



UNIVERSITETET I OSLO
DET HUMANISTISKE FAKULTET

Spoken language corpora, with special focus on The Nordic Dialect Corpus

—

Janne Bondi Johannessen

PhD training course: Infrastructural tools for the study of linguistic variation
Fefor Høifjellshotell, Gudbrandsdalen,
Norway, 2.-6. June, 2009



Spoken language corpora



- Purposes of a spoken language corpus
 - Study aspects of language as with written corpora
 - In addition, studies of:
 - Discourse
 - Phonology
 - Dialect
 - Diachronic change
 - Agelect
 - ...



How make spoken data into a corpus?


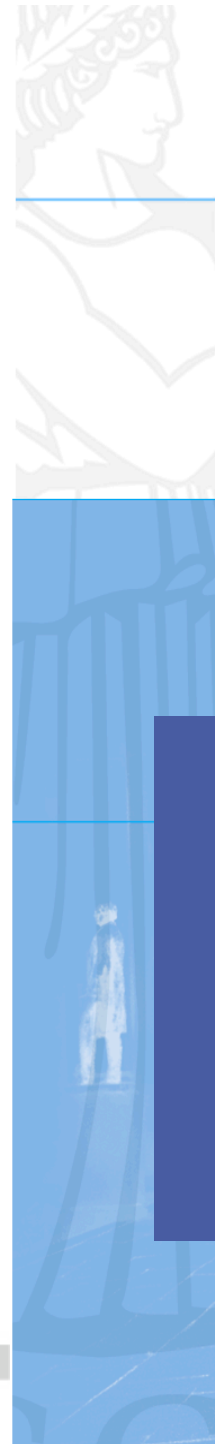
- Digitise sound/video
- Transcribe speech
 - Choose transcription
 - Phonetic
 - Orthographic
- Time code transcription for alignment with sound file
- Annotate
 - Extra-linguistic features (laughter, coughing)
 - Semi-linguistic features (meaningful sounds)
 - Gestures
 - Grammatically
 - POS tagging
 - Syntactic parsing
 - Other

Different situation and informants - different speech types


- The Big Brother Corpus
 - Pros
 - Lots of spontaneous speech data
 - Lots of dialogue and polylogue
 - Lots of emotional speech in different dialogue situations
 - Conflict, argument, love, irritation etc.
 - Cons
 - Not a dialect corpus
 - No representativity w.r.t. age, social class, education etc.
 - Not "controlled" recording situations
 - Small number of informants
- Nordic Dialect Corpus
 - Pros
 - Lots of spontaneous speech data
 - Lots of dialogue, **no** polylogue
 - **No** emotional speech
 - Cons
 - Is a dialect corpus
 - Partly representative
 - Controlled recording situation
 - Small AND large number of informants



Big Brother vs Nordic Dialect Corpus



[Trouble viewing video?](#)
context±
Offset
Left
Right
- Start +
- Stop +
>>



[Trouble viewing video?](#)
context±
Offset
Left
Right
- Start +
- Stop +
>>



Other dialect corpora?

We know of no comparable resource for any language

- *Sounds familiar? Accents and Dialects of the UK*
 - No grammatical search options
 - No results handling
- *The British National Corpus*
 - No audio
 - Orthographic transcription
 - Unreliable dialect categories
- *The DynaSand dialect database*
 - Few spontaneous utterances
- *The Spoken Dutch Corpus*
 - Not web-based
 - Orthographic transcriptions
 - Not dialect data
- The Scottish Corpus of Text and Speech
 - Not a dialect corpus
 - No searchable linguistic features
- Others under development:
 - Corpus of Estonian Dialects
 - Spoken Japanese Dialect Corpus
- Paul Thompson at the University of Reading: Posting at Corpora List 30 Nov. 2008 about linked audio or video files with transcripts: 15 answers, of which only one on dialects: ours



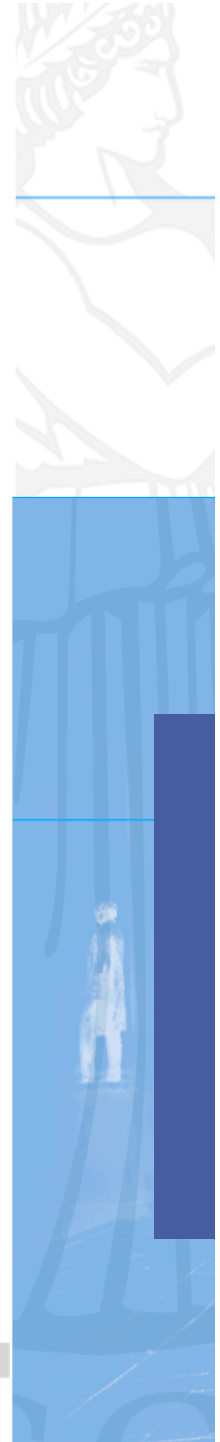
Comparison: NoTa-Oslo, Talesøk, BySoc, BNC, SCOTS

	NoTa	Talesøk	GSLC	BySoc	BNC	SCOTS
Transcription linked to audio	Yes	Yes	No	No	No	Yes
Transcription linked to video	Yes	No	No	No	No	Yes
User-friendly search without regular expressions	Yes	Yes	No	No	No	No
Possible to limit informant selection	Yes	Yes	Yes	Yes	Yes	Yes
Overlaps/ turntaking annotated	Yes	Yes	Yes	Yes	No	Yes
Transcription as standard orthography (or slightly modified)	Yes	Yes	Yes	Yes	Yes	Yes
POS tagged	Yes	No	No	No	Yes	No
POS tags can be used as the only search expressions	Yes	–	–	–	No	–



