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

June 14–16, 2023

Aspect-based Argument Generation in Russian

Valeriya Goloviznina
Vyatka State University, Kirov, Russia
golovizninavs@gmail.com

Irina Fishcheva
Vyatka State University, Kirov, Russia
fishchevain@gmail.com

Tatiana Peskisheva
Vyatka State University, Kirov, Russia
peskisheva.ta@gmail.com

Evgeny Kotelnikov
Vyatka State University, Kirov, Russia
kotelnikov.ev@gmail.com

Abstract

The paper explores the argument generation in Russian based on given aspects. An aspect refers to one of the sides or property of the target object. Five aspects were considered: "Safety", "Impact on health", "Reliability", "Money", "Convenience and comfort". Various approaches were used for aspect-based generation: fine-tuning, prompt-tuning and few-shot learning. The ruGPT-3Large model was used for experiments. The results show that traditionally trained model (with fine-tuning) generates 51.6% of the arguments on given aspects, with the prompt-tuning approach – 33.9%, and with few-shot learning – 10.6%. The model also demonstrated the ability to generate arguments on new, previously unknown aspects.

Keywords: argumentation mining; controlled texts generation; GPT; fine-tuning; prompt-tuning; few-shot learning

DOI: 10.28995/2075-7182-2023-22-117-129

Аспектно-ориентированная генерация аргументов на русском языке

Головизнина В. С.
Вятский государственный университет, Киров, Россия
golovizninavs@gmail.com

Фищева И. Н.
Вятский государственный университет, Киров, Россия
fishchevain@gmail.com

Пескишева Т. А.
Вятский государственный университет, Киров, Россия
peskisheva.ta@gmail.com

Котельников Е. В.
Вятский государственный университет, Киров, Россия
kotelnikov.ev@gmail.com

Аннотация

В статье исследуется генерация аргументов на русском языке с учетом аспектов. Под аспектом понимается одна из сторон или свойство целевого объекта. Рассматривались пять аспектов: «Безопасность», «Влияние на здоровье», «Надежность», «Деньги», «Удобство и комфорт». Для аспектно-ориентированной генерации применялись различные подходы: fine-tuning, prompt-tuning и few-shot learning. Для экспериментов использовалась модель ruGPT-3Large. Результаты показывают, что модель, дообученная традиционным способом (fine-tuning), генерирует 51.6% доводов по требуемым аспектам, при подходе prompt-tuning – 33.9%, а при few-shot learning – 10.6%. Также модель продемонстрировала способность генерировать аргументы по новым, ранее неизвестным аспектам.

Ключевые слова: анализ аргументации; управляемая генерация текстов; GPT; fine-tuning; prompt-tuning; few-shot learning

1 Introduction

One of the important directions in the field of controlled text generation is the generation of argumentative texts [2], [10], [12]. An argument is a combination of a claim and at least one premise supporting or refuting that claim [13] (see Figure 1). The claim expresses the author's point of view on the controversial issue. The point of view includes the author's stance and the topic (or target). For example, in the claim "Electric cars are better than ordinary cars", the target is electric cars and the stance is "for".

To support or refute the claim, premises[1] "for" or "against" can be given, respectively.

Each premise describes one or more aspects of target. Aspect is a word or phrase that indicates one of the sides or property of the target. For example, the rebuttal premise "Battery costs have more than halved in the last four years alone" mentions the "Money" aspect.

Aspect-based argument generation allows to tune the meaning of the generated premises. However, at present there are very few studies in the field of aspect-based argument generation for the English language [12], and, to the best of our knowledge, there are no such studies for the Russian language.

We are trying to fill this gap. Applying various approaches, we train the ruGPT-3Large model on the Russian-language corpus of arguments with annotated aspects. Five aspects were considered: "Safety", "Impact on health", "Reliability", "Money", "Convenience and comfort". For aspect-based generation, the following methods were used: fine-tuning, prompt-tuning and few-shot learning.

