LIINA LINDSTRÖM, MAARJA-LIISA PILVIK,
MIRJAM RUUTMA & KRISTEL UIBOAED
University of Tartu

On the use of perfect and pluperfect in Estonian dialects: Frequency and language contacts[1]

Abstract The present paper examines the use of two compound tenses – perfect and pluperfect – in Estonian. Perfect and pluperfect have emerged due to the influence of the Baltic and Germanic languages and are used frequently in Estonian. However, while looking at the usage frequency derived from the Corpus of Estonian Dialects, dialect areas display remarkable differences, which can be explained either by local language contacts with Swedish, Russian, Latvian, and Finnic languages (Votic, Ingrian, and Finnish) or by functional differences in the use of compound tenses. It appears that there are two main regions where the compound tenses are used more often compared to other areas: the Insular dialect and Mulgi dialect regions. The increase of compound tenses in the Insular dialect could be a result of contacts with Swedish. However, the Insular dialect also exhibits a high number of negated utterances using the perfect reflecting changes in the formation of negation more generally in this area. The Mulgi dialect shows a high number of pluperfect forms that can be related to the abundance of reported narratives in the data, but also as an increase of using pluperfect as an evidential strategy, which is probably a result of contacts with Latvian.

Keywords compound tenses, perfect, pluperfect, Estonian dialects, frequency, corpus-based dialectology, language contacts

1. The previous version of this study has been published in Estonian (Lindström, Pilvik, Ruutma & Uiboaed 2015). This study was supported by the European Union through the European Regional Development Fund (The Centre of Excellence in Estonian Studies).

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Multilingual Finnic. Language contact and change. 155–193.
Uralica Helsingiensia 14. Helsinki 2019.
‹https://doi.org/10.33341/uh.85035›

1. Introduction

In the present article, we examine the usage frequency of Estonian compound past tenses – perfect and pluperfect – in Estonian dialects, using the data from the Corpus of Estonian Dialects[2] (CED). We take a look at the links between the frequency of use and local language contacts to examine whether long-term language contacts explain differences between dialects with regard to their (dis)preference for the compound tense forms.

Perfect and pluperfect are past tense forms that have emerged in Estonian and other Finnic languages most likely through the influence of surrounding Indo-European languages. Ariste has claimed that the development of compound tense usage in Estonian mainly has followed the example of the Baltic languages (Ariste 1956), later Serebrennikov (1959) has stressed that in addition to the Baltic languages, the Germanic languages have also played an important role. The same assumption is shared more or less by several scholars: the role of the Baltic languages has been important in the initial emergence of compound past tenses in Estonian and other Finnic languages, but the later use and function of the perfect and pluperfect are attributed to Germanic influence (Ikola 1960; Laanest 1975: 152). Laakso (2001: 191) links the spread of the compound past tenses more generally with the emergence of copula constructions, which, in turn, bear the imprint of other (especially Germanic) languages.

The emergence of compound tenses can also be explained as a natural development of a language – grammaticalization. Likewise, Estonian compound past tenses can be seen as a universal grammaticalization path from the aspectually used verb *olema* 'to be' and the predicative active past participle (*-nud*) to the tense category where the *olema*-verb and a past participle form the core of the clause. (Metslang 1993.) Aspectual use is still present today and is sometimes referred to as the retrospective, expressing the present state as a result of a previous event (Erelt 2013: 86), e.g., *Poiss on üllatunud* 'The boy is surprised', *Koer on surnud* 'The dog is dead'.

2. ‹http://www.murre.ut.ee/murdekorpus/›

During the development of the construction, the *olema*-verb had auxiliarized and the past event (i.e., the expression of tense) had been foregrounded (Metslang 1993). A typical perfect concludes the past event from the present point of view; "the perfect indicates the continuing present relevance of a past situation" (Comrie 1976: 52). The semantic shift from resultative to perfect thus includes the generalization from "current result" to "current relevance" (Lindstedt 2000).

Estonian compound tenses still can carry both temporal and aspectual meanings simultaneously with one of them becoming more salient in a sentence. The Estonian perfect has all the main functions that are characteristic of typical perfects (cf. Comrie 1976; Lindstedt 2000): perfect of result, experiential perfect, perfect of a persistent situation, perfect of recent past, quotative (evidential) perfect (see Metslang 1997 for details). Example (1) represents the experiential perfect; ex. (2) represents the pluperfect of the result.

(1) Estonian

Ta	***on***	***käi-nud***	*mitme-s*	*ülikooli-s.*
s/he	be.3SG	go-APP	many-INE	university-INE

'He has gone to many universities.'

(2) Estonian

Lapse-d	***ol-i-d***	*toa*	*juba*	*ära*	***korista-nud***.
child-PL	be-PST-3PL	room.GEN	already	PFV	clean-APP

'The children had already cleaned the room'

Although the emergence of perfect and pluperfect has also been explained as a language-internal phenomenon, language contacts have clearly reinforced the influence on compound tense usage. For instance, literary Estonian has been under strong Russian and German influence, especially during 1890–1990 (Metslang 1997).

In Estonian dialects, the use of the past compound tenses has not gained much attention. Lonn & Niit (2002) note that in the Insular dialect, compound tenses are extensively used. As this grammatical phenomenon exists in all dialects, the topic is not discussed in other papers dealing with dialects. Differences among dialects, however, have been attested for verbal constructions in general. Uiboaed et al. (2013) have studied different constructions consisting of a finite and

a non-finite verb in the Corpus of Estonian Dialects (on which the present article also is based) and found that the western dialects tend to use such constructions more extensively than eastern dialects. The western, higher frequency group was formed by the Mid, Western, Insular, and sometimes also Coastal (depending on the method) dialects. The lower frequency areas in the eastern part were formed by the southern dialects (Võru, Seto, Tartu, Mulgi) as well as the Eastern and Northeastern dialects. (Uiboaed 2013: 89.) Uiboaed's study, however, did not include *olema* + *-nud* compounds and the present article extends the knowledge of this phenomenon.

In this article, we discuss the use of perfect and pluperfect forms in Estonian dialects based on the Corpus of Estonian Dialects. We concentrate on regional differences and explore whether there are significant (dis)similarities between geographical areas, which could be seen in previous studies (Uiboaed et al. 2013; Uiboaed 2013). We therefore observe the compounds formed by the *olema*-verb and *-nud* participle in Estonian dialects, focusing primarily on usage frequency. Quantitative findings are interpreted in the light of language contacts.

Our main hypothesis is based on Uiboaed's study about other verbal constructions, so we expect higher usage frequency of compound past tenses in the western dialects compared to eastern areas. Differences might emerge from the fact that in western dialects compound tenses have similar functions as the simple past tense (imperfect) and the latter is preferred in eastern dialects. Discrepancies in usage frequency might be induced by language contacts: long-term interaction with Germanic languages (especially with Swedish) in western Estonia and stronger contacts with Latvian and Russian in eastern Estonia.

We now turn to the overview of the relevant contact languages which most likely have had an impact on usage frequency of the perfect and pluperfect in dialects. Additionally, we briefly present corresponding tense forms in these languages.

2. Estonian dialects and language contacts

Estonian dialects are traditionally divided into northern and southern groups, which have also been thought of as having developed from separate languages. These two groups differ on phonological, morphological as well as on lexical levels. As the Northeastern and Coastal dialects have systematic differences from other northern dialects, they are often classified as a separate group. (Kask 1984.)

The present study follows the dialect classification by Pajusalu et al. (2009) that has also been the basis for the Corpus of Estonian Dialects (Map 1). According to this classification, there are three main dialect groups: the North Estonian dialect group consists of the Eastern, Insular, Central, and Western dialects, the Northeastern Coastal dialect group consists of the Northeastern and the Coastal dialects, and the South Estonian group consists of the Mulgi, Tartu, Seto, and Võru dialects. A minor difference concerns Võru and Seto which are considered to be one dialect by Pajusalu et al., whereas the CED separates these as two different geographical variants. This distinction is clearly justified on a syntactic level, as Seto has proved to be considerably different from Võru in various syntactic respects (see Uiboaed 2010; Lindström, Uiboaed & Vihman 2014; Lindström forthcoming).

The influence of language contacts on Estonian morphosyntax has been examined mostly in connection with two linguistic areas (*Sprachbund*): the Circum-Baltic (see Dahl & Koptjevskaja-Tamm 2001; Wälchli 2011) and the European linguistic area, whose characteristics are usually subsumed under the term *Standard Average European* (SAE[3]) (Haspelmath 2001). The main mediator of the SAE features into Estonian has probably been German (both Low and High German), which was the language of the upper class in the region for a long period of time. German has influenced Estonian on all linguistic levels, including the functions of the perfect and pluperfect (Metslang 1997), and this impact was constant from the 13th century up to the beginning of the 20th century. However, based on previous studies, it is hard to estimate the regularity and intensity of German influence on different Estonian dialects; there is no evidence that German influence

3. SAE features in Estonian have been discussed by Metslang (2009).

Map 1. The dialect classification in the Corpus of Estonian Dialects. Dialect names are highlighted in larger font. The map is the modified version of the parish map (EKI 2014). Place names marked on the map are referred to later in the text.

has been stronger on some areas than on others. Still, we can infer from geographical and historical facts that German influence has most likely been weaker in the southeastern part of Estonia, e.g., the Seto dialect area, due to its intense connections to Russian-speaking Pskov.

The Circum-Baltic language area includes the Finnic languages, Russian, the Baltic languages (Latvian and Lithuanian), the Germanic languages (particularly Swedish, German, Low German), and several other languages spoken in the region, see the review by Koptjevskaja-Tamm & Wälchli (2001) and Wälchli (2011). The Circum-Baltic

language area has been described as a complex linguistic community where mutual influences of individual languages are ongoing and there is no single dominating language; therefore, it is also referred to as a contact superposition zone rather than a *Sprachbund* (Koptjevskaja-Tamm & Wälchli 2001: 626; Wälchli 2011: 325). There are no isoglosses covering all the Circum-Baltic languages (Koptjevskaja-Tamm & Wälchli 2001: 728–732).

