Studying past and present Heritage Norwegian: Possibilities and pitfalls

Janne Bondi Johannessen

Studying an old heritage language

• Requires a wide range of knowledge on many different levels, not just in theoretical linguistics
  – Statistics
  – Dialectology (of baseline language as well as majority language)
  – History of language
  – Ethnography
  – Lexicography

… and this will be shown in this talk
Using corpora as evidence of people we have not met

• The talk:
  • Some good examples of corpus use
  • Some pitfalls of corpus use
  • Some knowledge of Scandinavian Heritage languages
Different types of heritage language

- A) the heritage language has a mother language elsewhere in the world, and it usually ends up being different from this baseline language along a number of dimensions, from the lexicon to some features of the grammar. Typical examples of heritage languages are Norwegian, Russian or Spanish in America.
- B) indigenous languages that live alongside a majority (often colonial) language (for example Cree, Navajo, Mohawk).
- Here: focus on the first type
Heritage languages: a very diverse phenomenon

• “[h]eritage speakers may become interested in further developing their skills in the heritage language for potential career advantages”…
  (Montrul 2016:3)

• (about adolescents and young adults)
Norwegian Immigration to the USA and Canada

- The first Norwegians arrived in New York in 1825.
- By 1930, 810,000 had arrived in the US and 40,000 in Canada. No country except Ireland had a higher rate of emigration.
- Einar Haugen (1953) writes that the 1800s was a century of huge population growth in Norway, and the number of emigrants from Norway equaled the whole 1800 population.
- Norwegians quickly built institutions that were important to them. They organized and built churches, hospitals, old peoples’ homes, and established Luther College (Decorah, Iowa) as early as in 1861, and St. Olaf College (Northfield, Minnesota) in 1875. There were Norwegian-language schools, and newspapers.

Swedish: Almost the same facts, and almost the same language
«Utvandrere» [emigrants], by Gustav Wentzel, 1903
Halvor Boe, with wife and four children by their first home in Edmore, North Dakota
How do we know about the language input for the present day heritage speakers?

• There are no recordings of their speech as children. When we study their language and we see that it differs from the baseline (European Norwegian) we do not know whether it is the language that has changed in their lifetime in these individuals or whether the language had already changed by the time they grew up. It therefore is a challenge for the researchers who want to study their language and understand what is going on.

• Use other people’s speech to determine something about what the general situation must have been like.
Einar Haugen, 1906–1994

Professor at University of Wisconsin Madison, then Harvard University.

Did fieldwork in the Midwest in the 1930s and 40s. Interviews with questionnaires and recorders, “they might last anywhere from ten to twelve hours” (Haugen 1953:326)

The recordings (54 hours) are available at the Text Laboratory, UiO, some with transcriptions

*The Norwegian Language in America: A Study in Bilingual Behavior* (1953)
Bertha Eds. Kondovi.

Har butt i byen siden ho var 15 år. Det ho seier på norsk, høyest ekte nok ut, men det er klart at ho er ute av trening meddi ho slår over i engelsk kvar augneblink.

Smare under enn over middels intelligent. Svært dårlig informant. Ei mengd spursmål svarar ho berre på med "I don't know", noko som førde til at eg like godt hoppa over fleire av dei spursmåla som eg av røymele veit det er vanskeleg å få svar på av dårlige informantar.
**Biographical Information**

1. Name: Bertha Ede
2. Address: Mønsted
3. Place of birth: Canton, Teppe
4. Year of birth: 1874
5. Father's name: Johannes Eriksen
6. Mother's name: Ingeborg Teigen
7. Date, place of father's birth: ca. 1832, Møreland (2) Norge
8. Date, place of mother's birth: ca. 1855
9. Date of immigration to America: which members of family came to America and when: For us, 1726-28, they came, visit upon visit with us. Did not feel at home. Gift... Goodwin Co., Ill.
10. Place(s) of origin in Norway:

   [List of places]