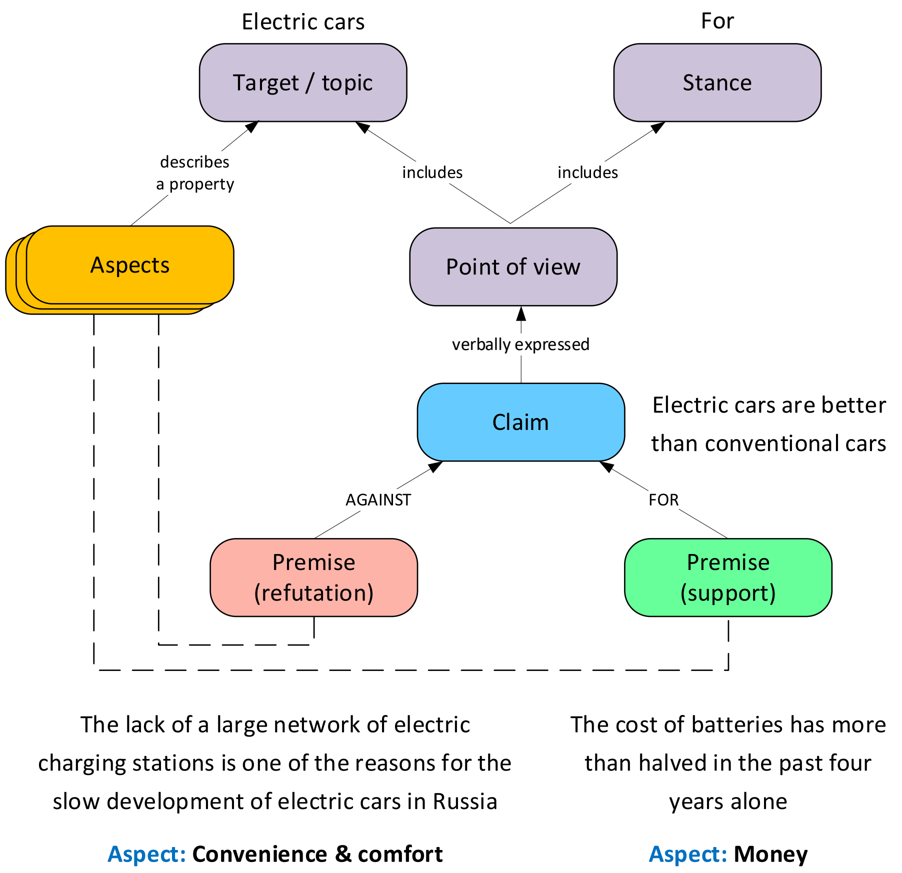

Figure 1: Argument structure. The claim "Electric cars are better than conventional cars", which expresses the stance "for" regarding the target (electric cars), is supported by the premise with the aspect "Money" and is refuted by the premise with the aspect "Convenience and comfort"

[1] Often a "premise" is called an "argument" when it is clear from the context which claim it is being referred to.

The contributions of our work are as follows:
- for the first time in the Russian language the methods of aspect-based argument generation are studied;
- the possibilities of models for arguments generation for new, unfamiliar aspects are analyzed;
- the best-scoring fine-tuned model is made publicly available [2].

2 Previous work

In this section, we first review general approaches to controlled text generation and then provide an overview of the work on aspect-based argument generation.

2.1 Controlled Text Generation

Controlled text generation refers to the task of generating text according to a given controlled element [14]. The main idea of controlled text generation based on pre-trained language models is to give the model a control signal in an explicit or implicit way to control the generation of text that satisfies given conditions. Zhang et al. [14] identify several approaches to controlled text generation.

Fine-tuning consists in tuning the parameters of the whole model or a part of it to generate text that meets specific conditions. In addition to traditional fine-tuning, there are other methods: adding an adapted module, using a prompt, and reinforcement learning.

Adding an adapted module is the construction of an additional module for solving a specific problem [14]. During the training process, the parameters of the language model are frozen, only a special module is trained.

Using a prompt is selecting an input sequence template and using it as a control hint for the language model to generate the required texts. Templates can be selected manually or automatically. The few-shot and zero-shot methods [1] involve manual selection of the prompt. The prefix tuning [5], p-tuning [6], or prompt tuning [4] methods allow to select the prompt automatically. In this case, the vectors corresponding to the prompt are tuned during the training process, while the parameters of the language model remain unchanged.

The main idea of methods based on reinforcement learning is to get feedback on whether the control conditions are achieved as a reward for fine-tuning of the language model [14].

Retrain or refactoring is a change in the original architecture of the language model or retraining of the model from scratch in accordance with the characteristics of a given task [14]. This approach can improve the quality and controllability of text generation, but is limited by the lack of tagged data and the high consumption of computing resources.

