

Kalbos dalių ir gramatinių formų pasiskirstymas lietuvių kalbos tekstynuose. Morfologiškai anotuoti tekstynai

Doc. dr. Erika Rimkutė

Vytauto Didžiojo universitetas

Lituanistikos katedra

Kompiuterinės lingvistikos centras

erika.rimkute@vdu.lt

11 kalbos dalių:

daiktavardžiai,
būdvardžiai,
skaitvardžiai,
įvardžiai,
veiksmažodžiai,
prieveiksmiai,
prielinksniai,
jungtukai,
dalelytės,
jaustukai ir ištiktukai.

Kalbos dalių pasiskirstymas

Kalbos dalis	Sakytinė k.	Rašytinė k.
Daiktavardis	16,2	39,37
Būdvardis	2,8	7,33
Įvardis	16,9	8,7
Skaitvardis	1,9	0,96
Veiksmažodis	22,8	20,5
Prieveiksmis	10,2	6,72
Prielinksnis	4,2	4,65
Dalelytė	12	1,97
Jungtukas	8,2	7,62
Ištiktukas, jaustukas	4,8	0,2

Kalbos dalys	Administracinis stilius	Grožinis stilius	Mokslinis stilius	Publicistinis stilius
Daiktavardžiai	45,14	26,2	42,6	39,76
Veiksmažodžiai	19,14	24,69	19,1	20,45
Būdvardžiai	6,8	5,56	9,68	7,2
Skaitvardžiai	0,7	0,73	0,67	1,12
Įvardžiai	7,36	13,16	7,82	8,25
Prieveiksmiai	3,9	10,39	5,12	6,54
Dalelytės	1,42	4,16	1,42	2,13
Jungtukai	7,76	8,83	7,69	7,37
Jaustukai	0,37	0,37	0,02	0,075
Ištiktukai	–	0,051	0,0008	0,013
Prielinksniai	4,28	5,77	3,89	4,8

Linksnis	Dažnumas rašytinėje k.	Dažnumas sakytinėje k.
Vardininkas	26,38	vns. 31,6, dgs. 30
Kilmininkas	40,88	vns. 22,9, dgs. 33,5
Naudininkas	4,07	vns. 2,6, dgs. 2,3
Galininkas	16,49	vns. 26, dgs. 22,8
Įnagininkas	6,25	vns. 5, dgs. 9
Vietininkas	5,37	vns. 8,6, dgs. 2,1
Šauksmininkas	0,48	vns. 3,3, dgs. 0,3
Iliatyvas	0,08	

Veiksmažodžių formų pasiskirstymas

Formos	Rašytinė kalba	Sakytinė kalba
Asmenuojamosios	50,7	82,1
Dalyviai	27,7	5,4
Padalyviai	3,2	0,4
Pusdalyviai	1,8	0,2
Bendratys	16,6	11,9

Giminės kategorijos pasiskirstymas

vyriškoji 57,2 proc.,

moteriškoji 36,6 proc.

Skaičiaus kategorijos pasiskirstymas

vienaskaita 65 proc.

daugiskaita 32 proc.

Laipsnio kategorijos pasiskirstymas

nelyginamasis 87,5 proc.,

aukštesnysis 8 proc.,

aukščiausiasis – 4,5 proc.

Nuosakos kategorijos pasiskirstymas

tiesioginė 91,5 proc.,

liepiamoji 2,7 proc.,

tariamoji 5,8 proc.

Laiko kategorijas pasiskirstymas

esamasis 52,8 proc.,

būtašis k. 31,2 proc.,

būtašis d. 1,1 proc.,

būsimasis 4,9 proc.

Asmens kategorijos pasiskirstymas

pirmasis 12,3 proc.,

antrasis 6,4 proc.

trečiasis 81,3 proc.

Gramatinių formų vartoseną

Rašytinės kalbos tekstyne vidutiniškai vartojama kaitybinių formų:

- 2,54 daiktavardžių,
 - 2,63 veiksmažodžių,
 - 8,23 įvardžių,
 - 4,72 būdvardžių,
 - 3,55 skaitvardžių,
 - 1,52 prieveiksmių.
-
- Vidutiniškai vartojama 2,34 kaitybinės formos visų kalbos dalių.

