RESEARCH

The Nordic research infrastructure for syntactic variation: Possibilities, limitations and achievements

Øystein Alexander Vangsnes[1,2] and Janne Bondi Johannessen[3]

[1] CASTL and AcqVA, Department of Language and Culture, UiT The Arctic University of Norway, Langnes, 9037 Tromsø, NO

[2] Department of Language, Literature, Mathematics and Interpreting, Western Norway University of Applied Sciences, NO

[3] MultiLing, ILN, University of Oslo, Blindern N-0317, Oslo, NO

Corresponding author: Øystein Alexander Vangsnes (oystein.vangsnes@uit.no)

The Scandinavian Dialect Syntax project was a collaboration between ten research groups from all of the five Nordic countries lasting for a period of about ten years. Besides resulting in a large number of scientific papers and theses on a range of different topics, a concrete outcome of the collaboration was the establishment of lasting research infrastructures in terms of two databases: the Nordic Dialect Corpus (NDC) and the Nordic Syntax Database (NSD). This paper first describes the two infrastructures and then proceeds to showcase how they may be used for the exploration of two selected dialect syntactic topics: the relative placement of sentence adverbs and infinitive markers (±split infinitives) across varieties of Mainland North Germanic, and the lack of Verb Second in *wh*-questions across Norwegian dialects.

1 Introduction

Since 2003 and for a period of about ten years a network of ten research groups in the five Nordic countries worked within the Scandinavian Dialect Syntax project (ScanDiaSyn) towards mapping syntactic variation across the North Germanic dialect continuum. Two major research tools grew out of the collaboration: the Nordic Dialect Corpus (NDC) and the Nordic Syntax Database (NSD), see Section 2.

In 2014 *The Nordic Atlas of Language Structures Online (NALS)* was launched with some 50 papers based on data from the two databases organised under the following six general topics: (i) Noun phrases, (ii) Verb phrase: Argument structure and verb particles, (iii) Verb placement, (iv) Middle field/TP: Subject placement, object shift, auxiliaries and tense marking, (v) Left and right periphery: complementisers, questions, extractions etc., and (vi) Binding and co-reference. NALS is formally organised as a journal, and has continued afterwards to publish papers that map variation across the North Germanic dialect continuum.

In this paper we will discuss two syntactic phenomena on the basis of data available from the dual NDC/NSD research infrastructure: (i) the relative placement of adverbs and infinitive markers (± split infinitive), and (ii) non-V2 in matrix *wh*-questions across Norwegian dialects. For each of these we discuss possibilities, limitations and achievements posed by the research infrastructure as promised by the title of the paper. Section 2 presents the two types of research infrastructure. Section 3 presents the investigation of placement of adverbs and infinitives, and also illustrates how the corpus and database can

be used to find the empirical evidence needed. Section 4 contains a thorough presentation of the variation of word order in *wh*-questions, while Section 5 concludes the paper.

2 The two research infrastructures

The ScanDiaSyn network of research groups consisted of a number of smaller and bigger projects funded by a variety of sources. In effect, ScanDiaSyn was therefore a project umbrella and there was also funding for the network itself from the Nordic research bodies NordForsk and NOS-HS. In the various countries national funding was obtained to carry out basic and systematic data collection in the projects DanDiaSyn, FinDiaSyn, IceDiaSyn, NorDiaSyn, and SweDiaSyn. The Norwegian national project was responsible for the technical solutions, including building the Nordic Dialect Corpus (Johannessen et al. 2009, 2014) and the Nordic Syntax Database (Lindstad et al. 2009) as well as getting the annotations (double transcriptions and tagging) done (Johannessen 2017). Furthermore, between 2005 and 2010 the ScanDiaSyn umbrella also included substantial funding for the Nordic Center of Excellence in Microcomparative Syntax (NORMS), which financed a number of postdoctoral visiting fellows, cross-institutional thematic research groups, fieldwork trips to targeted areas in the Nordic countries, as well as seminars and conferences. Between 2005 and 2010 a project blog was kept up with entries written both in Scandinavian and English (mainly), including quite extensive reports from the NORMS field trips, and these reports can be found at https://tinyurl.com/NORMSfieldwork.

In total, data from 228 different locations across the Nordic countries were collected in the project, and the data were of two types: (i) recordings of spontaneous speech from conversations between and interviews with dialect speakers, and (ii) judgment data on a list of prepared test sentences that probe a number of different syntactic constructions. These data formed the basis for the two different infrastructures in the project: the *Nordic Dialect Corpus* and the *Nordic Syntax Database*.

The Nordic Dialect Corpus (NDC) consists of spontaneous speech. Ideally, each location in the corpus should be represented by four speakers (two age groups and both genders), and each speaker would be recorded both in an interview and a free conversation. However, in practice, the different funding situations in the different countries meant that this could not always be achieved. For example, in Sweden, there was less funding, but fortunately the project was given access to interview data from a recently completed project, SweDia 2000. On the other end of the scale was Norway, where funding covered extensive fieldwork across the whole country in addition to the development of the technical research infrastructure. The NDC corpus contains transcribed recordings with over 3 Million words, by 823 dialect speakers from 228 locations in six countries (Denmark, Faroe Islands, Finland, Iceland, Norway and Sweden), mainly sampled between 2005 and 2010.

The Nordic Syntax Database (NSD) contains judgment data for a number of test sentences (140–240, depending on country) probing variation in various (morpho)syntactic phenomena from more than 900 informants – many of the same ones as in the NDC. The sentences were carefully selected to represent grammatical phenomena that we knew varied across dialects in one or more of the five languages. A major goal was to present sentences even in locations where a phenomenon was thought not to exist. This way we would obtain negative as well as positive data, making it possible to draw new syntactic isoglosses across the North Germanic dialect continuum.

The locations represented in the infrastructure were chosen to ensure a good geographical distribution and also to cover well-known local/regional dialect boundaries, and although most locations are in rural areas there are also some cities in the Norwegian and Danish list of locations. For the Swedish speaking area, the SweDia 2000 material (see

above) only included material from rural and semi-rural locations, and hence no Swedish
(and Finnish) cities are represented in the infrastructure. Since geographical distribution
was prioritised, the locations were not balanced for population size, hence there are both
"small" and "big" dialects in the sample.

The fieldworkers were a good mix of senior and junior researchers and student assistants
who travelled to the locations of the informants to carry out the data collection. In most
cases two fieldworkers would do the data sampling together. In some cases the fieldwork-
ers would speak a similar dialect as the informants, but in other cases not.

The test sentences in the questionnaire were pre-recorded by a speaker of the same
regional variety so that the pronunciation would be the same as or similar to that of the
informants. This was done to avoid the sentences being deemed unacceptable because of
pronunciation. The informants were instructed to judge the sentences on a Likert scale
from 1 (bad) to 5 (good) according to their own dialect intuitions, and they would give
their judgments after hearing each pre-recorded sentence. These questionnaire sessions
lasted about one to one and a half hours. The recording sessions consisted of an interview
of about 15–20 minutes with one of the fieldworkers and a conversation with another
informant of about 30–45 minutes. Sampling data from four informants at one location
normally took a full work day.

The informants were typically recruited through a local contact person according to
a set of criteria targeting traditional dialect speakers. Beyond age and gender informa-
tion sociolinguistic information about the informants was not recorded. The inform-
ants were not paid, but given a symbolic gift as a token of gratitude and they were
also served coffee, tea and soft drinks as well as fruit and (non-crunchy) candy at the
sessions. For further details about the project logistics, including methodologies, see
Vangsnes (2007a; b), Johannessen et al. (2008), Lindstad et al. (2009), Johannessen et
al. (2014).

The papers in the *NALS Journal*, which typically exploit both the the NDC (corpus) and
the NSD (judgment database), show that the dialect infrastructure developed under the
ScanDiaSyn umbrella does indeed allow researchers to investigate new isoglosses and dia-
lect phenomena across the Nordic countries. The two case studies to be presented in this
paper should serve to make the same point.

3 Case study 1: Relative placement of infinitival marker and negation (±split infinitives)

The differing relative placement of infinitival markers and adverbs across the Scandinavian
written languages is a well-known issue (see for example Hulthén 1947; Faarlund et al.
1997). The received wisdom is that Danish requires the adverb (especially the negative
adverb) to be placed before the infinitival marker, as in (1), whereas Swedish requires the
adverb to follow the infinitival marker, as in (2), and Norwegian is supposed to accept
both orders. The "Danish pattern" is what often is referred to as "unsplit infinitives"
since the adverb does not split the infinitival marker from the verb, and conversely the
"Swedish pattern" is generally referred to as "split infinitives" precisely since the adverb
separates the infinitival marker from the verb. A third pattern, given in (3), is the standard
word order in Icelandic infinitivals (with an infinitival marker), and this pattern was also
tested for the mainland languages (and Faroese). The sentences are all given here with
Bokmål Norwegian words and orthography.

