Kriterio de la verboklaso

Konsiderante ĉi tiun kriterion, la oracioj povas esti: senpersonaj, aktivaj, transitivaj, netransitivaj, refleksaj k.t.p.

Kiel ni vidas, ĉi tiu klasifo paralelas al tiu de la verboj, tiel ke ĝi preskaŭ ne havas intereson; sole nur, ni menciu ke paradokse povas ekzisti transitivaj oracioj kun netransitivaj verboj. Ekz.:

Petro ŝatas vivi sian vivon. /37/

luj misklasifoj

Eraro iel komuna estas inkluzi la elipsajn oraciojn en la kriterion de la verbo-modo. Ŝajne ili povas esti konsiderataj sub la struktura kriterio, sed ĉar elipso estas figuro prefere ĉi tiuj oracioj devas esti studataj en la Figura Sintakso.

Kelkaj gramatikoj mencias "tranĉitajn oraciojn" /franclingve: propositions incises/ kaj donas ekzemplojn kiel jenan:

Mi statas laca - diris Rikardo - pro tiel enuigaj laboroj dum la tuta tago. /38/

Kiu ja estas kompleksa oracio kun ŝanĝo en normala vortordo; do, ankaŭ ĝi devas esti pritraktata en la ĉapitro de la Figura Sintakso.

Notoj

[U Scienca Revuo Vol 18 No 2 /70/
[U] Scienca Revuo Vol 19 No 1 /73/

Resumo

KLASIFO DE ORACIOJ
/Felix Garcia Blázquez, Caracas, Venezuelo/

Oni ekzamenas la malsamajn kriteriojn por klasifi oraciojn cele fiksi la nomenklaturon kaj disklarigi iomajn konceptojn.

Če la klasifo de oracioj laŭ ilia strukturo, oni fiksas la koncepton de komplementita oracio fronte al tiu de kompleksa oracio.

Če la klasifo de la oracioj laŭ ilia interrilato interne de la periodo, oni eliminas la konfuzon kiu ekzistas inter sendependa, kunordigita kaj "čefa" oracio.

Estas komentataj plie, aliaj kriterioj nome tiu de la modo kaj tiu de la speco de verbo.

SCIENCA REVUO de Internacia Scienca Asocio Esperantista BEOGRAD, Jugoslavio | El Vol. 24 n-ro 2/3(100/101) 20.5.1973.

KVANTA KARAKTERIZO DE VORTARA RIČECO DE TEKSTOJ

/H. D. Maas, SAARBRÜCKEN, okc. Germanio/

La preferata sindedičo de la moderna lingvistiko al generativa transforma gramatiko facile forgesigas pri aliaj ne malpli allogaj esplorindaj problemoj de la ĝenerala lingvoscienco. Tiurilate ni volas atentigi pri kvantaj rilatoj inter la amplekso de teksto kaj de ties vortaro, kiuj estas konindaj en la sfero de komputera lingvo-operaciado (KLO). Kiel kaŭzon de la malofta interesiĝo de la lingvistoj je tiaj kvantaj esploroj oni povas supozi la fakton, ke rezonado pri la strukturo de lingvo surbaze de la propraj intuicioj estas pli facila ol la starigo de empirisme pruvendaj, matematike formaligitaj reguloj, kies fundamento ne estas sole intuicio, sed ekscioj rezulte de traesploro de donita materialo.

1. CELO DE LA ESPLORO.

En tiu ĉi artikolo ni uzos la terminon "teksto-vortaro" por signi la aron de ĉiuj vortoj uzitaj en difinita teksto. La amplekson de iu teksto T ni nomos N(T) aŭ simple N; ĝi estas la nombro de ĉiuj en ĝi entenataj vortoj. En tiu ĉi koncerno oni difinas la vorton ĝenerale kiel sinsekvon de literoj inter interspacoj aŭ interpunkciaj signoj.

La tekstovortaro povas esti difinita laŭ diversaj manieroj, depende de la difino de tiuj elementoj, kiuj apartenu al ĝi. Ni diferenciĝas ĝenerale inter du specoj da vortaraj elementoj:

a) vortoformoj - la vortoj kiel grafikaj unuoj, aperantaj en tiuj formoj en tekstoj. Eĉ se vortoformo aperas plurfoje en traktata teksto, ĝi troviĝas en la vortaro nur unufoje;

b) reduktitaj vortoformoj, kiujn ni nomos *vokabloj*. La vortotrunkon, kiun oni ricevas fortrančante la gramatikajn afiksojn de vorto-

formo, ni nomas *vokablo*. Se en la teksto ekzistas deversaj vortoformoj, redukteblaj al la sama vokablo, la koncerna tekstovortaro entenos tiun vokablon nur unufoje.

