universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

The Estonian Illative: Productivity and Formalization

verfasst von / submitted by

Johannes Hirvonen, BA BA BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 854
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Finno-Ugristik
Betreut von / Supervisor:	Univ.-Prof. Dr. Johanna Laakso

Contents

Abbreviations

ABL = ablative
ACC = accusative
ADE = adessive
ALL = allative
COM = comitative
DM = Distributed Morphology
EKG = Eesti Keele Grammatika
EKK = Eesti Keele Käsiraamat
ELA = elative
EMPH = emphatic particle
ESS = essive
GEN = genitive
ILL = illative
INE = inessive
INF1 = infinitive 1
INF2 = infinitive 2
L.ILL = long illative
NEG = negation
NOM = nominative
PL = plural
PRES = present tense
PRT = partitive
Q = quantity grade
SG = singular
S.ILL = short illative
TRM = terminative
TRS = translative
ÕS = Õigekeelsussõnaraamat
1 = first person
2 = second person

Persönliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbst erstellt und alle verwendeten Quellen ordnungsgemäß zitiert habe.

1 Problem statement

The illative singular takes a special position among the other local cases in the Standard Estonian case system because it has a high number of allomorphs while other local cases do not show any variation in exponence. Besides the productive *-sse* exponent, which will be called the "long" illative in what follows, there is a number of "short" allomorphs of the illative singular which are partly idiosyncratically connected to certain roots, and partly productive as well. The allomorphs whose distribution and analysis will be the focus of this thesis are *-sse*, a zero exponent, and an exponent which causes gemination of a root-internal consonant:[1]

Allomorph	Stem	ILL.SG	Gloss
-sse	*liiga*	*liigasse*	'into the league'
-ø	*k`arpi*	*k`arpi*	'into the box'
Gemination	*kino*	*k`inno*	'into the cinema'

Table 1: Three allomorphs of the Estonian illative

In traditional normative grammars like *Eesti Keele Käsiraamat* (EKK) 'Handbook of Estonian' (M. Erelt, T. Erelt, and Ross 2007) and *Eesti Keele Grammatika* (EKG) 'Grammar of the Estonian Language' (M. Erelt et al. 1995 and M. Erelt et al. 1993) the long illative has been treated as viable parallel form for almost any noun, and across declension classes it is indeed the most widespread exponent of the illative. Both classes that *karp* and *kino* belong to are to a different extent also used with the long illative in *-sse*. The geminated illative has been argued to involve outright root suppletion since only a small subclass of stems in this class can have a geminated illative form, while the zero illative has often been treated as a special marked exponent or a separate case altogether. The existence of parallel illative forms for the same nouns gives rise to another source of variation:

[1] A note on Estonian orthography: the graphemes <b,d,g> do not stand for voiced plosives, but lenis consonants, while <p,t,k> stand for their fortis counterparts, <š> stands for [ʃ] and <ž> for [ʒ]. Both vocalic and consonantal long segments are expessed by two graphemes. <õ> stands for an unrounded mid-high back vowel, often transcribed as [ɤ], <ä,ö,ü> stand for [æ,ø,y]. The grave accent ` as marker of "overlength" is not part of the standard orthography but is sometimes used in grammars and dictionaries to indicate the presence of overlength and is written here after the onset consonant of the relevant stressed syllable. Palatalisation of segments is not marked in the ortography and will be indicated in this thesis by ^{j} only if necessary.

The long illative behaves like any other local case with respect to the form of the stem and always attaches to the so called "weak" stem of a noun while the zero illative takes the "strong" stem. Consider the following example of parallel forms of the illative singular of *karp* 'box':

Allomorph	ILL.SG	Gloss
-sse	*karbisse*	'into the box'
-ø	*k`arpi*	'into the box'

Table 2: Stem and case allomorphy

The long illative attaches to a stem with a lenis consonant, and the short illative to a stem with a fortis consonant and prosodic overlength. In cases like the above, two problems pertaining to allomorphy arise when analyzing the distribution of the illative singular allomorphs: First, only roots with a certain prosodic structure allow a zero illative in the first place (in the table above the *karp* class), but the long illative is also encountered. Second, depending on the illative exponent that is chosen for these roots, the roots themselves exhibit allomorphy. In the above example we thus find allomorphy of the illative singular itself and at the same time stem allomorphy which manifests in the alternation of a fortis plosive with a lenis, and certain differences in foot-level prosody. This interplay of "allomorphies" is interesting especially from the viewpoint of modular and serialist grammar theories because they make very specific predictions about the steps involved in the derivation of morphologically complex word forms that can influence allomorph selection. The main research questions of this thesis are:

- What is the distribution of the illative allomorphs?
- How productive is the geminating illative?
- What is the link between illative allomorph choice and occurrence of stem alternations?

This thesis is structured as follows: In section 2 a short overview about the relevant properties of Estonian morphology and phonology and about Distributed

Morphology is given. The illative allomorphs are be argued to be genuine allomorphs of one and the same case with no difference in morphosyntax. Section 3 deals with the productivity of the illative allomorphs. After a review of previous studies and the notion of productivity, the results of an own corpus study about the geminating illative are presented. They confirm that the geminating illative is unproductive but also show the full extent of variation for individual roots. In section 4 an analysis which links the representation of Estonian roots to a general account of stem alternations is devised. According to this analysis, stem alternations are triggered by an autosegmental index carried by certain exponents. It is furthermore argued that the "long" quantity grade is morphophonologically marked, and not the "overlong" grade. With this background, section 5 is concerned with the analysis of the illatives themselves. The exponents are argued to vary with respect to whether they carry the autosegmental index to explain the stem alternations before different allomorphs. Finally, illative allomorph choice is linked to other instances of prosodically conditioned allomorph choice in the Estonian nominal system. Section 6 concludes.[2]

2 Introduction

2.1 Introductory remarks about Estonian phonology and morphology

2.1.1 Ternary quantity

Estonian famously has a phonemic ternary length distinction (marked by Q1, Q2 and Q3) in both consonants and vowels (for recent overviews see Kuznetsova 2018; Pajusalu 2015; Prillop 2013, 2018a; Türk 2019). Given the typological scarcity of ternary length contrasts in the languages of the world, very different approaches to explain this contrast have been proposed: it has been regarded as contrast on

[2]I am grateful for all the insights I gained from discussions with and the valuable advice I havve received in the course of writing this thesis from people around me. I especially want to thank (in alphabetical order) Laura Grestenberger, Heiki-Jaan Kaalep, Laura Kalin, Gabriel Pantillon, Markus Pöchtrager, Julia Scheiblauer, Marili Tomingas, and Albert Zak for taking their time and providing help and advice.

the segmental level, as property of syllables, or of even higher prosodic domains. Due to the special properties of the Q3:Q2 distinction, most recent accounts treat it as a phenomenon of higher prosodic domains like feet, and a similar approach will be followed in this thesis, too. The quantity contrast manifested in consonants can be observed in the table below (Q3 is marked by ` after the onset consonant of the relevant syllable).