(1)   Kjell hadde lenge prøvd **ikke å komme** for sent på jobb.
        kjell had   long tried not to come   too late at work
        'Kjell had for a long time tried not to come too late for work.'

(2)   Kjell hadde lenge prøvd **å ikke komme** for sent på jobb.

(3)   Kjell hadde lenge prøvd **å komme ikke** for sent på jobb.

The issue of "split infinitive" is well-known also from English grammar and is a source of controversy where prescriptivists favour the use of unsplit over split infinitives (see Huddleston & Pullum 2002: 581f, for references). This same prescriptivism can be witnessed also in the context of Norwegian as unsplit infinitives have traditionally been the recommended word order. Furthermore, the Norwegian reference grammar (Faarlund et al. 1997: 997) claims that while both orders are possible in Norwegian, the unsplit pattern is the most natural for many language users (Faarlund 1997: 997). We will see below that the NSD database does not support this claim.

Figure 1 is a screenshot from the NSD of sentence (1) as it was presented to the informants, and we see that the Swedish test sentence is worded differently but nevertheless probes the same word order as its Norwegian and Danish counterparts. Similar adjustments across the language-specific questionnaires were made in several cases, but each "bundle of test sentences" probing a specific phenomenon is always given a unique identity in the database, in this case the number 143.

Searching the database gives a long list with all the answers of all the informants, divided over several results pages. Each informant would grade this sentence (and all the others presented to them) on a scale from 1 to 5. These results are rendered as colour codes in the database, to enable the researcher to assess the results at a glance, see Figure 2.

Even more visually illustrative are the maps that can be generated from the result page. The results from the sentence evaluations performed across the three countries Denmark, Norway and Sweden show very clear and different results, see Maps 1, 2 and 3. A white marker means that a sentence has a mean score of 4 or higher at that geographical location, whereas a black marker means that it has a mean score of 2 or lower. In other words, a white marker indicates acceptability, a black one non-acceptability.

There is a clear acceptance of the unsplit, "Danish pattern" in Denmark. Furthermore, in Sweden the unsplit pattern is quite clearly not accepted, but more strikingly, at the great majority of Norwegian locations the pattern is also dismissed. The split, "Swedish pattern", on the other hand, is accepted not just in Sweden, but also in most of Norway, although there is an enclave in the central part of the country (the Trøndelag area) where the split pattern seems to be rejected at several locations. In turn, as Map 3 makes evident, the "Icelandic pattern" with the verb preceding negation is not accepted by anyone in the mainland countries (and not in the Faroe Islands either).

143	🇩🇰	Kjeld havde	længe	prøvet	ikke	at	komme	for	sent på		arbejde
		Kjeld have-Pret	long	try-PrtPast	not	to(Inf)	come-Inf	too	late	on	work
		For a long time, Kjeld had been trying not to come too late for work.									
	🇳🇴	Kjell ha-dde	lenge	prøv-d	ikke	å	komm-e	for	sent på		jobb.
		Kjell have-Pret	long	try-PastPrt	not	to(Inf)	come-Inf	too	late	on	work
		For a long time, Kjell had tried not to arrive late at work.									
	🇸🇪	Kjell ska		försök-a	inte		att	skäll-a på		barn-en vid	middagsbord-et.
	🇫🇮	Kjell shall.Pres		try-Inf	not		to(Inf)	yell	on	child-Pl by	dinner.table-Neut.Sg.Def
		Kjell will try not to yell at the children by the dinner table.									

Figure 1: The test sentence for unsplit infinitives in Danish, Norwegian and Swedish, assigned number 143 in the Nordic Syntax Database.

Figure 2: Some of the results of the evaluations of sentence no. 143 (the unsplit, "Danish pattern").
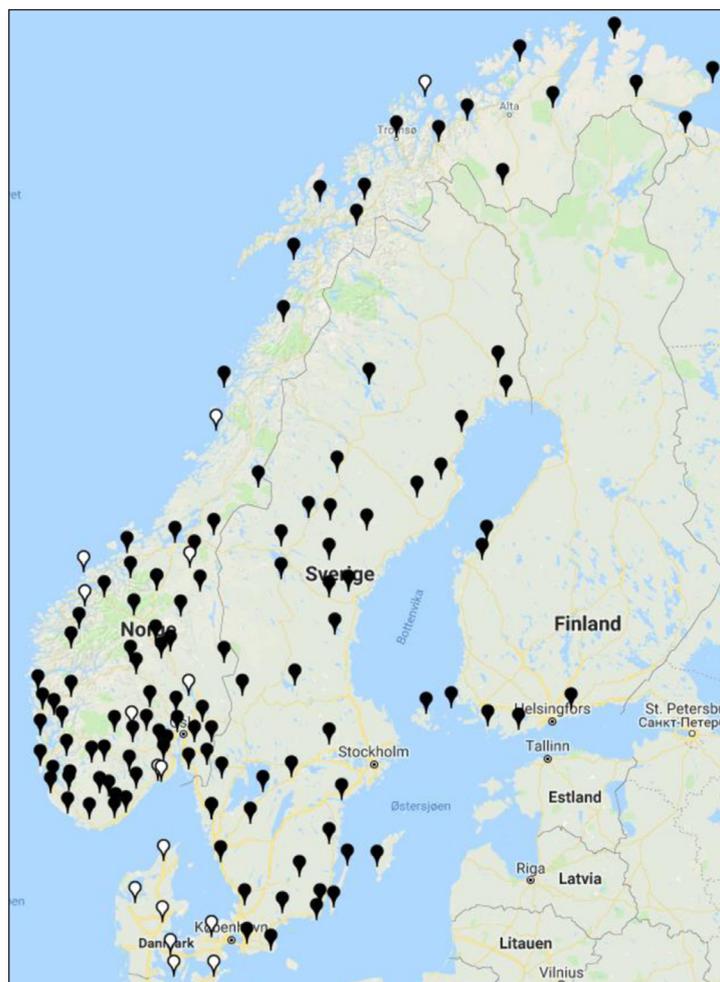
**Map 1:** Unsplit infinitive: neg – C – Vinf.
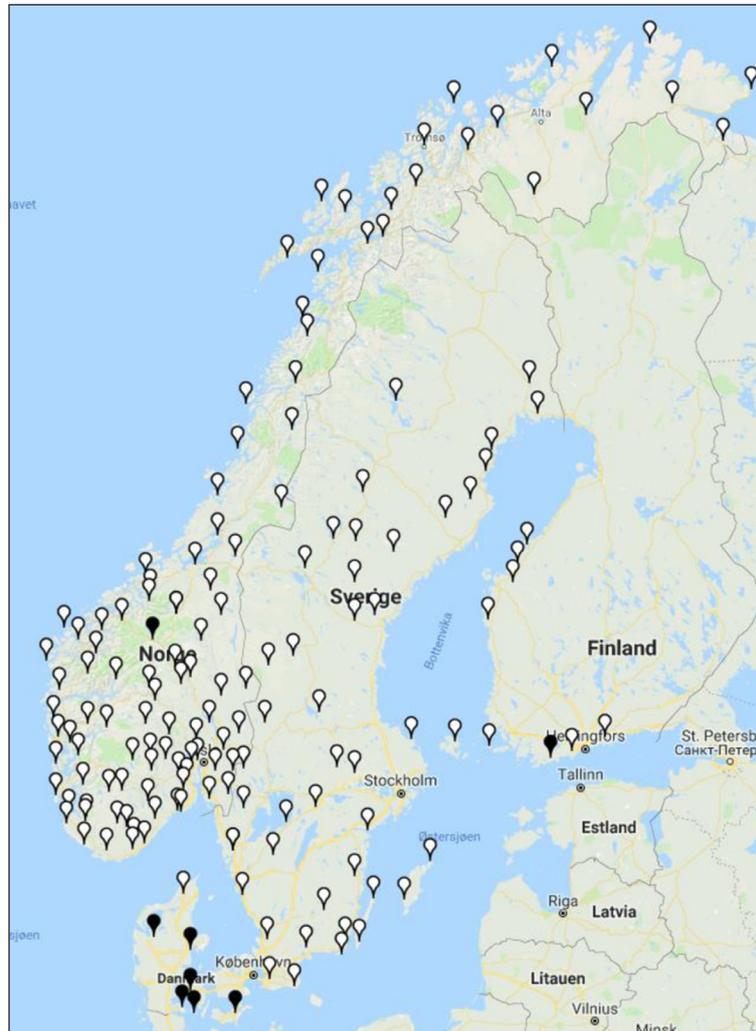
Map 2: Split infinitive: C – neg – Vinf.

