

School and Gender in Numbers: A Stylometric Insight into the Lexis of Teenagers’ Description Essays

Michal Místecký¹, Lucie Radková²

Abstract. The goal of the paper is to make use of four quantitative indicators (MATTR, ATL, Q, and VD) to study vocabulary richness, lexis complexity, text activity, and syntactic complexity of Czech schoolchildren’s writing tasks. The corpus comprises 60 texts written by elementary-school and secondary-school pupils, distributed equally according to the gender and the education level (30 each); the genre of the task was description (“my bedroom” for the elementary schoolers, and “class/school of the future” for the secondary ones). The research is carried out in three distinct comparisons (schools, genders, and the mixture of both), and the results are interpreted with the assistance of a pedagogical professional. At the end of the study, a detailed summary of the outcomes is provided.

Keywords: MATTR, ATL, Q, VD, stylometry, quantitative linguistics, pedagogy, school, gender, Czech.

1. Introduction

Recently, a lot of studies have appeared focusing on quantitative investigation of various linguistic discourses. With the literary and politics-oriented studies leading the way (cf. Andreev et al. 2018; Místecký 2018; Dai, Liu 2019), there are other fields in which a stylometric analysis can bear needed fruit (David et al. 2014; Čech 2016). In the present paper, the sphere of Czech schoolchildren’s writing tasks will be researched, with the goal to enrich both didactic scholarship, and to prove the usability of the quantitative methods in the domain.

The school writing has been given much attention in the last years (cf. Čechová at al. 2008, Holubová 2014, Štěpáník, Holanová 2017, Rysová 2017), which culminated in devoting an entire section in the new Czech handbook of stylistics (Hoffmannová et al. 2016) to the subject. However, to our knowledge, no paper has been published yet to study the matter from the viewpoint of quantitative measurement. This is why this article will try to fill the gap in the research, and may become a pioneering piece for other analyses to come.

¹ University of Ostrava, Ostrava, the Czech Republic; e-mail: mmistecky@seznam.cz.

² University of Ostrava, Ostrava, the Czech Republic; e-mail: lucie.radkova@osu.cz.

2. Methods

Out of the numberless methods used in stylometry, four have been selected to start the investigation off (MATTR, ATL, Q, and VD). The set is a combination of the pragmatic approach – all the indexes can be computed automatically, with no manual work needed –, and the endeavour to take various stylistic factors into account. Moreover, all the calculations have been proved to be independent of text length, and also of each other (cf. Zörnig, Místecký 2018).

First, MATTR (Moving-Average Type-Token Ratio) has been used, as it seems to be the most effective and widespread tool of assessing vocabulary richness (cf. Covington, McFall 2010). Its formula, based on TTR (Type-Token Ratio), but taking into consideration the division of a text into moveable sections (windows), is as follows –

$$MATTR(L) = \frac{\sum_{i=1}^{N-L} V_i}{L(N-L+1)}.$$

In the formula, L stands for the number of tokens in one window, V_i for the number of types in one window, and N for the total of the tokens in the text. Given the length of the studied tasks, the window size in the present research was set at 30. The basic unit of the research is, due to the workings of the software, the word-form.

The count will be exemplified upon the following micro-text:

*Tato škola je velmi moderní. V budově školy se nachází více než sto učeben.*³

The excerpt comprises 14 tokens; arbitrarily, the size of the window will be set at 10 tokens. It means that there are, in total, five windows (including words 1–10, words 2–11, words 3–12, words 4–13, and words 5–14). The count reads –

$$MATTR(10) = \frac{9 + 9 + 10 + 10 + 10}{10 * (14 - 10 + 1)} = 0.96.$$

The overall MATTR-calculated vocabulary richness of the excerpt is 0.96.

Second, the texts were investigated on the basis of the average tokens length (ATL). This indicator is supposed to provide some information on the complexity of the vocabulary the producer of a text tends to use. Its formula is –

$$ATL = \frac{1}{N} \sum_{i=1}^N p_i;$$

p_i stands for the number of graphemes in a word i , and N for the total of the words in the text. It is to be noted that such a count may be problematic in languages with no script, or in those where the written form does not correspond much with the spoken one (e.g., English and French).

For the sake of an example, let us have a Czech sentence

*Moje škola budoucnosti by měla být veliká.*⁴

³ “This school is very modern. In the school building, there are more than one hundred classrooms.”

