Yuri Tambovtsev, Juliana Tambovtseva, and Ludmila Tambovtseva,
Novosibirsk Pedagogical University, Novosibirsk, Russia

A phono-statistical approach to style: A comparison of the sound picture in German and English poets

Abstract. Quantitative methods utilizing the frequency of occurrence of phonemes have been employed to address problems in language typology and historical relatedness. The same methods applied to other linguistic objects have addressed questions of authorship. In this case measures of speech sound frequency are employed to address components of style. The sound picture occupies a position of great importance in poetry, as it is significantly involved in the aesthetic transaction. Thus the sound picture is intimately associated with considerations of an individual poet's style. The frequency of occurrence of phonemic groups serves as the basis to construct linguistic statistical models which allow us to differentiate poets stylistically. This article reports the phonemic distances between some German, Russian, and English poets obtained by computing the speech sound chains they utilize, and contrasting them with the sound picture of Virgil in the Latin.

Introduction

The sound picture is understood as the totality of the occurrence of speech sounds that occur in a text, in this case in some poems. Every poem has a unique sound picture depending on the frequency of occurrence of various phoneme groups in its sound chain.

The goal of this article is to consider the differences and similarities of some German, Russian, and English poems on the phonetic level. Difference is one side of the coin; similarity is the other side of the same coin. They may be defined metrically by the phono-statistical distance. Often a person may feel that the poetry of one writer is similar to some other poet's, but this is a subjective phenomenon. Different people may have different impressions of the same piece of poetry. These impressions are likewise subjective. However,

the distances in the phonetic sound chain which occur in the poems that we calculate should be considered quite objective since they do not depend on anyone's impressions. They exist outside a person's mind and do not depend on the person; some other person may investigate the same poem using these methods and should receive the same results. Therefore, this distance is objective and does not depend on anybody's impression.

It is necessary also to consider the problem of variability of the sound pictures in poems. Unfortunately, many linguists who discuss the problem of variation in language do not give a precise definition to determine how different a text should be in order to be considered a different text. In other words, they do not solve the problem of whether the linguistic object is the same or if the linguistic object became a different object because it accumulated too many changes.

In philosophy this problem is called the transition of quantity into quality. This problem was set up by the ancient Greek philosophers on the example of a ship. They reasoned whether, if we change one board in the ship in question, we should say that it is a different ship. The answer is obvious: no, it is not. What if we change 50% of the boards? Or more than 60%? Where is the threshold where we must consider that the changes sufficiently great to render the object a different object? The great German philosopher G.W.F. Hegel (1770-1831) tried to solve this philosophical problem by formulating one of the basic philosophical laws: the transition of quantity into quality **[Hegel 1978**: 219].

In linguistics it is hard to say if two linguistic objects are the same or if their variation goes too far for them to be considered the same. The objects may concern two phonemes, two words, two texts, or two dialects. Unfortunately, in linguistics this problem cannot be solved by linguistic methods. However, this problem is rather easily solved through the methods of mathematical statistics. We can easily enough detect the difference if we use such methods as the Chi-square criterion, Smirnov-Kolmogorov, or the t-test [**Tambovtsev, 2003**: 5-37]. In

fact, we used these criteria to distinguish between long and short phonemes, the functions of Gerund and Participle-1, and authorship by introducing the distances between the linguistic objects based on these criteria [**Tambovtsev, 2009-d**: 37-48]. In this investigation we used the Chi-square criterion. We shall not discuss it in detail here, since we have discussed the application of the Chi- square criterion in linguistics in general and the distances built on them in particular elsewhere [**Tambovtsev, 2008; 2009-a; 2009-b; 2009-c; 2009-d; 2010].**

The transcribed literary text is to be considered a linguistic object here. Any object is said to possess certain features. We treat the notion of feature as the basis on which the objects (in this case languages) are compared, i.e., in the same way it is done in systemic linguistics [**Melnikov, 2003**: 29]. It is the set of features which allows us to distinguish one object from the other. A scholar must be very careful in choosing the set of features. If a linguist chooses unimportant features, he may receive some strange typology. We agree with V. Vinogradov, who proposed to compare any linguistic objects by the important chosen features [**Vinogradov, 1973**: 230 – 236]. This is why we may regard a language or a poem as a bundle of features with different numerical values to which we can apply various statistical criteria to measure its distances from others.