A striking result from the database data, evident from the maps, is that the "Danish pattern" is hardly accepted at all by the informants from the Norwegian measure points. Only at ten locations in Norway have the informants accepted it, and these are spread quite evenly across the country.

The general picture thus is very clear. Norwegian dialects generally follow the "Swedish pattern", with negation between the infinitival marker and the infinitive. The "Danish pattern" is rejected in most cases, and there is only one place in which the "Danish pattern" gets a higher score than the Swedish one. The claim by Faarlund et al. (1997: 997) that the "Danish pattern" is more natural for many Norwegian language users is therefore severely weakened by the data in the NSD.

A note on the data from Denmark is in order. First, at one location in the North of Jutland ("Vendsyssel") both the split and the unsplit patterns are accepted. Pedersen (2017: 44ff) has looked more closely at these data, and she finds that four informants at this location accepts the test sentence. Furthermore, she also points out that there is at least one informant at all of the other locations in Jutland that accepts the sentence, and at the measure point Eastern Jutland ("Østjylland") three informants do so. Second, Pedersen (op. cit.) shows that the existence of the "Swedish pattern" has been mentioned and documented in the dialectological literature also for the insular parts of Denmark. In other words, even in Danish dialects, the "Danish pattern" does not seem to be as obligatory as one might think.
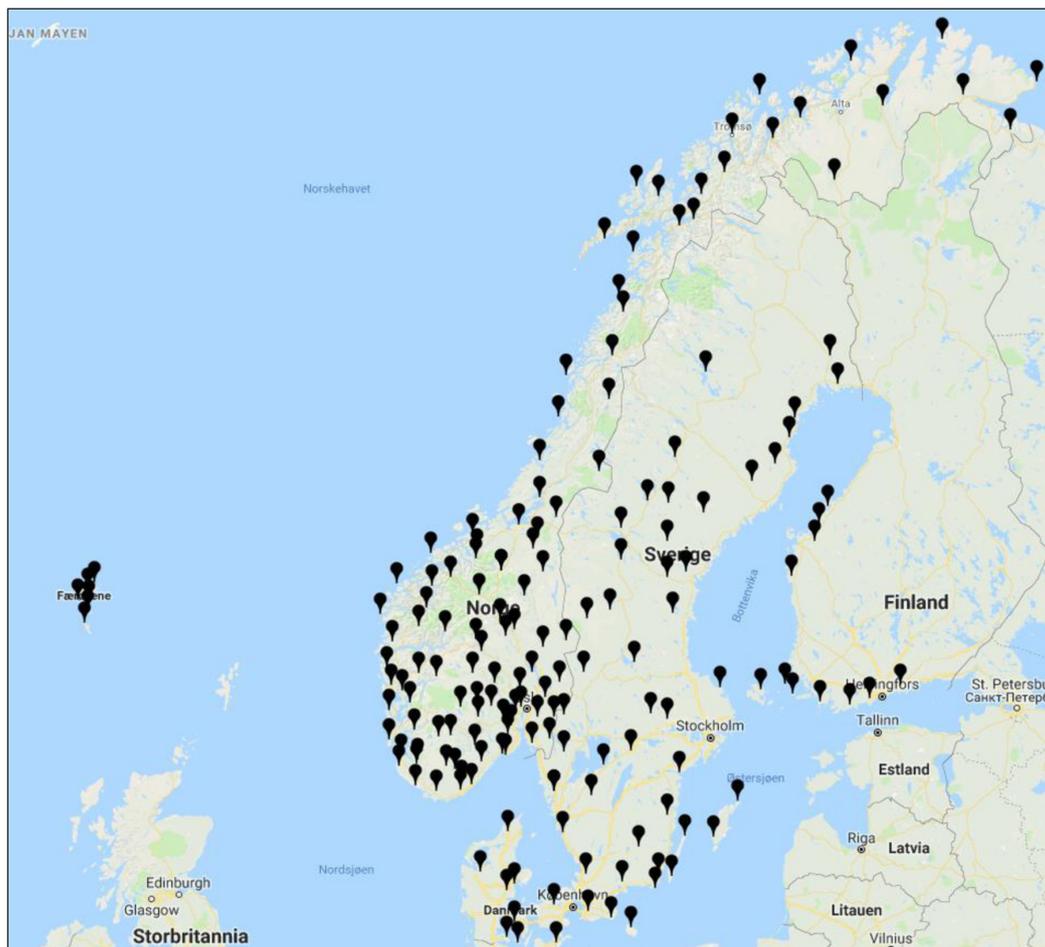
Map 3: The "Icelandic pattern": C – Vinf – neg.

On the basis of these considerations, it is worth investigating to what extent data from the corpus of spontaneous speech corroborate the results from the database. Defining a search for the "Danish pattern" is, however, not trivial since a negation preceding the infinitival marker may belong to the matrix predicate rather than to the infinitival clause: the example in (4) is ambiguous between a high and low attachment for the negation. This is not the case with the "Swedish pattern", see (5), repeated from (2).

(4)   Kjell prøver **ikke å komme** for sent på jobb.
        Kjell tries   not   to come   too late at work
        i. 'Kjell does not try to come too late for work.'
        ii. 'Kjell tries not to come too late for work.'

(5)   Kjell hadde lenge prøvd **å ikke komme** for sent på jobb.
        Kjell had   long tried to not   come   too late at work
        'Kjell had for a long time tried not to come too late for work.'

A search for the string [negation] + [infinitival marker], tailored to the "Danish pattern" is formulated as in Figure 3. The search specifies that the first word should not be a verb in the past or present tense (to try to avoid the ambiguous pattern exemplified in (4)), followed by the negation word and the infinitival marker.

This search, when limited to just the Norwegian part of the corpus, gives 13 relevant results. All of them turn out to involve the idiomatic phrase *for ikke å* 'in order not to/to

Figure 3: Search string for unsplit infinitives (the "Danish pattern") in the Nordic Syntax Database.

not even', i.e. only with this preposition and only in this meaning. An example is given in (6).

(6)   for **itt å snakk** om   syklinga   her
         for not to speak   about cycling.DEF her
         'to not even talk about the cycling around here' (bjugn_19)

Other than these 13, there are zero hits for unsplit infinitives (the "Danish pattern") in Norwegian dialects.

The search for the split "Swedish pattern", on the other hand, gives 60 hits from 41 different locations across all of Norway. All of the hits are relevant, and an example is given in (7).

(7)   Han sku   lær   oss **å ikkje ver** redd   uveret.
         he   should teach us to not   be afraid storm.DEF
         'He was going to teach us not to be afraid of the storm.' (stamsund_03gm)

The NDC has a map function which allows the user to generate a map to show which locations the hits are from, and Map 4 shows the distribution of the 60 split infinitives found in Norwegian dialects.

There are some important differences between the maps generated by the NSD, Maps 1–3, and the NDC, Map 4. The former only generates maps on the basis of sentence evaluations, while the latter generates hits from spontaneously produced speech. This means that while the locations in Map 4 show where the split infinitive (the "Swedish pattern") has been attested, we cannot draw the conclusion that the missing points on the map are places where it could not occur. We already know that the unsplit infinitive (the "Danish pattern") has only been attested for a sub-construction of all logically possible syntactic possibilities, and therefore that it is not one that would cover the unmarked places on the map. When a corpus does not attest a certain usage, it may be because the informants simply did not use that construction during the recorded conversation session.

The maps illustrate clearly why language research benefits from both a database and a corpus kind of infrastructure. The database data from the NSD has many more hits than the corpus data from NDC. However, a database based on evaluations of sentences can only answer questions that have been asked, and databases will therefore contain only a subset of variations of a construction. A corpus, on the other hand, where informants speak freely, will exemplify many different constructions, even ones that the researcher has not thought of beforehand. At the same time, it is to some extent arbitrary what constructions conversation partners use in recordings. This is exemplified in Map 4, which has far fewer locations than Maps 1–3.

Map 4: The 60 hits of split infinitives found in the NDC spread across 41 locations in all of Norway.

