

Variation and change in Norwegian *wh*-questions: The role of the complementizer *som**

Marit Westergaard¹, Øystein A. Vangsnes¹ & Terje Lohndal²

1. UiT The Arctic University of Norway, 2. NTNU - Norwegian University of Science and Technology

Abstract: In this paper, we consider variation in Verb Second (V2) word order in *wh*-questions across Norwegian dialects by investigating data from the Nordic Syntax Database (NSD), which consists of acceptability judgments collected at more than 100 locations in Norway. We trace the geographical distribution of the two main variables: phrasal vs. monosyllabic *wh*-elements (the latter argued to be heads) and subject vs. non-subject questions. In subject questions, non-V2 is realized by inserting the complementizer *som* in second position instead of the verb. We also discuss the connection between non-V2 and the possibility of inserting the complementizer *som* under extraction of a *wh*-subject from an embedded clause, i.e. in *that*-trace contexts. Based on synchronic data, we propose a diachronic account of the geographical distribution and argue that the development from V2 to non-V2 has started in subject questions, thus allowing us to relate the loss of the V2 requirement to changes in the properties of the complementizer *som*.

* This paper is a revised and considerably extended version of Westergaard, Vangsnes & Lohndal (2012). We are grateful to audiences in a number of venues between June 2010 and August 2011 for their helpful comments and to two anonymous reviewers for constructive feedback.

b. *Hvem hadde kjøpt sko til to tusen?*
who had bought shoes to two thousand
'Who had bought shoes for two thousand?'

Note that the word order in (1a) is *wh-S-V*, while non-V2 in the subject question in (1b) is realized by inserting the relative complementizer *som* (in the current dialect pronounced [so], rendered here as *så*) in second position (*wh-som-V*). In this paper, we investigate the variation and distribution of non-V2 word order across a number of Norwegian dialects, using judgment data from the Nordic Syntax Database (Lindstad, Nøklestad, Johannessen & Vangsnes 2009, Johannessen & Vangsnes 2010, Johannessen, Vangsnes, Laake, Lindstad & Åfarli 2010). Based on this contemporary variation, we sketch a possible historical development from V2 to non-V2, arguing that the process starts in subject questions and that the behavior of the element *som* plays a crucial role. In other words, we develop a diachronic account on the basis of synchronic variation observed in contemporary dialects of Norwegian.

The paper is organized as follows. In the next section, we outline some background on non-V2 word order in Norwegian dialects based on existing literature. We focus on the factors responsible for the variation, such as the function and category of the *wh*-element as well as on a correlation between non-V2 and the type of complementizer appearing in embedded clauses under extraction. In section 3, we provide the results of our investigation as a number of maps, showing the geographical distribution of the variation across the country. In section 4, we sketch a diachronic account of this synchronic geographical distribution, arguing that the non-V2 phenomenon has spread in small steps. Section 5 is a brief summary and conclusion.

2. Background and previous research

The non-V2 phenomenon in *wh*-questions in Norwegian dialects has been described and discussed to a notable extent in both the dialectological and theoretical literature, and different accounts have been proposed of how this deviation from generalized matrix V2 syntax should be understood; see e.g. Iversen (1918), Elstad (1982), Åfarli (1986), Taraldsen (1986), Lie (1992), Fiva (1996), Nilsen (1996), Westergaard (2003, 2005, 2009a, 2009b, forthcoming), Westergaard & Vangsnes (2005), Vangsnes (2005), Rognes (2011), and Vangsnes & Westergaard (2014).

Let us first point out that non-V2 in these cases entails lack of verb movement to the Left Periphery, which may be shown by considering the position of the verb relative to sentence adverbs: If a sentence adverb is present, such as for instance the negative element *ikkje* ‘not’, the finite verb must appear to the right of it. This is illustrated in (3a) for a non-subject question and (3b) for a subject question involving *som* in second position.

- (3) a. *Kem du <*skal> ikkje <skal> møte i baren?*
who you shall not shall meet in bar.DEF
‘Who will you not meet in the bar?’
- b. *Kem som <*er> ikkje <er> i baren no?*
who SOM is not is in bar.DEF now
‘Who’s not in the bar now?’

Dialects allowing non-V2 in *wh*-questions have normally been assumed to exhibit this word order in both subject and non-subject questions, but in this paper we show that this is not true across the board. Another difference between subject and non-

subject questions is that non-V2 is a purely syntactic requirement in subject questions (in the relevant dialects), while there is always variation between V2 and non-V2 in non-subject questions, dependent on both syntax and information structure (Vangsnes 2007, Westergaard 2009a, forthcoming).¹ The main factor responsible for this variation is the length of the *wh*-element: In a number of dialects a clear distinction is made between simple (monosyllabic) and complex (disyllabic and longer) *wh*-elements. This is the case in the Tromsø dialect, for example, as shown in (4)–(5a), where complex *wh*-elements require V2 while simple *wh*-words allow either word order in non-subject questions (examples (5a, b) are taken from a corpus of spontaneous dialect speech (Anderssen 2006)).

(4) *Korsn bil kjøpte du? / *Korsn bil du kjøpte?* (Tromsø dialect)

how car bought you / how car you bought

‘What kind of car did you buy?’

(5) a. *kor e skoan hannes henne?* (INV, file Ole.17)

where are shoe.PL.DEF his LOC

‘Where are his shoes?’

b. *kor dem e henne?*

where they are LOC

‘Where are they?’

This distinction between simple and complex *wh*-elements is accounted for by arguing that the complex ones are phrases while the simple ones are heads, the latter

¹ There are indications in the data that non-V2 in subject questions is not completely obligatory in the dialects that allow it. It is unclear whether this is due to dialect mixture or other (possibly sociolinguistic) factors.

thus occupying the position that the verb would normally move to in V2 constructions, i.e. the head of the Interrogative Phrase (Westergaard & Vangsnes 2005, Westergaard 2009a). This blocks verb movement to this head position in those cases. Furthermore, when the syntax allows both V2 and non-V2, as in (5a), information structure determines which word order is chosen in a particular situation. Non-V2 is preferred when the subject conveys given or accessible information (typically a pronoun, marked [-FOC]), while V2 is generally used when the subject is new and/or focused information, marked [+FOC] (Westergaard 2003, 2009a, b). V2 word order can be accounted for by verb movement to a lower functional head (referred to as the head of the Topic Phrase in Westergaard 2009b), which has information structural effects. The syntactic structure of a V2 *wh*-question with a simple *wh*-element is provided in (6), while a corresponding non-V2 structure is illustrated in (7).

(6) IntP

3

Int° TopP

wh 3

Top° ...

[-FOC] IP

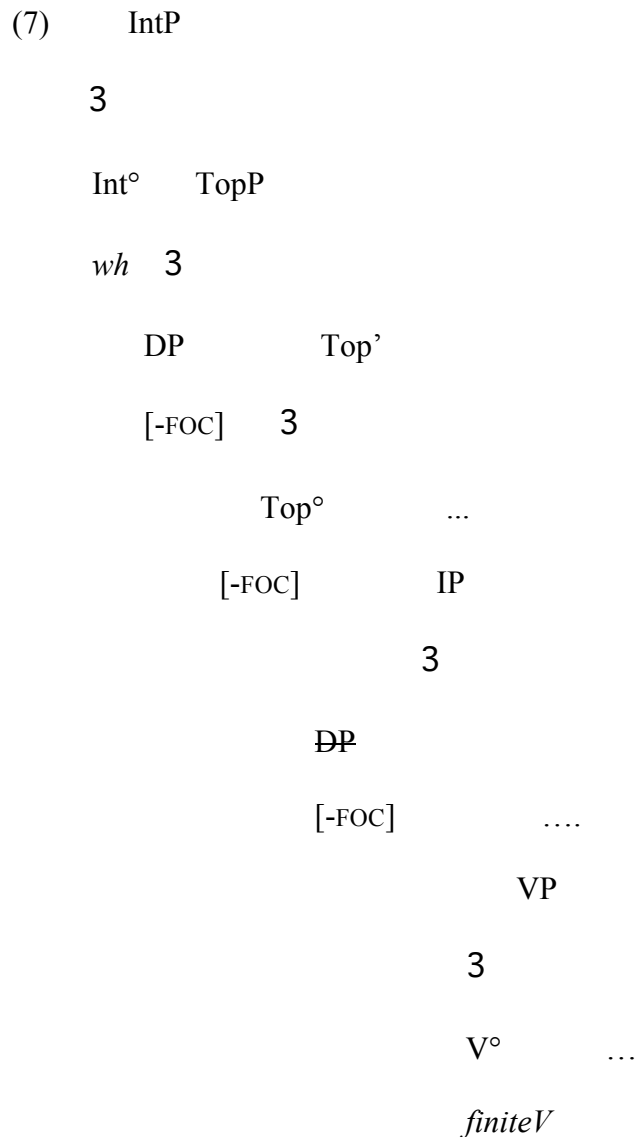
finiteV 3

DP I'

[+FOC] 3

I° ...

finiteV



Other dialects do not distinguish between simple/short and complex/long *wh*-elements and *wh*-phrases with respect to allowing non-V2 word order. In a study of the variety of Norwegian spoken in Nordmøre (northwestern part of the country), Åfarli (1986) shows that non-V2 in the equivalent of example (4) is perfectly acceptable in this dialect, as shown in (8). This has also been shown for other northwestern dialects by Vangsnes (2005, 2007); see also Westergaard & Vangsnes (2005).

(8) *Kåles bil du kjøpte?* (Nordmøre dialect)

how car you bought

‘What kind of car did you buy?’

Non-V2 word order after complex *wh*-elements has also been attested in a dialect ‘pocket’ in northern Troms, where this has been argued to be the result of language contact with Kven/Finnish and Saami, both of which are non-V2 languages (see Nilsen 1996, Westergaard 2005, forthcoming). An example of this is provided in (9), from Westergaard (forthcoming), where the second clause is a matrix *wh*-question introduced by the *wh*-item *katti*, ‘when’ (literally ‘what time’).

(9) *å korr fin du va på håre – katti du har årrna de?* (M4)

oh how nice you were on hair.DEF – what.time you have fixed it

‘Oh, how nice your hair looks – when did you have it done?’

