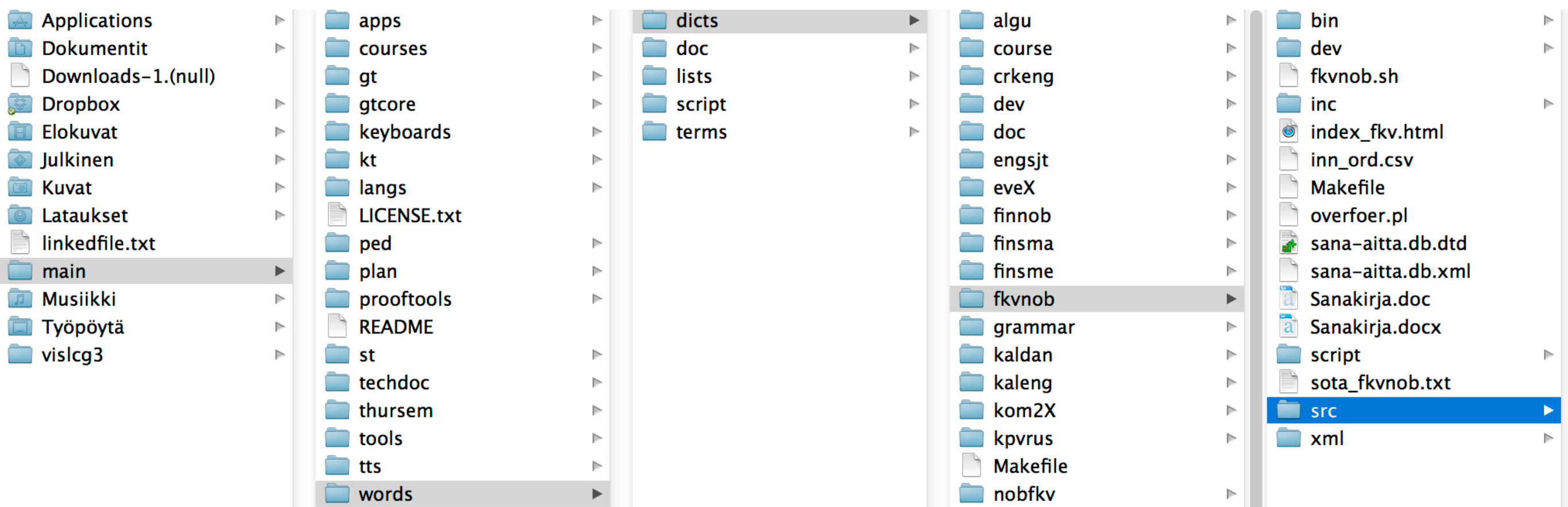


# *Kieliteknologian muistilista*

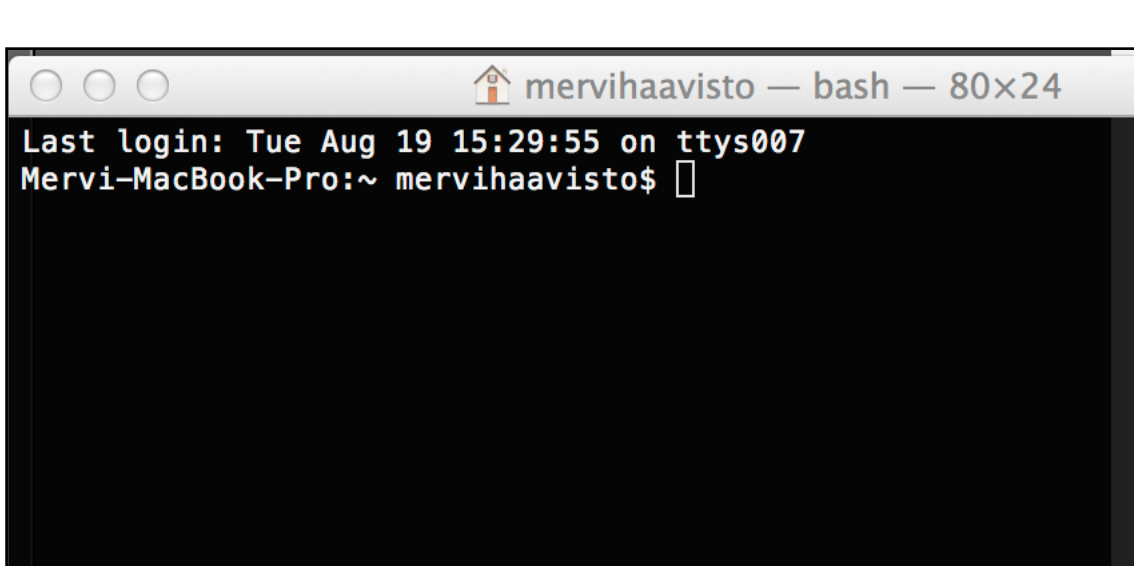


# Utgangspunkt til arbeidet er Eira Söderholms Kainun kielen grammatikki

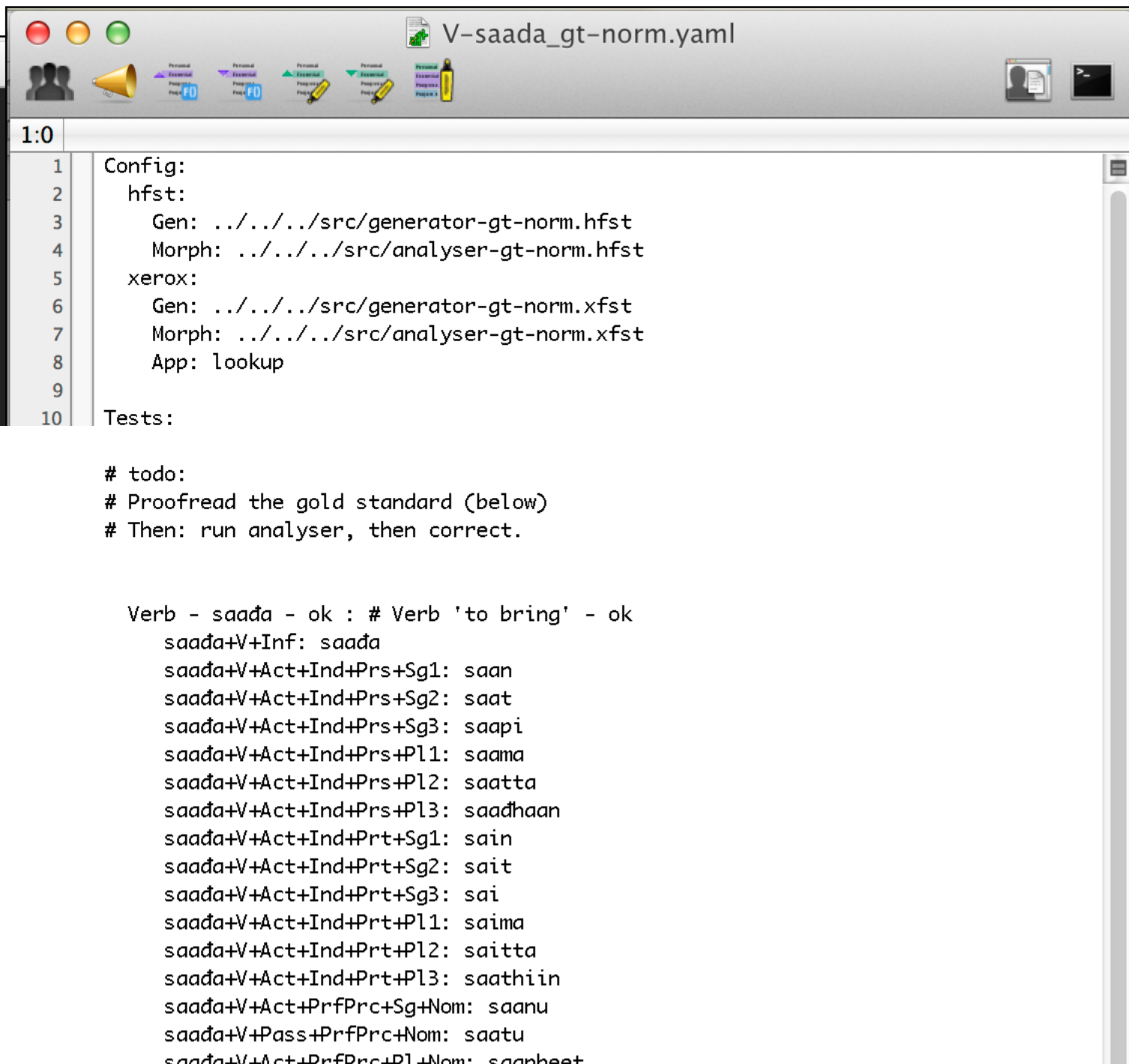
<

Den siste versjonen av boka:  
**main/langs/fkv/doc**

# *Filer redigeres med SubEthaEdit og brukes med Terminal/Pääte*



```
mervihaavisto — bash — 80x24
Last login: Tue Aug 19 15:29:55 on ttys007
Mervi-MacBook-Pro:~ mervihaavisto$
```

