

How variational acquisition drives syntactic change: The loss of verb movement in Scandinavian

Caroline Heycock
Joel Wallenberg

Abstract Although language acquisition is frequently invoked as a cause of syntactic change, there has been relatively little work applying a formal model of acquisition to an actual case of language change and testing its predictions empirically. Here we test the model of Yang (2000) on the historical case of the loss of verb movement to Tense (V-to-T) in Faroese and Mainland Scandinavian, using quantitative data from a number of corpora. We show that the model straightforwardly predicts the historical data, given minimal and uncontroversial assumptions about Scandinavian syntax. In contrast to a number of previous attempts to explain this repeated pattern of change, it is not necessary to appeal to any bias against learning structures involving V-to-T—a welcome result, given current evidence from acquisition. The newer V-in-situ parameter setting overcomes the original V-to-T grammar because it is more learnable in a language that also has embedded verb-second (EV2). Finally, we argue that the course of the diachronic change is evidence against a strong version of the "Rich Agreement Hypothesis" (RAH), but that under this account the stability of V-to-T in Icelandic provides evidence for the weaker version (cf. Bobaljik 2002).

Keywords Scandinavian, V-to-T, diachronic change, variational acquisition

Caroline Heycock
Linguistics & English Language
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
3 Charles Street
Edinburgh
EH8 9AD
Scotland
United Kingdom
email: heycock@ling.ed.ac.uk

Joel Wallenberg
School of English Literature, Language and Linguistics
Percy Building
Newcastle University
Newcastle upon Tyne
NE1 7RU
England
United Kingdom
email: joel.wallenberg@ncl.ac.uk

1 Introduction

While all the Scandinavian languages are robustly “verb-second” (V2) in their syntax of verb-movement¹, it is well known that there is variation in the placement of the finite verb in contexts where V2 is excluded (for example, relative clauses, indirect questions, complements to certain predicates). In Icelandic, the finite verb always precedes negation and sentence-medial adverbs in such contexts, while in the standard varieties of all the Mainland Scandinavian languages the finite verb follows these elements, as illustrated by the following contrast (taken from Jonas 1996, p. 50):

- (1) a. Ég spurði af hverju Jón (*ekki) hefði (ekki) lesið þessa bók. (Icelandic)
I asked why John (not) had (not) read that book
 ‘I asked why John had not read that book.’
 b. Jeg spurgte hvorfor Jon (ikke) havde (*ikke) læst bogen. (Danish)
I asked why John (not) had (not) read book.DEF
 ‘I asked why John had not read the book.’

This difference has been analysed as a split between languages in which the finite verb moves to T[ense] (or some other head in the “IP domain”) and languages in which it remains *in situ*. Icelandic falls into the first category; Danish and the other mainland Scandinavian languages into the second.² In Faroese the change from the former to the latter type has not yet gone to completion, but it appears to be in a very late stage (Heycock et al. 2010, 2012). As will be reviewed in more detail below, it has been established that V-in-situ is a innovation in each of these languages: all the Scandinavian languages developed from a system that is in the relevant respects like modern Icelandic, with V-to-T as well as V2. Thus the question arises: why is V-to-T repeatedly lost in the history of Scandinavian?

The most widely-adopted answer to date has related this diachronic change to another: the loss of agreement morphology on finite verbs. We will however argue that existing accounts in terms of the Rich Agreement Hypothesis (RAH) do not make the right predictions about the change, either predicting a loss of V-to-T that is earlier and less variable than what is actually observed, or failing to predict change at all.

Instead we will argue that the variational acquisition model of Yang (2000) can predict the loss of V-to-T, given only relatively uncontroversial assumptions about the syntax of the Scandinavian languages, and assumptions about the course of acquisition that are based on recent empirical studies. To summarize: we assume that the grammar that accounts for the vast majority of the input to children at the earliest stage of the change is in essential respects that like that of modern Icelandic—V-to-T is obligatory in all contexts,³ and stylistic fronting is available.⁴ Second, at some point, children are also exposed to some output of a grammar like that of modern Mainland Scandinavian; we will discuss whether this has to be assumed to be as a result of contact or whether it could be due to internal innovation.

¹Note that here and throughout, we do not use “V2” as a simple description of surface order. Rather, a “V2 clause” is one which we assume to exhibit movement of the finite verb to a head higher than T[ense] (typically diagnosed by the possibility of XP–V_{fin}–Subj order), probably in the C-domain. By “V2 language”, we mean a language which obligatorily exhibits this type of movement in matrix clauses. For a recent overview of the V2 phenomenon, see Holmberg (2011).

²References to “Danish,” “Swedish” etc in the text are to the standard varieties of these languages, unless specified otherwise.

³As discussed in Bobaljik and Thráinsson (1998), and in particular detail in Angantýsson (2001, 2007, 2011), Thráinsson (2003, 2010b), “verb third” order, in which the finite verb follows sentence-medial adverbs, does occur as a rare option in certain subordinate clauses in Icelandic. While Bentzen (2007) takes this to arise from lack of verb movement, we follow Angantýsson’s and Thráinsson’s proposal that this marked order is the result of exceptional placement of the adverb. See also footnote 25 below.

⁴As will be seen below, the availability of stylistic fronting is not in fact critical for the account.

Using the Icelandic Parsed Historical Corpus (Wallenberg et al. 2011) as a quantitative model for the original grammar, and estimates derived from corpora of modern Swedish for the incoming grammar, we demonstrate that when the input is of this mixed type, the V-in-situ grammar provides more unambiguous evidence for itself to the learner than the V-to-T grammar does. This gives a slight but sufficient advantage to the V-in-situ grammar in the competition to be acquired by learners. Perhaps counterintuitively, this advantage derives from the possibility of embedded V2 (EV2). In all environments in which embedded root phenomena are excluded (e.g. embedded interrogatives, relatives), wherever there is a diagnostic for the position of the verb (negation or medial adverb), each grammar produces unambiguous output signalling its presence to the learner. Thus for example (2a), cited from Falk (1993, p. 169), unambiguously signals a V-to-T derivation, while (2b) (Falk 1993, p. 172) equally unambiguously is V-in-situ:

- (2) a. æn min guþ brytar eigh niþar þin guþ
*if my god breaks **not** down your god*
 ‘if my god doesn’t break down your god’
 (Old Swedish, *Codex Bureanus*, date: c. 1350)
- b. om thenne her ærende ekki faa en snar ænda
*if these errands **not** get an immediate end*
 ‘if these errands do not come to an immediate end’
 (Old Swedish, date: c. 1470–1490)

However, in environments in which embedded V2 is possible—e.g. certain adverbial clauses and most declarative complements to verbs, as in (3a) and (3b)—all the output of the older V-to-T grammar can be parsed by the V-in-situ grammar (e.g. Verb–Neg order, as in (3a), being parsed as embedded V2), while the new V-in-situ grammar produces outputs that cannot be parsed at all by the V-to-T grammar (i.e. potential embedded root contexts with Neg/Adv–Verb order, as in (3b)).^{5,6}

- (3) a. Men iagh haffuer bidhit för tigh, at tijn troo skal icke om intet warda
*but I have prayed for you that your faith shall **not** on nothing become*
 But I have prayed for you, that your faith will not come to nothing.
 (Luke 22:32, Swedish *Gustav Vasa Bible*, date: 1526/1541)
- b. Tå quinnan sågh, at thet icke war lönlighit
*when woman.DEF saw that this **not** was secret*
 When the woman saw this this was not secret
 (Luke 9:47, Swedish *Gustav Vasa Bible*, date: 1526/1541)

Given the simple learning model of Yang (2000), this advantage of the V-in-situ grammar will inevitably lead to the eventual elimination of the other grammar. There is no need to posit any acquisitional bias against V-to-T, as in other proposals—a welcome result given the best current evidence about the acquisition of these languages. As we will discuss, this model also suggests how the variation between V-in-situ and V-to-T could arise even in the absence of language contact.

⁵It is important to be aware that in the Scandinavian languages, in contrast to e.g. German, embedded V2 is not in complementary distribution with the overt complementizer found in declaratives and glossed as *that*, so the presence or absence of this complementizer does not help resolve the ambiguity. Note also that, as already indicated, the Icelandic V-to-T grammar does allow the Neg/Adv–Verb (“verb third”) order in certain restricted circumstances; see footnote 25 below for a discussion of why this is not a problem for our account.

⁶All the *Gustav Vasa Bible* examples are from *Fornsvenska Textbanken 1.0*, Lars-Olof Delsing, University of Lund, project2.sol.lu.se/fornsvenska/.

However, we will also argue that the stability of Icelandic provides support for the “weak” version of the Rich Agreement Hypothesis.

We proceed by first discussing in Section 2 the Rich Agreement Hypothesis, the predictions that it makes for diachronic change, and the problems that these pose given the historical record. In Section 3, we briefly present the variational acquisition model of Yang (2000). We then discuss the predictions that it makes for the competition between V-in-situ and V-to-T in Section 4, and show how we tested these predictions using data from Icelandic and Swedish to model the properties of the two competing systems. In Section 5 we argue that alternative accounts that rely on an acquisitional bias against V-to-T are, at best, inconsistent with recent studies of acquisition; we also discuss how in our model the variation between V-in-situ and V-to-T could have arisen even without any kind of language contact. Before concluding, we discuss the apparent diachronic stability of V-to-T in the history of Icelandic in Section 6, and present new data from the Icelandic Parsed Historical Corpus (IcePaHC: Wallenberg et al. 2011) showing a temporary peak of V-in-situ orders in the historical record. We argue that this effect is due to a process of calquing written Danish, and that the stability of V-to-T in Icelandic provides evidence for the “weak” version of the RAH.

2 V-to-T and the RAH

2.1 The Rich Agreement Hypothesis

As reviewed in Bobaljik (2002), the split between Icelandic, which exhibits V-to-T, and the standard Mainland Scandinavian languages, which do not, has been ascribed to the difference in the agreement paradigms of the two sets of languages: Icelandic has a rich agreement paradigm (though there is little consensus as to exactly what “rich” means), while the modern standard Mainland Scandinavian languages have lost all agreement morphology on finite verbs (Kratzer 1984, Roberts 1985, Kosmeijer 1986, Holmberg and Platzack 1991). The hypothesis that the syntactic distinction between moving the finite verb to T or leaving it in situ derives from differences in agreement morphology came to be known as the Rich Agreement Hypothesis (RAH). In its strong form, developed in detail in Rohrbacher (1994, 1999) and more recently revived in a modified version in Koenenman and Zeijlstra (2010, 2012), the RAH states that rich agreement⁷ is in a biconditional relation with V-to-T: the finite verb moves out of the VP (in the absence of V2) *iff* the language has rich agreement on finite verbs. An alternative, weaker form, argued for in Bobaljik and Thráinsson (1998), Bobaljik (2002), is that the implication is only one way: rich agreement—for Bobaljik and Thráinsson, defined as the co-occurrence of distinct morphemes for Tense and Agreement—entails verb movement out of the VP, but this movement is also possible in languages without such morphology.

In addition to the correlation between rich agreement and V-to-T in the modern Scandinavian languages, further support for a causal relation between the two phenomena was argued to be provided by diachronic considerations (again, see the summary in Bobaljik 2002, pp. 137ff); as documented in Haugen (1982), Platzack (1988), Falk (1993), Vikner (1997), the loss of V-to-T in Swedish and Danish followed loss of agreement morphology.