Credits

- Collaboration between the research network ScanDiaSyn and Nordic Centre of Excellence in Microcomparative Syntax (NORMS)
- Some of the corpus is financed by national research councils in the individual countries
- The technical development has been financed by the University of Oslo, the Norwegian Research Council, and the Nordic research funds NOS-HS and NordForsk



Why the Nordic Dialect Corpus was developed

- Initiated by members of Nordic Centre of Excellence in Microcomparative Syntax and the ScanDiaSyn network
- Overarching goal: to study the dialects of the North-Germanic dialect continuum
 - The Nordic languages are closely related and have some mutually intelligibility
 - Studying the dialects within each national language is misguided from a theoretical and principled point of view
 - Difficult for each researcher to get hold of relevant data on their own in such a large area.
 - Many different kinds of data needed for syntax research



Corpus features

- Linguistic contents
 - Dialects from five closely related languages
- Annotation
 - POS tagging and two types of transcription
- Search interface
 - Advanced possibility to combine an array of search criteria and results presentation in an intuitive interface
- Many search variables
 - Linguistics-based, informant-based, time-based
- Multimedia display
 - Linking of sound and video to transcriptions
- Display of informant details
 - Number of words and other informant information
- Advanced results handling
 - Concordances, collocations, counts, statistics ...
- The corpus is available on the web



Linguistic contents and numbers

- The corpus contains dialect data from the national languages Danish, Faroese, Icelandic, Norwegian, and Swedish
- Contains speech data from app. 240 informants with 521 000 words (by 7 May 2009)
- The goal: 600 informants, and many more words
- All the recordings represent spontaneous speech
- Differences in data collection due to differences in financing
 - Norwegian, Oevdalian, and some Danish: two kinds of recordings per informant:
 - semi-formal interview
 - informal conversation between two informants
 - Recordings of both young and old informants, both genders
 - Both new and old recordings
 - Audio or both audio and video recordings



Annotation: transcription

- Each dialect has been transcribed by the standard official orthography of that country
 - In addition all the Norwegian dialects and some Swedish dialects have also been transcribed phonetically
 - The Norwegian phonetic transcription follows roughly that of Papazian and Helleland (2005). The transcription of the Oevdalian dialect follows the Oevdalian orthography (standardised in 2005 by the *Rådjörum* – The Oevdalian Language Council).
 - The phonetic transcription is translated to an orthographic transcription via a semi-automatic dialect transliterator



Annotation: tagging

- The corpus will be POS tagged, with selected morpho-syntactic features language by language
- Norwegian
 - POS tagged by a TreeTagger first developed for the Corpus of Oslo spoken Norwegian (Søfteland og Nøklestad 2008), and used unchanged for the dialect corpus. (Accuracy of 96.9%)
- Swedish
 - A TnT POS tagger developed by Sofie Johansson-Kokkinakis (2003) for written Swedish is being used to tag the Swedish dialect data. The corrected result will be used to train a TreeTagger.

Search interface – Glossa

Scandinavian Dialect
Corpus

Glossa ([my results](#) | [my annotations](#) | [statistics](#) | [full query](#) | [help](#))

criteria»

+
-



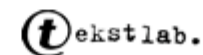
[Recording locations](#)
[Transcriptions](#)

Regular expressions: <input type="checkbox"/>	Hits per page: <input type="text" value="20"/>	Randomize <input type="checkbox"/>	Orthographic <input checked="" type="radio"/>	<input type="button" value="Search corpus"/>
Search within: <input type="text" value="s"/>	Max results : <input type="text" value="2000"/>	Skip tot. freq. <input checked="" type="checkbox"/>	Phonetic <input type="radio"/>	<input type="button" value="Reset form"/>
informant ⁺				<input type="button" value="Show texts"/>
<hr/>				<input type="button" value="Save subcorpus"/>
country ⁺	region ⁺	area ⁺	place ⁺	Choose subcorpus
<hr/>				
agegroup ⁺	sex ⁺	rec (year) ⁺	genre ⁺	
<hr/>				
Display: <input type="text"/>	Search within: <input type="text"/>			



Denne Glossaversjonen er under utvikling.
This version of Glossa is undergoing development.

[Please report bugs and errors here](#)



Searching for lemmas

gås

criteria»

- word » ▶ lemma
- occurrences » ▶ start of word
- pos » ▶ within word
- num » ▶ end of word
- degr » ▶ case sensitive
- case » ▶ exclude word
- sex » ▶ add word form
- nlex » ▶ add negated word form
- pers » ▶ add lemma
- temp » ▶ add negated lemma
- def » ▶
- descr » ▶
- type » ▶
- phonetic » ▶

Re Se

per page: Randomize


results : Skip tot. freq. Orthographic Phonetic

inf

country region area place

agegroup sex rec (year) genre

Display: Search within:



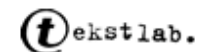
Recording locations
Transcriptions

[Choose subcorpus](#)



Denne Glossaversjonen er under utvikling.
This version of Glossa is undergoing development.

