Verb-second in spoken and written Estonian

VIRVE-ANNELI VIHMAN (iD)

GEORGE WALKDEN (iD)

Author affiliations can be found in the back matter of this article

ABSTRACT

This paper investigates clausal constituent order in Estonian, a language often described in the literature as exhibiting a verb-second "tendency". We present a corpus-based study of ordering in independent affirmative declarative clauses, drawing data from both written and spoken corpora. Our results show that, while written Estonian is robustly a verb-second language along the same lines as the modern Germanic standard languages, spoken Estonian exhibits much more variation. Our findings lead us to suggest that spoken Estonian patterns with the recently-established class of "verb-third" languages, and that syntactic analyses developed to account for these languages can also account for our spoken Estonian data.

CORRESPONDING AUTHOR:
Virve-Anneli Vihman

University of Tartu, Ülikooli 18, 50090 Tartu, Estonia

virve.vihman@ut.ee

1 INTRODUCTION

Verb-second (V2) constituent order, though a signature property of the Germanic languages, is very rare outside the Indo-European family. In his overview article, Holmberg (2015) lists only Estonian and potentially Karitiana; to this we might add Khoekhoegowab (den Besten 2002) and Dinka (van Urk & Richards 2015). An important question in recent literature on V2 is whether the properties that pretheoretically fall under this label are in fact ontologically uniform. Can they be accounted for in terms of a single rule, structure, or parameter, as traditionally assumed in generative syntax, or is V2 more of a conspiracy (Weerman 1989), the confluence of several smaller rules or properties that are in principle independent of one another (see recently e.g. Lohndal, Westergaard & Vangsnes 2020)? In answering questions like this one it is useful to investigate these non-Indo-European languages further, in order to determine what (if anything) is at the core of the V2 phenomenon and what is the result of historical or areal contingencies.

In this paper we focus on the question of V2 in Estonian. Though Estonian has been described as a V2 language, this claim is usually hedged somewhat (see section 2), and little empirical data beyond authors' intuitions has been brought to bear on the question. We use corpora that have become available in the last couple of decades to conduct an exploratory study (section 3) on declarative main clauses in written and spoken Estonian, in order to empirically establish the extent to which V2 characterises Estonian usage, and to investigate the character of violations of V2. Our results (in section 4) show that – while written Estonian is a well-behaved verb-second language for the most part – spoken Estonian displays substantially more variation.

Recent years have seen an increase in research on "verb-third" Germanic varieties, which exhibit systematic but nontrivial deviations from linear verb-second (e.g. Walkden 2015; 2017; te Velde 2017; Alexiadou & Lohndal 2018; Haegeman & Greco 2018) – and on deviations from verb-second more generally (Hsu 2017; 2021; Wolfe 2019a; b). A well-known example is the emerging variety of German generally known as Kiezdeutsch. Extensive research has shown that Kiezdeutsch and other, similar varieties – some of these heavily stigmatized – do not deviate randomly from verb-second. Rather, the "deviations", such as they are, are highly constrained syntactically and information-structurally: we are dealing here with full-fledged natural languages which differ grammatically from their corresponding standard varieties. In section 5 of the paper we consider whether the analytical approaches that have been advanced to deal with Germanic V3 are also appropriate to account for the spoken Estonian facts. Section 6 concludes.

2 ESTONIAN CLAUSAL SYNTAX: PREVIOUS RESEARCH
2.1 CONSTITUENT ORDER IN MAIN CLAUSES

Typologically, Estonian is usually classed as a Subject-Verb-Object (SVO) language (de Sivers 1969: 351–352; Tael 1988; Vilkuna 1998; Dryer 2013a; Lindström 2017: 547). However, all authors to have looked seriously at Estonian constituent order admit that this is not the full story. Vilkuna (1998: 178), for instance, mentions that as well as "V2 tendencies" Estonian exhibits OV in specific constructions, and some discourse-configurationality. These V2 tendencies set Estonian apart from the other Finnic languages and Uralic languages more broadly.[1] An example of a verb-second clause with inversion is given in (1).

(1)   *V2 declarative with inversion* (Erelt et al. 1997: 432)
        Kiiresti **lahku-s-id** ōpilase-d   koolimaja-st.
        quickly leave-PST-3PL student-NOM.PL schoolhouse-ELA
        'The students departed quickly from the schoolhouse.'

In contrast to the Germanic V2 languages, there are several unembedded contexts in which V2 is systematically absent (Remmel 1963; Erelt et al. 1997; and in particular Lindström 2005; 2007). These are *wh*-interrogatives, as in (2), exclamatives, as in (3), and non-subject-initial negative clauses, as in (4).

---

[1]   Written Finnish exhibited some V2 in the early years of the 20th century, but this is marginal today (Vilkuna 1998: 228, note 6). The even more closely related (but critically endangered) language Livonian shows no traces of subject-verb inversion of the V2 type (Remmel 1963: 356).

(2)   Wh-*question* (Lindström 2007: 228)
        Kes   mei-le täna   külla   **tule-b**?
        who.NOM we-ALL today village/visit.ILL come-PRS.3SG
        'Who will visit us today?'

(3)   *Exclamative* (Lindström 2007: 228)
        Küll ta   täna **tule-b**.
        PART s/he.NOM today come-PRS.3SG
        'S/he's sure to come today!'

(4)   *Non-subject-initial negative* (Lindström 2007: 228)
        Täna ta   mei-le külla   **ei tule.**
        today s/he.NOM we-ALL village/visit.ILL not come
        'Today s/he won't come to visit us.'

In this study we leave these clause types aside and focus on affirmative declarative main clauses, the most canonical environment for V2 in Estonian. Moreover, as in the Germanic V2 languages, V2 is typically not found in embedded clauses (Remmel 1963; Erelt et al. 1997; Vilkuna 1998; Ehala 2006; Lindström 2007).[2] These, too, we will set aside.

The only published formal analyses of Estonian clause structure that we are aware of are Ehala (2006) and Holmberg, Sahkai & Tamm (2020). Ehala adopts an X'-theoretic approach in which auxiliaries and finite lexical verbs occupy the head position of IP, while non-finite lexical verbs remain in situ in the VP. Based on clauses like (5), introduced by the complementizer *kui* 'if' which disallows V2, Ehala argues that Estonian has a head-initial IP and a head-final VP, thus instantiating the (cross-linguistically relatively rare) SIOV basic constituent order.
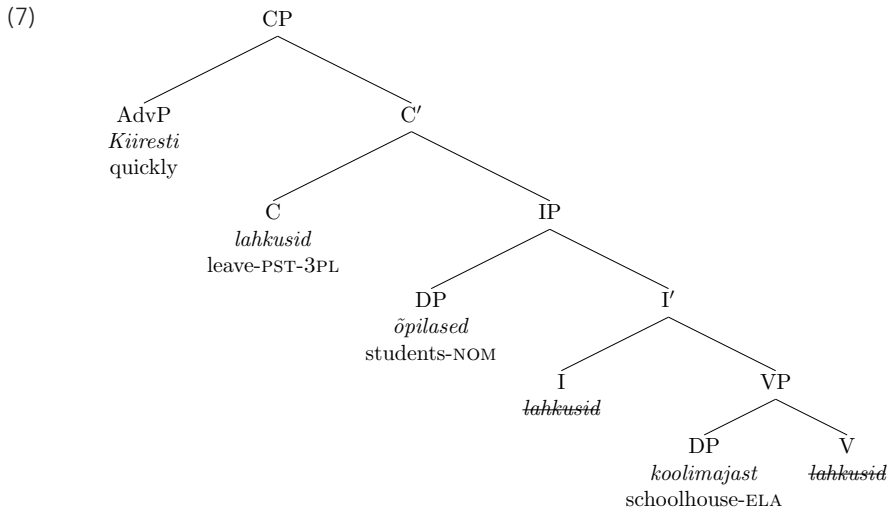
(5)   *SIOV in an 'if'-clause* (Ehala 2006: 68)
        kui lapse-d   on   lõpuks supi   ära söönud
        if child-NOM.PL have finally soup.GEN away eaten
        'if the children have finally eaten up the soup'

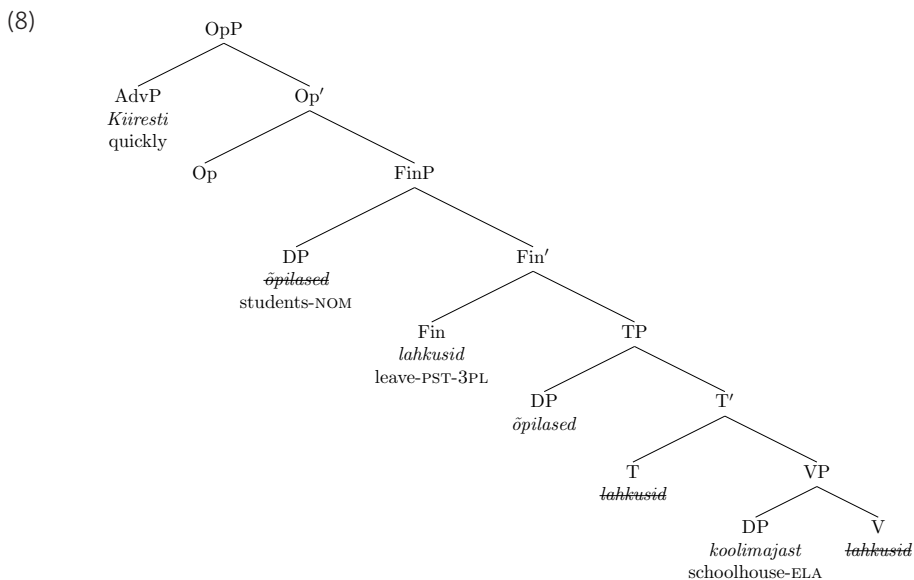This analysis is represented in tree form in (6).

(6)
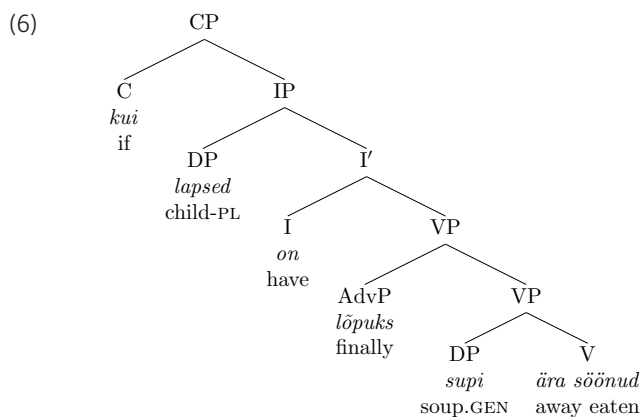
Ehala does not discuss the analysis of V2 in detail, but we can assume that he has in mind the standard analysis of den Besten (1989 [1983]), in which the finite verb uniformly moves to C. A sentence like (1) would thus be represented as in (7), abstracting away from irrelevant structure. In view of the fact that auxiliaries and lexical verbs are predicted to behave differently under Ehala's approach, since auxiliaries cannot occur lower than I, we code for verb type in our corpus investigation (see section 3).

