

ARTICLE

Children's acquisition of word order variation: A study of subject placement in embedded clauses in Norwegian

Tina RINGSTAD¹ and Marit WESTERGAARD²

¹NTNU Norwegian University of Science and Technology, Norway

²UiT The Arctic University of Norway & NTNU Norwegian University of Science and Technology, Norway

Corresponding author: Tina Ringstad; Email: tina.ringstad@ntnu.no

(Received 23 November 2020; revised 27 January 2023; accepted 24 March 2023)

Abstract

Norwegian embedded clauses give children two options for subject placement: preceding or following negation (S-Neg/Neg-S). In the adult language, S-Neg is the 'default' and highly frequent option, and Neg-S is infrequent in children's input. However, Neg-S may be argued to be the structurally less complex. We investigate whether children are aware of the existence of both subject positions, and if they prefer the more frequent or the less complex position. Through an elicited production task with monolingual Norwegian children (N=33, age 3;1-6;1) we find that children in general overuse the Neg-S option, and we suggest that children have an inherent preference for the less complex position, due to a principle of structural economy. We also find that a group of children display U-shaped development, first using only S-Neg, then only Neg-S and finally S-Neg again, and we relate this to structure building and economy of movement.

Keywords: First language acquisition; subject placement; word order variation; Norwegian; economy; frequency

Introduction

Children acquiring Norwegian face a challenge when learning subject placement: they must discover that both main and embedded clauses have two subject positions, that there are certain restrictions on their use, and that the generalizations for subject placement in main and embedded clauses are not identical. The task is not made any easier by the infrequency of one of the positions. In this paper we investigate children's production of subjects in embedded clauses to see whether they initially prefer the syntactically less complex position, which is very infrequent in their input, or the position that is more frequent, which is also argued to be more complex, involving more syntactic movement.

The two subject positions are visible in the presence of negation (and occasionally other adverbs): the subject may precede or follow negation (S-Neg or Neg-S), as shown in

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

(1a, b) for main clauses and (2a, b) for embedded clauses. We follow an assumption that the two positions are found above and below negation in a hierarchical structure, and therefore refer to them as high and low subject placement. On accounts of derivational economy, the low subject position is less complex than the high one, since it involves less movement (e.g., Westergaard, 2009).

- (1) a. I går spelte *han* **ikkje** xylofon.
Yesterday played he not xylophone
 ‘Yesterday he didn’t play the xylophone’
- b. I går spelte **ikkje** *Olav* xylofon.
Yesterday played not Olav xylophone
 ‘Yesterday, Olav didn’t play the xylophone’
- (2) a. Eg merka at *han* **ikkje** spelte xylofon.
I noticed that he not played xylophone
 ‘I noticed that he didn’t play the xylophone’
- b. Eg merka at **ikkje** *Olav* spelte xylofon.
I noticed that not Olav played xylophone
 ‘I noticed that Olav didn’t play the xylophone’

What distinguishes the two positions is not entirely clear, but there seems to be a distinction between subject types (pronominal and lexical NPs, the latter henceforth referred to as simply NPs), in that pronominal subjects are more often used in the high position, while NP subjects are more common in the low position (e.g., Johannessen & Garbacz, 2011; Munch, 2013; Westergaard, 2011). While S-Neg is argued to be the default word order (Faarlund et al., 1997), meaning that it is always permitted and that it is the more frequently used of the two alternatives (Westergaard, 2011), the Neg-S order is not allowed in all types of embedded clauses (Eide, 2002; Garbacz, 2005, 2014). This means that children rarely encounter the low subject position in embedded clauses in their input. The infrequency of the low subject position might make it unavailable to early acquisition, and even if it is noticed by the child, the child must encounter it sufficiently often to make the correct generalizations of its use.

The prominent distinction of high frequency (the high subject position) versus low complexity (the low subject position) makes this variability an ideal test case for children’s preferences when faced with optionality: the role of the input and the role of intrinsic biases. In this paper we implement a controlled experimental study, varying the subject type, to investigate whether children use both subject positions in embedded clauses, and whether they distinguish pronominal from NP subjects in their production. We also examine adults’ subject placement in embedded clauses in spontaneous speech, to address previous findings showing discrepancies in the distribution of NP subjects. We start our paper with an overview of current knowledge of target language generalizations and L1 acquisition studies of subject placement, before outlining research questions and predictions. We then present our studies on adult spontaneous production and child elicited production showing that *i*) children generally overuse the low subject position in embedded clauses, and *ii*) some children’s production resembles a U-shaped development in which they start out using only the high subject position, before categorically using the low position and finally the high position again.

Background and research questions

Target language

Two subject positions: Structure and distinguishing factors

The word order variation illustrated in Norwegian main and embedded clauses in (1)-(2) is generally assumed to be distinguished syntactically in that the subject occupies different hierarchical positions, one above and one below negation. Various accounts exist for the exact location of the two subject positions (e.g., Bentzen, 2009; Cardinaletti, 2004; Kiss, 1996; Mohr, 2005; Westergaard & Vangsnes, 2005; Wiklund et al., 2007). For the sake of clarity a brief outline of the syntactic structure follows, but since the precise location of the subjects in the syntactic structure is not at issue in the present paper, we refer to the two positions as high and low.

We adhere to an analysis where both subject positions are located in the I-domain (Holmberg, 1993; Westergaard & Vangsnes, 2005): the subject following negation (1b, 2b) is moved out of the VP (Bobaljik & Jonas, 1996) to the specifier of TP, and the subject preceding negation (1a, 2a) resides in a higher specifier position which we call SpecS(ubj)P (following Ringstad, 2021). The two subject positions then encapsulate negation, which can be analysed as adjunction to TP (following Eide, 2002; Holmberg, 1993; see also Brandtler, 2008 on this negation position for Swedish). Figure 1 shows a basic analysis of the embedded clauses in (2).¹ The structure in Figure 1 shows that the distinction between the high and low subject is that in the former, the subject has undergone a longer move.

The high and low subjects are typically distinguished by semantic, pragmatic and syntactic features. Subject type, i.e., pronominal vs. lexical NP subjects, is often discussed as a distinction, and so is information structure: the subject position preceding negation is argued to be reserved for discourse given subjects and the one following these elements for subjects expressing new information (e.g., Bentzen, 2009; Nilsen, 1997; Svenonius, 2002; Westergaard & Vangsnes, 2005). The former position is typically linked to pronominal subjects, having a familiar referent, and the latter to lexical NP subjects, where the referent varies according to whether it is indefinite (introducing new information) or definite (encoding familiar information). An additional factor is stress: the low subject position may be correlated with contrastively stressed subjects, but stress does not seem to be obligatory for low subjects (e.g., Westergaard & Vangsnes, 2005). Furthermore, specificity is often discussed in relation to the two subject positions: the high position is typically reserved for subjects with a specific reading, and correspondingly, the low position for subjects with a non-specific interpretation (e.g., Bentzen, 2009). Using different subject positions in questions may also have a pragmatic effect, as in 'Can you not/not you X?' signalling that the speaker holds either positive or pessimistic beliefs about the answer (Urbanik & Svennevig, 2018). Lastly, the low subject position might not be permitted in all environments: whereas Faarlund et al. (1997, p. 891) note that the low position is possible in embedded *wh*-questions, relative clauses and clauses with the complementizer *at* 'that', there are indications that not all speakers accept low subjects in 'because'-clauses (Eide, 2002) and 'when'-clauses (Garbacz, 2014). The low subject position is absent in

¹Norwegian is typically also argued to have a lower position for negation, at the edge of VP (Holmberg & Platzack, 1995). This negation is not generally discussed in relation to the two subject positions, but in studies focusing on verb placement in embedded clauses, where it is used as a diagnostic that the verb has not moved out of the VP if the word order is Negation-Verb. Working out the details of the two negation positions is not part of the present study, but we note that for children acquiring the language, two possible positions for negation may be confusing: Which position should be used when?

