

9

Punctuation in L2 English: Computational Methods Applied in the Study of L1 Interference

Olga Vinogradova, Anna Viklova, and
Veronika Smilga
HSE UNIVERSITY, MOSCOW

This chapter presents the results of learner corpus observations of punctuation misuses in English language writing by learners with Russian as their native language. The observations were carried out across research datasets extracted from the texts of essays from English examinations written by Russian university students. The research question concerned the possibility of first language (L1) interference as a cause of punctuation mistakes, so Russian punctuation conventions were introduced for comparison. The statistical results confirmed the strong influence of the native language in making decisions about the uses of punctuation marks. The conclusions highlight the importance for English as a Foreign Language (EFL) professionals of drawing their students' attention to the similarities and differences in applying punctuation marks in L1 and L2 starting with quite early stages in their acquisition of the new language.

“Loads of people hate punctuation, don't know why we should use it and don't really understand the purpose of it.” This is how Joanne Rudling, the author of the online course *Beginners Guide to Punctuation*, starts her introduction.¹ This attitude to punctuation has been around for a long time, as there have been arguments over the general importance of punctuation or over this or that convention ever since the appearance of the first guides and stylesheets for authors in all English-speaking countries (there are many references to such arguments throughout the book by the English journalist Henry Hitch-

1 This research was carried out as part of the HSE University Research Foundation project «2021 - Автоматизированная проверка текста, написанного на английском языке русскоязычными авторами»

ings *The Language Wars: A History of Proper English*, published in 2011). In the unpublished manuscript, *How to Punctuate*, by R. L. Trask and L. D. Wale (as cited in Jones, 1997), the authors start by addressing questions common in the 1990s—which are still relevant today in the age of texting and Twitter:

Why should you learn to punctuate properly? After all, many people have made successful careers without ever learning the difference between a colon and a semi-colon. Perhaps you consider punctuation to be an inconsequential bit of decoration, not worth spending your valuable time on. Or perhaps you even regard punctuation as a deeply personal matter—a mode of self-expression not unlike your taste in clothes or music. (p. 9)

Rudling, as well as Trask and Wale, attempt to refute such arguments, and they successfully demonstrate the importance of punctuation, as does Lynne Truss (2003) in *Eats, Shoots and Leaves: The Zero Tolerance Approach to Punctuation*. In spite of the popularity of Truss's book among the broader public, there are many papers both in the second-language acquisition (SLA) community and in computational linguistics whose authors emphasize the fact that punctuation—both the conventions of usage among native speakers and the teaching of them to language learners—still remains hugely understudied. In “What’s the Point? A (Computational) Theory of Punctuation,” Bernard Jones (1997) attempted to make English learners aware of English punctuation conventions by studying the distributions of native speakers' uses of punctuation through extensive corpus research. A similar approach was applied to study punctuation uses in a recent paper by Markov et al. (2018), but with advanced computational methods. It shows that punctuation uses in learner English were good predictors of a learners' native language, as native language continues to influence punctuation use even after a high proficiency level in a non-native language is achieved. The implications for SLA professionals are serious, suggesting that punctuation may need to be highlighted in English-language instruction in ways that it currently is not. Our research, the focus of which is not on identification of learners' first language (L1), but on relating the problems with English punctuation uses by Russian learners to Russian punctuation conventions, also confirms the insufficiency of attention paid to English punctuation in English as a foreign language (EFL) teaching traditions.

In the following chapter we review the punctuation approaches that are expressed in English punctuation guides and stylesheets, and as the main part of the research, we examine misuses of English punctuation in Russian learners' texts in English and decide whether they stem from the standard punctuation usage in Russian. At the same time, we look into and take into

consideration the cultural background and educational practices in Russia in comparison to those in English-speaking countries. We specifically look at the writing of Russian learners of English in academic settings because the need to stick to the set of conventions is stronger there than in less formal settings and the conventions themselves are more uniform than in informal production. Against the background of the well-known set of conventions (see reference sources), the research reveals which specific deviations from those conventions can be connected to the phenomenon of L1 transfer. This constituted the second goal for our study of punctuation misuses in learners' English, because EFL instructors in Russia need to be aware of which areas of punctuation their Russian learners of English have to be taught.

The research was carried out in three stages. First, we extracted research datasets for each of four major punctuation marks used in learner essays—commas, semicolons, colons, and dashes. Depending on the size and special features of each dataset, we worked out a separate method of analyzing the uses and misuses for each punctuation mark. In the set of comma errors, which was by far the largest, we implemented randomized sampling; in the set of sentences with dashes we excluded those used for numeric spans and other types of range (e.g., April–June period; 55–65 years old, etc.); in the dataset with colons, we did not take into consideration their mathematical uses between numbers. The analysis across the datasets, and the statistical results of our observations are presented in the next section. At the next stage, we annotated the uses of these four punctuation marks as correct or incorrect by comparing learner uses with the rules outlined in our reference materials. Sets of sentences with erroneous uses of the four punctuation marks were analyzed and classified. We also provide information on the corresponding Russian punctuation rules and on the ways Russian students study punctuation at the secondary school level while discussing the possibility of L1 interference among the identified punctuation deviations from the accepted norms in the English language. Methodologically, we followed the directions outlined in the seminal work of Sylvian Granger (2012). The last part summarizes all the conclusions concerning violations of punctuation rules which Russian learners of English most frequently make in their academic writing. We also point out the possible implications for approaches to teaching English punctuation.