Oni vidas, ke povas ekzisti pluraj difinoj de vokablo, depende de tio, kion oni komprenas sub gramatika afikso. En nia laboro pri tekstoj en Esperanto ni rigardis -j kaj -n kiel gramatikajn afiksojn, sed estas tute klare, ke en la kazo de Esperanto la redukto de vortoformoj povas esti multe pli ekstrema.

Estis la celo de niaj esploroj trovi funkcian interrilaton inter la teksto-amplekso N kaj ties vortaro-amplekso V. Se oni atingus la celon, oni povus facile taksi, ĉu unu aŭtoro disponas pri pli granda vortotrezoro, ol alia. En la sfero de komputera lingvo-operaciado oni ŝatus antaŭscii, kiom da loko oni devas rezervi por la vortaro de traktota teksto aŭ en kioma grado kreskos la vortaro, kiam oni aldonas pluan tekston al jam prilaborita teksto-amaso.

2. SOLVOMETODO.

Jam delonge oni scias, ke tia interrilato inter V kaj N ekzistas. Charles Muller [1] citas, ke validas proksimume

$$V = \sqrt{N} ,$$

sed almenaŭ por ne tro grandaj tekstoj tiu ĉi formulo ne taŭgas. Aliloke troveblas alia pli preciza rilato:

$$V = N^k$$

Oni asertas, ke k estas konstanta, sendependa de la tekstolongo N. Per elnombrado de multaj tekstoj oni povas kontroli, ĉu la formulo jes aŭ ne validas. Tiukaze k estas kalkulebla laŭ

$$k = \frac{\lg V}{\lg N} .$$

Ni asertas, ke k malkreskas kun kreskanta N.

Jen tabelo el Herdan [2] :

N	V	k
1213	437	0,856
1258	493	0,868
1220	579	0,893
1196	456	0,863
1191	494	0,874
1193	412	0,850

Ĉar tiuj ĉi tekstoj hovas preskaŭ saman longon N, ni mezumis ricevante N_m = 1212 kaj V_m = 478. El tio sekvas k = 0,870.

Se oni kunigas ĉiujn ses tekstojn, oni hovos teksto kun longo N = 7271, kies elnombrado montris teksto-vortaron V = 1527. El tio sekvas la eksponento k = 0,824, kio signifas sufiĉan malkreskon. Se validus k = 0,870 universale, la kunigita teksto devus enhavi ĉ. 2300 vortoformojn. Ni do vidas, ke la konstanteco de k ne estas pruvebla, almenaŭ se temas pri vortoformovortaro.

La materialo de Ch. Muller [1] , kiu esploris la verkaron de la franca dramisto Corneille kaj okupiĝis speciale pri lia vokablaro, montras plue, ke k ankaŭ en la kazo de vokablo-vortaro ne estas konstanta.

Ni elkalkulis k por diversaj tekstoj, elnombritaj de Muller, konstatis, ke en la verkaro de Corneille regas proksimume la jena leĝo

$$\lg \frac{1}{k} = 0{,}0137 \sqrt{\lg N^3}$$

aŭ

$$\lg(-100 \lg k) = 0{,}135 + 1{,}50 \lg \lg N$$

Tio signifas, ke la interrilato inter lg(-100 lg k) kaj lg lg N estas proksimume lineara. Por pruvi tion, ni difinas x = lg lg N kaj y = lg(-100 lg k) kaj desegnas la konstatitajn punktojn (x,y) en diagramon.

Jen unue la tabelo:

N	V	k	x	y
21	18	0,950	0,120	0,342
105	70	0,913	0,310	0,600
192	117	0,905	0,358	0,638
196	101	0,874	0,360	0,767
434	222	0,890	0,421	0,708
482	212	0,866	0,428	0,796
506	236	0,876	0,432	0,756
577	254	0,872	0,441	0,778
635	272	0,867	0,447	0,785
666	259	0,855	0,450	0,832
724	291	0,860	0,456	0,813
913	328	0,850	0,471	0,846
1.162	395	0,846	0,487	0,857
1.752	494	0,833	0,510	0,898
2.438	690	0,836	0,530	0,888
3.177	808	0,832	0,544	0,910
3.399	751	0,817	0,548	0,954
10.010	1477	0,790	0,602	1,000
16.150	1620	0,762	0,625	1,072
20.268	1713	0,748	0,634	1,100
128.813	3728	0,698	0,708	1,193
139.906	3759	0,695	0,712	1,198
174.681	3299	0,673	0,720	1,236
218.213	3983	0,674	0,727	1,235
303.343	4022	0,658	0,739	1,260
532.800	5347	0,652	0,757	1,270

La tabelo jam pruvas nerefuteble, ke k malkreskas konsiderinde de preskaŭ 1 por malgranda N ĝis 0,65 por grandega teksto (en nia kazo ĝi estas la tuta verkaro de Corneille). La valoro y samtempe preskaŭ seninterrompe kreskas kaj atingas sian maksimumon por N = 532.800.

