



Promoting the demoted

Jensen, Torben Juel; Christensen, Tanya Karoli

Published in:
Lingua

DOI:
[10.1016/j.lingua.2013.08.005](https://doi.org/10.1016/j.lingua.2013.08.005)

Publication date:
2013

Document Version
Peer reviewed version

Citation for published version (APA):
Jensen, T. J., & Christensen, T. K. (2013). Promoting the demoted: The distribution and semantics of “main clause word order” in spoken Danish complement clauses. *Lingua*, 137, 38-58. DOI: 10.1016/j.lingua.2013.08.005

Promoting the demoted: The distribution and semantics of “main clause word order” in spoken Danish complement clauses

Torben Juel Jensen & Tanya Karoli Christensen

1 Introduction

Like many other Germanic languages, modern Danish distinguishes between what are traditionally described as ‘main clause’ and ‘subordinate clause’ word orders (Diderichsen, 1974; Heycock et al., 2010; Meinunger, 2006; Wiklund et al., 2009). Contrary to what these terms imply, both word orders occur in subordinate clauses (henceforth ‘subclauses’), but seemingly with different restrictions. It is these restrictions, and thereby indirectly the reasons for using one or the other word order, the present article is concerned with. Here, we restrict our attention to complement clauses¹, partly because they display more variation (i.e. more main clause word order) than most other types of subclauses², partly in order to enable comparisons with previous studies of word order in Scandinavian languages, which have also focused on complement clauses (or, more restrictedly, on *at*- ‘that’-clauses). The data consist of a very large, extensively annotated corpus of sociolinguistic interviews in modern spoken Danish. Due to the unique size of this material, which contains almost 3,500 complement clauses in which the two word orders can be distinguished, we have been able to test hypotheses regarding the use of ‘main clause word order’ in subclauses through statistical analyses, supplemented with in-depth, qualitative analyses where needed. Such hypotheses have so far primarily been proposed and tested on the basis of introspective evidence,

¹ Complement clauses in this study comprise *at* ‘that’-clauses, clauses where *at* ‘that’ is omitted, *hv*- ‘wh’-clauses and *om* ‘if, whether’-clauses. Examples are given throughout this article.

² A notable exception is *fordi* ‘because’-clauses, cf. Jensen (2011).

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

often within the frames of a syntactic model like generative grammar. In distinction to this, we take a variationist approach where distributions in observed language use are the object of study, and where explanations may be based on both extra- and intra-linguistic variables. We follow current trends in the paradigm by including semantico-pragmatic variables in the account of intra-linguistic variables (e.g. Bresnan and Ford, 2010; Pichler, 2010; Terkourafi, 2011).

In Danish grammatical tradition the distinction between the two word orders has been portrayed as a linear model called the 'sentence frame', where one frame was attributed to main clauses and the other to subclauses (Diderichsen, 1974). The so-called 'main clause frame' is characterized by having a first position where any constituent but the finite verb, negation and modal particles may be placed. The finite verb occupies second position, and always precedes any sentence adverbial which is not placed initially, including the negation. The so-called 'subclause frame' only allows placement of the subject in initial position, which is then followed by sentence adverbials, relegating the finite verb to third position. In this study, the word order distinction is operationalized as the relative orders of finite verb and sentence adverbials, abbreviated as **V>Adv** for the type in (1) and **Adv>V** for the type in (2), where Adv is meant to include negation (other studies have focused solely on the relative order of verb and negation, cf. Heycock et al., 2010; Wiklund et al., 2009). As mentioned above, both may be used in subclauses; see examples of Danish subclauses in (1-2) below.

- (1) **Conj** > **1st** > **V_{fin}** > **(Subj)** > **Adv** **(Objects etc.)** 'main clause schema'
a. *at* *hun* *købte* *nok ikke* *blomster* *i dag*

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

that she bought probably not flowers today

b. *at i dag købte hun nok ikke blomster*

that today bought she probably not flowers

(2) **Conj > Subj > Adv > V_{fin} (Objects etc.) - 'subclause schema'**

at hun nok ikke købte blomster i dag

that she probably not bought flowers today

'...that she did not buy flowers today.'

Some studies place an emphasis on the distinction between (1a) and (1b), where the latter has a non-subject in initial position, making it a clear case of embedded V2 (see, e.g. Heycock, 2007).

Others also consider clauses of the type in (1a), with an initial subject and a negation succeeding the finite verb, a variant of V2 (e.g. Julien, 2007; Wiklund et al., 2009). As can be seen from the examples, both are instances of V>Adv, and are counted as such in this study (more on this in Section 3). Note that it is only possible to distinguish the two word order types in a minority of subclauses. A study of a small sample of the conversations from the LANCHART corpus has shown that just about 25% of all subclauses can be unambiguously determined with respect to word order (in 20% of the cases based on presence of a sentence adverbial and in 5% of the cases on initial placement of a non-subject only) (Jensen, 2011).

In accordance with variationist principles of accountability (see e.g. Tagliamonte, 2012: 9-11), we do not restrict our attention to the placement of negation but include a range of other types of sentence adverbials (cf. Section 3), since they conform to the same relative-to-finite pattern as negation.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

In Danish subclauses, Adv>V is the historically younger word order type. Up to the fifteenth century, V>Adv was used widely in both main clauses and in subclauses (Heltoft, 2005), but according to Gregersen and Pedersen (1997, 2000) Adv>V became close to obligatory in subclauses over the following centuries. Prescriptive advice is seldom given in grammars and handbooks, with Høysgaard (1979 [1747]: § 392 & 394) as a notable exception, but Gregersen and Pedersen argue that the norm has been taught “as a matter of course” (2000: 427). However, at least as far as spoken language is concerned, V>Adv word order is regularly reported in certain types of subclauses in modern Danish, as well as in the other Mainland Scandinavian languages (Andersson, 1975; Gregersen and Pedersen, 2000; Julien, 2007; Sundquist, 2003). A survey of early 20th-century rural dialects indicates that V>Adv in subclauses is much more common in the western Danish dialects than in the eastern dialects, regarding both frequency and types of subclauses in which V>Adv is found (Gregersen and Pedersen, 2000; Pedersen, 1996).

1.1 Purpose of the study

The study presented in this article is based on a wide-spread conception of the semantics and pragmatics of V>Adv, and (at least some) other so-called main clause phenomena (MCP; see Aelbrecht et al., 2012). In short, this conception assumes that a subclause exhibiting MCP retains much of the semantic value of a regular main clause, even though it has been demoted to subordinate (syntactic) status. Or put somewhat differently, a ‘MCP word order’ used in a subclause acquires more focus or attention than a ‘non-MCP word order’ (it is marked in this context; cp. Andersen, 2001; Christensen and Heltoft, 2010). This effect has been described as ‘emphasis’ (Hooper and Thompson, 1973), ‘assertion’ (Hooper and Thompson, 1973; Julien, 2007; Meinunger, 2006), ‘main point of utterance’ (Simons, 2007; Wiklund et al., 2009), or ‘more

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

important information' (Christensen and Heltoft, 2010). For the purposes of this study, we prefer the term *foreground* vs. *background*, although no substantial difference is intended to the presently more widespread 'main point of utterance'. On the other hand, we do not find a description as assertion promising for subclauses that are syntactically dependent on another clause, i.e. the matrix clause. In our view it is often the entire complex of matrix and subclause which is asserted, not only the subclause (for a somewhat different discussion of assertion vs. foreground, see Cristofaro, 2003: 34-38; Ogle, 1981). An advantage of having a terminological pair like this, is that it allows us to discuss also the opposite effect of the one attributed to V>Adv (i.e., background information). Note, though, that we do not mean to imply that the alternative word order type, Adv>V, necessarily signals background information; it may be promoted to the foreground by other means (such as modal particles; cf. Christensen, 2007). The notions of fore- and backgrounding apply just as much to the matrix clause of which the subclause is a constituent, as we will see later. In contrast to the subclause, however, the matrix clause may be considered foregrounded per default. It should also be noted that the hypothesis of V>Adv as a foregrounder only applies to subclauses, as V>Adv word order is obligatory in main clauses unless they are initiated by a subjective particle such as *gid*, roughly translatable as 'I wish that' or *sikke* 'how, such' (as in *sikke de larmer* 'how noisy they are'). In the latter context, Adv>V acquires a completely different, modal meaning (Christensen and Heltoft, 2010).

The purpose of this study, then, is to determine the distribution of V>Adv in complement clauses in a corpus of modern spoken Danish, and test the hypothesis that V>Adv is a foregrounding signal in such clauses. In continuation of this, it is crucial to note that even though sociolinguistic methodology requires semantic equivalence between variants (Labov, 1978), it is

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

possible to compare variants that differ on one semantic level (here, the semantico-pragmatic level of information status), if they are comparable on another. An obvious candidate for such comparability is truth-conditional semantics, and as shown in (1) and (2), the same state-of-affairs may be denoted by either word order (cf. Weiner and Labov, 1983, for a similar approach; but see also Dines, 1980; Hasan, 2009[1989]; Lavandera, 1978; Pichler, 2010; Romaine, 1984; Sankoff, 1973; Terkourafi, 2011, for more critical views on semantic equivalence as a methodological principle).

In some respects, the present study may be seen as a complement to the studies of other Scandinavian languages than Danish (Icelandic, Faroese, Norwegian and Swedish) presented in Julien (2007) and especially Wiklund et al. (2009), as they basically test the same hypothesis. However, our study differs not only with respect to the language studied, but also in the very different methodologies employed. The methodological approach of Wiklund et al. is mainly intuition- and judgment-based, whereas we base our conclusions first and foremost on observational data from a spoken language corpus of sufficient size to enable statistical treatment. Our qualitative analyses necessarily draw on intuitions about the meaning of an utterance, but the utterances in question are all part of our data set, and can be analysed according to the preceding and succeeding context, interlocutors' reactions etc. Importantly, such data invariably contain occurrences one could not have come up with when constructing examples for judgment tasks or other types of experiments. Julien's study is corpus-based but with a much more restricted scope than ours, as it only includes subclauses initiated by *at(t)* 'that' which contain a negation and with maximum four words between the subordinator and the finite verb. In contrast to this our study

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

includes *all* complement clauses occurring in the data material containing *any* sentence adverbial (see definition below).

2 Theoretical background

In this section we will review a set of phenomena that are related directly or indirectly to foregrounding/backgrounding of subclauses. Some have to do with elements of the subclause itself, others with the matrix clause governing it.

2.1 Expected *V>Adv* contexts

First of all, there are two types of clauses in which *V>Adv* is the expected order, even though tradition categorises them as subclauses. For many purposes they do not count as true subclauses, but nonetheless they are included in this study because we consider it an empirical question whether they display word order variation (and derived from this, whether they are better categorised as main clauses).

One type is direct quotes, where *V>Adv* may be simply 'inherited' from the portrayed discourse as in *he said "why don't you come over?"*

Another type is found in the context of what has been called 'epistemic phrases' or 'fragments' (Thompson and Mulac, 1991; Thompson, 2002). These are (a subtype of) clauses containing complement-taking predicates (CTP; see Thompson, 2002; Boye and Harder, 2007), which do not behave as regular matrix clauses and do not contribute with propositional content, but rather modify the content expressed in the complement clause. Examples from Boye and Harder (2007: 572, 579) are found in (3-4):

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

(3) *I think* she loves me.

