

ADWISER

Automated Detection of Writing Inaccuracies for
Students of English in Russia

Developed by:

**Olga Vinogradova
Anna Viklova
Ksenia Pospelova
Eduard Klyshinsky**

**Elizaveta Ershova
Anton Buzanov
Sofia Generalova
Darya Matyash**

**Darya Overnikova
Nika Smilga
Ivan Torubarov
Nikita Login
Irina Panteleeva**

In cooperation with REALEC

- Models are roughly based on the most common types of mistakes in REALEC
- Models are developed and adjusted using real examination essays



**Total 14500
texts/3620994
words**

**Exam 5965
texts/1412169
words**

**Total 13592
annotated
texts**

Экосистема ADWISER



- Экспертная разметка
- Наблюдения
- Выявление паттернов ошибок
- Сбор данных для тестирования моделей



- преподаватели
- студенты
- аннотаторы
- администраторы корпуса
- исследователи лингвисты

- тестирование гипотез
- автоматическая детекция ошибок
- сбор данных



- программисты
- преподаватели
- студенты
- аннотаторы

ADWISER Создание моделей

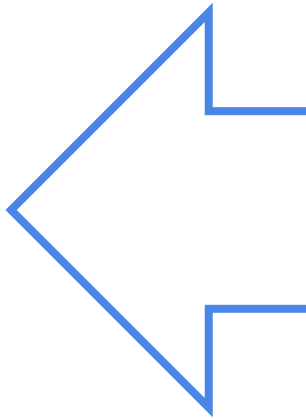
- 1) Выявление паттернов ошибок, составление списка подозрительных контекстов
- 2) Частеречная разметка текстов с помощью TreeTagger
- 3) Составление регулярных выражений для Python3
- 4) Экспертная разметка выявленных случаев и улучшение моделей, уточнение списка подозрительных контекстов

```
def patt():  
    #context Verb/aux + ,(?) + wh-word(+ whether + if) + aux + NP + VP  
    start = '(?:<[^>]+?\sV..>)(?:,\sPUN)?(?:<[^>]+?\s(?:AVQ|PNQ|DTQ)>|<[whether|if]\s.+?>)' + \  
            '(?:<[^>]+?\sV[B|D|H|M]..>'  
    #1: all verbs, 2: comma, 3: wh-words + whether&if, 4: auxiliaries (be, do, have, modal)  
    noun_phrase = nounp()  
    verb = '(?:<[^>]+?\sV..>'  
    pattern = start + noun_phrase + verb  
    return pattern
```

sentence	words between IN and NUM	
From 5% in 1940 the percentage of serious began to fall and only in the period between 2000 and 2020 it has returned to this previous point	0	1
For the next ten years in Sweden the quantity of the old people have been falling and then, in 1990, rised to 20% in 2010	0	1
While in 50-s there were a few of obese - only 4%, in 2010 the percentage of thick old people has reached an awful number of 60%	0	1
Overall, total amount of old people significantly have increased in times from 1940 to 2040 in three countries	0	1
As we see it on one of the charts in 1999 France have visited about 11 millions of UK visitors	0	1
To summarize, in the last 60 years the proportion of obese people have started to grow, while ideal or underweight category have become remarkably lower	1	0
Overall, within the information provided it can be said that in the last hundred years the proportion of population 65-year-olds and older in Japan, Sweden and in the USA has increased	1	0
This graphs might be construed to define, how the way of life have changed in the last 60 years	2	0
142 Japan and Sweden part of elders has grown up in period between 1980 and nowadays, nevertheless USA keeps the same level	2	0

ADWISER Mistakes in tense forms

- From 5% in 1940 the percentage of serious began to fall and only in the period between 2000 and 2020 it has returned to this previous point
- Sweden and USA had a similar figures, but the gap between them has become larger during the period between 2000 and 2010.
- While in 50-s there were a few of obese - only 4%, in 2010 the percentage of thick old people has reached an awful number of 60%



It looks as though you have used the wrong form of the verb together with time indication

Model description:

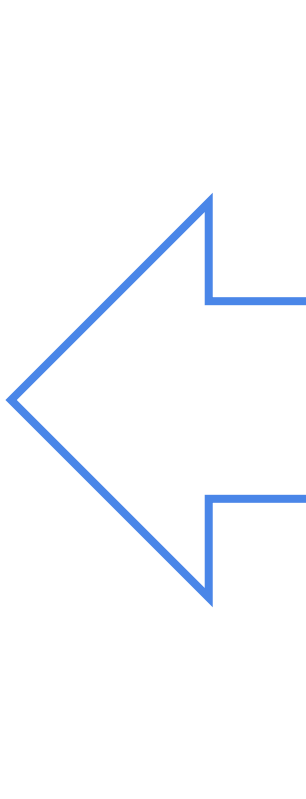
Present Perfect contexts followed by an indication of a precise moment in the past

Present Perfect clause + <0-4 words> + in +

((last|few|recent)* + NUM + (years|days)) etc

ADWISER Mistakes in tense forms

- To sum up, the world level of unemployment **was fluctuating** between 2014 and 2015, but in areas of developing countries the rate **was decreasing** during the period.
- To conclude, we may see that in 1950 inhabitants of Newtown **were tending** to have an ideal weight over their whole life in contrast to people from 2010, which had tendency to have an obese weight.
- In 2012 about 150 million users **were browsing** Facebook on their desktop computers.



The usage of Past Continuous might be erroneous

Model description:

Past Continuous contexts followed by an indication of a precise moment in the past

Past Continuous clause + .. + NUM + .../during the period/this year etc

ADWISER Mistakes in Past Continuous

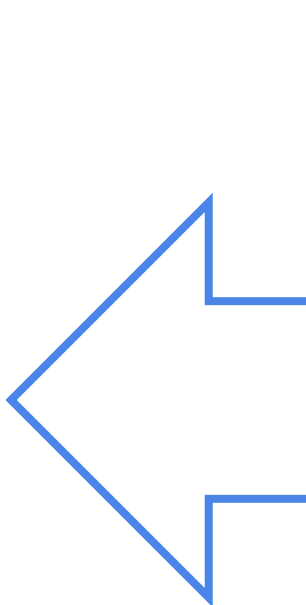
1st results: 42 contexts, 70% accuracy

- expanding the list of possible time indications
- expansion of the model scope

2nd results: 48 contexts, 80% accuracy

ADWISER Mistakes in punctuation

- It should be mentioned, that the increasing of the life duration can lead to huge ecological problem.
- Speaking about our experience and its role in our life, I can say, that sometimes we are able to avoid our genes and to be as well as we want to be.
- And most of us do not understand, why it is so expensive.



You may have used a redundant comma in this sentence

Model description:

Extra comma before a subordinate clause

Noun phrase + .. + verb + , + conj

ADWISER Mistakes in punctuation

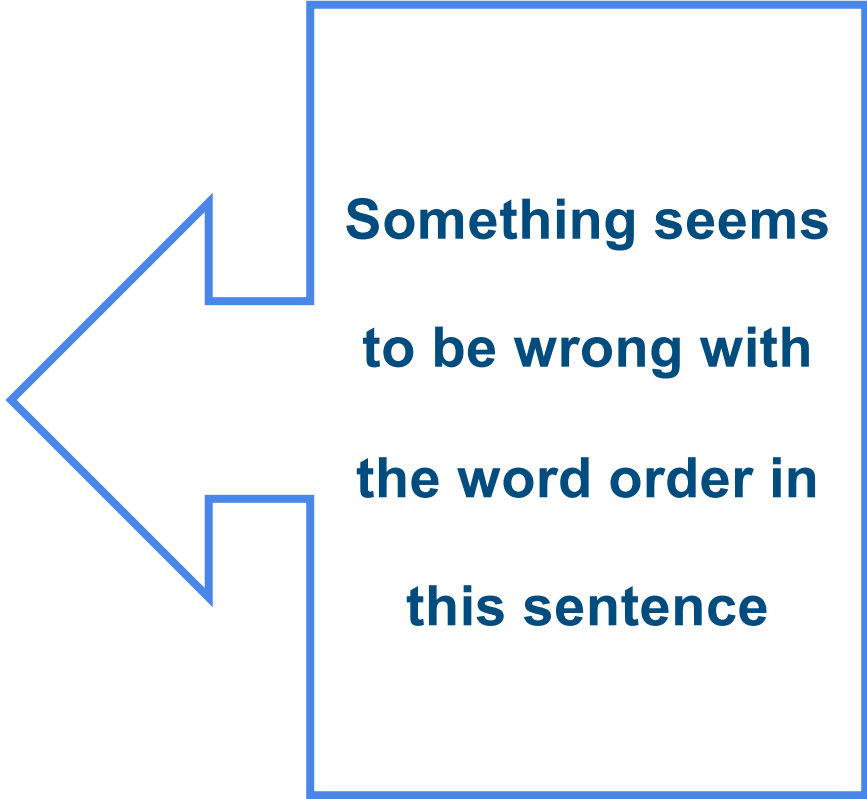
1st results: 659 contexts, 94% accuracy

- expanding the list of triggers
- several subordinate clauses

2nd results: 648 contexts, 96% accuracy

ADWISER Mistakes in Word Order

- For several reasons which I will mention below I explain **why do I think so**.
- Additionally, no one can decide **what should people do or not do**.
- I don't know **what can I do now**.



**Something seems
to be wrong with
the word order in
this sentence**

Model description:

Inversion in subordinate clause

Wh-word + V + NP + V_aux - *We need to know why did the level of crime boost up*

ADWISER Mistakes in Word Order

- People who obtain important posts just have no alternative but jet or plane if they have to get **to**, **for example**, the USA from Russia.
- The second argument is that oftenly one can find in the internet things, which one can't buy, **for example**, old publications or old films and music, etc.
- Also it is possible to reach their, **for example**, workplaces not using private cars but going there by feet.

, for example,

**It looks as though
occupies the
wrong position in
your phrase**

Model description:

Misplaced logical connectors

NP + PREP + logical connector + Noun - *I went from, for example, St. Petersburg to Moscow*

ADWISER Mistakes in Word Order

1st results: 21 sentences, 67% accuracy

- semicolons

2nd results: 24 sentences, 75% accuracy

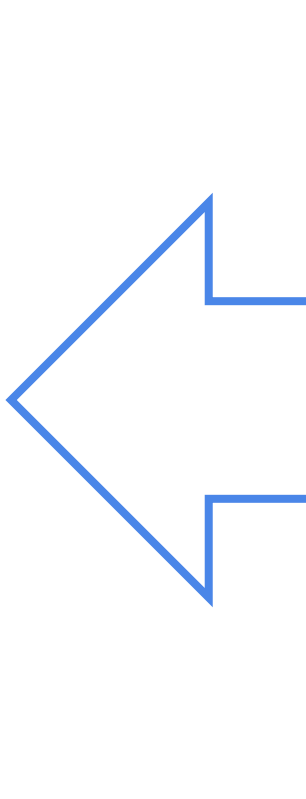
- *what* in NP

- expansion of the model scope

3rd results: 27 sentences, ~100% accuracy

ADWISER Mistakes in Tense Forms

- In conclusion I want to say that moving business and employing workers in developing countries have a big number of disadvantages and **if we will** think not only about nearest future we will understand that it is bad for people who live in both types of countries when companies make unequally between developing and developed countries at the situation like this.
- For example, the polar bears would have no place to hunt, **if icebergs slide**.
- There is no need to leave criminals in prisons forever, but it would be better **if we care** of our safety more.



You may have used the wrong form of the verb in the condition.

Model description:

Conditionals

if + NP + will/would -It would be better if the herbalist makes a remedy specially for you.

if + NP + VP + NP + would/V1 -If we achieve it, this problem would be solved.

ADWISER Mistakes in Word Order

1st results: 27 sentences, 85% accuracy

- *what if, because if*
- *if* in reported questions
- *would like*

2nd results: 29 sentences, 69% accuracy

- more accurate regular expression

3rd results: 211 sentences, 96% accuracy

ADWISER Mistakes in Word Order

Barely you can, but the main character of the book I read, Sherlock Holmes, does it easily and the method he used to detect such thing was later called Holmesian deduction.

You may need
inverted word
order in this
sentence.

Model description:

Inversion in constructions

Never/nowhere/barely/scarcely/few + NP + VP

Had + NP + V3 + wh-/that/if? + would + NP + V1/have/V3

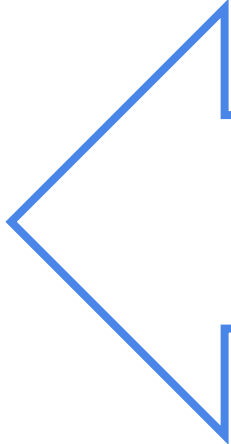
ADWISER Mistakes in Punctuation

To sum up, it can be concluded , that during the 1940 and 2040 the percentage of population in three countries rise differently, espassially in Japan, the nation become **more aged, than** in the USA and Sweden.