In this paper, we present a proposal of the development of the non-V2 phenomenon in Norwegian dialects, which draws on elements from different previous accounts. Central to our analysis is the novel suggestion that changes in the properties of the functional item *som* play a key role. This non-inflecting item has many functions: For example, it introduces relative clauses and it also obligatorily appears after *wh*-subjects in embedded *wh*-clauses. This holds generally for all varieties of Norwegian and is illustrated by the example in (10), rendered in standard Norwegian *Bokmål*.

(10) *Han spurte hvem *(som) hadde kjøpt sko til to tusen.*

he asked who SOM had bought shoes to two thousand
'He asked who had bought shoes for two thousand.'

In certain dialects, the element *som* may also appear as a complementizer in an embedded clause under extraction of a *wh*-subject, as illustrated in (11b). This is ungrammatical in Standard Norwegian, where the complementizer has to be deleted in these cases (11a). The example in (11c) illustrates the fairly well-known fact that certain varieties of Norwegian (and Swedish) allow violations of the *that*-trace effect. As we show in section 3, sentences such as (11b) and (11c) are largely in complementary distribution across the country.

- (11) a. *Hvem tror du ___ er i baren nå?*
who think you is in bar.DEF now
- b. *Kem tror du **som** er i baren nå?*
who think you SOM is in bar.DEF now
- c. *Kem tror du **at** er i baren nå?*
who think you that is in bar.DEF now

All: 'Who do you think is in the bar now?'

In a questionnaire-based study of twenty-five individuals divided into six crudely defined regions, Nordgård (1985, 1988) makes an interesting observation regarding the item *som*. He finds that there is overlap between dialects that allow *som* under extraction (11b) and dialects that allow non-V2 in matrix *wh*-clauses, e.g. (1a, b). This correlation leads him to postulate the following condition (Nordgård 1988: 35, somewhat adapted here).

(12) Nordgård's Condition: A dialect may have non-inverted word order in matrix *wh*-questions iff the dialect allows insertion of the complementizer *som* under extraction of the embedded subject.

Note the role of *som* in Nordgård's Condition: The same element that replaces the declarative complementizer *at* under extraction is the one that appears in the matrix Left Periphery in subject *wh*-questions, instantiating the non-V2 word order in that particular syntactic context.

In this paper, we consider Nordgård's Condition more carefully. Whereas Nordgård based it on judgments obtained from a very low number of informants, we show that new and systematic data from around 400 informants by and large confirm his findings. We argue that this correlation is due to a diachronic development (from V2 to non-V2) that started out with subject *wh*-questions. More specifically, we propose that the puzzling lack of subject-verb inversion encountered in Norwegian dialects has in fact arisen from a minor change in the complementizer system, whereby the item *som* has partially taken over lexicalization properties of the declarative complementizer *at*, and this has ultimately opened up for a grammar with non-V2 in matrix *wh*-clauses. Furthermore, we argue that the microvariation that can now be observed across contemporary dialects reflects a series of stepwise changes in the properties of the Norwegian Left Periphery following the initial changes in the properties of *som*.

3. The data

3.1. Methodology

In the Scandinavian Dialect Syntax project, both questionnaire data and recordings were sampled from more than 100 locations spread across all of Norway during the period 2007–2011. Four individuals were consulted at each measuring point. The questionnaire consisted of approximately 140 test sentences, which were presented aurally with a prerecorded local/regional voice, and the informants were asked to judge them on a scale from 1 (bad) to 5 (good). The results from the survey have been assembled in an electronic online database, the *Nordic Syntax Database* (henceforth NSD, cf. Lindstad et al. 2009), along with the results from partially overlapping questionnaires from Swedish, Danish, Faroese, and Icelandic dialects.

The questionnaire investigated a large number of different phenomena, and only four sentences tested non-V2 in matrix *wh*-questions. The four test sentences are the following, rendered here in a northern Norwegian form. In (13a) the *wh*-constituent is short (monosyllabic) and functions as the subject of the clause: In this case the realization of non-V2 is due to the occurrence of *som* between the fronted subject and the finite verb. In (13b) the *wh*-constituent is also short, but it is a non-subject, in this case a predicative. In (13c) the *wh*-constituent is also a non-subject, in this case a temporal adverbial, and it is more complex. Example (13d) involves a phrasal, hence complex, subject *wh*-constituent.²

² The following two test sentences were also added to the questionnaire after the data collection had started.

- (i) a. *Koffør han va så sur, egentli?*
 what-for he was so grumpy really
 ‘Why was he so grumpy?’
 b. *Korsn du ska feire 17. mai i år?*
 how you shall celebrate 17th May in year
 ‘How will you celebrate 17th of May this year?’

They were tested only in about half of the Norwegian measuring points, and we therefore do not have exhaustive information about the geographical distribution, but by and large the data acquired comply with the results for (13c) and (13d) insofar as the clear positive judgments are found within the same areas, see below. For (ia) there are some individual positive judgments at measuring points where other complex *wh*-phrases are not accepted, and although this invites further investigation, the overall picture seems quite clear.

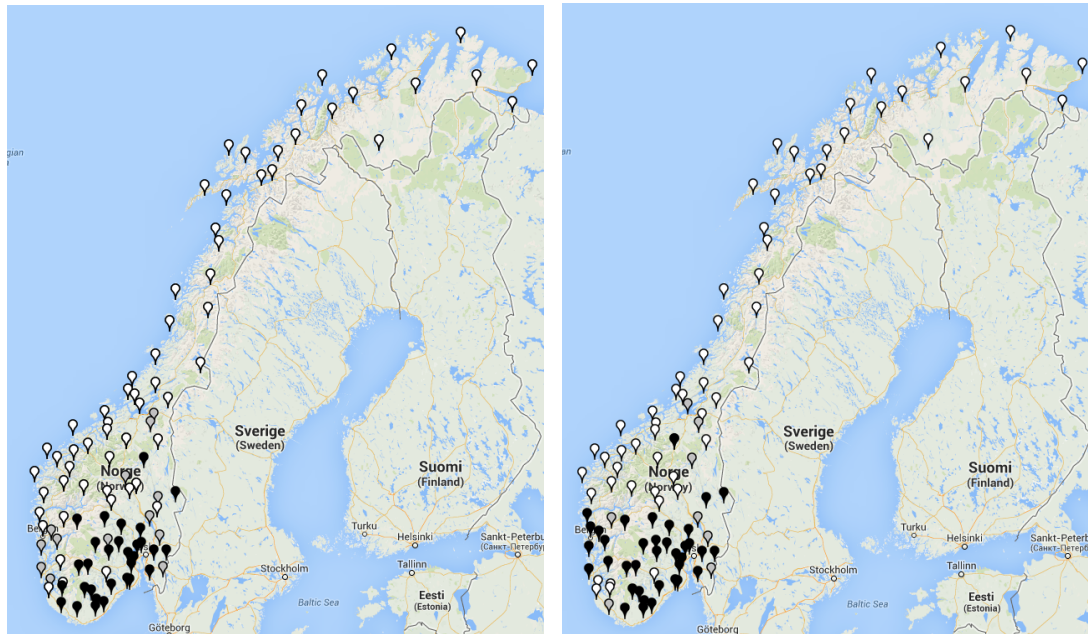
- (13) a. *Kem som sæll fiskeutstyr her i bygda?*
 who SOM sells fishing.gear here in village.DEF
 ‘Who sells fishing gear in this village?’
- b. *Ka du heite?*
 what you are.called
 ‘What is your name?’
- c. *Ka ti du gjekk ut a ungdomsskoln?*
 what time you went out of secondary-school
 ‘When did you finish secondary school?’
- d. *Kor mange eleva som går på den her skoln?*
 how many pupils SOM go on the here school
 ‘How many pupils go to this school?’

Versions of these examples *with* verb second were not included in the questionnaire, and hence we have no information on preferences for V2 in these cases. However, the overall picture is that in judgment tasks V2 is generally accepted by everyone, especially in non-subject questions (see Vangsnes 2007, Westergaard forthcoming). This is presumably due to influence from the standard language in such tasks. The questionnaire also included several examples of subject extraction from an embedded *wh*-question, which we return to in section 3.3.

3.2 Non-V2 word order

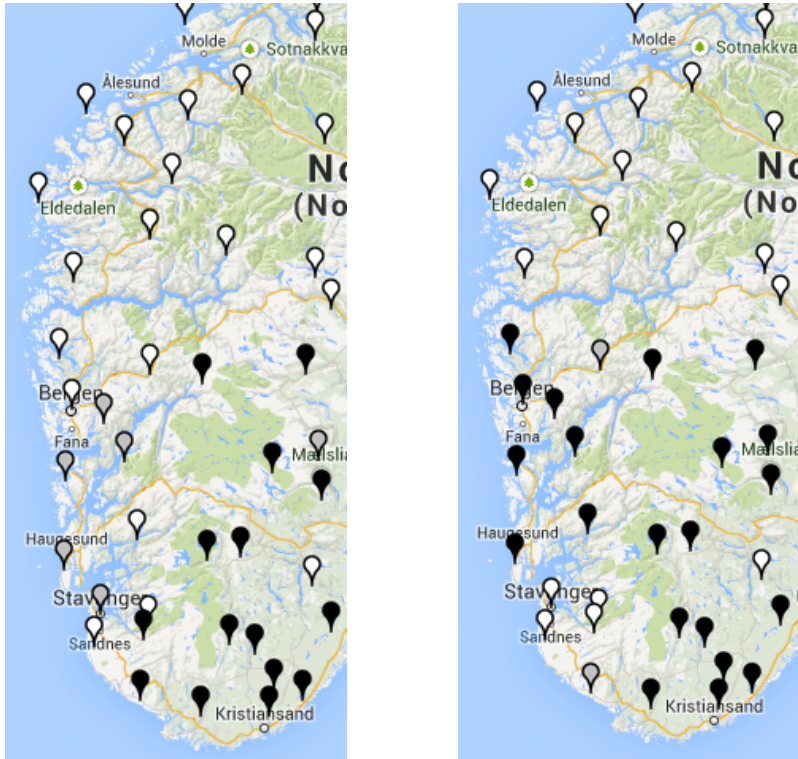
Non-V2 word order in the examples in (13) would not be allowed in Standard Norwegian. Maps 1a and 1b, which are generated from the NSD, show how subject and non-subject questions with monosyllabic *wh*-elements and non-V2 word order

(sentences 13a and 13b) were judged in Norwegian dialects: The white markers indicate a mean score of 4–5 on a 5-point scale (i.e. a positive judgment), the grey markers a mean score of about 3, and the black markers a negative score of 1–2.