Estonian belongs to the eastern part of the Circum-Baltic languages together with the Baltic, Finnic, and Slavic languages (especially Russian). Although these languages belong to different language families, many morphosyntactic similarities have been attested among them (see, e.g., Dahl & Koptjevskaja-Tamm 2001; Klaas-Lang & Norvik 2014; Serzant 2012, 2015; Vaba 2011). All of the above-mentioned studies are characterized by the fact that at least when addressing the morphosyntax, Estonian is treated as a whole and the intralanguage variation is not accounted for. However, there are a multitude of phenomena in Estonian displaying significant regional differences that can be explained by contacts that have taken place locally, that is, they have affected only certain areas. Thus, for example, due to the Swedish influence in Swedish contact areas the use of the nominative subject instead of the partitive has increased (Juhkam 1998, 2012), but the same has also been observed in the bilingual Estonian-Russian region in Iisaku[4] (Must 1965). Swedish influence has increased the use of the nominative object (Juhkam 1983), and again the same tendency has been noted in Iisaku (Must 1965), where the Swedish influence could not be the main factor addressing this change. Thus, similar phenomena may occur in a number of regions, but the languages driving these developments may vary.

The primary contact languages for dialects are Swedish, Latvian, Russian, and Votic, to a lesser extent also Finnish and Ingrian, in the earlier period also Livonian. These languages have a clear regional spread, long-term contacts, and bilingual speakers in corresponding areas. We now turn to the individual contact languages and briefly describe the history of these contacts.

4. Place names mentioned here and in the rest of the article are marked on Map 1 in smaller font.

Swedish. Estonian has had long-standing contacts with Swedish and the Scandinavian languages, but more intensive and direct contacts were established from 12th century, when the Swedes moved to Estonia's western and northern coasts. Since then, the contacts between Swedish and Estonian have been strong and persistent, indicated by both the number of loanwords as well as the mutual interactions in the pronunciation and morphosyntax (Juhkam 1998). Expanding during the Medieval period and reaching its peak in the 15th century, Swedish settlement was continuous between Ruhnu and Viimsi (Juhkam 1992).

Until World War II, the Swedish-speaking population of western Estonia (Vormsi, Noarootsi, Riguldi, Osmussaar), northern Estonia (Vihterpalu, Kurkse, the Pakri Islands, Tallinn), and on Ruhnu Island was rather stable. Estonian and Swedish dialect contacts, therefore, particularly affected the islands and the Western dialect area (mainly the western Estonian coastal parishes) as well as the northwestern part of Estonia (Risti, Harju-Madise, Keila Parishes) (Juhkam 1998: 26–29). During World War II, most of the Estonian Swedes moved to Sweden.

Russian. The most long-standing Russian influence has been present in southeastern Estonia, which has long been in the sphere of influence of Pskov. The main Russian influence has been attested in the Seto area, which is visible in the material culture as well as in lexical loans (Must 2000), pronunciation (Pajusalu 1999), and grammar (Kask 1984: 27).

Close and long-term contacts have also been present in the Northeastern dialect, in the area of Iisaku, where during the 17th–18th century there arose a Russian-Estonian bilingual population (Must 1965, 2000). There has been long-term Russian settlement on the west coast of Lake Peipus, which dates back to the end of the 16th century, when a large number of Russians migrated to Estonia. At the end of the 17th century, most of the current villages already existed. In the beginning of the 18th century, Old Believers, who had left the Russian Orthodox Church, moved to Estonia and the former Estonian villages on the coast of Peipus became Russian villages. (Must 2000: 8.) Russian contacts on the west coast of Lake Peipus, therefore, are not as long-standing as in Seto dialect area; however, they have been intensive and influenced especially the Eastern dialect of Estonian and the southern part of the Northeastern dialect.

Latvian. Estonian-Latvian language contacts have been long and many corresponding structures can be found in both languages. Latvian influence is the strongest in the Estonian enclaves in Latvia (Leivu, Lutsi) and in the dialect areas near Estonia's southern state border (Vaba 2011). The number of Latvian loanwords is the highest in the Võru and Mulgi dialects as well as in Saarde and Häädemeeste (the southern part of the Western dialect) (Vaba 1997).

Latvian contacts are also noticeable on the islands; this has been described as the *Sprachbund* surrounding Irbe Strait – more specifically common traits have been attested in western Saaremaa, western Hiiumaa, the Latvian Dundaga dialect, the Ruhnu Swedish language, and Livonian. It is likely that Livonian has been an intermediary language at least on the lexical level. (Vaba 1997: 475–479.)

Votic, Ingrian, and Finnish. Votic influences are apparent in the Eastern (Pall 1982; Univere 1988) and Northeastern dialects (Pall 1982; Alvre 2000). Pall (1982) explains Votic traits in the Eastern dialect with direct contacts that have arisen as a result of the Votes' immigration. Alvre (2000) has listed common features in the Northeastern dialect and Votic and highlights various phonetic and grammatical similarities. Ariste has foregrounded Vaivara Parish as a complex contact area where the eastern part has been exposed to a number of other Finnic languages – Votic, Ingrian, and Ingrian Finnish. The Votic-Ingrian-Finnish mixed villages have been situated on the other side of the Narva River. (Ariste 1962.) Contacts between the Northeastern dialect of Estonian, Votic, Ingrian, and Ingrian Finnish have implied many loanwords from one Finnic language to another in the area (see Björklöf 2019, in this volume).

Based on the distribution of Finnish loanwords, it has been argued that the Finnish influence is evident also in some parishes of the Coastal dialect, mainly Kuusalu, Haljala, Jõelähtme (e.g., Söderman 1996; Norvik 2000; Björklöf 2012, 2018). This is a result of intensive oversea trading and fishing contacts (Must 1987: 13–15; Björklöf 2012, 2018).

3. Perfect (and pluperfect) in the contact languages

Next we will examine the use of compound tenses in the contact languages in order to identify the nature of their potential effects on Estonian dialects.

Swedish. In modern Swedish, the perfect is mainly formed with an auxiliary verb *ha(va)* 'to have' and the main verb in its supine form, which historically is the neuter form of the perfective participle (Dahl 1995: 61), examples (3)–(4). Additionally, a more constrained perfect-like construction is comprised of the auxiliary *vara* 'to be' and the perfective participle (example 5). In modern Swedish, the use of the construction with *vara* is restricted and it expresses mainly resultativity. In older Swedish, *vara* as an auxiliary was more widespread compared to contemporary Swedish, and its use was less limited. (Larsson 2009: 156.) The gradual disappearance of the perfect formed with the *vara*-verb is related to the development of the perfect construction with the *ha(va)*-verb: *vara* dominated until the 17th century. A remarkable decline in its frequency of use occurred sometime around the turn of the 17th–18th century, when the *ha(va)*-perfect started to become significantly more frequent. (Larsson 2009.)

(3) Swedish (Larsson 2009: 144)

Han har kommit hit.
he have:PRS come:SUP here
'He has come here.'

(4) Swedish (Larsson 2009: 144)

Artikeln har blivit färdig.
article:DEF have:PRS become:SUP ready:SG
'The article has become ready.'

(5) Swedish (Larsson 2009: 272)

De är redan anlända.
they be:PRS already arrive:PTCP.PL
'They have already arrived.'

Thus, Swedish has developed the so-called HAVE-perfect or the possessive perfect, which is characteristic of the European linguistic area (Haspelmath 2001; Heine & Kuteva 2006), but the former BE-perfect is still preserved to some extent. Also, in Estonian Swedish dialects the *ha(va)*-perfect is more frequent than the *vara*-perfect (Henrik Rosenkvist, p.c.). As perfect and pluperfect are extensively used in Swedish and they have a wide functional range, one could assume that the contact with Swedish has increased the use of compound tenses also in the contact area.

Russian. Russian has no special form for either the perfect or pluperfect, but only one finite past tense, which historically is a form of the participle (Tommola 2000: 442). General past has functions that are filled by the perfect or pluperfect in other languages. Perfect has, however, existed in Old Church Slavonic (in example 6) where it was formed with the copula 'be' and the participle.

(6) Old Church Slavonic (Tommola 2000: 459)
čast' post-a preminu-l-a est'
part fast-GEN pass:PFV-PP-F be.3SG
'A part of the fast-time has passed.'

Additionally, Russian has developed perfect-like constructions: a resultative passive construction (in example 7) that, when occurring with the past copula, is ambiguous between a perfect (dynamic) and a present resultative (stative) reading (Tommola 2000: 463), and the so-called possessive perfect (in example 8), in which the actor is expressed with the construction *u* + GEN (*u syna* 'son'). This construction is particularly common in the northern and northwestern Russian dialects, and it has spread also to the Finnic languages and Latvian (Seržant 2012). In Estonian, the corresponding construction is formed with an adessive agent, the verb *olema* 'to be', and a passive past participle (Lindström & Tragel 2010; Lindström 2015), more frequently used in southern and eastern Estonia (Uiboaed 2013: 176, ex. 9), i.e., in the contact area of Russian.

(7) Russian (Tommola 2000: 463)

Magazin by-l otkry-t (v vosem').
shop be-PST.M open:PFV-PPP (PREP eight)

'The shop was open/opened (at eight o'clock).'

(8) Russian (Tommola 2000: 464)

U syn-a institut za-konč-en/-o.
PREP son-GEN institute PFV-finish-PPP:M/-NT

'My son has graduated from the institute.'

(9) Estonian (Seto dialect)

Salmõ-l oll' ka ostõ-t purgitäis me-tt
Salme-ADE be.PST.3SG also buy-PPP jarful honey-PRT

'Salme had also bought a jarful of honey.'

The focus of the current article lies on *olema* + *nud*-participle compounds and therefore the passive constructions (including the possessive perfect) are not discussed further.

Russian lacks perfect and pluperfect forms and uses a general past instead, which leads us to assume that similar tendencies appear also in the Russian contact areas: simple past is used instead of perfect and pluperfect and the compound past tenses have a lower usage frequency in comparison to other areas.

Latvian. The Latvian perfect is formed by much the same means as the Estonian perfect, i.e., using the auxiliary *būt* 'to be' and the active past participle, except that the participle agrees with the subject in both number and gender (example 10). The auxiliary verb can occur either in present, past, or future form. In the present perfect (the auxiliary is in the present tense) when the subject is in third person singular, the auxiliary verb can be omitted from the sentence. (Kalnača 2015: 82.)

(10) Latvian (Kalnača 2015: 82)

Šodien Bille ar vecomāti ir gājušas ilgi.
today Bille with grandmother:INS.F be.PRS.3 walk:PTCP.PST.F long

'Today Bille and her grandmother have been walking for a long time'

As the Latvian perfect and pluperfect are formally and functionally quite similar to those in Estonian, it can be assumed that the use of the perfect and pluperfect in the contact area is more or less the same as the Estonian average or has increased the use of compound tenses in the Latvian contact area of Estonian.