During **post-processing**, the parameters of the language model are fixed [14]. For the input sequence, the language model creates an initial distribution of tokens, the post-processing module re-ranks this distribution, ensuring that the model selects the desired token, thus controlling the generation of text.

2.2 Aspect-based Argument Generation

The problem of aspect-based generation of arguments has not yet been studied enough. Schiller et al. [12] apply fine-tuning of the CTRL model on sequences that include control codes [Topic][Stance][Aspect] (for example, *Nuclear Energy CON radioactive waste*) and a premise (for example, *Nuclear reactors produce radioactive waste…*) for the controlled generation of premises on a given topic, stance and aspect.

In paper [2], to generate premises on economic topics, the original ruGPT-3Large model and the same model fine-tuned on an argument corpus containing 3,500 sentences were used. As a result of manual evaluation, 63.2% of the sentences generated by the fine-tuned ruGPT-3Large model turned out to be premises, while the original model without fine-tuning was able to generate only 42.5% of premises.

In our work, in contrast to [2] and [12], in addition to the traditional fine-tuning of the whole model, the following methods for controlled argument generation are studied:
- traditional fine-tuning of the whole model and fine-tuning of the last layer only;
- using of prompt-tuning;
- using a few-shot manual prompt.

[2] https://tinyurl.com/452euk4w.

In [2], the aspect of the premise is not taken into account, and in [12], the aspect is part of the premise. In our work, the aspect reflects the semantic orientation of the premise and is not part of the sentence containing the premise. For example[3]:

- topic: school uniforms,
- premise: outsiders who do not belong to the campus are easy to identify and therefore do not pose much of a threat to students.

In [12], the following aspects are indicated: [outsiders, easy to identify, threat]; in our work, such an premise would have the "Security" aspect.

3 Materials and Methods

3.1 Corpora

We use the existing corpus of premises with aspects specified for them[4].

The corpus contains 548 premises that have from 1 to 3 aspects from the list containing 20 aspects, such as "Safety", "Living standard", "Quality", etc. A complete list of aspects with their frequency is given in Appendix A. We have combined the most similar aspects "Impact on health" and "Impact on the psyche", "Price" and "Profitability", replacing them with aspects "Impact on health" and "Money" respectively. From this corpus, we have identified the most frequently aspects (they met more than 80 times) and the sentences corresponding to them. Thus, we have formed a corpus for the generation, which includes 418 unique argumentative sentences. For each sentence, from 1 to 3 aspects are selected from the following list: "Safety", "Impact on health", "Reliability", "Money", "Convenience and comfort". Since one premise can have several aspects, we used 507 argumentative sentences (with repetitions) to train the models.

The corpus contains 14 topics. Each topic is reflected in the claim about which the argument is built. For example, the topic "Cryptocurrency" in the corpus is represented by the claim "We need to use and invest in cryptocurrencies", the topic "Children's video blogs" is represented by the claim "Children should be encouraged to create vlogs", and the topic "Esports" is represented by the claim "Esports should be made an Olympic sport". A complete list of topics, claims to them and the distribution of topics by aspects are presented in Appendix B.

3.2 Language Model

Models of the GPT (Generative Pre-trained Transformer) family consist of a Transformer decoder with a different number of layers [8]. The ruGPT-3 model is a Russian-language model from Sber, based on GPT-2 [9], available in five versions of different sizes (ruGPT-3Small, ruGPT-3Medium, ruGPT-2Large, ruGPT-3Large, ruGPT-3XL) [11]. The model was trained on 80 billion tokens. In our experiments, we used the ruGPT-3Large model (760M parameters).

3.3 Training Methods

For controlled argument generation, we explore several methods of model training: fine-tuning, prompt-tuning, and few-shot learning.

In traditional fine-tuning, the weights of the model change, adjusting to the required task – generating premises. The ruGPT-3Large model is fine-tuned on text sequences containing a claim, an aspect, and a premise. The input of the model is a sequence of the form:

Claim: {*claim*}; **Aspect:** {*aspect*}; **Premise:** {*premise*}

Parts of the sequence in bold are keywords; instead of parts in curly brackets, real claims, aspects, and premises are substituted. When testing, the input of the model is the following sequence:

Claim: {*claim*}; **Aspect:** {*aspect*}; **Premise:**

[3] Example is taken from the UKP corpus [12].
[4] https://github.com/kotelnikov-ev/RuArgumentMining/tree/main/AspectCorpus.

