

Norwegian *som*: The complementizer that climbed to the matrix Left Periphery and caused Verb Second violations*

MARIT WESTERGAARD
ØYSTEIN A. VANGSNES

University of Tromsø¹
marit.westergaard@uit.no
oystein.vangsnes@uit.no

TERJE LOHNDAL

Norwegian University
of Science and Technology
terjeloh@gmail.com

In this paper we consider the connection between Verb Second violations in Norwegian dialects and the possibility for insertion of the complementizer *som* under extraction of a *wh*-subject from an embedded clause, i.e. in **that*-trace contexts. We argue that it is the latter grammatical property that has led to the lifting of the Verb Second requirement in *wh*-clauses in Norwegian dialects. This deviates from earlier proposals which argue that this historical development starts with short *wh*-items being analyzed as X^0 s and inserted in the head position that otherwise attracts the finite verb.

Luigi: Thank you so much for truly inspiring academic papers as well as face-to-face interaction on various occasions in Siena, in Tromsø, and elsewhere.

1. Introduction

Based on his empirical findings regarding the absence of Verb Second (V2) effects in Norwegian dialects, Nordgård (1985, 1988) postulates the following condition (somewhat adapted here).

- (1) 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.

The Norwegian complementizer *som* is the element which introduces, among other things, relative clauses, and the essence of Nordgård's condition is that only dialects that allow sentences like the one in (2c) should allow sentences like the ones in (3), with non-V2 word order, which deviate from the examples in (4) from

* This paper has grown out of presentations given on a number of occasions in the period between June 2010 and August 2011. We are grateful to the many audiences for their feedback.

Standard Norwegian (see also Iversen 1918, Elstad 1982, Åfarli 1986, Taraldsen 1986, Lie 1992, Fiva 1996, Nilsen 1996, Rice and Svenonius 1998, Westergaard 2003, 2005, 2009a, 2009b, Westergaard and Vangsnes 2005, Vangsnes 2005).

- (2) a. Kem tror du — er i baren nå?
 who think you — is in bar.DEF now
 b. Kem tror du *at* er i baren nå?
 who think you that is in bar.DEF now
 c. Kem tror du *som* er i baren nå?
 who think you SOM is in bar.DEF now
 All: ‘Who do you think is in the bar now?’

- (3) a. Kem du skal møte i baren?
 who you shall meet in bar.DEF
 ‘Who will you meet in the bar?’
 b. Kem som er i baren no?
 who SOM is in bar.DEF now
 ‘Who’s in the bar now?’

- (4) a. Hvem skal du møte i baren?
 who shall you meet in bar.DEF
 ‘Who will you meet in the bar?’
 b. Hvem er i baren nå?
 who is in bar.DEF now
 ‘Who’s in the bar now?’

The example in (2b) is in itself an interesting one since it illustrates the fairly well-known fact that certain varieties of Norwegian (and Swedish) allow violations of the **that-trace*.

Notice that, in the case of a *wh*-subject, the manifestation of non-V2 is the insertion of the complementizer *som* (cf. (3b)): if it were absent, the finite verb would of course appear in second position. This fact highlights 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.

But non-V2 in these cases in fact entails lack of verb movement to the Left Periphery: if a sentence adverb is present, such as for instance negation, the finite verb must appear to the right of it. This is illustrated in (5) – example (5b) shows that *som* is obligatory if the finite verb occurs to the right of negation and (5c) that negation cannot appear to the right of the finite verb when *som* is present.

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

In this paper we consider Nordgård's condition more carefully. Whereas Nordgård based it on judgments obtained from a very low number of informants – twenty-five individuals divided into six crudely defined regions – we show that new and systematic data from around 400 informants by and large confirm his findings. However, we argue that this is due to a diachronic development (from V2 to non-V2) that started out with subject *wh*-questions. Our data were collected in the Scandinavian Dialect Syntax project and are available through the Nordic Syntax Database.

Furthermore, we propose that the puzzling lack of subject/verb inversion encountered in Norwegian dialects in fact has arisen from a minor change in the complementizer system, whereby the item *som* has partially taken over lexicalization properties of the otherwise declarative complementizer *at*, and 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.