Votic, Ingrian, and Finnish. In all Finnic languages, including Votic, Ingrian, and Finnish, the perfect and pluperfect are both formed with the verb 'be' and the past participle.

In Votic, it has been noted that compound tenses are rare and instead the simple past tense is used. (Ikola 1960; Markus & Rozhanskiy 2011: 233). The tendency seems to apply when observing Votic texts that are incorporated into the Corpus of Estonian dialects.

Ikola (1960) has argued that the development of the Finnish tense system (including the perfect and pluperfect) in the old literary language has been influenced by Swedish. However, a comparative study on the use of the perfect, based on translations of the Bible, has revealed that in the first Finnish translations of New Testament texts (1776), the perfect occurs less frequently than in the Swedish version, and its functions also differ significantly from the Swedish perfect (Swedish representing a typologically "usual" use of perfect) (Dahl 2014). Following that, Finnish dialects that have been in contact with Estonian dialects may have used compound tenses even less than what is attested in the Swedish-influenced translations of the Bible.

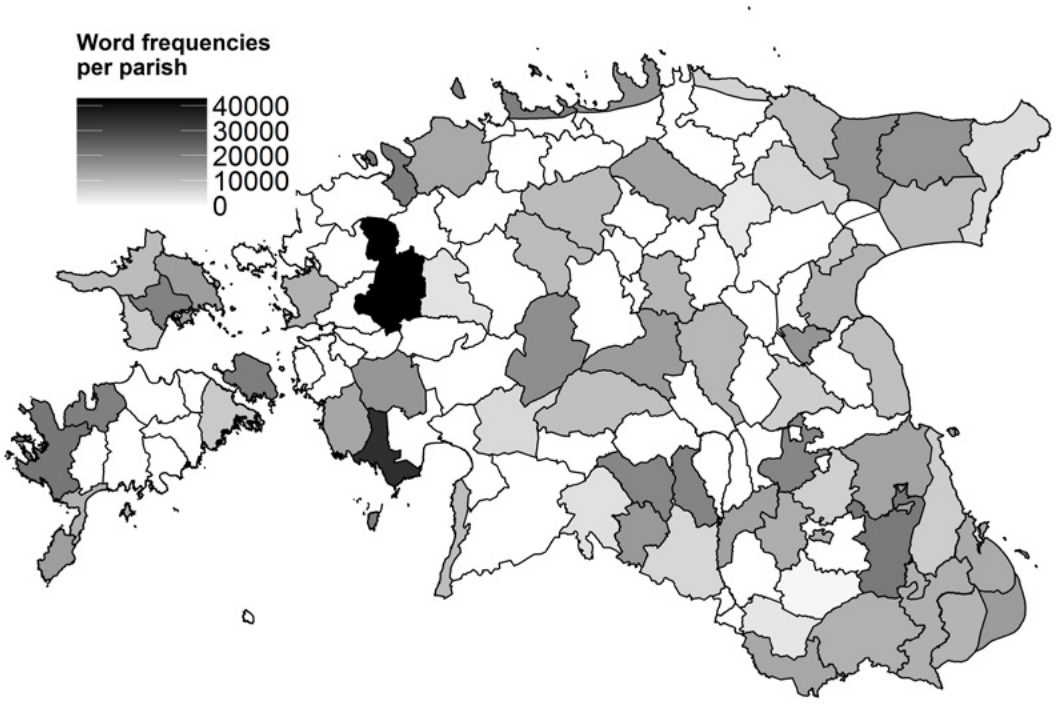
As the Votic, Ingrian, and Finnish forms of the perfect and pluperfect are similar to Estonian, it can be expected that contacts with those languages have not significantly affected the use of the perfect and pluperfect in the contact area.

4. The usage frequency of compound past tenses in Estonian dialects

4.1. Data and method

The current study presents frequency data from the Corpus of Estonian Dialects (CED). The role and impact of frequency on the development and change of linguistic phenomena has been studied quite thoroughly in recent decades (see, e.g., Arppe et al. 2010; Divjak & Gries 2012; Gries & Divjak 2012). The importance of frequency has been in central focus especially in usage-based approaches to language (Barlow & Kemmer 2000; Bybee 2010). In studies concerning dialect syntax, it has been found that syntactic variation usually cannot be described via categorical distinctions (the phenomenon X occurs in some area and Y does not), but rather in terms of frequential typicalities and atypicalities (Kortmann 2010). Moreover, text frequencies have been claimed to reflect the perception of the language system more adequately than for instance the atlas data used widely in traditional dialectology (Szmrecsanyi 2013: 4). Frequency effects can hardly be ignored when we talk about language contacts: regional idiosyncrasies can be exposed by frequency. The usage pattern already existing in a language (or dialect) can acquire a higher usage frequency due to a similar pattern in a contact (or model) language with which the speakers are constantly confronted (Heine & Kuteva 2005: 47). Even if a particular phenomenon exists in several areas, its usage frequency can be substantially lower outside the core of the observed contact region (see Koptjevskaja-Tamm & Wälchli 2001: 627). Change in the usage frequency of some (previously existing) pattern has been considered the typical contact-induced grammatical transfer (Heine & Kuteva 2005: 48).

We now turn to the usage frequency of perfect and pluperfect tense forms in the CED, which consists mainly of spoken dialect interviews recorded in 1960–1970s. Speakers of the CED have been mostly elderly people considered to be good representatives of the old local dialects. (For an overview of the CED, see Lindström et al. 2019, in this volume.) The research material has been collected from the morphologically annotated part of the corpus; at the time of data



Map 2. Word frequencies per parish in dialect data. White indicates the absence of data from that parish.

retrieval (01.10.2014) the corpus contained 834 311 morphologically annotated words. The overall number of speakers in the data was 389 and the number of interviews (= files) of various length was 323.

The data were automatically retrieved with R (R Development Core Team 2013) scripts, searching for the lemma of the verb *olema* with a 7-word context from both left and right of the verb. Only observations with the *-nud* participle within the given context were selected for further inspection. These criteria were met by 13 385 context units, which in turn were inspected manually for true compound past tenses. The final dataset for further analysis contained 6 242 utterances with either perfect or pluperfect forms.

The CED does not cover all Estonian parishes and therefore the research data are not fully representative of the whole country. Map 2 presents the parishes which are covered with data for the present analysis. The darker the region, the more material has been included from that area. White areas indicate parishes where no data was available.

In the present study we examine the usage frequency of the compound tense forms consisting of the verb *olema* 'to be' and the *-nud* participle. Different moods have not been distinguished except in Section 5.1. where we only discuss examples of indicative mood in order to compare the proportions of perfect and pluperfect; however, moods other than indicative are used infrequently and thus do not considerably distort the overall picture. At this stage, the functions of the compound past tenses have not been tracked separately. We have, however, differentiated affirmative and negative sentences as well as perfect and pluperfect in order to explore the patterns which potentially affect the overall usage frequency of the phenomenon in the data.

We carried out analyses on three levels: dialects, parishes, and villages. Since the number of words from different dialects is represented unevenly in the corpus we utilized frequency normalizations in order to level out frequency fluctuations and achieve comparability across areas. In the first two analyses (dialects and parishes), we use the mean size of the corpus as a normalization base[5].

4.2. Results

4.2.1. Dialects

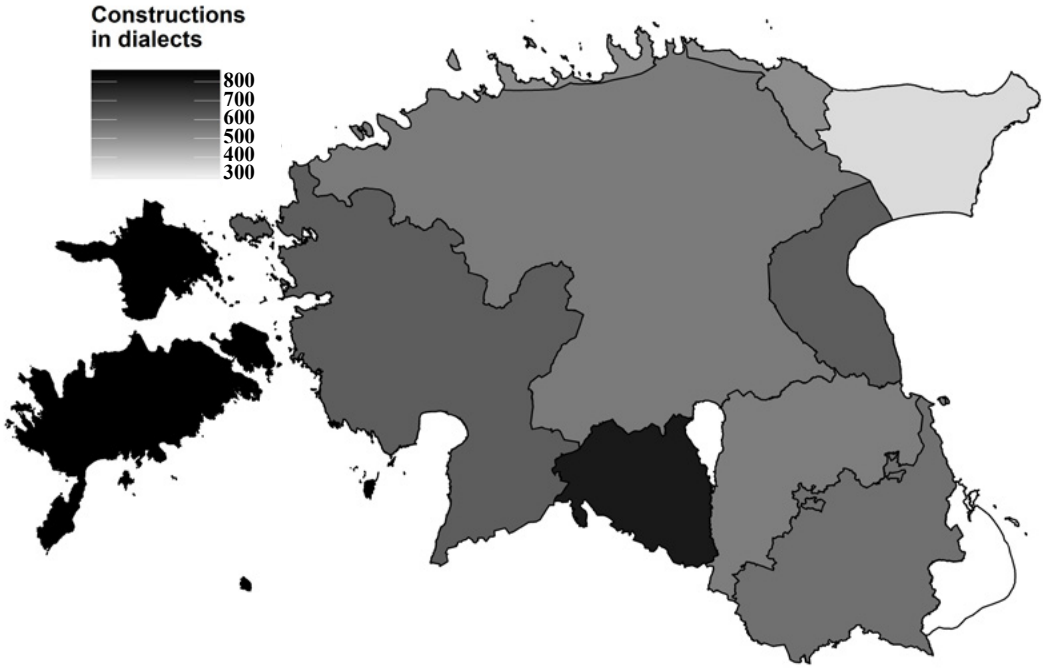
First, we examine the usage frequencies of the compound past tenses across dialects. Table 1 presents the absolute and normalized frequencies of compound tense forms. Here, the frequencies are summarized within a single dialect, i.e., the total number of occurrences across all the parishes belonging to the respective dialect is presented. The normalized frequencies, therefore, reflect the usage frequencies of the compound past tenses as if the amount of annotated words from all the dialects were equal.

Map 3 gives the same data spatially. Darker areas indicate higher frequency and lighter areas indicate lower frequency.

5. The formula of normalization: (form frequency/words in dialect or parish)*corpus mean.

Dialect	Words	Absolute frequency	Normalized frequency
Coastal	51 667	316	510
Eastern	45 280	339	625
Insular	166 898	1 723	861
Mid	130 086	860	552
Mulgi	63 516	617	810
Northeastern	47 660	193	338
Seto	39 175	123	262
Tartu	65 591	428	544
Võru	70 038	486	579
Western	154 400	1 157	625

Table 1. The absolute and normalized frequencies of the compound tense constructions across dialects.



