UiT The Arctic University of Norway

Two Nordic research infrastructures for syntactic variation

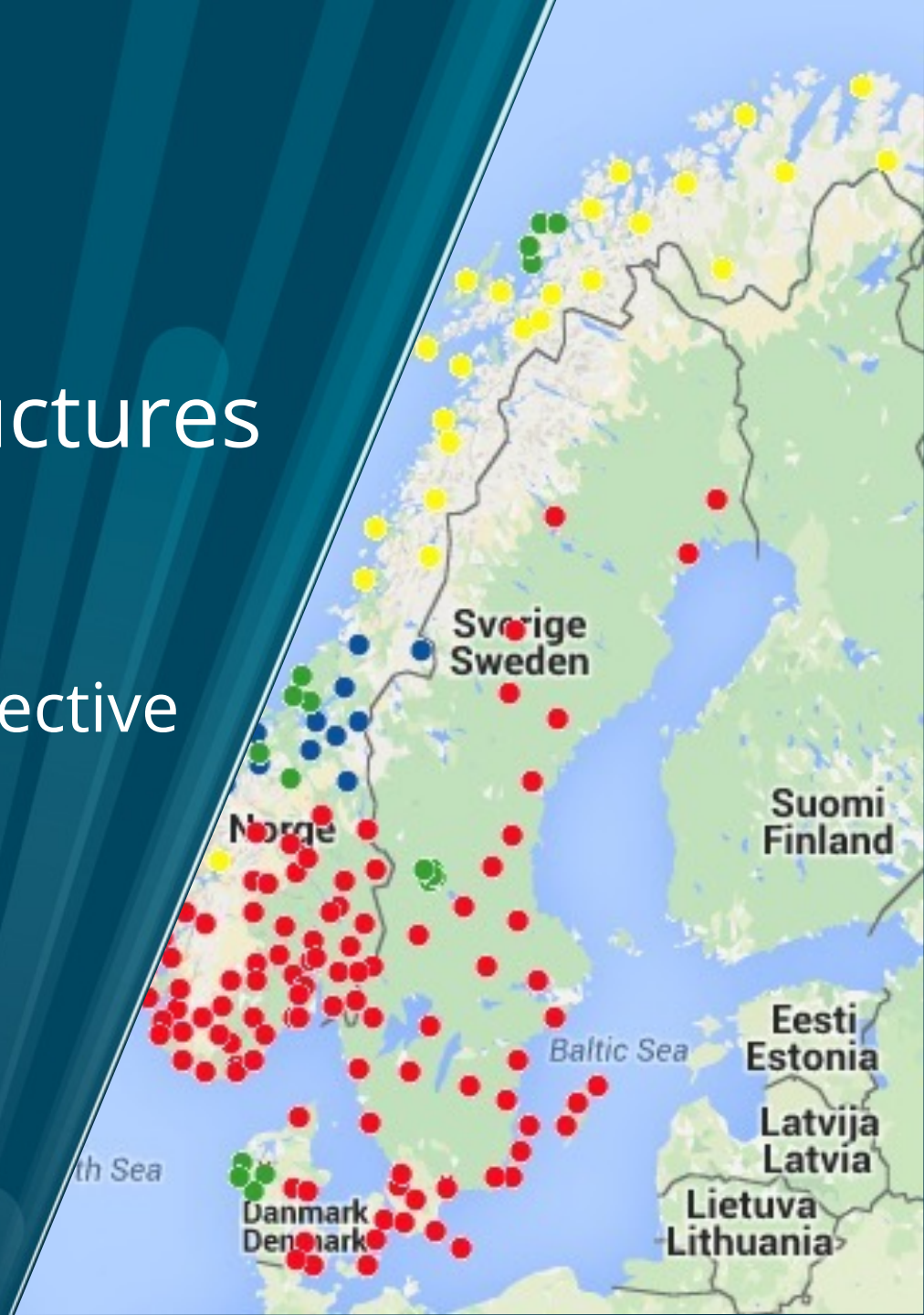
Results, limitations and a future perspective

