

morf2morf.py

**Programm morfoloogilise märgenduse
teisendamiseks**

Sisukord

Skripti üldine kirjeldus.....	3
Otstarve.....	3
Kasutamine.....	3
Asukoht svn-s.....	3
Skripti käsurea parameetrid.....	3
Skripti sisend ja väljund.....	4
Sisendi formaat.....	4
Väljundi formaat.....	4
Standardsisend.....	5
Sisend failide nimekirjana.....	6
Sisend ja väljund -1 režiimis.....	6
Seadete fail.....	8
Asukoht	8
Seadistatavad parameetrid.....	8
Seadete faili formaat.....	9
Sõnastik ja reeglid.....	11
Sõnastiku faili struktuur.....	11
Sõnastiku faili näide.....	11
Reeglid ja prioriteedid.....	11
Skripti veateated ja logi.....	13
Logi näide.....	13
Parameetrite väärtused.....	13
Sõnastikust reeglite lugemine.....	13
Sisendi töötlemine.....	13
Veateated	14
Puuduvad morfoloogilised kategooriad.....	14
(P)Arendamise mõtted.....	16
Reeglite prioriteedid/skoorid käsitsi seadistatavaks.....	16
Lisa sõnastiku võimalus.....	16

Skripti üldine kirjeldus

Otstarve

Skript *morf2morf.py* on mõeldud morfoloogiliste kategooriate teisendamiseks ühest standardist teisse etteantud sõnastiku alusel. Skript on kirjutatud keeles Python ning on testitud Python 2.7 ning 3.5 keskkonnas.

Kasutamine

morf2morf.py [-1] [-c conffile] [-d dictfile] [-a dictfile] [-e fileextension] [-f foldername] [files]

Asukoht svn-s

svn+ssh://svnrpvpkp@imbi.at.mt.ut.ee/svnrpvpkp/trunk/rabauti/morf2morf

Skripti käsurea parameetrid

Parameeter	Kohustuslik	Kirjeldus	Kommentaar
-1	ei	skript töötab režiimis "üks rida korraga"	Vt " Sisend -1 režiimis ".
-c	ei	skripti seadete fail tsv formaadis	Vaikeväärtus on <i>conf/conf.tab</i> .
-d	ei	sõnastiku fail tsv formaadis	Vaikeväärtus on <i>dictionary/giella2cg.tab</i> .
-a	ei	lisaõnastiku fail tsv formaadis	Lisa-reeglite fail.
-f	ei	väljundi failide kaust	Vaikeväärtus <i>./</i> (jooksev kataloog).
-e	ei	väljundi failide laiend	Vaikeväärtus on <i>mrf.inforem</i> .

Skripti sisend ja väljund

Sisendi formaat

Sisendiks on tekstifail XML-vormingus ja UTF-8 formaadis. Lause eraldajaks on `<s>` ja `</s>` märgendid, sõne ja sõne morfoloogiline info asuvad ühel real. Osalause eraldajaks on märgendid `<kindel_piir/>` ning `<kiil>`, `</kiil>`. Kõiki teisi märgendeid ignoreeritakse.

Näide:

```
<doc id="269429" length="10k-100k" crawl_date="2013-01-10"
url="http://www.tehnokratt.ee/" web_domain="www.tehnokratt.ee" langdiff="0.18"
texttype="unknown">
<p heading="0">
<s>
Tulemuseks... tulemus+ks // _S_ Sg Tra, //
saad... saa+d // _V_ Pers Prs Ind Sg2 Aff, //
teksti... tekst+0 // _S_ Sg Gen, //
<kindel_piir/>
mille... mis+0 // _P_ Sg Gen, //
võid... või+d // _V_ Pers Prs Ind Sg2 Aff, //
kopeerida... kopeeri+da // _V_ Inf, //
nt... nt+0 // _Y_ ?, //
hyphenator-et.js... hyphenator-et.js+0 // _Q_ //
failiks... fail+ks // _S_ Sg Tra, //
. .... // _Z_ //
</s>
</p>
</doc>
```

