

## 2013. korpuste automaatse märgendamise väljundi analüüs

### Koondkorpuse reeglipõhise morfoloogilise ühestamise tulemused

2013. aastal on reeglipõhise (CG) morfoloogilise ühestajaga märgendatud Koondkorpusest tehtud uus versioon (v.a. uue meedia allosa ja seadusetekstid, saadaval projekti kodulehel portaalis [www.keeletehnoloogia.ee](http://www.keeletehnoloogia.ee)).

Uuendusena kasutati ühestamise sisendina morfoloogiliselt analüüsitud ja oletatavate pärisnimede algvormide suhtes eelanalüüsitud tekstid, kus teksti piires sisalduvate pärisnimede ja nende pakutud algvormide põhjal on mõnede morfoloogilise analüsaatori pakutud rohkete analüüsitud hulgast valitud iga teksti konteksti sobivad algvormid.

Kuna analüüsi algstaadiumil tehtud vead mõjutavad ka kõiki järgnevaid analüüsietappe, põhjustades omakorda pindsüntaktilisi ja sõltuvusseoste vigu, on morfoloogilise ühestamise reegleid jätkuvalt täiendatud kasutades 94-tuhandesõnalist treeningkorpust. Morfoloogilise kirjelduse saagis treeningkorpusel on 97.2%, täpsus 90.0%, vigu on 2.8%, mitmesus 6.3%.

Morfoloogilise kirjelduse saagis 27-tuhandesõnalisel testkorpusel on 97.1%, täpsus 90.4%, vigu on 2.9%, mitmesus 6.2%. Võrdluseks morfoloogiliselt analüüsitud, kuid ühestamata testkorpuse morfoloogilise kirjelduse saagis on 99.5%, täpsus 35.5%, vigu on 0.5%, mitmesus 57.7%. Kui me vaatleme morfoloogilise ühestamise nõ „omavigu“, st lahutades juba morfoloogilise analüsaatori väljundis puuduvad käsitsi ühestatud analüüsitud, siis reeglipõhise ühestaja omavigu on 2.4%.

Morfoloogilise ja süntaktilise kirjelduse vigade arvutamisel ei loetud vigadeks selliseid erinevusi, kus käsitsi märgendatud tekstis oli süntaktiliseks funktsiooniks märgitud @??, need on peamiselt kasutatud võõrkeelsetes fraasides. Samuti ei loetud vigadeks üldnime (common noun) ja pärisnime (proper noun) märgendi erinevused, kui morfoloogilise vormi kirjeldus langes kokku.

Tabel 1. Morfoloogilise reeglipõhise ühestaja treenimisel ning testimisel kasutatud tasakaalus korpuse ilu-, ajakirjanduse ning teadus tekstid.

Fail	Sõnade arv
Ilu_kanep	7515
Ilu_kivirahk	7980
Ilu_sauter	7532
Ilu_valton_kana	8174
Ilu_vilep_lohe	5935
Aja_EPL_2007_08_12	2293
Aja_pm_2000_10_04	17454
Aja_slõleht_2007_12_17	12798
Tea_dr8020	8128

Tea_dr8047	7651
Tea_geofyysika	8766
Treeningkorpus kokku	94226
Testkorpus: aja_ml_2002_47	26727

Tabelites 2 ja 3 on toodud morfoloogilise ühestamise tulemused treening- ja testkorpuses. Täpsus näitab korrektsete tõlgenduste osakaalu tekstis kõikide allesjäänud tõlgenduste hulgas, mida lähem on see 100%-le, seda ühesem on väljund. Saagis näitab, mitu protsenti õigeid tõlgendusi on teksti alles jäänud, mida lähem see on 100%-le, seda vähem korrektseid tõlgendusi on tekstist kustutatud.

Tabel 2. Treeningkorpuse morfoloogilise ühestamise tulemused tasakaalus korpuse valitud ilu-, ajakirjanduse ning teadus tekstides.