Øystein A. Vangsnes & Maud Westendorp

UiT The Arctic University of Norway

REEDS 2023, Amsterdam

29 June 2023



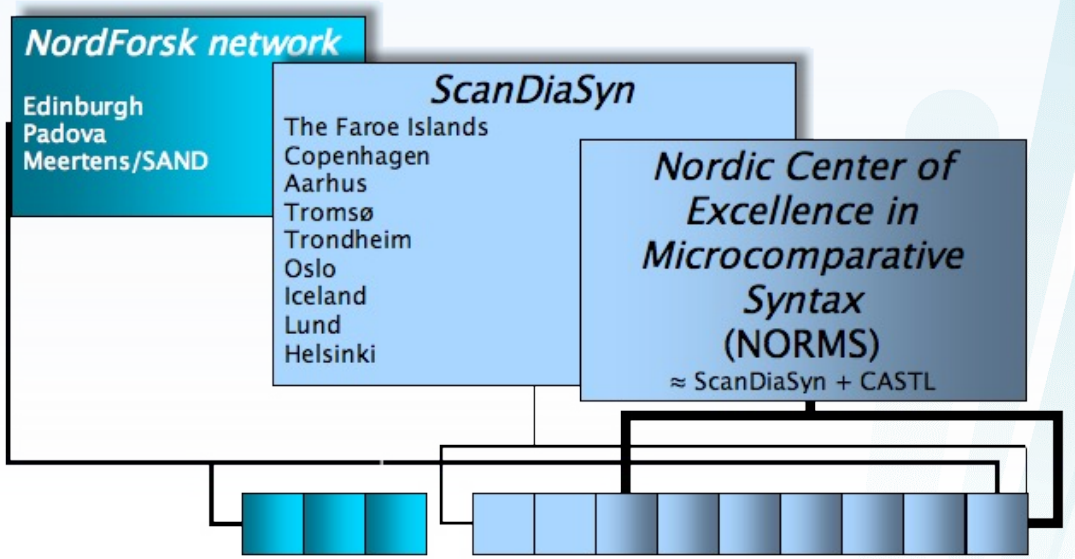
ScanDiaSyn

NORDISK DIALEKTSYNTAKS

Scandinavian Dialect Syntax

Mállýskutilbrigði í norrænni setningagerð

- ScanDiaSyn was a **project umbrella** of **10 research groups** across all five Nordic countries and the Faroe Islands
- Operative phase: **2005–2012**
- Main objective: Map and study the **syntactic variation** across the North Germanic (Scandinavian) dialect continuum, a **largely unexplored** field
- Dialectologists, theoretical syntacticians, computational linguists
- National and Nordic funding (complex situation), including a Nordic CoE (NORMS)
- Three liaised non-Nordic groups (Edinburgh, Padova/ASIT, Meertens Institute/SAND); part of the Edisyn network (European Dialect Syntax)



ScanDiaSyn

NORDISK DIALEKTSYNTAKS

Main achievements:

- significantly improved knowledge of syntactic variation across the North Germanic language area;
- disseminated through a large number of publications, including the **Nordic Atlas of Language Structures (NALS) Journal**
- operative research infrastructure in the form of (i) the **Nordic Dialect Corpus (NDC)** and (ii) the **Nordic Syntax Database (NSD)**, developed at the Text Laboratory at the University of Oslo

Current status:

- the corpus search interface continues to be updated; the database does not

This presentation

Display the ScanDiaSyn research infrastructure by looking at:

- the phenomenon of non-V2 in *wh*-questions across Norwegian dialects
  - based on the corpus data
  - based on the database data (questionnaires)
- COMP trace effects across North Germanic

Invite feedback on

- how the ScanDiaSyn infrastructure may be developed in the future
- and how it may feed into the goals of REEDS (infrastrcture, methodologies, interdisciplinary research, sustainable research collaborations)

Non-V2 in matrix *wh*-questions

- A well-known and much studied phenomenon (early source Iversen 1918)
- No verb movement
- Manifests as *som*-insertion (COMP) in subject questions
- Subject to micro-variation: (i) ± short *wh*-constituents only, (ii) ± subject questions only, (iii) long *wh*-constituents in subject questions only (short otherwise)

(1) Ka / #[kor mange] du (faktisk) kjøpte (*faktisk)?
what / how many you (actually) bought (actually)
'What/how many did you (actually) buy?'

(2) Kem / #[kor mange] som (faktisk) kom (faktisk)?
who / how many SOM (actually) came (actually)?
'Who / how many (actually) came?'

All 737 speakers (2754289 tokens) selected from 183 places in 5 countries

Informant code
Recording year
Birth year
Gender
Age
Age group
Place
Area
Region
Country
Genre