2   For Germanic this is an oversimplification: a lively debate over the years has focused precisely on typological variation in the availability of embedded V2 in Germanic, starting with Rögnvaldsson & Thráinsson (1990) on Icelandic and Diesing (1990) on Yiddish. The nature and limits of this variation is still not fully understood, though in at least some Germanic languages the availability of embedded V2 may be linked to assertion (see Vikner 1995; Holmberg 2015: §3.4; Gärtner 2016; and Walkden & Booth 2020 for discussion). Examining exactly how Estonian fits into this typology is a desideratum for future work. Remmel (1963: 243–244) distinguishes two types of embedded clauses in Estonian with reference to communicative prominence: ordinary subordinate clauses, and "subordinate clauses with the weight (function) of a main clause", with V2 available only in the latter. Lindström (2007) also shows that different types of embedded clauses display verb-finality at different rates. This strongly suggests that similar if not identical factors may be at play in Estonian and in the Germanic languages that permit embedded V2.

(7)

```
                        CP
                   ┌────┴──────────┐
                 AdvP              C′
               Kiiresti      ┌─────┴──────────┐
               quickly       C                IP
                           lahkusid       ┌────┴────────┐
                        leave-PST-3PL     DP            I′
                                       õpilased    ┌────┴──────┐
                                      students-NOM  I          VP
                                                 lahkusid   ┌───┴─────┐
                                                            DP        V
                                                        koolimajast  lahkusid
                                                       schoolhouse-ELA
```

Holmberg, Sahkai & Tamm (2020) propose that Estonian has a left periphery consisting of OpP and FinP above TP. Spec,OpP fulfils the same function as CP in standard approaches, hosting wh-phrases etc.; Spec,FinP is where the EPP is satisfied in Estonian, either by a subject (normally) or by some other constituent (if the subject is absent or remains low). Holmberg, Sahkai & Tamm (2020) capture V2 by assuming that i) the finite verb occupies Fin, ii) the subject moves to Spec,FinP as normal, but iii) a lower copy of the subject is spelled out in the normal case, for reasons that are ultimately prosodic. A sentence like (1) would thus be represented as in (8).

(8)

```
                     OpP
                ┌─────┴──────┐
              AdvP           Op′
            Kiiresti    ┌─────┴─────────┐
            quickly     Op             FinP
                              ┌─────────┴──────────┐
                              DP                   Fin′
                           õpilased          ┌──────┴────────┐
                         students-NOM        Fin             TP
                                           lahkusid     ┌─────┴──────┐
                                        leave-PST-3PL   DP           T′
                                                     õpilased   ┌────┴──────┐
                                                                T          VP
                                                             lahkusid  ┌────┴─────┐
                                                                       DP         V
                                                                   koolimajast  lahkusid
                                                                  schoolhouse-ELA
```

2.2 THE SYNTAX OF SUBJECTS IN ESTONIAN

Two further properties of Estonian subject syntax are worth mentioning at this point, as they bear directly on our method and analysis.

First, Estonian – like most of the world's languages, but unlike the prototypical Germanic V2 languages – exhibits null referential subjects (Dryer 2013b, following de Sivers 1969: 47–48). The exact details of the Estonian null subject system remain to be established: Kivik (2010: 66) suggests that Estonian is a "mixed null-subject language" in the sense of Vainikka & Levy (1999), like Finnish and Hebrew (cf. also Holmberg 2017: 366). Grammars state that null subjects are possible in the first and second person, but third person null subjects are also possible in certain contexts and registers (Lindström 2001; Keevallik 2003; Duvallon & Chalvin 2004). Importantly, null subjects are not restricted to (surface) V1 clauses, as they are in most present-day Germanic languages (Ross 1982; Trutkowski 2016).

For our purposes, null subjects are important because subject-verb inversion, a typical diagnostic for V2 in Germanic languages, is of course not detectable when subjects are null. This reduces the unambiguous evidence for V2 syntax available in corpus studies.

Second, and unlike most Germanic languages, Estonian exhibits a double system of personal subject pronouns, as shown in *Table 1*; these are usually referred to as "short" and "long" forms (Pajusalu 2017: 569–570).

PERSON/NUMBER	SG		PL	
	SHORT	LONG	SHORT	LONG
1	ma	mina	me	meie
2	sa	sina	te	teie
3	ta	tema	nad	nemad

Table 1 Forms of pronominal subjects in Estonian.

The consensus about these forms is that the short form is unmarked, whereas if the subject pronoun is focused, contrastive, stressed or accented the long form is used. The long form can also be used without any particular accent or information-structural prominence, i.e. solely anaphorically, but it is less frequent in this role (Pajusalu 2017: 569, 577). The short form, on the other hand, can only be unstressed.[3] The general preference for short forms is more pronounced in written than spoken Estonian (Pajusalu 2005). The distinction between short and long forms will play a role in our discussion of deviations from V2 in section 5.

2.3 SPEECH, WRITING, VARIATION AND CHANGE

In a well-behaved V2 language, one and only one constituent may precede the finite verb in V2 clauses. Precisely this has been called into question for Estonian, however. Vilkuna (1998: 180), for instance, while observing that Estonian has a V2 "character", adds that "informants are not willing to exclude sentences that violate the V2 constraint, and exceptions are found, especially in spoken language … and with weak [i.e. short] pronominal subjects" (see also Lindström 2005). On the latter point, a clear contrast is expected between (9), an entirely grammatical V2 violation with a preverbal short pronoun as subject, and (10), with a long pronoun or full NP subject. Full NPs pattern with long pronouns in being dispreferred in this position, yet speakers often do not exclude (10), as noted by Vilkuna.

(9) Täna ta   tule-b   mei-le külla.
     today s/he.NOM-SHORT come-PRS.3SG we-ALL village/visit.ILL
     'Today s/he is going to visit us.'

(10) ?Täna tema /   vanaema   tule-b   mei-le külla.
     today s/he.NOM-LONG grandma.NOM come-PRS.3SG we-ALL village/visit.ILL
     'Today s/he / grandma is going to visit us.'

While XVS (i.e. non-subject-initial V2) occurs at a rate of 24% in Tael's (1988) corpus of written Estonian, it is rarer in spoken language. Lindström (2000) found XVS in 5–7% of all main clauses in a corpus of spoken discourse in standard Estonian and two dialects. However, subject omission is frequent in spoken varieties of Estonian; if we also include subjectless XV(X) clauses, the percentage of all non-subject-initial clauses with V2 constituent order in Lindström's data shows cross-dialectal differences, with non-subject-initial V2 in 21–30% of clauses. Her analysis of written data gives still higher proportions of non-subject-initial V2 (40%); a corpus of anecdotes exhibits V2 in only 13% of clauses, preferring a verb-initial order typical of narratives.

The syntactic difference between spoken and written Estonian, though it has not been thoroughly investigated empirically, is often alluded to in the literature (e.g. Lindström 2017: 555). A diachronic explanation for the difference can be adduced from the particular sociohistorical circumstances of speakers of Estonian in the modern era. Estonian has been in close contact with German, Danish and Swedish – all V2 languages – from the Middle Ages onwards, and German in particular enjoyed lengthy contact and high overt prestige. Vilkuna (1998: 180), for example, suggests that this is the origin of V2 in Estonian: the prestigious V2

---

3   It is tempting to equate short and long pronouns with the weak and strong pronouns of e.g. Cardinaletti & Starke (1999). Indeed, they seem to support this distinction, as the short pronouns cannot be coordinated, modified or used in marked positions, whereas the long pronouns can. Since this is not the focus of our work, however, we continue to use the pretheoretical terms here.

Germanic languages spoken and written in the area served as a partial model, and source of translations, during the emergence of an Estonian standard language during the modern era (see also Ziegelmann & Winkler 2006: 65). If so, it is possible that V2 simply never became a more systematic part of colloquial Estonian, or at least not as rigidly as in the standard written language. Lindström (2017: 555) indicates that this may be due to the spoken language being more sensitive to information structure. We do not address the diachronic question directly in this paper, but note here that the idea of explicit prescriptive pressure having fundamentally shaped the Estonian written standard is well established in the literature: Ehala (1998), for instance, argues that reforms led by Johannes Aavik in the early 20th century (in this case aiming to *reverse* German influence) were instrumental in changing this variety's basic constituent order from SOVI to SIOV (see also Raag 1998).

2.4 VERB-SECOND DEVIATIONS IN KIEZDEUTSCH AS COMPARATOR

The starting point for our investigation is the existence of deviations from verb-second. In a "well-behaved" V2 language like German and the Scandinavian languages, as captured by the V-in-C analysis of den Besten (1989 [1983]), these would not be expected to occur at all.[4] In this paper, therefore, we aim to establish what it means for there to be a V2 "tendency" in (spoken or written) Estonian, and in particular what the exceptions to V2 look like. We take our lead from recent research on emerging Germanic varieties such as Kiezdeutsch (Freywald et al. 2015; te Velde 2017; Walkden 2017; Alexiadou & Lohndal 2018), which has focused on precisely these exceptions and how best to account for them. An example of a deviation from verb-second in Kiezdeutsch is given in (11).

(11)   *Kiezdeutsch* (KiDKo, transcript Mu9WT; Freywald et al. 2015: 83)
         GEStern   isch **war** KUdamm.
         yesterday I   was Ku'damm
         'Yesterday I was at Kurfürstendamm.'

As already stated, these deviations are not random. The prototypical example of so-called "verb-third" clauses (te Velde 2017; Walkden 2017) involves some non-subject constituent in initial position (*GEStern* "yesterday" in (11)), followed by an element which is nearly always a pronominal subject (*isch* "I" in (11)). Freywald et al. (2015) establish that this immediately preverbal element is virtually always unaccented, and that it has the information-structural profile of a familiar topic, i.e. it is given information. By and large, there is a consensus about these facts in the literature at this point, which more or less also holds for comparable emerging varieties of Danish, Norwegian and Swedish (Freywald et al. 2015) and Dutch (Meelen, Mourigh & Cheng 2020).[5] Formal analyses differ in details, but agree on the necessity for two available specifier positions before the finite verb.[6]

In light of these facts it is clear why the short vs. long pronoun asymmetry in Estonian is of interest: short pronouns are unaccented, whereas long pronouns need not be. Therefore, if the deviations from V2 observed in Estonian are of the same nature as those observed in Kiezdeutsch, we expect to see a predominance of short rather than long pronouns in the immediately preverbal position. More generally, we expect this preverbal position to be occupied by pronominal subjects rather than other constituents, on the whole.

In the quantitative analysis we carried out, reported on in the subsequent sections, we were interested in determining: (i) how prevalent V2, V3, and other orders are in the corpus; (ii) whether we find differences between written and spoken language; and (iii) what factors best predict the use of V2 and other orders. We expected to find a dominance of verb second order, and to find a difference between written and spoken data. Based on the perceived contrast between (9) and (10),

---

4   With a few minor and well-understood apparent exceptions, such as the Swedish adverb *kanske* 'maybe' (Platzack 1986).

5   Disagreements focus mainly on whether elements are allowed in preverbal position that are a) not subjects and b) not pronouns; see Walkden (2017) for discussion. Whether or not they are grammatical in absolute terms, though, such cases are rare enough that they are likely to have little relevance to a quantitative study.

6   Another example of a Germanic language which has been argued to display V3 syntax is Old English, though the patterns here are more diverse. The literature on the syntax of Old English V2 and V3 is substantial, and not all relevant to this paper: see van Kemenade (1987), van Kemenade, Milićev & Baayen (2008), Haeberli (2002), Speyer (2010), and Walkden (2014: 67–89) for detailed discussion.