Many linguists whose point of view is expressed by Yurij S. Stepanov think that on the phonetic level it is easier to detect similarities and differences because it is the only language stratum which has a material shape. He proposes to call these common language phenomena ’postulates’. He discusses the postulate of language identities, equalities, and language differences [**Stepanov, 1975**: 301]. On the one hand, we can speak about an identity of a language to itself at some point in the history of its development. On the other hand, we can speak about the identity of this language to some other genetically related or unrelated languages. We propose to measure this language identity as the value of the distance between languages.

Discussing the application of typological methods at different language levels (phonetic, morphological, and syntax), Boris A. Serebrennikov underlines the importance of the chosen features. He believes that the features comprise the main characteristics of the language type. We agree with his remark that it is possible to build typology on one or several features. According to Serebrennikov even one feature can show typological difference or similarity, for instance, the ratio of the frequency of occurrence of the consonants to vowels in the speech sound chain [**Serebrennikov, 1983**: 289 – 291]. We called this ratio "consonantal quotient" and built the typological classification of Finno-Ugric languages on it, along with some other basic features [**Tambovtsev, 2006**].

Our usual typological approach is to use the phonetic level of a linguistic object to build a typological model on the basis of the values of the chosen phonetic features. In fact, the same typological approach is used by Vladimir D. Arakin, who compared both genetically related and unrelated languages, for instance English and Russian, or the languages from Slavonic, Germanic and Austronesian taxa [**Arakin, 2000**: 5 – 8]. It is important to remark that this typological approach allows us not only to analyse the linguistic objects in some language, but also to compare language matter cross-linguistically. This gives the most interesting results [**Refromatskij, 2005:** 455 — 456].

Typological Features

While analysing languages a linguist should take up universal features, i.e. features which can be found in every world language. And in doing so he must keep to the principle of commensurability. In this case our typological research keeps to the comparison of languages on the level of phonetics where the universal groups of consonants and vowels are defined by the work of the active organs of speech (labial, front, palatal, guttural), the manner of impeding the air stream (sonorant, occlusive, fricative), and the activity of vowel cords (voiced). At the first stage of investigation the vowels are taken as a whole without being

divided into groups. Thus we chose nine features for comparison, though it is possible to take up more features if we divide vowels into some groups.

Labial
Front
Palatal
Guttural
Sonorant
Sonorant
Occlusive
Fricative
Voiced
Vowels

Nevertheless, it was proved that these nine features are sufficient to measure the typological distances. Let us emphasize once again that the features mentioned above are universal because of the construction of the speech producing apparatus of humans. To obtain our samples we transcribed German, English, Latin and Russian poetic texts according to the general rules accepted in these languages [e.g. **Borovskij et al., 1949; Jarho et al., 2005; Kozmin et al., 2004; Suntsova, 1958; Tambovtsev, 1976; 1977].**

Poetry is said to be different from prose because the sound picture of the poetic linguistic object plays a greater role. When hearing poetry we can always state it is not prose. We can also distinguish one poet from the other on the phonetic level, i.e., by its sound picture. So, one can say that in poetry the sound picture is the most important part, since it transacts the aesthetic information. That is, poetry transfers its images not only by the ideas but significantly by the sound picture. Many poets have unique sound pictures. This allows us to distinguish one poet from another. Yuri M. Lotman points out that a certain emotional tenor of a poetic object is embodied in the distinct pattern of the frequency of occurrence of certain groups of phonemes [**Lotman 1996**: 14]. Oleg S. Shirokov calls this the "sound colouring" of a poem [**Shirokov 1985**: 17 – 22]. So, we come to a conclusion that any poetic

creation transfers not only semantic but also aesthetic meaning, and the latter prevails. We should bear in mind that this aesthetic meaning is closely connected with the only material matter in a language, i.e., its sounds.

Often a reader requires the aesthetic content more than the semantic. Estimating the beauty of sounds, the poet subconsciously chooses certain sounds more often than others. The choice of sounds is of course embodied in the choice of words, and the choice of words in poetry is rather different from the choice of words elsewhere. Linguists noticed this while comparing poetry and prose. Some of them employed mathematical methods to verify the difference, [**Fucks 1975**: 282 – 283], for instance. Wilhelm Fucks investigated the length of words in the works of such German poets as Goete and Rilke and found them to be quite different according to their statistical characteristics. In particular, he stated that Goete uses 49% monosyllabic words while Rilke employs more – 63%. The other features also show sharp differences [**Fucks 1975**: 315 – 320]. Fucks also discovered that different Latin authors demonstrate different values in their statistical characteristics [**Fucks 1975**: 287]. W. Fucks explains the difference between German and Latin texts by the differences in the structure of the German and Latin languages, which enter different groups of the Indo-European language family. Comparing the intensity and length of the connectivity of German poetry (Goete, Schiller, Rilke) and those in Russian (Pushkin), English (Shakespear) and Old Greek (Homer), W. Fucks concludes that they are quite different because of the structure of these languages [**Fucks 1975**: 365]. It is important to note that his investigation is both typological and multilingual. One should be aware of the fact that investigations of this sort have a brilliant future and attract attention of many linguists.