La koncerna diagramo montras la rilaton inter x kaj y kaj krome la regresan rekton

$$y = 0{,}135 + 1{,}50\ x$$

Por donita N ni povas kalkuli nun teorian eksponenton k kaj sekve la atendeblan vortaro-amplekson V_t. La devion de la reala V for de la teoria V_t ni kalkulas laŭ

$$d = (V_t - V)^2 / V_t$$

Jen tabelo por V_t kaj d:

N	V	V_t	d
21	18	18	0,0
105	70	70	0,0
192	117	111	0,3
196	101	115	1,7
666	259	272	0,6
1752	494	500	0,1
2438	690	602	12,9
128813	3728	3600	4,4

Se d estas pli granda ol 2, oni povas esti certa, ke tiu devio nešuldiĝas al nura hazardo.

2.1 RELATIVA KRESKO DE V.

Se ni supozas - konraŭe al niaj ĵusaj ekkonoj -, ke k estas konstanta, ni povas diferencii la ekvacion

$$V = N^k$$

kaj ricevas

$$V' = \frac{dV}{dN} = k\, N^{k-1}$$

aŭ

$$V' = k\, \frac{V}{N}\,.$$

Ni nun povas respondi al la demando, kiom kreskas V, se al havata teksto N kun vortaro V estas aldonita plua teksto kun longo N_o: La kresko V_o de V estas

$$V_o = k\, \frac{V}{N}\, N_o.$$

Se ni konsideras la tutan verkaron de Corneille, la formulo liveras la rezulton, ke

$$V' = 0{,}0065.$$

Tio signifas, ke oni povus atendi ses ĝis 7 novajn vokablojn inter ĉiuj mil tekstovortoj, se oni trovus ankoraŭ ne konatan verkon de la dramisto.

Konsiderante, ke k ne estas konstanta, sed varias laŭ nia trovita regresa formulo, ni ekhavas la sekvantan kreskon:

$$V' = k\, \frac{V}{N}\, (1 - 0{,}0473 \sqrt{\lg N^3}).$$

El tio sekvas $V' = 0{,}00475$ - do valoro iom pli malgranda. La formulo donas pluan informon: La kresko ĉesos, kiam

$$1 - 0{,}0473 \sqrt{\lg N^3} = 0,$$

do por N = 80.000.000, el kio sekvas k = 0,495 kaj V_{maks}=8350. Se la formulo estus valida por ĉiuj N, ni povus konkludi, ke la aŭtoro atingus maksimume vortaron de 8350 vokabloj, se li estus povinta skribi sufiĉe longe.

Sed oni vidas facile, ke la formulo ne validas por grandegaj N, ĉar por ili la kresko de la vortaro fariĝus negativa, kio komprenble ne povas okazi.

2.2 PLUA EKSPERIMENTO.

La rilatumon $\frac{N}{V}$ oni nomas *frekvenco* f. Por ĝi Werner Muller [3] konstatis la sekvantan leĝon por germanaj tekstoj:

(2.2a) $$\lg f = (0{,}179 \lg N + 0{,}026)^2$$

Por N = 100 ni ricevas f = 1,4 kaj sekve V = 71, por N = 11.000.000 rezultiĝas f = 45,6 kaj V = 241.000. Komparante la grandegan nombradon de Kaeding [4], kiu trovis en tekstaro de 11.000.000 da vortoj 258.000 malsamajn, kun tiu ĉi teoria rezulto, oni povas konstati tre bonan akordon kun la realeco.

Por la tekstoj de Corneille ni elkalkulis

(2.2b) $$\lg f = \lg \frac{N}{V} = (0{,}026 \lg N - 0{,}083)^2$$

Elirante de la funkcia rilato

$$\lg \frac{N}{V} = (a \lg N + b)^2$$

ni ekhavas

$$\lg V = \lg N - (a \lg N + b)^2$$

kaj post diferenciado

$$V' = \frac{V}{N}\, (1 - 2ab - 2a^2 \lg N)$$

El tiu formulo la fina kresko de la Corneille' a vortaro doniĝas nun kiel

$$V' = 0{,}0027$$

do malpli ol laŭ la antaŭaj kalkuloj.