Hide filters | Reset form

Nordic Dialect Corpus v. 4.0

Nordic Dialect Corpus
ScanDiaSyn/NORMS

Simple | **Extended** | CQP query | Search

min max Lemma Start End Middle Phonetic form Segment initial

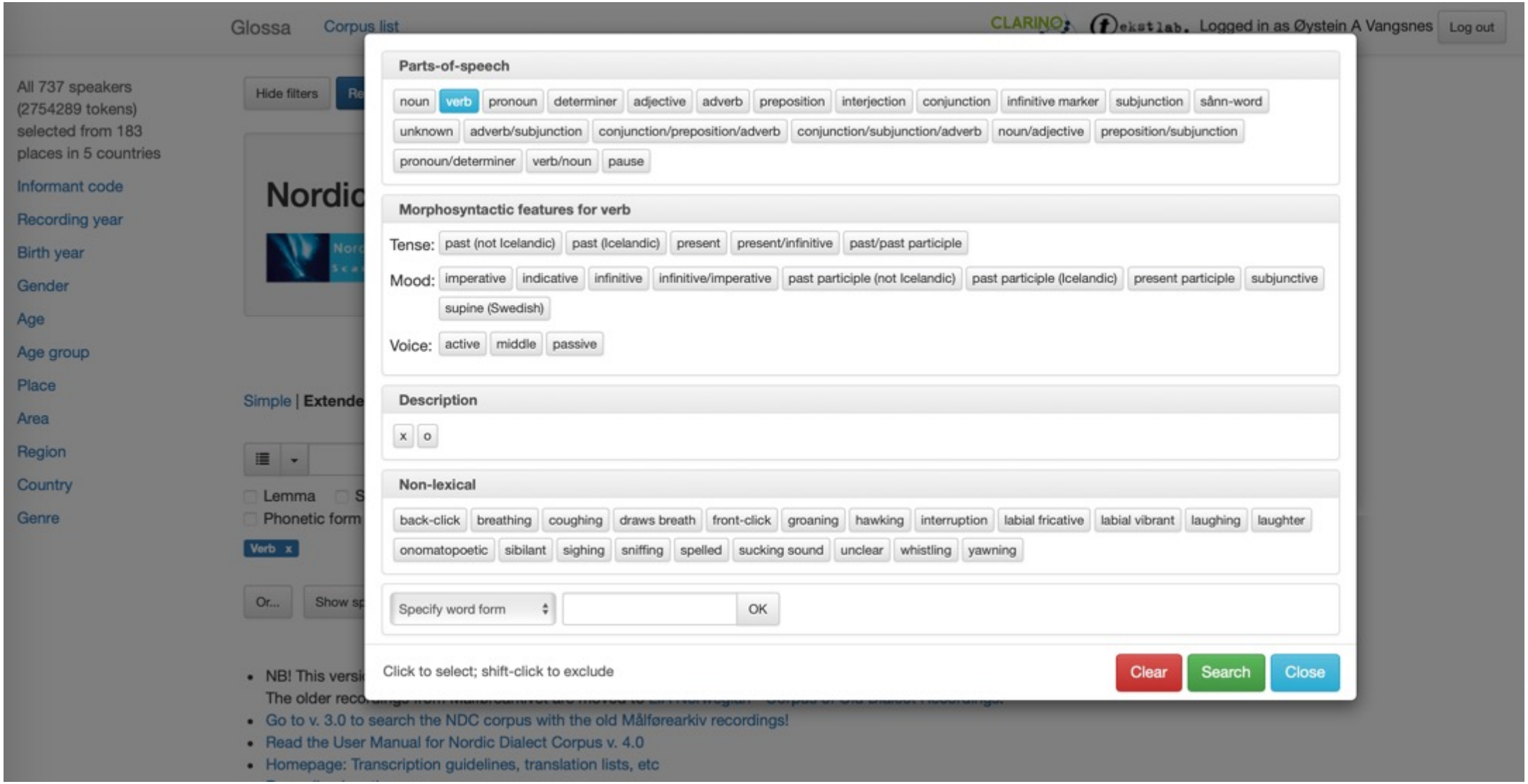
min max Lemma Start End Middle Phonetic form

min max Lemma Start End Middle Phonetic form Segment final

Or... | Show speakers | random results (with seed:)

- Recording locations
- User license for the corpus
- Report errors in the corpus
- Read or download the transcriptions:
  - User license for the transcriptions
  - In html-format
  - In txt format with informant codes: - Orthographic transcriptions - Phonetic transcriptions
  - In txt format without informant codes: - Orthographic transcriptions - Phonetic transcriptions
- The old search interface
- There is a technical issue with audio/video playback in Safari. We recommend using another browser.
- How to refer to the corpus

Searchable categories (±linguistic)



Search strings for *wh*-questions

- ‘Sentence’ has no status in this *speech* corpus.
- ‘Segment’ comes closest.
- ‘#’ marks pauses/breaks in the conversation.

a.	<segment initial ‘WH’> + <‘not verb’> (0 words between)	(non-V2)
b.	<#> + < ‘WH’> + <‘not verb’> (0 words between)	(non-V2)
c.	<segment initial ‘WH’> + <‘verb’> (0 words between)	(V2)
d.	<#> + < ‘WH’> + <‘verb’> (0 words between)	(V2)

438 of 737 speakers (1997920 of 2754289 tokens) selected from 111 places in 1 country

Informant code

Recording year

Birth year

Gender

Age

Age group

Place

Area

Region

Country

× Norway

Genre

Hide filters | Reset form

Simple | **Extended** | CQP query | Search

hva | min | max

Lemma | Start | End | Middle | Phonetic form | Segment initial

Lemma | Start | End | Middle | Phonetic form | Segment final

!Verb x

Or... | Show speakers | random results (with seed:)

Concordance | Map | Statistics

Found 424 matches (9 pages)

Sort by position | Download

« ‹ 1 › »

aal_02uk Trans		**hva jeg**	skal til sommeren ja
translated by Google	what I'm going to do in the summer yes		
		kå e	ska te somarn ja
aal_04gk Trans		**hva det**	var jeg kunne ikke skjønne hva det var da # så e nei det hadde jeg ikke tenkt at det var noe hva det var heller
translated by Google	what it was I couldn't figure out what it was then # so e no I didn't think it was anything what it was either		
		ko de	va e kunna kji sjønne ko re va då # så ee næi de hadde e kji tenngkt att de va nokko ko de va hellde
aaseral_02uk Trans		**hva ?**	
translated by Google	what?		
		kå ?	
aaseral_01um Trans		**hva han**	fikk i den ?
translated by Google	What did he get in it?		
		ko hann	fe i denn ?