Please report bugs and errors [here](#)



Searching for more than one word

ig	interval:		+
criteria»	min	criteria»	-
end of word	3 max	konj	



[Recording locations](#)
[Transcriptions](#)

Regular expressions: <input type="checkbox"/>	Hits per page: <input type="text" value="20"/>	Randomize <input type="checkbox"/>	Orthographic <input checked="" type="radio"/>
Search within: <input type="text" value="s"/>	Max results : <input type="text" value="2000"/>	Skip tot. freq. <input checked="" type="checkbox"/>	Phonetic <input type="radio"/>

Search corpus

Reset form

Show texts

Save subcorpus

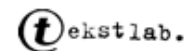
[Choose subcorpus](#)

Display: Search within:



Denne Glossaversjonen er under utvikling.
This version of Glossa is undergoing development.

[Please report bugs and errors here](#)



Search results

CWB expression: "([((word=".*ig" %c))] []{0,3} [((pos="konj"))]) ;"

Informants: 240

scandiasyn:

CWB expression: "([((word=".*ig" %c))] []{0,3} [((pos="konj"))]) ;"

Action :

: 474

Results pages: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [23](#) [24](#)

- [i](#) [M](#) [K](#) alvdal_02 påbygning # eller det er ikke så **veldig herlig det men**
- [i](#) [M](#) [K](#) alvdal_01 det er godt å bli **ferdig snart da ja men** (uforståelig) jævlig mye å gjøre i år
- [i](#) [M](#) [K](#) alvdal_01 ikke sett n # så nå har jeg jo **plutselig fire sjøl da for** jeg hadde jo tre # tre sjø
- [i](#) [M](#) [K](#) alvdal_02 tjener ikke så **jævlig mye men** # det blir jo litt
- [i](#) [M](#) [K](#) alvdal_01 ja det s- ja synes jeg ser så **herlig ut * men** jeg ...
- [i](#) [M](#) [K](#) alvdal_02 * m * yes * (latter) * det ser **herlig ut ja men** det er vel ett nålhue å komme gj

Some results presented as frequency list

occurences	match
8	artig og
6	rolig og
6	egentlig men
3	mulig og
3	vanskelig og
3	egentlig # men
3	viktig og
3	mulig men
2	tidlig for
2	veldig kjekt og
2	frodig og
2	veldig fin oppvekst og
2	tidlig e begynne og
2	vanlig skole og
2	artig det og
2	veldig men
2	egentlig og
2	egentlig heile # oppveksten og
2	forferdelig råhølja heter det og
2	veldig gode venner i_hvert_fall og
2	veldig fin natur og
2	grådig nå at men

Searching for part of speech



Recording locations
Transcriptions

Search corpus
Reset form
Show texts
Save subcorpus
Choose subcorpus

Re Se
inf
country +
agegroup +
Display: []

criteria»
word »
occurrences »
pos »
num »
degr »
case »
sex »
nlex »
pers »
temp »
def »
descr »
type »
phonetic

(spm)
adj
adv
cbl
det
inf-merke
interj
konj
pause
pause2
prep
pron
pron/det
sbu
subst
subst:adj
sänn
ukjent
verb
verb:subst
ambiguous

Hits per page: 20
Max results: 2000
Randomize
Skip tot. freq.
Orthographic
Phonetic

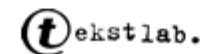
area +
place +
genre +

Search within: []



Denne Glossaversjonen er under utvikling.
This version of Glossa is undergoing development.

Please report bugs and errors [here](#)



Phonetic querying

Scandinavian Dialect
Corpus

Glossa ([my results](#) | [my annotations](#) | [statistics](#) | [full query](#) | [help](#))

um +
criteria» -
phonetic ▾
end of word ▾



[Recording locations](#)
[Transcriptions](#)

Regular expressions:

Search within:

Hits per page:

Max results:

Randomize

Skip tot. freq.

Orthographic

Phonetic

informant +

country +

region +

area +

place +

agegroup +

sex +

rec (year) +

genre +

[Choose subcorpus](#)

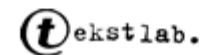
Display:

Search within:



Denne Glossaversjonen er under utvikling.
This version of Glossa is undergoing development.

[Please report bugs and errors here](#)



Displaying results

aasen_35 ja var det inte det ?

aasen_48 **ojoj och hur fin (uforstaelig) tank att i gar # sa pratade M1 och jag just om ... # for da hade vi bröllopsdag i gar**

aasen_35 jaha ?



[Trouble viewing](#)

[video?](#)

context

± - ▾

- Start +

- Stop +

>>

Informants: 240

scandiasyn:

CWB expression: "`(((phon="um.*" %c)))` ;"

Action :

: 242

Results pages: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#)

- i 🗨 🗨 aasen_48 ojoj och hur fin (uforstaelig) tank att i gar # sa pratade M1 och jag just **om** ... # for da hade vi bröllopsdag i gar
- i 🗨 ls men vad gjorde **om** om rasterna ?
- i 🗨 ls men vad gjorde om **om** rasterna ?
- i 🗨 🗨 aasen_35 och eh # till slut var det någon som var in där och tit- skulle se efter **om** hon hade haft varit darinne och där stod en slik # stor # brödkartong
- i 🗨 ls var det göra det_där byta **om** ? lära sig något nytt ?
- i 🗨 🗨 aasen_35 det det var ju så det var ju många # stockholmare som var här **om** sommaren # över i gårdarna och hade ju lite # lärde ju oss lite av dem
- i 🗨 🗨 aasen_35 ja det var verkligen riktig bocksvenska **om** de förstod det vet jag inte