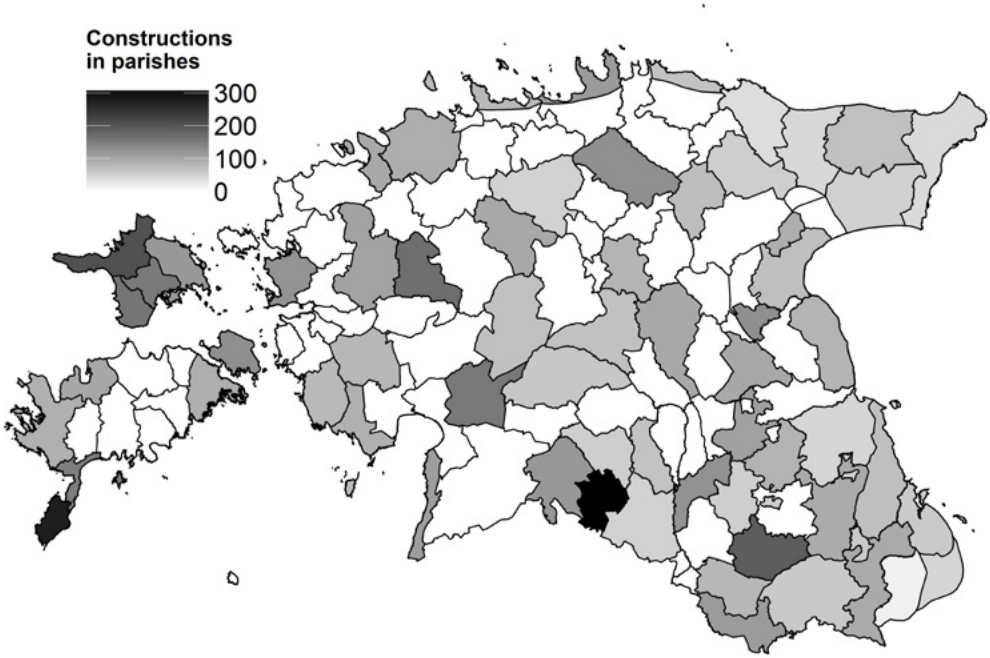
Map 3. Normalized frequencies of compound tense constructions in the dialect data.

Both the table and the map reveal that the usage frequency of the compound tenses is particularly high in the Insular and Mulgi dialects, whereas in the Seto dialect it is the lowest. The frequency is low also in the Northeastern dialect. Thus, the differences between the dialects are remarkable and at least partially correspond to the language contact situation: high frequency in the contact area of Swedish (Insular dialect), low frequency in the contact area of Russian and the Finnic languages (Seto, Northeastern dialects). However, the Eastern dialect also has had long-standing contacts with Russian as well as with Votic, which, however, are not reflected in decreased use of perfect and pluperfect. Likewise, the high frequency of compound tenses in the Mulgi area is hard to explain with Latvian contacts as we lack evidence concerning the higher prevalence of compound tense usage in Latvian compared to Estonian. However, based on current data and taking into account that there exist corresponding compound tenses in Latvian, we may conclude that the influence of Latvian may be responsible for increased use of compound tenses in Mulgi.

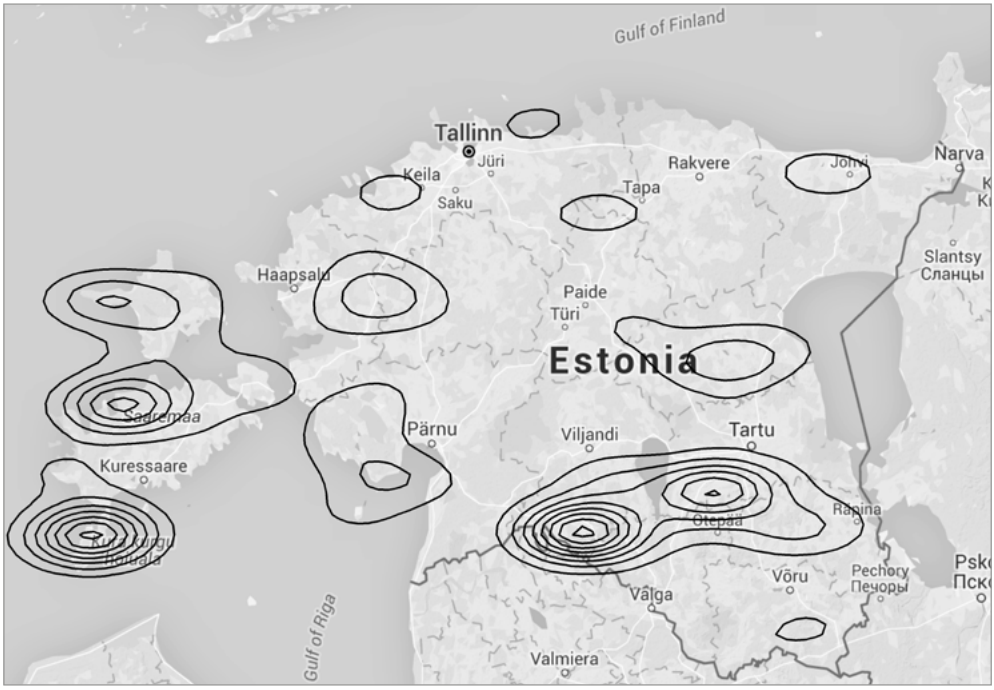
4.2.2. Parishes

We now turn towards a lower level of generalization, and have a look at the compound past tense frequencies by parish. The sparseness of the data in the CED prevents us from getting a more coherent geographical picture: the recordings do not cover all parishes and the map, therefore, is fragmented by white areas, indicating the absence of data from those parishes (see also Map 2 above).

Next, we take a look at the normalized frequencies, this time the normalization base is the mean number of words from parishes. Map 4 does not show clearly distinguishable larger coherent areas, but similarly to dialects, the construction frequency is higher on the islands, particularly in Hiiumaa and in the southwestern part of Saaremaa. On the mainland, the usage frequency of compound past tenses is higher in Karksi Parish, which belongs to the Mulgi dialect, and in Urvaste Parish in the Võru dialect region.



Map 4. Normalized frequencies of compound tense constructions across parishes.



Map 5. Distribution of data points across villages (Google (2016) as a base map).

4.2.3. Villages

Next, we will narrow the geographical unit and take a look at individual villages. This time we use absolute frequencies per village and visualize the density of data points on Map 5 (see the previous page). The interpretation of the map is similar to one known from regular geographical maps visualizing landscape surfaces. The number of lines on the map indicates the proportion of constructions in the region. The higher the concentration of lines, the more data points (i.e., occurrences of compound tenses) we have in our dataset for that particular area (these areas are "higher"). The data reveal that the frequency of the compound tenses is very high in the southwestern part of Saaremaa, northern Saaremaa, and western Hiiumaa (Reigi Parish), i.e., in the Insular dialect. The second distinctive core area is formed by the villages in the Karksi Parish, in the Mulgi dialect area, where the frequency of compound tenses is very high, and in the Tartu dialect (Nõo and Rõngu Parishes). When interpreting the map, however, one has to keep in mind that the construction frequencies depicted on the village map are not normalized and the density of data points is at least partially also affected by the amount of raw data available from different regions (see Map 2).

4.2.4. Language contacts and the usage of compound past tenses: summary

When the information from the three maps is converged, differences in usage frequencies of the compound tenses imply the influence of individual contact languages. Frequency of the compound past tenses is higher in areas where the Swedish language contacts have been stronger and longer; particularly in the Insular dialect (in the southwestern part of Saaremaa and the western part of Hiiumaa). The parish map (Map 4) shows that the frequency of compound tenses is fairly high on the entire island of Hiiumaa, especially in Reigi Parish, a major Swedish populated region in the past. Relatively high construction frequency in the Western dialect also illustrates the influence of Swedish, although the smaller areas with long-term Swedish settlement were not included in the analysis.

Surprisingly, the Mulgi dialect (especially Karksi) stands out with a very high frequency of compound tenses (Maps 4 and 5). The

frequency is rather high also in Halliste, Rõngu, and Urvaste, which could be indicating that we are not dealing with the idiosyncratic behavior of only one parish (or one village/speaker). One could assume that the usage frequency of the compound tense forms has been influenced by contacts with Latvian where similar constructions exist.

The Seto and the Northeastern dialects are low frequency areas. Low frequency in Seto can be explained by the long-term contacts with Russian, which lacks the corresponding tense constructions. Historically, Seto has had less contacts with the German-speaking upper class, which definitely has reduced the overall German influence in that area and, which could serve as an additional factor leading to less frequency in compound tense usage. In contrast, the second Russian contact area – the Eastern dialect – exhibits rather the opposite tendencies (Map 3). The newer and more restrained nature of Russian contacts in the Eastern dialect could have prevented the extensive structural impact on the dialect.

The Northeastern dialect is characterized as a contact area with Russian (especially Iisaku Parish, see Must 1965), on the one hand, and with other Finnic languages (Votic, Ingrian, Finnish), on the other hand. The usage frequency of the compound tenses is relatively low in the Northeastern dialect. Due to the complex past historical and geographical situation in this dialect area it is difficult to assess whether the rare use of compound tenses is a result of language contacts or some other factors, since compound tenses do exist in neighboring Finnic languages. The contacts with Russian have taken place in a relatively restricted area and thus have probably not influenced the whole region; however, the Russian influence may have come *via* neighboring Finnic languages since there has been a strong influence of Russian especially on Votic and Ingrian. Thus, the low frequency of compound tenses in the Northeastern dialect can be a result of contacts with other Finnic languages, indirect Russian influence, or it may be an indication of weaker influence from Germanic languages, i.e., this area is more conservative in this respect.

We can thus conclude that the language contacts have influenced the usage frequency of the compound tenses in Estonian dialects. Intense contacts with Swedish have been constant on the northwest coast and on the islands since the 12th century (or earlier), active overseas communication is known also from earlier times. The increase in

usage frequency in the islands and in western Estonia can therefore be attributed at least partly to long-term contacts. The Seto area has long contacts with Russian, whereas similar contacts in the Eastern dialect region on the coast of Lake Peipus have been significantly shorter. This could possibly explain why the usage frequencies of the compound tenses differ significantly in the two regions with Russian contacts. Another possible explanation that applies more to the Seto and the Northeastern dialects is that both regions could possibly have been much less affected by German than other areas in Estonia.

