ESUKA – JEFUL 2023, 14–2: 297–320

TÄITSA LAMBIST:[1] EVIDENCE OF ONGOING CASE GRAMMATICALISATION FROM TEENAGERS' AND ADULTS' SPOKEN LANGUAGE

Mari Aigro, Virve-Anneli Vihman
University of Tartu, EE
mari.aigro@ut.ee, virve.vihman@ut.ee

Abstract. Eleven out of fourteen Estonian nominal cases are generally regarded as semantic cases, rather than grammatical, but some are more grammaticalised (i.e. less semantic) than others. These can be said to be (semi-)grammatical cases, carrying both semantic and syntactic functions. This study focusses on elative and allative case, which stand out among semantic cases in terms of multifunctionality and a higher degree of grammaticalisation. We investigate whether we can find evidence that the grammaticalisation of these cases is ongoing through an apparent time study. While previous approaches have used written data to investigate diachronic change in cases, the present study uses spontaneous spoken data by teenagers and adults to study the nature of grammatical change. This is the first study to diachronically investigate case characteristics related to multifunctionality and grammaticalisation through a quantitative distributional approach.

Keywords: case, elative, allative, grammaticalisation, spoken corpora, youth language

DOI: https://doi.org/10.12697/jeful.2023.14.2.12

1. Introduction

Many core assumptions in the field of morphological case have been contested in recent decades. One major hypothesis which seems not to hold in the simple way in which it was originally proposed has to do with the bifurcation of case systems, according to which some cases are used for semantic (typically, adverbial) functions and others for core grammatical (typically subject and object) functions. Grammatical case

1 *Täitsa lambi-st* (completely lamp-ELA.SG), 'totally randomly', lit. 'totally from [a/the] lamp'. This is a colloquial phrase used by contemporary Estonian teenagers.

is often viewed as a tool for expressing a dependency (or government) relationship between a head and its arguments, e.g. in Estonian, genitive marking an object (*Priidu*) in (1a). Semantic case, on the other hand, is seen as having more semantic content and being less closely connected to the predicate structure, e.g. Estonian elative case, marking a separative spatial relation in (1a, *kodust* 'from home') or (1b, *koolist* 'from school').

(1) a. *viska-s* ***Priidu*** ***kodu-st*** *välja*
throw-PST.3SG Priit.GEN home-ELA out
'[S/he] threw Priit out of his home.' (M, 16)[2]
b. *kõndi-s* ***kooli-st*** *koju* *vaata*
walk-PST.3SG school-ELA home.ILL look.IMP.SG
'S/he walked home from school, you see.' (F, 15)

However, both grammatical and semantic cases can be multifunctional. For instance, in addition to spatial adverbials, as in (1), the Estonian elative also marks participants of events described by verbs, e.g. *sõltuma* 'depend on' in (2):

(2) ***selle-st*** ***välis-materjali-st*** *sõltu-b* *palli* *hind*
this-ELA external-material-ELA depend-PRS.3SG ball.GEN price
'The price of the ball depends on its outer material.' (M, 12)

There is some debate regarding whether semantic cases can be seen as *truly* multifunctional. While it is not a conceptual issue for grammatical cases to have semantic as well as grammatical functions (e.g. the Estonian partitive case signalling a partitive quantitative function, *pool õuna* 'half an apple', in addition to marking objects), semantic cases are often considered unable to have 'true' grammatical functions (Blake 2001). Hence, the genitive argument *Priidu* in (1a) is viewed as inherently different from the elative *välismaterjalist* in (2) the elative case being selected by the verb (Erelt 1989; Erelt 2003; Veismann, Erelt & Metslang 2017).

Aigro (2022), following Kuryłowicz (1964), argues that there are no empirical grounds for regarding grammatical uses of elative (as in 2) and genitive (in 1a) as rendering qualitatively distinct argument types.

2 All examples originate from the Estonian Teen Speech Corpus, unless otherwise stated. The speaker's gender and age is given after each corpus example.

The only systematic distinction between elative and genitive arguments is that elative arguments are much more likely to co-occur with verbs expressing non-dynamic events and states (Aigro 2022). However, a connection to state verbs cannot be viewed as an indication of weaker or different argument status, because argument status cannot be directly linked to predicate semantics; that would lead to, for instance, partitive direct objects of states being excluded from the object category as well (e.g. *mõistan probleemi* 'I understand the problem'). Furthermore, Aigro (2022) also finds a lack of distinction between the cognitive status of arguments marked in spatial cases or partitive case. Estonian speakers participating in a judgement task did not differ in their judgements of these relations. Based on these results, this study assumes the Estonian case system to have 'true' functional overlap in the sense that cases labelled as 'semantic' may have grammatical functions essentially analogous to those of cases seen as 'grammatical' – it just happens less frequently.

Case multifunctionality may be addressed by looking at its diachronic origins; the framework of case grammaticalisation gives us a useful tool for approaching multifunctionality from this perspective (Lehmann 1985; Hopper & Traugott 2003; Heine 2008; Lestrade 2010; Aigro 2022). Entirely semantic cases are less grammaticalised than grammatical cases, often serving as the source categories from which grammatical cases arise (Heine 1990; Heine, Claudi & Hünnemeyer 1991; Larjavaara 1991; Heine & Kuteva 2002; Rice & Kabata 2007). The greater degree of grammaticalisation for grammatical cases can be seen in their being more bleached of semantics. For instance, genitive markers in Estonian tend to mark more abstract, semantically bleached relations (such as direct objects or possession) than abessive ('without') markers.

In this view, multifunctional 'semantic' cases which mark both semantically specific relations (elative with spatial reference, as in 1b) and abstract, semantically bleached relations (elative marking an argument, as in 2) may actually constitute semigrammaticalised cases in the interim stages of a grammaticalisation pathway leading from semantic to grammatical case. Broadly, this is in line with the framework of Nichols (1983, 1984), as a case may theoretically be multifunctional, with no primary or secondary functions, when it is on this grammaticalisation

pathway. It may mark true syntactic relations (2), even if such relations do not constitute all, or even the majority of uses of that case.