- Phonetic transcription

... **kuogeðum** wi:ð å: gamtkuorteð og då: fann ig att # iet n ... brauðkuort frá dett djæstbuoð
ja # før at då: wår femtisjæks se: so fyøddes ju pojken og **djifteðum** uoss åreð firiað ja # ja
ojojoj og ur fin kommentar tænk at i går # so **prateðum** M1 og ig just um ... # før då: addum
ojojoj og ur fin kommentar tænk at i går # so prateðum M1 og ig just **um** ... # før då: addum

Orthographic transcription

... **tittade** vi på gammal_kortet och då fann jag åter # ett n ... brudkort från det bröllopet
ja # för att då vår femtiosex sedan så föddes ju pojken och **gifte** oss året dessförinnan ja # ja
ojojoj och hur fin (uforståelig) tänk att i går # så **pratade** M1 och jag just om ... # för då hade v
ojojoj och hur fin (uforståelig) tänk att i går # så pratade M1 och jag just **om** ... # för då hade vi

Display of transcription and tagging

CWB expression: "(((word="bil.*" %c))) ;"

Action:

Hits found: 124

Results pages: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#)

- alvdal_02** så skal jeg ta førerkort på bil nå # når jeg bli atten # så er jo blir litt godt å ## endelig få kjøre **bil**
- alvdal_02** ja jeg # kjøpte meg **bil** i sommer # som jeg har pussa opp att
- alvdal_02** har du traktor så # eller **bil** så kjører du heller det enn å sparke # går litt fortere
- alvdal_04** nei da så er da mye men nå er det vel så du kommer ikke innover med **bil** heller nå vil jeg tru
- alvdal_03** nye veien utover er så stille og så utoverbakke det er så at **bilene** bare trille utover jeg hører dem ikke i det hele tatt
- alvdal_04** ja # ja når du skal ut med **bil** så er det vel det
- alvdal_03** men sa at jeg tar da ikke **bil** for å reise på skitur # hvis jeg i
- alvdal_03** og nå har jeg blitt for lat så jeg gidder ikke å gå opp her heller kjøre på Klebersvanga det er helst der jeg kjører for jeg tør ikke kjøre opp
- alvdal_03** og jeg tenkte at jeg kommer aldri hjem att med denne derre de type: appell er
- alvdal_03** og e så var jeg en ta disse postfolkene så da # " har du fått pro defn: be **bilen** ? "
- alvdal_03** " ja " sa " jeg har gjort det " men e ja vi har skulle ringe på n M nlex: 100 en det er ikke han som er # e der jeg går med **bilen** " sa jeg # men det v
- alvdal_03** ja jeg hadde ikke lovt han noe men han ville gjerne ha **bilen**
- alvdal_04** ja dem er aldeles sjuk på gamle **biler** den # er m hun ene der a F9 hun kjøpte denne gamle Ladaen som jeg hadde vet du jeg hadde hadde Lada jeg
- alvdal_04** men så byn- ja ja en gikk den men hun var aldeles sjuk på n vet du denne **bilen** så hun # kjøpte nå for en billig penge der
- alvdal_04** men så byn- ja ja en gikk den men hun var aldeles sjuk på n vet du denne bilen så hun # kjøpte nå for en **billig** penge der
- alvdal_03** hva_for en **bil** du har nå ?
- alvdal_04** han driver på med **bilforretning** inni Oslo # så han skaffa meg denne der er en totusenmodell så den er det er en svær # firehjulstrekker og

lemma: bil
phon: bilan
pos: subst
sex: mask
num: fl
type: appell er
defn: be **bilen** ? "
descr: me
nlex: 100

Informant-based querying



criteria»

[Recording locations](#)
[Transcriptions](#)

Regular expressions:
Search within:

Hits per page:
Max results :

Randomize
Skip tot. freq.

Orthographic
Phonetic

Search corpus

Reset form

informant

aasen_35
aasen_48
aeroe2
aeroe3
als1
alvdal_01
alvdal_02

[>] [

choose

country

Denmark
Faroe
Iceland
Norway
Sweden

[>] [

choose

region

Gotland
Göteborg
Norland
Svealand

[>] [

choose

area

Blekinge
Bohuslän
Dalarna
Dalsland
Gotland
Gästrikland
Halland

[>] [

choose

place

als
alvdal
andøya
Ankarsum
Anundsjö
Arjeplog
Asby

[>] [

choose

agegroup

A
 B

sex

F
 M

rec (year)

1959
2000
2007
2008

[>] [

choose

genre

ikke_udskrevet
intervju
klip
samtale
-

[>] [

choose

Show texts

Save subcorpus

[Choose subcorpus](#)

Display information on informants 1

CWB expression: "(((word="bil.*" %c))) ;"

Action:

Hits found: 124

Results pages: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#)

[i](#) [M](#) [K](#) alvdal_02 så skal je

[i](#) [M](#) [K](#) alvdal_02 ja jeg # k

[i](#) [M](#) [K](#) alvdal_02 har du tra

[i](#) [M](#) [K](#) alvdal_04 nei da så

[i](#) [M](#) [K](#) alvdal_03 nye veien

[i](#) [M](#) [K](#) alvdal_04 ja # ja nå

[i](#) [M](#) [K](#) alvdal_03 men sa at

[i](#) [M](#) [K](#) alvdal_03 og nå har

[i](#) [M](#) [K](#) alvdal_03 og jeg ter

[i](#) [M](#) [K](#) alvdal_03 og e så v

[i](#) [M](#) [K](#) alvdal_03 " ja " sa "