⁷For Rohrbacher, rich agreement is defined as distinctive marking of 1st and 2nd person in at least one number of one tense in the regular agreement paradigm; for Koenenman & Zeijlstra, rich agreement is defined as a paradigm that is at least as specified as the least specified pronominal system in the world’s languages, which, following Greenberg’s Universal 42 (Greenberg 1963), they take to involve at least the distinctions [+/-speaker], [+/-participant], [+/-plural]

2.2 Problem for a diachronic account: the strong RAH is too strong

Despite the temporal link between the loss of agreement morphology and of V-to-T, the course of the diachronic loss of V-to-T in Scandinavian is actually problematic for the strong (biconditional) version of the RAH, as already argued in Sundquist (2002, 2003), Bobaljik (2002). For example, Sundquist shows that by 1350 all person distinctions in the agreement paradigm in Danish had been lost; but V-to-T in subordinate clauses in texts from the first half of the 16th century occurs at an overall rate of over 40%, and even in the second half of the 17th century it still occurs at a rate of over 12%. Falk (1993) demonstrates that there is also a lag in Swedish (although she does not draw the same conclusions).⁸ The lag of more than two centuries between the total loss of person agreement morphology in Danish and the total loss of V-to-T means that many generations of speakers acquired a system that allowed V-to-T (at frequencies as high as 45%) in the absence of the person agreement that all proponents of the strong RAH have argued is the necessary condition for the movement.

A case in which V-to-T can be generated at all by a system that has already lost agreement morphology invalidates the strong version of the RAH. The fact that speakers of historical Scandinavian could acquire and produce V-to-T in the absence of rich agreement (even if it was produced variably) is therefore a serious counterexample to both proponents of a strong derivational relationship between agreement and V-to-T (e.g. Rohrbacher 1999) and the acquisitional version of the strong RAH (Koenenman and Zeijlstra 2010, 2012).⁹ Further, the fact that the loss is gradual, and that as far as we can tell it proceeds via changes in the relative frequency of two variants *within* the usage of individuals, is also incompatible with the strong RAH.¹⁰

2.3 The weak RAH is too weak

The weak version of the rich agreement hypothesis (a one-way implication, according to which rich agreement entails verb movement out of the VP, but not vice versa) is, as Bobaljik and Thráinsson point out, consistent with gradual change. If supplemented with further assumptions (e.g. the grammar competition approach of Kroch 1989, 2001, as suggested in Thráinsson 2003, p. 180), it is also consistent with variation within a single speaker. However, given that under this hypothesis V-to-T can be acquired without morphological evidence, we now face the problem of explaining why verb movement is ever lost—*a fortiori*, why it has been lost repeatedly in the history of Germanic. Even though, according to Bobaljik & Thráinsson's analysis, rich morphology is an acquisitional cue to the existence of an elaborated functional structure (Bobaljik and Thráinsson 1998, Bobaljik 2002), their analysis does not provide an explanation for why verb movement is gradually lost when syntactic evidence for it is present, even if only variably. That is, while the

⁸Koenenman and Zeijlstra (2012) suggest that Falk does not in fact show that V-to-T outlives rich morphology in Swedish because she does not separate out data from subordinate clauses where embedded V2 is potentially available, leaving open the possibility that in the intermediate aftermath of the loss of rich agreement what we are actually seeing is a high rate of embedded V2, rather than V-to-T. However, in his work on the history of Danish, Sundquist is careful to avoid this confound, showing that the rates of V-Neg order in the Danish texts hold both in declarative *that*-complements, and—crucially—in other subordinate clauses (Sundquist 2003, pp. 237–238, 241–242). Heycock et al. (2012) also go to some lengths to distinguish between V-to-T and embedded V2 in modern Faroese, *contra* the discussion of Faroese in Koenenman and Zeijlstra (2012).

⁹The same argument could also be made for the loss of V-to-T in English; see, for example, the classic account in Kroch (1989).

¹⁰An anonymous referee points out that for older manuscripts that may have been copied by different scribes it is often hard to be sure whether even a single text reflects the usage of a single speaker. However, a number of Sundquist's sources for Early Modern Danish are series of letters that have not been recopied, and Sundquist is careful to give figures for these individual sources, as well as the aggregated figures for the different time periods.

strong RAH incorrectly predicts an abrupt and early change, the weak RAH incorrectly predicts stasis.

It is worth noting that this problem remains even when additional causes are hypothesized. Thus, for example, Sundquist (2003, p. 246) argues that Danish loses V-to-T because the frequency of stylistic fronting (SF) in Early Modern Danish results in a high degree of structural ambiguity. That is, the negative marker and other sentence-medial adverbs that delimit the left edge of the VP can be fronted via SF to the left of the finite verb, so that an example like (4) is ambiguous between the two parses (V-to-T, with negation fronted even higher via SF; or V-in-situ, with negation in situ above the VP):

- (4) som icke kan skriffuiss paa denne gang
which not can written-be at this time
 ‘which cannot be written at this time’

We believe that Sundquist is right in pointing to the crucial contribution of structural ambiguity and the paucity of unambiguous data for learners, and we will be expanding on this in what follows. But the explanation as it stands has two problems. First, although Sundquist proposes that the prevalence of stylistic fronting is a “contributing factor [that] may have introduced the possibility of embedded-clause word order without verb raising” (p. 246), there is no explanation given for why the combination of V-to-T and stylistic fronting should not be acquired successfully. In fact, Sundquist shows that the frequency of stylistic fronting remains high throughout the Early Modern Danish period that he studies (up through the 16th and 17th centuries). But since this shows that stylistic fronting is successfully acquired, something more is needed to explain why learners should misanalyse some instances of it as V-in-situ, even when evidence for verb movement remains elsewhere. Second, even if we allow for the possibility of such a misanalysis introducing a V-in-situ option, again more is needed to explain why this option takes over and eventually drives out V-to-T altogether, rather than simply remaining as a possible variant.

It should be clear from the above discussion of the strong and weak versions of the RAH that the change in verb movement cannot be straightforwardly explained in terms of the relationship between morphology and syntax; the explanation for the change must lie elsewhere. First, as already argued in Bobaljik (2002), the motivation for V-to-T movement in a given syntactic derivation cannot depend on the presence of specific morphology, as it does in the strong RAH of Rohrbacher (1999). Secondly, the two syntactic variants must be able to coexist for long periods of time after morphology is lost, even if this coexistence is ultimately unstable. In this way, it cannot be the case that V-to-T is unacquirable, as in the strong RAH of Koenenman and Zeijlstra (2010). However, clearly some pressure causes the variation between V-to-T and V-in-situ to eventually resolve in favour of the V-in-situ grammar. The weak RAH of Bobaljik (2002) does not identify such a pressure, so we hypothesize that the pressure is related to the word orders produced by the variants, not to the morphology. If the pressure in favour of V-in-situ is attributable to word order alone, then we can explain how it is that verbal morphology is lost in mainland Scandinavian long before the variation resolves.

The variational acquisition model of Yang (2000) provides the basis for such a word order-based explanation. This model asks the question: does either the V-to-T grammar or a V-in-situ grammar have a selectional advantage (in a Darwinian sense) in terms of how learnable its syntax is, independently of morphological cues? In the next section we give an introduction to Yang’s model of syntactic acquisition and change, and show how it applies to the case of the loss of V-to-T in North Germanic. We also show how Yang’s model predicts an advantage of a V-in-situ grammar

over a V-to-T grammar on the basis of the word order evidence to the learner, thus sidestepping the need for an explanation in terms of morphology.

3 The variational acquisition model of Yang (2000)

Once a new variant has been introduced into the linguistic environment surrounding the learner (however it is introduced) the model in Yang (2000) provides a way to evaluate the “fitness”, in an evolutionary sense, of the two variants. Hence it predicts whether the new variant will deterministically win out over the older variant, or whether the older variant will withstand invasion by the new one (given enough time, and holding other environmental factors equal). Yang adapts a classical “instrumental conditioning” model of learning (Bush and Mosteller 1951, 1958) to syntactic acquisition in a situation of intraspeaker syntactic variation, and evaluates the evolutionary fitness of the syntactic variants in terms of the amount of evidence each variant, or “grammar”, makes available to the learner for acquiring it. Since acquisition during a syntactic change in progress involves a speaker learning, and choosing between, multiple syntactic variants which are all compatible with the same context (or use), the state of affairs Yang describes is identical to the notion of “grammar competition” in Kroch (1989).¹¹ A brief description of Yang’s syntactic learning model for the case of grammar competition follows (for details and the relevant mathematical proofs, see the full description in Yang 2000). For simplicity, the discussion is framed in terms of two competing grammars, but the logic would be the same for multi-way competition.

First, given a mixture of output from two grammars in the data for the learner, G_1 and G_2 , a child is expected to acquire both grammars. In fact, we assume with Yang and other acquisition theorists that children try out many grammars whose output is not present in the data at all, so a child could arrive at this state of variation between G_1 and G_2 regardless of the exact composition of the input. There is solid empirical evidence for such an assumption, going back at least to Labov and Labov (1978); see also Fritzenschaft et al. (1990) for a good example from German. For simplicity’s sake, we will focus on the case where output from both G_1 and G_2 is actually present in the input to the child, and where the child has already narrowed her hypothesis space of possible grammars down to these two.

In acquiring the two grammars, the child assigns some probability (weight) to each, and then continues to update these weights dynamically throughout the learning process, as the child processes each piece of input data she is exposed to in the speech community. Both grammars G_1 and G_2 generate some sentences that unambiguously identify them to the learner; if they did not, then there would be no basis for the learner (or the linguist) for distinguishing them.¹² However, the two grammars most likely also generate some ambiguous sentences, in the sense that they could be the product of either grammar: for example, in our case, all root clauses, since they exhibit V2, require finite verb movement to a functional head higher than T (usually assumed to be C), and so all of these clauses are ambiguous with respect to the V-to-T parameter.¹³ In fact, even subordinate clauses are ambiguous with respect to V-to-T vs. V-in-situ if they do not contain any adverbial elements which could aid the learner in determining the structural position of the verb. Thus, both

¹¹Note that the idea that something like grammar competition must be occurring in the acquisition of syntax goes back at least to Fritzenschaft et al. (1990)

¹²We are here laying aside the issues that arise where one grammar generates a set of strings which is a proper subset of the strings generated by the other grammar. This is an important question for some cases, however.

¹³Note that in the Scandinavian languages—in contrast to English—the kind of movement that we are here calling V-to-T is *not* a prerequisite for the kind of movement involved in V2, as evidenced by the fact that the modern standard Mainland Scandinavian languages have the latter without the former.

clauses in the sentence *Holmes asked if the dog barked* are ambiguous, and the learner has no basis for determining whether a V-to-T or a V-in-situ grammar generated them.

Yang's model states that when the child hears a sentence, she picks a grammar to analyze the sentence, choosing blindly based only on the preexisting weights associated with each grammar (i.e. the weights based on sentences the child heard prior to the current one). If the child hears an unambiguous sentence—e.g. only G_1 could have produced the sentence—then if the child picked G_1 beforehand and used it to try to analyze the sentence, G_1 will be rewarded; otherwise, G_2 will be punished and G_1 will be indirectly rewarded. Either way, G_1 ends up with an augmented weight. However, if the child encounters an ambiguous input—i.e. either G_1 or G_2 , can analyze the string the child hears—then the child will reward whichever grammar she happened to have picked. Ultimately, as the task is iterated many times, the child will most frequently augment the grammar that was most successful in analyzing unambiguous inputs. And, of course, this process can be iterated over a number of generations of learners as well, so the winner of a diachronic competition between two grammars over a long period of history will also be the one that can analyze the most unambiguous sentences in the input to learning.