What this investigation shows, is that when the researcher is fortunate enough to have a database of evaluations for a particular structure, there will be hits for all the locations investigated. A corpus does not necessarily cover all locations if a construction is not among the most common ones. Still, the corpus can be used to check whether informants in the database have given answers that are indeed compatible with their own language production. Our investigation in this particular case shows that this is indeed the case. The placement of the adverb with respect to the infinitival marker in Norwegian turns out to follow the Swedish pattern, illustrated in Maps 1–3, and the production data from the corpus show the same, even more convincingly, in Map 4. And the whole infrastructure together shows that the claim made in the Norwegian reference grammar (Faarlund et al. 1997) is not supported by our empirical investigations.

We now turn to look at a different and far more complex issue, namely the lack of Verb Second in matrix *wh*-questions in Norwegian dialects.

4 Case study 2: Non-V2 in matrix *wh*-questions across Norwegian dialects

4.1 Previous knowledge

The traditional portrayal of Norwegian and the North Germanic languages in general is that they are well-behaved Verb Second languages, i.e. with the finite verb in a fronted Wackernagel position in matrix clauses, always occurring before the subject whenever a non-subject introduces the clause. This is exemplified by the declarative clause in (8), and the comparison with the idiomatic English translation serves to make the point.

(8)   a.   I morgon skal studentane   ta   eksamen.
              tomorrow will students.DEF take exam
              'Tomorrow the students will take the exam.'

        b.   *I morgon studentane skal ta eksamen

In Lohndal et al. (forthcoming) an overview of exceptions to the V2 requirement in Norwegian is given. The exceptions are more than one tends to acknowledge in general descriptions of the language. In short, the main message in the paper is that Verb Second in Norwegian cannot be an effect of a single macro-parameter, but is rather due to several minor rules which conspire to give the impression that V2 is almost omnipresent (cf. also Weerman 1989).

  One phenomenon which all the same has received considerable attention is the lack of V2 in matrix *wh*-questions, see e.g. Iversen (1918), Elstad (1982), Nordgård (1985), Åfarli (1986), Taraldsen (1986), Lie (1992), Fiva (1996), Nilsen (1996), Westergaard (2003; 2005; 2009a; b; 2017), Westergaard & Vangsnes (2005), Vangsnes (2005), Rognes (2011), Reite (2011), Vangsnes & Westergaard (2014), Westendorp (2017; 2018), and Westergaard, Vangsnes & Lohndal (2017). Iversen (1918: 37) is an early source commenting on the phenomenon. In his study of the syntax of the city dialect of Tromsø, he notes that the interrogative pronouns *kæm* 'who' and *ka* 'what' are associated with "a quaint word order" in that direct questions (i.e. matrix ones) show the same word order as indirect questions (i.e. embedded ones) with these *wh*-pronouns.

(9)   a.   *Tromsø dialect* (Iversen 1918)
              Kæm du   træfte?
              who you met
              'Who did you meet?'

        b.   Ka   du   hete?
              what you are-called
              'What is your name?'

The 'quaintness' of these structures is, of course, that the standard written language would require V2 in the corresponding cases, exemplified here by standard Nynorsk Norwegian examples.

(10)   a.   Kven trefte du?/*Kven du trefte?
        b.   Kva heiter du?/*Kva du heiter?

At the outset of the ScanDiaSyn network it was already well established that this lack of Verb Second is widespread across Norwegian dialects. The only area for which no instance of the phenomenon had been reported seemed to be Central Eastern Norway around the capital Oslo and the adjacent coastal areas to the south. It had also been established that there is considerable variation across the dialects. The first four points below are basic and shared characteristics of the phenomenon.

a. The finite verb is in a position to the right of sentence adverbs (i.e. has not moved to C).
b. In subject *wh*-questions the complementiser *som* appears in second position (and the finite verb in a position to the right of sentence adverbs).
c. All dialects that allow non-V2 in *wh*-questions, also allow V2: the choice of ±V2 appears to be governed by information structure in that V2 is preferred when the subject is given information and non-V2 when the subject is new information.
d. Dialects that allow non-V2 in matrix *wh*-questions also allow the complementiser *som* before the trace position of an extracted *wh*-subject, hence seemingly violating a COMP trace effect.

Points a. and b. entail that there is a strong parallelism between matrix *wh*-questions with non-V2 and embedded *wh*-questions: in embedded *wh*-questions the finite verb also appears to the right of a sentence adverb and the appearance of the complementiser *som* is obligatory in subject questions.

(11)   Eg lurer   på …
        I   wonder on
        a.   … kven du   eigentleg trefte./… *kven trefte du   eigentleg.
                  who you   actually   met   who met   you actually
              'I wonder who you actually met.'

        b.   … kven som eigentleg bestemmer./… *kven bestemmer eigentleg.
                  who SOM actually decides   who decides   actually
              'I wonder who actually decides.'

The complementiser *som* has no one-to-one English equivalent: it also appears in relative clauses and clefts (corresponding to *that*) and in small clauses and comparatives (corresponding to *as*) (see Stroh-Wollin 2002; Vangsnes 2004: 22f for details).

   The insight in point c. is due to Westergaard (2003; 2005), who studied V2 vs. non-V2 quantitatively in a corpus of the Tromsø dialect. Point d. can be credited to Nordgård (1985) and is illustrated by the example in (12).

(12)   Kem du   trur   **som** __ e i baren?
        who you think SOM   is in bar.DEF
        'Who do you think is in the bar?'

The following four points concern issues of *variation* across the dialects that allow non-V2 in matrix *wh*-questions in the first place.

e. Many dialects allow non-V2 only with the short, monosyllabic items *kven* 'who', *kva* 'what' and *kor* 'where'.
f. Some dialects also allow complex *wh*-items and *wh*-phrases with non-V2.
g. Some dialects allow non-V2 only in subject *wh*-questions. i.e. with *som* in second position.
h. Some dialects allow complex *wh*-subjects but only simple non-subject *wh*-constituents, hence make a ±subject distinction with regard to the complexity of the *wh*-constituent.

Point e. was noted for the Tromsø city dialect already by Iversen (1918: 37), and stated more broadly as a trait of Northern Norwegian dialects by Elstad (1982). Point f. was shown by Nordgård (1985) and Åfarli (1986) for Northwestern Norwegian dialects. Point

g. is due to Lie (1992: 66) who noted that some Western Norwegian dialects seem to only allow non-V2 with *wh*-subjects and insertion of *som*. Point h. can be attributed to Fiva (1995) who reported that in a survey of the Tromsø dialect many informants found complex *wh*-subjects acceptable followed by *som* but would still only accept short *wh*-constituents in non-subject questions with non-V2.

4.2 The questionnaire data

These various bits of knowledge informed the development of test sentences for the questionnaire to be used in the project. Since the questionnaire was to probe a long list of different constructions, some corners inevitably had to be cut. One of them was to check if the informants allowed both V2 and non-V2 in matrix *wh*-questions, and along with that, to what extent the choice was governed by information structure. In the Norwegian version of the questionnaire, which at the outset had about 130 sentences, we ended up with the following four sentences regarding non-V2 in matrix *wh*-questions.

(13)   **Ka**   du   hete?
        what you are.called
        'What are you called?'

(14)   **Ka   tid**   du   gikk ut   av ungdomsskolen?
        what time you went out of secondary.school.DEF
        'When did you graduate from secondary school?'

(15)   **Kem som** sæll fiskeutstyr   her i   bygda?
        who SOM sells fishing.gear here in village.DEF
        'Who's selling fishing gear in this village?'

(16)   **Kor mange** eleva som går på den her   skolen?
        how many   pupils SOM go on the here school.DEF
        'How many pupils go to this school?'

(13) has a short *wh*-predicative, (14) has a complex *wh*-adverb, (15) has a short *wh*-subject and (16) has a complex *wh*-subject, and in sum these sentences should serve to probe the issue of short versus complex *wh*-constituents, the ±subject condition and, furthermore, whether subject and non-subject questions differ with respect to allowing complex *wh*-constituents with non-V2.

   However, the sentences would not serve to detect whether there would be a difference between arguments and non-arguments, or if different *wh*-adverbs would give different results. After the data collection had begun, an additional test sentence with a different *wh*-adverb was added to the questionnaire in order to possibly obtain more information about other relevant factors.

(17)   **Koffer**   han va   så sur,   egentli?
        what-for he   was so grumpy actually
        'Why was he so grumpy?' ('What was the actual reason for his grumpiness?')