[i](#) [M](#) [K](#) alvdal_03 ja jeg hadde ikke lovt han noe men han ville gjerne ha **bilen**

[i](#) [M](#) [K](#) alvdal_04 ja dem er aldeles sjuk på gamle **biler** den # er m hun ene der a F9 hun kjøpte denne gamle Ladaen som jeg hadde vet du jeg

[i](#) [M](#) [K](#) alvdal_04 men så byn- ja ja en gikk den men hun var aldeles sjuk på n vet du denne **bilen** så hun # kjøpte nå for en billig penge der

[i](#) [M](#) [K](#) alvdal_04 men så byn- ja ja en gikk den men hun var aldeles sjuk på n vet du denne bilen så hun # kjøpte nå for en **billig** penge der

[i](#) [M](#) [K](#) alvdal_03 hva_for en **bil** du har nå ?

Mozilla Firefox

http://omilia.uio.no/glossa/html/profile.php?tid=alvdal_03&corpus=scandiasyn

Informant details for *alvdal_03* in the Scandiasyn corpus

SELECT tid,sex,agegroup,country,region,area,place,wc,rec FROM SCANDIASYNauthor WHERE tid = 'alvdal_03'

Code	Sex	Age group	Country	Region	Area	Place	Word count	Recorded
alvdal_03	F	B	Norway			alvdal	4876	2008

Fullført

ele tatt

t der jeg kjører fo

jeg går med **bile**

Display information on informants 2

Word count for selected informants: 521179 (total for Scandiasyn corpus: 521179)

Informant details for in the Scandiasyn corpus

<u>Code</u>	<u>Sex</u>	<u>Age group</u>	<u>Country</u>	<u>Region</u>	<u>Area</u>	<u>Place</u>	<u>Word count</u>	<u>Recorded</u>
ankarsrum_om1	M	B	Sweden	Götaland	Småland	Ankarsum	2379	2000
ankarsrum_om3	M	B	Sweden	Götaland	Småland	Ankarsum	2641	2000
ankarsrum_ym1	M	A	Sweden	Götaland	Småland	Ankarsum	2374	2000
anundsjo_om2	M	B	Sweden	Norrland	Ångermanland	Anundsjö	4399	2000
anundsjo_ow3	F	B	Sweden	Norrland	Ångermanland	Anundsjö	6336	2000
arjeplog_om1	M	B	Sweden	Norrland	Norrbottnen	Arjeplog	4894	2000
arsunda_ow1	F	B	Sweden	Norrland	Gästrikland	Årsunda	2933	2000
arsunda_om1	M	B	Sweden	Norrland	Gästrikland	Årsunda	1395	2000
arsunda_om2	M	B	Sweden	Norrland	Gästrikland	Årsunda	1690	2000
arsunda_ow2	F	B	Sweden	Norrland	Gästrikland	Årsunda	2094	2000
asby_om2	M	B	Sweden	Götaland	Småland	Asby	2186	2000
asby_om3	M	B	Sweden	Götaland	Småland	Asby	1591	2000
asby_ym1	M	A	Sweden	Götaland	Småland	Asby	2873	2000
bara_om1	M	B	Sweden	Götaland	Skåne	Bara	3119	2000
bara_ow1	F	B	Sweden	Götaland	Skåne	Bara	1239	2000
bara_om3	M	B	Sweden	Götaland	Skåne	Bara	1238	2000
bara_om4	M	B	Sweden	Götaland	Skåne	Bara	3004	2000
bara_ow3	F	B	Sweden	Götaland	Skåne	Bara	1617	2000
bengtsfors_om1	M	B	Sweden	Götaland	Dalsland	Bengtsfors	1833	2000
bengtsfors_om2	M	B	Sweden	Götaland	Dalsland	Bengtsfors	2211	2000
bengtsfors_ym2	M	A	Sweden	Götaland	Dalsland	Bengtsfors	2981	2000
boda_om1	M	B	Sweden	Götaland	Öland	Böda	6116	2000
boda_om3	M	B	Sweden	Götaland	Öland	Böda	3507	2000
boda_ow1	F	B	Sweden	Götaland	Öland	Böda	2674	2000
boda_ow3	F	B	Sweden	Götaland	Öland	Böda	2294	2000
boda_ow2	F	B	Sweden	Götaland	Öland	Böda	1904	2000
bredsättra_om2	M	B	Sweden	Götaland	Öland	Bredsättra	3314	2000
bredsättra_om3	M	B	Sweden	Götaland	Öland	Bredsättra	2241	2000
bredsättra_ow1	F	B	Sweden	Götaland	Öland	Bredsättra	1046	2000
faro_om1	M	B	Sweden	Gotland	Gotland	Fårö	2879	2000
faro_om3	M	B	Sweden	Gotland	Gotland	Fårö	1392	2000

Action menu

CWB expression: "`((((word="bil.*" %c)))`);"

Action :