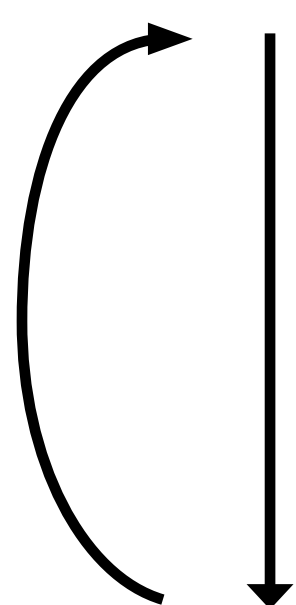


```
1:0
1 Config:
2   hfst:
3     Gen: ../../../../src/generator-gt-norm.hfst
4     Morph: ../../../../src/analyser-gt-norm.hfst
5   xerox:
6     Gen: ../../../../src/generator-gt-norm.xfst
7     Morph: ../../../../src/analyser-gt-norm.xfst
8     App: lookup
9
10 Tests:

# todo:
# Proofread the gold standard (below)
# Then: run analyser, then correct.

Verb - saada - ok : # Verb 'to bring' - ok
saada+V+Inf: saada
saada+V+Act+Ind+Prs+Sg1: saan
saada+V+Act+Ind+Prs+Sg2: saat
saada+V+Act+Ind+Prs+Sg3: saapi
saada+V+Act+Ind+Prs+Pl1: saama
saada+V+Act+Ind+Prs+Pl2: saatta
saada+V+Act+Ind+Prs+Pl3: saadhaan
saada+V+Act+Ind+Prt+Sg1: sain
saada+V+Act+Ind+Prt+Sg2: sait
saada+V+Act+Ind+Prt+Sg3: sai
saada+V+Act+Ind+Prt+Pl1: saima
saada+V+Act+Ind+Prt+Pl2: saitta
saada+V+Act+Ind+Prt+Pl3: saathiin
saada+V+Act+PrfPrc+Sg+Nom: saanu
saada+V+Pass+PrfPrc+Nom: saatu
saada+V+Act+PrfPrc+Pl+Nom: saanbeet
```

Arbeidsprosess



oppdater  
rediger  
test  
oppdater  
last opp



*Stammer og endelser tilsettes til lexc-filer*  
*stamme (stems): 'saa' -> endelse (affixes): 'da'*  
*- i mapper morphology/stems og affixes*

Nye ordboksord tilsettes til analysatoren:

- i stems-filer i ordklasser (n\_x v\_x osv)
- og klassifiseres etterpå (n\_32 v43\_odd).

# *Yaml-testfiler lages av visse typer av ord, og ordbøyning testes i Terminalen*

```
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'
[12/34] [PASS] saada+V+Act+Ind+Prt+Pl2 => saittavn up
[13/34] [PASS] saada+V+Act+Ind+Prt+Pl3 => saathiinvn up
[14/34] [PASS] saada+V+Act+PrfPrc+Sg+Nom => saanu st
[15/34] [PASS] saada+V+Pass+PrfPrc+Nom => saatu st
[16/34] [PASS] saada+V+Act+PrfPrc+Pl+Nom => saanheet st
[17/34] [PASS] saada+V+Act+PrsPrc+Sg+Nom => saava st
[18/34] [PASS] saada+V+Act+Cond+Sg1 => saaisvn st
[19/34] [PASS] saada+V+Act+Cond+Sg2 => saaitst
[20/34] [PASS] saada+V+Act+Cond+Sg3 => saaitst
[21/34] [PASS] saada+V+Act+Cond+Pl1 => saaisima
[22/34] [PASS] saada+V+Act+Cond+Pl2 => saaittima
[23/34] [PASS] saada+V+Act+Cond+Sg3 => saais 490 history
[23/34] [FAIL] saada+V+Act+Cond+Sg3 => Unexpected results: sapais, sasais, sakais,
satais [FAIL] saada+V+Act+Cond+Sg3 => Unexpected results: sapais, sasais, sakais,
satais
[24/34] [PASS] saada+V+Act+Cond+Pl1 => saaisima MacBook-Pro:fkvnob mervihaavisto$ svn t
[24/34] [FAIL] saada+V+Act+Cond+Pl1 => Unexpected results: sataisima, sakaisima, s
```

- make check
- kommandoen finner man under testord

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/generator-gt-norm.xfst ../../gt-norm-yamls/V-saada\_gt-norm.yaml; popd

*Stadieveksling lages i twolc-filen, f.eks for-  
kortninger: 'saa : sain : saisin' eller endringer 'k-> j'*