	Ühestamata morf. analüüsitud tekst			CG ühestatud tekst		
	Saagis	Täpsus	Algmitmesus	Saagis	Täpsus	Morfoloogiline mitmesus
Treeningkorpus	98.99	35.52	57.44	97.21	90.02	6.28

Juba ühestamata treeningkorpuse saagis on 99%, mis annab treeningkorpuse morfoloogilise ühestamise nn „omavigu“ ainult 1.8%. Sõnaliigi tasemel vigu kõikidest vigadest on 50.7%, seega sõnaliigi täpsusega on korrektne väljund saadud 98.6% sõnadel. Käänevigu on 43.3% kõikidest vigadest ja ülejäänud 6% on eri verbivormide vahelised vead.

Tabel 3. Testkorpuse morfoloogilise ühestamise tulemused.

	Ühestamata morf. analüüsitud tekst			CG ühestatud tekst		
	Saagis	Täpsus	Algmitmesus	Saagis	Täpsus	Morfoloogiline mitmesus
Testkorpus	99.52	35.50	57.67	97.11	90.42	6.15

Morfoloogilise kirjelduse vigu on 2.9%, neist „omavigu“ 2.4%. Sõnaliigi tasemel vigu kõikidest vigadest on 44.1%, seega sõnaliigi täpsusega on korrektne väljund saadud 98.6% sõnadel. Käänevigu on 48.2% kõikidest vigadest ja ülejäänud 7.7% on eri verbivormide vahelised vead.

Tabel 4. Morfoloogilise ühestamise vigade jaotus sõnaliigiti treeningkorpuse ilu-, ajakirjanduse ning teadus tekstides (protsentides kõigi vigade arvu suhtes). Kui nii õige kui vea sõnaliik on sama (peadiagonaali väärtused), siis tegemist on algvormi või morfoloogilise vormi veaga.

	V	S	A	P	N	D	K	J	I	G	Y
--	---	---	---	---	---	---	---	---	---	---	---

Viga Õige											
V	5.78	3.46	1.79	0.15	0.04	0.08	0.11				0.04
S	2.59	37.62	2.85	0.30	0.42	0.84	0.91	0.04	0.19	0.11	1.48
A	4.83	5.33	0.76	0.11		0.61					
P	0.08	1.22	0.04	3.73	1.14	0.19					0.08
N	0.08	0.99		0.91	1.48	0.04					0.04
D	0.34	3.92	0.95	0.46	0.08		2.66	0.76	0.15		0.11
K		0.65	0.04		0.04	1.14					
J		1.64				0.27	0.04				0.08
I	0.08	0.65	0.04			0.57					0.04
G	0.04	0.19									
Y		0.38		0.08	0.08				0.19		
T	0.04	3.39		0.15	0.27	0.08					

Tabel 5. Sõnaliigitähiste seletused.

Sõnaliigitähis	Seletus
V	Tegusõna e verb
S	Nimisõna e substantiiv
A	Omadussõna e adjektiiv
P	Asesõna e pronoomen
N	Arvsõna e numeraal
D	Määrsõna e adverb
K	Kaassõna e pre- v postpositsioon
J	Sidesõna
I	Hüüdsõna
G	Genitiivatribuut

Y	Lühend
T	Tundmatu sõnaliik (võõrsõna)

Tabel 6. Morfoloogilise ühestamise vigade jaotus sõnaliigiti testkorpuse ajakirjandustekstis (protsentides kõigi vigade arvu suhtes). Kui nii õige kui vea sõnaliik on sama (peadiagonaali väärtused), siis tegemist on algvormi või morfoloogilise vormi veaga.

Viga	V	S	A	P	N	D	K	J	Y
Õige									
V	7.77	1.55	2.85				0.78		
S	2.59	44.30	3.63	0.26		1.17	1.30		
A	6.61	3.89	0.26			0.26			
P		1.68		2.98	1.81				
N		0.13		0.78	0.39				0.52
D	0.39	3.50	0.78	1.42			2.46	0.91	
K	0.13	1.17	0.13		0.13	1.55		0.13	
J	0.26	0.13				0.78			
Y		0.13						0.26	
								ET	
T		0.13							

Tabel 7. Sagedasemad vigade liigid treeningkorpuses.

Sagedus	Protsent	Vea liik
219	8.33	Õige nimetav, valitud omastav kääne.
185	7.04	Õige omastav, valitud nimetav kääne.
118	4.49	Õige omastav, valitud osastav kääne.
116	4.41	Õige osastav, valitud omastav kääne.
89	3.39	Õige lühike sisseütlev, valitud osastav kääne.
86	3.27	Mineviku partitsiip: õige omadussõna, valitud verbivorm.