Komparante la formulojn (2.2a) kaj (2.2b), ni estas tentitaj supozi, ke la konstanto b devus esti 0, el kio sekvas

$$\lg f = \lg N - \lg V = a^2\, (\lg N)^2 \quad \text{aŭ}$$

(2.2c) $$\lg V = \lg N - a^2\, (\lg N)^2$$

Tio finfine supozigas al ni, ke la vera rilato inter V kaj N devus havi la formon

$$\lg V = \sum_{n=1}^{\infty} a_n\, (\lg N)^n \qquad \text{kie } a_1 = 1,\ a_2 < 0$$

Ĉu tio estas vera, ni ne povas pruvi; tial ni iom okupiĝos pri la formulo (2.2c). Se ĝi estas preciza, oni povas kalkuli la koeficienton a^2 laŭ

$$a^2 = \frac{\lg N - \lg V}{(\lg N)^2}\,.$$

Jen tabelo:

	Teksto	N	V	k	a
a)	*Germana lingvo*				
	FAZ-tekstaro 1)	97.792	19.871	0,864	0,167
	rde-tekstaro 1)	104.918	20.025	0,857	0,168
	Kaeding [4]	11.000.000	258.200	0,770	0,182
	Susi (el [5])	16.269	1.497	0,754	0,242

	Legolibro (el [5])	10.373	1.839	0,813	0,215
	Trakl (el [8]) [2]	32.730	3.808	0,792	0,214
b)	*Rusa lingvo*				
	Puškin (el [2])	28.591	4.783	0,824	0,198
	Puškin "	4.703	1.672	0,878	0,183
	Puškin "	4.952	1.567	0,865	0,192
	Puškin "	9.147	2.432	0,855	0,192
	Steinfeldt-statistiko [6]	387.211	24.224	0,785	0,215
c)	*Angla lingvo*				
	teksto el [2]	1.212	478	0,870	0,206
	"	7.271	1.527	0,824	0,214
	gazetara lingvo [7]	43.989	6.001	0,815	0,200
d)	*Esperanto* (propraj esploroj)				
	(vortoformoj)	500	299	0,919	0,174
	"	900	491	0,912	0,173
	"	449	246	0,903	0,192
	"	2.183	925	0,888	0,183
	"	25.500	4.748	0,833	0,194
	(vokabloj)	25.500	2.800	0,782	0,222

1) Temas pri tekstaroj komplitaj de Departemento por Komputera Lingvistiko en Universitato de Saarlando.

2) La vortaro estis parte aŭtomate vokabligita.

Oni vidas, ke la koeficiento a same kiel k malmulte varias. Malgranda a signalas, ke la teksto enhavas relative grandan vortotrezoron, dum granda a signifas la malon. Sub la kondiĉo, ke la formulo (2.2c) estas ĝusta, ni povas taksi la maksimume atingeblan vortoprovizon. La funkcia rilato estas

$$\lg V_{maks} = 1/(4a^2)$$

Por atingi tiun vortaro-amplekson, oni bezonas tekston kun amplekso donita per la sekvanta formulo

$$\lg N_o = 1/(2a^2)$$

La koeficiento k en $V = N^k$ tiam atingus la ekzaktan valoron 0,5. Jam per alia metodo ni trovis preskaŭ la saman valoron por la verkaro de Corneille. Supozeble en tio kuŝas la pli profunda kaŭzo de la formulo de P. Guiraud

$$R=V/N^{0,5}$$

kiu rigardas la koeficienton R kiel mezuron por la vortoprovizo-amplekso.

Kiel gravan stilan indikilon oni komprenas la ripetfaktoron

$$f = \frac{N}{V} .$$

Laŭ niaj ĝisnunaj ekscioj ĝi forte dependas de la tekstolongo N. Tamen Josef Mistrik [9] asertas, ke por fakaj, profesiaj artikoloj validas f = 6,3 kaj por beletra prozo f = 2,42. El tiuj nombroj ni povas konkludi, ke liaj fakaj arikoloj estis multe pli longaj ol liaj beletrajoj, ĉar f kreskas kun kreskanta N.

Nia formulo (2.2c) ne povas kontentigi nin, ĉar ĝi asertas, ke tekstoj kun ampleksoj pli grandaj ol N_o liveras malpli grandan vortaron, ol V_{maks}, kiu jam estis atingita por tekstolongo N_o. Se V estas reprezentebla kiel funkcio de N, do

$$V = F(N),$$

ni devas postuli, ke F konstante kreskas kaj konverĝas al reala nombro.

2.3 LASTA EKPROVO.

Ni ekvidis, ke la formulo $V = N^k$ aŭ $k = \frac{\lg V}{\lg N}$ kun kelkaj rezervoj estas bone utiligebla. Se k estus konstanta, ĝi liverus la kreskon de la vortaro depende de la tekstolongo:

$$V' = \frac{dV}{dN} = k \frac{V}{N} .$$

Ĉar $k = \frac{\lg V}{\lg N}$, sekvas

$$\frac{dV}{dN} = \frac{\lg V}{\lg N} \cdot \frac{V}{N}$$

Ni pliĝeneraligas tiun ĉi diferencialan ekvacion al

(2.3a) $$V' = \frac{dV}{dN} = \left(\frac{\lg V}{\lg N}\right)^2 \cdot \frac{V}{N} .$$