The database results for the four initial non-V2 *wh*-questions have recently been presented in Westergaard et al. (2012; 2017). The results by and large confirm the findings reported in the earlier literature, including statements e. to h. above, but now on the basis of a much more comprehensive and systematic data collection (four informants from 107

locations spread out across Norway). Four types of dialects allowing non-V2 emerge from
the NSD data:

A. Complex *wh*-constituents are allowed with non-V2 in both subject and non-subject
   questions (i.e. all four test sentences)
B. Only short *wh*-constituents are allowed with non-V2 in subject and non-subject
   questions alike (i.e. (13) and (15)).
C. Complex *wh*-subjects are allowed with non-V2, but only short *wh*-constituents are
   allowed in non-subject questions (i.e. all but (14) are accepted).
D. Only *wh*-subjects are allowed with non-V2 (i.e. (15) and (16) are accepted).

The distribution of these dialect types is given in Map 5 which is also published in
Westergaard et al. (2017: 26).

In the area labeled with '*' non-V2 is dismissed by the informants consulted, and for
the area labeled with '?' no clear pattern emerges as far as the authors can see (see also
Westendorp 2017; 2018, for an assessment of the data from this area).

4.3 The spontaneous speech data

Vangsnes & Westergaard (2014) present data regarding the phenomenon based on searches
in the Nordic Dialect Corpus (NDC). The searches were optimised to target matrix *wh*-
questions, and a gross number of 2273 hits were trimmed down to 1332 relevant ones
after fragments, exclamatives and embedded clauses had been sorted out.

The distribution of the remaining relevant examples over different *wh*-items and phrases
were as given in Table 1 (Vangsnes & Westergaard 2014: 142). The figures show three
things in particular. First, for the short *wh*-items 'what', 'who' and 'where', there is a more
or less even distribution between V2 and non-V2. Second, the *when* questions come in
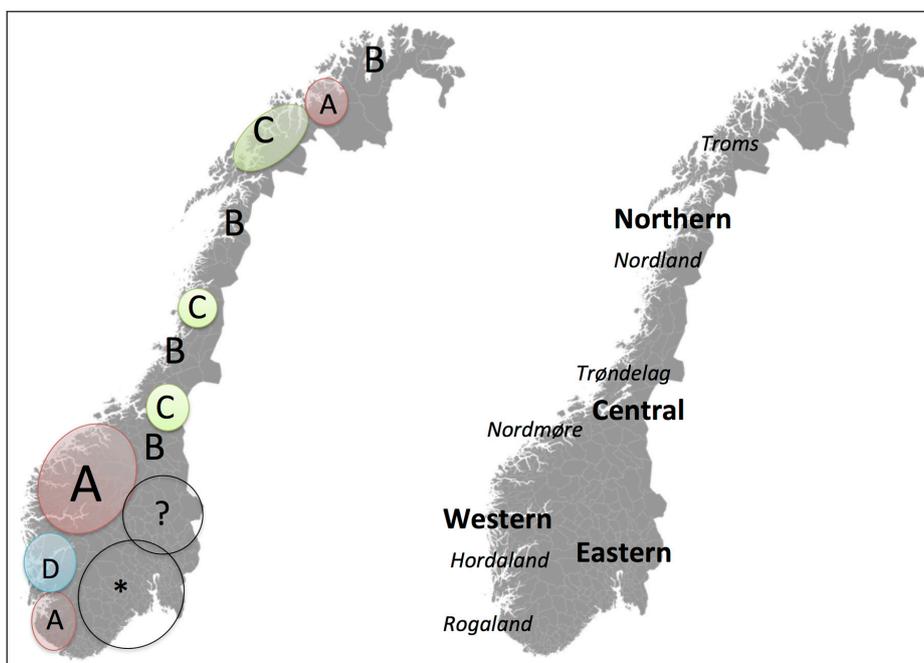
Map 5: Map taken from Westergaard et al. (2017). The distribution of different types of grammars
for matrix *wh*-questions across Norway – A = allows both simple and complex *wh*-constituents
with non-V2, B = allows only simple *wh*-constituents with non-V2, C = allows both simple and
complex *wh*-subjects but only simple non-subjects with non-V2, D = allows only subject *wh*-
questions with non-V2, ? = unclear pattern(s), * = no non-V2 is allowed.

Table 1: V2 and non-V2 in matrix *wh*-questions in the Norwegian part of the Nordic Dialect Corpus
by type of *wh*-constituent.

	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[1]	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

a middle position with about every fourth instance having non-V2. Third, for the other adverbial *wh*-items and the *wh*-phrases, very few appear with non-V2.

Concerning the first observation, the distribution of V2 versus non-V2 varies across different parts of the country. Vangsnes & Westergaard (2014: 143) show that for the three short *wh*-items 'what', 'who', and 'where' non-V2 is far more frequent than V2 in Northern Norwegian, and that the picture gradually shifts to the opposite when one moves through Central Norwegian and Western Norwegian to Eastern Norwegian. The figures they provide can be summarised as in Table 2.[2]

On this issue the corpus data complement the questionnaire data in NSD as the latter only provide information about the acceptance of non-V2: the informants were never asked to judge matrix *wh*-questions with V2. Although we therefore do not know the relative preference of V2 versus non-V2, the production data from the corpus suggest that non-V2 is the unmarked option in Northern Norwegian dialects for the three short items 'what', 'who' and 'where' and that there is a gradual shift in preference as we move south.

Concerning the second observation, there exist both short and long variants for 'when' in Norwegian dialects. Some dialects use the monosyllabic variant *når*, which is the one used in the standard varieties, but the complex *hva tid*, literally 'what time' is widespread, and the variants *når tid* 'when time' and *hvor tid* 'where time' are also found. We should therefore consider what variants are used in the 21 instances of 'when'-questions with non-V2. In this case we are also in the fortunate situation that one of the four *wh*-questions probing non-V2 in NSD is a 'when'-clause (see above), thus allowing us to compare production and judgments directly at the level of the individual.

In Table 3 each informant is listed with whichever 'when'-variant they used and how they judged the NSD 'when'-question (#33 in the questionnaire). As we see, there are five instances with the short, monosyllabic item *når* in non-V2 matrix *wh*-questions, produced by five different informants from three different locations in Central Norway. The informants' judgments of the NSD test sentence vary, but crucially the complex variant *hva tid* (adjusted for local pronunciation) was used during data collection also in this area, and the lesser acceptance of the test sentence may be due to the fact that the *wh*-expression used in the test is not the short variant the informants spontaneously use themselves.

---

[1] Vangsnes & Westergaard (2014) report the number 22 for *when*-clauses with non-V2, but on closer examination it turns out that one of the hits were counted twice, and the number is adjusted accordingly here.

[2] Vangsnes & Westergaard (op. cit.) separates out the district Agder, but we include it here in Western Norway to get the four main Norwegian dialect areas described in Mæhlum and Røyneland (2012).

Table 2: The balance between V2 and non-V2 in matrix *wh*-questions with short *wh*-constituents in the Norwegian part of the Nordic Dialect Corpus.

	[what/who/ where] + V2		[what/who/ where] + non-V2	
	n	%	n	%
Northern Norwegian	81	21.3%	299	78.7%
Central Norwegian	50	43.1%	66	56.9%
Western Norwegian	124	56.6%	95	43.4%
Eastern Norwegian	147	79%	39	21%
Whole country	402	44.6%	499	55.4%

Table 3: Non-V2 'when'-question in the Nordic Syntax Database.

Region	Informant code	used by informant	NSD #33 score	n
Central N (Trøndelag, Nordmøre)	inderoey_01um (young male)	*når …*	3	1
	inderoey_02uk (young female)	*når …*	4	1
	oppdal_10 (young female)	*når …*	1	1
	surnadal_29 (young female)	*når …*	2	1
	surnadal_28 (young female)	*når …*	2	1
Northwestern N (Sunnmøre, Nordfjord)	heroeyMR_03gm (older male)	*ka ti …*	5	4
	stryn_01um (young male)	*ka ti …*	5	4
	stryn_02uk (young female)	*ka ti …*	5	6
Southwestern N (Hordaland, Rogaland)	hjelmeland_01um (young male)	*ka ti …*	5	1
	bergen_02uk (young female)	*korr ti …*	1	1

There are, furthermore, 14 examples produced by three informants from two locations in Northwestern Norway, all of whom give the NSD test sentence the highest score. This is an area known for allowing complex *wh*-items with non-V2, and also here the production and judgment data are in harmony. The two remaining examples both involve complex *wh*-expressions. They are uttered by two informants from two places in Southwestern Norway: Hjelmeland and Bergen. The Hjelmeland informant gives the test sentence a high score, whereas the Bergen informant gives it a low score.

Of the 21 'when'-clauses there is therefore only one case where there is a clear incompatibility between production and judgment. The seemingly intermediate position of 'when'-clauses is thus partly due to the fact that some of them involve a short, monosyllabic variant of the *wh*-expression and partly to the fact that only three informants produced most of the non-V2 cases (14 of 21).

