



UNIVERSITETET I OSLO
DET HUMANISTISKE FAKULTET

The Nordic Dialect Corpus

Janne Bondi Johannessen

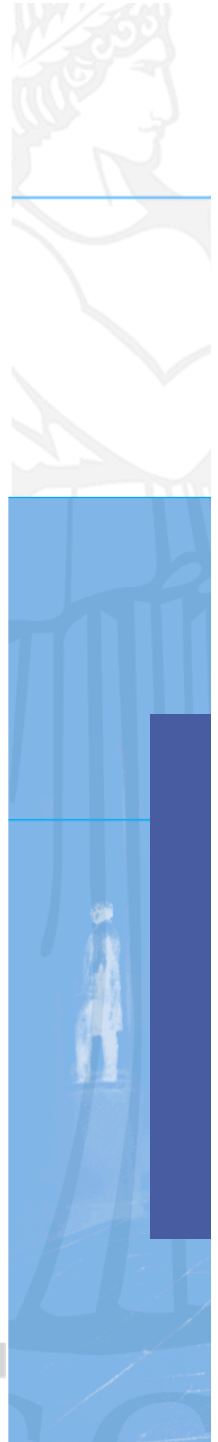
RILIVS, September 17th-18th 2009, University of Oslo



Credits

- Collaboration
 - The research network ScanDiaSyn and the Nordic Centre of Excellence in Microcomparative Syntax (NORMS)
- Financing of linguistic contents
 - National research councils and universities in the individual countries
- Financing of technical development
 - University of Oslo,
 - The Norwegian Research Council
 - 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

Two research tools

- Nordic Dialect Corpus
- Nordic Syntactic Judgment Database





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 – app. 916 841 words (by 10 September 2009)
- Still growing quickly
- 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



Individual texts from

- Danish
 - DanDiaSyn and NORMS
- Faroese
 - NORMS field work and Nordforsk
- Icelandic
 - Ásta Svavadottir, NORMS
- Norwegian
 - NorDiaSyn, NORMS and Målførearkivet
- Swedish
 - Swedia 2000, SweDiaSyn, NORMS field work and UiO

Contents by country and date



UNIVERSITETET I OSLO
DET HUMANISTISKE FAKULTET

Country	No of words May 2009	No of words September 2009
Denmark	19 088	123 187
Faroes	22 207	48 427
Finland	0	0
Iceland	10 287	10 287
Norway	165 176	424 443
Sweden	304 421	310 497
Sum	521 179	916 841





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 been transcribed phonetically
 - The Norwegian phonetic transcription
 - follows roughly that of Papazian and Helleland (2005).
 - The transcription of the Övdalian dialect follows the Övdalian orthography (standardised in 2005 by the *Råðdjärum* – The Övdalian 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.
- Icelandic
 - IceTagger available in the IceNLP toolkit (Hrafn Loftsson 2008)



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
- La phonologie du français contemporain (PFC)
 - Web-based dialect corpus with audio links
 - No grammatical annotation
- 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

æøå...»

criteria»

+

-



Regular expressions: ☐
Search within:

Hits per page:
Max results :

Randomize ☐
Skip tot. freq. ☒

Orthographic ☐
Phonetic ☐
Both ☒

Search corpus

Reset form

informant +

country +

region +

area +

place +

agegroup +

sex ☐
☐ F
☐ M

rec (year) +

genre +

Show texts

Save subcorpus

[Choose subcorpus](#)

Display:

Search within:

- [Transcription guidelines, translation lists, etc](#)
- [Recording locations](#)
- [Transcriptions](#)

Searching for lemmas

Scandinavian Dialect
Corpus

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



gås

criteria»

word »

occurrences »

pos »

num »

degr »

case »

sex »

nlex »

pers »

temp »

def »

descr »

inf type »

phonetic

lemma

start of word

within word

end of word

case sensitive

exclude word

add word form

add negated word form

add lemma

add negated lemma

per page: 20

Randomize ☐

Orthographic ☒

Phonetic ☐

results : 2000

Skip tot. freq. ☒

Search corpus

Reset form

Show texts

Save subcorpus

Choose subcorpus

country +

region +

area +

place +

agegroup +

sex +

rec (year) +

genre +

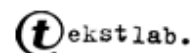
Display:

Search within:



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

Please report bugs and errors [here](#)



Searching for more than one word

Scandinavian Dialect
Corpus

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

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



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

Regular expressions: ☐

Search within: s

Hits per page: 20

Max results : 2000

Randomize ☐

Skip tot. freq. ☒

Orthographic ☒

Phonetic ☐

Search corpus

Reset form

Show texts

Save subcorpus

[Choose subcorpus](#)

