

Variation in negation in Seto

Liina Lindström¹, Maarja-Liisa Pilvik¹ and Helen Plado^{1,2}

¹ University of Tartu | ² Võro Institute

Seto is an exceptional language in the Uralic family due to its systematic use of postverbal negation, although preverbal and double negation marking are also used. Postverbal negation is still the most frequent and unmarked pattern occurring in about 74% of negative clauses in Seto. This paper analyzes variation between pre- and postverbal negation in East Seto (spoken in present-day Russia), based on data gathered during fieldwork trips in 2010–2013. By applying quantitative methods that are used in variationist studies (regression modelling, conditional inference trees, and random forests), we determine the variables affecting the choice between pre- and postverbal negation. Marked preverbal negation occurs more likely with first and third person, cognition verbs, and present tense, all of which are often used in fixed expressions like *I don't know*. We also found a strong structural persistence effect in the data and remarkable differences between individual speakers.

Keywords: standard negation, word order, language variation, Uralic languages, South Estonian Seto

1. Introduction

There is a typological tendency for standard negation markers to be preverbal, i.e. to occur before the main verb, with postverbal negation markers being much more rare (Dahl 2010: 23, Dryer 2013, Vossen 2016). Thus, it is not surprising that most Uralic languages have only preverbal negation. These markers originate from the negative auxiliary verb of Proto-Uralic, which was conjugated for mood, tense, and person (Janhunen 1982). Indeed, in contemporary Uralic languages spoken dispersedly in a large area in Northern Eurasia, preverbal negation and the negative auxiliary strategy dominate (Miestamo et al. 2015).

With this in mind, the South Estonian dialect Seto is exceptional as a Uralic language in terms of its predominantly postverbal negation (Pajusalu, to appear), while all of the main contact languages surrounding it (Standard Estonian, Russ-

ian, Latvian) only use preverbal negation. There is no good historical explanation for the emergence and development of postverbal negation in Seto. Similar phenomena, although to a lesser degree, can also be found in a few other Finnic varieties, e.g. Veps and Lude. In these languages, however, postverbal negation is only a minor pattern, while in Seto it is the most common pattern for negation.

When we look more closely at the Seto data, we find a lot of variation in the position of the negation marker: the negator may occur preverbally (1), postverbally (2) or it can be doubled (3). Seto uses different negators for present and past tense – *ei(?)*¹ for present tense (examples 1 and 3), *es* for past tense (example 2). Both may also attach to the end of the main verb.

Dryer (2013) has noticed the following tendency: if a negation paradigm includes different morphemes, then the position of the negator depends on semantic or grammatical factors. However, if the same morpheme is used, the negator's position is determined based on extragrammatical factors. In Seto, both present (*ei*) and past negators (*es*) occur both pre- and postverbally but it is not clear whether they themselves can synchronically be treated as two versions of the same negation marker. Moreover, there are no studies on what drives the variation in their syntactic position. Is it due to language-internal and functional factors, i.e. dependent on the linguistic context, or is it the result of language contact and an indicator of language change?

- (1) *ei vōi? jättä? vällä*
 NEG can.CNG leave.INF out
 ‘(one) can not leave (it) outside’
- (2) *ku marju saa es syss...*
 when berry.PL.PRT get.CNG NEG.PST then
 ‘when (we) did not get any berries, then...’
- (3) *inne ei tulõ vällä ei? ku keskpäivä*
 before NEG come.CNG out NEG than midday.PRT
 ‘(S)he doesn’t come out before midday.’

In this paper, we take a closer look at the variation between pre- and postverbal negation in East Seto (spoken in present-day Russia), based on data gathered during fieldwork trips in 2010–2013. By applying quantitative methods widely used in variationist studies (regression modelling, conditional inference trees, and random forests; see e.g. Tagliamonte & Baayen 2012, Baayen et al. 2013), we look for the variables that affect the choice between pre- and postverbal negation. We also

1. In most cases, *ei* is used preverbally and *ei?* with a glottal stop postverbally. However, the stop *?* is quite often left unpronounced, especially if the following word begins with a vowel.

compare the data from 2010–2013 to the data recorded in the same area in the 1970s and 1980s, currently held in the Corpus of Estonian Dialects (CED)², in order to outline how the position of negation words has changed over time. Since East Seto is a highly endangered variety we assume that the use of postverbal negation has decreased under the influence of the main contact languages Standard Estonian and Russian. Thus, we expect to see more variation in the contemporary data, compared to what has been described in previous dialect overviews. We also assume that the variation in Seto can be explained mainly by extralinguistic factors, i.e. individual adaptation of the preverbal negation order found in contact languages.

The paper is structured as follows: we first give a sociolinguistic overview of Seto with a special focus on East Seto (Section 2). We then discuss the negation system in Seto and also provide a short historical background of negation marking in Uralic (Section 3). This is followed by introducing our sample data, speakers, and methods (Section 4). Then, we present the results of this study (Section 5): first, we show the general distribution of negation patterns in the sample data and provide comparisons with earlier data from the same area as well as from other nearby dialect regions; this is followed by the presentation and discussion of the results of the statistical analyses. Finally, we draw conclusions about the East Seto negation system and the factors conditioning the variation between pre- and postverbal negation (Section 6).

2. Sociolinguistic overview of Seto

Seto is a South Estonian variety, which is spoken in the border area of Estonia and Russia. It belongs to the Finnic branch of Uralic languages. South Estonian diverged from Proto-Finnic before other Finnic languages (Sammallahti 1977, Viitso 1985, Kallio 2012) and thus it remains rather distinct.

Linguistically South Estonian is a separate language from North Estonian, which has served as a basis for Standard Estonian. Crucial differences between South and North Estonian can be found on all linguistic levels (Pajusalu et al. 2009, see also overviews on South Estonian Seto and Võro by Pajusalu, to appear; Plado, Lindström & Iva, to appear). The Ethnologue catalogue of world's languages includes the biggest variety of South Estonian – Võro – as a separate language (ISO 639-3 code vro), and mentions Võro-Seto as its alternative name (Eberhard et al. 2020). The Glottolog database of world's languages, especially the lesser known languages (Hammarström et al. 2020), uses Kallio (2014) to further divide South

2. <https://doi.org/10.1515/1-00-0000-0000-0000-00076L>

Estonian into Leivu-Lutsi, Seto-Kraasna, and Võro-Mulgi-Tartu varieties, based on the country where they are considered to be spoken (Latvia, Russia, and Estonia, respectively).

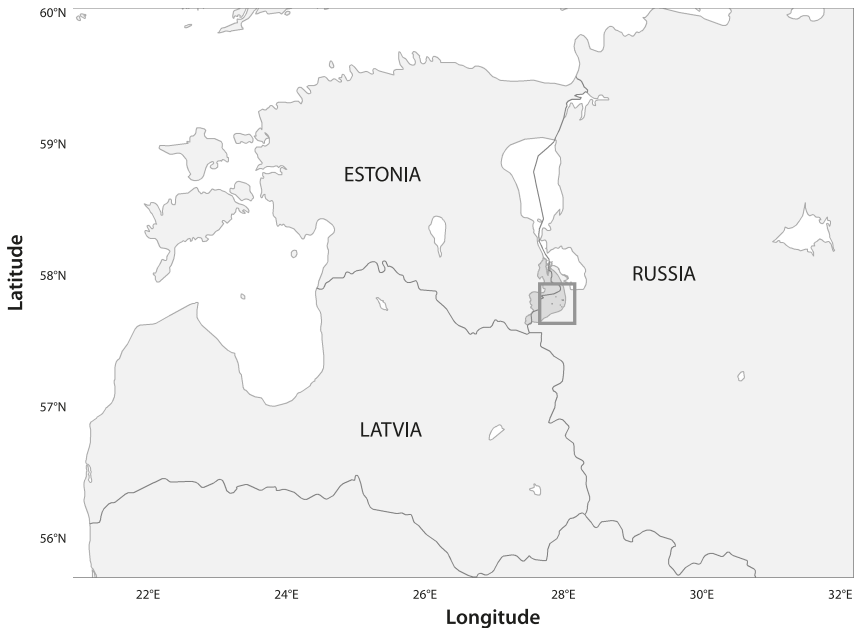
Due to historical reasons (namely close contacts with North Estonian and the development of written standard Estonian on the basis of North Estonian in the 19th century), South Estonian has been significantly influenced by North Estonian and has become more and more assimilated to it. In Estonian dialect classifications, South Estonian has been classified as a dialect group of Estonian, consisting of Tartu, Mulgi, Võro, and Seto dialects (Kask 1984). Tartu and Mulgi, spoken on the border area with North Estonian, have had more contact with North Estonian, which is why they have become linguistically closer to Standard Estonian than Seto and Võro. Võro and especially Seto have retained many old South Estonian features that have been lost in Mulgi and Tartu.

Seto is therefore especially closely related to South Estonian Võro and in most classifications (e.g. Pajusalu et al. 2009, Keem & Käsi 2002) it is regarded as a variety of Võro. The main differences between Võro and Seto have emerged due to the strong linguistic, cultural and religious influence that Russian has had on Seto. Linguistically, Russian influence is mainly lexical (Seto has a large number of Russian loanwords, e.g. *paaba* ‘woman’ < Russian *баба*) and phonological (Seto has adopted the velar *l* /*L*/). However, as more recent studies in Estonian dialect syntax (e.g. Uiboed 2013, Lindström et al. 2015, 2019, Lindström & Uiboed 2017, Lindström, to appear) have shown, Seto and Võro also differ syntactically.

Although Seto and Võro are linguistically close to each other, speakers of the two varieties regard themselves as clearly distinct communities, mainly due to cultural reasons. The most significant difference lies in religion: Setos are typically members of the Orthodox church whereas Võro people are usually Lutheran. Religion, folklore, traditions, language, and regional identity in the border area between Estonia and Russia constitute the main pillars of Seto identity. While Võro speakers identify as Estonians, Setos regard themselves as being neither Estonians nor Russians (Paas 1927: 8–9). Jääts (2000: 651) claims that nowadays, “[...] Setos are somewhere between being a regional subdivision of Estonians and a separate ethnos”. Since 2010 Setos are officially regarded as an individual ethnicity in the Russian Federation (Juhkason et al. 2012: 12).

Seto is traditionally spoken on both sides of the Estonian-Russian border (Map 1). According to the 2011 Population and Housing Census there are 12,500 Seto language speakers in Estonia, both in the traditional language area as well as in bigger towns outside of the traditional region. Additionally, there are Seto speakers in the Pskov region of the Russian Federation (the Pechory district). In 1944, shortly after World War II, there were ~5,600 Setos in the Pechory district (Jääts 1998: 1531), but their numbers have decreased continuously ever since. Dur-

ing the Soviet period (1944–1991) there was an extensive emigration of Setos to the Estonian SSR (Jääts 1998: 1531–1532, Lõuna 2003: 129) with many others joining them after Estonia regained its independence in 1991. According to the 2002 Russian census,³ there are 170 Setos in the Pechory district. During our fieldwork trips in 2010 and 2011 we were told that there are about 210 Setos living in the Pechory district and Pskov (Juhkason et al 2012: 11). Seto is thus a highly endangered language variety. This especially applies to East Seto, which is spoken in the Pechory district of the Russian Federation. Its other varieties, i.e. North Seto and South Seto, are spoken in present-day Estonia. These varieties have more speakers and they have been influenced by Standard Estonian more than East Seto.



Map 1. The Seto language area (in darker shade) and the area of the study (villages marked within the rectangular borders)

In the Pechory district, Russians and Setos have lived side by side for a long time, traditionally in different villages. Even during the period between the two world wars the separation of villages persisted and there were almost no mixed marriages (Jääts 1998: 1152). The situation changed after World War II when Russian became the language of power and bureaucracy. The number of Seto-Russian marriages increased and younger Setos knew Russian very well (Jääts

3. <http://www.afunrf.ru/narodyi/seto/>; last accessed on 29.06.2021.

1998:1536–1537). Nowadays, Setos in that area do not constitute as strong a community, with typically only 1–3 Setos living in each village. The chances to use the Seto language in everyday life have thus significantly diminished.

In 1919 primary education became obligatory and a large number of schools were founded in the rural areas during the 1920s (Jääts 1998:1148–1150, Lõuna 2003: 46–47), providing education in Estonian. Seto has never been the language of instruction nor a subject in the schools of the Pechory district. Estonian continued to be the language of instruction in some schools of the Pechory district until the 1970s (Juhkason et al. 2012:15). Pechory also had an Estonian high school, where the teaching language was Estonian until 2005. After the Estonian schools were closed in the rural areas of the Pechory district, most Seto children attended the Estonian school in Pechory. Some also attended the nearby schools in the Estonian Soviet Republic.

