

The Art of Debate in the Age of the Internet: Computational Text Analysis Resources at the RSU

Alise Tīfentāle, PhD

Communication Studies Department

DIGITAL LATVIAN LANGUAGE RESOURCES AND TOOLS IN
COMMON RESEARCH INFRASTRUCTURE

Riga, March 5, 2020

We, the peoples of United Nations
Determined to save succeeding generations from the scourge of war,
which twice in our lifetime has brought untold sorrow to mankind, and
To reaffirm faith in fundamental human rights, in the dignity
and worth of the human person, in the equal rights of men and
women and of nations large and small . . .

Charter of the United Nations



Maria Bordy, untitled (assembly hall of the United Nations), undated.
Edward Steichen, ed., *The Family of Man* (New York: Museum of Modern Art, 1955), 184–85.

The Index of Internet Aggressiveness

<http://barometrs.korpuss.lv/>

Authors:

Communication Studies Department, Riga Stradins University (Sergejs Kruks, Anda Rožukalne, Klāvs Sedlenieks, Ruta Siliņa, Ilva Skulte)

Institute of Mathematics and Computer Science of University of Latvia, Artificial Intelligence Lab (Ginta Garkāje, Normunds Grūzītis, Valters Rostoks)

Supported by: Ombudsman of the Republic of Latvia; Friedrich Ebert Foundation

Partners: Delfi, TVNET, Apollo

Dataset:

~19.80 M comments, ~515.19 M words, ~160,060 articles

Interneta agresivitātes indekss

eksperimentāla versija

Datu kopa:

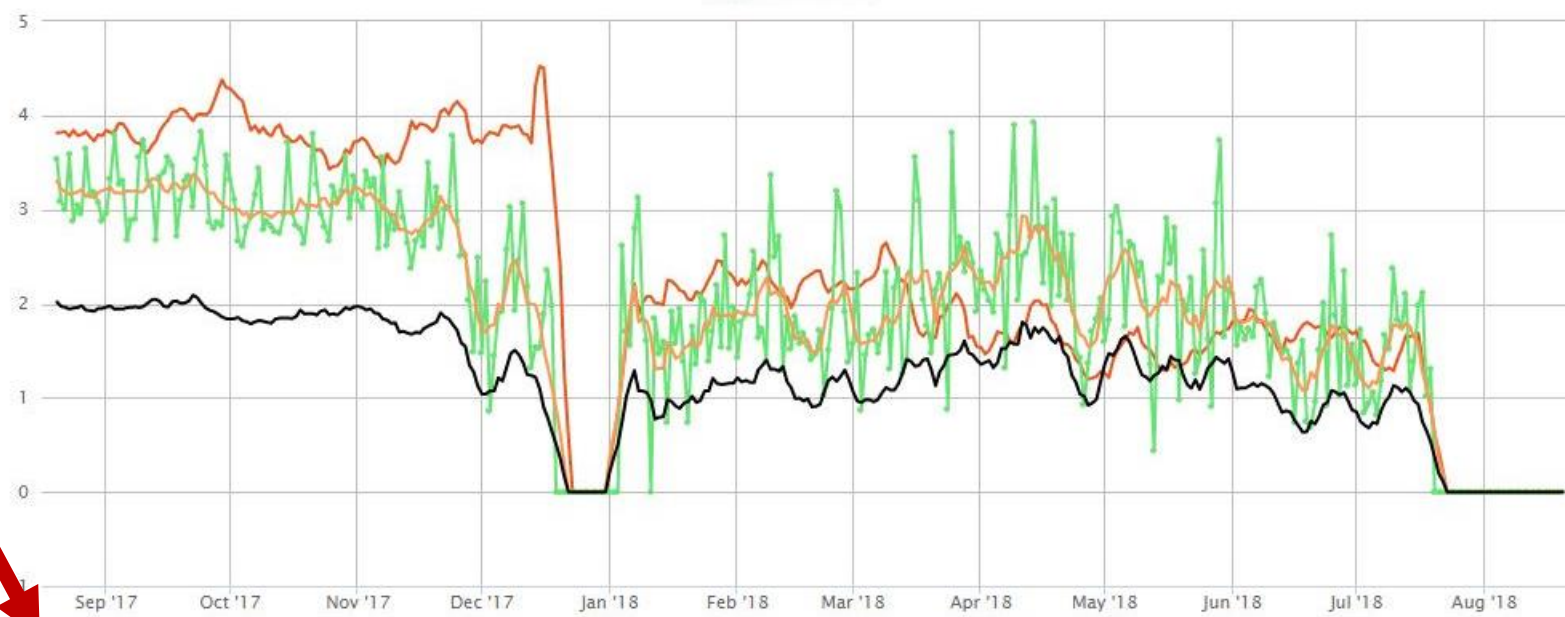
Pēdējās 365 dienas ▾ 2017-08-20 - 2018-08-20 LV ▾