11. Name of wife or husband, age, place of birth: John Ede, 15, Engen, Siden, F. Mønsted, Teppe
12. When was or is Norwegian spoken: Fell 15, yes, most every time. In 1905, yes. Not always. Used Norwegian, and sometimes English. Has, both, his children,
   born, here.
13. What type of Norwegian (dialect) heard from parents, schoolmates, neighbors, and yourself: ...
14. Which dialect do you speak: ...
(Mitt namn er) Jakåp T. Seljestå, føtt den niande februar, 1866, i Odda, Hardanger. (My name is) Jacob T. Seljestad, born February 9, 1866, in Odda, Hardanger.

Kå va de du ville ha ßekst? (Litt om livet der heime?) Ja, de me kadla' heima var what did you want at home? (A little about your life at home?) Well, what we called den. Ðe laeg oppi ogst bygdene som fanst i Hardanger. Sëljestå, de låg oppi ei bygd-- siste home was in the highest valley of Hardanger. Seljestad lay up in a valley-- the last garen oppi Hardanger so djekk øve te Røldal, fårstår du, å der var vinteren omtrete farm in Hardanger where one crossed over to Røldal, you see, and there the winter syv månner eller seks å en hall månne i åre, så kalt var de. Å der va eg oppfødde me lasted for seven months, or six and a half, ever year, that's how cold it was: Where snøskavlen så var opp te en tyve fot høg om vintren, skavlar so bles i sammen, får-I was raised, the snow drifts lay up to twenty feet high in the winter, drifts that står du. Å når me skulle ut i skogcn, så måtte me gå på ski, me kunne ikkje gå får blow together, you see. And when we were going out in the woods, we had to use skis. de var så mye snø atte me måtte gå på ski, så når me sku te skulen, så måtte me ha couldn't walk, for it was than

(http://tekstlab.uio.no/norskiamerika/opptak/haugen.html
Index of /media/lydobilde/amerikanorsk/einar_haugen/DVD1/CD10
And Index of /media/lydobilde/amerikanorsk/einar_haugen/Transkripsjoner)
Norwegians in America

Almost two hundred years ago, the first Norwegians went to America to start a new life. In the one hundred years from 1825 to 1925 800 000 Norwegians emigrated. At present, most Norwegian Americans only speak American, but there still are some who learned Norwegian at home and who have continued to speak Norwegian as a heritage language as adults. These Americans are most often well up in their eighties and nineties.

In the project on Norwegian in America, we have gone to America and Canada a number of times to interview the heritage language speakers. The recordings are integrated in a new speech corpus of American Norwegian.

Norwegian in America

The American recordings and transcriptions are freely available for research. The material is valuable for interesting studies, not only on the particular American Norwegian language variety, but also on how immigrant languages live and die, and how Norwegian Americans have lived their lives in a multilingual reality.

The project on Norwegian in America is part of the national LIA Project, (Language Infrastructure made Accessible) and is one of the research areas at Multiling (Centre for Multilingualism) at the University of Oslo.

Call for papers: Seventh Annual Workshop on Immigrant Languages in the Americas

Seventh Annual Workshop on Immigrant Languages in the Americas will be arranged at University of Georgia, Athens October 27–29, 2016.

Read more
Case 1: Word order in Norwegian subordinate clauses


• Subordinate clause word order in Heritage Norwegian and Swedish differs from the one in European Scandinavian in that there seems to be free variation between

a. SBU–SUBJECT– ADVERB– VERB (H Norw, E Norw)  \( \text{No verb-movement} \)
b. SBU–SUBJECT– VERB– ADVERB (H Norw)  \( \text{V-to-T movement} \)

• In European Scandinavian = baseline language, only the former is used and grammatical.