This is the same thing as saying that the most successful grammar is the one which generates the highest frequency of unambiguous outputs, compared with all the outputs that it generates. When the learning process is iterated over a number of generations, the first generation of learners becomes the second generation's parents (or adult speech community, more generally) and determines the composition of the linguistic input to the second generation's learning. The grammar which can analyze the highest proportion of unambiguous inputs, from the learner's perspective, is by definition also the one producing the highest proportion of unambiguous outputs from the adult speaker's perspective (out of all the outputs it produces). For example, if 30% of the clauses that G_1 generates unambiguously signal G_1 to the learner (and 70% of its clauses are ambiguous), but G_2 generates 40% unambiguous clauses and 60% ambiguous, then ultimately G_2 will win as the learning process is iterated over generations of learners. Yang (2000) also shows mathematically that the proportion of unambiguous sentences a grammar generates is decisive independently of the initial weights of G_1 and G_2 ; that is, the initial frequencies of G_1 and G_2 in the linguistic environment when the competition begins are not relevant for determining the winner. Thus, even if a given grammar begins as an extreme minority variant, if it is detectable to the learner at all, it will eventually win out over the majority variant if it generates/analyzes a higher proportion of unambiguous sentences of its own type than its competitor does. In fact, something like this has to be the case for any internal linguistic change to occur, since a new variant always begins as a minority variant in the speech community.¹⁴

In terms of Darwinian selection, each grammar has a “fitness” which determines how likely it is to “reproduce” itself in a given learner's probability weights and in the acquisition process of future generations of learners:

Fitness(G) = proportion of unambiguously “ G ” clauses it generates out of all
the clauses it generates.

If a grammar, G_2 , has a higher fitness than another grammar, G_1 , i.e. it generates more unambiguous clauses which signal, “I'm a G_1 clause”, then G_2 has an “advantage” over G_1 :

$$\text{Advantage}(G_2 \text{ over } G_1) = \text{Fitness}(G_2) - \text{Fitness}(G_1)$$

¹⁴We say “internal” change, because it would be theoretically possible for a new variant to be introduced as a majority variant in a situation of language contact where a massive migration of new-variant-speakers overwhelmed an existing speech community.

If $\text{Fitness}(G_2) > \text{Fitness}(G_1)$, then G_2 must win in the long run (and vice-versa). In this way, the outcome of any syntactic change is fixed (to the extent that the fitnesses are fixed), once the change begins.

The goal of the remainder of the paper is to test the model in Yang (2000) against the empirical facts of the loss of V-to-T movement in North Germanic. We already know the result of the competition. In all standard varieties of mainland Scandinavian, the V-in-situ grammar won out; in Faroese the process is not yet complete, but V-in-situ is dominant and by at least some measures seems close to totally replacing the V-to-T option (Heycock et al. 2010, 2012).¹⁵ So, reasoning backwards, Yang’s model predicts that the V-to-T grammar must have been less fit than the V-in-situ grammar. That is, the V-in-situ grammar should generate a higher proportion of unambiguously V-in-situ sentences than the V-to-T grammar generates of unambiguously V-to-T sentences. As we will see in the next section, this prediction is borne out, but only if some amount of embedded V2 is an option for at least the V-in-situ grammar.

4 Testing the Variational Acquisition Model

4.1 Design

As just discussed, the fitness of a V-to-T grammar on the one hand, and a V-in-situ grammar on the other, is the proportion of the clauses in the output from each that are unambiguously attributable to that grammar, out of all the clauses in the output from each. Note, crucially, that we are dealing with tokens rather than types of clauses. In our case, as already mentioned, no root clause provides unambiguous data in favor of either grammar, as the effect of V2 is to obliterate the evidence. We will assume that the relative frequency of root to subordinate clauses is constant between the outputs of the two grammars; hence we can simply ignore root clauses, as they will have no effect on the acquisition of V-to-T/V-in-situ. Similarly, subordinate clauses that contain no potential marker for the left edge of the *vP/VP*—that is, no negation or sentence-medial adverb—are also ambiguous between the two grammars. And again, it seems reasonable to assume that the frequency of negation at least should be independent of the choice of grammar, so again we do not need to consider subordinate clauses without negation or a sentence-medial adverb.

The crucial cases for the acquisition of V-to-T/V-in-situ are thus the subordinate clauses that contain negation or a sentence-medial adverb. Of these there are two main sets of clauses to consider. First are the subordinate clauses which do not allow for the possibility of embedded V2 (EV2).¹⁶ From what has been established on the basis of modern Mainland Scandinavian, we know that these include at least all clauses that involve *A'*-movement; for example relative clauses and embedded interrogatives (see Heycock et al. 2010, 2012 for discussion of the distribution of

¹⁵ A survey carried out by a group led by Höskuldur Thráinsson based on questionnaires completed by over 300 Faroese speakers of different ages seems to indicate a higher rate of acceptance of V-to-T than would be expected from the results reported in Heycock et al. (2010, 2012): examples in Thráinsson (2010a) are cited with rates of acceptance at 25%–30%, although it is not entirely clear how these figures are calculated on the basis of the 3-point scale used.

¹⁶ Note that we are assuming that children can learn from data in subordinate clauses, *contra* the strongest possible statement of the degree-0 hypothesis of Lightfoot (1991). We also make the simplifying assumption that children make the same distinctions between EV2-permitting and non-EV2 clause types as adult speakers do. Waldmann (2008)’s data from children acquiring Swedish provides some evidence that this is a reasonable assumption: at around 4 years, children show a similar effect of clause type on EV2 as adults, though the children show more verb-movement overall. A weaker version of degree-0, where children slowly come to use subordinate clause evidence at some point in the acquisition period, is potentially consistent with our analysis and findings; see further discussion in section 5.

	V-to-T output		V-in-situ output	
Root All outputs ambiguous wrt V-to-T. Stalemate	V–Neg Hún borðar ekki fisk (Ice) she eats not fish <i>She doesn't eat fish</i>		V–Neg Hon äter inte fisk (Swe) she eats not fish <i>She doesn't eat fish</i>	
Subordinate Non-EV2 contexts All outputs of each grammar unambiguous. Stalemate	V–Neg ... hvort hún borðar ekki fisk ...whether she eats not fish ... <i>whether she doesn't eat fish</i>		Neg–V ... om hon inte äter fisk ... if she not eats fish ... <i>if she doesn't eat fish</i>	
Subordinate Possible EV2 contexts V-in-situ has an advantage	V–Neg ... að hún borðar ekki fisk ... that she eats not fish ... <i>that she doesn't eat fish</i>		Neg–V ... att hon inte äter fisk ... that she not eats fish ... <i>that she doesn't eat fish</i>	V–Neg ... att hon äter inte fisk ... that she eats not fish ... <i>that she doesn't eat fish</i>
		Fitness of V-to-T grammar		
		Fitness of V-in-situ grammar		

Figure 1: The advantage of a V-in-situ grammar with optional EV2

EV2 in Scandinavian). Also included here are conditionals, and at least a subset of “adverbial” clauses, such as temporal and locative clauses.¹⁷ Such non-EV2 clauses that contain a marker for the left edge of the VP provide unambiguous evidence for the grammar that has produced them: if the finite verb precedes e.g. negation in such a clause, it is the result of a derivation involving V-to-T; if it follows negation it must be in situ (for the exceptional cases of “verb third” discussed earlier see the discussion in section 4.3). Thus in such clauses, neither the V-to-T nor the V-in-situ grammar has a greater fitness than the other: in all cases, whichever grammar generated the clause can be unambiguously determined from the string.

The second main class of subordinate clauses to consider is the type in which EV2 is possible. The cases that are typically considered here are declarative complements to non-factive, non-negative verbs of communication or cognition (e.g. *say*, *assert*, *think*, *believe*); but certain “adverbial” clauses also allow EV2, in particular some rationale, concessive, and result clauses, the latter including the possibly special case of clauses expressing consequence of degree (e.g. *He was so tall that he could not enter the room*). In all these cases, EV2 appears to be optional. In the case of a V-in-situ grammar, if EV2 does not apply the output will have the verb following negation/adverb, and so will be unambiguously attributable to this grammar. If EV2 does apply, on the other hand, the output will be ambiguous as to whether its derivation involved V-to-T or not. So here it is only the cases where EV2 has not applied that will contribute to the fitness of the V-in-situ grammar. In the case of a V-to-T grammar, EV2 is actually string vacuous if the subject is initial; that is, the verb will always precede negation/adverb. But this means that **no** clauses in this context contribute to the fitness of a V-to-T grammar, since all are parseable under a V-in-situ grammar as cases of EV2. Thus the V-in-situ grammar has an advantage if it allows EV2 (making the outputs of the V-to-T grammar parseable) but only employs it for a proportion of cases (still generating structures that cannot be parsed by the V-to-T grammar). This situation is set out schematically in Figure 1.¹⁸ The left column represents outputs of the V-to-T grammar in the three different sets of clause types just discussed; the right column, outputs of the V-in-situ grammar. The shaded areas represent outputs that are parseable *only* by the relevant grammar. The size of the cells is not

¹⁷It may well be that all these cases involve A' movement, see e.g. Haegeman (2010).

¹⁸Note that although the diagram labels the shaded portions as the “fitness” of the two grammars, in fact the fitness of each grammar is the *ratio* of the shaded portion to the unshaded, for each grammar (the ratio of shaded to unshaded in the left and right halves of the diagram).

		V-to-T output	V-in-situ output		
<p>Root</p> <p>All outputs ambiguous wrt V-to-T. Stalemate</p>	V-Neg	<p>Hún borðar ekki fisk (Ice)</p> <p>she eats not fish</p> <p><i>She doesn't eat fish</i></p>	V-Neg	<p>Hon áter inte fisk (Swe)</p> <p>she eats not fish</p> <p><i>She doesn't eat fish</i></p>	
	<p>Subordinate</p> <p>Non-EV2 contexts</p> <p>SF (V-to-T grammar)</p> <p>parseable as V-in-situ</p> <p>Some output of V-in-situ grammar parseable as SF.</p>	<p>V-Neg (overt S)</p> <p>... ef hún borðar ekki fisk</p> <p>...if she eats not fish</p> <p>... <i>if she doesn't eat fish</i></p>	<p>Neg-V (null S)</p> <p>...sem ekki borðar fisk</p> <p>C-REL not eats fish</p> <p>... <i>who doesn't eat fish</i></p>	<p>Neg-V (overt S)</p> <p>... om hon inte áter fisk</p> <p>... if she not eats fish</p> <p>... <i>if she doesn't eat fish</i></p>	<p>Neg-V (null S)</p> <p>... som inte áter fisk</p> <p>... C-REL not eats fish</p> <p>... <i>who doesn't eat fish</i></p>
		<p>V-Neg (null S)</p> <p>... sem borðar ekki fisk</p> <p>... C-REL eats not fish</p> <p>... <i>who doesn't eat fish</i></p>			
	<p>Subordinate</p> <p>Possible EV2 contexts</p> <p>V-in-situ has an advantage</p>	<p>V-Neg</p> <p>... að hún borðar ekki fisk</p> <p>... that she eats not fish</p> <p>... <i>that she doesn't eat fish</i></p>	<p>Neg-V</p> <p>... att hon inte áter fisk</p> <p>... that she neg eats fish</p> <p>...<i>that she doesn't eat fish</i></p>	<p>V-Neg</p> <p>... att hon áter inte fisk</p> <p>... that she eats neg fish</p> <p>...<i>that she doesn't eat fish</i></p>	

	Fitness of V-to-T grammar
	Fitness of V-in-situ grammar

Figure 2: The fitness of a V-in-situ grammar with optional EV2 vs V-to-T with optional SF

intended to correspond to any estimate of the actual proportions of these outputs in the data, but even so Figure 1 shows that given these assumptions the V-in-situ grammar has an advantage.

