Open Access Library Journal

About the Proximity of the Languages on the Example of German Languages (Part 2)

Tonko Tonkov

ul. "Taras Shevchenko" № 20, vhod "V", 1113, Sofia, Bulgaria
Email: tonkov@mail.mgu.bg

Received 8 April 2014; revised 18 May 2014; accepted 28 June 2014

Copyright © 2014 by author and OALib.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

Proximity of languages is defined as an application of a special function distance between sets of points. It is discussed the languages of the German group in the text "Universal Declaration of the human Rights", translated from the English original, only as an example.

Keywords

Distance between Languages, Distance between Sets of Points

Subject Areas: **Linguistics, Statistics**

1. Introduction

This paper is a continuation of our paper [1], where we defined the proximity of the languages of the German group on the base of a text, translated from English to the others languages, using a special function distance between sets of points under some conditions, defined firstly in [2]. In this paper we will apply the same

Definition. Let $A_1=\{a_1(i_1)\}_{i_1=1}^{n_1}$ and $A_2=\{a_2(i_2)\}_{i_2=1}^{n_2}$ be two sets of points $a_j(i_j)=(x_j(i_j),y_j(i_j))$, $j=1,2$; let φ_j, $j=1,2$, be two angles of rotation around the points of Fermat (or Fermat-Torricelli) $F_j(F_jx,F_jy)$, $j=1,2$. Then by definition under distance between the sets A_1 and A_2 we understand the number

$$\operatorname{Re}x=\operatorname{Re}x(A_1,A_2)=\frac{1}{n_1n_2}\min_{\varphi_1,\varphi_2}\left|\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sqrt{(X_2(i_2)-X_1(i_1))^2+(Y_2(i_2)-Y_1(i_1))^2}-1\right|$$

where

$$X_j(i_j)=\frac{1}{p_j}\left((x_j(i_j)-F_jx)\cos\varphi_j-(y_j(i_j)-F_jy)\sin\varphi_j\right),\quad j=1,2;$$

How to cite this paper: Tonkov, T. (2014) About the Proximity of the Languages on the Example of German Languages (Part 2). *Open Access Library Journal*, **1**: e722. http://dx.doi.org/10.4236/oalib.1100722

$$Y_j\left(i_j\right)=\frac{1}{p_j}\left(\left(x_j\left(i_j\right)-F_jx\right)\sin\varphi_j-\left(y_j\left(i_j\right)-F_jy\right)\cos\varphi_j\right),\ \ j=1,2;$$

$$p_j=\frac{1}{n_j^2}\sum_{i_1=1}^{n_j}\sum_{i_2=1}^{n_j}\sqrt{\left(x_j\left(i_2\right)-x_j\left(i_1\right)\right)^2+\left(y_j\left(i_2\right)-y_j\left(i_1\right)\right)^2},\ \ j=1,2.$$

We will apply the computer system Mathematica, in which the formula for Re*x* looks like:

(1) $\mathrm{Re}\,x\left(A_1,A_2\right)\rightarrow$

```
NMinimize[Abs[(1/(n1*n2)  {Sum[Sqrt[(a2x[j]-a1x[i])^2+(a2y[j]-a1y[i])^2]/.
{a1x[i] → 1/p1((x1[i] -F1x)Cos[fi] - (y1[i]-F1y) Sin[fi]), a1y[i]→ 1/p1((x1[i]
-F1x) Sin[fi] + (y1[i]-F1y) Cos[fi]),a2x[j]→ 1/p2((x2[j]-F2x) Cos[psi]-(y2[j]
-F2y)Sin[psi]), a2y[j]→1/p2((x2[j]-F2x) Sin[psi] + (y2[j] -F2y) Cos[psi])},
{i, n1},{j, n2}]/.{<...>}}-1], {fi, psi}]
```

where <…> is the list of the coordinates of all points.