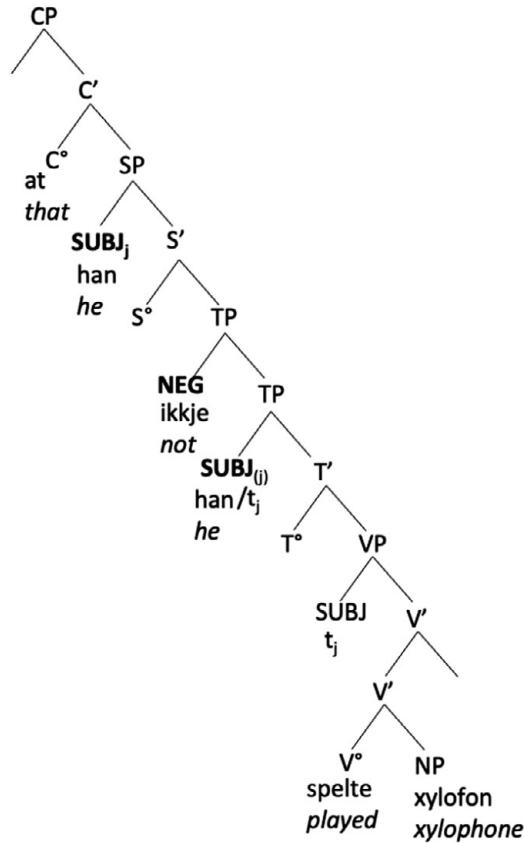


Figure 1. A basic analysis of the embedded clause in (2), showing negation as adjoined to TP, the low subject position in the specifier of TP, and the high subject position in the specifier of a S(ubj)P.

some clause types (such as *fordi* ‘because’, and *selv om* ‘even though’) in corpora of spontaneous speech (Garbacz, 2005; Munch, 2013).

The two subject positions can be informative of the role of structural/derivational complexity in acquisition. Economy-based approaches to acquisition hold that children have a general preference for the lowest available syntactic positions and avoid structures involving more syntactic operations (Clark & Roberts, 1993; Jakubowicz, 2011; Westergaard, 2009). Such a view contends that children only undertake syntactic operations that are obligatorily required (Jakubowicz, 2011), or, as formulated by Westergaard’s (2009, p. 2016) principle of structural economy: children ‘a) only build as much structure as there is evidence for in the input’, and ‘b) only move elements as far as there is evidence for in the input’. If children are inclined to use less complex structures before more complex structures, and longer moves are more complex, children are expected to start out using the low subject position in embedded clauses. There is also variation present showing that the longer move is not obligatory. Children will thus be expected to expand their analysis of the target grammar, moving subjects to the high position, as they encounter sufficient evidence that they should.

Two subject positions: Frequency and distribution in numbers

In order to investigate the distribution of subject types in main and embedded clauses, Westergaard (2011) has analyzed two relatively large corpora, one of eight adults producing child-directed speech (CDS) (The Tromsø corpus, Anderssen, 2006) and one of adult-to-adult dialogues involving altogether 166 speakers (*NoTa*, 'The Norwegian corpus of spoken language', Tekstlab, 2004). The main findings are the following: NP subjects are rare in both clause types, accounting for only 6.1% (87/1435) and 5.9% (17/290) of all subjects in main and embedded clauses respectively in CDS and as little as 1.3% (29/2199) and 5.9% (38/640) in the *NoTa* corpus. In main as well as embedded clauses, pronominal subjects tend to appear in the high position, 87.9% (1185/1348) and 90.1% (246/273) in CDS, and 84.7% (1839/2170) and 88.2% (531/602) in *NoTa*. This entails that across corpora and clause types, approximately 10-15% of pronouns are used in the low position. NP subjects, on the other hand, appear almost exclusively in the low position in main clauses, 97.7% (85/87) and 96.6% (28/29) in child-directed speech and *NoTa*, while in embedded clauses this position is considerably less common, 64.7% (11/17) and only 26.3% (10/38) in the two corpora respectively. Note that these percentages are based on very low raw numbers, especially from CDS, where 7/11 examples are due to one speaker.

Westergaard (2011) also shows that the word order chosen is generally based on information structure, as represented by the two subject types (pronouns and NPs), but that other factors may play a role as well, e.g., category type, length and prosodic weight of the subject. Furthermore, it is worth noting that the high position seems to be highly preferred in embedded clauses, regardless of the category of the subject (88.2% for pronouns and 73.7% for NPs in *NoTa*), possibly as a result of a subject-first strategy (this does not apply in main clauses, as the two subject positions are only visible in non-subject-initial clauses). The preference of the high subject position means that the low position in embedded clauses is attested to an almost negligible extent in adult speech: embedded clauses with negation are found to make up less than one percent of the total number of clauses (0.45% in Ringstad, 2019; and 0.8% in Westergaard, 2009), and the low subject position comprises only a small part of these numbers. We summarise the findings from Westergaard (2011) for embedded clauses in Table 1.

This distribution of high vs. low subjects in adult speech can be informative of the role of input in child language acquisition. According to studies and models of acquisition grounded in input frequency, high(er) frequency forms or structures are important, as *i*) when faced with inconsistencies, children tend to regularize inconsistent forms based on the higher frequency form (Hudson Kam & Newport, 2005; Schwab et al., 2018), *ii*) they are acquired early, or earlier, than competing forms (Ambridge et al., 2015), and *iii*) they cause children to assume that this form is more likely to be found in the target language (Pearl, 2021). The numbers in Table 1 show that the high subject position is by far the

Table 1. The occurrence of high subjects (S-Neg) in embedded clauses in two corpora of adult spontaneous speech (data from Westergaard, 2011)

	Tromsø corpus (CDS)	NoTa corpus
Pronominal subjects	90.1% (246/273)	88.2% (531/602)
NP subjects	35.3% (6/17)	73.7% (28/38)
Total	86.9% (252/290)	87.3% (559/640)

more frequent in adult speech (this is confirmed in study 1, see below), and thus in children's input. Therefore, if children rely solely, or mostly, on input frequency when acquiring word order variation such as subject placement, they would be expected to initially only use, or overuse, the high subject position in embedded clauses.

Previous studies of subject placement in child production

Very few studies have investigated the acquisition of subject placement in Norwegian, although children's acquisition of optional argument placement has been studied in a number of languages, especially with regards to objects (e.g., Anderssen et al., 2010 on Norwegian; Barbier, 2000; Schaeffer, 2000; Unsworth, 2005 on Dutch; Penner et al., 2000 on German; Mykhaylyk & Ko, 2010 on Ukrainian). The general findings are that, when faced with the option of two object positions, children tend to overuse the low position as compared to the target language.

To our knowledge, there are only two data sets providing information about children's production of subject placement in Norwegian, one based on corpus data from three young children aged 1;8-3;3 (Anderssen, 2006), reported in Westergaard (2008, 2011) and Anderssen and Westergaard (2010), and another based on experimental data from four somewhat older children aged 3;8-5;8, reported in Anderssen et al. (2010). These studies focus on subject placement in main clauses, and the findings show the following: young children make a distinction between pronominal and NP subjects from the beginning of relevant utterances, always placing NPs in the low position, as shown in (3), while pronominal subjects appear either high or low; see examples (4)-(5) (from Westergaard, 2008). However, the proportion of pronouns in the high position is much lower than in adult data, indicating that children have an early preference for the low position, i.e., the one that involves less syntactic movement.

- (3) komte **ikke** *reven* med mæ # i senga mi? (Ina.18, age 2;8.12)
come.PAST not fox.DEF with me in bed.DEF mine
 'Didn't the fox come with me in my bed?'
- (4) no kan **ikke** *han* sove mer. (Ann.10, age 2;3.9)
now can not he sleep more
 'Now he can't sleep any more.'
- (5) og no kan æ **ikke** drikke det. (Ole.19, age 2;10.0)
and now can I not drink it
 'And now I can't drink it'

The developmental data from the corpus show that the distribution of high and low pronominal subjects in main clauses corresponds to adult proportions already around age 2;6-3;0, and the experimental data in Anderssen et al. (2010) confirm that subject placement has been acquired by the older children. Westergaard (2008, 2011) argues that the children's early preference for the low subject position is due to a principle of economy of movement, which has been attested also in other languages, e.g., German, as illustrated with the subject following negation in example (6) from Clahsen et al. (1993/94), or in other linguistic properties of Norwegian, e.g., Object Shift, V2 word order or the word order of possessives (Anderssen et al., 2010; Anderssen & Westergaard, 2010; Westergaard, 2009; Westergaard & Anderssen, 2015).

Table 2. The distribution of pronominal and NP subjects in embedded clauses; corpus data of three children (age 1;8-3;3)

	High position	Low position
Pronominal subjects	9	8
NP subjects	3	4
Total	12	12

- (6) darf **nich** Julia haben. (Mathias, stage II) (German)
may not Julia have
 'Julia may not have that.'
 Target form: Das darf Julia nicht haben.

There is even less data on Norwegian children's production of subject placement in embedded clauses. The only exceptions are Anderssen and Westergaard (2010) and Westergaard (2011), who have investigated the corpus data of the three young children mentioned above (Anderssen, 2006). Given the structural complexity and the general infrequency of embedded clauses with negation (or adverbs), it is to be expected that children below age 3;3 would not produce many examples, and the data are consequently extremely sparse. Altogether, the three children produce only 24 examples, and the distribution of subject types across the two positions, which is displayed in Table 2 (adjusted from Westergaard, 2011), shows the following: the children use both subject positions from early on, the two positions are used with the same frequency (50%, 12/24), and the category of the subject does not seem to play a role, as pronouns and NPs appear in either position (9 vs. 8 for pronominal subjects, 3 vs. 4 for NPs).