The screenshot shows a computer screen with two windows. The left window is a text editor displaying Finnish morphological rules in TWOLC format, with line numbers 741-773. The right window shows a phonological map for the word 'talo' (house) in Finnish, with line numbers 423-444. The map shows the word 'talo' in its base form and its inflected forms: 'talo>ss^A' (genitive singular), 'talo>ss^A' (genitive plural), and 'taloOssa' (locative singular). The map also shows the word 'talo' in its base form and its inflected forms: 'talo>ss^A' (genitive singular), 'talo>ss^A' (genitive plural), and 'taloOssa' (locative singular).

```
mappen fkv/src/phonology/fkv-phon.twolc
```

lexc talo+N+Sg+Ine MORFOLOGI - ENDRINGER - LEXC

talo>ss^A

twolc talo>ss^A

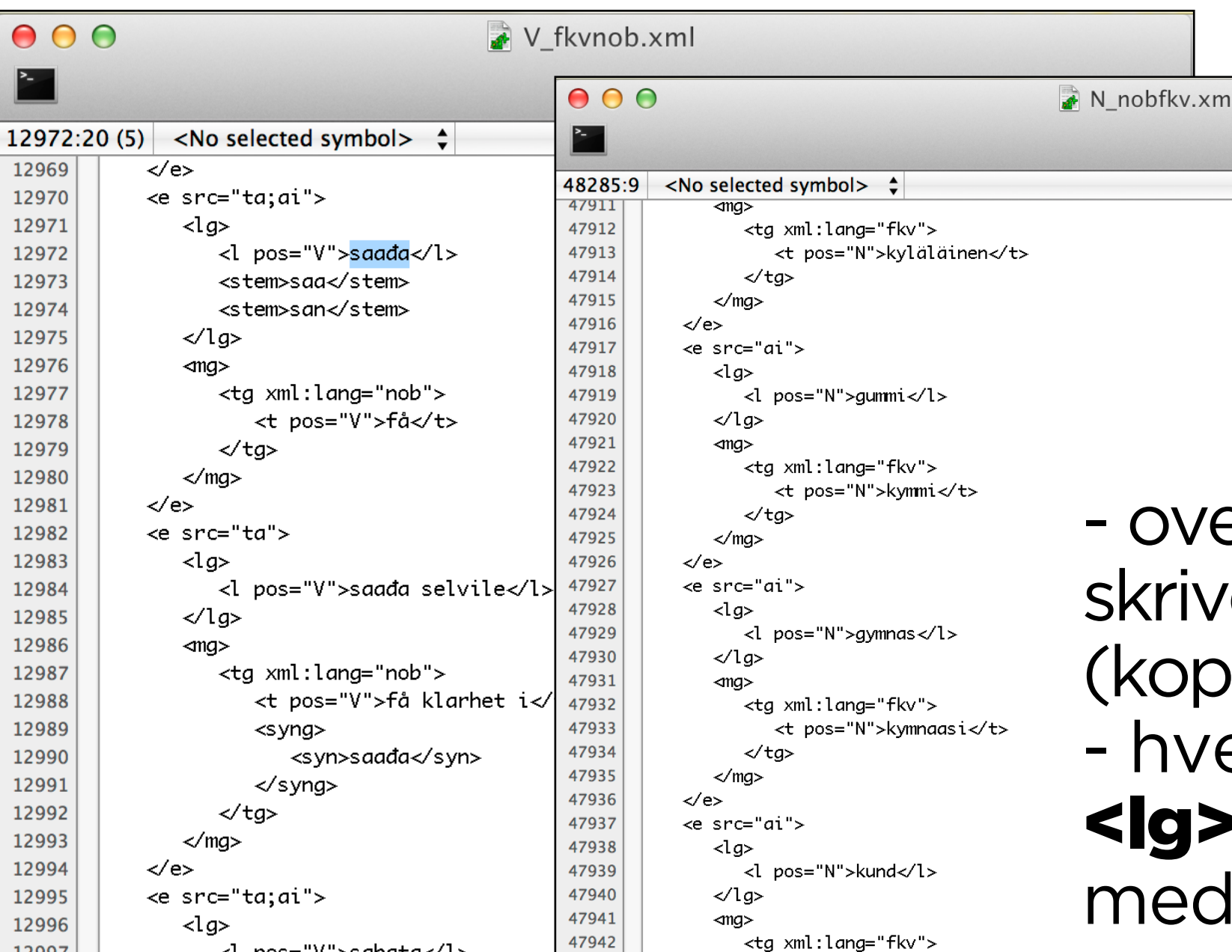
ta|o0ssa

MORFONOLOGI - PROSESSER - TWOLC



# Ordboksfiler finner i ordklasser på begge språk **fkvnob** og **nobfkv**

- i mapper `main/words/dicts/` **fkvnob** / `src/` **nobfkv**



The screenshot shows two XML editor windows. The left window, titled 'V\_fkvnob.xml', displays lines 12969 to 12997. The right window, titled 'N\_nobfkv.xml', displays lines 47911 to 47942. Both windows show XML code for dictionary entries. In the left window, the word 'saada' is highlighted in blue. The right window shows a similar entry for 'saada selvile'.