The majority of Setos in the Pechory district follow Estonian media. Most of them watched Estonian TV until 2010, when their access to these channels became restricted. However, they still listen to Estonian radio and read the Seto-Estonian bilingual newspaper *Setomaa*.

In conclusion, Seto, and especially East Seto spoken in present-day Russia, is a highly endangered South Estonian variety with a constantly decreasing number of speakers. The East Seto variety is linguistically interesting because it has had less contact with Standard Estonian than the Seto varieties spoken in Estonia. As a result, East Seto has retained many features that have been lost in the other varieties of Seto due to the strong influence of Standard Estonian.

3. Negation in Seto and its historical background

3.1 Negation in Seto

Seto marks standard negation with a negation word and a connegative form of the main verb. The distinction between present and past tense is made with the negation word, with *ei* being used for present tense and *es* for past tense. Both the present and past negation words may exhibit some phonological variation due to coarticulation in postverbal position, i.e. it is sometimes assimilated with the verb, becoming a suffix (e.g. *ei* ~ *õi* ~ *ai* ~ *i*; *es(s)* ~ *õs(s)* ~ *as(s)* ~ *is(s)* ~ *s(s)*). The lexical verb occurs in the connegative form, i.e. as a verb stem, which is unmarked for person and tense and ends with a glottal stop (ʔ). The use of the glottal stop is inconsistent: it can also be moved to the end of the negation construction or left out altogether. The formation of standard negation in Seto is summarized in Table 1.

Table 1. Standard negation in Seto

	Affirmative	Negative
Present tense	<i>maʔ kōnōlō</i> ‘I speak’ <i>saʔ kōnōlōt</i> ‘you speak’ <i>tä kōnōlōs</i> ‘(s)he speaks’	<i>maʔ/saʔ/tä kōnōlō-iʔ ~ ei kōnōlōʔ</i> ‘I/you don’t speak / (s)he doesn’t speak’
Past tense	<i>maʔ kōnōli</i> ‘I spoke’ <i>saʔ kōnōlit</i> ‘you spoke’ <i>tä kōnōl</i> ‘(s)he spoke’	<i>maʔ/saʔ/tä kōnōlō-s ~ es kōnōlōʔ</i> ‘I/you/(s)he didn’t speak’

As already mentioned in Section 1, the position of the negation word in Seto is subject to variation: most commonly it occurs postverbally (Pajusalu et al. 2009: 189, Pajusalu, to appear), with preverbal negation being less frequent. In either case, the negation word generally occurs immediately adjacent to the main verb, which is also a typological tendency (Dahl 2010: 23). However, our data includes nine observations (three postverbal and six preverbal) in which an unstressed particle (4) or a pronoun occurs between the main verb and the negation word.

EAST SETO⁴

- (4) *a tiijäʔ ka ei nä-i-de tüü asʔju*
but know.CNG PTCL NEG they-PL-GEN work.GEN thing.PL.PRT
‘I don’t know about their work.’

There is also an option to double the negation marker: the first negation word precedes or follows the main verb (compare examples 3 and 5) and it is later repeated at the end of the clause or at the end of the scope of negation (Lindström 1997). Double negation occurs both in Võro and Seto. The use of double negation is analysed in Section 5.1.2.

EAST SETO

- (5) *olõ = õss hüä-ʔ ubina-ʔ õss*
be.CNG NEG.PST good-PL apple-PL NEG.PST
‘These were not good apples.’ (CED)

Prohibitives (negative imperatives) are formed either with the negator (*e*)*i*(ʔ), as in (6), or the negation verb *är*(ʔ) ~ *ar*(ʔ) (7) in South Estonian. The chosen strat-

4. Examples mostly follow the transcription which is used in the Corpus of Estonian Dialects (CED): <˘> at the beginning of the word stands for the third quantity degree, <˘> for palatalization of the preceding consonant, <(.)> for short pauses, <(…)> for long pauses, <=> for coarticulations of words. The glottal stop is marked here with ʔ, instead of *q* as in the CED. Both the negation word and the finite verb are in bold.

egy restricts the negation word's position: *ei* (or its variants) occurs only postverbally, *är(?) ~ ar(?)* only preverbally. In both cases the main verb is conjugated in third person imperative (optative) form with the suffix *gu/ku*. Double negation can also occur with prohibitives as a combination of both strategies, but it is rather rare. The formation of prohibitives is summarized in Table 2.

EAST SETO

- (6) *put-ku ei säl midägi apparaattõ*
 touch-IMP NEG there anything.PRT machine.PL.PRT
 'Don't touch the machines there.'

EAST SETO

- (7) *siss=ku maa omm viil k'ülm (...) siss ärä külba-kku linna*
 then=when ground be.3SG still cold then NEG.2SG sow-IMP flax.PRT
 'Don't sow flax when the ground is cold.' (CED)

Table 2. The imperative, standard negation, and prohibitive forms of the verb *lugõma* 'read' in Seto

Person	Imperative	Standard negation	Prohibitive
Sg2	<i>Loe?!</i> 'Read!'	<i>loe-i?</i> '(You) don't read'	<i>Lugõgu-i?! ~ Ar?/Är? lugõgu!</i> 'Don't read!'
Pl2	<i>Lugõgõ?!</i> 'Read!'	<i>loe-i?</i> '(You) don't read'	<i>Lugõgu-i?! ~ Ar?/Är? lugõgu!</i> 'Don't read!'

In regard to the variation in word order of prohibitives, the postverbal negator is used in the Seto and Eastern Võro areas, while preverbal prohibitives are widespread in the rest of the Võro area and rare in Seto. Preverbal prohibitives are structurally closer to Standard Estonian, which uses the preverbal imperative negation verb *ära* with a person suffix on the main verb (see Table 3 in Section 3.2.2). In Section 5.1.3 we look more closely at the tendencies related to the use of pre- and postverbal negation words in prohibitives.

In the next sections, a historical background for Uralic negation is provided, with special reference to the Finnic languages.

3.2 Negation in Uralic/Finnic: A general background

3.2.1 Standard negation

Most Uralic languages have asymmetric negation, i.e. their negative constructions exhibit more differences from the corresponding affirmative constructions than just an added negation marker (Miestamo 2005: 351). The historic reconstruction

of Proto-Uralic has been attributed a negative auxiliary verb (NegAux) that was conjugated for mood, tense and person. The negative auxiliary stem has been reconstructed as **e* for indicative mood (Janhunen 1982). Present-day Uralic languages still featuring a negative auxiliary include Forest Enets, Tundra Nenets, Nganasan, Komi, Udmurt, Mari, Erzya, South Saami, Skolt Saami, Livonian, Finnish, Votic (Miestamo et al. 2015:17), and Pite Saami (Joshua Wilbur, p.c.). Hence, most of the contemporary Uralic languages exhibit the A/Fin/NegVerb type (according to the asymmetric negation typology in Miestamo 2005), manifesting some degree of finiteness of the negator. Symmetric negation, where only a negative marker is added to affirmative constructions, can be found in Selkup, Hungarian, Eastern Khanty, Mansi, Komi, Udmurt, and Erzya (Miestamo et al. 2015:17).

Most languages in the Finnic branch of the Uralic language family have retained at least some verbal features of the NegAux. For instance, it is conjugated for person in Finnish, Karelian, Veps, Lude, Votic, Ingrian, and Livonian (Laanest 1975:158–159), e.g. Finnish *en lue* ‘I don’t read’, *et lue* ‘you don’t read’, *ei lue* ‘he/she doesn’t read’, etc. The distinction between the present and past forms of the NegAux is still present in South Estonian (including Seto) and Livonian (Laanest 1975:158). Livonian also uses person suffixes in some of the past tense forms of NegAux (compare: *äb kiz* ‘I don’t ask’, *äd kiz* ‘you don’t ask’ in present tense, *iz kiz* ‘I didn’t ask’, *izt kiz* ‘you didn’t ask’ in past tense, Metslang et al. 2015:435–436), while all person suffixes on the NegAux have been lost in South Estonian and both *ei loe?* ‘don’t read’ in present tense and *es loe?* ‘didn’t read’ in past tense are used to refer to all grammatical persons. *Es*, the South Estonian past tense form of the NegAux, has retained the historical past tense marker **ś* (Künnap 2007). Verbal features on the NegAux have been completely lost only in North (Standard) Estonian, where, after losing its finiteness, the NegAux has become a negative particle without any verbal inflection (Tamm 2015:401).

Finnic languages have a special connegative form for the present tense variant of the negated main verb. In most of them it is merely the bare verb stem, e.g. North Estonian (*ma*) *ei loe*, Finnish *en lue* ‘I don’t read’. South Estonian, however, has also retained the historical suffix **k*, manifested as a glottal stop: (*ma?*) *ei loe?* ‘(I) don’t read’. In most Finnic languages the connegative form is the same for all persons, with Livonian being the exception as it uses person suffixes on plural forms of the main verb, compare (*mēg*) *sīe-mō* [(we) eat-1PL] ‘we eat’ with *äb sīe-mō* [NEG eat-1PL] ‘we don’t eat’, and (*tēg*) *sīe-tō* [(you) eat-2PL] ‘you eat’ with *ät sīe-tō* [NEG.2PL eat-2PL] ‘(you) don’t eat’ (Viitso 2003). A similar tendency can be found in the South Estonian Tartu dialect, where person suffixes are also used in negation, with the whole system being closer to symmetric negation (e.g. *me ess jõvva-me* [we NEG.PST reach-1PL] ‘we didn’t reach’). In Veps and Lude, the suffix

goi/ koi (Veps) or *guoi/ güöi* (Lude) occurs on the plural forms of verbs, e.g. *emai lugegoi* ‘we don’t read’ (Grünthal 2015: 202). Historically, this suffix is related to the third person imperative suffix **go/ko* (Viitso 2003). Veps has an additional suffix on the main verb, *ške* (singular) or *šket* (plural), which is assumed to be historically related to the NegAux in the past tense, incorporating the same past tense morpheme (*š*) as the negation marker *es* in South Estonian (Kährik 1978, Viitso 2003). This means that the historical postverbal negation auxiliary has become a suffix on the main verb in Veps.

For the past tense form of the negated main verb, all Finnic varieties except South Estonian and Livonian use the past participle. Finnish distinguishes between the singular and plural past participle, e.g. *e-n luke-nut* [NEG-1SG read-PST.PTCP.SG] ‘I didn’t read’, *e-mme luke-neet* [NEG-1PL read-PST.PTCP.PL] ‘we didn’t read’. North Estonian, on the other hand, only uses one past participle form: *ma/me ei luge-nud* [I/we NEG read-PST.PTCP] ‘I/we didn’t read’. South Estonian uses the connegative form of the main verb for negating both the present and the past, the distinction being expressed with the negation word instead: *ma? ei loe?* [I NEG read.CNG] ‘I don’t read’ (present tense), *ma? es loe?* [I NEG.PST read.CNG] ‘I didn’t read’ (past tense).

3.2.2 Prohibitives

In the majority of the world’s languages, including the Uralic languages, the negation strategy of imperatives differs from standard negation (see van der Auwera 2006, Miestamo 2007). The negative imperative (prohibitive) stem has been reconstructed as **eIV* for Proto-Uralic (Janhunen 1982). In Finnic languages the negation auxiliary is used, with the imperative marker attached to both the auxiliary and the main verb (Laanest 1975: 160), e.g. Finnish *älä lue*, Estonian *ära loe* (imperative Sg2, ‘don’t read!’) and Finnish *äl-kää luke-ko* or Estonian *är-ge luge-ge* (imperative Pl2, ‘don’t read!’).

According to the typology of prohibitives by van der Auwera et al. (2013), which is based on second person singular prohibitives, most of the Uralic languages, especially its Finnic branch, belong to Type2, where the negation marker of the second person singular imperative is different from that of standard negation, but the verb form does not differ from that of declaratives. However, the same does not hold for second person plural. Compare the Estonian formations of imperative, standard negation and prohibitive for the verb *lugema* ‘read’ in Table 3 (see also Tamm 2015 for more details).

In South Estonian one can find more options for forming prohibitives. In addition to the strategy using the preverbal negation verb *ära* (~ *ar(?)* ~ *är(?)*, see example 7), there is also the option to use the standard (postverbal) negation marker to form prohibitives (6). In both cases the main verb bears imperative

Table 3. The imperative, standard negation, and prohibitive forms of the verb *lugema* ‘read’ in Standard Estonian

Person	Imperative	Standard negation	Prohibitive
Sg2	<i>Loe!</i> ‘Read!’	<i>ei loe</i> ‘(You) don’t read’	<i>Ära loe!</i> ‘Don’t read!’
Pl2	<i>Luge-ge!</i> ‘Read!’	<i>ei loe</i> ‘(You) don’t read’	<i>Är-ge luge-ge!</i> ‘Don’t read!’

morphology. South Estonian thus also uses the Type3 strategy of the typology of van der Auwera et al. (2013), where

the second singular prohibitive employs a sentential negative that is used in declaratives or, if the notion of indicative is relevant, in indicative declaratives, but the verbal construction is not the one found in second singular imperatives.