(4) The weather is getting better, *I think*

Boye and Harder (2007) argue that not all CTP-clauses are necessarily such secondary (epistemic) phrases, and list a range of formal criteria for distinguishing which is which. They mention a set of criteria for secondary phrase status, e.g., NEG-raising (such as *I don't think she loves me* where negation is read as pertaining to 'she loves me', not to 'I think') and adverbial distribution (as in (4) where the CTP-clause is placed clause finally like an adverbial might have been), and adverbial modification (as in *I never thought she loved me*) as a criterion for true matrix clause status (see further in Boye and Harder, 2007: 578-580). Where such secondary, adverbial-like phrases are clearly identified, it is argued, the accompanying 'complement clause' is actually a main clause—and would therefore be expected to have regular main clause word order (V>Adv, in our case).

2.2 Matrix clause predictors

In most other cases, the word order seems to vary substantially, but with certain restrictions. One restriction, taken up by several studies of MCP (not only V>Adv), is based on the type of matrix clause predicate governing the subclause in question. Hooper and Thompson's (1973) five-way distinction is arguably the most influential (cf. Julien, 2007; Meinunger, 2006; Wiklund et al., 2009). The five predicate classes are not exhaustive of complement-taking predicates, but they have the advantage of being defined separately from the identificational criteria of MCP, here V>Adv, and of specifying the (truth-conditional) semantic relations to their complements. It is thus possible to give principled predictions about the distributional patterns of Adv>V and V>Adv in subclauses under

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

different predicate types. Classes A-C are called ‘nonfactive’ by Hooper and Thompson (1973: 473)³, while classes D-E are ‘factive’, a distinction related to both assertion and foregrounding, as will be explained below.

2.2.1 Classes A and B

Many of the predicates in class A are ‘verba dicendi’ (communicative predicates) and may be used for direct quotes (an expected V>Adv context as explained above), but also for indirect quotes (where Adv-V is commonly used in Danish). Much the same goes for class B, mostly ‘verba cogitandi’ (cognitive predicates), which may introduce imagined quotes of both kinds. The two types of predicates are typically referred to as ‘strongly’ and ‘weakly assertive’ predicates, respectively. As mentioned by Hooper and Thompson (1973: 475), predicates of both of these classes may occur in parenthetical function when used in a matrix clause (Urmson, 1952), i.e. much as the secondary phrases mentioned above. It is important to note, though, that not all such clauses display any of the formal criteria mentioned by Boye and Harder (2007), and when not,

³ Hooper and Thompson’s predicate classes (1973: 473-474):

<i>Nonfactive</i>			<i>Factive</i>	
<i>Class A</i>	<i>Class B</i>	<i>Class C</i>	<i>Class D</i>	<i>Class E</i>
say	suppose	be (un)likely	resent	realize
report	believe	be (im)possible	regret	learn
exclaim	think	be (im)probable	be sorry	find out
assert	expect	doubt	be surprised	discover
claim	guess	deny	bother	know
vow	imagine		be odd	see
be true	it seems		be strange	recognize
be certain	it happens		be interesting	
be sure	it appears			
be obvious				

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

they are ambiguous between a parenthetical (backgrounded) reading and a foregrounded reading of the matrix clause (a point made by Boye and Harder as well).

All in all, predicates of class A and B have functions that predict a fairly high degree of co-occurrence with V>Adv subclauses while not precluding Adv>V.

2.2.2 Class C

Predicates of Hooper and Thompson's class C are typically called 'non-assertive' (Wiklund et al., 2009: 1917). Examples are *deny*, *doubt*, *be (un)likely*. According to both Julien (2007) and Wiklund et al. (2009), they do not allow V>Adv (or, more specifically, V>Neg). Already Andersson (1975; quoted by Heycock, 2007) noted that the non-negated variants belong in class B (while negated variants of B, such as 'do not suppose' should be placed in class C). If this principle is followed through, class C becomes a class of (inherently) negative predicates (cf. Meinunger, 2006: 462), a fact which presents some difficulty for the conventional test for logical-semantic presupposition, which tests for so-called truth-preservation under negation (see section 2.2.3).

According to the findings in Hooper and Thompson, Julien and Wiklund et al., V>Adv should not occur in the context of class C predicates. We will return to this in the discussion (section 5).

2.2.3 Classes D and E

Classes D and E consist of *factive* and *semi-factive* predicates, respectively. Typically, factivity is explained by recurrence to presupposition, but as implied above, there is more than one definition of presupposition in the tradition. What unites the two types of predicates, is logical-semantic presupposition (Cristofaro, 2003: 30), which has to do with the entailed truth value of the subclause. Some predicates entail the truth of their complements, a characteristic tested by

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

inserting (or removing) a negation in the matrix clause (Kiparsky and Kiparsky, 1970, the conventional test for presupposition mentioned above), compare the a. and b. versions of (5-6):

- (5) a. It is sad that Anna is leaving
b. It is not sad that Anna is leaving
- (6) a. She realized that Anna is leaving
b. She did not realize that Anna is leaving

In all versions, it is entailed – or logically presupposed – that Anna is leaving.

Semi-factive predicates may 'lose their factivity' when used in questions and conditionals (Hooper and Thompson, 1973; Karttunen, 1973), or as parentheticals (Urmson, 1952), or when used with simple, present tense and a first-person singular subject (e.g. Boye and Harder, 2007; Glismann, 1978; Simons, 2007). Compare (7a) and (7b):

- (7) a. I realize that she has left for France
b. I do not realize that she has left for France

When the speaker explicitly states a lack of knowledge (as in (7b)), she cannot simultaneously believe the proposition to be true, and thus truth is not preserved under negation for these constructions (1sg, present, without further qualifications).

For other constructions with semi-factives the content of the subclause is logically presupposed, just as with true factive predicates. However, for both types of predicates, logical

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

presupposition does not necessarily lead to pragmatic presupposition, or what we term

backgrounding; cf. Ogle (1981) and Cristofaro (2003). Compare the dialogues in (8) and (9), both with semi-factive predicates (see also Simons, 2007: 1036).

- (8) a. Who was Louise with last night?
b. Henry thinks that she was with Bill.
- (9) a. What is bothering Henry?
b. He discovered that Louise was with Bill last night.

The foregrounded part in (8.b) is the subclause, concerning the whereabouts of Louise (i.e., that she was with Bill). In (9.b), on the other hand, the matrix clause, concerning the discovery Henry has made, is foregrounded.

Pragmatic presupposition (=backgrounding) is contextually determined, but will be found more often with true factive than with semi-factive predicates because of the different semantics of these two categories. True factive predicates “express some emotion or subjective attitude about a presupposed complement” (Hooper and Thompson, 1973: 479). Typically, such attitudes and evaluations are proffered about propositions whose contents are already known to both speaker and hearer, thus forming the ‘common ground’ between them (Stalnaker, 2002), and therefore also belonging to the backgrounded part of the utterance. In other words: The expression of stance in the matrix clause predicate will most likely be what is at stake for the speaker and therefore foregrounded information. Semi-factive predicates, on the other hand, typically denote situations of ‘coming to know something’ (e.g. *realize*, *find out*, *learn*). Informing the addressee of such a

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

changed mental state, may be the main point of an utterance, but not necessarily, as shown above.

It should be clear from this that factive predicates are expected not to license V>Adv in the subclause, because a foregrounding signal (V>Adv) should not be possible in a backgrounded context, whereas semi-factive predicates are expected to occur more often with V>Adv, depending on other contextual factors.

2.3 Subclause predictors

2.3.1 The effect of subordinators

Subordinators appear to play an important role with respect to the distribution of V>Adv. An analysis of this word order type as a signal of assertion (as in Julien, 2007; Meinunger, 2006; and indirectly Hooper and Thompson, 1973) would predict that the Danish equivalents of *if/whether* do not allow of V>Adv, since these subordinators clearly present the information in the subclause as non-asserted. In contrast, this prediction does not follow from an analysis of V>Adv as signal of foregrounded information (or 'main point of utterance'), unless one takes the somewhat roundabout argument of Hooper and Thompson's that only asserted subclauses may be emphasized, or here: foregrounded (1973: 472-473)⁴.

As regards the more frequent *at* 'that', Vikner (1995: 84) claims that in Danish (as well as English, Faroese, Frisian, Norwegian and Swedish) "[e]mbedded V2 is only grammatical if *that/at* is present". Wiklund et al. (2009: 1918) make a more fine-grained distinction whereby "the

⁴ The argument as phrased by Hooper and Thompson is circular, though: "RTs [i.e. root transformations] that produce emphasis are restricted to asserted clauses because emphasis would be unacceptable in clauses that are not asserted, e.g. embedded clauses which are presupposed, or clauses which are questions or imperatives." ((Hooper and Thompson, 1973: 472-473))

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

complementizer is obligatory in non-subject-initial V2 clauses and preferred in subject-initial V2 clauses (disregarding cited assertions)".

Thompson and Mulac (1991), on the other hand, have argued that only complements containing a subordinator (*that, if, whether*) are really complements, and that those without a subordinator are simply modified by the adverbial-like secondary phrases mentioned above. Thompson and Mulac's generalization concerns English, but if it holds for Danish too, it would give the opposite prediction of Vikner and Wiklund et al., since absence of a subordinator would promote a main clause reading.

2.3.2 Sentence adverbials

As mentioned, not only negation but also a range of sentence adverbials may be placed relative to the finite verb. Sentence adverbials form a rather heterogeneous group that require subdivision in order to bring about any predictions regarding word order. One type of subdivision is based on the layered model of utterances found in, e.g., Dik's Functional Grammar (Dik, 1989). Sentence adverbials always pertain to the upper layers of an utterance, since they do not modify the state-of-affairs described (such as manner adverbials and time/place adverbials do), but rather comment on it or place it relative to other utterances. On the basis of such layered descriptions of Danish adverbials (Andersen, 1986; Jensen, 2000), we propose a rough distinction between **textual** adverbials, which either describe how one utterance is related to another or place it in a chain of reasoning (e.g., *også* 'also', *alligevel* 'nonetheless'), **epistemic/evaluative** adverbials, commenting on the speaker's degree of knowledge about or attitude towards the proposition (*måske* 'maybe', *heldigvis* 'luckily'), **modal particles**, which carry interactional meaning and signal relations between speaker and hearer (*jo* 'you know, of course', *da* 'really, surely') and **swear words** (*fandeme*, lit.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

'may the devil take me', *kraftedeme*, lit. 'may cancer eat me'), some of which in Danish behave syntactically like sentence adverbials. Few analyses have been made regarding the use of sentence adverbials with the two subclause word orders investigated here, but Christensen (2007: 169) claims that all modal particles are compatible with V>Adv subclauses, and further that modal particles such as *jo* and the more infrequent *skam* 'really, to be sure' promotes subclauses with Adv>V word order to the foreground. This suggests that at least these modal particles may be found more frequently with the semantico-pragmatically concordant V>Adv word order.

3 Material and methods

3.1 Corpus

The data material for the study presented in this article consists of transcribed recordings of sociolinguistic interviews with a subgroup of the informants from the extensive LANCHART study of language change in 20th-century Danish (Gregersen, 2009). The speakers come from either the capital, Copenhagen, or the small rural town Vinderup in the western Jutland peninsula. The 91 different informants have been divided into age cohorts, referred to in the following as Generation 1 (born 1942-63), Generation 2 (born 1964-74) and Generation 3 (born 1989-92).⁵ The distribution of speakers for generation and locality appears from Table 1. The informants in each cell are equally dispersed with respect to gender and the two social classes working and middle class. In total, the corpus subsection under study comprises transcriptions of 132 recordings, each about 1-3 hours in duration, totalling 1,547,032 words.