Ni povas rigardi n kiel nekonatan variablon empirisme konstateblan. Logaritmigante la formulon kaj uzante la mallongigojn

$$k = \frac{\lg V}{\lg N} \quad \text{kaj} \quad f = \frac{N}{V},$$

ni ricevos

$$n = \frac{\lg V' + \lg f}{\lg k} .$$

Komputere statistikante Esperanto-tekston de longo N=900, ni konstatis la sekvantajn rezultojn:

N	V	V'	lg f	k	n
50	37	0,74	0,130	0,923	0,00
100	67	0,64	0,174	0,913	0,5
150	101	0,67	0,172	0,920	0,05
200	134	0,58	0,174	0,926	1,83
250	159	0,58	0,196	0,916	1,05
300	192	0,61	0,194	0,922	0,58
350	220	0,54	0,202	0,920	1,83
400	246	0,55	0,212	0,919	1,30
450	275	0,53	0,214	0,921	1,72
500	299	0,53	0,222	0,917	1,42
550	328	0,53	0,225	0,920	1,42
600	352	0,43	0,232	0,918	3,53
650	371	0,45	0,244	0,914	2;64
700	397	0,48	0,247	0,914	1,87
750	419	0,46	0,253	0,912	2,12
800	443	0,50	0,258	0,913	1,05
850	469	0,48	0,259	0,911	1,46
900	491	0,44	0,263	0,910	2,27

Oni vidas, ke n estas sufiĉe ŝancelema; se ni tamen kalkulas la aritmetikan mezumon de n (neglektante la tri komencajn valorojn), ni ekhavos valoron de ĉ. 1,74. Se ni volus difini n per tiu ĉi me-

todo, ni bezonus multe pli ampleksajn kaj detalajn statistikojn. Tial ni iros alian vojon.

Por la tekstaro de Corneille ni jam konstatis la regresan leĝon

$$\lg k = -0{,}0137\sqrt{\lg N^3}$$

el kiu sekvas la proksimuma kresko de la vortaro:

$$V' = k\,\frac{V}{N}\,(1 - 1{,}5\cdot 0{,}0137\sqrt{\lg N^3}).$$

Rememorigante la difinon de n kiel

$$n = \frac{\lg(V'\cdot f)}{\lg k},$$

ni hovos post kelka kalkulado

$$n = 1 - \frac{\lg(1 - \frac{3}{2}a\sqrt{\lg N^3})}{a\sqrt{\lg N}} \qquad (a=0{,}0137)$$

Ĉar $\lg(1-x) < -0{,}434\cdot x$, sekvas

$$n > 1 + \tfrac{3}{2}\cdot 0{,}43 = 1{,}65$$

Post tiuj anticipaj taksoj de n, ni volas okupiĝi pri la solvo de la diferenciala ekvacio (2.3a). Post transformo ĝi aspektas tiel

$$\frac{dV}{V\,(\lg V)^n} = \frac{dN}{N\,(\lg N)^n}$$

Tion oni povas facile integri; sekvas por $n \neq 1$

$$\frac{1}{(\lg V)^{n-1}} = \frac{1}{(\lg N)^{n-1}} + c$$

kaj por n = 1 la jam konata

$$\lg V = k \lg N \text{ aŭ } V = N^k.$$

Ni devas postuli, ke V konverĝu al V_∞, kiam N kreskas senfine. Tio okazas nur, se n estas pli granda ol 1. Tial ni pritraktos nur la kazojn n = 2 kaj n = 3; anstataŭ n-1 ni skribos m.

Por senfina N doniĝas vortaro-amplekso V_∞, kie

$$\frac{1}{(\lg V_\infty)^m} = c$$

aŭ

$$\lg V_\infty = \sqrt[m]{c}$$

Post iom da transkalkulo oni ekhavas

$$\lg V_o = \frac{\lg V \cdot \lg N}{\sqrt[m]{(\lg N)^m - (\lg V)^m}}$$

aŭ

$$\lg V_o = \frac{\lg V}{\sqrt[m]{1 - \left(\frac{\lg V}{\lg N}\right)^m}}$$

aŭ

$$\lg V_o = \frac{\lg V}{\sqrt[m]{1 - k^m}}$$

El tio facile sekvas formulo por k

(2.3b) $$k^m = 1 - \left(\frac{\lg V}{\lg V_o}\right)^m.$$

Plue validas

(2.3c) $$\lg V = \lg N \sqrt[m]{1 - \left(\frac{\lg V}{\lg V_o}\right)^m}$$

(2.3d) $$\lg N = \frac{\lg V}{\sqrt[m]{1-\left(\frac{\lg V}{\lg V_o}\right)^m}}$$

(2.3e) $$\lg V = \frac{\lg N}{\sqrt[m]{1 + \left(\frac{\lg N}{\lg V_o}\right)^m}}$$

Ni nun esploros tiujn ĉi formulojn por m = 1 kaj m = 2.