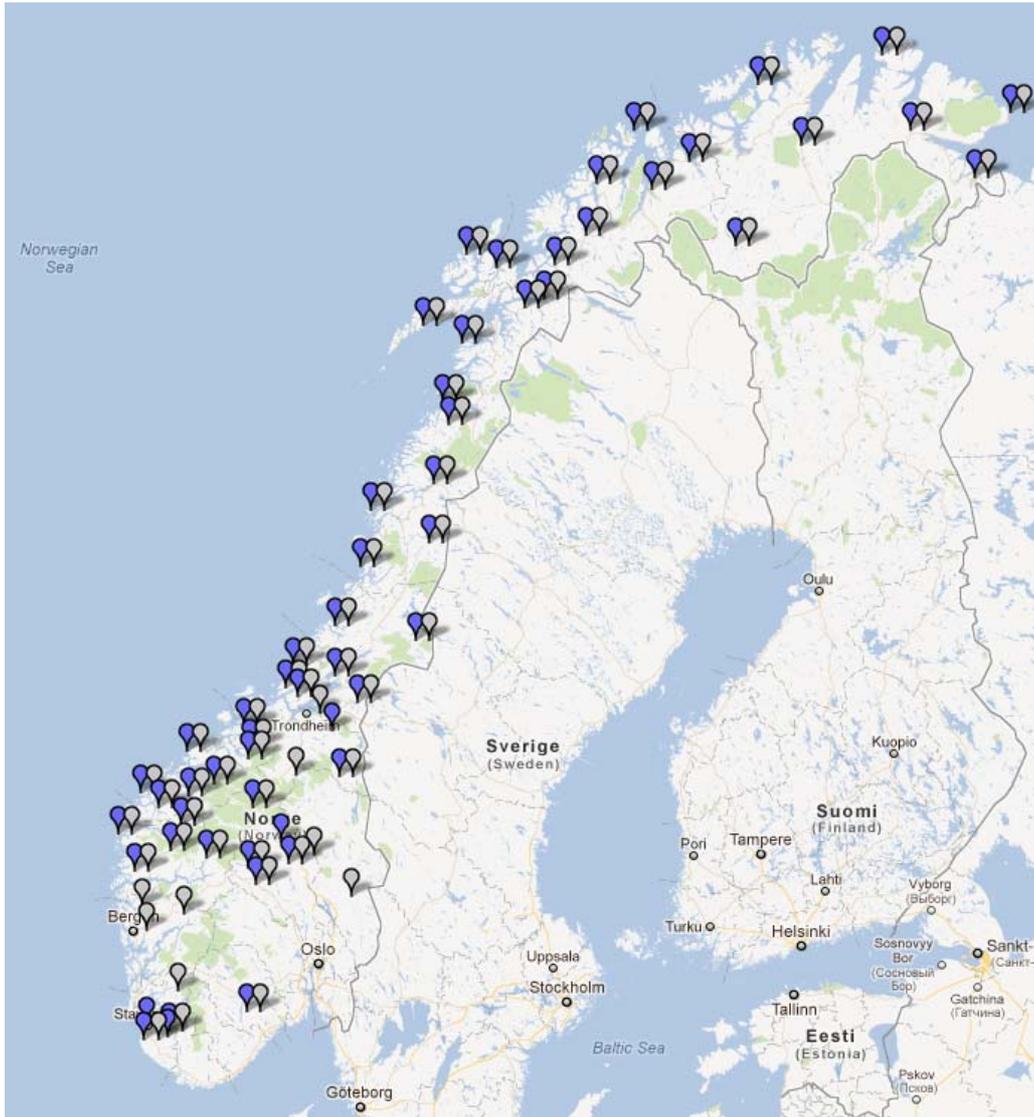
2. The data

In the Scandinavian Dialect Syntax project both questionnaire data and recordings were sampled from well over 100 locations spread across all of Norway in the period 2007 to 2011. Four individuals were consulted at each measure 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), along with the results from partially overlapping questionnaires from Swedish, Danish, Faroese, and Icelandic dialects. (For further information, see Johannessen et al. 2009.)

The questionnaire investigated a large number of different phenomena, and only six sentences tested the non-V2 phenomenon in matrix *wh*-questions. There were also six sentences testing extraction (including fronting) of an embedded *wh*-subject. Two of the non-V2-related test sentences are provided in (6a-b), rendered here in a North Norwegian dialectal form.

- (6) a. Ka du hete?
 what you are.called
 'What are you called?'
 b. Kem som sæll fiskeutstyr her i bygda?
 who SOM sells fishing.gear here in village.DEF
 'Who's selling fishing gear in this village?'