Hence, each case can be described in terms of its (synchronic) degree of grammaticalisation. Traditionally, grammaticalisation has been studied through diachronic approaches, using historical texts from different periods. Unfortunately, this line of research requires a longer period of documentation than is available for Estonian cases.

However, one may also use synchronic data as a tool for investigating grammaticalisation (Aigro 2022). Grammaticalisation has implications for the expected distribution of lexemes occurring with various case affixes. Simply put, the more grammaticalised the case, the wider range of lexemes it occurs with. One may therefore outline the grammaticalisation of cases by investigating their quantitative distribution in a language corpus.

Operationalised in this way, the synchronic approach opens up promising possibilities for the study of Estonian cases. Cases are viewed as language-specific phenomena with a potentially wide range of functional profiles. Determining the role of individual cases in the syntax of a given language requires in-depth investigation of each case together with its distribution, in order to render a description of the synchronic case system and the possible gradience in grammaticality of case functions. Instead of assigning 'semantic' and 'grammatical' labels to cases, the researcher is now tasked with assessing and comparing their multifunctionality, the particular ways in which some of them have assumed grammatical functions and the factors leading to the increased use of a case in grammatical function. By measuring synchronic variables, Aigro (2022) showed that elative and allative are much more grammaticalised than other spatial cases, outlining them as multifunctional between semantic (adjunct-marking) and syntactic (argument-marking) domains.

In addition to the synchronic perspective, in which cases are described in terms of their relative degree of grammaticalisation at a certain period, case systems are also expected to vary in a diachronic perspective. From this perspective, the relevant question is whether a multifunctional case is mapped to its functions in a stable way or is in the midst of ongoing grammaticalisation and change. The availability of the former possibility has not always been accepted by grammaticalisation researchers: the notion that a case may be stably used with multiple functions over an extended period of time, and we need not expect it to further grammaticalise any time soon, is explored by Nichols and

Timberlake (1991). They note that the concept of grammaticalisation involves two assumptions: 1) that grammatical and non-grammatical phenomena are clearly distinct, and 2) that grammaticalisation refers to a change "towards ossification, towards idiomatization, towards a kind of semiotic entropy" (1991: 129). To counter this assumption, they describe a range of distinct, yet similarly grammaticalised usage contexts of the Russian instrumental case, outlining its multifunctionality in Old Russian. They note that instead of gradually becoming increasingly grammaticalised, the case is "remarkably stable over time", both phonologically and syntactically (Nichols & Timberlake 1991: 142). Hence, we do not assume that a multifunctional case always constitutes a temporary category on the functional pathway from semantic to grammatical. Cases need not be in the midst of teleological change, directed toward functionally uniform end points. Instead, conceptualising the active/passive nature of grammatical change as a variable opens up an essential dimension relevant to both synchronic and diachronic analyses. The dual questions of whether (and to what degree) cases are multifunctional and whether grammatical change is ongoing can be researched by means of comparing the language usage of speakers of different ages (see section 4).

2. Spatial cases in Estonian

The nominal inflection system of Estonian includes fourteen morphological cases, of which only partitive, genitive and nominative are viewed as grammatical cases marking syntactic roles. The rest constitute semantic cases that have been described as marking various adverbials (Erelt et al. 1995; Veismann, Erelt & Metslang 2017). Both grammatical and semantic cases exhibit multifunctionality. In addition to core arguments, as in (1a), genitive and partitive may mark adverbials, as in the genitive time phrase in (3):

(3) *nii* *igav* *ma* *pea-n* *lihtsalt* *nagu* *Käteka-t*
so boring 1SG must-PRS.1SG simply like Kätekas-PAR
vaata-ma ***terve*** ***päeva***
watch-INF whole.GEN day.GEN
'So boring I have to just like, watch Kätekas [*Kättemaksukontor*, a TV show] all day.' (F, 11)

Of the semantic cases, all spatial cases and comitative are regarded as the most multifunctional. Erelt, Erelt and Ross (2007) note five different functions for illative, six for inessive and thirteen for elative; for the external spatial cases, eight functions are listed for allative, seven for adessive and three for ablative. In this perspective, elative and allative are highlighted as the spatial cases with the widest range of functions (Nurka 2014; Lindström & Vihman 2017; Viht & Habicht 2019; Aigro 2022; Aigro & Vihman 2023). For instance, in addition to its concrete spatial function (1b) and argument-marking function (2), elative also marks partitive (4a) and topic (4b) functions.

(4) a. *üks* ***nei-st*** *taht-is* *mind* *kallista-da*
one 3PL-ELA want-PST.3SG 1SG.PAR hug-INF
'One of them wanted to hug me.' (F, 15)

b. mina *ja* *Tommi* *rääki-si-me* ***selle-st***
1SG and Tommi talk-PST-1PL this-ELA
'Tommi and I talked about this.' (F, 15)

Some spatial semantics is available in the interpretation of (4a), *neist* 'of them' indicating a group, defining a metaphorical space, from which one element is extracted. Example (4b), on the other hand, is bleached of both literal and metaphorical spatial semantics.

Allative functions include its spatial meaning 'onto' as well as marking the thematic roles of Recipients (5a), which have metaphorical spatial semantics (as the end points or Goals of the action), Experiencers (5b) which may also be conceptualised as metaphorical Goals of emotions or experience, as well as semantically non-spatial Themes (5c):

(5) a. *siis* *ta* *kirjuta-s* ***mu-lle***
then 3SG write-PST 1SG-ALL
'Then s/he wrote to me.' (M, 14)

b. *või-n* *joonista-da* *küll* ***mu-lle*** *täiega* *meeldi-b*
can-1SG draw-INF FOC 1SG-ALL really like-PRS.3SG
'I can draw, I really like it.' (F, 15)

c. *ta* *ei* *ole* *nagu* *avatud* ***millegi-le*** ***uue-le***
3SG NEG COP.CNG like open something-ALL new-ALL
'S/he's not, like, open to anything new.' (F, 13)

In an in-depth analysis of the six spatial cases in Estonian, Aigro (2022) shows that all six mark verbal arguments (Themes), but they do so to differing extents. Allative and elative are shown to mark arguments with the largest number of verbs, and they are shown to be the most grammaticalised spatial cases. Furthermore, it is empirically misleading to label either of them as a semantic case, because 81% of allative tokens and 51% of elative tokens in written language express non-spatial grammatical functions, marking bleached argument relations.[3] Based on a written language corpus, elative also stands out as the spatial case used with the widest range of functions (n = 10), while allative and adessive have the smallest number of functions used by speakers (n = 5).