The model generates a premise by continuing the sentence. Two variants of this approach are considered:

- fine-tuning of the whole model,
- fine-tuning of only the last layer.

In the prompt-tuning method, the vectors that serve as the prompt for the generation control are trained, the model weights are frozen. The prompt vectors obtained during the training process are fed to the input of the model along with embeddings representing text tokens. The input sequence looks like:

<P**n*>**{*claim*}**<P****m*>**{*aspect*}**<P****k*>**,

where $m, n, k \geq 0$ are numbers that indicate the number of special <**P**> tokens in a specific prompt format, claims and aspects are substituted for curly braces.

In the few-short learning method, a prompt is supplied to the model input, including generation examples that describe the task. In our work, these are examples of arguments that include a claim, an aspect, and a premise. In this method, neither the model nor additional vectors are trained. The model input in the few-shot learning contains a prompt that includes 16 examples (limited by GPU memory) written as:

Claim: {*claim*}; **Aspect:** {*aspect*}; **Premise:** {*premise*}

and ending with the sequence:

Claim: {*claim*}; **Aspect:** {*aspect*}; **Premise:**

The model is asked to generate a premise on the last specified claim and aspect.

In this method, we use two types of prompts:

- the prompt contains examples of premises for all aspects in accordance with the distribution of aspects in the original corpus,
- the prompt contains examples of premises only on the aspect of the generated premise.

4 Experiments

4.1 Experimental Setup

We considered several options for formats of prompt in prompt-tuning:

- <P*100>{claim}<P*20>{aspect}<P*100>,
- <P*100>{claim}<P*4>{aspect}<P*20>,
- <P*60>{claim}<P*4>{aspect}<P*60>,
- <P*20>{claim}<P*20>{aspect}<P*20>,
- <P*60>{claim}<P*1>{aspect}.

To implement this approach, we used the ru-prompts library[5]. When choosing the number of special tokens, we were guided by training examples provided by the library developers, which used sequences of 100, 20, and 4 special tokens. We also added variants of the prompt formats, with the same or close total value of the number of special tokens, but arranged differently in the sequence. The second and third formats have the same number of special tokens, but they have a different arrangement in the sequence, similarly for the third and fourth options.

With the help of 5-fold cross-validation we selected the best two prompt formats:

- <P*100>{claim}<P*20>{aspect}<P*100>,
- <P*100>{claim}<P*4>{aspect}<P*20>.

[5] https://github.com/ai-forever/ru-prompts.

For experiments, the NVIDIA RTX A6000 video card and transformers library[6] were used. For each method, we selected a number of training epochs on a 5-fold cross-validation from the following ranges:

- fine-tuning the whole model = [1…5],
- fine-tuning the last layer = [1…20],
- prompt-tuning = [1…300].

The best were 2 epochs for training the whole model, 20 epochs for training the last layer, and 300 epochs for prompt-tuning. The learning rate $5 \cdot 10^{-5}$ and batch size 4 were the same for all the models. We used the following parameters to generate[7]: top_p=0.95, top_k=50, do_sample=True, max_new_tokens=150, no_repeat_ngram_size=3. The generated sequence was segmented into sentences using the natasha library[8]. The first sentence was used for annotation.

Thus, we test six models:

- a fine-tuned whole model,
- a model with fine-tuned last layer,
- an original model with 220 special tokens in prompt,
- an original model with 124 special tokens in prompt,
- an original model with a prompt containing various aspects,
- an original model with a prompt containing one aspect of interest.

4.2 Results and Discussion

Using each of the six models, 254 sentences were generated for 5 aspects of the corpus, that is, 1,524 sentences were obtained for annotation. We determined the number of sentences generated in accordance with the distribution of aspects by claims in the original corpus (Appendix B), excluding rare <claim-aspect> pairs, which accounted for only 1 or 2 premises.

The annotation was done by three annotators, for which the method of generating sentences was hidden. The annotator evaluated the generated sentence according to several parameters:

- whether the sentence is a premise for the specified claim (0/1);
- if the sentence is a premise, then what aspect is expressed in it;
- whether the sentence contains grammatical errors (0/1);
- whether the sentence contains a violation of logic (0/1).

For each generated premise, the annotator could indicate 3 aspects from the full list of aspects (Appendix A).

