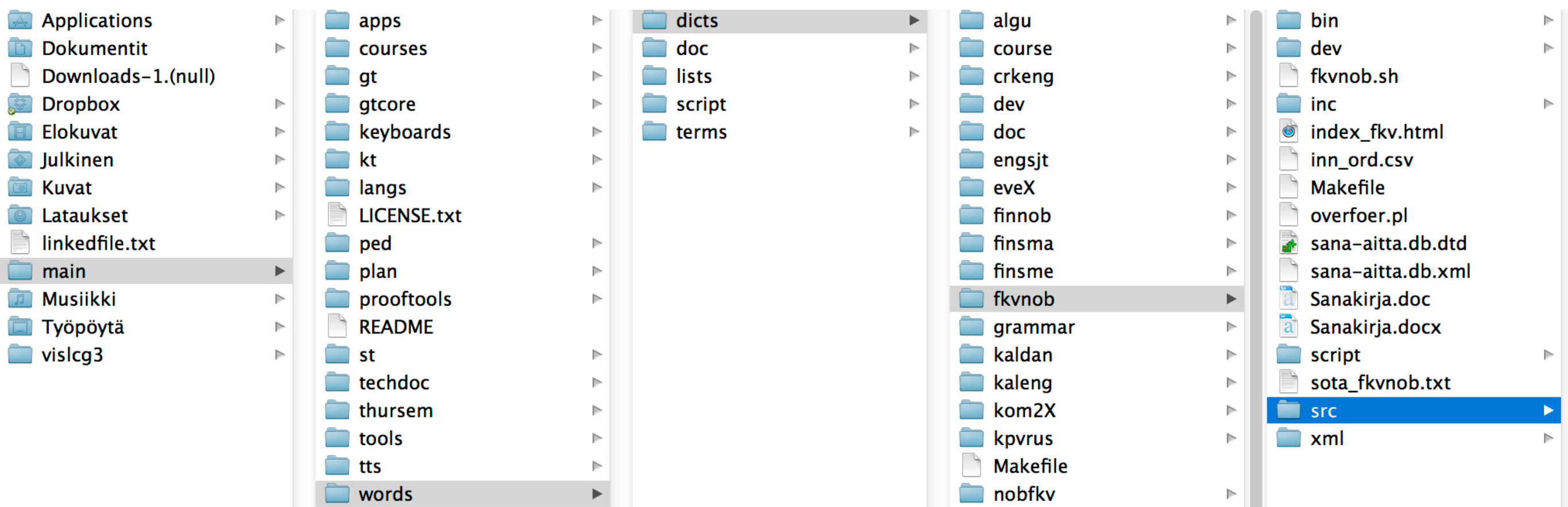


# *Kieliteknologian muistilista*

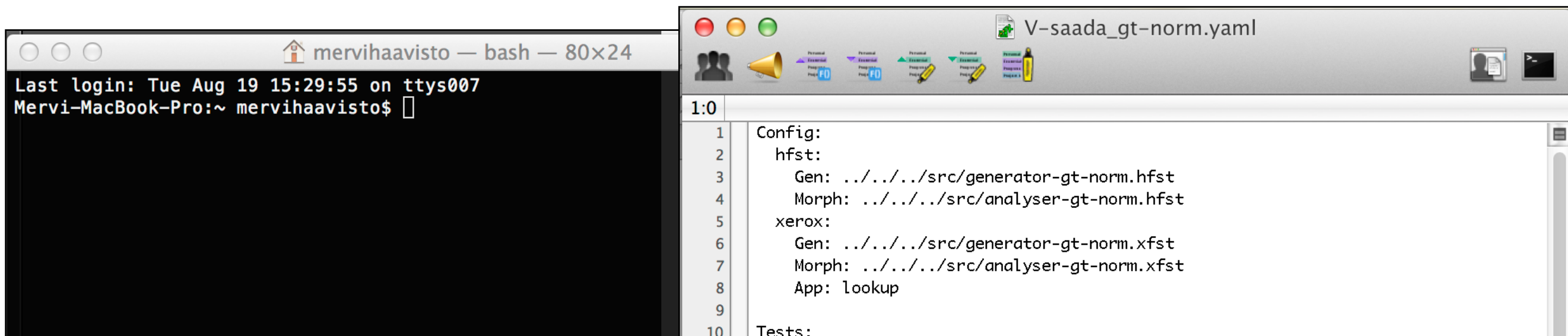


# Lähtökohtana on Eira Söderholmin kirjoittama Kainun kielen grammatikka

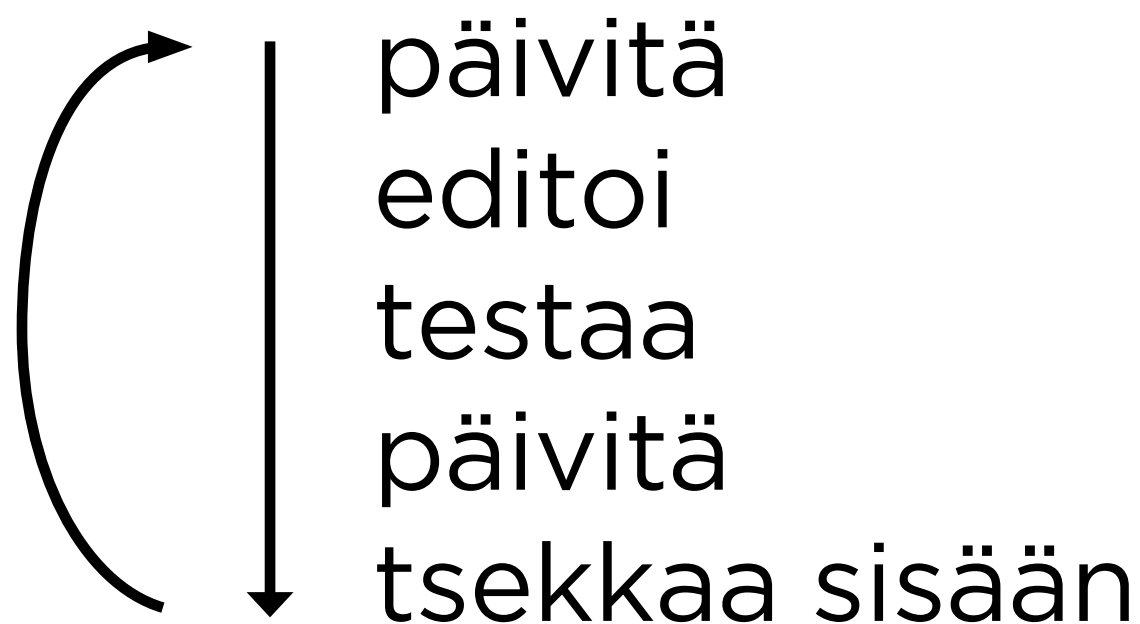
<

Viimeisin versio kansiossa  
**main/langs/fkv/doc**

# *Tiedostoja muokataan SubEthaEditissä ja niitä käytetään Terminaalilla/Päätteellä*

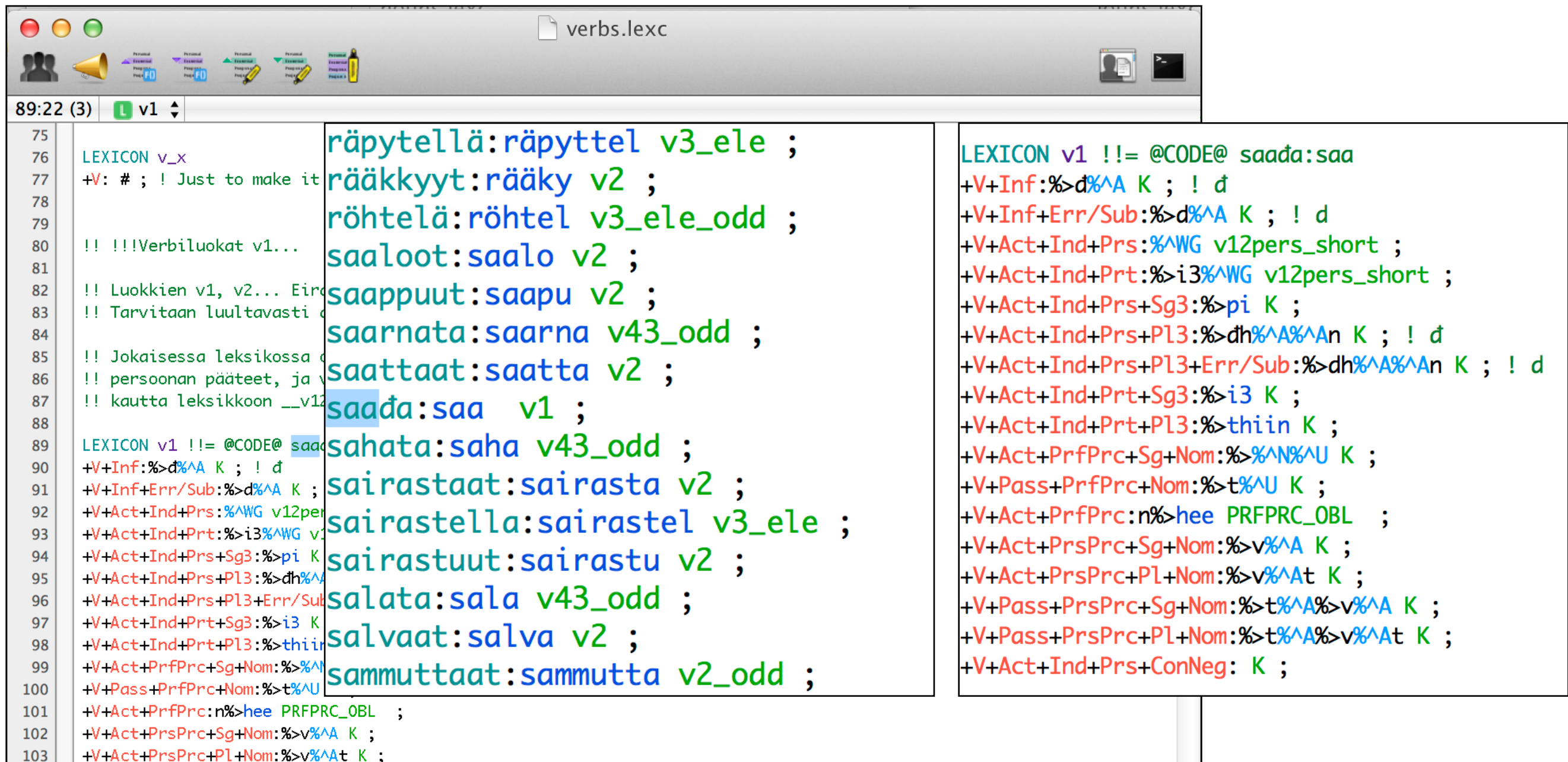


## Työnkulku





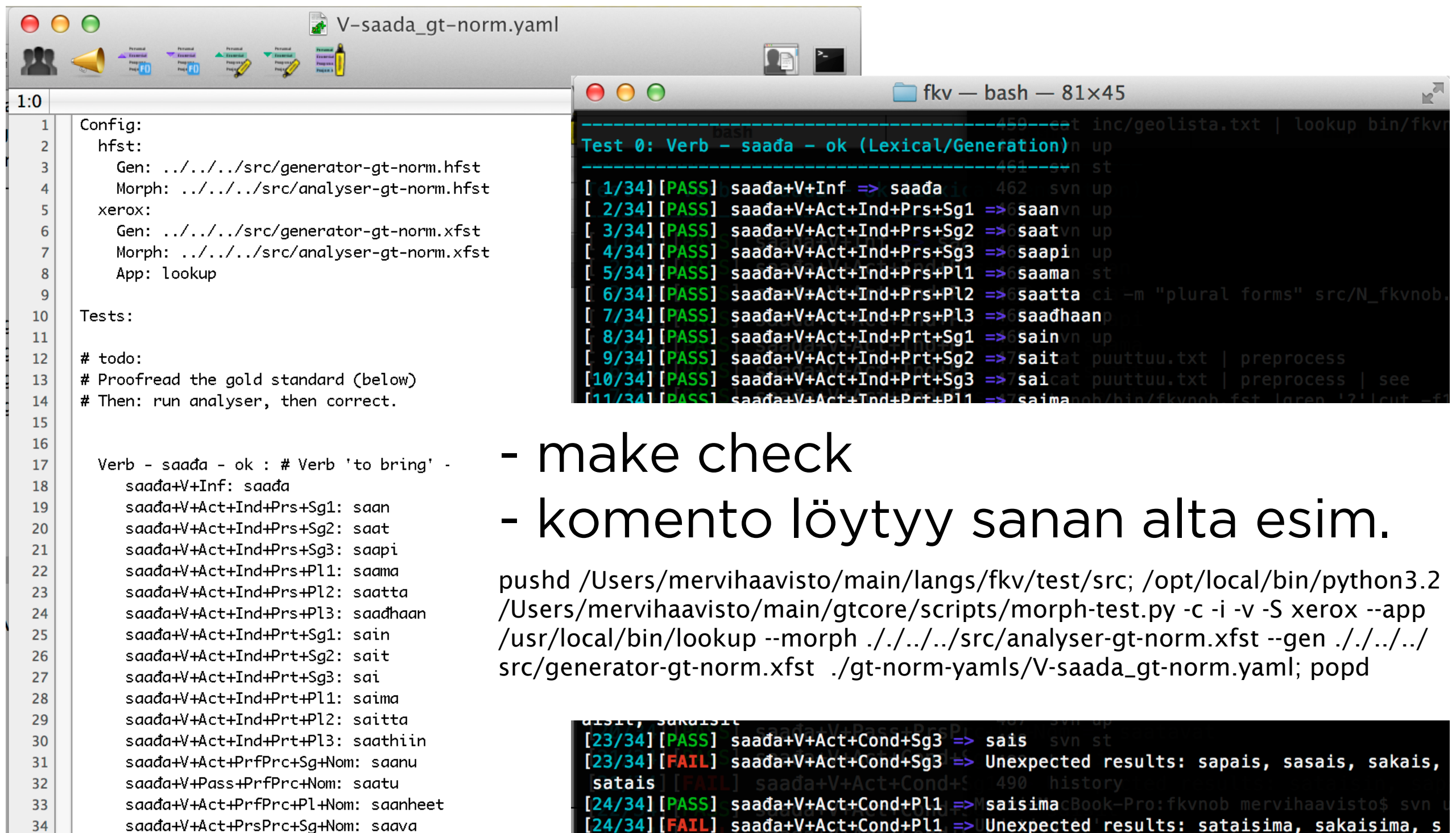
# *Sanojen vartalot ja päätteet lisätään lexc-tiedostoihin vartalo: 'saa' -> pääte: 'da' - kansiossa morphology/stems ja affixes*



```
89:22 (3) L v1
75 LEXICON v_x
76 +V: # ; ! Just to make it
77
78
79
80 !! !!!Verbiluokat v1...
81
82 !! Luokkien v1, v2... Eir
83 !! Tarvitaan luultavasti
84
85 !! Jokaisessa leksikossa
86 !! persoonan päätteet, ja
87 !! kautta leksikkoon __v1
88
89 LEXICON v1 != @CODE@ saada: saa
90 +V+Inf:%>d%AA K ; ! d
91 +V+Inf+Err/Sub:%>d%AA K ; ! d
92 +V+Act+Ind+Prs:%^WG v12pers_short ;
93 +V+Act+Ind+Prt:%>i3%^WG v12pers_short ;
94 +V+Act+Ind+Prs+Sg3:%>pi K ;
95 +V+Act+Ind+Prs+Pl3:%>dh%AA^An K ; ! d
96 +V+Act+Ind+Prs+Pl3+Err/Sub:%>dh%AA^An K ; ! d
97 +V+Act+Ind+Prt+Sg3:%>i3 K ;
98 +V+Act+Ind+Prt+Pl3:%>tiin K ;