It is obvious, that first subway was opened in capital of Great Britain, than it take **more, than** one hundred and thirty years to open it in Los Angeles.

But in fact, we face architecture much **more often, than** other kinds of art.

The US have the biggest average health spending per person, it equals 6 719 dollars, and the life expectancy is only two years **less, than** in Netherlands.



**It looks as though
you have wrongly
used the
highlighted
comma in
comparative
constructions**

Model description:

(more/less {+ ADJ/ADV}) / ADJ-er / ADV-er + {[1,7] words + } comma + than

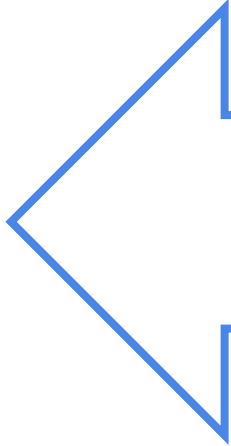
ADWISER Mistakes in Punctuation

Than .. than* - incorrect structure with *than

So, it can be seen that **than** further go technologies, **than** more people will use them.

Using *than* instead of *then*:

The curves, belonged to the USA and Sweden went similarly until 1990, **than** there was a significant growth of elderly people rate in Sweden, which fluctuated from 2010 to 2020 and then continued to rise.



It looks as though
you have wrongly
used the
highlighted
comma in
comparative
constructions

Model's accuracy:

~94% of the output was correct, other "wrong" answers are not wrong from the English grammar's point of view, but there is some work to make the model more accurate, still

ADWISER Mistakes in Punctuation

For instance, 56 and over aged people seemed to work more than in 1998 in all spheres, except **such** hard work, **as** building.

At this point, your view on life is challenged, and **such** stress, **as** bullying, may lead to frustration and depression in this period of life, which is not what you want from a school.

For example, **such** activities, **as** running and doing morning exercises do not involve some extra facilities.

Those people, who consider that we can improve our health to increase the number of **such** facilities, **as** gyms, stadiums, swimming pools and so on, say that if there were right amount of such facilities, people would go to them more frequent than now.

It looks as though
you have wrongly
used the
highlighted
comma in
constructions
such/as/like ... as

Model description:

***such/as* + [[0,20] words +] comma + *as* + [0,20] words {+ comma}**

OR

comma + *like* + [1,10] words {+ comma}

ADWISER Mistakes in Punctuation

Examples of the sentences that helped us make the decision of omitting comma + *like*+ [1,10] words {+ comma} model

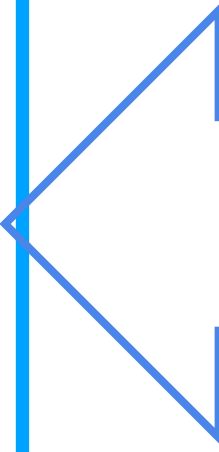
Sentences that are correct:

There are reasons for throwing the idea away, like social environment.

Sometimes child can grab or taste something, while parents do not see, and it may lead to a bad consequences, like stomachache, hand cutting or he can hurt his leg.

Usage of *like* instead of *namely*:

The line graph illustrate the number of people aged 65 and over in some countries, like Japan, Sweden and USA in the percontage from 1940 to 2040.



It looks as though
you have wrongly
used the
highlighted
comma in
constructions
such/as/like ... as

Model's accuracy

after the first check of the programme's output we decided to delete the model comma + *like*+ [1,10] words {+ comma} because of the uncertainty of punctuation rules in English

the output reduced by 9 times, but became ~98% accurate

ADWISER Mistakes in Punctuation

Moreover the part of young people will be reduce in both countries.

However is it right?

On the other hand there are ather group of people who claim that ut will have a little effect on public healt.

It looks as though
you have wrongly
used the
highlighted
comma with
introductory
phrases

Model description at the beginning of the sentence:

(From+ [1,3/5] words + point of view/viewpoint/perspective) / (In + + [1,3] words + opinion) / (To + my/your/his/her/our/their + mind) / For example/For instance/However/Nevertheless/Consequently/Moreover/On the other hand/In other words/Surprisingly/Unsurprisingly/Hopefully/Interestingly/Obviously/To sum up/Thus/Of course

ADWISER Mistakes in Punctuation

Stealing about rural areas, we can see, that **however** number of people, who use Internet access in their places, less than in urban areas, it still increases.