70	2.66	Õige nimetav, valitud osastav kääne.
----	------	--------------------------------------

Tabel 8. Sagedasemad vigade liigid testkorpuses.

Sagedus	Protsent	Vea liik
116	15.03	Õige nimetav, valitud omastav kääne.
44	5.70	Õige omastav, valitud nimetav kääne.
40	5.18	Õige osastav, valitud omastav kääne.
38	4.92	Mineviku partitsiip: õige omadussõna, valitud verbivorm.
31	4.02	Õige omastav, valitud osastav kääne.
31	4.02	Õige lühike sisseütlev, valitud osastav kääne.
23	2.98	Õige nimetav, valitud osastav kääne.

Tabel 9. Viie sagedasema käände vigade jaotus treeningkorpuses (arv, protsent kõikide vigade hulgas).

Viga	Nimetav	Omastav	Osastav	Lühike sisseütlev	Seestütlev
Õige					
Nimetav		219, 8.33	70, 2.66	5, 0.19	2, 0.08
Omastav	185, 7.04		118, 4.49	36, 1.37	
Osastav	55, 2.09	116, 4.41		54, 2.05	3, 0.11
Lühike sisseütlev	5, 0.19	50, 1.90	89, 3.39		
Seestütlev			5, 0.19		

Tabel 10. Viie sagedasema käände vigade jaotus testkorpuses (arv, protsent kõikide vigade hulgas).

Viga	Nimetav	Omastav	Osastav	Lühike sisseütlev	Seestütlev
Õige					
Nimetav		116, 15.03	23, 2.98		
Omastav	44, 5.70		31, 4.02	11, 1.42	
Osastav	18, 2.33	40, 5.18		13, 1.68	1, 0.13

Lühike sisseütlev		14, 1.81	31, 4.02		
Seestütlev	1, 0.13				

### Tasakaaluscorpuse süntaktilise analüüsi tulemused

Reeglipõhise pindsüntaktilise (CG) ja sõltuvussüntaktilise (CG) analüsaatoriga märgendati Tasakaalus korpus (15 miljonit sõna, saadaval projekti kodulehel portaalis [www.keeletehnoloogia.ee](http://www.keeletehnoloogia.ee)).

Märgenduse kvaliteedi mõõtmiseks valitud testkorpus koosnes ajalehetekstist (sama, mis morfoloogilise ühestamise testimisel) kogumahuga 26727 sõna. Süntaktilise kirjelduse saagis on keskmiselt 92.9%, täpsus 69.3%, vigu on 7.1%. Sõltuvussüntaktilise analüüsi tulemuse saagis on tõusnud 62%-st 78.5%-ni.

Kui me vaatleme süntaksianalüüsi nõ „omavigu“, st lahutades juba morfoloogilise ühestamise käigus tehtud vead morfoloogilisel tasandil, siis ilma morfoloogiliselt ühestamata teksti puhul on see 12.3% (kokku 12.8%), reeglipõhise ühestajaga süntaksi omavigu on 4.0% (kokku 7.4%).

Peamised mitmeseks jäävad süntaktiliste funktsioonide kombinatsioonid on: adverbiaal ja nimisõnaline täiend (24%), adverbiaal ja määrsõnaline täiend (14%), objekt ja nimisõnaline täiend (8%) ning finiidne põhiverb ja finiidne abiverb (5%).

Tabel 11 Testcorpuse pindsüntaktilise ning sõltuvussüntaktilise analüüsi tulemused, kus sisendina antud morfoloogiliselt ühestamata ja reeglipõhiselt ühestatud tekst.

	Ühestamata morf. analüüsitud tekst			CG ühestatud tekst		
	Saagis	Täpsus	Sõltuvusseoste saagis	Saagis	Täpsus	Sõltuvusseoste saagis
Testkorpus	87.9	46.8	67.3	92.9	69.3	78.5

Süntaktilistest funktsioonidest kõige enam esineb tekstides adverbiaal ning selle äratundmise protsent on ka keskmisest kõrgem. Enim vigu tehakse subjekti, objekti, nimisõnalise ja omadussõnalise täiendi ja adverbiaali suhtes.