Research Materials, Datasets, and Methodological Approach

This study was conducted on a learner corpus, REALEC (Russian Error-Annotated Learner English Corpus), of about 6,000 essays (roughly 1,463,000

words total) written in English by university students of HSE University, Moscow as part of their second-year English examinations.² The examination includes two essay tasks: an argumentative essay of about 250–350 words and a 150–250-word description of graphic materials. The native language of the vast majority of the authors is Russian.³ Most of the writers were between 19 and 21 years old at the time they sat for the examination. Some essays were written by hand, and some were typed on the computer during the examination. The errors made by students in their texts were initially annotated manually by students specially trained to identify errors in English grammar with the help of reference materials, and the total number of annotations in the corpus is over 106,400.

The initial datasets were set up by extracting all sentences from the corpus in which student authors used the following punctuation marks—commas, semicolons, colons, and dashes. The distribution of the data across the four punctuation marks is shown in Table 9.1. It is clear from the summary statistics that the available data with commas exceed those with colons, semicolons, and dashes many times over, so to make all datasets comparable, we had to randomly select five sentences out of each hundred with commas (5%). As a result, we obtained 1,930 sentences with commas for the initial comma research dataset.

A different approach had to be applied to the set of all sentences with dashes. Close to half of the 5,340 sentences with this mark used a dash with spaces around it in place of a hyphen (as in “up – to – date” instead of “up-to-date”), which we excluded, so the initial number was 3,187 dashes in 2,643 sentences. Of those, we then chose to exclude the cases in which dashes introduce spans for numeric or other values (i.e., 1940–2040, 15%–59%, or A – Z). The resulting number of sentences in the dataset was thus only about a sixth of the total count of uses of this mark.

In the set with colons another elimination was applied—we decided not to consider the uses of colons in cases like “the ratio was 3:2” (there were five of them).

2 The essays are available in a learner corpus called REALEC (Russian Error-Annotated Learner English Corpus), which is provided open access at <http://realec.org/index.xhtml/#/exam/>.

3 For a few residents of the RF, Russian may be their second language, but as they will have studied at least some subjects in Russian at the secondary school level, and as the language of instruction in their studies at HSE University is Russian, their proficiency in Russian will be very close to that of native speakers; besides, overseas students with different native languages do not take the English examination from which we get written essays for the learner corpus.

The dataset with semicolons was made up of all uses of this mark in the corpus except for eight sentences in which this punctuation mark was used in misspelled words.

For each of the four punctuation marks under observation, we counted the number of sentences in which a certain punctuation mark was included in the error span which annotators identified with the Punctuation tag, the fourth most frequent type of error annotated in REALEC (8,056 occurrences out of 106,401 manual annotations in the corpus). In Table 9.1, summary statistics are shown for each mark.

Table 9.1. Statistics of Data Extracted from REALEC and of the Sets Chosen for the Research

	Comma	Dash	Colon	Semicolon
# marks in the corpus	60,165	3,187	1,779	474
# sentences in the corpus with the mark	38,599	2,643	1,729	342
# sentences extracted for the research dataset	1,929 (5% random selection)	926 (not in spans)	1,724 (not between numbers)	334
# sentences with Punctuation tag	87	11	20	27

However, we have to admit that annotating with the Punctuation tag in the corpus—at least for some annotators—was not completely consistent, and the corrections conform to the rules in the reference materials only in clear and unambiguous cases, as the annotators themselves were students and learners of English and thus prone to the same weaknesses in the use of punctuation. Therefore, for the sake of bolstering the reliability of the data, the research team of three EFL experts annotated the sentences in each dataset to identify correct and incorrect uses of punctuation on the basis of the authoritative reference sources—Jones (1997), Huddleston and Pullum (2002), and Straus and Kaufman (2008). For unclear cases, we used additional reference sources on the following websites:

- <https://dictionary.cambridge.org/grammar/british-grammar/punctuation>
- <https://www.ef.com/wwen/english-resources/english-grammar/punctuation/>
- https://www.grammarbook.com/punctuation_rules.asp

- <https://linguapress.com/grammar/english-punctuation.htm>

While making decisions about whether or not a particular use of a punctuation mark is appropriate, we also worked out classes of punctuation misuses. The statistics on errors identified by our team of EFL instructors on the basis of reference sources is shown in Table 9.2 along with the percentage of sentences that were considered erroneous.

Table 9.2. Statistics on Comma, Dash, Colon, and Semicolon Uses and Misuses in Research Datasets.

	Comma	Colon	Dash	Semicolon
# sentences extracted for the research dataset	1,929	1,724	926	334
# sentences chosen to be in the research dataset ⁴	1,873	1,638	863	269
# sentences marked correct in the use of the punctuation	1,300	1204	574	171
# sentences marked incorrect in the use of the punctuation	573	434	289	98
percentage of errors	31%	36%	33%	36%

Categorization of punctuation errors in Russian learner texts

In all four datasets, we noted cases when a different punctuation mark was required instead of the one used by the author of the text; we also counted redundant uses of each punctuation mark; whenever possible, we noted absent punctuation where the mark was necessary. By far the largest set of punctuation errors is composed of comma misuses. The analysis of the comma research dataset resulted in identifying five classes of the most frequent mistakes and five more classes of mistakes that appeared with lower frequency, but we can say that there were nevertheless enough errors represented in the corpus to consider them characteristic for Russian learners. We also identified two classes of errors which border on a variation of the norm rather than being considered erroneous, but because the rules related to these last two classes are included in the majority of the reference materials (see the discussion in the last section of this paper), we chose to take them into consideration along with the other ten. As a result, we worked out 11 different classes of

⁴ We eliminated sentences with a low level of proficiency in English in which the meaning was totally incomprehensible to our team.

punctuation errors that are listed in Table 9.3 together with examples from the essays in the learner corpus.