Most of them interact and coexist, **for instance** in Moscow, we can see old churches preserved till now near modern skyscrapers.

I'm convinced that transport in future will be able to transport people as fast as information, **for example** e-mails of messages in Skype, now.

It looks as though
you have wrongly
used the
highlighted
comma with
introductory
phrases

**Model description in the middle or in the end of the sentence:
(from+ [1,3/5] words + point of view/viewpoint/perspective) / (in + [1,3] words +
opinion) / (to + my/your/his/her/our/their + mind) / for example/for
instance/however/nevertheless/consequently/to start with/moreover/on the other
hand/in other words/hopefully/in conclusion/to sum up/thus/of course**

ADWISER - Automated Detection of Writing Inaccuracies for Students of English in Russia

sentence	words between IN and NUM	
From 5% in 1940 the percentage of serious began to fall and only in the period between 2000 and 2020 it has returned to this previous point	0	1
For the next ten years in Sweden the quantity of the old people have been falling and then, in 1990, rised to 20% in 2010	0	1
While in 50-s there were a few of obese - only 4%, in 2010 the percentage of thick old people has reached an awful number of 60%	0	1
Overall, total amount of old people significantly have increased in times from 1940 to 2040 in three countries	0	1
As we see it on one of the charts in 1999 France have visited about 11 millions of UK visitors	0	1
To summarize, in the last 60 years the proportion of obese people have started to grow, while ideal or underweight category have become remarkably lower	1	0
Overall, within the information provided it can be said that in the last hundred years the proportion of population 65-year-olds and older in Japan, Sweden and in the USA has increased	1	0
This graphs might be construed to define, how the way of life have changed in the last 60 years	2	0
142 Japan and Sweden part of elders has grown up in period between 1980 and nowadays, nevertheless USA keeps the same level	2	0

Model description:

Present Perfect contexts followed by an indication of a precise moment in the past

Present Perfect clause + <0-4 words> + in +

((last|few|recent)* + NUM + (years|days)) etc

ADWISER WEBSITE

<https://linghub.ru/adwiser/>

ADWISER

Check your text for errors!

Annotation result:

Kids can not understand, why their parents not always be with them.

So, we can't say, that these types of transport hurms our environment less, than planes.

If artist will continue use their enormous freedom, whole world will go to chaos and darkness.

If we will inclined to believe this forecast, the electronic books will never replace the printed one on the market.

Never in my life I've been happier with technical progress than nowadays, when I am far away from home and have only one option to see my family and friends - by Skype.

Barely you can, but the main character of the book I read, Sherlock Holmes, does it easily and the method he used to detect such thing was later called Holmesian deduction.

For the next ten years in Sweden the quantity of the old people have been falling and then, in 1990, rised to 20% in 2010.

The tendency of investing money by developed countries is different : from 2006 to 2008 the sum The tendency of investing money by developed countries is different: from 2006 to 2008 the sum was increasing, in 2009 the number dropped, from 2010 to 2011 the sum was growing again and from 2011 to 2013 the sum was decreaing. When the data of the USA and Sweden

was growing up to 15%, the 65 years old population in Japan was levelling

out to 3% and stayed approximately the same from 1960 to 1990. Another interesting fact is that at the begining the population aged 65 was less in

China and the most in USA, but in 2040 the situation changed completely and the population aged 65 was the biggest in Japa