3.2.3 Word order

Although Proto-Uralic has been reconstructed as having SOV basic word order (Janhunen 1982, Vilkuna 1998: 178), present-day Uralic languages show a lot of word order variation. In the Finnic branch the basic order has changed from SOV to SVO. According to Vilkuna (1998: 178), the Eastern SVO languages (Karelian, Veps) show extensive word order variation and spontaneously use SOV to focus the verb or the object. In the Western SVO languages (Estonian, Finnish), the OV order is restricted to specific constructions and tendencies towards verb-second word order can be found (for word order variation in Estonian see e.g. Tael 1990, Lindström 2005, 2017a; for Finnish, e.g. Huumo 1993, Lehtinen 1992, Vilkuna 1989).

As for negation word order, the negator generally precedes the main verb in Uralic languages (Vilkuna 1998: 211). Exceptional postverbal negation occurs in South Estonian (e.g. Seto and Eastern Võro), Lude (example 8), Veps, some Karelian dialects (Zaikov 2000) and the dialect of the island of Kihnu, which belongs to the Insular dialect group of North Estonian. In Kihnu postverbal negation can be found only in past tense and only with the negation morpheme *es* (example 9), which is attached as a suffix on the verb and does not get any stress in speech (differently from Seto where it can be sometimes stressed) (Lonn & Niit 2002: 41, Sang 1975). Some of the newer negation markers of the Insular dialect may also occur postverbally (e.g. *mitte* < **mitään* ‘anything’), in which case they often occur in the clause-final position or in combination with other negative markers or polarity items (Klaus 2009). A typical example of postverbal negation in combination with several polarity items (*enäm* ‘anymore’, *kiegi* ‘anybody’) and *mitte* (pronounced *mtte*) is presented in (9).

CENTRAL LUDE

- (8) *hän kestä-nü ei*
 he/she resist-PST.PTCP NEG
 'he/she didn't resist' (Kehayov et al. 2013: 80)

NORTH ESTONIAN INSULAR DIALECT, KIHNU ISLAND

- (9) *kui kevāde jälle paast tul-i siis lähā-ss enām jälle*
 when spring.GEN again feast come-PST.3SG then go-NEG.PST anymore again
kiegi=mtte
 anybody=NEG
 'When the feast started in spring nobody went there anymore.' (CED)

Veps has been described as having the exceptional Verb + NegAux order. However, postverbal negation is an infrequent pattern in Veps (Kettunen 1943, Kährik 1978). The same applies to Lude (Kehayov et al. 2013) and the Insular dialect of Kihnu. Thus, Seto is the only Finnic variety where postverbal negation is used systematically and where it is the most common position for the negation marker.

The origin of postverbal negation in Seto and other Finnic languages is unclear. It may be related to earlier basic SOV word order. However, typologically the correlations between basic word order and placement of negator seem to be more complex. Dryer's study on word order correlations (1992) found that both OV and VO languages tend to place negation markers before the verb and therefore the order of negator and verb are not correlated with the order of O and V. Dahl has also concluded that the order of the verb and the negator does not depend on basic word order in languages and "[...] there is a preverbal tendency which is fairly independent of the order between object and verb" (Dahl 2010: 23). On the other hand, it has been found that languages with the negative auxiliary show stronger correlations between the Verb + NegAux order and basic word order, meaning that OV languages with a negative auxiliary tend to have postverbal negation (Dahl 1979, Dryer 1992, Dahl 2010, Miestamo 2011). Hence, it is still possible that postverbal negation in Veps, Lude and Seto is related to the historical OV basic word order, which is accompanied by preserving a certain degree of finiteness of the inherited negative auxiliary. In Seto, this is manifested mainly in the distinction between present and past (*ei* and *es* accordingly); in Lude and Veps, conjugation for person has been retained.

4. Data and methods

In the next sections, we take a closer look at the data from East Seto, spoken in the Pskov region of present-day Russia. Our main aim is to compare the older

data from the 1970–1980s to the data collected during fieldwork in the 2010s, in order to discover possible changes in the speakers' preferences towards pre- and postverbal negation. We then compare it to Võro and North Seto data in order to outline the transition zone from an area that mostly uses preverbal negation to one that mostly uses postverbal negation. We also take a closer look at the variation in pre- and postverbal variation in East Seto and use multifactorial statistical modelling to detect the linguistic and extralinguistic factors affecting this variation. This enables us to provide a detailed account of the nature of the variation and determine whether the use of pre- and postverbal negation is more likely to be related to functional or sociolinguistic factors, with the latter potentially indicating (contact-induced) language change.

4.1 Data

The data used in this study mainly comes from spoken interviews recorded in the East Seto area in the Pechory district in Russia between 2010 and 2013. The interviews were conducted in an informal setting in the consultants' homes. We asked them about everyday life, both in the past and present, the schools, Seto traditions related to religion, etc. The interviews were later transcribed. All the occurrences of negation were extracted manually from the transcribed texts. In this study we use the data from eight speakers, all of whom can be characterised as older fluent speakers of Seto who have grown up in a Seto-speaking family and have been surrounded by a Seto-speaking community for most of their lives. Data were extracted from the recordings and consist of 5 hours and 44 minutes of spoken data.

A total of 1083 instances of negation occurred in our dataset. This sample is relatively small compared to what is common in corpus linguistics, but it provides sufficient material for sociolinguistic analysis and allows for statistical modelling of extra- and intralinguistic constraints regulating the position of the negation word, if the aim is restricted to explanation rather than prediction. However, as the speaker sample is quite limited even for a very small speaker community such as East Seto, the individual speakers in the sample inevitably rise to the focus of the analyses. It is therefore important to give a brief overview of the speakers' profiles (Table 4).

We compare the distribution of negation patterns in contemporary East Seto data to earlier data from the same region and data from neighbouring areas by using the materials in the Corpus of Estonian Dialects (CED). The CED data are of a similar nature, consisting of spoken unstructured interviews on various topics such as speakers' life stories, everyday life in the past and present, traditions and

Table 4. Profiles of the eight East Seto speakers in the sample

Speaker	Education	Family	Contacts with Estonia
SP1 Male B. 1938	7 years, local school	Wife is Russian, but she speaks Estonian. Relatives (sisters) live in Estonia. His children live in Russia. Communicates with them mostly in Russian.	Travelled often to Estonia during the Soviet era. Visits his sisters in Estonia several times a year. Watches Estonian TV and listens to Estonian radio.
SP2 Male B. 1950	High school in Estonian	Relatives in Estonia, wife is Russian.	Has lived in Estonia. Does not visit his relatives in Estonia. Listens to Estonian radio.
SP3 Female B. 1932	4 years, local school	Sister used to live in the same village. Both husbands were Setos. One son currently lives in Estonia, other two in Russia. Communicates with her children and their families in Seto.	Has not lived in Estonia, nor does she visit Estonia. Watches Estonian TV and listens to Estonian radio.
SP4 Female B. 1933	6 years, local school	Communicates with her daughter in Seto or Estonian, speaks Russian to her grandchild. Relatives live in Estonia.	Spends winters in Kaliningrad (Russia). Lived in Estonia for 8 years in the 1950s, often visits her relatives in Estonia. Sometimes listens to Estonian radio.
SP5 Female B. 1948	Local school, higher education in Estonia	Husband is Seto, no children.	Has studied in Estonia. Visits Estonia rarely. Sometimes listens to Estonian radio. Occasionally communicates in Estonian.
SP6 Female B. 1942	Local school, Pskov university (in Russian)	Mother and husband were Russians. Speaks Russian to her son's family.	Lived in Estonia for 1 year. Generally does not follow Estonian media.
SP7 Female B. 1921	4 years, local school	Husband was Seto, communicates with her children in Seto. Daughter lives in Estonia.	Has been spending winters in Estonia in her old age. Good command of Estonian. Follows Estonian media.
SP8 Female B. 1927	4 years, local school	Husband was Seto, from the same village. Son lives in Russia, communicates with him in Seto.	Grandson lives in Estonia.

customs, etc. These interviews were recorded between 1960 and 1993, but mostly in the 1970s. The data is introduced and analysed in more details in 5.1.1.

4.2 Variables

In order to find relevant factors (predictors) which could explain the variation in contemporary East Seto, we encoded the dataset consisting of 1083 observations of negation for a number of sociolinguistic, discourse-related, semantic, and morphosyntactic variables. The variables were mainly selected based on previously published research on Estonian negation and studies about other morphosyntactic alternations in Estonian dialects (see Klavan et al. 2015, Lindström & Uibo 2017). Hence, potential predictors include the time reference (manifested lexically by the negation word), structural persistence/priming (see e.g. Szmrecsanyi 2005, Tamminga 2016, Travis 2007, Torres Cacoullos & Travis 2019), grammatical person, the form of the subject (partly accounting for clause type in Estonian, since normal transitive and intransitive clauses take only nominative subjects, but existential, possessive, resultative, and experiential clauses can also take partitive subjects [Erelt & Metslang 2006, Huumo & Lindström 2014, Lindström 2017b]), and the semantic class of the verb. We also coded the ID and gender of speakers. Table 5 presents a summary of the encoded variables and their levels.

After excluding observations where it was not possible to determine the values for one or several variables, either due to the ambiguity of the form/context or the quality of the recordings, we were left with a final dataset consisting of 1033 observations. The variables included in the statistical models were required to demonstrate co-occurrence with both preverbal and postverbal negation because exclusive preference for only one of them would have been problematic for some of the methods used for the statistical analysis. For instance, grammatical mood was not used as a variable because the few imperative and conditional forms could only be found with postverbal negation (prohibitives were discussed in Sections 3.1 and 3.2.2). The speakers' gender was disregarded in the statistical analysis because it was highly correlated with the speaker. This is because the number of consultants in our sample is small and their gender is fixed. Hence, the two variables turned out to explain the same aspects of variation.

4.3 Methods

To explain the variance between preverbal and postverbal negation in East Seto, we used a well-established combination of classification methods for studying linguistic variation in sparse and unbalanced data such as ours (see Janda et al. 2010, Tagliamonte & Baayen 2012, Baayen et al. 2013, Klavan et al. 2015). We started our

Table 5. Coded variables

Variable	Description	Levels/values
POS*	Position of the negation word (the dependent/response variable)	<i>pre</i> = before the verb <i>post</i> = after the verb
Explanatory/independent variables conditioning the variation		
NEG_WORD	Negation word	<i>ei</i> = negation word in the present tense <i>es</i> = negation word in the past tense
PREV	Position of the negation word in the previous use of negation	<i>pre</i> = before the verb <i>post</i> = after the verb
PERSON	Person reference of the verb form	1 = first person 2 = second person 3 = third person <i>IMPS</i> = impersonal voice
SUBJ	Form of the grammatical subject	<i>nom</i> = nominative case <i>part</i> = partitive case <i>other</i> = infinitive construction or subordinate clause <i>no</i> = no explicit subject
VERB_TYPE	Semantic class of the main verb	<i>be</i> = the verb <i>olema</i> ‘be/have’ <i>cognition</i> = cognition/psych verbs (‘know’, ‘think’, ‘feel’ etc.) <i>modal</i> = modal and phase verbs (‘can’, ‘may’ etc.) <i>motion</i> = motion verbs (‘go’, ‘come’, etc.) <i>other</i> = all other verbs
GENDER	Gender of the speaker	<i>F</i> = female <i>M</i> = male
SPEAKER	Individual speakers (<i>N</i> =8)	SP1, SP2, SP3, SP4, SP5, SP6, SP7, SP8

* The cases with double negation were joined with the binary classes based on the position of the first negation word.

analysis by running conditional random forest models (Hothorn et al. 2006a) in order to determine the most influential variables. Random forest models are complex non-parametric models, based on growing a given number of (e.g. 2000) conditional inference trees (Hothorn et al. 2006b), which are essentially a group of classification algorithms with a tree-like structure. The method works by partitioning the observations (= individual cases of negation; rows in a data frame)

in the sample recursively into two distinct groups based on the explanatory variables which are most strongly associated with the response variable (position of the negation word). Partitioning continues until no further statistically significant splits can be made, i.e. there are no more explanatory variables whose levels significantly differ from each other in terms of evoking a preference for either preverbal or postverbal negation in that specific subgroup of observations. The resulting groups of observations each constitute a specific context of negation where either word order is more likely to occur (see examples in Figures 2b and 3b in Sections 5.2.1 and 5.2.2, respectively) and these contexts can be used for predicting the negation word order for any new observation with specific values for the relevant explanatory variables. Each actual observation in the sample data also gets a predicted negation pattern based on the context values. This prediction is then matched with the actual value and the model is classified as accurate if it is able to predict the values correctly. While single trees can be slightly unstable in their predictions and sensitive to smaller deviations in the sample data, random forests considering predictions from a large number of single trees ensure more stable and plausible results by introducing two sources of randomness. Firstly, each tree in the forest is grown on a random sample of the data drawn with replacements. This means that some observations (approx. 30%) are left out of each subsample, while other observations get represented multiple times. Therefore, each tree has a slightly different structure, since it is grown on a slightly different sample. However, the predictors that really are strongly associated with the response variable make it to most of the trees because their effect can be clearly observed even if some random changes happen in the data. Secondly, while single conditional inference trees always select the strongest predictors for the splits, random forest models enable us to test only a given number (usually the square root of the number of all predictors) of randomly selected variables in each node of the tree. This means that if the predictor usually most closely associated with the response variable does not make it to the selection, other important variables, whose influence may otherwise be overshadowed by the dominating variable, also get a chance to demonstrate their effect. This results in more reliable predictions for future observations. Random forests also enable us to rank the predictors according to their overall importance, i.e. their contribution to explaining the variation (Hothorn et al. 2006, Strobl et al. 2009a).