⁵ One of the speakers (from Vinderup 3) utters only four subclauses in the interview; none of these are included in the main results presented in this article because they either occur in a parenthetical tag or have only a fronted sentence adverbial.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

	Generation 1	Generation 2	Generation 3
Copenhagen (recorded in 1987-89 and 2005-07)	24	18	-
Vinderup (recorded in 2006-07)	-	19	30

Table 1: Informants in the LANCHART word order study

This design makes it possible to study variation and change in language use in both apparent and real time, as well as geographically.

3.2 Methodology

Using variationist methodology, we have defined the word order distinction $V > Adv$ vs. $Adv > V$ as the dependent variable under study, while a range of both linguistic and non-linguistic factors are evaluated as independent variables potentially influencing it.

3.2.1 The envelope of variation

A subclause is here defined very broadly as any clause which is a constituent of another clause and dependent on this other clause (this definition is thus in line with traditional descriptions of subclauses, and not a functional definition like the one in Cristofaro, 2003). Our intention has been to include all contexts in which both $V > Adv$ and $Adv > V$ word order is possible, along with contexts where $V > Adv$ is (much) more frequent than $Adv > V$ – in other words, the whole “envelope of variation” (cf. e.g. Labov, 2004).

Clauses outside the envelope of variation have been excluded for reasons of principle (they do not vary between the two word orders). These are first and foremost ‘subclauses’ with what we call *V > Subj tags*. Such tags have an unambiguously parenthetical reading and the fixed structure

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

Verb-Subj-(Adv). They are therefore better analysed as interjectionals or tags than as matrix clauses, for which reason the 'subclause' is not a true subclause either. They are typically placed sentence finally (11.a), but may also be found sentence internally (11.b).

(11) a. *de var jo ikke lige så gode som os syntes vi vel*
they were ADV not as good as us thought we probably
'they were just not as good as us, we thought, I guess'

b. *der er vist kommet lidt mere fokus på det tror jeg inden for
altså læreruddannelsen*
you-know the-teacher-education
'there has probably been a bit more focus on it, I think, within educational studies'

All 641 tokens of this construction in our data display V>Adv word order, and are not included in the results presented below. Also excluded are subclauses which for some reason are indeterminable with respect to word order even though they contain a sentence adverbial: interrupted clauses, some elliptical clauses, and sentence 'intertwinings'.⁶

⁶ Sentence intertwining is a special case of extraction, where a constituent of a subclause occurs in the first position of the matrix clause. As in cleft constructions, an 'empty slot' is left behind in the subclause. Compare (i) and (ii) below, where the second is an intertwined version of the first where the pronoun *det* 'it' is extracted from the subclause to a fronted position in the matrix clause:

(i) *jeg tror ikke det har haft nogen indflydelse*
I think not it has had any influence
'I don't think it has had any influence'

(ii) *det tror jeg ikke har haft nogen indflydelse*
that think I not has had any influence
'That I don't think has had any influence.'

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

As mentioned in the introduction, the variation with respect to the placement of sentence adverbial before or after the finite verb is closely related to the possibility of having non-subjects placed initially in the clause. As expected from the 'main clause schema', all the non-subject initial clauses (including the V1 clauses) have V>Adv word order.⁷ It may therefore be argued that also clauses without sentence adverbials could be included in the analyses, counting non-subject initial clauses as instances of V>Adv (V2). However, this would violate the Labovian 'principle of accountability' (e.g. Tagliamonte, 2012), which states that also contexts where a variant did not occur, but might have, should be counted - so-called non-occurrence. When a non-subject does not occur initially, what we find is a subject (or an empty slot, in the 'clause schema' tradition). But here a problem arises: Initial placement of a subject is compatible with both of the word order types that we are interested in, and does not in itself distinguish the traditional 'main clause' word order from 'subclause' word order. Where no sentence adverbial is present, it is impossible to tell if a subject-initial clause represents one or the other word order, and thus non-subjects fail as a

The reason for excluding such constructions is that when the extracted element is the subject, as in (ii), it is not possible to determine whether the sentence adverbial is placed in the subclause or in the matrix clause, making the word order ambiguous. All 354 complement clauses in intertwined constructions containing a sentence adverbial in the data material were excluded from the analyses presented in this article.

⁷ A reviewer argues that non-subject initial clauses should be excluded from the analysis, as it is a context which is non-variable with respect to Adv>V / V>Adv word order. We agree that this would be an obvious way of "circumscribing the variable context" by excluding the "don't count cases" as the variationist procedure prescribes (Tagliamonte, 2012: 10). However, this would imply seeing V>Adv in these clauses as *caused* by the topicalization of a non-subject, an analysis with which we do not agree. Instead, we see the *possibility* of non-subject topicalization and the placement of the sentence adverbial as two phenomena caused by the same underlying factors. The *actual* topicalization of a non-subject, though, is probably based on completely different factors, having to do with textual coherence, and the exclusion of this part of the data would therefore, in our opinion, blur the picture of the factors influencing the choice of 'main' versus 'subordinate clause word order'.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

sociolinguistic variable, simply because its counterpart does not identify this main clause phenomenon correctly. On the other hand, adverbial placement always distinguishes between the two word orders. This does not mean, of course, that a study of the distribution of non-subject initial word order on different types of subclauses would not be interesting, but it would address a different phenomenon than we do here (cf. also footnote 7).

3.3 Coding

The actual coding process was semi-automatic. For the 132 conversations included in the study, all occurrences of text strings potentially constituting a sentence adverbial⁸ were tagged automatically; subsequently, homographic or polysemic forms were manually discarded.

The remaining subclauses were categorized and coded with respect to a number of syntactic and semantico-pragmatic factors in addition to word order. The coding scheme is shown in Table 2, and examples are given in Appendix A. All the factors are more or less directly related to the semantico-pragmatic claim that V>Adv is a foregrounding signal (cf. section 2).

3.3.1 Information status

Most directly related to the semantico-pragmatic claim is the factor INFORMATION STATUS. This factor was intended to capture the distinction between expected V>Adv subclauses (for which the status as subclauses may be questioned, cf. section 2.1) and subclauses which are more open (: variable) regarding word order.

⁸ Based on previous investigations of Danish sentence adverbials (Nimb, 2004) as well as a pilot study of a smaller part of the LANCHART corpus a list of 107 potential candidates, including multiword strings, such as *lige så godt* 'just as well', was compiled. A detailed instruction manual was provided for coders to help them decide which instances were sentence adverbials and which were not.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

The context of direct (real or imagined) quotes of speech or thought has been very strictly defined to include only occurrences containing an explicit signal that a quote has begun, such as a change in tense or person (see (12)). The reason for this is to avoid uncertain or even spurious inclusions into the category, since a great deal of verbs may be argued to be communicative in certain contexts.

- (12) *så sagde hun nu kommer han sikkert igen* (V>Adv)
then said she now comes he probably again
'then she said now he's probably coming back'

The context of secondary, epistemic phrases was identified mainly as clauses containing simple, present tense epistemic predicates and a first person subject (cf. section 2.1). Apart from such clauses, we have added a few construction types that contribute very little information (of which some have epistemic meaning because of the presence of a modal verb or the like), see (13).

- (13) *det kan godt være at det ikke er så klart jeg husker det* (Adv>V)
it can well be that it not is so clearly I remember it
'it may be the case that I don't remember it very clearly'

Note that in spite of the clear secondary status of the matrix clause, the subclause may display Adv>V word order, as in (13), preventing an analysis as a main clause.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

3.3.2 Type of matrix clause predicate

The factor 'type of matrix predicate' is intended as an operationalization of Hooper and Thompson's (1973) five classes of predicates (cf. section 2.2), but note that their Class C does not form part of our quantitative analyses, since virtually no occurrences of this type of predicate were found in this corpus.⁹ As mentioned in Section 2.2.2, some researchers refer negated occurrences of Class B predicates (i.e. our 'cognitive' predicates) to Class C. Instead, we have included presence of a negation in the matrix clause as a separate factor in the analysis (see below).

The wide range of predicates found in our data gave rise to two additional categories. The first we have called 'causatives' (14) (see also Meinunger, 2006).

(14) a. *det er også en af grundene til at vi tør næsten ikke at flytte*

that is also one of the-reasons to that we dare almost not to move
'that's also one of the reasons that we almost don't dare to move away'

b. *så er det ikke hans skyld at han ikke kan svare*

then is it not his fault that he not can answer
'then it isn't his fault that he cannot answer'

Causatives present the content of the subclause as contingent upon some external fact, often mentioned in the previous context. They seem to hold also under negation, but are not typical examples of factive predicates in that they do not offer an evaluation. On the other hand, in stating that something is the reason why the subclause holds, they present the content of the subclause

⁹ The single token we found (containing the predicate *udelukke* 'exclude') was included in the category 'cognitive predicates'.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

as given and the preceding reason as the new information. For this reason, we expect few occurrences of V>Adv in such contexts.

Additionally, we formed a residual group, simply called OTHER. This group contains a range of predicates and constructions that do not fit into the other categories. Note that this category also includes contexts in which semi-factive predicates are mitigated by a 1sg subject plus present tense (since these do not give rise to logical presuppositions and thus do not belong with the other semi-factive constructions, cf. section 2.2.3).

3.3.3 Other matrix clause factors

Sentence adverbials may also be found in the matrix clause, and if such adverbials attract attention in and of themselves, they may influence the relative prominence of matrix and subclause, and hence also the disposition of the latter for V>Adv word order. Sentence adverbials were therefore registered in subclauses as well as their matrix clauses.

The relative position of matrix clause and subclause was coded for, since initial versus final placement of the subclause is likely to be connected with foregrounding, due to the regular theme-rheme order (Halliday and Matthiessen, 2004: 64): Previous studies have shown that initially placed subclauses most often represent GIVEN background information serving as a pivotal point in the local organization of the text (Ford and Thompson, 1986; Jensen, 2007; Schiffrin, 1992). Initially placed subclauses would therefore not be expected to be foregrounded.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

Information status of subclause	
INFORMATIONSTATUS	PRIMARY (co-occurring w/ epistemic, secondary phrase) QUOTED (explicitly signalled quote) OPEN (variable contexts)

Type of matrix predicate	
MATRIXPREDTYPE	COM (communicative, cf. H&T Class A) COG (cognitive, cf. H&T Class B) FACTIVE (H&T Class D) ¹⁰ SEMIFACTIVE (H&T Class E) CAUSATIVE OTHER

Sentence adverbials in matrix clause	
MATRIXADVTYPE	MODALP (modal particles) SWEARING (swear words) EPISTEVAL (epistemic/evaluative adverbials) TEXTUAL (cohesive/argumentative adverbials) NEGATION No sentence adverbials (0)

Sentence adverbials in subclause	
ADVTYPE: ¹¹	MODALP (modal particles) SWEARING (swear words) EPISTEVAL (epistemic/evaluative adverbials) TEXTUAL (cohesive/argumentative adverbials) NEGATION
MULTIADV (number of adverbials)	NO (not multiadverbial, i.e., only one adverbial present) YES (multiadverbial, i.e., more than one adverbial present)

Subordinator type	
--------------------------	--

¹⁰ We have excluded from the category of factive clauses what may be called 'equative constructions' or 'predicates of identity' (see e.g. Heycock and Kroch, 1997), of the structure 'the x is (that) subclause', e.g. *the funny thing is that...* In such constructions, the predicate is simply a copula (*være* 'be') and has no effect in itself on the factivity of the subclause. Such clauses were placed in the residual category OTHER.