We have calculated the frequency of occurrence of speech sounds which depict sound pictures in German, English, Latin, and Russian according to the nine features listed previously.

Table 1. The Frequency of Occurrence of Phonemic Groups in German, English, Russian and Latin Poetry.

№	Phonemic Groups	Heine	Gldrn	Goete	Rilke	Schill	Byron	Moor	Ahmat	Virgili
1.	Labial	10.34	9.35	10.69	9.92	10.42	13.37	13.62	11.47	12.15
2.	Front	42.32	42.51	42.65	43.91	44.11	42.01	42.32	35.52	37.95
3.	Palatal	0.32	0.21	0.21	0.22	0.22	0.68	0.58	5.75	0.41
4.	Guttural	9.26	7.89	9.04	7.88	8.48	6.39	5.79	6.14	6.82
5.	Sonorant	25.87	25.18	25.30	26.69	26.29	22.56	23.22	24.85	22.48
6.	Occlusive	20.86	20.39	20.07	19.75	21.47	20.24	18.26	20.42	23.00
7.	Fricative	15.51	14.39	17.22	15.75	15.47	19.65	20.83	13.61	11.85
8.	Voiced	13.20	12.31	11.15	11.77	12.32	17.65	18.20	11.06	7.4
9.	Vowels	37,76	40.04	37.41	38.07	36.77	37.55	37.69	41.12	4.67

Sample volumes of phoneme classes surveyed are as follows:
German: Goete “Faustus” – 327,694 phonemes; Hoelderlin “Poetry Collection” – 146,558 phonemes; Heinrich Heine. “Deutschland” – 52,630 phonemes; Rilke – 82,443 phonemes; Schiller “Poetry Collection” – 36,055 phonemes;
English: George Gordon Byron “Childe Harold's Pilgrimage” – 31,125 phonemes; Thomas Moore “Poetry Collection” – 31,111 phonemes.
Russian: Anna Ahmatova “Poetry Collection” – 69,112 phonemes.
Latin: Virgilius “Aeneida” – 358,121 phonemes.

Discussion of the Results

Let us begin our phono-typological analysis with considering the distances between the sound pictures in the German poetry. The ordered series of phono-metrical distances are:

1)Heine-Schiller (0.94); 2)Schiller-Rilke (0.95); 3)Hoelderlin-Rilke (1.40); 4)Goete-Rilke (1.64); 5)Goete-Schiller (1.74); 6)Heine-Rilke (1.91); 7)Heine-Hoelderlin (1.97); 8)Goete-Heine (2.05); 9)Schiller-Hoelderlin (2.19); 10)Goete-Hoelderlin (6.45). One can see by this ordered series that the sound picture of Heine is most similar to that of Schiller (0.94). Schiller and Rilke are also very close (0.95). It tells us that their poetry sounds very similar. If we juxtapose the life span of the German poets in question, we can understand that Goete

could have influenced all of them (Goete: 1710-1782). Schiller lived and wrote later (1759-1805), followed by Hoelderlin (1770-1843), Heine (1797-1856), and Rilke (1875-1926).

The distribution of certain phonemic groups which we have defined is quite different in the English poetry. Thus, the phono-metric distance between Goete and Byron is much greater – 18.00. It is several times greater than between the German poets. We can say this is a typological distance. It may be explained by differences in structure of the English and German languages, though both languages belong to the same Indo-European language group, Germanic. The phono-metric distance between Byron and Moore is not so great, 1.26, as both poets are English.

The difference in the sound pictures is more vivid when the languages enter different language groups. For instance, the phono-typological distance between Goete and Virgilius, 21.99, is greater than that between Goete and Byron. It is possible to explain this by the fact that German and Latin occupy different language groups: Germanic against Italic. Now let us compare the sound pictures of Goete and a Russian poet, Anna Ahmatova. The distance is much greater – 45.06. So, Russian shows a greater phono-typological distance since it belongs to the Slavonic group of the Indo-European language family. In Russian, the phono-metrical distances between poets are not so great. The least is between Anna Ahmatova and Aleksander Blok, 0.44. Distances among authors calculated appear in Table 2.