Non-V2 word order in these cases would not be allowed in Standard Norwegian, as we saw above (example (4)). Map 1 shows the distribution of positive results for sentences (6a-b) in the NSD: The blue markers represent sentence (6a) and grey markers (6b). 'Positive result' means that the mean score for all informants is 4 or higher (on a 5-point scale).

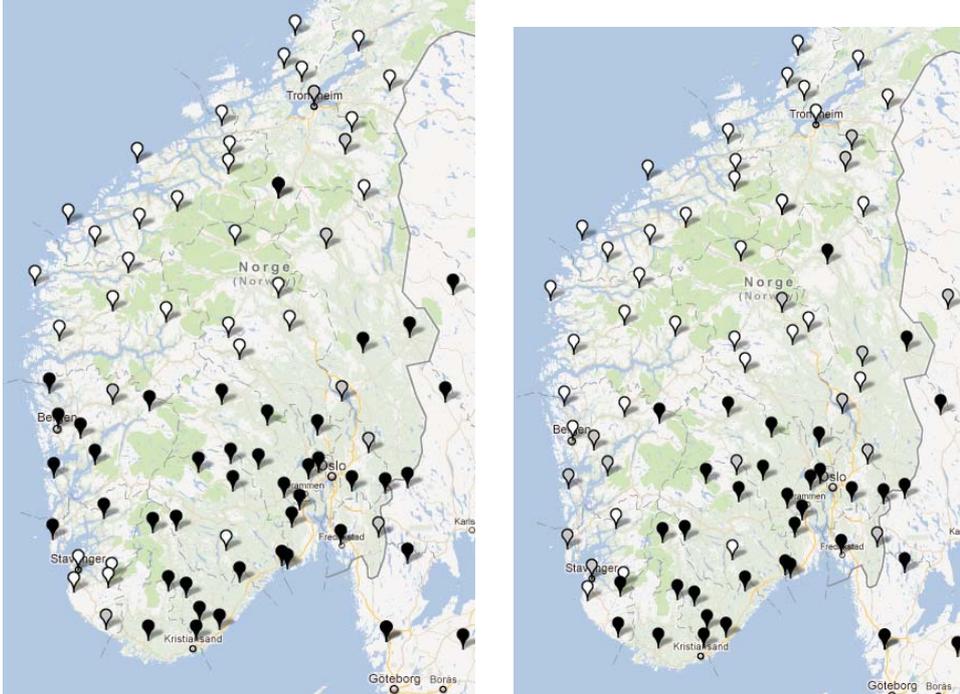


Map 1: Examples of non-V2 matrix *wh*-sentences with and without *som* (grey markers, cf. (6b) vs. blue markers, cf. (6a)).

The map clearly illustrates that both sentences are accepted in large parts of Norway, from North Norway through Central Norway (the area around the city Trondheim) and further down along the western coast in the south. In geographical terms a distinction between east and west is only made for the southern, “thicker” part of Norway, and we see that positive mean results are generally absent from Eastern Norway and from parts of Western Norway.

Furthermore, we also see that there are more grey markers than blue ones in the south. This entails that there are several dialects that accept sentence (6b) but not (6a). In other words, there are some dialects that accept non-V2 with a *wh*-subject but not with a non-subject. There are a few places where the situation is the opposite, but importantly, the grey-only markers outnumber the blue-only ones by 10 to 4, and furthermore, the grey-only markers form a contiguous area (in the west). The latter point is especially clear when we consider measure points with a medium result (i.e. a mean score of approximately 3). In the following two maps, high scores (>4) are shown by white markers, medium scores ($2 < 4$) by grey markers, and low scores (< 2) by black markers for sentences (6a) and (6b),

respectively. We see that the area in which (6b) (*wh*-subject question with *som*-insertion) gets either a high or medium score is much larger than the corresponding area for (6a) (non-subject *wh*-question). We consider this to be an important finding.



Maps 2 and 3: Mean scores in South Norway for sentences (6a) (left map; non-subject *wh*-clause) and (6b) (right map; subject *wh*-clause); white markers indicate high score, grey markers medium score, and black markers indicate low score.

Turning now to extraction of *wh*-subjects to the matrix Left Periphery, the Norwegian questionnaire contained the following test sentences (rendered here in Standard *Bokmål* Norwegian orthography).

- (7) 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?’



Map 4: The distribution of high (white markers), medium (grey markers), and low (black markers) scores for sentence (7b): *som*-insertion under *wh*-extraction of an embedded subject.