Table 9.3. Categorization and Examples of Errors in the Use of Commas in REALEC⁵

	Class of errors	Examples
1	A redundant comma after the main clause in front of a subordinate clause	But we can't ignore the fact, that the main aim of high education is to bring up great professionals. And the problem raises, when it concerns young people who are truly believe in their own independence. For them it is normal, when there are more male or female students in subjects.
2	Confusion with commas around relative clauses	From the graph it can be noted that the proportion of the population aged 65 and over increases from 1940 to 2040 years in all countries, which are presented. Thirdly, such equality would negatively affect not only a person, who was not accepted by university, but also a person, who was. Firstly, there are faculties, where men's qualities are incredibly needed.
3	Confusion with commas around discourse-organizing units	All in all_ the graph shows, that the amount of people aged 65 and over is not static between 1940 and 2040, it changes dramatically and in whole is going up. If universities ever would count gender as a decisive factor, the whole education system would be broken_ in my opinion. Comparing the proportion of populations_ it can be seen that since 1940, the average levels had grown up from 5 to 10% to more than 25%.
4	Absence of comma to coordinate independent clauses with <i>and/but/or etc.</i>	In conclusion, I would like to emphasize that both man and woman should be able to choose their subject independly_ and the amount of people to every discipline by no means should be determined by universities. First of all, in modern world every person should have the same rights as others_ and any discrimination is prohibited.
5.1	Colon required instead of comma	All in all the graph shows, that the amount of people aged 65 and over is not static between 1940 and 2040, it changes dramatically and in whole is going up.

5 NB: All words in the sentences in Tables 9.3 and 9.5–9.7 are represented with the author's spelling, grammar, and vocabulary fully preserved; punctuation marks regarded as incorrect are bolded; for erroneous absence of a punctuation mark, we underline the space where the mark is needed.

	Class of errors	Examples
5.2	Semicolon required instead of comma	To draw a conclusion, we may say that not only social or economic tendencies, but also government policies might cause certain changes in social proportions, consequently, these policies and tendencies should be accurately governed. So we may notice that all of the lines go up, it means that proportion of population aged 65 and over within years will be higher and higher. Some people believe that it's inappropriate, to make universities accept equal numbers of male/female students, others agree that it will improve current situation in education.
5.3	Dash required instead of comma	Maybe this way we will miss a gender equality, but we will achieve a much more important thing, the people equality. What is worth doing, is worth doing well, it is unwise to forbid it.
6	Redundant comma in different contexts (not the same as in Class 1)	That's why I suppose that universities have to try to accept equal numbers of representatives of both sex, for every subject. Some people believe that it's inappropriate, to make universities accept equal numbers of male/female students, others agree that it will improve current situation in education. What is worth doing, is worth doing well, it is unwise to forbid it. To specialize on a type of work you're best at, is the most efficient way to organise society.
7.1	Confusion with commas around appositives	To sum up, everyone_ both males and females_ has a right and opportunity to choose their profession and study in a place they want.
7.2	Confusion with commas around attributive participial construction	Compared to LinkedIn, Instagram is more used by young people, aged from 18 to 29. Some people, living in far regions, haven't got a cinema near. Nowadays when peoples health is affected by different dangers, caused by the development of humanity, this topic is especially problematic.
8	Confusion with commas in comparative constructions	First of all, in that situation all of students should feel better, than in the situation, when in class there are more girls or boys. It is interesting that LinkedIn is as popular among 65+ users, as among 18-29 users.
9	Oxford comma	So, you should stop drinking a lot, spend holidays with your son or daughter_ and try to be a good example for him.
10	Absence of comma after a subordinate clause in front of the main clause	If the weather is good all over the year_ then it is not the problem to provide the population with food, because you can seed all the year. Secondly, most music fans, if they like the group_ will buy the T-shirts or something like that with photos of favourite groups.

	Class of errors	Examples
11	Ellipsis	The persantage of unemployment stood unchanged in N. Africa at 12,5%, in S. Asia_ at 3,9%.

For comparison with similar prior research, Table 9.4 gives a list of common rules taught in the American EFL tradition; it was compiled by Israel et al. (2012, p. 286) by looking through guidebooks and stylesheets available for American EFL professionals (rules concerning use of punctuation in special categories like numerals, titles, and proper nouns rather than in the sentences are omitted in our representation).

Table 9.4. Common Comma Uses (in American EFL Practices)

Rule	Example	Corresponding error classes from Table 9.3
Elements in a list	Paul put the kettle on, Don fetched the teapot, and I made tea.	Absence of comma to coordinate independent clauses with <i>and/but/or, etc.</i> , class 4
Initial word/phrase	Hopefully, this car will last for a while.	Confusion with commas around discourse-organizing units, class 3
Dependent Clause	After I brushed the cat, I lint-rollered my clothes.	Absence of comma after a subordinate clause in front of the main clause, class 10
Independent Clause	I have finished painting, but he is still sanding the doors.	Absence of comma to coordinate independent clauses with <i>and/but/or, etc.</i> , class 4
Parentheticals	My father, a jaded and bitter man, ate the muffin.	Confusion with commas around appositives, class 7,1
Conjunctive adverbs	I would be happy, however, to volunteer for the Red Cross.	Confusion with commas around discourse-organizing units, class 3
Contrastive Elements	He was merely ignorant, not stupid.	Confusion with commas in comparative constructions, class 8

It is clear that practically all the classes of errors we identified in Russian learners' writing were the same as the classes of recommended uses of commas on the list compiled by the American authors, whose research aim was to achieve highly efficient automated identification of punctuation errors in EFL learners' writing.

The distribution of the occurrences of erroneous uses of commas in our research dataset is given in Figure 9.1. The numbers next to the lines for each class of errors demonstrate how many sentences contained this type of error

from the 573 sentences that were regarded as erroneous by the three EFL experts. This subset of 573 sentences was a little under a third of the original 1,980 sentences. The rest of the sentences were 1,300 sentences with correct uses of punctuation (two-thirds of the research dataset) and 56 erroneous sentences in which the level of the author’s proficiency was clearly so low that we had enough ground for doubts about his or her awareness of any punctuation in the English language at all.