5. Other factors influencing construction frequency

In addition to language contacts, the overall increase in frequency is associated with other factors, as well. In our data, two areas were especially highlighted with high frequency: the Mulgi (especially Karksi Parish) and Insular dialects. In both areas, the perfect and pluperfect have a variety of different functions, which explains the higher usage frequencies. In the next sections, we inspect differences between the use of 1) perfect and pluperfect, 2) negative and affirmative clauses to explore the potential influence of these aspects on frequency.

5.1. Perfect vs. pluperfect

While a typical perfect expresses a past event from the present point of view, the reference point of the pluperfect is in the past, i.e., it describes a past event from the point of view of an observation time which could be also in the past. One of the most salient functions of the pluperfect is its quotative (evidential) use. In contemporary (spoken) Estonian, the pluperfect is frequently used in reported narratives whereas the perfect is rather rare in this function. (Lindström & Toomet 2000.)

To examine the proportions of the two compound tenses we only included the instances which were morphologically marked for indicative mood, since only indicative makes a distinction between past tense forms (imperfect, perfect, and pluperfect) (Erelt et al. 1995: 74–79). Table 2 and Figure 1 present the distribution of the perfect and pluperfect

across dialects. The proportion of the pluperfect is highest in the Mulgi dialect (53% of all compound tense constructions). Other South Estonian dialects (Võru and Tartu) have a higher proportion of pluperfects, as well, implying that there might be a wider functional range for pluperfect in this region. Seto, however, differs from other southern dialects with a lower proportion of pluperfect tense usage. As Table 1 and Map 3 already illustrated, Seto makes less use of the compound tenses in general.

Dialect	Perfect	Pluperfect	Pluperf (%)
Mulgi	263	301	53
Võru	226	219	49
Tartu	206	143	41
Eastern	190	93	33
Western	697	325	32
Mid	556	220	28
Coastal	198	76	28
Northeastern	124	42	25
Seto	91	31	25
Insular	1 327	298	18

Table 2. Proportions of indicative perfect and pluperfect forms in the dialect data[6].

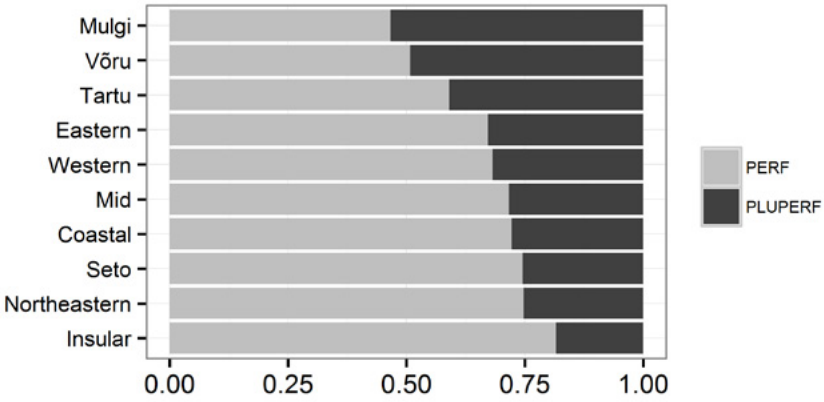


Figure 1. Proportions of indicative perfect and pluperfect forms in the dialect data.

6. Note that numbers in this table differ compared to other tables, as only indicative mood is included to calculate the proportions of perfect type.

Compared to other dialects, Mulgi exhibits a disproportionately high use of pluperfects, which express evidentiality (the reported nature of the information) and are particularly characteristic of reported narrative texts (example 11).

(11) Mulgi, Karksi Parish

miu	*vanaemä*	***olli***	***elä-nu***	*(.) vanast*	*Püügle*
I.GEN	grandmother	be.PST.3SG	live-APP	formerly	Püügle
Savimäe-l	*ja (.) ja*	***olli***	***asute-n***	*sõnna*	
Savimäe-ADE	and and	be.PST.3SG	establish-APP	there	
ende-l (.)	*maja*	*tetä*	*ja (...)*		
self-ADE/ALL	house.GEN	do.INF	and		

'My grandmother lived in Püügle Savimäe and
had established herself a house there'

The abundance of the quotative uses of pluperfect in the Mulgi area is at least partially related to the fact that the Mulgi sample contains more narrative texts (such as legends and reported narratives) than general descriptions of everyday work and customs, which are characteristic of most CED texts. Among the texts from Karksi, there are two recordings collected by Paul Ariste in 1938, which mainly contain relatively short reported narratives (stories). The expansion of usage frequency may therefore be related to the higher-than-average amount of narrative texts, which motivate the frequent use of the pluperfect for evidential purposes.

However, the large number of reported narratives in the data does not alone explain the general frequent use of pluperfect in Mulgi or in other South Estonian dialects (Tartu and Võru) since narrative texts are not exceptional in other areas included in the sample. Moreover, the pluperfect is not the only means for marking reported evidentiality and therefore it is not clear why the Mulgi dialect prefers the pluperfect to such an extent for expressing this meaning. Estonian has a variety of evidential constructions conveying evidential meaning, e.g., quotative mood, bare past participles, infinitives, evidential *pidama* 'must' + infinitive constructions, etc. (Kask 1984; Kehayov 2008). The use of bare past participles (example 12), in particular, is common across all of Estonia (Pilvik & Uiboaed 2014).

(12) Võru, Põlva Parish

ja ***tul-nuq*** *sõss hainaruga (.) läve-st* *sisse* *ja*
and come-APP then haystack threshold-ELA in and
ast-no *sinnäq sängü* *ette* *ja* *ütel-nü* *et (...)*
step-APP there bed.GEN ahead and say-APP that
kule *Jaan' et (.)*
listen.IMP.2SG Jaan that

'And (it has been told that) the haystack came in across the threshold and stepped by the bed and said that "listen, Jaan" /…/'

Much as in other areas, in Mulgi the bare past participle has a high usage frequency; however, it often alternates with the pluperfect and perfect tenses. In example (13), the narrative starts with the pluperfect (first line), followed by a bare past participle (line 2, *tahten* 'want-APP'), which is replaced with a perfect in its negative form, and later the speaker continues again with a bare past participle. Similar patterns can be found in other dialects, as well, but for some reason the Mulgi dialect clearly prefers the pluperfect for expressing evidentiality.

(13) Mulgi, Karksi Parish

nuu vanast ***olli*** *(.)* *talu* *esä* ***ollu*** *ja kangest*
PTCL earlier be.PST.3SG farm.GEN father be.APP and very
rikass (...) siss ***tahte-n*** *(.) et* ***eij ol*** ***tahte-n***
rich then want-APP that not be.CNG want-APP
las-te-l *se-dä* *raha (.)* *anda (.)* *ja* ***mõtel-nu***
child-PL-ADE/ALL that-PRT money.PRT give.INF and think-APP
t (.) *ma* *mata* *ta* *maha* *või* *kohekil* *äräde*
that I bury.1SG it.GEN ground.ILL or somewhere PFV
seant ihne ***ollu*** *kangest (...)*
such stingy be.APP very

'Earlier, there was a farmer and (he was) very rich. Then (he) wanted (repairs:) didn't want to give the money to (his) children, and thought that "I will bury it in the ground or somewhere", (he) was so stingy'

A previous study on synthetic and analytic quotative constructions in the Corpus of Estonian Dialects has also noted that quotative constructions in general are quite frequent in Mulgi data and analytic quotative constructions (mainly indicative perfect and pluperfect) in particular are used there more frequently than in other dialects (Pilvik & Uiboaed 2014).

The main contact language of the Mulgi dialect region, Latvian, has a special mood for marking reported evidentiality (*oblique mood*) as well as bare past participles that are used for the same purpose (Kalnača 2015; Muižniece, Metslang & Pajusalu 1999). The contacts with Latvian may have reinforced the need for expressing reportedness more generally in South Estonian, especially in the Mulgi area. The alternation between the compound tenses and bare participle is also common in Latvian (see Siegl & Kehayov 2006 for discussion on the relationship between the bare participle and compound tenses).

5.2. Affirmative vs. negative constructions

The data analysis revealed that there are remarkable differences in proportions of negative compound tenses between dialects.

The inspection of the ratio between affirmative and negative utterances (Table 3 and Figure 2) reveals that the Insular dialect exhibits considerably more negative sentences among the *olema* and *-nud* compounds (47%) than other dialects. The number of negated sentences is slightly higher also in the Western dialect (27%). Higher usage frequency of the compound tense forms in the Insular and Western dialects is, therefore, at least partly related to their increased use in negative sentences.

Dialect	Affirmative	Negative	Negation (%)
Insular	912	811	47
Western	838	319	28
Tartu	334	94	22
Eastern	266	73	22
Mid	676	184	21
Coastal	257	59	19
Võru	397	89	18
Mulgi	516	101	16
Northeastern	166	27	14
Seto	109	14	11

Table 3. The proportion of negative sentences among the uses of compound past tense forms.

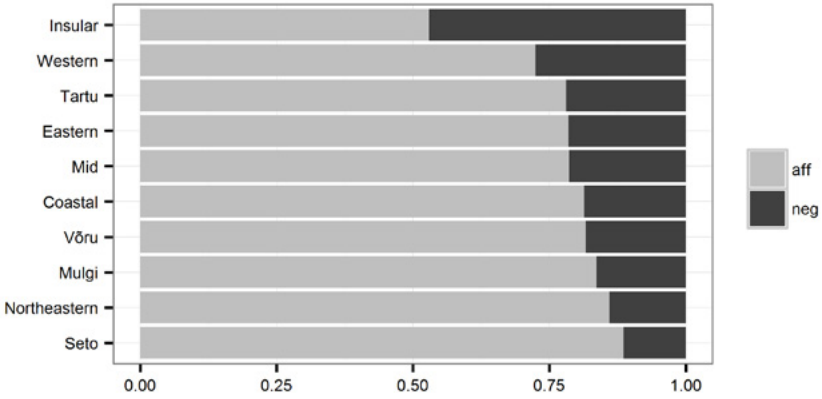


Figure 2. Proportions of affirmative and negative compound tense forms in the dialect data.