Väljundi formaat

Väljundiks on tekstifail inforem vormingus. Kodeeringuks on UTF-8.

```
"<s id="1">"
"
"<Tulemuseks>"
Δ "tulemus" Lks S com sg tr
"<saad>"
Δ "saa" Ld V main indic pres ps2 sg ps af
"<teksti>"
Δ "tekst" L0 S com sg gen
"<mille>"
Δ "mis" L0 P PLIIK sg gen
"<võid>"
Δ "või" Ld V main indic pres ps2 sg ps af
"<kopeerida>"
Δ "kopeeri" Lda V main inf
"<nt>"
Δ "nt" L0 Y nominal ?
"<hyphenator-et.js>"
Δ "hyphenator-et.js" L0 _Q_
"<failiks>"
Δ "fail" Lks S com sg tr
"<.>"
Δ ". " Z Fst
"</s>"
```

Standardsisend

Standardsisendist sisendi lugemisel kirjutab skript väljundi standardväljundisse.

```
$ python morf2morf.py < test/sample.gtmrf
```

```
Reading conf from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab
```

morf2morf.py is running with the following params:

local_conf file	/Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab
outout_folder	.
output_extension	mrf.ext
default_dictionary	/Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab

Imported 455 rules from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab

```
1.. "<s id="1">"
"<Las>"
  "las" L0 D
"<nemad>"
  "tema" Ld P pers ps3 pl nom
"<ka>"
  "ka" L0 D
"<teenivad>"
  "teeni" Lvad V main indic pres ps3 pl ps af
"<,>"
  "," Z Com
"<ega>"
  "ega" L0 D
"<mul>"
  "mina" Li P pers ps1 sg ad
"<kahju>"
  "kahju" L0 S com sg part
"<pole>"
  "ole" L0 V main indic pres ps neg
"<.>"
  " ." Z Fst
"</s>"
```

Standardväljundi saab suunata tekstifaili.

```
$ python morf2morf.py < test/sample.gtmrf > result.inforem
```

```
Reading conf from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab
```

morf2morf.py is running with the following params:

local_conf file	/Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab
outout_folder	.
output_extension	mrf.ext
default_dictionary	/Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab

Imported 455 rules from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab

```
1..
$
$ cat result.inforem
"<s id="1">"
"<Las>"
  "las" L0 D
"<nemad>"
  "tema" Ld P pers ps3 pl nom
"<ka>"
  "ka" L0 D
"<teenivad>"
  "teeni" Lvad V main indic pres ps3 pl ps af
"<,>"
  "," Z Com
"<ega>"
  "ega" L0 D
"<mul>"
  "mina" Li P pers ps1 sg ad
"<kahju>"
  "kahju" L0 S com sg part
"<pole>"
  "ole" L0 V main indic pres ps neg
"<.>"
  " ." Z Fst
"</s>"
```

Sisend failide nimekirjana

Sisendfailid saab skriptile ette anda nimekirjana. Sellisel juhul salvestatakse iga ette antud faili

teisendamise tulemus eraldi failina. Tulemusfaili nimi tuletatakse sisendfaili nimest järgnevalt:

[väljundi_kataloog]+sisendfaili nimi+väljundi_laiend.

Väljundi kataloogi ja laiendi seadistamise parameetrid on kirjeldatud peatükkides "[Skripti käsurea parameetrid](#)" ja "[Seadete fail](#)".

Sisendiks on failid:

samples/269429.url.olp.gtmrf-1
samples/269420.url.olp.gtmrf-1

Väljundiks on failid:

./269429.url.olp.gtmrf-1.mrf.inforem
./269420.url.olp.gtmrf-1.mrf.inforem

```
$ python morf2morf.py samples/269429.url.olp.gtmrf-1 samples/269420.url.olp.gtmrf-1  
Reading conf from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab
```

```
morf2morf.py is running with the following params:  
local_conf file      /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab  
outout_folder        ./  
output_extension     mrf.ext  
default_dictionary   /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab  
Input files:  
samples/269429.url.olp.gtmrf-1  
samples/269420.url.olp.gtmrf-1
```

```
samples/269429.url.olp.gtmrf-1 -----> ./269429.url.olp.gtmrf-1.mrf.inforem  
samples/269420.url.olp.gtmrf-1 -----> ./269420.url.olp.gtmrf-1.mrf.inforem
```

Mitme faili teisendamiseks saab kasutada ka meta-märkide abi.