Some of the labels we have applied in the distribution in Figure 9.1 constitute specific classes of errors identified in the comma dataset—namely, the four types of errors most frequently made by Russian learners of English:

- a redundant comma after the main clause and in front of the subordinate clause;
- confusion with commas around different types of relative clauses;
- confusion with commas around different discourse-organizing units;
- absence of a comma before a conjunction coordinating independent clauses (and, but, or).

One more frequently made mistake was using a comma where another punctuation mark (e.g., dash, colon, or semicolon) was required instead of the comma. Other types of errors were less frequent in comparison with the first five, and four types out of six included a few specific types of a similar nature, so the count was carried out for all of them as a class.

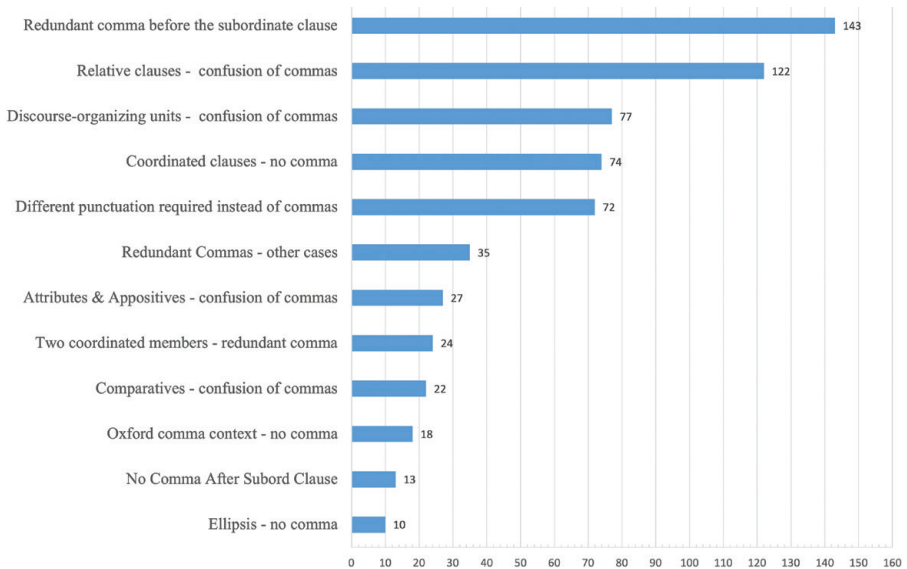


Figure 9.1. Distribution of classes of punctuation errors in the use of commas.

Examples concerning the other three punctuation marks are presented in Tables 9.5, 9.6, and 9.7.

Among the incorrect uses from the research datasets for the three other punctuation marks for many contexts, there was often a question of the variability of dashes and colons, or colons and semicolons, or even of all three marks. For two examples from Huddleston and Pullum (2002), for example, three native English speakers regarded the following punctuation uses as equally appropriate:

- He had forgotten the thing he needed most: a map.
- He had forgotten the thing he needed most – a map.
- He had forgotten the thing he needed most; a map would have saved him a lot of trouble.
- He had forgotten the thing he needed most – a map would have saved him a lot of trouble.
- All students had to take a language: Sue took French.
- All students had to take a language; Sue took French.
- All students had to take a language – Sue took French.

Our decision was to regard any sentence similar to those in the examples here as correct. The only choice we applied was the following (and it may possibly look suspicious to some experts, but it consistently proved preferable in our search in corpora of writing produced by native speakers): in cases of general expression followed by a list of specific entities (noun groups, verb groups, etc.), a dash has to introduce the illustration, and a colon has to be used for explanation. Prototypical examples of erroneous uses of a colon are represented in Table 9.5 with a number of similar occurrences out of 434 sentences (see Table 9.1) given in parentheses. The misused colons in the examples are given in red.

Table 9.5. Examples of Errors in the Use of Colons in REALEC

	Type of error (#sentences out of 434)	Examples
1	Dash required instead of colon (206)	From 194 to 1960 the number rose steadily in both countries: in the USA and in Sweden, while Japan experienced a slight fall in the number of old people by 4 per cent.
2	Comma required instead of colon (58)	On my opinion: the main purpose of social media is communication. For example: 70% males and only 30% females got postgraduate diploma.

	Type of error (#sentences out of 434)	Examples
3	Semicolon required instead of colon (64)	A good worker should share the interests of the company: if he does not, he will not work for success. If the government decided to increase the number of sports facilities it does not mean that all people decided to go in for sport: some of them, moreover, are not able to do it.
4	Redundant colon (106)	From position of airlines we can argue that: air market is not very high profitable, attentional taxes will decrease amount of passengers and increase price of tickets. And there is a task: how to grow the level of people health?

Next we give examples of dash misuses in Table 9.6, and then in the last section we go on to discuss why Russian learners produced strange sentences with redundant dashes, which we counted separately in our annotation of errors (21 out of 104 total redundant uses of dashes).

Table 9.6. Examples of Errors in the Use of Dashes

	Type of error (#sentences out of 289)	Examples
1	Comma required instead of dash (74)	The more reposts — the more audience, listeners and followers. When you make copies of film or a music without author’s permission — you can be punished.
2	Colon required instead of dash (61)	That is why import is necessary — it allows people to choose that they want to eat from the wide range of products.
3	Semicolon required instead of dash (14)	To my mind, the last position is not right at all — it sounds like a feminism.
4	Redundant dash (83+21)	The last group — of children was only 14,3% and is going to descend to 11,5%. Using hi-tech to motivate people be more active — is a possible solution. In addition, the first small step for public health — restrictions on fast food.