Closer inspection of negated perfect cases in the Insular dialect shows that the use of this construction has been functionally extended and does not express only the typical meaning for perfects. In example (14), the first two affirmative clauses are in the simple past, while for the following negated clause the perfect is used instead of the simple past (as it is in Standard Estonian). In this context, however, the negated clause may be interpreted also as an experiential perfect, but such an interpretation is not evident in all cases. In example (15), the first affirmative clause (*kis juuwa täis mees oli* 'who was a drunken man') employs the simple past, while in the following negative clause, the perfect is used (*see pole sääld tulngid* 'that didn't leave from there'), followed by the affirmative clause again in the simple past (*see=li teise umiguni veel ja* 'that was until next morning'). There is no evident functional reason (other than negation) to use perfect in the second clause; in standard Estonian, in this context the simple past would be used. Thus it seems that for negation, often the perfect is preferred in contexts where (at least in standard Estonian) the simple past occurs.

(14) Insular, Jämaja Parish

moo *isa* *surr-i* *nuoorelt* *ää (.)* *naad* *ela-sid* *neli*
I.GEN father die-PST.3SG young PFV (.) they live-PST.3PL four
oasta-d *kuoos (...)* *rohkem* *naad* ***põle*** ***ela-n***
year-PRT together more they NEG.be.CNG live-APP
'My father died young. They lived four years
together, they didn't live any more'

(15) Insular, Emmaste Parish

kiss *juuwa* *täis* *mees* *ol-i (.)* *see* ***pole*** *sääld*
who drink.INF full man be-PST.3SG this NEG.be.CNG there_from
tul-n-gid *(.)* *see= li* *teise* *umigu-ni* *veel* *ja*
come-APP-CLI this be-PST.3SG another morning-TER still and
'Who was a drunken man, that didn't leave from
there, that was until next morning'

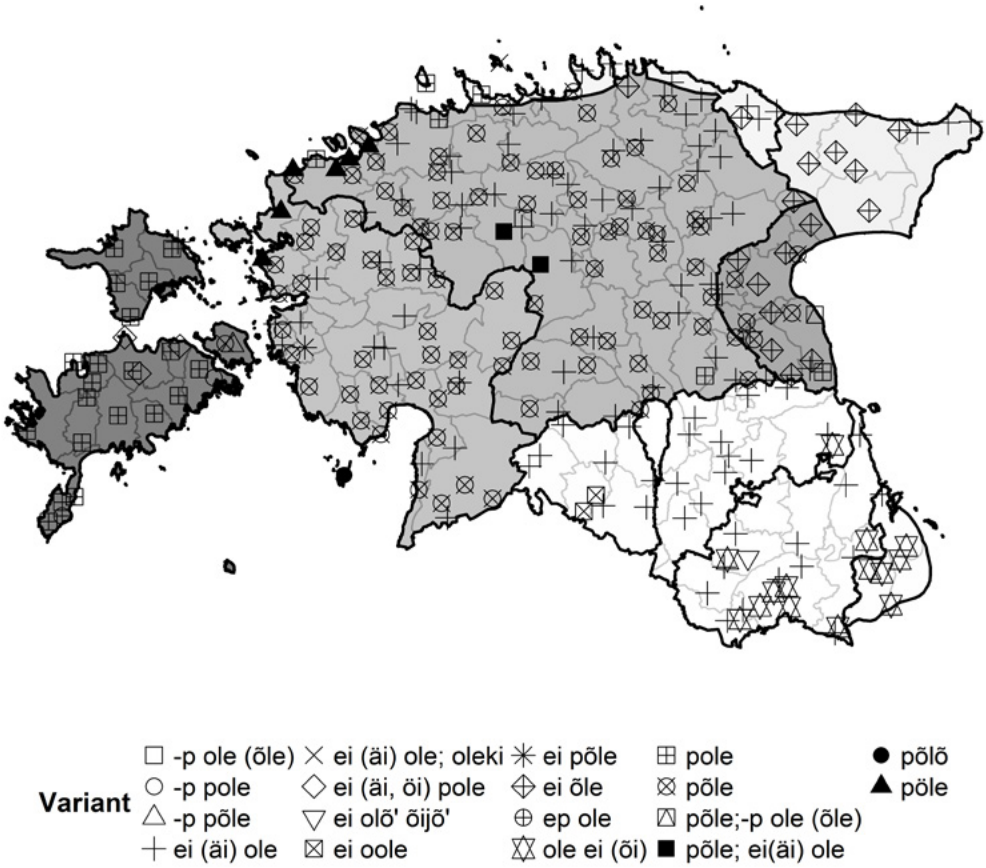
The negative perfect also alternates with a bare participle, which is used as an evidential strategy in Estonian. In example (16), in the first clause only a bare participle appears, but the second clause (syntactically coordinated with the first clause) uses the perfect tense. The bare participle does not have a negative counterpart of its own; in the negative clause, however, the simple past could be expected instead of the negative perfect in standard Estonian.

(16) Insular, Kihnu Parish

aga *rahvas* *ol-n* *ikka* *jõrmu* *täüs* *ja*
but people be-APP PTCL fear.PRT full and
põlõ ***toeht-n*** *mitte* *uata-ma* *tulla*
NEG.be.CNG may-APP NEG look-SUP come.INF
'But people were very afraid and were not allowed to come and look'

It is not clear why the compound tenses are used more often in negated sentences in the Insular and Western dialects, or whether these tendencies are driven by contact-induced extension of the compound tense functions or by a language-internal development. Information about clear parallels in the main contact language – Swedish – is not available to the authors. However, it is worth noting that the vast majority of the negative compound tense forms in the Insular dialect are formed with the synthetic negation word *pole* 'is not' (as in examples 14–15)

(703 cases out of the 811 negated sentences, i.e., 87%); in comparison, the corresponding proportions in the Western dialect were 123 out of 319 (39%). Koit (1963) has referred to the negation word *pole* (historically: *ep* 'not' *ole* be.CNG) as a highly productive means of negation in the Insular dialect. Also, Map 6, combining Andrus Saareste's (1955) data on the alternation of *ei ole* and *pole* in Estonian dialects (displayed utilizing symbols) and dialect corpus frequency data on the occurrences of *pole* from a study on analytic and synthetic parallel constructions (Uiboaed & Pilvik 2014), show the clear preference for *pole* in the Insular dialect.



Map 6. The spread and frequency of *pole* and *ei ole* in Estonian dialects.

This leads us to assume that in the Insular dialect, the compound *pole* + *-nud* has generalized into the regular negative construction referring to the general past also in contexts where otherwise the simple past would be used. The former past tense negative construction in the Insular dialect is formed with the negation word *es* and the connegative verb form (e.g., *es ole* 'was not'), but its use has somewhat diminished (especially in Hiiumaa, Koit 1963), and given way to the general North Estonian *ei* 'no(t)' + *-nud* participle compound, e.g., *ei ol-nud* 'was not' (Lonn & Niit 2002) and *pole* + *-nud* participle (*pole ol-nud* 'has not been; was not'). One can argue that the changes of negation in the tense system have lead the *pole* + *-nud* compound to express general past tense negation. This has potentially given rise to the increased frequency of compound tense usage in the Insular dialect. One might also be tempted to connect higher usage frequency of the negative compound tenses to the lower proportion of pluperfects in the data, as seems to be the case at least for the Insular dialect where the proportion of pluperfect forms is relatively low compared to other dialects (as indicated in Table 2 and Figure 1). However, negative pluperfect forms (e.g., *polnud* / *ei olnud ol-nud* 'had not been') can be considered a rather rare and marginal phenomenon in both formal and informal uses of Estonian. Due to the rarity of these forms in actual language use, it is impossible to argue that the amount of pluperfect forms in the data is also responsible for the proportions of affirmative and negative sentences. In our data, altogether there were only 21 negative indicative pluperfect forms, with only 4 of them from the Insular dialect. Even when we take into account all possible moods (in addition to the indicative), the corresponding figures rise to a mere 85 and 14; i.e., negative pluperfect forms are still infrequent in the data. The majority of negative compound tense forms, therefore, consists of indicative perfect forms with dialectal differences still holding.

6. Summary

The perfect and pluperfect as contact-induced categories exist in all Estonian dialects, but differences emerge in their usage frequency. By examining the data in the Corpus of Estonian Dialects on three levels (by dialects, parishes, and villages), we found that high compound tense frequency is particularly characteristic of the Insular and Mulgi dialects (especially in Karksi Parish), low frequency, in turn, is common in the Seto and Northeastern dialects. These areas have also had intensive local language contacts: the Insular dialect with Swedish, the Mulgi dialect with Latvian, the Seto dialect with Russian, and the Northeastern dialect area with other Finnic languages and, to some extent, also with Russian. Thus, the contact language influences are at least partially reflected in the frequency: contacts with languages, which make extensive use of similar categories (Swedish, Latvian) have reinforced the use of compound past tenses in contact regions; languages, which either only have simple past forms (Russian) or in which the compound tense forms exist, but are less used (Votic, to some extent also Finnish and Ingrian), have had an opposite effect on the usage frequency of the perfect and pluperfect in Estonian dialects. Although German has also heavily influenced all levels of the Estonian language for a long period of time, it is difficult to assess the impact of German on a more local level; it can only be stated that due to historical reasons, the Seto dialect is less influenced by German.

When comparing the frequencies of compound tenses in different dialects to those of other constructions consisting of a finite and a non-finite verb form (Uiboaed 2013; Uiboaed et al. 2013), the former do not show the distinctions, which are otherwise strong and clear-cut between the western and eastern (including southern) dialects.

The high frequency regions, the Insular and Mulgi dialects in particular, are characterized also by additional functions of compound tense constructions. In the Insular dialect, higher frequency is related to a large proportion of negated sentences. The majority of these sentences are negated using a synthetic form *pole* (instead of the analytic *ei ole* ‘is not’), which has become entrenched as a general negation word; instead of the simple past, the perfect is used in negative contexts, often receiving a general past tense reading. Texts from Karksi

Parish in the Mulgi region are characterized by an abundance of quotative uses of the compound past tenses which is mainly reflected in the frequency of the pluperfect. The same tendency is also present though to a lesser extent in other South Estonian dialects (Tartu and Võru). This may be related to the large amount of narrative texts in the research material as well as the overall tendency towards expressing reported evidentiality more often; the latter may also be induced by contacts with Latvian.