Maps 1 a–b: *Questionnaire results from the Nordic Syntax Database for a matrix subject wh-question (left) and a matrix non-subject wh-question (right) without V2 across Norwegian dialects, in both cases with a short, monosyllabic wh-word (= (13a) versus (13b)).*

The maps show that both sentences are accepted in Northern and Central Norway, but not in the (south)eastern parts of the country. While there is considerable overlap, the judgments of the two sentences differ in an area between the cities Bergen and Stavanger in Western Norway, where (13a), the subject question, gets a medium or high score, whereas (13b), the non-subject question, gets a low score. In Maps 1c and 1d we have zoomed in on the area in question.



Maps 1 c–d: *Questionnaire results from the Nordic Syntax Database for a matrix subject wh-question (left) and a matrix non-subject wh-question (right) without V2 in Western Norway, in both cases with a short, monosyllabic wh-word (= (13a) versus (13b)).*

Both (13a) and (13b) involve short, monosyllabic *wh*-words, whereas (13c) and (13d) have more complex *wh*-phrases. The relevance of this distinction is well known from the literature, as it has been observed that many dialects allow non-V2 only if the *wh*-phrase is short/monosyllabic (see section 2). This condition is sometimes referred to as *Elstad's Generalization* after Elstad (1982) (see e.g. Fiva 1996). The issue is extensively discussed in e.g. Westergaard & Vangsnes (2005), Vangsnes (2005), and Westergaard (2009a). It is also well known that the generalization does not hold for all dialects, and (13c) and (13d) were included in the questionnaire to provide more detailed information on this issue. Ideally, several more test sentences should have been included, but since non-V2 in *wh*-questions was just one of many phenomena investigated by the questionnaire, this was not possible.

Sentence (13c) involves a non-subject (adverbial) disyllabic *wh*-phrase, repeated here for convenience. Map 2 shows that this sentence receives a high or medium mean score mainly in the northern part of western Norway (Nordvestlandet) and also in a small area in northern Norway (Nord-Troms, cf. section 2).

- (13) c. *Ka ti **du gjekk** ut a ungdomsskoln?*
what time you went out of secondary-school
‘When did you finish secondary school?’



Map 2: *Questionnaire results from the Nordic Syntax Database for a matrix complex non-subject (adverbial) wh-question without V2 across Norwegian dialects (=13c).*

Sentence (13d) involves a complex subject *wh*-phrase, and the results for this sentence is given in Map 3a. This sentence gets a medium or high score at more measuring points than (13c), both in western and northern Norway. In other words, it seems that complex *wh*-subjects, such as *kor mange eleva* ‘how many pupils’ in (13d), are accepted to a greater extent than complex *wh*-non-subjects, such as *ka ti* ‘when’ (lit. what time) in (13c). This observation also corresponds to previous findings: Fiva (1996) reports that Elstad’s Generalization does not hold for *wh*-subjects in the Tromsø dialect, and the same has been found in dialects spoken at Senja (just southwest of Tromsø), according to fieldwork carried out in connection with the NORMS project in October/November 2006.

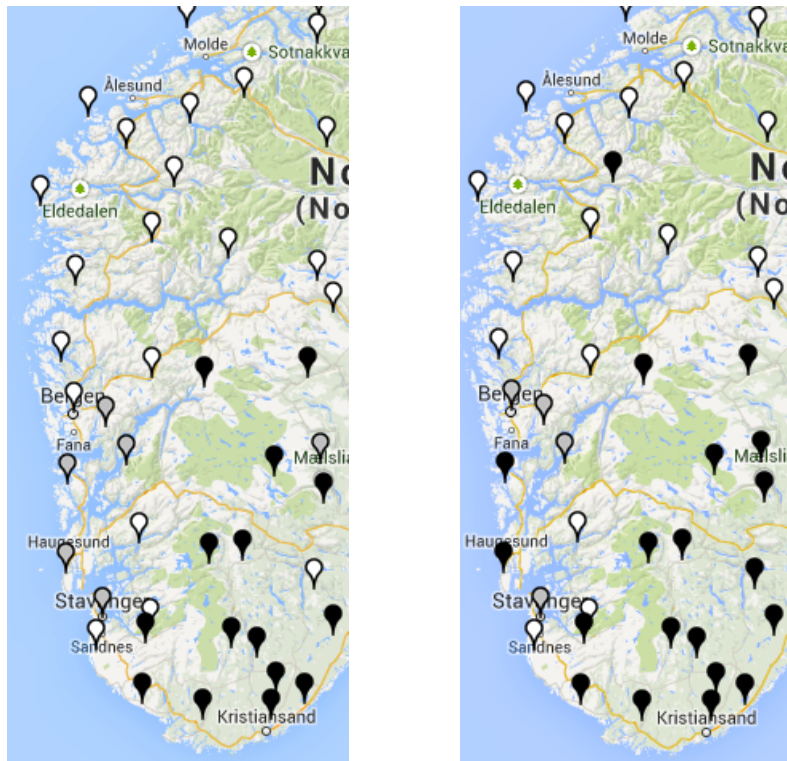
- (13) d. *Kor mange eleva som går på den her skolen?*
 how many pupils SOM go on the here school
 ‘How many pupils go to this school?’



Map 3a: *Questionnaire results from the Nordic Syntax Database for a matrix complex subject wh-question without V2 across Norwegian dialects (=13d).*

When we zoom in on western Norway, the picture regarding simple vs. complex *wh*-subjects reveals a clear tendency: *wh*-subjects with non-V2 seem by and large to be accepted regardless of their complexity in this area. This is illustrated by Map 1c and Map 3b which allow a comparison of sentences (13a) and (13d), i.e. a subject question with a short *wh*-element and a subject question with a complex *wh*-element: At the two locations between Stavanger and Bergen (southwestern Norway) where (13d) (complex *wh*-subject) gets a low mean score, (13a) (simple *wh*-subject) only gets a medium score, and in areas where (13d) gets a medium score, (13a) gets a medium or a high score. We noted above that dialect speakers in this area (Hordaland)

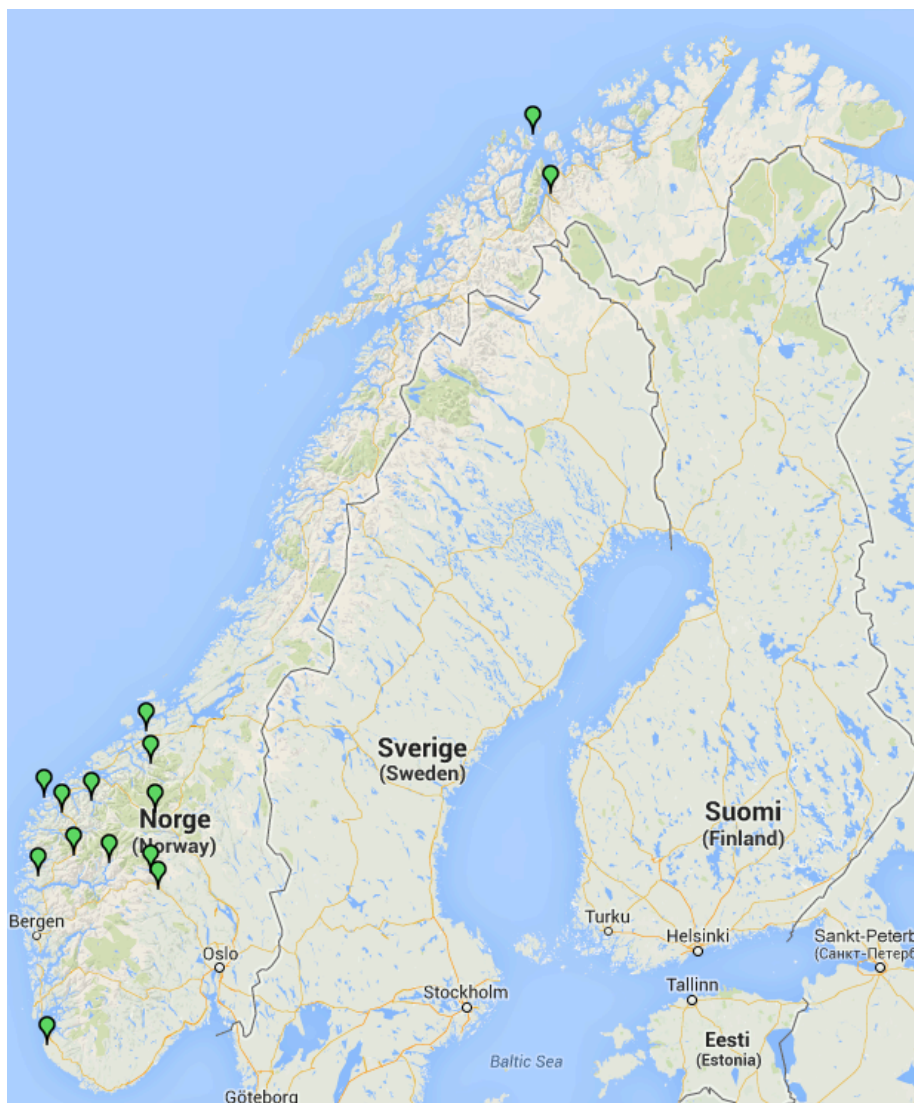
allow (13a) (simple *wh*-subject) but not (13b) (simple *wh*-non-subject), and although (13d) (complex *wh*-subjects) seems to be judged slightly less good than (13a), it does get better judgments than the non-subject questions. Our conclusion is that subject *wh*-questions are accepted with non-V2 regardless of their complexity in this area, and this is relevant for the proposal we present in section 4.