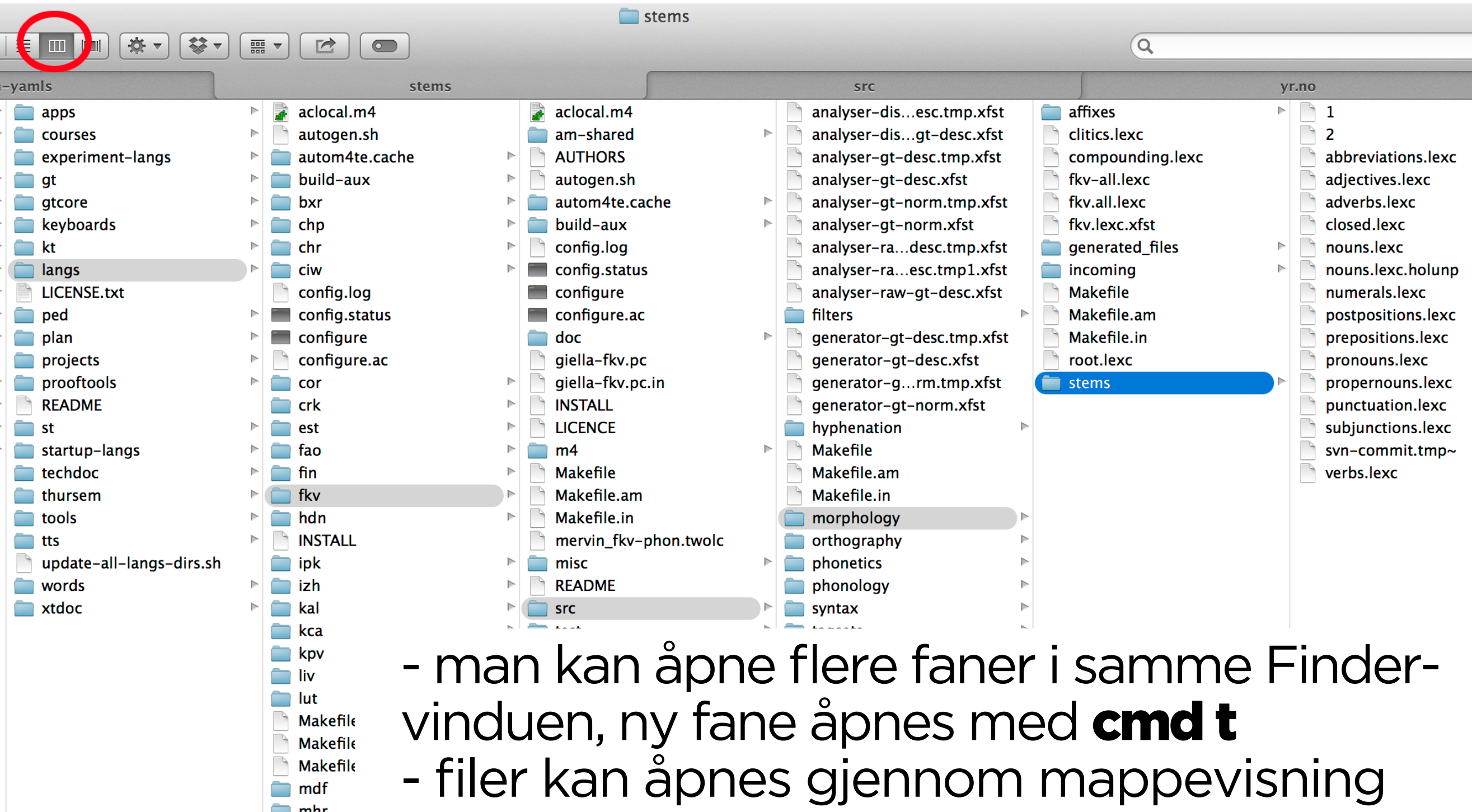
```
12969 </e>
12970 <e src="ta;ai">
12971   <lg>
12972     <l pos="V">saada</l>
12973     <stem>saa</stem>
12974     <stem>san</stem>
12975   </lg>
12976   <mg>
12977     <tg xml:lang="nob">
12978       <t pos="V">få</t>
12979     </tg>
12980   </mg>
12981 </e>
12982 <e src="ta">
12983   <lg>
12984     <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">saata</l>
```

```
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>
```

- oversettelse **<e></e>** skrives mellom entry-tagger (kopier og lim inn)
- hvert tag begynner f.eks. **<lg>** og taggen må slutes med **</lg>**