¹¹ In clauses with more than one sentence adverbial, the type refers to the first sentence adverbial in the clause. The subclauses only very rarely (in less than 2 % of the cases) contain more than two sentence adverbials, and in 84 % of the cases they only have one.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

SUBORDINATOR	AT (<i>at</i> 'that') HV/OM (<i>hv</i> -words 'wh-' or <i>om</i> 'if/whether') ZERO (no subordinator)
Position of subclause in matrix clause	
POSITION	INITIAL (before the matrix clause) FINAL (after the matrix clause, i.e., clause finally)

Table 2: Coding scheme.

4 Results

4.1 Two steps of analyses

After the exclusion of the various irrelevant clauses (cf. Section 3.2.1), the data material contains 3464 complement clauses with sentence adverbials. We have performed multivariate statistical analyses of the data, in this case mixed effects models.¹²

The term 'mixed' refers to the fact that this type of modelling in addition to the so-called 'fixed effects' or 'non-random variables' such as GENDER and LOCALITY, also includes 'random effects', such as the individual speaker (PARTICIPANT). The latter effects are characterized by being non-repeatable: They are sampled randomly from populations of speakers (i.e. they are intended to represent a larger population), and if we were to replicate the sampling we would have to choose other speakers. Hence, this type of statistical model takes into account the non-repeatable effect of the individual speaker by assigning a baseline mean (called an intercept) to each speaker with respect to the dependent variable, as an adjustment for the fact that the behaviour of individuals cannot be expected to be completely determined by the (non-random) social and linguistic factors. A mixed effects model is therefore more conservative than other types of multivariate analyses as

¹² We used the *glmer*-function in R, version 3.0.1, Package *lme4* version 0.999999-2.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

the social and linguistic factors (the fixed effects) are only chosen as significant when they are strong enough to rise above the inter-speaker variation (Baayen, 2008; Johnson, 2009). In addition to the speaker, we also include the specific sentence adverbial(s) (FULLADVERBIAL) in the subclause as a random effect; in clauses with more than one sentence adverbial, each specific sequence of adverbials counts as a separate adverbial (i.e. *ikke, jo ikke* and *jo ikke lige* are treated as three different adverbials).

Reflecting the analyses mentioned in section 3.3.1, we divided the statistical analyses into two main steps. In the first step, including all 3464 complement clauses, the primary factor evaluated was the semantico-pragmatic factor INFORMATIONSTATUS. This factor was included in the analysis as a fixed effect in addition to the four linguistic factors POSITION, SUBORDINATOR, ADVTYPE and MULTIADV and the non-linguistic factors of GENDER and SOCIALCLASS, as well as their interactions.

In the second step, quoted and semantically primary subclauses were discarded in order to focus on the subclauses with OPEN information status (constituting 57% of the complement clauses in the corpus)¹³. The latter have a close to fifty-fifty distribution with respect to word order (45% V>Adv), and this is thus the truly variable context. In the analyses of these clauses, two

¹³ QUOTED subclauses displayed 97% V>Adv, while PRIMARY subclauses had a somewhat lower tendency to V>Adv, 79%. We may tentatively conclude that even a fairly restricted definition of secondary phrases do not suggest that the co-occurring Primary subclauses are necessarily main clauses (because Adv>V word order does occur). This is an important point when comparing analyses of English, like Thompson (2002), with data from other languages, here Danish.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

additional factors were included, as described in section 3.3: MATRIXPREDTYPE and

MATRIXADVTYPE.¹⁴

The resulting best models in no cases include LOCALITY or TIMEOFRECORDING, indicating that the variation is stable within the period from the 1980s to the 2000s, and that no geographical variation is found anymore, either (contrary to earlier findings of dialectal differences, cf. Pedersen (1996)). YEAROFBIRTH was only significant in Copenhagen as an interaction with the linguistic factor INFORMATIONSTATUS (indicating that young Copenhageners produce more QUOTED subclauses with V>Adv than older Copenhageners).

The analyses of linguistic factors and the remaining background variables GENDER and SOCIALCLASS are therefore based on the corpus as a whole, i.e., both Vinderup and Copenhagen informants in both old and new recordings.

4.2 Mixed models analysis including all complement clauses

The results of the first-step analysis, comprising all complement clauses, show that the remaining background variables, GENDER and SOCIALCLASS, are not statistically significant, neither are the linguistic variables POSITION and MULTIADV, but SUBORDINATOR, INFORMATIONSTATUS and ADVTYPE are significant. All in all, we are dealing with a linguistically determined phenomenon.

The best model is shown in Appendix B; and Figure 1 shows the model effects as regards the factors INFORMATIONSTATUS and SUBORDINATOR which interact significantly with each other. The model effect is the probability of V>Adv estimated by the model as a function of a (statistically

¹⁴ The primary reason for not including these factors in the first analysis including also the quoted and semantically primary subclauses is that it would cause massive collinearity with the factor INFORMATIONSTATUS: Quoted and semantically primary subclauses are primarily governed by communicative and cognitive matrix predicates.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

significant) factor when all the other factors are kept constant.¹⁵ As the two factors interact with each other, they are shown together.¹⁶ The results as regards the factor ADVTYPE will be described in relation to the analysis of subclauses with OPEN information status (Section 4.3), since the effects of this factor are essentially the same in the two models.

INSERT FIGURE 1 ABOUT HERE [ALL FIGURES ARE PLACED AT THE END]

When including quoted and primary subclauses, 64% of the subclauses have V>Adv word order (i.e. what is traditionally referred to as 'main clause word order'). When excluding them, the share of V>Adv amounts to 45%.

The plot in Figure 1 shows that the information status of the subclause interacts with the type of subordinator in a very significant way: As regards clauses without subordinator (ZERO) and clauses initiated by *at*, PRIMARY clauses and QUOTED clauses are significantly more disposed to V>Adv word order than clauses of OPEN information status. While this might be construed as proof that the subclauses are in fact main clauses, the fact that PRIMARY clauses have a fair amount of Adv>V (cf. note 13) goes against such a suggestion (note, again, that Adv>V cannot be used in declarative main clauses. When used in main clauses, Adv>V shifts clearly into subjective, typically emotive, meaning). Nonetheless, the finding that PRIMARY and QUOTED clauses (in most cases also being primary information) are much more disposed for V>Adv than other (OPEN) subclauses clearly supports the hypothesis that V>Adv is related to foregrounding (cf. section 2.1).

¹⁵ The predicted probabilities presuppose that the third factor, AdvType is kept at its default level "OTHER".

¹⁶ The plots are made using the plotLMER.fnc-function (Package languageR version 1.4)

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

Clauses initiated by *hv*-words or *om* behave differently as they rarely (in only about 8% of the cases) occur with V>Adv word order except in QUOTED subclauses. In quotes there is no statistically significant difference between *hv/om* and the two other types of subordinators (*zero* and *at*) with respect to the disposition for V>Adv.¹⁷

Contrary to what would be expected from Vikner (1995) and Wiklund et al. (2009), clauses without subordinator are significantly *more* disposed to V>Adv word order than clauses initiated by *at* (and *hv-/om*, except in quotes) regardless of the information status¹⁸.

4.3 Quantitative analyses of clauses of 'open' information status

Before reviewing the mixed-models analyses of OPEN subclauses, we will have a look at the overall distribution of V>Adv word order in this subsection of the data. Figure 2 plots all the individual factors and their tendency for V>Adv.

¹⁷ There are no occurrences of QUOTED subclauses without subordinator displaying Adv>V word order in the data material. Such contexts may therefore be considered 'cited main clauses' and hence outside the envelope of variation. We have kept them in this step of the analyses anyway as neither QUOTED subclauses nor clauses without subordinator (ZERO) are outside the envelope of variation generally.

¹⁸ It may be argued that many of the subclauses included in this step of the analyses would be seen as "cited main clauses" or "cited assertions" by Vikner and Wiklund et al., and these are explicitly excluded from the generalisation advanced by Wiklund et al. (cf. section 2.3.1). This obviously pertains to the QUOTED clauses, but also some of the PRIMARY and OPEN clauses may be seen as "cited" in a broad sense. As described in section 3.3.1, subclauses have only been coded as QUOTED if there is an explicit signal of direct quote in addition to a communicative matrix predicate subclause. This means that a subclause like *han sagde nu gik det jo galt* 'he said now it went wrong' is counted as OPEN even though it may be seen as a quote. However, even when excluding subclauses under a communicative matrix predicate (cf. section 3.3.2), subclauses *without* subordinator are more disposed for V>Adv than clauses with *at* or *hv/om*. The shares of V>Adv are: ZERO 53%, AT: 43% and HV/OM 7%.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

INSERT FIGURE 2 ABOUT HERE

Again, the type of subordinator plays an important role, in that clauses without a subordinator are clearly disposed to V>Adv while HV/OM clauses rarely occur with this word order. However, when including the factor SUBORDINATOR in the same model as the factor MATRIXPREDTYPE, a problem of collinearity arises, as embedded questions (HV/OM clauses) do not occur under factive and causative matrix predicates (indicating that the two factors are correlated). To avoid such collinearity, we exclude the factor SUBORDINATOR from the mixed-models analysis presented below.

Again, GENDER and SOCIALCLASS are not statistically significant, but all the linguistic factors except MULTIADV are. An analysis of the effect of adverbials in the subclause showed that a binary opposition of DIALOGICAL (comprising modal particles and swear words) vs. OTHER (all the rest) resulted in as good a model as the previously mentioned categorisation (called ADVTYPE1 in Figure 2). The simpler categorisation (AdvType) was therefore chosen. For sentence adverbials in the matrix clauses negations were held separate from the other adverbials as they have a very high effect on the word order in the subclause. The best model is shown in Appendix C, and Figure 3 shows the model effects.¹⁹

INSERT FIGURE 3 ABOUT HERE

¹⁹ The predicted probabilities presuppose that all the other factors are kept at their default levels which is MATRIXPREDTYPE=OTHER, POSITION=FINAL, ADVTYPE=OTHER, MATRIXADVTYPE=0.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

From the model effects, it is clear that the most important factors with respect to explaining the word order variation are the type of matrix predicate (MATRIXPREDTYPE) and the type of adverbial in the subclause (ADVTYPE).

Regarding matrix predicates (cf. the upper left diagram in Figure 3), the model shows that both FACTIVE and CAUSATIVE matrix predicates, as expected from the hypothesis of V>Adv as a foregrounding signal, clearly disfavour V>Adv word order: Subclauses of both types are significantly less disposed to V>Adv than any of the other types. As can be seen from Figure 2, however, they do actually have V>Adv word order in about 10% of the occurrences, so the word order distribution is not categorical in this case either.