Coordinates of the points of Fermat F[Fx,Fy] we find by the formula:

(2) Point of Fermat>F[Fx,Fy] NMinimize[(1/n)Sum[Sqrt[(Fx-x[i])^2+(Fy-y[i])^2],{i, n}]/.{<…> }, {Fx,Fy}]] and for the normalizing divisor *p* – by the formula:

(3) `Divisor>p→(1/(n*n))Sum[Sqrt[(ax[j]-ax[i])^2+(ay[j]-ay[i])^2] /.{ax[i]→ x[i], ay[i]→y[i],ax[j]→x[j], ay[j]→y[j]}/.{< … >},{i,n}, {j,n}]//N,`

where <…> contains the list of coordinates of the *n* points of the set with this point of Fermat (Point of Fermat for given set of points is such a point, which sum of distances to all points of the set is minimal).

The object of our research here is the same text "Universal Declaration of the human Rights", proclaimed by the United Nations General Assembly in Paris on 10 December 1948 and translated at present in 437 translations and can be seen in Internet [3] on Official UN **Universal Declaration of Human Rights** Home Page.

In [3] we studied the translated texts on this Home Page in Slavic languages.

Now we continue the study the corresponding translations in Germanic languages of the two groups: West German Languages with the following 9 representatives (in alphabetical order): Afrikaans, Deutsch (German), English, Frysk (Frisian), Nedelands (Dutch), Niederdeutsche, Pijin, Scots, Yiddish, and North German Languages with 6 translations in Dansk (Danish), Islenska (Icelandic), Føroyskt (Faroese), Norsk (Norwegian) in two forms: Bokmål and Nynorsk, and Svenska (Swedish).

2. Statistics of the Letters in Different Translations of "Universal Declaration of the Human Rights" and the Distances between Languages

2.1. West German Languages

The text in Deutsch "**Allgemeine Erklärung der Menschenrechte**" contains 29 letters (including ß) with total absolute frequency 10 025, distributed in increasing range as it follows in numbers and figures:

{1-x-1, 2-y-2, 3 -ß -31, 4- ö -34, 5-j-36, 6-p-40, 7- ü -52, 8- ä-56, 9-v-84, 10-w-109, 11-z-117, 12-k-126, 13-b-130, 14-m-163, 15-o-173, 16-f-212, 17-c-323, 18-g-378, 19-l-413, 20-u-461, 21-s-492, 22-h-504, 23-a-530, 24-d-574, 25-t-606, 26-i-765, 27-r-790, 28-n-1069, 29-e-1754}

The letter **x** has the smallest frequency 1, the letter **e**—the greatest frequency of 1754.

Arranged in decreasing order (ranked), the same data forms the following sequence:

De{1754,1069,790,765,606,574,530,504,492,461,413,378,323,212,173,163,130,126,117,109,84,56,52,40,36, 34,31,2,1}

Applying the Formulas (2) and (3), we get Point of Fermat {FDex→15., FDey→173.}} and divisor pDe→378.83.

By similar way we get:

In English text "Universal Declaration of the human Rights":

En{1077,803,714,706,705,699,609,465,446,398,324,292,224,196,189,166,165,152,109,99,75,18,16,15,7,4}

FEnx→14.9011, FEny→210.188; pEn→325.326.

In Frysk (Frisian) text "**Universele ferklearring fan de rjochten fan de minske**":

Fry{1461,1105,736,651,639,632,528,454,390,387,348,309,288,236,228,224,159,154,145,136,131,104,71,61, 36,33,29,21,18,10,5,2,2,1,1} with Ffryx→18., Ffryy→154; pfry→327.098.

In Afrikaans text **“UNIVERSELE VERKLARING VAN MENSEREGTE”**:

Afr{1694,776,753,607,568,491,446,442,414,378,373,301,246,181,143,135,113,112,98,77,72,6,2} with FAfrx →12., FAfry→301; pAfr→351.508.

In Nedelands (Dutch) text **“UNIVERSELE VERKLARING VAN DE RECHTEN VAN DE MENS”**:

Ne{2144,1187,764,684,681,645,622,545,390,388,352,324,306,205,178,176,166,152,147,142,100,96,1,1} with FNex→13., FNey→306; pNe→417.932.