In summary, Estonian has a functionally mixed case system. All fourteen cases exhibit some multifunctionality, with many expressing both semantic and grammatical functions (Erelt, Erelt & Ross 2007). For some of them, the semantic function dominates, e.g. illative occurs in spatial (*majja* 'into the house') or metaphorically spatial (*hinge* 'into the soul') functions in over 90% of its instances. However, the proportion of spatial functions is lower for the other spatial cases; for elative and allative, it is so low that they ought to be regarded as semigrammatical cases (Aigro 2022).

Nevertheless, several gaps remain in what we know about multifunctionality in Estonian cases; this study addresses two of these gaps. First, semigrammatical status has been described for elative and allative based on a corpus of written language (Aigro 2022). Spoken language data needs to be analysed as well, for a more comprehensive understanding of the way these two cases compare to grammatical cases in language usage. Second, there are no studies on the nature of diachronic change in these two cases. We do not know whether multifunctional cases in Estonian are in the midst of ongoing grammaticalisation, where speakers use them in an increasingly broad range of contexts over time, or whether they have settled in their multifunctional roles with no significant ongoing change. Our research questions set out to investigate these gaps.

3 These token proportions regard the allative Experiencer, Recipient and Beneficiary functions as grammatical functions (Aigro 2022: 138). They were excluded from the bleached argument set in that study, where they were categorised as encoding metaphorical (semantic) Goals of events. If these semantic roles are categorised as spatial functions, allative marks grammatical functions in 38% of its instances.

3. Research questions

This study seeks answers to the following two research questions (RQs):

> RQ1: How do the degrees of grammaticalisation of elative and allative compare to those of the conventionally grammatical cases, partitive and genitive, in spoken language? (Synchronic dimension)
>
> RQ2: How static or dynamic is the grammaticalisation of elative and allative? (Diachronic dimension)

RQ1 is answered by comparing the usage of elative and allative to partitive and genitive case. In this study, partitive and genitive take the role of a control group of sorts, because their predominant function as markers of syntactic relations gives reason to assume that they are also the most grammaticalised cases in Estonian. The method partly aligns with the method used in Aigro (2022), using the same basic productivity measures to assess grammaticalisation (see section 4.1).

For RQ2, instead of using diachronic corpora, we tackle the question using the sociolinguistic method of charting apparent time, investigating language change through examining the language usage of speakers of different ages. This study includes data from adult speech (Phonetic Corpus of Estonian Spontaneous Speech, Lippus et al. 2021) and teenagers' speech (TeKE Corpus, Mandel et al. 2022; Koreinik et al. 2023). "Most sociolinguists agree that adolescence is the focal point of linguistic innovation and change", due to incrementation or the active use of novel forms characteristic of adolescents (Tagliamonte 2016: 3). In the transition to adulthood, individuals are believed to dramatically decrease their use of novel forms, resulting in a more stable grammar as adults (Tagliamonte 2016: 5).

Hence, the linguistic behaviour of teenagers can be expected to differ from that of adults in incorporating more linguistic innovations, including novel lexemes, novel inflected lexeme combinations and novel case functions. Such innovations also underlie the concept of case grammaticalisation, as the process is driven by an increased proportion of novel contexts, but also bleached semantics, which in its turn facilitates the introduction of an even wider range of novel contexts. Teenagers' language usage provides a context in which innovations are

especially likely to occur. If a case shows innovative uses and broader applicability (use with a greater number of lexemes or functions) more often in the teenagers' than the adults' spoken language, it may be regarded as being in the midst of ongoing grammaticalisation. If, however, the teenagers' and adults' data pattern similarly in terms of case usage, the case may be regarded as being more stable in its functionality and not in the midst of ongoing change.

4. Data and method

Evidence of grammaticalisation can be found through comparing synchronic language data from two corpora of spoken language, because the measures indicating grammaticalisation involve a number of predictable changes in terms of case distribution in a corpus. Among the features associated with grammaticalisation, the one which is easily quantifiable in corpus data is productivity (see section 4.1). Hence, this study uses productivity measures as a tool for assessing and comparing the extent of grammaticalisation.

The study is based on two spoken language corpora. Adult speech is represented by the Phonetic Corpus of Estonian Spontaneous Speech (Lippus et al. 2021), which contains 913,543 tokens. Most of the data (75%) in that corpus is made up of transcriptions of dialogues between speakers between 20–85 years of age. Of the total token count, 13% originate from monologues and 12% from conversations between three people.

Teenagers' spoken language is represented by the Teen Speech Corpus (Koreinik et al. 2023), which comprises conversations among friends, mostly dialogues, like the Phonetic Corpus. Speakers are aged 9–18. The form of the corpus included in this study consists of 531,179 morphologically annotated tokens (as of September 2022). Both corpora primarily consist of spontaneous speech, and they provide a useful point of comparison between age groups.

This study investigates the use of allative and elative, two highly multifunctional cases in Estonian (Aigro 2022). In order to answer RQ1 (see section 3), we also investigate the use of partitive and genitive case. Table 1 describes the frequency of these four cases in the two corpora.

Table 1. Number of tokens of lexemes occurring in four cases in two corpora of spoken Estonian.

	Phonetic Corpus of Estonian Spontaneous Speech	**Teen Speech Corpus**
Elative	5,607	2,548
Allative	5,165	3,646
Partitive	29,071	20,864
Genitive	34,221	22,371

The samples described in Table 1 were used to create eight random samples, each comprising 2,500 tokens of one case in a particular corpus (matched to the smallest sample in Table 1). We found the values of two measures of productivity in each of the eight datasets – realised and potential productivity (section 4.1).

For RQ1, the productivity measures for elative, allative and the two grammatical cases will be compared separately within each corpus. This allows us to assess how these cases compare to each other in terms of grammaticalisation, and how results from spoken data compare to the results from written data reported in Aigro (2022).