We complemented the analysis with a mixed-effects logistic regression model where the odds of the occurrence of one of the response levels (postverbal negation, in our case) is assessed based on certain configurations and combinations of predictors, i.e. different contexts of negation. One of the main differences between the ‘tree & forest’ models and standard regression models is that the former are built on the interplay between different predictor sets, while the latter make pre-

dictions over the entire sample space, assessing the global effect of one predictor (or combination of predictors, when interactions are included) at a time. Another distinction between the methods is that the ‘tree & forest’ models cannot handle the non-independence of data points very well. For example, it is well-known that speakers have their individual preferences or habitual patterns in choosing between alternative morphosyntactic constructions (e.g. Baayen et al. 2008, Baayen et al. 2013), causing observations from the same speaker to be correlated. Mixed-effects regression models allow introducing random effects, which enable us to assess the global impact of predictor variables while adjusting these assessments according to some potential random variation. The random effects of this study include the speakers who are well known to exhibit idiosyncratic behaviour in grammatical alternation studies. While these idiosyncrasies are usually not of particular interest to the researcher, they should still be taken into account (Pinheiro & Bates 2002, Baayen et al. 2008).

5. Results

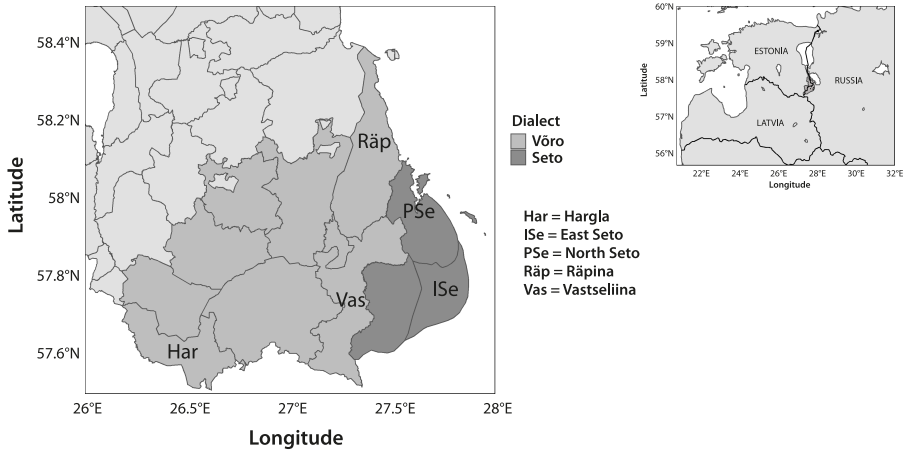
5.1 General findings and change over time

In this section, we examine the usage frequency of negation patterns (preverbal, postverbal, and double negation) in both temporal and spatial dimension by comparing available data in East Seto to its nearby varieties. In addition, we zoom in on ongoing changes in East Seto with the focus on changes in using double negation and prohibitives because despite providing some interesting insights, their relatively infrequent use prevents these phenomena from being included in the statistical analysis on variation in pre- and postverbal negation in Section 5.2.

5.1.1 Distribution of negation patterns

We first look at the general changes in negation word placement in East Seto and compare the data with its close varieties, i.e. North Seto, Rāpina and Vastseliina (the Eastern varieties of Võro, spoken close to the Seto area), and Hargla (the Western variety of Võro, spoken close to the Estonian-Latvian border). The two dialect regions along with the abbreviations of the relevant varieties are shown in Map 2. The data for comparison come from the CED and have been analysed in recent student theses: Rāpina in (Laan 2009), Hargla and North Seto in (Klaus 2009); the data from Vastseliina were analysed by ourselves.

The initial comparison involves the distribution of preverbal, postverbal and double negation in these five samples (Table 6 and Figure 1). As we can see from Figure 1, the differences between Seto and Võro varieties are obvious: postver-



Map 2. The Võro and Seto language areas

bal negation is the main pattern in Seto (both East and North Seto), while in the Võro varieties of Vastseliina, Räpina, and Hargla preverbal negation dominates. Vastseliina is the closest to Seto (with 26% of the instances of standard negation occurring postverbally). Räpina is spoken in an area bordering on North Seto, but preverbal negation dominates nevertheless. In the Hargla variety postverbal negation is almost never used. Thus, a clear distinction in the negation order emerges along the border between the traditional Võro and Seto areas, with Vastseliina constituting a transition zone.

Table 6. The use of preverbal, postverbal, and double negation in South Estonian varieties

Variety	Preverbal	Postverbal	Double	Total
East Seto (1970s)	91	228	26	345
East Seto (2010s)	260	806	17	1083
North Seto (1970s)	58	197	6	261
Räpina (1970s)	329	26	30	385
Hargla (1970s)	259	2	23	284
Vastseliina (1970s)	191	75	19	285

Interestingly, double negation is possible in all of the five areas, although it is not very frequent. This supports the claim in the dialectal overview in Keem & Käsi (2002: 53), which states that double negation has been used all over the Võro and Seto areas, but is vanishing.

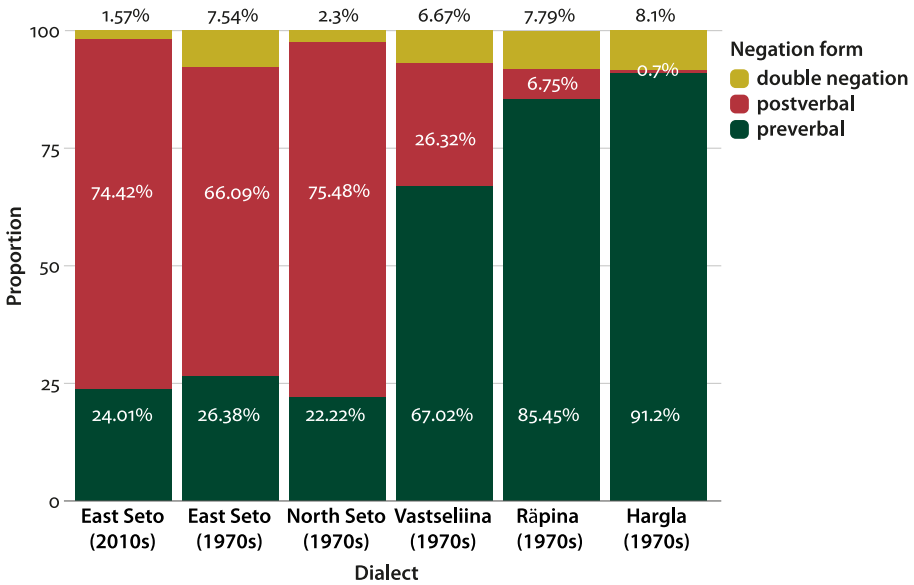


Figure 1. The use of preverbal, postverbal, and double negation in South Estonian varieties

When focusing on the variation in East Seto, we can see that there have been no substantial changes in the relative frequency of negation patterns during the past 40 years. The proportion of preverbal negation has remained more or less the same, while the use of postverbal negation has slightly increased, mostly at the expense of double negation. Overall, the use of postverbal negation has been surprisingly stable and at this point there does not appear to be much pressure on the East Seto negation system from the dominant contact languages. The situation would probably be different if we had comparable data from present-day Seto varieties spoken in Estonia (North and South Seto), where Estonian influence is probably much stronger and preverbal negation is likely to be more widespread than in East Seto spoken in Russia.

In addition, the East Seto negation system has been stable in regard to the symmetric/asymmetric negation typology. In his study about changes related to mood and modality in the language death situations of four Finnic varieties (including East Seto), Kehayov (2017: 161–169) found that during the process of language death, the tendency to use symmetric negation increases. His study, however, involved forgetters (see Kehayov 2017: 104–107) and we observed no such change in our data from older fluent speakers.

5.1.2 Double negation

Dahl (1979) and Payne (1985) have shown that double negation markers are generally placed on each side of the verb. As a counterexample, Afrikaans is often mentioned, with two identical negation markers following the main verb (see also Dahl 2010: 20). Seto double negation demonstrates similar options with more variation: the first negation word may occur preverbally, but it is more likely to be manifested postverbally; the second negation word, identical to the first negation word, occurs either after the verb or later in the clause. In older East Seto data (from the 1970–1980s), double negation occurs 26 times out of 345 negation instances (~7.5%), with the first negation word occurring preverbally 6 times (10) and postverbally 20 times (11, 12, 13).

EAST SETO

- (10) *lää=i täämbä külb-mä (...) eij=olō säänes-t ilma*
 go.CNG=NEG today sow-INF2 NEG=be.CNG such-PRT weather.PRT
ōij=ōt (...) täämbä midä her'n'-i-t
 NEG=that today some.PRT pea-PL-PRT
 'I'm not going to sow today, there is no such weather that [I could sow] peas.'
 (CED)
- (11) *olō=ōss inäp (.) mu `aigo sus's'i ess*
 be.CNG=NEG.PST anymore I.gen time.PRT wolf.PL.PRT NEG.PST
 'In my time there were no wolves anymore.'
 (CED)
- (12) *tiijä=ei=ma no-i-d joht õi?*
 know.CNG=NEG=I those-PL-PRT PTCL NEG
 'I don't know those at all.'
 (CED)
- (13) *ol'l tsurrō=ni `tütrekka (...) mahu=ōss `tarrō=ōss*
 be.PST.3SG boy.PL.PRT=and girl.PL.PRT fit.CNG=NEG.PST room.ILL=NEG.PST
 'There were [a lot of] boys and girls, [they] didn't fit into the room.'
 (CED)

Double negation is used for intensifying negation (Keem & Käsi 2002: 53) and/or delimiting the scope of negation (Lindström 1997). The intensifying function of double marking is evident in our earlier data as it is typically used in strong statements, as in examples (10–12). A typical context for double negation is that of existential clauses with partitive subjects, where the subject is focused and the second negation marker therefore also delimits the scope of negation, occurring after the focused part of the negated clause (10–11). Other parts of the sentence may also belong in the focus of negation, for example the object *noid* 'those', which is accompanied with another focus particle *joht* in (12), or the adverbial

tarrõ ‘into the room’ in (13). In the examples (10–13) it is evident that the same maker is doubled (either *ei* or *es*).

In newer data from 2010–2013, double negation is used remarkably less frequently and also less systematically. There were only 17 instances of double negation (~1.6% of all 1083 occurrences of negation). In 7 instances, the negation marker directly preceded and directly followed the main verb (14). This indicates that the use of double negation has become inconsistent and it is no longer used for delimiting the scope of negation. In addition, the use of negation markers was inconsistent as there was a mismatch of tenses/negation words in 5 instances (15), which was not found in earlier data. Thus we can conclude that the system of East Seto double negation is disappearing since its overall frequency is decreasing.

EAST SETO

(14) *tä-l om poig om nigu miis-t ma=ei tiijä=ei*
 she-ADE be.3SG son be.3SG PTCL man-PRT I=NEG know.CNG=NEG
 ‘She has a son, I don’t know her husband.’

(15) *S'aksa aigu õ=õ nuu? na=i lää=es Saksa sõtta*
 German time PTCL=PTCL PTCL they=NEG go.CNG=NEG.PST German war.ILL
 ‘During the German time, they didn’t go to the German war.’

5.1.3 Prohibitives

Initially we expected to see a lot of variation in East Seto, especially in prohibitives, because in forming them one can combine the two main prohibitive strategies, one with the preverbal negation verb *ärä* ~ *är(?)* ~ *ar(?)*, characteristic for Võro in general (Plado et al., to appear) and for Standard Estonian (*ära*), and the other one with the postverbal negation marker *ei*, characteristic for Seto and East Võro (Pajusalu, to appear, Lindström, to appear). Such combined constructions have been found in North Seto, spoken in Estonia (16).