Maps 1c and 3b: *Monosyllabic (left) and complex (right) wh-subject with non-V2 in western Norwegian dialects (= (13a) versus (13d)).*

Map 4 shows the measuring points where all four test sentences get a high score. It is clear that northwestern Norway is the focal area of non-V2 in matrix *wh*-questions, with one outlier in Northern Norway (Kåfjord and Vannøya in Nord-Troms) and another in the southwest (Time in Rogaland). The fact that these areas/locations do not conform to Elstad's Generalization confirms reports in the literature (cf. section 2): Nordgård (1985, 1988) and Åfarli (1986) describe dialects of northwestern Norway in this way, and Nilsen (1996) does so for dialects of Nord-

Troms. Furthermore, Nordgård (1985, 1988) reports that his two informants from Rogaland county allow both simple and complex *wh*-constituents with non-V2, and in a comprehensive study Rognes (2010: 75ff) confirms Nordgård's finding and concludes that there is no complexity constraint on non-V2 matrix *wh*-questions in the Rogaland dialect area.³



Map 4: Areas where non-V2 is allowed in all *wh*-questions (sentences (13a–d)).

³ Notice that this entails that the acceptance of complex *wh*-constituents in matrix *wh*-clauses is more widespread in Rogaland than the single measuring point (Time) suggests, and this is indeed what Rognes (2010) documents through examples excerpted from both the Nordic Dialect Corpus and the World Wide Web.

3.3 Extraction

Turning now to extraction of *wh*-subjects out of a declarative complement clause to the matrix Left Periphery, the Norwegian questionnaire contained the following test sentences (rendered here in Standard Norwegian *Bokmål* orthography). In (14a) there is no complementizer introducing the embedded clause from which the fronted *wh*-constituent is extracted, in (14b) the embedded clause is introduced by *som*, and in (14c) the embedded clause is introduced by *at*.⁴

- (14) a. *Hvem tror du har gjort det?*
who think you have done it
'Who do you think has done it?'
- b. *Hvem tror du som har gjort det?*
who think you SOM has done it
'Who do you think has done it?'
- c. *Hvem tror du at har gjort det?*
who think you that has done it
'Who do you think has done it?'

The sentence in (14a), i.e. the only one with no complementizer preceding the subject gap, is considered good by virtually all informants in the survey. The interesting variation relates to (14b) on the one hand and (14c) on the other, i.e. to the difference between insertion of *at* or *som* in the embedded Left Periphery. These data give important indications about the distribution of *that*-trace violations in Norwegian dialects (cf. section 2).

⁴ The questionnaire also included a couple of other extraction sentences for which the data do not bear on the discussion here.



Map 5: *The distribution of high (white markers), medium (grey markers), and low (black markers) scores for sentence (14b): som-insertion under wh-extraction of an embedded subject.*

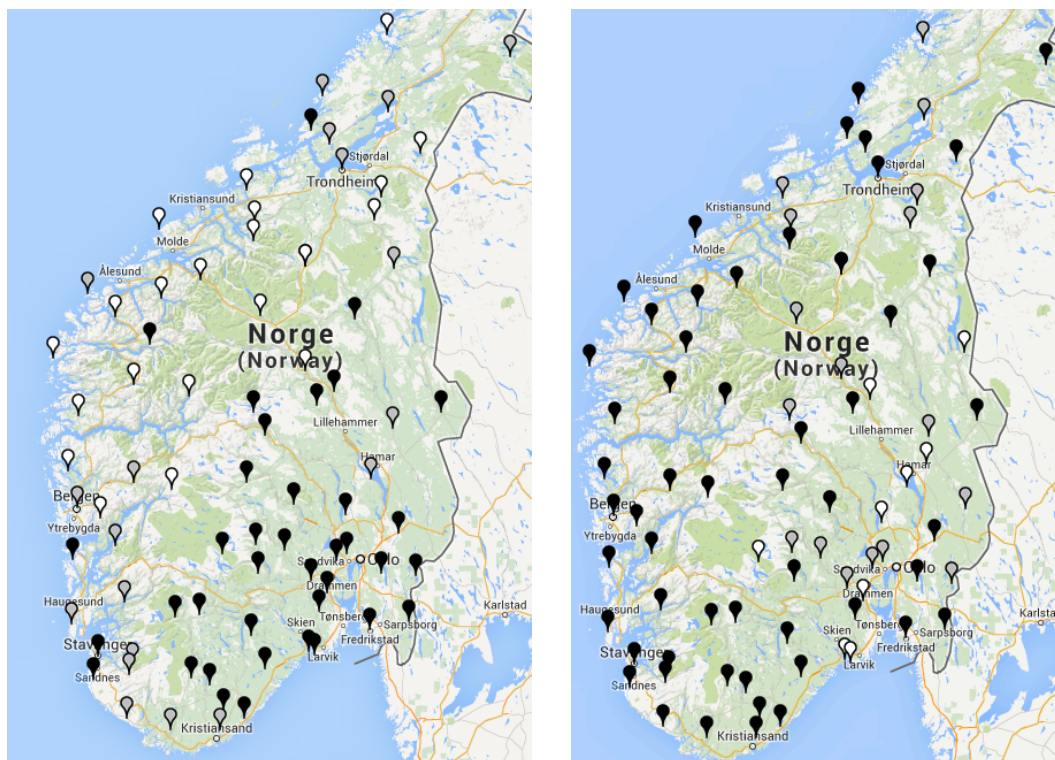
Map 5 shows the distribution of high (white markers), medium (grey), and low (black) scores for (14b), which is the sentence with insertion of *som*. When we compare this with the previous maps showing the distribution of non-V2 word order (in particular Map 1a), we see that there is a considerable overlap between disallowing non-V2 in matrix questions and disallowing embedded *som*-insertion

under *wh*-subject extraction. That is, the concentration of black markers in Map 5 is in eastern Norway, which is also where non-V2 is most clearly disallowed in matrix *wh*-questions. This observation thus points in the direction of Nordgård's Condition (cf. section 2), which emphasizes the correspondence between non-V2 word order and the complementizer *som* under extraction. Nevertheless, we see that this condition does not hold across the board, as there are some locations in western and central Norway where non-V2 is generally accepted, but where (14b) gets a low score.



Map 6: *The distribution of high (white markers), medium (grey markers), and low (black markers) scores for sentence (14c): at-insertion under wh-extraction of an embedded subject.*

Map 6 provides the corresponding information about sentence (14c), which has *at*-insertion under extraction. Comparing Maps 5 and 6, we may conclude that there is a considerable degree of complementarity between *som*-insertion and *at*-insertion, which is clear for northern Norway. In Maps 7a–b, we compare the findings from Maps 5 and 6 for southern Norway. Again we see that areas where *som*-insertion is generally allowed correspond to areas where *at*-insertion is disallowed.



Maps 7a-b: The acceptance of *som*-insertion (left = sentence (14b)) versus *at*-insertion (right = sentence (14c)) under extraction and fronting of a *wh*-subject from an embedded clause. White markers indicate a high average score, grey markers a medium average score and black markers a low average score.

This also means that Nordgård's Condition is to a large extent confirmed, now by a sample of more than 400 informants at over 100 locations: Dialects that allow *som* in the embedded Left Periphery under extraction of a *wh*-subject from the embedded clause also allow instantiations of non-V2 in the matrix Left Periphery.

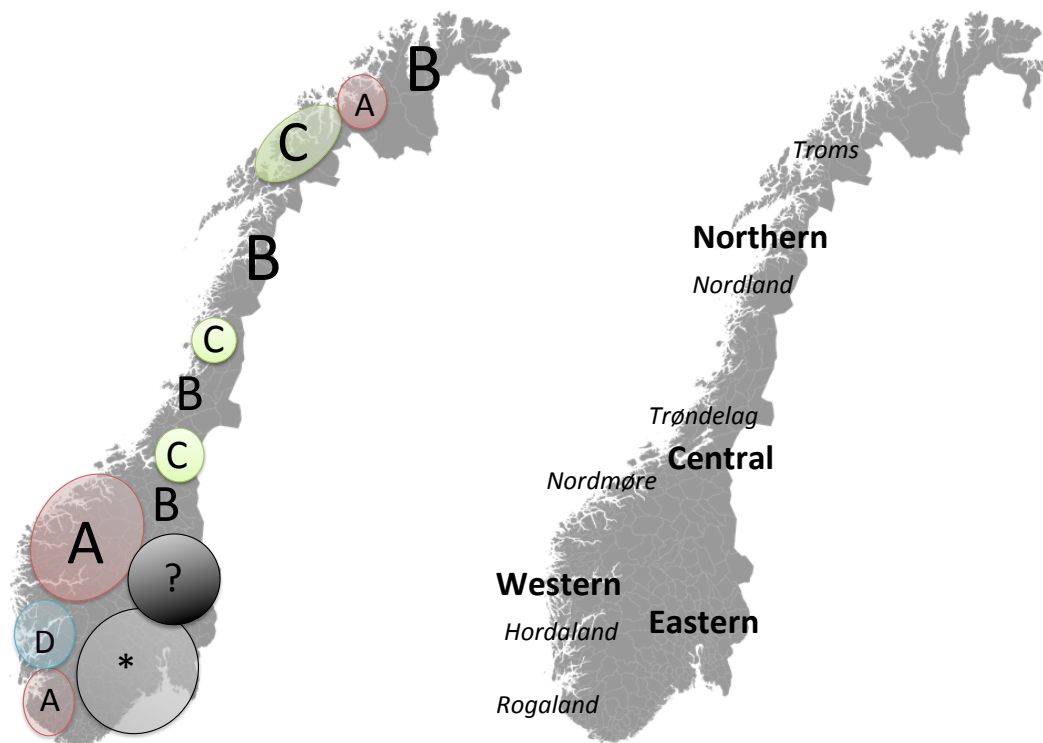
3.4 Intermediate summary

The broad picture falling out of the data in the *Nordic Syntax Database* – considered along with earlier reports in the literature – is that there are basically four types of non-V2 grammars in *wh*-questions in Norway, as outlined in (15). The dialects corresponding to these grammar types are plotted in Map 8.