We considered as premises those sentences that were annotated as premises by at least two annotators, the same rule was applied to identify sentences with grammatical errors and violations of logic. Examples of premises with grammatical and logical errors (for the claim: "Need to save for retirement"):

- *быть осторожным. это означает: иметь накопления и ни в коем случае не делать пенсионные сбережения втом банке, в котором они, по крайней мере в ближайшие пять лет являются гарантированным вкладом и,если такая ситуация возникнет, по истечении такого же времени с целью последующего ихсчета в банке, где они будут гарантированы на все случаи жизни, но ни в каких других.*
  [*to be careful. this means: to have savings and in no case make pension savings in a bank in which they are, at least for the next five years, a guaranteed deposit and,if such a situation arises, after the same time with a view to their subsequent account in a bank where they will be guaranteed for all occasions, but in no other.*];
- *Следует делать пенсионную сбережения, которые вернут в виде процентов, полученных за активы, в том числе активы, которые были начислены в результате обесценения;*
  [*One should make retirement savings, which will return in the form of interest received on assets, including assets that have been accrued as a result of impairment;*].

[6] https://huggingface.co/docs/transformers.
[7] https://huggingface.co/docs/transformers/main_classes/text_generation.
[8] https://github.com/natasha/razdel.

If at least one annotator indicated among the three aspects the one on which the sentence was generated, then we considered this sentence as a premise generated by the required aspect.

The inter-annotator agreement, calculated on the basis of Krippendorff's alpha, was 0.5781 for premise annotation, and 0.4911 for aspect annotation (whether the required aspect is present in the premise). The most difficult topics for premise annotation were "Retirement savings" (0.3601) and "Online shopping" (0.4578); the easiest topics were "Distant work" (0.8059) and "Shooters" (0.7648). The most difficult topics for aspect annotation were also "Retirement savings" (0.2356) and "Esports" (0.2639); the easiest topics coincided with premises: "Distant work" (0.6290) and "Shooters" (0.7185).

Table 1 shows the number of trainable parameters for each model and the statistics of the generated premises: the number of premises generated by the model; the number of premises generated for the required aspect and the number of premises generated for the required aspect that do not contain defects (grammatical errors or violation of logic). The table also shows the proportion (in percent) of such premises among all sentences generated by the model (indicated in brackets).

Model	# trainable parameters	# generated sentences (%)	Premises		Premises on aspect		Premises on aspects without defects	
			#	%	#	%	#	%
Fine-tuned whole model	760,300,032	254 (100%)	**158**	**62.2**	**131**	**51.6**	33	13.0
Fine-tuned last layer	77,194,752		36	14.2	26	10.2	9	3.5
Prompt-tuned-220	337,920		125	49.2	86	33.9	**75**	**29.5**
Prompt-tuned-124	190,464		101	39.8	72	28.4	63	24.8
Few-shot learning all aspect	0		49	19.3	27	10.6	21	8.3
Few-shot learning selected aspect	0		41	16.1	26	10.2	21	8.3

Table 1: Number of trainable parameters and statistics of generated premises for each model

The largest number of premises (column 3) and premises on the required aspects (column 4) was generated by fine-tuned whole model. The quality of the generated premises (column 5) is higher for prompt-tuned and few-shot learning models than for fine-tuned models. Perhaps this is due to the modification of the model weights during fine-tuning, which to some extent violates the grammatical correctness of the generated texts.

Table 2 shows examples of generated premises. Each example was labeled by three annotators as a premise on the required aspect.

Claim: Video games have a positive effect on people	
Aspect: Impact on health	
Model	Generated premise
Fine-tuned whole model	*Однако исследования, в которых участвовали 15-летние школьники, показали, что многие из них играли в видеоигры в отсутствие реальной угрозы жизни или здоровью и таким образом получали возможность выплеснуть свою агрессию.* [*However, studies involving 15-year-old high school students showed that many of them played video games in the absence of a real threat to life or health and thus got an opportunity to express their aggression.*]
Fine-tuned last layer	*это связано с тем, что многие подростки проводят время за видеоиграми, в которых им дают шанс использовать способности мозга и сознания в определенных областях своего восприятия реальности, повышая степень своего развития и, соответственно, шанс успешного получения результата в том или ином из многих случаев, о котором в них говорится.* [*this is due to the fact that many teenagers spend time playing video games in which they are given a chance to use their brain and consciousness abilities in certain areas of their perception of reality, increasing their degree of development and thus their chance of success in one or another of the many cases they are talking about.*]
Prompt-tuned-220	*В исследовании было показано, что те, кто играет в видеоигры, в 2,5 раза чаще страдают заболеванием сосудов головного мозга по сравнению с теми, кто не играет в игры.* [*The study showed that those who play video games are 2.5 times more likely to suffer from cerebral vascular disease compared to those who do not play games.*]
Prompt-tuned-124	*В ходе исследования медики изучили влияние игр на людей и пришли к выводу, что люди, игравшие в видеоигры, страдают от различных болезней чаще.* [*During the study, doctors have studied the effect of games on people and concluded that people who played video games, suffer from a variety of diseases more often.*]
Few-shot learning all aspect	*ьы часто используют видеоигры для развития моторики и координации движений;* [*often use video games to develop motor skills and motor coordination;*]
Few-shot learning selected aspect	*ьысбые 3D-экранные видеоигры могут привести к психическим заболеваниям. ...* [*3D screen video games can lead to mental illness. ...*]