Kopējās tendences Atslēgvārdu tendences Biežumi Novirzes Konteksti Komentāri Atslēgvārdi Stopvārdi Topvārdi Datumi Par projektu

1. grafiks. Potenciāli agresīvo vārdlietojumu īpatsvars, analizējot ziņu portālu komentārus kā vienotu datu kopu

[Lejupielādēt datus](#)



— Faktiskā vērtība lv — Tendence lv — Vidējā vērtība lv — Tendence, ņemot vērā atslēgvārdu svarus lv — Faktiskā vērtība ru — Tendence ru — Vidējā vērtība ru — Tendence, ņemot vērā atslēgvārdu svarus ru

Highcharts.com



The Index of Internet Aggressiveness

<http://barometrs.korpuss.lv/>

Initially over 65 thousand used words
(in comments on portals)



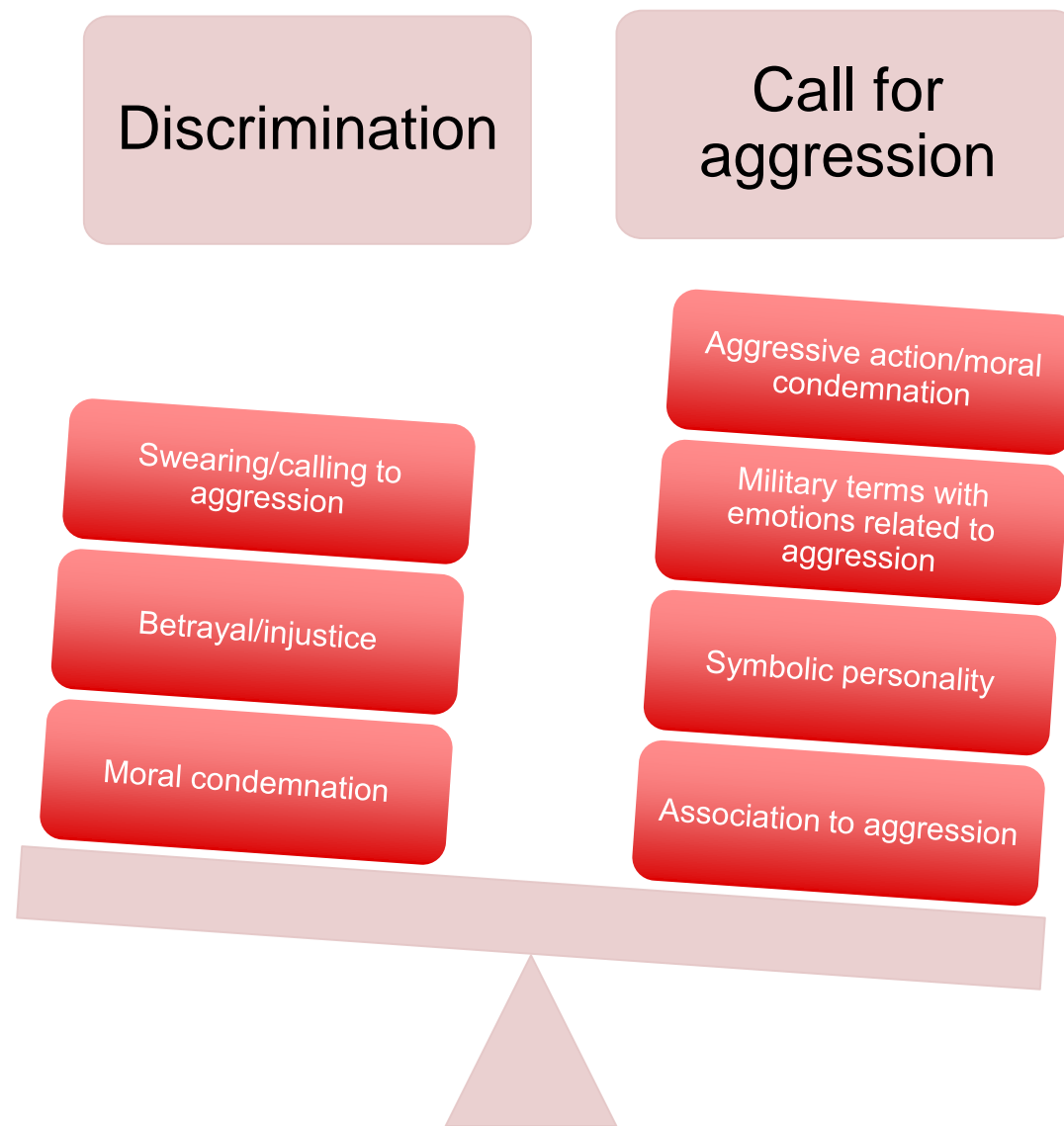
3.5 thousand “aggressive” word forms



900 basic word stock in all forms and
spellings

The Index of Internet Aggressiveness

<http://barometrs.korpuss.lv/>



The Index of Internet Aggressiveness

<http://barometrs.korpuss.lv/>

Most often used words 28.05.-31.10.2011.

- | | |
|-------------------|--------------------|
| 1. Kārt – 555 435 | 6. Sist – 168 360 |
| 2. Uts – 310 806 | 7. Ciest – 151 686 |
| 3. Šaut – 296 917 | 8. Karš – 131 191 |
| 4. Auns – 269 912 | 9. Līķis – 112 213 |
| 5. Āzis – 216 267 | 10. Āksts – 106 58 |

The Index of Internet Aggressiveness

<http://barometrs.korpuss.lv/>

Some expressions found in comments can be seen as incitement to aggression

Calling for public death penalty

Use of aggressive expressions picked from politics or media

Large part of the content consists of reciprocal miscalling, reflections on the comments of other commentators

Aggressive statements trend to deviate quickly from content of the message on which the comments are made

Self-marginalization and denial of the different