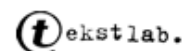
Display: ▼

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

Scandinavian Dialect
Corpus

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



Recording locations
Transcriptions

Re Se

criteria»

word »

occurrences »

pos »

num »

degr »

case »

sex »

nlex »

pers »

temp »

def »

descr »

inf type »

phonetic

country +

agegroup +

Display:

(spm)

adj

adv

cbl

det

inf-merke

interj

konj

pause

pause2

prep

pron

pron/det

sbu

subst

subst:adj

sann

ukjent

verb

verb:subst

ambiguous

Hits per page: 20

Max results : 2000

Randomize ☐

Skip tot. freq. ☒

Orthographic ☒

Phonetic ☐

Search corpus

Reset form

Show texts

Save subcorpus

Choose subcorpus

area +

place +

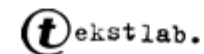
genre +

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](#))

æøå...»

um

criteria»

end of word
phonetic

+

-



Regular expressions: ☐
Search within:

Hits per page:
Max results :

Randomize ☐
Skip tot. freq. ☒

Orthographic ☐
Phonetic ☐
Both ☒

informant +

country -

Denmark
Faroe
Iceland
Norway

[>
[<

Sweden

choose ▾

region +

area +

place +

agegroup + sex + rec (year) + genre +

[Choose subcorpus](#)



Informants: 126
















scandiasyn:

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

Action :

: 449

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](#)

- i    **aasen_35** ... **tittade** vi på gammal_kortet och då fann jag åter # ett n ... brudkort från det bröllopet
... **kuogeðum** wi:ð å: gamtkuorteð og då: fann ig att # iet n ... brauðkuort frå dett djæstbuoð
- i    **aasen_48** ja # för att då vår femtiosex sedan så föddes ju pojken och **gifte** oss året dessförinnan ja # ja
ja # før at då: vår femtisjæks se: so fyøddes ju pojken og **djifteðum** uoss åreð firiað ja # ja
- i    **aasen_48** ojoj och hur fin (uforståelig) tänk att i går # så **pratade** M1 och jag just om ... # för då hade vi bröllopsdag i går
ojjoj og ur fin kommentar tenk at i går # so **prateðum** M1 og ig just um ... # før då: addum wi:ð brøllopsdag i går
- i    **aasen_48** ojoj och hur fin (uforståelig) tänk att i går # så pratade M1 och jag just **om** ... # för då hade vi bröllopsdag i går
ojjoj og ur fin kommentar tenk at i går # so prateðum M1 og ig just **um** ... # før då: addum wi:ð brøllopsdag i går
- i    **aasen_48** ojoj och hur fin (uforståelig) tänk att i går # så pratade M1 och jag just om ... # för då **hade** vi bröllopsdag i går
ojjoj og ur fin kommentar tenk at i går # so prateðum M1 og ig just um ... # før då: **addum** wi:ð brøllopsdag i går




















































Display of transcription and tagging

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

Action:

Hits found: 124

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

- 1    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**
- 1    alvdal_02 ja jeg # kjøpte meg **bil** i sommer # som jeg har pussa opp att
- 1    alvdal_02 har du traktor så # eller **bil** så kjører du heller det enn å sparke # går litt fortere
- 1    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
- 1    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
- 1    alvdal_04 ja # ja når du skal ut med **bil** så er det vel det
- 1    alvdal_03 men sa at jeg tar da ikke **bil** for å reise på skitur # hvis jeg il
- 1    alvdal_03 og nå har jeg blitt for lat så jeg gidder ikke å gå opp her heller **sex** kjøre på Klebersvanga det er helst der jeg kjører for jeg tør ikke kjøre opp
- 1    alvdal_03 og jeg tenkte at jeg kommer aldri hjem att med denne derre de **num** type: appell er
- 1    alvdal_03 og e så var jeg en ta disse postfolkene så da # " har du fått pro **defn**: be **bil**en ? "
- 1    alvdal_03 " ja " sa " jeg har gjort det " men e ja vi har skulle ringe på n M **descr**: me **nlex**: 100 en det er ikke han som er # e der jeg går med **bil**en " sa jeg # men det v
- 1    alvdal_03 ja jeg hadde ikke lovt han noe men han ville gjerne ha **bil**en
- 1    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
- 1    alvdal_04 men så byn- ja ja en gikk den men hun var aldeles sjuk på n vet du denne **bil**en så hun # kjøpte nå for en billig penge der
- 1    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
- 1    alvdal_03 hva_for en **bil** du har nå ?
- 1    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

