **Fine-tuned paraphraser**. We additionally fine-tune the paraphrase model on the train set to check the hypothesis of the capabilities combinations that the model learn (both paraphrasing and simplification);
- **Fine-tuned ruT5-Large** [8]. We fine-tune the row ruT5-Large model on the simplification train set.
- **ParaGeDi**. We train GeDi-classifier on the train part of the RuSimpleSentEval-2021 set and use the paraphrase model described above as the main model for ParaGeDi controllable approach.

In our work, all models we use are derived from the pre-trained RuT5-Large[9] model, which is a T5 model (Raffel et al., 2020) pre-trained for the Russian language. The fact that we derive both components from the same model allows us to avoid problems with the difference in the model vocabulary.

As for the GeDi-classifier model, we fine-tune RuT5-Large on the RuSimpleSentEval-2021 Shared Task train set. We use the Adam optimizer with the learning rate $1e-4$, three epochs, and the weight of the discriminative loss $\lambda = 0.3$.

We evaluate several style power coefficients ($w = 5, 10, 15, 20$). It should also be noted that we do not evaluate $w = 0$ as, in this case, the influence of the GeDi-calssifier is neglected, and the result is equal to the original paraphrase model, which is our baseline. To avoid randomness, we use the following generation parameters:

[4] https://toloka.ai/tolokers/
[5] https://github.com/pymorphy2/pymorphy2
[6] `https://habr.com/ru/company/sberdevices/blog/667106/`
[7] `https://huggingface.co/Helsinki-NLP/opus-mt-en-ru`
[8] https://huggingface.co/sberbank-ai/ruT5-large
[9] `https://huggingface.co/sberbank-ai/ruT5-large`

- $do\_sample = False$,
- $num\_returned\_sequences = 1$,
- $max\_len = 128$.

4.3 Metrics

We evaluate the model on public and private test sets of RuSimpleSentEval-2021 Shared Task using the following metrics:

- **BertScore**(Zhang et al., 2019), which is computed between the original (complex) sentences and model predictions.
- **SARI** (Xu et al., 2016), which is commonly recognized as a metric for evaluating automatic text simplification systems. The metric compares the model predictions against the references and the original (complex) sentences.
- **BLEU score**(Papineni et al., 2002), which in our case is computed between the reference answers and predictions
- **iBLEU** (Sun and Zhou, 2012) which is computed as follows:

$$iBLEU = a * BLEU(preds, refs) + (1 - \alpha) * BLEU(preds, source),$$

  where α is the parameter responsible for the balance between adequacy and dissimilarity. In our work, we follow the methodology from the original paper and use $\alpha = 0.9$.
- **Diversity** We report a degree of diversity measured using the mean number of distinct n-grams, normalized by the length of text (Li et al., 2015). We report dist-1, dist-2, and dist-3 scores for distinct uni-, bi-, and trigrams, respectively.

5 Results

Results on public and private test sets are presented in Tables 1 and 2, respectively. The results reveal that the GeDi-based approach with style power coefficients of 5 and 10 shows quality comparable with the standard fine-tuning approach. Larger values of the style power coefficient lead to a decrease in quality as the classifier influence becomes too strong, which negatively affects the generated output. Thus, the ParaGeDi-based approach can be considered a good alternative to standard fine-tuning. In addition, as long as it does not change the initial model and can be used with different main models, it gives more freedom for its usage.

Model	**BertScore**	**SARI**	**BLEU**	**iBLEU 0.9**	**dist 1**	**dist 2**	**dist 3**
Golden testset	0.816874	**66.106573**	1.0	0.916141	0.971855	**0.940157**	**0.882364**
Paraphraser	0.925663	41.004799	0.314653	0.342387	0.964854	0.923054	0.855773
FT paraphraser	**0.970198**	41.594171	0.367276	0.412937	0.974326	0.932282	0.866955
FT ruT5-Large	0.969541	41.819602	**0.369884**	**0.415395**	**0.974098**	0.931853	0.866066
ParaGeDi (sp 5)	0.914065	40.792974	0.310180	0.332548	0.965152	0.919561	0.848917
ParaGeDi (sp 10)	0.888886	40.501325	0.295284	0.307751	0.969362	0.911230	0.831918
ParaGeDi (sp 15)	0.826108	38.539389	0.256159	0.255457	0.882723	0.815006	0.731320
ParaGeDi (sp 20)	0.659992	33.045052	0.081489	0.075360	0.401245	0.356622	0.307940

Table 1: The results on the public test set of the RuSimpleSentEval-2021. ParaGeDi is evaluated with different Style Power coefficients (sp in shortly). *FT* stands for fine-tuned. Detailed metrics descriptions are given in subection 4.3.