- (15) a. Dialects which allow non-V2 regardless of the function and complexity of the *wh*-phrase, i.e. all the sentences in (13a–d) (northwestern Norway plus outliers, indicated as A on map 8).
- b. Dialects which allow only short, monosyllabic *wh*-words with non-V2 word order in subject as well as non-subject questions, i.e. (13a–b), but not (13c–d) (the dominating pattern in Trøndelag and northern Norway, which conforms to Elstad’s Generalization, indicated as B on map 8).
- c. Dialects which allow non-V2 with all *wh*-subjects, i.e. (13a) and (13d), but only with simple *wh*-elements in non-subject questions, i.e. (13b) but not (13c) (Parts of Troms county and also in Nordland and Trøndelag counties, indicated as C on map 8).
- d. Dialects which allow only subject *wh*-questions with non-V2, simple and complex, i.e. (13a) and (13d) but not (13b) and (13c) (southwestern Norway, indicated as D on map 8).

In addition to these four dialect types, there is a fifth type which does not allow non-V2 in *wh*-questions at all, just like the written language (*Bokmål* and *Nynorsk*). These dialects are spoken in (coastal) eastern Norway, in the area marked as * on Map 8.

Moreover, for the dialects in the northeastern part of eastern Norway the data from the Nordic Syntax Database do not provide any clear pattern.



Map 8 a-b: *Different dialect types with non-V2 in matrix wh-questions (left) and indications of main regions mentioned in the exposition (right).*

A generalization that emerges from this investigation is that if non-V2 is allowed in questions with complex *wh*-elements, it is also allowed in questions with short *wh*-elements, but not vice versa (as recognized in previous work, e.g. Westergaard 2009a). What our current investigation shows is that if non-V2 is allowed in non-subject questions, it is also allowed in subject questions, but not vice versa. We take these generalizations to indicate that non-V2 in questions with short *wh*-elements is historically prior to non-V2 with complex *wh*-elements, and that non-V2 in subject questions is historically prior to non-V2 in non-subject questions. That is, the development has started in subject questions and has then spread to non-subject

questions. Furthermore, there has been a spread of non-V2 from simple to complex *wh*-elements (at least in non-subject questions). Note that the grammar types do not form contiguous areas. This distribution indicates that these developments are partly related and to some extent independent, and that the timing of the independent changes is somewhat different across the dialects.

An alternative scenario, suggested to us by an anonymous reviewer, would be a development from non-V2 to V2. Obviously, this is in principle also a possible development, and we know that this is currently taking place in northern Troms, as these contact dialects are undergoing dialect leveling and thus becoming more similar to the dialects spoken in the larger region (Sollid 2003, Westergaard 2005, forthcoming; see also section 4.7). A development from non-V2 to V2 across the whole country would entail that the process has started in non-subject questions and then spread to subject questions, and in questions with complex *wh*-elements before questions with simple *wh*-elements. One major problem with this analysis is the theoretical rationale behind such a change. According to an economy principle proposed by van Gelderen (2004), historical processes typically first affect the least complex elements. We discuss this principle further in section 4.2.

In the next section we develop an analysis according to which the contemporary variation reflects stages in a possible diachronic development from V2 to non-V2 in Norwegian *wh*-questions. We focus on Nordgård's Condition and argue that the correspondence of extraction and non-V2 word order is the result of a historical development which starts with a change in the properties of the functional item *som*. This change makes *som* suitable for merger in an embedded *wh*-CP under subject extraction and subsequently eligible for merger in a matrix *wh*-CP. According to our

proposal the dialect types just listed emerge in the reverse order by successive changes in the grammatical system, starting with this change in *som*.

4. A diachronic account of the synchronic variation

In this section, we provide a diachronic account of the synchronic variation, first presenting a few theoretical preliminaries and then describing each stage of the development.

4.1. Theoretical preliminaries

We hypothesize that the following diachronic development has led to the synchronic microvariation just described. We specify the properties of each stage as well as the geographical areas affected by the change plus the contemporary language varieties that reflect the various stages.

(16) **Stage 0:** General V2 (currently attested in Standard (written) Norwegian and in (coastal) eastern Norwegian dialects)

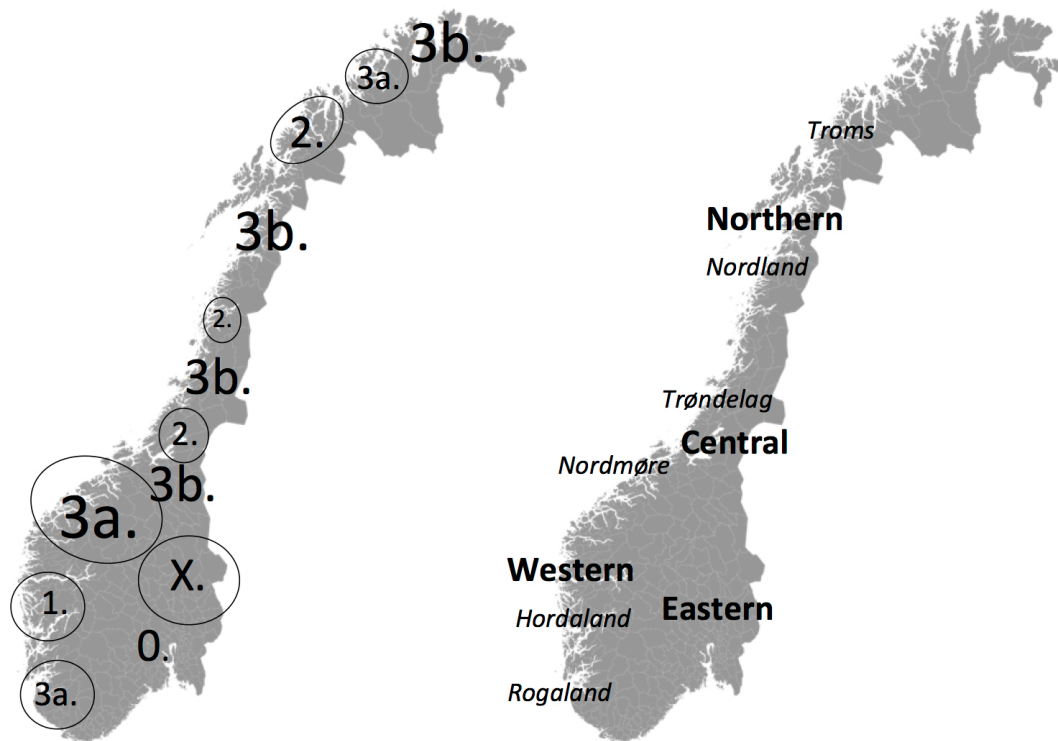
Stage 1: Non-V2 in all subject questions with short and long *wh*-elements (change has affected all areas except those mentioned under Stage 0; attested in present-day Hordaland and southwestern dialects).

Stage 2: Non-V2 spreads to non-subject questions with short *wh*-elements (change has affected all areas except those mentioned under Stages 0 and 1; attested in Tromsø and some other northern Norwegian dialects)

Stage 3a: Non-V2 spreads to non-subject questions with complex *wh*-elements (change has affected northwestern dialects and northern Troms dialects; still attested there)

Stage 3b: Non-V2 is restricted to short *wh*-elements in subject questions; cf. Stage 2. (This stage is somewhat questionable both theoretically and empirically, cf. section 4.6 below, but this system is the most common one in northern Norwegian dialects, corresponding to Elstad’s Generalization.)

In Map 9a the letters in Map 8a have been replaced by numbers corresponding to the postulated stages in our diachronic analysis, and for the sake of convenience we repeat the map naming the relevant regions.



Map 9a-b: Map of Norway with main regions and some counties of relevance to our account, where numbers indicate contemporary dialects with grammar types of the different diachronic stages.

We now propose an account of how these varieties have developed, emphasizing the early stages. Our proposal is that the non-V2 phenomenon as such has arisen from a minor change in the complementizer system, whereby the item *som* has partially taken over the lexicalization properties of the declarative complementizer *at* ‘that’.

Subsequently *som* has become associated also with the matrix Left Periphery, leading to a situation with non-V2 in the input to children, which has primed further developments resulting in the elaborate microvariation pattern that we see in contemporary dialects.

This proposal is different from other proposals pursued in the literature, where the change from V2 to non-V2 is argued to start with the monosyllabic *wh*-words, extending to the disyllabic *wh*-words and *wh*-phrases (Vangsnes 2005, Westergaard & Vangsnes 2005, Westergaard 2009a). In the present model, these changes constitute the development from Stage 1 to 2 and from 2 to 3a, i.e. after the development from Stage 0 (V2) to Stage 1 (the first stage with non-V2 syntax).

As a starting point for our analysis, Table 1 gives a simplified overview of the distribution of *that*, *at* and *som* in English, eastern Norwegian, and western/northern Norwegian dialects.⁵ Varieties like western/northern Norwegian that may use *som* under extraction have a variant of *som* that we call SOM-X, found here in the column labeled ‘Extraction C’.

Table 1: *at* and *som* in Norwegian dialects and the corresponding English words for four syntactic contexts.

	Declarative C	Extraction C	Relative/cleft C	Comparatives and small clauses
English	<i>that</i>	<i>that</i>	<i>that</i>	<i>as</i>
eastern Norw.	<i>at</i>	<i>at</i>	<i>som</i>	<i>som</i>
western/northern Norw.	<i>at</i>	<i>som</i>	<i>som</i>	<i>som</i>

⁵ For elaborate discussion of clause types with (and without) *som* in Swedish and North-Germanic, see Stroh-Wollin (1997, 2002), who argues that *som* has the feature [-FORCE].

Notice that we have also included comparative and small clause constructions in this table so as to give a broader typology of the environments in which *som* can occur.⁶ The element *som* (Icelandic and normalized Old Norse *sem*, Faroese *sum*) appears to be derived from the same root as the English adjective ‘same’ (Old Norse *samr*, Greek *homos*, Latin *simul-*), and its first use as a functional element was in correlative constructions in the Old Scandinavian languages. The use of *som* as a relativizer emerged in the Old Norse period, and against that background the use in extraction environments and in the matrix Left Periphery can be viewed as a continuation of the grammaticalization process. That is also highlighted by the comparison with English in Table 1.

We envisage that the differing properties of *som* (and *as*) across these varieties can be captured in terms of feature composition. We do not provide a very detailed analysis here, but the essential difference is that eastern Norwegian *som* is equipped with one feature more than English *as*, a feature which makes it suitable for use also in relative clauses (in addition to its use in comparatives and small clauses).⁷ For the sake of exposition let us call this feature [FIN], focusing on the expansion from non-finite to finite clauses. In western and northern Norwegian dialects *som* carries an additional feature making it suitable for insertion also in cases of extraction of a subject. We call this feature [NOM], alluding to the fact that it is a strongly subject-

⁶ An example of an equative comparative construction is given in (i) and of a small clause construction in (ii). See Åfarli and Eide (2000) for more information on the latter construction.

