Distribution of attributes as a feature of individual style*

Sergey Andreev
smol.an@mail.com

Smolensk State University, Smolensk
Russian Federation

Abstract

The distribution of two types of attributes (adjectives and nouns in genitive construction) is studied. Busemann's coefficient reveals different types of relationship of the adjectival and nominal attributes in the texts of 6 Russian female authors. At the same time it was found that power function fits well the data irrespective of the peculiarities of the authors' individual style revealing a general order of the distribution of the attribute.

Keywords: *distribution, Busemann's coefficient, attributes, the power function, individual style.*

1 Introduction

To analyze individual styles quite a big list of characteristics is used to adequately reveal author's speech peculiarities and establish reliable bases to differentiate styles. This list includes a substantial number of speech properties, both formal and semantic [Juola, 2006; Holms, 1994; Rudman, 1998]. One of such properties whose prognostic value in this respect should be tested is the frequency of attributes in the texts of different authors in general and of certain attributive types, in particular [Köhler, Altmann, 2014].

The syntactic position of an attribute (adnominal) has at least one important peculiarity—it is not obligatory in verbal syntactic structure and thus is highly optional, depending on the author's inclinations and literary taste. On the other hand attributives play a highly important role in elaborating topics.

As a result one can suppose that the frequency and the patterns of the distribution of different types of attributes can serve as an important feature of an author's style. In other words this can serve as an explicit feature for comparing and/or discrimination of the styles of different authors.

According to the part of speech of the word used as an attribute different types attributes can be established: adjectives (green leaves), pronouns (my friend, this book, and other types of pronouns), participle (dancing people), infinitive (a wish to win), adverb (a room upstairs) and some others. One of the most frequent and semantically important is the genitive construction which in Russian

is formed by a noun in genitive case, corresponding to the English genitive of-construction (the book of Peter). Such genitive constructions (N) reflect the nominal strategy of description, opposing it to a more standard strategy of the use of adjectives (A).

Here some questions may arise—are the relations between the frequencies of both types (N and A) constant for all the authors and if not-to what extent they differ in different texts? These questions are viewed on the material of the data-base which includes 6 feminine Russian authors (Tokareva, Ulitskaya, Tolstaya, Marinina, Ustinova, Polyakova). The choice was motivated by the following reasons:

- all the authors are of the same gender;
- they are very popular in their genres;
- the genres of their novels are rather different: the first three authors are writers of belle lettres style, the last three ones are detective writers.

2 Methods

For the analysis the samples from 3 books by each author were chosen for the analysis. The samples were of 1000 words length and were taken from the beginning of each book. The list of the novels is given in the appendix. Adjectival and genitive attributes were marked in the samples.

To find out the proportions between these two types of attributes Busemann's coefficient was used [Altmann, 2015]:

$$C = \frac{G}{A+G}\ , \tag{1}$$

where C is Busemann's coefficient, A stands for all the adjectival attributes, G stands for all the genitive attributes.

The coefficient values can vary between 0 (genitive attributes are absent completely) and 1 (no adjectival attributes were registered). High values of C ($C > 0.5$) show that G—constructions play a more important role in description, low values of the coefficient ($C < 0.5$) indicate the predominance of A in the style of the author. To test the results chi-square statistic was used [Andreev, Místecký, Altmann, 2018]:

$$\chi^2 = \frac{(A-G)^2}{A+G}\,. \tag{2}$$

Busemann's coefficient is statistically significant with 1 degree of freedom and $p < 0.05$ if $\chi^2 > 3.4$.

3 Results & Discussion

The results of the analysis are shown in Table 1. In all cases the results proved to be statistically significant. Ranking the values of C in increasing order one can get the following graphical image. As seen from Figure 1 texts form a gently rising curve. It is noteworthy that in many cases texts by the same author are positioned close to one another. Thus three novels by Ustinova (13–15) are placed next to each other and besides are characterized by nearly the same values of the coefficient. Close to one another are T1 and T3 (Tokareva), T5 and T6 (Ulitskaya), T7 and T9 (Tolstaya), T11 and T12 (Marinina). This demonstrates a comparatively similar relations of two types of attributes in the works of the same author.