Results (after data cleaning)

(Vangsnes & Westergaard 2014; Vangsnes & Johannessen 2018)

	V2		non-V2		Total
	n	%	n	%	n
hva ‘what’	284	43%	376	57%	660
hvem ‘who’	50	45%	61	55%	111
hvor ‘where’	68	52.3%	62	47.7%	130
når + hva tid ‘when’	58	73.4%	21	26.6%	79
hvorfor ‘why’	46	97.9%	1	2.1%	47
hvordan ‘how’ (manner)	119	93.0%	9	7.0%	128
‘wh-XP’	169	95.3%	8	4.7%	177
Total	794	59.6%	538	40.4%	1332

Geography

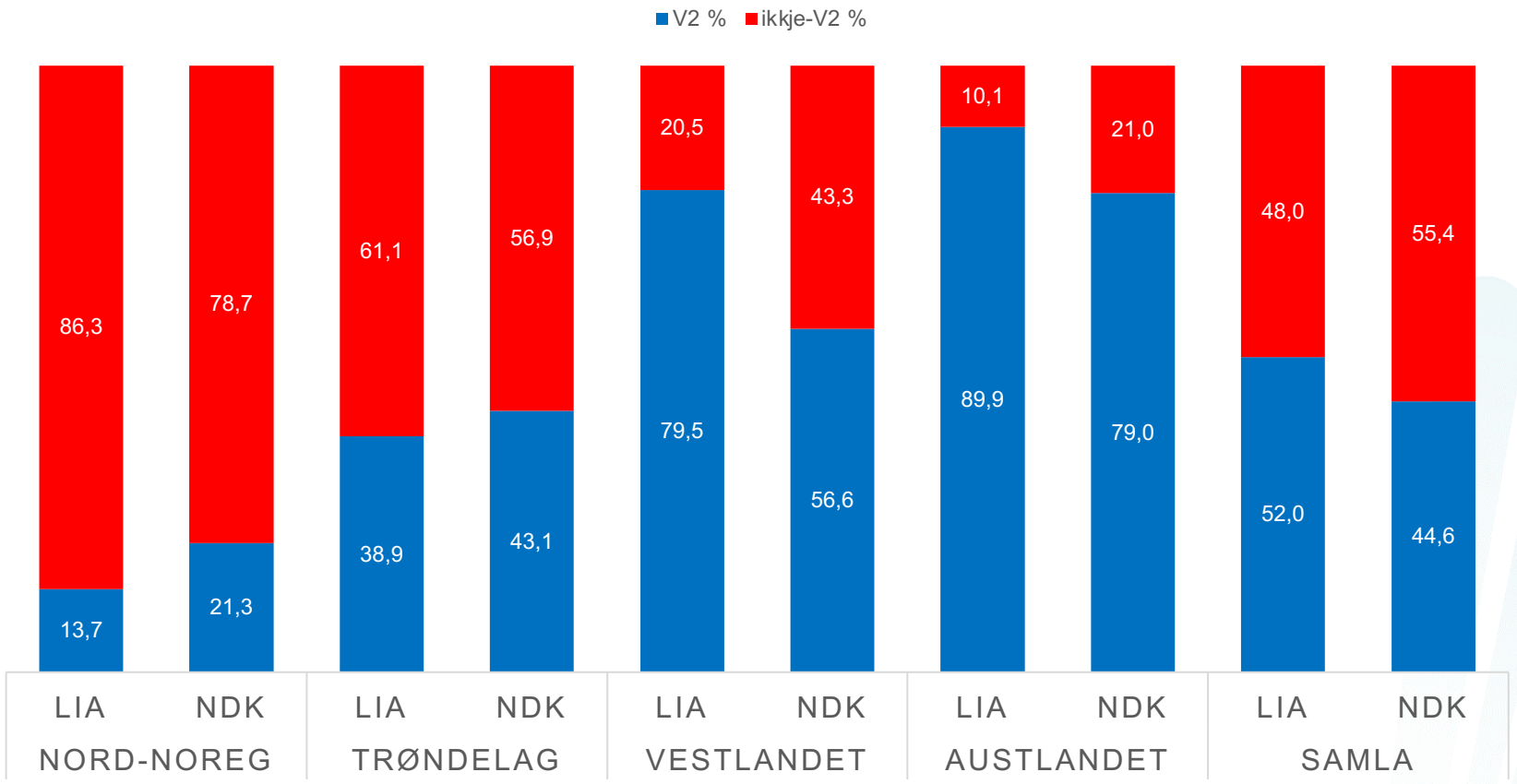
(Vangsnes & Westergaard 2014; Vangsnes & Johannessen 2018)

	North	Central	West	East	Totalt
what + V2	65 (22,6%)	37 (42%)	85 (53,8%)	97 (76,4%)	284 (43%)
what + non-V2	222 (77,4%)	51 (58%)	73 (46,2%)	30 (23,6%)	376 (57%)
who + V2	8 (17,4%)	5 (41,6%)	14 (56%)	23 (82,1%)	50 (45%)
who + non-V2	38 (82,6%)	7 (58,4%)	11 (44%)	5 (17,9%)	61 (55%)
where + V2	8 (17%)	8 (50%)	25 (69,4%)	27 (87,1%)	68 (52,3%)
where + non-V2	39 (83%)	8 (50%)	11 (30,6%)	4 (12,9%)	62 (47,7%)