The sentence in (7a), 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 the other two sentences, i.e. the difference between insertion of *at* or *som* in the embedded Left Periphery. In (7b) we find *som*-insertion and in (7c) *at*-insertion. These data give important indications with respect to the distribution of **that*-trace violations in Norwegian dialects.

Map 4 shows the distribution of high (white markers), medium (grey), and low (black) scores for (7b), which is the sentence with insertion of *som*. When we compare this map with the previous maps, in particular map 1, we immediately see that there is a considerable match between not allowing non-V2 in matrix questions and not allowing embedded *som*-insertion under *wh*-subject extraction: the concentration of black markers is in Eastern Norway, which is also where non-V2 is most clearly not allowed in matrix *wh*-questions. The observation thus points very clearly in the direction of Nordgård's condition, although we see that

there are also some locations in Western and Central Norway where (7b) gets a low score.



Map 5: The distribution of high (white markers), medium (grey markers), and low (black markers) scores for sentence (7c): at-insertion under *wh*-extraction of an embedded subject.

Map 5 provides the corresponding information about sentence (7c) with *at*-insertion under extraction. Comparing maps 4 and 5, we may conclude that there is a considerable degree of complementarity between *som*-insertion and *at*-insertion, and furthermore, that Nordgård's condition to a large extent is confirmed, now with a sample from over 100 locations and more than 400 informants: 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.

Having noted that, we now return to the observation made initially that there are some dialects that allow non-V2 only with *wh*-subjects, manifested as insertion of *som* in second position. We believe that this gives us an indication of

how the phenomenon of non-V2 in *wh*-clauses has arisen, and in the analysis that we propose in the next section, we argue that it all starts with *som*.

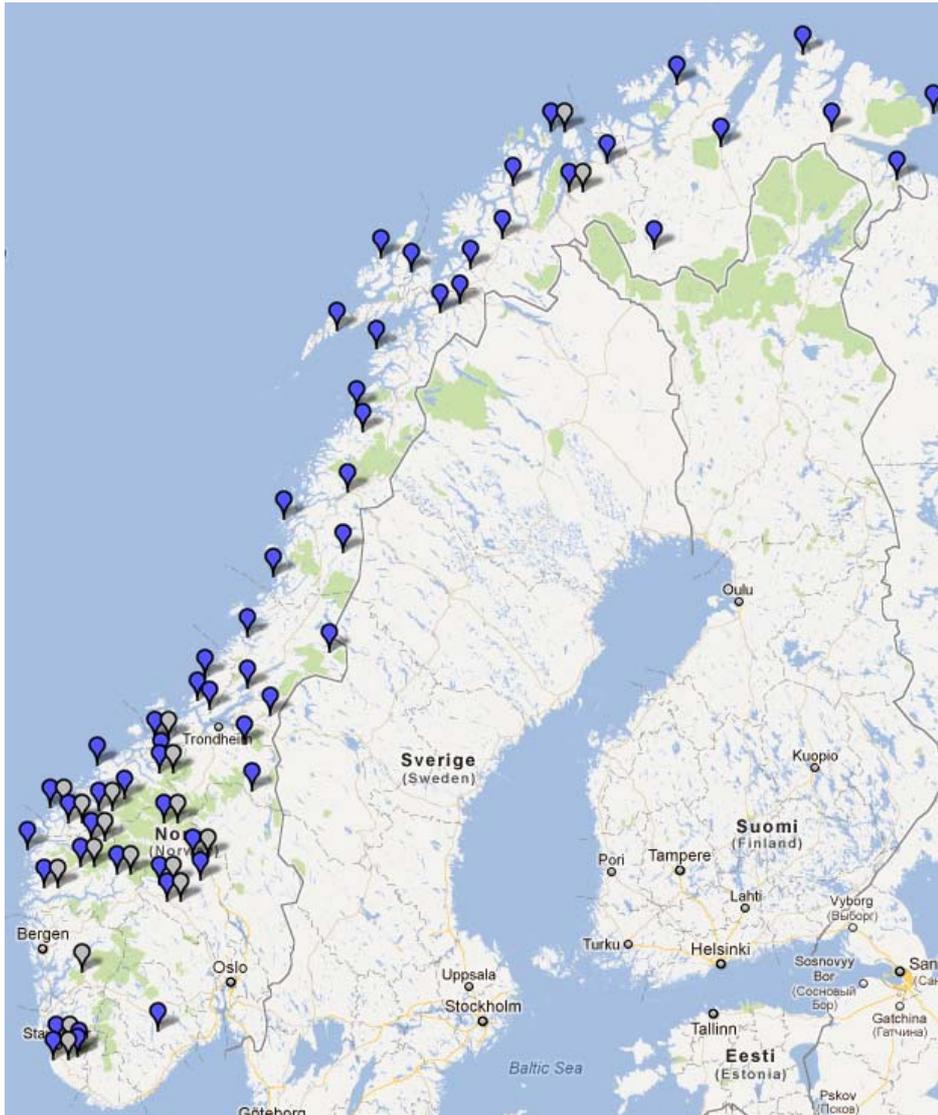
Before we turn to the analysis, it is necessary to add a note about different kinds of *wh*-elements. As has been extensively discussed in Westergaard and Vangsnes (2005), Vangsnes (2005), and Westergaard (2009a), dialects differ as to whether they accept only short, monosyllabic *wh*-words with non-V2 or whether V2-violations are acceptable also with more complex *wh*-constituents. Map 6 illustrates this for the subject *wh*-sentences in (8) and map 7 illustrates it for the non-subject *wh*-sentences in (9). ((8a) = (6b) and (9a) = (6a).) Blue markers indicate a high average score for the sentences with short *wh*-constituents and the grey markers indicate high average score for the sentences with long non-subject *wh*-constituents.