Informant-based querying

Scandinavian Dialect
Corpus

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

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](#) [H](#) [K](#) alvdal_02 så skal je

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

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

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

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

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

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

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

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

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

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

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

[i](#) [H](#) [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](#) [H](#) [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](#) [H](#) [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](#) [H](#) [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

der jeg kjører fo

jeg går med **bile**

Display information on informants 2

Word count for selected informants: 916841 (total for Scandiasyn corpus: 916841)

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>Record</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	Norrbottn	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

Action menu

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

Action :

Hits four

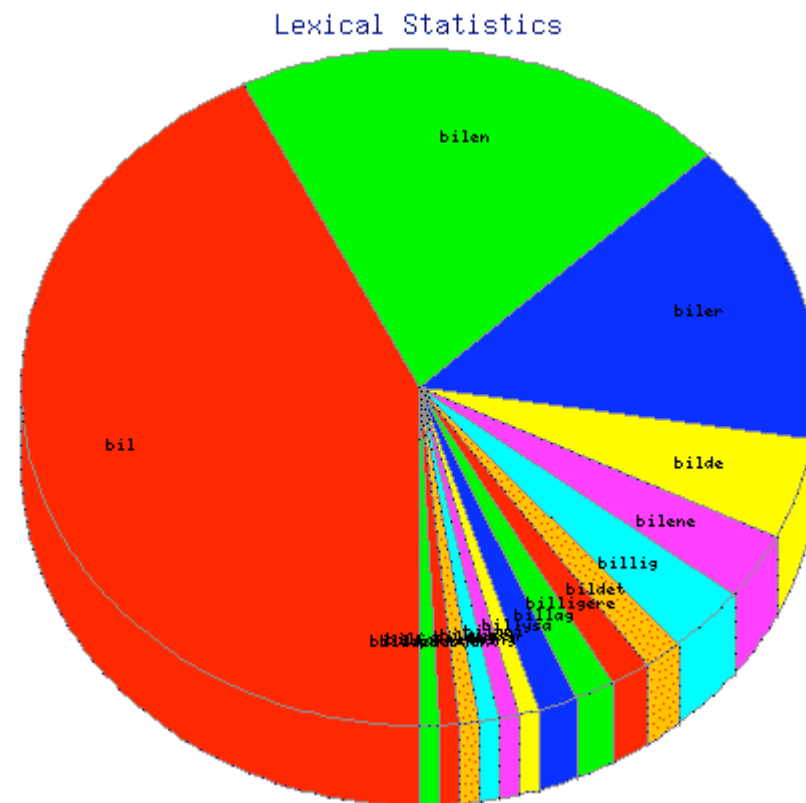
Results f

- 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**
- pte 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	bilene
18	biler
6	bilde
4	billig
4	billigere
2	billag
2	bilysa
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 **me** **fjordvaer_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 jo

*

i **me** **fjordvaer_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 **me** **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 **me** **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:

Microsoft Excel - 1241429978_22838-1.xls [Read-Only]

File Edit View Insert Format Tools Data Window Help

ABC

Gå til Office Live | Åpne | Lagre

B43 å så var m # var vel e i kj- e vi måtte over elva med

	A	B	C	D
1	SENT_ID	LEFT CONTEXT	MATCH	RIGHT CONTEXT
2	1645		bilen	går ? (latter)
3	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 CONTEXT	01645	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 # er kjører jo så l								
	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 er pen bil da # jækla fin	bil	# det er det	01703	har du				
	pen bil da # jækla fin bil # det er det	01685	nei # men det er pen bil da # jækla fin	bil	# det er det						
	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				
	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 å ## endelig få kjøre bil								
	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 bilen der	02990	og e så var jeg en ta disse postfolkene så da # har du fått probleme								
	bilen ?	02991	ia " sa " jeg har gjort det " men e ia vi har skulle ringe nå n M4 han # " ia men det er ikke han som er # e der ien								

Saving results

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

Action :

Hits four

Results p

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

? (latter)

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

her 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

her 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://www.tekstlab.uio.no/nota/scandiasyn/>



References

- Loftsson, H. 2008. Tagging Icelandic text: A linguistic rule-based approach. *Nordic Journal of Linguistics* 31:47-72.
- La phonologie du français contemporain (PFC) .
http://www.projet-pfc.net/index.php?option=com_wrapper&view=wrapper&Itemid=184