3 METHOD

We compiled a sample including an equivalent number of main clauses from written and spoken corpora. The written data were extracted randomly from the Fiction subcorpus of the University of Tartu's Balanced Corpus of Written Estonian (5 million words total in Fiction), using an online search engine (cl.ut.ee/korpused). The spoken data were randomly drawn from the University of Tartu's Corpus of Spoken Estonian, maintained by the research group of Spoken Estonian (not publicly available). Our spoken language selection derives from a subset of everyday (face-to-face and telephone) conversations. The written corpus includes 751 clauses, and the spoken corpus includes 758 clauses. Only clauses with a finite verb and at least one overt argument were included. Each clause in the initial sample which did not match these criteria was replaced by a new, randomly drawn clause. We look only at independent (main) clauses, leaving aside subordinate clauses.

Determining constituent order involves discriminating grammatical relations which are sometimes ambiguous; semantic or pragmatic judgments may be required to distinguish between two potential analyses. Automatic parsing is unavailable and would be unreliable for this task. Instead, we coded the clauses in the two language samples manually. Codes were checked by two linguistically trained coders and disagreements were resolved through discussion. In addition to clause type (declarative, exclamative, imperative, interrogative) and polarity (affirmative, negative), which were used to exclude clauses not matching our inclusion criteria (affirmative, declarative main clauses), each main clause was coded for:

1. Corpus: WRI – written, SPO – spoken

2. Constituent order:

   a. Verb position in clause: V1, V2, V3, V# (clause-final), and Vx (for any other linear position). Clauses in which the verb was in third *and* final position were coded as V3.

   b. Subject position relative to the verb (SV, VS)

3. Subject form (lexical noun/NP, pronoun)

4. Subject pronoun type (short, long; only applies to personal pronouns)

5. Finite verb type (lexical verb, auxiliary, copula, modal verb)

First, in section 4.1, we examine the coded variables in the dataset, in order to give a general picture of their distribution. Second, in order to examine the data statistically and reveal which predictors have the most influence on verb position in Estonian, we use two non-parametric classification methods, recursive partitioning trees (Hothorn, Hornik & Zeileis 2006) and random forests (Breiman 2001; Strobl et al. 2008), reported in section 4.2. The first of these, in the conditional inference framework, performs binary splits of the data locally, each time making the split based on which variables best classify the data. The model splits the data recursively and stops when no further significant splits can be made based on the predictors; hence, it does not include any non-significant predictors in the final model. One advantage of this method, in contrast to others used for similar purposes like regression models, is that the output is presented in easily interpretable visualisations.

The random forests method (Breiman 2001) complements binary recursive partitioning. The random forests model derives from a large number of conditional inference trees, each one constructed based on a random permutation of the predictor variables. Prediction accuracy is measured before and after each permutation, thus assessing the extent to which each predictor improves the model (Strobl et al. 2008). Based on these trees and the prediction accuracy measures, the model chooses the best variables for classifying the data and assigns relative "importance" to each variable. These methods have been successfully used in linguistic studies of corpus data as an alternative to regression models (e.g. Tagliamonte & Baayen 2012; Baayen et al. 2013; Janda 2013, Lindström & Vihman 2017; Walkden & Rusten 2017). The analyses were performed using R (R Development CoreTeam 2013). We use the partykit package for both conditional inference tree (*ctree*) and random forests (*cforest*) analyses.

Following the results of the quantitative analysis in the next section, we look more closely at the set of examples which do not follow V2 constituent order and ask whether these are systematic, how they can be characterised, and whether V3 examples exhibit similarities to Germanic V3 found in languages such as Kiezdeutsch.

4 RESULTS
4.1 OVERVIEW OF THE FINDINGS

Because of systematic differences in constituent order across clause types, we include in the analysis below only affirmative, declarative main clauses, comprising 569 clauses in the written sample and 498 in the spoken sample. Our dataset revealed a preponderance of V2 constituent order across both datasets (82.8% of affirmative declarative clauses follow V2 order).

We also found significant differences between the written and spoken data. As can be seen in *Figure 1*, V2 is prevalent in both corpora, but the overwhelming preference for V2 is tempered in the spoken corpus (76% V2), with V3 constituent order making up 14% of affirmative declarative clauses, followed by verb-first (5%) and verb-final (4%) order. In the written corpus, not only is V2 constituent order found in the vast majority of affirmative declarative main clauses (89%), but exceptions to V2 also differ from those in the spoken language data. Most of the exceptions are verb-initial clauses (6%), with V3 accounting for only 4% of all the affirmative declarative clauses in our written data sample.[7]
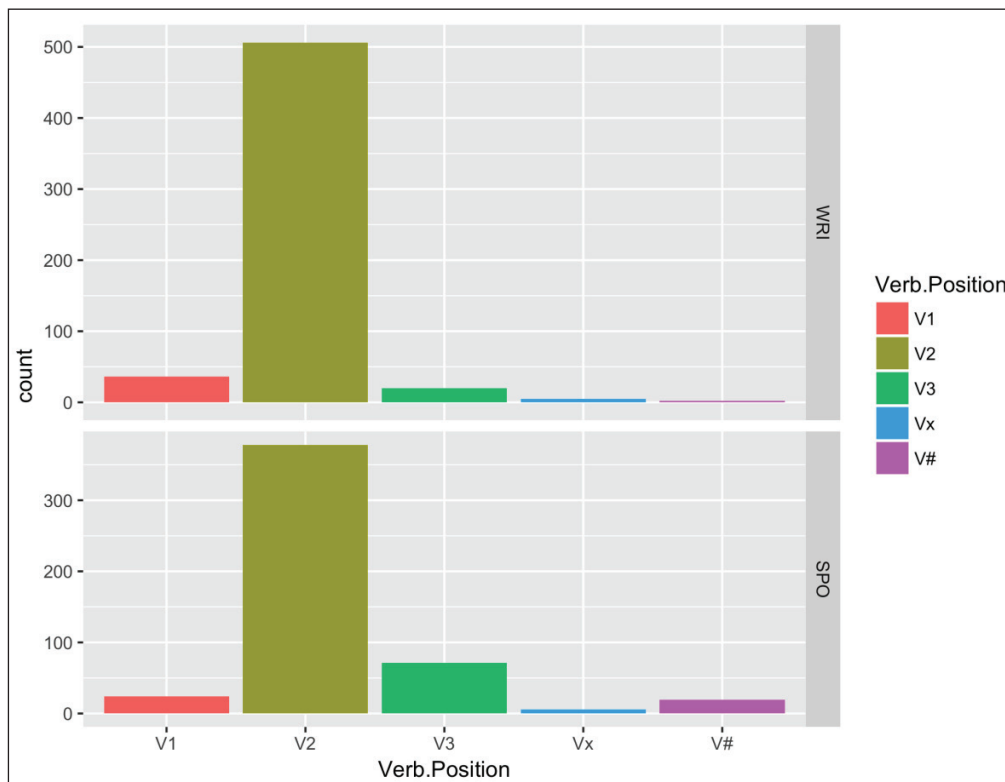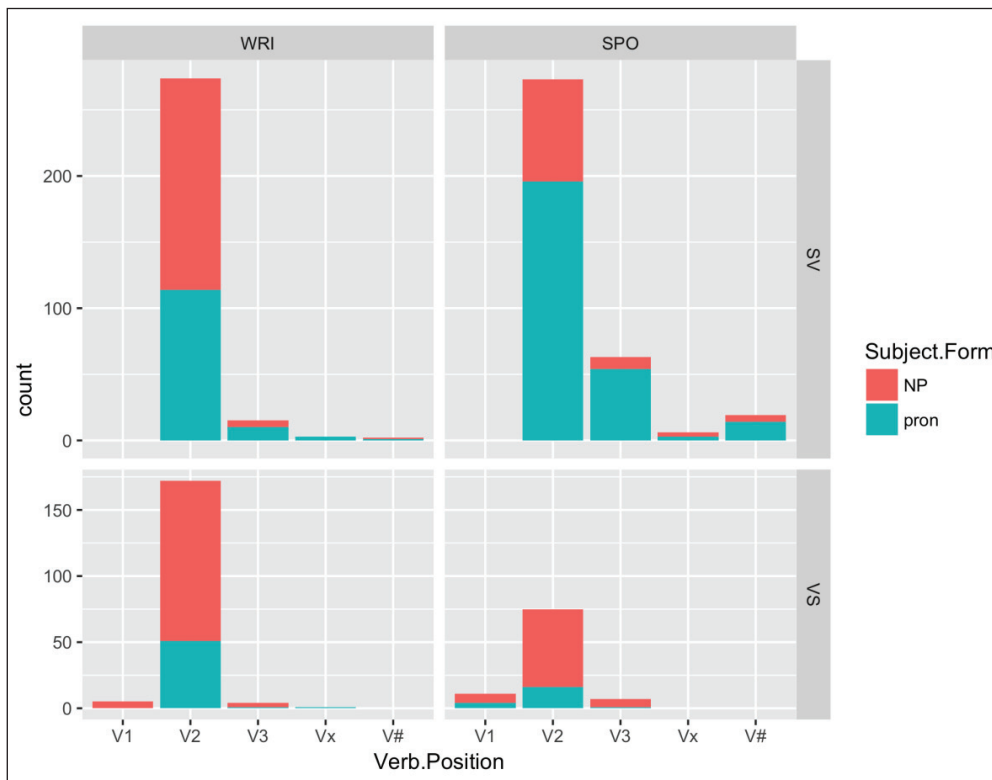

Figure 1 Verb position in affirmative declarative main clauses, by corpus (WRI = written; SPO = spoken). Note (here and throughout) that V# indicates verb-final clauses; any linear position besides first, second, third or final is coded as Vx.

Figure 2 examines this distribution more closely, plotting word order by register (written and spoken, left and right panels) and relative order of subject and verb (SV, top, and VS, bottom panels). SV order marks all clauses with preverbal subjects, regardless of other, intervening constituents (hence, (X)S(X)V(X)), and VS includes all clauses with postverbal subjects, similarly disregarding other constituents. In addition to clausal verb position, *Figure 2* also shows the subject form (lexical nouns vs pronouns).

We see here that written Estonian (left panels) can be characterised as a fairly well-behaved V2 language. Noting that spoken language usage also follows the general V2 "tendency", we may ask how to characterise the exceptions. Looking first at verb position in spoken data, we see that the majority of V3 structures are found with preverbal subjects (including both XSV and

---

7   This dataset does not allow us to investigate individual differences, but it is possible, as suggested by a reviewer, that the spoken data contains variation across speakers.

Figure 2 Verb position and subject form (lexical NP or pronoun) in affirmative declarative main clauses with overt subjects, by corpus and Subject-Verb relative order.

SXV). Conversely, although inversion is used less in spoken language, by and large, clauses with inverted subjects co-occur overwhelmingly with V2 across both corpora (V2 accounts for 94.5% of clauses with postverbal subjects, VS, in the written corpus and 80.6% in the spoken corpus; in both corpora, the exceptions to V2 with postverbal subjects are mostly V1, and slightly more rarely V3).