The dataset for semicolons in REALEC was the smallest dataset in this study, as there were only 98 sentences in which the researchers identified misuses of semicolons. In many sentences there was no ground whatsoever for using a semicolon, so in the end we considered the errors in the use of semi-

colons to be typos rather than inappropriate decisions to use a semicolon. This must have happened, for one, in this case: “However; in airbus usually people sleep.” Besides, Russian punctuation conventions for semicolons cover fewer cases than English rules do. As a result, there was little ground to suspect L1 interference. However, to make a certain claim, we need to be confident that we have sufficient data, and the dataset was too small. Out of 98 sentences with erroneous uses of a semicolon, in the majority—61—we came to the decision that a comma was required instead of a semicolon.

Table 9.7. Categorization and Examples of Errors in the Use of Semicolons

	Type of error (# sentences out of 98)	Examples
1	Comma required instead of semicolon (61)	<p>The number desktop users accounted for approximately 130 million in 2012; whereas it fell slightly by 2013 accounting about 130 million users.</p> <p>As we know, sport can normilise blood pressure; help your heart work and prevent the development of obesity.</p> <p>The positive point of starting university studying after finishing school is that you will learn new science; and get new information.</p>
2	Dash required instead of semicolon (14)	<p>The population has been broken into two main types; general and prison.</p> <p>In 2000 the highest amount of children without access to primary school education was in Africa; 23,7 million girls and 20 million boys.</p>
3	Colon required instead of semicolon (10)	<p>In 2015 there was a slight decrease of this rate in the Middle East; it decreased by 1,4%.</p> <p>The second place was occupied by 26 - 40 year olds people; their figure accounted 30 %.</p>
4	Redundant semicolon (13)	<p>The positive point of starting university studying after finishing school is that you will learn new science; and get new information.</p> <p>And I ‘m absolutely sure that only strong measures in each field; may lead us to result, which consists in unpol-luted nature and health nation.</p> <p>The tendency towards such a ‘passive’ way of participat-ing is growing more and more popular, which has led to a great amount of agrument; whether it is worth doing at all.</p>

Discussions Concerning English Punctuation Acquisition

While trying to establish the types of errors that may have been made under the influence of learners’ native language, the researchers compared the systems of punctuation rules for English and for Russian. Table 9.8 summarizes the results of the comparison with short made-up examples with the same meaning in English and Russian. Columns “Eng” and “Rus” have a plus if a rule that mandates the use of punctuation exists in English or Russian correspondingly, and they have a minus if there is no such rule. The third symbol used in columns “Eng” and “Rus” is “+/-”, and it means that there is a rule prescribing when to apply and when not to apply a punctuation mark in the corresponding context. However, Table 9.8 only includes cases in which the distribution in using or not using punctuation is different for English and Russian. The table is followed by detailed explanations of the different conventions in each language.

In the examples in the last column, the punctuation mark or its absence—according to the rule—is enclosed in brackets, and when the rule prescribes no use of a mark, there is just space in brackets. So, it is clear which areas correlate in the English and in the Russian equivalents. The optional uses of a punctuation mark are enclosed in parentheses.

Table 9.8. Differences in English and Russian Punctuation Conventions for the Same Contexts

	Presence of punctuation		Examples
	Eng	Rus	
A comma after the main clause in front of the subordinate clause	-	+	I will tell you this[] when you come. Я расскажу тебе это[,] когда ты придешь.
Commas around relative clauses	+/-	+	I know a rule[,] which works here. Я знаю правило[,] которое здесь работает. I know this rule[,] which is hard to apply. Я знаю это правило[,] которое применять трудно.
Punctuation around discourse-organizing units	comma/ semi-colon	comma/-	I know this[,] however[,] I don’t know what to do. OR I know this[;] however[,] I don’t know what to do. Я это знаю[,] однако[,] я не знаю, что делать. It is difficult to work with him[,] however[,] it is also nice. Работать с ним трудно[,] однако[] и приятно.

	Presence of punctuation		Examples
	Eng	Rus	
Comma after discourse-organizing units introducing lists	+	-	You will want to bring many backpacking items[./;] for example, sleeping bags, pans, and warm clothing. Вам следует взять много походных вещей[,] например[] спальные мешки, сковородки и теплую одежду.
Commas after fronted units	+/-	-	Right from the start[(,)] I knew the rules. С самого начала[]я знал эти правила.
Comma for coordinating independent clauses with <i>and/but/or</i> , etc.	+/-	+/-	I learned about these rules[(,)] and it was helpful. Я узнал про эти правила[,] и это было полезно.
Punctuation after subject clause introduced with <i>wh</i> -word or expressed by the infinitive.	-	comma/dash	What you know[] is true. Что (То, что) ты знаешь [, /—] (это) правда. To tell her[] is the hardest thing to do. Сказать ей [—] это самое трудное.
Comma after the infinitive of purpose after the predicate	-	+	He came there[] to see her. Он пришел туда[,] чтобы ее увидеть.
Commas around attributive participial constructions	+/-	+	People[] living in remote regions[] haven't got electricity. У людей[,] живущих далеко[,] нет электричества. Living in remote regions[,] people travel more. Живя в отдаленных районах[,] люди ездят больше.
Commas in comparisons with <i>than</i>	-	+	To tell her is much harder[] than to conceal it from her. Сказать ей гораздо труднее[,] чем скрыть это от нее.
Punctuation in cases of ellipsis	comma	dash	In 1920, the highest percentage was shown by the elderly, but in 2010[,] by the young. В 1920 самый высокий процент показали пожилые, а в 2010 [—] молодые.