- (i) *Hun løper like fort som Mary.*
 she runs as fast as Mary
 ‘She runs as fast as Mary.’
- (ii) *Jeg kjenner John som lærer.*
 I know John as teacher
 ‘I know John as a teacher.’

⁷ There are also varieties of British English that use *as* in relative clauses. These varieties are mainly found in the central Midlands, cf. Herrmann (2002) and references cited there. This points towards a similar grammaticalization path to that of Norwegian *som*.

related element: a functional element with the features [FIN] and [NOM] seems an appropriate item for introducing embedded clauses with a subject gap.

Expanding this sketch would lead us too far here, and various alternatives are viable. However, we firmly believe that *som* is the item undergoing grammaticalization, and we would furthermore like to point out that our approach seems to be well suited for a nanosyntactic analysis whereby *som* acquires additional structure (features) without *at* necessarily losing any: The items may overlap in terms of lexicalization properties, but *som* will always win over *at* since the full lexicalization range of *at* is bigger than the structure at hand. With reference to Table 1, the idea would be that *at* in principle could lexicalize all four syntactic functions (by the Superset Principle, Caha 2009:55), but that it competes with other items on some of them and loses whenever the competing item cannot lexicalize one of the higher functions of *at*. To exemplify, on the assumption that both *at* and *som* may lexicalize Relative/Cleft C but only *at* Declarative C, *som* will be picked for Relative/Cleft C since *som* does not contain Declarative C. In other words, *som* is a better match for lexicalizing Relative/Cleft C than *at*, since it does not contain the irrelevant structure Declarative C. (See Vangsnes 2013 for an application of these ideas in a comparative study of *wh*-syncretisms across varieties of Germanic.)

4.2. Stage 0

The initial stage, represented by Standard Norwegian and eastern Norwegian dialects, is characterized by a requirement for V2, i.e. verb movement to the matrix Left Periphery, triggered by the head of IntP, as suggested by Westergaard (2009a, 2009b), based on work by Rizzi (1997, 2001). In other words, this is our implementation of the V2 requirement in matrix *wh*-clauses as we traditionally know it.

(17) [IntP *wh*P verb [TP (subject) T ...]]

That is, the verb moves to the head and the *wh*-item to the specifier of IntP. We assume that subject-initial *wh*-questions are derived in the same way as non-subject questions, following the analysis in Travis (1984) (see also Vangsnes 2008 for Norwegian). We also adopt the following assumption:

(18) *Lexicalization requirement for the Interrogative head*

Int^o has to be filled with overt material

The overt material in second position may consist of a verb, a complementizer, or a simple *wh*-item that can be analyzed as a head (see below). If Int^o is empty, the derivation will crash.

4.3. Stage 1

At some point between stage 0 and stage 1, SOM-X emerges in the language, and this is the first step towards allowing non-V2 structures in matrix *wh*-questions. As discussed above, SOM-X differs from “regular” *som* in that it can replace the complementizer *at* under extraction of a *wh*-subject, on our account due to the addition of the feature [NOM].

Importantly, the addition of the [NOM] feature is not sufficient to yield matrix *wh*-questions without V2: There exist varieties of North Germanic known to allow insertion of a relative complementizer in the embedded Left Periphery under *wh*-extraction without allowing non-V2. This is the case in Övdalian, spoken in Sweden

(Henrik Rosenkvist, personal communication), and it has been documented for western Jutlandic and Faroese by Øystein A. Vangsnes during NORMS fieldwork there in January and August 2008, respectively. Examples (19) and (20) were accepted by either most or a substantial proportion of the informants interviewed.

(19) *Hu manne trowe du **dæ** snakke æ dialekt hæ ve æ Minn?* (W. Jutlandic)

how many think you there speak the dialect here at the Minn

‘How many do you think speaks the dialect here at Thorsminde?’

(20) *Hvør tryrt tú (%ið / %**sum**) hevur gørt tað?* (Faroese)

who think you REL/SOM has done it

‘Who do you think has done it?’

The % sign in the Faroese example (20) indicates that, whereas all informants accept the version without any complementizer, a subset of these also allow the presence of one of the complementizers *ið* and *sum*. (For western Jutlandic all informants accepted the presence of *dæ* ‘there’ in the extraction clause, hence no % sign.) Notice furthermore that the item introducing the western Jutlandic embedded clause in (19) is the (otherwise) locative element *dæ* ‘there’, i.e. the equivalent of Danish *der*, which may introduce subject relatives but not non-subject relatives.

Given such examples in varieties that do not allow *som* (or *der*) in matrix subject *wh*-questions, it is clear that the mere presence of *som* under subject extraction is not sufficient to trigger non-V2 structures: By a further development, *som* must also

acquire the ability to lexicalize heads in the matrix Left Periphery.⁸ When this process is completed, the item has developed into SOM-X, and using SOM-X to meet the lexicalization requirement on matrix Int^o will then be more economical than moving the finite verb there. By assumption, merging a new lexical item is more economical than moving another element from elsewhere in the structure (cf. the Late Merge Principle in van Gelderen 2004; see also van Gelderen 2008, 2011, Westergaard 2009a, b).

We have now reached stage 1 in the development. This makes it possible for non-V2 structures of the sort ‘whP–som–verb’ to emerge. In such a grammar, non-V2 is licensed in subject *wh*-questions only. A simplified syntactic representation of the relevant part of the clause looks as follows; cf. sentences (13a) and (13d):

(21) [IntP *wh*P_{subject} SOM-X [TP verb ...]]

In (21), SOM-X has been merged in the head of IntP and the *wh*-item has been merged in SpecIntP. Due to the presence of SOM-X in the head of IntP, it is impossible for the verb to move to this head position. Since the *wh*-item has to be the subject, SpecTP will remain empty, and TP is licensed by the trace of the subject.

In order to understand how *som* may have acquired the ability to be merged also in the matrix Int^o (i.e. has become SOM-X), we can consider an argument from Lie (1992: 72) as to what may have triggered the change. Lie argues that non-V2 develops from clefts, along the following pattern:

⁸ Again we would deal with this in terms of feature acquisition, for instance by adding a [FORCE] feature to SOM-X: On the assumption that FORCE is absent in embedded clauses, this would account for why SOM-X equipped with the [FORCE] feature could be used in the matrix Left Periphery. A parallel issue would be how to deal with closely related varieties in which a (normally) embedded interrogative complementizer can also be used to introduce matrix yes/no questions. A case in point is the Rogaland dialect of Norwegian where the item *om* ‘if, whether’ is used to introduce both embedded dependent questions and matrix yes/no-questions (see Vangsnes 1996, Rognes 2011: 121ff for further discussion).

- (22) a. *Hå e de du si?*
 what is it you say
- b. *Hå e du si?*
 what is you say
- c. *Hå du si?*
 what you say
 ‘What do you say?’

According to Lie, all of these variants are viable in his native east Norwegian dialect spoken in Hedalen about 140 km to the northeast of Oslo. Sentence (22a) is an example of a cleft construction where the object of *say* has moved. The pronoun *det* ‘it’ satisfies the subject requirement in the main clause, whereas *du* ‘you’ satisfies the subject requirement in the embedded clause. In (22b), the expletive pronoun in the main clause has disappeared (deleted/not merged).⁹ In support of his own native intuitions Lie (1992: 72) provides evidence for (22b) from the following sources: a 16th century dialect text from Gudbrandsdalen (which is within the contiguous area of Stage 3a dialects displayed in Map 4 above) as well as more recent examples and mentionings for both eastern and western Norwegian dialects in the dialectological literature.¹⁰

Lie does not discuss subject questions in relation to his own proposal, but his account of non-subject questions may be easily transposed to subject questions. The trajectory would look as follows.

⁹ It is unclear what causes the disappearance of the expletive pronoun. See also Engdahl (2012) on optional expletive subjects in Swedish.

¹⁰ Lie (1992: 72f) provides some further arguments to back up his suggestion that the non-V2 structure has developed from clefts, but these are rather convoluted from our perspective, and we therefore chose not to present them here.

- (23) a. *Hå va de som skjedde?* (Hedalen dialect)
 what was it SOM happened
 ‘What was it that happened?’
- b. *Hå va som skjedde?*
 what was SOM happened
 ‘What was it that happened?’
- c. *Hå som skjedde?*
 what SOM happened
 ‘What happened?’

Sentence (23a) is an ordinary cleft structure with the expletive *det* ‘it’. In (23b) the expletive has been deleted, yielding a structure where the subject requirement has not been satisfied. The final structure is (23c), where SOM-X is lexicalizing Int°. This means that, from only being allowed in embedded clauses, SOM-X may now also appear in the matrix Int° head, cf. (23a) vs. (23c). The structures for (23b) and (23c) are provided in (24).

- (24) a. [IntP *Hå va* [TP ... [CP *som* [TP *skjedde*]]]]
- b. [IntP *Hå som* [TP *skjedde*]]

In terms of features, the change can be analyzed as a change from *som* with [FIN] and [NOM] to also include [FORCE]; that is, in (24b) *som* has the feature bundle [FIN, NOM, FORCE].¹¹

The intermediate stage in (23b) can in fact be encountered among present-day speakers. The following four authentic examples with an empty subject position have been excerpted from the Nordic Dialect Corpus. Note that the Fredrikstad and Jevnaker dialects are outside the area where non-V2 in matrix *wh*-questions have been documented, which suggests that this is an independent process. (Additional examples may be found in *Norsk Talespråkskorpus* (NoTa), a comprehensive corpus of the Oslo dialect which is also outside the non-V2 area.)

- (25) a. *va ær såmm spellær ?* (fredrikstad_01um)
 what is SOM plays
 ‘What’s playing?’
- b. *vemm ær såmm ikke kåmmer a ?* (jevnaker_02uk)
 who is SOM not come then
 ‘Who’s not coming, then?’
- c. *høkk æ såmm dirrigere no a ?* (alvdal_03gm)
 who is SOM conduct now then
 ‘Who’s conducting now, then?’
- d. *ja kemm e så he dei ?* (sokndal_03gm)
 yes who is SOM have them
 ‘Yes, who has them?’