Table 2: Examples of generated premises. Defects are highlighted in red, the symbol "..." means that the continuation of the sequence does not refer to the premise and contains errors

When evaluating generation methods, it is important to understand whether the model replicates the training data. To determine the similarity between the generated premises and the training data, we calculated the average and maximum ROUGE-L and cosine similarity between the generated premises of each model and the premises of the training corpus (Table 3). Cosine similarity was calculated using the RuBERT model [3] and the Sentence Transformers library[9].

[9] https://www.sbert.net/

Model	ROUGE-L		Cosine similarity	
	mean	max	mean	max
Fine-tuned whole model	0.0468	**0.4800**	0.6255	0.9400
Fine-tuned last layer	0.0445	0.3590	0.5859	0.9288
Prompt-tuned-220	0.0487	0.4000	0.6107	**0.9501**
Prompt-tuned-124	0.0448	0.3478	0.5913	0.9332
Few-shot learning all aspect	0.0459	0.2632	0.5998	0.9245
Few-shot learning selected aspect	0.0487	0.2500	0.6018	0.9279

Table 3: ROUGE-L and cosine similarity between the generated premises and the premises of the training corpus

With the maximum mean cosine similarity of 0.6255 across the models, the mean ROUGE-L of the fine-tuned whole model does not exceed 0.05. The maximum value of ROUGE-L among all models was shown by the fine-tuned whole model, while the maximum cosine similarity value was shown by the prompt-tuned model. However, even similar premises have clear differences. We give examples of premises that obtained maximum ROUGE-L and cosine similarity scores:

- **Corpus sentence**: *В 2014 году глава Банка Эстонии осторожно отмечал отсутствие доказательства того, что Биткойн не является финансовой пирамидой* [*In 2014, the Governor of the Bank of Estonia was careful to point out the lack of proof that Bitcoin is not a pyramid scheme*].
  Generated sentence: *заявил, что биткойн не является самостоятельной финансовой пирамидой* [*stated that Bitcoin is not a financial pyramid in its own right*].
- **Corpus sentence**: *Если ты будешь платить ренту в биткойнах, ты можешь обанкротиться в случае если он сильно пойдет вверх, а твой доход привязан к фиатным деньгам* [*If you pay your rent in bitcoins, you can go bankrupt if it goes up strongly and your income is tied to fiat money*].
  Generated sentence: *Это приводит к риску вывода активов и денежных сумм с кратной целью, т.е. при первоначальной сделке с биткойнами инвестор рискует получить убыток и подвергнуть риску свои накопления* [*This leads to the risk of withdrawing assets and sums of money in multiples, i.e. in an initial bitcoin transaction an investor risks making a loss and putting his savings at risk*].

Thus, we can conclude that the models generate different premises from the training examples.

We also tested the ability of the resulting models to generate premises for unfamiliar claims and aspects. This means that the training data for the fine-tuned and prompt-tuned models did not contain these claims and aspects, and the prompt of the few-shot model did not contain examples for these claims and aspects. To test this, we generated three five-aspect sentences for two claims using each model:

- claim: "Children's camp affects the child positively",
  aspects: "Intellectual development", "Moral aspect", "Communication with people", "Popularity", "Entertainment";
- claim: "Free money is better spent than saved",
  aspects: "Standard of living", "Legal aspect", "Efficiency", "Attitudes of the authorities", "Prospects".