has always stayed

Back



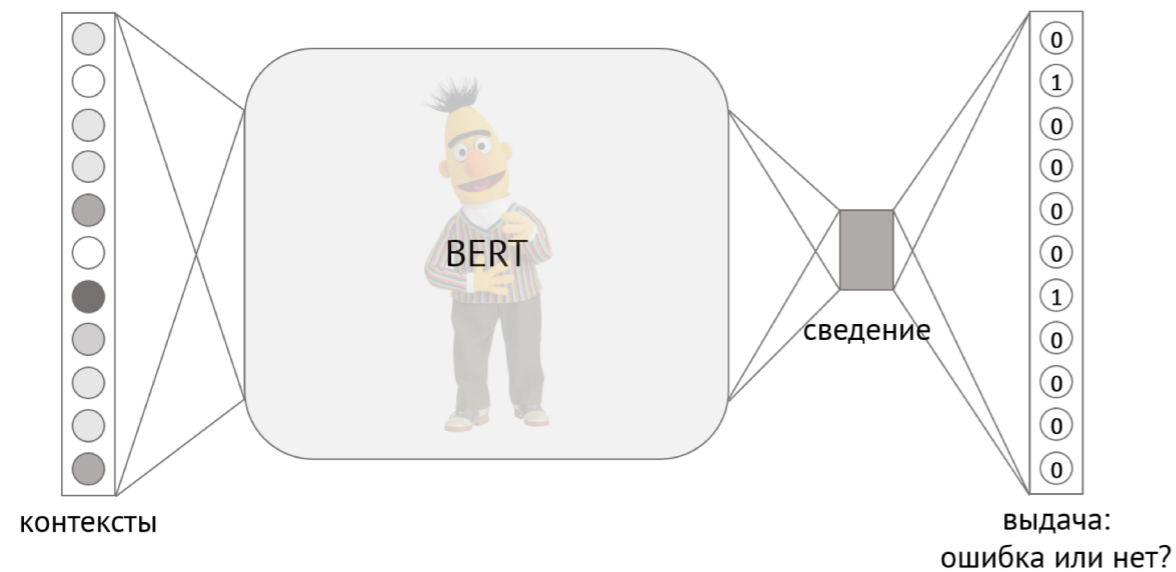
ADWISER - Automated Detection of Writing Inaccuracies for Students of English in Russia

RESEARCH PROSPECTS

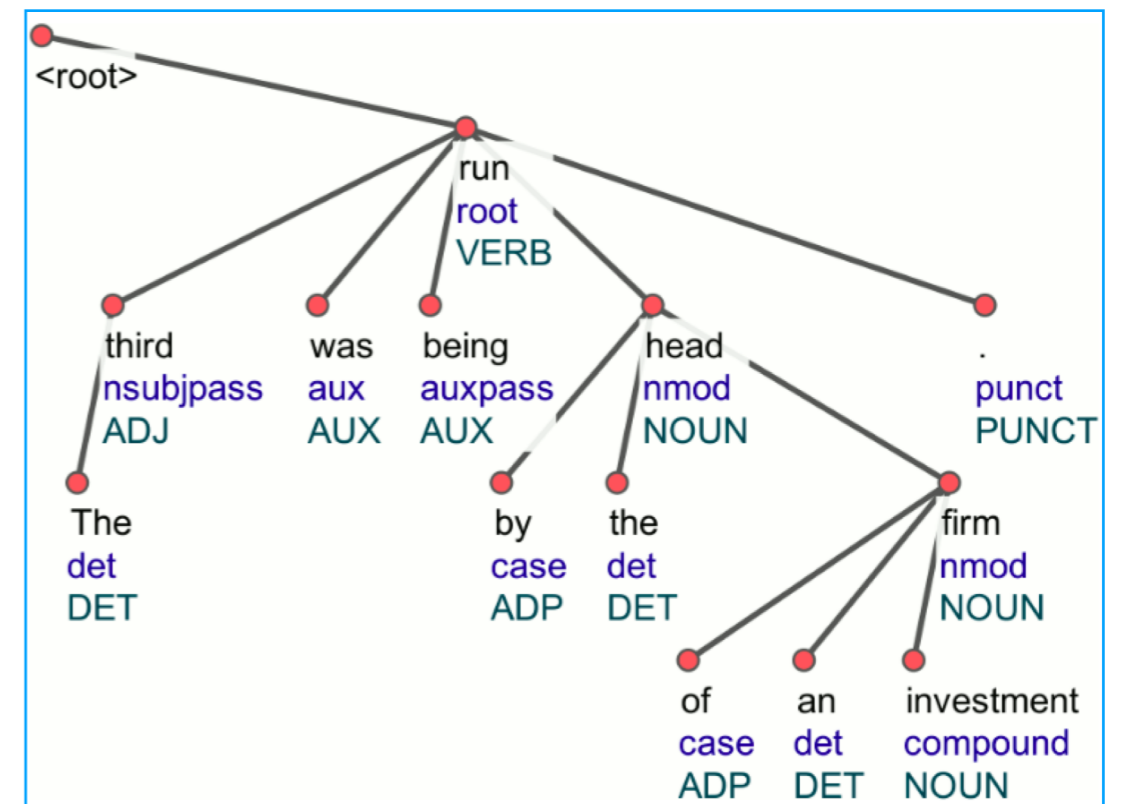
1. Machine Learning Based Models using BERT developed by Ivan Torubarov

...with daughter and son. As for me, my family [MASK] time with me. Also, more important...