La nombrado de Kaeding [4] liveris N = 11.000.000 kaj V=258.200. El tio sekvas

por m = 1 $\lg V_o = 23{,}4$

por m = 2 $\lg V_o = 8{,}455$

Tio jam montras, ke m = 1 ne povas esti ĝusta, ĉar ĝi liveras maksimume atingeblan vortoprovizon de pli 10^{23} vortoj. Ni kontrolas ĉi tiujn rezultojn kalkulante V_o surbaze de la jam menciitaj FAZ-kaj rde-tekstaroj. Unue FAZ-tekstaro:

N = 97.792 V = 19.871

El tio sekvas

por m = 1 $\lg V_o = 31{,}2$

por m = 2 $\lg V_o = 8{,}455$

Due rde-tekstaro kun N = 104.918 kaj V = 20.025:

por m = 1 $\lg V_o = 30{,}6$

m = 2 $\lg V_o = 8{,}336$

Tiuj ĉi elnombradoj estas inter si kompareblaj, ĉar la tri vortaroj konsistas nur el vortoformoj kaj la tekstaroj estis kompilitaj el tekstoj de multe da aŭtoroj. Nia kalkulado do evidente pruvas, ke la proponita diferenciala ekvacio por m = 2 aŭ n = 3 bonege respegulas la realecon. Oni povas atendi, ke V_o por diversaj unuopaj aŭtoroj estas malpli granda.

Ĉar la kvociento $\lg V / \lg V_o$ varias inter 0 kaj 1, ni povas substitui

$$\frac{\lg V}{\lg V_o} = \sin x\ .$$

Tio modifas la formulon (2.3d):

$$\lg N = \frac{\lg V_o \cdot \sin x}{\sqrt{1-\sin^2 x}} = \lg V_o \cdot \operatorname{tg} x$$

$$\frac{\lg N}{\lg V_o} = \operatorname{tg} x \qquad x = \operatorname{arctg}\left(\frac{\lg N}{\lg V_o}\right)$$

$$\arcsin\left(\frac{\lg V}{\lg V_o}\right) = \operatorname{arctg}\left(\frac{\lg N}{\lg V_o}\right)$$

Por la sekvantaj tekstoj ni kalkulis la naturan logaritmon de V_o per komputero.

teksto	N	V	$\log_e V_o$	lg V_o
angla lingvo				
uson-angla gazetara laŭ [7]	43.989	6.001	14,97	6,48
artikolo de Hitler laŭ [2]	1.220	579	14,27	6,18
" de Stresemann "	1.191	494	12,85	5,56
" de Churchill "	1.258	493	12,52	5,43
" de Halifax "	1.196	456	12,16	5,27
" de Beneš "	1.213	437	11,77	5,09
" de Stalin "	1.193	412	11,43	4,96
mezumo de la antaŭaj ses	1.212	478	12,50	5,41
la antaŭaj ses kune laŭ [2]	7.271	1.527	12,96	5,61
ĉina lingvo				
laŭ [7]	13.248	3.332	15,62	6,77
Esperanto				
teksto de Lapenna	900	491	15,20	6,58
teksto de Zamenhof	449	246	12,80	5,55
teksto de Lapenna [10]	2.183	925	14,87	6,43
tekstaro I (el [11] kaj [12])	25.492	4.746	15,36	6,65
tekstaro II "	26.795	4.870	15,34	6,64
franca lingvo (vokabligita vortaro, laŭ [13] kaj [14]				
Pulchérie	16.669	1.355	10,75	4,66
tekstaro de Corneille	174.681	3.299	10,93	4,74
tekstaro de Corneille	303.343	4.022	11,02	4,77
Théodore	17.173	1.498	11,05	4,78
Rodogune	16.867	1.498	11,08	4,79
La Place Royale	13.807	1.385	11,11	4,81
Le Cid	16.424	1.536	11,21	4,86
tekstaro de Corneille	218.213	3.983	11,23	4,87
tekstaro de Corneille	20.268	1.713	11,27	4,88
tuta verkaro de Corneille	532.800	5.347	11,31	4,89
La Suite du Menteur	17.611	1.687	11,43	4,95
tekstaro de Corneille	139.906	3.759	11,45	4,96
Menteur	16.677	1.667	11,48	4,98
tekstaro de Corneille	128.813	3.728	11,50	4,99
Phèdre	14.217	1.653	11,73	5,09
Clindor	1.447	494	11,86	5,14
Médée	14.256	1.715	11,87	5,15
germana lingvo				
nombrado el Kaeding [4]	11.10^6	258.200	19,48	8,55
FAZ-tekstaro	97.792	19.871	19,48	8,55
rde-tekstaro	104.918	20.025	19,20	8,32
Heisenberg laŭ [3]	68.562	15.050	19,09	8,28
Strittmaier "	106.785	18.050	18,41	7,97
Heimpel "	42.446	9.130	17,63	7,64
Heisenberg "	25.797	6.330	17,25	7,48
Mann "	23.941	5.940	17,13	7,44
Bollnow "	70.316	9.685	16,13	7,00
Frisch "	56.695	8.533	16,10	6,99
Jung "	38.863	5.800	15,14	6,56
Heisenberg "	13.514	2.934	14,69	6,37
Bergengruen "	8.460	2.060	14,22	6,16
infano Hilde laŭ [5]	11.187	2.037	13,22	5,74
legolibro "	10.373	1.839	12,91	5,60
infano Susi laŭ [5]	16.269	1.497	11,13	4,83
latina lingvo				
Plautus laŭ [7]	33.094	8.437	18,25	7,92
rusa lingvo				
Steinfeldt [6]	387.211	24.224	16,28	7,06
Puškin laŭ [2]	4.703	1.672	15,49	6,72
Puškin "	28.591	4.785	15,02	6,51
Puškin "	9.147	2.432	15,02	6,51
Puškin "	4.952	1.567	14,65	6,34