(We exclude the subjunction at, 'that', which allows the order in b as well, which is understood by most linguists today as having a semantic explanation, and is therefore different.)
Subordinate order, Present Heritage Norwegian: V movement

"om du finner ikke ut så"
"if you find not out then"
‘If you don’t find out, then’
(Norwegian, Colleen, sunburg_MN_04gk)

"Om ja sätter inte en sten på farfars grav"
"if I put not a stone on granddad’s grave"
‘så kommer det aldrig en sten där"
"so comes there never a stone there"
‘If I don’t put a stone on granddad’s grave there will never be a stone there.
(Swedish, Konrad, mn11_m013)

(From Larsson & Johannessen 2015b)
Subordinate order, Present Heritage Norwegian: ... but also no V movement

"jeg kan gjøre det hvis om det ikke regner"
I can do it if whether it not rains
‘I can do it if it does not rain.’
(Norwegian, Tip, coon_valley_WI_06gm)

"det er mange som ikke har slutta og"
there are many who not have stopped too
‘There are also many who have not stopped.’
(Norwegian, Elnor, coon_valley_WI_02gm)

(From Larsson & Johannessen 2015b)
Subordinate order, Past Heritage Norwegian: V movement

"Då di kamm ti detti landi då settla dæ på en when they came to this country the then settled they on a homstedde som e no Taon åv Farmington homestead that is now town of Farmington ‘When they came to this country, they settled in a homestead that is now the town of Farmington.’
(Norwegian, Winfield Krostu, Waupacs co., born in Wis 1884, rec.1942)

"Då me ha no blitt jipte, […] så kjæm hornaran when we have now become married so came musicians the ‘When we had been married came the musicians’
(Norwegian, Winfield Krostu, Waupacs co., born in Wis 1884, rec.1942)"
So...

• The present day recordings show that Heritage Scandinavian speakers have two kinds of word order for embedded clauses.
  – (Also: Taranrød 2011 and L & J 2015b show that they are in free variation; the same speaker can use both.)

• The past recordings, from 1930s–40s by Einar Haugen show that even previous Heritage Norwegians could use the non-European word order.

• But is it certain that this is a new development? Could it perhaps be that this same variation exists in European Scandinavian?
• So far:
  – Some modern recordings
  – Recordings and transcriptions from Einar Haugen’s work

• How about facts about European Norwegian? Can we be sure that it is as we think? What if there is variation in Scandinavia, too?

• Answer: Use a corpus of dialects! The Nordic Dialect Corpus
What is a speech corpus?

• A collection of recorded spoken language
• Transcriptions of the recordings
• The transcriptions and the audio and video are linked to each other
• It is searchable by words (or subword units) and grammatical categories
• It delivers concordances
• It can provide maps
• It can provide spectrograms
ja # ja det var ø det var første tingen de e spurt det. # kom hjem at (uninterpretable)
når du # kamm heim att ... ær du æ re

ja # ja de var ee de va fysste tingen domm ee sporde det. nå ru # ær du æ re

yes ... yes it was e it was the first thing e asked it when you ... came home at (uninterpretable) ... are you a Catholic? ... _ ... En

bensin å ja # og # jeg skal kjøpe bi- e bensin. før jeg reiser hjem # i_kveld for det er
e bensin å ja # å je ska tjøpe bi- ee bensin. førre je reiser hemm # i_kveld før d e r

gasoline oh yes ... and ... I'll buy bi- e petrol before returning home ... i_kveld it is much cheaper here than it is Oppi Eau_Claire translated by

hele tida # hver gang jeg kommer til Blair så kjøper jeg bensin her. før jeg reiser hjem (laughter)

hele tia # her gonng je kammør te BleR så tjøpe je bensin hær. før jeg reiser hemm _ translated by Google as Norwegian

all the time ... every time I come to Blair so I buy gasoline here ... before I go home (laughter) translated by Google as Norwegian

but tha- det er mange år but tha- det er mange år. siden det og

bø dhæ- de æ mange år bø dhæ- de æ mange år. sid' n de å

but tha- it is many years but tha- it is many years since it and translated by Google as Norwegian

får du vondt i øra får du vondt i øra. når du # kommer ned ?