Next, we examined whether subjects preceding or following the verb, in each corpus, were more likely to be pronouns or NPs (NP here representing lexical nouns and noun phrases, including quantifier and number phrases). Overall, the spoken dataset includes a much greater proportion of pronominal subjects (63%) than the written data (38% pronouns). Yet the spoken data also shows different word order patterns with lexical or pronominal subjects. In both corpora, post-verbal subjects tend to be full NPs, more so than preverbal subjects, but the contrast with preverbal subjects is especially striking in the spoken data. V3 and verb-final clauses are shown to occur with notably greater frequency in the spoken corpus with preverbal, pronominal subjects. We will look more closely at examples, and at what other constituents appear preverbally with V3, in Section 4.3.
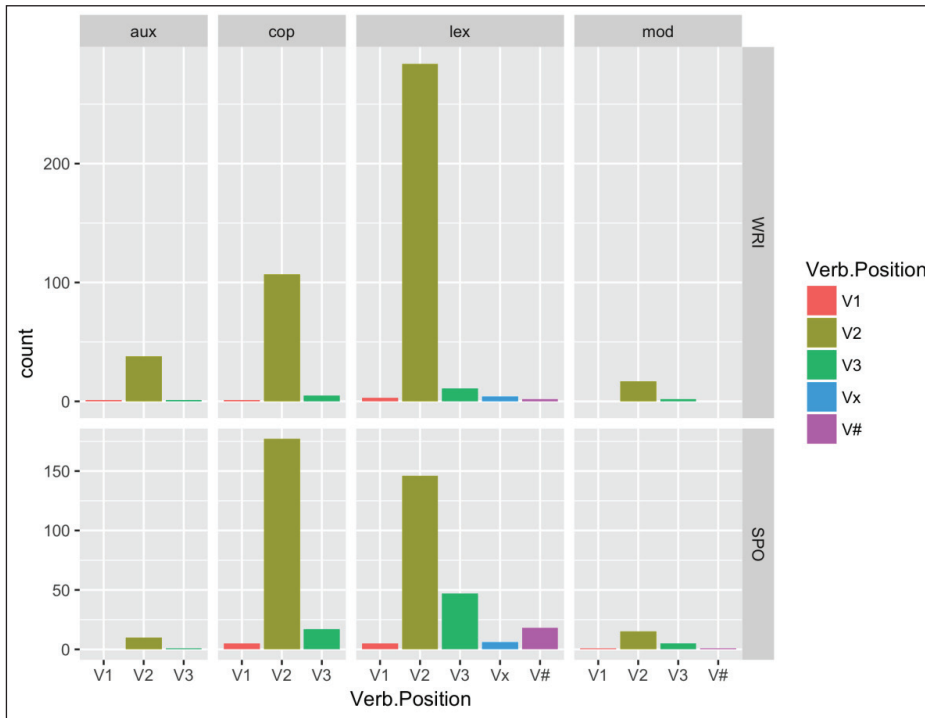
We also asked whether different types of verbs are used in differing positions (*Figure 3*).

Figure 3 shows word order by register (top and bottom panels) and four types of verbs: auxiliary, copula, lexical and modal verbs. Visual inspection reveals that the greatest variability in word order occurs in the spoken corpus with lexical verbs, where V2 accounts for only 66% of a total of 222 clauses with lexical verbs, and V3 accounts for 21%. V3 also occurs with 8.3% of copula clauses (n = 200) in the spoken data.
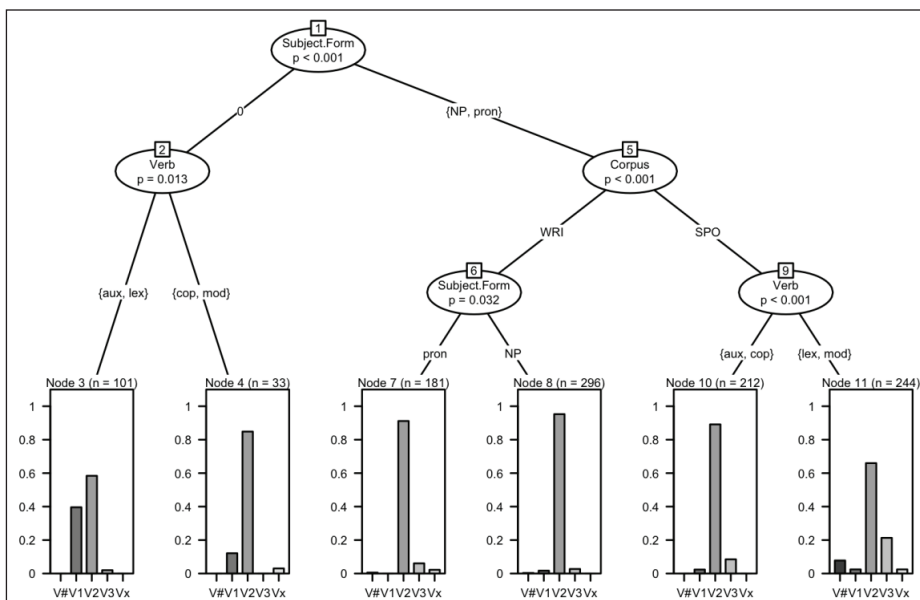
4.2 STATISTICAL MODELS

Using the recursive partitioning tree model in the conditional inference framework, we analysed factors affecting verb position in all the affirmative declarative clauses in our corpus, including corpus, subject form, and verb type as predictors. The model was not improved by either subject pronoun form, differentiating long and short pronouns, or subject position (SV, VS); in other words, the difference between long and short pronouns, and the difference between SV and VS, do not give the model any additional predictive power on top of the other predictors included. These were left out of the final model. The following formula was used in the final model: ctree(Verb.Position ~ Corpus + Subject.Form + Verb, data = ADV2, controls = ctree_control(minbucket = 25)). The model output is shown in *Figure 4*.

As the model in *Figure 4* shows, subject form was selected as the most significant factor determining verb position (Node 1, at the top of the tree), yet the first split is not made between

lexical and pronominal subjects, but rather between subjectless clauses, which exhibit more verb-first order, and those with overt subjects. For subjectless clauses, the next split (Node 2) is made by verb form: here, auxiliary and lexical verbs are grouped together, with increased verb-initial order, contrasting with copulas and modals.

However, the right branch of Node 1 includes many more clauses than the left branch. Corpus emerges as a highly significant predictor (Node 5), with a clear split between the written and spoken corpus, the former exhibiting almost exclusively V2 in affirmative declarative clauses. Nevertheless, subject form appears again as a significant predictor within the written corpus data, for clauses with overt subjects, with a significant difference between pronouns and NPs. The split made at Node 6 shows that in the written corpus, pronominal subjects slightly increase the likelihood of exceptions to V2 (Node 7). Considering the greater number of V2 deviations in the spoken corpus, it is surprising that this split is not found under the right branch of Node 5 (SPO). This may be due to the much greater proportion of pronouns overall in the spoken corpus (recall ***Figure 2***).

Finally, verb type significantly affected verb position in the spoken data (Node 9), with auxiliary and copula clauses (Node 10) showing a stronger preference for V2 than clauses with lexical and modal verbs (Node 11).

In order to assess the model's accuracy, we first compared predictions to actual observations. The model correctly predicts 82.85% of cases, but all the correct predictions were for V2, and this percent exactly matches the proportion of V2 clauses. The model does not predict any other verb position because of the preponderance of V2. Therefore, we also examined the Area Under the ROC curve (AUC, or C-index), a more flexible measure which assesses the model's ability to distinguish between classes based on their predicted probabilities instead of the predicted classes themselves. The multiclass.roc() function in the pROC package returns an AUC of 0.739, meaning the model's discriminative ability is satisfactory (0.8–0.9 is good and > 0.9 is excellent). This measure is slightly less affected by the dominance of V2 in our data because it takes into account distinctions between the other verb positions.

To confirm these findings and gain a more robust picture of the effect of our predictors on verb position in the corpus, we also performed an analysis using random forests. The output of the random forests model is shown in **Figure 5**. We used the same predictors as in the previous model, and applied the following formula: cforest(Verb.position ~ Corpus + Sform + Verb, data = ADV2, control = ctree_control(minbucket = 25)).
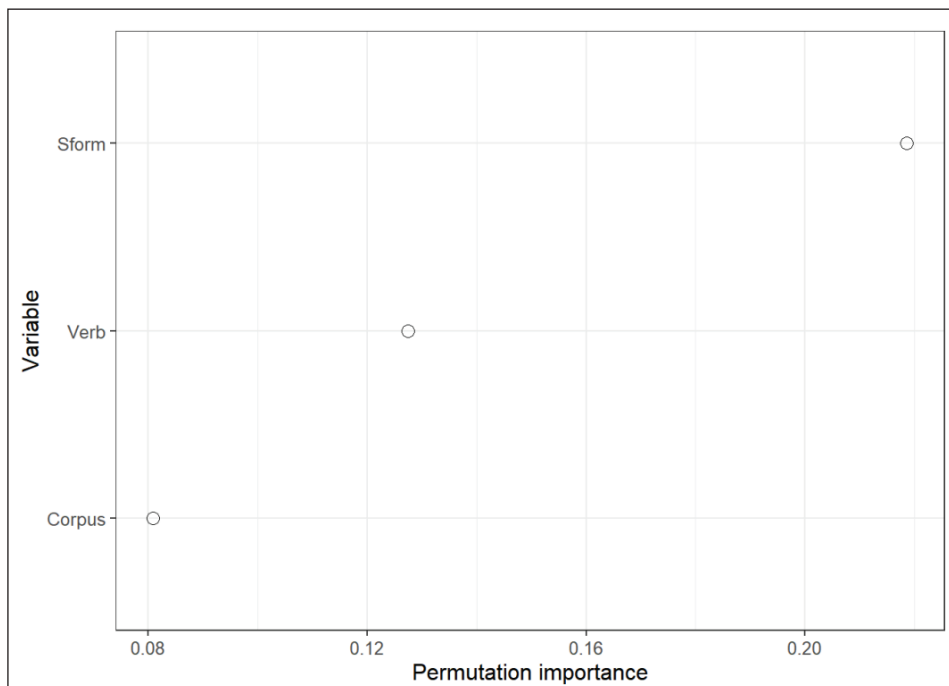
Figure 5 Random forests output. Variable importance of predictors affecting verb position.

The random forests model confirms the high importance of subject form in predicting verb position in the clause, followed by verb type and register, or corpus. As seen above, clauses with subject pronouns are less likely to exhibit V2, and clauses in the written corpus are more likely to do so. Subject form splits the data first between clauses with null and overt subjects, and then between those with pronominal and lexical subjects, with pronouns allowing more V3 exceptions to the predominant V2 order. Lexical and modal verbs are more likely to appear in V3 than auxiliaries and copulas. The discriminative ability of the forest is better than that of the single tree, with an AUC/C-index of 0.759. Although the forest analysis takes into account the probability of the other classes, it is still V2 which contributes most to this high value.

4.3 EXCEPTIONS TO V2 IN THE DATA

This section zooms in on the exceptions to V2 discussed above. In this section we briefly discuss V1 clauses such as (12), verb-final clauses such as (13), and the very rare "other" category,[8] before focusing on the verb-third examples.

---

8   Such examples are simply those in which the verb is in later than third position (i.e. preceded by more than two constituents) but not in absolute clause-final position. Though the category is thus a pretheoretical one, some of these cases correspond to the "verb-medial" category in Sahkai & Tamm (2019) and Holmberg, Sahkai & Tamm (2020).

(12)   (V1)

   a.   *(written corpus: clausal argument predicate)*
          Tundus   nagu   oleks   sõjaväe   hävitus-lennuk   köögi-põranda-le
          felt   like   be-CND   army.GEN destroyer-plane.NOM kitchen-floor-ALL
          häda-maandumise   sooritanud.
          emergency-landing.GEN   performed
          '[It] seemed like a fighter plane had made an emergency landing on the kitchen
          floor.'

   b.   *(spoken corpus: omitted subject)*
          sa-d=   'aigla-st   oma   tõved   kallale
          get-PRS.2SG   hospital-ELA REF.GEN diseases attack-ALL
          '[You'll] get your diseases attacking you from the hospital.'

   c.   *(written corpus: existential/presentational construction)*
          Avane-s   jalg-värav.
          open-PST.3SG   foot-gate.NOM
          'The pedestrian gate opened.'