Table 1. Busemann's coefficient of the proportion of genitive Constructions in all attributives, used for the purpose of description

Text	Busemann's coefficient (C)	Chi-Square
T1	0.286	23.02
T2	0.304	11.67
T3	0.293	9.80
T4	0.305	35.67
T5	0.274	14.30
T6	0.263	54.82
T7	0.240	36.84
T8	0.303	27.33
T9	0.250	31.84
T10	0.229	33.88
T11	0.205	40.56
T12	0.206	42.51
T13	0.185	25.35
T14	0.186	36.36
T15	0.194	20.84
T16	0.209	37.57
T17	0.231	17.31
T18	0.203	45.47

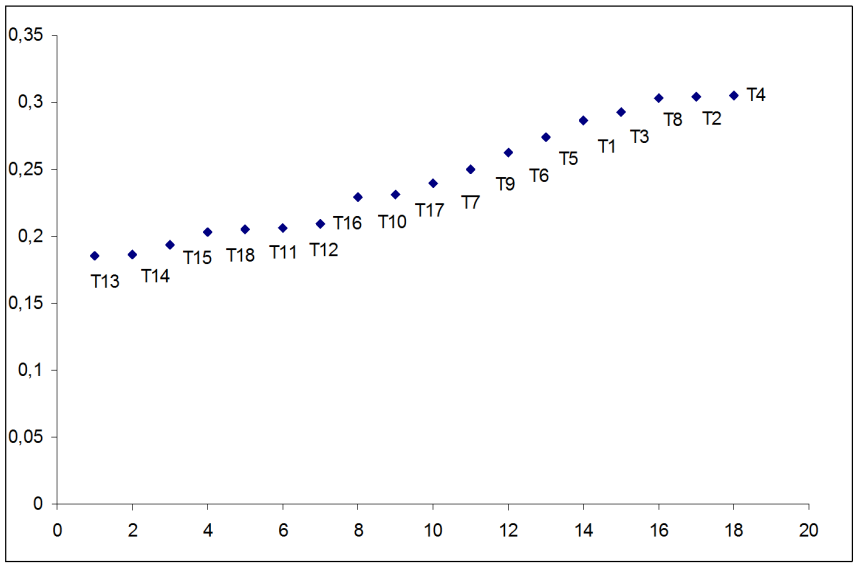

Figure 1: Graphic projection of Busemann's coefficients of genitive and adjectival attributes in 18 novels

It should also be noted that the authors of detectives have somewhat lower values of the coefficient.

The next step was to analyze the relationship of genitives and adjectival attributes in its development from the beginning of the samples to the end. For this purpose the number of all adjectival attributes found before each genitive in the text were counted on a cumulative basis.

As an example, let us consider the development of the relations of these two attribute types over

Table 2: Increase of adjectival attributes against genitives over the text (T1, T16)

Tokareva "Skazat'–ne skazat' " T1			Polyakova "Moye vtoroye ya" T16		
Genetive	Observed frequencies of A	Expected frequencies of A	Genetive	Observed frequencies of A	Expected frequencies of A
1	2	4.27	1	2	1.64
2	7	7.50	2	4	3.96
3	12	10.43	3	8	6.65
4	14	13.17	4	8	9.59
5	17	15.79	5	8	12.75
6	21	18.31	6	14	16.08
7	22	20.76	7	23	19.57
8	25	23.13	8	28	23.21
9	25	25.45	9	28	26.97
10	27	27.73	10	33	30.84
11	30	29.96	11	34	34.83
12	30	32.15	12	39	38.91
13	34	34.31	13	42	43.09
14	36	36.44	14	45	47.36
15	37	38.54	15	46	51.71
16	39	40.62	16	57	56.15
17	40	42.67	17	63	60.66
18	42	44.69	18	68	65.25
19	45	46.70	19	69	69.90
20	51	48.69			
21	51	50.65			
22	58	52.60			

the text in two novels: "Skazat'-ne skazat' " by Tokareva (T1) and "Moye vtoroye ya" by Polyakova (T16). The results of the counts in these two texts are represented in table 2. In the first column the ordinal number of genitive each construction in the text is given. In the second column the number of all adjectival attributes which come in the text before this given genitive construction are summarized. The third column contains theoretically expected (according to the formula) frequencies of adjectival attributes.