99 +V+Act+PrfPrc+Sg+Nom:%>t%AU K ;
100 +V+Pass+PrfPrc+Nom:%>t%AU K ;
101 +V+Act+PrfPrc:n%>hee PRFPRC_OBL ;
102 +V+Act+PrsPrc+Sg+Nom:%>v%AA K ;
103 +V+Act+PrsPrc+Pl+Nom:%>v%AA^t K ;

räpytellä:räpyttel v3_ele ;
rääkkyt:rääky v2 ;
röhtelä:röhtel v3_ele_odd ;
saaloot:saalo v2 ;
saappuut:saapu v2 ;
saarnata:saarna v43_odd ;
saattaat:saatta v2 ;
saada:saa v1 ;
sahata:saha v43_odd ;
sairastaat:sairasta v2 ;
sairastella:sairastel v3_ele ;
sairastuut:sairastu v2 ;
salata:sala v43_odd ;
salvaat:salva v2 ;
sammuttaat:sammutta v2_odd ;
```

# *Tietyistä sanoista tehdään yaml-testitiedosto, ja sanan taivutus testataan Terminaalissa*



```
1:0
1 Config:
2   hfst:
3     Gen: ../../src/generator-gt-norm.hfst
4     Morph: ../../src/analyser-gt-norm.hfst
5   xerox:
6     Gen: ../../src/generator-gt-norm.xfst
7     Morph: ../../src/analyser-gt-norm.xfst
8     App: lookup
9
10 Tests:
11
12 # todo:
13 # Proofread the gold standard (below)
14 # Then: run analyser, then correct.
15
16
17 Verb - saada - ok : # Verb 'to bring' -
18   saada+V+Inf: saada
19   saada+V+Act+Ind+Prs+Sg1: saan
20   saada+V+Act+Ind+Prs+Sg2: saat
21   saada+V+Act+Ind+Prs+Sg3: saapi
22   saada+V+Act+Ind+Prs+Pl1: saama
23   saada+V+Act+Ind+Prs+Pl2: saatta