↑ previous sentence ↑ token in question ↑ following sentence



2. Syntactic structure annotation tools for greater precision developed by Irina Panteleeva



1. Models developed for detection of Punctuation mistakes:
 - «, that» mistakes
 - «, than» in comparative constructions
2. Models developed for detection of Word Order mistakes:
 - questions in reported speech with inversion
 - no inversion in direct speech questions

REALEC Testmaker & Testing platform

- We have a large collection of English texts
- The texts are error-annotated
- Why not use it as a source of exercise material?

REALEC Testmaker & Testing platform

Step 1. Generate exercises from REALEC texts

REALEC Testing Platform

[Quizzes](#) [Grades](#) [Questions](#) [Users](#) [IELTS](#) [Log Out](#)

Enter path to folder:

Create new question folder with name:

[Generate questions](#)

Select error tag from REALEC ▾

Select all

Unselect all

Punctuation

Spelling

Capitalisation

Grammar

Determiners

Articles

Quantifiers

REALEC Testmaker & Testing platform

Step 2. Create quiz from generated exercises

REALEC Testing Platform

[Quizzes](#) [Grades](#) [Questions](#) [Users](#) [IELTS](#)

[Log Out](#)

Select error tag from REALEC ▾

Select folder ▾

Apply tag and folder filter

Page:

<< To First 1 2 3 >> To Last

- 1. The issue of illegal copying and sharing *music or films* on the internet is the issue that is very popular nowadays. Some people **hold a view** that people doing it should be punished , while others do not support them. From my point of view, the second position is right.
- 2. First of all, some people, for example, children or teenagers, cannot afford to buy music albums or films to themselves. Therefore, they do not have **chance** to listen to or watch it. What is more, musicians and film producers can lose money not because of illegal pirate copies , but because people do not know about them.
- 3. This is actually about musicians and producers who are at the begining of their careers. They need to use all possible channels to become more popular and the lack of access to them will not **let them to do it**. However, there are some problems with illegal copying
- 4. The first one is that popular film makers and musicians do not get all the income they can get. Therefore, the cannot make new films or music because **they are lack of money**. The second one is that sometimes they do not understand how popular they really are , as there is no evidence or statistics about people downloading it.
- 5. These problems can actually be solved. As a solution, producers and musicians can allow to download some of their products for free and **ask money** for some other. What is about the second problem, to my mind, it can be solved after the first one.
- 6. People will listen to music or watch films for free and want to buy more of them. To sum up, illegal copying and sharing music and films is not such a big crime **for which** people should be punished .
- 7. The issue of illegal copying and sharing music or films on the internet is the issue that is very popular nowadays. Some people **hold a view** that people doing it should be punished , while others do not support them. From my point of view, the second position is right.
- 8. The issue of illegal copying and sharing music or films on the internet is the issue that is very popular nowadays. Some people **hold a view** that people doing it should be punished , while others do not support them. From my point of view, the second position is right.
- 9. First of all, some people, for example, children or teenagers, cannot afford to buy music albums or films to themselves. Therefore, they do not have **chance** to listen to or watch it. What is more, musicians and film producers can lose money not because of illegal pirate copies , but because people usually do not know about them.
- 10. This is actually about musicians and producers who are at the begining of their careers. They need to use all possible channels to become more popular and the lack of access to them will not **let them to do it**. However, there are some problems with illegal copying.
- 11. The first one is that popular film makers and musicians do not get all the income they can get. Therefore, the cannot make new films or music because **they are lack of money**. The second one is that sometimes they do not understand how popular they really are as there is no evidence or statistics about people downloading it.
- 12. These problems can actually be solved. As a solution, producers and musicians can allow to download some of their products for free and **ask money** for some other. What is about the second problem, to my mind, it can be solved after the first one.
- 13. People will listen to music or watch films for free and want to buy more of them. To sum up, illegal copying and sharing music and films is not such a big crime **for which** people should be punished .
- 14. Others claim that the Internet is a place where you can get all the media content you want without paying for it because everyone does exactly the same thing. To my mind, it's fair and morally justified to pay for **someone's else work** , no matter what type of work you're dealing with. Firstly, it takes a lot of time for musicians and film producers to create a piece of art.
- 15. They do a lot of work, spend money on the process of making a product and advertisement, try to understand their audience. Secondly, intellectual property is protected by the law and everyone who doesn't follow the **judiciary system's rules** should be punished