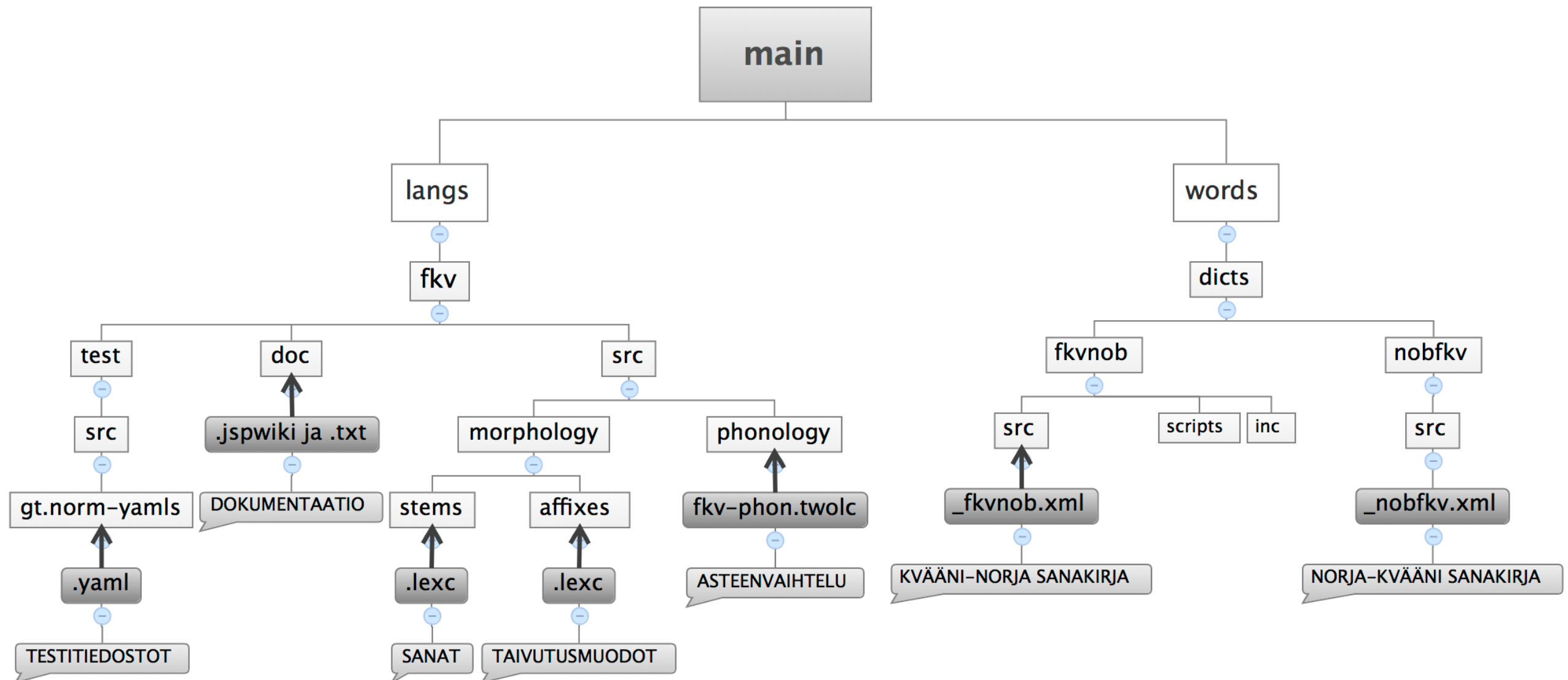
*Mapstruktur i maskinen ser slik ut i Finder, her vises det stien til stammer-mappen*



- man kan åpne flere faner i samme Finder-vinduen, ny fane åpnes med **cmd t**
- filer kan åpnes gjennom mappevisning