(13)   *(spoken corpus: V-final)*
          'ilmselt   sa   'selle-pärast   '**os-si-d-ki**.
          probably   you.NOM   because-of-this   buy-PST-2SG-FOC
          'That's probably why you even bought [it]'

Recall from *Figure 1* that V1 is the only alternative to V2 patterns occurring with any notable
frequency in the written corpus. As shown by the conditional inference tree in *Figure 4*, V1 clauses
are only frequently found without overt subjects (both subjectless constructions like (12a) and
omitted topics like (12b)), and most of the exceptions to this involve zero objects. Example (12a)
illustrates that Estonian does not require expletive subjects with clausal-argument predicates,
which is typical for null-subject languages (Rizzi 1982; Gilligan 1987).[9] As regards (12b), V1 in "topic
drop" configurations is commonly found in all known V2 languages (see e.g. Mörnsjö 2002 on
Swedish, Nygård 2013 on Norwegian, and Trutkowski 2016 on German). Whatever analysis works
for these languages, then, can presumably be straightforwardly transferred to our Estonian data.
As for (12c), this is a VS example of an existential/presentational construction with no preverbal
element; most strict V2 languages would have a prefield expletive in clauses like this one, and
Estonian would often have an initial adverbial constituent, but again the absence of an overt
expletive is no surprise given Estonian's ability to omit subjects generally. Existential/presentational
clauses tend to occur with unaccusative verbs and are found in both speech and writing.

Verb-final and "other" clauses, meanwhile, are rare in absolute terms, though more frequent in
the spoken corpus than V1. Those that do occur (such as (13)) have a strongly discourse-non-
neutral flavour. It is possible that these can be assimilated to the classes of exceptions to V2 in
matrix clauses discussed by Lindström (2007): exclamatives and negated clauses. Our written
data sample has only one example (which has a marked, poetic or nursery-rhyme feel), and
more than a third of the 19 examples in the spoken dataset are marked with a focus clitic on the
verb, as in (13), or the clause-initial emphatic particle *küll*. These exceptions require further study,
and we leave them aside in the rest of this paper; Remmel (1963), Tael (1988), Lindström (2017)
and Sahkai & Tamm (2019) all suggest that the verb is accented or in focus in such examples.[10]

In contrast to the written corpus, the spoken Estonian dataset exhibits a number of verb-third
clauses; these are second to V2 in frequency. As discussed in section 2.4, the deviations from V2
found in Kiezdeutsch and emerging varieties of Scandinavian languages are prototypically V3,
with an adverbial element followed by a pronominal subject in preverbal position; the preverbal
subject in second position is almost always unaccented and given (Freywald et al. 2015). Short
pronominal subjects tend to occur in V2 deviations in Estonian as well (Vilkuna 1998: 180; see
also Lindström 2005).

---

9   Though Holmberg & Nikanne (2002) show that Finnish seems to be a counterexample to the generalization
that null-subject languages do not have expletive subjects: with clausal extraposition, expletive *se* is optional in this
language (cf. their example (9)). In their analysis, the Finnish expletive is an expletive topic rather than a subject.

10   In addition to exclamatives, *wh*-questions and non-subject-initial negative clauses, Lindström (2017: 558)
lists as conditions under which finite verbs may appear at the end of a main clause: focussed predicates and
clauses with an initial epistemic adverb expressing the likelihood of an event taking place.

We therefore take a closer look here at the V3 clauses relatively prevalent in the spoken corpus, at 14% of affirmative declarative clauses (71 out of 498). We expected them to behave similarly to the V3 clauses in Kiezdeutsch, with time adverbials and subject pronouns as the prototypical preverbal constituents. We also expected the subject pronoun to be preverbal (in second position in the clause) and typically in the short form, indicating information-structural familiarity, or givenness, and unaccented prosody. An example of this sort of V3 clause is given in (14), which appears to be structurally identical to the Kiezdeutsch example in (11).

(14) [eile]   [ma]   rääki-s-in   lihtsalt
      yesterday 1SG.NOM.SHORT talk-PST-1SG simply
      'Yesterday I was only talking.'

Of the V3 clauses with preverbal subjects, 84% (53/63) are pronominal. Looking only at those pronouns which have a short/long contrast occurring in V3 clauses, they are overwhelmingly in short form (43/46). In the written data, V3 occurs only with short subject pronouns, and overall, long forms are used much less frequently. Again including only clauses with the pronouns allowing a short/long contrast, long subject pronouns occur in 6.5% of V3 and 10.9% of V2 in the spoken corpus; in the entire written data sample, long subject pronouns are found only in 6.7% of V2 clauses. Overall, however, the long forms occur too rarely to allow statistical comparisons.

Six of eight V3 clauses in the spoken data without preverbal subjects have full, lexical postverbal subjects. One has a pronominal, postverbal subject and one is coded as a subjectless impersonal; this has a topicalised, short object pronoun in second position which is syncretic with the third person nominative subject pronoun (15):[11]

(15) ['minu=tea-da] [ta] 'peide-ti   sealt   ülevalt   toast   lihtsalt 'ära.
      1SG.GEN know-INF s/he hide-PST.IMPL there-ABL above-ABL room-ELA just   away
      'As far as I know they just hid her/him away from that room up there.'

As for the sentence-initial constituent, we find some variation. Adverbs are the most frequent clause-initial elements in V3 examples. These include temporal adverbs, as in (14) above, as well as locative adverbs (16a), and discourse-pragmatic adverbs, such as those in (16b-c). No manner adverbs are attested among the first two constituents in V3 clauses in the sample. Note that (16a) is a V3 example with a full NP subject in preverbal position.[12]

(16) a. *(spoken corpus)*
          [seal] ['tava]   on niimodi
          there custom.NOM is   thus
          'That's the custom there.'

      b. *(spoken corpus)*
          [muidu=] [siuke   'seismise ja 'vahtimise 'töö]   on vist   kõige
          otherwise this-kind standing and guarding work.NOM is probably most
          'hullem
          worse
          'Anyway, this kind of standing and watching job has got to be the worst.'

      c. *(written corpus)*
          [Küllap] [niisugune seadus] on tehtud sellepärast, et…
          surely   such.NOM   law.NOM is made   because   that
          'Surely such a law has been written because…'

11   See Holmberg & Nikanne (2002) and Manninen & Nelson (2004) for analyses of topicalised arguments in clause-initial position in Finnish impersonals. The analysis of Kiezdeutsch in Walkden (2017) predicts pronominal objects to be possible in the preverbal position, similarly to this example.

12   As a reviewer notes, some examples, such as (15) and (16b), are amenable to another analysis: they involve attitudinal adverbials ("as far as I know", (15)) or discourse-connective adverbs ("anyway", (16b)) which take propositional scope. Syntactically, these elements could be outside the V2 clause altogether; Swan (1994) and Lenker (2000) show that Old English *soþlice, witodlice* "truly" and similar adverbials seem to behave this way. In German, too, elements like *freilich* "admittedly" and many others may occur initially preceding a V2 clause (Pasch et al. 2003: 504–509). They note that such elements may, however, also occur with inversion; the same is true for Old English, and Estonian (Lindström 2017: 553). Moreover, only some of our examples display this ambiguity; (14) and (16a), for instance, do not.

Unlike Kiezdeutsch, in which object-initial clauses are unattested, the spoken Estonian data does include object-first V3 clauses (as does Old English), albeit very infrequently (four clauses in this dataset with OSV), as demonstrated in (17).

(17) [niukst   jutu-d]   [ma]   loe-n   'ikka läbi
      this-kind.PL story-NOM.PL 1SG.NOM.SHORT read-PRS.1SG still through
      'I always read those kinds of stories.'

Subject-initial clauses are found in the data (18/71), usually with short subject pronouns co-occurring with adverbs in second linear position, as in (18),[13] but NPs also occur, as shown in (19). Example (19) also shows that, in addition to adverbs, arguments such as experiencers in oblique (locative) cases appear in V3 clauses in preverbal position; in information-structural (and prosodic) terms, the short locative pronoun in (19) is equivalent to the short subject pronouns; these short, subject-like oblique pronouns often participate in inversion, as noted by Lindström (2017: 552).

(18) [ma]   ['üks-päev] 'mõtle-s-in sinu   'peale.
      1SG.NOM one-day   think-PST-1SG 2SG.GEN onto
      'I was thinking about you one day.'

(19) [see 'Kairi] [mu-lle]   'elista-s enne   millalgi
      that Kairi   1SG.SHORT-ALL call-PST.3SG before sometime
      'That Kairi called me earlier sometime.'

In summary, the vast majority of V3 exceptions to V2 have a pronominal subject in preverbal position. Most of these are short pronouns, and tend to co-occur with temporal, locative or discourse-pragmatic adverbs. In addition to adverbs, objects and oblique experiencer arguments are attested in the spoken data, where most of the V3 deviations are found. While the bulk of the V3 examples look similar to those found in Kiezdeutsch, the subject-initial clauses are different; we defer discussion of these until 5.1.

5 DISCUSSION

In our quantitative analysis, we confirmed the prevalence of V2 order in Estonian corpus data. We also found differences between written and spoken language, as expected, with spoken language diverging from V2 more than written language. Finally, we determined that subject form (null vs overt *and* pronouns vs NPs), verb type and corpus accounted for a fair amount of the variation in verb position. Because of the infrequent use of long pronouns, we did not find an effect of pronominal form (long vs short), but did find slightly increased use of V3 order with subject pronouns even in the written data.

Examining the exceptions to V2 more closely, we found that the vast majority of V3 clauses include preverbal short subject pronouns, with some V3 clauses with postverbal lexical noun subjects. We now turn to possible analyses and explanations of these findings.

5.1 SYNTACTIC ANALYSIS

In this subsection we sketch how a particular formal analysis, that of Walkden (2017), can account for the facts presented in the previous section. The discussion is illustrative, and not intended to imply that this is the only possible or plausible analysis; some alternative possibilities are mentioned at the end of the subsection. Generally, our findings speak against analysing V2 as a unified phenomenon from a theoretical perspective, and in favour of the view that V2 effects may have subtly different ontologies in different languages.

The classic generative analysis of V2 in German and languages like it, based on den Besten (1989 [1983]), derives the V2 restriction from the nature of the highest functional

---

13   Note that examples such as (18) demonstrate that the short pronoun is not a proclitic attaching to the finite verb – at least not in all cases. This means that an analysis along the lines of van Kemenade's (1987) classic account of Old English, attributing V3 to pronominal cliticization, is not available. We are grateful to a reviewer for pointing this out.

projection in the clause, CP: only one constituent may occupy the specifier of CP, and the finite verb occupies the C head position. This account derives the fact that V2 is asymmetric and does not apply in embedded clauses: finite complementizers and the finite verb are in complementary distribution, with the former blocking the movement of the latter to C.

Walkden (2017: 60–65) departs only minimally from this classic analysis in accounting for V3 varieties like Kiezdeutsch. In this approach, the CP-domain is split into two: CP1 and CP2. The higher position, CP2, is multifunctional, and its specifier can host all the same elements as Spec,CP in the classic analysis. The specifier of CP1 is choosier: only familiar topics may occupy this position, the canonical instance of which is a pronominal subject. The finite verb occupies the lower head position, C1. Only one phrasal element may move to the CP-domain (the "bottleneck" effect: Haegeman 1996; Roberts 2004).