- (8) a. Kem som sæll fiskeutstyr her i bygda?
 who SOM sells fishing.gear here in village.DEF
 ‘Who’s selling fishing gear in this village?’
- b. 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?’
- (9) a. Ka du hete?
 what you are.called
 ‘What are you called?’
- b. Ka tid du gikk ut av ungdomsskolen?
 what time you went out of secondary-school.DEF
 ‘When did you graduate from secondary school?’

We see from these maps that there are more dialects that allow short *wh*-constituents than long *wh*-constituents with non-V2. We also see – again – that there are more dialects that allow subject *wh*-questions without V2, and when comparing the two maps, the careful viewer will also observe that there are more dialects that allow long *wh*-subjects than dialects that allow long *wh*-non-subjects.



Map 6: Short (blue) versus long (grey) wh-subjects in matrix sentences with non-V2.



Map 7: Short (blue) versus long (grey) *wh*-non-subjects in matrix sentences with non-V2.

We can therefore summarize these distributions as follows: If non-V2 is allowed in non-subject questions, then non-V2 is allowed in subject questions through the presence of *som*. If non-V2 is possible with long *wh*-elements, then non-V2 is also possible with short *wh*-elements. If non-V2 is possible in subject questions with long *wh*-elements, then non-V2 is possible in subject questions with short *wh*-elements. And finally, we have seen above that *som* and *at* under extraction are in complementary distribution across Norwegian dialects: Generally, only dialects that allow *som* in the embedded Left Periphery allow non-V2 in the matrix Left Periphery.

We hypothesize that the following diachronic development has led to the synchronic microvariation just described (contemporary language varieties that reflect the various stages are given in parentheses):

- (10) a. **Stage 0:** V2 (Standard Norwegian)
 b. **Stage 1:** Non-V2 in subject questions only (Hordaland dialects)
 c. **Stage 2:** Non-V2 also in non-subject questions with short *wh*-elements (several dialects in North Norway, e.g. Tromsø)

- d. **Stage 3a:** Non-V2 in questions also with long *wh*-elements (North Western dialects and Northern Troms dialects)
- e. **Stage 3b:** Non-V2 in subject questions restricted to short *wh*-elements only (Nordland dialects)

For the sake of clarity we include a map where the locations mentioned are indicated.



Map 8: Map of Norway with indications of main regions and some counties and places of importance to the discussion in the paper.

In the next section, we propose an analysis of how these four stages have developed, emphasizing stages 1 and 2. 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 otherwise 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 and then extending to the disyllabic *wh*-words and *wh*-phrases (Vangsnes

2005, Westergaard and Vangsnes 2005, Westergaard 2009a). In the present model, these changes constitute the development from Stage 1 to 2 and from 2 to 3.

3. The proposal

Table 1 gives a simplified overview of the distribution of *that*, *at* and *som* in English, Eastern Norwegian, and West/North Norwegian dialects. Varieties like West/North Norwegian that may use *som* under extraction have a variant of *som* that we will call *som-x*.

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

	Decl. C	Extraction C	Relative/cleft C
English	<i>that</i>	<i>that</i>	<i>that</i>
East Norw.	<i>at</i>	<i>at</i>	<i>som</i>
West/North Norw.	<i>at</i>	<i>som</i>	<i>som</i>

The initial stage – stage 0, represented by Standard Norwegian and East Norwegian dialects – is characterized by a requirement for 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 a particular take on the Verb Second requirement as we traditionally know it.