Table 2
Distances between language objects of different authors based on the TMB (Tambovtsev) coefficient

№	Authors	Distance
1.	Ahmatova – Blok	0.44
2.	Heine – Schiller	0.94
3.	Schiller – Rilke	0.95
4.	Byron – Moore	1.26
5.	Hoelderlin – Rilke	1.40
6.	Goete – Rilke	1.64

7.	Goete – Schiller	1.74
8.	Heine – Rilke	1.91
9.	Heine – Hoelderlin	1.97
10.	Goete – Heine	2.05
11.	Schiller – Hoelderlin	2.19
12.	Goete – Hoelderlin	6.45
13.	Goete – Byron	18.00
14.	Goete – Virgilius	21.99
15.	Goete – Ahmatova	45.06

Conclusions:

1. The phono-typological distance between the sound pictures in different languages is usually greater than in one and the same language. This may be a product of the different structures imposed by these languages.

2. The investigation of the five German poets considered shows that the most similar sound pictures are those of Heine and Schiller, with the phono-metric distance of 0.94.

3. Perceptible stylistic differences in the sound picture of the works of these poets can be seen from the following ordered series: 1) Heine-Schiller (0.94); 2) Schiller-Rilke (0.95); 3) Hoelderlin-Rilke (1.40); 4) Goete-Rilke (1.64); 5) Goete-Schiller (1.74); 6) Heine-Rilke (1.91); 7) Heine-Hoelderlin (1.97); 8) Goete-Heine (2.05); 9) Schiller-Hoelderlin (2.19); 10) Goete-Hoelderlin (6.45), which are measurable by statistical methods. Thus, the greatest phono-metric distance is between Goethe and Hoelderlin (6.45), while the distance between Byron and Moore is less, 1.26. This measurable distance may be much less, e.g. between Ahmatova and Blok, 0.44, with a less salient distinction in the sound pictures of their works.

3. Employing statistical methods to the language sound chains in the works of poets as objects of comparison offers a means to quantify the stylistic difference in the sound pictures of the authors being compared.

References

Arakin, 2000 – Arakin Vladimir D. Sravnitel'naja tipologija anglijskogo i russkogo jazykov. [Comparative typology of English and Russian] - Moskva, 2000.

Borovskij et al., 1949 – Borovskij J.M., Boldyrev A.V. Latinskij jazyk. [The Latin language]. - Moskva, 1949.

Fucks, 1968. Nach Allen Regeln der Kunst: Diagnosen ueber Literatur, Musik, bildende Kunst – die Werke, ihre Autoren und Schoepfer. Deutscher Verlags-Anstalt, Stuttgart,.

Fucks, 1975 – Fucks, Wilhelm. Po vsem pravilam iskusstva: Tochnye metody v issledovanijah literatury, muzyki I izobrazitel'nogo iskusstva. (Russian translation). [Wilhelm Fucks. Nach Allen Regeln der Kunst: Diagnosen ueber Literatur, Musik, bildende Kunst – die Werke, ihre Autoren und Schoepfer. Deutscher Verlags-Anstalt, Stuttgart, 1968]. - Isskustvo i EVM. - Moskva: Izdatel'stvo "Mir", 1975, p.281 – 442.

Hegel 1978 – Hegel, Georg Wilhelm Friedrich. Gasammelte Werke. Bd 11. - Hamburg: Felix Meiner Verlag, 1978.

Jarho et al., 2005 – Jarho V.N., Katsman N.L., Lifshits I.A. et al. - Latinskij jazyk. [The Latin Language]. - Moskva, 2005.

Kozmin et al., 2004 – Kozmin O.G., Sulemova G.A. Fonetika nemetskogo jazyka. [Phonetics of the German Language]. - Moskva, 2004.

Lotman, 1996 – Lotman Yuri M. Analiz poeticheskogo teksta. [Analysis of the poetic text]. - Sankt-Peterburg, 1996.

Melnikov, 2003 – Melnikov, Gennadij P. Sistemnaja typologija jazykov: Printsipy, metody, modeli. [Systemic typology of languages: Principles, methods, models]. - Moskva: Nauka, 2003.

Reformatskij 2005 – Reformatskij A.A. Vvedenie v jazykovedenie. [Introduction into Linguistics]. - Moskva, 2005.