Let us now see how this analysis can be applied to spoken Estonian. V2 itself is easy to derive under this analysis: it is found whenever only one of the two specifier positions is filled. Thus, in a non-subject-initial V2 clause like (1), the initial constituent occupies Spec,CP2, and Spec,CP1 remains empty, since there is no appropriate familiar topic to move there: see (20a). Similarly, V1 is derived straightforwardly: it is found when either both positions are unfilled or the material in them is not pronounced. In the latter case, the result is topic-drop V1 clauses of the type in (12b), schematized in (20b).[14] In the former case, when both Spec,CP1 and Spec,CP2 remain empty and there is no clausal aboutness topic or framesetter, we derive existential/presentational V1 clauses of the type in (12c), schematized in (20c).

(20)   a. $[_{CP2}$ Kiiresti $[_{C2'} [_{CP1} [_{C1'}$ **lahku-s-id** $[_{IP}$ õpilased   koolimaja-st. $]]]]]$
           quickly   leave-PST-3PL   students schoolhouse-ELA
           'The students departed from the schoolhouse quickly.' (=1)

       b. $[_{CP2}$ Ø $[_{C2'} [_{CP1} [_{C1'}$ sa-d= $[_{IP}$ 'aigla-st   oma   tõved   kallale $]]]]]$
           (you)   get-PRS.2SG   hospital-ELA REF.GEN diseases attack-ALL
           '[You'll] get your diseases attacking you from the hospital.' (=12b)

       c. $[_{CP2} [_{C2'} [_{CP1} [_{C1'}$ Avane-s $[_{IP}$ jalg-värav $]]]]]$
                           open-PST.3SG   foot-gate.NOM
           'The pedestrian gate opened.' (=12c)

XP-S-V clauses with an unstressed pronominal subject, of the kind given in (14) above that made up the majority of examples of V3 in our data, are derived equally straightforwardly, as in the tree in (21).
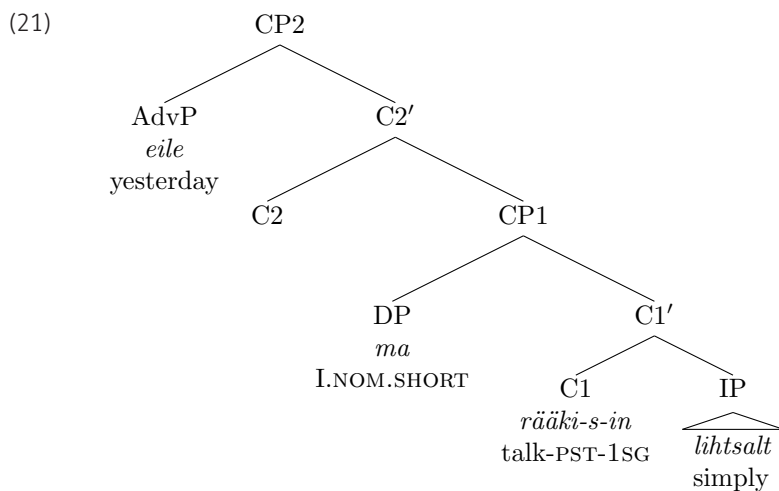
(21)

       'Yesterday I was only talking.' (=14)

Since Spec,CP1 is not restricted to pronominal subjects, the same analysis applies to examples like (15) and (16a).[15]

There are two types of clause that this analysis does not derive comfortably, both relatively rare: subject-initial V3 clauses (S-XP-V) like (18), constituting 4% of our spoken data, and verb-final and "other" clauses, together constituting 5% of our spoken data. For the verb-final and verb-late clauses, we can posit that the verb exceptionally fails to move to C1 (or at any rate is not spelled out there). Some or all of our S-XP-V clauses may also submit to such an analysis; however, here there is another option. Subject-initial V3 is found in at least one otherwise consistently V2 language, namely Dutch, as in (22) (Barbiers 1995, his example 1a):

(22)   [De krant] [gisteren] meldde het voorval niet.
        the paper   yesterday   reported the incident not
        'Yesterday's newspaper did not report the incident.'

According to Barbiers, the two preverbal elements here form a single "pseudo-DP" constituent in Dutch, with the DP moving to the specifier of the adverbial projection before this new pseudo-DP is moved to the left periphery. Without dwelling on the details, one prediction resulting from Barbiers' analysis is that manner adverbs should be ruled out in this configuration, since they are first Merged below the subject; all our Estonian examples are in line with this prediction. If this analysis is correct, these examples are not true exceptions to V2.

This concludes our brief examination of how Walkden's (2017) analysis fits the spoken Estonian data. As alluded to above, this analysis is not the only option available to us. Te Velde (2017) presents an alternative in which the verb only moves as high as the IP domain, occupying I (see also Nistov & Opsahl 2014); rather than Spec,CP1 and Spec,CP2, the specifier positions in question are Spec,IP and Spec,CP respectively. Minor and conceptual issues aside, these two families of analysis make one crucially different prediction: while the split-CP analysis predicts a substantive asymmetry between embedded and unembedded clauses (since the complementizer and finite verb compete for the CP1 position), all else being equal, the IP analysis predicts no such asymmetry. We cannot resolve this question here, since embedded clauses lie outside the scope of this paper. We can note, however, that the finite verb clearly does not move to clause-medial position in all subordinate clauses, as it would be predicted to under the V-to-I approach: this is shown by examples like (23) (from Ehala 2006: 68, his (21b)), in which the finite verb remains below objects and IP adverbs. This is a point in favour of the split-CP analysis.

(23)   kui lapse-d lõpuks supi   ära on   söönud
        if   child-PL finally soup.GEN away have eaten
        'if the children have finally eaten up the soup'

Another line of thinking ties Estonian V3 to prosody. In Walkden's (2017) analysis, deaccenting of the constituent in Spec,CP1 is a byproduct of the fact that it is a familiar topic. Holmberg, Sakhai & Tamm (2020) also argue for two specifier positions preceding the finite verb: Spec,OpP and Spec,FinP. The subject moves to Spec,FinP under this analysis, but is not necessarily spelled out there: instead, a PF condition requires that an intonation phrase (ι) immediately dominates no more than two prosodic phrases (φ), and this normally causes a lower copy in the subject chain to be spelled out. Weak pronominal subjects are able to be spelled out preverbally, however, since they do not constitute a prosodic phrase of their own. Again, it is not possible to tease these analyses apart here, a task we leave for future research.

5.2 THE SPOKEN-WRITTEN DIVIDE

Our results clearly confirm that the difference between speech and writing in Estonian, hinted at in research at least as far back as Remmel (1963), and found in corpus data examined by Lindström (2005), is real and substantial. Written Estonian is essentially a well-behaved

---

15   We also occasionally find OSV examples such as (17), which do not seem to be productive in Kiezdeutsch, but which are found in otherwise similar languages such as Old English. *Prima facie*, the bottleneck restriction ought to rule these out. Walkden (2017: 73) speculates that these involve a type of Hanging Topic construction, with the "object" in such structures first Merged in the CP-domain and the true argument of the verb being a silent clause-internal object. Since Estonian allows objects to remain unexpressed, and the initial constituent in (17) is in nominative case, this analysis seems to fit well here.

strict V2 language, at least as far as affirmative declarative clauses are concerned: 95% of our data consists of V1 and V2 clauses that are straightforwardly compatible with a German- or Swedish-style verb-second grammar. Spoken Estonian is a different kettle of fish, exhibiting 14% V3, and the analysis in 5.1 is only intended to apply to this spoken variety. Judging by this variable alone, there seem to be multiple grammars (Kroch 1994; Roeper 1999) at play in Estonian. Moreover, it would not be inappropriate to term this a situation of diglossia in the sense of Ferguson (1959).

Independently of the precise syntactic analysis, how can this situation be explained? One possibility is that the strict V2 found in written Estonian is not part of core grammar at all: instead it's a grammatical "virus" overlaid onto core competence (Sobin 1997). Characteristic of such viruses is that they are absent from the usage of the youngest children, represent prestige variants, and incur a processing cost. Against this, however, it can be noted that strict V2 does not display other hallmarks of viruses according to Sobin (1997), such as lexical specificity (sensitivity to particular lexical items) and nonlocality (insensitivity to constituent structure): V2 in written Estonian, as elsewhere, applies to all finite verbs and is sensitive to the extent of the first constituent.

Another possible hypothesis is that there is only one grammatical system at work, and that the difference between speech and writing relates to prosodic conditioning of V2. Under this hypothesis, prosodic cues that are available in speech are not available in writing, and hence other strategies must be used there. For the closely related language Finnish, Vilkuna (1989) suggests that information structure-driven constituent order principles are operative in written language, whereas in spoken language the same categories are expressed prosodically. This hypothesis is also supported by Lindström's (2005) finding that, in spoken Estonian data, constituent order varies less than in written data, a finding that is also consistent with the evidence that we have collected. More generally, there is a long tradition of linking prosody and constituent order variation in the literature on Estonian, in one way or another, e.g. recently Sahkai & Tamm (2019). Moreover, many of the deviations from V1 and V2 in our written sample, drawn from a fiction corpus, come from dialogue: 7 out of 20 of the V3 clauses in the written data include first or second-person pronouns or verbs, indicating that V3 may be used as a literary device used to convey the prosody of spoken language.

It is not obvious which of these hypotheses is correct (or whether the truth is some combination of the two, or neither); more research is needed.

5.3 SOCIOLINGUISTIC AND HISTORICAL FACTORS

Lurking in the background of any discussion of V2 syntax in Estonian is the influence of German. From the thirteenth century onwards, speakers of Low German settled in the Estonian-speaking area and achieved substantial social and economic power, joined and gradually supplanted by High German as of the fifteenth century. At the time of the emergence of an Estonian written standard in the eighteenth and nineteenth centuries, this influence was still strong, and Germanophone intellectuals played a key role in the process of standardization, when most texts were translated or written by Germans speaking L2 Estonian and Estonians educated in L2 German (Metslang 2009: 50).

Language reformers in the 20th century, particularly the highly prolific Johannes Aavik, often advocated ridding the language of German influence. Aavik (1912) considered Estonian word order to be riddled with German influence, pointing to clause-final placement of predicate complements and clause-final finite verbs in embedded clauses, as well as inversion caused by the V2 principle (1912: 356). He considered the first two to be worthier battles to fight, as V2 was common to all Germanic languages, rather than signalling specifically German influence. Reformers took various views but usually focussed on the differences between embedded and unembedded clauses; Tauli (1959: 244–245) followed Aavik's advocacy for maintaining basic word order in subordinate clauses, appealing to the fact that closely related languages like Finnish, Votic and Ingrian do not use distinct constituent order in main and embedded clauses.

An open question is whether V2 constituent order was really a direct transfer from German (or Germanic): cross-linguistic transfer of clausal constituent order is not unheard of, but also not particularly common, tending to be found only in intense contact situations (see e.g. Thomason

2001: 67–74). Another possibility, suggested by a reviewer, is that existing constituent order patterns (e.g. strict V2) were amplified and others (e.g. non-V2 orders) were suppressed, initially through conscious monitoring; this fits well with the idea that strict V2 in Estonian is a "virus" in the sense discussed in the previous subsection.