Although it is virtually impossible to conclude anything from such meagre data, Westergaard (2011) speculates that these results indicate that the children do make a distinction between main and embedded clauses, in that the distinction between pronominal and NP subjects that is clear in main clauses is not made in embedded clauses, thus following patterns in the input (recall that NP subjects are also to a large extent found high in embedded clauses in the adult language). Furthermore, the lower proportion of subjects in the high position compared to the adult data (50% in child production vs. approximately 87% in adult production) might indicate that children are affected by the principle of economy of movement also in this context. The sparsity of data makes it necessary to conduct more research in order to understand children's behavior with this highly complex and extremely infrequent structure.

Summary and research questions

We have shown the two subject positions in embedded clauses to be an ideal case for testing whether children rely more on input frequency or syntactic complexity in acquisition: the high subject position is by far the more frequent one in adult speech, and thus in children's input, whereas the low subject position is less complex since it involves a shorter move. Furthermore, the two positions are discerned by a distinction

between NP and pronominal subjects, allowing us to study whether and when children are sensitive to this distinction.

As shown above, more data are needed on adult subject placement in embedded clauses in order to get a clearer picture of the adult system. Since there are indications that subject distribution might differ across embedded clause types, we focus on *that*-clauses.

Our first research question is therefore:

RQ1: Where do adults place the subject (high or low) in embedded *that*-clauses, when a) the subject is an NP, and b) the subject is a pronoun?

Furthermore, we are interested in children's placement of subjects, and our second research question is:

RQ2: Which subject position (high or low) do children aged 3-6 years use in embedded *that*-clauses?

Related to this question we are particularly interested in knowing

- a) whether children distinguish between pronominal and lexical NP subjects, and
- b) whether children's production varies as a function of age, and at what age the adult subject distribution is acquired.

Even though both main and embedded clauses allow word order variation with subject placement, restrictions on their use differ. Children must therefore pay attention to clause-specific input to learn the appropriate generalizations. Above, we described previous studies showing children's overuse of the low position for subject placement in main clauses. The tentative findings for embedded clauses (based on scarce data, Anderssen & Westergaard, 2010; Westergaard, 2011) showed that children do not distinguish subject types and positions in embedded clauses. This means that children's production is different from the input and indicates that they may have a preference for the low position also in embedded clauses, although at a later stage. It also provides some evidence that children do make a distinction between main and embedded clauses, using the two subject positions differently in the two clause types. In main clauses, children were found to have an adult-like distribution of the two subject positions at age 2;6-3;0, distinguishing between subject types. Since embedded clauses add an extra level of complexity, and since they are more infrequent in the total amount of input, we hypothesize the following:

H1: Subject placement in embedded clauses is more difficult to acquire than in main clauses, so children are likely to be older than 3 years when subject placement in embedded clauses is target-like.

In main clauses, children were found to distinguish between pronominal and lexical NP subjects from early on. However, regarding the distinction between subject types in embedded clauses, we hypothesize the following:

H2: Children at age 3 distinguish between pronominal and lexical NP subjects as a syntactic category, but their input in embedded clauses is likely so complex and/or infrequent that they will not use the two subject types according to the target grammar. (If children do distinguish pronouns and NPs we expect that they will distribute subjects of these two types differently across the two subject positions, pronouns high and NPs low).

The answer to this descriptive question (RQ2) should ultimately contribute to answering the following research question:

RQ3: How can children's subject placement be explained?

If children's production deviates from the target grammar, we focus on two hypotheses for RQ3, related to input frequency and syntactic complexity:

H3: If children are guided mainly by input frequencies when acquiring subject placement, we expect them to initially overuse the high subject position as compared to the target grammar.

On the other hand:

H4: If children are guided mainly by principles of economy (less movement and syntactic structure) when acquiring subject placement, we expect them to initially overuse the low subject position as compared to the target grammar.

Study 1: Adult production

We investigated spontaneous speech in three large corpora of adult speech: the Big Brother corpus (Tekstlab, 2009), the Norwegian part of the Nordic Dialect Corpus (NorDiaCorp) (Johannessen et al., 2009) and NoTa (Tekstlab, 2004). These corpora are all online resources found at the Textlab hosted by the University of Oslo. The Big Brother corpus consists of transcriptions from almost 100 episodes of the TV show Big Brother, comprising approximately 440,300 words. The NorDiaCorp and NoTa corpora both consist of transcribed recordings of dialogues, the former comprising 438 informants and 1,997,920 tokens, the latter comprising 166 informants and 957,000 words.² Given the aggregated size of these corpora, the large span of demographic variation, and the variety of dialects and speech situations, we consider them representative of Norwegian (following recommendations in e.g., Stefanowitsch, 2020, p. 28ff). As mentioned above, different clause types allow low subject placement to a varying extent. To obtain clear generalizations for one clause type, with the possibility of later comparisons with other clause types, we only searched the corpora for embedded clauses with the complementizer *at* 'that', a clause type known to allow low subject placement (e.g., Faarlund et al., 1997, p. 891).

The three corpora can be searched for part of speech, and across the three corpora, we used the following search strings: *at* 'that' + *ikke* 'not' + noun/pronoun + verb and *at* 'that' + noun/pronoun + *ikke* 'not' + verb. In the search we defined that no element could occur between the specified lemmas. For the sake of coherence, we restrict our study to syntax, and use syntax as a proxy for subjects' function and information structure. More detailed studies of the subjects' pragmatic and information structural properties could yield further insights of what triggers the two subject positions.

²The three corpora have been utilized separately in previous studies addressing low subject placement: the Big Brother corpus in Garbacz (2005), ScanDiaSyn (NorDiaCorp) in Johannessen and Garbacz (2011) and NoTa in Westergaard (2011). In this study we combine data from the three corpora to get higher representativity of dialects and speech situations, and unlike these previous studies we isolate only one embedded clause type.

Results

Our search returned a total of 793 relevant utterances from the three corpora.³ The high subject position made up 84% (N=665) of these, meaning that 16% (N=128) had the subject in the low position. These numbers are similar to findings in Garbacz (2005), Anderssen and Westergaard (2010) and Westergaard (2011). NPs made up 7.4% (59/793) of all subjects. Pronouns were overwhelmingly found in the high position (87%, 638/734), while NPs were more evenly distributed, 54% (32/59) in the low and 46% (27/59) in the high position. An overview of these findings is given in Table 3.

In the background section we saw that the low subject position is often assumed to be reserved for subjects that are new or focused. We checked if this included obligatory prosodic stress, by using the researchers' perception through listening through the 32 low NP subjects and 96 low pronominal subjects. Emphasis was found on 10 NP subjects and 11 pronominal subjects, while five and seven occurrences for NPs and pronouns respectively were unclear with regards to stress. This shows that the minority of low subjects have prosodic emphasis, aligning with findings in Johannessen and Garbacz (2011).

The general pattern of subject distribution outlined in the background section (given and specific subjects in the high position, new and non-specific subjects in the low position) predicts that we should find definite NPs high and indefinite NPs low. This prediction was not borne out either. Of the 32 NP subjects in the low position, only six were indefinite (this included two quantifiers, *alle* 'everyone', three of the generic plural *folk* 'people' and one indefinite plural, *fler kostymer* 'more costumes'). Four were proper nouns and 22 were definite, a surprisingly high number. We further examined the definite NPs, and found that, unexpectedly, 11 of them had a specific reading, e.g., *datteren min* 'my daughter'. For two of the definite NPs, the specificity was not clear. The distribution of the NP subjects is shown in Table 4.

Of the 27 NP subjects in the high position, only ten were definite, three were proper nouns and 13 were indefinite. Of the definite NPs, one was clearly generic, as it was referring to a kind of animal (*hjorten* 'the deer'). Seven had specific reference and two were

Table 3. The distribution of pronominal and lexical NP subjects in high and low subject positions broken down for each corpus

Corpus	Subject type	Neg-S	S-Neg	Total
BigBrother	NP	37% (N = 7)	63% (N = 12)	19
	Pron	20% (N = 44)	80% (N = 176)	220
NoTa	NP	42% (N = 5)	58% (N = 7)	12
	Pron	6% (N = 11)	94% (N = 183)	194
NorDiaCorp	NP	71.5% (N = 20)	28.5% (N = 8)	28
	Pron	13% (N = 41)	87% (N = 279)	320
Total		16% (N = 128)	84% (N = 665)	793

³Some of the utterances involve a complex complementizer such as *fordi at* 'because that' and *sånn at* 'so that'. Crucially, the complementizer *at* 'that' is always included. These utterances are included on the assumption that the presence of *at* ensures uniformity of the word order phenomenon under scrutiny.