The formula is as follows [Naumann et al., 2012]: $y = a * x^b$, where a and b are parameters.

The results are shown in Figures 2 and 3 in graphical form. A shown in the figures the observed frequencies of adjectival attributes (dots) are very near to those theoretically expected, shown as a full line (curve).

The values of the parameters a and b are as follows. For T1 $a = 4.274$, $b = 0.812$; for T16 $a = 1.638$, $b = 1.275$. If $b < 1$ the curve is concave (figure 2), if $b > 1$ the curve is convex (figure 3) [Naumann et al., 2012: 26–27].

Table 3 contains the values of the parameters a and b of the power function and the coefficient of determination $\left(R^2\right)$ for all 18 texts.

R^2 for all the novels is very high which proves good fitting. Parameter b showing the increase or decrease of adjectival attributes towards the end is rater different even in the novels of the same writer. Only in case (Ulitskayay) all the novels of the same author show the same tendency of

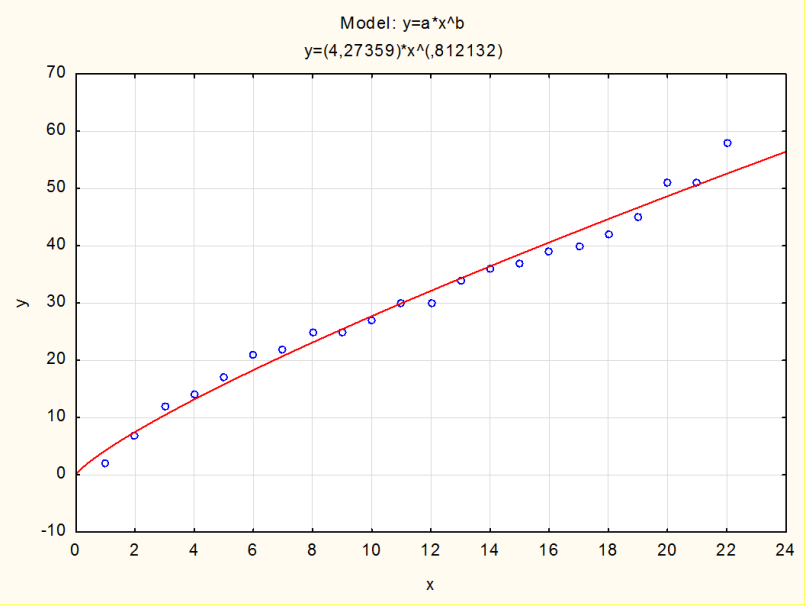

Figure 2: Tokareva T1

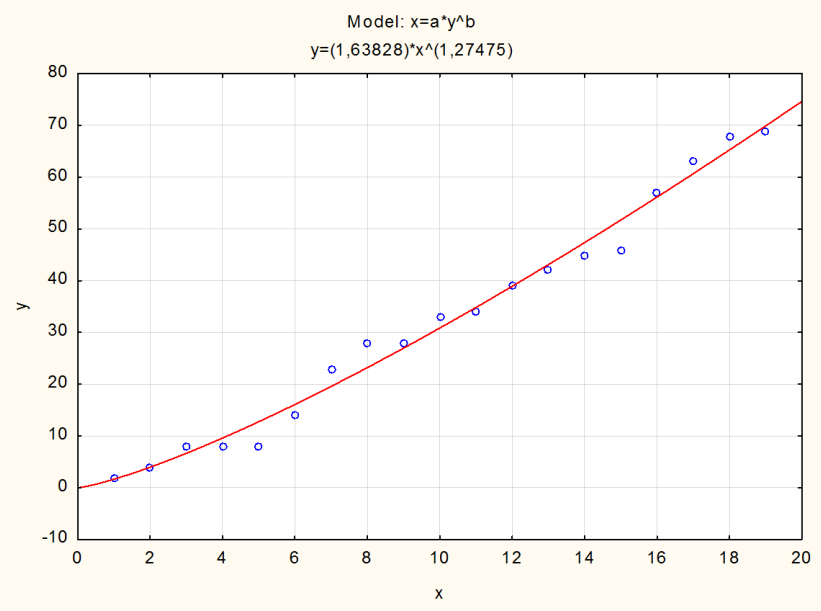

Figure 3: Polyakova T16

gradually decreasing the number of adjectives over the text as in all fer novels (T4—6) $b < 1$. The biggest increase in the number of adjectival attributes and, correspondingly decrease of the genitives is seen in T2 and T3 (Tokareva). Vice versa, the largest increase of the number of genitives takes place in the novel of Ulitskaya (6). Marinina (10—12) demonstrates highly balanced relationship of adjectival attributes and genitives from beginning to the end of her novels.

On the whole the analysis revealed the existence of general tendencies as well as certain differences

Table 3: Parameters of the power function and the coefficient of determination

Text	*a*	*b*	R^2
T1	4.274	0.812	0.9815
T2	0.300	1.725	0.9837