Sniffy, disdainful attitude towards ourselves and our nation

We are a shitty nation

Intolerance to the different – nationalism and extreme trends

Saeima (Parliament) Corpus Analysis

<http://saeima.korpuss.lv/>

Authors:

Ilze Auziņa, Roberts Dargis, Sergejs Kruks, Guna Rābante-Buša, Ilva Skulte

Communication Studies Department, Riga Stradins University

Institute of Mathematics and Computer Science of University of Latvia, Artificial Intelligence Lab

Funding:

State research program project EKOSOC-LV Nr. 5.2.5.

Dataset:

Corpus includes all words that parliament, ministers, and other participants have said (and also shouted from the hall) since 1993 – **more than 13 M words.**

Par vietni

Saeimas debašu korpus ietver visus vārdus, ko deputāti, ministri un citi Saeimas sēžu dalībnieki ir teikuši no tribīnes (un arī klieguši no zāles) kopš 1993. gada – kopā 13 miljoni vārdu. Tik liels vārdu skaits ļauj skaidrot politiskās runas kopsakarības. Noskaidrot ne tikai deputāta vai partijas nostāju kādā jautājumā, bet arī atbildības pieņemšanos par pausto vērtējumu vai priekšlikumu.

Vietni ir radījuši Rīgas Stradiņa universitātes Komunikācijas studiju katedras un Latvijas Universitātes Matemātikas un informātikas institūta Mākslīgā intelekta laboratorijas pētnieki, īstenojot Valsts pētījumu programmas projektu EKOSOC-LV Nr. 5.2.5.

Ar šo digitālo rīku žurnālisti, politiologi, valodnieki un citi interesenti var noskaidrot, ko deputāti ir teikuši par kādu noteiktu problēmu.

Saeimas korpusa vietnes veidotāji:

Ilze Auziņa, Roberts Darģis, Sergejs Kruks, Guna Rābante-Buša, Ilva Skulte

Lai labāk saprastu Saeimas debašu korpusa iespējas politikas un valodas analīzē, piedāvājam lasīt agrāk veiktos [pētījumus](#) un [lietošanas pamācību](#).