Meta-märkide abil kogu kataloogi sisu teisendamise näide:

```
$ python morf2morf.py ~/svnkorpuskasiyh/trunk/ettentenist-valik/gtmrf-1/*.gtmrf-1  
Reading conf from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab
```

```
morf2morf.py is running with the following params:  
local_conf file      /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab  
outout_folder        ./  
output_extension     mrf.ext  
default_dictionary   /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab  
Input files:  
/Users/rabauti/svnkorpuskasiyh/trunk/ettentenist-valik/gtmrf-1/100006.url.olp.gtmrf-1  
/Users/rabauti/svnkorpuskasiyh/trunk/ettentenist-valik/gtmrf-1/100024.url.olp.gtmrf-1  
/Users/rabauti/svnkorpuskasiyh/trunk/ettentenist-valik/gtmrf-1/100026.url.olp.gtmrf-1  
/Users/rabauti/svnkorpuskasiyh/trunk/ettentenist-valik/gtmrf-1/100084.url.olp.gtmrf-1  
/Users/rabauti/svnkorpuskasiyh/trunk/ettentenist-valik/gtmrf-1/100108.url.olp.gtmrf-1  
/Users/rabauti/svnkorpuskasiyh/trunk/ettentenist-valik/gtmrf-1/100307.url.olp.gtmrf-1  
/Users/rabauti/svnkorpuskasiyh/trunk/ettentenist-valik/gtmrf-1/100439.url.olp.gtmrf-1  
/Users/rabauti/svnkorpuskasiyh/trunk/ettentenist-valik/gtmrf-1/100495.url.olp.gtmrf-1  
/Users/rabauti/svnkorpuskasiyh/trunk/ettentenist-valik/gtmrf-1/100575.url.olp.gtmrf-1
```

Sisend ja väljiund -1 režiimis

Sisendit teisendatakse ühe rea kaupa, ignoreeritakse lause struktuuri infot (nt osalausete piire). Mõeldud selleks, et saaks skripti kasutada ka üksikute ridade teisendamiseks, ilma et sisend oleks XML-vormingus.

NB! -1 režiimis ignoreeritakse kõiki XML-märgendeid. Enamik [teisendusreegleid](#) selles režiimis ei tööta.

Väljundiks on inforem-laadne kuju, milles puuduvad lausete piire määravad märgendid (<s>, </s>)

```
$ echo "et  et+0 //_ //" | python morf2morf.py -1
```

```
"<et>"  
"et" L0 J sub
```

Seadete fail

Seadete fail on mõeldud skripti vaikimisi väärtuste muutmiseks, ilma et neid peaks iga kord käsurealt ette andma (nt tulemusfailide kataloogi nimi ja laiend). Lisaks saab seadete failis üle kirjutada väärtusi, mida käsurealt ei saagi ette anda, nagu reeglite skoorid.

Asukoht

Vaikimisi seadete fail *conf.tab* asub skripti kataloogis *conf*. Kui tahetakse kasutada mujal asuvat seadete faili, siis tuleb faili asukoha skriptile ette anda käsurealt parameetri *-c* väärtusena.

```
$ python morf2morf.py -c otherconf/conf.tab test/sample.gtmrf
Reading conf from otherconf/conf.tab
```

```
morf2morf.py is running with the following params:
local_conf file      otherconf/conf.tab
output_folder        ./
output_extension     mrf.ext
default_dictionary   /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab
Input files:
test/sample.gtmrf
```

Seadistatavad parameetrid

Parameetrid, mille väärtused saab seadete failis üle kirjutada.