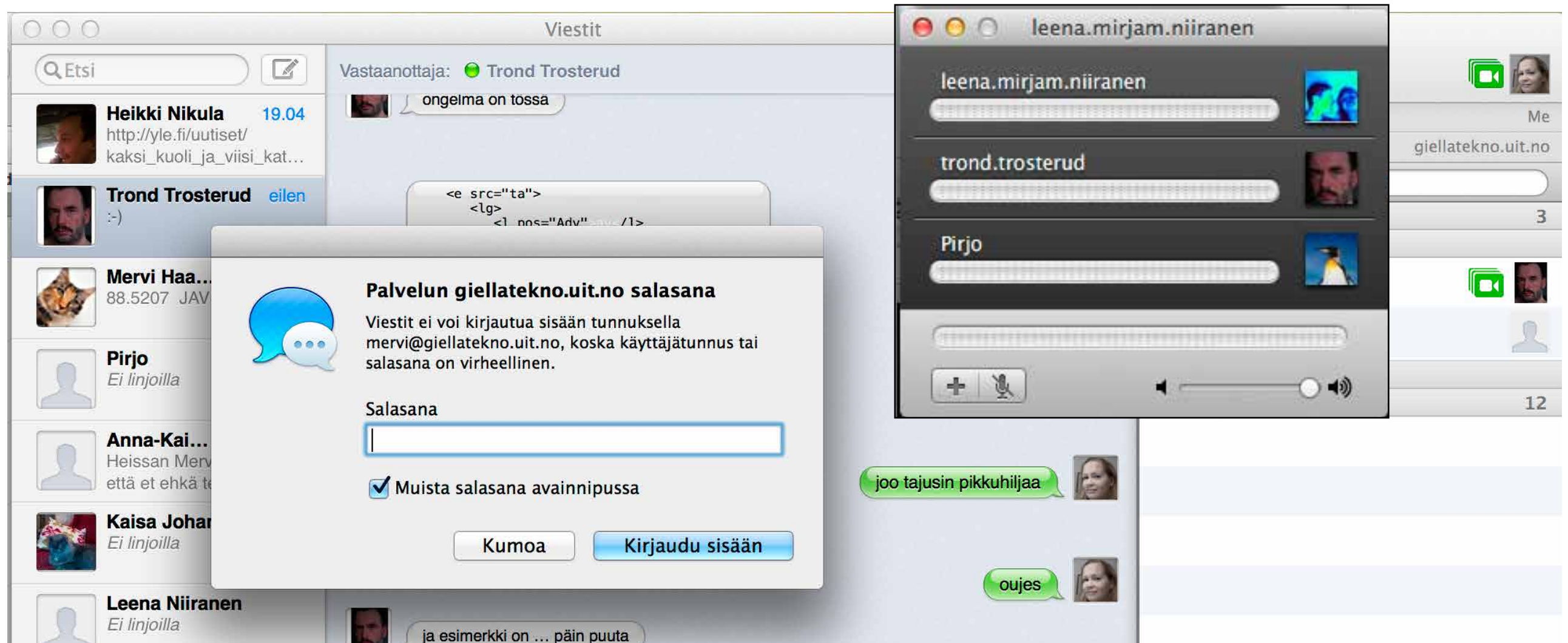
*Mapstrukturen i prosjektet vårt ser vanligvis slik ut:*



***Analysator***

***Ordbok***

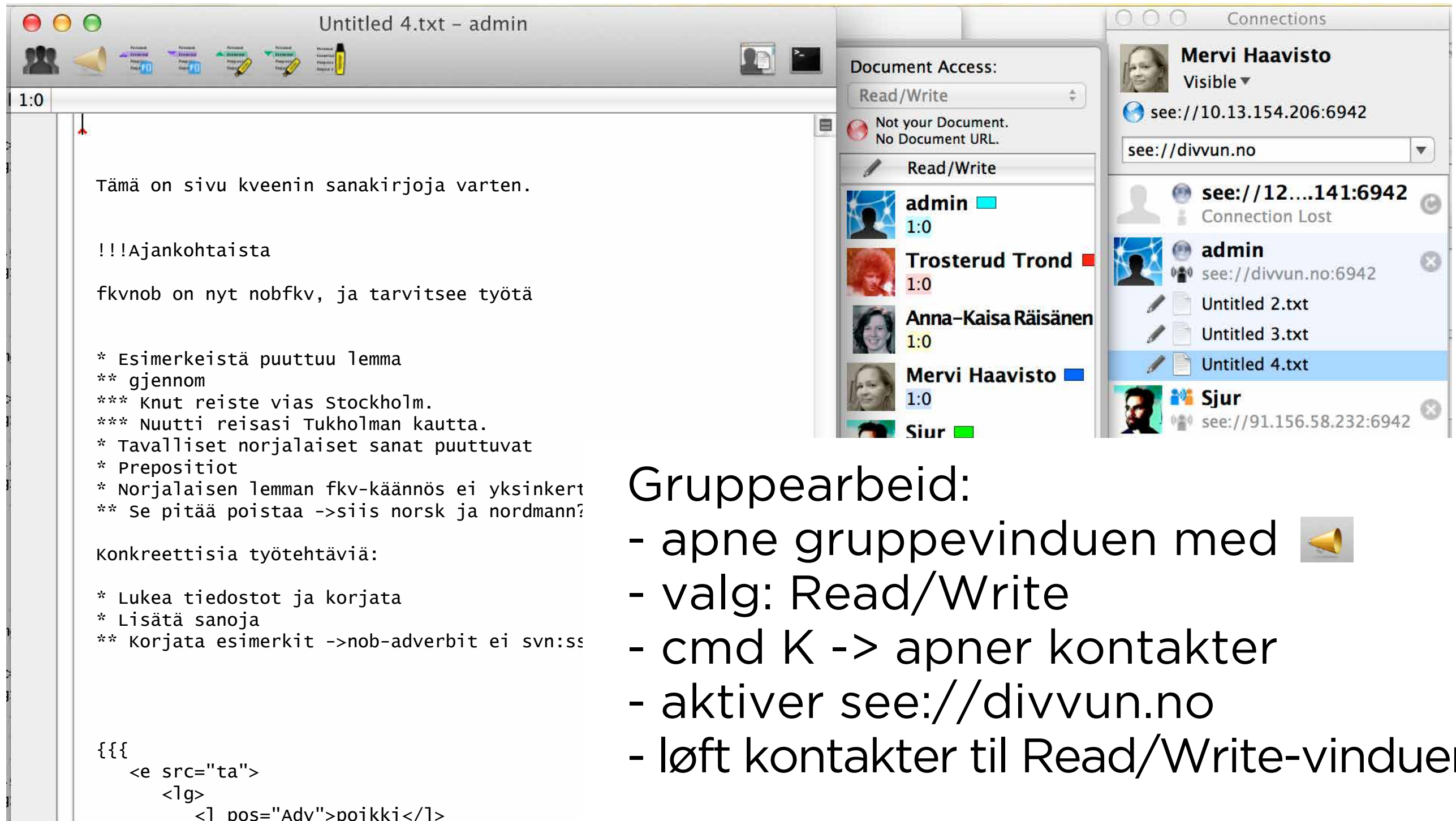
# *Meldinger/Viestit i Mac er Giellateknos kommunikasjonskanal: chat, samtaler, møter*