The first convention in Table 9.8 covers the class of comma misuses most frequently demonstrated in Russian student essays written in English (see Table 9.3, class “Redundant comma in front of the subordinate clause after the

main clause”). These errors constitute the first area in which the phenomenon of language transfer begs to be explored, and the following considerations are the reasons for thinking that it is indeed L1 interference. First, in Russian a comma introduces and completes any subordinate clause, no matter whether the dependent clause comes before or after the main clause. This rule is rigid, and absence of a comma before or after the dependent clause in a Russian sentence is seen as a blunder by native speakers of Russian. This punctuation rule is taught in Russian classes at the secondary school level. Russian punctuation gets even more important in Russian grammar classes in the final years in Russian high schools, as all students take their obligatory school-leaving examination in the Russian language, and violations of punctuation conventions are strictly penalized. The results of this examination are submitted (along with the results of a few other examinations) in applications for university studies. As a result, knowledge of Russian grammar, including punctuation, is of much importance, at least for those students who apply for university programs. This means, in turn, that students graduating from secondary schools remember the main punctuation rules very well—as the Russian saying has it, “they know it well even if they are woken up in their sleep.” The result for the acquisition of English as a foreign language is that it has become very difficult for Russian learners of English not to automatically separate any subordinate clause with commas in their writing in English—hence a lot of erroneous commas in contexts like “think, that” or “fact, that” or “do it, when”.

The additional proof of L1 interference is the fact that a comma is always recommended after a subordinate clause of longer than four to five words introducing a main clause (see Dependent Clause type among “American EFL comma rules” in Table 9.4). Russian learners almost never omit this comma—out of 573 sentences with errors in the use of commas, there were only 13 in which the dependent clause was not separated from the main clause with a comma. Compare this number with 143 sentences with a redundant comma in front of the subordinate clause after the main clause. In our view, this has to do with the same strict rule in Russian—a subordinate clause always has to be separated from the main clause with a comma, on both sides if applicable, so students overwhelmingly do it in English too, not differentiating the use of commas for subordinate clauses positioned before and after the main clause, nor looking at how many words there are in the subordinate clause. In the end, the phenomenon of L1 transfer in this particular area appears to be positive for one sequence of clauses but clearly negative for the other.

The next row in Table 9.8 also relates to the frequent error, “Confusion with commas in relative clauses” (the second in frequency of occurrence) in Table 9.3. The reason for the frequency is the fact that relative clauses con-

stitute an especially difficult grammar area for Russian learners, and again, interference with L1 is certainly at play. In Russian, there is no differentiation of relative clauses, and the Russian equivalents to English relative clauses (traditionally called attributive subordinate clauses) are always separated from the main clause with a comma. Furthermore, they all are introduced by one and the same conjunction in Russian equivalents of defining and non-defining relative clauses, and the Russian equivalents of coordinate clauses are introduced by a completely different conjunction. Accordingly, acquisition of the last type of relative clause poses no difficulty at all for Russian learners, while the notion of the division of clauses into defining and non-defining types presents an entirely new concept. It should not be surprising, then, that it requires more time and much more exposure to authentic English sentences with relative clauses than most students usually get at the university level. Moreover, relative clauses also require sensitivity towards linguistic subtleties on the part of teachers of English so that they can draw their students' attention to all the differences between defining and non-defining relative clauses when they come across them in authentic texts, tests, and other materials. Furthermore, sensitivity towards subtle differences is sometimes lacking even in texts written by native speakers of English⁶—misused commas around relative clauses can be identified in authentic English written production. A simple experiment proved this: the search for “which” in a corpus of British academic writing in English⁷ returned among the first 20 sentences (out of 29,167) two sentences with errors in the use of commas around a relative clause. The sentences are copied here with the error spans for the comma and the conjunction in brackets:

1. In conclusion racism, which can be defined as an ideology which categorises people as inferior according to their race, and is put into practice through many policies and actions[,] which seek to exclude non-whites from many areas of society, can still be seen in today's society.
2. The few hospitals set up to provide for the working classes were poor quality, and the women were subjected to harsh treatments, which the upper classes[] who could afford to pay for superior treatment[] were not.

6 There is a comment on that feature in relation to punctuation around relative clauses in Moore, 2016, p. 2.

7 The search was carried out on the platform called SketchEngine, which hosts many language corpora and is equipped with tools for searching contexts for separate words, phrases, or sophisticated combinations of elements of the text, and the results of the search were received from the British Academic Written English Corpus (BAWE).

In sentence (1), the relative clause is in our view defining because it restricted the set of all policies and actions to those that seek to exclude non-whites, so no comma was required in front of the relative clause; however, the comma after the relative clause was justified because it is also the end of the long non-defining relative clause, starting with “which can be defined.” In sentence (2), the relative clause introduced with “who” was non-defining, in our opinion, as “upper classes” in this sentence was the opposite of “working” classes, namely, people who are not poor and who can thus “afford to pay,” so the relative clause required commas both before and after it.

The second conclusion concerning relative clauses is that, just as in the prior case involving subordinate clauses, Russian learners appear to benefit from applying the rule of obligatory commas around relative clauses transferred from Russian punctuation when they apply it to English non-defining and coordinate clauses. However, the same Russian rule causes redundant commas in the case of defining clauses, which—unfortunately for Russian learners—are more frequent than the first two types.