Language Infrastructure made Accessible

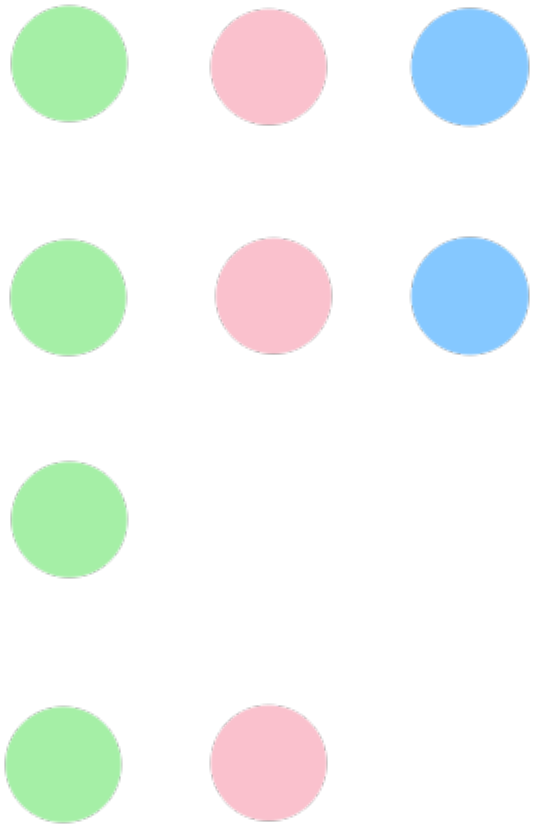
- Corpus project (2014–2019) digitizing and transcribing old recordings from university dialect archives

Comparison of the distribution of non-V2 with short *wh*-items in LIA and NDC



Word order variation in *wh*-questions

Nordic Syntax Database (NSD) – Westendorp (2018)



1. Kva du **heiter**?
   what you **called**
   'What is your name?'
2. Kven som **sel** fiskeutsyr her i bygda?
   who COMP **sells** fishing.gear here in village
   'Who sells fishing equipment around here?'
3. Kva tid du **gjekk** ut av skolen, då?
   what time you **went** out of skole then
   'When did you leave school?'
4. Kor mange elever som **går** på skulen?
   how many students COMP **go** on skole
   'How many students go to this school?'