In Niederdeutsche (Low German) text **“Allgemeen Verklaren vun de Minschenrechten”**:

UD{1652,1243,903,722,709,561,455,455,454,453,378,285,242,238,214,174,172,168,151,129,120,109,99,16, 11,7,3,2} with FUDx→15.5077, FUDy→221.865; pUD→383.368

In Pijin text **“Universol Declarason lo Hiuman Raits”**:

Pi{961,878,709,673,513,495,486,447,394,351,314,184,166,154,153,149,134,123,122,93,67,15,3,1} with FPix→14.2154, FPiy→174.743; pPi→295.649.

In Scots text **“Universal Declaration o Human Richts Inning”**:

Sc{1002,801,779,777,689,549,533,467,429,364,319,304,196,191,136,134,128,127,111,105,56,41,15,11,6} wit FScx→13., FScy→196; pSc→318.109.

In Yiddish text **“Universal deklaratsie fun mentshnrekht hagdome”**:

Yi{895,821,658,621,546,520,465,398,387,362,338,224,222,210,196,184,166,163,161,155,132,119,116,116,1 14,72,56,52,26,25,15,12,12,1} with FYix→18.2901, FYiy→163.804; pYi→249.707

Now we are ready to apply Formula (1) for Deutsch and Frysk (for instance), where X-coordinate means the place (the rank) in the sequence of frequencies; Y-coordinate is the value of the frequency:

```
Rex[Deifry]→ NMinimize[Abs[(1/(29*35)) {Sum[Sqrt[(a2x[j]-a1x[i])^2+
(a2y[j]-a1y[i])^2]/.{a1x[i]→ 1/pDe ((xDe[i] -FDex) Cos[fi] - (yDe[i]-FDey)
Sin[fi]), a1y[i]→ 1/pDe ((xDe[i] -FDex) Sin[fi] + (yDe[i]-FDey) Cos[fi]), a2x[j]
→ 1/pfry ((xfry[j]-Ffryx) Cos[psi]-(yfry[j] -Ffryy) Sin[psi]), a2y[j]→1/pfry
((xfry[j]-Ffryx) Sin[psi] + (yfry[j] -Ffryy) Cos[psi])}, {i,29}, {j,35}]/.
{xDe[1]→1,  yDe[1]→1754,xDe[2]→2,yDe[2]→1069,xDe[3]→3,  yDe[3]→790,  xDe
[4]→4,yDe[4]→765,xDe[5]→5,yDe[5]→606,xDe[6]→6,yDe[6]→574,xDe[7]→7,yDe[
7]→530,xDe[8]→8,yDe[8]→504,xDe[9]→9,yDe[9]→492,xDe[10]→10,yDe[10]→461,
xDe[11]→11,yDe[11]→413,xDe[12]→12,yDe[12]→378,xDe[13]→13,yDe[13]→323,xD
e[14]→14,yDe[14]→212,xDe[15]→15,yDe[15]→173,xDe[16]→16,yDe[16]→163,xDe[
17]→17,yDe[17]→130,xDe[18]→18,yDe[18]→126,xDe[19]→19,yDe[19]→117,xDe[20
]→20,yDe[20]→109,xDe[21]→21,yDe[21]→84,xDe[22]→22,yDe[22]→56,xDe[23]→2
3,yDe[23]→52,xDe[24]→24,yDe[24]→40,xDe[25]→25,yDe[25]→36,xDe[26]→26,yDe
[26]→34,xDe[27]→27,yDe[27]→31,xDe[28]→28,yDe[28]→2,xDe[29]→29,yDe[29]→
1,FDex→15.000000000000158`,FDey→173.00000000000128`,pDe→378.8302790211779
4`,xfry[1]→1,yfry[1]→1461,xfry[2]→2,yfry[2]→1105,xfry[3]→3,yfry[3]→736,
xfry[4]→4,yfry[4]→651,xfry[5]→5,yfry[5]→639,xfry[6]→6,yfry[6]→632,xfry[
7]→7,yfry[7]→528,xfry[8]→8,yfry[8]→454,xfry[9]→9,yfry[9]→390,xfry[10]→
10,yfry[10]→387,xfry[11]→11,yfry[11]→348,xfry[12]→12,yfry[12]→309,xfry[1
3]→13,yfry[13]→288,xfry[14]→14,yfry[14]→236,xfry[15]→15,yfry[15]→228,xf
ry[16]→16,yfry[16]→224,xfry[17]→17,yfry[17]→159,xfry[18]→18,yfry[18]→15
4,xfry[19]→19,yfry[19]→145,xfry[20]→20,yfry[20]→136,xfry[21]→21,yfry[21]
→131,xfry[22]→22,yfry[22]→104,xfry[23]→23,yfry[23]→71,xfry[24]→24,yfry[
24]→61,xfry[25]→25,yfry[25]→36,xfry[26]→26,yfry[26]→33,xfry[27]→27,yfry
[27]→29,xfry[28]→28,yfry[28]→21,xfry[29]→29,yfry[29]→18,xfry[30]→30,yfr
y[30]→10,xfry[31]→31,yfry[31]→5,xfry[32]→32,yfry[32]→2,xfry[33]→33,yfry
[33]→2,xfry[34]→34,yfry[34]→1,xfry[35]→35,yfry[35]→1,Ffryx→17.999999999
999993`,Ffryy→154.`,pfry→327.09812272488006`}}-1], {fi, psi}]
```

