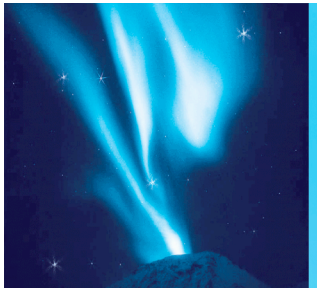


ScanDiaSyn

NORDISK DIALEKTSYNTAKS

Scandinavian Dialect Syntax

Mállýskutilbrigði í norrænni setningagerð

Generating a lexicon of Scandinavian modal verbs from a parallel corpus

Gunnar Hrafn Hrafnbjargarson

ScanLex

Lexicon with ‘one-to-one’ correspondences

hana – hun, ho, henne – hende – hana – ho, henne – henne – her

burde – burde – eiga, skula – eiga – skola – ought, shall

Needed for the multi-language lemmatization of the ScanDiaSyn database

For automatic generation of the lexicon, large parallel corpora are needed

- Such corpora do not exist (at least not for Icelandic and Faroese).
- We need to develop methods for extracting data from large corpora.

Relying on morphology

Until now, pronouns have been fed manually into the database according to their case, number, person, gender (and sometimes meaning).

This is possible because of the morphological/phonological similarities in the pronominal systems of Scandinavian. *Hun* corresponds to *hún*, and *henni* corresponds to *hende.*

Relying on morphology

Although the modal verbs look alike and/or sound alike, they do not always correspond to each other:

- Norwegian *skulle* doesn't always have the same meaning as Danish *skulle*.
- Danish *ville* doesn't always correspond to Icelandic *vilja.*

Since we cannot rely on morphology, we have to rely on something else, e.g. parallel corpora.

Scandinavian parallel corpora

Not many parallel corpora cover all of the Scandinavian languages.

- One is the Sophie Treebank (hf.uio.no/tekstlab/prosjekter/SOFIE.htm).

- The *KDE part* of the **OPUS corpus** (logos.uio.no/opus/) contains KDE system messages from a.o. Bokmål, Danish, Icelandic, Nynorsk and Swedish, but no Faroese.

The Sophie Treebank

The Sophie Treebank is a parallel treebank that consists of material from nine 'North' European languages.

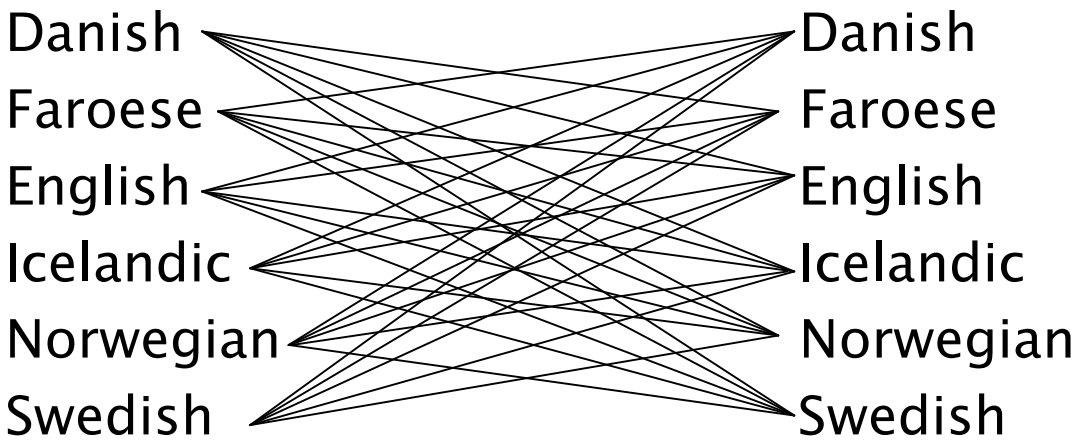
Sentences with modal verbs/Total in corpus:

Danish	126/547
Faroese	85/556
Icelandic	101/542
Norwegian	124/543
Swedish	48/215
English	84/530

171 aligned blocks (with one to six members).

The method (first try)

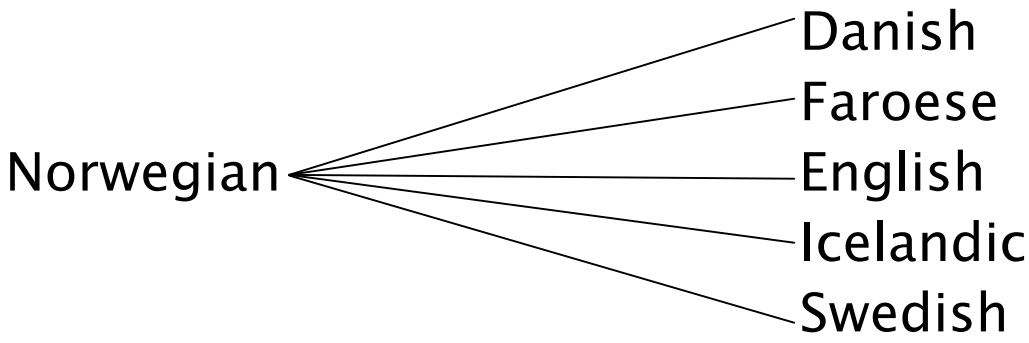
First, I wanted to find all possible matches:



After three weeks this had become a mess …

The method (second try)

Then, I analyzed the Norwegian modal verbs (using the definitions of the Norwegian Reference Grammar), and found the correspondences in the other languages:



Suddenly, everything worked much better … still, the database still needs heavy proof-reading.