Thus, we obtained 150 sentences for annotation. This annotation was conducted similarly to the previous one. The results of the analysis are presented in Table 4.

Inter-annotator agreement, calculated with the help of Krippendorff's alpha, is 0.4538 for the annotation of premises, and 0.3006 for the annotation of aspects (whether there is a required aspect in the premise).

Model	# generated sentences (%)	Premises		Premises on aspect		Premises on aspects without defects	
		#	%	#	%	#	%
Fine-tuned whole model	30 (100%)	**18**	**60.0**	**11**	**36.7**	1	3.3
Fine-tuned last layer		3	10.0	0	0.0	0	0.0
Prompt-tuning_220		7	23.3	6	20.0	**6**	**20.0**
Prompt-tuning_124		6	20.0	4	13.3	3	10.0
Few-shot learning all aspect		8	26.7	2	10.0	1	3.3

Table 4: Statistics of generated premises on unfamiliar aspects for each model

In this experiment, the fine-tuned whole model also showed the best result.

Table 5 shows examples of generated premises for unfamiliar aspects. For each example, the number of annotators who considered the sentence to be a premise and whether the premise expresses the required aspect are given.

Claim: Children's camp has a positive effect on the child			
Aspect: Communication with people			
Model	Generated premise	Voting	Aspect
Fine-tuned whole model	*Но в реальности дети зачастую лишены такой возможности, потому что находятся в стесненных условиях, их постоянно окружают чужие люди, от которых они не могут укрыться.вольте думать позитивно!...* [*But in reality, children are often deprived of this opportunity, because they are in cramped conditions, they are constantly surrounded by strangers, from whom they can not hide.*]	3 from 3	+
Fine-tuned last layer	*Утвержение: В лагере есть работа и льготы для детей;* [*The camp has jobs and benefits for children;*]	2 from 3	–
Prompt-tuned-220	*Коротко, в отличие от традиционных лагерей, для детей проводят так называемое коммуникативно ориентированное пространство или пространство деятельности.* [*Briefly, unlike traditional camps, a so-called communication-oriented space or activity space is held for children.*]	2 from 3	+
Prompt-tuned-124	*Влияние на детей в детском возрасте может быть позитивным, когда люди, с которыми он сталкивается, являются нормальными, но может быть и негативным, когда он сталкивается с ненормальными.* [*The impact on children at an early age can be positive when the people he encounters are normal, but it can also be negative when he encounters the abnormal.*]	3 from 3	+
Few-shot learning	*Детский лагерь является лучшим местом для раскрытия творческих способностей у детей, что позволяет им чувствовать себя более комфортно;* [*Children's camp is the best place to unleash creativity in children, making them feel more comfortable;*]	3 from 3	–

Table 5: Examples of generated premises on unfamiliar aspects. Defects are highlighted in red, the symbol "..." means that the continuation of the sequence does not refer to the premise and contains errors.

5 Conclusion

The paper explored aspect-based argument generation in Russian. Five aspects were considered: "Safety", "Impact on health", "Reliability", "Money", "Convenience and comfort". The models were trained using different approaches (fine-tuning, prompt-tuning and few-shot learning). The best model was wholly fine-tuned on the aspect-based corpus of premises. This model generated 51.6% of the premises on the given aspects, the model obtained using the prompt-tuning approach gives 33.9% of the premises on the given aspects, and with the few-shot learning approach – 10.6%.

The problem of fine-tuned models is the low level of grammatical correctness of the generated premises compared to the prompt-tuning and few-shot learning models. For example, for the best fine-tuned model out of 131 generated premises without grammatical errors and logic violations, there were 33 premises (25.2%), and for the prompt-tuned-220 model, 75 out of 86 premises (87.2%) were correct. When the post-processing of the generated sentences is complicated or impossible for some reason, prompt-tuned models become preferable.

It is important to note that fine-tuned models are able to generate premises on new, unfamiliar aspects. For example, the fine-tuned whole model was able to generate 36.7% of premises (11 premises for the required aspect out of 30 generated sentences). This allows us to hope for the potential application of such models for a wide range of topics.

In the future, we plan to expand the annotated corpus of premises and aspects and use reinforcement learning [7].

We made the best fine-tuned model for generating premises on the given aspects available to the public.

Acknowledgements

This work was supported by Russian Science Foundation, project № 22-21-00885, https://rscf.ru/en/project/22-21-00885/.