Stenogrammas iegūtas no [Saeima.lv](#). Tehnisko nodrošinājumu līdzfinansē [Clusterpoint](#).



RĪGAS STRADIŅA
UNIVERSITĀTE



Latvijas Universitātes
Matemātikas un informātikas institūts



Teksts



No

Līdz

Periods

Sēdes tips

 kārtējā ārkārtas sesijas

Kārtot pēc

Rezultāts

... Godātie deputāti! Prezidija locekļi un komisiju priekšsēdētāji! Šodien beidzas priekšlikumu iesniegšanas termiņi ļoti nozīmīgiem likumprojektiem, bet nestrādā **internets**. Un tāpēc es lūdzu šoreiz pieņemt šos ierosinājumus rakstiskā formā, nenoformējot tos **interneta** variantā.

... . Viens no pašiem populārākajiem masu saziņas līdzekļiem ir **internets**, ko mēs visi lietojam, taču patiesībā vēl joprojām nav nekādu reālu iespēju šo saziņas līdzekli kontrolēt

... .gadā, pieņemot šo likumu, mēs, tā sacīt, aizskrējām nedaudz laikiem priekšā un acīmredzot pieņemām, ka jau 2011.-2012.gadā visās mājās būs **internets** un visi cilvēki būs apmācīti **interneta** lietošanā, kā arī apguvuši elektronisko deklarēšanas sistēmu izjūt nesamērīgu likumdošanas slogu, jo viņi ir spiesti iegādāties datorus, spiesti iegādāties **internetu** vai rast kādu citu variantu elektroniskās deklarēšanas principa īstenošanai

... . Un attiecībā uz lētu un efektīvu pieejamību jāteic, ka Latvijā nekas nav lēts un lēti pieejams parastam cilvēkam. Pūķa kungs, paskatieties, cik maksā **internets**: 20 stundas ilga lietošana Latvijā maksā 70 dolārus, bet Igaunijā - 17 dolārus

... . Un, ja bērns mācās otrajā maiņā, tas nozīmē, ka viņš var aiziet uz **interneta** salonu, piemēram, pulksten 12. 00 - tajā laikā, kad viņa vienaudži būs mācības Tas, ka pašvaldības policisti ķers mūsu bērnus **interneta** salonos, teiksim tā, dienas laikā, - tas ir nekorekti Šeit ir rakstīts: "Bērnām aizliegts atrasties sabiedriskās vietās, kur pieejams **internets**..." Vai jūs zināt, kur ir pieejams **internets** tagad? Tas ir pieejams, piemēram, mūsu Saeimas bibliotēkā, tas ir pieejams katrā Latvijas bibliotēkā, jau mazu pagastu bibliotēkās ir **internets**. Katrā skolā arī ir **internets**. Vai tas nozīmē, ka bērnam nav jābūt skolā, kurā ir **internets**? Tas ir pilnīgi absurds! Man šķiet, ka cilvēki, kas rakstīja šo priekšlikumu, nesaprot, kas ir **internets**, kas ir skolēns, kas ir bērns. Šis jēdziens attiecas arī uz pusaudžiem, kuriem ir 17 gadu

... . Tagad vairs nebrauc ar zirgiem tik daudz, ar pajūkiem, bet ir jau **internets**, un nav nekur jābrauc, jo var sazināties tīri labi

Meklēšanas sintakse

Noklikšķinot uz rezultāta, tiek parādīts plašāks konteksts.

māja	vārds tiek meklēts dažādos tā locījumos
māja puse	vārdi tiek meklēti dažādos locījumos, ievērojot, ka tie atrodas viena runātāja runā (ne obligāti blakus)
"māja"	vārds tiek meklēts dotajā formā
"mājas puse"	frāze tiek meklēta dotajā formā
med*	meklē visus vārdus, kas sākas ar burtu kopu līdz simbolam *
med???	jautājuma zīme aizvieto jebkuru vienu simbolu
atalgojums ~mediķi	meklē doto, izņemot ar tildi apzīmēto vārdu

Aizvērt

Teksts

internet*

Sēdes tips

kārtējā a

Rezultāts

... ir savairojušas
izsūtīšanas avotu

.....