⁴ “My school of the future should be big.”

The ATL value of its words is calculated as –

$$ATL = \frac{4 + 5 + 11 + 2 + 4 + 3 + 6}{7} = 5.$$

This means that the average length of the tokens in the example is 5 graphemes.

Next, we will focus on measuring to what extent a text is story-oriented, or, inversely, description-based. To this end, we will make use of an index originally devised by Busemann (1925), and later on employed in stylometry (cf. Andreev, Místecký, Altmann 2018). Activity (Q), as it is called, is a ratio of the number of the verbs in the text (V) and the total of the adjectives (A) and the verbs in it; formally –

$$Q = \frac{V}{A + V}.$$

The count will be exemplified upon a sample task (p_M_1). In it, there are 13 verbs and 8 adjectives; the calculation thus yields –

$$Q = \frac{13}{13 + 8} = 0.6190.$$

If $Q > 0.5$, the text may be considered active; if $Q < 0.5$, it is taken as descriptive. In our case, the text is active.

The last index to be counted is verb distances (VD). It is a simple indicator of syntactic complexity of a text, which takes into account the number of words to be found in between two verbs. Mathematically –

$$VD = \frac{1}{D} \sum_{i=1}^D d_i,$$

D signifying the number of the distances between the verbs, and d_i the number of the words between the verbs.

The count will be presented upon the sample sentence

*Vchod vás nejdříve skenuje a poté se rozjede pás, který vás doveze do šatny.*⁵

Here, there are three verbs (“skenuje”, “rozjede”, and “doveze”), which accounts for two distances; the formula thus calculates –

$$VD = \frac{3 + 3}{2} = 3.$$

The average verb distance in the given sentence is 3 words.

To conclude, in order to be able to compare the results of text groups, we will employ the statistical u-test, which is a traditional tool in quantitative linguistics (cf. Kubát 2016). Its formula reads –

⁵ “First, the entrance will scan you, and then, a line will start moving, taking you to your locker room”.

$$u = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}};$$

X_1, X_2 stand for the average values of the two sets of values, s_1, s_2 for the standard deviations, and n_1, n_2 for the numbers of texts in the two groups. The standard deviations are counted on the basis of the given values. If $u > 1.96$, the difference between the two sets is considered statistically significant.

3. Material

The study material comprises 60 texts in total; 30 texts were written by pupils at elementary school, another 30 by the grammar-school ones. The age of the participants was 12–13, which corresponds to the “sixth grade” of the elementary school (and the first grade of the Czech upper elementary education), and the first grade (called “prima”) of the eighth-year grammar secondary school. The pupils at the latter have passed a written entrance exam. The gender was represented equally, each group consisting of 15 boys and 15 girls. The genre of the written task was description; the elementary-school pupils were supposed to describe their bedrooms, whereas the grammar-school ones focused on their class of the future, or on the school of the future.

Before the very task, both groups were given instructions on how to write the description; they were confronted with various examples, and were asked to produce sample writings on their own. The attention was paid to the choice of suitable parts-of-speech, with adjectives (including those used in recommended comparisons) prevailing over the number of verbs, and to keeping the unified structure of the description (the left-right, centre-periphery, or bottom-up directions, etc.); the pupils were told not to repeat words, especially verbs “be” and “have”. As to the personal intake of the teacher, some elementary-school boys followed the option to describe their bedrooms from unusual viewpoints (e.g., the cell phone, the second person plural, the things in the bedroom, etc.).

Each of the 60 texts is tagged according to whether it was written by an elementary-school pupil (“e”), or a grammar-school one (“g”); next, the gender is indicated, “M” standing for the males, “F” for the females; and finally, it is allocated a number, the boys occupying the upper first half (1–15), and the girls the lower one (16–30). Examples will be presented in the forthcoming section.

As to the interpretations, we have made use, besides our own pedagogical experience, of the insights provided by PhDr. Věra Podhorná, psychologist and the head of the Advisory Centre of Pedagogy and Psychology at Karviná, Czechia.

4. Results

The results will be commented upon in various constellations. First, the exhaustive list of all the outcomes will be presented (see Tables 1a–b). Next, three sets of text groups will be contrasted to each other (schools, genders, and the combination of the two). To conclude, the

indexes will be investigated as to their capacity to provide statistically significant outcomes in comparing the studied sets. All the figures are rounded to the nearest hundredth.