In addition, we compared our results with the top-3 solutions of the RuSimpleSentEval-2021 competition (Sakhovskiy et al., 2021), which include *qbic* solution based on Multilingual Unsupervised Sentence Simplification (Martin et al., 2020) and fine-tuned GPT-based solutions by *orzhan*, *ashatilov*, and *alenusch*. To complete the picture, we also included mBART-based (Liu et al., 2020) baseline presented by the organizers. Results are presented in Table 3. First, it can be seen that all our solutions (which are RuT5-based) surpass the baseline. Second, most of them, including the ParaGeDi method with reasonable style power coefficient of 5 and 10, outperform competition winners (mostly GPT-based) showing

Model	BertScore	SARI	BLEU	iBLEU 0.9	dist 1	dist 2	dist 3
Golden testset	0.816874	**66.106573**	1.0	**0.967823**	0.940655	0.883676	**0.882364**
Paraphraser	0.92467	40.418701	0.301265	0.330843	0.961526	0.922913	0.857691
FT paraphraser	**0.968782**	41.643578	**0.358353**	0.404432	0.968473	**0.931082**	0.866247
FT ruT5-Large	0.965881	41.517535	0.357556	0.402777	**0.969426**	0.929413	0.863115
ParaGeDi (sp 5)	0.912825	40.859850	0.300608	0.324721	0.961111	0.918092	0.848473
ParaGeDi (sp 10)	0.887088	40.240902	0.274954	0.289805	0.960448	0.907891	0.830453
ParaGeDi (sp15)	0.824515	38.249361	0.255155	0.255730	0.873924	0.810920	0.730028
ParaGeDi (sp 20)	0.668402	33.238699	0.098595	0.091794	0.432894	0.389271	0.339774

Table 2: Simplification results on the private test set. ParaGeDi is evaluated with different Style Power coefficients (sp in shortly). *FT* stands for fine-tuned. Detailed metrics descriptions are given in subection 4.3.

higher SARI scores. Such results can be regarded as another proof of the quality of the ParaGeDi approach. In addition, such results indicates that RuT5 is a better backbone for the text simplification task than the GPT-based models. We observe the same trends on the TS task in the GEM benchmark [10]. The T5-small model shows the best performance on the analogous datasets for English, among which are wiki auto, asset turk, and test turk datasets (Xu et al., 2016)).

Model	SARI
Golden testset	**66.106**
FT ruT5-Large	**41.819**
FT paraphraser	**41.594**
Paraphraser	**41.004**
ParaGeDi (sp 5)	**40.792**
ParaGeDi (sp 10)	**40.501**
№1 orzhan	40.233
№2 alenusch	38.870
№3 ashatilov	38.843
ParaGeDi (sp 15)	38.539
ParaGeDi (sp 20)	33.045
BASELINE	30.152

Model	SARI
Golden testset	**66.106**
FT paraphraser	**41.643**
FT ruT5-Large	**41.517**
Paraphraser	**40.418**
ParaGeDi (sp 5)	**40.859**
ParaGeDi (sp 10)	**40.240**
№1 qbic	39.689
№2 orzhan	39.279
№3 ashatilov	38.491
ParaGeDi (sp 15)	38.249
ParaGeDi (sp 20)	33.238
BASELINE	-

Table 3: Simplification results comparison with the top-3 solutions of RuSimpleSentEval-2021 on public (left) and private (right) test sets respectively. Following the original competition approach, we compare results using the SARI metric. *sp* stands for style power. *FT* stands for fine-tuned. *BASELINE* stands for the mBART-based baseline from RuSimpleSentEval-2021 competition.