... Godātie kolēģi

... - Juridiskā bir

... Tatad redziet,
citām valstīm, be

... Augsti godāta
jūnijā vai jūlijā, tā

... Godātie kolēģi! Patiešām ir problēma... Tā ir liela problēma, kas notiek **internetā**. Šis nevēlamais saturs, ko mūsu bērni un jaunieši tur var redzēt, brīžiem ir patiešām šausminošs. **Internetā** var ne tikai pornogrāfiskus materiālus... pornogrāfiska rakstura materiālus redzēt, bet arī vardarbību, smagu vardarbību, tajā skaitā slepkavības. Aktuāla problēma **internetā** ir jauniešu novešana līdz pašnāvībai. Tā ka šis spektrs ir ļoti plašs

... Mums šodien ir problēma ne tikai ar mikrofonu, bet arī ar **internetu**. Man ir prieks, ka tieši pirms kāda laika, lai taupītu papīru, mēs nolēmām, ka neprintēsim likumprojektus un varēsīm strādāt ar dokumentiem elektroniski, bet šodien pusstundas laikā ir lielas problēmas un tiešām nevar izmantot dokumentus, par kuriem mēs balsojam

Sēdes tips

kārtējā ārkārtas sesijas

Kārtot pēc

Atbilstības

Dzēst

Vairāk opciju

Meklēt

Rezultāti



Latvijas ceļš

2002/02/28 kārtējā sēde

... ļoti nozīmīgiem likumprojektiem, bet nestrādā
... ntā.

... . Viens no pašiem populārākajiem masu saziņas līdzekļiem ir **internets**, ko mēs visi lietojam, taču patiesībā vēl joprojām nav nekādu reālu iespēju šo saziņas līdzekli kontrolēt

... .gadā, pieņemot šo likumu, mēs, tā sacīt, aizskrējām nedaudz laikiem priekšā un acīmredzot pieņemām, ka jau 2011.-2012.gadā visās mājās būs **internets** un visi cilvēki būs apmācīti **interneta** lietošanā, kā arī apguvuši elektronisko deklarēšanas sistēmu izjūt nesamērīgu likumdošanas slogu, jo viņi ir spiesti iegādāties datorus, spiesti iegādāties **internetu** vai rast kādu citu variantu elektroniskās deklarēšanas principa īstenošanai

... . Un attiecībā uz lētu un efektīvu pieejamību jāteic, ka Latvijā nekas nav lētā un lēti pieejams parastam cilvēkam. Puķa kungs, paskatieties, cik maksā **internets**: 20 stundas ilga lietošana Latvijā maksā 70 dolārus, bet Igaunijā - 17 dolārus

... . Un, ja bērns mācās otrajā maiņā, tas nozīmē, ka viņš var aiziet uz **interneta** salonu, piemēram, pulksten 12. 00 - tajā laikā, kad viņa vienaudži būs mācībās Tas, ka pašvaldības policisti ķers mūsu bērnus **interneta** salonos, teiksim tā, dienas laikā, - tas ir nekorekti Šeit ir rakstīts: "Bērnā aizliegts atrasties sabiedriskās vietās, kur pieejams **internets**..." Vai jūs zināt, kur ir pieejams **internets** tagad? Tas ir pieejams, piemēram, mūsu Saeimas bibliotēkā, tas ir pieejams katrā Latvijas bibliotēkā, jau mazu pagastu bibliotēkās ir **internets**. Katrā skolā arī ir **internets**. Vai tas nozīmē, ka bērnam nav jābūt skolā, kurā ir **internets**? Tas ir pilnīgi absurds! Man šķiet, ka cilvēki, kas rakstīja šo priekšlikumu, nesaprot, kas ir **internets**, kas ir skolēns, kas ir bērns. Šis jēdziens attiecas arī uz pusaudžiem, kuriem ir 17 gadu

... . Tagad vairs nebrauc ar zirgiem tik daudz, ar pajūgiem, bet ir jau **internets**, un nav nekur jābrauc, jo var sazināties tīri labi