Table 4. Detailed overview of the type of NP subjects in the low (Neg-S) and the high (S-Neg) position. Specificity could not be determined for all definite NPs.

Type of NP	Neg-S	S-Neg
Proper noun	4	3
Definite	22	10
Generic	9	1
Specific	11	7
Indefinite	6	13
Total	32	27

unclear with respect to specificity. Among the indefinite NPs in the high position, we found one occurrence of a quantifier (*noen* ‘some’), two of regular plurals (*attenåringer* ‘eighteen-year-olds’), and nine of the generic plural *folk* ‘people’.⁴ Lastly, the weight of the NP subjects (measured by the number of syllables) was evenly distributed across the two positions. Thus, there is no clear distinction between the two subject positions with respect to definiteness/specificity or weight.

Summing up, our corpus investigation of adult production shows that low subjects are only used in 16% of *that*-clauses. Pronominal subjects are found in the high subject position in 87% of occurrences, reflecting numbers from previous studies (Westergaard, 2011). NP subjects are almost evenly distributed between the high and the low position. The restriction on the low position is not clear. Low pronominal subjects are not always stressed, and apart from the clear tendency that pronominal subjects are high, our data do not show a correspondence between the information value of different subject types and their placement in the syntactic structure. The distribution of NP subjects shows considerably more variation than previously assumed. The somewhat blurry picture of subject patterns in embedded clauses indicates that the most reliable observation is the distinction between subject types (also pointed out by Westergaard, 2011). Consequently, we must assume that what children rely on when acquiring embedded clause subject placement is the syntactic category of the subject.

Study 2: Experimental study

In the experimental study we used an elicited production task to collect spontaneous production of subject placement relative to negation in embedded clauses.

Participants

We recruited and tested a total of 41 children. Two children turned out to be bilingual and were excluded from data analysis. Six children (age 3;6-4;1) were excluded due to lack of relevant responses (no embedded clauses). The results are thus based on 33 typically

⁴Since the NP *folk* ‘people’ appears so frequently as a subject in our material, we should note that its occurrence is attested across all three corpora and in different situations, so it is not the case that it is repeatedly mentioned in the same speech situation.

developing monolingual children (age 3;1-6;1) raised in the city of Trondheim, Norway, and acquiring the local dialect. The children were recruited through daycare centers and schools in the area. Ten adult controls (age 20-28) also participated in the tasks. These were all speakers of the local dialect, recruited through social media and acquaintances.

Methodology

Our experimental tasks were based on a ‘shy puppet’ design (Crain & Thornton, 1998). The children were shown a picture and heard a pre-recorded short story about two popular children’s book characters, Karsten and Petra. Pre-recording the stories ensured that all participants received identical stimuli. Before starting the experiment, the children were introduced to a toy turtle, and told that this turtle wanted the children’s help to remember what happened in the stories. The turtle would not talk to adults, and therefore the children were asked to help answering his questions. The procedure of the lead-in stories and questions is shown in (7), eliciting an NP, and in (8), eliciting a pronoun.⁵

- 7) **Voice-over:** Petra is being watched by her baby-sitter. The baby-sitter often cooks meat loaf for dinner.

I dag lagar ho ikkje kjøttkake
Today makes she not meatloaf

‘Today she does not make meatloaf.’

Petra is happy about that.

Turtle asks: What is Petra happy about?

Target answer:

Petra er glad for at (ikkje) barnevakta (ikkje) lagar kjøttkake.

Petra is happy for that (not) babysitter.DEF (not) makes meatloaf

‘Petra is happy that the baby-sitter isn’t making meatloaf’

- 8) **Voice-over:** Miss Bunny (Frøken Kanin) tasted lion food once. She thought the taste was awful.

Ho et ikkje løvemat til frukost
She eats not lionfood for breakfast

‘She does not eat lion food for breakfast’.

She is happy about that.

Turtle asks: What is Miss Bunny happy about?

Target answer:

Frøken Kanin er glad for at (ikkje) ho (ikkje)

Miss Bunny is happy for that (not) she (not) eats

et løvemat til frukost.

lionfood for breakfast

‘Miss Bunny is happy that she doesn’t eat lion food for breakfast’

⁵The AE raises the question of whether the word order in the lead in stories could influence the word order in children’s (particularly producing only S-Neg) elicitation. All relevant context orders are Subj-Verb-Neg (with one exception containing topicalization: XP-Verb-Subj-Neg, in (7)). Since the preceding word order (generally) contains a verb between the subject and negation, we think it is unlikely that children’s production is a result of direct priming.

The child, the experimenter, and the turtle could all see the computer screen at all times. The intro prompts were presented to all those who took part in the situation, so when the puppet asked the participant a question, all information was already known to everyone present. This ensured that all contexts were identical with respect to the familiarity of the subject.

Task design and analysis

The task contained 12 experimental items in addition to 6 filler items. The test items were *that*-clauses with negation, whereas the filler items were *that*-clauses with various adverbs. The test items were grouped into two conditions by subject type: NP and pronoun. To avoid children pointing to the screen, and instead using pronominal subjects referentially, all pronominal subjects were co-referent with the matrix subject (MS) – as in (8) where *she* (embedded subject, ES) = *miss Bunny* (MS). The setup was such that replacing NP subjects by a pronoun would be odd: NP subjects were not co-referent with matrix subjects, as in (7) where *Petra* (MS) ≠ *babysitter* (ES). Replacing NPs with pronouns would require pointing to the screen, which was placed out of reach of participants, to resolve ambiguity, and it would not be pragmatically consistent with the puppet's wish of a reminder, as it would presuppose knowledge of who the pronoun referred to.

Certain predicates facilitate a third possible word order not of relevance to the present study (S-Verb-Neg). To avoid too many responses involving this word order, the matrix predicate was not one known to freely allow V-Neg (e.g., assertive predicates) (see e.g., Wiklund et al., 2009). The matrix predicates were all assumed to be familiar to children in the age group: *glad for* 'happy about', *passe på* 'watch (out)', and *lei seg for* 'sad about'. The age appropriateness was ensured by checking the lexical database 'Norwegian Words' (Lind et al., 2015). In order to avoid contrastive embedded subjects, the lead-in story always revolved around an action rather than a person, making it felicitous to negate the predicate rather than the subject, and there were no constituent negations in the contexts, thus no comparisons.

Participant responses were coded as (a) S-Neg, (b) Neg-S, or (c) other. The latter category comprised responses with the embedded clause word order Verb-Neg (N=35), responses with a different complementizer (N=23), as well as failures to respond with a subordinate clause (N=69) or at all (N=27).

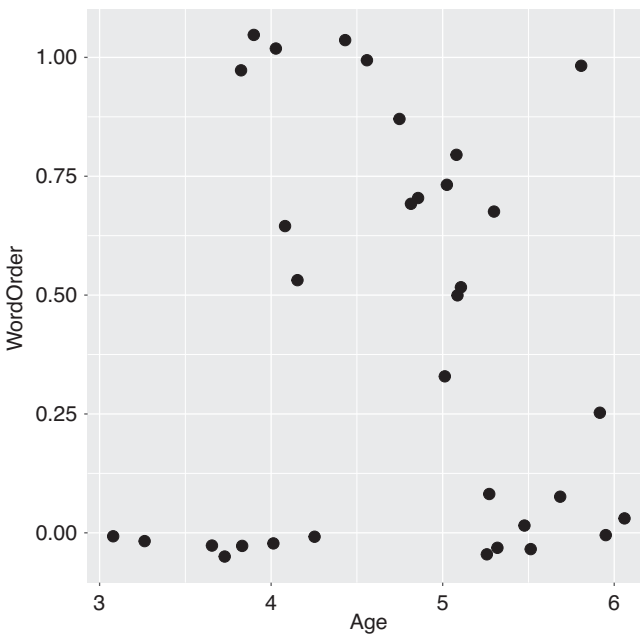
Results

Adults produced relevant word orders on 109 of 120 trials. All 109 word orders were of the type S-Neg, meaning that adults categorically produced subjects in the high position. They used both pronominal (N=44) and NP subjects (N=65). Children produced relevant word orders on 275 of 396 trials and also used both pronominal (N=160) and NP subjects (N=115). Child participants used the high subject position slightly more than the low position (153 vs. 122 occurrences). The results (Table 5) show that children used the low subject position more often with NPs than pronouns: while 59% of NP subjects appear in the low subject position, only 34% pronominal subjects appear low.