Also for the manner *how* questions with non-V2 found in the corpus the simple∼complex issue plays a role. The form of manner 'how' varies to a considerable extent across Norwegian dialects (see Vangsnes 2008), and Vangsnes & Westergaard (2014: 145) report that eight of the nine examples of non-V2 involves monosyllabic variants (*korr, koss, høss*). Only the ninth example has a disyllabic variant (*kelles*).

In other words, the number of non-V2 questions with complex *wh*-expressions is even lower than it seems at first sight in Table 3. The single 'why'-clause involves a disyllabic *wh*-item (as is always the case in Norwegian dialects),[3] and if we put together the 'when'

---

[3] The word for 'why' in Norwegian dialects is always a complex item literally corresponding to 'what-for' or 'where-for'. The single 'why'-example with non-V2 in the corpus has the variant *koffor.*

(21), '(manner) how' (8), 'why' (1), and *wh*-phrases (9) – which amount to 39 – and subtract
the ones with monosyllabic *wh*-expressions (5 *when* and 7 *how*), we are left with 27 non-V2
matrix questions with a complex *wh*-expression out of a total of 539, in other words 5%.

   Map 6 indicates the locations of all of the 27 examples, including the six mismatches
(see Vangsnes & Westergaard 2014: 145ff, for details).

   19 of the 27 examples are from four locations in Northwestern Norway, indicated by the
light blue markers. Three of the cases are from three locations in Southwestern Norway,
indicated by purple markers. Both of these areas are roughly the ones indicated by the
letter A in Map 5 above, hence where the NSD data suggests that informants by and large
accept complex *wh*-phrases in matrix non-V2 questions.

   The yellow icon marks the single example from a location in the northwest corner of
the Eastern Norwegian dialect area, more specifically from the place Lom, which by the

Map 6: Locations at which the Nordic Dialect Corpus provides examples of non-V2 with complex
   wh-constituents.

NSD data is part of the northwestern A-area: both of the test sentences with a complex *wh*-phrase (*wh*-subject and *when*) receive a high score at this location.

The remaining three cases run counter to the data in NSD. Two of them are uttered by informants at two locations in Central Norway, Oppdal and Røros, indicated by green markers, and at both locations the relevant test sentences receive a low score both in general and by the two specific individuals who uttered the corpus sentences in particular. The same holds for the final example from Bergen in Western Norway, indicated by a dark blue marker. Accordingly, these three examples constitute noise in the data that it would be worth following up in future studies of the topic.

Despite this slight discrepancy (3 out of 27 cases) and despite the low total number of complex *wh*-questions with non-V2, the overall picture we are left with when scrutinising the corpus data in the Nordic Dialect Corpus versus the judgment data in the Nordic Syntax Database is that there is a very good match between the two sources. Attempts at analysing these data from a more formal, generative perspective can be found in Westergaard et al. (2017). (See also Rognes 2011, for a study focusing on one particular dialect area; and Vangsnes 2005; Westergaard & Vangsnes 2005; for slightly older theoretical approaches.)

4.4 Subject wh-extraction and Nordgård's Generalisation

In the background section above we noted that Nordgård (1985) established a correlation between dialects that allow non-V2 in matrix *wh*-questions and dialects that allow the insertion of the complementiser *som* before the trace of an extracted *wh*-subject. The questionnaire based data in the Nordic Syntax Database offers information on this issue.

As detailed in Bentzen (2014), eight test sentences for *wh*-extraction were also included in the questionnaire, probing both subject and object extraction, with and without the presence of either of the complementisers *som* and *at* 'that' and also with a resumptive subject pronoun. The following three sentences tap into the Nordgård's generalisation (1985) and more broadly the so-called that-trace effect or COMP trace effect (see Pesetsky 2016 for an overview).

(18) a. Hvem tror du har gjort det?
     who think you has done it

   b. Hvem tror du som har gjort det?
     who think you SOM has done it

   c. Hvem tror du at har gjort det?
     who think you that has done it
     All: 'Who do you think has done it?'

The following three maps (Map 7) details the acceptance of the three sentences in Central, Western and Eastern Norway, with white markers indicating high average scores, grey markers medium average scores, and black markers low average scores (cf. section 2).

Map 7a (the leftmost one) clearly indicates that the sentence with no overt complementiser is accepted as good by everyone. Furthermore, the map in the middle shows that *som*-insertion mostly gets a high or medium score in Western and Central Norway but is largely rejected in Eastern Norway. The sentence with *at*-insertion is in contrast only fully accepted at some measure points in Eastern Norway with some medium scores further to the North in Central Norway.

The data visualised here partly support Nordgård's (1985) generalisation insofar that sentence (18b) is only accepted in those parts of the country where non-V2 is accepted.
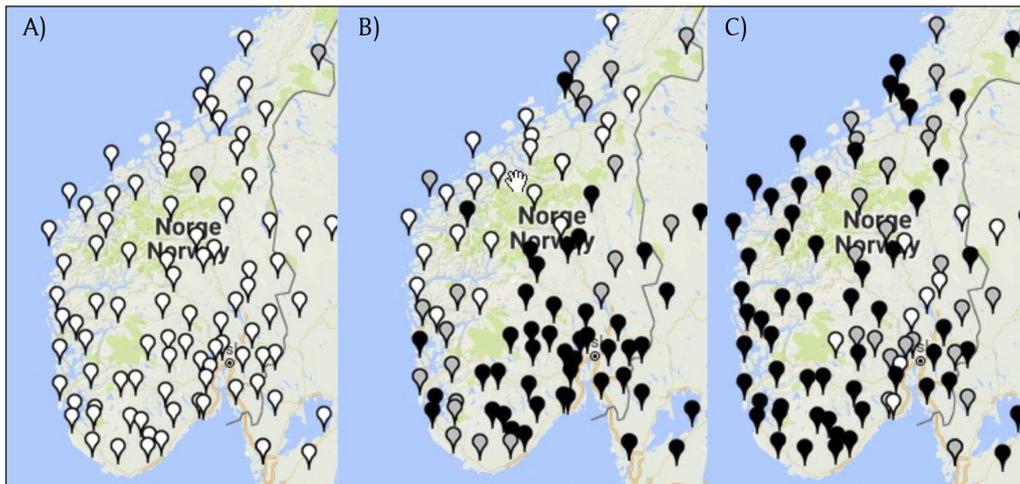
Map 7: Average scores for (long) extraction of a *wh*-subject with: **a)** no complementiser in the embedded clause (18a); **b)** presence of the complementiser *som* in the embedded clause (18b); **c)** presence of the complementiser *at* in the embedded clause (18c).

At the same time, it is also clear that far from all speakers who allow non-V2 allow *som*-insertion with extraction of a *wh*-subject. Still, the preference for no complementiser before a subject trace position over versions with an overt complementiser is a quite well-known fact from previous studies of Germanic languages, and it is also documented for object extraction (see Cowart 1997; Hawkins 2004; Bentzen 2014; Schippers 2017). On the presumption that the informants in the NSD survey were able to contrast the cases with and without complementiser when consulted, the overall lower acceptance for the COMP trace sentences should therefore come as no great surprise.

Furthermore, although the complementarity between *som*-insertion and *at*-insertion is not perfect given the many medium score locations in Central Norway in particular, Map 8 shows that when we just compare measure points with a high score, the complementarity is quite clear: the grey dots mark locations with a high score for *som*-insertion and the blue ones a high score for *at*-insertion.

This map also shows that *som*-insertion is widely accepted in Northern Norway and that *at*-insertion is widely accepted in Finland Swedish.

4.5 Discussion

This exposition of how the issue of non-V2 in Norwegian matrix *wh*-questions was researched in the ScanDiaSyn project should have revealed some significant achievements and also some limitations. In many ways the data confirm what had already been established if one pieces together information from various sources in the existing literature, but they now establish this in a much more systematic and complete way. The questionnaire and corpus data furthermore also by and large confirm each other and thus strengthen the empirical basis.

One very clear limitation with the questionnaire data is that the number of test sentences is low. Relevant additional variables may not have been detected because of this. In particular, given the scarcity of non-V2 questions with complex *wh*-constituents in the corpus compared to the abundance of non-V2 with short *wh*-constituents, it would have been desirable to test out a broader range of complex *wh*-constituents so as to compare different kinds of *wh*-adverbials or *wh*-adverbials versus complex *wh*-arguments.