Which modals occur in which meaning?

MÅTTE

Epistemic (probability, necessity)
Deontic (permission, order)

KUNNE

Epistemic (possibility, dynamic)
Deontic (offer, order/request)

VILLE

Epistemic (future, prediction)
Deontic (volition, intention)

SKULLE

Epistemic (roomer, necessity)
Deontic

BURDE

Deontic (obligation, advise)

FÅ

Deontic (dynamic, order)

MÅTTE

Altså ***måtte*** verdensrommet en eller annen gang ha blitt til av noe annet. (48,166N)

Himingeimurinn ***hlaut*** því einhvern tímann að hafa orðið til úr einhverju öðru. (48,166I)

Tí mátti alheimurin onkuntíð vera blivin til úr onkrum øðrum. (48,171F)

So space must sometime have been created out of something else. (48,156E)

Altså måtte verdensrummet en eller anden gang være blevet til af noget andet. (48,167D)

Värdsrymden måste altså en gång i tiden ha skapats ur någonting annat. (48,157S)

MÅTTE

Var det ikke trist at folk flest **måtte** bli syke før de innså hvor fint det er å leve? (35,121N)

Var það ekki sorglegt að flestir **þurftu** að verða veikir til að átta sig á hvað það er gott að lifa? (35,121I)

Var tað ikki syrgiligt, at tey flestu **máttu** gerast sjúk, áðrenn tey fataðu, hvussu gott tað er at liva? (35,125F)

Nog var det väl synd att folk **måste** bli sjuka för att begripa hur skönt det var att leva ? (35,115S)

MÅTTE

-- Neimen Sofie da, du ***må*** ikke snakke sånn. (163,516N)

Þetta ***máttu*** ekki segja. (163,516I)

- Nei, Suffía, soleiðis ***mást*** tú ikki snakka! (163,530F)

Nej Sofie sådan ***må*** du da ikke snakke. (163,521D)

MÅTTE

Da ***må*** du eventuelt legge en beskjed til meg I postkassen. (155,476N)

Þá ***verðurðu*** að setja skilaboð til mín í póstkassann. (155,475I)

Gevst tú, ***mást*** tú leggja mær eini boð í postkassan. (155,490F)

In that case you ***must*** leave a message for me in the mailbox.
(155,465E)

I så fald ***må*** du lægge en besked til mig I postkassen. (155,479D)

KUNNE

Kunne det være fra pappa? (56,195N)

Gat þetta verið frá pabba? (56,195I)

Kundi tað vera frá pápanum? (56,202F)

Could it be from Dad? (56,187E)

Kunne det være fra far? (56,196D)

Kunde det vara från pappa? (56,187S)

KUNNE

Altså: Hvis et spedbarn hadde ***kunnet*** snakke, ville det sikkert sagt noe om hvilken forunderlig verden det hadde kommet til. (122,389N)

Sem sagt: Ef ungabarn ***kynni*** að tala myndi það örugglega segja eitthvað um þennan undarlega heim sem það hefur hafnað í. (122,389I)

If a newborn baby ***could*** talk, it would probably say something about what an extraordinary world it had come into. (122,377E)

Altså: ÷ Hvis et spædbarn havde ***kunnet*** tale, ville det sikkert sige noget om hvilken forunderlig verden det er kommet til. (122,391D)

KUNNE

Du ***kan*** godt si at en filosof forblir like tynnhudet som et lite barn hele livet. (148,464N)

Segja ***má*** að heimspekingurinn sé alla ævi álíka næmur [...] og lítið barn. (148,463I)

You ***might*** say that [...] a philosopher remains as thin-skinned as a child. (148,453E)

Du ***kan*** godt sige at en filosof [...] forbliver lige så tyndhudet som et lille barn. (148,467D)

KUNNE

Hvis jeg er opptatt av hester […], ***kan*** jeg ikke forlange at alle andre skal være like opptatt av det samme. (84,285)

Hafi ég áhuga á hestum […] ***get*** ég ekki krafist þess að allir aðrir séu sama sinnis.(84,285I)

Um eg eri hugtikin av rossum […}, kann eg ikki krevja, at øll onnur skulu vera líka hugtikin av teimum. (84,294F)

If I happen to be interested in horses […], I ***can***not expect everyone else to share my enthusiasm. (84,275E)

Hvis jeg er optaget av heste […], ***kan*** jeg ikke forlange at alle andre skal være lige så optaget af det. (84,286D)