Morfologiniai anotatoriai

- <http://tekstynas.vdu.lt/page.xhtml?id=morphological-annotator>
- <http://semantika.lt> (paslauga *Lietuviško teksto analizė ir taisymas*)

Morfologiškai anotuoti tekstynai

- Morfologiškai anototas lietuvių kalbos tekstynas (1,6 mln. žodžių)
<https://clarin.vdu.lt/xmlui/handle/20.500.11821/9>.
- <http://158.129.51.247:8080/annis-gui-3.4.4/>
- Morfologiškai anototas lietuvių kalbos tekstynas (208 mln. žodžių; su paieškos sistema) <http://corpus.vdu.lt/lt/>
- Sakytinės lietuvių kalbos tekstynas (226 tūkst. žodžių)
<http://donelaitis.vdu.lt/sakytines-kalbos-tekstynas/>.

Morfologinės pažymos <http://corpus.vdu.lt/lt> (1)

- Morfologinės pažymos sudarytos remiantis MULTEXT-East (*Multilingual Text Tools and Corpora for Central and Eastern European Languages*) standartu.
- Morfologinių pažymų kodai suderinami su kitų kalbų gramatinių rinkinių žymėjimais (pvz., *noun* – n, *verb* – v, *feminine gender* – f).
- Įvesti keli papildomi lietuviškai gramatikai būdingi kodai (pvz., *būdinys* – b).

Morfologinės pažymos <http://corpus.vdu.lt/lt> (2)

- Kiekviena lietuvių kalbos dalis turi skirtingą morfologinių kategorijų skaičių (nuo 2 iki 14).
- Net ir tos pačios kalbos dalies gramatinės formos dažnai turi skirtingą morfologinių kategorijų skaičių.
- Tos pačios kalbos dalies gramatinėms formoms taikomos vienodo ilgio pažymos.

Morfologinės pažymos <http://corpus.vdu.lt/lt> (3)

- *vertinimų*: lema „vertinimas“, Ncmpgn-;
- *turi*: lema „turėti“, Vgmp3s--n--ni-;
- *nebūti*: lema „nebūti“, Vgi-----y--n--;
- *nesiprausiančiojo*: lema „nesiprausti“, Vgpp-smayygy--;
- *bendresnio*: lema „bendras“, Agcmsgn;
- *ant*: lema „ant“, Sgg;
- *bei*: lema „bei“, Cg;
- *vieno*: lema „vienas“, Mcmsgdn;
- *pvz.*: lema „pvz.“, Ys.

Morfologinių pažymų kombinacijos

- Lietuvių kalboje apie 1500 morfologinių pažymų variantų (60 mln. žodžių tekstyne), pvz.:

dkt. mot. g. vns. K.

dkt. mot. g. dgs. V.

vks. teig. nesngr. tiesiog. n. būt. k. l. vns. 1 asm.

vks. teig. sngr. tiesiog. n. būt. k. l. vns. 1 asm.

vks. teig. nesngr. tiesiog. n. būs. l. vns. 1 asm.

vks. teig. nesngr. tiesiog. n. būs. l. dgs. 1 asm.

ANNIS (ANNIS Corpus Search) | 158.129.51.247:8080/annis-gui-3.4.4/#_q=bGVtbWE9IsSvc2FreW1hcyI&c=TUFUQVNFZG9rdW1lbnRhaQ&cl=10&cr=10&s=0&l=10 | 110% | leškoti | not logged in | Login

Help/Examples | Query Result | Base text | Token Annotations | 1 / 13 | Displaying Results 1 - 10 of 126 | Result for: lemma="įsakymas"

lemma="įsakymas"

Query Builder

Search | More | History

126 matches in 21 documents

Corpus List | Search Options

Visible: All

Name	Texts	Tokens		
Alksnis_paula_1.0	63	39 754		
MATAS_dokumentai	39	420 206		
MATAS_grozine_literatura	9	319 958		
MATAS_moksliniai_tekstai	13	516 903		
MATAS_periodika	30	649 501		
pcc2	2	399		