and we get: Rex [De,fry] {0.0086215,{fi 0.477283, psi 0.483977}}.

i.e. the distance between German and Frisian text of “Human Rights” is equal to 0.0086215 with angles of rotations $\varphi = 0.477283$ and $\psi = 0.483977$.

In similar way in [4] we get the distance between:

Frisian and Dutch: Rex[Fry, Nl]→{0.00989766,{fi→0.355909,psi→0.329401}};
English and Frisian: Rex[GB, fry]→{0.0133242,{fi→0.499621,psi→0.509012}};
German and English: Rex[De, GB]→{0.0159066,{fi→0.585008,psi→0.583676}};
German and Dutch: Rex[De, Nl]→{0.0161782,{fi→0.444233,psi→0.430987}};
Afrikaans and Dutch: Rex[Afr, Nl]→{0.0190324,{fi→0.478443,psi→0.481322}};
Frisian and Afrikaans: Rex[Fry, Afr]→{0.0265862,{fi→0.471716,psi→0.446935}};
English and Dutch: Rex[GB, Nl]→{0.0282393,{fi→0.523854,psi→0.516264}};
English and Afrikaans: Rex[GB, Afr]→{0.0361154,{fi→0.488544,psi→0.48138}};
German and Afrikaans: Rex[De, Afr]→{0.0453406,{fi→0.514139,psi→0.503535}}.
Those data are visible in the symbolic diagram (**Figure 1**).

2.2. North German Languages

The letter data (rang-letter-frequency) for the texts of this group of languages are:
In the Danish text "**VERDENSERKLÆRINGEN OM MENNESKERETTIGHEDERNE**"
Dk{1493,826,697,611,607,597,534,502,473,443,393,278,245,237,224,195,115,103,86,80,71,58,52,34,15}
In the Icelandic text "**Mannréttindayfirlýsing Sameinuðu þjóðanna"**:
Ic{803,729,633,621,508,494,476,415,384,381,306,296,231,201,168,154,128,126,113,104,104,93,93,80,74,74,60,53,49,22,17}
In Føroyskt (Faroese) text "**HEIMSYVIRLÝSING UM MANNARÆTTINDI**":
Fa{835,708,679,564,555,511,398,391,380,356,335,301,221,185,183,173,165,159,147,115,108,82,78,64,58,41,35,25,11}
In Norvigian bokmol text "**VERDENSERKLÆRINGEN OM MENNESKERETTIGHETENE**":
Nb{1372,762,718,690,567,561,478,463,391,376,308,298,234,213,208,204,102,101,100,80,51,49,46,22}
In Norvigian nynorsk text "**DEN INTERNASJONALE FRÅSEGNA OM MENNESKERETTANE**":
Nn{1161,660,656,653,622,600,519,438,386,370,336,320,248,184,173,165,105,103,98,85,65,63,49,11})
In the Swedish (Svenska) text "**Allmän förklaring om de mänskliga rättigheterna Ingress**":
Sv{878,875,807,797,760,659,575,514,366,341,311,303,258,245,240,218,191,156,119,112,97,94,80,45,39,1}
As before, we get the distance between (**Figure 2**):
Norwegian and Swedish – Rex[No,Sv]→{0.00710435,{fi→0.4599,psi→0.448084}};
Norwegian and Icelandic – Rex[No,Ic]→{0.00762928,{fi→0.479103,psi→0.485782}};
Danish and Norwegian – Rex[Dk,No]→{0.00864969,{fi→0.557609,psi→0.576521}};
Swedish and Icelandic – Rex[Sv,Ic]→{0.0130824,{fi→0.547709,psi→0.563627}};
Danish and Swedish – Rex[Dk,Sv]→{0.0162672,{fi→0.451468,psi→0.452169}};
Danish and Icelandic–Rex[Dk,Ic]→{0.019497,{fi→0.625088,psi→0.647884}}.