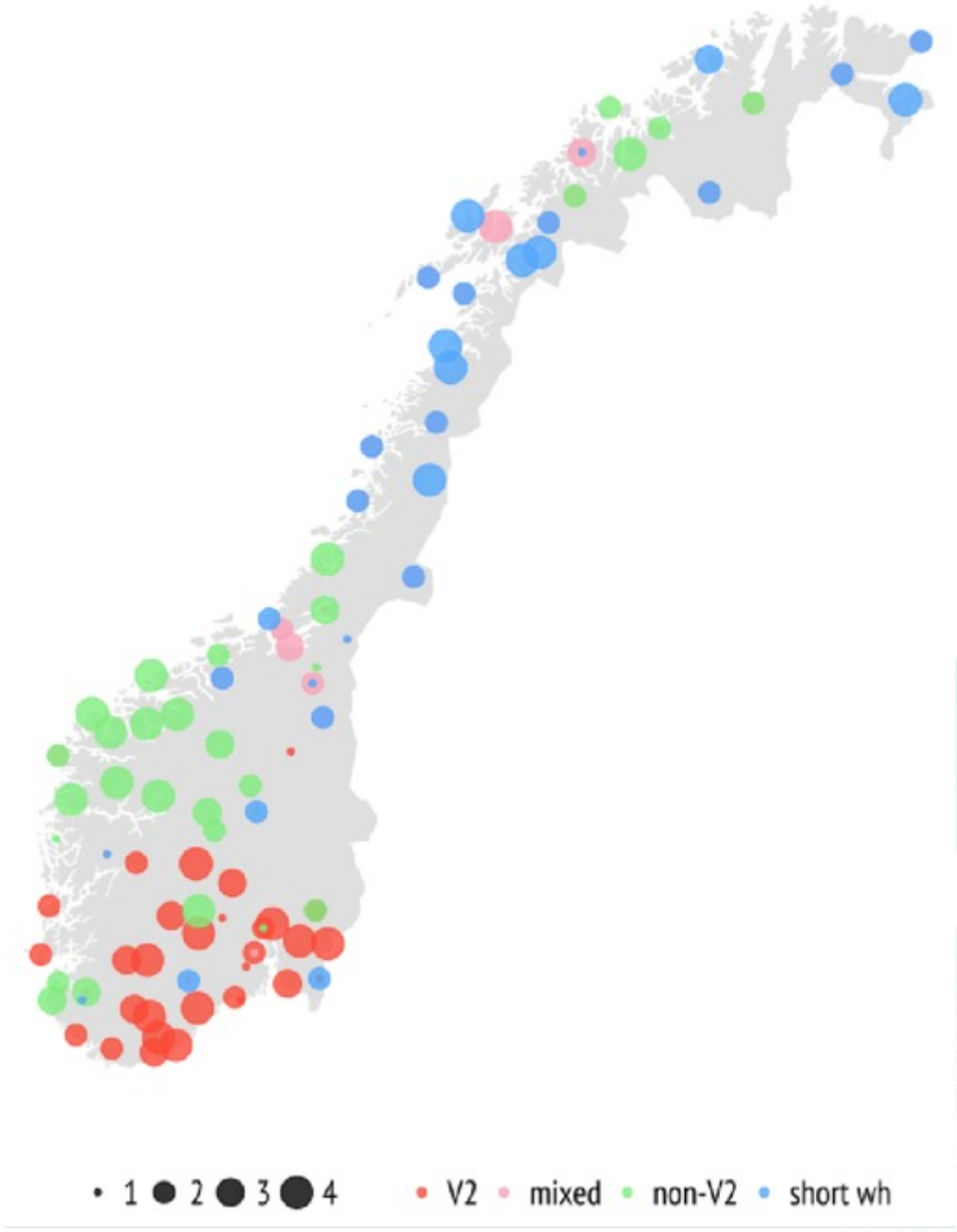
Dialectal *wh*-grammars

Nordic Syntax Database (NSD) – Westendorp (2018)

Four types of *wh*-grammars:

- only V2
- non-V2 but only with short *wh*-elements
- non-V2 with *wh*-subjects and short *wh*'s
- non-V2 possible with all *wh*-elements

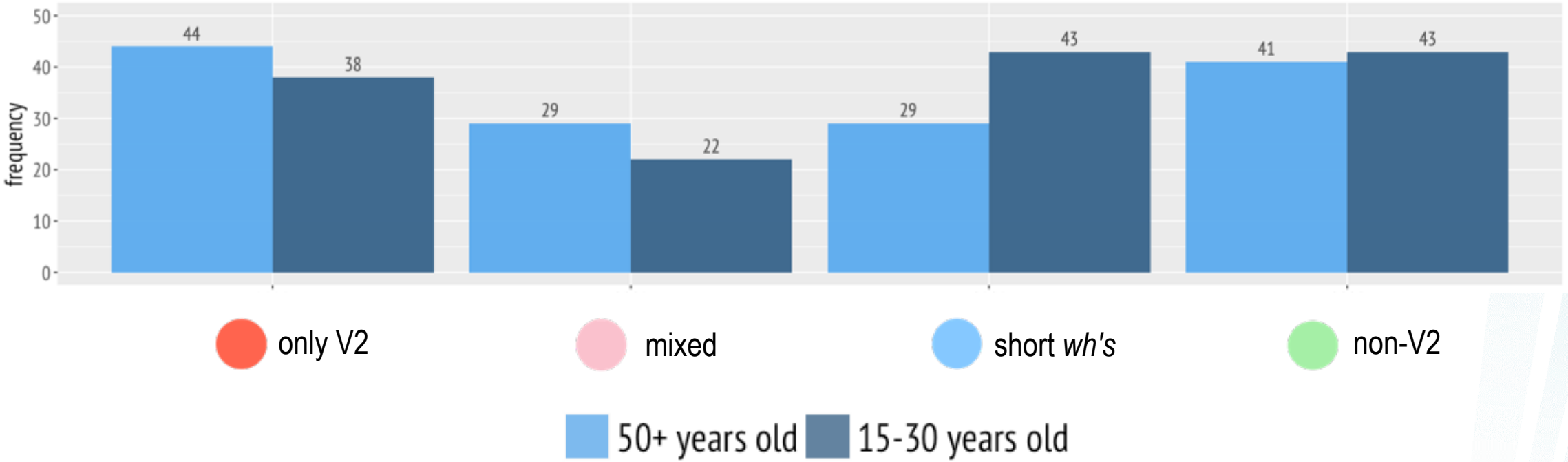
NB! All dialects always allow V2.



Dialectal *wh*-grammars

Nordic Syntax Database (NSD) – Westendorp (2018)

Figure 8. Frequency of use of different dialect types split by age group.



Dialectal *wh*-grammars

Nordic Syntax Database (NSD) – Westendorp (2018)