Our study does not directly speak to the historical questions, but does highlight some areas where further research is needed. Investigations using a comparable methodology and historical texts may be able to shed light on the historical development of V2 and V3 in Estonian. Here the lack of direct evidence of spoken Estonian before a certain point will of course be a major limitation, but looking at genres that are closer to speech, or which represent speech (personal letters and other egodocuments; dramas), may be revealing. In view of the previous subsection, one important question is the extent to which spoken Estonian was *ever* a V2 language in the strict sense. Establishing this may inform the broader question of whether it is possible for superstratal influence such as that of German on Estonian to lead to shifts in basic constituent order.

Another important question here is comparative. The other Finnic languages spoken in the region, such as Livonian, Ingrian and Votic, stood in a similar historical relation to Baltic German throughout their histories, as did Indo-European Baltic varieties such as Latvian. Finnish was not in such intense contact with German, but has consistently been in close contact with Swedish, another strict V2 language. None of these languages appear to show verb-second effects to the same extent as Estonian – even relaxed V2/V3 effects of the kind documented here for spoken Estonian. In a comparative, cross-linguistic corpus study of constituent order in written language, Mandel (p.c., in prep) found 68% of affirmative declarative clauses in Finnish to exhibit V2 order and only 46% in Latvian, compared to 88% in her Estonian sample. Latvian uses V3 in 37% of the affirmative declarative clauses included in her study, V4 in 11% and even V5 in 2%, while Finnish exhibits V3 in 22% of the sample. Why should these languages be so different? Was there something special about the Estonian-German contact scenario? Or could the developments in Estonian be autochthonous after all? More research is necessary in order to answer these questions.

6 CONCLUSION

This paper set out to establish the extent to which spoken and written Estonian can be characterized as verb-second languages. Drawing data from affirmative declarative main clauses in two corpora of Estonian, we were able to show that written Estonian is, to a first approximation, a well-behaved strict V2 language, whereas spoken Estonian must be characterized differently due to the large number of V3 clauses found here. A recursive partitioning tree model showed that the strongest predictor of word order (across both written and spoken data) was whether or not the subject was overt, with subjectless clauses much more likely to be V1. Among the clauses with an overt subject, written vs. spoken was the strongest predictor, with the spoken corpus containing many more deviations from V2. A random forests model additionally showed that a strong effect of subject form (both null vs. overt and pronominal vs. full NP) was present.

With the difference between spoken and written Estonian established, we showed that Walkden's (2017) analysis of V3 languages such as Kiezdeutsch and Old English was able to account straightforwardly for the majority of our spoken examples (section 5.1). Many open questions remain, such as the precise nature of written Estonian verb-second on a cognitive level (section 5.2), the role of prosody, and the historical trajectory of V2 and V3 (section 5.3) – some of which we hope to address in future research.

ABBREVIATIONS

ABL   ablative

ALL   allative

CND   conditional

ELA   elative

FOC   focus

GEN	genitive
ILL	illative
IMPL	impersonal
INF	infinitive
NOM	nominative
PART	emphatic particle
PL	plural
PRS	present
PST	past
3PL	third person plural
1/2/3SG	first/second/third person singular
V1/2/3/#/x	verb-first/second/third/final/other-position

ACKNOWLEDGEMENTS

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Virve-Anneli Vihman ⓘ *orcid.org/0000-0002-0874-1730*
University of Tartu, Ülikooli 18, 50090 Tartu, Estonia

George Walkden ⓘ *orcid.org/0000-0001-5950-9686*
University of Konstanz, PO Box D 175, 78457 Konstanz, Germany

REFERENCES

Aavik, Johannes. 1912. Kõige suurem germanismus Eesti keeles [The greatest Germanism in Estonian]. *Eesti Kirjandus [Estonian Literature]* 9. 353–369.

Alexiadou, Artemis & Terje Lohndal. 2018. V3 in Germanic: A comparison of urban vernaculars and heritage languages. In Mailin Antomo & Sonja Müller (eds.), *Non-canonical verb positioning in main clauses*, 245–264. Hamburg: Helmut Buske.

Baayen, R. Harald, Anna Endresen, Laura A. Janda, Anastasia Makarova & Tore Nesset. 2013. Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics,* 37(3), 253–291. DOI: *https://doi.org/10.1007/s11185-013-9118-6*

Barbiers, Sjef. 1995. Another case of scrambling in Dutch. *Algemene Vereniging voor Taalwetenschap* 12. 13–24. DOI: *https://doi.org/10.1075/avt.12.04bar*

Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1). 5–32. DOI: *https://doi.org/10.1023/A:1010933404324*

Cardinaletti, Anna & Michal Starke. 1999. The typology of structural deficiency: a case study of the three classes of pronoun. In Henk van Riemsdijk (ed.), *Clitics in the Languages of Europe*, 145–235. Berlin: Mouton de Gruyter. DOI: *https://doi.org/10.1515/9783110804010.145*

de Sivers, Fanny. 1969. *Analyse grammaticale de l'estonien parlé*. Clermont-Ferrand: G. de Bussac.

den Besten, Hans. 1989 [1983]. On the interaction of root transformations and lexical deletive rules. In: Werner Abraham (ed.), *On the formal syntax of the Westgermania*, 47–131. Amsterdam: John Benjamins. Reprinted in Hans den Besten (1989), *Studies in West Germanic syntax*. Tilburg: Katholieke Universiteit Brabant dissertation. Amsterdam/Atlanta, GA: Rodopi. DOI: *https://doi.org/10.1075/la.3.03bes*

den Besten, Hans. 2002. Khoekhoe syntax and its implications for L2 acquisition of Dutch and Afrikaans. *Journal of Germanic Linguistics* 14. 3–56. DOI: *https://doi.org/10.1017/S1470542702046020*

Diesing, Molly. 1990. Verb movement and the subject position in Yiddish. *Natural Language & Linguistic Theory* 8(1). 41–79. DOI: *https://doi.org/10.1007/BF00205531*

Dryer, Matthew S. 2013a. Order of Subject, Object and Verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (*http://wals.info/chapter/81*) (Accessed 2019-12-08).

Dryer, Matthew S. 2013b. Expression of pronominal subjects. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (*http://wals.info/chapter/101*) (Accessed 2019-12-15).

Duvallon, Outi & Antoine Chalvin. 2004. La réalisation zéro du pronom sujet de première et de deuxième personne du singulier en finnois et en estonien parlés. *Linguistica Uralica* 40(4). 270–286.

Ehala, Martin. 1998. How a man changed a parameter value: the loss of SOV in Estonian subclauses. In Richard Hogg & Linda van Bergen (eds.), *Historical Linguistics 1995*, vol. 2: *Germanic linguistics*, 73–88. Amsterdam: John Benjamins. DOI: *https://doi.org/10.1075/cilt.162.07eha*

Ehala, Martin. 2006. The word order of Estonian: implications to universal language. *Journal of Universal Language* 7. 49–89. DOI: *https://doi.org/10.22425/jul.2006.7.1.49*

Erelt, Mati, Tiiu Erelt & Kristiina Ross. 1997. *Eesti Keele Käsiraamat* [Handbook of Estonian]. Tallinn: Eesti Keele Sihtasutus. (*www.eki.ee/books/ekk09/*) (Accessed 2019-12-08).

Ferguson, Charles. 1959. Diglossia. *Word* 15. 325–340. DOI: *https://doi.org/10.1080/00437956.1959.11659702*

Freywald, Ulrike, Leonie Cornips, Natalia Ganuza, Ingvild Nistov & Toril Opsahl. 2015. Beyond verb second – a matter of novel information-structural effects? Evidence from German, Swedish, Norwegian and Dutch. In Jacomine Nortier & Bente A. Svendsen (eds.), *Language, youth and identity in the 21st century: linguistic practices across urban spaces*, 73–92. Cambridge: Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9781139061896.006*

Gärtner, Hans-Martin. 2016. A note on the Rich Agreement Hypothesis and varieties of embedded V2. *Working Papers in Scandinavian Syntax* 96. 1–13.

Gilligan, Gary. 1987. *A cross-linguistic approach to the pro-drop parameter*. Los Angeles, CA: University of Southern California dissertation.

Haeberli, Eric. 2002. Observations on the loss of verb second in the history of English. In C. Jan-Wouter Zwart & Werner Abraham (eds.), *Studies in Comparative Germanic Syntax: Proceedings from the 15th Workshop on Comparative Germanic Syntax*, 245–272. Amsterdam: John Benjamins. DOI: *https://doi.org/10.1075/la.53.15hae*

Haegeman, Liliane. 1996. Verb second, the split CP, and null subjects in early Dutch finite clauses. *GenGenP* 4. 135–175.

Haegeman, Liliane & Ciro Greco. 2018. West Flemish V3 and the interaction of syntax and discourse. *Journal of Comparative Germanic Linguistics* 21(1). 1–56. DOI: *https://doi.org/10.1007/s10828-018-9093-9*

Holmberg, Anders. 2015. Verb-second. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – theory and analysis: an international handbook* 1. 342–382. Berlin: Mouton de Gruyter.

Holmberg, Anders. 2017. Linguistic typology. In Ian Roberts (ed.), *The Oxford handbook of universal grammar,* 355–376. Oxford: Oxford University Press. DOI: *https://doi.org/10.1093/oxfordhb/9780199573776.013.15*

Holmberg, Anders, Heete Sahkai & Anne Tamm. 2020. Prosody distinguishes Estonian V2 from Finnish and Swedish. *Proceedings of the 10th International Conference on Speech Prosody 2020,* 439–443. DOI: *https://doi.org/10.21437/SpeechProsody.2020-90*

Holmberg, Anders & Urpo Nikanne. 2002. Expletives, subjects and topics in Finnish. In Peter Svenonius (ed.), *Subjects, expletives, and the EPP*, 71–106. Oxford: Oxford University Press.

Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674. DOI: *https://doi.org/10.1198/106186006X133933*

Hsu, Brian. 2017. Verb second and its deviations: an argument for feature scattering in the left periphery. *Glossa: A Journal of General Linguistics* 2(1). 35. DOI: *https://doi.org/10.5334/gjgl.132*

Hsu, Brian. 2021. Coalescence: a unification of bundling operations in syntax. *Linguistic Inquiry* 52(1). 39–87. DOI: *https://doi.org/10.1162/ling_a_00372*

Janda, Laura A. 2013. Quantitative methods in Cognitive Linguistics: An introduction. In Laura A. Janda (ed.), *Cognitive linguistics: The quantitative turn. The essential reader*, 1–32. Berlin & Boston, MA: De Gruyter Mouton. DOI: *https://doi.org/10.1515/9783110335255.1*

Keevallik, Leelo. 2003. Colloquial Estonian. In Mati Erelt (ed.), *Estonian language*, 342–378. Tallinn: Estonian Academy Publishers.

Kivik, Piibi-Kai. 2010. Personal pronoun variation in language contact: English in the United States. In Cornelius Hasselblatt, Bob de Jonge & Muriel Norde (eds.), *Language contact: new perspectives*, 63–86. Amsterdam: John Benjamins. DOI: *https://doi.org/10.1075/impact.28.05kiv*

Kroch, Anthony S. 1994. Morphosyntactic variation. In Katherine Beals (ed.), *Proceedings of the thirtieth annual meeting of the Chicago Linguistics Society*, 180–201. Chicago: Chicago Linguistics Society.