Analyzing model's performance on the concrete examples, we explored the behavior of Fine-tuned paraphraser, which showed best performance on the private test set (see Table 4). It can be seen that the model preserves the original meaning of the sentence and does not change facts. However, it only slightly simplifies sentences replacing individual words rather than rewriting the whole sentence in a more simplistic style. Thus, there is still room for improvement.

As for the limitations of the ParaGeDi method, with the growth of the style power coefficient, ParaGiDi starts to lose coherence and forget the original sentence's meaning. While the generations with the style power coefficients 5 or 10 yields reasonable quality, sentences generated with the coefficient 20 often lack

[10]https://gem-benchmark.com/results

Original sentence	Prediction
Дорожка следов у волка ровнее и образует почти ровную строчку, а у собак — извилистую линию.	Дорожка следов у волка ровная, а у собак — извилистая линия.
В том же году компания открыла офис в Казани; главная страница портала, некоторые службы и поиск стали доступны как на русском, так и на татарском языке.	В том же году компания открыла офис в Казани. Многие службы и поиск стали доступны как на русском, так и на татарском языке.
В связи с появившейся необходимостью медицинского обследования врачи приняли решение направить его в одну из московских клиник.	В связи с возникшей необходимостью медицинского обследования врачи отправили его в одну из московских клиник.
В рационе древних египтян присутствовали горох, бобы и нут, огурцы, в больших выращивался салат-латук.	В рационе древних египтян были горох, бобы и огурцы, в большом количестве выращивался салат-латук.
Атлантические течения, разогретые Гольфстримом, приносят мягкие зимы; иногда зимой и ранней весной здесь бывают снегопады, хотя снег обычно лежит недолго.	Атлантические течения приносят мягкие зимы, и иногда зимой и ранней весной здесь бывают снегопады, хотя снег обычно лежит недолго.

Table 4: Fine-tuned paraphraser examples from the test set.

meaning. In addition, as long as the ParaGeDi approach uses two language models, it works slower and requires more computational resources during the inference stage compared to the fine-tuned language models.

6 Conclusion

In this paper, we dealt with the text simplification problem regarding it as a special case of text style transfer task. We adopted the ParaGeDi method, which uses the idea of controlled text style transfer. We used the combination of two RuT5-Large models (paraphrase model and GeDi-classifier) to solve this task. In the experiments, that approach proved quite promising; the results are comparable to fine-tuning for the single style class. The ruT5-based simplification models surpassed the best results on the RuSimpleSentEval-2021 shared task.

As a part of future research, we plan to consider the reverse problem of making the text more complex and official. Thus, we plan to explore the capabilities of the models, which can work in both directions: simplifying the text or making it more complex and official.

References

Suha S Al-Thanyyan and Aqil M Azmi. 2021. Automated text simplification: a survey. *ACM Computing Surveys (CSUR)*, 54(2):1–36.

David Dale, Anton Voronov, Daryna Dementieva, Varvara Logacheva, Olga Kozlova, Nikita Semenov, and Alexander Panchenko. 2021. Text detoxification using large pre-trained neural models. *arXiv preprint arXiv:2109.08914*.

Anna Dmitrieva, Antonina Laposhina, and Maria Yuryevna Lebedeva. 2022. Creating a list of word alignments from parallel russian simplification data. *Frontiers in Artificial Intelligence*, 5:984759.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2020. Language-agnostic bert sentence embedding. *arXiv preprint arXiv:2007.01852*.

Alena Fenogenova. 2021. Russian paraphrasers: Paraphrase with transformers. // *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, P 11–19, Kiyv, Ukraine, April. Association for Computational Linguistics.

Farit Galeev, Marina Leushina, and Vladimir Ivanov. 2021. rubts: Russian sentence simplification using back-translation. // *Komp'juternaja Lingvistika i Intellektual'nye Tehnologii*, P 259–267.