Q1	Q2	Q3	Gloss		
lina	*linna*	*l`inna*	'linen.NOM'	'city.GEN'	'city.PART/S.ILL'
kabi	*kapi*	*k`appi*	'hoof.NOM'	'closet.GEN'	'closet.PART/S.ILL'
lugu	*luku*	*l`ukku*	'story.NOM'	'lock.GEN'	'lock.PART/S.ILL'

Table 3: The three quantity grades

The three quantities are distinguishing different case forms and thus are phonologically contrastive. As can be noted from the paradigm above, there is a quantitative contrast between Q3 and Q2 within the paradigm of a noun, and the Q1 examples originate from other paradigms, while they are segmentally identical. Crucially, there is only ever a quantitative contrast between Q3 and Q2 within one nominal paradigm (Blevins 2008), with the exception of the geminated illative where a Q1 consonant is lengthened and surfaces with Q3 prosody. Length alternation is discussed at length in section 4. For now, note that the morphologically relevant distinction obtains between Q3 and Q2 and that this is a prosodic property.

2.1.2 Noun morphology and the illative

Estonian has a nominal system with rich morphology. Apart from its numerous cases, the system is characterized by a mixture of analytic and synthetic morphology and various alternations of stems within a single paradigm. The basic template for an Estonian noun is root-vowel-number-case,[3] but there are some

[3]A note on terminology: the term "root" is used in this thesis strictly in the sense of Distributed Morphology, namely as uncategorized open class vocabulary item in contrast to functional morphemes (Embick 2015). The usage of this term does not imply any synchronic or diachronic relevance of phonestemic roots or indo-european style roots in Estonian (see Laakso 2017 for an overview). The term "stem" is used in this thesis as a cover term for the phonological

featural combinations like partitive plural in which this template has become opaque and the features are expressed by portmanteaus. Diachronically, Estonian has lost many segments through apocope (Laanest 1975) and some nouns show case syncretism of two to four of the core cases nominative, accusative,[4] genitive, and partitive. Therefore, Estonian has been described as currently undergoing a grammaticalization change from an agglutinating to a more Standard Average European fusional type (H. Metslang 2009). For example, the genitive is characterized not by a specific suffix but only by the root's stem vowel which is not visible in the nominative of many noun classes and which for the vast majority of nouns is unpredictable. Similarly, the null illative neither shows any analytic morphology but only a null exponent and occurs only with certain roots. All other case suffixes in the singular except the nominative, partitive and short illative are always analytic and get suffixed to the same stem which the null genitive uses.

For an overview about the case system consider the paradigm of *k`app* 'cupboard' in Table 4, according to *Eesti Keele Õigekeelsussõnaraamat* 2018 (ÕS) 'Dictionary of Correct Estonian'.[5] The first four cases are usually grouped as "grammatical" cases, the illative up to the ablative as local cases and the rest are various semantic cases. With respect to overt marking, the grammatical cases in this paradigm which belongs to a noun class which will be called the *sepp* class in what follows, differ only with respect to which stem is used and whether the theme vowel is present, and they do not exhibit any overt suffixes. The other cases in the singular are clearly marked by suffixes added to the stem of the genitive, besides the short illative which is overtly null and takes the same stem as the partitive. The distribution in the plural follows a similar pattern. The partitive case is characterized by portmanteau suffixes for number and case, the nomina-

form of a root plus possibly other segments like additional vowels where case suffixes attach to. From the viewpoint of morphology, stems have no relevance and are simply the result of other morphological processes like Vocabulary Insertion (Embick and Halle 2005).

[4]In traditional descriptions like M. Erelt, T. Erelt, and Ross (2007) and M. Erelt et al. (1993), the accusative is not recognized as separate case since it is always syncretic with another case form – with the genitive in the singular and with the nominative in the plural. For discussion about a separate accusative case in Estonian see e.g. Caha 2009; den Dikken and Dékány 2018; Hiietamm 2003; Norris 2018a,b.

[5]Accessible at `http://www.eki.ee/dict/qs/`.

tive, genitive and accusative are null, and all other cases build upon the plural suffix *-de* with the same exponents like in the singular. While there is allomorphy among some of the grammatical cases like the partitive singular which has three productive suffixes, *-ø*, *-t*, and *-d*, and the partitive plural with *-V* (here *-e*), *-id* and *-sid*, the other cases do not exhibit allomorphy at all – with the exception of the illative singular. Note in particular that if there are stem changes in a paradigm, in this case length alternation between Q3 and Q2, the long illative in *-sse* uses the same stem like all other local cases in the singular, while the short null illative uses the same stem like the partitive and is in this paradigm homophonous with the partitive. Color coding refers to the two length grades of the stem in this paradigm: the Q3 stem forms are marked white, the Q2 stems grey.

	SG		PL	
NOM	k`app-ø		kapi-d-ø	
ACC	kapi-ø		kapi-d-ø	
GEN	kapi-ø		k`appi-de-ø	
PRT	k`appi-ø		k`appe	k`appi-sid
ILL	kapi-sse	k`appi-ø	k`appi-de-sse	
INE	kapi-s		k`appi-de-s	
ELA	kapi-st		k`appi-de-st	
ALL	kapi-le		k`appi-de-le	
ADE	kapi-l		k`appi-de-l	
ABL	kapi-lt		k`appi-de-lt	
TRS	kapi-ks		k`appi-de-ks	
TRM	kapi-ni		k`appi-de-ni	
ESS	kapi-na		k`appi-de-na	
ABE	kapi-ta		k`appi-de-ta	
COM	kapi-ga		k`appi-de-ga	

Table 4: Paradigm of *k`app*

Since this thesis is concerned with allomorphy of the illative singular which is sensitive to properties of the root, a short overview about a few productive declension classes in Estonian shall be given. The declension classes themselves are ultimately predictable from the structure of the root (see section 4.2). This overview is based mainly on Kaalep (2009, 2010, 2012). To describe the Estonian

nominal system, traditional descriptions like those in EKK assume three basic stems from which, depending on the declension class, all other case forms can be deduced. These three stems are the nominative, genitive, and partitive singular. In what follows, four paradigms of relevant root types will be juxtaposed in Table 5–8. Only the nominative, genitive, partitive and illative are given since the other cases show no allomorphy. The first paradigm is from the same class as *k`app* above.

	SG	PL
NOM	s`epp	sepa-d
GEN	sepa	s`eppa-de
PRT	s`eppa	s`eppi
ILL	s`eppa (∼ sepa-sse)	s`eppa-de-sse

Table 5: *s`epp* 'blacksmith'

	SG	PL
NOM	seminar	seminari-d
GEN	seminari	seminari-de
PRT	seminari	seminare
ILL	seminari ∼ seminari-sse	seminari-de-sse

Table 6: *seminar* 'seminar'

	SG	PL
NOM	liiga	liiga-d
GEN	liiga	liiga-de
PRT	liiga-t	liiga-sid
ILL	liiga-sse	liiga-de-sse

Table 7: *liiga* 'league'

The three traditional stems in the bisyllabic *sepp* class show that this class is characterized by a nominative singular which lacks the stem vowel, has a null partitive and Q3:Q2 alternations within the paradigm. A short illative is possible here and long illatives are unusual (see section 3.1). The trisyllabic *seminar* class

	SG	PL
NOM	saba	saba-d
GEN	saba	saba-de
PRT	saba	sabu ∼ saba-sid
ILL	saba-sse ∼ s`appa	saba-de-sse

Table 8: *saba* 'tail'

similarly lacks the stem vowel in the nominative and also has a null partitive but does not show any stem alternations. A null illative is possible, and thus there is syncretism between genitive, (accusative,) partitive and illative. The bisyllabic *liiga* class retains its stem vowel in all forms, has a partitive in *-t*, and no stem alternations and no short illative. The bisyllabic *saba* class, a subclass of what will very descriptively be called CVCV class in this thesis, also retains its stem vowel in all cases and shows syncretism between all core cases (other roots with an identical structure have a partitive in *-t*). It shows no stem alternations. That is, the Q3:Q2 alternation is only encountered in a subclass of paradigms of bisyllabic stems. However, the geminated Q3 illative which is characterized by a geminated root-internal consonant is also possible for a number of roots in this otherwise Q1 class. Note that one fact unites all the declension classes, namely the condition that the genitive must end in a vowel. Now, before turning to the allomorphs of the illative, the treatment of allomorphy in Distributed Morphology will be introduced.