¹¹ Recall here from section 4.1 that the addition of the feature [NOM] by hypothesis takes place when the use of *som* extends from relative clauses to cases of subject extraction. The strong subject relation induced by this feature is sustained also when the use of *som* extends to the matrix Left Periphery, and hence *som*-insertion is only found with matrix *wh*-subjects.

4.4. Stage 2

Stage 1 gives rise to non-V2 structures in the input. This makes it possible for learners to generalize this word order from subject questions to non-subject questions. For this development to take place, learners must analyze the *wh*-items as heads which may lexicalize Int°, thereby making it impossible for the verb to move to this position. According to an economy principle proposed by van Gelderen (2004), the Head Principle, there is a tendency (in acquisition and in language more generally) to treat any element as a head, and this is more likely to affect the simple than the complex *wh*-elements (cf. also the distinction between different types of pronouns in Cardinaletti & Starke 1996), cf. also Westergaard & Vangsnes (2005), Vangsnes (2005) and Westergaard (2009a). This will give us stage 2, where non-V2 is allowed in subject as well as non-subject questions, but only with short *wh*-elements in the latter question type. This change has taken place everywhere except in southwestern Norway (stage 1) as well as eastern Norway (stage 0), and this stage of the historical development is currently attested in the dialect spoken in Tromsø and some other Northern dialects, possibly also Trøndelag.

The reason for the complexity constraint in these dialects may also be due to simple *wh*-elements being considered separate categories in the acquisition process. It has been shown that Norwegian children have no problem distinguishing between verb movement in questions with simple and complex *wh*-elements in the relevant dialects (Westergaard 2009a). Such a development also corresponds to a model where diachronic development takes place in small steps (e.g. Lightfoot & Westergaard 2007, Westergaard 2009b, c).

Since only short (monosyllabic) *wh*-elements are heads in these dialects, it follows that complex *wh*-items cannot meet the lexicalization requirement on Int^o, and at Stage 2 only short *wh*-items can appear with non-V2 in non-subject questions. This is illustrated in (26), corresponding to sentence (13b).

(26) [IntP *wh* [TP subject verb ...]]

For subject questions, we assume the structure in (21) above, which licenses both simple and complex *wh*-items.

Note that V2 is still possible in questions with short *wh*-items in these dialects, but the verb is not assumed to move to Int^o in these cases. Rather, it moves to a lower functional head with information structure effects, illustrated in the structures in (6) and (7) above: V2 only occurs when the subject is discourse new or focused (cf. Westergaard 2009a for further details).

4.5. Stage 3a

At this stage, non-V2 word order extends from short to more complex *wh*-expressions in non-subject questions. This means that non-V2 is allowed in all *wh*-questions, irrespective of the function or complexity of the *wh*-element, cf. map 4. This could be argued to be due to the high frequency differences between simple and complex *wh*-elements, the former being attested approximately 90–95% in the input to children (cf. Westergaard 2009b). As argued in Westergaard (2009a), this frequency difference makes the complex *wh*-elements vulnerable to change. Syntactically, this means that the lexicalization requirement on Int^o is lifted, and there is no longer a V2

requirement in *wh*-questions, in the sense of verb movement to Int°. This is illustrated in the following structure, corresponding to sentences (13c) and (13d):

(27) [_{IntP} *wh*P Int [_{TP} (subject) verb ...]]

This pattern is especially prominent in northwestern dialects, the core dialect area for non-V2. It is also found in northern Troms, where it has been argued to be the result of language contact with a non-V2 language (Nilsen 1996, Westergaard 2005, forthcoming). That is, this pattern seems to be the result of two similar but independent developments. In the stage 3a dialects, non-V2 word order is optional in all *wh*-questions; i.e. both V2 and non-V2 word orders are possible. This means that the V2 word order that still appears in these dialects should be the result of verb movement to a lower head than Int°, e.g. the Topic head mentioned above. See Westergaard & Vangsnes (2005) and Westergaard (2009a, b) for further discussion.

4.6. Stage 3b

The pattern found in the Nordland dialects could be argued to constitute a separate stage in the historical development. In these dialects there seems to be a complexity constraint on the *wh*-element in both subject and non-subject questions, i.e. these dialects follow Elstad's Generalization. In Westergaard, Vangsnes & Lohndal (2012), we suggested that this is the result of a development from stage 2 dialects, resulting in non-V2 in subject questions now being *restricted* to contexts with short *wh*-elements only.

However, it is very difficult to motivate such a change syntactically: Since *som* is in Int° in subject questions, blocking verb movement to this position and the

corresponding V2 word order, the complexity of the *wh*-element should have no effect on the presence or absence of *som*. That is, if *som* and a short *wh*-element do not compete for the same syntactic (head) position, then there is no reason that *som* should be competing with a long *wh*-element.

In this article we would therefore like to suggest that the acceptability judgments found in Nordland do not reflect a separate stage with clear syntactic reflexes. Instead, we would argue that the difference between long and short *wh*-elements in subject questions is related to different preferences for cleft constructions in these contexts. Both Lie (1978) and Svenonius (1998) claim that clefts are especially frequent in questions in Norwegian, and in Søfteland's (2014) study she finds that clefts make up 10% in questions, compared to only 5% in non-questions. Søfteland (2014) also discusses possible differences between cleft and non-cleft questions and shows that there is considerable disagreement among linguists: While some argue that the two constructions are generally identical (e.g. Faarlund, Lie & Vannebo 1998), Lie (1978) shows that the two variants may differ with respect to presupposition in the following way: In the non-clefted question in (28a) nothing is presupposed (meaning that in addition to the answer to the question being what was said, it could also be "nothing"), while the question in (28b) presupposes that something was in fact said (examples cited in Søfteland 2014:70).

- (28) a. *Hva sa han om Hedda?*
what said he about Hedda?
- b. *Hva var det han sa om Hedda?*
what was it he said about Hedda
What did he say about Hedda?

Furthermore, both Lie (1978) and Gundel (2002) claim that non-clefted questions are virtually ungrammatical in Norwegian if there is a strong presupposition in the dependent clause in the clefted version of the question. This may in fact be what is causing the difference in grammaticality between the two subject questions in our study (13a, d), repeated here for convenience.

- (13) a. *Kem som sæll fiskeutstyr her i bygda?*
who SOM sells fishing.gear here in village.DEF
‘Who sells fishing gear in this village?’
- d. *Kor mange eleva som går på den her skoln?*
how many pupils SOM go on the here school
‘How many pupils go to this school?’

That is, while (13a) is a completely natural question also if the speaker does not know if anybody in fact does sell fishing gear in the village, (13d) is odd on the interpretation that the speaker does not know whether there are any pupils at all at the school. Thus, given the strong presupposition in (13d) that there are in fact some pupils at the school, the cleft version of the question (rendered in (29)) would be highly preferred.

- (29) *Kor mange eleva er det som går på den her skoln?*
how many pupils is it SOM go on the here school
‘How many pupils go to this school?’

This analysis is supported by the fact that the tendency to judge (13a) as more acceptable than (13d) is not only found in Nordland, but also in other dialects, especially in Hordaland (cf. section 3.2 above), although the score differences for the two sentences in these dialects are considerably smaller (cf. maps 1c and 3b). Nevertheless, the difference in acceptability judgments is relatively consistent across dialects. If our current suggestion is on the right track, this means that Elstad's Generalization may not be due to any syntactic requirement relevant to only some dialects, but could simply be a result of somewhat different preferences for cleft constructions in different types of subject questions. Clearly, much more extensive research is necessary to resolve this issue.

4.7. Summary

The diachronic scenario proposed in the previous sections relies mainly on the patterns that we see in the synchronic data and it accounts for the variation across Norwegian dialects with respect to *wh*-movement and the distribution of the complementizer *som* in interrogative clauses. As in previous work, e.g. Westergaard (2009a), the transition between the stages is generally argued to be due to economy principles or frequency issues in the acquisition process.

Furthermore, we have identified SOM-X and its important role in starting the development, now understood as the change from stage 0 to stage 1. The combination of *som* being used under extraction of *wh*-subjects plus a gradual “erosion” of function words in subject clefts, has, on our view, led to the association of *som* with the matrix Left Periphery. While there is a robust distinction between simple and complex *wh*-elements in non-subject questions in many dialects, our investigation has shown that this difference is much less clear in subject questions. We have therefore

suggested that the different grammaticality judgments obtained for questions with simple and complex subject *wh*-elements in certain dialects (Elstad's Generalization) are not due to a syntactic difference between the two, but simply a stronger preference for a cleft construction in the latter case.

The proposal made in this paper is that the contemporary geographical distribution is the result of stepwise diachronic change. In this development we see a tension between changes creating microvariation and changes that minimize distinctions within the relevant dialects. Examples of the former are the changes from stage 0 to stage 1 and from stage 1 to stage 2. In the change from stage 0 to 1, a difference between subject and non-subject questions emerge, while the change from stage 1 to 2 involves a difference between simple and complex *wh*-questions. An example of the latter is the change from stage 2 to 3a, resulting in loss of the difference between simple and complex *wh*-elements. Furthermore, the development from stage 1 to 2 could also be argued to be of the latter kind, ensuring that non-V2 spreads from subject to non-subject questions. Both types of processes are attested in child language, i.e. both under- and overgeneralization (see e.g. Snyder 2007, Westergaard 2014). In other words, while the historical changes are presumably the result of dialect contact and other sociolinguistic factors, the *nature* of the changes may be argued to be the result of the dynamics of language acquisition.

Finally, we would like to add a note of caution regarding the limitations of the data. Despite the fact that the NSD contains judgments from more than 400 speakers across the country, there were very few sentences in the questionnaire testing the relevant non-V2 properties. This means that our proposal is based on a small number of sentences, basically just one example of each non-V2 phenomenon investigated.

Our conclusions must therefore be considered tentative until further research has been carried out on more extensive data from the various dialect areas.