For RQ2, the productivity measures from each corpus are compared to each other, to assess the differences between adults' and teenagers' linguistic behaviour. The study refrains from merely comparing individual numerical productivity values, because one must be careful with a comparison across corpora. In our case, a relevant expected distinction is that adults generally use a wider range of lexemes compared to teenagers (see also Tagliamonte 2016: 25). As productivity measures involve a comparison of lexical variance, they are not able to distinguish between increased lexicon size and increased proportion of innovations. Hence, reporting that, for instance, elative is more productive in adult speech than in teen speech may not be relevant to grammaticalisation, because one would expect this result based on the wider lexical range used by adults.

As a way around this natural difference in lexicon size and lexical diversity between speaker populations, we investigate differences in the relative productivity of the individual cases inside each speaker group separately. If the lexicon is larger among adults, this would be expected to render higher productivity values for all cases in the adult corpus,

rather than affecting any one case in particular. Hence, if a spatial case is more similar to grammatical cases in the teenage than adult data, we take this as evidence of innovations in the adolescent data. This would show that teenagers are applying the case to a broader range of lexical material, which, in turn, indicates ongoing grammaticalisation for that case.

Because both corpora contain spontaneous spoken language, spoken language artefacts such as repetition and self-correction (6a–6b) affect both datasets in similar ways:

(6) a. *mis* *sa* *arva-d* *selle-st* *et* *ää* (...) *noh*
what 2SG think-PRS.2SG this-ELA that um well
selle-st *et* *ää*
this-ELA that um
'What do you think about the fact that, um, well, the fact that um…' (M, 10)

b. *noh* *selle-st* *ju* *selle-st* *ei* *tul-nud*
well this-ELA FOC this-ELA NEG come-CNG
keegi *nei-le* *rõõmu-ga* *teata-ma*
noone 3PL-ALL joy-COM announce-INF
'No one came and happily told them about this, well, this.' (Phonetic Corpus of Estonian Spontaneous Speech)

However, there are also differences between the two corpora. Differences in lexicon size are discussed above, but in addition to that, words in languages other than Estonian are differently coded in the two corpora, including the decisions regarding what to count as non-Estonian. The differences in the two coding systems render non-Estonian data in the two corpora incomparable. For this reason, and because the English lexemes are not fully parsed morphologically, we have excluded them from the data. English nouns do, however, occur in the data with Estonian case affixes. Their omission will have a greater effect on the Teen Speech Corpus than the adult corpus, as it removes a sizable lexicon of items unlikely to occur in adult speech (according to Vihman et al. 2022, 3% of tokens in the teen speech corpus were in English). We discuss examples of English tokens in the Teen Speech Corpus in section 5.3.

4.1. Productivity

Productivity is a useful measure for assessing case grammaticalisation. The grammaticalisation of cases includes predictable changes in overall distribution, expected to pattern with increased productivity. The more grammaticalised a case, the wider the range of lexical material it is expected to occur with.

Productivity has been measured in a number of ways (Zeldes 2012), two of which are used in this study – realised and potential productivity. Realised productivity is the type count of unique lexemes occurring with a case in a sample (Baayen & Lieber 1991; Baayen 1993; Baayen & Renouf 1996). Cases with a greater lexeme type count can be said to have a wider range of application. This implies a broadened functional range and greater grammaticalisation. However, realised productivity has its limitations as a measure of language change. It describes the productivity of a case during the time period covered by data, i.e. productivity which has been already implemented. The concept of productivity, however, inherently implies an expectation of future behaviour, referring to the availability of the case for extension to novel contexts. To measure this future dimension, another measure is required.

Potential productivity is measured via the proportion of a morpheme's *hapax legomena* among all its tokens (Baayen 1993). Hapaxes are instances where a pattern may be found with a particular lexeme only once in the data. In this study, hapaxes are case-marked lexemes which are represented with only one token in the corpus. For instance, the lexeme *aju* 'brain' has one allative token (*ajule*) in the Phonetic Corpus of Estonian Spontaneous Speech: *ajule* constitutes an allative hapax in that corpus, expressing a more peripheral area of allative usage. Extension to a new function may be expected to introduce novel lexemes for a case, meaning that cases with more peripheral uses may be considered to have a wider range of functions and to be more productive than those with more frequent conventional uses and fewer peripheral uses.

For calculating productivity, this study takes a variable-corpus approach (Gaeta & Ricca 2006). Instead of comparing case distribution in all lexemes marked in these cases, analysis is conducted on equally-sized samples. Hence, a random sample of 2500 case tokens was extracted from each of the eight token lists represented by the eight cells

in Table 1, rendering the eight datasets on which this study is based. An equal number of tokens will prevent the overestimation of productivity in the case of lower-frequency cases (those with smaller samples). Taking this approach, potential productivity is measured by a straightforward hapax count rather than the standard formula where hapaxes are divided by the number of morpheme tokens. As the samples are equal, the divisor would be the same for each dataset.

In summary, we measure realised and potential productivity in all eight datasets, representing four cases in two corpora. To answer RQ1, we compare the values of each case within each corpus. To answer RQ2, we compare productivity-based case rankings between the two corpora.

5. Results

This section outlines the lexical distribution of elative and allative in the two corpora. First, we provide results that address RQ1 and RQ2 (section 5.1). Second, we discuss the lexemes occurring with elative and allative most frequently, offering explanations to the differences observed in 5.1 (section 5.2). Third, we elaborate briefly on the way the Estonian elative and allative case markers are used with English items in the teenagers' data (section 5.3).

5.1. Grammaticalisation of elative and allative

In order to investigate the synchronic dimension of case grammaticalisation (RQ1), elative and allative are compared to partitive and genitive case use in terms of two productivity measures.

Genitive clearly emerges as the most productive of the four cases in adult speech, as it is found with the most diverse lexeme types (Figure 1A) as well as occurring in unique contexts more frequently than the other cases (Figure 1B). The difference between genitive and elative is statistically significant for both realised ($\chi^2 = 30.352$, $p < .0001$) and potential ($\chi^2 = 38.843$, $p < .0001$) productivity measures. Moreover, elative shows greater realised productivity (Figure 1A) than the partitive, occurring with a significantly wider range of lexeme types ($\chi^2 = 6.272$, $p = .01$). There is no statistically significant difference between elative and partitive in potential productivity. Allative is the least productive among the four cases in the adult data.