Tabel 12. Süntaktiliste vigade jaotus sõnaliigiti: 1. veerg - sõnaliigi märgend, 2. veerg – sõnaliigi nimetus, 3. veerg – sõnaliigi esinemise arv testkorpuses, 4. veerg – sõnaliigi esinemise protsent testkorpuses, 5. veerg – süntaktiliste vigade arv antud sõnaliigiga sõnadel, 6. veerg – süntaktiliselt vigaselt analüüsitud sõnade protsent antud sõnaliigiga sõnade hulgas ('+' tähendab vigade suhtelist suurenemist võrreldes eelmise aasta tulemustega ja '-' - vähenemist).

Sõnaliik	Seletus	Esineb kordi	Osakaal tekstis %	Süntaksi vigade arv	Vigade % sõnaliigiti
V	Verb	5407	20.23	169	3.13 -
S	nimisõna	10436	39.05	894	8.57 -
A	omadussõna	1920	7.18	172	8.96 -
P	asesõna	2234	8.36	162	7.25 -

N	arvsõna	826	3.09	177	21.43 -
D	määrsõna	3087	11.55	221	7.16 -
K	kaassõna	623	2.33	31	4.98 +
J	sidesõna	1812	6.78	11	0.61 +
I	hüüdsõna	25	0.09	12	48.0 -
G	genitiivatribuut	16	0.06	0	0 -
X	ainult ühendverbi koosseisus esinev määrsõna	8	0.03	3	37.50 +
Y	lühend	332	1.24	37	11.14 -
T	tundmatu sõnaliik	1	0	1	100

Sõnaliigiti suurenes ainult kaassõna (täiendi või adverbiaali valikul), sidesõna ('ehk', 'kui', 'kuid', 'või' - eksliku määrsõna tõlgenduse valiku korral morfoloogilise ühestamise käigus) ning väga harva esinevate ainult ühendverbi koosseisus esinevate määrsõnade ('pärit') süntaktilise funktsiooni määramise vigade hulk, kõikide muude sõnaliikide korral süntaktilise funktsiooni vigade hulk vähenes.

Tabel 13. Süntaktiliste vigade jaotus süntaktilise funktsiooni järgi: 1. veerg – süntaktilise funktsiooni märgend, 2. veerg – süntaktilise funktsiooni nimetus, 3. veerg – süntaktilise funktsiooni esinemise arv testkorpuses, 4. veerg – süntaktilise funktsiooni esinemise protsent testkorpuses, 5. veerg – süntaktiliste vigade arv antud süntaktilise funktsiooniga sõnadel, 6. veerg – süntaktiliste vigade protsent antud süntaktilise funktsiooniga sõnadel kõikide süntaktiliste vigade hulgas ('+' tähendab vigade suhtelist suurenemist võrreldes eelmise aasta tulemustega ja '-' - vähenemist).

Süntaktiline funktsioon	Seletus	Esineb kordi	Osakaal tekstis %	Vigade arv	% kõigist vigadest
@FMV	finiitne öeldis	3161	11.83	40	2.12 -
@IMV	infiniitne öeldis	749	2.80	28	1.48 -
@FCV	finiitne abiverb	545	2.04	13	0.69 -
@ICV	infiniitne abiverb	14	0.05	2	0.11 +
@NEG	verbi eitus	292	1.09	0	0 -
@SUBJ	subjekt	3298	12.34	308	16.3 -
@OBJ	objekt	1848	6.91	235	12.43 -
@PRD	predikatiiv	577	2.16	136	7.20 +
@ADVL	adverbiaal	6670	24.96	303	16.03 +
@INFN>	infinitiiv eestäiendina	1	0.00	0	0
@<INFN	infinitiiv järeltäiendina	81	0.30	11	0.58 +
@P>	eessõnafraasi kuuluv nimisõna	511	1.91	43	2.28 -
@<P	tagasõnafraasi kuuluv nimisõna	130	0.49	33	1.75 -
@NN>	nimisõna eestäiendina	3771	14.11	275	14.55 +
@<NN	nimisõna järeltäiendina	591	2.21	161	8.52 +
@AN>	omadussõna eestäiendina	1715	6.42	161	8.52 +
@<AN	omadussõna järeltäiendina	7	0.03	5	0.26 -
@DN>	määrsõna eestäiendina	156	0.58	45	2.38 +
@<DN	määrsõna järeltäiendina	11	0.04	3	0.16 +
@KN>	kaassõna eestäiendina	6	0.02	0	0 -