Hits four

Results p

- count
- download
- sort
- collocations
- annotate
- show metadata
- metadata distribution
- delete hits
- save hits

a førerkort på bil nå # når jeg bli atten # så er jo blir litt godt å ## endelig få kjøre **bil**

ste meg **bil** i sommer # som jeg har pussa opp att

or så # eller **bil** så kjører du heller det enn å sparke # går litt fortere

da mye men nå er det vel så du kommer ikke innover med **bil** heller nå vil jeg tru

alvdal_03 nye veien utover er så stille og så utoverbakke det er så at **bilene** bare trille utover jeg hører dem ikke i det h

alvdal_04 ja # ja når du skal ut med **bil** så er det vel det

alvdal_03 men sa at jeg tar da ikke **bil** for å reise på skitur # hvis jeg ikke skal inn i S1

alvdal_03 og nå har jeg blitt for lat så jeg gidder ikke å gå opp her heller jeg nå jeg må kjøre på Klebersvanga det er helst

alvdal_03 og jeg tenkte at jeg kommer aldri hjem att med denne derre derre **bilen** der

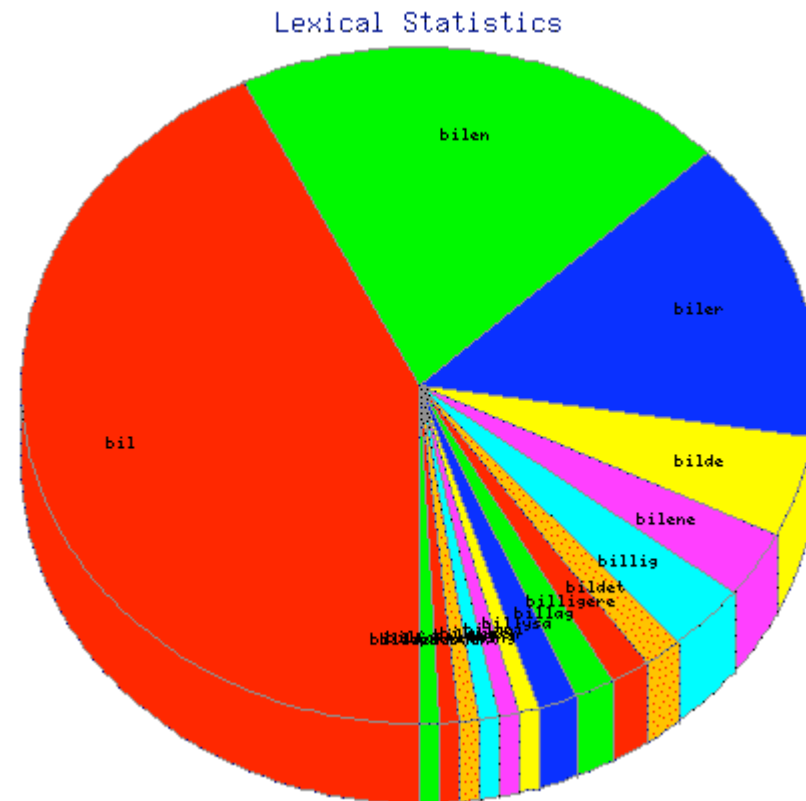
alvdal_03 og e så var jeg en ta disse postfolkene så da # " har du fått problemer med **bilen** ? "

alvdal_03 " ja " sa " jeg har gjort det " men e ja vi har skulle ringe på n M4 han # " ja men det er ikke han som er # e der je

alvdal_03 ja jeg hadde ikke lovt han noe men han ville gjerne ha **bilen**

Count

occurences	match
53	bil
25	bilen
18	biler
6	bilde
4	bilene
4	billig
2	bildet
2	billigere
2	billag
2	billysa
1	bilvei
1	bilveien
1	bilveg
1	bilforretning
1	bildeprosjekt
1	bildeler



Deleting or selecting some results

CWB expression: "(((word="bil.*" %c))) ;"

Action :

Hits found: 124































Results pages: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#)

Delete selection

select all

unselect all

Finished deleting

-    alvdal_02 så skal jeg ta førerkort på bil nå # når jeg bli atten # så er jo blir litt godt å ## endelig få kjøre **bil**
-    alvdal_02 ja jeg # kjøpte meg **bil** i sommer # som jeg har pussa opp att
-    alvdal_02 har du traktor så # eller **bil** så kjører du heller det enn å sparke # går litt fortere
-    alvdal_04 nei da så er da mye men nå er det vel så du kommer ikke innover med **bil** heller nå vil jeg tru
-    alvdal_03 nye veien utover er så stille og så utoverbakke det er så at **bilene** bare trille utover jeg hører dem ikke i det hele tatt
-    alvdal_04 ja # ja når du skal ut med **bil** så er det vel det
-    alvdal_03 men sa at jeg tar da ikke **bil** for å reise på skitur # hvis jeg ikke skal inn i S1
-    alvdal_03 og nå har jeg blitt for lat så jeg gidder ikke å gå opp her heller jeg nå jeg må kjøre på Klebersvanga det er helst der jeg (uforståelig)
-    alvdal_03 og jeg tenkte at jeg kommer aldri hjem att med denne derre derre **bilen** der
-    alvdal_03 og e så var jeg en ta disse postfolkene så da # " har du fått problemer med **bilen** ? "

Annotating results

CWB expression: "`((((word="hun" %c)))[(pos="subst"))]` ;"

Action :

Hits found: 96

Results pages: [1](#) [2](#) [3](#) [4](#) [5](#)

Save annotations (* indicates: no reviewed value stored.)