NORTH SETO

(16) *arr sa tul-kku i?*
 NEG.2SG you come-IMP NEG
 ‘Don’t come!’ (CED)

We found only one occurrence of a prohibitive in the older East Seto data from the CED. Interestingly, it was preverbal, formed with the negation word *ärä* (7). In our East Seto data from 2010–2013, we found only 16 instances of prohibitives. They were all postverbal, formed with the negation word *ei*, as in (17). We therefore conclude that at least among the older speakers of East Seto, prohibitives do not vary.

EAST SETO

- (17) *put-ku ei säl midägi apparaattō*
 touch-IMP NEG there anything.PRT machine.PL.PRT
 ‘Don’t touch the machines there.’

It is difficult to claim anything about the past or ongoing changes on this basis since the data set is too small for substantial conclusions. At least in our newer data the prohibitives are formed consistently with postverbal constructions. Kehayov (2017) in his study on language death (which also included East Seto) found an example of a preverbal prohibitive formed with the negation word *ei*; this constitutes an example of mixing the two basic constructions (18). Hence, the system might not be as stable as it appears from our data.

EAST SETO

- (18) *ei joos-ku nii*
 NEG run-IMP SO
 ‘Don’t run like this!’ (Kehayov 2017: 258)

5.2 Factors conditioning the variation between pre- and postverbal negation

In the next few subsections we investigate the variation between pre- and postverbal negation in East Seto by focusing on the factors triggering any preference. Our main aim is to discover whether it is language-internal (e.g. morphosyntactic and semantic) or language-external (sociolinguistic and discourse-related) factors which mainly drive this variation. Language-internal factors would tell us more about the contrasts in the functional settings of pre- and postverbal negation while language-external factors would point towards the processes of contact-induced language change in general.

5.2.1 Implications for the importance of the individual speaker

We started the analysis by constructing a conditional inference forest with the individual speaker included as an explanatory variable, in addition to other aforementioned variables. In order to avoid overestimating the importance of correlated variables,⁵ conditional importances were calculated (see Strobl et al. 2009b).

5. The formula used was $POS \sim NEG_WORD + PREV + PERSON + SUBJ + VERB_TYPE + SPEAKER$. The number of random variables tried for each split was set to the ceiling of the square root taken from the number of explanatory variables in the model, as recommended in Breiman (2001). As our model contained 6 explanatory variables, the number of variables tested in each split was 3.

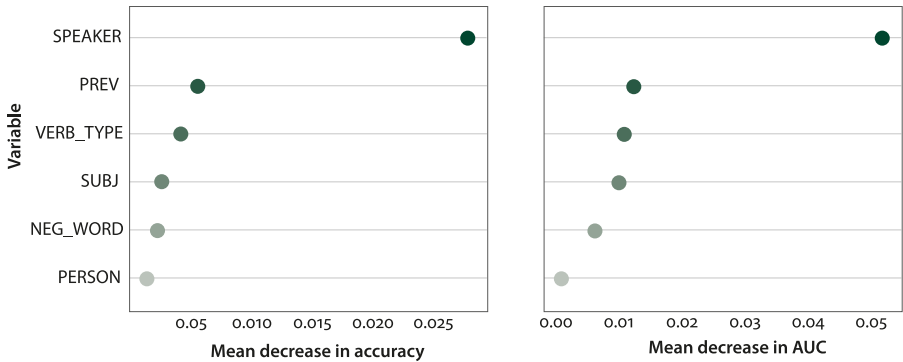
We arrived at a model which suggested that the most important variable determining the choice between preverbal and postverbal negation was indeed the individual speaker, as shown in Figure 2.

The ranking in the left graph in Figure 2a is based on the mean decrease in model classification accuracy⁶ when the values of each specific variable are randomly permuted. The ranking in the right graph is based on the mean decrease in what is called an area under the ROC curve (AUC⁷ for short), which shows the model's capability to distinguish between negation word position and is better suited for unbalanced response variables. For more details on the calculation and difference between the two see Janitza et al. 2013.

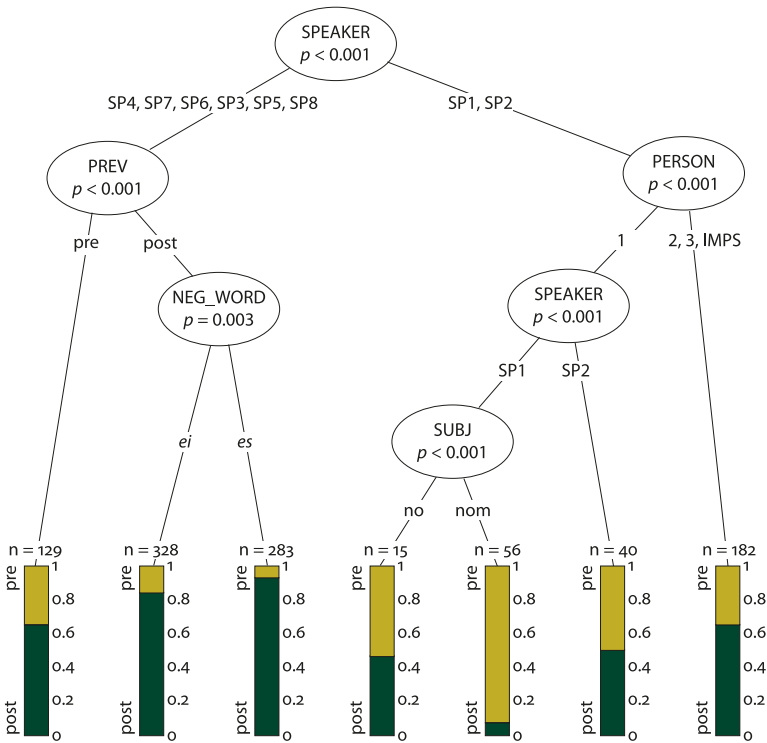
Figure 2a tells us that individual speakers and their preferences help to explain by far most of the variation in our data, while the person reference of the verb form fails to contribute much to the fit of this model when averaged over 2000 individual trees. In other words, given information about the individual speaker, we could quite easily disregard all information about grammatical person and still have a statistical model that can fairly well predict contexts where either pre- or postverbal negation is used. The idiosyncrasies of speakers are striking and they are probably related to their individual background, i.e. factors such as exposure to Estonian and Russian, level of education, the nature of employment held throughout their life, etc. (see Table 4 in Section 4.1). When looking at a single conditional inference tree (Figure 2b) with the same parameters as the conditional random forest model, it is evident that there is a gender distinction in the sample, with the two male speakers (SP1 & SP2) clearly exhibiting a higher likelihood of using preverbal negation than the female speakers. This may be due to the fact that both male speakers in our sample have had strong contacts with Estonia and Estonians during the Soviet era, and now have closer contacts with Russians. This means that they are overall less Seto-oriented than the female speakers in our sample. In addition, different factors seem to condition the use of pre- and postverbal negation for male and female speakers: PERSON for male speakers and a more complex interaction between PREV, NEG_WORD and VERB_TYPE for female speakers. However, this strong gender effect for the whole population of East Seto speakers seems unlikely and can rather be considered an artefact of sample size and gender bias.

6. Classification accuracy is the proportion of actual observations in the sample data that have the same class (either “pre” or “post”) as what the model predicts for those particular observations.

7. AUC can be interpreted as the proportion of cases where a randomly chosen actual observation of postverbal negation in the sample data gets a higher predicted probability for “post” than a randomly chosen actual observation of preverbal negation.



a. Permutation predictor importance



b. Conditional inference tree

Figure 2. Relative variable importances in a conditional random forest model (panel a) and a single conditional inference tree (panel b). The darker shade on the upper graph indicates high importance ranking, the lighter shade low importance ranking

The classification accuracy of this model was 0.84 and the AUC was 0.88. This means that this model has very strong predictive power. However, the model would not be generalizable to new data, meaning that we can use it to predict how our eight speakers would use negation in new utterances, but not how the negation system in East Seto is used in general. In order to assess which of the other variables strongly associate with the position of the negation word, we removed the speaker factor from the model.

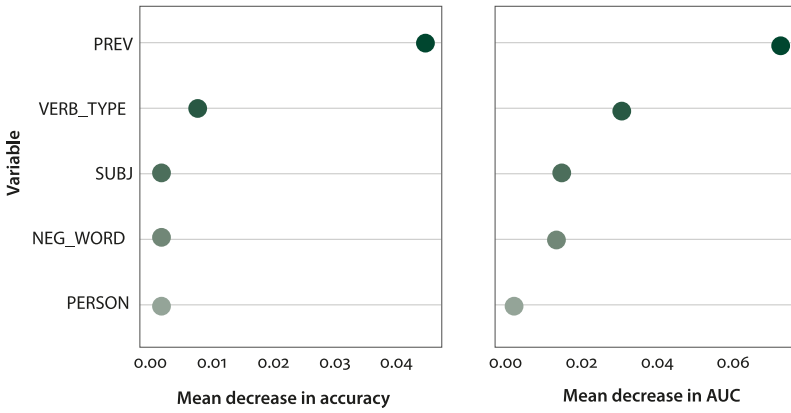
5.2.2 Effects of other variables

Excluding the speaker from the model results in a drop in the model's predictive accuracy (classification accuracy 0.79, AUC 0.81). However, it now establishes PREV (the negator's position in the previous use of negation) and VERB_TYPE (the semantic class of the main verb) as the strongest predictors (Figure 3a).

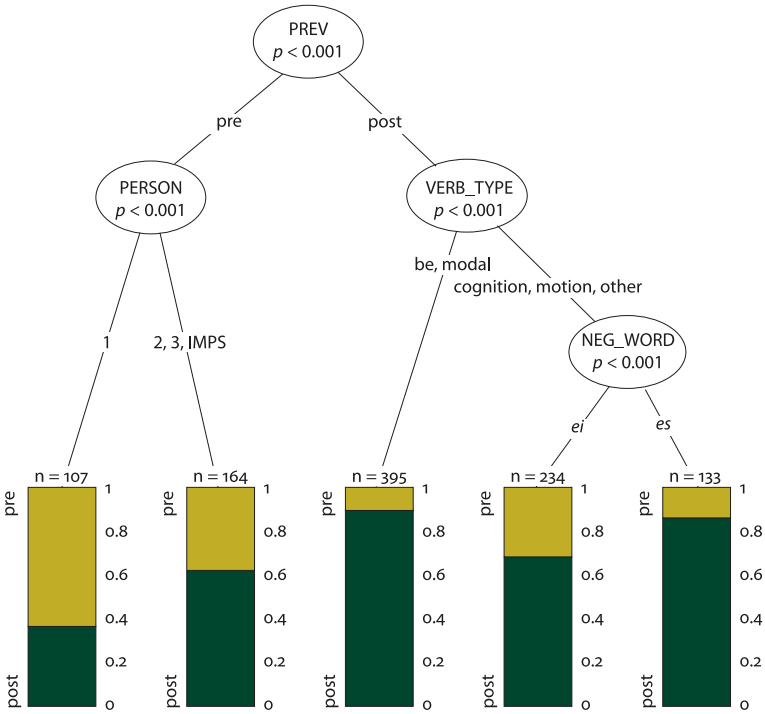
PREV and VERB_TYPE were influential in previous models as well, but were overshadowed by individual speaker preferences. This means that structural persistence is the best determiner of the outcome, suggesting that a recently experienced syntactic pattern strongly influences the speaker's subsequent language performance and that it is easier for the speakers to repeat structures that are similar to those previously produced/heard. The semantic class of the verb is a somewhat weaker predictor, while the tense distinction in the negation word and the other two morphosyntactic variables SUBJ (the form of the subject) and PERSON (the grammatical person of the verb form) probably only participate in interactions with other predictors in most of the single trees, i.e. affect the choice of negation word order only in some specific contexts already determined by the two stronger predictors. Since PERSON has been assigned the lowest conditional importance in both models, despite showing up in single conditional trees (e.g. 2b and 3b), it can be assumed that it is correlated with or conditioned by other predictors in the model.

We can assess the direction and strength of the relationship between the predictors and the position of the negation marker by constructing partial dependence plots (Figure 4), demonstrating how the probability of the negation word position partially depends on one or more predictors, while at the same time averaging the effect of other predictors in the model.