får du vondt i øra får du vondt i øra. nær du # kammør ne ?

you get pain in my ears get sore ears when you ... come down? translated by Google as Norwegian

gå for noen dager ja # kan ikke bare gå for noen dager. når du go- reiser til Norge du må reise-
gå først noen dager ja # kan itte bære gå først noen dager. nårr du gå- reiser te Nærg du må reise å

før de kom hit men de var ikke kjent før de kom hit. som jeg veit om ## så

translate 28.09.16
ja # ja det var e det var første tingen de e spurte det når du # kom hjem att (uninterpretable) ”er du er du katoli”

CWB expression: ”([[(pos="subjunc")]])([[pos="pron")]])”

Action: Map

Hits found: 1249

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 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48

ja # ja det var e det var første tingen de e spurte det nær du # kom hjem att (uninterpretable) ”er du er du katoli”

ja # ja de var ee de va fysste tingen domm ee sporde de nå ru # kàmm heim att ___

yes ... yes it was e it was the first thing the e asked it when you ... came home att (uninterpretable) ___ are you are you a Catholic? ___

ee bensin å ja # og # jeg skal kjøpe bi- e bensin før jeg reiser hjem # i_kvel

ee bensin å ja # å # je ska tjøpe bi- ee bensin førje reiser hemm # i_kvel

e gasoline oh yes ... and ... I’ll buy bi- e petrol before returning home ... i_kvel’d it is much cheaper here than it is Oppf’Eau_Clai
blair WI_07gm nei da # (uninterpretable) katolikk
blair WI_04gk (laughter) * tør
blair WI_04gk ikke gå ut med en katolikk å
blair WI_07gm * no no no kunne ikke gå med
blair WI_07gm katolikkene # jeg veit ikke hva det er # forskjell det var men
blair WI_04gk * ja
blair WI_04gk veit du
blair WI_07gm ja # ja det var e det var første tingen de e spurt det når du #
blair WI_04gk * nei
blair WI_07gm visste ikke # noe forskjell vi det var det merke ingen forskjel
blair WI_04gk * nei * ja
blair WI_04gk du e du e ble gift ?
blair WI_07gm ha ?
blair WI_04gk du var gift ?
blair WI_07gm * ja jeg
blair WI_07gm gift e e kona hun e døde to år nå # to år (uninterpretable)
blair WI_04gk * ja

CWB expression: "(((pos="subjunc")))(((pos="pron")))"
Action: 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53
Results found: 1249

blair WI_07gm ja # ja det var e det var første tingen de e spurt det når du # kom hjem att (uninterpretable)"
ja # ja de var ee de va fysste tingen domm ee sporde de nå ru # kjimm heim att er du ære kjom

Yes ... yes it was e it was the first thing the e asked it when you ... came home att (uninterpretable) are you are you a Catholic? ... En translated

blair WI_04gk e bensin å ja # og jeg skal kjøpe bi - e bensin før jeg reiser hjem # i kveld for det er my
ee bensin å ja # å # je ska tjope bi- ee bensin førre je reiser hemm # i kveld før de my

e gasoline oh yes ... and ... I'll buy bi - e petrol before returning home ... i_kveld it is much cheaper here than it is Oppi Eau_Claire translated by C
Nordic Dialect Corpus

• Contains about 2,8 million words from conversations and interviews by dialect speakers from five North Germanic languages.
• It is transcribed and linked to audio and video
• Has a map function, and can be searched in a large variety of ways.
• Can be filtered by age, gender and importantly, place (options go from small place via county to region of country and the whole country)
Taranrød 2011: Findings in the Nordic Dialect Corpus

• Heritage Norwegian
  – Out of twelve relative clauses containing an adverb, six. i.e. half, had verb-adverb order. She also looked at her findings in relation to the overall size of the text material. Her six non-standard word order relative clauses were found in a material of altogether 60 000 words. (Taranrød 2011:53)

• European Norwegian
  – Only three subordinate clauses had verb movement in a material of a total of 1.5 million words (Taranrød 2011:64).