*

i **⌘** **⌘** [_mefjordvaer_19](#) ja men det hender jo at # at vi slår til og så bruker uttrykk og d- **hun datter** vår som bor i # Oslo hun em # hun jot

*

i **⌘** **⌘** [_mefjordvaer_19](#) " ja hvor fikk **hun tak** i de jernene ? " " jo det var smeden i Været " sa han # " som laga jernet til hun " sa han

*

i **⌘** **⌘** [_stange_03](#) og e # ja jeg vet ikke åssen jeg skal si det # nei jeg har akkurat e s- sagt åt **hun dama** de skrev at det jeg kan tenk

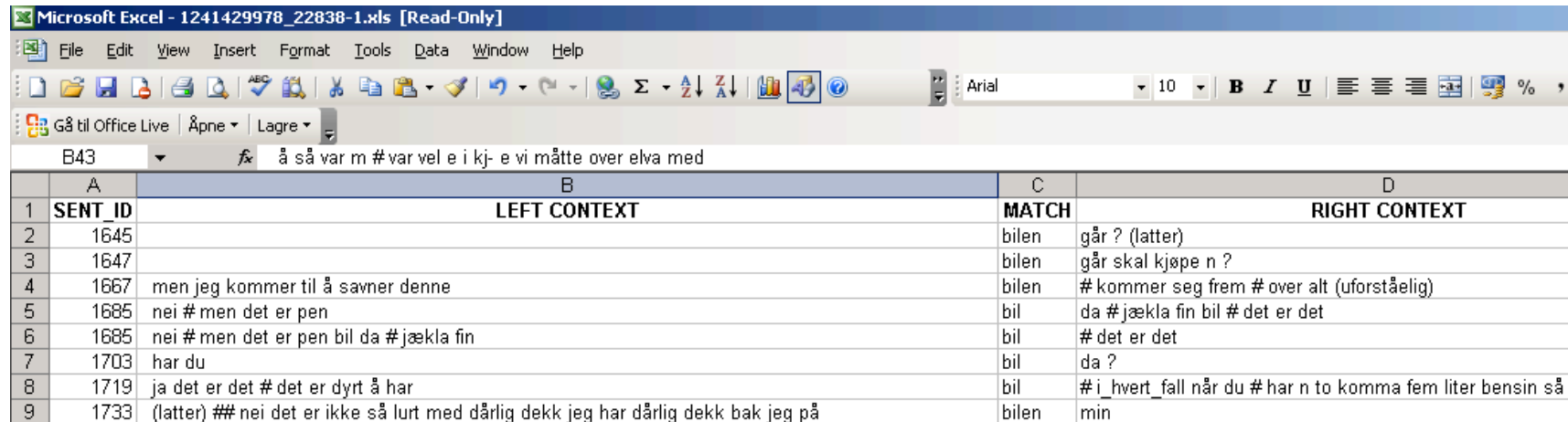
*

i **⌘** [_stonglandseide_47](#) når du kommer dit som du skulle tømme henne så har jo **hun hålt** lekke tom på veien og sånn blir det med skatten
veien

*

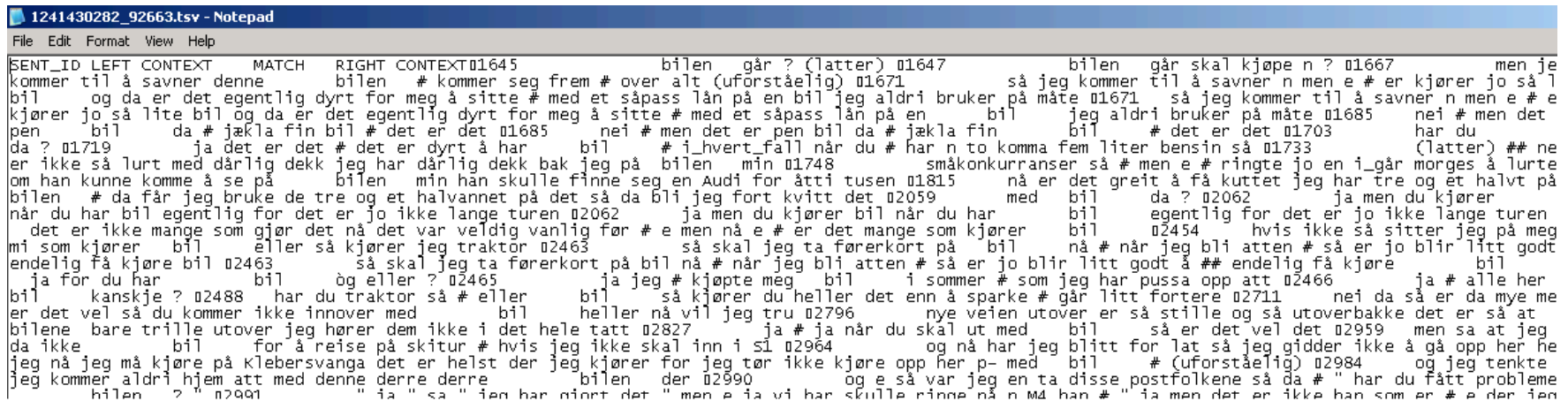
Downloading results, examples:

Excel:



	A	B	C	D
	SENT_ID	LEFT CONTEXT	MATCH	RIGHT CONTEXT
1	1645		bilen	går ? (latter)
2	1647		bilen	går skal kjøpe n ?
4	1667	men jeg kommer til å savner denne	bilen	# kommer seg frem # over alt (uforståelig)
5	1685	nei # men det er pen	bil	da # jækla fin bil # det er det
6	1685	nei # men det er pen bil da # jækla fin	bil	# det er det
7	1703	har du	bil	da ?
8	1719	ja det er det # det er dyrt å har	bil	# i hvert fall når du # har n to komma fem liter bensin så
9	1733	(latter) ## nei det er ikke så lurt med dårlig dekk jeg har dårlig dekk bak jeg på	bilen	min