Quiz name:

Strip punctuation and parenthesis from submitted answers

[+ Add questions from REALEC](#)

Folder name:

[Create folder](#)

REALEC Testmaker & Testing platform

Step 4. Review test results

Results of user info hidden in Tense choice quiz

Download: [Excel file](#)

Question	Submitted answer	Correct answers	Mark	Fraction of correct answers
Local landline calls reached a peak in 1999, having steadily risen before that year and losing position at almost the same rate afterwards. In general, we can see an increase in the total number of call minutes, and mobile calls have become significantly more popular in 2002 than in 1995, while other types of calls either had a relatively moderate increase, as national/international telephone calls, or maintained almost the same number of minutes in 2002 as in 1995 (local landline calls).	became <input type="checkbox"/> Mark as new correct	became	1.0	0.72727272727273
The lines of population in the USA and Sweden are going to be parallel again from 2027. American population increases from this moment from 16% to 23% in 2040 , and the population of Sweden will grow from 18% to 25% in 2040. All in all, from this graph it can be concluded that the levels of the elderly population in Japan, Sweden and the USA is predicted to grow from 2030 rapidly	will increase <input type="checkbox"/> Mark as new correct	will increase; will be increasing; is expected to increase; will have increased; is predicted to increase	1.0	0.6875
The graph shows the information about proportions of the population aged 65 and over in three countries: Japan, Sweden , and the USA. The graph shows the period from 1940 to 2040. We can see that both lines go up from 7-9% to 15% in 1980. Overall proportion grows up at the end of the period. There is a tendency towards sharp increase in Japan and in the USA.	went <input type="checkbox"/> Mark as new correct	went; had gone	1.0	0.8571428571428571
The graph provides data about the percentage of of people people 65 years old and over. It demonstrates a in this proportion from 1940 to 2040 in Japan, Sweden and the USA. It's clear from the graph that in every country in this comparison the percentage of people 65 years old and over would increase from 5%-10% in 1940 to approximately 25% in 2040. However, the time and pace of growing are different.	will increase <input type="checkbox"/> Mark as new correct	is expected to increase	0.0	0.0
It means that a person of a particular temperament will never be happy in the common meaning of the word. So, happiness is a thing that is normal for one person and that another may not have got To conclude, society needs happy people because they made our existense better in many spheres. Nevertheless, the pursuit of happiness is unique for each person , and not everyone knows which path to follow.	make <input type="checkbox"/> Mark as new correct	make	1.0	1.0
The group of people who are over 60, is more than the children group around 10%. as for the year 2050, the per cent of people who are over 60 rese gradually and will be equal to 42,3%, while the percent of children will be only 11,5. To sum up, the population of Yemen will have more young people and children while old part of population of Italy in period from 2000 to 2050 will totally increase from 24,1% to 42,3%.	will rise <input type="checkbox"/> Mark as new correct	will rise	1.0	0.5909090909090909
Also we are always in a hurry and we want to eat fast. Fastfood become very popular in our life but it is not a healthy food. In my opinion, I think that there are a lot of ways to improve public health.	has become <input type="checkbox"/> Mark as new correct	has become	1.0	0.7142857142857143
There is another situation in Italy. The percentage of people aged 15-59 dominates in 2000 (61,6%) and will dominane constantly (46.2%) in the future.	dominated <input type="checkbox"/> Mark as new correct	dominated	1.0	0.8636363636363636
My cousin suffered from different diseases in his childhood. So that his parents enroll him to the voleyball club. As a result, he is so strong and never ill now.	enrolled <input type="checkbox"/> Mark as new correct	enrolled	1.0	0.7333333333333333

REALEC Testmaker & Testing platform

Our testing platform is currently available at <http://realectestingplatform.pythonanywhere.com/>

The source code can be found at https://github.com/nicklogin/realec_testing_platform

The testing platform relies on following instruments:

- Django web engine & database ORM
- REALEC Testmaker - a tool for extracting exercises from annotated texts previously developed by our research group (<https://github.com/nicklogin/REALEC-English-Test-Maker>) - the code was modified a bit to be used as API
- NicEdit.js - very simple and lightweight rich text editor for web (<http://nicedit.com/>)