Figure 2 shows how the production of subject placement is distributed for all 33 children. In the plot, each dot corresponds to one participant's total proportion of low subjects for both subject types combined. Nineteen children are categorical in their

Table 5. The percentage of low subjects (Neg-S word order) in embedded clauses by adult and child participants. The remainder of the items are S-Neg

		Neg-S
Adults	NP	0% (0/65)
	Pronoun	0% (0/44)
Children	NP	59% (68/115)
	Pronoun	34% (54/160)

**Figure 2.** The proportion of Neg-S (100% Neg-S = 1.00, 100% S-Neg=0.00) in each child participant's production by age.

production – some only use the low subject position (proportion=1.00), whereas some only use the high position (proportion=0.00). Fourteen children use both subject positions. Figure 3 shows a similar plot, but here each participant's total proportion of low subjects is shown for each subject type (NP and pronoun). The plot reflects the findings in Table 5, with more NP subjects in the low position and more pronominal subjects in the high position. Additionally, the plot shows that the slightly older children use the high subject position almost exclusively with pronouns, whereas the low subject position is used with both subject types, suggesting some knowledge of the adult pattern. From around age 5, children only use the low subject position with NP subjects. In what follows, we examine the production of the children with categorical production, before addressing the production of the children who use both word orders.

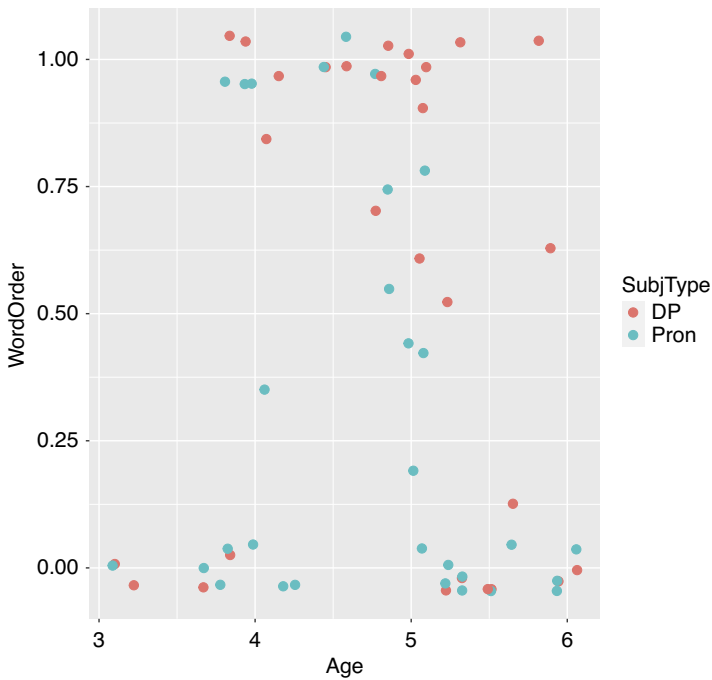


Figure 3. The proportion of Neg-S (100% Neg-S = 1.00, 100% S-Neg=0.00) in each child participant's production by age and subject type.

Children with categorical production

In this section we examine the responses of the 19 children who categorically produce only one word order, either Neg-S or S-Neg. Table 6 shows the number of each participant's response with either high or low subjects, the subject type used, and each participant's total responses to the experimental items.

Figure 3 showed that a group of the youngest participants categorically use the high subject position (S-Neg). This group of children consists of seven participants (age 3;1-4;3), and their aggregated production comprises 28 occurrences. The response rate in this youngest group is quite low, likely an effect of the task being challenging for them.

The next categorical group of children is slightly older. In contrast to the youngest children, these only use the low subject position, Neg-S. The Neg-S children comprise a group of six participants, where five are aged 3;10-4;7, with an age gap up to the oldest participant (5;10). This slightly older participant seemed to struggle with embedded clauses and only produced three usable responses in the task. This group's accumulated production of low subjects is 44 occurrences. Lastly, there is a group of six older categorical children (age 5;3-6;1). Like the youngest categorical children, these children produce only high subjects (N=57). The high number of responses combined with the use of only the high position seems to reflect a target-like state. In general, children with categorical production used both pronominal and NP subjects, reflecting that their use of a single word order does not stem from them using only one subject type. A few participants used only one subject type (P4, P6 P7 and P10: pronominal subjects, P2,

Table 6. Categorical children's production in raw numbers, grouped by age: 3;1-4;3 only S-Neg, 3;10-5;10 only Neg-S, and 5;3-6;1 only S-Neg

Participant	Age	N of S-Neg	N of Neg-S	Subject type, NP/Pronoun	Total responses
P1	3;1	3		1/2	8
P2	3;3	4		4/0	9
P3	3;8	6		3/3	7
P4	3;9	1		0/1	4
P5	3;10	11		2/9	11
P6	4;0	1		0/1	2
P7	4;3	2		0/2	4
P8	3;10		9	4/4	9
P9	3;11		11	6/5	11
P10	4;0		3	0/3	9
P11	4;5		10	3/7	10
P12	4;7		8	1/7	8
P13	5;10		3	3/0	3
P14	5;3	11		4/7	12
P15	5;4	11		4/7	11
P16	5;6	12		6/6	12
P17	5;6	2		2/0	2
P18	5;11	10		2/8	10
P19	6;1	11		5/6	11

Note. The remaining productions included in the column 'total responses' were of a third possible word order (Verb-Neg), not under consideration in this study. Participant P8 produced one utterance with an unclear subject, thus only 8 subjects were counted even though 9 relevant examples were produced.

P13 and P17 NP subjects). These participants displayed an overall low number of relevant responses, and the use of one subject type is therefore likely to be random.

Children who use both subject positions

A total of 14 children (age 4;1-5;11) produced altogether 147 relevant utterances, using both subject positions. They also seemed to have a slight preference for the low one (79 low vs. 68 high subjects). NP subjects were used more in the low position than pronominal subjects (51 vs. 28), and pronominal subjects were used more than NP subjects in the high position (54 vs. 14); see Table 7.

To investigate the possible effects and interactions of subject type and age on subject placement within this group of children, we used a mixed effect logistic model with the glmer function in the lme4 package (Bates et al., 2015) in R, version 1.3.959 (R Core Team, 2018). In the model, fixed effects were subject type, participant age (scaled), and their interaction. The model included random intercepts for items and participants. The maximal converging model further included random slopes for subject type by participant,

Table 7. Overview of the production of the 14 children who produced both word orders.

	S-Neg	Neg-S	Total
NP	14 (21.5%)	51 (78.5%)	65
Pronoun	54 (66%)	28 (34%)	82
SUM	68 (46%)	79 (54%)	147

Table 8. Results of the logistic mixed effects model used to analyze the production of children using both word orders. Fixed effects: Subject type, participant age and their interaction.

Call: glmer(WordOrder ~ SubjType*scale(Age) + (1 +SubjType Participant) + (1+ Item)			
	Estimate	z-value	p-value
Subject type	-3.458	-3.113	0.00185 **
Age	-2.065	-2.139	0.03247 *
Subject type:Age	-0.296	-0.289	0.76652

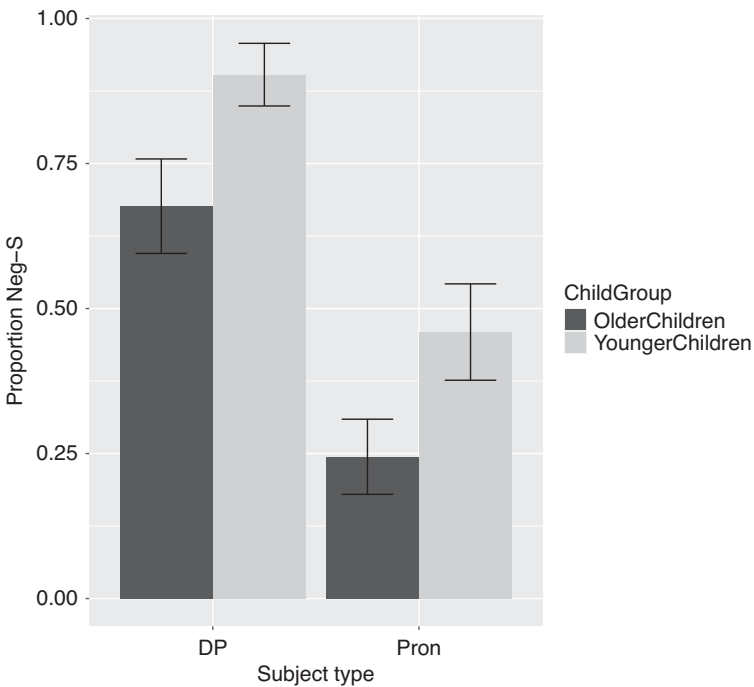


Figure 4. Proportion of Neg-S (low subject) by younger children (<4;11) and older children (≥5) as a function of subject type.

but not for subject type by item, as the high collinearity between these two variables (0.99) caused the model to not converge. Our model showed a significant main effect of subject type on subject position, such that Neg-S was produced significantly more with NP subjects than with pronominal subjects ($z = -3.113, p < .01$). Age was also found to be a significant predictor of subject placement ($z = -2.139, p < .05$), in that the production of Neg-S decreased with age. There was no Subject Type \times Age interaction, so older children's production of word order by subject type was not significantly different than younger children's production. A likelihood ratio test of our maximal model against simplified models without subject type and age confirmed the significance of subject type ($p < .005$), but not of age ($p = .07$).