Parameeter	Vaikimisi väärtus	Kirjeldus
Üldised seaded		
conf.default_dictionary	dictionaries/giella2cg.tab	vaikimisi sõnastik
conf.additional_dictionary	[pole määratud]	lisa-sõnastik
conf.output_extension	mrf.inforem	tulemusfailide laiend
conf.output_folder	.	tulemusfailide kataloog
Sõnastiku reeglite seaded	(skoor skriptis)	reegel (vt “Sõnastik ja reeglid”)
dict_weight.lemma+	100	l+
dict_weight.lemma-	100	l-
dict_weight.prev_lemma+	70	pl+
dict_weight.prev_lemma-	70	pl-
dict_weight.next_lemma+	70	nl+
dict_weight.next_lemma-	70	nl-
dict_weight.clause_lemma+	50	cl+
dict_weight.clause_lemma-	50	cl-
dict_weight.prev_morf+	65	pm+
dict_weight.prev_morf-	65	pm-
dict_weight.next_morf+	65	nl+
dict_weight.next_morf-	65	nl-
dict_weight.clause_morf+	50	cm+
dict_weight.clause_morf-	50	cm-

dict_weight.gramcat+	25	g+
dict_weight.gramcat-	25	g-
dict_weight.prev_gramcat+	20	pg+
dict_weight.prev_gramcat-	20	pg-
dict_weight.next_gramcat+	20	ng+
dict_weight.next_gramcat-	20	ng-

Seadete faili formaat

Seadete fail on tsv formaadis. Ridu, mis algavad # märgiga ignoreeritakse.

Näide seadete failist (rohelistega on märgitud aktiivsed read):

```
#comment line starts with #
#remove '#' to activate or override

#####
#           Default configuration values
#####
#conf.default_dictionary      /dict/folder/and/filename.tab
#conf.additional_dictionary  /dict/folder/and/filename.tab
conf.output_extension      mrf.ext
conf.output_folder        result

#####
# Dictionary rules weights
#####

dict_weight.lemma+        300
#dict_weight.lemma-        100
#dict_weight.prev_lemma+   70
#dict_weight.prev_lemma-   70
#dict_weight.next_lemma+   70
#dict_weight.next_lemma-   70
#dict_weight.clause_lemma+ 50
#dict_weight.clause_lemma- 50

#dict_weight.prev_morf+     65
#dict_weight.prev_morf-     65
#dict_weight.next_morf+     65
#dict_weight.next_morf-     65
#dict_weight.clause_morf+   50
#dict_weight.clause_morf-   50

#dict_weight.gramcat+      25
#dict_weight.gramcat-      25
#dict_weight.prev_gramcat+ 20
#dict_weight.prev_gramcat- 20
#dict_weight.next_gramcat+ 20
#dict_weight.next_gramcat- 20
```

Skripti käivitamisel kuvatavatest parameetritest saab näha, kas parameetrile omistati soovitud väärtus.

```
$ python morf2morf.py test/sample.gtmrf
Reading conf from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab
```

```
morf2morf.py is running with the following params:
local_conf file      /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab
output_folder      result
output_extension  mrf.ext
default_dictionary   /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab
Input files:
    test/sample.gtmrf
```


Konfiguratsiooni failis määratud seaded saab üle kirjutada skripti käivitamisel käsurealt.

```
python morf2morf.py -e muu.laiend test/sample.gtmrf
```

Reading conf from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab

morf2morf.py is running with the following params:

local_conf file	/Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab
-----------------	---

output_folder	result/folder
---------------	---------------

output_extension	muu.laiend
-------------------------	-------------------

default_dictionary	/Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab
--------------------	--

Input files:

test/sample.gtmrf

Sõnastik ja reeglid

Sõnastiku faili struktuur

Sõnastiku fail on TSV (*tab-separated values*) fail UTF-8 kodeeringus, väljade eraldajaks on tabulaator ning jutumärk (“).

Kõik morfoloogiliste kategooriate teisendused tehakse sõnastiku faili ridade alusel.