Tab separated values:



```
1241430282_92663.tsv - Notepad
File Edit Format View Help
SENT_ID LEFT CONTEXT MATCH RIGHT CONTEXT01645 bilen går ? (latter) 01647 bilen går skal kjøpe n ? 01667 men je
kommer til å savner denne bilen # kommer seg frem # over alt (uforståelig) 01671 så jeg kommer til å savner n men e # er kjører jo så l
bil og da er det egentlig dyrt for meg å sitte # med et såpass lån på en bil jeg aldri bruker på måte 01671 så jeg kommer til å savner n men e # e
kjører jo så lite bil og da er det egentlig dyrt for meg å sitte # med et såpass lån på en bil jeg aldri bruker på måte 01685 nei # men det
pen bil da # jækla fin bil # det er det 01685 nei # men det er pen bil da # jækla fin bil # det er det 01703 har du
da ? 01719 ja det er det # det er dyrt å har bil # i hvert fall når du # har n to komma fem liter bensin så 01733 (latter) ## ne
er ikke så lurt med dårlig dekk jeg har dårlig dekk bak jeg på bilen min 01748 småkonkurranser så # men e # ringte jo en i går morges å lurte
om han kunne komme å se på bilen min han skulle finne seg en Audi for åtti tusen 01815 nå er det greit å få kuttet jeg har tre og et halvt på
bilen # da får jeg bruke de tre og et halvannet på det så da bli jeg fort kvitt det 02059 med bil da ? 02062 ja men du kjører
når du har bil egentlig for det er jo ikke lange turen 02062 ja men du kjører bil når du har bil egentlig for det er jo ikke lange turen
det er ikke mange som gjør det nå det var veldig vanlig før # e men nå e # er det mange som kjører bil 02454 hvis ikke så sitter jeg på meg
mi som kjører bil eller så kjører jeg traktor 02463 så skal jeg ta førerkort på bil nå # når jeg bli atten # så er jo blir litt godt
endelig få kjøre bil 02463 så skal jeg ta førerkort på bil nå # når jeg bli atten # så er jo blir litt godt å ## endelig få kjøre bil
ja for du har bil og eller ? 02465 ja jeg # kjøpte meg bil i sommer # som jeg har pussa opp att 02466 ja # alle her
bil kanskje ? 02488 har du traktor så # eller bil så kjører du heller det enn å sparke # går litt fortere 02711 nei da så er da mye me
er det vel så du kommer ikke innover med bil heller nå vil jeg tru 02796 nye veien utover er så stille og så utoverbakke det er så at
bilene bare trille utover jeg hører dem ikke i det hele tatt 02827 ja # ja når du skal ut med bil så er det vel det 02959 men sa at jeg
da ikke bil for å reise på skitur # hvis jeg ikke skal inn i sl 02964 og nå har jeg blitt for lat så jeg gidder ikke å gå opp her he
jeg nå jeg må kjøre på Klebersvanga det er helst der jeg kjører for jeg tør ikke kjøre opp her p- med bil # (uforståelig) 02984 og jeg tenkte
jeg kommer aldri hjem att med denne derre derre bilen der 02990 og e så var jeg en ta disse postfolkene så da # " har du fått probleme
bilen ? " 02991 " ja " så " jeg har gjort det " men e ja vi har skulle ringe nå n M4 han # " ja men det er ikke han som er # e der ien
```

Saving results

CWB expression: "`[[(lemma="bil" %c)]] ;`"

Action:

Hits four

Results

- count
- download
- sort
- collocations
- annotate
- show metadata
- metadata distribution
- delete hits
- save hits

? (latter)

ner til å savner denne **bilen** # kommer seg frem # over alt (uforståelig)

ner til å savner n men e # er kjører jo så lite **bil** og da er det egentlig dyrt for meg å sitte # med et såpass lån på en bil jeg aldri bruker på måte

ner til å savner n men e # er kjører jo så lite bil og da er det egentlig dyrt for meg å sitte # med et såpass lån på en **bil** jeg aldri bruker på måte

alvdal_02 ja det er det # det er dyrt å har **bil** # i_hvert_fall når du # har n to komma fem liter bensin så

alvdal_01 småkonkurranser så # men e # ringte jo en i_går morges å lurte på om han kunne komme å se på **bilen** min han skulle finne seg en Audi for åtti tusen

alvdal_01 ja men du kjører **bil** når du har bil egentlig for det er jo ikke lange turen

alvdal_01 ja men du kjører bil når du har **bil** egentlig for det er jo ikke lange turen

alvdal_01 det er ikke mange som gjør det nå det var veldig vanlig før # e men nå e # er det mange som kjører **bil**

alvdal_02 hvis ikke så sitter jeg på meg mor mi som kjører **bil** eller så kjører jeg traktor

alvdal_02 så skal jeg ta førerkort på **bil** nå # når jeg bli atten # så er jo blir litt godt å ## endelig få kjøre bil

alvdal_02 så skal jeg ta førerkort på bil nå # når jeg bli atten # så er jo blir litt godt å ## endelig få kjøre **bil**



- Nordic dialect corpus:

http://omilia.uio.no/glossa/html/index_dev.php?corpus=scandiasyn