VILLE

Hun kunne være helt sikker på at ingen ***ville*** finne henne her. (42,150N)

Þar var hún örugg um að enginn ***myndi*** finna hana. (42,150I)

Hun kunne være helt sikker på, at ingen ***ville*** finde hende der. (42,155D)

Ville hun vært en annen da? (14,57N)

Would she then have been someone else? (14,53E)

Skulle hon då ha varit en annan? (14,55S)

Kanskje ***vil*** hun trenge legebehandling … (143,437N)

She ***may*** even neeed medical attention … (143,426E)

VILLE

Det er nettopp hvordan han har klart det, vi gjerne ***vil*** avsløre. (108,339N)
Það sem við ***viljum*** komast að er einmitt hvernig honum tókst það. (108,340I)
Men vit ***vilja*** fegin avdúka, hvussu hann ber seg at. (108,351F)
What we ***would*** like to know is just how he did it. (108,327E)
Men vi ***vil*** gerne afsløre, hvordan han kan gøre det. (108,340D)

Som du skjønner, ***vil*** jeg gi deg en gave som du kan vokse på. (57,204N)
Eins og þú veist ***ætla*** ég að gefa þér gjöf sem getur orðið þér til heilla.(57,204I)
Sum tú skilir, ***vil*** eg geva tær eina gávu, tú kannst búnast av. (57,211F)
As I'm sure you'll understand, I ***want*** to give you a present that will help you grow. (57,194E)
Som du forstår, ***vil*** jeg give dig en gave du kan vokse på. (57,205D)
Du vet ju att jag ***vill*** ge deg en present som du kan växa med. (57,196S)

SKULLE

[...] kom hun også til å tenke på at hun ikke ***skulle*** være her bestandig. (27,101N)

[...] fór hún líka að hugsa um það að hún ***myndi*** ekki vera hér alla tíð. (27,101I)

[…] kom hon eisini í tankar um, at hon ikki ***skuldi*** vera her altíð. (27,105F)

[…] kom hun også i tanke om, at hun ikke ***skulle*** være her altid. (27,101D)

[...] så kom hon att tänka på at hon inte alltid ***skulle*** finnas till. (27,96S)

Hvorfor ***skulle*** det være så vanskelig å være opptatt av [...] (75,264N)

Af hverju ***þurfti*** að vera svona erfitt að hugsa um [...] (75,264I)

Hví ***skuldi*** tað vera so trupult at vera upptikin av [...] (75,272F)

Hvorfor ***skulle*** det være så svært at være optaget af [...] (75,266D)

SKULLE

[...] når det helt åpenbart ***skulle*** et ganske annet sted? (58,216N)
[...] þegar það ***átti*** augljóslega að fara eitthvert annað? (58,216I)
[...] tá lð hann visti, at tað ***skuldi*** til eitt heilt annað stað? (58,223F)
[...] når det helt åbenbart ***skulle*** et andet sted hen? (58,217D)
[...] när det uppenbarligen ***skulle*** någon annanstans (58,206S)

For sikkerhet skyld ***skal*** vi derfor gjøre et par tankeeksperimenter [...] (129,403N)
Til öryggis ***skulum*** við gera nokkrar tilraunir í huganum [...] (129,402I)
Fyri at eingin ivi skal vera, ***skulu*** vit gera nokkrar tilraunir í huganum [...] (129,416F)
For en sikkerheds skyld ***skal*** vi derfor gøre et par tankeeksperimenter [...] (129,404D)

BURDE

Og fremfor alt: Hvordan ***bør*** vi leve? (93,310N)
Og ekki síst: Hvernig ***eigum*** við að lifa? (93,310I)
Hvussu ***eiga*** vit at liva? (93,320F)
And most important, how ***ought*** we to live? (93,299E)
Og fremfor alt: ÷ Hvordan ***bør*** vi leve? (93,310D)

Kanskje ***burde*** hun se etter om det lå noe mer der? (37,123N)
Kannski ***ætti*** hún að gá að því hvort þar væri eitthvað meira að finna? (37,123I)
Kanska ***skuldi*** hon farið og hugt eftir, um har lá okkurt afturat? (37,127F)
Perhaps she ***should*** go and see if any more letters had arrived. (37,114E)
Måske ***skulle*** hun se efter, om der lå noget mere? (37,124D)
Hon ***skulle*** kanske se efter om det fanns någonting mer I brevlådan? (37,117S)

FÅ

Hun ***fikk*** ofte høre at hun hadde vakre mandeløyne […] (16,74N)

Hún hafði oft ***fengið*** að heyra að hún hefði falleg möndlulaga augu […] (16,74I)

Hon ***fick*** ofta höra att hennes ögon var vackert mandelformade […] (16,70S)

-- Nei, sånn ***får*** du ikke lov å snakke til meg, Sofie (165,529N)

- Nej sådan ***må*** du ikke snakke til mig, Sofie. (165,533D)

http://omilia.uio.no/scanlex