@<KN	kaassõna järeltäiendina	57	0.21	7	0.37 -
@<Q	kvantori järellaiend	400	1.50	35	1.85 -
@Q>	kvantori eeslaiend	4	0.01	4	0.21 +
@J	sidend	1813	6.78	12	0.63 +
@B	hüüatus	6	0.02	6	0.32 +
@??	tundmatu, nt võõrkeelne sõna	0	0	0	pole viga
@Vpart	ühendverbi afiksaaladverb	313	1.17	24	1.27 -

Lauseehituse seisukohalt tähtsaimate süntaktiliste funktsioonide (öeldis, subjekt, objekt, kaassõnafraasi kuuluv noomen, kaassõna täiendina, kvantori järellaiend) äratundmine on võrreldes eelmise aasta tulemustega paranenud; mõnevõrra suurenes predikatiivi, adverbiaali j täiendite määramise vigade hulk.

Tabel 14. Eri sõnaliigiga ja süntaktilise funktsiooniga 1) sõnade arv testkorpuses, 2) esinemise protsent testkorpuses ning 3) automaatselt õigesti määratud süntaktilise funktsiooni protsent (saagis).

Süntaktiline funktsioon / sõnaliik	V	S	A	P	N	D	K	J	I	G	Y	X
@FMV	3160 11.82 98.77	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	1 0 0	0 0 0
@IMV	748 2.80 96.39	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	1 0 0	0 0 0
@FCV	545 2.04 97.61	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
@ICV	14 0.05 85.71	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
@NEG	292 1.09 100	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
@SUBJ	141 0.53 86.52	2314 8.66 91.18	8 0.03 12.5	704 2.63 93.9	71 0.27 94.4	19 0.07 0	6 0.02 0	0 0 0	0 0 0	0 0 0	35 0.13 82.9	0 0 0
@OBJ	169 0.63 78.70	1374 5.14 89.88	6 0.02 0	223 0.839 1.9	49 0.18 69.4	18 0.07 5.56	3 0.01 0	0 0 0	0 0 0	0 0 0	6 0.02 83.3	0 0 0
@PRD	27 0.10 77.78	184 0.69 70.11	314 1.17 82.8	24 0.09 70.8	27 0.10 51.8	0 0 0	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
@ADVL	227 0.85 93.83	2573 9.63 95.53	119 0.45 93.3	319 1.19 92.8	213 0.80 95.8	2575 9.63 95.4	550 2.06 97.4	0 0 0	13 0.05 100	0 0 0	75 0.28 98.7	5 0.02 100
@INFN>	1 0 100	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
@<INFN	81 0.30 86.42	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
@P>	0 0 0	381 1.43 95.54	3 0.01 0	91 0.34 93.4	32 0.12 59.4	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	4 0.01 0	0 0 0