Subclauses governed by communicative predicates (COM) are significantly more disposed to V>Adv word order than subclauses governed by cognitive predicates (COG), which are again significantly more disposed to V>Adv than OTHER predicates. This, again, supports the hypothesis of V>Adv as a foregrounding signal, since we would expect communicative predicates to frequently govern subclauses that are foregrounded, even when they do not contain any explicit signals of being quotes (as in the QUOTED subclauses, discarded from this step of the analysis). Cognitive predicates will often introduce something important that the speaker or some other person knows or has learned, and these would then be foregrounded. The residual group OTHER mainly serves as a point of comparison. Note, that there is no statistically significant difference between communicative and semi-factive predicates with respect to word order of their subclauses, and both are significantly more disposed for V>Adv than clauses under all other types of predicates. This suggests that semi-factive predicates in spoken language 'lose their factivity' more often than might be expected from their traditional label.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

As regards the effect of the type of adverbial in the subclause (ADVTYPE), the analysis shows that subclauses with dialogical adverbials (primarily modal particles) are much more disposed to V>Adv word order than clauses with other types of adverbials. This result also supports the semantico-pragmatic hypothesis since dialogical adverbials have been suggested in themselves to foreground subclauses and hence to be correlated with V>Adv word order (cf. section 2.3.2).

The distribution of V>Adv word order on subclauses with one versus multiple adverbials (the factor MULTIADV) shown in Figure 2 suggests that the latter have a higher tendency for V>Adv. However, this factor does not come out as significant in the model containing the specific string of sentence adverbial(s) as a random effect, and the reason is a large variation within the different adverbial combinations as regards their disposition for V>Adv: Some, such as *i hvert fald ikke* 'at any rate not' and *bare ikke* 'just/simply not' (which are also quite frequent) are very disposed to V>Adv, while others, such as *nu egentlig* 'after all', only very rarely occur with V>Adv word order.

The effect of adverbials in the matrix clause (MATRIXADVTYPE) is that subclauses under a negated matrix clause are less disposed to V>Adv word order than subclauses under matrix clauses containing other types of adverbials or no adverbial. There are no statistically significant differences between the levels 0, DIALOGICAL and OTHER matrix clause adverbials.

Finally, the POSITION of the subclause in the matrix clause has a large effect on word order (cf. the lower left diagram in Figure 3), initially placed subclauses being much less disposed to V>Adv word order than clauses placed finally. This is also expected from the hypothesis as the former are not likely to be foregrounded (cf. section 3.3.3)

In order to evaluate the relative importance of the various significant factors presented above, a random forest analysis was performed (cf. e.g. Tagliamonte, 2012; Tagliamonte and

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

Baayen, 2012). Here it is possible to include both MATRIXPREDTYPE and SUBORDINATOR due to the fact that a random forest analysis is much less sensitive to collinearity (i.e. correlated factors) (cf. Tagliamonte and Baayen, 2012). The result is shown in Figure 4.²⁰ The analysis shows that SUBORDINATOR is by far the most important factor, followed by MATRIXPREDTYPE and ADVTYPE. The influence of the individual speaker (PARTICIPANT), i.e., the speaker's individual tendency for using V>Adv word order, are intermediate with respect to importance—clearly less important than three most important factors mentioned above, but more than MATRIXADVTYPE and POSITION. The very low importance of the factor POSITION is explained by the fact that the subclauses are only very rarely placed initially (cf. Figure 2), why POSITION only explains a very small part of the variation in the data.

INSERT FIGURE 4 ABOUT HERE

5 Discussion

As described above, we generally find much more V>Adv word order in our subclauses than expected from the tradition of describing V>Adv as 'main clause word order' and from studies using other methodological approaches and other types of data (cf. Julien, 2007; Wiklund et al., 2009; Vikner, 1995). We find variation in all types of complement clauses, but overall, the distributions support the hypothesis that V>Adv word order in subclauses functions as a foregrounder: It is much more common in subclauses otherwise marked as representing direct quotes and in subclauses that are clearly primary in relation to their matrix clauses than in other types of subclauses. Conversely, but also in accordance with the hypothesis, subclauses under

²⁰ The plot was made in R using the *cforest*- and *varimp-functions* (Package *party* version 1.0-6).

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

factive and causative predicates only rarely occur with V>Adv word order. Additionally, subclauses with dialogical sentence adverbials are much more disposed for V>Adv word order than clauses with other types of adverbials, in accordance with previous claims stating that such adverbials have a foregrounding effect on subclauses in themselves. Finally, the effect of the placement of the subclause in relation to its matrix clause supports the hypothesis since initially placed subclauses, a position normally indicating that the proposition of the subclause is not the main point of the utterance but the theme of it, are less disposed to V>Adv word order than subclauses placed sentence finally.

Nonetheless, several results warrant further discussion. We will go through factive predicates, subordinators and adverbials below, before considering an alternative hypothesis.

5.1 Factive predicates

As mentioned, we find 10% V>Adv under factive predicates which is unexpected from the description of pragmatic presupposition in the literature. Along the lines of Simon's (2007) analysis of semi-factive predicates we will argue that also true factive predicates may co-occur with foregrounded subclauses. We shall give just two examples here (15-16).

- (15) *det er jo så meget sjovt at hun har jo været meget ung da*
it is ADV then very funnythat she has ADV been very young when
jeg øh startede hos hende i første klasse
I erm started with her in first grade
'it is very funny that she of course must have been very young when I, erm, started in with her in first grade'

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

- (16) *så det blev vi lidt sure over # at det skulle lige være*
so that became we a-little mad about that it should just be
vores klasse
our class
'so we were quite upset that it had to be exactly our class'

In isolation, it is impossible to ascertain which clause is foregrounded and which is backgrounded, but with additional context it becomes clear that in both cases the subclause is foregrounded, since this is what the speaker follows up on (cf. 17-18).

- (17) *jeg var meget glad for den klasselærerinde jeg fik der hed fru Tram # og det er jo så*
meget sjovt at hun har jo været meget ung da jeg øh startede hos hende i første klasse
fordi da jeg så selv blev lærer der mødte jeg hende faktisk

'I was very happy with the teacher I had, who was called Mrs. Tram, # and it is very funny
that she of course must have been very young when I, erm, started in her first grade
class, because when I later became a teacher myself, then I actually met her'

- (18) *så skulle de vælge hvad for en af dem der skulle skilles # og så har de taget den der*
kunne bedst sammen fordi at # de troede at alle inde fra E-klassen stadigvæk # ville
snakke sammen selv om de ikke gik i klasse sammen # så det blev vi lidt sure over # at
det skulle lige være vores klasse # når det var den der fungerede bedst

'then they chose which one of them [the pre-existing classes] had to be divided # and then
they took the one who got along better because # they thought that everybody from the E

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

Class would still # speak to each other even if they didn't attend the same class anymore
so we were quite upset # that it had to be exactly our class # when that was the one
which worked the best'

In (17) it seems that the relevant matrix clause ('it is funny that') is backgrounded in comparison to the subclause ('she must have been young'), while this is less clear in (18). Here, we would argue that the speaker employs a strategy of foregrounding both matrix and subclause. The slight pause indicated by # suggests that she is focusing her attention separately on each of the relevant clauses. This raises questions for the analysis of foregrounding and backgrounding generally, but does not impede on our conclusion that V>Adv is a foregrounding signal.

5.2 Subordinators

Contrary to Vikner (1995) who seems to suggest that the subordinator *at* is required for V>Adv to be grammatical, we find that subclauses containing this subordinator are *less* disposed to V>Adv than subclauses without an explicit subordinator, also in subclauses which cannot be analysed as 'cited main clauses' in any reasonable sense of the concept.

As already mentioned, we do not find general support for Thompson and Mulac's (1991) claim that subordinator-lacking subclauses are in fact main clauses, either. Several of our examples disprove this for Danish. Here is an example containing two subclauses without *at*, both displaying Adv>V word order, the first governed by a communicative predicate (*sige* 'say'), the second by a semi-factive (*forstå* 'understand').

(19) *hvis der er en af kunderne der ringer og siger de ikke*

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

if there is one of the customers who call and say they not

kan forstå deres girokort ikke er betalt hh så er det mig

can understand their giro_form not is paid hh then is it me

de henvender sig til

they address themselves to

'if there is one of the customers who call and **say** that they cannot understand that their

Giro form isn't paid then it is me they approach'

This is one of the frequent examples of a construction that is difficult to translate into English without making substantial changes of elements that are central to the analysis. In Danish, it is clear from the Adv>V word order that both subclauses involved are subordinate complements to *siger* 'say' and *forstå* 'understand', respectively, even though no subordinators are present. In English, it seems that omission of *that* would render at least the first of the two (*de ikke kan forstå* 'they cannot understand') more direct quote-like. This is not the case in Danish, because the word order is Adv>V. Note, again, that if Adv>V were used in a main clause (i.e., if an analysis as a direct quote was attempted), it would acquire a very specific emotive meaning—and an initial, subjective particle would be required.²¹

On the other hand, we also find typical examples of epistemic, secondary phrases with a V>Adv subclause containing the subordinator *at*.

²¹ A constructed example of such a main clause use of Adv>V word order is:

Gid de dog kunne forstå det

I-wish they ADV could understand it

'I so wish they could understand it'

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

- (20) *når man tænker tilbage på at at man var ikke mere end seksten*
when you think back on that that you were not more than sixteen
år eller sådan noget
years or such something
'when you **think** about the fact that that you were no more than sixteen years old or
something'

In (20), the other linguistic cues support an analysis of the first clause as a secondary phrase: The verb *tænker* 'think' is in the present tense and the subject is a generic pronoun, *man* 'man', which in this context refers to the speaker and thus substitutes for the 1sg pronoun. Nevertheless, we do find (two consecutive instances of) the subordinator *at* 'that' in the subclause.

The diagnostic of Thompson and Mulac (1991) thus does not hold up for Danish. The presence of *at* (or another subordinator) is not necessary for subordinating a clause syntactically in Danish, and on the other hand *at* does not necessarily subordinate the subclause semantico-pragmatically (=background it). However, we would argue that *at* is indexical of subclause status, whereby Adv>V is selected more often than the main clause-like V>Adv in clauses initiated by *at*.

5.3 Adverbials

Our interpretation of the effect of the type of adverbial in the subclause is that modal particles and swear words in themselves function as foregrounders, and that this explains the association between such adverbials and V>Adv word order. Another interpretation is that such adverbials are

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

just congruent with a subclause which is already foregrounded by other means, such as V>Adv word order.

However, arguing that (some) sentence adverbials function as foregrounders in themselves, it is compelling to consider the consequences of having sentence adverbial position as the dependent variable in our study of the distribution of 'main clause' word order. Most importantly of course, only subclauses containing a sentence adverbial (or a negation) are included, and this only covers about a fifth of all subclauses, according to an additional study made on another subcorpus of the LANCHART data (Jensen, 2011). Subclauses without a sentence adverbial may of course also be foregrounded, either by having an initial non-subject, by prosody or otherwise, but it is not unlikely that subclauses with sentence adverbials are foregrounded more often than subclauses in general, and hence also more disposed to 'main clause' word order. Other studies are thus necessary in order to determine whether our findings are representative of other subclauses.

Nonetheless, the different effects of the adverbial types is an important finding because other studies typically only consider the negation – which is probably also part of the reason why these studies generally find a lower proportion of V>Adv than we do.