References

[1] Brown T., Mann B., Ryder N., Subbiah M., Kaplan J.D. et al. (2020), Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems, Vol. 33, pp. 1877–1901.

[2] Fishcheva I., Osadchiy D., Bochenina K., Kotelnikov E. (2022), Argumentative Text Generation in Economic Domain // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue-2022", Issue 21, pp. 211–222.

[3] Kuratov Y., Arkhipov M. (2019), Adaptation of deep bidirectional multilingual transformers for Russian language, Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue -2019", Issue 18, pp. 333–340.

[4] Lester B., Al-Rfou R., Constant N. (2021), The Power of Scale for Parameter-Efficient Prompt Tuning // Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 3045–3059.

[5] Li X. L., Liang P. (2021), Prefix-Tuning: Optimizing Continuous Prompts for Generation // Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, pp. 4582–4597.

[6] Liu X., Zheng Y., Du Z., Ding M., Qian Y., Yang Z., Tang J. (2021), GPT Understands, Too. arXiv:2103.10385.

[7] Ouyang L., Wu J., Jiang X., Almeida D., Wainwright C. et al. (2022), Training language models to follow instructions with human feedback // NeurIPS 2022.

[8] Radford, A., Narasimhan, K., Saliman, T., Sutskever I. (2018), Improving Language Understanding by Generative Pre-Training. OpenAI Technical report.

[9] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. (2019), Language Models are Unsupervised Multitask Learners, OpenAI Technical report.

[10] Ruckdeschel M., Wiedemann G. (2022), Boundary Detection and Categorization of Argument Aspects via Supervised Learning // International Conference on Computational Linguistics. Proceedings of the 9th Workshop on Argument Mining, pp. 126–136.

[11] Sber: Russian GPT3 models. Retrieved March 15, 2023, from https://github.com/sberbank-ai/ru-gpts.

[12] Schiller B., Daxenberger J., Gurevych I. (2021), Aspect-Controlled Neural Argument Generation // Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 380–396.

[13] Stede M., Schneider J. (2018), Argumentation Mining, Synthesis Lectures on Human Language Technologies, San Rafael: Morgan and Claypool Publishers.
[14] Zhang H., Song H., Li S., Zhou M., Song D. (2022), A Survey of Controllable Text Generation using Transformer-based Pre-trained Language Models // arXiv:2201.05337.

Appendix A

Aspect	Frequency in the corpus, sentences
Safety	**133**
Impact on health (Impact on health + Influence on the psyche)	**107 (56 + 51)**
Reliability	**90**
Money (Price + Profitability)	**89 (55 + 34)**
Convenience and comfort	**88**
Attitude of the authorities	78
Prospects	72
Efficiency	39
Standard of living	26
Legal aspect	26
Environmental friendliness	23
Communication with people	22
Popularity	15
Quality	12
Career	9
Intellectual development	7
Entertainment	7
Moral aspect	4

Table 6: Frequency of aspects in the corpus[10]

[10] The arguments in the table are not unique, since one sentence can have several aspects.

Appendix B

Topic	Claim	Aspect					Total
		Safety	Impact on health	Reliability	Money	Convenience and comfort	
Paper and e-books	Paper books are better than e-books	1	0	0	0	7	8
Children's video blogs	Children should be encouraged to create vlogs	1	11	0	2	0	14
Children's gadgets	Gadgets have a positive effect on children	12	59	0	0	0	71
Blood donation	Donation is necessary for society and safe	6	14	0	0	0	20
Esports	Esports should be made an Olympic sport	0	1	1	4	0	6
Cryptocurrency	You need to use cryptocurrency and invest in it	87	1	59	30	24	201
Online education	Online education can compete with traditional education	0	0	0	5	2	7
Retirement savings	Need to save for retirement	0	0	15	4	1	20
Online shopping	Should shop online	11	0	1	0	10	22
Supermarkets and food markets	It is better to buy products in the supermarket, not in the market	6	0	2	7	11	26
Distant work	Remote work is preferable to office work	3	1	0	6	11	21
Freelance	Freelancing is better than being hired	0	1	1	4	2	8
Shooters	Video games have a positive effect on a person	0	19	0	0	0	19
Electric cars	Electric cars are better than regular cars	6	0	11	27	20	64
Total		133	107	90	89	88	507

Table 7: Distribution of corpus topics by aspects[11]

[11] The arguments in the table are not unique, since one sentence can have several aspects.