

Tekstynas **STENOGRAMOS_INDV** skirtas individualių autorių autorystės nustatymo tyrimams.

Tekstyną sudaro parlamentarų pasisakymai Seime
Laikotarpis: 1990 m. kovo mėn. – 2013 m. gruodžio mėn.
Aurių kiekis: 147 autoriai
Minimalus žodžių kiekis tekste: 100 žodžių

Eilutės kiekviename iš tekstų atitinka teksto savybes, naudojamas autorystės nustatymo tyrimuose (daugiau apie savybes skaitykite [1]):

- 1) vidutinis sakinio ilgis tekste
- 2) vidutinis žodžio ilgis tekste
- 3) skirtingų žodžių ir visų žodžių tekste santykis
- 4) tekstas
- 5) lemuotas tekstas (lemuota automatinio įrankiu "Lemuoklis"[2]: atpažįstamas žodis verčiamas lema, bendrinis žodis taip pat verčiamas mažosiomis raidėmis; jeigu žodis neatpažįstamas – paliekama neliesta jo forma)
- 6) kalbos dalys (nustatytos "Lemuoklio" pagalba)
- 7) priklausomybių gramatikos sintaksinės žymos kiekvienam iš žodžių (apdorota automatinio įrankiu MaltParser¹, prieš tai jį apmokius su lietuviškuoju medžių banku² bei didžiausią tikslumą duodančiais parametrais [3])
- 8) funkciniai žodžiai (jungtukai, jaustukai, dalelytės, prielinksniai, įvardžiai: nustatyti automatinio lemavimo įrankiu "Lemuoklis")
- 9) simbolių 2-gramos (tekstinio dokumento mastu). Tekstas suskaidomas į elementus po du simbolius (pritaikius slenkančio lango metodą per vieną simbolį), tarpo simboliai prieš tai keičiami "_". Pvz.: "visas_tekstas": "vi", "is", "sa", "as", "s_", "_t", "te" ir t.t.
- 10) simbolių 3-gramos (tekstinio dokumento mastu)
- 11) simbolių 4-gramos (tekstinio dokumento mastu)
- 12) simbolių 5-gramos (tekstinio dokumento mastu)
- 13) simbolių 6-gramos (tekstinio dokumento mastu)
- 14) simbolių 7-gramos (tekstinio dokumento mastu)
- 15) tekstas junginiais: žodžiais ir jų kalbos dalimis
- 16) tekstas junginiais: lemomis ir žodžių kalbos dalimis
- 17) tekstas junginiais: žodžiais, jų kalbos dalimis bei kita morfologine informacija (morfologinė informacija, pvz. linksnis, skaičius, laipsnis, laikas ir kt. nustatyta "Lemuoklio" pagalba)
- 18) tekstas junginiais: lemomis, žodžių kalbos dalimis bei kita morfologine informacija

Nuorodos:

[1] Kapočiūtė-Dzikiienė, Jurgita; Utkā, Andrius; Šarkutė, Ligita. 2014. Seimo posėdžių stenogramų tekstynas autorystės nustatymo bei autoriaus profilio sudarymo tyrimams. *Kalbotyra*, 66: 27–45.

[2] Zinkevičius, Vytautas. 2000. Lemuoklis – morfologinei analizei. *Gudaitis, L. (red.) Darbai ir Dienos*, 24: 246–273.

[3] Kapočiūtė-Dzikiienė, Jurgita; Nivre, Joakim; Krupavičius, Algis. 2013. Lithuanian Dependency Parsing with Rich Morphological Features. *Empirical Methods in Natural Language Processing – 4th Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL'2013)*, psl. 12–21.

¹ Įrankis MaltParser parsisiųstas iš <http://www.maltparser.org/>.

² Medžių bankas sukurtas VDU vykdyto Lietuvos valstybinio mokslo ir studijų fondo Lituaniistikos mokslinių tyrimų prioriteto įgyvendinimo 2007–2008 metų programos projekto "Internetiniai ištekliai: anototas lietuvių kalbos tekstynas ir anotavimo priemonės (ALKA 2)" metu.