2.2 Allomorphy in Distributed Morphology

The morphology theory this thesis will follow is Distributed Morphology (DM) (Bobaljik 2015; Embick 2015; Halle and Marantz 1993, 1994). DM is a modular, realizational, generative theory of morphology. In DM, the role of morphology is to map the output of syntax to phonology. It is realizational since the role of morphology in this theory is to phonologically realize the abstract structures syntax has built. That is, morphology interprets syntactic structure rather than projecting it. Two basic tenets of DM are (Bobaljik 2015):

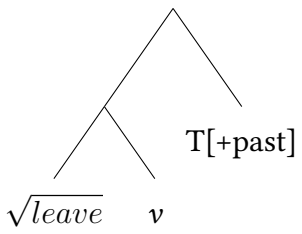
- The main mechanism of meaningful composition in language is syntax and

it operates above and below the word level. Thus, morphosyntax is "syntactic hierarchical structure all the way down" (Halle and Marantz 1994, pp. 276–277). DM is a non-lexicalist theory of morphology since it dispenses with the notion of the lexicon as locus of pre-syntactic word formation and views word formation as syntactic.

- The pieces that syntax operates on are abstract and devoid of phonological content. These abstract pieces are paired with phonological content only post-syntactically in a process called Vocabulary Insertion, which associates the pieces of morphosyntactic computation, morphemes, with exponents. Vocabulary Insertion rules supplant part of what is known as the "lexicon" in lexicalist theories and are one of the language specific parts of morphology.

Concretely, the derivation of any surface form which spells out morphosyntactic structure proceeds as follows (Embick 2015; Kalin 2019): First, the abstract morphosyntactic structure which consists of syntactic primitives and features is built. For example, the structure underlying *left*, the past tense of *leave*, looks as follows:

(1) Past Tense of $\sqrt{leave}$ (Embick 2015, p. 185)



The most embedded node in this complex structure is the uncategorized root which is categorized here as verb by the verbalizer *v* and is contained within another projection of a node for tense. This T head carries a [+past] feature. After this structure is built, the nodes are concatenated and transformed from a hierarchical to a linear structure (that is, linearized), starting from the most embedded node:

(2) Concatenation Statement from example 1 (Embick 2015, p. 185)
$\sqrt{leave} \frown v$, $v \frown$T[+past]

Exponent choice and allomorphy happen after linearization in the process of Vocabulary Insertion. Here, the nodes are realized as exponents from a list of vocabulary items with phonological content. At this point the conditions and context for allomorph choice become relevant. The choice of one allomorph over another is based on the precedence of vocabulary items and the most specified viable canditate wins this selectional process. Specificity is stated via the Subset Principle:

(3) Subset Principle (Halle 1997)
"The phonological exponent of a Vocabulary Item is inserted into a position if the item matches all or a subset of the features specified in the terminal morpheme. Insertion does not take place if the Vocabulary Item contains features not present in the morpheme. Where several Vocabulary Items meet the conditions for insertion, the item matching the greatest number of features specified in the terminal morpheme must be chosen."

In the absence of any vocabulary item specified for the context at hand, the default or "elsewhere" allomorph is chosen (Halle and Marantz 1994; Kiparsky 1973). Vocabulary insertion rules may be specified for the "inner" or "outer" local context of a morpheme within a derivation. That is, the vocabulary insertion rule can make reference to features of a hierarchically higher or lower node. Concretely, since the syntactic structure and morphosyntactically relevant features are present prior to vocabulary insertion, and insertion of the phonological content of morphemes proceeds from the inside out, the vocabulary insertion rule can access morphosyntactic features from both directions but phonolological structure only of an already inserted and thus hierarchically lower structure (Bobaljik 2000; Bonet and Harbour 2012). The locality definition that I will adhere to in this thesis is the one by Embick (2010, 2015). It states that allomorphy is only triggered by linear adjacence. That is, only the previously hierarchically next node can trigger allomorphy. Turning back to vocabulary items, consider as a very simple example the vocabulary items for the English past tense:

(4) Vocabulary Items for the English Past Tense (Embick 2015, p. 184)

a. T[+past] ⇔ -t/$\sqrt{bend}$, $\sqrt{leave}$, . . .⌒__

b. T[+past] ⇔ -Ø/$\sqrt{hit}$, $\sqrt{quit}$, . . .⌒__

c. T[+past] ⇔ -ed

There are three vocabulary items which pair the syntactic head T carrying the feature [+past] with an exponent. All three rules include an exponent (*-t*, *-ed* or *-ø*) and a context which specifies that the respective item is chosen only when this context is encountered. In the case of the first vocabulary item, it is the position concatenated to the right of a certain list of roots, including $\sqrt{leave}$. The last rule does not include any context and is thus the elsewhere exponent if no condition for another exponent is met. In the case of the first two vocabulary items, note however, that these call for linear adjacence between the root and the past tense morpheme which is not present given the concatenation statement in example 2 since *v* intervenes. To solve this issue, there is another postulated mechanism concerning locality, namely Pruning (Embick 2010, 2015). It states that morphemes without an overt exponent are either deleted or irrelevant for locality. Locality for allomorphy then means that two nodes α and β are local to each other iff no overt node intervenes, and can trigger allomorphy. In the case of the English past tense, *v* is null and therefore deleted. Now *leave*[6] and the past tense morpheme become linearly adjacent and enter the necessary locality condition for the right past tense allomorph to be inserted:

(5) *leave*⌒T[+past]

This short overview served as the introduction of the theoretical morphological apparatus which will be employed in the analysis of the Estonian illative allomorphs in this thesis.

[6]I will assume, following Kalin (2019), that at this point the root (and *v*) have already been inserted.

2.3 The illative case

The Estonian illative case has a rather large number of different allomorphs, as listed in Table 9 below. This is the maximal list compiled from M. Erelt, T. Erelt, and Ross (2007), Kaalep (2009), and Kio (2006). Judging by their strongly differing phonological shapes (besides *-se* whose status is controversial) these are true suppletive allomorphs and can not be explained by mere surface allomorphy since there is no phonological rule which could link their phonetic shapes. In the table below, the nouns are given in the genitive or partitive simply for ease of exposition, and their respective illative forms.