If we consider the possibility of stylistic fronting (SF), however, the situation becomes more complex. As discussed above with respect to the Early Modern Danish example in (4), SF can front negation or a sentence-medial adverb (among other elements) above a finite verb that has moved out of the VP. The example in (5), cited from Jónsson (1991), is from modern Icelandic, a language for which we have independent evidence that V-to-T is obligatory.

- (5) Þetta er tilboð [sem ekki er ____ hægt að hafna].
this is offer that not is possible to reject
‘This is an offer that cannot be rejected.’

The availability of stylistic fronting has two effects. On the one hand, an example in which negation/adverb has been moved leftward by SF is now parseable by a V-in-situ grammar as a straightforward instance of V-in-situ (this is clearly the kind of situation that Sundquist had in mind in his account of the loss of V-to-T in Danish). So outputs of the V-to-T grammar in which SF of negation/adverb *actually has* taken place are ambiguous, reducing the fitness of the V-to-T grammar; in Figure 2 there is now an unshaded portion of the cell representing non-EV2 subordinate clauses output by the V-to-T grammar. On the other hand, some outputs of the V-in-situ grammar are, equally, parseable by the V-to-T grammar as instances of SF. SF is limited to clauses in which the canonical subject position lacks any phonetic realization (see Holmberg 2006 for an overview), whether in an impersonal construction, as in (4), or as a result of A'-movement, as in (5). Thus exactly those subordinate clauses produced by the V-in-situ grammar that have an “empty” subject position followed immediately by negation are also parseable by the V-to-T grammar as instances of SF. That is, outputs of the V-in-situ grammar in contexts in which SF of negation *could* take place are ambiguous, reducing the fitness of the V-in-situ grammar. If there is an overt subject, there is no such ambiguity:

- (6) a. ... [_{TP} ____ Negation Verb_{fin} ...] b. ... [_{TP} Subject Negation Verb_{fin} ...]
Ambiguous: V-in-situ or V-to-T+SF *Unambiguous: V-in-situ*

Thus, now there is also an unshaded portion of the cell in Figure 2 that represents the output of the V-in-situ grammar in EV2 contexts. The hypothesis that we are pursuing here is that the fitness of the V-in-situ grammar (the proportion of the right-hand-side that is shaded—unambiguously attributable to the V-in-situ grammar) must be greater than the fitness of the V-to-T grammar (the proportion of the left-hand-side that is shaded—unambiguously attributable to the V-to-T grammar).

Recall that the fitness of each grammar is calculated in terms of the proportions of its outputs that are unambiguous; that is, we are concerned with tokens rather than types. In order to determine whether in fact the fitness of the V-in-situ grammar is greater than that of the V-to-T grammar, we need to look at production data associated with the two grammars. The next section sets out how we did this.

4.2 Method

In order to test Yang’s model against the situation where a V-to-T grammar comes into contact, and hence competition, with a V-in-situ grammar, we compiled two samples from naturally occurring data which could represent the two varieties. In order to estimate the fitness of the V-to-T grammar, we used a sample of Icelandic subordinate clauses from the Icelandic Parsed Historical Corpus (IcePaHC: Wallenberg et al. 2011), as Icelandic is a Scandinavian language with obligatory V-to-T movement¹⁹ (see Thráinsson 2007, for an overview of verb-movement in Icelandic and classic references on the topic). While there is no comparable parsed corpus for a mainland Scandinavian language, we assembled samples of modern Swedish subordinate clauses from three different sources: a sample from one of the novels published by *Bonnier* during the years 1976-77 (from *Korp*, (unparsed) corpus of modern Swedish, part of *Språkbanken*—Borin et al. 2012), a sample of clauses from a contemporary blog (also from *Korp*), and clauses from the dataset of caregiver speech cited in Waldmann (2008).²⁰ We also restricted the Icelandic sample to clauses from narrative genre texts, so that the Icelandic sample would better match the *Bonnier* sample, and generally be as colloquial as possible (given the literary nature of all of the texts in IcePaHC).

The samples consisted of subordinate clauses only, for the reasons we mentioned in the last section, and the clauses were categorized into types based on whether they typically allow EV2 or exclude EV2. Because of the ability of Stylistic Fronting to mask the V-to-T grammar in the absence of an overt subject, we also categorized clauses based on whether the subject of the clause was overt or null (i.e. extracted, reduced in conjunction, null expletive, and a few other very rare null subject cases). The estimated fitness and advantage of the grammars could then be calculated based on the sample. Essentially, referring back to the schematic representation in Figure 2, we set aside the top row (main clauses) and calculated the fitness of each grammar as the ratio of instances falling in the shaded and unshaded portions of each column.

Every subordinate clause in the sample contained enough material to potentially diagnose the position of the verb: a preverbal subject (or null subject)²¹, a finite verb (lexical or auxiliary), and

¹⁹Or movement to the relevant head in the I-domain, which we assume is T for the purposes of this discussion. See Bobaljik and Thráinsson (1998) for further discussion of the details.

²⁰Our intention was to use a sample from across a range of novels and blogs within the *Språkbanken* corpus; we relied on the randomization of sentences that had already been carried out by the creators of the corpus. However, after we had extracted and categorized the relevant clauses from this corpus, we discovered that the randomization (for the creators of the corpus, motivated for reasons of copyright) was only carried out per source. As a result, our samples may not be representative of the Swedish of a whole speech community, though they do give us a more comprehensive view of the system of two individual speakers. In the future we intend to check this result by repeating the study with a random selection across a number of different authors.

²¹We did not include sentences with postposed subjects, which can also license stylistic fronting in the absence of an overt

a medial adverb or negation. The negation/adverb diagnostic requires some further explanation. The Swedish clauses all contained negation; given that the Swedish data from Språkbanken was unparsed, we could reliably extract clauses with negation by automatic methods, but could not reliably distinguish the set of medial adverbs in our initial (automatic) extraction of data from the *Bonnier* and blog data. The caregiver data that we received from Waldmann also only included examples with negation. For the Icelandic sample, on the other hand, we included clauses if they contained either negation or a medial adverb. The set of medial adverbs was determined by using the parse (and lemmatization) to extract all the adverbs in IcePaHC which occurred at least once between an auxiliary and a nonfinite verb. Then, we extracted a list of adverbs which occurred following a nonfinite verb in the corpus (i.e. the set of adverbs that can appear clause-finally). We took only medial adverbs which did **not** also occur clause-finally to constitute the set of medial adverbs (at least for the Icelandic in the corpus, since it was an exhaustive list of such adverbs for IcePaHC). We also restricted the Icelandic sample to texts written prior to the year 1600 or after the year 1850, for reasons that we will discuss below in section 6.

The clauses were categorized into different types, by hand for the Swedish data, and automatically (using the corpus annotation) for the Icelandic data: declarative complement clauses,²² consequence of degree clauses, other adverbial clauses, comparative clauses, relative clauses,²³ and indirect questions. The first two types clearly permit EV2 as a possibility across Scandinavian, the latter three types exclude EV2, and the adverbial clauses are something of a mixed bag. We left the adverbial category mixed for reasons to do with the annotation conventions of the Icelandic corpus (the “CP-ADV” category), and counted these clauses as non-EV2 clauses. This means that our estimate of the V-in-situ advantage may be artificially low.²⁴ But as we will show in the next section, this did not affect the result in terms of the existence of an advantage for V-in-situ.

In the end, the Icelandic sample included a total of 1199 subordinate clauses. The Swedish samples included all of the subordinate clauses in the Waldmann caregiver data set (211 clauses), and a randomly selected set of clauses from the *Bonnier* novel (285 clauses) and the blog entries (290 clauses): 786 in total. Space constraints do not permit us to include examples of the different

expletive. Perhaps it would have been more precise to include these, but they are difficult (and sometimes impossible) to identify in clauses containing only a finite lexical verb. They were excluded across the board by specifying that all clauses either have a null subject or an overt subject which precedes the finite verb. In any case, we do not think that their frequency is high enough to be likely to make a difference.

²²There is of course considerable debate as to whether certain predicates (for example, factives, do or do not permit EV2 in their declarative complements. See, among others, Julien (2007, 2010), Wiklund et al. (2009), Wiklund (2010), Heycock et al. (2010, 2012). As this debate is not yet resolved, we took the decision not to attempt to subdivide the declarative complement clauses.

²³Including clauses parsed as “clause-adjoined relatives”, CP-CAR, in IcePaHC. These are clauses containing a gap which have the syntax of non-restrictive relative clauses, but which modify an entire preceding clause rather than a DP. A simple English example would be:

- (i) One critical linguist questioned every assumption we made, which I didn’t enjoy very much.

²⁴We know that some of these clauses are definitely EV2 environments, as shown by the existence of some examples with a finite verb preceding a diagnostic in the Swedish data for this clause type in Tables 1 and 2. However, note that this decision does not affect the calculation of fitness for the Swedish samples at all, since all Diagnostic > V_{fin} orders in all overt-subject clause types are counted as unambiguously V-in-situ clauses, and any V_{fin} > Diagnostic orders in the Swedish data are counted as ambiguous with respect to the V-to-T parameter. The decision of how to count the different clause types as EV2 or non-EV2 only affects the fitness of Icelandic: a V-Neg order is only unambiguous evidence for a V-to-T grammar in a non-EV2 context. Thus, counting all adverbial clauses as non-EV2 for the purposes of this study has the effect of slightly overestimating the Icelandic fitness by however many true embedded V2 clauses this category contains. Since this overestimation does not significantly impact the comparison between Icelandic and Swedish (Swedish has more unambiguous clauses even given this overestimation of Icelandic fitness), we do not feel this decision has had an adverse effect on the results here.

Clause Type	Icelandic		Swed. Novel			Swed. Blogs		
	$V_{fin}>Diag$	$Diag>V_{fin}$	$V_{fin}>Diag$	$Diag>V_{fin}$?	$V_{fin}>Diag$	$Diag>V_{fin}$?
Overt Subject								
Declarative (EV2)	470	5	0	76	4	0	57	6
Degree (EV2)	75	2	0	10	0	1	6	0
Adverbial	264	7	2	116	3	55	95	20
Comparative	31	0	0	0	0	0	0	0
Relative	58	12	0	22	0	0	21	5
Indirect Q	26	1	0	9	0	0	3	3
Null Subject								
Declarative (EV2)	20	44	0	0	0	0	0	0
Degree (EV2)	3	19	0	0	0	0	0	0
Adverbial	7	38	0	0	0	0	0	0
Comparative	0	2	0	0	0	0	0	0
Relative	31	79	0	42	1	0	14	4
Indirect Q	2	3	0	0	0	0	0	0
Unambiguous Clauses:	419		233			182		
Total Clauses:	1199		285			290		
% Unambiguous:	34.9%		81.8%			62.8%		

Table 1: Comparison of V-to-T (Icelandic) and V-in-situ (Swedish) Fitness.
Gray cells: clauses that are unambiguous for the relevant parameter setting.

clause types from the corpora here, but we include a selection in the online appendix.