Tables 1a–b
The results of the counts for all the samples

	MATTR	ATL	Q	VD
e M 1	0.90	4.18	0.62	4.75
e M 2	0.82	4.06	0.50	5.72
e M 3	0.85	4.12	0.47	4.96
e M 4	0.87	4.22	0.69	4.93
e M 5	0.90	4.18	0.85	4.31
e M 6	0.81	4.30	0.72	4.97
e M 7	0.89	4.14	0.52	4.39
e M 8	0.85	4.31	0.59	4.83
e M 9	0.93	4.10	0.64	3.89
e M 10	0.91	4.31	0.76	5.00
e M 11	0.88	4.34	0.79	5.86
e M 12	0.91	4.70	0.52	4.50
e M 13	0.90	4.42	0.56	5.56
e M 14	0.80	4.00	0.67	4.06
e M 15	0.85	4.43	0.63	6.80
e F 16	0.87	4.25	0.22	5.55
e F 17	0.85	4.23	0.45	5.41
e F 18	0.88	4.10	0.53	5.84
e F 19	0.87	4.34	0.51	5.27
e F 20	0.88	3.94	0.64	5.24
e F 21	0.88	4.34	0.42	4.82
e F 22	0.87	3.87	0.67	5.41
e F 23	0.89	4.20	0.61	4.59
e F 24	0.86	3.98	0.53	5.14
e F 25	0.86	3.97	0.72	5.20
e F 26	0.87	4.06	0.74	5.00
e F 27	0.87	4.00	0.44	5.18
e F 28	0.86	4.00	0.71	4.17
e F 29	0.90	4.48	0.34	8.56
e F 30	0.87	4.32	0.44	5.33

	MATTR	ATL	Q	VD
g M 1	0.86	4.70	0.39	5.72
g M 2	0.87	4.95	0.41	6.25
g M 3	0.93	5.11	0.38	6.43
g M 4	0.85	4.62	0.50	3.89
g M 5	0.94	5.03	0.54	5.37
g M 6	0.90	4.75	0.51	5.64

*School and Gender in Numbers:
A Stylometric Insight into the Lexis of Teenagers' Description Essays*

g M 7	0.94	4.76	0.39	6.51
g M 8	0.88	5.01	0.45	6.20
g M 9	0.92	4.79	0.44	5.70
g M 10	0.90	4.89	0.38	6.73
g M 11	0.91	4.62	0.38	6.95
g M 12	0.86	4.49	0.29	6.17
g M 13	0.94	5.10	0.58	4.77
g M 14	0.87	4.67	0.32	3.70
g M 15	0.93	4.86	0.50	4.40
g F 16	0.95	4.92	0.45	5.82
g F 17	0.94	4.78	0.27	8.67
g F 18	0.91	4.62	0.46	6.21
g F 19	0.93	5.55	0.35	6.25
g F 20	0.87	4.78	0.32	8.89
g F 21	0.91	4.76	0.40	5.41
g F 22	0.88	4.98	0.30	7.19
g F 23	0.93	5.08	0.53	7.61
g F 24	0.94	4.93	0.59	5.42
g F 25	0.90	4.72	0.41	6.60
g F 26	0.92	4.68	0.50	5.95
g F 27	0.93	4.94	0.42	6.97
g F 28	0.91	5.05	0.46	7.25
g F 29	0.91	4.74	0.43	5.38
g F 30	0.90	4.35	0.53	3.94

4.1 Elementary School vs. Grammar School

In this part, we are going to confront the outcomes of the pupils on the basis of the type of school they attend. The results are presented in Table 2; here, the averages of the elementary-school and grammar-school values are listed, altogether with the u-test calculations. The statistically significant results are marked with asterisks.

Table 2
The results of the school confrontation

	Elementary School	Grammar School	u-test
MATTR	0.87	0.90	5.02*
ATL	4.20	4.84	12.20*
Q	0.58	0.43	5.24*
VD	5.17	6.07	3.28*

It is visible that the schools significantly differ in all the indicators. In activity, the grammar-school pupils manifest, on average, descriptiveness ($Q < 0.5$), which is in line with the requirements of the genre they have produced; on the other hand, the elementary-school participants tend towards activity, even though the average is not very high above 0.5. It is

thus to be inferred that the grammar-school attendants respect the pre-defined norms of the genre, whilst elementary-school children may not go by them so strictly.