La statistiko pri la verkaro de Corneille evidente montras, ke lg V_o versajne ne dependas de la tekstolongo. La aliaj statistikoj tamen indikas ioman kreskon de lg V_o kun kreskanta tekstolongo. La kaŭzo de tio eble estas, ke la pli longaj tekstoj efektive estas tekstaroj, al kiuj kontribuis pluraj aŭ eĉ multaj aŭtoroj. La vortotrezoro de pluraj kompreneble estas pli granda ol tiu de unuopulo.

Ni ankoraŭ interesiĝas pri la demando, kian amplekson V_d havas teksto de la longo $N = V_o$. La respondo sekvas facile el formulo (2. 3e):

$$\lg V_d = \frac{1}{\sqrt{2}} \lg V_o$$

aŭ

$$\frac{\lg V_d}{\lg V_o} = \frac{\lg V_d}{\lg N} = \frac{1}{\sqrt{2}} \approx 0{,}707$$

Tio signifas: Se por iu teksto la koeficiento k havas la valoron 0,707, tiam ni povas esperi, ke la teksto-amplekso N egalas la teorie maksimuman vortaro-amplekson V_o. Pro la statistika variado tiu ĉi metodo de difino de V_o kompreneble ne estas konsilinda. En la kazo de la verkaro de Corneille ni disponas pri la jam menciita regresa formulo por k

$$\lg k^{-1} = 0{,}0137 \sqrt{\lg N}^{\,3}$$

kiu permesas kalkuli N por donita k. Por k = 0,707 ni ricevas lgN= 4,92 - do lg V_o devas esti 4,92, kio bone akordiĝas kun la nombro 4,89, kiun ni kalkulis por la tuta verkaro.

Finfine ni volas komuniki kelkajn relativajn kreskojn de la vortaro, kiuj devus validi ĉe la fino de la nomitaj tekstoj:

teksto	N	V	kresko V'
Kaeding	11.10^6	258.200	0,0107
FAZ-tekstaro	97.792	19.871	0,132
Esperanto-tekstaro	25.492	4.746	0,108
" "	25.492	2.800	0,052 (vokabloj)
Trakl (germana)	32.730	3.808	0,058 (vokabloj)

Tio informas nin, ke se ni aldonas al la supre citita Esperanto-teksto pluan artikolon de 1000 vortoformoj, tiu teksto liveros 108 novajn vortoformojn aŭ 52 novajn vokablojn. Tiun aserton ni te-

stis aldonante tekston kun 1303 vortoformoj kaj elnombris 124 novajn vortoformojn, dum oni povis atendi 140. Samtempe montriĝis, ke la teksto enhavis 60 novajn vokablojn, dum teorie atendeblaj estis 67. La rilatimo 60:124 estas tamen preskaŭ precize 67:140.

Por la Esperanto-tekstaro de 25.500 vortoformoj ni konstatis vortaran amplekson de 4746 vortoformoj respektive 2800 vokabloj. El tiuj donitaĵoj sekvis

$$\lg V_o = 6{,}65 \qquad \text{(vortoformoj)}$$
$$\lg V_o = 5{,}55 \qquad \text{(vokabloj)}$$

La diferenco kompreneble šuldiĝas al tio, ke el ĉiu vokablo oni povas formi regule plurajn vortoformojn. La diferenco inter ambau $\lg V_o$, kiu estas 1,10, estas komprenebla, se oni supozas, ke ĉiu vokablo povas liveri $10^{1,1} = 12{,}6$ vortoformojn meze. Se oni konsideras, ke speciale el verba radiko estas formeblaj tiom da participoj kun diversaj gramatikaj finaĵoj, tiu faktoro ne šajnas troigita. Utiligante la statistikon de V. Sadler [15] oni povas taksi, ke ĉiu vokablo estas pluformebla al ĉ. 9 vortoformoj.