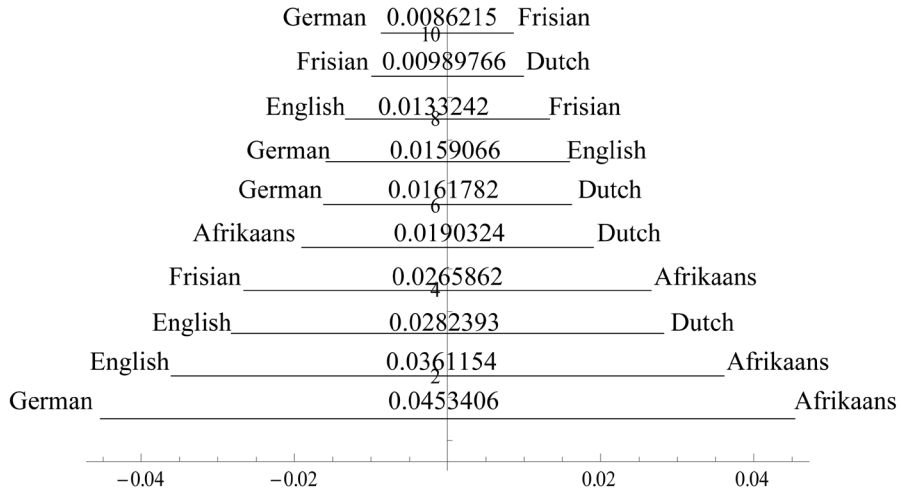

Figure 1. The distances between 10 pairs of West German languages in increasing order (top-down).

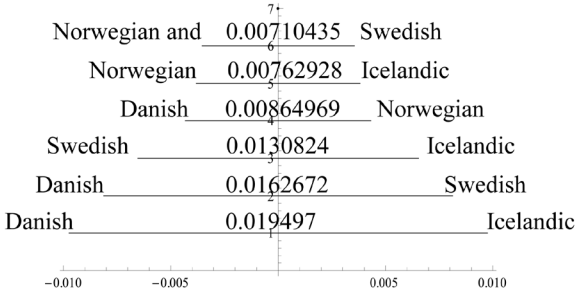

Figure 2. The distances between 6 pairs of North German languages in increasing order (top-down).

3. Other Point of View about the Distance between Languages of the German Group

In this paper we introduce the term *imaginary* or *virtual languages* in such a way:

Virtual west germen language, briefly VWGL has the sequence of absolute frequencies of letters as mean values of corresponding frequencies of the West germen languages shown above. This sequence is:

VWGL{{1754,1069,790,765,606,574,530,504,492,461,413,378,323,212,173,163,130,126,117,109,84,56,52,40,36,34,31,2,1,0,0,0,0,0,0}+{1077,803,714,706,705,699,609,465,446,398,324,292,224,196,189,166,165,152,109,99,75,18,16,15,7,4,0,0,0,0,0,0,0,0,0,0}+{1694,776,753,607,568,491,446,442,414,378,373,301,246,181,143,135,113,112,98,77,72,6,2,0,0,0,0,0,0,0,0,0,0,0,0}+{1461,1105,736,651,639,632,528,454,390,387,348,309,288,236,228,224,159,154,145,136,131,104,71,61,36,33,29,21,18,10,5,2,2,1,1}+{2144,1187,764,684,681,645,622,545,390,388,352,324,306,205,178,176,166,152,151,147,142,100,96,1,1,0,0,0,0,0,0,0,0,0,0}+{1652,1243,903,722,709,561,455,455,454,453,378,285,242,238,214,174,172,168,151,129,120,109,99,16,11,7,3,2,0,0,0,0,0,0,0}+{961,878,709,673,513,495,486,447,394,351,314,184,166,154,153,149,134,123,122,93,67,15,3,1,0,0,0,0,0,0,0,0,0,0,0}+{1002,801,779,777,689,549,533,467,429,364,319,304,196,191,136,134,128,127,111,105,56,41,15,11,6,0,0,0,0,0,0,0,0,0,0}+{895,821,658,621,546,520,465,398,387,362,338,224,222,210,196,184,166,163,161,155,132,119,116,116,114,72,56,52,26,25,15,12,12,1,0}}/9