It is probable that the high level of descriptiveness is linked to the difference in verb distances, too. The multitude of adjectives prolongs the sentences, raising the complexity of their structures; this is also supported by the use of comparisons and examples. It is not be forgotten, though, that the grammar-school pupils wrote about the class/school of the future; this topic necessitates more explanations, which may have complicated the used phrasing even further.

To conclude, there is a significant difference in the average length of tokens, elementary-school children using, on average, shorter words than their grammar-school peers. This is also connected to the same situation in the sphere of the MATTR-measured vocabulary richness. Both indicate that grammar-school people may be more sophisticated in the use of lexis than the elementary schoolers. It is an open question whether the factor behind these results is the intellectual capacity of the children, or whether they are attributable to the motivated family environments. Another possible explanation stems from, once again, the difference in the topics – the sci-fi-like description of the school/classroom of the future may require the use of a lot of loanwords (names of appliances, specialized vocabulary, etc.), these being longer than the basic words of the Slavic origin.

4.2 Boys vs. Girls

Regardless of the schools, the gender will be investigated in this subchapter. The general results are listed in Table 3.

Table 3
The results of the gender confrontation

	Boys	Girls	u-test
MATTR	0.88	0.89	1.10
ATL	4.54	4.50	0.41
Q	0.53	0.48	1.53
VD	5.30	5.94	2.27*

The results confirm the idea that gender itself is not a sufficient discriminant of style; only one difference – the one in verb distances – is statistically significant. We may presuppose variation in the stylistic manners of boys and girls at different schools; the average values are thus not very indicative of general trends. For instance, if the standard deviations of the activity values are counted, we arrive at the interval of 0.39–0.67 for the boys, and 0.35–0.61 for the girls; it means that the two genders manifest, on average, the same figures. As to verb distances, however, girls tend to be more syntactically complex, which may hint at the fact that they may respect the standards of the genre more than the boys, since the principles of the description writing (use of adjectives, exemplifications, comparisons, a certain amount of precision, etc.) favour employment of complicated phrase structures. Nonetheless, a detailed viewpoint may shed more light on this interpretation.

4.3 Combining Factors

In the present research, the corpus will be divided into four groups, with respect to both gender and school. The elementary-school boys (“e_M”), the elementary-school girls (“e_F”), the grammar-school boys (“g_M”), and the grammar-school girls (“g_F”) will be treated separately. For each index, the counted values of the u-test will be summed up, presented, and visualised in the forms of scatter plots. This is a well-established procedure in statistics-driven research (cf. Kubát, 2016).

First, the vocabulary richness results will be commented upon.