Allomorph	Productive?	Examples
-sse	✓	*redeli* : *redelisse* 'ladder', *sama* : *samasse* 'same', *žel`ee* : *žel`eesse* 'jelly', *liiga* : *liigasse* 'league'
-se(?)	✓	*t`öölise* : *t`ööl`isse* 'working', *`alguse* : *`alg`usse* 'start', *`endise* : *`end`isse* 'former'
-ø	✓	*d`ušši* : *d`ušši* 'shower', *k`aela* : *k`aela* 'neck', *p`aati* : *p`aati* 'boat', *seminari* : *seminari* 'seminar'
-µ	✘	*rida* : *r`itta* 'row', *pere* : *p`erre* 'family', *talu* : *t`allu* 'farm', *kino* : *k`inno* 'cinema'
-hV	✘	*s`uu* : *suhu* 'mouth', *`öö* : *öhe* 'night', *s`oo* : *sohu* 'bog'
-de	✘	*s`uur* : *suurde* 'big', *k`eel* : *keelde* 'language', *l`ääs* : *läände* 'west'
-te	✘	*käsi* : *k`ätte* 'hand', *vesi* : *v`ette* 'water'
-da	✘	*mõni* : *m`õnda* 'some'

Table 9: Allomorphs of the illative

The first allomorph, *-sse*, often referred to as "long illative" (M. Erelt, T. Erelt, and Ross 2007; M. Erelt et al. 1995) is the elsewhere allomorph. It is the allomorph which is used for the plural illatives of all stem classes and for the singular forms of many other classes which do not share any prosodic properties. Furthermore, if certain roots exhibit parallel forms of the long and the short illative, one of them is always the elsewhere *-sse* (Kaalep 2009). All the other allomorphs are referred to as "short illatives" and their occurence is constrained to only the singular illatives of certain classes of stems. Concretely, their distribution follows

certain prosodic requirements like foot structure, which also is a cue for declination class membership of a noun, or simply a list of roots in the case of the unproductive allomorphs. The productive null suffix is one of the first kind: it is only available for a certain class of stems. Stems ending in *-se* like *t`öölise-* 'working' also show a kind of shortened illative where the long form attaches to the stem and the *-se* of the stem gets deleted, thus yielding *t`ööl`isse* 'working.ILL', and not *t`öölisesse*. The status of this allomorph is controversial and it is unclear whether this is a genuine case of real allomorphy or simple haplology (M. Erelt et al. 1995; Kaalep 2009). Therefore, this illative allomorph will not be included in this thesis.

A number of roots exhibit another interesting strategy of illative singular formation which involves gemination of a stem internal consonant segment marked here as *-μ* allomorph. In how far this strategy of illative formation is still productive and whether it can be argued to synchronically involve lengthening will be discussed in section 3. In any case, the appearance of this allomorph is also constrained to certain roots with a Q1 CVCV structure. The four last allomorphs *-hV*, *-de*, *-da*, and *-te* are no longer productive and only used with a small number of roots.[7] Thus, the allomorphy of the illative is in these cases clearly triggered by the root it attaches to. For example, although the suffixes which are subsumed here as *-hV* all attach to a single (CV:) syllable, most other roots of this structure do not select this allomorph. For example, *l`uu* 'bone' uses *l`uusse*, and *intervj`uu* 'interview' with main stress on the last syllable uses *intervj`uusse*. The context for the respective vocabulary item simply is the root, and these allomorphs will not be part of this thesis either. Therefore, the rest of this thesis is concerned with the distribution and analysis of the suffixes *-sse*, *-ø*, and *-μ*. In what follows, the term "short illative" will refer to both the null suffix and *-μ* and "long illative" to *-sse*.

Recall from the considerations of the paradigm of *k`app* above that there is another factor differentiating the illative allomorphs, namely which stem they attach to. Irrespectively of what changes occur in the stems, the short illative always builds on the same stem as the partitive and the long illative on the same

[7]The diachronic relationship between the last three can still be observed in Finnish: *suur-ta* 'big.PRT', *vet-tä* 'water.PRT', *mon-ta* 'many.PRT'.

stem as the genitive. This raises an interesting question: How can different allomorphs of a single morpheme select different stems or trigger stem changes? If parallel forms for the null and the long illative do indeed exist, this has far-reaching consequences for the analysis of both the illative and the stem allomorphs. Before turning to this question more thoroughly in this thesis, a possible alternative approach involving two separate cases and/or free variation of the allomorphs will be rejected in the next section to delineate the problem more clearly.

2.3.1 No free variation and only one feature

A problem for any formal account of the illative allomorphs is the purported rather free variation of the short and long allomorphs. Normatively oriented grammars like M. Erelt, T. Erelt, and Ross (2007) claim that the long illative can be used with almost any noun in the singular and that the short illative is the morphological exception which requires a certain phonological structure. According to these views, variation between the zero illative and the long illative for words such as *k`arpi* vs. *karbisse* is rather free and both variants are normatively correct. However, while the short illative is indeed constrained to certain stem classes, the long illative is not as widely used as claimed by M. Erelt, T. Erelt, and Ross (2007). If there were free variation of the exponents in certain stem classes, one way of formalizing this variation would be positing two different illative features selected under certain conditions or two different cases altogether.

If the elsewhere allomorph was in principle available for any stem without being overridden by the more specified vocabulary item for -ø, this variation would have to be explained not by reference to the root but by different featural triggers outside of the root. Recall from section 2.2 that in DM, there are three possible triggers for allomorphy of the same morpheme: morphosyntactic or phonological features on the "inside" or morphosyntactic features on the "outside". For the illative allomorphs, an allomorph trigger on the inside can be rejected since we are dealing with purported variation of one and the same root which only has one possible phonological exponent in the relevant cases discussed here and only one root identifier. A trigger from outside is neither viable since it is un-

clear what that trigger should be, as for example, number is hierarchically lower than case in Estonian. What is left is the possibility that the allomorphs of what has been called the illative are actually the exponents of different features or cases. At the same time, these different cases could with their morphosyntactic features trigger stem allomorphy from the outside. The main point to be made in this section is that since there are no reasons to assume that there are different morphosyntactic features which are responsible for stem and allomorph selection, allomorphy of the illative cannot be explained by two separate cases or features.

There have in fact been ideas to implement the short illative as separate case, namely the "aditive" (Viitso 1976; Viks 1982a,b). It is argued to differ from the long illative in semantics, verbal government and in that it is excluded from the regular formation of case forms (e.g. Viks 1982b).[8] In DM, a separate morphological case would for the most part be regarded as spelling out different structure or a different case feature. One argument for two different features would be differing semantic interpretations of the long and short illative. However, while some short illative forms, especially geminated illative forms, are lexicalized,[9] there is no semantic difference between many pairs like *peresse* und *p`erre* 'family.ILL' or *talusse* and *t`allu* 'farmhouse.ILL' in isolation. This is most straightforwardly explained by only one feature. With respect to syntax, the existence of a morphosyntactic feature could be motivated by showing that it has morphosyntactic effects and argued against if there are none. In the case of a possible short and long illative feature, some other observable differences like selection should be observable. M. Erelt, T. Erelt, and Ross (2007) indeed claim that besides stylistical and minor semantic differences (e.g. preferred usage of the short illative in idioms and multi verb constructions), a syntactic divide between the forms can be found in the selection by certain verbs and nouns derived from them. It is claimed that in a non-local meaning, often when selected by certain verbs like *kiinduma* 'attach', *suhtuma* 'relate', or *armuma* 'fall in love' (or their participles), only the long illative can be used. The following example is given (M. Erelt, T.

[8]Note however that not all authors who argue for syntactic and semantic differences between the allomorphs view the aditive as separate case, e.g. M. Erelt et al. (1995). See Kio (2006) and Raag (1998) for a history of the aditive.

[9]For example, *kätte*, lit. 'into the hand', 'into exposure'; or *ellu* lit. 'into life', "into realization".

Erelt, and Ross 2007, §M55):[10]

(6) *Ta oli **jõesse** / ***jõkke** lausa kiindunud.*
s/he was river.L.ILL river.S.ILL quite attached

S/he was quite attached to the river.