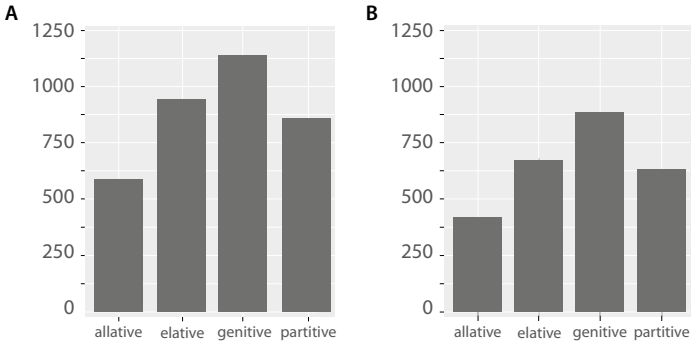

Figure 1. Productivity in adult data. Realised productivity (A) showing number of lexemes (types) occurring with each case, and potential productivity (B), showing number of tokens occurring only once with each case.

Quite a different picture emerges from the teen language sample (Figure 2). While allative is still the least productive of the four cases, elative has the highest productivity according to both measures in the teen data. It is not statistically distinct from genitive, neither in terms of realised (Figure 2A) nor potential productivity (Figure 2B). Both elative and genitive are significantly more productive according to both measures than the partitive case (genitive vs partitive, realised productivity: $\chi^2 = 5.421$, $p = .02$; potential productivity: $\chi^2 = 4.715$, $p = .03$).

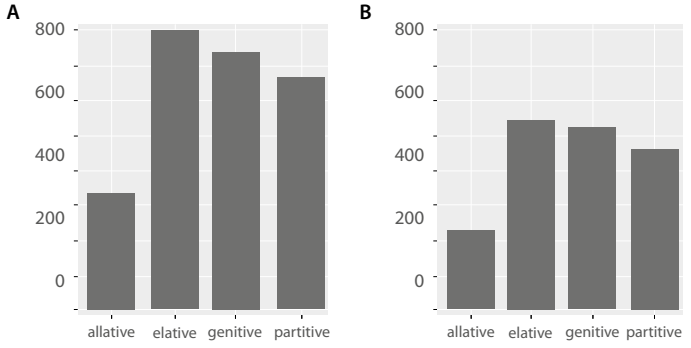

Figure 2. Productivity in teenager data. Realised productivity (A) showing number of lexemes (types) occurring with each case, and potential productivity (B), showing number of tokens occurring only once with each case.

RQ2 is addressed by comparing the productivity rankings in Figure 1 and Figure 2. The adult and teen speech data are similar in several respects. First, the allative remains the least productive case among the four inspected cases in both samples. Second, elative is more productive than partitive in both registers. In adult speech, this applies only to realised productivity, while in teen speech this applies to both realised and potential productivity.

However, one major difference exists between the adult and teen data. Elative ranks higher in productivity in teen usage than adult usage. Its range of uses and functions in the teen data are analogous to those of genitive, a highly grammaticalised grammatical case (Miljan & Cann 2013). In the adult data, however, while elative exceeds allative in terms of productivity overall and exceeds partitive in terms of realised productivity, it is significantly more restricted than genitive case.

5.2. Predominant case-marked lemmas

Section 5.1 demonstrates that teenagers use elative differently from adults, while their use of allative is more similar to that of adults. When we analyse the most frequent tokens in teenager and adult data, it becomes apparent what may drive this distinction. Table 2 shows the fifteen most frequent tokens occurring with elative case marking.

We find differences in the lexemes occurring most frequently in elative in the two data samples. Tokens such as *klassist* 'from class' and *koolist* 'from school' are not as frequent in the discussion topics in adult dialogues for obvious reasons. In addition to differences related to the topic of conversation, the teenagers' sample also has a higher token frequency of idiomatic colloquialisms (*lambist* 'randomly', lit. 'from the lamp', *jumalast* 'totally', lit. 'from God'; for the latter, see Veismann et al. 2017: 339), which in itself constitute an illustrative example of widened case usage in teen speech compared to adult speech.

Table 2. The fifteen most frequent elative tokens in adult and teenager data; raw counts in random samples of 2500 tokens of elative case.

Teenager speech		Adult speech	
Token	**Frequency**	**Token**	**Frequency**
selle-st 'this-ELA'	519	*selle-st* 'this-ELA'	493
mille-st 'what-ELA'	85	*mille-st* 'what-ELA'	77
minu-st '1SG-ELA'	72	*eest* 'front-ELA'	58
tema-st '3SG-ELA'	40	*ühe-st* 'one-ELA'	34
klassi-st 'class-ELA'	34	*tema-st* '3SG-ELA'	34
aja-st 'time-ELA'	32	*asja-st* 'thing-ELA'	34
kooli-st 'school-ELA'	28	*mingi-st* 'some-ELA'	27
asja-st 'thing-ELA'	27	*keele-st* 'language-ELA'	27
sinu-st '2SG-ELA'	26	*teise-st* 'other-ELA'	25
ühe-st 'one-ELA'	25	*kahe-st* 'two-ELA'	19
lambi-st 'lamp-ELA'	25	*aja-st* 'time-ELA'	18
käe-st 'hand-ELA'	24	*minu-st* '1SG-ELA'	15
jumala-st 'God-ELA'	21	*iga-st* 'every-ELA'	15
enda-st 'oneself-ELA'	20	*otsa-st* 'end-ELA'	14
koha-st 'place-ELA'	19	*külje-st* 'on-ELA'	14

However, there are further semantic differences between the lexemes in the two columns in Table 2, especially regarding animacy. As described in section 2, even though case inflections are expected to be *able* to occur with all nouns, certain cases are more semantically compatible with some lexemes than they are with others. Spatial cases are expected to mark animate entities less frequently than, for instance, comitative or grammatical cases. When animate lexemes do occur with a spatial case, this marks extended functionality for that case. Because the original semantic meaning of the spatial cases typically involves a place or spatial relation, compatible with concrete, inanimate referents, their greater use with animate referents suggests a degree of semantic bleaching in case usage. Such uses introduce new types of metaphorical meaning for spatial cases. For instance, using allative ('onto') in *ema-le* 'mother-ALL' creates a metaphorically spatial interpretation ('to mom') rather than a directly spatial interpretation ('onto mom') (Aristar 1997).