5.4 Assertion as an alternative hypothesis – or realis as a necessary but not sufficient condition?

Several times we have mentioned an alternative hypothesis for the semantics of V>Adv (or MCP, more generally), i.e. 'assertion'. Assertion and foregrounding may be compatible in many respects, but we would point to just two facts supporting the foregrounding perspective better. Embedded questions of the type with the subordinator *om* 'if, whether' occur with V>Adv, not often, but they do occur. Such embedded questions cannot be assertions, but they can be foregrounded; cf. 21 below representing a foregrounded subclause with V>Adv word order.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

- (21) a. *det var noget med om psykologen kunne ligesom bringe os*
it was something with whether the-psychologist could like bring us
ud af fatning
out of composure
'it had to do with whether the psychologist could like disconcert us'

Secondly, subclauses of predicates such as *think* or *believe* occur very often with V>Adv but are hardly assertions. In both cases the notion of foregrounding seems more promising as an explanation of choice of word order.

On the other hand, if foregrounding is the only motivation for using V>Adv, the scarcity of this word order in clauses initiated by *hv*-words or *om* would imply that they are very rarely foregrounded. We do not have a ready explanation as to why this should be the case.

Another result which does not follow from the hypothesis of V>Adv as a foregrounder, is that presence of a negation in the matrix clause makes the subclause less disposed for V>Adv word order. However, this effect could be interpreted as caused by the conversion of class A and B predicates to class C predicates (i.e. non-asserting predicates) described in Section 2.2.2.

All in all, it seems that assertion, or preferably the less speech act-invoking term *realis* (see, e.g., Palmer, 2001), may be a conditioning factor in addition to foregrounding, but further analyses are needed to ascertain this.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

5.5 Social variables and non-categorical distributions

As described, we hardly find any effect of extra-linguistic variables on the distribution of V>Adv in the data. This suggests that we are dealing with a semantico-pragmatic rather than a sociolinguistic opposition, at least as far as spoken language is concerned. Our study has thus not reproduced the very clear geographical patterning shown in Pedersen's (1996) study of traditional Danish dialects (cf. Section 1). This finding is probably a consequence of the thorough standardization, which has taken place in 20th century Danish (cf. e.g. Maegaard et al., 2013). We might speculate that the reason for the lack of social and regional distribution is that the semantico-pragmatic difference between the two variants makes them unsuitable as social indexes.

On the other hand, there seems to be a large difference between the language use in our data material and written Danish. To all appearances, there is a much tighter association between syntactic subordination and Adv>V word order in written Danish, irrespective of the discursive status of the subclause (Gregersen and Pedersen, 2000; Heycock et al., 2011). Our impression is that the educational system in many cases enforces Adv>V word order in subclauses, at least as far as the written language is concerned. The mere tradition of calling Adv>V 'subordinate clause word order' and V>Adv 'main clause word order' is also likely to leave the impression that it is ungrammatical to use V>Adv in subclauses. However, our study has demonstrated that spoken language use, at least as far as complement clauses are concerned, is quite far from the prescriptive norm, and moreover that usage is much closer to a V>Adv norm when the content of the subclause is foregrounded. It seems that the semantico-pragmatic function of V>Adv in subclauses is allowed to flourish in spoken language, which is not laid victim to the same degree of prescriptivism as written language is.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

Part of the explanation of the non-categorical patterning we see in the data may be found in this norm conflict. Another part may be that speakers change strategy while uttering the complex matrix-subclause construction. In some cases, the speaker may initially plan to formulate the utterance as a matrix-subclause construction, hence using a complement-taking predicate, but in the transition between the two clauses (even after using a complementizer) change strategy and formulate the content of what was originally planned as a subclause as an independent main clause. This would result in what is really a non-completed 'matrix clause' followed by a main clause, even though it may also formally be analysed as a matrix-subclause construction. An important perspective for future research would be to code for symptoms of change in strategy between the matrix clause and the subclause, such as pauses, hesitations, self-corrections, or changes in intonational patterns, and correlate their occurrence with the word order in the (putative) subclause.

6 Conclusion

We have presented the results of a study of word order in spoken Danish complement clauses based on a corpus of sociolinguistic interviews. The complement clauses in the corpus were coded for a number of linguistic factors with the purpose of testing the semantico-pragmatic hypothesis that V>Adv word order is used as a foregrounding device. We find that the results support this hypothesis.

In contrast with traditional descriptions, the results show that complement clauses in Danish are not in general characterized by having Adv>V word order (i.e. 'subordinate clause word order'). Seen as a whole, the complement clauses display 64% V>Adv. Some of these may be analysed as 'cited main clauses' or main clauses with a sentence formed adverbial or interjectional phrase,

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

but even when sorting out clauses which may functionally be analysed as main clauses (although syntactically subordinated), the share of V>Adv is much higher than would be expected if V>Adv in subclauses was a mere product of changes in communicative strategy or the like.

Acknowledgements

We wish to thank Stine Hartmann Bierre, Marie Herget Christensen, Line Dalberg, Cecilie Meldgaard Goth, Liva Hyttel-Sørensen, Sune Sønderberg Mortensen and Randi Skovbjerg Sørensen for help in different stages of the coding process. Thanks also to Kasper Boye, Frans Gregersen, Jeffrey Keith Parrott and three anonymous peer reviewers for valuable suggestions and comments to the manuscript. This article is based on research funded by The Danish National Research Foundation (DNRF63) and The Danish Council for Independent Research | Humanities.

Appendix A: Coding categories

Subordinator (the factor SUBORDINATOR)

Three categories: 1. clauses with *at* 'that' (i), 2. clauses without an overt conjunction (ii), and 3. clauses with *om* 'if, whether' or a *hv*-word 'wh-word' (embedded questions) (iii).

(i) *det var min engelsklærer der sagde til mig at hun syntes*
it was my English teacher who said to me that she thought
ikke at jeg skulle gøre det
not that I should do it
'it was my English teacher who said to me that she did not think I should do it'

(ii) *men jeg vidste hun ville ikke sige det til min far*
but I knew she would not say it to my father
'but I knew she would not tell my father'

(iii) *jeg ved ikke om man måtte egentlig give øretæver endnu*
I know not if one might really give ear-beatings still
men de gjorde det
but they did it
'I do not know whether one was still allowed to box your ears, but they did it'

Position in the matrix clause (the factor POSITION)

Initial: The subclause is placed before the finite verb of the matrix clause (iv):

(iv) *at det så kun har været de sidste to minutter der har været*
that it then only has been the last two minutes which have been

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

hyggelige det er sådan set ligegyldigt

pleasant that is such seen unimportant

'that only the last two minutes were pleasant is just unimportant'

Final: The subclause is placed finally in the matrix clause (see i-iii above), including post-positioning (v):

(v) *det var ikke sådan et sted man sådan egentlig opholdt sig særlig*

it was not such a place *one* like actually stayed self *especially*

meget'

much

'it wasn't a place where you like actually hang out a lot'

Type of matrix predicate (the factor MATRIXPREDTYPE)

Communicative (verba dicendi; Hooper & Thompson class A)

(vi) *så er det så jeg siger til Birger at han ikke skal sejle mere*

'and then I say to Birger that he should not sail any more'

Cognition (verba cogitandi; Hooper & Thompson class B)

(vii) *jeg mener måske at vi øh nok er mere voksne end så mange andre*

'I think maybe that we are probably more grown up than many others'

Factive (Hooper & Thompson class D)

(viii) *så kan man selvfølgelig godt ærgre sig over at man ikke har modtaget mere*

'then of course you may be disappointed that you didn't receive more'

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

Semi-factive (Hooper & Thompson class E; excluding tokens containing both first-person singular (1. sg) and present tense in the matrix clause, if not qualified in some other way - for instance with an explicit mention of the source of the knowledge)

(ix) *altså de ved det kun er for skæg ikke*

'I mean they know it's just for fun right'

(x) *jeg fandt ud af at jeg skulle slet ikke have fokus på det*

'I found out that I shouldn't focus on that at all'

Causative

(xi) *så det er også en af grundene til at vi tør næsten ikke at flytte herfra*

'so that's also one of the reasons that we almost don't dare to move away from here'

Other

(xii) *at det så blev det det var så noget andet*

'that it became that, that was another matter'

man skal også passe på man ikke maler fortiden op som sådan meget idyllisk

'you need to be careful that you don't portray the past as very idyllic

Appendix B: Best model, all complement clauses

Factor	Level	Estimate	p
Intercept		-0.19663	0.208268
Subordinator	hv/om	-2.03118	2.44e-15
	Zero	0.50687	5.05e-05
InformationStatus	Primary	1.09858	9.90e-08
	Quoted	1.53294	0.000387
AdvType	Dialogical	2.21072	5.34e-12
Subordinator by InformationStatus	hv/om : Primary	-0.87977	non-sig
	Zero : Primary	0.01253	non-sig

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

	hv/om : Quoted	3.19891	1.57e-05
	Zero : Quoted	2.02511	6.66e-05
AdvType by	Primary : Dialogical	-0.38152	non-sig
InformationStatus	Quoted : Dialogical	-1.60861	0.017206
AdvType by	hv/om: Dialogical	-1.49428	0.019915
Subordinator	Zero: Dialogical	0.95505	0.026725
Number of observations: 3464, Participants: 90. Fulladverbials: 255			
Random effects:			
Participant (intercept), Variance 0.1505, Std. deviation 0.38795			
Fulladverbial (intercept), Variance 0.6923, Std. deviation 0.83204			
The goodness of fit of the model is good with a concordance statistic C of 0.9059902 and a Somers' D _{xy} of 0.8119804			

Appendix C: Best model, complement clauses with “open” information status

Factor	Level	Estimate	p
Intercept		-0.7555	0.000677
AdvType	Dialogical	2.3846	0.000100
Position	Initial	-1.8461	0.031354
MatrixPredType	Causative	-1.4772	0.007672
	Cog	0.4758	0.003024
	Com	1.2664	3.77e-15
	Factive	-1.5097	1.74e-06
	Semifactive	0.9620	1.33e-07
MatrixAdvType	Dialogical	-0.1240	non-sig
	Negation	-1.0190	0.000125
	Other	0.1326	non-sig
Number of observations: 1979, Participants: 89, Fulladverbials: 168			
Random effects:			
Participant (intercept), Variance 0.20701, Std. deviation 0.45498			
Fulladverbial (intercept), Variance 1.41705, Std. deviation 1.19040			
The goodness of fit of the model is good with a concordance statistic C of 0.8592429 and a Somers' D _{xy} of 0.7184858			