6:60 000 vs. 3:1 500 000

- CANS: 49 000 words
- NDC: a subgroup that amounts to the same number of total words 49 000 words: old informants from Oppland county

- Both corpora contain long stretches conversations between two informants. Searching the corpus will usually give more than one hit per informant.
CANS – Heritage Norwegian

<table>
<thead>
<tr>
<th>Complementizer</th>
<th>Adverb</th>
<th>Verb-Adverb</th>
<th>Adverb-Verb (word-order in Norway)</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>om 'whether'</td>
<td>ikke 'not'</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>som</td>
<td>ikke 'not'</td>
<td>6</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>'which/who/that'</td>
<td>ikke 'not'</td>
<td>15</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>at 'that'</td>
<td>ikke 'not'</td>
<td>22</td>
<td>10</td>
<td>32</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
NDC - European Norwegian

<table>
<thead>
<tr>
<th>Complementiser</th>
<th>Adverb</th>
<th>Verb-Adverb</th>
<th>Adverb-Verb</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>om ‘whether’</td>
<td>ikke ‘not’</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>som ‘which/who/that’</td>
<td>ikke ‘not’</td>
<td>1</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>at ‘that’</td>
<td>ikke ‘not’</td>
<td>13</td>
<td>19</td>
<td>32</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>14</td>
<td>40</td>
<td>54</td>
</tr>
</tbody>
</table>
What is the next step?

- The numbers are small...
- Using statistic significance tests is tempting...

... but
"Et viktig premiss for svært mange hypotesetester - de fleste som er i vanlig bruk blant lingvister - er at observasjonene er uavhengige. Dette premisset er ofte brutt i lingvistisk forskning, særlig ved at samme individ er observert flere ganger, f.eks. i et korpus, og hver observasjon er benyttet som en observasjon i testen."

[An important premise for numerous hypothesis tests (and most of those used by linguists) is that the observations must be independent. This premise is often broken in linguistic research, especially in that the same individual is observed several times, e.g. In a corpus, and where every observations is used as one observation in the test.]
• "Premisset gjelder vanlige tester vi ofte bruker, som kjikvadrat-test, Fisher-test, t-test, Wilcoxons rangsum-test (også kalt Mann-Whitney-test), anova, Pearsons korrelasjonstest, Spearmans korrelasjonstest, lineær regresjon, logistisk regresjon."
• “as noted in Saito (1999) and Hatch and Lazaraton (1991), chi-square is a much-abused test in second language research studies, and often one of its assumptions (that of independence of data) is violated as a matter of course.” (Larson-Hall, 2010, s. 206)

• “It is not uncommon to find cases in which the assumption of independence of observations [of the chi-square test] is violated, usually by having the same participant respond more than once.” (Howell, 2010, s. 152)
• “many statistical techniques assume that the individual data points are independent of each other, which is usually not the case in corpus data where one speaker/writer may provide many concordance examples;” (Gries, 2015, s. 14)

• So: If you have few items or unclear data and wish they were clearer, the simple significance tests won’t help you.
Lesson: Be careful when using statistics!

• L & J (2015b):
• “We have calculated whether the differences between the tables for each subjunction type is significant. Using Fisher’s Exact Test, we find that the difference between Heritage and European Norwegian for *that* clauses and *som* clauses is highly significant, with p-value = 0.0004649 and p-value = 0.02473.”
• The corpus speakers: 11 ... No. of examples: 32 ...
• The significance calculation is not valid. Each occurrence is not an independent finding, since there is more than one occurrence per speaker.
• So the result is valid, but the statistical measurement is not...
So, even with no direct knowledge of the parents or grandparents of our heritage speakers...