4.3 Results

Given that the historical change went in the direction of V-in-situ overtaking V-to-T in mainland Scandinavian, Yang’s acquisition model makes the prediction that the V-in-situ grammar has a higher fitness than the V-to-T grammar. Thus, the hypothesis we set out to test was that the proportion of unambiguous clauses generated by the V-in-situ grammar, as estimated by the naturally occurring Swedish samples, is higher than the proportion of unambiguous clauses generated by the V-to-T grammar, as estimated by the Icelandic sample. This hypothesis is confirmed by the data, regardless of which Swedish sample is considered. We had additionally hypothesized that EV2 contexts formed a crucial part of V-in-situ’s advantage. This hypothesis was also borne out: there is no statistically significant difference between the estimated fitness of V-to-T and V-in-situ once these contexts are removed from consideration.

The results by clause type are shown in Tables 1 and 2 for the Icelandic sample and the three Swedish samples. The data is organized by clause type and order of the finite verb with respect to the diagnostic element (i.e. negation, or medial adverb for Icelandic—all abbreviated to *Diag[nostic]*). The Swedish data requires an additional ordering possibility, also illustrated in the online appendix, labeled “?”, which represents a Swedish perfect construction—absent from Icelandic—in which the perfect auxiliary is silent. Since there is no overt finite verb in these clauses, they are all ambiguous with respect to the V-to-T parameter. Gray-shaded cells represent clauses that unambiguously show one of the parameter settings (V-to-T for Icelandic, V-in-situ for Swedish).

The V-in-situ shows an advantage of 0.28–0.47 over the V-to-T grammar, depending on which

Clause Type	Swed. Caregiver		
	$V_{fin}>Diag$	$Diag>V_{fin}$?
Overt Subject			
Declarative	3	26	0
Degree	10	0	0
Adverbial	41	97	0
Comparative	0	0	0
Relative	1	14	0
Indirect Q	0	3	0
Null Subject			
Declarative	0	0	0
Degree	0	0	0
Adverbial	0	0	0
Comparative	0	0	0
Relative	0	16	0
Indirect Q	0	0	0
Unambiguous Clauses:	140		
Total Clauses:	211		
% Unambiguous:	66.4%		

Table 2: V-in-situ Fitness based on Swedish Caregiver data (Waldmann 2008)
Gray cells: clauses that are unambiguous for the relevant parameter setting.

Swedish sample is used for the comparison. For example, using the Swedish Novel data as representative of Swedish, we calculate the Swedish advantage over Icelandic by subtracting the proportion of unambiguous clauses in the Icelandic sample ($419/1199 = 0.349$) from the proportion of unambiguous clauses in the Swedish sample ($233/285 = 0.818$), arriving at an advantage in favor of Swedish of 0.469. The distributions of ambiguous vs. unambiguous clauses are clearly statistically different between Icelandic and Swedish, regardless of which Swedish sample is used ($72.6 \leq \chi^2 \leq 203$ on 1df, depending on Swedish sample, $p \approx 2.2 \times 10^{-16}$). This confirms our hypothesis: according to the comparison between the Icelandic and Swedish samples, the V-in-situ grammar has a greater fitness than the V-to-T grammar.²⁵

The data also show that the EV2 contexts do indeed form a crucial part of the V-in-situ grammar’s advantage. If we only consider clause types which disallow EV2, i.e. indirect questions and relative clauses, and then calculate the fitness of V-to-T and V-in-situ based only on the resulting Icelandic and Swedish samples, the Icelandic sample now appears to have an advantage over Swedish. Depending on which Swedish sample is used, V-to-T now shows an advantage of

²⁵The discerning reader may have noticed that there are 27 Icelandic clauses in Table 1 which show $Diag[nostic] > V_{finite}$ order in clauses without subject gaps, even though we have been assuming that this order is not generable by the V-to-T grammar in a non-Stylistic Fronting context. We take these clauses to be examples of the exceptional high position for adverbs in Icelandic which is discussed in Angantýsson (2011) and mentioned above in footnote 3. Angantýsson (2011, 67-68, 108) shows that these exceptionally high adverbs are restricted in the syntactic and information-structural contexts in which they can occur, and are generally accented in this position. That being the case, we assume that an Icelandic (i.e. V-to-T) speaker would identify this construction as exceptional, and thus it does not provide a way for the V-to-T grammar to parse the $Diag. > V_{finite}$ orders which are generated by the V-in-situ grammar (or, at the most, only those which happen to have the appropriate syntactic/information-structural/prosodic properties). However, when this construction is produced by the V-to-T grammar, it would always be parsable by the V-in-situ grammar. Therefore, we have counted these clauses in the Icelandic sample as ambiguous with respect to the V-to-T parameter, but continue to count $Diag. > V_{finite}$ orders in the Swedish samples as unambiguously V-in-situ.

0.05–0.13 over V-in-situ on the basis of indirect questions and relative clauses alone (based on the same calculation as above, but this time removing the first 4 rows of each table from consideration). However, the advantage is small enough that the distributions of ambiguous vs. unambiguous clauses are no longer significantly different between Icelandic and any of the Swedish samples ($0.14 \leq \chi^2 \leq 3.4$ on 1df, $0.71 \geq p \geq 0.066$, depending on Swedish sample). This means that there may be a small advantage for the V-to-T grammar on the basis of the non-EV2 samples, but it is also possible that the grammars are equivalent in terms of fitness. In either case, it is clear that the large advantage initially shown by the V-in-situ grammar entirely disappears when EV2 contexts are removed.

4.4 Discussion

The comparison of the Icelandic and Swedish sample corpora confirms the prediction of Yang’s model based on the historical facts: since V-in-situ drove out the V-to-T grammar in the history of a number of North Germanic varieties, the fitness of V-in-situ should be greater than that of V-to-T in terms of how often it signals its presence to the learner with unambiguous clauses. This is the case in the sample corpora, if and only if embedded verb-second is a possibility within the V-in-situ grammar. Assuming the sample corpora to be a fair model of variation between V-to-T and V-in-situ parameter settings, any contact between these two grammars at a level that language learners could observe should lead to the V-in-situ grammar replacing the V-to-T grammar. However, when the possibility of EV2 is removed, the advantage of V-in-situ disappears.

Some of the V-in-situ advantage shown by the comparison of Swedish and Icelandic samples appears to be due to factors which might not be directly related to the parametric difference²⁶. First, there is a prescription in Swedish against fronting the finite verb over negation in subordinate clauses that is salient enough to have a name: the *BIFF rule*²⁷. This prescription may be the cause of the very low rate of EV2 in the *Bonnier* novel data (2/204 or 0.9% out of all Declaratives, Degree, and Adverbial clauses, excluding the “?” category). Assuming the prescription is at work here, driving down the production of EV2 when the V-in-situ grammar can nevertheless parse EV2, this has the effect of inflating the estimated advantage of V-in-situ: it now maximizes its advantage by generating more unambiguous exemplars for the hypothetical learner, but it can still parse any amount of $\text{Diag.} > V_{\text{finite}}$ strings in EV2 contexts produced by the V-to-T grammar. This is an unexpected consequence of Yang’s variational model, and one which should be investigated further. However, for the purposes of this study, it is likely that the other Swedish samples are more characteristic of Scandinavian V-in-situ production than the novel data. We would expect blogs to show less influence from prescription since they tend to be more colloquial in style than formally published prose, and the caregiver sample—as spontaneous speech to children—should be the least affected. Indeed, those samples show a much higher rate of EV2 across Declarative,

²⁶It is probably clear to the reader that we have been assuming the existence of some sort of V-to-T parameter in UG, which creates the difference between Icelandic-type languages and Swedish-type languages. While this is a simplifying assumption for our application of Yang’s Variational Learning model, the model itself does not depend on any particular assumptions about whether V-to-T is a parameter or what the parameter space of UG looks like. However, the model does depend on the learner’s ability to identify the “linguistic variable” (in the sense of e.g. Weinreich et al. 1968 Labov 1972) and the possible range of variation within the variable. If, for a particular variable, the child does *a priori* know a parameter space and possible parameter settings from UG, then the child already knows the range of possible variation that she has to consider. Thus, while the model does not depend on the notion of parameters, any particular application of the learning model may be complicated or simplified depending on one’s assumptions about UG and how parameters are defined.

²⁷“I Bisats kommer Inte Före det Finita verbet” (in a subordinate clause negation comes before the finite verb). See <http://sv.wikipedia.org/wiki/Biff-regeln> for a popular description.

Degree, and Adverbial clauses (54/177 or 31% for blogs, 56/214 or 26% for caregivers), and they show rates of EV2 which are not significantly different from each other ($\chi^2 = 0.7$ on 1df, $p = 0.4$).

The estimate for the V-to-T fitness also seems to be diminished by the Icelandic sample having a higher percentage of EV2 clause types (out of all clause types) than the Swedish samples do. We are not currently sure what has produced this effect, or whether it is at all related to the V-to-T movement parameter. For now, we would simply like to note that this type of difference in the distribution of syntactic contexts for a parameter can have a serious effect on the estimated fitness of a parameter setting, and it would be worthwhile in the future to look into whether such a skewed distribution could actually be connected to a particular parameter setting, or whether it is always orthogonal. This is another way of stating the question: are grammatically independent features always statistically independent? The assumption that linguistic features which are thought to apply independently of each other in the grammar are also statistically independent in their application, at least until proven otherwise, underlies much work in the quantitative investigation of syntactic change (cf. Kroch 1989, Santorini 1993, for some examples of building on this assumption).

When the clauses which allow EV2 were removed from consideration, we saw that the V-to-T grammar showed a slight, but not statistically significant advantage over the V-in-situ grammar. It is worth pointing out that much of this effect is due to the possibility of Stylistic Fronting, which allows the V-to-T grammar a possible parse for *Diag. > V_{finite}* strings when a subject gap is present, which is 61% of all relative clauses for the Icelandic sample, and 66%, 41%, and 52% of relative clauses in the Swedish novel, blog, and caregiver samples, respectively. Allowing for SF is certainly a correct step historically, as the attested mainland Scandinavian varieties allowed SF during the time period when the loss of V-to-T took place (see discussion above and in Sundquist 2003). However, we did not assume that the application of SF in order to parse such examples incurred any kind of additional cost to the learner due to the complexity of the derivation, as would be assumed under an acquisition model such as Clark and Roberts (1993). As we mention in Section 5 below, we believe it to be an empirical question whether such a cost really exists in acquisition, and hope to pursue the effects of such a bias on the formal model in future research.

But as the model currently stands, the grammars are more or less tied for fitness without EV2. If this were correct, and there were really no selective pressure on the system one way or another, we would expect a population of V-to-T grammars (i.e. a V-to-T speech community) to be able to withstand invasion from the V-in-situ grammar. Since a new innovation will almost always be introduced into the speech community at an initially low frequency, the V-in-situ grammar could never overtake an established V-to-T grammar unless V-in-situ had a selectional advantage. Under some assumptions about the population of speakers involved, it could be the case that V-in-situ could persist as a minority variant, but under other assumptions, a new V-in-situ variant wouldn't survive at all.