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

References

- Aelbrecht, L., Haegeman, L., Nye, R., 2012. *Main Clause Phenomena*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Andersen, H., 2001. Markedness and the theory of linguistic change, in: Andersen, H. (Ed.), *Actualization: Linguistic change in progress*. John Benjamins, Amsterdam/Philadelphia, pp. 21-57.
- Andersen, J.E., 1986. Adverbier – sprogvidenskabens stedbørn. *Beiträge zur nordischen Philologie* 15, 77-88.
- Andersson, L.-G., 1975. *Form and function of subordinate clauses*. Department of linguistics, University of Göteborg, Göteborg.
- Boye, K., Harder, P., 2007. Complement-taking predicates: Usage and linguistic structure. *Studies in Language* 31, 569-606.
- Bresnan, J., Ford, M., 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86, 168-213.
- Baayen, R.H., 2008. *Analyzing Linguistic Data*. Cambridge University Press, Cambridge.
- Christensen, T.K., 2007. *Hyperparadigmer – en undersøgelse af paradigmatiske samspil i danske modussystemer*, Institut for kultur og identitet. Roskilde Universitetscenter, Roskilde.
- Christensen, T.K., Heltoft, L., 2010. Mood in Danish, in: Rothstein, B., Thieroff, R. (Eds.), *Mood in the Languages of Europe*. John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 85-102.
- Cristofaro, S., 2003. *Subordination*. Oxford University Press, Oxford.
- Diderichsen, P., 1974. *Elementær dansk grammatik*, 3. udg. ed. Gyldendal, København.
- Dik, S.C., 1989. *The theory of functional grammar, Part 1, The structure of the clause*. Mouton de Gruyter, Dordrecht.
- Dines, E.R., 1980. Variation in Discourse: "And Stuff like That". *Language in Society* 9, 13-31.
- Ford, C., Thompson, S.A., 1986. Conditionals in discourse: a text-based study from English, in: Closs Traugott, E., Meulen, A.T., Reilly, J.S., Ferguson, C.A. (Eds.), *On conditionals*. Cambridge University Press, Cambridge, pp. 353-372.
- Glismann, O., 1978. On factives and semi-factives, in: Weinstock, J. (Ed.), *The Nordic languages and modern linguistics / 3*. University of Texas, Austin, pp. 360-365.
- Gregersen, F., 2009. The data and design of the LANCHART study. *Acta Linguistica Hafniensia* 41, 3-29.
- Gregersen, F., Pedersen, I.L., 1997. Hovedsætningsordstilling i underordnede sætninger. *Danske Folkemål* 39.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

- Gregersen, F., Pedersen, I.L., 2000. A la Recherche du Word Order Not Quite Perdu, in: Herring, S.C., Reenen, P.T.v., Schøsler, L. (Eds.), *Textual Parameters in Older Languages*. John Benjamins, Amsterdam, pp. 393-431.
- Halliday, M.A.K., Matthiessen, C.M.I.M., 2004. *An introduction to functional grammar*, 3. ed. Arnold, London.
- Hasan, R., 2009[1989]. *Semantic variation and sociolinguistics*, in: Webster, J.J. (Ed.), *The Collected Works of Ruqaiya Hasan*. Equinox, London/Oakville.
- Heltoft, L., 2005. Ledsætning og letled i dansk. OV-sætningens rester, in: Heltoft, L., Nørgaard-Sørensen, J., Schøsler, L. (Eds.), *Grammatikalisering og struktur*. Museum Tusulanum, København.
- Heycock, C., 2007. Embedded root phenomena, in: Everaert, M., van Riemsdijk, H. (Eds.), *The Blackwell Companion to Syntax*. Blackwell Publishing, Malden, pp. 174-209.
- Heycock, C., Kroch, A., 1997. *Inversion and Equation in Copular Sentences*, Workshop on (Pseudo)Clefts, Berlin.
- Heycock, C., Sorace, A., Hansen, Z.S., 2010. V-to-I and V2 in subordinate clauses: an investigation of Faroese in relation to Icelandic and Danish. *Journal of Comparative German Linguistics* 13, 67-97.
- Heycock, C., Sorace, A., Hansen, Z.S., Vikner, S., Wilson, F., 2011. Residual V-to-I in Faroese and its lack in Danish: detecting the final stages of a syntactic change. *Working papers in scandinavian syntax* 87, 137-165.
- Hooper, J.B., Thompson, S.A., 1973. On the applicability of root transformations. *Linguistic Inquiry* 4, 465-497.
- Høysgaard, J., 1979 [1747]. *Accentuered og raisonered grammatica*, in: Bertelsen, H. (Ed.), *Danske grammatikere*. Det danske sprog- og litteraturselskab, København, pp. 249-488.
- Jensen, E.S., 2000. *Danske sætningsadverbialer og topologi i diakron belysning*. University of Copenhagen.
- Jensen, T.J., 2007. *Betingelse, generalisering og realitet. Hvis-konstruktioner i dansk talesprog*. C. A. Reitzel, København.
- Jensen, T.J., 2011. *Ordstilling i ledsætninger i moderne dansk talesprog*. *Ny Forskning i Grammatik* 18, 123-150.
- Johnson, D.E., 2009. Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass* 3, 359-383.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

Julien, M., 2007. Embedded V2 in Norwegian and Swedish. *Working papers in scandinavian syntax* 80, 103-161.

Karttunen, L., 1973. Some observations on factivity. *Papers in Linguistics* 4, 55-69.

Kiparsky, P., Kiparsky, C., 1970. Fact, in: Bierwisch, M., Heidolph, K.E. (Eds.), *Progress in Linguistics*. Mouton, The Hague, pp. 143-173.

Labov, W., 1978. Where Does the Linguistic Variable Stop? A Response to Beatriz Lavandera. *Working Papers in Sociolinguistics* 44.

Labov, W., 2004. Quatitative analysis of linguistic variation, in: Ammon, U., Dittmar, N., Mattheier, K.J. (Eds.), *Sociolinguistics (Soziolinguistik): International Handbook of the Science of Language and Society*. Mouton de Gruyter, Berlin, pp. 6-21.

Lavandera, B.L., 1978. Where does the sociolinguistic variable stop? *Language in Society* 7, 171-182.

Maegaard, M., Jensen, T.J., Kristiansen, T., Jørgensen, J.N., 2013. Diffusion of Language Change: Accommodation to a Moving Target. *Journal of Sociolinguistics* 17, 3-36.

Meinunger, A., 2006. On the discourse impact of subordinate clauses, in: Molnár, V., Winkler, S. (Eds.), *The architecture of focus*. Walter de Gruyter, Berlin.

Nimb, S., 2004. *Danske adverbier mellem leksikon og syntaks*, Det humanistiske fakultet. Københavns Universitet, København.

Ogle, R., 1981. Redefining the scope of root transformations. *Linguistics* 19, 119-146.

Palmer, F.R., 2001. *Mood and Modality*, 2nd ed. Cambridge University Press, Cambridge.

Pedersen, I.L., 1996. 'Der kan jo være nogen der kan itte tåle det' - Om hovedsætningsordstilling i bisætninger i danske dialekter, in: Ejskjær, I., Jul Nielsen, B., Pedersen, I.L. (Eds.), *Studier i talesprogsvariation og sprogkontakt. Til Inger Ejskjær på halvfjerdsårsdagen den 20. maj 1996*. C.A. Reitzel, København, pp. 242-251.

Pichler, H., 2010. Methods in discourse variation analysis: Reflections on the way forward. *Journal of Sociolinguistics* 14, 581-608.

Romaine, S., 1984. On the problem of syntactic variation and pragmatic meaning in sociolinguistic theory. *Folia Linguistica* XVIII, 409-437.

Sankoff, G., 1973. Above and beyond phonology in variable rules, in: Bailey, C.-J.N., Shuy, R.W. (Eds.), *New Ways of Analyzing Variation in English*. Georgetown University Press, Washington D.C., pp. 42-62.

Schiffrin, D., 1992. Conditionals as topis in discourse. *Linguistics* 30.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

Simons, M., 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117, 1034–1056.

Stalnaker, R., 2002. Common Ground. *Linguistics & Philosophy* 25, 701-721.

Sundquist, J.D., 2003. The Rich Agreement Hypothesis and Early Modern Danish embedded-clause order. *Nordic Journal of Linguistics* 26, 4-34.

Tagliamonte, S.A., 2012. *Variationist sociolinguistics - Change, observation, interpretation*. Wiley-Blackwell, Malden.

Tagliamonte, S.A., Baayen, R.H., 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24, 135-178.

Terkourafi, M., 2011. The pragmatic variable: Toward a procedural interpretation. *Language in Society* 40, 343-372.

Thompson, S.A., 2002. Object Complements and Conversation. *Studies in Language* 26.

Thompson, S.A., Mulac, A., 1991. The discourse conditions for the use of the complementizer that in conversational English. *Journal of Pragmatics* 15, 237–251.

Urmson, J.O., 1952. Parenthetical verbs. *Mind* 61, 480-496.

Vikner, S., 1995. *Verb movement and expletive subjects in the Germanic languages*. Oxford University Press, New York.

Weiner, J.E., Labov, W., 1983. Constraints on the agentless passive. *Journal of Linguistics* 19, 29-58.

Wiklund, A.-L., Bentzen, K., Hrafn Hrafnbjargarson, G., Hróarsdóttir, T., 2009. On the distribution and illocution of V2 in Scandinavian that-clauses. *Lingua*, 1914-1938.

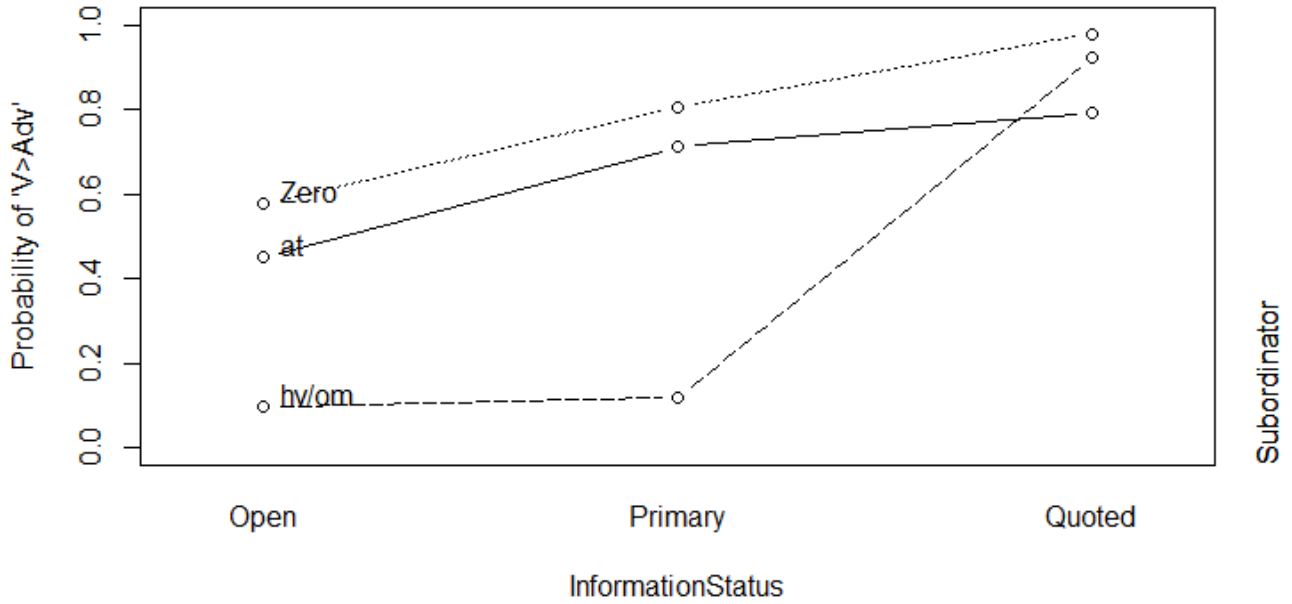


Figure 1: Partial effects of Information status and subordinator, all complement clauses.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