Shirokov 1985 – Shirokov Oleg S. Vvedenie v jazykoznanie. [Introduction into Linguistics]. - Moskva: Moscow State University, 1985.

Serebrennikov, 1983 — Serebrennikov Boris A. O materialisticheskom podhode k javlenijam jazyka. [On the Materialistic Approach to Language Phenomena]. - Moskva, 1983.

Stepanov, 1975 – Stepanov Yurij S. Metody i printsipy sovremennoj lingvistiki. [Methods and Principles of Modern Linguistics]. - Moskva, 1975.

Suntsova, 1958 – Suntsova I.P. Vvodnyj kurs fonetiki nemetskogo jazyka.[Introductory course of phonetics of the German language]. - Moskva, 1958.

Tambovtsev, 1976 — Tambovtsev, Yuri. Raspoznavanie fonem chelovekom. [Recognition of phonemes by man]. - In: Empiricheskoe predskazanie i raspoznavanie obrazov. Vychislitel'nye sistemy. Tomus 67. - Novosibirsk: Institute of Mathematics, 1976, p.161 — 164.

Tambovtsev, 1977 — Tambovtsev, Yuri. Nekotorye harakteristiki raspredelenija fonem mansijskogo jazyka. [Some characteristics of distribution of phonemes in the Mansi language]. - In: Soviet Finno-Ugric Studies, XIII , 1977, № 3, p.195-198.

Tambovtsev, 2003 – Tambovtsev, Yuri. Tipologija funktsionirovanija fonem c zvukovoj tsepochke indoevropejskih, paleoaziatskih, uralo-altajskih i drugih jazykov nira: kompaktnost' podgrupp, grupp. semej i drugih jazykovyh taksonov. [Typology of functioning of phonemes in the sound chain of Indo-European, Paleo-Asiatic, Uralo-Altaic and other world languages: compactness of subgroups, groups, familes and other language taxa]. - Novosibirsk: Sibirskij Nezavisimyj Institut, 2003, 143 pages.

Tambovtsev, 2006 - Tambovtsev, Yuri. Degrees of Dispersion of Various Language Taxa as the Index of Their Typological Similarity. - In: Favete Linguis. Studies in Honour of Viktor Krupa. - Bratislava: Institute of Oriental Studies of Slovak Academy of Sciences, 2006, p. 219 – 260.

Tambovtsev, 2008 – Tambovtsev, Yuri. Typological Distance Between Korean and other Languages of Asia. - In: Lingua Posnaniensis, # 50 (L), 2008, p. 177 – 186

Tambovtsev, 2009-a - Tambovtsev, Yuri. Chislennoe modelirovanie tipologicheskoj shozhesti jazykov v nekotoryh jazykovyh taksonah. [Numerical modelling of the typological similarity of languages in some language taxa]. - In: Acta Linguistica, Vol. 2, 2009, p.37 – 48.

Tambovtsev, 2009-b Tambovtsev, Yuri. Variativnost' dolgih i kratkih glasnyh v mansijskom jazyke. [Variation of the long and short vowels in the Mansi language]. - In: Finno-Ugorskij Mir, №1, 2009, p. 22 – 27.

Tambovtsev, 2009-c Tambovtsev, Yuri. Estonskij i vyro: dva otdel'nyh jazyka ili dva dialekta odnogo jazyka? [Estonian and vyro: two separate languages or two dialects of the same language?] - In: Fenno-Ugristica, # 28, 2009, p. 86 – 92.

Tambovtsev, 2009-d – Tambovtsev, Yuri. The Imitation of Style: Frequency of Occurrence of Linguistic Elements. - In: California Linguistic Notes, Vol. XXXIV, #2, Spring, 2009, p. 3 – 11.

Tambovtsev, 2010 – Tambovtsev, Yuri. Fonotipologicheskaja blizost' lingvisticheskih objektov po kriteriju “Khi-kvadrat”. [Phono-typological closeness of linguistic objects by “Chi-square” criterion]. - In: Vestnik Novosibirskogo gosudarstvennogo universiteta. Serija: Informatsionnye Technologii, Volume 8, #3, p.46 – 54.

Vinogradov, 1973 – Vinogradov V.V. Metody tipologii. [Methods of typology]. - Obshchee jazykoznanie: metody lingvisticheskih issledovanij. [General linguistics: methods of linguistic investigations]. - Moskva, 1973, p. 224 – 256.