T3	0.970	3.113	0.8922
T4	3.709	0.879	0.9824
T5	4.750	0.824	0.9894
T6	10.556	0.661	0.9878
T7	9.223	0.735	0.9905
T8	4.690	0.825	0.9740
T9	2.056	1.126	0.9308
T10	2.275	1.123	0.9640
T11	4.302	0.968	0.9609
T12	2.373	1.162	0.9350
T13	1.425	1.444	0.9241
T14	3.007	1.148	0.9774
T15	4.032	0.999	0.9662
T16	1.638	1.275	0.9853
T17	5.315	0.868	0.9274
T18	1.860	1.310	0.9551

in style. Different aspects of relations between two main types of attributes in text makes it possible to estimate the role of different kinds of descriptiveness in an author's style and can be used as objective criteria for the differentiation and classification of styles. It should be noted that to get a more complete picture of such relationship of different types of attributes further steps are needed.

References

[Juola, 2006] Juola, P. (2006) Authorship attribution // Foundations and Trends in Information Retrieval. December 2006. Vol. 1. Is. 3. Hanover, MA, USA: Now publishers Inc., 2006. P. 233–334.

[Holms, 1994] Holmes, D.I. (1994) Authorship attribution // Computers and the Humanities. 1994. Vol. 28, No. 2, P. 87–106.

[Rudman, 1998] Rudman, J. (1998) Non-traditional authorship attribution studies in the Historia Augusta: Some caveats // Literary and Linguistic Computing. 1998. Vol. 13, No. 3. 1998. P.151–157.

[Köhler, Altmann 2014] Köhler R., Altmann G. (2014) Problems in Quantitative Linguistics. Lüdenscheid: Ram-Verlag, 2014. – 148 p.

[Altmann, 2015] Altmann G. (2015) Problems in Quantitative Linguistics. 2015. Vol. 5. Lüdenscheid: RAM-Verlag.

[Andreev, Místecký, Altmann, 2018] Andreev S., Místecký M., Altmann G. (2018) Sonnets: Quantitative Inquiries. Studies in Quantative Linguistics, 29. Lüdenscheid: RAM-Verlag, 2018. – 130 p.

[Naumann, et al., 2012] Naumann S., Popescu I.-I., Altmann G. (2012) Aspects of nominal style // Glottometrics. 2012. V. 23. P. 23–55.