(11) $[_{\text{IntP}} \text{whP verb } [_{\text{TP}} \text{T} \dots]]$

There are some varieties of North Germanic that are known to allow insertion of *som* in the embedded Left Periphery under *wh*-extraction without allowing non-V2. Cases in question are Övdalian, spoken in Sweden (Henrik Rosenkvist, personal communication), Western Jutlandic (NORMS fieldwork, January 2008), and Faroese (NORMS fieldwork, August 2008).

We nevertheless argue that acquiring *som-x* is the first step towards allowing non-V2-structures: The next step is when *som-x* is allowed to lexicalize the head of IntP. When that step has been taken, using *som-x* to meet the lexicalization requirement on Int^o will be more economical than moving the finite verb there.

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 this grammar, non-V2 is licensed in subject *wh*-questions only, without any constraints on the size of the *wh*-element. This grammar corresponds to the Hordaland dialects, cf. the discussion at the end of section 2. A simplified syntactic representation of the relevant part of the clause looks as follows:

(12) $[_{\text{IntP}} \text{whP}_{\text{subject}} \text{som } [_{\text{TP}} \text{T} \dots]]$

In technical terms we think of the ability of *som-x* to lexicalize matrix Int^o as (direct) merge being preferred over move. Although we will leave aside here the exact formulation of what allows *som-x* to acquire this ability, 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:

- (13) a. Ka **va det** du sa?
 what was it you said
 b. Ka **va** du sa?
 what was you said
 c. Ka du sa?
 what you said
 'What did you say?'

Lie provides evidence for (13b) from dialect texts from the 16th century, from quotes in the dialectological literature, and from his own dialect (Hedalen, northwest of Oslo). He does not discuss subject questions, but his argument could be transferred to such cases in the following way:

- (14) a. Ka **va det** som skjedde?
 what was it SOM happened
 'What was it that happened?'
 b. Ka **va** som skjedde?
 what was SOM happened
 'What was it that happened?'
 c. Ka som skjedde?
 what SOM happened
 'What happened?'

This means that, from only being allowed in embedded clauses, *som-x* may now also appear in the matrix Int° head ((14a) vs. (14c)). The stage in (14b) can in fact be encountered among present-day speakers. The following four authentic examples have been excerpted from the Nordic Dialect Corpus (additional examples can be found in the NoTa Oslo corpus of spoken Norwegian, see <http://www.tekstlab.uio.no/nota/oslo/index.html>). In all these examples, the subject position is empty.

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

What is important is that Stage 1 gives rise to non-V2 structures in the input. Subsequently, this makes it possible for a learner to analyze short *wh*-items as heads lexicalizing Int° (cf. van Gelderen 2004, Westergaard and Vangsnes 2005,

Vangsnes 2005, Westergaard 2009a). That will give us stage 2, where non-V2 is allowed in both subject and non-subject questions.

Since only short *wh*-elements are heads, 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. Complex *wh*-subjects are still allowed at this stage since *som-x* will meet the requirement on Int^o. This stage is reflected by the Tromsø dialect and a few others. For an analysis of this stage, see Vangsnes (2005) or Westergaard (2009a).

The last stage is split into two different developments. We only give a brief outline of them here, as a more detailed account is currently being prepared by the authors. At Stage 3a, the lexicalization requirement on Int^o is lifted and non-V2 word order extends to more complex *wh*-expressions, also as non-subjects. This pattern is especially prominent in Northwestern dialects.

Stage 3b is different, in that *som-x* here loses its ability to lexicalize Int^o. Verb Movement to Int^o is therefore required unless the *wh*-element is a head (in both subject and non-subject questions). The dialects in Nordland county in North Norway reflect this stage.

4. Conclusion

To conclude, two general observations are crucial for our proposed account:

- (16) a. *som*-insertion and *that*-insertion under *wh*-extraction are by and large in complementary distribution, geographically speaking.
- b. there are more places where only subject *wh*-questions are allowed with non-V2 than only non-subject questions (mainly the Hordaland dialects).

The observation in (16b) arguably supports the general idea that non-V2 starts with a change in the complementizer system/inventory, affecting subject questions first. This idea is in harmony with the proposal in Lie (1992) that the 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 the complementizer *som* and that results in a distribution that generally supports Nordgård's (1985) original observation.