For example, if we assume a finite population of grammars (i.e. a finite population of speakers and utterances), a fixed population size,²⁸ and neutral drift (i.e. no selection or mutation), then contact between two variants will always lead to one variant replacing the other over time. In this special case of neutral drift, the probability of one variant replacing the other is the same as the proportion of that variant in the whole population (this case is a *Moran process* with no selection; Nowak 2006, 97; Moran 1958). In other words, if a V-to-T population were invaded by V-in-situ to the point where 10% of the grammars were V-in-situ, there would be a 0.9 (i.e. 90%) probability that V-to-T would drive out all of the V-in-situ over time by drift alone (a purely

²⁸Of course this is an idealization, but it could apply to a society which has reached its growth limit given environmental circumstances. For instance, this could describe the Icelandic population prior to the 20th century.

random process). In this way, not only could a change to V-in-situ not get off the ground without some advantage, but in fact V-in-situ would be likely to disappear if it were innovated, or if it were introduced by contact at anything less than a massive frequency. If some amount of continuing mutation (e.g. repeated spontaneous innovation of V-in-situ in acquisition) were introduced into the model, then the outcome would be different, but it is unlikely from acquisition evidence that this type of spontaneous innovation would be in favor of the V-in-situ grammar. Indeed, acquisition evidence points in precisely the opposite direction, as we discuss below. For these reasons, then, we conclude that the possibility of embedded V2 is a crucial ingredient in the loss of V-to-T in Scandinavian.

5 Evidence against an acquisitional bias in favour of V-in-situ

A number of different researchers, working with a range of assumptions about language acquisition and language change, have proposed that the crucial driver for the replacement of a V-to-T grammar by a V-in-situ grammar is an acquisition bias of some kind in favor of V-in-situ. This idea has an intuitive appeal, but we believe that at least in the variants with which we are familiar, it makes some wrong predictions.

Watanabe (1994) points out that the difference between an Icelandic-type grammar (with V2 and V-to-T) and a Swedish-type grammar (with V2 and V-in-situ) is potentially unlearnable under the assumption that children only learn from data in root clauses, or root clauses “plus a little bit” (the degree-0 learnability of Lightfoot 1991). Watanabe upholds the constraint of degree-0 learnability and indeed argues that it “explains the [loss of verb movement] that took place in Mainland Scandinavian” (p. 157). He follows the RAH in arguing that agreement morphology is the acquisitional trigger for the verb movement that is observed in Icelandic, which he analyses as movement to AgrS. In the absence of this trigger, the default is for no movement to occur, hence movement to AgrS is predicted to be lost as soon as rich agreement is lost. However, for reasons largely to do with his account of diachronic change in English, Watanabe also argues for the existence for a shorter verb movement option, to Tense. The trigger for the acquisition of this movement is word order, not morphology: if the child observes that the finite verb precedes negation, she will set the parameter in favor of V-to T. This is where two crucial assumptions come into play. The first is degree-0 learnability. The children can only make use of the information about word order that they get from root clauses; but given the V2 nature of Scandinavian, as we have already seen above, root clauses completely destroy word-order evidence concerning V-to-T. How then can the parameter be set? This requires a second assumption: that a parameter setting that involves verb movement is marked (p. 168); in the absence of unambiguous data, the unmarked, default value will be assigned.

This account however has the same problem as the Strong RAH: it predicts a loss of verb movement in the Mainland Scandinavian languages that occurs immediately after the loss of rich agreement, rather than lagging by centuries, and it also predicts that there will be no intraspeaker variability. As discussed in section 2.2, both of these predictions are falsified by the data from Danish, Swedish, and Faroese. As far as we can see, Watanabe is however correct in his analysis of the options available under the assumption of degree-0 learnability; if children cannot make use of information from non-root clauses, there has to be some other way to set the parameter that will distinguish Icelandic from modern Danish in terms of the presence vs. absence of verb movement out of VP, thus forcing what is effectively a strong version of the RAH. But as we have seen, this makes the wrong predictions about the unfolding of the change. As a result, we are led to conclude

that the pattern of the loss of verb movement in Mainland Scandinavian set out in Sundquist (2002, 2003) is evidence against degree-0 learnability itself.

However, this evidence only counts against degree-0 learnability, in the strong form of that hypothesis that Watanabe assumes, in which children are restricted to degree-0 evidence for the duration of the acquisition period. As suggested by a reviewer, one could alternatively hypothesize that children begin their acquisition under the degree-0 restriction, but that this constraint is lifted (perhaps gradually) at some later stage of the critical period. Waldman's 2008 study demonstrates that Swedish children produce a high proportion of adult-like V-in-situ structures from their earliest production of clauses introduced by a subordinator. Given the evidence from the acquisition of Swedish, the constraint would have to be completely lost before the age of 4, at least for the clauses most clearly marked as subordinated (e.g. indirect questions, relatives, etc.), as Waldmann finds that by 4 children no longer produce non-adult-like V-Neg orders in these contexts (Waldmann 2008, p. 235). Such a version of the degree-0 hypothesis would be consistent with our analysis, and in fact, would make an account like ours necessary in order to explain the diachronic change in Scandinavian. One might even go so far as to speculate that differences in the ease of identifying subordination in different clause types (depending on how clearly marked they are) lead to the differences between the clause types with regard to embedded root phenomena. If children initially assume that matrix and subordinate clauses have the same syntax, and then have to learn the matrix/subordinate distinction, they might learn the distinction to varying degrees of success depending on how many cues to subordination exist in each subordinate clause type, leading to different amounts of matrix clause syntax in each type. Of course, this scenario assumes that children would initially show (or expect to hear, if this takes place pre-production) more EV2 than adults, rather than less, on analogy with matrix clauses.

As Watanabe notes, his assumption that V-in-situ is a "default" parameter setting is related to the proposal in Clark and Roberts (1993) that also involves an acquisitional bias in favour of V-in-situ. For Clark & Roberts, there is a bias toward parameter settings that maximize "elegance," which in this context means derivations with the least number of movements. This bias is not strong enough to outweigh unambiguous evidence in favour of movement, but if there is little unambiguous evidence, it is predicted that the more "elegant" parameter setting will win out: V-to-T will lose to V-in-situ, possibly over a number of generations. Like the filtered learning model of Kirby (1999), as well as a number of other models of acquisition and change in the literature, a weakness of this model is that it does not predict the observed intra-speaker variation during the course of the change. Clark et al. (2008), however, show that it is possible to add an acquisitional bias as a "filter" in an extended version of Yang's variational acquisition model, and successfully predict both intra-speaker variability and gradual change. Such a filtered variational learning model could then potentially incorporate the bias against verb movement assumed by Watanabe and Clark & Roberts.

However, there are two arguments against making this move. First, the filtered variational learning model *adds* the possibility of acquisition biases to the original variational learning model described in Yang (2000). But as we have shown, in the case at hand, the original model already predicts the loss of V-to-T when it is in competition with V-in-situ; so adding a bias in favor of V-in-situ is an unnecessary complication.

Second, while it seems intuitively plausible that V-in-situ should be "easier to learn" than V-to-T, or could be a "default" setting that can only be overridden if there is robust evidence for the learner, in fact current studies of the acquisition of verb placement in Scandinavian do not appear to support this view. We know for example that modern standard Swedish has invariant V-in-situ.

If this were also the “default” setting, this would predict that V-to-T should never show up in the course of acquisition. But Waldmann (2008) has shown, in a study of the speech of four children acquiring Swedish, that they do in fact produce examples of V_{fin} -Neg orders in contexts where V2 is excluded, at least until a little more than three years old. An earlier study by Håkansson and Dooley-Collberg (1994) also found the same non-targetlike V-Neg orders in the speech of children, although only with auxiliary verbs. And Heycock et al. (2013) provide evidence that Faroese children up until at least the age of 7 also produce and accept V-to-T at a rate much higher than expected on the basis of the input.²⁹ These data are, on the other hand, compatible with the variational acquisition model; children “try out” the different parameter settings, which are adjusted on the basis of their success in parsing the input. There is no question that the unambiguous data in favour of V-in-situ are infrequent: only embedded clauses in which EV2 is excluded, and of those, only the ones which contain a marker for the left edge of the VP such as negation. Hence it is to be expected that it will take a relatively long time for the V-to-T grammar to be rejected and the parameter set definitively to V-in-situ even when the input is entirely generated by a T-in-situ grammar. We conclude that there are both theoretical and empirical reasons not to include in our model a bias in favour of V-in-situ, *contra* much of the earlier literature.

The scarcity of unambiguous data also opens the possibility that the V-in-situ variant could arise spontaneously, even in the absence of language contact (contact between Faroese and Danish being potentially relevant for the decline of V-to-T in Faroese—an explanation however which cannot straightforwardly be extended to the prior change in Danish itself). Our account does not rely on any acquisitional filter favouring V-in-situ, or a default setting of V-in-situ that can only be overcome by exposure to unambiguous evidence for V-to-T. Instead we assume that initially children try out both V-to-T and V-in-situ, *regardless* of the input they receive. If all the input data is consistent in their word order with V-to-T, then given enough time the probability of the child using the grammar that is not represented in the input will sink to zero, and the parameter will be set in favour of V-to-T. Similarly if all the input data is consistent in their word order with V-in-situ. However, as we have seen, it is the unambiguous data that are crucial for altering the probability of a grammar being used. In a circumstance in which such data are very scarce, it is possible that neither probability will have reached zero at the point where the critical period for language acquisition is over. Thus a child exposed only to data from a V-to-T grammar might nevertheless acquire a variable system with some non-zero frequency of V-in-situ. This system would then produce input for the next generation of learners, and given the advantage of the V-in-situ grammar, its frequency would increase. Of course, the innovation would still need to escape into the adult grammars of a few speakers of a given generation for the change to have a chance to spread.

At this point it could be asked whether our system actually predicts that Icelandic should lose (or perhaps should already have lost) V-to-T; we address this question in the next section.

6 The History of V-to-T in Icelandic and the Weak RAH

Given the possibility that any child, and so any generation, could innovate both the V-in-situ grammar and the V-to-T grammar, we should ask ourselves why the V-to-T system of Icelandic has

²⁹A referee notes that Thráinsson (2010a) reports *adult* acceptance of V-to-T at a rate higher than expected on the basis of the input. Even so, Heycock et al. (2013) includes a direct comparison between older and younger children, with respect to both acceptance and production, and shows that the 10-year olds accept and produce V-to-T at a significantly lower rate than the younger children.

been so diachronically stable up to the present day. The question is particularly worth asking since the V-in-situ grammar could potentially have been introduced into Icelandic either through internal innovation or through contact. Iceland has never been entirely isolated, and there are longstanding familial ties between Iceland and Norway, Denmark, and the Faroe Islands, three places where the change did take place (or is at least well under way), not to mention the political and social ties with Denmark beginning with Danish-Norwegian rule in 1262.³⁰ We see two possible explanations for the lack of change in Icelandic: the first is chance, and the second is the weak version of the Rich Agreement Hypothesis (Bobaljik 2002, Bobaljik and Thráinsson 1998).

First, there is the possibility that the V-in-situ grammar was in fact innovated (or introduced) in Icelandic at some point, once or multiple times, but it simply never spread beyond one or two people due to chance. In evolutionary dynamics, the chance of a variant spreading, even if it is an advantageous variant, is partially determined by the process of “neutral drift”, as briefly mentioned above in 4.4. Neutral drift is the idea that in small, finite populations, there is a significant chance that variants can die off or otherwise fail to spread, even if they have an advantage over the majority variant in a population (see Nowak 2006, Chapter 6, for a mathematical introduction and references on the topic). Depending on the population size, new variants have a certain chance of accidentally being killed off by randomly applied processes of death, and even an advantageous variant will need to be introduced over and over again until it eventually has a chance to catch on and spread through the population (see the discussion of drift and linguistic change in Nettle 2012, and references therein). In linguistic terms, we must consider the population of utterances (where grammars are expressed), as well as the population of humans who produce the utterances and learn the grammars.