24   saada+V+Act+Ind+Prs+Pl3: saadhaan
25   saada+V+Act+Ind+Prt+Sg1: sain
26   saada+V+Act+Ind+Prt+Sg2: sait
27   saada+V+Act+Ind+Prt+Sg3: sai
28   saada+V+Act+Ind+Prt+Pl1: saima
29   saada+V+Act+Ind+Prt+Pl2: saitta
30   saada+V+Act+Ind+Prt+Pl3: saathiin
31   saada+V+Act+PrfPrc+Sg+Nom: saanu
32   saada+V+Pass+PrfPrc+Nom: saatu
33   saada+V+Act+PrfPrc+Pl+Nom: saanheet
34   saada+V+Act+PrsPrc+Sg+Nom: saava
```

```
fkv — bash — 81x45
-----459---eet inc/geolista.txt | lookup bin/fkv
Test 0: Verb - saada - ok (Lexical/Generation)n up
-----461---svn st
[ 1/34] [PASS] saada+V+Inf => saada 462 svn up
[ 2/34] [PASS] saada+V+Act+Ind+Prs+Sg1 => saanvn up
[ 3/34] [PASS] saada+V+Act+Ind+Prs+Sg2 => saatvn up
[ 4/34] [PASS] saada+V+Act+Ind+Prs+Sg3 => saapin up
[ 5/34] [PASS] saada+V+Act+Ind+Prs+Pl1 => saaman st
[ 6/34] [PASS] saada+V+Act+Ind+Prs+Pl2 => saatta ci -m "plural forms" src/N_fkvnob.
[ 7/34] [PASS] saada+V+Act+Ind+Prs+Pl3 => saadhaanp
[ 8/34] [PASS] saada+V+Act+Ind+Prt+Sg1 => sainvn up
[ 9/34] [PASS] saada+V+Act+Ind+Prt+Sg2 => saitat puuttuu.txt | preprocess
[10/34] [PASS] saada+V+Act+Ind+Prt+Sg3 => saicat puuttuu.txt | preprocess | see
[11/34] [PASS] saada+V+Act+Ind+Prt+Pl1 => saima nob/bin/fkvnob.fst | grep '!?!cut -f'
```