Table 4

MATTR: the values of u-test of the studied text groups

MATTR	e M	e F	g M
g F	5.74*	10.00*	2.40*
g M	3.24*	4.44*	
e F	0.31		

Table 5

The average MATTRs and the sums of the u-test values for the studied text groups

	MATTR	u-test Sum
e M	0.87	9.30
e F	0.87	14.75
g M	0.90	10.09
g F	0.92	18.15

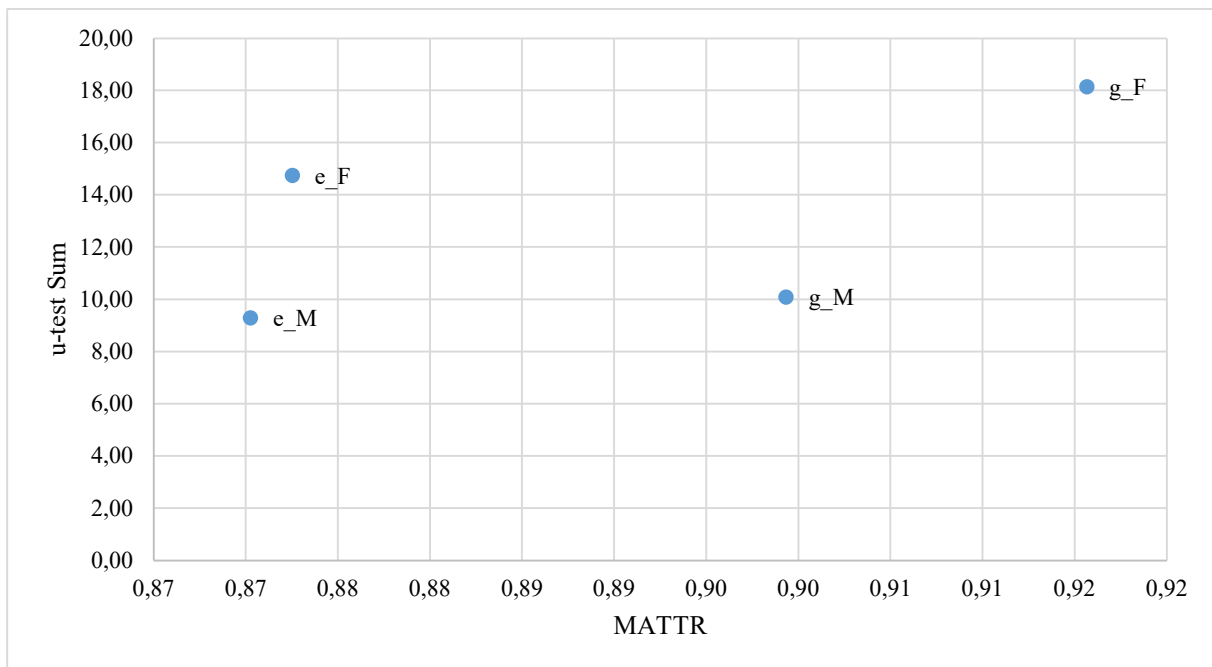


Figure 1. The scatter plot of the MATTR values for the studied text groups.

The values of the MATTR seem to respect the school division, providing one aspect as a premium – the significant difference between the grammar-school boys and girls. The values of the g_F group seem to be high very consistently, as the increased value of the u-test also means a low figure of the standard deviation (see the corresponding formula). The wide vocabulary range of the secondary-school girls is probably due to the aforementioned respect of theirs to the genre requirements. The use of adjectives and the emphasis the Czech stylistic tradition puts on lexical variety may have been the determining factors behind these figures. It is to be noted that the elementary-school girls do also manifest a slightly higher (and more consistent) score than the one of the boys, without opening such a gap as the grammar-school goers do.

Second, the outcomes for ATL will be interpreted.

Table 6

ATL: the values of u-test of the studied text groups

ATL	e_M	e_F	g_M
g_F	7.56*	8.99*	0.44
g_M	8.74*	10.46*	
e_F	1.84		

Table 7

The average ATLs and the sums of the u-test values for the studied text groups

	ATL	u-test Sum
e_M	4.25	18.14
e_F	4.14	21.29
g_M	4.82	19.64
g_F	4.86	17.00

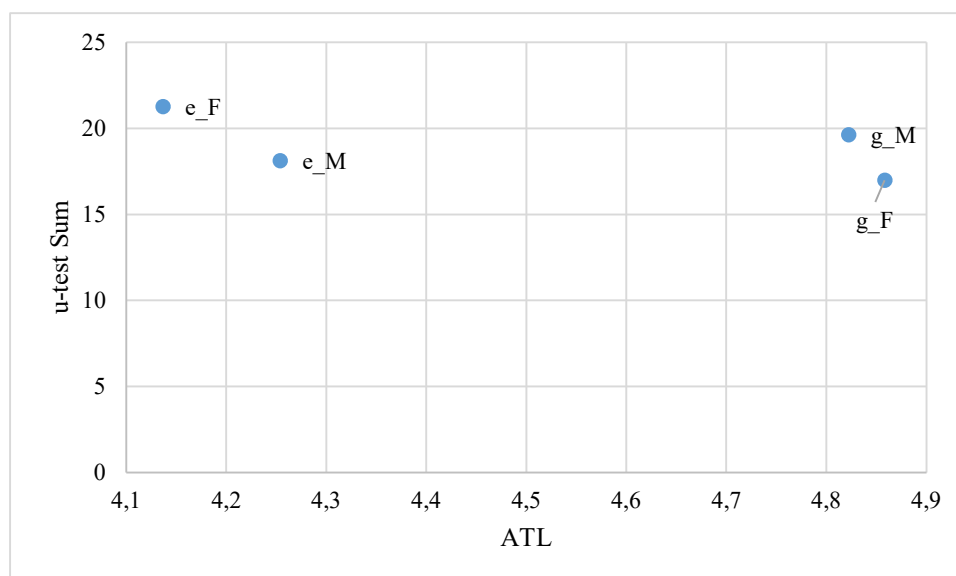


Figure 2. The scatter plot of the ATL values for the studied text groups.