In the case of pre-20th century Iceland, we could expect both of these populations to be relatively small. First, the human population of Iceland has always been small compared to mainland Scandinavia, and high mortality in pre-20th century Iceland could easily have killed off individual children who had innovated the new grammar. Additionally, families in pre-20th century Iceland tended to be widely dispersed around the perimeter of the island, which means that children would have had a limited number of conversational partners to interact with (primarily their parents), which limits the number of utterances playing a role in acquisition (*cf.* the discussion of New Zealand in Trudgill 2004). While it is beyond the scope of this article to develop a fully articulated mathematical model of this situation, we recognize that this should be done before we rule out the possibility that chance stopped the spread of V-in-situ in Iceland. In principle, it would be possible to calculate the number of times V-in-situ would need to have been innovated before it could spread and take over, given estimates for the size of the population of speakers and the population of utterances. Note also that one consequence of a drift-based explanation for the stability of V-to-T in Icelandic is that V-to-T could cease to be stable under current conditions in Iceland, where the population is growing much faster than in previous generations and is much more concentrated (around Reykjavík) than it was prior to the 1940s.

Alternatively, the weak version of the RAH, while it did not explain why V-to-T was lost in so many varieties, does predict its retention in Icelandic. If the system of Tense and Agreement affixes in Icelandic provides evidence for the “Split-IP” structure in the way that Bobaljik (2002) and Bobaljik and Thráinsson (1998) argue, and this structure requires V-to-T, then Icelandic children

³⁰In fact, the Danish scholar Rasmus Rask was so concerned about the effect of Danish-Icelandic contact on Icelandic that he predicted in a letter (c. 1813) that hardly anyone in Iceland would understand Icelandic 200 years later, unless a conscious effort was made to preserve Icelandic (quoted in Ottósson 1990, 52, citing Rask 1941). Danish rule continued until Iceland’s full independence in 1944.

have a powerful, unambiguous cue for the existence of this kind of movement. In fact, since this cue would be present in a large number of verb forms, it would be a much more frequent source of evidence than the diagnostics we have been discussing in this article, which are only present in a subset of subordinate clauses. From this perspective, the fact that V-in-situ provides more word order evidence to learners than V-to-T does would pale in significance compared to the constant stream of morphology-based evidence signaling the V-to-T grammar.

If the weak RAH is correct, it also forces our hand in explaining a potentially problematic set of data from the history of Icelandic. We mentioned above in section 4.2 that we excluded texts from the Icelandic sample which were written between 1600-1850 (inclusive). The reason for this decision should be apparent from the data in Figure 3 below. Surprisingly, given what we have said thus far, the texts in IcePaHC written between 1600-1850 show a mixture of V-to-T and V-in-situ grammars, rather than the uniform V-to-T system that one would expect from modern Icelandic and Old Icelandic. Obviously, such texts could not be used as part of a sample to represent a pure V-to-T grammar. But the historical fact shown below raises a much more serious concern for our argument: if a V-in-situ grammar truly came into widespread contact with a V-to-T grammar in 17th century Icelandic, how is it possible that the V-to-T grammar recovered and remains the system of modern Icelandic? One might argue that the V-in-situ grammar did escape into the Icelandic speech community at an appreciable frequency at some point, but that the weak RAH explains why V-in-situ did not go on to replace V-to-T. However, if the weak RAH is correct, then it is surprising that the V-in-situ grammar could ever reach the levels suggested by the right-hand facet of Figure 3 below (clauses with overt subjects). (Note that the left facet of the graph is clauses with subject gaps, which is irrelevant for the discussion except to show that stylistic fronting has a serious effect on the position of diagnostics with respect to the finite verb.)

While we can only really speculate as to the interpretation of this dataset, we suggest that the apparent V-in-situ in 17th and 18th century Icelandic was a phenomenon of the very literary written language, and never affected the spoken language. This idea makes some sense sociologically, since most of the Icelandic authors during that period were educated in Copenhagen, and it is possible that some of them consciously tried to imitate the Danish authors they would have been exposed to, as well as the language of the local Danish colonial aristocracy. It is clear, for instance, that a number of Danish loanwords appear in Icelandic texts from the 17th and 18th centuries which are rare (or nonexistent) in the earlier or later traditions of Icelandic writing. Certainly this is the only obvious explanation of the data which is compatible with the weak RAH: if Tense and Agr morphology is a cue for V-to-T, then we would not expect any period of decline in V-to-T in spoken Icelandic.

A possible objection to this idea is the way in which clause type interacts with the rate of verb movement over diagnostics during this period, which might indicate that there is some effect from native (or near-native) competency in the V-in-situ grammar. The clause types in the chart are divided into EV2 contexts and non-EV2 contexts in the same way we discussed above, with the difference that adverbial clauses are kept separate since they contain a mixture of EV2 and non-EV2 clauses (see discussion at the end of section 4.2). As one can see in the right facet of the graph, the frequency of verb movement over negation and diagnostic adverbs follows the expected pattern for a V-in-situ grammar with EV2. The EV2 contexts show the highest frequency of verb-movement across the diagnostics, the non-EV2 clauses show the least, and then adverbial clauses fall between the other two types, which is expected if they contain some EV2 and some non-EV2 clauses. Of course, the overall frequency of verb-movement is much higher than we would expect for a V-in-situ grammar, in all clause types, since most of the observed behavior of these texts is

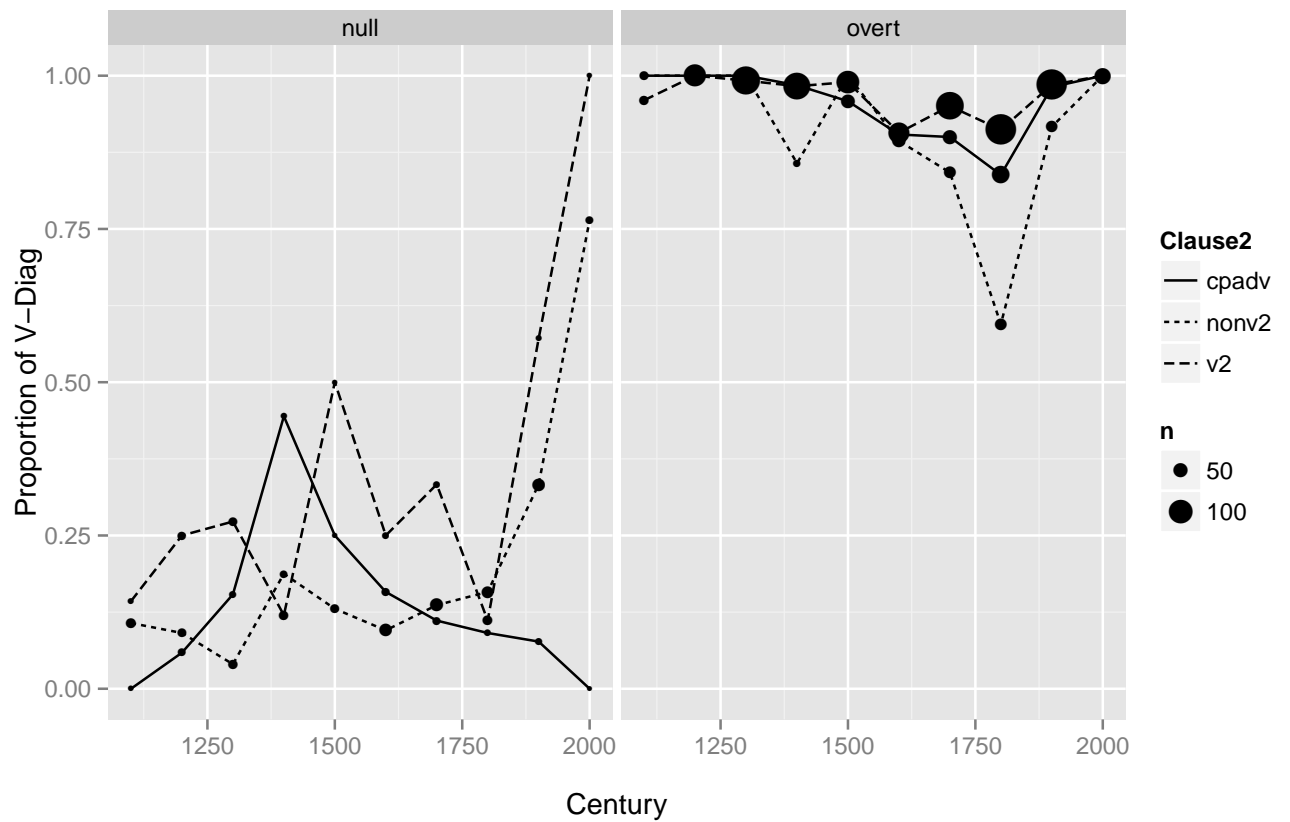


Figure 3: Proportions of Stylistic Fronting and V-to-T over the history of Icelandic (subordinate clauses only, from all text genres). Left facet: null subjects; Right facet: overt subjects.

still from the native Icelandic V-to-T grammar. If all of the Diag. $> V_{fin}$ order were simply calqued on the Danish order by authors with no native competence in a V-in-situ grammar, then we might not expect the statistical effect of clause type to be quite so regular.

However, this is only an objection if the interference of the native Icelandic grammar on the authors' production is rather irregular. Icelandic does show EV2 which is distributed by clause type in a similar way as it would be in a V-in-situ grammar (see Angantýsson 2011, and references), though EV2 would be string-vacuous for these particular diagnostics in Icelandic. It could be the case that the authors are really trying to calque Danish word order in all subordinate clause types, but they make mistakes in the calquing which are distributed in a way that reflects their native competence in EV2. In other words, it is harder for speakers to regularly calque Diag. $> V_{fin}$ order in clause types which favor EV2, because the interference from their native grammar is in the direction of more verb-movement (EV2 plus V-to-T, instead of the simple pressure from their native V-to-T). Of course, this is only currently a thought-experiment concerning how applying a written prescription might interact with a statistical affect from one's native grammar. More detailed, statistical research on how calquing and translation effects distribute in written language is needed in order to decide if this scenario is plausible or not.

7 Conclusion

Many previous accounts of the loss of V-to-T in Faroese and Mainland Scandinavian Scandinavian have pointed to the scarcity of unambiguous evidence available to the child as a crucial factor. However, we have argued that none of the available accounts has a mechanism that allows it to predict both the change and its gradual spread through the grammars of individual speakers over the course of generations. In this paper we have argued that the variational acquisition model of Yang (2000) provides such a mechanism, given the assumption that the V-to-T grammar was in competition with a V-in-situ grammar that optionally allows embedded V2, and taking data from corpora of Icelandic and Swedish as model for the outputs of a V-to-T and V-in-situ grammar respectively. This competition could have arisen through contact, or by spontaneous innovation resulting from the scarcity of the unambiguous evidence for the V-to-T grammar. We have shown, further, that this account does not require us to posit a "default" setting or other acquisitional bias in favour of V-in-situ; we take this as a significant advantage of our account, given evidence from the acquisition of Swedish and Faroese that even with input containing no (Swedish) or few (Faroese) unambiguous cases of V-to-T, children go through a stage where they postulate this as an option for the language they are acquiring. Finally, we have argued that because of the advantage the V-in-situ grammar has in acquisition, a V-to-T grammar could only be truly stable diachronically if it showed some evidence to the learner other than word order, e.g. agreement morphology. In this way, our results not only constitute an argument against the strong version of the RAH, but also indirectly support the weak version over an alternative where there is no link between morphology and the possibility of V-to-T.