or briefly

VWGL{12640/9,8683/9,6806/9,6206/9,5656/9,574,1558/3,4177/9,3796/9,3542/9,351,289,2213/9,1823/9,1610/9,1505/9,1333/9,1277/9,1165/9,350/3,293/3,568/9,470/9,29,211/9,50/3,119/9,77/9,5,35/9,20/9,14/9,14/9,2/9,1/9}

Analogically, *Virtual north germen language*, briefly VNGL, is presented by the sequence:

VNGL{6542,4560,4190,3936,3619,3422,2980,2723,2380,2267,1989,1796,1437,1265,1196,1109,806,748,663,576,496,439,398,256,186,116,95,78,51,22,17}/6

or briefly

VNGL{3271/3,760,2095/3,656,3619/6,1711/3,1490/3,2723/6,1190/3,2267/6,663/2,898/3,479/2,1265/6,598/3,1109/6,403/3,374/3,221/2,96,248/3,439/6,199/3,128/3,31,58/3,95/6,13,17/2,11/3,17/6},

and finally, *Virtual west-north germen language*, briefly VWNGL, with average value of frequencies of the all 15 west and north germen language, with the sequence:

VWNGL{6394/5,13243/15,10996/15,10142/15,1855/3,8588/15,7654/15,460,6176/15,5809/15,1716/5,4397/15,730/3,3088/15,2806/15,2614/15,713/5,135,1828/15,542/5,275/3,1007/15,868/15,517/15,397/15,266/15,214/15,31/3,32/5,19/5,37/15,14/15,14/15,2/15,1/15}.

(The subgroup of East German languages has no any official text of “Universal Declaration of the human Rights”.)

We calculated the distances of the German languages from the 3 Virtual languages and they are (without angles of rotations) arranged in increasing order (**Table 1**):

and also: Rex [VWNGL,VWGL] →{0.000974309,{fi→0.453276,psi→0.457978}};

Rex [VWNGL,VNGL] →{0.00993812,{fi→0.622273,psi→0.630315}};

Rex [VWGL,VNGL] →{0.0100142,{fi→0.566539,psi→0.575076}}.

Finally, we show a diagram (**Figure 3**) of the letter frequencies in the cited text of Universal Declaration (left