Acknowledgements

This study was supported by the Estonian Research Council (project PUT90 "Estonian Dialect syntax") and by the (European Union) European Regional Development Fund (Centre of Excellence in Estonian Studies).

Abbreviations

1, 2, 3	person	M	masculine
ADE	adessive	NEG	negation
ALL	allative	NT	neuter
APP	active past participle	PFV	perfective
CED	Corpus of Estonian Dialects	PL	plural
CLI	clitic	PP	past participle
CNG	connegative verb form	PPP	passive past participle
DEF	definite	PREP	preposition
ELA	elative	PRS	present tense
F	feminine	PRT	partitive
GEN	genitive	PST	past
ILL	illative	PTCL	particle
IMP	imperative	PTCP	participle
INE	inessive	SG	singular
INF	infinitive	SUP	supine
INS	instrumental	TER	terminative

References

Alvre, Paul 2000: Kirderannikumurde ja vadja keele ühisjooni [Common features of the Estonian Northeastern Coastal dialect and the Votic language]. – Jüri Viikberg (ed.), *Inter dialectos nominaque. Pühendusteos Mari Mustale 11. novembril 2000*. Eesti Keele Instituudi Toimetised 7. Tallinn: Eesti Keele Sihtasutus. 1–13.

Ariste, Paul 1956: Läänemere keelte kujunemine ja vanem arenemisjärk [The development and the earlier stage of the Finnic languages]. – H. Moora (ed.), *Eesti rahva etnilisest ajaloost*. Tallinn: Eesti Riiklik Kirjastus. 5–23.

—— 1962: Teiste läänemere keelte elemente Vaivara murrakuis [Elements of other Finnic languages in the Vaivara dialects]. – *Emakeele Seltsi aastaraamat* VIII: 11–18. Tallinn.

Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert & Arne Zechsel 2010: Cognitive Corpus Linguistics: five points of debate on current theory and methodology. – *Corpora* 5(1): 1–27.

Barlow, Michael & Suzanne Kemmer 2000: *Usage-based models of language*. CSLI publications Stanford.

Björklöf, Sofia 2012: *Viron rantamurteen länsiryhmän sanaston alkuperä suomalaislainojen valossa* [The origin of the vocabulary of the western subgroup of the Estonian Coast dialect in light of Finnish loanwords. MA thesis]. Pro gradu -tutkielma. Helsingin yliopiston suomen kielen, suomalais-ugrilaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos. Available at: ‹http://urn.fi/URN:NBN:fi-fe2012121410304›

—— 2018: Uusia lainaetymologioita viron koillisrannikon rantamurteen länsiryhmässä [New loan etymologies in the western subgroup of the Estonian Coast dialect of the northeastern coast]. – Sampsa Holopainen & Janne Saarikivi (eds), *Περὶ ὀρθότητος ἐτύμων. Uusiutuva uralilainen etymologia*. Uralica Helsingiensia 11. Helsinki: Suomalais-Ugrilainen Seura. 357–487. Available at: ‹https://journal.fi/uralicahelsingiensia/issue/view/uh11/489›

—— 2019: Mutual contacts and lexical relations among the Finnic varieties of western Ingria and northeastern Estonia. – Sofia Björklöf & Santra Jantunen (eds), *Multilingual Finnic. Language contact and change*. Uralica Helsingiensia 14. Helsinki: Finno-Ugrian Society. 89–153. Available at: ‹https://doi.org/10.33341/uh.85034›

Bybee, Joan 2010: *Language, usage and cognition*. Cambridge University Press.

Comrie, Bernard 1976: *Aspect. An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Dahl, Östen 1995: The Tense System of Swedish. – Rolf Thieroff (ed.), *Tense Systems in European Languages II*. De Gruyter. 59–68.

—— 2014: The perfect map: Investigating the cross-linguistic distribution of TAME categories in a parallel corpus. – Benedikt Szmrecsanyi & Bernhard Wälchli (eds), *Aggregating Dialectology, Typology, and Register Analysis*. Berlin: De Gruyter. 268–289.

Dahl, Östen & Maria Koptjevskaja-Tamm (eds) 2001: *Circum-Baltic Languages. Typology and Contact*. Volume 1: Past and Present, Volume 2: Grammar and Typology. Studies in Language Companion Series 54–55. Amsterdam: John Benjamins Publishing Company.

Divjak, Dagmar & Stefan Th. Gries 2012: *Frequency effects in language representation*. Vol. 2. Walter de Gruyter.

EKI 2014 = *Eesti Keele Instituudi kohanimeandmebaasi kihelkonnapiiride andmestik* [Map of the Place Name Database created by the Institute of Estonian Language].

Erelt, Mati 2013: *Eesti keele lauseõpetus. Sissejuhatus. Öeldis* [Estonian syntax. Introduction. Predicate]. Tartu Ülikooli eesti keele osakonna preprindid 4. Tartu.

Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael & Silvi Vare 1995: *Eesti keele grammatika I. Morfoloogia. Sõnamoodustus* [Estonian grammar I. Morphology. Word formation]. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.

Gries, Stefan Th. & Dagmar Divjak 2012: *Frequency effects in language learning and processing*. Vol. 1. Walter de Gruyter.

Haspelmath, Martin 2001: The European Linguistic Area: Standard Average European. – Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds), *Language Typology and Language Universals / Sprachtypologie und sprachliche Universalien / La Typologie des langues et les universaux linguistiques: An International Handbook / Ein internationales Handbuch / Manuel international, 1–2*. Berlin: de Gruyter. 1492–1510.

Heine, Bernd & Tanja Kuteva 2005: *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.

—— 2006: *The Changing Languages of Europe*. Oxford: Oxford University Press.

Ikola, Osmo 1960: Perfektin ja pluskvamperfektin synnystä [On the origin of the perfect and pluperfect]. – *Virittäjä* 64: 364–368. Available at: ‹https://journal.fi/virittaja/article/view/34251›

Juhkam, Evi 1983: Rootsipärane objekt eesti murretes [The Swedish-origin object in Estonian dialects]. – *Keel ja Kirjandus* XXVI (3): 122–125.

—— 1992: Eestirootslastest [On the Estonian Swedes]. – *Keel ja Kirjandus* XXXV (7): 394–402.

—— 1998: *Eesti-rootsi murdekontaktid* [Estonian–Swedish dialect contacts. Unpublished MA thesis]. Tallinn: Tallinn University, Faculty of Philology.

—— 2012: *Harju-Madise murrak* [The Harju-Madise dialect]. Mari-Liis Kalvik & Helmi Neetar (eds). Tallinn: Eesti Keele Sihtasutus.

Kalnača, Andra 2015: *A Typological Perspective on Latvian Grammar*. De Gruyter Mouton.

Kask, Arnold 1984: *Eesti murded ja kirjakeel* [Estonian dialects and the standard language]. Tallinn: Valgus.

Kehayov, Petar 2008: *An areal-typological perspective to evidentiality: the cases of the Balkan and Baltic linguistic areas*. Dissertationes Linguisticae Universitatis Tartuensis 10. Tartu: Tartu Ülikooli Kirjastus. Available at: ‹http://hdl.handle.net/10062/6524›

Klaas-Lang, Birute & Miina Norvik 2014: Balti areaali tüpoloogilisi sarnasusi morfosüntaksi valdkonnas [Typological similarities of morphosyntax in the Baltic linguistic area]. – *Keel ja Kirjandus* LVII (08–09): 590–608. Available at: ‹http://kjk.eki.ee/ee/issues/2014/8-9/535›

Koit, Enn 1963: Eitus saarte murdes [Negation in the Insular dialect]. – *Nonaginta. Johannes Voldemar Veski 90. sünnipäevaks 27. juunil 1963*. Emakeele Seltsi Toimetised 6. Tallinn. 136–147.

Koptjevskaja-Tamm, Maria & Bernhard Wälchli 2001: The Circum-Baltic Languages: An Areal-Typological Approach. – Östen Dahl & Maria Koptjevskaja-Tamm (eds), *Circum-Baltic Languages. Typology and Contact. Volume 2: Grammar and Typology*. Studies in Language Companion Series 55. Amsterdam: John Benjamins Publishing Company. 615–750.

Kortmann, Bernd 2010: Areal variation in syntax. – Peter Auer & Jürgen E. Schmidt (eds), *Language and Space: An International Handbook of Linguistic Variation. Theories and Methods*. Berlin: De Gruyter. 837–864.

Laakso, Johanna 2001: The Finnic languages. – Östen Dahl & Maria Koptjevskaja-Tamm (eds), *Circum-Baltic Languages. Typology and Contact. Volume 1: Past and Present*. Amsterdam: John Benjamins Publishing Company. 179–212.

Laanest, Arvo 1975: *Sissejuhatus läänemeresoome keeltesse* [Introduction to the Finnic languages]. Tallinn: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Larsson, Ida 2009: *Participles in Time. The Development of the Perfect Tense in Swedish*. Nordistica Gothoburgensia 29. Göteborg: Göteborgs Universitet.

Lindstedt, Jouko 2000: The perfect-aspectual, temporal and evidential. – *Empirical Approaches to Language Typology* (6): 365–384.

Lindström, Liina 2015: Subjecthood of the agent argument in Estonian passive constructions. – Marja-Liisa Helasvuo & Tuomas Huumo (eds), *Subjects in Constructions – Canonical and Non-Canonical*. Constructional Approaches to Language 16. Amsterdam: John Benjamins Publishing Company. 141–173.

—— forthcoming: Seto lause põhijooned [Basic features of the Seto clause]. – Karl Pajusalu, Andreas Kalkun & Ergo-Hart Västrik (eds), *Setomaa III: Keel, folkloor ja tänapäeva kultuur*. Seto Instituut.

Lindström, Liina, Pärtel Lippus & Tuuli Tuisk 2019: The online database of the University of Tartu Archives of Estonian Dialects and Kindred Languages and the Corpus of Estonian Dialects. – Sofia Björklöf & Santra Jantunen (eds), *Multilingual Finnic. Language contact and change*. Uralica Helsingiensia 14. Helsinki: Finno-Ugrian Society. 327–350. Available at: ‹https://doi.org/10.33341/uh.85040›

Lindström, Liina, Maarja-Liisa Pilvik, Mirjam Ruutma & Kristel Uiboaed 2015: Mineviku liitaegade kasutusest eesti murretes keelekontaktide valguses [Abstract: The use of the compound past tenses in Estonian dialects in the light of language contacts]. – Jüvä Sullõv (ed.), *Aig õdagumeresoomõ keelin. Aeg läänemeresoome keeltes. Time and tense in Finnic languages*. Publications of Võro Institute 29. Võro: Võro Instituut´. 40–71. Available at: ‹https://wi.ee/et/voru-instituudi-toimetised-28/›

Lindström, Liina & Piret Toomet 2000: Eesti suuliste narratiivide keelelisi erijooni [Linguistic features of oral narratives in Estonian]. – Tiit Hennoste (ed.), *Eesti keele allkeeled*. Tartu Ülikooli Eesti Keele Õppetooli Toimetised 16. Tartu. 174–203.