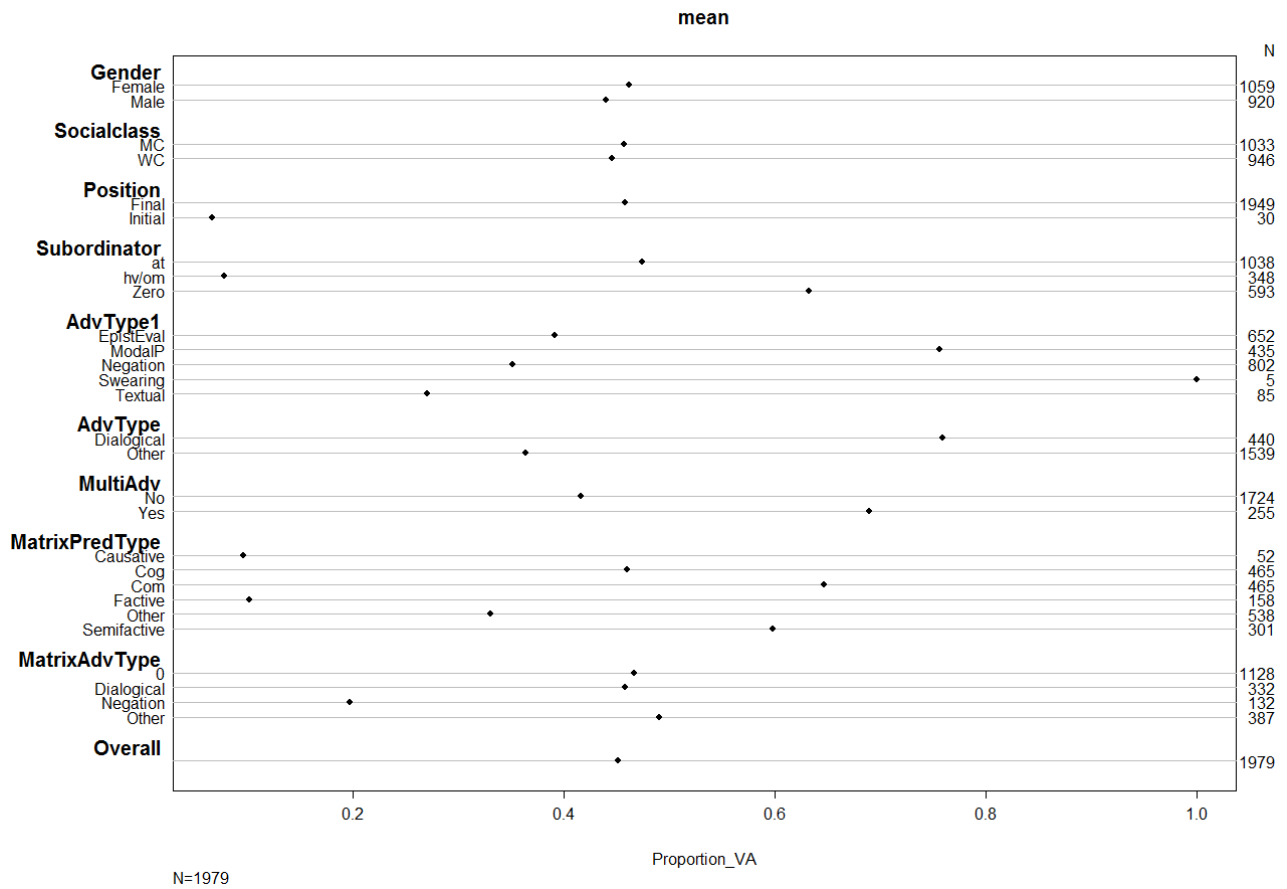


Figure 2: Distribution of V>Adv word order with linguistic and non-linguistic factors, only including complement clauses with 'Open' information status. The factor and level names are explained in Table 2.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

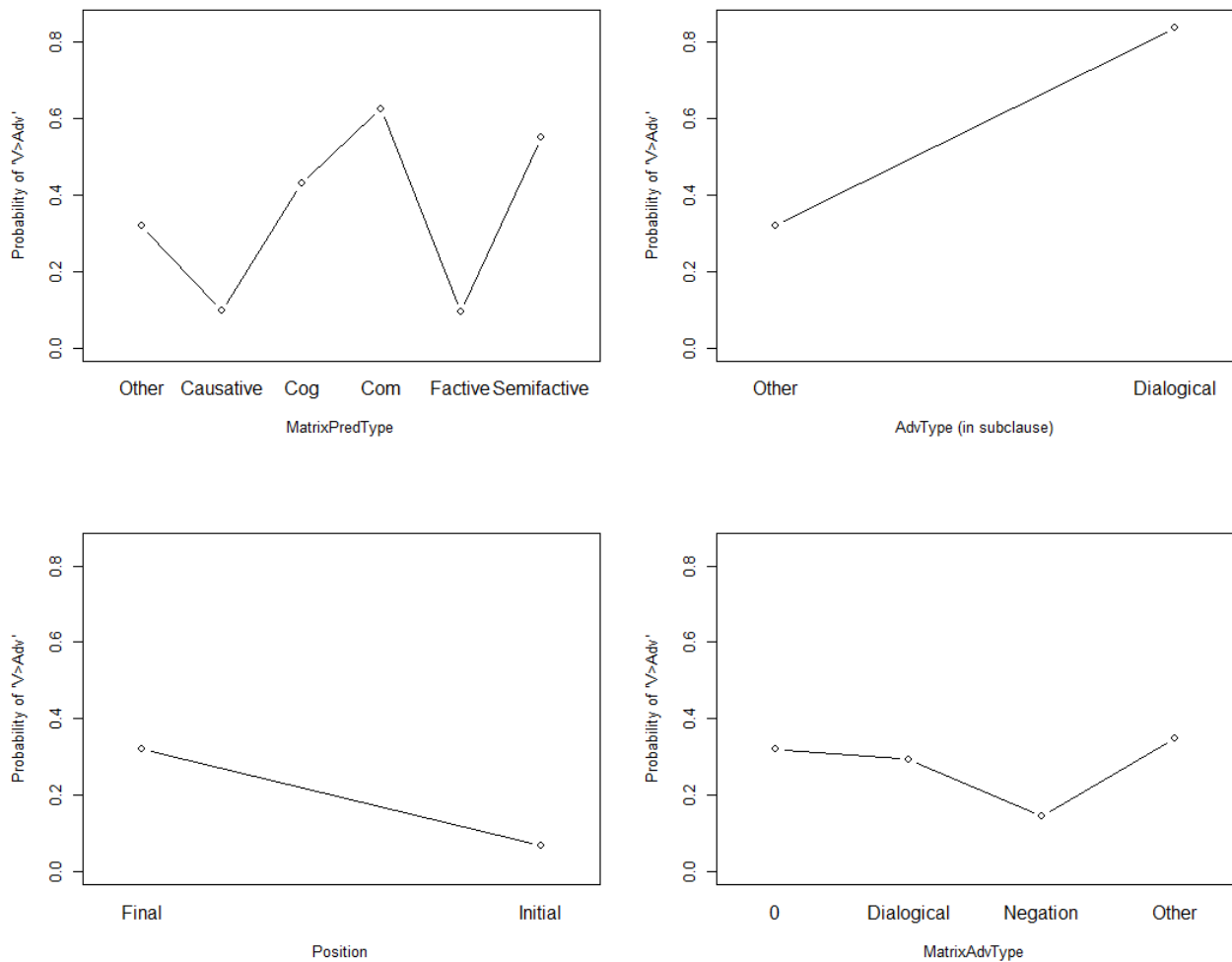


Figure 3: Mixed model effects, complement clauses with OPEN information status.

NOTICE: this is the author's version of a work that was accepted for publication in *Lingua*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Lingua*, Volume 137, Dec 2013, pp. 38-58, <http://dx.doi.org/10.1016/j.lingua.2013.08.005>.

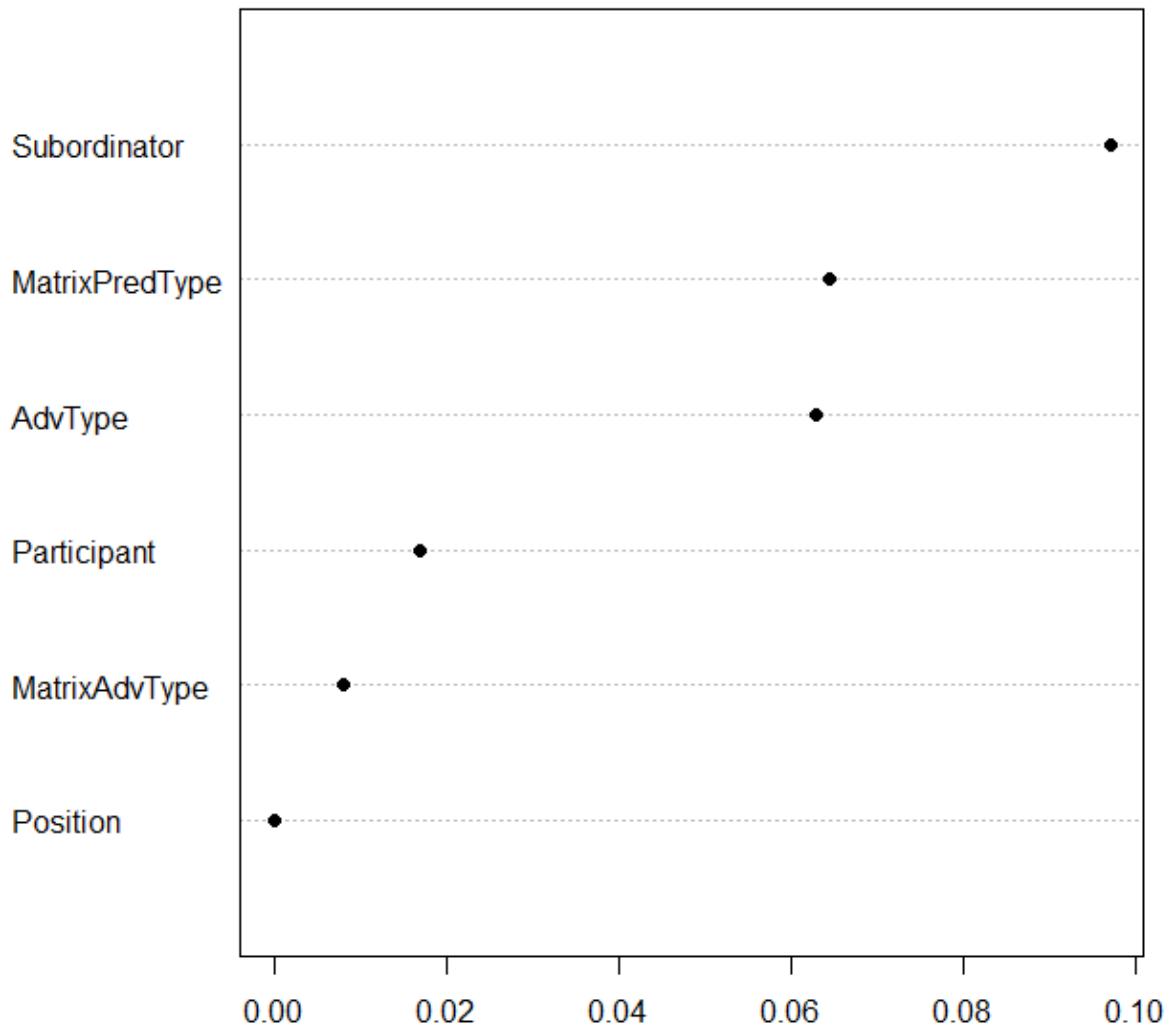


Figure 4: Random forest, complement clauses with OPEN information status.