1 Path: MATAS_dokumentai > 2007_penkioliktasis_seimo_posedis (tokens 1993 - 2013) left context: 10 right context: 10

inės vartotojų teisių apsaugos tarnybos vadovo įsakymu , dalyvavo įvairių

~POS.~NDEF.F.SG.GEN. .M.PL.GEN. .F.PL.GEN. .F.SG.GEN. .F.SG.GEN. .M.SG.GEN. .M.SG.INS. .NNEG.IND.PST.DEF.PL.3. .NNEG.~POS.~NDEF.F.PL.GEN.

nis vartotojas teisė apsauga tarnyba vadovas įsakymas , dalyvauti(-uja,-vo) įvairus

NOUN NOUN NOUN NOUN NOUN NOUN NOUN PUNCT VERB ADJ

NOUN NOUN NOUN NOUN NOUN NOUN PUNCT VERB ADJ

2 Path: MATAS_dokumentai > 8012 (tokens 14027 - 14047) left context: 10 right context: 10

 žemės ir miškų ūkio ministro įsakymu ir skelbiami Lietuvos

ASS.PRS.~NDEF.M.PL.NOM. .F.SG.GEN. .M.PL.GEN. .M.SG.GEN. .M.SG.GEN. .M.SG.INS. .PTCP.NNEG.PASS.PRS.~NDEF.M.PL.NOM. .F.SG.GEN. .F.

 žemė ir miškas ūkis ministras įsakymas ir skelbti(-ia,-ė) Lietuva

NOUN CONJ NOUN NOUN NOUN NOUN NOUN CONJ VERB PROPN

NOUN CCONJ NOUN NOUN NOUN NOUN NOUN CCONJ VERB PROPN

3 Path: MATAS_dokumentai > 8012 (tokens 34885 - 34905) left context: 10 right context: 10

cijos klausimais ; 9 . 4 . leidžia įsakymus ir įsakymais patvirtintus kitus

.M.PL.INS. .NNEG.IND.PRS.SG.3. .M.PL.ACC. .M.PL.INS. .PTCP.NNEG.PASS.PST.~NDEF.M.PL.ACC. .M.PL.

cija klausimas ; 9 . 4 . leisti(-džia,-do) įsakymas ir įsakymas patvirtinti(-a,-o) kitas

NOUN PUNCT X NUM PUNCT NUM PUNCT VERB NOUN CONJ NOUN VERB PRON

NOUN PUNCT X NUM PUNCT NUM PUNCT VERB NOUN CCONJ NOUN VERB PRON

4 Path: MATAS_dokumentai > 8012 (tokens 34887 - 34907) left context: 10 right context: 10

klausimais ; 9 . 4 . leidžia įsakymus ir įsakymais patvirtintus kitus

.M.PL.INS. .NNEG.IND.PRS.SG.3. .M.PL.ACC. .M.PL.INS. .PTCP.NNEG.PASS.PST.~NDEF.M.PL.ACC. .M.PL.ACC.

klausimas ; 9 . 4 . leisti(-džia,-do) įsakymas ir įsakymas patvirtinti(-a,-o) kitas

NOUN PUNCT X NUM PUNCT NUM PUNCT VERB NOUN CONJ NOUN VERB PRON

NOUN PUNCT X NUM PUNCT NUM PUNCT VERB NOUN CCONJ NOUN VERB PRON

Lemma="įsakymas"