Since children's subject placement seems to develop over time, we wanted to probe further into younger vs. older children's patterns. We therefore split the 14 children into two groups by the mean age, younger children ($\leq 5;0, N = 7$) and older children ($\geq 5;1, N = 7$). Figure 4 demonstrates how younger children use the low subject position more than older children, 66% (45/68) vs. 43% (34/79). Both groups use Neg-S more often with NP subjects than pronominal subjects.

Summarising these results, adults only use the high subject position, whereas children use both positions almost to an equal extent, but the low position more than the high. Some children categorically use only one of the subject positions. Younger and older children only place subjects high, and 'middle' children only place subjects low. Furthermore, some children use both subject positions, and NP subjects are more often placed low than pronominal subjects. Children start categorically using only the high position around age 5. This is identical to adults' performance in this task. Also around age 5 children using both word orders start using the low position for NP subjects only, a pattern resembling that of adults in spontaneous speech. Thus, both categorical production of high subjects and production of both positions with only NP subjects low may be considered target-like, since the former matches adult production in the elicitation task and the latter matches adult production in spontaneous speech. On this view, children seem to have acquired subject placement in embedded clauses around age 5.

Discussion

This paper concerns the influence of complexity and frequency on children's acquisition of two different subject positions in embedded clauses in Norwegian. In the background section we asked three research questions, repeated here. In what follows we give a brief overview of the results our studies returned and discuss these results with reference to our research questions.

RQ1: Where do adults place the subject (high or low) in embedded *that*-clauses, when a) the subject is an NP, and b) the subject is a pronoun?

RQ2: Which subject position (high or low) do children aged 3-6 years use in embedded *that*-clauses?

Related to this question we are particularly interested in knowing

- a) whether children distinguish between pronominal and lexical NP subjects, and
- b) whether children's production varies as a function of age, and at what age the adult subject distribution is acquired

RQ3: How can children's subject placement be explained?

Adults' subject placement

Results from the three corpora showed that the low subject position in embedded clauses was only used 16% of the time. Pronominal subjects were predominantly used in the high position (87%) by adults, aligning with previous findings (Westergaard, 2011). NP subjects were almost evenly distributed across the two positions (46% in the high position). This is in between the proportions from previous findings (cf. Table 1, with 35.3% and 73.7% NP subjects in the high position, notably based on very sparse data), indicating that there might be considerable variation across speakers with respect to the position of NP subjects in embedded clauses. Adding to this observation was our finding that both definite and indefinite, specific and generic NPs, were used in both subject positions. Consequently, current results are inconclusive with respect to the factors responsible for the subject distribution in embedded clauses, and children's input thus cannot be said to display a clear pattern, with the exception of the tendency of pronominal subjects appearing in the high position.

Next, we performed an elicitation task with children and adults. Here, adult production patterned differently than in the corpus investigation. In the elicitation task, adults placed all subjects in the high position regardless of subject type. This might be explained in several ways. Firstly, the choice of only one word order could be a task effect. Adults are known to be prescriptive in test situations, consciously or subconsciously using what they believe to be the 'correct' grammar (e.g., Cornips & Poletto, 2005). Since the high subject position is by far the more frequently used in embedded clauses, it might be perceived as 'correct'. Secondly, throughout the experiment, the familiarity of the subject was kept constant. If this factor is more likely to facilitate use of the high subject position, adults could be adhering to a pragmatic principle in their production. Thirdly, adult participants could be self-priming and therefore producing only this word order. We therefore assume that adults' production pattern from the corpus investigation is more representative of the type of input children typically receive in spontaneous speech.

Children's subject placement

In the elicitation task, the children (as a group) used both subject positions and also distinguished between pronominal and NP subjects by more often using the high position for pronominal subjects. The children used the low position slightly more than the high position, and younger children used the low subject position significantly more than older children. Some children, among the youngest and the oldest participants, categorically used the high subject position. A group of children age-wise in between, only used the low position. This means that our findings differ from observations from previous research (Anderssen & Westergaard, 2010; Westergaard, 2011) that found that young children used both subject positions in embedded clauses to an equal amount. This discrepancy is likely due to these studies investigating only very young children (up to age 3) and thus containing few relevant clauses. Furthermore, our finding that children producing both subject positions use the low position more than adults aligns well with findings from studies on subject placement in main clauses, as well as with the vast literature on children's preference for low object positions (e.g., Anderssen et al., 2010; Mykhaylyk & Ko, 2010; Schaeffer, 2000).

Explaining children's subject placement

In the background section we entertained two hypotheses for what could be guiding children's acquisition of the two subject positions: input frequency and syntactic

complexity. We noted that the high position was the more frequent one in children's input, whereas the low position was less complex, and that children's preference of subject placement would indicate whether frequency or complexity was the chief factor guiding acquisition. In the following sections, we first address the preference for low positions, then discuss a U-shaped acquisition pattern, and finally make some observations on frequency.

Before moving on, a note on children's ability to distinguish pronominal and lexical NP subjects is in order: as pointed out by one reviewer, our results speak to research showing that referring expressions are harder for children to process when they are realized lexically than as pronouns (see e.g., Arnon, 2010 for children's comprehension of pronominal/lexical objects in relative clauses). It is therefore possible that children are able to produce pronominal subjects in the complex position earlier than lexical subjects, since they have a lower processing load than lexical subjects.

Preference for low positions

The general finding that children use the low subject position to a much higher extent than adults, despite it being highly infrequent in their input, seems to reflect some intrinsic bias towards low positions. Since this is the structurally less complex position, it is plausible that this is caused by children being economical, not moving elements further than necessary (Westergaard, 2009). If children use both subject positions, but overuse the low position as compared to adults, it suggests that *i*) they have not acquired the full details of the variation and therefore assume the low position to be suitable more often than is the case, and *ii*) they have not encountered a sufficient amount of the relevant variation in their input for the bias towards low positions to be overridden. Two additional factors point towards children having a preference for the low position: *i*) the acquisition trajectory of subject placement in embedded clauses vs. main clauses, and *ii*) some children displaying categorical production. In the following we discuss these points in the mentioned order.

A comparison of our experimental data to children's acquisition of subject placement in main clauses shows that the acquisition trajectories in both clause types seem fairly similar: as shown by previous studies (see background section), children initially place all NP subjects in the low position in main clauses, while varying their placement of pronouns (although pronominal subjects are more often used in the low than in the high position). Thus, both in main and embedded clauses, children use the low position more than what is found in the target language, and in both clause types children more often use the low position for NP subjects than pronominal subjects. Additionally, the initial occurrences of subjects in both clause types are in the high position, before the onset of highly frequent use of the low position: in the longitudinal data of three children reported in Anderssen et al. (2010), Anderssen and Westergaard (2010), and Westergaard (2008, 2011), the first relevant utterances with high subject placement are actually attested before the first utterances with low subjects (although note that results from main clauses come from longitudinal data, whereas our data on embedded clauses are cross-sectional).

Importantly, at an age where children have acquired the distribution of subjects in main clauses (around age 3, Anderssen et al., 2010), they continue to overuse subjects in the low position in embedded clauses. We take this to mean that children must first entertain an hypothesis for subject placement in main clauses, and subsequently (since embedded clauses are later acquired in general) a similar hypothesis for subject placement

in embedded clauses. Thus, even though the syntactic positions for subjects are in principle the same for both clause types, this finding highlights how children seem to be aware that different restrictions and generalizations for their use apply. An initial hypothesis for each clause type seems to be that the low position should be used more than what is found in the target language.

A U-shaped acquisition pattern

We now turn to the children who have a categorical production. The categorical production of high subject placement is made by the youngest participants (3;1-4;3) and the oldest participants (5;3-6;1), whereas the children producing only low subjects are age-wise in between these two groups (3;10-4;7). This resembles a U-shaped learning pattern. The best-known example of U-shaped development is children's acquisition of past tense morphology of irregular verbs (Marcus et al., 1992; Pinker, 1999), where children start out using target-like versions of past tense, e.g., *went*, before transitioning through a phase of overgeneralizing the regular rule, *goed*, and finally converging on the target form *went*. The categorical children in our study have a production comparable to this, since it seems they can be grouped into three stages. Stage 1 would correspond to the youngest children, using only S-Neg, Stage 2 is made up of the 'middle' children using only Neg-S, and Stage 3 is older children using S-Neg.⁶

Although children at Stages 1 and 3 both have categorically high subject placement, we have some reason to assume that only the older children have reached an adult state and settled on a rule for subject placement in embedded clauses, but that this is not the case for the younger children: the older children produce target-like responses in (close to) 100% of all possible items in the experiment, while the younger children in general produce far fewer.