When we inspect the lexemes in Table 2 in terms of their animacy, we find four typically animate pronouns among the elative tokens in the teen corpus (1SG, 2SG, 3SG and reflexive), compared to two in the adult sample (1SG and 3SG). These are the only tokens typically referring to animate, human referents in the list. Moreover, elative forms of the main animate pronouns, *minu-st* 'I-ELA', *sinu-st* 'you-ELA' and *tema-st* 's/he-ELA' make up 6% of the data in our teenage sample of 2500 case tokens, while they only make up 3% of the adult sample (*sinu-st* 'you-ELA' also occurs in the adult sample, with a frequency of 13 tokens). This shows an increased use of elative to mark animate referents in the teen corpus. This phenomenon suggests semantic bleaching, increased productivity and extension in functionality.

Table 3. The fifteen most frequent allative tokens in teenage and adult data; raw counts in random samples of 2500 tokens of allative case.

Teenager speech		**Adult speech**	
Token	**Frequency**	**Token**	**Frequency**
minu-le '1SG-ALL'	803	*minu-le* '1SG-ALL'	423
sinu-le '2SG-ALL'	385	*sinu-le* '2SG-ALL'	195
tema-le '3SG-ALL'	221	*tema-le* '3SG-ALL'	175
enda-le 'REFL-ALL'	209	*selle-le* 'this-ALL'	124
selle-le 'this-ALL	80	*enda-le* 'oneself-ALL'	101
töö-le 'work-ALL'	37	*töö-le* 'work-ALL'	48
ema-le 'mother-ALL'	32	*koha-le* 'place-ALL'	41
kõigi-le 'everyone-ALL'	29	*teise-le* 'other-ALL'	29
teise-le 'other-ALL'	26	*oma-le* 'oneself-ALL'	29
inimese-le 'person-ALL'	24	*lapse-le* 'child-ALL'	21
ühe-le 'one-ALL'	18	*kelle-le=gi* 'someone-ALL=CL'	20
kelle-le 'who-ALL'	18	*inimese-le* 'person-ALL'	20
kelle-le=gi 'someone-ALL=CL'	16	*iga-le* 'every-ALL'	18
oma-le 'oneself-ALL'	14	*kõigi-le* 'everyone-ALL'	17
iga-le 'every-ALL'	13	*ühe-le* 'one-ALL'	15

While teenagers and adults share only seven of the top fifteen lexemes in elative in Table 2, in the allative data (Table 3) they share almost all of the top 15 lexemes (n = 13).This, too, illustrates a way in which teenagers are more similar to adults in terms of allative than elative use.

Human animacy is encoded by a wider range of lexemes in the allative case (Table 3) than elative (Table 2). This is to be expected, considering the greater concentration of allative functions compatible with human referents (e.g. Recipients, Experiencers: Erelt, Erelt & Ross 2007; Aigro 2022: 138). Given the overlap in the most frequent allative forms in the two corpora and the greater use of lexemes with human referents, animacy does not constitute a difference between teenager and adult allative use. However, teenagers use allative 1st and 2nd-person pronouns almost twice as frequently as adults (Table 3). As with elative, we take this to suggest an increase in semantic bleaching in case usage.

5.3. English in the teenagers' language

Because the notion of potential productivity is based on a count of hapaxes, it is worth considering what underlies the unique forms. These may derive from a grammaticalised case becoming more generally applicable, and hence being used with an increasingly wide range of lexemes, but they may also be linked to new lexemes previously unused in the language: teens may demonstrate both of these sorts of innovations. In the teen data, much lexical innovation comes from contact with English and the use of code-switching or English loans.

The quantitative analysis in this paper excludes English-language words for two reasons. First, there are differences in the language tagging in the two corpora. Second, it is not possible at this point to automatically extract morphologically integrated English words with a particular inflection. If we were to extract cases simply based on character strings (*-st*, *-le*), the list would include words like *past* and *pale*. However, the exclusion of English words is likely to affect the teen corpus more than the adult dataset: the proportion of words tagged as English-language is considerably larger in the teen spoken corpus (3.3%, see Vihman et al. 2022) than the adult language in the Phonetic Corpus (0.22%, see Aasa et al. 2022).

Deciding which insertions to code as English is far from trivial, especially for lexemes which have gained ground and are in use more generally in colloquial language. But if we are interested in unique forms, then newer, English-language lexemes are likely candidates, as words which are not used frequently overall. These are less likely to be controversial in terms of language tagging, not having been adopted into the language at large. Here, we take an initial look at elative and allative marking on English words, focussing only on the teen data.

Using data extracted from an earlier version of the Teen Spoken Corpus, we found 20 English lexical items (words and phrases) in elative case, only one of which (*playboxi-st* 'playbox-ELA', a term related to computer games) was repeated. In this list, we find context-specific terms such as *skin* and *diamond blocks* (7a), social-media related terms such as *chat*, *challenge* and *hater* (7b), and topics which have greater frequency or different connotations in English-language entertainment media than Estonian contexts, such as (7c).

(7) a. *maailm* *kus* *ma* *ehita-sin (...)*
world where 1SG build-PST.1SG
diamond blocki-de-st *maja*
diamond.block-PL-ELA house
'A world where I built a house out of diamond blocks.' (M, 13)

b. *noh* *ta* *teeni-b* *miljon-eid* *enda*
well 3SG earn-PRS.3SG million-PL.PAR REFL.POSS
hateri-te-st
hater-PL-ELA
'Well s/he is earning millions from his/her haters.' (M, 12)

c. *mingi* *hakka-b* *jälle* *enda* *mingi*
PARTICLE start-PRS.3SG again REFL.POSS PARTICLE
bisexual girli-de-st rääki-ma
bisexual.girl-PL-ELA speak-INF
'Like, [s/he] starts talking again about his/her, like, bisexual girls.' (F, 16)

Similarly, we found 22 examples of English-tagged items in allative case, but these employ only 13 lexemes. The lexemes which were repeated at least twice in allative form in the corpus were *bot*, *mute*, *podcast*, *recording* and *read*. Unique forms include gaming and social media terminology as in (8a) and English-context-derived terms such as *mental health* and *road trip* (8b-c).