*School and Gender in Numbers:
A Stylometric Insight into the Lexis of Teenagers' Description Essays*

In a way, the ATL situation is the easiest to grasp interpretatively. There are two distinct groups – the elementary-school participants, who manifest lower ATL values, and the grammar-school pupils, who boast higher numbers. The clear-cut situation points at the type of school as the decisive factor for the use of vocabulary; moreover, a textbook may also play part, as both the institutions use different ones. This is corroborated by the fact that the differences in the values of the same-school gender groups are not significant when compared to each other (e.g., “e_M” to “e_F”; see Table 6).

Third, an interpretation of the activity figures will be provided.

Table 8

Q: the values of u-test of the studied text groups

Q	e M	e F	g M
g F	5.76*	2.38*	0.10
g M	5.91*	2.37*	
e F	2.22*		

Table 9

The average Qs and the sums of the u-test values for the studied text groups

	Q	u-test Sum
e M	0.64	13.89
e F	0.53	6.97
g M	0.43	8.38
g F	0.43	8.24

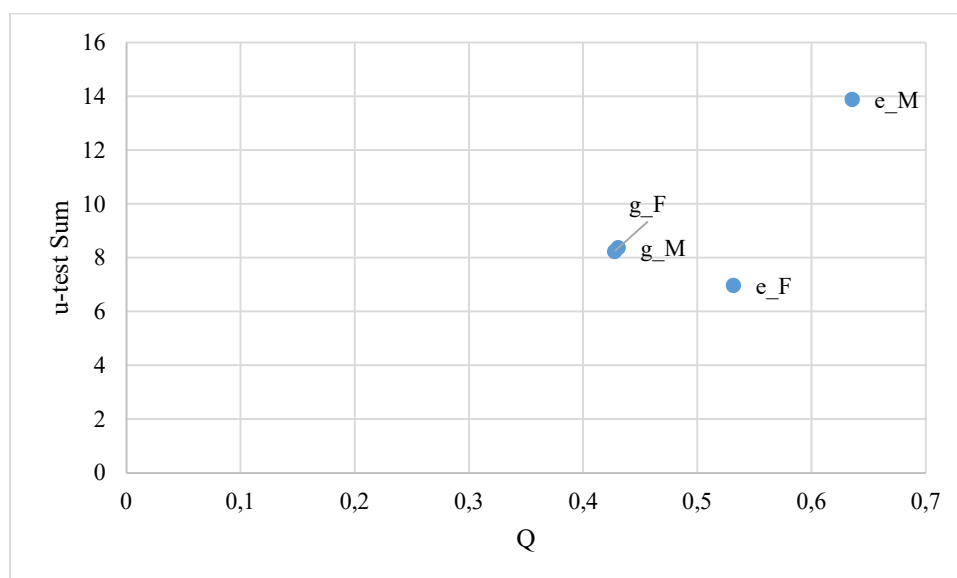


Figure 5. The scatter plot of the Q values for the studied text groups.

The activity values have produced an idiosyncratic picture. The values of the grammar-school people almost overlap; there is some variation to their data (the u-test sums being moderate), but in general, they represent the descriptive part of the corpus. The situation in the elementary-school results is more complicated: the girls seem to show uneven activity figures,

whereas in the boys' essays, verbs prevail – considerably and steadily – over adjectives. This is a crucial outcome, as it may hint at the fact that the elementary-school males do not follow the principles of writing descriptions very much, as they prefer telling a “story of description” over the description itself. The reasons behind this outcome (e.g., lack of abstract thinking, a sort of disrespect to rules, etc.) will be more likely uncovered as soon as more research has been done in the field. Be that as it may, their descriptions present the outlier of the activity research.

Fourth and last, we will have a look at the verb distances values.