When examining the individual predictor effects in Figure 4, we observe that the predicted probability of the occurrence of postverbal negation (depicted by the green line) is always higher and never falls below 0.5, no matter which conditioning variable we are looking at. However, some conditions increase the likelihood of the occurrence of the otherwise less frequent preverbal negation (the yellow line). As noted earlier, the most significant of them is structural persistence (PREV), meaning that preverbal negation is significantly more likely to occur



a. Permutation predictor importance



b. Conditional inference tree

Figure 3. Relative variable importances in a conditional random forest model (a) and a single conditional inference tree (b)

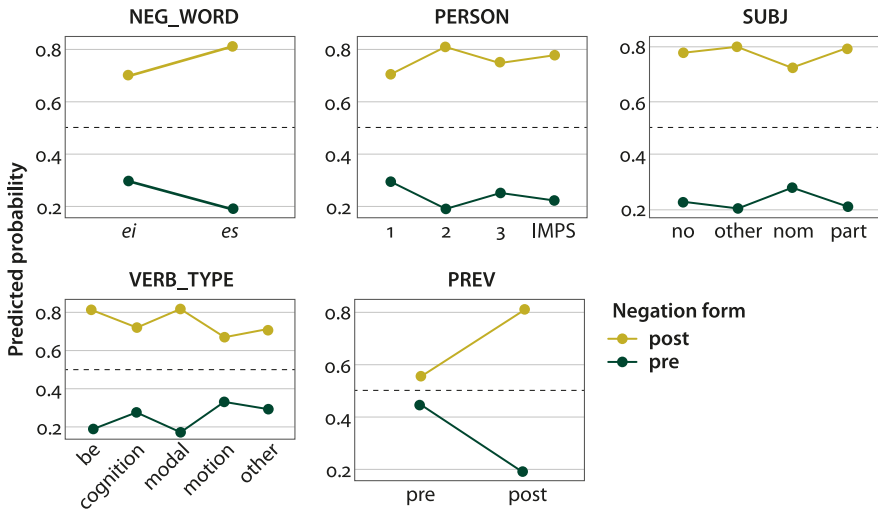


Figure 4. Partial dependence plots for individual predictor variables

when it has also been used in the previous instance of negation. Other factors that are more likely to evoke preverbal negation include the negation word *ei*, which is only used in the present tense, 1st person verb forms (both singular and plural), nominative subjects, cognition and motion verbs (and also the undefined verb class ‘other’). Hence, the most likely context for preverbal negation with all the individual effects occurring simultaneously is exemplified in (19), where the motion verb referring to the first person is in the present tense, and occurs with the negator *ei* and a subject in the nominative case. In addition, the most recently activated negation pattern also included a preverbal marker.

EAST SETO

- (19) *minä nende tüdrukki-dō-ga ei lää inäp*
 I they.PL.GEN girl-PL-COM NEG go.CNG anymore
 ‘I won’t go with those girls anymore.’

However, for individual observations, the model predicts that preverbal negation is even more likely to occur if a cognition verb (e.g. *teadma* ‘know’) was used in a similar context (20). Note that in this example the negated clause forms a fixed expression, used as a parenthetical or even a particle (*ma=i tiijä* ‘I don’t know’). Such expressions are often grammaticalized in languages and are used to show the speaker’s epistemic stance or hedge in interaction (see e.g. Kärkkäinen 2003, Weatherall 2011 for English). The same has happened in Estonian and Russian (Keevallik 2006, 2011), meaning that preverbal negation in Seto can be a result of

copying this discourse unit into Seto, with the same word order as it occurs in Estonian and Russian.

EAST SETO

(20) *ja siss ma=i tiijä mänegi s=naisterahvass*
 and then I=NEG know.CNG some PTCL=woman
 ‘... and I don’t know, some woman...’

Interactional effects are relevant as well (Figure 5). In such cases, the effect of one explanatory variable on the predicted negation form is conditioned by another explanatory variable.

In Figure 5 we are interested in the plots that depict either the green or the yellow line differently in two or more facets of the plot. These are the cases where the effect of one predictor on choosing between pre- or postverbal negation is conditioned by the values of another predictor. Such plots include ‘VERB_TYPE by NEG_WORD’, ‘PERSON by PREV’ and ‘SUBJ by PREV’. Two of these three plots illustrate interactions that can also be seen in the single conditional inference tree in Figure 3b. The effect of VERB_TYPE on the probability of using preverbal negation is much stronger with the negation word *ei* (present tense) for cognition verbs, while motion verbs and the undefined class ‘other’ are more likely than others to occur with preverbal negation in both tenses. The tendency of cognition verbs co-occurring with preverbal negation in the present tense might result from the same use of borrowed structures for fixed expressions as discussed earlier and illustrated in (20). Expressions such as *ei tiijä* ‘(I) don’t know’, *ei mälehtë* ‘(I) don’t remember’, are frequent in spoken language and they are therefore very susceptible to the effects of language contact and likely to be subject to variation.

Other interaction effects are somewhat more difficult to explain. For instance, the effect of PERSON is conditioned by PREV,⁸ meaning that while the use of preverbal negation in the previous instance of negation alone predicts preverbal negation to be reused, the likelihood of this happening is even higher when the speakers are referring to the first person, i.e. probably talking about themselves and their own experiences. Again, this can be linked to the effect of frequency in grammatical organization: speakers are frequently referring to themselves during the interviews, likely causing them to choose the contact-induced preverbal variant for negation. A similar effect seems to apply to the third person, also frequent

8. Note that in Figure 5 both the response variable POS as well as the predictor PREV have levels *pre* and *post*. In such cases, the lines depict the variable POS (negation word position in the analysed instance) while the axis text shows the variable PREV (negation word position in the previous instance of negation).

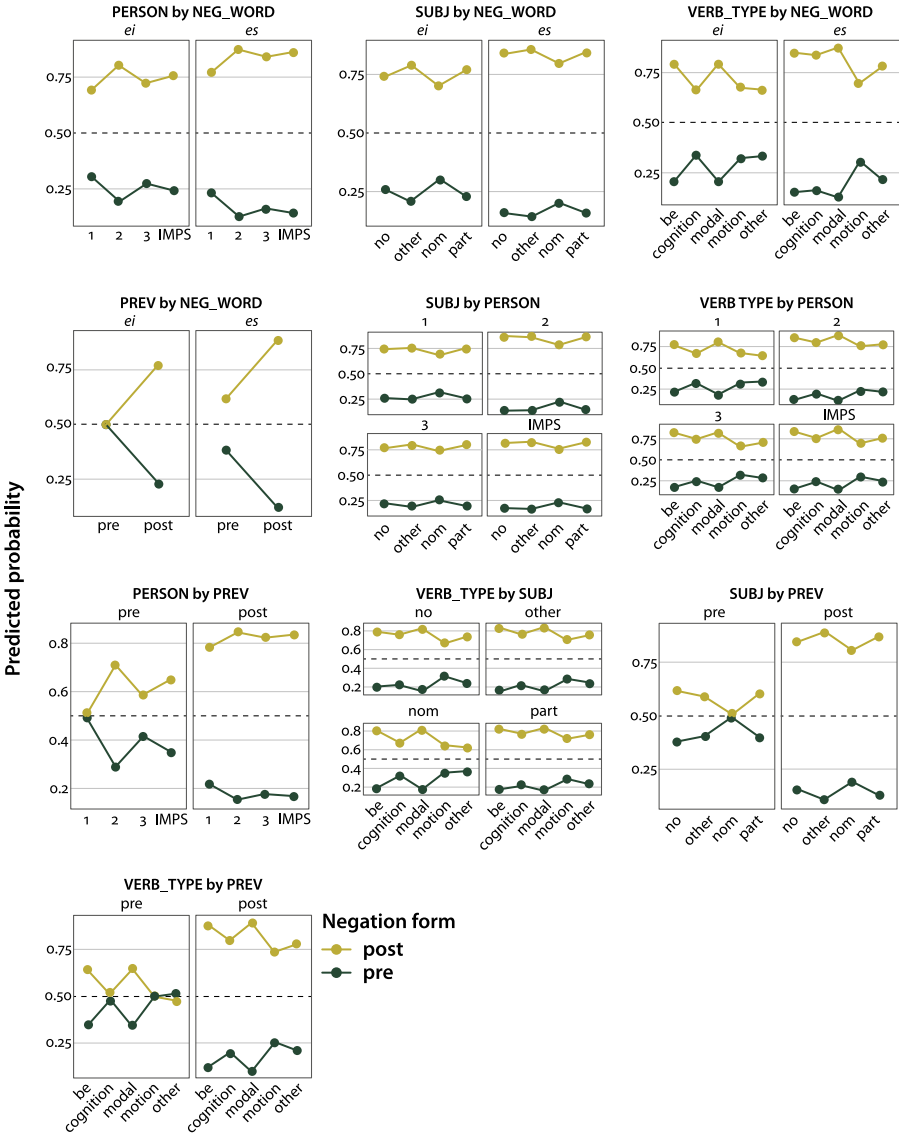


Figure 5. Partial dependence plots for interaction effects

in discourse. However, if the previous instance of negation included a postverbal marker, PERSON appears to have no such effect on negation word order.

PREV also seems to condition the effect of the subject: while nominative subjects are more likely to occur with preverbal negation than other subject forms no matter the previous negation form (this can also be associated with the use of fixed expressions), then for subjectless sentences, the relative likelihood for

the occurrence of preverbal negation increases slightly if the previous negation marker was postverbal. This deviates from the general structural persistence pattern, according to which speakers tend to repeat already produced or heard constructions. It is also possible that it is only a chance result based on this small data set, and not in fact significant.

While all the analyses above seem to point towards structural persistence and frequent expressions in communication playing an important role in explaining the negation word order, it still remains unclear how to treat the individual speaker. If so much of the variation can be attributed to individual preferences, it would be wrong to dismiss this effect completely. However, we would also like to be able to generalize the model results to the whole population. This is where mixed-effects regression modelling proves to be a more efficient tool than tree-based classification models. It allows us to assess both the individual and joint effects of the predictors, while also accounting for the idiosyncrasies of individual speakers.

The effects from logistic regression analysis (Figure 6) that are adjusted for the preferences of individual speakers are similar to the partial dependence effects in a random forest analysis. The only significant interaction effect was that of ‘VERB_TYPE by NEG_WORD’, while the interactions ‘PERSON by PREV’ and ‘SUBJECT by PREV’ were considered by the model as something that could have happened simply by chance. However, the individual effects of PERSON, PREV, and SUBJ were all identified as significant. In short: postverbal negation is considerably more likely in all situations in East Seto. However, speakers do have the tendency to reuse the negation pattern they have previously produced or heard. Verb forms referring to the first and third persons slightly raise the probability of using preverbal negation, as does the use of nominative case on the subject. Cognition verbs are far more likely to be used with preverbal negation in the present tense (with *ei*) than in the past tense (with *es*). This is in full accordance with the results from the ‘tree & forest’ analysis and the effects can be explained in the same manner as before.

Regression models also enable us to assess how much of the variation can be attributed to the random effect of the individual speaker, given the information about all other predictors. In this model, the speaker explains around 12% of the word order variation in East Seto negation forms, suggesting that a lot can still be attributed to the effects of other variables and their combinations. This is where regression modelling proves more beneficial than the ‘tree & forest’ models with the SPEAKER included as a fixed predictor. There are certainly differences that can indeed be attributed to individual preferences, but the ‘tree & forest’ models overestimate this effect and cannot distinguish it very well from the (socio)linguistic principles guiding the speakers more generally. Figure 7 shows the global individual preferences for the speakers.

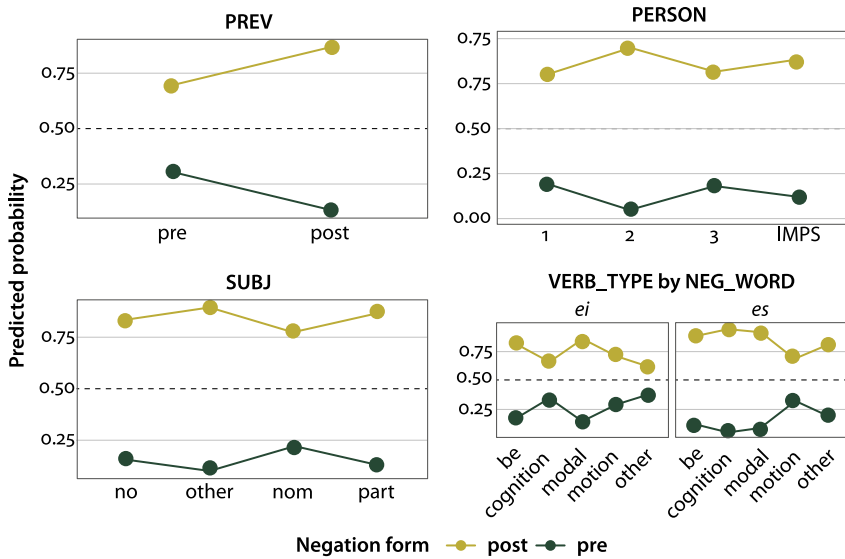


Figure 6. Predictor effects from mixed-effects logistic regression model

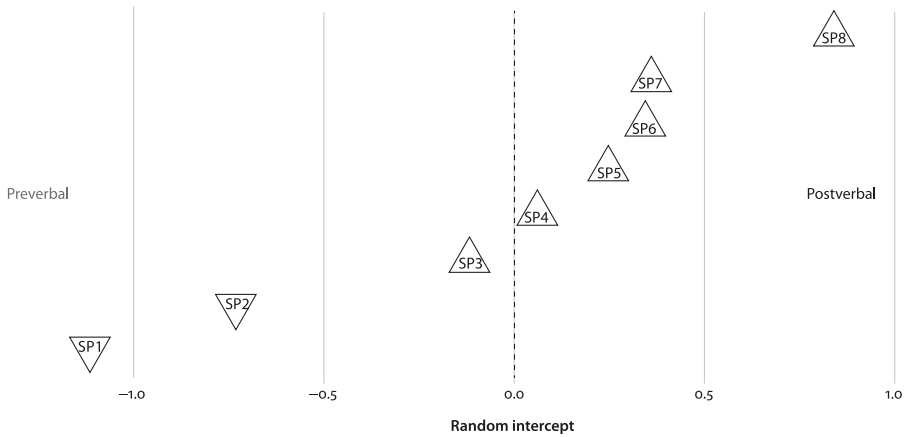


Figure 7. Random intercepts for individual speakers (male speakers marked with ▽, female with △)

For the speakers positioned on the right side of the graph, the model predictions are adjusted towards the likelihood of using (even) more postverbal negation, while for the speakers positioned on the left side, the predictions are adjusted towards using more preverbal negation. It can be seen that the two male speakers in the sample (marked with upside down triangles) are far more likely to use preverbal negation in any situation than the female speakers (marked with regu-

lar triangles). The two male speakers have had strong and long-standing contacts with Estonia, are married to Russians and do not use much Seto in their everyday communication (see Table 4 in Section 4.1). Therefore, it can be expected that their use of East Seto is subject to more influences from Estonian and Russian.