M. Erelt, T. Erelt, and Ross (2007) claim that the short illative cannot be used when selected by *kiindunud*. However, counterexamples to this claim are easy to find. In the following sentences, *kiindunud* appears with a short illative as its complement. Note in example 7b that Estonian shows concord in the nominal domain:[11]

(7) a. *Ning kas te olete kõik Tehnika **maailma** kiindunud*
and whether you.PL are all technology world.S.ILL attached
vä??
EMPH

And you all are very attached to the world of technology, or what??

b. *nad on kiindunud **ühte** ja **samasse** **asja***
they are attached one.S.ILL and same.L.ILL thing.S.ILL

They are attached to one and the same thing.

c. *kes oli **Ernsti** tohutult kiindunud*
who was Ernst.S.ILL incredibly attached

who was very fond of Ernst

This means that the statement from M. Erelt, T. Erelt, and Ross (2007) should rather be read normatively than as genuine synchronic description, as the short illative occurs in this context in forums, newspapers and books, and therefore in very different registers. The examples show that speakers use the normal form of

[10] Simplified by me, JH.

[11] All examples were retrieved by using the corpus "Tasakaalus korpus" at `www.keeleveeb.ee`. 7a is from an internet forum, 7b from belletristic literature and 7c from the 17.3.1993 issue of the newspaper "Eesti Päevaleht".

the illative they would use in any other environment without taking any selection by *kiindunud* into account (Hasselblatt 2000; Kaalep 2009; Kio 2006). Many of the purported parallel forms do not even surface in corpora at all. In the corpus in Tšernova (2014) there were 680 instances of *jõkke*, but not a single one of *jõesse*, although the latter is explicitly used in the example given in EKK above. Example 7b is especially telling with respect to insensitivity to selection as it shows forms of the long and the short illative in the same DP that is selected by *kiindunud* (instead of *ühesse ja samasse asjasse*). Without going through other cases of putative selection of a specific form of the illative and thus the necessity for two separate features, I will conclude that variation cannot be explained by different illative features. An account with only one case or feature can also easily explain the "mismatches" in noun-adjective(-determiner) concord as in 7b above and in the examples below.[12]

(8) a. *sured ja lambist kohe laksti uuesti* ***uude kehasse.***
die.2SG and lamp.ELA soon swoop again new.S.ILL body.L.ILL

you die and, swoop, randomly get into a new body again

b. *Kaalusime mõtet, kas ei peaks* ***sellesse uude***
weigh.PST.1PL thought.PRT whether NEG should this.L.ILL new.S.ILL
majja *kolima*
house.S.ILL move

We pondered whether we should move into this new house

The above examples show different allomorphs in noun-adjective concord. In 8a, the adjective *uus* is marked by the unproductive illative suffix *-de* and the noun *keha* is in the long illative. In 8b, *selle-*, which is used similarly to a determiner here, is in the long illative, and the adjective and noun are both in the short illative, albeit with the different allomorphs *-de* and *-μ*. Assuming only one feature can easily explain why examples like these exist – it is simply the same illative feature or structure in concord (see Norris 2014). If two features were

[12]The first example was retrieved from the corpus "Foorumid" accessible at `www.keeleveeb.ee` and is from a now defunct internet forum, the second example is from `https://runnelsimo.blogspot.com/2012/09/`.

involved here, the description of the allomorphs would have to resort to the same two vocabulary items for each illative feature, if both were in fact possible, and to complicated blocking mechanisms. I conclude on the basis of these findings that there is only one illative case. Therefore, the allomorphy of the illative morpheme is not caused by different illative features but only by different contexts for the same feature, and therefore can only be inward conditioned allomorphy triggered either by phonological or prosodical or stem class features. Thus it is clear that what we are dealing with here are in fact allomorphs of the same morpheme and that for the most part the context for their choice can be clearly defined. The next section will address the question of productivity of the illative allomorphs in more detail, especially of the geminated illative.

3 Productivity of the illative allomorphs

This section deals with the productivity of the different illative allomorphs. First, the statements from the normatively oriented EKK will be given as "baseline" for evaluating different claims made in the literature. After that, a number of corpus studies pertaining to the usage of the different allomorphs will be reviewed with the question about their productivity in mind. The notion of productivity itself will then be discussed. Then results and shortcomings of these studies are evaluated and consequences of these studies for the thesis at hand will be discussed before turning to the results of the corpus study conducted for this thesis.

3.1 Corpus studies about the illative

Since *-sse* is the elsewhere form of the illative, it is undoubtedly productive. The view about (all) the short illatives held by EKK and EKG is that there is a null suffix for the *sepp* and *seminar* classes and certain prosodically heavy suffixes, and suppletive stem allomorphy in the case of the geminating illatives of the CVCV class. Crucially, it is claimed in concordance with ÕS that basically for all stems with possible short illatives there exist parallel long illative forms. Under this view, *-sse* is not only the unspecified elsewhere form, but can also override the more specified allomorphs, whereas the short illatives are somehow exceptional

(Kaalep 2009). These sources claim that the long illative can in principle be used with every stem and only in some noun classes or single instances of nouns it is possible to use the short illative. The main problem with this statement is that parallel forms indeed do exist, but they are far from being evenly distributed. In recent years, these different claims have generated discussion about how to best describe (and formalize) the short illative within the Estonian noun system and how to predict the usage of one form over another. To clarify the actual situation of illative usage in Estonian, a number of corpus studies has been conducted in the course of the last two decades. These will be reviewed in what follows.

Raag (1998) conducted a corpus study monitoring illative usage over time. The focus of her study was to investigate the impact of language planning on Estonian morphology including illative singular usage, hence the diachronic focus. The material is taken from newspapers and spans the time frame from 1887 to 1991. In total, the usage of the short illative increased during this time from about 30% in (north Estonian) Tallinn newspapers and 45% in (south Estonian) Tartu newspapers to over 60% in Estonia in total. Concerning the separate allomorphs, Raag (1998) notes that the usage of geminated illatives follows the general distribution of short illatives over time in that geminated illatives became more common to use in general. The null illatives of the respective stem classes (*sepp*), however, dominate throughout the whole period of the corpus study and the long illative of this stem class had become a marginal form already in the 1923 sample. Raag (1998) does not find a clear correlation between verbal government and long illative usage and attributes apparent effects of this kind to normative pressure, caused by the desire of language planners and normativizing institutions to systemize the parallel forms. The few long illative forms in later occurences of nouns of the *sepp* class occur precisely after the verbs which are normed to govern long illative usage. She concludes her study by stating that the actual usage of the different illatives in the corpus does not reflect the normative statement that the long illative can be used with any stem.

Hasselblatt (2000) reviewed (among others) the questions whether the *-sse* allomorph is really the "usual" form to use and whether syntactic differences as involving different selection do in fact exist. He conducted a small-scale corpus study with 2000 total illative singular tokens where the material was constrained

to a few 1999 issues of the newspapers *Eesti Päevaleht*, *Sirp*, and *Postimees*. The results showed that at least from the standpoint of frequency the short illative was the preferred form to use. Of the 2000 tokens in the corpus, 1138, or 56.9% were in the short, and 862, or 43.1% in the long illative. Not taking doublets into account, the results were slightly more balanced: of a total of 970 illatives, 515, or 53.1% were short and 455, or 46.9% were long.[13] Without going into the details here about which lemmata actually were used with which allomorph, Hasselblatt (2000) concludes that if it is possible to use the short illative, it is used. Only 19 lemmata were found to be used with both the long and the short illative, which clearly speaks against widespread variation. For example, long parallel forms like *majasse* (versus *m`ajja*) 'house.ILL' or *j`õesse* (versus *j`õkke*) 'river.ILL' as in example 6 from EKK above could not be encountered at all. Concerning possible differences in selection, Hasselblatt (2000) checked all verbs which cooccurred with the illatives in his corpus and found no clear tendencies towards selection. The numbers are rather small but provide a preliminary outlook.