- make check
- komento löytyy sanan alta esim.

```
pushd /Users/mervihaavisto/main/langs/fkv/test/src; /opt/local/bin/python3.2
/Users/mervihaavisto/main/gtcore/scripts/morph-test.py -c -i -v -S xerox --app
/usr/local/bin/lookup --morph ../../src/analyser-gt-norm.xfst --gen ../../src/
src/generator-gt-norm.xfst ../gt-norm-yamls/V-saada_gt-norm.yaml; popd
```

```
distc, sakaisit
[23/34] [PASS] saada+V+Act+Cond+Sg3 => sais svn st
[23/34] [FAIL] saada+V+Act+Cond+Sg3 => Unexpected results: sapais, sasais, sakais,
satais [FAIL] saada+V+Act+Cond+S 490 history
[24/34] [PASS] saada+V+Act+Cond+Pl1 => saisima MacBook-Pro:fkvnob mervihaavisto$ svn t
[24/34] [FAIL] saada+V+Act+Cond+Pl1 => Unexpected results: sataisima, sakaisima, s
```

# *Geminaatio tehdään twolc-tiedostossa, esim. sana lyhenee: 'saa : sain : saisin' tai muuttuu 'k-> j'*

756:50 <No selected symbol>

741 "Shortening of long vowel in front of i"

742 !! \_\_@RULENAME@\_\_

743 Vx:0 <=> \_ Vy %>: ([Dummy: - %^AE:]) [ i | i2: | i3: | i4: ] ;

744 where Vx in ( a ä u y u y i )

745 Vy in ( a ä u y o ö e )

746 matched ;

747

748

749 !€ pää>i2>ss^A

750 !€ pöä>i>ssä

751

752 ! saada : sain : saisin

753 ! ottaa : otan : ottaisin i4 ei passaa kaikkialle!

754

755

756 !"Shortening of long vowel when vowel lengthening"

757 !! \_\_@RULENAME@\_\_

758 ! %^V:0 <=> Vowel %>: \_ %>: [ i3:i | i

759

760 ! Tämä sääntö aiheuttaa paljon ratkais

761 ! Mikä tämä sääntö on ja miten se eroa

762 ! soppiit: sopi: sopisi

763 ! juttaat: juti : jutaisi?

764 ! saada: sai : saisin

765 ! jäädä : jäin : jäisin

766 ! Poistettu toistaiseksi

767

768

769 "e:i in nom.sg. of e-stems"

770 !! \_\_@RULENAME@\_\_

771 e:i <=> Cns: \_ %^E2I:0 [ #: | .#. | %

772

773

423

424 ! jälki: jäljen ; järki; järjen; kulkkeet: kuljen; särkkeet : särjen

425

426 "Gradation k:j"

427 !k:j <=> [alelili6:] \_ [ole:li] (Dummy:) (:i) (Dummy:) %^WG:0 ; ! pako - pajon

428 k:j <=> [alelili6:] \_ [e:li] (Dummy:) (:i) (Dummy:) %^WG:0 ; !

429 i \_ [ö:lo:] (Dummy:) (:i) (Dummy:) %^WG:0 ; !