(8) a. *lähe-n* *ta* ***for-you-pagei-le***
go-PRS.1SG 3SG.POSS for-you-page-ALL
'I am going to his/her for-you page.' (F, 16)

b. *see* *rõhu-b* *nii* *hullult* *mu*
this weigh-PRS.3SG so badly 1SG.POSS
mental healthi-le *et* *ma* *ei* *jaksa*
mental health-ALL that 1SG NEG have.energy.CNG
'This affects my mental health so badly, that I don't have the energy.' (M, 14)

c. *me* *läh-me* *sellise-le* ***road tripi-le***
1PL go-PRS.1PL such-ALL road.trip-ALL
'We are going on a kind of road trip.' (F, 14)

It is likely that the elative and allative-marked English-language insertions include more hapaxes in the teen than the adult data; on the other hand, genitive and partitive are also used productively with English-language insertions, and so including English words may not change the proportions vis-à-vis the questions under consideration, namely the grammaticalisation of elative and allative, compared with the grammatical cases. In the teen data, both occur with English items and both can be seen to mark hapaxes, indicating that they would contribute globally to measures which are taken to indicate realised and potential productivity.

6. Discussion

For RQ1, spoken language based results indicate that elative case is significantly more grammaticalised than allative. Furthermore, although we would expect grammatical cases to be more productive than semantic cases, the multifunctionality of elative ranks it alongside the grammatical cases, even exceeding partitive in both spoken language samples.

This is an interesting result, in that the spoken data presents a somewhat different picture from the written language data analysed by Aigro (2022). In the written data, allative was found to be as productive as partitive, and both were exceeded by elative and genitive in terms of realised and potential productivity (Aigro 2022; 'lexical variance' was used instead of 'productivity' in that study). In spoken language,

however, the distribution of allative appears to be more lexically restricted than that of partitive. This leads us to conclude that elative is less sensitive to context and register than allative, again indicating higher overall productivity and greater grammaticalisation.

Regarding RQ2, we find that teenagers use elative with a wider range of lexemes than adults, based on a comparison of elative-genitive rankings in adult and teenager data. The top 15 elative tokens in the two registers give reason to hypothesise that this difference stems from the fact that teenagers use elative with lexemes expressing human entities more often than adults; this may be one of the forces driving the increased productivity and grammaticalisation of elative in teenagers' language use compared to adults. The expanded use of elative in teenagers' language also includes frozen colloquial expressions such as *lambist* 'randomly, out of the blue' (lit. 'from the lamp') and *jumalast* 'totally' (lit. 'from God'). All in all, elative may be viewed as being diachronically in flux, with innovations and usage in novel contexts in the speech of young people demonstrating and driving its ongoing grammatical change.

The same does not apply to allative, which appears to have achieved a more stable *status quo* in terms of grammaticalisation. It ranks lowest of the four cases in the study, in both teen and adult language. This similarity between the two registers was also underscored by the overlap in the most frequent allative tokens. Teenagers use allative with a similar distribution to that of adults, with an apparently similar proportion of lexemes with human referents. This lack of innovations in allative usage by teenagers indicates that diachronically, the allative case may be regarded as more static and settled than the elative, despite its rather multifunctional profile.

This is further supported by the comparison between spoken and written data. In Aigro's (2022) analysis of written data, elative ranks lower than genitive in terms of both realised and potential productivity, but higher than partitive in both measures. Hence, written data patterns with adult spoken language, rather than teen spoken language, in terms of case distribution rankings. This is to be expected: written language using Standard Estonian is where one would look to see already established language usage rather than case usage innovations; adult spoken language is also expected to represent earlier established language usage. Teen language, however, is precicely the register in which innovations

are expected to occur. The fact that these innovations in elative usage are found in adolescent language indicates a high likelihood that elative is undergoing further grammatical change. Future research on teenagers' case usage in online communication, and a more thorough analysis of English usage in the various datasets will help fill out this picture of grammaticalisation and innovation.

Acknowledgements

This study was funded by the Teenspeak in Estonia project (grant no. EKKD3), for which we gratefully acknowledge support from the "Estonian Language and Culture in the Digital Age (2019–2027)" R&D programme of the Estonian Ministry of Education and Researh.

References

Aasa, Triin, Korkus, Mari-Liis, Lindström, Liina, Mandel, Aive, Pilvik, Maarja-Liisa & Vihman, Virve. 2022. Kui hukas on tänapäeva noorte keelekasutus? *Postimees*, 08.10.2022.

Aigro, Mari. 2022. *In any case? Estonian spatial cases as argument markers* (Dissertationes Linguisticae Universitatis Tartuensis 44). Tartu: University of Tartu Press.

Aigro, Mari & Virve-Anneli Vihman. 2023. Oblique complements in Estonian: A corpus perspective. *Journal of Linguistics*. 1–31. https://doi.org/10.1017/S0022226722000536.

Aristar, Anthony Rodrigues. 1997. Marking and hierarchy types and the grammaticalization of case markers. *Studies in Language* 21(2). 313–368. https://doi.org/10.1075/sl.21.2.04ari.

Blake, Barry J. 2001. *Case*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139164894.

Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael & Silvi Vare. 1995. *Eesti keele grammatika I. Morfoloogia. Sõnamoodustus.* (ed.) Mati Erelt. Tallinn: Eesti Teaduste Akadeemia, Eesti Keele Instituut.

Erelt, Mati, Tiiu Erelt & Kristiina Ross. 2007. *Eesti keele käsiraamat.* Tallinn: Eesti Keele Instituut.