Table 10

VD: the values of u-test of the studied text groups

VD	e M	e F	g M
g F	4.07*	2.77*	2.12*
g M	2.08*	0.71	
e F	1.34		

Table 11

The average VDs and the sums of the u-test values for the studied text groups

	VD	u-test Sum
e M	4.97	7.49
e F	5.38	4.82
g M	5.63	4.91
g F	6.50	8.95

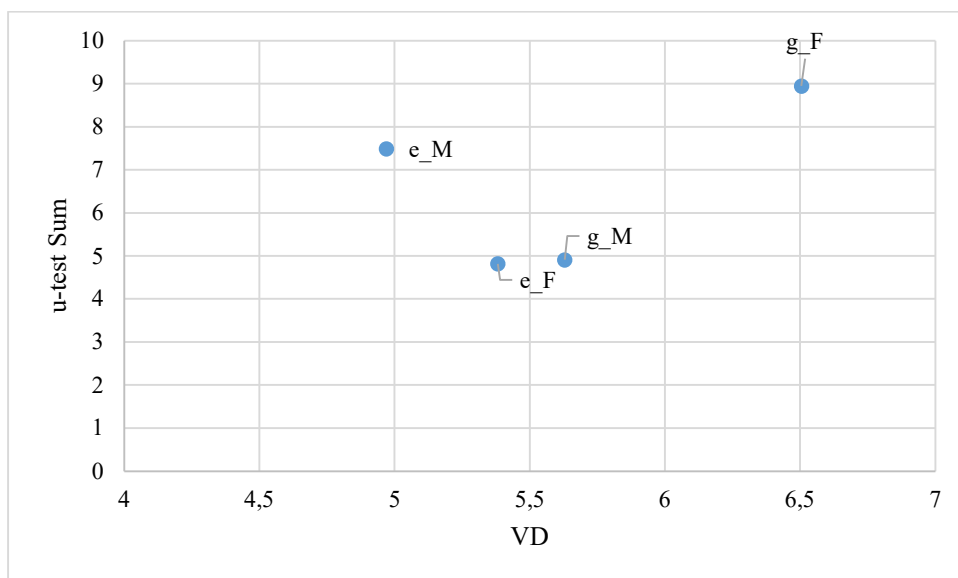


Figure 6. The scatter plot of the VD values for the studied text groups.

The results of VD do not seem to follow either of the factors; the samples are grouped haphazardly, in a constellation that has not been spotted in the preceding parts of the research. There are two outliers: the elementary-school boys, who use shorter verb distances, and the grammar-school girls, who (very consistently) boast complex syntactic structures. The results

of the girls were awaited, as VD was the only index in which the boys and the girls significantly differed on a general basis. A more amount of variation is present in the elementary-school girls and the grammar-school boys, who, on average, manifest moderate numbers. The outcomes show that the biggest difference is between the syntax of e_M and g_F, or that there may be other factors than school and gender that determine the level of sentence complexity.

4.4 Research on Statistical Significance

Throughout the research, we have tested the statistical significance of the obtained results; however, these counts can also be regarded from the viewpoint of the individual indexes. Each of them has been used to calculate eight tests – one in the cross-school comparison, one in the cross-gender one, and six in the combined investigation. Table 12 sums up the proportion of the statistically significant tests to the totals of them for each index; it seems that they can be divided into two groups, according to the decreasing values.

As to our research, the texts appear to differ mostly on the grounds of activity and vocabulary richness; especially the former is surprising, as the description is a rather prescriptive genre when it comes to the use of adjectives and verbs. The differences indicate varied approaches to the principles of writing, and may be connected to the psychology of gender as well.

Table 12
The indexes from the viewpoint of statistical significance

	Number of significant differences	Proportion
Q	6	75.00%
MATTR	6	75.00%
ATL	5	62.50%
VD	5	62.50%

To conclude, the statistical significance will be investigated from the perspective of the studied groups. Each category (e.g., the elementary-school boys) has undergone three comparisons with the others per index; there being four indexes, this accounts for 12 comparisons in total. The upcoming table shows the percentage of those of these tests which yielded significant values.

Table 13
The categories from the viewpoint of statistical significance

	Number of significant differences	Proportion
e_M	9	75%
e_F	8	66.67%
g_M	9	75%
g_F	10	83.33%

The table ranks the text groups according to the amount of their differences from each other; in general, this is almost the same (9 ± 0.7), with slightly elevated numbers in case of the grammar-school children (they show 9.5 significant differences on average, compared to 8.5

differences of the elementary-school pupils). This can be explained by the special situation of the grammar-school girls, who occupy outlying positions in MATTR and VD. Their style of writing can thus be deemed the most discernible.

5. Conclusions

The summarizing remarks will be presented in the following points.

1. It has been found out that there are considerable differences between the researched schools. The statistical significance appeared in the values of the indexes of MATTR, ATL, Q, and VD. We may thus conclude that in general, the grammar-school children have more complex vocabulary and syntax, use longer words, and tend to be more descriptive.