5.2.3 Summary and discussion of modelling results

The effect of individual speaker word order preferences is hard to overlook in such a small sample. The negation system of some East Seto speakers has levelled considerably under the influence of standard Estonian and possibly Russian, making the effect of other variables harder to detect. The two male speakers in the sample are more likely to use preverbal negation than the female speakers. Nevertheless, this implied gender effect is probably an artefact of sample size because it is highly unlikely that negation word order actually depends on gender. There is also no reason to think that the negation systems of male and female speakers are determined by different internal and external factors, as suggested by some of the models. In fact, when we tested the same models on the older East Seto data collected in the 1970s, gender exhibited an opposite effect with female speakers being more likely to use preverbal negation than male speakers. While keeping in mind that it is difficult to assign predictive power to models including fixed speaker effects, the significant differences between the speakers cannot be disregarded either. They can be attributed to the speakers' different exposure to contact languages, habits of reading and media consumption, level of education, and other social factors.

While postverbal negation in East Seto does seem to be the default pattern, certain factors significantly increase the likelihood of preverbal negation, a pattern common in Estonian and Russian, the two main Seto contact languages. The strongest effect was demonstrated for the position of the negation marker most recently activated in discourse. This is linked to structural persistence (Szmrecsanyi 2005) or structural priming, which “is the repetition of the same syntactic structure across clauses without pragmatic motivations” (Torres Cacoullos & Travis 2019: 674). As persistence effects occur across speakers and speech turns, they can be considered mediators of contact-induced change in a language system. The strong persistence patterns in the East Seto negation system indicate its high susceptibility to the effect of contact. Structural persistence is more related to short-term memory, to the specific interaction situation as well as to a series of such situations, rather than to any grammatical function, i.e. it can be considered extragrammatical. In this respect, the strong effect of the previously used/heard negation form on the negation system in East Seto is in accordance with Dryer's (2013) observations about what conditions variation in negation systems involving the repetition of the same morphemes. The two negative mor-

phemes that are used in Seto – *ei* (in present tense) and *es* (in past tense) – can both be used pre- and postverbally with the strongest conditioning factor being the construction used in the previous instance of negation. Furthermore, while the two markers themselves do differ from each other in terms of tense and the present tense form *ei* appeared to be a more likely candidate for preverbal position, this effect is probabilistic, not categorical, and also changes according to other contextual factors. While the tense distinction does not play an important role for motion verbs, which are always more likely than average to occur with preverbal negation, it is highly relevant to cognition verbs, which are far more likely to be used with a preverbal negation word in the present tense than in the past. This effect could be attributed to frequent fixed expressions such as *ma=i tiijä* ‘I don’t know’ or *ma ei taha* ‘I don’t want to’, which are often treated as a single syntactic unit like a particle in interactions, their preverbal negation being a result of copying the unit with Estonian or Russian word order – another instance of interpersonal structural persistence.

Regarding other factors affecting the variation of negation patterns in East Seto, the switch from the default postverbal to preverbal negation often happens in the case of first and third person reference, as well as with nominative subjects, all of which are the most common representatives of their functional categories. We therefore conclude that even the functional features conditioning the negation system are closely associated with repetition in discourse or in the language as a whole.

6. Conclusions

In this paper we investigate the variation in the syntactic position of the negation marker in South Estonian Seto, focusing on East Seto, which is spoken in present-day Russia. Seto is exceptional among Uralic languages due to its systematic use of postverbal negation. However, there is considerable variation in its negation marker position. Postverbal position is the most common variant, with the comparison of data from the 1970–1980s and 2010s revealing no change in this regard. This in itself is surprising because East Seto is a highly endangered variety and both of its main contact languages, Estonian and Russian, use preverbal negation. It may be due to our sample being made up of utterances from older fluent speakers whose relations with Estonia and Estonian are not very frequent (compared to those of the speakers of Seto varieties spoken in Estonia). The word order in prohibitives (negative imperatives) was also surprisingly consistent: in prohibitives, only postverbal negation was used in our data.

Although Seto is linguistically close to South Estonian Võro, we found substantial differences between them. Võro mainly uses preverbal negation. However, while the use of double negation previously occurred relatively frequently in both varieties, the form has decreased significantly in contemporary East Seto. In our data the double negation markers are used inconsistently and with limited functionality, inferring a general decline of the pattern.

Statistical analyses of factors conditioning the choice between pre- or postverbal negation in East Seto revealed that while postverbal negation is the default preference (used in about 74% of the cases), there are conditions which significantly increase the likelihood of the speaker using preverbal negation. The results suggest that the choice between the negation patterns in East Seto is first and foremost conditioned by such probabilistic factors as the frequency and repetition of specific structural forms (persistence), while functional or grammatical factors are less important. This implies that the choice between preverbal and postverbal negation depends on the speakers' individual linguistic background and the intensity of contact with Estonian and Russian. On a broader level the nature of the variation is linked to language change. Structural persistence is common in spoken discourse and it is an important factor in variation studies (Szmrecsanyi 2005, Travis 2007, Tamminga 2016, Torres Cacoullos & Travis 2019), which is why its role here is not surprising. The copying of fixed expressions with preverbal negation ('I don't know') shows that these frequent discourse items are easily transferred in language contact situations, that they are probably among the first constructions to include new syntactic forms and that they therefore could constitute one of the driving forces behind linguistic innovation.

Funding

This research has been supported by the Centre of Excellence in Estonian Studies (European Union, European Regional Development Fund) and is related to the research project "Interdisciplinary Corpus of Seto" (Estonian Ministry of Education and Research, EKKD41).

Abbreviations

1,2,3	person
ADE	adessive
CED	Corpus of Estonian Dialects
COM	comitative
CNG	connegative
GEN	genitive
ILL	illative

IMP imperative
INF infinitive
INF2 second infinitive (supine)
NEG negation
PL plural
PRT partitive
PST simple past
PTCL particle
PTCP participle
SG singular

References

- Baayen, R. Harald, Anna Endresen, Laura A. Janda, Anastasia Makarova & Tore Nettet. 2013. Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics* 37. 253–291. <https://doi.org/10.1007/s11185-013-9118-6>
- Baayen, R. Harald, Doug Davidson & Douglas Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59. 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Breiman, Leo. 2001. Random Forests. *Machine Learning* 45. 5–32. <https://doi.org/10.1023/A:1010933404324>
- Dahl, Östen. 1979. Typology of sentence negation. *Linguistics* 17. 79–106. <https://doi.org/10.1515/ling.1979.17.1-2.79>
- Dahl, Östen. 2010. Typology of negation. In Laurence R. Horn (ed.), *The expression of negation*, 9–38. Berlin: De Gruyter Mouton.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138. <https://doi.org/10.1353/lan.1992.0028>
- Dryer, Matthew S. 2013. Order of negative morpheme and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <http://wals.info/chapter/143> (last access 29 June 2021).
- Eberhard, David M., Gary F. Simons, & Charles D. Fennig (eds.). 2020. *Ethnologue: Languages of the world*. 23rd edn. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com> (last access 29 June 2021).
- Erelt, Mati & Helle Metslang. 2006. Estonian clause patterns – from Finno-Ugric to Standard Average European. *Linguistica Uralica* XLII(4). 254–266.
- Grünthal, Riho. 2015. *Vepsän kieliooppi* [The grammar of Veps] (Apuneuvoja suomalais-ugrilaisten kielten opintoja varten 17). Helsinki: Suomalais-Ugrilainen Seura.
- Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2020. Glottolog 4.2.1. Jena: Max Planck Institute for the Science of Human History. Available at: <https://glottolog.org/> (last access 29 June 2021).
- Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006a. Unbiased recursive partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674. <https://doi.org/10.1198/106186006X133933>

- Hothorn, Torsten, Peter Buehlmann, Sandrine Dudoit, Annette Molinaro & Mark Van Der Laan. 2006b. Survival Ensembles. *Biostatistics* 7(3). 355–373. <https://doi.org/10.1093/biostatistics/kxj011>
- Huumo, Tuomas. 1993. Suomen ja viron kontrastiivista sanajärjestysvertailuja [‘Contrastive comparison of Estonian and Finnish word order’]. Valma Yli-Vakkuri (ed.), *Studia comparativa linguarum orbis Maris Baltici 1. Tutkimuksia syntaksin ja pragmasyntaksin alalta* (Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 43), 97–158. Turku: Turun yliopisto.
- Huumo, Tuomas & Liina Lindström. 2014. Partitives across constructions: on the range of uses of the Finnish and Estonian “partitive subjects”. In Silvia Luraghi & Tuomas Huumo (eds.), *Partitive cases and related categories* (Empirical Approaches to Language Typology 54), 153–176. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110346060.153>
- Jääts, Indrek. 1998. Setude etnilise identiteedi ajalugu [‘History of the Ethnic Identity of Setos’]. *Akadeemia* 6–7. 1127–1152, 1520–1546.
- Jääts, Indrek. 2000. Ethnic identity of the Setus and the Estonian–Russian border dispute. *Nationalities Papers* 28(4). 651–670. <https://doi.org/10.1080/00905990020009665>
- Janda, Laura A., Tore Nessel & R. Harald Baayen. 2010. Capturing correlational structure in Russian paradigms: A case study in logistic mixed-effects modeling. *Corpus Linguistics and Linguistic Theory* 6(1). 29–48. <https://doi.org/10.1515/cllt.2010.002>
- Janhunen, Juha. 1982. On the structure of Proto-Uralic. *Finnisch-ugrische Forschungen* 44. 23–42.
- Janitza, Silke, Carolin Strobl & Anne-Laure Boulesteix. 2013. An AUC-based Permutation Variable Importance Measure for Random Forests. *BMC Bioinformatics* 14. 119. <https://doi.org/10.1186/1471-2105-14-119>
- Juhkason, Grethe, Andreas Kalkun, Liina Lindström & Helen Plado. 2012. Petserimaa setodest ja nende keelest 2010.–2011. aasta välitööde põhjal [‘On Pechory district Setos and their language based on fieldwork conducted in 2010 & 2011’]. In Jivä Sullöv (ed.), *Õdagumeresoomõ piiriq* (Võro Instituudi toimõndusõq 26), 11–29. Võro: Võro Instituut.
- Kährlik, Aime. 1978. Das Problem der Herkunft des verneinenden Präteritums im Wepsischen [‘The problem of the origin of the negating past tense in Veps’]. *Sovetskoje Finno-ugrovedenje / Soviet Finno-Ugric Studies* 14(3). 161–169.
- Kallio, Petri. 2012. The prehistoric Germanic loanword strata in Finnic. In Riho Grünthal & Petri Kallio (eds.), *A linguistic map of prehistoric Northern Europe*, 225–238. Helsinki: SKS.
- Kallio, Petri. 2014. The diversification of Proto-Finnic. In Joonas Ahola & Clive Tolley (eds.), *Fibula, fabula, fact. The Viking age in Finland*, 155–168. Helsinki: SKS.
- Kärkkäinen, Elise. 2003. *Epistemic stance in English conversation: A description of its interactional functions, with a focus on ‘I think’*. Amsterdam: John Benjamins. <https://doi.org/10.1075/pbns.115>
- Kask, Arnold. 1984. *Eesti murded ja kirjakeel* [Estonian dialects and the standard language]. Tallinn: Valgus.
- Keem, Hella & Inge Käsi. 2002. *Võru murde tekstid. Eesti murded VI* [Texts from Võro dialect]. Tallinn: Eesti Keele Instituut.
- Keevallik, Leelo. 2006. From discourse pattern to epistemic marker: Estonian (*ei tea* ‘don’t know’). *Nordic Journal of Linguistics* 29(2). 173–200. <https://doi.org/10.1017/S0332586506001570>