Statistika

klasė	pavyzdžių kiekis	žodžių kiekis	skirtingų žodžių kiekis	vid. pvz. ilgis žodžiais
ABIŠALA_A_A	873	181729	19494	208.17
ALBERTYNAS_A	229	58229	12842	254.28
ALEKNAITĖ-ABRAMIKIENĖ_V	390	90152	16610	231.16
ANDRIUKAITIS_V_P	3932	926247	54282	235.57
ANTANAVIČIUS_K	981	286707	28937	292.26
ASTRAUSKAS_V	203	40522	8366	199.62
AUŠTREVICIUS_P	619	125572	19699	202.86
AŽUBALIS_A	207	47941	10876	231.60
BALEŽENTIS_A	570	95483	13859	167.51
BALTRAITIENĖ_V	418	79584	11711	190.39
BALČYTIS_Z	528	123344	16233	233.61
BASKAS_A	430	101262	15797	235.49
BEINORTAS_J	1105	244656	28032	221.41
BEKINTIENĖ_D	243	49000	9504	201.65
BENDINSKAS_A	238	40153	8436	168.71
BERNATONIS_J	1223	227944	20213	186.38
BIČKAUSKAS_E	704	128212	16084	182.12
BLINKEVIČIUTĖ_V	774	192638	18070	248.89
BOBELIS_K	322	72530	12981	225.25
BOGUŠIS_V	275	54305	11865	197.47
BRADAUSKAS_B	695	131421	19638	189.09
BRAZAUSKAS_A_M	826	305606	30658	369.98
BURBIENĖ_S	506	118664	15598	234.51
BUTKEVIČIUS_A	761	208524	24655	274.01
DAGYS_R_J	1934	406769	35046	210.33
DAUKŠYS_K	796	161541	20576	202.94
DEGUTIENĖ_I	2491	484498	29686	194.50
EINORIS_V	273	55133	11114	201.95
ENDRIUKAITIS_A	543	121442	23024	223.65
ENDZINAS_A	281	44689	9811	159.04
GAPŠYS_V	229	49108	9902	214.45
GEDVILAS_V	1131	213084	16088	188.40
GENTVILAS_E	1221	231640	20337	189.71
GLAVECKAS_K	764	213584	22730	279.56
GRAKAUSKAS_E	308	80593	12061	261.67
GRAUŽINIENĖ_L	862	164964	17815	191.37
GRAŽULIS_P	1442	287855	30870	199.62
GRICIUS_A	485	93202	14435	192.17
GRUBLIAUSKAS_V	211	44177	9612	209.37
GYLYS_P	320	79762	15396	249.26
JACKŪNAS_Ž_J	276	81894	12209	296.72
JAKUČIONIS_P	389	84861	15535	218.15
JANKAUSKAS_D	468	95453	12981	203.96
JARAŠIŪNAS_E	277	77521	12204	279.86
JASKELEVIČIUS_L_K	274	68862	12387	251.32
JUKNEVIČIENĖ_R	660	147449	21607	223.41
JUKNEVIČIUS_Z	512	106924	15400	208.84
JURŠENAS_Č	7779	1454762	52595	187.01
KAKTYS_S	210	49687	10293	236.60
KAROSAS_J	354	77662	14287	219.38
KATKUS_J_A	253	54071	10503	213.72
KAŠĖTA_A	442	80757	13048	182.71
KIRKILAS_G	575	164584	21007	286.23
KLUMBYS_E	1252	264553	27781	211.30
KUBILIUS_A	5093	1076084	53593	211.29
KUNEVIČIENĖ_E_J	707	172178	19205	243.53
KUNČINAS_A	223	44472	8930	199.43
KUPČINSKAS_R	244	40825	9129	167.32
LANDSBERGIS_V	1732	595325	53521	343.72
LAPINSKAS_K	287	103365	13632	360.16
LINKEVIČIUS_L_A	285	65164	11171	228.65
LIONGINAS_J	366	79571	11924	217.41
LISTAVIČIUS_J	1402	271738	24894	193.82