2. As to the gender comparison, the girls have been proved to be more syntactically complex than the boys. This may be due to the attention they pay to observing the principles of the genre (exemplifications, use of adjectives, precision, etc.).

3. Concerning the research with the combined factors, each category will be treated separately.

a) The elementary-school boys tend to use shorter words, and less complex vocabulary and syntax; on the other hand, they prefer using verbal over adjectival description. This contradicts the expectations of the genre, and may have various reasons, which will be studied further.

b) The elementary-school girls, too, limit themselves to shorter words, and simple lexis and sentence structure. As to activity, they manifest middle values with a lot of variation.

c) The grammar-school boys, on the other hand, tend to score high in the length of words and vocabulary richness, though in the latter, the figures do variate. They are never outliers and mostly team up with various categories, sharing lower activity values with the grammar-school girls and middle figures in verb distances with the elementary-school girls.

d) The grammar-school girls share the employment of long words with the grammar-school boys, surpassing them, however, on the grounds of vocabulary richness. Given their low values of activity and an outstanding figure in verb distances, they seem to stick to the rules of the genre most firmly. Moreover, the high scores of the u-test sums in MATTR, ATL, and VD show compactness of the measured values.

4. Regarding the effectivity of the indexes, activity and MATTR are of the highest discriminatory value. Furthermore, the grammar-school groups display more statistically significant differences than the elementary-school ones, this being due to the specific style used by the grammar-school girls (see 3d).

Finally, it has to be stated that all the outcomes and their interpretations are but the first attempts to use stylometry in pedagogy; more research would be needed to come up with general results.

References

- Andreev, S., Místecký, M., Altmann, G.** (2018). *Sonnets: Quantitative Inquiries*. Lüdenscheid: RAM-Verlag.
- Busemann, A.** (1925). *Die Sprache der Jugend als Ausdruck der Entwicklungsrhythmik. Sprachstatistische Untersuchungen*. Jena: Verlag von Gustav Fischer.
- Covington, M. A., McFall, J. D.** (2010). Cutting the Gordian Knot: The Moving Average Type-Token Ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94–100.
- Čech, R.** (2016). *Tematická koncentrace textu v češtině*. Praha: ÚFAL.
- Čechová, M., Krčmová, M., Minářová, E.** (2008). *Současná stylistika*. Praha: Lidové noviny.
- Dai, Z., Liu, H.** (2019). Quantitative Analysis of Queen Elizabeth II and American Presidents' Christmas Messages Over 50 Years (1967-2018). *Glottometrics*, 45, 63–88.
- David, J., Davidová Glogarová, J., Radková, L., Šústková, H., Čech, R.** (2014). *Slovo a text v historickém kontextu: perspektivy historickosémantické analýzy jazyka*. Brno: Host.
- Hoffmannová, J., Homoláč, J., Chvalovská, E., Jílková, L., Kaderka, P., Mareš, P., Mrázková, K.** (2016). *Stylistika mluvené a psané češtiny*. Praha: Academia.
- Holubová, P.** (2014). *Komparativní analýza jazykové a komunikační kompetence v psaných komunikátech studentů 4. ročníku SŠ. Hodnocení maturitních slohových prací ve vztahu k jejich jazykové úrovni*. Praha: Univerzita Karlova v Praze.
- Kubát, M.** (2016). *Kvantitativní analýza žánrů*. Ostrava: Ostravská univerzita.
- Místecký, M.** (2018). Belza Chains in Machar's *Letní sonety*. *Glottometrics*, 41, 46–56.
- Popescu, I.-I.** (2007). Text ranking by the weight of highly frequent words. In: Grzybek, P., Köhler, R. (eds.). *Exact methods in the study of language and text*. Berlin / New York: Mouton de Gruyter, 557–567.
- Rysová, K.** (2017). Charakteristika výsledků písemných maturitních prací z českého jazyka. *Český jazyk a literatura*, 68(4), 157–168.
- Štěpáník, S., Holanová, R.** (2017). K jazykovým a stylistickým dovednostem budoucích češtinářů. *Český jazyk a literatura*, 68(5), 230–239.
- Zörnig, P., Místecký, M.** (2018). Quantifying the Importance of Stylometric Indicators: A Principal Component Approach to Czech Sonnets. *Glottometrics*, 43, 11–30.