@<P	0	107	0	7	15	0	0	0	0	0	1	0
	0	0.40	0	0.03	0.06	0	0	0	0	0	0	0
	0	74.77	0	85.7	73.3	0	0	0	0	0	0	0
@NN>	0	2695	4	830	117	0	0	0	5	0	120	0
	0	10.08	0.01	3.11	0.44	0	0	0	0.02	0	0.45	0
	0	94.51	0	93.4	59.8	0	0	0	0	0	86.7	0
Sõnaliik	V	S	A	P	N	D	K	J	I	G	Y	X
@<NN	0	422	5	33	77	0	0	0	0	0	54	0
	0	1.58	0.02	0.12	0.29	0	0	0	0	0	0.20	0
	0	76.30	20.0	78.8	40.3	0	0	0	0	0	92.6	0
@AN>	0	54	1453	2	190	0	0	0	0	16	0	0
	0	0.20	5.44	0.01	0.71	0	0	0	0	0.06	0	0
	0	0	94.9	50.0	86.3	0	0	0	0	100	0	0
@<AN	0	0	6	1	0	0	0	0	0	0	0	0
	0	0	0.02	0	0	0	0	0	0	0	0	0
	0	0	33.3	0	0	0	0	0	0	0	0	0
@DN>	1	0	0	0	0	151	0	0	1	0	0	3
	0	0	0	0	0	0.56	0	0	0	0	0	0.01
	0	0	0	0	0	73.5	0	0	0	0	0	0
@<DN	0	0	0	0	0	11	0	0	0	0	0	0
	0	0	0	0	0	0.04	0	0	0	0	0	0
	0	0	0	0	0	72.7	0	0	0	0	0	0
@KN>	0	0	0	0	0	0	6	0	0	0	0	0
	0	0	0	0	0	0	0.02	0	0	0	0	0
	0	0	0	0	0	0	100	0	0	0	0	0
@<KN	0	0	0	0	0	0	57	0	0	0	0	0
	0	0	0	0	0	0	0.21	0	0	0	0	0
	0	0	0	0	0	0	87.7	0	0	0	0	0
@<Q	0	1047	5	0	124	0	0	0	0	0	70	0
	0	1.21	0.01	0	0.14	0	0	0	0	0	0.08	0
	0	84.72	0	0	93.5	0	0	0	0	0	2.9	0
@Q>	0	4	0	0	0	0	0	0	0	0	0	0
	0	0.01	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
@J	0	0	0	0	0	0	0	1812	0	0	1	0
	0	0	0	0	0	0	0	6.78	0	0	0	0
	0	0	0	0	0	0	0	99.4	0	0	0	0
@B	0	0	0	0	0	0	0	0	6	0	0	0
	0	0	0	0	0	0	0	0	0.02	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
@??	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0
@Vpart	0	0	0	0	0	313	0	0	0	0	0	0
	0	0	0	0	0	1.17	0	0	0	0	0	0
	0	0	0	0	0	92.3	0	0	0	0	0	0

Tabel 15. Sõltuvusseoste vigade jaotus testkorpuses süntaktilise funktsiooni järgi.

Süntaktiline funktsioon	Seletus	Esineb kordi	Osakaal tekstis %	Vigade arv	% kõigist vigadest
@FMV	finiitne öeldis	3161	11.83	699	12.2
@IMV	infiniitne öeldis	749	2.80	227	4.0
@FCV	finiitne abiverb	545	2.04	29	0.5
@ICV	infiniitne abiverb	14	0.05	3	0.1

@NEG	verbi eitus	292	1.09	0	0
@SUBJ	subjekt	3298	12.34	489	8.5
@OBJ	objekt	1848	6.91	414	7.2
@PRD	predikatiiv	577	2.16	128	2.2
@ADVL	adverbiaal	6670	24.96	1759	30.7
@INFN>	infinitiiv eestäiendina	1	0.00	1	0
@<INFN	infinitiiv järeltäiendina	81	0.30	35	0.6
@P>	eessõnafraasi kuuluv nimisõna	511	1.91	46	0.8
@<P	tagasõnafraasi kuuluv nimisõna	130	0.49	37	0.6
@NN>	nimisõna eestäiendina	3771	14.11	711	12.4
@<NN	nimisõna järeltäiendina	591	2.21	454	7.9
@AN>	omadussõna eestäiendina	1715	6.42	172	3.0
@<AN	omadussõna järeltäiendina	7	0.03	5	0.1
@DN>	määrsõna eestäiendina	156	0.58	110	1.9
@<DN	määrsõna järeltäiendina	11	0.04	10	0.2
@KN>	kaassõna eestäiendina	6	0.02	6	0.1
@<KN	kaassõna järeltäiendina	57	0.21	52	0.9
@<Q	kvantori järellaiend	400	1.50	55	1.0
@Q>	kvantori eeslaiend	4	0.01	2	0
@J	sidend	1813	6.78	280	4.9
@B	hüüatus	6	0.02	5	0.1
@??	tundmatu, nt võõrkeelne sõna	0	0	0	pole viga
@Vpart	ühendverbi afiksaaladverb	313	1.17	8	0.1