Sebastian Gehrmann, Tosin Adewumi, Karmanya Aggarwal, Pawan Sasanka Ammanamanchi, Anuoluwapo Aremu, Antoine Bosselut, Khyathi Raghavi Chandu, Miruna-Adriana Clinciu, Dipanjan Das, Kaustubh Dhole, Wanyu Du, Esin Durmus, Ondřej Dušek, Chris Chinenye Emezue, Varun Gangal, Cristina Garbacea, Tatsunori Hashimoto, Yufang Hou, Yacine Jernite, Harsh Jhamtani, Yangfeng Ji, Shailza Jolly, Mihir Kale, Dhruv Kumar, Faisal Ladhak, Aman Madaan, Mounica Maddela, Khyati Mahajan, Saad Mahamood, Bodhisattwa Prasad Majumder, Pedro Henrique Martins, Angelina McMillan-Major, Simon Mille, Emiel van Miltenburg, Moin Nadeem, Shashi Narayan, Vitaly Nikolaev, Andre Niyongabo Rubungo, Salomey Osei, Ankur Parikh, Laura Perez-Beltrachini, Niranjan Ramesh Rao, Vikas Raunak, Juan Diego Rodriguez, Sashank Santhanam, João Sedoc, Thibault Sellam, Samira Shaikh, Anastasia Shimorina, Marco Antonio Sobrevilla Cabezudo, Hendrik Strobelt, Nishant Subramani, Wei Xu, Diyi Yang, Akhila Yerukola, and Jiawei Zhou. 2021. The GEM benchmark: Natural language generation, its evaluation and metrics. // *Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*, P 96–120, Online, August. Association for Computational Linguistics.

Vadim Gudkov, Olga Mitrofanova, and Elizaveta Filippskikh. 2020. Automatically ranked russian paraphrase corpus for text generation. *arXiv preprint arXiv:2006.09719*.

Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. 2019. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*.

Mikhail Korobov. 2015. Morphological analyzer and generator for russian and ukrainian languages. // Mikhail Yu. Khachay, Natalia Konstantinova, Alexander Panchenko, Dmitry I. Ignatov, and Valeri G. Labunets, *Analysis of Images, Social Networks and Texts*, volume 542 of *Communications in Computer and Information Science*, P 320–332. Springer International Publishing.

Ben Krause, Akhilesh Deepak Gotmare, Bryan McCann, Nitish Shirish Keskar, Shafiq Joty, Richard Socher, and Nazneen Fatema Rajani. 2020. Gedi: Generative discriminator guided sequence generation. *arXiv preprint arXiv:2009.06367*.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2015. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.

Alisa Liu, Maarten Sap, Ximing Lu, Swabha Swayamdipta, Chandra Bhagavatula, Noah A Smith, and Yejin Choi. 2021. Dexperts: Decoding-time controlled text generation with experts and anti-experts. *arXiv preprint arXiv:2105.03023*.

Mounica Maddela, Yao Dou, David Heineman, and Wei Xu. 2022. Lens: A learnable evaluation metric for text simplification. *arXiv preprint arXiv:2212.09739*.

Louis Martin, Angela Fan, Éric De La Clergerie, Antoine Bordes, and Benoît Sagot. 2020. Muss: multilingual unsupervised sentence simplification by mining paraphrases. *arXiv preprint arXiv:2005.00352*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. // *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, P 311–318.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

Andrey Sakhovskiy, Alexandra Izhevskaya, Alena Pestova, Elena Tutubalina, Valentin Malykh, Ivan Smurov, and Ekaterina Artemova. 2021. Rusimplesenteval-2021 shared task: evaluating sentence simplification for russian. // *Proceedings of the International Conference "Dialogue*, P 607–617.

Yves Scherrer. 2020. Tapaco: A corpus of sentential paraphrases for 73 languages. // *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association (ELRA).

Matthew Shardlow. 2014. A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications*, 4(1):58–70.

Askhat Sitdikov, Nikita Balagansky, Daniil Gavrilov, and Alexander Markov. 2022. Classifiers are better experts for controllable text generation. *arXiv preprint arXiv:2205.07276*.

Hong Sun and Ming Zhou. 2012. Joint learning of a dual smt system for paraphrase generation. // *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, P 38–42.

Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3:283–297.

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415.

Kevin Yang and Dan Klein. 2021. Fudge: Controlled text generation with future discriminators. *arXiv preprint arXiv:2104.05218*.

Xingxing Zhang and Mirella Lapata. 2017. Sentence simplification with deep reinforcement learning. *arXiv preprint arXiv:1703.10931*.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.