Sõnastiku esimeses veerus on morfoloogilised kategooriad, mida teisendatakse. Teises veerus on morfoloogilised kategooriad millega asendatakse esimeses veerus oleva info. Järgnevates veergudes (veergude arv on piiramata) on teisendust täpsustavad/kitsendavad [reeglid](#) või siis kommentaarid. Kommentaarid on vabas vormis ning neid ignoreeritakse skripti poolt.

Sõnastiku faili näide

```
_D_      D
_K_      K pre
_K_      K post   l+kaudu
_J_      J
_J_      J sub    l+et
_A_      A pos
_A_ Sg Nom      A pos sg nom
_A_ Pl Nom      A pos pl nom
_A_ Sg Gen      A pos sg gen
_A_ Pl Gen      A pos pl gen
_A_ Sg Par      A pos sg part
_A_ Pl Par      A pos pl part
_A_ Sg Ill      A pos sg ill
_A_ Pl Ill      A pos pl ill
```

D	D	
K	K pre	
K	K post	l+kaudu
J	J	
J	J sub	l+et

Reeglid ja prioriteedid

Mida täpsem on reegel, seda suurem prioriteet sellel on.

reegel		**skoor	näide
l+	lemma on	100	l+mina
l-	lemma pole	100	l+mina
pl+	*eelnev lemma on	70	pl+ei
pl-	*eelnev lemma pole	70	pl-ei
nl+	*järgnev lemma on	70	nl+ei
nl-	*järgnev lemma pole	70	nl-ei
cl+	*samases osalauses esineb lemma	50	cl+ei
cl-	*samases osalauses ei esine lemma	50	cl-ei

pm+	*eelneva sõne morfoloogiline märgend on	65	pm+_H_ Sg Gen
pm-	*eelneva sõne morfoloogiline märgend pole	65	pm-_H_ Sg Gen
cm+	*samas osalauses esineb morfoloogiline märgend	50	cm+_H_ Sg Gen
cm-	*samas osalauses ei esine morfoloogiline märgend	50	cm-_H_ Sg Gen
g+	grammatiline kategooria on	25	g+pl
g-	grammatiline kategooria pole	25	g+pl
pg+	*eelnev grammatiline kategooria on	24	pg+pl
pg-	*eelnev grammatiline kategooria pole	24	pg-pl
ng+	*järgnev grammatiline kategooria on	24	ng+pl
ng-	*järgnev grammatiline kategooria pole	24	ng-pl
* ei tööta skripti -1 režiimis ** reeglite skoori saab muuta seadete faili kaudu (vt "Seadete fail")			

Reegli sobimisel omistatakse tulemusele skoor ülal äratoodud tabeli põhjal ning tulemuseks valitakse kõige suurema skoori saanud reegli rida. Kui tulemuseks on mitu rida sama skooriga, siis valitakse neist suvaline.

Nt:

1. _Y_ ? Y l+jne skoor matchimisel on 1+100 = 101
2. _Y_ ? Y cl+koer skoor matchimisel on 1+50 = 51
3. **_Y_ ? Y l+jne cl+ei skoor matchimisel on 1+100+50 = 150**

Kui teisendatava sõne morfoloogiline info on “_Y_”, sõne lemma on “jne” ning samas osalauses leidub sõne, mille lemma on “ei”, siis sobivad teisenduseks kõik kolm sõnastiku rida, kuid suurema skoori saab kõige täpsem vastavus - rida number 3.

Juhul kui sõnastikus ei leidu teisendamiseks sobivat rida, siis teisendatakse teisendamata jäänud morfoloogilised kategooriad ülaregistrisse ning skripti töö lõpus väljastatakse vastav nimekiri morfoloogilistest kategooriatest, mille jaoks ei leitud reeglit sõnastikufailis. Vt “[Skripti veateated ja logi](#)”

Skripti veateated ja logi

Skripti töö logi ja veateated kirjutatakse STDERR väljundisse.