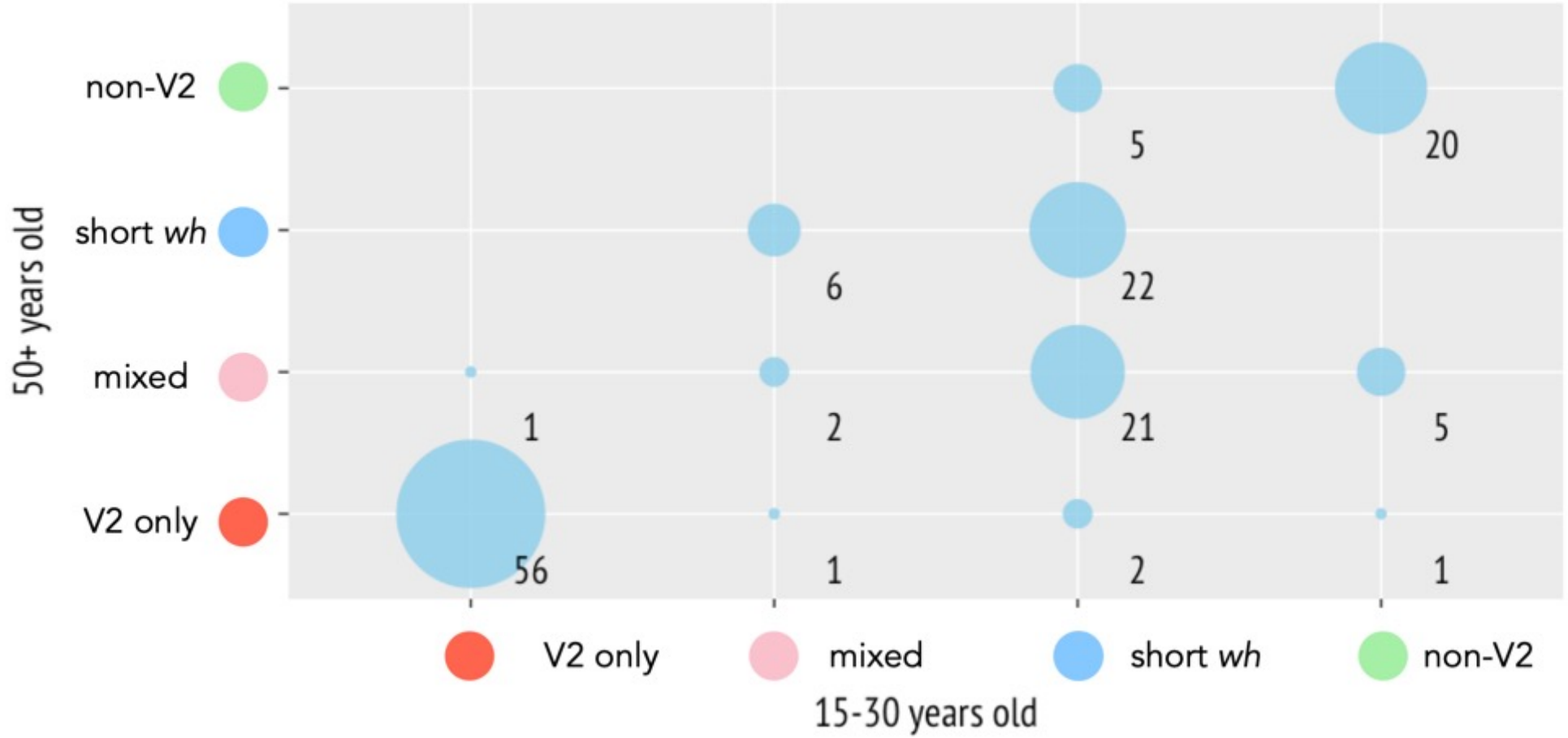


Figure 9. Cross tabulation of different dialect type combinations between young and old age group per location (without medium scores).

COMP *trace effects across North Germanic

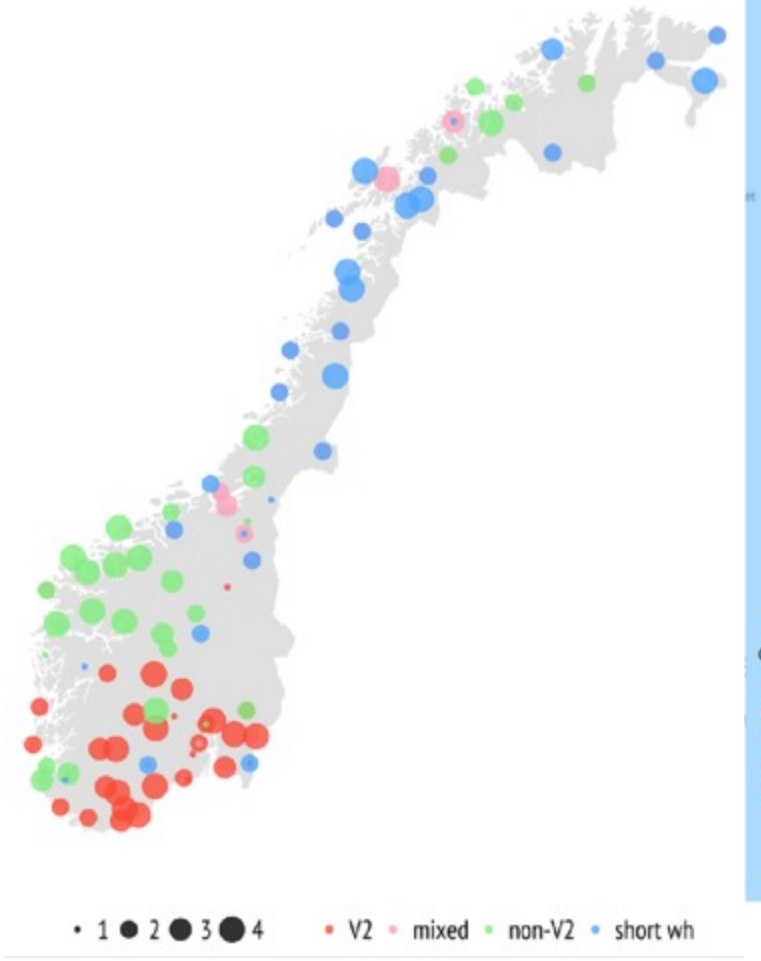
Nordgård's Condition (Nordgård, 1985):

A dialect allows non-inverted word order in matrix wh-*questions iff the dialect allows insertion of the complementizer* som *under extraction of the embedded subject.*

(5)	a.	Kven	trur	du	__	har	gjort	det?
		who	*think*	*you*	*Ø*	*has*	*done*	*it*
	b.	Kven	trur	du	at	har	gjort	det?
		who	*think*	*you*	*that*	*has*	*done*	*it*
	c.	Kven	trur	du	som	har	gjort	det?
		who	*think*	*you*	*SOM*	*has*	*done*	*it.*

'Who do you think [Ø/that/SOM] has done it?'