430 [älælelili6:] \_ [e:li] (Dummy:) (:i) (Dummy:) %^WG:0 ; ! maki - mäj

431 [elili6:] \_ [e:lilälæla:] (Dummy:) (:i) (Dummy:) %^WG:0 ; ! aika - ajan

432 [rll] \_ e: (Dummy:) (:i) (Dummy:) %^WG:0 ; ! jälki - jäljen

433 Cns u \_ e: (Dummy:) (:i) (Dummy:) %^WG:0 ; ! lukkeet - lujen

434

435

436 !€ maki^WG>n

437 !€ mäj^WG>n

438

439 !€ aika^WG>n

440 !€ aija^WG>n

441

442 !€ jälke^WG>n

443 !€ jälje^WG>n

444

kansio fkv/src/phonology/fkv-phon.twolc

lexc talo+N+Sg+Ine MORFOLOGIA - PÄÄTTEET - LEXC

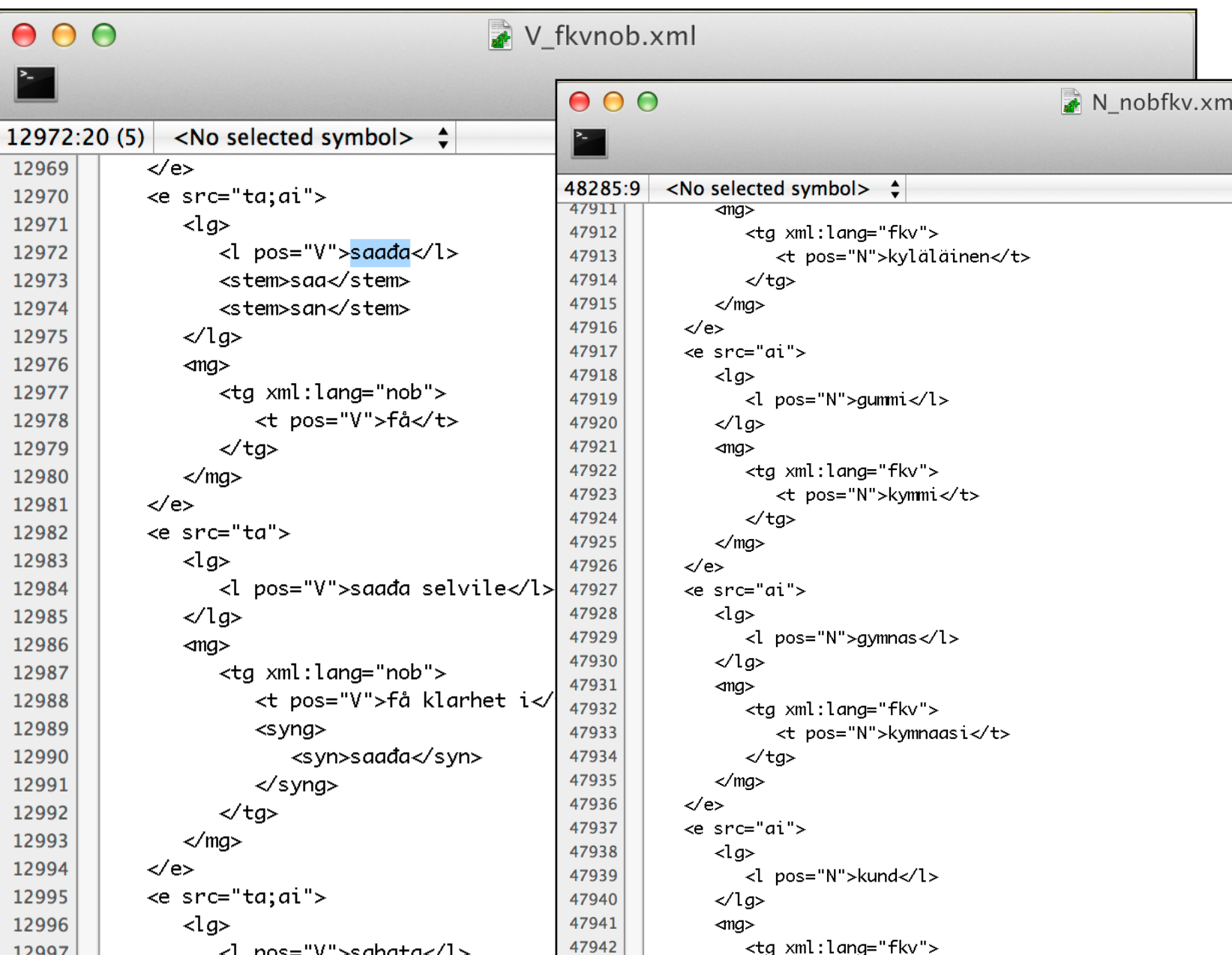
taloss^A

twolc taloss^A MORFONOLOGIA - PROSESSIT - TWOLC

talossa



*Sanakirja-tiedostot ovat sanaluokittain  
molemmilla kielillä **fkvnob** ja **nobfkv**  
- kansiossa *main/words/dicts/fkvnob/src/*  
*nobfkv**



The image shows two overlapping XML editor windows. The top window, titled 'V\_fkvnob.xml', displays XML data for Finnish words. The bottom window, titled 'N\_nobfkv.xml', displays XML data for Norwegian words. Both windows show a list of lines with line numbers and XML tags. The Finnish window has a search bar at the top with the text '<No selected symbol>'. The Norwegian window also has a search bar with the same text. The XML data includes tags like <e>, <lg>, <l>, <stem>, <mg>, <tg>, <t>, <syng>, and <syn>.

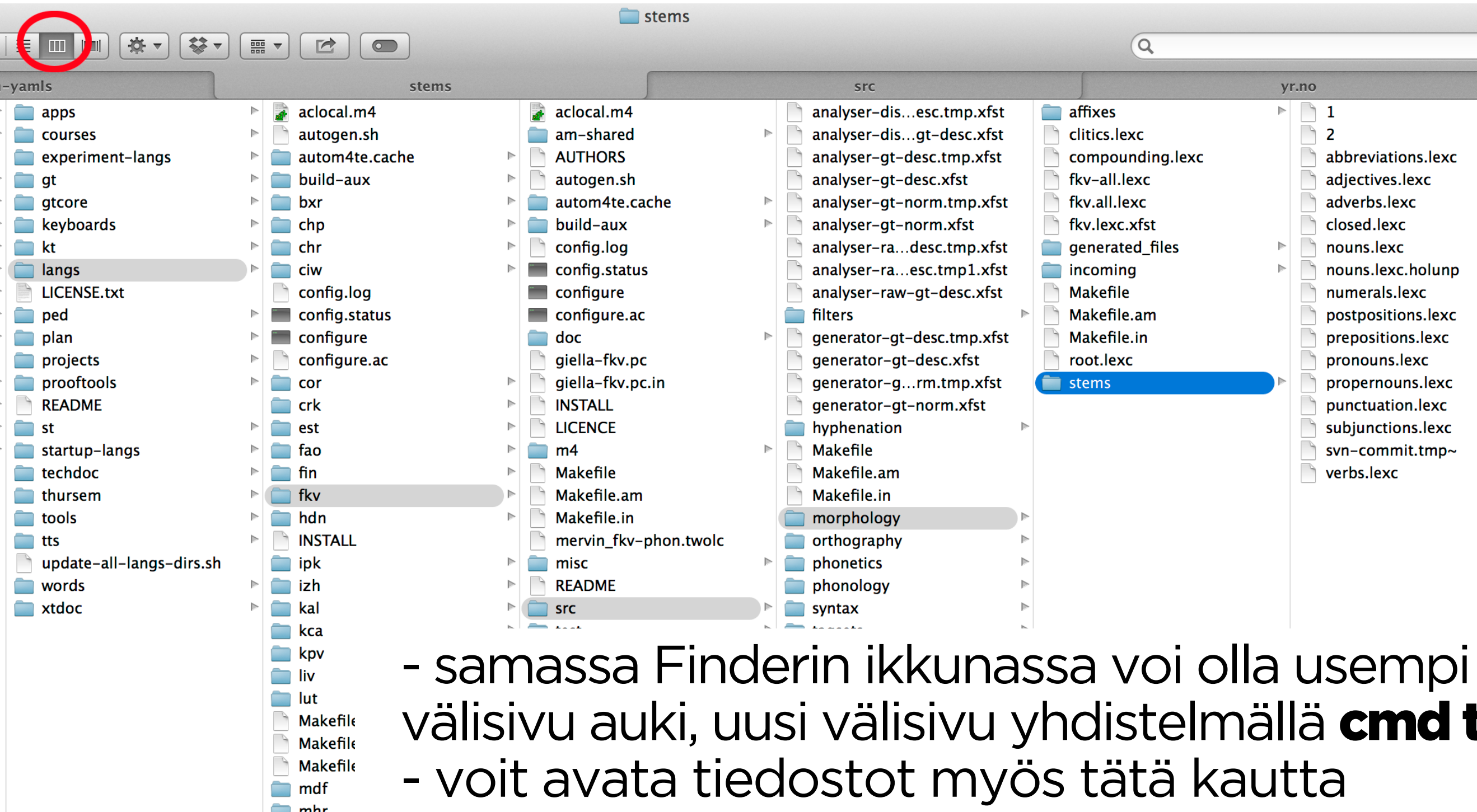
```
12969 </e>
12970 <e src="ta;ai">
12971   <lg>
12972     <l pos="V">saada</l>
12973     <stem>saa</stem>
12974     <stem>san</stem>
12975   </lg>
12976   <mg>
12977     <tg xml:lang="nob">
12978       <t pos="V">få</t>
12979     </tg>
12980   </mg>
12981 </e>
12982 <e src="ta">
12983   <lg>
12984     <l pos="V">saada selvile</l>
12985   </lg>
12986   <mg>
12987     <tg xml:lang="nob">
12988       <t pos="V">få klarhet i</t>
12989     <syng>
12990       <syn>saada</syn>
12991     </syng>
12992   </tg>
12993 </mg>
12994 </e>
12995 <e src="ta;ai">
12996   <lg>
12997     <l pos="V">sabata</l>
```

```
48285:9 <No selected symbol>
47911 <mg>
47912   <tg xml:lang="fkv">
47913     <t pos="N">kyläläinen</t>
47914   </tg>
47915 </mg>
47916 </e>
47917 <e src="ai">
47918   <lg>
47919     <l pos="N">gummi</l>
47920   </lg>
47921   <mg>
47922     <tg xml:lang="fkv">
47923       <t pos="N">kymmi</t>
47924     </tg>
47925   </mg>
47926 </e>
47927 <e src="ai">
47928   <lg>
47929     <l pos="N">gymnas</l>
47930   </lg>
47931   <mg>
47932     <tg xml:lang="fkv">
47933       <t pos="N">kymnaasi</t>
47934     </tg>
47935   </mg>
47936 </e>
47937 <e src="ai">
47938   <lg>
47939     <l pos="N">kund</l>
47940   </lg>
47941   <mg>
47942     <tg xml:lang="fkv">
```

```
<e src="ta">
  <lg>
    <l pos="N">höyhen</l>
    <stem>höyhene</stem>
  </lg>
  <mg>
    <tg xml:lang="nob">
      <t pos="N" gen="n">dun</t>
      <t pos="N" gen="f">fjær</t>
    </tg>
  </mg>
</e>
```