Furthermore, as pointed out above, there are quite clearly some cases of discrepancy between production and judgment data. A positive angle to that is that they identify areas and/or locations that need to be studied more carefully, and the northern part of Eastern
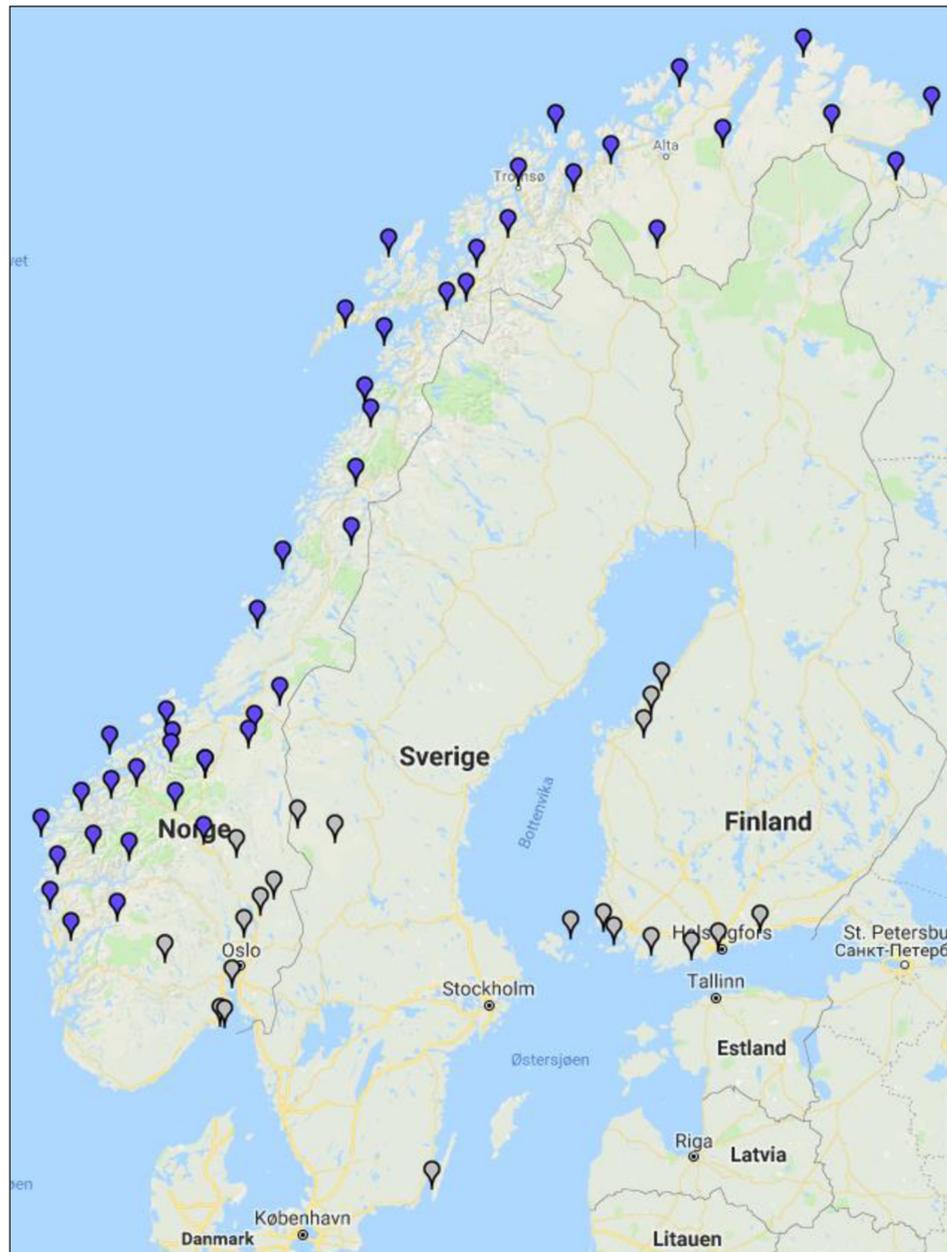
Map 8: Locations with high average scores for presence of complementiser *som* under subject wh-extraction (blue markers) versus locations with high average scores for presence of complementiser *at* under subject *wh*-extraction (grey markers).

Norway, marked as "?" in Map 5, in particular stands out as an area with an unclear pattern.

In this paper we have done little to put the data to statistical scrutiny. With data from over 100 locations and almost 400 individuals on a number of test sentences, the possibilities for doing so is certainly there, and one study which has approached the phenomenon in a systematic way by employing statistical methods is Westendorp (2017; 2018). Without going into details, she argues that the statistics do not support all of the diachronic speculations put forth in Westergaard et al. (2017) as to how Norwegian dialects – as the only ones across North Germanic – have developed this particular violation of Verb Second. Westendorp's statistical findings do however support the general idea that the phenomenon has started with short *wh*-constituents and later spread to questions with complex *wh*-expressions.

5 Conclusion

In this paper we hope to have demonstrated the assets of having access to both a database of syntactic judgments and a searchable corpus of free speech when researching topics in dialect syntax. In the case of the two Nordic dialect infrastructures, the Nordic Syntax Database and the Nordic Dialect Corpus, the data have been collected from a well-distributed set of locations, and crucially both kinds of data have been collected from largely the same set of informants. The well-known shortcoming of a corpus that particular constructions or variables may be scarce is counterbalanced by the way a questionnaire can ensure information from all participants on the selected constructions/variables. On the other hand, the closed nature of a questionnaire, where topics must be decided beforehand can be contrasted to the more dynamic nature of a corpus in which data one had not thought of in advance may occur. Moreover, data from the two sources (judgment versus production) for the same individuals may confirm each other, but they may also be contradictory. The latter kind of situation may help to identify issues that need to be further investigated.

For the concrete topics that we have chosen to base our exposition on, we have seen that the data from dialect infrastructures may represent both a correction of received knowledge and a strengthened confirmation of existing knowledge. In the case of split infinitives both the judgment data and the corpus data clearly show that in spoken Norwegian placing a sentence adverb between the infinitival marker and the infinitive (split version) is much more preferred and used than placing it before the infinitival marker (unsplit version). This runs counter to the received knowledge that Norwegian allows both structures and in fact prefers the unsplit version: the data show that in Norwegian dialects the split version is both preferred and most commonly used, placing them more in line with Swedish than with Danish.

In the case of matrix *wh*-questions with non-V2 word order in Norwegian, the judgment data serve to confirm the rather complex pattern of variation that can be pieced together on the basis of the existing literature going back to the early 20th century. Furthermore, the corpus data serve to complement the questionnaire data insofar that the abundance of non-V2 *wh*-questions in all dialect regions but Eastern Norwegian confirms that the phenomenon is widespread, and we also see an increase in the use of it as we move northwards through the country. However, the abundance of examples applies just to questions with short *wh*-constituents: only 5% of the non-V2 *wh*-questions in the corpus contain complex *wh*-constituents, and they are also very few compared to V2 *wh*-questions with the same kind of constituents. To some extent this finding squares with the insight that complex non-V2 *wh*-questions are judged acceptable in two rather restricted areas (northwestern dialects and southwestern dialects), and almost all of the cases are indeed produced by speakers from these areas. But even within these restricted areas the complex non-V2 cases are fewer than complex V2 cases, and this entails that weight or complexity plays a role in production even in these dialects.

One clear shortcoming with the questionnaire data regarding the issue of *wh*-questions and V2, is that the number of test sentences were few. This is a direct effect of the topic being part of a general questionnaire that probed a wide range of topics. When administering data collection by questionnaires there is a limit to how much time one can keep the informants' attention and willingness to respond. Since the questionnaire and production data were collected at the same time by fieldworkers visiting the various locations this is a limitation that it is hard to come around unless one sets up ways of getting back to each individual informant at later points in time. In turn, administering a system for that is laborious and resource demanding and was not viable in the case of the ScanDiaSyn project.

In any event, it seems quite clear that the establishing of the Nordic research infrastructure for syntactic variation has moved the field of North Germanic dialect syntax several steps forward, and although others may be in a better position to judge it objectively, we also believe that the output of the research collaboration has served to revitalise the field of Nordic dialectology.

Funding Information

The infrastructure was built with support from national and Nordic funding bodies: The Research Council of Norway, The Swedish Research Council, The Danish Research Council for Culture and Communication, and The Icelandic Research Fund, and the Nordic funding bodies NordForsk and NOS-HS which also provided important funding for network activities.

Competing Interests

The authors have no competing interests to declare.

References

Åfarli, Tor A. 1986. Some syntactic structures in a dialect of Norwegian. *Working Papers in Linguistics* 3. 93–111. Trondheim: University of Trondheim.

Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments.* Thousand Oaks, CA: SAGE Publications.

Elstad, Kåre. 1982. Nordnorske dialektar. In Tove Bull & Kjellaug Jetne (eds.), *Nordnorsk: Språkarv og språkforhold i Nord-Noreg,* 11–100. Oslo: Det Norske Samlaget.

Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk.* Oslo: Universitetsforlaget.

Fiva, Toril. 1996. Spørresetninger i Tromsødialekten. *Nordica Bergensia* 9. 139–155.

Hawkins, John A. 2004. *Efficiency and complexity in grammars.* Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199252695.001.0001

Huddleston, Rodney D. & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language.* Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/9781316423530

Hulthén, Lage. 1947. *Studier i jämförande nunordisk syntax II.* Göteborg: Wettergren & Kerbers Förlag.

Iversen, Ragnvald. 1918. *Syntaksen i Tromsøbymaal.* Kristiania: Bymaalslagets forlag.

Johannessen, Janne Bondi. 2017. Annotations in the Nordic Dialect Corpus. In Nancy Ide & James Pustejovsky (eds.), *Handbook of linguistic annotation,* Chapter 49. Springer. DOI: https://doi.org/10.1007/978-94-024-0881-2_50

Johannessen, Janne Bondi, Joel Priestley, Kristin Hagen, Tor Anders Åfarli & Øystein Alexander Vangsnes. 2009. The Nordic Dialect Corpus – An advanced research tool. In Kristiina Jokinen & Eckhard Bick (eds.), *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009* (NEALT Proceedings Series Volume 4), 73–80. Northern European Association for Language Technology (NEALT). Electronically published at Tartu University Library (Estonia). http://hdl.handle.net/10062/9206.

Johannessen, Janne Bondi, Lars Nygaard, Joel Priestley & Anders Nøklestad. 2008. Glossa: A multilingual, multimodal, configurable user interface. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis & Daniel Tapias (eds.), *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08),* 617–621. Paris: European Language Resources Association (ELRA).

Johannessen, Janne Bondi, Øystein Alexander Vangsnes, Joel Priestley & Kristin Hagen. 2014. A multilingual speech corpus of North-Germanic languages. In Tommaso Raso

& Heliana Mello (eds.), *Spoken corpora and linguistic studies*, 69–83. Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/scl.61.02joh

Lie, Svein. 1992. Ka du sei? *Maal og Minne*, 62–77.

Lindstad, Arne Martinus, Anders Nøklestad, Janne Bondi Johannessen & Øystein Alexander Vangsnes. 2009. The Nordic Dialect Database: Mapping microsyntactic variation in the Scandinavian languages. In Kristiina Jokinen & Eckhard Bick (eds.), *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009* (NEALT Proceedings Series Volume 4), 283–286. Northern European Association for Language Technology (NEALT). Electronically published at Tartu University Library (Estonia). http://hdl.handle.net/10062/9206.

Lohndal, Terje, Marit Westergaard & Øystein A. Vangsnes. Forthcoming. Verb Second in Norwegian: Variation and acquisition. In Rebecca Woods, Sam Wolfe & Theresa Biberauer (eds.), *Rethinking verb second*. Oxford: Oxford University Press.

Nilsen, Hilde. 1996. *Koffør dem sir det?* Tromsø: University of Tromsø: Cand.philol. thesis.

Nordgård, Torbjørn. 1985. *Word order, binding and the empty category principle*. Trondheim: University of Trondheim cand. philol. thesis.

Nordic Atlas of Language Structures Online Journal (NALS) 1. https://journals.uio.no/index. php/NALS.

Nordic Atlas of Language Structures Online Journal (NALS), 2 [old site with thematic structure]. http://www.tekstlab.uio.no/nals#/project_info.

Nordic Dialect Corpus. http://www.tekstlab.uio.no/nota/scandiasyn/.

Nordic Syntax Database. http://www.tekstlab.uio.no/nota/scandiasyn/.

Pedersen, Karen Margrethe. 2017. Syntaktiske oplysninger i dialektordbøger og store nationale ordbøger. In Jan-Ola Östman, et al. (eds.), *Ideologi, identitet, intervention: Nordisk dialektologi 10*. Helsinki: Nordica, University of Helsinki.

Reite, André. 2011. *Spørjing i skedsmokorsmålet: Undersøking og analyse av leddstelling og spørjeord i interrogative hovudsetningar i talemålet på Skedsmokorset*. Trondheim: Norwegian University of Science and Technology (NTNU) MA thesis.

Rognes, Stig. 2011. *V2, V3, V4 (and maybe even more): The syntax of questions in the Rogaland dialects of Norway*. Oslo: University of Oslo MA thesis.

ScanDiaSyn network. http://websim.arkivert.uit.no/scandiasyn/index.html%3fLanguage=en.

Schippers, Ankelien. 2017. On the variability of COMP-trace effects: A processing explanation. *Paper presented at Comparative Germanic Syntax Workshop 32*. Trondheim.

Stroh-Wollin, Ulla. 2002. *Som-satser med och utan som [Som-sentences with and without som]*. Uppsala: Uppsala University doctoral dissertation.

Taraldsen, Knut Tarald. 1986. *Som* and the Binding Theory. In Lars Hellan & Kirsti Koch Christensen (eds.), *Topics in Scandinavian syntax*, 149–184. Dordrecht: Reidel. DOI: https://doi.org/10.1007/978-94-009-4572-2_8

Vangsnes, Øystein A. 2004. On *wh*-questions and V2 across Norwegian dialects: A survey and some speculations. *Working Papers in Scandinavian Syntax* 73. 1–59. Lund: University of Lund.

Vangsnes, Øystein A. 2005. Microparameters for Norwegian *wh*-grammars. *Linguistic Variation Yearbook* 5. 187–226. Amsterdam: John Benjamins.

Vangsnes, Øystein A. 2007a. Scandinavian Dialect Syntax (before and after) 2005. *Nordlyd – Tromsø University Working Papers in Language & Linguistics*, 7–24.

Vangsnes, Øystein A. 2007b. ScanDiaSyn: Prosjektparaplyen Nordisk dialektsyntaks. In T. Arboe (ed.), *Nordisk dialektologi og sociolingvistik*, 54–72. Århus: Peter Skautrup Centeret for Jysk Dialektforskning, Århus Universitet.

Vangsnes, Øystein A. & Marit Westergaard. 2014. *Ka korpuse fortæll?* Om ordstilling i *hv*-spørsmål i norske dialekter. In Janne Bondi Johannessen & Kristin Hagen (eds.), *Språk i Norge og nabolanda. Ny forskning om talespråk*, 133–151. Oslo: Novus.

Westendorp, Maud. 2017. *Development and variation of non-V2 order in Norwegian wh-questions*. Amsterdam: University of Amsterdam MA thesis.

Westendorp, Maud. 2018. New methodologies in the Nordic Syntax Database: Word order variation in Norwegian *wh*-questions. In *Nordic Atlas of Language Structures Online Journal*, 3(1). 1–18.

Westergaard, Marit. 2003. Word order in *wh*-questions in a North Norwegian dialect: Some evidence from an acquisition study. *Nordic Journal of Linguistics* 26. 81–109. DOI: https://doi.org/10.1017/S0332586503001021

Westergaard, Marit. 2005. Optional word order in *wh*-questions in two Norwegian dialects: A diachronic analysis of synchronic variation. *Nordic Journal of Linguistics* 28. 269–296. DOI: https://doi.org/10.1017/S0332586505001459

Westergaard, Marit. 2009a. Microvariation as diachrony: A view from acquisition. *Journal of Comparative Germanic Linguistics* 12. 49–79. DOI: https://doi.org/10.1007/s10828-009-9025-9

Westergaard, Marit. 2009b. *The acquisition of word order: Micro-cues, information structure, and economy*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.145

Westergaard, Marit. 2017. Word order and verb movement in Norwegian *wh*-questions: A comparison of production and judgment data. In Bettelou Los & Pieter de Haan (eds.), *Word order change in acquisition and contact: Essays in honour of Ans van Kemenade*, 35–56. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.243.03wes

Westergaard, Marit & Øystein A. Vangsnes. 2005. *Wh*-questions, V2, and the left periphery of three Norwegian dialects. *Journal of Comparative Germanic Linguistics* 8. 119–160. DOI: https://doi.org/10.1007/s10828-004-0292-1

Westergaard, Marit, Øystein A. Vangsnes & Terje Lohndal. 2017. Variation and change in Norwegian *wh*-questions: The role of the complementizer *som*. *Linguistic variation* 17(1). 8–43.