MALKEVIČIUS_S	344	89696	15839	260.74
MARGEVIČIENĖ_V_V	284	45360	9043	159.72
MASIULIS_E	930	187508	23354	201.62
MASIULIS_K	320	68581	14002	214.32
MATULAS_A	910	185155	20558	203.47
MATULEVIČIUS_A	729	151747	20479	208.16
MAZURONIS_A	289	53754	11066	186.00
MAZURONIS_V	560	119346	17819	213.12
MIKUTIENĖ_D	427	81991	13900	192.02
MILČIUS_L	362	96203	16903	265.75
MIŠKINIS_P_A	314	71143	14422	226.57
MONKEVIČIUS_A	253	60799	10857	240.31
MUNTIANAS_V	2241	384398	22249	171.53
NAVICKAS_V	314	82654	12789	263.23
OLEKAS_J	1614	317478	32332	196.70
OZOLAS_R	216	58654	12862	271.55
PALAITIS_R	347	67820	12554	195.45
PANGONIS_J	277	59772	10741	215.78
PAPOVAS_P	697	155169	14840	222.62
PATACKAS_A_V	250	57002	13576	228.01
PAULAUSKAS_A	1808	361556	32174	199.98
PAUŽA_B	203	41827	8273	206.04
PAVIRŽIS_G_A	492	124205	18595	252.45
PEKELIŪNAS_A	277	40863	5542	147.52
PETRAUSKAS_V	448	102260	15435	228.26
PEČELIŪNAS_S	2430	458652	35146	188.75
PRAPIESTIS_J	484	118672	15444	245.19
PRONCKUS_M	524	133542	18518	254.85
PRUNSKIENĖ_K_D	971	292414	31034	301.15
PUPINIS_E	559	97951	14553	175.23
RAMONAS_J	201	51177	9427	254.61
RASIMAVIČIUS_L_N	207	52116	10228	251.77
RAZMA_J	1719	316059	32168	183.86
RAŠKINIS_A_J	287	60240	12012	209.90
RINKEVIČIUS_V	279	51047	9921	182.96
RUDYS_A	704	189548	22065	269.24
RUDZYS_R	316	60426	12165	191.22
RUPEIKA_B_V	305	62829	14106	206.00
SABATAUSKAS_J	840	165027	19158	196.46
SABUTIS_L	1142	241574	21457	211.54
SADECKAS_A	228	49780	9385	218.33
SAKALAS_A	2024	400853	32606	198.05
SALAMAKINAS_A	553	93212	14095	168.56
SAULIS_V	233	42663	8763	183.10
SKARDŽIUS_A	1298	211002	15637	162.56
SMETONA_R	366	81533	13586	222.77
STANCIKIENĖ_A	270	52548	10632	194.62
STANKEVIČIUS_Č_V	880	197587	22631	224.53
STARKEVIČIUS_K	243	48846	10502	201.01
STASIŠKIS_A_N	601	137910	19982	229.47
STEPONAVIČIUS_G	2539	456598	28934	179.83
STOMA_S	200	41199	9294	206.00
STUNDYS_V	360	66216	12259	183.93
SYSAS_A	2130	451444	34044	211.95
TAMULIS_J	241	57118	9916	237.00
TAURANTAS_A	846	178287	16803	210.74
TREINYS_M	304	109141	17361	359.02
VAGNORIUS_G	1237	517318	36131	418.20
VAIŠNORAS_A	455	100213	16098	220.25
VALINSKAS_A	511	88656	10740	173.50
VARAŠKA_M	209	47556	10198	227.54
VESELKA_J	3018	672644	51589	222.88
VIDŽIŪNAS_A	2654	455756	27287	171.72
VISOKAVIČIENĖ_B_T	255	65582	9934	257.18
VĖSAITĖ_B	503	102037	17538	202.86

ZASČIURINSKAS M	594	114568	15924	192.88
ZIMNICKAS_V_V	308	58160	9540	188.83
ZINGERIS_E	262	63668	13557	243.01
ČAPLIKAS_A	779	165450	18501	212.39
ČEKUOLIS_J	413	67995	12763	164.64
ČEPAS_V	408	81582	16714	199.96
ČIGRIEJENĖ_V_M	277	48416	11429	174.79
ŠEDBARAS_S	542	111373	14048	205.49
ŠEDŽIUS_A	260	50452	9711	194.05
ŠIAULIENĖ_I	354	83213	15835	235.06
ŠILGALIS_Ž	300	63785	10505	212.62
ŠIMAŠIUS_R	332	80244	13210	241.70
ŠIMĖNAS_A	474	157179	18532	331.60
ŠIMĖNAS_J	541	120778	18069	223.25
ŠLIČYTĖ_Z	335	86908	14469	259.43
ŠUKYS_R	1106	259971	23437	235.06
ŽAKARIS_E	273	49315	8234	180.64
ŽIEMELIS_V	601	130369	19100	216.92
ŽUKAUSKAS_H	314	52734	9672	167.94
Visi	110908	23908302	279494	215.57