Kio (2006) conducted a detailed corpus study taking into account lemmata of three noun types as defined in ÕS for which, according to ÕS, two illatives were possible allomorphs. These classes were the *pesa* class with variation between *-sse* and gemination, the *tööline* class with variation between *-sse* and *-se*, and the *sepp* class with variation between *-sse* and *-ø*. The first class is a subclass of Q1 CVCV and basically the noun class which will be of interest in the corpus study conducted in this thesis as well. The material was collected from mixed corpora (newspapers, belletristic texts, . . .) and spanned a large time frame from the year 1890 to 2001. The results indicate that of the *pesa* class, 88.7% of the illative tokens were short, that is, geminated illatives like *p`essa* 'nest.ILL'. All but three stems were preferrably used in the short illative (including *kino* 'cinema' with 96.2% short illatives), with *eru* 'retired' to a 100% (although occuring mostly in idiomatic structures), and with *kava* 'scheme' to 67.2% as extremes. The three outliers were *klubi* 'club', which was used to 98.8% in the *long* form instead of the short one, *toru* 'tube' with 73.3% long forms, and *haru* 'branch' with 72.5% long illatives. While there was no high variation between long and short forms

[13]Note that this is not entirely equal to total lemma frequency, since possible parallel forms of one lemma with the short illative and one with the long could have been counted multiple times.

of one stem (preference of one allomorph over the other was clearly over 50% in every case), rather widespread variation among the whole class of roots could be observed.

For the *sepp* class the results are very clear. Of all illative tokens of the lemmata in this class, a total of 97.2% were short forms, that is, with a zero allomorph. The lemmata *järv* 'lake' and *tund* 'hour' were used in the short illative to a 100%, and the only instances of long forms of *hing* 'spirit' and *riik* 'state' were from around 1900. There were no outliers like in the *pesa* class above and it can be deduced that the null illative is highly productive since this is an open class. The *tööline* class was the most heterogenous class with respect to short and long illative usage. Of all controlled illative tokens, only 40.2% were short, that is, with *-se* as allomorph. For example, *`eelmine* 'previous' was used in the short illative in 73.7% of cases whereas *h`aigus* 'sickness' only in 34.2%. I conjecture that the heterogenity may be a result of the grouping of rather different stems into one noun class by ÕS. For example, the class includes derivations by *-mine*, *-line*, and *-us* and lemmata of possibly different syllable structures like *pr`aegune* and *s`uuruse* (Bye 1997; Kaalep 2010). Additionally, it is not clear whether the short illative of this class (*-se*) is an instance of haplology or genuine prosodically conditioned allomorphy (M. Erelt, T. Erelt, and Ross 2007; Kaalep 2012). Furthermore, Kio (2006) suggests that the lack of the short illative can be explained by (output-oriented constraints) avoiding two Q3 syllables after another as in words like *`alg`usse* 'beginning.s.ILL'. I suspect that the heterogenous results can be traced back to the interplay of (some of) these factors. Kio (2006) also shortly discusses whether selection by verbs or their nominal derivatives has any effect on the illative allomorph and concludes that this is not the case. Only in the case of *suhtuma* 'relate' a slight preference for the long form was noted. This case can possibly be attributed to the register of the texts in the corpus.

Since the focus in Kio (2006) is to study variation between the illative forms in stems where this is theoretically possible, only lemmata with a minimum of 10 occurrences of both illatives were selected. This excludes an unknown number of stems which could serve as possible input to short illative formation (or can not, in fact) and thus skews the results in favor of frequent stems when the total percentage of short illatives of a noun class is calculated. The sample was

not balanced and there were great differences in frequency, with a range from 4588 tokens of *elu* 'life' versus 29 tokens of *haru* 'branch' in the *pesa* class. This is not per se problematic but skews the distribution of the short illatives in favor of highly frequent lemmata if only token frequency of short versus long illatives is calculated for the whole stem class in the end. That is, the amount of short illatives could be exaggerated for the stem classes due to overrepresentation of some short illative forms. Similarly, the low variation between the two illative allomorphs for some single stems could be attributed to them being frequent stems. Highly frequent words are usually irregular, and vice versa, irregular forms are often highly frequent. They are therefore predicted to retain idiosyncrasies in their paradigms more easily, even if from a viewpoint of regular formation this does not reflect productive rules of formation. In any case, the numbers corroborate the findings in Raag (1998) and Hasselblatt (2000) and make a stronger statement for the point that there is no free variation and that the null illative is the preferred form to use in the *sepp* class.

While the main focus of Kaalep (2009) is not a corpus study, some numbers from a search in the corpus "Morfoloogiliselt ühestatud korpus"[14] from the 1990's and early 2000's are given, showing the ratio of long and short illative usage for different stem classes. It is not clear, however, whether the whole class was checked or only a subclass of the stems. Anyways, the distribution of the null and the long illative in the *sepp* class is a clear 297 : 7 and of all derived nouns in *-ik* which can also have a null illative, 34 : 4 which clearly speaks for the productivity of the null suffix for certain stem classes. The *seminar* type showed a relation of 14 : 2. The CVCV class was found to behave less uniformly than previous studies might have suggested. This is probably because in contrast to previous studies no preliminary selection of stems according to possible parallel forms as suggested by normative sources like ÕS was undertaken and simply all stems in the sample were counted. The ratio between geminate and *-sse* illatives was 47 : 43 in this class, and slightly differed in its subclasses: in the open *pesa* class the ratio was 31 : 39 while in the closed *jõgi* class it was 6 : 1. Kaalep (2009) concludes from these results that the application of the rule for geminated illative formation (contra EKK who assume root suppletion) is unproductive.

[14] Accessible at `https://www.cl.ut.ee/korpused/morfkorpus/`.

A large-scale study similar to Kio (2006) was conducted by Tšernova (2014). The study differed from Kio (2006) mainly in that some other highly frequent closed class lemmata with a CVCV structure like *käsi* 'hand', *jõgi* 'river', and *tuba* 'room' were taken into account which all occurred in the corpus to a 100% in the short illative, and that the newspapers did only span the years 1995–2007. In an again unbalanced sample, of all illative tokens of the selected classes (*pesa*, *jõgi*, *seminar* and, *tööline* class) 93.2% were short. Of course, with the even larger number of illative tokens in general (115309), the frequent stems skew the relation even more. For example, the short illative of *käsi* 'hand', *k`ätte*, had a token frequency of 15847 while the short illative *j`ärve* 'lake' only occured 44 times. Irrespectively of these concerns, the data one again shows that frequently used CVCV stems are available basically only in the geminated short illative and that the null illative is the preferred allomorph in the *sepp* class.

The papers A. Metslang (2015) and Siiman (2016, 2018) form a series of statistical analyses about which morphosyntactic factors and which semantic factors of a root or factors pertaining to the sentential context are relevant for long versus short illative choice. The methods vary slightly: A. Metslang (2015) uses a non-balanced corpus and univariate analysis with an unbalanced corpus, which however results in some numbers too small to be significant and not being able to clarify what variables predict the use of one illative form over another. Rather, the univariate analysis checks with what variables one form is used in general. Siiman (2016) also uses an univariate analysis but this time the corpus is balanced, i.e. the number of short and long illatives is identical. Siiman (2018) is a multivariate analysis of the sample in the previous studies. For a summary and contextualization of A. Metslang (2015) and Siiman (2016, 2018) see Siiman (2019).