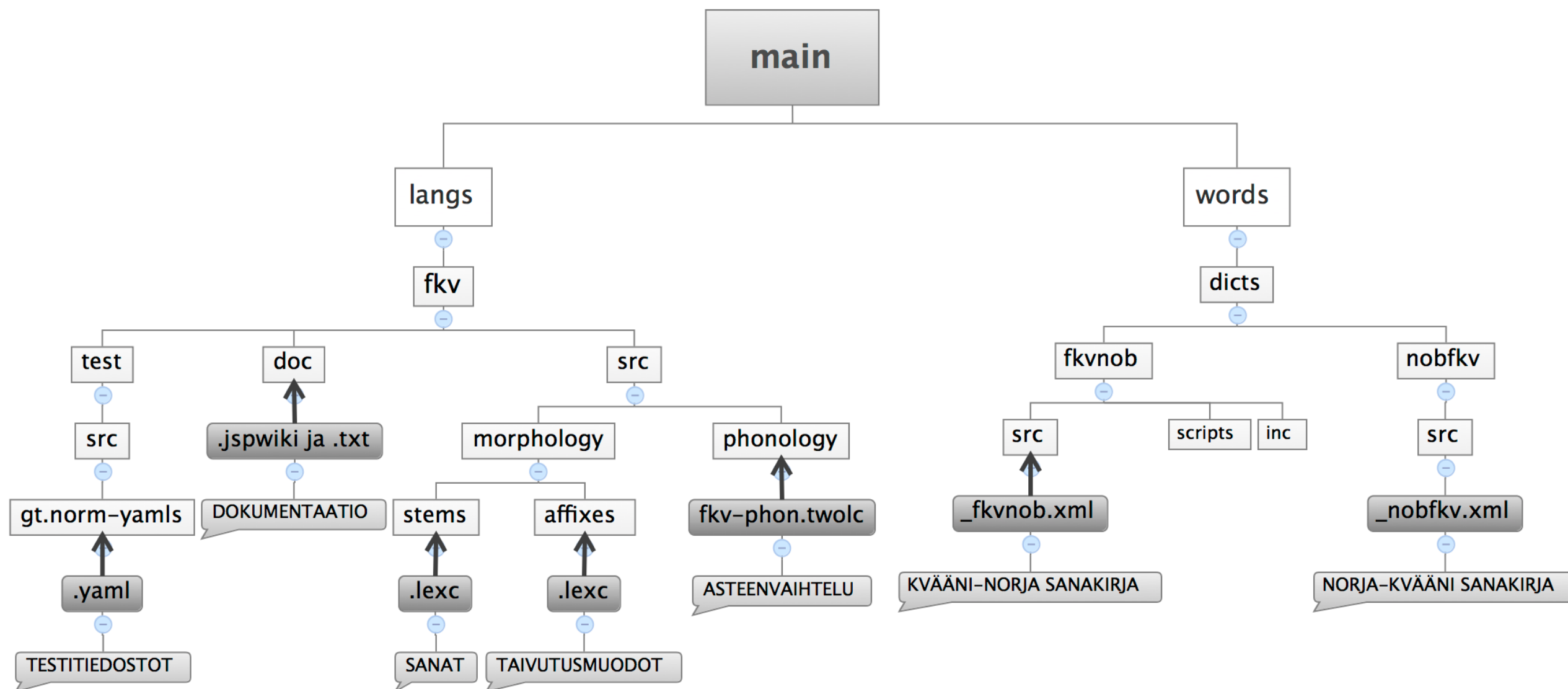
- sanat ovat **<e></e>** tagien välissä
- jokainen tagi alkaa esim. **<lg>** ja ne pitää lopettaa **/-merkillä: </lg>**