Logi näide

```
[Katrins-MBP:morf2morf rabauti$ python morf2morf.py -f result samples/269429.url.olp.gtmrf-1 test/sample.gtmrf
Reading conf from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab

morf2morf.py is running with the following params:
  local_conf file      /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab
  output_folder       result/
  output_extension     mrf.inforem
  default_dictionary   /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab
  Input files:
    samples/269429.url.olp.gtmrf-1
    test/sample.gtmrf

Imported 455 rules from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab

samples/269429.url.olp.gtmrf-1 -----> result//269429.url.olp.gtmrf-1.mrf.inforem

1..2..3..4..5..6..7..8..9..10..11..12..13..14..15..16..17..18..19..20..21..22..23..24..25..26..27..28..29..30..31..32..
33..34..35..36..37..38..39..40..41..42..43..44..45..
Unrecognized format on line 1408: ja      ja+0 //_J_ /
46..47..48..49..50..51..52..53..54..55..56..57..58..59..60..61..62..63..64..65..66..67..68..69..70..71..72..73..74..75..
76..77..78..79..80..81..82..83..84..85..86..87..88..89..90..91..92..93..94..95..96..97..98..99..100..101..102..103..
104..105..106..107..108..109..110..111..112..113..114..115..116..
Missing morf tags:
_Q_
_V_ Pers Prt Prc
_Y_ Ela

test/sample.gtmrf -----> result//sample.gtmrf.mrf.inforem

1..
```

Parameetrite väärtused

```
Reading conf from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab

morf2morf.py is running with the following params:
  local_conf file      /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/conf/conf.tab
  output_folder       result/
  output_extension     mrf.inforem
  default_dictionary   /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab
  Input files:
    samples/269429.url.olp.gtmrf-1
    test/sample.gtmrf
```

Sõnastikust reeglite lugemine

Imported 455 rules from /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf/dictionaries/giella2cg.tab
Sõnastiku nimi ja asukoht: /Users/rabauti/svnkrpsoft/trunk/rabauti/morf2morf//dictionaries/giella2cg.tab
Sisse loetud sõnastikufaili ridade arv: 455.

Sisendi töötlemine

```
samples/269429.url.olp.gtmrf-1 -----> result//269429.url.olp.gtmrf-1.mrf.inforem
```

Sisendiks on fail *samples/269429.url.olp.gtmrf-1*. Väjundiks on fail *result/269429.url.olp.gtmrf-1.mrf.inforem* .

```
Parsing input and writing output.
1..2..3..4..5..6..7..8..9..10..11..12..13..14..15..16..17..18..19..20..21..22..23..24..25..26..27..28..29..30..31..32..33..3
4..35..36..37..38..39..40..41..42..43..44..45..
Unrecognized format on line 1408 ja      ja+0 //_J_ /
46..47..48..49..50..51..52..53..54..55..56..57..58..59..60..61..62..63..64..65..66..67..68..69..70..71..72..73..74..75..76..
77..78..79..80..81..82..83..84..85..86..87..88..89..90..91..92..93..94..95..96..97..98..99..100..101..102..103..104..105..10
6..107..108..109..110..111..112..113..114..115..116..
```

Numbrid 1..2..3 tähistavad mitmenda lause töötlemise juures parasjagu ollakse.

Veateated

```
.33..34..35..36..37..38..39..40..41..42..43..44..45..  
Unrecognized format on line 1408: ja      ja+0 //_J_ /  
46..47..48..49..50..51..52..53..54..55..56..57..58..59..60..61..62..63..64..65..66..67..68..69..70..71..72..73..74..75
```

Veateade koosneb vea kirjeldusest, vigase rea numbrist sisendfailis ning vigase rea sisust. Lisaks on võimalik näha, et viga tekkis 45. lause töötlemisel.

Puuduvad morfoloogilised kategooriad

Skripti töö lõpus väljastatakse morfoloogilised kategooriad, mille jaoks ei leitud sobivat reeglit ning mis jäid teisendamata.

```
Missing morf tags:  
_Q_  
_V_ Pers Prt Prc  
_Y_ Ela
```