A. Metslang (2015) undertook a statistical analysis of the reasons for the usage of the short illative versus the long illative, caused by the structure of the root. The corpus under consideration was the same as in Kaalep (2009), namely "Morfoloogiliselt ühestatud korpus". Similarly to Kio (2006) the starting point to choose roots were those which could, according to ÕS, be used with both the long and the short illative. The stems themselves were randomly sampled from a corpus and parallel forms were included in the sample. The difference between the most frequent form *k`ätte* 'hand.ILL' to single instances of a few lemmata

in the illative was 48 : 1. Now, both the short and the long illative forms were controlled for correlations with a number of characteristics of the stem, namely the presence of length or grade alternations ("weakening" versus "strengthening"), the quantity of the nominative stem, stem-final alternations like *inimene* : *inimese* 'human', whether the last segment of the nominative stem was a vowel or a consonant, and finally the number of syllables in the genitive. Of these properties all but two turned out to be significant. The two non-significant properties were the kind of stem alternation (*astmevaheldusmall*) and the quantity grade of the stem. For example, nouns undergoing stem alternations in general favored the short illative and those without the long illative (A. Metslang 2015).

Siiman (2016) checked the short and long illatives for syntactic and semantic factors governing allomorph choice. The data stems from all corpora accessible at `www.keeleveeb.ee` with the exception of Tasakaalus Korpus and Murdekorpus (so etTenTen which will be used in the corpus study for this thesis was included too). The morphosyntactic factors which were checked this time were syntactic category (e.g. noun vs. adverb), sentential status (argument vs. adjunct), government by verbs, and whether the illative was part of a fixed expression. The semantic factors were whether the lemma was a common or a proper noun and semantics of the verb. Of these were found to be significant: government, fixed expression, and the semantic categories of common and proper noun. This time, the selection of forms was more balanced: 420 short illatives were contrasted to 420 long illatives.

For verbal government, which has been argued not to play a role in the choice of the illative allomorph by other authors (Hasselblatt 2000; Kaalep 2009; Kio 2006; Raag 1998), Siiman (2016) found a significant effect that long illative usage after a certain class of verbs was preferred. However, Siiman (2019) remarks that it is rather a tendency that in a government structure the long illative is more likely to be used, and that not all governed nouns of all verbs behave the same. With respect to status as idiom, the short illative is significantly more frequent, especially with idiomaticized body part expressions like *s`ilma* 'eye.ILL', *k`ätte* 'hand.ILL', or *pähe* 'head.ILL'. Regarding the semantic properties, common nouns were used more frequently with the short illative and proper nouns with the long illative. This is in line with the observation that homophonous common nouns

like *maja* 'house' and proper nouns like *Maja* do not share the same inflectional properties: while *maja* has a geminated illative and a null partitive, *Maja* only uses the long illative and a partitive in *-t* (Blevins 2008).

Siiman (2018) is a multivariate reanalysis of the material of her previous papers. The results are basically the same as in the previous papers and as main reasons for short illative usage the structural properties of the root surface (Siiman 2018, 2019). Controlling for the interplay between different variables, Siiman found that the most significant predictor for choosing the short illative was the presence of weakening length alternation, while for stems with no or strengthening length alternation, three more predictors emerged: quantity degree of the base form, number of syllables in the genitive stem, and the stem-final alternation pattern. The significant semantic factors were the same as in Siiman (2016), namely: status as common noun, in which case the short illative was preferred, government by certain verbs like *suhtuma* or *armuma*, where the long illative was preferred, and multi-word expressions, where mostly short illatives were used. At least the first results are in line with previous research and the claim in this thesis that the choice between the illative allomorphs is mostly conditioned by prosodic properties of the root.

3.1.1 Discussion

In summary, previous corpus studies converge in a number of facts about the (short) illative. Starting with Raag (1998) and Hasselblatt (2000) basically all other studies have found or agree with the fact that if it is possible to use the null illative with a root, it is preferrably used. While the long *-sse* illative clearly is the elsewhere form, parallel forms of long and null illatives with the same roots are neither in free variation nor evenly distributed and the long illatives occur rather marginally in the *sepp* class. The same authors also conclude that the distribution of the different forms under government of verbs or derived nouns is marginal at best and that actually occuring instances of government can be attributed to normative language usage. The only outlier with respect to this claim is Siiman (2016) who finds a statistically significant correlation between governing verb and illative allomorph. This will be commented on below. While some of the studies

(Hasselblatt 2000; A. Metslang 2015; Siiman 2016, 2018) treat the null illative and the geminated illative together (that is, as the short illative), others (Kaalep 2009; Kio 2006; Tšernova 2014) explicitly separate the two allomorphs. The clearest indication of the actual difference in productivity of these allomorphs can be found in Kio (2006) where the stems of the open *pesa* class show rather inconsistent behavior across the whole stem class while the whole open *sepp* class shows very consistent results. This is an expected result if the null illative suffix is productive for the *sepp* class. On the other hand, having high discrepancies between 100% of short illatives for one lemma and only a few instances for others in the *pesa* class despite the same prosodic structure is also expected under the assumption that the context for the geminating illatives of this class is a list of roots which can in principle vary among speakers, as also indicated by marginal attestation of new forms like *klubi* : *kluppi* 'club', *pubi* : *puppi* 'pub', and *varu* : *varru* 'reserve' in a subgroup of speakers when the forms are frequently used (Kaalep 2009). Finally, all studies which differentiate between the null and the geminated illative consequently take into account the role of the prosody of the stem or its stem class in determining the selection of the illative allomorphs. Besides the studies mentioned above, A. Metslang (2015) and Siiman (2018) in a way prove this indirectly not by reference to stem classes but by reference to morphological and phonological properties of the stems which can be grouped into stem classes.

Turning to the desiderata of these studies, the first fact worth mentioning is that all corpus studies only check a certain number of the stems of one declination class. This is of course due to various limitations of the studies and the size of the declination classes. One effect of this selectiveness is that across different studies, there was some disagreement about how widespread the usage of the geminated illative actually is, owing to different choices of material. While the *pesa* class roots in Kio (2006) are for the most part preferrably used with the geminated illative, the results in Kaalep (2009) are more balanced between the geminated and long illative. This can plausibly be traced back to the selection of the roots. Kio (2006) used frequent lemmata which, according to ÕS, were possible to use with both allomorphs while Kaalep (2009) did apparently not make this selection (although it remains unclear which roots were included in his sample). The results in Tšernova (2014) are even more "extreme" than in Kio (2006) since

even more frequent stems were included. The thesis at hand attempts to bridge this gap by actually looking at all possible roots of this class. This is possible since the class of unequivocal Q1 CVCV roots is rather small. While the corpus study in this thesis will not reveal anything new about the productivity of the geminating illative, it will show the whole distribution of the geminated illative across all possible Q1 CVCV roots for a better evaluation of this class as a whole.