In the third and fourth rows in Table 9.8, we again have contexts in which errors in the use of commas occur very frequently—commas around discourse-organizing units. It is true that Russian and English separate these expressions with commas in very similar ways, and in both languages, there are words used as discourse-organizing elements that sometimes have commas around them and sometimes do not. The problem is that the rules prescribing different comma behavior with such words differ in English and in Russian. Furthermore, the rules are sophisticated, and applying them, again, requires linguistic sensitivity, and for some speakers they become a sort of grey area in English punctuation conventions. We can quote here Straus and Kaufman’s (2008) well-known American grammar book with a large chapter devoted to punctuation:

Use either a comma or a semicolon before introductory words such as *namely*, *that is*, *for example*, *e.g.*, *for instance* when they are followed by a series of items. Use a comma after the introductory word. Examples: You may be required to bring many items, for example, sleeping bags, pans, and warm clothing. OR You may be required to bring many items; for example, sleeping bags, pans, and warm clothing. OR You may be required to bring many items, e.g., sleeping bags, pans, and warm clothing. (pp. 57-58)

In the end, a reader is left with the idea that it does not matter whether or not the discourse-organizing units given before the list of items are followed

by a comma, but a corpus search in any large collection of native English writing demonstrates that at least “namely,” “that is,” “for example,” and “for instance” are indeed overwhelmingly followed by a comma.

The differences between English and Russian comma conventions with discourse-organizing units are shown in Table 9.8 with examples of English “however” and Russian “*однако*.” The former has to have commas around it “when they are used as interrupters” (Straus & Kaufman, 2008, p. 57) and does not require commas when used in the same way as “how” (“however you do it”), or with an adjective or adverb following it (“however difficult it may seem”; “however smartly you solve it”). The latter, the Russian equivalent, is always introduced with a comma, but is followed by a comma only when there is a complete clause after it and is not followed by a comma when there is a coordinated member, and not a complete clause, after it (literally: “hard however nice”). Given that Russian students had substantial difficulty differentiating the two uses in Russian, it is not surprising that they have overgeneralized both the “Russian” and the “English” conventions by applying them in writing in English and getting, as a result, confusion with commas of the following two types—with a comma or semi-colon in front of “however,” but lacking a comma after it: “In other words, some people find happiness in their family, however others prefer to be alone;” and with a redundant comma like “He tries to solve a problem, however, difficult it may seem.”

Another example of the difficulty with punctuating discourse-organizing units is that words like “thus” get a comma when they appear at the beginning of a sentence but can do without commas in the middle of a sentence like in this example: “He overcame his incertitude and thus won the respect of his classmates.” In Russian, there are two equivalents of “thus”—one is a word combination “*таким образом*,” which always requires commas after it or on both sides, and the other is just one word “*так*,” which can be used in exactly the same two ways as “thus” in English—as a discourse-organizing unit at the beginning followed by a comma, or as an adverbial modifier without commas around it. If a certain learner bears the first equivalent in mind, he or she will develop redundant commas in “He overcame his incertitude and[,] thus[,] won the respect of his classmates.” If the learner thinks of a one-word equivalent of “thus,” their use of commas with this word in English may turn out correct.

The next context from Table 9.8 that needs explaining is when two clauses get coordinated with the help of conjunctions “and,” “but,” “or,” “as well as,” and some others. According to Straus and Kaufman (2008, p. 55), the comma is optional when two coordinating clauses are independent in English: “Use a comma to separate two strong clauses joined by a coordinating conjunction—

and, or, but, for, nor. You can omit the comma if the clauses are both short.” When two subordinate clauses get coordinated, the English convention is not to have a comma in front of “and,” “but,” “or,” and “as well as,” but if the first clause is long, it often has other constructions in it, and a comma at the end of some constructions often falls on the position in front of the coordinating conjunction. This may explain the fact that not many sets of punctuation tips include the rule that a comma should not be placed between the two coordinated subordinate clauses. As a result, in English coordinating clauses of both types—dependent and independent—a comma is somewhat optional. In Russian, on the contrary, the rule for a comma in coordinating clauses is very strict: if the two clauses are independent, then the comma is obligatory, while the comma is forbidden in coordinating two subordinate clauses. The same rule, but with different variations, makes coordinating English clauses an error-prone area for Russian learners.

One more class in Table 9.8—“Comma after subject clause introduced with ‘wh’-word or expressed by the infinitive”—correlates with the classes “Comma required instead of dash” and “Redundant dash” in Table 9.6. Errors in these two classes are clearly caused by interference with Russian practices. The first of the examples in Table 9.8 represents a clause introduced by “what” to express the subject of a larger clause. The rule in English forbids ever separating a subject from a predicate with a comma. However, the same construction in Russian obeys the rule that requires separating any clause with commas, including a subject clause. Nevertheless, the Russian language allows for the alternative of omitting the predicate when it is expressed by the verb “BE” in the present tense (and it is omitted in almost all such cases); as a result, a dash appears instead of the omitted verb, and it may be followed by the Russian equivalent to “this.” The same thing happens to subjects expressed by infinitive phrases, and again in Russian sentences with the verb “BE” in present tense as a predicate, this predicate may get omitted, causing the appearance of a dash. When a Russian learner of English constructs the direct equivalent of the Russian sentence of this type, it looks completely un-English (even though his and or her level of English proficiency may otherwise be high enough), as in this sentence: “In addition, the first small step towards public health—restrictions on easy availability of fast food.” (It means that the first small step towards public health IS to impose restrictions on . . .)

Still another point of difference between English and Russian punctuation rules concerns specific constructions for which these two languages choose different punctuation marks. Quite a few comparative constructions operate differently punctuation-wise: comparison with “than” never requires

a comma in front of “than” in English, but in its Russian equivalent a comma is always required before “*чем*,” in the construction with two forms of the comparative degree of adjectives or adverbs, the common punctuation mark in English is a comma, which is sometimes used in Russian too, but the use of dashes between the two groups with the comparative degrees is more common in Russian, as in this: “The more money, the better for them.” “*Чем больше денег – тем лучше для них.*” OR “*Чем больше денег, тем лучше для них.*”

Finally, the choice of the punctuation mark when something has been omitted from the second (or third) clause is also different in the two languages—a comma in English, but a dash in Russian (see examples in the last row in Table 9.8).