Erelt, Mati. 1989. *Eesti lauseliikmeist.* Tallinn: Eesti NSV Teaduste Akadeemia.

Erelt, Mati. 2003. Süntaks. *Emakeele Seltsi Aastaraamat* 48. 63–97.

Heine, Bernd, Ulrike Claudi & Friederike Hünnemeyer. 1991. *Grammaticalization: A conceptual framework*. Chicago: Chicago University Press.

Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511613463.

Heine, Bernd. 1990. The dative in Ik and Kanuri. In William A. Croft, Suzanne Kemmer & Keith Denning (eds.), *Studies in typology and diachrony*, 129–149. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.20.09hei.

Heine, Bernd. 2008. Evolution of case systems. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*, 458–469. Oxford: Oxford University Press.

Hopper, Paul & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139165525.

Koreinik, Kadri, Aive Mandel, Maarja-Liisa Pilvik, Kristiina Praakli & Virve-Anneli Vihman (2023). Outsourcing teenage language: A participatory approach for exploring speech and text messaging. *Linguistics Vanguard*. https://doi.org/10.1515/lingvan-2021-0152.

Kuryłowicz, Jerzy. 1964. *The inflectional categories of Indo-European*. Heidelberg: Carl Winter Universitätsverlag.

Larjavaara, Matti. 1991. Aspektuaalisen objektin synty. *Virittäjä* 95(4). 372–408.

Lehmann, Christian. 1985. Synchronic variation and diachronic change. *Lingua e Stile* 20. 303–318.

Lestrade, Sander. 2010. *The space of case*. Radboud University, Nijmegen. Ph.D. thesis.

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

Lippus, Pärtel, Kätlin Aare, Anton Malmi, Tuuli Tuisk & Pire Teras. 2021. Phonetic Corpus of Estonian Spontaneous Speech version 1.2. https://datadoi.ee/handle/33/351.

Mandel, Aive, Kristiina Praakli, Virve-Anneli Vihman & Kadri Koreinik. 2022. Rahvateadusega noortekeelt kogumas. *Oma Keel* 44. 16–21.

Miljan, Merilin & Ronnie Cann. 2013. Rethinking case marking and case alternation in Estonian. *Nordic Journal of Linguistics* 36(3). 333–379. https://doi.org/10.1017/S0332586513000309.

Nichols, Johanna & Alan Tiberlake. 1991. Grammaticalization as retextualization. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, 129–146. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.19.1.07nic.

Nichols, Johanna. 1983. On Direct and Oblique Cases. In Amy Dahlstrom, Claudia Brugman, Monica Macaulay, Inese Civculis, Michele Emanatian, Donna Sakima & Raquel Teixeira (eds.), *Proceedings of the Ninth Annual Meeting of the Berkeley Linguistics Society*. 170–192. Berkeley, USA.

Nichols, Johanna. 1984. Direct and oblique objects in Chechen Ingush and Russian. In Frans Plank (ed.), *Objects. Towards a theory of grammatical relations*, 183–209. London: Academic Press Inc.

Nurka, Maiu. 2014. *Eesti keele elatiivi semantika kognitiivse keeleteaduse vaatenurgast*. Tartu Ülikool, eesti keele osakond. Magistritöö.

Pajusalu, Renate. 2006. Multiple motivations for meaning of an elative wh-construction in Estonian. *Trames* 10(4). 324–340. https://doi.org/10.3176/tr.2006.4.02.

Rice, Sally & Kaori Kabata. 2007. Crosslinguistic grammaticalization patterns of the allative. *Linguistic Typology* 11(3). 451–514. https://doi.org/10.1515/LINGTY.2007.031.

Tagliamonte, Sali. 2016. *Teen talk: The language of adolescents*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139583800.

Veismann, Ann, Anni Jürine, Karl Pajusalu, Renate Pajusalu & Ilona Tragel. 2017. Eesti taeva tähendused. *Journal of Estonian and Finno-Ugric Linguistics* 8(1). 323–347. https://doi.org/10.12697/jeful.2017.8.1.17.

Veismann, Ann, Mati Erelt & Helle Metslang. 2017. Määrus. In Mati Erelt & Helle Metslang (eds.), *Eesti keele süntaks* (Eesti keele varamu III), 300–375. Tartu: Tartu Ülikooli Kirjastus.

Vihman, Virve-Anneli, Kristiina Praakli, Maarja-Liisa Pilvik & Mari-Liis Korkus. 2022. Kas noored on inglise keelega *obsessed*? Millest räägivad korpusandmed? In Anna Verschik (ed.), *Research in children's and youth bilingualism* (Philologia Estonica Tallinensis), 292−322. Tallinn: Tallinn University Press. https://doi.org/10.22601/PET.2022.07.11.

Viht, Annika & Külli Habicht. 2019. *Eesti keele sõnamuutmine* (Eesti keele varamu IV). Tartu: Tartu Ülikooli Kirjastus.

Kokkuvõte. Mari Aigro, Virve-Anneli Vihman. *Täitsa lambist*: elatiivi ja allatiivi grammatiseerumise tõendeid noorte ja täiskasvanute suulises keeles. Neljateistkümnest eesti keele käändest peetakse ühtteist semantilisteks kääneteks, kuid mõned neist paistavad olevat teistest enam grammatiseerunud. Neid võiks ka pidada (pool)grammatilisteks, kuna nad markeerivad muuhulgas verbide argumente. Käesolev uuring keskendub elatiivile ja allatiivile, mis on varasemates uuringutes esile tõusnud multifunktsionaalsuse ja grammatiseerumise poolest. Uuringus võrreldakse elatiivi ja allatiivi distributsiooni spontaanses täiskasvanute ja teismeliste suulises keeles, et uurida pseudodiakroonilisel meetodil, kas nende käänete grammatiseerumine on eesti keeles aktiivne või on tegu funktsionaalselt passiivsete kategooriatega. See on esimene uuring, mis kvantifitseerib käänete multifunktsionaalsusega seotud distributiivsed omadused, et hinnata käändeid hõlmavate grammatiliste muutuste intensiivsust.

Märksõnad: kääne, elatiiv, allatiiv, grammatiseerumine, suuline keel, noorte keel