Ĉar la prezentita materialo bone akordiĝas kun la eldiroj ĉerpeblaj el nia hipoteza diferenciala ekvacio, ni povas esti certaj, ke la derivitaj formuloj estas universale aplikeblaj kaj kontentige respegulas la statistikajn leĝojn inter la ampleksoj de taksto kaj vortaro.

BIBLIOGRAFIAJ NOTOJ

[1] Charles Muller, Initiation à la statistique linguistuque. Paris 1968.
[2] G. Herdan, Language as Choice and Chance. Groningen 1956.
[3] W. Muller, Gedanken zur automatischen Analyse von Normen und Normabweichungen. En: Muttersprache, kajero 9/10, 1969.
[4] Kaeding, Häufigkeitswörterbuch der deutschen Sprache. Steglitz bei Berlin, 1898.
[5] H. Meier, Deutsche Sprachstatistik. Hildesheim 1964.
[6] E. Steinfeldt, Russian Word Count, Moscow (sen jaro).
[7] G. K. Zipf, The Psycho-Biology of Language. Cambridge (USA) 1965.
[8] W. Klein, H. Zimmermann, Trakl. Versuche zur maschinellen Analyse von Dichtersprache. En: Sprachkunst, kajero 1/2, 1970.
[9] Josef Mistrik, Quantitative Methods and Stylistic Typology. En: Recueil linguistique de Bratislava, Volume II, Bratislava 1968.
[10] I. Lapenna, La internacia lingvo kiel esprimo kaj antaŭeniganto de universalismaj tendencoj. En: I. Lapenna, Elektitaj paroladoj kaj prelegoj, Rotterdam 1966.
[11] I. Lapenna, Elektitaj paroladoj kaj prelegoj, Rotterdam 1966.
[12] L.L. Zamenhof, Originala Verkaro. Leipzig 1929.
[13] Ch. Muller, Essai de statistique lexicale, Paris 1964.
[14] Ch. Muller, Etude de statistique lexicale, Paris 1967.
[15] V. Sadler, Relativaj oftecoj de kelkaj lingvaj elementoj en Esperanto. En Scienca Revuo, kajero 2/3, 1959.

SCIENCA REVUO de Internacia Scienca Asocio Esperantista BEOGRAD, Jugoslavio

El Vol. 24 n-ro 2/3(100/101) 20.5.1973.

LA LINGVONIMIKO

- ĝiaj esenco kaj problemoj -

/A.D.Duličenko, AŠHABAD, Sovetio/

En la lingvistika scienco de la lasta periodo intensive evoluiĝas iuj novaj branĉoj, al kiuj apartenas ankaŭ onomastiko. La amasigo kaj la plenigo de materialo pri la nomita temaro postulis diferenciĝon kaj apartiĝon de kelkaj malargaj onomastikaj branĉoj. Tiel, el la toponimiko apartiĝis hidronimiko, urbonimiko /aŭ polisonimiko/ ktp., el la antroponimiko - etnonimiko.

En la artikolo estas entreprenita unu provo pli diferencigi etnonimian materialon, apartiginte novan branĉon /aspekton/ pri nomoj de lingvoj de la mondo. Bone estas konate, ke lingvo estas unu el la plej gravaj signoj de nacio aŭ popolo. Pro tio ni konsideras la materialon pri nomoj de lingvoj kiel sendependa branĉo /aspekto/ de la etnonimiko[1/].

La nuntempa mondo parolas proksimume 2500 - 3000 lingvojn. La kompleteco de lingva karto estas kondiĉita /krom iuj aliaj kaŭzoj/ de la malkontenta esploreco de apartaj lingvogrupoj; ofte pro tio estas embarase diferencigi unu dialekton aŭ subdialekton de lingvo mem. En la historio de esploro de la lingvoj en Afriko kaj en Azio estas konataj eventoj, kiam esploristoj nomigis unu priskribatan lingvon, baziĝante a/ sur la nomo de tribo aŭ gento /t.e. sur la bazo de etnonimo/, b/ sur la nomo de loko, c/ sur la nomo, kiu uzis najbaraj triboj rilate al alia tribo, d/ sur la nomo de unu el kelkaj dialektoj de la sama lingvo ktp. Pro tio lingvistika karto de la mondo suferas de defektoj kaj de konfuzaĵo. Pro tio ni konstatas, ke nuntempe la homoj de nia terglobo parolas ĉirkaŭ 2500--3000 lingvojn.

1/ Ni agnoskas ankaŭ koheron de nomoj de la lingvoj kun lingvistika terminaro, sed en la artikolo, pro la manko de loko, ni aparte la problemon ne pritraktas.