... . Un attiecībā uz piedalīšanos... uz parakstīšanos vai kādu citu rīcību tur, **internetā**... Ne jau par valsts valodu ir runa! Vai jums ir zināms, cik daudziem Latvijas iedzīvotājiem ir pieejami datori? (No zāles dep Čepāne: „70 procentiem!”) Es ļoti gribu cerēt, ka jūsu optimistiskās prognozes piepildīsies un ka platjoslas **internets** būs pieejams visur, bet es ceru, ka jums ir zināmi arī dati par nabadzības līmeni, par cilvēku izglītības līmeni un par to, cik viegli ir apgūt **internetu**. Vai mēs tiešām varam apgalvot, ka visi pilsoņi, visi 100 procenti, būs līdztiesīgi, ka visiem būs līdztiesīga pieeja un piekļuve **internetam**, kā arī nepieciešamās

Statistika

Atrasti 17 no 303 818 (0.01%) fragmentiem laika posmā no līdz .

Sadalījums pēc dzimuma

Dzimums	Skaitis	Grupas lielums	Procenti no rezultātu skaita ⓘ	Procenti no kategorijas kopskaita ⓘ
Vīrieši	7	174 290	41.18%	0.00%
Sievietes	4	37 249	23.53%	0.01%

Sadalījums pēc grupas

Sadalījums pēc runātāja

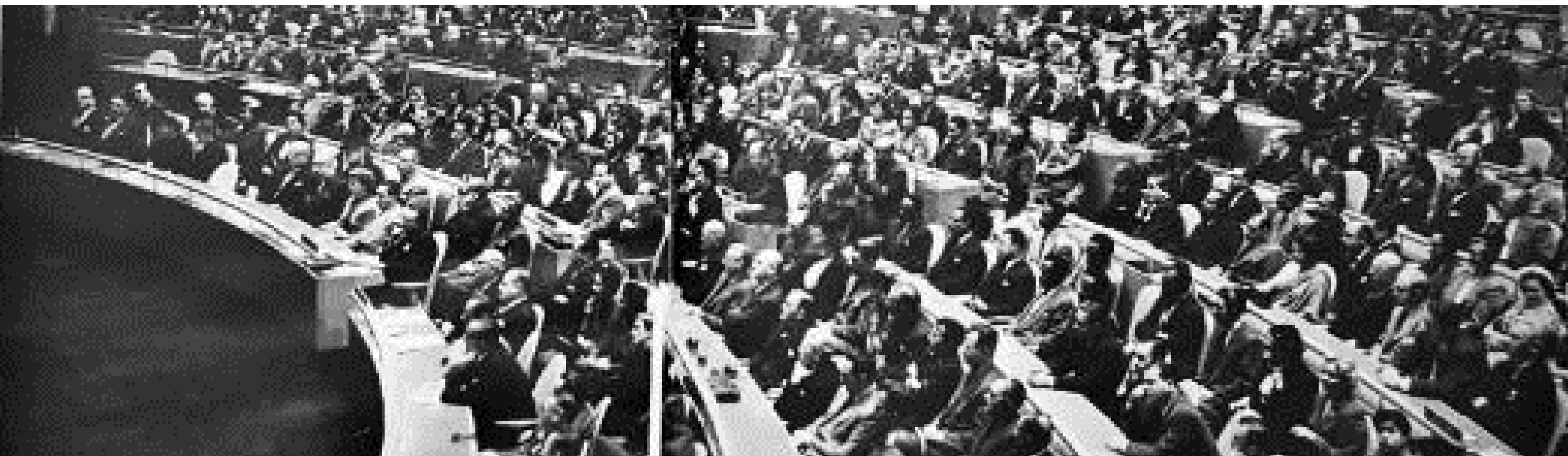
Runātājs	Skaitis	Grupas lielums	Procenti no rezultātu skaita ⓘ	Procenti no kategorijas kopskaita ⓘ
Ilma Čepāne (1947)	1	1 055	5.88%	0.09%
Ingrīda Circene (1956)	1	1 133	5.88%	0.09%
Anna Seile (1939)	1	3 239	5.88%	0.03%
	1	334	5.88%	0.30%

The Index of Internet Aggressiveness

<http://barometrs.korpuss.lv/>

Saeima (Parliament) Corpus Analysis

<http://saeima.korpuss.lv/>



Thank you!
Paldies!