The younger children's production of only high subjects might reflect two interacting factors: *i*) the extreme infrequency of low subjects in embedded clauses in their input: it is possible that the children at Stage 1 have not encountered (sufficient) examples of this subject position, and therefore are not aware that this is a possibility in the target grammar, *ii*) the youngest children might have a less detailed syntactic structure than the target language requires, with only one subject position, thus following a route predicted by one part of the structural economy principle of Westergaard (2009, p. 216), that children 'only build as much structure as there is evidence for in the input'. In this case, children will have developed a structure where the low subject position (SpecTP) is the only position for subjects, combined with the low negation (adjoined to VP), which gives the correct surface word order but not the more complex structure found in the target grammar.

At Stage 2, children seem to move through a phase of re-analysis of the non-target-like hypothesis. We would argue that the reanalysis is caused by children having encountered the low subject in their input enough to entertain it as a possibility in their target grammar. Thus, it is plausible that children at Stage 2 use the low subject position (SpecTP), combined with the high negation, yielding the Neg-S order, as in the target grammar. Further, we would argue that children at Stage 2 have expanded their analysis to

⁶Since our study does not contain longitudinal data, in the following we use age as a proxy for individual children's development when discussing possible causes of this U-shaped pattern. In order to investigate this properly, longitudinal data of embedded clause subject placement should be collected in a future study.

include the two subject positions, but that they prefer to use the low one, as it is more economical. Again, this development follows a route predicted by the second part of the structural economy principle (Westergaard, 2009 p. 216), that children 'only move elements as far as there is evidence for in the input'. Whereas earlier findings of children's preference of low argument positions have been explained by economy of movement, to explain the U-shaped development both parts of the economy principle need to be invoked, i.e., economy of movement and economy of structure building.

Our suggested analysis of the U-shaped development reflects a conservative learner, gradually building and sophisticating syntactic structure: it seems children at Stage 3 have the full target-like structure involving two subject positions, and the different features associated with each, presumably as a result of sufficient exposure to both positions and an adult-like mapping of features and position.

We finally note that there seems to be some individual variation in the use of subject positions, since some children use both positions and others use only Neg-S or S-Neg. As noted with adult speakers' production, categorical responses may be the result of a task effect or self-priming.

Frequency vs. complexity

Children overuse the low subject position even though they have an overwhelming amount of evidence that the high position exists and is the more common one in their target language. Our corpus data show that the ratio of high vs. low subject placement in embedded clauses is 84% to 16% in the adult language. This observation indicates the strength of the principle of economy: even when faced with over 80% high subjects, children's tendency of structural economy lingers for some time. Children's behaviour thus reflects that what is less complex is preferred over what is more frequent.

Our findings also contribute some indication of thresholds in input frequency: for the young categorical children who only use the high subject position, we suggest that the proportion of low subjects in embedded clauses is so small that these young children have not had enough relevant material to hypothesize over. In addition, the relatively late acquisition of embedded subject placement might be an effect of the general infrequency of embedded clauses with negation, the structure that is necessary to disambiguate the two positions.

Conclusion

In this paper we have examined children's (age 3;1-6;1) and adults' production of subject placement in embedded clauses in Norwegian. Two subject positions are available in such clauses, one preceding and one following negation (high and low positions). The low position is less complex, while the high position is more frequent in children's input. The study investigated whether children are able to discover these two subject positions and distribute subjects across them in a target-like manner, with the aim to distinguish between two important approaches to L1 acquisition: one grounded in input frequency and the other in complexity.

Our first research question (RQ1) regarded adults' production of subject placement in embedded clauses, which we investigated through spontaneous and elicited production. In spontaneous speech, adults were found to almost always place the subject high (84%). Furthermore, pronominal subjects are almost only used in the high position, whereas NP

subjects are evenly distributed across the two positions, with considerable variation in their internal properties. The somewhat unclear function of subjects in the high and low positions, in addition to the infrequency of low subjects in children's input, was predicted to make acquisition challenging for children, and RQ2 concerned children's elicited production of subjects in embedded clauses. The answer to RQ2 is that children generally use the low subject position considerably more than adults, but more so with NP than pronominal subjects, demonstrating a systematicity resembling that of the target language. Additionally, a group of children display U-shaped development, where the youngest children only use the high subject position, slightly older children only the low position, and the oldest children only the high position. Children seem to acquire subject variability in embedded clauses around age 5.

Our third research question (RQ3) asked how to explain children's subject placement. Since we find that children use the low subject position to a much higher extent than adults, despite the infrequency of this position in their input, we suggest that children display an intrinsic bias towards less complex structures. We suggest that children's U-shaped development can be explained by a two-part principle of economy, where they adhere to economy of movement and economy of structure building. We also noted that the extreme infrequency of low subjects in embedded clauses in the total amount of children's input is likely to delay some children's onset of the Neg-S word order; thus, there may be a frequency threshold that could explain some young children's early categorical behavior.

Acknowledgments. We thank Kristin Melum Eide, Dave Kush, Ingrid Bondevik, members of the AcqVA community, two anonymous reviewers, the journal's action editor, Ana Santos, and the audience at MONS 18 (2019) for their valuable help and comments at various stages of this research.

Competing interest. the authors declare none.

References

- Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*, 239–273.
- Anderssen, M. (2006). *The Acquisition of Compositional Definiteness in Norwegian*. (Doctoral dissertation). University of Tromsø.
- Anderssen, M., Bentzen, K., Rodina, Y., & Westergaard, M. (2010). The acquisition of apparent optionality: Word order in subject and object shift constructions in Norwegian. In M. Anderssen, K. Bentzen and M. Westergaard (Eds.), *Variation in the Input: Studies in the Acquisition of Word Order*, 241–270. Dordrecht: Springer Netherlands.
- Anderssen, M., & Westergaard, M. (2010). Frequency and economy in the acquisition of variable word order. *Lingua* *120*(11), 2569–2588. doi: 10.1016/j.lingua.2010.06.006
- Arnon, I. (2010). Rethinking child difficulty: The effect of NP type on children's processing of relative clauses in Hebrew. *Journal of Child Language*, *37*(1), 27–57. doi:10.1017/S030500090900943X
- Barbier, I. (2000). An experimental study of scrambling and object shift in the acquisition of Dutch. In S. M. Powers & C. Hamann (Eds.), *The acquisition of scrambling and cliticization*, 41–69. Dordrecht: Springer Netherlands.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01.
- Bentzen, K. (2009). Subject positions and their interaction with verb movement. *Studia Linguistica* *63*(3), 261–291. doi: 10.1111/j.1467-9582.2009.01162.x
- Bobaljik, J. D., & Jonas, D. (1996). Subject Positions and the Roles of TP. *Linguistic Inquiry* *27*(2), 195–236.
- Brandtler, J. (2008). On the Structure of Subordinate Clauses in Swedish. *Working Papers in Scandinavian Syntax*, Vol 81. Lund University.