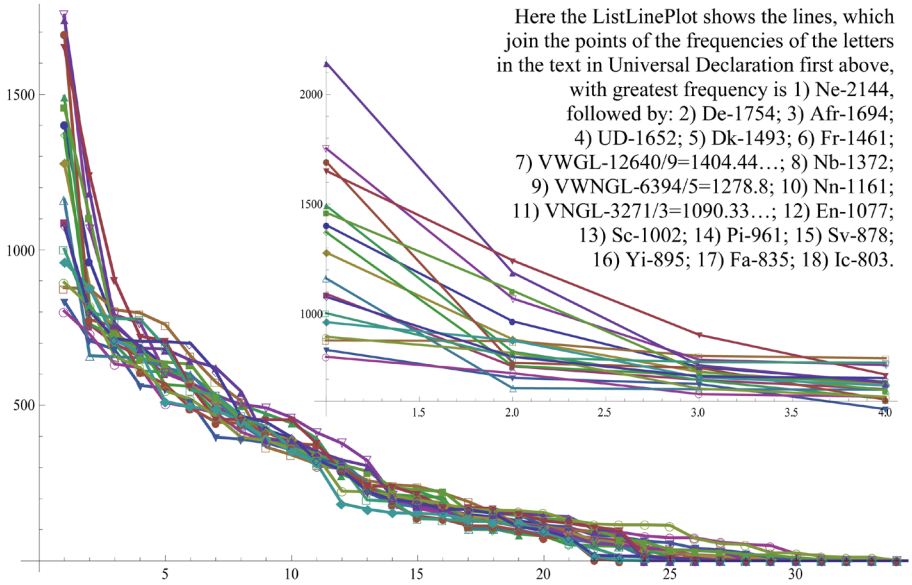

Figure 3. The ListLinePlot of the frequencies of the letters in the text of Universal Declaration of all 15 German languages and 3 Virtual German languages for the 35 frequencies (left) and for the first 4 frequencies (right).

Table 1. The distances between all 15 languages of the German group having the text of Universal Declaration and the Virtual WGL (left), NGL (in middle) and WNGL (right) (top-down).

Rex[VWGL,fry] →{0.00247601}	Rex[VNGL,Fa] →{0.00610147}	Rex[VWNGL,fry] →{0.0027904}
Rex[VWGL,Yi] →{0.00811229}	Rex[VNGL,Dansk] →{0.00800209}	Rex[VWNGL,Ic] →{0.0068969}
Rex[VWGL,Fa] →{0.00927121}	Rex[VNGL,Sv] →{0.00751063}	Rex[VWNGL,Yi] →{0.00760748}
Rex[VWGL,Ic] →{0.00928276}	Rex[VNGL,Yi] →{0.0068453}	Rex[VWNGL,Fa] →{0.00834023}
Rex[VWGL,De] →{0.00963576}	Rex[VNGL,En] →{0.00873952}	Rex[VWNGL,De] →{0.00846554}
Rex[VWGL,UD] →{0.00981895}	Rex[VNGL,Nb] →{0.00880051}	Rex[VWNGL,UD] →{0.00991607}
Rex[VWGL,Ne] →{0.0118554}	Rex[VNGL,fry] →{0.010511}	Rex[VWNGL,Sc] →{0.0113349}
Rex[VWGL,Sv] →{0.0132834}	Rex[VNGL,UD] →{0.0111175}	Rex[VWNGL,Sv] →{0.0118391}
Rex[VWGL,Dansk] →{0.0133069}	Rex[VWNGL,Ic] →{0.0139702}	Rex[VWNGL,En] →{0.0127977}
Rex[VWGL,Sc] →{0.0141489}	Rex[VNGL,Sc] →{0.0122394}	Rex[VWNGL,Dansk] →{0.0137273}
Rex[VWGL,En] →{0.0145651}	Rex[VNGL,Ne] →{0.0140147}	Rex[VWNGL,Pijin] →{0.0167792}
Rex[VWGL,Pijin] →{0.0196085}	Rex[VNGL,Afr] →{0.0172831}	Rex[VWNGL,Ne] →{0.0140889}
Rex[VWGL,Nb] →{0.0222891}	Rex[VNGL,De] →{0.0200066}	Rex[VWNGL,Nb] →{0.0233984}
Rex[VWGL,Afr] →{0.0289614}	Rex[VNGL,Pijin] →{0.0227993}	Rex[VWNGL,Afr] →{0.0318355}
Rex[VWGL,Nn] →{0.0506215}	Rex[VNGL,Nn] →{0.0307185}	Rex[VWNGL,Nn] →{0.0532962}

—for the all 35 points; right—for the first 4 points in decreasing order of the absolute frequencies).

References

[1] Тонков, Т. Върху близостта на езиците от групата на германските езици [About the Proximity of the Languages of the Group of German Languages].—в: Дни на науката`2013, Велико Търново (in print).

[2] Тонков, Т. (2009) Разпознаване на образи—обща теория и примери [Pattern Recognition—General Theory and

Examples]. Математически форум, том XI, брой 1, 2-16.

[3] Universal Declaration of Human Rights (in Internet: Official UN Universal Declaration of Human Rights Home Page).

[4] Тонков, Т. (2013) За близостта между естествените езици [On the Proximity between Natural Languages]—в: Дни на науката`2012, Велико Търново, 575-583.