Lindström, Liina & Ilona Tragel 2010: The possessive perfect construction in Estonian. – *Folia Linguistica: Acta Societatis Linguisticae Europaeae* 44 (2): 371–400.

Lindström, Liina, Kristel Uiboaed & Virve-Anneli Vihman 2014: Varieerumine tarvis-/vaja-konstruktsioonides keelekontaktide valguses [Variation of the tarvis/vaja-constructions in light of language contacts]. – *Keel ja Kirjandus* LVII (08–09): 609–630. Available at: ‹http://kjk.eki.ee/ee/issues/2014/8-9/536›

Lonn, Varje & Ellen Niit 2002: *Saarte murde tekstid* [Insular dialect texts]. Eesti murded VII. Tallinn: Eesti Keele Instituut.

Markus, Elena B. & Fedor I. Rozhanskiy 2011: *Sovremennyj vodskij jazyk. Teksty i grammatičeskij očerk. V 2-h tomah* [Contemporary Votic language. Texts and grammar. In 2 volumes]. Sankt-Peterburg: Nestor-Istorija.

Metslang, Helle 1993: Verbitarind ajatähendust väljendamas [Verbal construction in expressing temporality]. – *Virittäjä* 97: 203–221. Available at: ‹https://journal.fi/virittaja/article/view/38574›

—— 1997: On the use of the Estonian past tense forms through the last one hundred years. – Mati Erelt (ed.), *Estonian: typological studies II*. Publications of the Department of Estonian of the University of Tartu 8. Tartu. 98–145.

—— 2009: Estonian grammar between Finnic and SAE: some comparisons. – *Sprachtypologie und Universalienforschung* 62, 1/2: 49–71.

Muižniece, Liena, Helle Metslang & Karl Pajusalu 1999: Past participle finitization in Estonian and Latvian. – Mati Erelt (ed.), *Estonian. Typological studies III*. Publications of the Department of Estonian of the University of Tartu 11. Tartu. 128–157.

Must, Mari 1965: Vene-eesti kakskeelsus Kirde-Eestis [Russian–Estonian bilingualism in northeastern Estonia]. – H. Moora & L. Jaanits (eds), *Slaavi-läänemeresoome suhete ajaloost*. Tallinn: Eesti Raamat. 107–131.

—— 2000: *Vene laensõnad eesti murretes* [Russian loanwords in Estonian dialects]. Tallinn: Eesti Keele Sihtasutus.

Norvik, Piret 2000: Kirderannikumurde sõnavaralisi suhteid [Lexical relations in the Estonian Northeastern Coastal dialect]. – Jüri Viikberg (ed.), *Inter dialectos nominaque. Pühendusteos Mari Mustale 11. novembril 2000*. Eesti Keele Instituudi Toimetised 7. Tallinn: Eesti Keele Sihtasutus. 184–201.

Pajusalu, Karl, Tiit Hennoste, Ellen Niit, Peeter Päll & Jüri Viikberg 2009: *Eesti murded ja kohanimed* [Estonian dialects and place names]. 2. ed. Tallinn: Eesti Keele Sihtasutus.

Pall, Valdek 1982: Idamurde vahekorrast naabermurretega ja vadja keelega [The Relation of the Estonian Eastern dialect with neighboring dialects and the Votic language]. – *Keel ja Kirjandus* XXV (5): 246–251.

Pilvik, Maarja-Liisa & Kristel Uiboaed 2014: Grammatical evidentiality in Estonian: areal evidence of variation. Paper presented at the Syntax of the World's Languages VI, Pavia.

R Development Core Team 2013: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Available at: ‹http://www.R-project.org/›

Saareste, Andrus 1955: *Väike eesti murdeatlas. Petit atlas des parlers estoniens* [Small dialectal atlas of Estonian]. Uppsala: Kungliga Gustav Adolfs Akademien.

Serebrennikov, B. A. 1959: Pluskvamperfekti ja perfekti päritolu probleemist läänemeresoome keeltes [On the problem of the origin of the pluperfect and perfect in the Finnic languages]. – *Emakeele Seltsi aastaraamat* IV (1958): 249–255.

Seržant, Ilja A. 2012: The so-called possessive perfect in North Russian and the Circum-Baltic area. A diachronic and areal account. – *Lingua* 122: 356–385.

—— 2015: The Independent Partitive as an Eastern Circum-Baltic isogloss. – *Journal of Language Contact* (8): 341–418.

Siegl, Florian & Petar Kehayov 2006: The evidential past participle in Estonian reconsidered. – *Études finno-ougriennes* 38: 75–117.

Söderman, Tiina 1996: *Lexical characteristics of the Estonian North Eastern Coastal dialect*. [PhD thesis.] Studia Uralica Upsaliensia 24. Acta Universitatis Upsaliensis. Uppsala.

Szmrecsanyi, Benedikt 2013: *Grammatical variation in British English dialects: a study in corpus-based dialectometry*. Cambridge: Cambridge University Press.

Tommola, Hannu 2000: On the perfect in North Slavic. – Östen Dahl (ed.), *Tense and Aspect in the Languages of Europe*. Empirical Approaches to Language Typology 20. Berlin: De Gruyter. 441–478.

Uiboaed, Kristel 2010: Ühendverbid eesti murretes [Compound verbs in Estonian dialects]. – *Keel ja Kirjandus* LIII (1): 17–36.

—— 2013: *Verbiühendid eesti murretes* [Verb compounds in Estonian dialects]. Dissertationes Philologiae Estonicae Universitatis Tartuensis 34. Tartu: Tartu Ülikooli Kirjastus. Available at: ‹http://hdl.handle.net/10062/34499›

Uiboaed, Kristel, Cornelius Hasselblatt, Liina Lindström, Kadri Muischnek & John Nerbonne 2013: Variation of verbal constructions in Estonian dialects. – *Literary & Linguistic Computing* 28 (1): 42–62.

Uiboaed, Kristel & Maarja-Liisa Pilvik 2014: Exploring the borders of analyticity and syntheticity. Paper presented at the conference Methods in Dialectology XV, Groningen, August 11–15 2014.

Univere, Aili 1988: Idamurre [The Estonian Eastern dialect]. – *Emakeele Seltsi aastaraamat* 32 (1986): 59–93.

Vaba, Lembit 1997: *Uurimusi läti-eesti keelesuhetest* [Studies on Latvian–Estonian language relations]. Tallinn – Tampere: Eesti Keele Instituut & Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.

—— 2011: Kuidas läti-eesti keelekontakt on mõjutanud eesti murdekeele grammatikat ja sõnamoodustust [How Latvian–Estonian language contact has influenced grammar and word formation in Estonian dialects]. – *Emakeele Seltsi aastaraamat* 56 (2010): 204–246.

Wälchli, Bernhard 2011: The Circum-Baltic languages. – Bernd Kortmann & Johann van der Auwera (eds), *The Languages and Linguistics of Europe. A Comprehensive Guide*. De Gruyter Mouton. 325–340.

Perfekti ja pluskvamperfekti kasutusest eesti murretes: sagedus ja keelekontaktid

Liina Lindström, Maarja-Liisa Pilvik, Mirjam Ruutma & Kristel Uiboaed

Artiklis vaadeldakse perfekti ja pluskvamperfekti esinemissagedust eesti murretes korpusandmete põhjal. Murretevahelistele sageduserinevustele otsitakse seletusi eelkõige lokaalsetest keelekontaktidest. Liitajad arvatakse olevat eesti keelde tekkinud balti keelte mõjul, nende edasiste funktsioonide arengut on mõjutanud peamiselt germaani keeled (Serebrennikov 1959, Ikola 1960). Eesti murrete puhul on aga liitaegade kasutusele mõju avaldanud lisaks ka lokaalsed kontaktid rootsi, vene, läti ja läänemeresoome keeltega (peamiselt vadja, soome ja isuri keelega). Eesti murrete korpuse andmete põhjal kasutatakse perfekti ja pluskvamperfekti teiste murretega võrreldes eriti sagedasti saarte ja Mulgi murdes, kasutussagedus on aga oluliselt väiksem Seto ja kirdemurdes. Sagedused peegeldavad vähemalt osaliselt keelekontaktide mõju: kontaktid keeltega, kus samad kategooriad on olemas ning laialdaselt kasutuses (rootsi ja läti keel), suurendavad liitaegade kasutussagedust ka kontaktalal (*resp.* saarte ja Mulgi murdes); tihedad kontaktid keeltega, milles vastav kategooria puudub (vene) või on küll olemas, ent harvemini kasutuses (vadja ja isuri keel), aga pigem vähendavad liitaegade kasutussagedust ning kontaktalal võivad liitaegade asemel olla enam levinud teised väljendusvahendid (nt lihtminevik).

Lisaks keelekontaktide mõjule tulevad sagedusi uurides esile ka piirkondlikud funktsionaalsed erinevused liitaegade kasutuses. Saarte murde andmestikus oli eitavate lausete proportsioon liitaegade kasutusjuhtudest märkimisväärselt kõrgem kui teistes murretes. Valdav enamik eitavatest lausetest olid omakorda *pole*-eitussõnaga moodustatud perfektivormid, mis annab alust oletada, et *pole* + *nud*-partitsiibi ühend on saarte murdes kujunenud üldminevikku tähistavaks eitusvormiks, kuna seda kasutatakse ka kontekstides, milles vastav jaatav vorm oleks lihtminevikus. Mulgi murde suur liitaegade kasutussagedus on seotud kvotatiivsete pluskvamperfekti vormide suure osakaaluga andmestikus. See võib tuleneda sellest, et Mulgi murde materjalis esines keskmisest enam narratiivseid tekste (muistendeid, pärimusi), ent ka kontaktimõjulisest evidentsiaalsete strateegiate laialdasest kasutamisest kontaktalal läti keelega.