Lenker, Ursula. 2000. *Soþlice* and *witodlice*: Discourse markers in Old English. In Olga Fischer, Anette Rosenbach & Dieter Stein (eds.), *Pathways of change: Grammaticalization in English*, 229–249. Amsterdam: John Benjamins. DOI: *https://doi.org/10.1075/slcs.53.12len*

Lindström, Liina. 2000. Narratiiv ja selle sõnajärg [Narrative and its word order]. *Keel ja Kirjandus [Language and Literature]* 3. 190–200.

Lindström, Liina. 2001. Verb-initial clauses in narrative. *Estonian: Typological Studies* 5. 138–168.

Lindström, Liina. 2005. *Finiitverbi asend lauses. Sõnajärg ja seda mõjutavad tegurid suulises eesti keeles [The position of the finite verb in the sentence. Word order and factors affecting it in spoken Estonian]*. Tartu: University of Tartu dissertation.

Lindström, Liina. 2007. Verb-final clauses in spoken Estonian. In Márta Csepregi & Virpi Masonen (eds.), *Grammar and context: New approaches to the Uralic languages*, 227–247. Budapest: Eötvös Loránd.

Lindström, Liina. 2017. Lause infostruktuur ja sõnajärg [Sentential information structure and word order]. In Mati Erelt & Helle Metslang (eds.), *Eesti keele süntaks [Estonian syntax]*, 547–565. Tartu: University of Tartu Press.

Lindström, Liina & Virve-Anneli Vihman. 2017. Who needs it? Variation in experiencer marking in Estonian "need"-constructions. *Journal of Linguistics* 53 (4), 789–822. DOI: *https://doi.org/10.1017/S0022226716000402*

Lohndal, Terje, Marit Westergaard & Øystein A. Vangsnes. 2020. Verb second in Norwegian: variation and acquisition. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 770–789. Oxford: Oxford University Press. DOI: *https://doi.org/10.1093/oso/9780198844303.003.0033*

Mandel, Aive. (in prep). *Sõnajärje võrdlus eesti, soome ja läti keeles [A comparison of word order in Estonian, Finnish and Latvian]*. Masters thesis (in preparation), University of Tartu.

Manninen, Satu & Diane Nelson. 2004. What is a passive? The case of Finnish. *Studia Linguistica*, 58(3), 212–251. DOI: *https://doi.org/10.1111/j.0039-3193.2004.00115.x*

Meelen, Marieke, Khalid Mourigh & Lisa Lai-Shen Cheng. 2020. V3 in urban youth varieties of Dutch. In András Bárány, Theresa Biberauer, Jamie Douglas & Sten Vikner (eds.), *Syntactic architecture and its consequences II: Between syntax and morphology*, 335–363. Berlin: Language Science Press.

Metslang, Helle. 2009. Estonian grammar between Finnic and SAE: some comparisons. *Sprachtypologie und Universalienforschung* 62. 49–71. DOI: *https://doi.org/10.1524/stuf.2009.0004*

Mörnsjö, Maria. 2002. *V1 declaratives in spoken Swedish. Syntax, information structure and prosodic pattern*. Lund: Lund University dissertation.

Nistov, Ingvild & Toril Opsahl. 2014. The social side of syntax in multilingual Oslo. In Brit Mæhlum & Tor Åfarli (eds.), *The sociolinguistics of grammar*, 91–116. Amsterdam: John Benjamins. DOI: *https://doi.org/10.1075/slcs.154.05nis*

Nygård, Mari. 2013. *Discourse ellipsis in spontaneously spoken Norwegian*. Trondheim: Norwegian University of Science and Technology dissertation.

Pajusalu, Renate. 2005. Anaphoric pronouns in Spoken Estonian: Crossing the paradigms. In R. Laury (ed.) *Minimal reference: The use of pronouns in Finnish and Estonian discourse*, 107–134. Helsinki: SKS.

Pajusalu, Renate. 2017. Viiteseosed [Referential relations]. In Mati Erelt & Helle Metslang (eds.), *Eesti keele süntaks [Estonian syntax]*, 566–589. Tartu: University of Tartu Press.

Pasch, Renate, Ursula Brauße, Eva Breindl & Ulrich H. Waßner. 2003. *Handbuch der deutschen Konnektoren*, vol. 1: *Linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüpfer (Konjunktionen, Satzadverbien und Partikeln)*. Berlin: de Gruyter. DOI: *https://doi.org/10.1515/9783110201666*

Platzack, Christer. 1986. COMP, INFL, and Germanic word order. In Lars Hellan & Kirsti Koch Christensen (eds.), *Topics in Scandinavian Syntax*, 185–234. Dordrecht: Kluwer. DOI: *https://doi.org/10.1007/978-94-009-4572-2_9*

R Development CoreTeam. 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Accessible at *http://www.R-project.org/*.

Raag, Virve. 1998. *The effects of planned change on Estonian morphology*. Studia Uralica Upsaliensia, 29. Uppsala: Acta Universitatis Upsaliensis.

Remmel, Nikolai. 1963. Sõnajärjestus eesti lauses. Deskriptiivne käsitlus [Word order in the Estonian sentence. A descriptive treatment]. In Richard Kress (ed.), *Eesti keele süntaksi küsimusi [Issues in Estonian syntax]*, 216–271. Tallinn: Eesti Riiklik Kirjastus.

Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris. DOI: *https://doi.org/10.1515/9783110883718*

Roberts, Ian. 2004. The C-system in Brythonic Celtic languages, V2 and the EPP. In Luigi Rizzi (ed.), *The cartography of syntactic structures*, vol. 2: *The structure of CP and IP*, 297–328. Oxford: Oxford University Press.

Roeper, Tom W. 1999. *Universal bilingualism: Bilingualism, language and cognition*. Cambridge: Cambridge University Press. DOI: *https://doi.org/10.1017/S1366728999000310*

Rögnvaldsson, Eiríkur & Höskuldur Thráinsson. 1990. On Icelandic word order once more. In Joan Maling & Annie Zaenen (eds.), *Syntax and semantics 24: modern Icelandic syntax*, 3–40. San Diego, CA: Academic Press.

Ross, John Robert. 1982. Pronoun-deleting processes in German. *Paper presented at the annual meeting of the Linguistic Society of America*, San Diego, California.

Sahkai, Heete & Anne Tamm. 2019. Verb placement and accentuation: does prosody constrain the Estonian V2? *Open Linguistics* 5, 729–753. DOI: *https://doi.org/10.1515/opli-2019-0040*

Sobin, Nicholas. 1997. Agreement, default rules, and grammatical viruses. *Linguistic Inquiry* 28. 318–343.

Speyer, Augustin. 2010. *Topicalization and stress clash avoidance in the history of English*. Berlin: de Gruyter. DOI: *https://doi.org/10.1515/9783110220247*

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9(1). 307. DOI: *https://doi.org/10.1186/1471-2105-9-307*

Swan, Toril. 1994. A note on Old English and Old Norse initial adverbials and word-order with special reference to sentence adverbials. In Toril Swan, Endre Mørck & Olaf Jansen (eds.), *Language change and language structure: older Germanic languages in a comparative perspective*, 233–270. Berlin: de Gruyter.

Tael, Kaja. 1988. *Sõnajärjemallid eesti keeles (võrreldduna soome keelega)* [Word order patterns in Estonian (in comparison with Finnish)]. Tallinn: TA Keele ja Kirjanduse Instituut.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178. DOI: *https://doi.org/10.1017/S0954394512000129*

Tauli, Valter. 1959. Eesti kirjakeele sõnajärje probleeme [Issues of word order in standard Estonian]. *Virittäjä*, 241–251.

te Velde, John. 2017. Temporal adverbs in the Kiezdeutsch left periphery: combining late merge with deaccentuation for V3. *Studia Linguistica* 71(3). 301–336. DOI: *https://doi.org/10.1111/stul.12055*

Thomason, Sarah Grey. 2001. *Language contact: an introduction*. Edinburgh: Edinburgh University Press.

Trutkowski, Ewa. 2016. *Topic drop and null subjects in German*. Berlin: Walter de Gruyter.

Vainikka, Anne & Yonata Levy. 1999. Empty subjects in Finnish and Hebrew. *Natural Language & Linguistic Theory* 17(3). 613–671. DOI: *https://doi.org/10.1023/A:1006225032592*

van Kemenade, Ans. 1987. *Syntactic case and morphological case in the history of English*. Dordrecht: Foris. DOI: *https://doi.org/10.1515/9783110882308*

van Kemenade, Ans, Tanja Miliev & R. Harald Baayen. 2008. The balance between discourse and syntax in Old and Middle English. In Marina Dossena, Maurizio Gotti & Richard Dury (eds.), *English historical linguistics 2006,* volume I: *Syntax and morphology*, 3–22. Amsterdam: John Benjamins. DOI: *https://doi.org/10.1075/cilt.295.04kem*

van Urk, Coppe & Norvin Richards. 2015. Two components of long-distance extraction: successive cyclicity in Dinka. *Linguistic Inquiry* 46(1). 113–155. DOI: *https://doi.org/10.1162/LING_a_00177*

Vikner, Sten. 1995. *Verb movement and expletive subjects in the Germanic languages*. Oxford: Oxford University Press.

Vilkuna, Maria. 1998. Word Order in European Uralic. In Anna Siewierska (ed.), *Constituent order in the languages of Europe*, 173–233. Berlin: Mouton de Gruyter. DOI: *https://doi.org/10.1515/9783110812206.173*

Walkden, George. 2014. *Syntactic reconstruction and Proto-Germanic*. Oxford: Oxford University Press. DOI: *https://doi.org/10.1093/acprof:oso/9780198712299.001.0001*

Walkden, George. 2015. Verb-third in early West Germanic: a comparative perspective. In Theresa Biberauer & George Walkden (eds.), *Syntax over time: lexical, morphological, and information-structural interactions*, 236–248. Oxford: Oxford University Press. DOI: *https://doi.org/10.1093/acprof:oso/9780199687923.003.0014*

Walkden, George. 2017. Language contact and V3 in Germanic varieties new and old. *Journal of Comparative Germanic Linguistics* 20. 49–81. DOI: *https://doi.org/10.1007/s10828-017-9084-2*

Walkden, George & Hannah Booth. 2020. Reassessing the historical evidence for embedded Verb Second. In Rebecca Woods & Sam Wolfe (eds.), *Rethinking verb second*, 536–554. Oxford: Oxford University Press. DOI: *https://doi.org/10.1093/oso/9780198844303.003.0022*

Walkden, George & Kristian Rusten. 2017. Null subjects in Middle English. *English Language and Linguistics* 21(3). 439–473. DOI: *https://doi.org/10.1017/S1360674316000204*

Weerman, Fred. 1989. *The V2 conspiracy: a synchronic and diachronic analysis of verbal positions in Germanic languages*. Berlin: de Gruyter. DOI: *https://doi.org/10.1515/9783110250442*

Wolfe, Sam. 2019a. Redefining the typology of V2 languages: the view from Medieval Romance and beyond. *Linguistic Variation* 19(1). 16–46. DOI: *https://doi.org/10.1075/lv.15026.wol*

Wolfe, Sam. 2019b. *Verb second in medieval Romance*. Oxford: Oxford University Press.

Ziegelmann, Katja & Winkler, Eberhard. 2006. Zum Einfluß des Deutschen auf das Estnische. In Anne Arold, Dieter Cherubim, Dagmar Neuendorff & Henrik Nikula (eds.), *Deutsch am Rande Europas,* 44–70. Tartu: University of Tartu Press.