Opprette et konto med Gmail og søk kontakter



# *Editorprogram SubEthaEdit kan også brukes som gruppeverktøy*



The screenshot displays the SubEthaEdit application interface. The main window, titled "Untitled 4.txt - admin", contains a text document with 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 highlighted, showing a list of documents: Untitled 2.txt, Untitled 3.txt, and Untitled 4.txt.

## Gruppearbeid:

- apne gruppevinduen med 
- valg: Read/Write
- cmd K -> apner kontakter
- aktiver see://divvun.no
- løft kontakter til Read/Write-vinduen



# *Mest vanligste kommandoer i analysator- og ordboksarbeid*

**cd** Change Directory, tilbake hjem

**cd ..** et steg opp i filsystem

## **fkv analysator**

cd main/langs/fkv/...

## **kvääni-norja ordboksfiler**

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

**svn up** den siste versjonen av

versjonskontrollsystemet, gjøres alltid i begynnnelse

**svn diff** viser forskjellen mellom to filer eller mapper

**svn st** viser redigerte filer, som ikke har blitt legget inn i versjonskontrollsystemet, f.eks:

**A** addition

# *Mest vanligste kommandoer i analysator- og ordboksarbeid*

**M** modification (local modification)

**U** updated

**svn ci -m “Tähän kirjoitat viestin.” src/morphology/affixes/nouns.lexc** kommandoen oppdaterer nouns.lexc-filen til versjonskontroll, skriv inn en kommentar om dine endringer

**history** viser tidligere kommandoer

f.eks. **!531** gjentar kommandoen nr. 531

**pwd** Print Working Directory, sjekk hvor du er

**make** lager en ny analysator

**make check** en ny analysator og testning av ord

**see** åpner filen i SubEthaEdit, oppgi også navnet av filen i kommandoen

# *Mest vanligste kommandoer i analysator- og ordboksarbeid*

**ls** list, lister innholdet av mappen

**man ls** og **cd --help** help-kommandoer

**mkdir** lager en ny mappe

**svn log | less tiedostonnimi** viser logfilen (en skjermvisning)

**svn up -r 12345** tilbake til en tidligere versjon

**xmllint** (og filnavnet) viser xml-koden til *ordboka*  
f.eks. for sjekking

**ufkv** 'up fkv', paika1a gir bøyning til ord:

paika1a paikka+N+Sg+Ade

paika1a paika1a+Adv

(forlat med **ctrl c**)

**dfkv** 'down fkv', f.eks. paikka+N+Sg+Ade

gir bøyet ord: paikka+N+Sg+Ade paika1a



# OPPLASTNING AV FILER (I FKV-MAPPEN)

**save**

**svn up**

**make** (hvis i analysatoren)

**svn st** (hvis nødvendig)

**svn ci -m “*Tähän kirjoitat viestin.*” src/morpho-  
logy/affixes/nouns.lexc** oppdaterer denne filen til  
versjonskontroll

**svn ci -m “*Tähän kirjoitat viestin.*” src/morpho-  
logy/** oppdaterer alle redigerte filer i denne  
mappen til versjonskontroll

# KONFLIKT I FILEN

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** ordne konflikten senere

**diff-full** viser lokale endringer

**edit** rediger filen i SEE

**mine-conflict** hvis du vil lagre egne endringer

**theirs-conflict** hvis du vil få andres endringer

Du kan også kopiere engringene til en annen fil, og sjekke forskjellene og slå dem sammen

```
cp tiedostonnimi.lexc varatiedosto.lexc (cp kopioi)  
diff tiedostonnimi.lexc varatiedosto.lexc
```

# AVSLUTNING AV PROGRAMMER

ufkv/dfkv **ctrl c**

ut av søk, help osv. **q**

Emacs-editor **ctrl x** & **ctrl c**

SubEthaEdit **cmd q**