5. Conclusion

Based on acceptability judgments from about 400 speakers in the Nordic Syntax Database, we have investigated the distribution of non-V2 word order across a number of Norwegian dialects. Our findings show that if non-V2 is allowed in questions with complex *wh*-elements, it is also allowed in questions with simple *wh*-elements, but not vice versa. Furthermore, if a dialect allows non-V2 in non-subject questions, this word order is also found in subject questions (with the complementizer *som* in second position), but not vice versa. We have also seen that *som*-insertion and *that*-insertion under *wh*-extraction are more or less in complementary distribution, geographically speaking, which in turn is in line with Nordgård's Condition in that dialects allowing *that*-insertion by and large are outside the area where dialects allow non-V2 in matrix *wh*-questions. This distribution supports our proposal that non-V2 starts with a change in the complementizer inventory, affecting subject questions first. This idea can be harmonized with the proposal made by Lie (1992) that the non-V2 phenomenon has developed from a cleft structure. This also means that the present-day microvariation in Norwegian dialects with respect to V2/non-V2 word order in *wh*-questions is the result of a diachronic development that starts with a change in the complementizer *som*, resulting in non-V2 in subject questions as the first stage. This change then spreads in a step-wise fashion to include non-V2 word order also in other contexts.

References

- Åfarli, Tor A. 1986. Some syntactic structures in a dialect of Norwegian. *Working Papers in Linguistics* 3. 93–111.
- Åfarli, Tor A. & Kristin Melum Eide. 2000. Subject requirement and predication. *Nordic Journal of Linguistics* 23. 27–48.
- Anderssen, Merete. 2006. *The Acquisition of Compositional Definiteness in Norwegian*. University of Tromsø dissertation.
- Caha, Pavel. 2009. The Nanosyntax of Case. University of Tromsø dissertation.
- Cardinaletti, Anna & Michal Starke. 1996. Deficient pronouns: a view from Germanic. In Höskuldur Thráinsson, Samuel David Epstein & Steve Peter (eds.), *Studies in Comparative Germanic Syntax, Vol. 2*, 21–65. Dordrecht: Kluwer.
- Elstad, Kåre. 1982. Nordnorske dialektar. In Tove Bull & Kjellaug Jetne (eds.), *Nordnorsk: Språkav og språkforhold i Nord-Noreg*, 11–100. Oslo: Det Norske Samlaget.
- Engdahl, Elisabet. 2012. Optional expletive subjects in Swedish. *Nordic Journal of Linguistics* 35. 99–144.
- Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Fiva, Toril. 1996. Spørresetninger i Tromsødialekten. *Nordica Bergensia* 9. 139–155.
- van Gelderen, Elly. 2004. *Grammaticalization as Economy*. Amsterdam: John Benjamins.
- van Gelderen, Elly. 2008. *The Linguistic Cycle: Language Change and the Language Faculty*. Oxford: Oxford University Press.
- van Gelderen, Elly. 2011. Where did Late Merge go? Grammaticalization as feature economy. *Studia Linguistica* 62. 287–300.

- Gundel, Janet K. 2002. *It-clefts in English and Norwegian*. In Hilde Hasselgård, Bergljot Behrens, Cathrine Fabricius-Hansen & Stig Johansson (eds.), *Information structure in a cross-linguistic perspective*, 113–128. Amsterdam: Rodopi.
- Herrmann, Tanja. 2002. *Relative Clauses in Dialects of English: A Typological Approach*. Albert-Ludwigs-Universität Freiburg dissertation.
- Iversen, Ragnvald. 1918. *Syntaksen i Tromsø bymaal*. Kristiania: Bymaalslagets forlag.
- Johannessen, Janne Bondi, Joel Priestley, Kristin Hagen, Tor A. Åfarli & Øystein A. Vangsnes. 2009. The Nordic Dialect Corpus – an Advanced Research Tool. *NEALT Proceedings Series* 4. 73–80.
- Johannessen, Janne Bondi, Øystein A. Vangsnes, Signe Laake, Arne Martinus Lindstad & Tor A. Åfarli. 2010. The Nordic Dialect Corpus and Database: Methodological challenges in collecting data. In Barry Heselwood & Clive Upton (eds.), *Proceedings of Methods XIII Papers from the Thirteenth International Conference on Methods in Dialectology*, 113–122. Frankfurt: Peter Lang.
- Johannessen, Janne Bondi & Øystein A. Vangsnes. 2010. Nye nordiske isoglossar: Dialektforskning ved bruk av Nordisk syntaksdatabase (NSD). *Svenska landsmål och svenskt folkliv*, 2011. 69–78.
- Lie, Svein. 1978. Cleft wh-questions in Norwegian and their presuppositions. In Kirsten Gregersen, Hans Basbøll & Jacob Mey (eds.), *Papers from the fourth Scandinavian conference of linguistics*, 62–77. Odense: Odense University Press.
- Lie, Svein. 1992. Ka du sei? *Maal og Minne* 1992. 62–77.
- Lightfoot, David & Marit Westergaard. 2007. Language Acquisition and Language Change: Inter-relationships. *Language and Linguistics Compass* 1. 396–415.

- Lindstad, Arne Martinus, Anders Nøklestad, Janne Bondi Johannessen & Øystein A. Vangsnes. 2009. The Nordic Dialect Database: Mapping Microsyntactic Variation in the Scandinavian Languages. *NEALT Proceedings Series* 4. 283–286.
- Nilsen, Hilde. 1996. *Koffør dem sir det?* University of Tromsø Cand.philol-thesis.
- Nordgård, Torbjørn. 1985. *Word Order, Binding and the Empty Category Principle*. University of Trondheim Cand.philol-thesis.
- Nordgård, Torbjørn. 1988. Omkring ordstilling i hv-spørsmål i norske dialekter. *Skriftserie fra Institutt for fonetikk og lingvistikk, Universitetet i Bergen* 33. 26–37.
- Norsk talespråkskorpus [NoTa corpus], The text lab, ILN, University of Oslo. <http://www.tekstlab.uio.no/nota/oslo/index.html>
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of Grammar: A Handbook of Generative Syntax*, 281–337. Dordrecht: Kluwer.
- Rizzi, Luigi. 2001. On the position “(Int)errogative” in the left periphery of the clause. In Guglielmo Cinque & Giampaolo Salvi (eds.) *Current Studies in Italian Syntax*, 287–296. Oxford: Elsevier.
- Rognes, Stig. 2011. V2, V3, V4 (and maybe even more): The syntax of questions in the Rogaland dialects of Norway. University of Oslo MA thesis.
- Snyder, William. 2007. *Child Language: The Parametric Approach*. Oxford: Oxford University Press.
- Sollid, Hilde. 2003. Dialektsyntaks i Nordreisa: Språkdannelse og stabilisering i møtet mellom kvensk og norsk [Dialect syntax in Nordreisa: language creation and stabilization in a contact situation between Kven-Finnish and Norwegian]. University of Tromsø dissertation.

- Stroh-Wollin, Ulla. 1997. Framväxten av obligatoriska *som*. In Patrik Åström (ed.), *Studier i svensk språkhistoria* 4, 171–178. Stockholm: Institutionen för nordiska språk vid Stockholms universitet.
- Stroh-Wollin, Ulla. 2002. *Som*-satser med och utan *som* [*Som*-sentences with and without *som*]. Uppsala University dissertation.
- Svenonius, Peter. 1998. Clefts in Scandinavian: An investigation. *ZAS Papers in Linguistics* 10. 163–190.
- Søfteland, Åshild. 2014. Utbrytingskonstruksjonen i norsk spontantale [The cleft construction in Norwegian spontaneous speech]. University of Oslo 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.
- Travis, Lisa. 1984. Parameters and Effects of Word Order Variation. MIT dissertation.
- Zwart, Jan-Wouter. 1997. The Morphosyntax of Verb Movement: A Minimalist Approach to Dutch Syntax. Dordrecht: Kluwer.
- Vangsnes, Øystein A. 1996. *Om*-spørsmål i rogalandsk. *Nordica Bergensia*, 9. 166–188.
- Vangsnes, Øystein A. 2005. Microparameters for Norwegian *wh*-grammars. *Linguistic Variation Yearbook* 5. 187–226.
- Vangsnes, Øystein A. 2007. Pinning down fluctuating grammars: On main clause *wh*-syntax across Norwegian dialects. Talk given at NLVN seminar *Dialogue between paradigms*, Schæffergården 12–15 October.

- Vangsnes, Øystein A. 2008. On peripheral doubling in Scandinavian. In Sjef Barbiers, Olaf Koenenman, Marika Lekakou & Margreet van der Ham (eds.), *Microvariation in Syntactic Doubling*, 419–441. Bingley: Emerald.
- Vangsnes, Øystein A. 2013. Syncretism and functional expansion in Germanic *wh*-expressions. *Language Sciences* 36. 47–65.
- Vangsnes, Øystein A. & Marit Westergaard. 2014. *Ka korpuse fortæll? Om ordstilling i hv-spørsmål i norske dialekter*. [What does the corpus tell us? On word order in *wh*-questions in Norwegian dialects]. In Janne Bondi Johannessen & Kristin Hagen (eds.), *Språk i Norge og nabolanda. Ny forskning om talespråk*, 133–151. Oslo: Novus.
- 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.
- 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.
- Westergaard, Marit. 2009a. Microvariation as diachrony: A view from acquisition. *Journal of Comparative Germanic Linguistics* 12. 49–79.
- Westergaard, Marit. 2009b. *The Acquisition of Word Order: Micro-cues, information structure, and economy*. Amsterdam: John Benjamins.
- Westergaard, Marit. 2009c. Many Small Catastrophes: Gradualism in a Microparametric Perspective. In Paola Crisma & Giuseppe Longobardi (eds.), *Historical Syntax and Linguistic Theory*, 75–90. Oxford: Oxford University Press.
- Westergaard, Marit 2014. Linguistic variation and micro-cues in first language acquisition. *Linguistic Variation* 14. 26–45.

- Westergaard, Marit. Forthcoming. 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*. Amsterdam: John Benjamins.
- 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.
- Westergaard, Marit, Øystein A. Vangsnes & Terje Lohndal. 2012. Norwegian *som*: The complementizer that climbed to the matrix Left Periphery and caused Verb Second violations. In Valentina Bianchi & Cristiano Chesi (eds.), *Enjoy Linguistics! Papers offered to Luigi Rizzi on the occasion of his 60th birthday*, 329–343. Siena: CISCL Press.