Conclusions

Data from the learner corpus clearly indicate that Russian learners had serious problems in the use of punctuation while writing in English, as about a third of the sentences on average may at best look strange and at worst confuse a reader to the extent of becoming incomprehensible. Even one redundant comma in a simple sentence like “We couldn’t figure out, when the change started—it was unbearable.” may lead the reader to the meaning that the people in the sentence couldn’t figure out something that was discussed before this sentence, while the clause introduced with “when” may be understood as referring to the time of change as unbearable, while the author clearly meant that they couldn’t figure out when the change started, and that was unbearable. As a result, the Russian rule of separating all clauses in the sentence with commas may be the cause of confusion when writing in English.

To consider punctuation an area that should only be taught to students mastering the highest levels of proficiency in English means to seriously hinder students’ development in the acquisition of the target language. Being accustomed to rigorous observance of punctuation rules in their mother tongue, Russian learners are probably compelled to pay more attention to the use of punctuation marks in the foreign language they are trying to master than many other learners. Besides, when some conventions in the use of punctuation are the same, learners may well generalize the sameness to the areas that are subject to rules different from those in the native language. Instead of meeting those needs, the Russian tradition of teaching English as a foreign language has largely neglected many areas where punctuation plays a role in overcoming ambiguity and confusion in order to get the idea through to others efficiently. The result is that Russian learners suffer from insufficient

exposure to existing standards and conventions and therefore develop lower resistance to interference coming from their native language.

Punctuation similarities and differences between the native and the target languages have to be revealed to the learners explicitly, and the mere exposure to authentic materials is hardly enough to acquire the necessary skills, as the percentage of errors in writing production at the intermediate to upper-intermediate level of proficiency in English shows. In Elkılıç et al. (2009), the authors state in their conclusion that in order to avoid L1 interference, Turkish EFL students should be taught English punctuation explicitly and in comparison with Turkish punctuation conventions. Similar to the goals of our research, the authors made an attempt to identify which areas of punctuation uses have to be taught in comparison with and in contrast to the Turkish punctuation rules.

We do not mean to say that the problem of teaching too little punctuation is only a problem in the Russian EFL tradition. Learners of English with many different L1s have complained about poor exposure to punctuation conventions in the national traditions of teaching English: there has been a discussion on the internet over the last few years about the fact that Chinese EFL professionals teach almost no punctuation. We can see it, for example, in the following Quora thread, “First of all, it’s a language education failure—no systematic course to emphasise the importance of punctuation in English, only a few students self-study after class” (Li, 2015). It does not take much effort to find similar complaints expressed online by learners of English with other native languages.

The attitude to punctuation on the part of English-speaking cultures also plays a role in insufficient attention to punctuation for learners of English. As was demonstrated in the beginning of this paper, many speakers of English think that the context is enough for the reader to work out the meaning without thinking about punctuation marks in the text. Some say that punctuation is gradually dying or fading out in English, especially in the era of very short messages that the new communication technologies dictate. But one should not forget that the people saying it are those who, back when they were young children, had years of constant processing of conventionally used punctuation without even realizing what the conventions themselves were. This is what learners of English need to work very hard at if they ever want to catch up on what they did not get as children. So, learners need help, they need focused exposure, they need EFL professionals patiently pointing out all the commas, colons, semicolons, and dashes used in authentic texts and misused by learners. The teaching traditions in Russia—and all around the world—will do more good for learners of English if they become more punctuation-friendly!

References

- Elkılıç, G., Han, T., & Aydin, S. (2009). Punctuation and capitalisation errors of Turkish EFL students in composition classes: An evidence of L1 interference. In Padem, H. (Ed.), *Proceedings of International symposium on sustainable development* (pp. 279-284). International Burch University. <http://tara.sdu.edu.tr/vufind/Record/95849>
- Granger, S. (2012). How to use foreign and second language learner corpora? In A. Mackey & S. G. Gass (Eds.), *A guide to research methods in second language acquisition* (pp. 7-29). Basil Blackwell.
- Hitchings, H. (2011). *The language wars: A history of proper English*. John Murray Publishers.
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.
- Israel, R., Tetreault, J., Chodorow, M. (2012, June 3). *Correcting comma errors in learner essays, and restoring commas in newswire text* [Paper presentation]. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 284-294). <https://dl.acm.org/doi/10.5555/2382029.2382065>
- Jones, B. (1997). *What's the point? A (computational) theory of punctuation* [Unpublished doctoral dissertation]. Centre for Cognitive Science, University of Edinburgh.
- Markov, I., Nastase, V., & Strapparava, C. (2018). Punctuation as native language interference. *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 3456-3466). Association for Computational Linguistics. <https://aclanthology.org/C18-1000/>
- Mingjie, L. (2015, May 1). Why do Chinese people constantly mix up English style punctuation with Chinese style? [Online forum comment]. Message posted to <https://www.quora.com/Why-do-Chinese-people-constantly-mix-up-English-style-punctuation-with-Chinese-style>
- Moore, N. (2016). What's the point? The role of punctuation in realising information structure in written English. *Functional Linguistics*, 3, 6. <https://functional-linguistics.springeropen.com/articles/10.1186/s40554-016-0029-x>
- Rudling, J. (n.d.). "Beginner's Guide to English Punctuation." *Curious.com*. Retrieved September 6, 2021 from <https://curious.com/howtospell/series/beginners-guide-to-english-punctuation>
- Straus, J., & Kaufman, L. (2008). *The blue book of grammar and punctuation*. Jossey-Bass.
- Truss, L. (2003). *Eats, shoots & leaves: The zero tolerance approach to punctuation*. Gotham Books.