• We can find evidence for what is indeed the baseline, by using:
  – Old recordings and transcriptions
  – A modern dialect corpus

• Finding that Heritage Scandinavian is indeed different from baseline European Norwegian

• But be careful with statistics!
Case 2: Grammatical gender


• Issue: Is there traditional gender agreement in Norwegian Heritage Language? If there is, then the indefinite article should have three forms:
  – En (M. SG.INDEF)
  – Ei (F. SG.INDEF)
  – Et (N. SG.INDEF)
• So whenever other forms are found, it is tempting to conclude that traditional gender agreement is missing.

• If, for example, we find something other than *et* with neuter nouns we could quickly conclude that our informant does not have the traditional system.
et realized as /i/, /ei/ or /æi/ (CANS corpus)

Are the speakers in these six places in MI and WI all attrited? Has the gender system changed?
Check with baseline, European Norwegian

Places in Norway where *et* has been realized as /iː/, /eɪː/, æi/ and /æ/ in the transcribed corpus.
• Checking the heritage speakers in America against baseline European speakers is important.

• Compare dialect background in American Norwegian speakers with the same speakers of the same background in Norway.

• Lesson learnt: Know your dialectology! (And use corpora to check your data.)
Case 3: Loan words

- Ekspekte (expect), farm, figgure ut (figure out), fil (field), grævel (gravel), råd (road), risess (break), mil (mile) (Johannessen & Laake 2012)
- Æker (acre), kaonti (county), enjin (engine), kændi (candy), fær (fair), casj (cash), geim (game) og gråsseri (grocery) (Johannessen & Hjelde to appear)

... but
• To study loan words in a language, such as Heritage Norwegian, you must make sure you are actually dealing with loan words.

English: *train*
Heritage Norwegian: *træn*
European Norwegian: *tog*
Is *træn* borrowed into Heritage Norwegian?

- Check older versions of Norwegian, and dialects.
- Norsk Ordbok (Dictionary of dialects, with many old sources, written and spoken):
  - *tren* (jarnvegstog, tog)
  - First example from 1878 (from the weekly magazine Fedraheimen).

Lesson:
- Learn your lexicography!
- Learn your history
Tren in Norway (Norsk Ordbok)
Problem 4: Yet another word

Heritage Norwegian: *Travle* ‘walk’

English: *travel* ‘to go on a trip or journey : to go to a place and especially one that is far away’ (Mariam-Webster)

Norwegian: Non-existent in standard dictionaries

(The discussion is based on Johannessen & Laake 2012, 2017, Johannessen & Hjelde, to appear)
Travle: Coon Valley, WI, USA, Westby, WI, USA, Sunburg, MN, USA, Webster, SD, USA, Fargo, ND, USA, Hatton, ND, USA, Billings, Mt, USA, Archerwill, SK, Canada, North Battleford, SK, Canada, Outlook, SK, Canada. Altogether: 3678 km.
• Documented early on: Flom (1901, 1926), Flaten (1901), Haugen (1953), Hjelde (1992), Johannessen & Laake (2012), plus letters from golddiggers Onon Bjørnsen Dahle & Knud Halvorsen Dahle (1852).

• Puzzle: Also documented in Danish (Kjær, Iver & Mogens Baumann Larsen 1987) and Swedish (Hasselmo 1974).

• Where does this non-standard meaning come from?
  – British dialects in immigration population? (Haugen 2953)
  – DARE (Dialect Atlas over Regional English states North Carolina
  – Norwegian dialects?
Norsk Ordbok
• Lesson: Search as many sources as you can. But it may not be enough…
Conclusion

- Heritage studies are exciting, but demanding
- There are many resources that can be used to support one’s ideas
- But one can never be too careful
  - statistical tests
  - dialectology
  - lexicography
  - language history
  - the dialects of other languages
Thank you!
References


• Flom, George T. (1926): English Loanwords in American Norwegian. As spoken in the Koshkonong settlement, Wisconsin. *American Speech* Vol. 1, s.541-558


• Jensen, Bård Uri. 2016. Note: The assumption of independent observations.


• Åfarli. T.A.. M.B. Grimstad & K.V. Subbarao (2013) Dakhini and