Acknowledgments

We would like to thank Lars Borin for his help in accessing and manipulating the datasets from Språkbanken, and Christian Waldmann for making available to us his dataset from the speech of Swedish caregivers. We would also like to thank Kajsa Djärv for her work in collating and

categorizing the relevant tokens from the Språkbanken datasets. Finally, we are grateful for the feedback that we have had from audiences at the Workshop on Verb Movement (Edinburgh 2012), the 27th Comparative Germanic Syntax Workshop (Yale 2012), the 14th Diachronic Generative Syntax Conference (Lisbon 2012), the Formal Ways of Analyzing Variation workshop at the 25th Scandinavian Conference of Linguistics, from Anthony Kroch who commented on an early draft, and from Jonathan Bobaljik and the three referees who helped us improve the final version of this manuscript.

References

- Angantýsson, Ásgrímur. 2001. Skandinavísk orðaröð í íslenskum aukasetningum. *Íslenskt mál* 23.
- Angantýsson, Ásgrímur. 2007. Verb-third in embedded clauses in Icelandic. *Studia Linguistica* 61:237–260.
- Angantýsson, Ásgrímur. 2011. The syntax of embedded clauses in Icelandic and related languages. Doctoral Dissertation, University of Iceland.
- Bentzen, Kristine. 2007. Order and structure in embedded clauses in Northern Norwegian. Doctoral Dissertation, University of Tromsø.
- Bobaljik, Jonathan. 2002. Realizing Germanic inflection: Why morphology does not drive syntax. *Journal of Comparative Germanic Linguistics* 6:129–167.
- Bobaljik, Jonathan, and Höskuldur Thráinsson. 1998. Two heads aren't always better than one. *Syntax* 1:37–71.
- Borin, Lars, Markus Forsberg, and Johan Roxendal. 2012. Korp – the corpus infrastructure of Språkbanken. In *Proceedings of LREC 2012*.
- Bush, Robert, and Frederic Mosteller. 1951. A mathematical model for simple learning. *Psychological Review* 58:313–323.
- Bush, Robert, and Frederic Mosteller. 1958. *Stochastic models for learning*. New York: Wiley.
- Clark, Brady, Matthew Goldrick, and Kenneth Konopka. 2008. Language change as a source of word order correlations. In *Variation, selection, development: Probing the evolutionary model of language change*, ed. Regina Eckardt, Gerhard Jäger, and Tonjes Veenstra, 75–102. Berlin: Mouton de Gruyter.
- Clark, Robin, and Ian Roberts. 1993. A computational model of language learnability and language change. *Linguistic Inquiry* 24:299–345.
- Falk, Cecilia. 1993. Non-referential subjects in the history of Swedish. Doctoral Dissertation, University of Lund.
- Fritzenschaft, Agnes, Ira Gawlitzek-Maiwald, Rosmarie Tracy, and Susanne Winkler. 1990. Wege zur komplexen syntax. *Zeitschrift für Sprachwissenschaft* 9:52–134.

- Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of grammar*, ed. Joseph Greenberg, 73–113. Cambridge, MA: MIT Press.
- Haegeman, Liliane. 2010. The movement derivation of conditional clauses. *Linguistic Inquiry* 41:595–621.
- Håkansson, G., and S Dooley-Collberg. 1994. The preference for modal + neg: an L2 perspective applied to Swedish L1 children. *Second Language Research* 10:95–124.
- Haugen, Einar. 1982. *Scandinavian language structures*. Minneapolis: University of Minnesota Press.
- Heycock, Caroline, Antonella Sorace, and Zakaris Svabo Hansen. 2010. V-to-I and V2 in subordinate clauses: an investigation of Faroese in relation to Icelandic and Danish. *The Journal of Comparative Germanic Linguistics* 13:61–97.
- Heycock, Caroline, Antonella Sorace, Zakaris Svabo Hansen, and Frances Wilson. 2013. Acquisition in variation (and vice versa): V-to-T in Faroese children. *Language Acquisition* 20:5–22.
- Heycock, Caroline, Antonella Sorace, Zakaris Svabo Hansen, Frances Wilson, and Sten Vikner. 2012. Detecting the late stages of syntactic change: The loss of V-to-T in Faroese. *Language* 88:558–600.
- Holmberg, Anders. 2006. Stylistic fronting. In *The Blackwell companion to syntax*, ed. Martin Everaert and Henk van Riemsdijk, volume IV, chapter 65, 532–565. Oxford: Blackwell.
- Holmberg, Anders. 2011. Verb second. In *Syntax—an international handbook of contemporary syntactic research*, ed. T. Kiss and Artemis Alexiadou. Walter de Gruyter, 2 edition.
- Holmberg, Anders, and Christer Platzack. 1991. On the role of inflection in Scandinavian syntax. In *Issues in Germanic syntax*, ed. Werner Abraham, Wim Kosmeijer, and Eric Reuland, 93–118. Berlin: Mouton de Gruyter.
- Jonas, Dianne. 1996. Clause structure and verb syntax in Scandinavian and English. Doctoral Dissertation, Harvard University.
- Jónsson, Jóhannes Gísli. 1991. Stylistic fronting in Icelandic. *Working Papers in Scandinavian Syntax* 48:1–43.
- Julien, Marit. 2007. Embedded V2 in Norwegian and Swedish. *Working Papers in Scandinavian Syntax* 80:103–161.
- Julien, Marit. 2010. Embedded clauses with main clause word order in Mainland Scandinavian. Published on *LingBuzz*: (<http://ling.auf.net/lingBuzz/000475>), December 2010.
- Kirby, Simon. 1999. *Function, selection, and innateness*. Oxford: Oxford University Press.
- Koenenman, Olaf, and Hedde Zeijlstra. 2010. Resurrecting the rich agreement hypothesis: Weak isn't strong enough. In *Movement in Minimalism: Proceedings of the 12th Seoul Conference on Generative Grammar*, ed. Duk-Ho An and Soo-Yeon Kim. Seoul: Hankwukmwunhwasa.

- Koenenman, Olaf, and Hedde Zeijlstra. 2012. One law for the rich and another for the poor: The Rich Agreement Hypothesis rehabilitated. Ms, University of Amsterdam. Accessed in lingbuzz archive, May 2012.
- Kosmeijer, Wim. 1986. The status of the finite inflection in Icelandic and Swedish. *Working Papers in Scandinavian Syntax* 26:1–41.
- Kratzer, Angelika. 1984. On deriving syntactic differences between English and German. Ms. Technical University, Berlin.
- Kroch, Anthony. 2001. Syntactic change. In *The handbook of contemporary syntactic theory*, ed. Mark Baltin and Christopher Collins, 629–739. Malden, MA: Blackwell.
- Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William, and Teresa Labov. 1978. Learning the syntax of questions. In *Recent advances in the psychology of language*, ed. R. Campbell and P. Smith, volume III, 1–44. New York: Plenum Pres.
- Lightfoot, David. 1991. *How to set parameters: Arguments from language change*. Cambridge, MA: MIT Press.
- Moran, Patrick Alfred Pierce. 1958. Random processes in genetics. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 54, 60–71. Cambridge Univ Press.
- Nettle, Daniel. 2012. Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367:1829–1836.
- Nowak, Martin A. 2006. *Evolutionary dynamics: exploring the equations of life*. Cambridge, MA: Belknap Press of Harvard University Press.
- Ottósson, Kjartan G. 1990. *Íslensk málhreinsun. sögulegt yfirlit*. Number 6 in Rit Íslenskrar málnefndar. Reykjavík, Iceland: Íslensk málnefnd.
- Platzack, Christer. 1988. The emergence of a word order difference in Scandinavian subordinate clauses. In *Special issue on comparative Germanic syntax*, ed. D. Fekete and Z. Laubitz, 215–238. McGill.
- Rask, Rasmus Kristian. 1941. *Breve fra og til Rasmus Rask*, volume 1. Copenhagen: Munksgaard.
- Roberts, Ian. 1985. Agreement parameters and the development of the English modal auxiliaries. *Natural Language and Linguistic Theory* 3:21–58.
- Rohrbacher, Bernhard. 1994. The Germanic languages and the full paradigm. Doctoral Dissertation, University of Massachusetts.
- Rohrbacher, Bernhard. 1999. *Morphology-driven syntax: A theory of V to I raising and pro-drop*, volume 15 of *Linguistik Aktuell*. Amsterdam: John Benjamins.

- Santorini, Beatrice. 1993. The rate of phrase structure change in the history of Yiddish. *Language Variation and Change* 5:257–283.
- Sundquist, John D. 2002. Morphosyntactic change in the history of the mainland Scandinavian languages. Doctoral Dissertation, Indiana University.
- Sundquist, John D. 2003. The rich agreement hypothesis and Early Modern Danish embedded-clause word order. *Nordic Journal of Linguistics* 26:233–258.
- Thráninsson, Höskuldur. 2003. Syntactic variation, historical development, and minimalism. In *Minimalist syntax*, ed. Randall Hendrick, chapter 4, 152–191. Oxford: Blackwell.
- Thráninsson, Höskuldur. 2007. *The Syntax of Icelandic*. Cambridge: Cambridge University Press.
- Thráninsson, Höskuldur. 2010a. How can V2 vary? URL <http://malvis.hi.is/sites/malvis.hi.is/files/HowCanV2Vary.pdf>, handout of a talk given at the Workshop on Verb Movement, Amsterdam 2010.
- Thráninsson, Höskuldur. 2010b. Predictable and unpredictable sources of variable verb and adverb placement in Scandinavian. *Lingua* 120:1062 – 1088.
- Trudgill, Peter. 2004. *New Dialect Formation: the inevitability of colonial Englishes*. Oxford: Oxford University Press.
- Vikner, Sten. 1997. V⁰-to-I⁰ movement and inflection for person in all tenses. In *The new comparative syntax*, ed. Liliane Haegeman, chapter 8, 189–213. Edinburgh: Longman.
- Waldmann, Christian. 2008. Input och output: Ordföljd i svenska barns huvudsatser och bisatser. Doctoral Dissertation, Lund University.
- Wallenberg, Joel C., Anton K. Ingason, Einar F. Sigurðsson, and Eiríkur Rögnvaldsson. 2011. Icelandic Parsed Historical Corpus (IcePaHC). Version 0.9. Size: 1 million words. URL http://www.linguist.is/icelandic_treebank.
- Watanabe, Akira. 1994. The role of triggers in the Extended Split-Infl Hypothesis: Unlearnable parameter settings. *Studia Linguistica* 48:156–178.
- Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. Empirical foundations for a theory of language change. In *Directions for historical linguistics*, ed. W. Lehmann and Y. Malkiel, 95–195. Austin, TX: University of Texas Press.
- Wiklund, Anna L. 2010. In search of the force of dependent verb second. *Nordic Journal of Linguistics* 33:81–91.
- Wiklund, Anna-Lena, Kristine Bentzen, Gunnar Hrafn Hrafnbjargarson, and Þorbjörg Hróarsdóttir. 2009. On the distribution and illocution of V2 in Scandinavian *that*-clauses. *Lingua* 119:1914–1938.
- Yang, Charles. 2000. Internal and external forces in language change. *Language Variation and Change* 12:231–250.