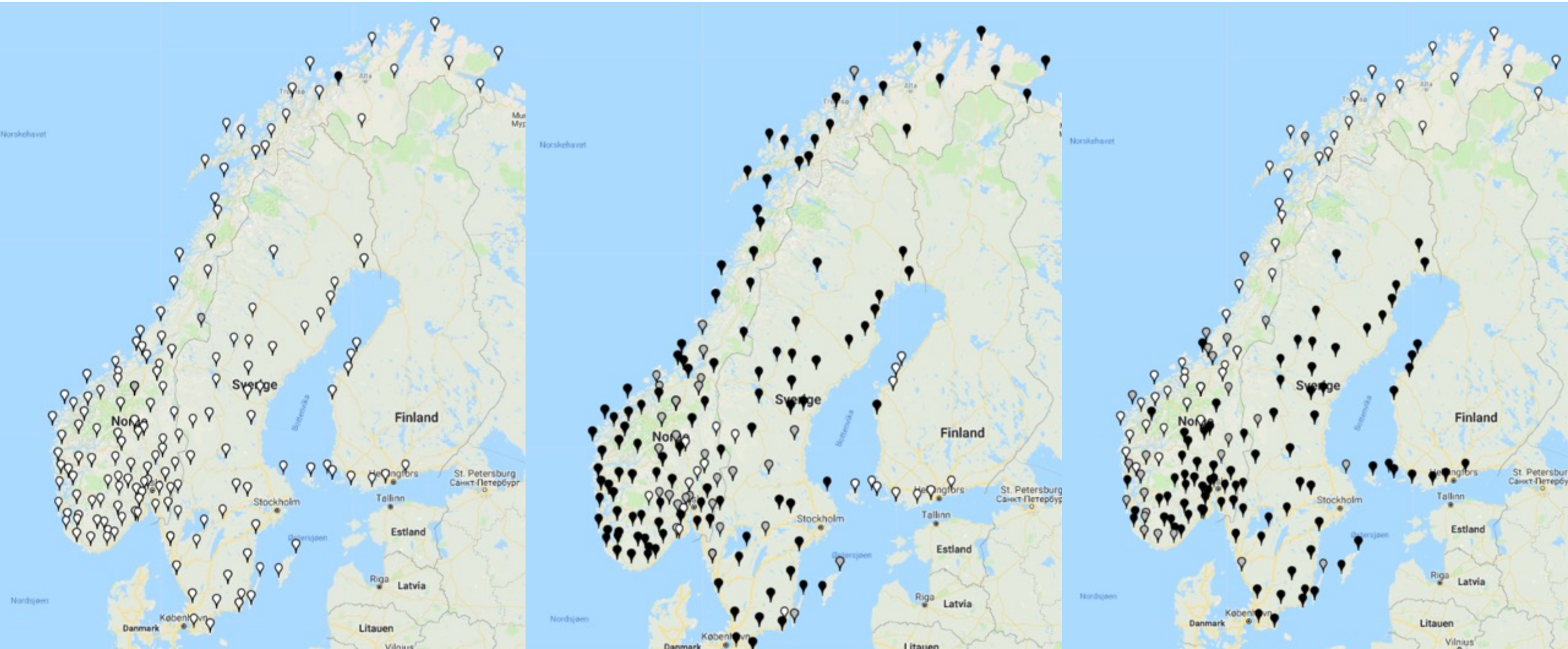
COMP *trace effects across North Germanic



Rightmost map:
Black markers indicate a low score for *som*-insertion under extraction of a *wh*-subject, white markers a high score

Leftmost map:
Red dots indicate places where non-V2 in matrix *wh*-clauses are dismissed.

no COMP *at* ('that') *som* (REL)



Danish and Faroese

“Relativizers” may also be inserted under *wh*-extraction in Danish and Faroese:

(6) Hvem tror du (at) der har gjort det?
who think you that there has done it
‘Who do you think has done it?’

(7) Hvør trýrt tú ið/??sum/*at hevur gørt tað?
who think you IÐ / SOM / that has done it
‘Who do you think has done it?’

Vangsnes (2019)

COMP trace effects across North Germanic varieties

A typology of COMP *trace* effects across North Germanic:

(i) a (declarative) complementizer in the east (Fenno-Swedish, Eastern Norwegian),

(ii) a resumptive complementizer in the west (most of Norwegian, Faroese), and

(iii) a resumptive XP element in the south (Danish)

Summary

- Infrastructure:
  - Nordic Dialect Corpus
  - Nordic Syntax Database
  - (LIA norsk)

- Limitations:
  - New methodologies using data from the infrastructures

- Future perspective:
  - how the ScanDiaSyn infrastructure may be developed in the future
  - and how it may feed into the goals of REEDS (infrastrcture, methodologies, interdisciplinary research, sustainable research collaborations)

Nå, ka dokker trur?

now what you.PL think

Takk for oss!

thanks for us