# *Koneessa oleva kansiorakenne näyttää tältä Finderissä, tässä ollaan vartalo-kansiossa*





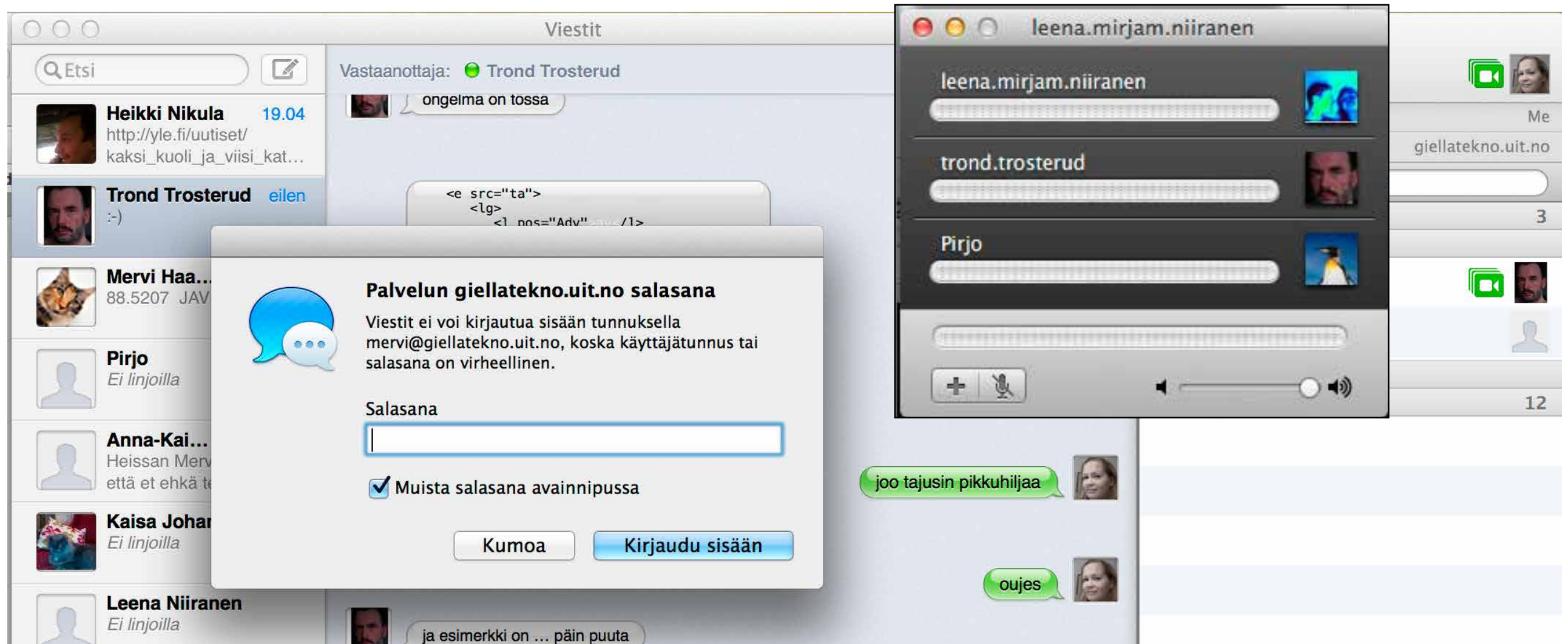
*Tavallisesti projektissamme käytettävä kansiorakenne näyttää tältä:*



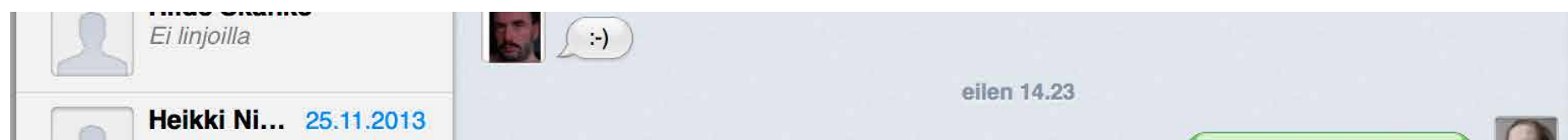
***Analysaattori***

***Sanakirja***

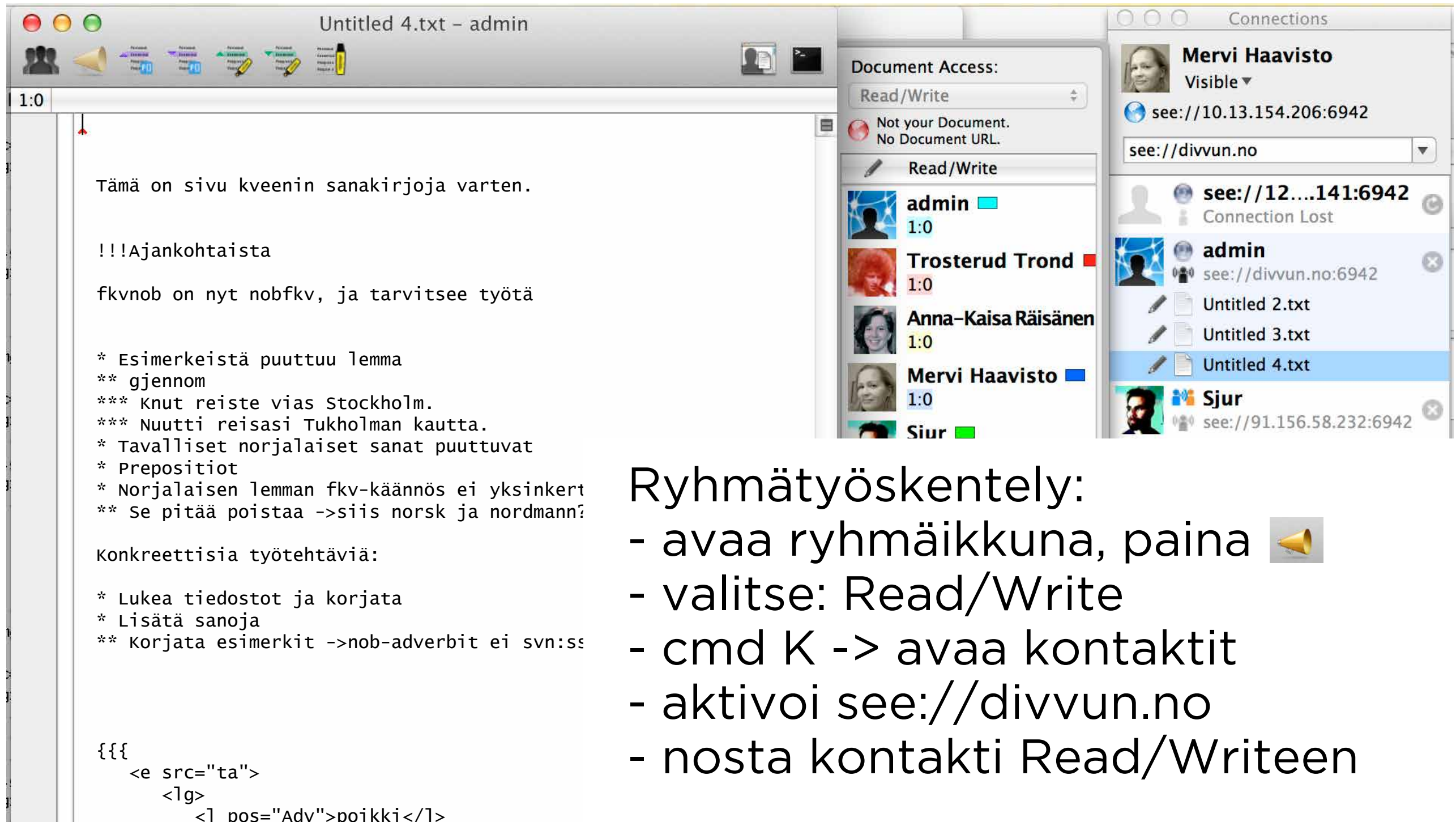
# *Macin Viestit/Meldinger on Giellateknon käyttämä yhteydenpitoväline: chat, puhelut*