Query Builder

Search More History

9 matches in 5 documents

Corpus List Search Options

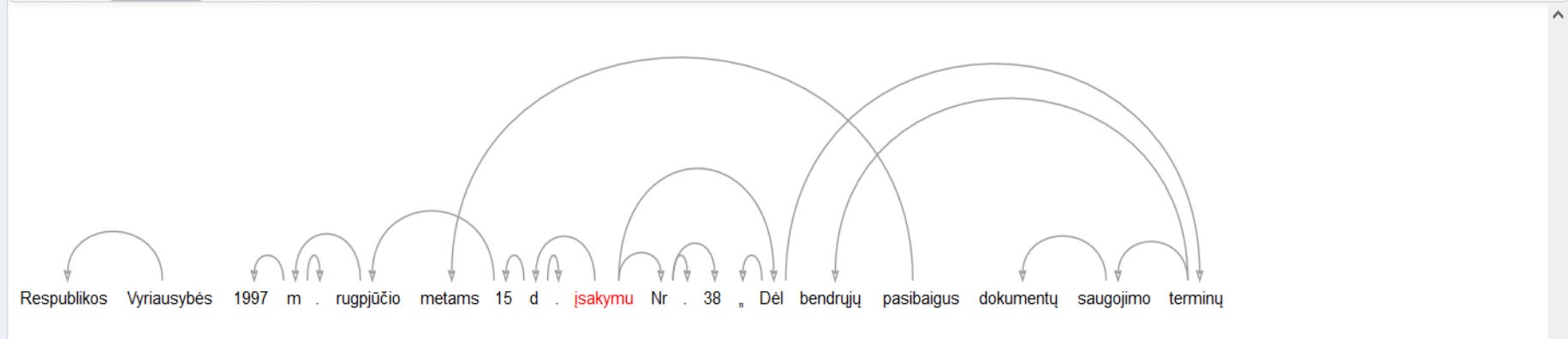
Visible: All

Name	Texts	Tokens		
Alksnis_paula_1.0	63	39 754	i	📄
MATAS_dokumentai	39	420 206	i	📄
MATAS_grozine_literatura	9	319 958	i	📄
MATAS_moksliniai_tekstai	13	516 903	i	📄
MATAS_periodika	30	649 501	i	📄
pcc2	2	399	i	📄

Help/Examples Query Result

Base text Token Annotations

1 / 1 Displaying Results 1 - 9 of 9 Result for: lemma="įsakymas"



9 Path: Alksnis_paula_1.0 > general_periodicals > tb1-10_V2 (tokens 136 - 156) left context: 10 right context: 10

atšauktų	daug	diskusijų	visuomenėje	sukėlusį	įsakymą	,	nurodantį	senesnių
M.-NDEF	.3.SG.COND.NNEG	.-POS	.F.PL.GEN	.F.SG.LOC	.PTCP.PST.DEF.SG.M.-ACT.NNEG.ACC	.M.SG.ACC	.PTCP.PRS.SG.M.-ACT.NNEG.ACC	.M.PL.GE
atšaukti	daug	diskusija	visuomenė	sukelti	įsakymas	,	nurodyti	senas
VERB	ADV	NOUN	NOUN	VERB	NOUN	PUN	ADJ	ADJ
Pred_Adj	Adj	Atr	Adj	Atr	Obj	Aux	Atr	Atr

deps (tb1-10_V2)



Paieškos galimybės per ANNIS sistemą

- lemma="kalba" (ieškoma lemos *kalba*);
- pos="ADJ" (ieškoma būdvardžių);
- gram=".M.SG.GEN." (ieškoma vienskaitos vyriškąja gimine kilmininko forma pavartotų žodžių);
- tok="pasakė" arba "pasakė";
- syfun="Atr" (ieškoma atributų).

Nuorodos

- *Dabartinės lietuvių kalbos tekstynas* (neanotuotas)
<http://tekstynas.vdu.lt/tekstynas/>.
- I. Dabašinskienės straipsnis *Šnekamosios lietuvių kalbos morfologinės ypatybės*. *Acta Linguistica Lithuanica*, LX (2009), p. 1–15.
<http://etalpykla.lituanistikadb.lt/fedora/get/LT-LDB-0001:J.04~2009~1367167159055/DS.002.0.01.ARTIC>
- E. Rimkutės disertacijos *Morfologinio daugiareikšmiškumo ribojimas kompiuteriniame tekstyne*, 2006, 2.4 poskyris.
http://fcim.vdu.lt/~erika_rimkute/straipsniai/disertacija.pdf.
- Lithuanian Treebank ALKSNIS
<https://clarin.vdu.lt/xmlui/handle/20.500.11821/10>