- Cardinaletti, A.** (2004). Towards a cartography of subject positions. In L. Rizzi (Ed.), *The structure of CP and IP: The cartography of syntactic structures*, Vol. 2, 115–165. New York: Oxford University Press.
- Clahsen, H., Penke, M., & Parodi, T.** (1993/94). Functional categories in early child German. *Language Acquisition* 3(4), 395–429. doi: [10.1207/s15327817la0304_3](https://doi.org/10.1207/s15327817la0304_3)
- Clark, R., & Roberts, I.** (1993). A Computational Model of Language Learnability and Language Change. *Linguistic Inquiry* 24(29), 299–345.
- Cornips, L., & Poletto, C.** (2005). On standardizing syntactic elicitation techniques (part 1). *Lingua* 115(7), 939–957. doi: [10.1016/j.lingua.2003.11.004](https://doi.org/10.1016/j.lingua.2003.11.004)
- Crain, S., & Thornton, R.** (1998). *Investigations in Universal Grammar: A guide to experiments on the acquisition on syntax*. Cambridge, MA: MIT Press.
- Eide, K.** (2002). Adjunction sites for negation in Norwegian: Modals and negation. *Nordic Journal of Linguistics* 25(2), 225–252.
- Faarlund, J. T., Lie, S., & Vannebo, K. I.** (1997). *Norsk referansegrammatikk [Norwegian Reference Grammar]*. Oslo: Universitetsforlaget.
- Garbacz, P.** (2005). Bisatsstrukturer som ‘man kan inte’ bortse från (om bisatsorfoeljd i de fastlandsskandinaviske språken). [Embedded clause structures which cannot be ignored (on embedded clause word order in the Mainland Scandinavian languages)]. Ms. University of Lund. Retrieved from <[https://portal.research.lu.se/portal/sv/publications/bisatsstrukturer-som-man-kan-inte-bortse-fracan-om-bisatsorfoeljd-i-de-fastlandsskandinaviske-spraaken\(c2c71d2c-6bf4-4fc6-92ca-3e5402f4ca6a\)/export.html](https://portal.research.lu.se/portal/sv/publications/bisatsstrukturer-som-man-kan-inte-bortse-fracan-om-bisatsorfoeljd-i-de-fastlandsskandinaviske-spraaken(c2c71d2c-6bf4-4fc6-92ca-3e5402f4ca6a)/export.html)>
- Garbacz, P.** (2014). Negation-Subject Embedded Word Order. *Nordic Atlas of Linguistic Structures (NALS) Journal* 1. <http://www.tekstlab.uio.no/nals/#/chapter/39>
- Holmberg, A.** (1993). Two Subject Positions in IP in Mainland Scandinavian. *Working Papers in Scandinavian Syntax*, 29–41.
- Holmberg, A., & Platzack, C.** (1995). *The Role of Inflection in Scandinavian Syntax*. Oxford University Press.
- Hudson Kam, C., & Newport, E.** (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development* 1(2), 151–195.
- Jakubowicz, C.** (2011). Measuring derivational complexity: New evidence from typically developing and SLI learners of L1 French. *Lingua* 121(3), 339–351.
- Johannessen, J. B., & Garbacz, P.** (2011). Fältarbete med Nordic Dialect Corpus. In M. Reinhammar, L. Elmevik, & L-E. Edlund (Eds.), *Studier i svenska språkets historia 11. Förhandlingar vid Elfte sammankomsten för svenska språkets historia i Uppsala 23–24 april 2010*. Acta Academiae Regiae Gustavi Adolphi 116.
- Johannessen, J. B., Priestly, J., Hagen, K., Áfarli, T. A., & Vangsnes, Ø. A.** (2009). The Nordic dialect corpus - an advanced research tool. *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009. NEALT Proceedings Series Volume 4*. <http://tekstlab.uio.no/nota/scandiasyn/>
- Kiss, K. E.** (1996). Two subject positions in English. *Linguistic Review* 13(2), 119–142. doi: [10.1515/tlir.1996.13.2.119](https://doi.org/10.1515/tlir.1996.13.2.119)
- Lind, M., Simonsen, H. G., Hansen, P., Holm, E., & Mevik, B.-H.** (2015). Norwegian words: A lexical database for clinicians and researchers. *Clinical Linguistics & Phonetics* 29(4), 276–290. doi: [10.3109/02699206.2014.999952](https://doi.org/10.3109/02699206.2014.999952)
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F.** (1992). Overregularization in Language Acquisition. *Monographs of the Society for Research in Child Development* 57 (4). Doi: [10.2307/1166115](https://doi.org/10.2307/1166115)
- Mohr, S.** (2005). *Clausal Architecture and Subject Positions: Impersonal Constructions in the Germanic Languages*. Amsterdam/Philadelphia: John Benjamins.
- Munch, C. B. Ø.** (2013). *North Germanic Negation. A microcomparative perspective*. (Doctoral dissertation). University of Tromsø.
- Mykhaylyk, R., & Ko, H.** (2010). Optional scrambling is not random: Evidence from English-Ukrainian acquisition. In M. Anderssen, K. Bentzen and M. Westergaard (Eds.), *Variation in the Input: Studies in the Acquisition of Word Order*, 241–270. Dordrecht: Springer Netherlands.
- Nilsen, Ø.** (1997). Adverbs and A-shift. *Working papers in Scandinavian Syntax*, 59, 1–31.
- Pearl, L.** (2021). Theory and predictions for the development of morphology and syntax: A Universal Grammar + statistics approach. *Journal of Child Language* 48, 907–936.
- Penner, Z., Tracy, R., & Weissenborn, J.** (2000). Where scrambling begins: Triggering object scrambling at the early stage in German and Bernese Swiss German. In S. M. Powers & C. Hamann (Eds.), *The acquisition of scrambling and cliticization*, pp. 127–165. Dordrecht: Springer Netherlands.

- Pinker, S.** (1999). *Words and Rules: the ingredients of language*. New York: Basic Books.
- R Core Team.** (2018). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL <https://www.R-project.org/>.
- Ringstad, T. L.** (2019). Distribution and function of embedded V-Neg in Norwegian: A corpus study. *Nordic Journal of Linguistics* 42, 329–363. doi: 10.1017/S0332586519000210
- Ringstad, T. L. S.** (2021). *Acquiring variation. A study of children's and adults' embedded clause word orders in Norwegian – distribution, development and effects*. (Doctoral dissertation, Norwegian University of Science and Technology). Available from <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2767200>.
- Schaeffer, J.** (2000). Object scrambling and specificity in Dutch child language. In S. M. Powers & C. Hamann (Eds.), *The acquisition of scrambling and cliticization*, 71–93. Dordrecht: Springer Netherlands.
- Schwab, J. F., Lew-Williams, C., & Goldberg, A. E.** (2018). When regularization gets it wrong: Children over-simplify language input only in production. *Journal of Child Language*, 25(5), 1054–1072.
- Stefanowitsch, A.** (2020). *Corpus linguistics. A guide to the methodology (Textbooks in Language Sciences 7)*. Berlin: Language Science Press.
- Svenonius, P.** (2002). Subject positions and the placement of adverbials. In P. Svenonius (Ed.), *Subjects, Expletives and the EPP*, 201–242. New York: Oxford University Press.
- Tekstlab.** (2004). Norsk talespråkkorpus – Oslodelen [Norwegian speech corpus – the Oslo part]. Tekstlaboratoriet, ILN, University of Oslo.
- Tekstlab.** (2009). BigBrother-korpuset [The BigBrother corpus]. Tekstlaboratoriet, ILN, University of Oslo.
- Unsworth, S.** (2005). *Child L2, adult L2, child L1: Differences and similarities: A study on the acquisition of direct object scrambling in Dutch* (Doctoral dissertation, Utrecht University). Retrieved from https://www.lotpublications.nl/Documents/119_fulltext.pdf.
- Urbanik, P., & Svennevig, J.** (2018). Managing contingencies in requests: The role of negation in Norwegian interrogative directives. *Journal of Pragmatics* 139, 109–125.
- Westergaard, M.** (2008). Verb movement and subject placement in the acquisition of word order: Pragmatics or structural economy? In P. Guijarro-Fuentes, P. Larranaga & J. Clibbens (eds.), *First language acquisition of Morphology and Syntax: Perspectives Across Languages and Learners*, 61–86, Amsterdam: John Benjamins.
- Westergaard, M.** (2009). *The Acquisition of Word Order. Micro-cues, information structure and economy*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Westergaard, M.** (2011). Subject positions and information structure: the effect of frequency on acquisition and change. *Studia Linguistica*, 65(3), 299–332.
- Westergaard, M., & Anderssen, M.** (2015). Word order variation in Norwegian possessive constructions: Bilingual acquisition and attrition. In J. B. Johannessen & J. Salmons (Eds.), *Germanic heritage languages in North America: Acquisition, attrition and change [Studies in Language Variation]*, 21–45. Amsterdam: John Benjamins.
- Westergaard, M. R., & Vangsnes, Ø. A.** (2005). Wh-questions, V2, and the left periphery of three Norwegian dialects. *Journal of Comparative Germanic Linguistics* 8(1), 117–158.
- Wiklund, A.-L., Bentzen, K., Hrafnbjargarson, G. H., & Hróarsdóttir, T.** (2009). On the distribution and illocution of V2 in Scandinavian *that*-clauses. *Lingua* 119(12), 1914–1938. doi: 10.1016/j.lingua.2009.03.006
- Wiklund, A.-L., Hrafnbjargarson, G. H., Bentzen, K., & Hróarsdóttir, T.** (2007). Rethinking Scandinavian verb movement. *Journal of Comparative Germanic Linguistics* 10(3), 203–233. doi: 10.1007/s10828-007-9014-9

Cite this article: Ringstad T., & Westergaard M. (2023). Children's acquisition of word order variation: A study of subject placement in embedded clauses in Norwegian. *Journal of Child Language* 1–25, <https://doi.org/10.1017/S0305000923000259>