Tee tili Gmail-tunnuksillasi ja hae kontakteja



# Editoriohjelma SubEthaEdit on myös ryhmätyökalu



The screenshot displays the SubEthaEdit application window titled "Untitled 4.txt - admin". The main text area contains the following content:

Tämä on sivu kveenin sanakirjoja varten.

!!!Ajankohtaista

fkvnob on nyt nobfkv, ja tarvitsee työtä

\* Esimerkeistä puuttuu lemma  
\*\* gjennom  
\*\*\* Knut reiste vias Stockholm.  
\*\*\* Nuutti reisasi Tukholman kautta.  
\* Tavalliset norjalaiset sanat puuttuvat  
\* Prepositiot  
\* Norjalaisen lemmän fkv-käännös ei yksinkert  
\*\* Se pitää poistaa -> siis norsk ja nordmann?


konkreettisia työtehtäviä:

\* Lukea tiedostot ja korjata  
\* Lisätä sanoja  
\*\* Korjata esimerkit -> nob-adverbit ei svn:ss

{{{  
 <e src="ta">  
 <lg>  
 <l pos="Adv">poikki</l>

On the right side, there are two panels. The "Document Access:" panel shows a dropdown menu set to "Read/Write" and a message "Not your Document. No Document URL.". Below this is a list of users: admin (1:0), Trosterud Trond (1:0), Anna-Kaisa Räisänen (1:0), Mervi Haavisto (1:0), and Siur. The "Connections" panel shows a list of connections, including Mervi Haavisto (Visible, see://10.13.154.206:6942), see://12....141:6942 (Connection Lost), admin (see://divvun.no:6942), and Sjur (see://91.156.58.232:6942). The "admin" connection is currently selected, showing a list of open documents: Untitled 2.txt, Untitled 3.txt, and Untitled 4.txt.

## Ryhmätyöskentely:

- avaa ryhmäikkuna, paina 
- valitse: Read/Write
- cmd K -> avaa kontaktit
- aktivoi see://divvun.no
- nosta kontakti Read/Writteen



# *Tärkeimpiä komentoja analysaattorin ja sanakirjan kanssa työskennellessä*

**cd** Change Directory, takaisin alkuun

**cd ..** ylöspäin hakemistossa

**fkv analysaattori**

cd main/langs/fkv/...

**kvääni-norja sanakirjan tiedostot**

cd main/words/dicts/fkvnob/src/...

**svn up** viimeisin versio versionhallinnasta, tehdään aina työtä aloittaessa

**svn diff** näyttää paikalliset muutokset

**svn st** näyttää muokatut tiedostot, joita ei ole päivitetty versionhallintaan, esim:

**A** addition

# *Tärkeimpiä komentoja analysaattorin ja sanakirjan kanssa työskennellessä*

**M** modification (local modification)

**U** updated

**svn ci -m “Tähän kirjoitat viestin.” src/morphology/affixes/nouns.lexc** komento päivittää muokatun nouns.lexc-tiedoston versionhallintaan, kirjoita tähän myös viesti mitä olet tehnyt

**history** näyttää aikaisemmat komennot

esim. **!531** toistaa komennon nr. 531

**pwd** Print Working Directory, tarkista missä olet

**make** uusi analysaattori

**make check** uusi analysaattori ja sanojen testit

**see** avaa tiedoston SubEthaEditissä, anna myös tiedoston nimi samassa komennossa

# *Tärkeimpiä komentoja analysaattorin ja sanakirjan kanssa työskennellessä*

**ls** listaa hakemisto

**man ls** ja **cd --help** help-komennot

**mkdir** uusi hakemisto

**svn log | less tiedostonnimi** näyttää lokitiedoston (näytöllisen)

**svn up -r 12345** palaaminen tiettyyn versioon

**xmllint** (ja tiedoston nimi) näyttää *sanakirjan* xml-koodin esim. tarkistusta varten

**ufkv** 'ylös fkv', *paikala* antaa sanan taivutuksen:

*paikala* *paikka+N+Sg+Ade*

*paikala* *paikala+Adv*

**dfkv** 'down fkv', esim. *paikka+N+Sg+Ade* saat taivutetun sanan: *paikka+N+Sg+Ade* *paikala* (poistu **ctrl c**)



# TIEDOSTON TSEKKAUS SISÄÄN FKV-KANSIOSSA

**save**

**svn up**

**make** (jos olet analysaattorissa)

**svn st** (tarvittaessa)

**svn ci -m “*Tähän kirjoitat viestin.*” src/morpho-  
logy/affixes/nouns.lexc** päivittää yhden tiedoston  
versionhallintaan

**svn ci -m “*Tähän kirjoitat viestin.*” src/morpho-  
logy/** päivittää kansion kaikki muokatut tiedostot  
versionhallintaan

# KONFLIKTI TIEDOSTOSSA

Conflict discovered in ‘/Users/mervihaavisto/main/langs/fkv/doc/adjectives-stems.jspwiki’.

Select: (p) postpone, (df) diff-full, (e) edit,  
          (mc) mine-conflict, (tc) theirs-conflict,  
          (s) show all options: mc

**postpone** selvittää konflikti myöhemmin

**diff-full** näyttää paikalliset muutokset

**edit** muokkaa tiedostoa SEE:ssä

**mine-conflict** jos haluat säilyttää omat muutokset

**theirs-conflict** jos haluat tilalle toisten muutokset

Voit myös kopioida muutokset varatiedostoon,  
tarkistaa eroavaisuudet ja yhdistää tiedot

cp tiedostonnimi.lexc varatiedosto.lexc (**cp** kopioi)

diff tiedostonnimi.lexc varatiedosto.lexc

# OHJELMIEN LOPETTAMISKOMENNOT

ufkv/dfkv **ctrl c** eli ^c

ulos hausta, helpistä jne. **q**

Emacs-editori **ctrl x** & **ctrl c**

SubEthaEdit **cmd q**