- Keevallik, Leelo. 2011. The terms of not knowing. In: Tanya Stivers, Lorenza Mondada & Jakob Steensig (eds). *The morality of knowledge in conversation*, 184–206. Cambridge: University Press. <https://doi.org/10.1017/CBO9780511921674.009>
- Kehayov, Petar. 2017. *The fate of mood and modality in language death: Evidence from Minor Finnic*. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110524086>
- Kehayov, Petar, Eva Saar, Miina Norvik & Andres Karjus. 2013. Hääbuva kesklüüdi murde jälgedel suvel 2012 [‘In the footsteps of vanishing Central Lude in summer 2012’]. *Emakeele Seltsi aastaraamat* 58 (2012). 58–101.
- Kettunen, Lauri. 1943. *Vepsän murteiden lauseopillinen tutkimus* [Syntactic investigation of Veps dialects] (Suomalais-ugrilaisen Seuran Toimituksia 86). Helsinki: Suomalais-Ugrilainen Seura.
- Klaus, Anneliis. 2009. Eitus eesti murretes [Negation in Estonian dialects]. Tartu: University of Tartu MA thesis.
- Klavan, Jane, Maarja-Liisa Pilvik & Kristel Uiboaed. 2015. The use of multivariate statistical classification models for predicting constructional choice in spoken, non-standard Varieties of Estonian. *SKY Journal of Linguistics* 28. 187–224.
- Künnap, Ago. 2007. Veel läänemeresoome ja eesti eituspartiklite *ei, ep, es* päritolust [‘More on Finnic and Estonian negation particles *ei, ep, and es*’]. *Keel ja Kirjandus* 12. 968–975.
- Laan, Triin. 2009. Eitus Räpina murrakus [Negation in Räpina]. Tartu: University of Tartu BA thesis.
- Laanest, Arvo. 1975. *Sissejuhatus läänemeresoome keeltesse* [Introduction to Balto-Finnic languages]. Tallinn: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Lehtinen, Tapani. 1992. Suomen ja viron verbioppuisista sivulauseista. [‘On verb-final subordinated clauses in Finnish and Estonian’] In Tapani Lehtinen, Jyrki Kalliokoski & Kirsti Siitonen (eds.) *Nordens språk i Baltikum. Pohjoismaiden kielet Baltiassa. Baltian maiden yliopistojen pohjoismaisten kielten opettajien kokous, Riika 26.–30.11.1991* (Nordisk språksekretariats rapporter 18), 70–82. Oslo: Nordisk språksekretariat.
- Lindström, Liina. 1997. Eitus Võru murde suulises kõnes [‘Negation in Spoken Võro’]. In Karl Pajusalu & Jüvä Sullöv (eds.), *Läänemeresoome lõunapiir* (Võro Instituudi toimõtiseq 1), 143–154. Võro: Võro Instituut.
- Lindström, Liina. 2005. *Finiitverbi asend lauses. Sõnajärg ja seda mõjutavad tegurid suulises eesti keeles* [The position of the finite verb in a clause: Word order and the factors affecting it in spoken Estonian]. (Dissertationes philologiae estonicae universitatis tartuensis 16). Tartu: Tartu University Press.
- Lindström, Liina. 2017a. Lause infostruktuur ja sõnajärg. [‘Information structure and word order’]. In: Mati Erelt, Helle Metslang (eds.), *Eesti keele süntaks* (Eesti keele varamu 3), 537–565. Tartu: Tartu University Press.
- Lindström, Liina. 2017b. Partitive subjects in Estonian dialects. *Journal of Estonian and Finno-Ugric Linguistics* 8(2). 191–231. <https://doi.org/10.12697/jeful.2017.8.2.07>
- Lindström, Liina. To appear. Seto lause põhijooned [‘Main features of Seto syntax’]. In Andreas Kalkun, Karl Pajusalu & Ergo-Hart Västriik (eds.), *Setomaa* 3. Keel, rahvaluule ja tänapäeva kultuur.
- Lindström, Liina, Maarja-Liisa Pilvik, Mirjam Ruutma & Kristel Uiboaed. 2015. Mineviku liitaegade kasutusest eesti murretes keelekontaktide valguses. [‘The use of the compound past tenses in Estonian dialects in light of language contacts’]. In Jüvä Sullöv (ed.) *Aig õdagumeresoomõ keelin. Time and tense in Finnic languages*. (Võro Instituudi toimõndusõq 29), 39–70. Võro: Võro Instituut.


- Lindström, Liina, Maarja-Liisa Pilvik, Mirjam Ruutma & Kristel Uiboaed. 2019. On the use of perfect and pluperfect in Estonian dialects: frequency and language contacts. In Sofia Björklöf & Santra Jantunen (eds.), *Multilingual Finnic – Language contact and change*, 155–193. Helsinki: Finno-Ugrian Society.
- Lindström, Liina & Kristel Uiboaed. 2017. Syntactic variation in ‘need’-constructions in Estonian dialects. *Nordic Journal of Linguistics* 40(3). 313–349. <https://doi.org/10.1017/S0332586517000191>
- Lonn, Varje & Ellen Niit. 2002. *Saarte murde tekstid. Eesti murded VII* [Texts from insular dialects. Estonian dialects VII]. Tallinn: Eesti Keele Instituut.
- Lõuna, Kalle. 2003. *Petserimaa. Petserimaa integreerimine Eesti vabariiki 1920–1940* [Integration of the Petseri District into the Estonian Republic in 1920–1940]. Tallinn: Eesti Entsüklopeediakirjastus.
- Metslang, Helle, Karl Pajusalu & Tiit-Rein Viitso. 2015. Negation in Livonian. In Matti Miestamo, Anne Tamm & Beáta Wagner-Nagy (eds.), *Negation in Uralic languages*, 433–456. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.108.16met>
- Miestamo, Matti. 2005. *Standard negation. The negation of declarative verbal main clauses in a typological perspective* (Empirical Approaches to Language Typology 31). Berlin: De Gruyter Mouton.
- Miestamo, Matti. 2007. Negation – An overview of typological research. *Language and Linguistic Compass* 1/5. 552–570. <https://doi.org/10.1111/j.1749-818X.2007.00026.x>
- Miestamo, Matti. 2011. A typological perspective on negation in Finnish dialects. *Nordic Journal of Linguistics* 34(2). 83–104. <https://doi.org/10.1017/S0332586511000126>
- Miestamo, Matti, Anne Tamm & Beáta Wagner-Nagy. 2015. Negation in Uralic languages – Introduction. In Matti Miestamo, Anne Tamm & Beáta Wagner-Nagy (eds.), *Negation in Uralic languages*, 1–41. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.108.01int>
- Paas, Friedrich-Eugen. 1927. *Sega-abiellud ja nende mõju rahvusele piiriäärsetes maakondades Eestis* [Mixed marriages and their impact on nation in border counties in Estonia]. Tartu: Tartu Ülikool.
- Pajusalu, Karl. To appear. Seto South Estonian. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *Oxford guide to the Uralic languages*. Oxford: Oxford University Press.
- Pajusalu, Karl, Tiit Hennoste, Ellen Niit, Peeter Päll & Jüri Viikberg. 2009. *Eesti murded ja kohanimed* [Estonian dialects and place names]. Tallinn: Eesti Keele Sihtasutus.
- Payne, John R. 1985. Negation. In Timothy Shopen (ed.), *Language typology and syntactic description, Vol. I: Clause structure*, 197–242. Cambridge: Cambridge University Press.
- Pinheiro, José C. & Bates, Douglas M. 2002. *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- Plado, Helen, Liina Lindström & Sulev Iva. To appear. South Estonian Võro. In Daniel Abondolo & Riitta-Liisa Valijärvi (eds.), *The Uralic languages*. London: Routledge.
- Sammallahti, Pekka. 1977. Suomalaisten esihistorian kysymyksiä [‘Questions about the prohistory of Finns’]. *Virittäjä* 81. 119–136.
- Sang, Joel. 1975. Eitus Kihnu murrakus [‘Negation in Kihnu’]. *Keel ja Kirjandus* 3. 155–162.
- Strobl, Carolin, James Malley & Gerhard Tutz. 2009a. An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychological Methods* 14(4). 323–348. <https://doi.org/10.1037/a0016973>

- Strobl, Carolin, Torsten Hothorn & Achim Zeileis. 2009b. Party on! A New, Conditional Variable Importance Measure for Random Forests Available in the party Package. *Technical Report* no. 050. Department of Statistics, University of Munich.
- Szmrecsanyi, Benedikt. 2005. Language users as creatures of habit: a corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1. 113–149. <https://doi.org/10.1515/cllt.2005.1.1.113>
- Tael, Kaja. 1990. *An approach to word order problems in Estonian*. Tallinn: Eesti Teaduste Akadeemia humanitaar- ja ühiskonnateaduste osakond. Preprint KKI-66.
- Tagliamonte, Sali & 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. <https://doi.org/10.1017/S0954394512000129>
- Tamm, Anne. 2015. Negation in Estonian. In: Matti Miestamo, Anne Tamm & Beáta Wagner-Nagy (eds.), *Negation in Uralic languages*, 399–431. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.108.15tam>
- Tamminga, Meredith. 2016. Persistence in phonological and morphological variation. *Language Variation and Change* 28(3). 335–356. <https://doi.org/10.1017/S0954394516000119>
- Torres Cacoullou, Rena & Catherine E. Travis. 2019. Variationist typology: Shared probabilistic constraints across (non-)null subject languages. *Linguistics* 57(3). 653–692. <https://doi.org/10.1515/ling-2019-0011>
- Travis, Catherine E. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19(2). 101–135. <https://doi.org/10.1017/S0954394507070081>
- Uiboaed, Kristel. 2013. *Verbiühendid eesti murretes* [Verb constructions in Estonian dialects] (Dissertationes Philologiae Estonicae Universitatis Tartuenssis 34). Tartu: Tartu Ülikooli Kirjastus.
- van der Auwera, Johan. 2006. Why languages prefer prohibitives. *Journal of Foreign Languages* 1. 2–26.
- van der Auwera, Johan & Ludo Lejeune (with Valentin Goussev). 2013. The prohibitive. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <http://wals.info/chapter/71> (last access 29 June 2021).
- Viitso, Tiit-Rein. 1985. Läänemeresoome murdeliigenduse põhihood [‘Main features of Finnic dialect classification’]. *Keel ja Kirjandus* 7. 399–404.
- Viitso, Tiit-Rein. 2003. Põhiverbi muutumine eitussõna järel, lingua franca ja algkeel. [‘Inflection of the Main Verb after a Negation Word, the Protolanguage and Lingua Franca’]. *Keel ja Kirjandus* 1. 24–31.
- Vilkuna, Maria. 1989. *Free word order in Finnish. Its syntax and discourse functions*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Vilkuna, Maria. 1998. Word order in European Uralic. In Anna Siewierska (ed.), *Constituent order in the languages of Europe*, 173–233. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110812206.173>
- Vossen, Frens. 2016. Towards a typology of the Jespersen Cycles. Antwerp: University of Antwerp PhD dissertation.
- Weatherall, Ann. 2011. I don’t know as a prepositioned epistemic hedge. *Research on Language and Social Interaction* 44(4). 317–337. <https://doi.org/10.1080/08351813.2011.619310>

Zaikov, Pekka M. 2000. *Glagol v karel'skom jazyke* (grammaticeskie kategorii lica-cisla, vremeni i naklonenija) [Verb in Karelian]. Petrozavodsk: Izdatelstvo Petrozavodskogo gosudarstvennogo universiteta.

Address for correspondence

Liina Lindström
Institute of Estonian and General Linguistics
University of Tartu
Jakobi 2
Tartu 51005
Estonia
liina.lindstrom@ut.ee

 <https://orcid.org/0000-0003-1458-6682>

Co-author information

Maarja-Liisa Pilvik
Institute of Estonian and General Linguistics
University of Tartu
Tartu, Estonia
maarja-liisa.pilvik@ut.ee

Helen Plado
Institute of Estonian and General Linguistics
University of Tartu
Tartu, Estonia
helen.plado@ut.ee

Publication history

Date received: 31 July 2019

Date accepted: 1 September 2020

Published online: 24 August 2021