Second, the corpora in the different studies were quite heterogenous with respect to their time frame and text genres. In a way, this is desirable for cross-comparison of different registers and genres and in fact shows that the general preference for the short illative in some declension classes persists irrespectively of the material. However, using material which mainly contains language from "official" registers like newspapers is likely to show a preference for forms which are oriented on the literary norm. In a way, it is not surprising that by checking texts which are subject to normativized language usage, government by a verb will turn out to be a significant factor as in Siiman (2016, 2019), since normative grammars demand its usage in these structures. A similar point was raised already by Tauli (1968, p. 198) who stated that it is essentially meaningless to try and examine the usage of morphology by Estonian authors, especially since the second quarter of the 20th century, since it does merely reflect the language of the editors of linguistic correctness adhering to official norms of that time. Raag (1998, p. 80) also concludes that newspaper articles, especially from the Soviet period, can not be taken to reflect the writers' original usage of the illative because they have linguistically been cleaned up according to the norms. I suspect that given the normative orientation and advice in ÕS and EKK this is to some extent still a factor today. In Kio (2006), notably, at least 3 of the 15 long illatives of *kino* were used after *puutuma* or *suhtuma* (not all 15 examples are given) while of the five short illatives of *klubi*, three are from belletristic literature, and according to Kio (2006) with the intent of replicating spoken language which may indicate a difference in usage relating to registers.

Register and normativity therefore seem to be relevant factors for illative allomorph choice. In Raag (1998) there indeed are some indications of how illative usage has diachronically been influenced by the language norms of that time. The most striking example of this is the usage of the geminated illative for the words

küla ‘village’, *elu* ‘life’, and *talu* ‘farm’. Up to Raag’s (1998) newspaper sample of the year 1907, these stems are almost exclusively used with the long illative, 1923 already 70% short illatives can be found, and in 1938 almost all illatives of these stems were geminated. The turning point coincides with the efforts by Johannes Aavik, arguably the most influential Estonian language reformer, to promote the usage of the short illative. Another similar development which can be tied to a certain point in time is, again, the distribution of geminate illatives in general. In the 1961 sample geminate forms make out 86% of illatives, but only 83% in the 1991 sample. The difference is small but unexpected since short illative usage has steadily been on the rise over the whole sampling time frame (see also Kio 2006). Raag (1998) traces this difference back to the publication of the rather liberal 1976 edition of ÕS which in contrast to the two previous editions no longer normed the geminated illative for each stem separately but rather for stem types and granted speakers and writers more freedom. Since these flucuations can be explained by the prevalence of different language norms over time the corpus search at hand should first, try to take into account nonstandard language to abstract away from these sociolinguistic and language policy factors, and second, include as recent data as possible to get a good synchronic picture of illative usage. With these facts and desiderata in mind, we can now turn to the corpus study for this thesis.

3.2 Corpus study: Productivity of the geminating illative

One question this thesis sets out to answer is in how far a geminating short illative is productive for Estonian CVCV roots. The circumstance which gives rise to this question is the apparent discrepancy between the treatment of the geminated illatives in EKK as suppletive stem allomorphy (M. Erelt, T. Erelt, and Ross 2007, p. M55) and the homomorphism between all geminated illatives (all bisyllabic Q3 with a single vowel and long consonant) as well as the (admittedly marginal) attestation of new forms like *varu* : *varru* ‘reserve’, *klubi* : *kluppi* ‘club’ or *pubi* : *puppi* ‘pub’. This root class also shows some variation in the applicability of a geminated illative for loans: *kino* : *kinno* ‘cinema’ is highly preferred over the long form *kinosse*, *klubi* : *kluppi* seems to be tied to a colloquial register (vs. the long illative *klubisse*), and other similar stems are not used in the short illa-

tive at all. Furthermore, no complete dataset covering the whole Q1 CVCV class of roots is available from which some general tendencies for geminated illative choice could possibly be inferred. For the discussion about the productivity of the geminated illative it is necessary to assess how widespread variation among the same class of stems actually is. The purpose of this corpus study is to provide a clear picture of geminated illative usage for Q1 CVCV nouns today by checking all possible nouns in this class. This is possible because the number of roots in this class is rather small. While the corpus study will not provide any new insights into whether the geminated illative singular is productive (it is not), it will provide this dataset and illustrate the current amount of variation found in this class. I first provide a short overview of the notion of morphological productivity and I then motivate a method for measuring the productivity of the geminating illative for a corpus study. After that, the results of the corpus study and the implications for the analysis of the geminated illative are discussed.

3.2.1 Morphological productivity

Productivity is a feature of morphological processes or rules. According to Bauer (2001), productivity can be described as the potential of a morphological process to coin new words repetitively, and if this is the case, the extent to which this process is exploited in language use (Bauer 2001, p. 211). While productivity usually goes hand in hand with frequency of word forms produced by the process or its semantic coherence, these factors are prerequisites for productivity rather than its description. Consequently, "attestation of a new word must be taken as prima facie evidence that the morphological process giving rise to the coinage is productive" (Bauer 2001, p. 58). Bauer (2001) is mostly concerned with derivational morphology, and not inflectional, which is the focus of this thesis. I assume that, since DM does not make a categorical distinction between these types of morphology, the insights can for the most part be carried over to the productivity of inflectional morphology. Furthermore, it will be discussed in section 3.3.1 that some criticisms pertaining to morphological productivity do not hold for inflectional morphology in the study at hand.

The corpus study in this thesis is concerned with the distribution of the long

and the geminated illative among Q1 CVCV roots. In the case of vocabulary insertion rules as employed in DM, competition between exponents for one morpheme is modeled by specificity and/or precedence within the vocabulary insertion rule. If two exponents compete for insertion for one morpheme, the more specified one wins. If there is no item specified enough, the elsewhere case applies (see section 2.2). Since the elsewhere is not specified for any context it has to be productive by definition since the number of bases it can apply to is potentially infinite. However, while all elsewhere cases must be productive, they neither need to exhaustively subsume all productive rules in a language nor be predominant by token frequency. This can be illustrated by the example of German plural formation (Bauer 2001). Abstracting away from the availability of the allomorphs for different genders, all but the second pattern are productive at least for one gender of nouns. The last allomorph *-s* is the elsewhere today (Wiese 1996, p. 138), but it is far from being the most common one. One study found that forms with *-s* constitute only 0.02% of plural tokens in an average text (Mugdan 1977, p. 97), while the number of types is claimed to lie between 1% and 4% (Marcus et al. 1995, pp. 227–228).

SG	PL	Gloss	Allomorph
das Fenster	die Fenster	'window'	*-ø*
der Vater	die Väter	'father'	Umlaut
der Wald	die Wälder	'wood'	Umlaut and *-er*
das Schaf	die Schafe	'sheep'	*-e*
der Stuhl	die Stühle	'chair'	Umlaut and *-e*
die Schule	die Schulen	'school'	*-en*
das Baby	die Babys	'baby'	*-s*

Table 10: German Plural Allomorphs (Bauer 2001, p. 61)

Thus, frequency of a word form alone cannot be directly equated with productivity, since the number of bases for a possible derivational affix may be small, but if the affix can attach to every of its possible bases, the process is productive, albeit in a very restricted domain. Conversely, a vocabulary item may also be productive in a certain context although it is not the elsewhere. This is the case in Estonian: the zero illative is productive in the *sepp* class since it is repetitively

used to form illatives of roots in this class, and the class itself is open, so the vocabulary insertion rule can be applied in this context across the board to produce "new" forms. The elsewhere is the *-sse* exponent, however. This means that a vocabulary item in inflectional morphology is unproductive either if the item is specified for a certain class of possible roots which is closed (however the class itself is specified) or the item itself is specified only for a number of roots. The Estonian geminated illative will turn out to be an unproductive vocabulary item of the latter kind.

Productivity usually refers to "societal productivity" (Bauer 2001). That is, a morphological process is considered to be productive if it is used productively in a speaker community. Here, a methodological conflict arises when measuring societal productivity quantitatively via corpus data which has been produced by a large number of individuals is used to justify a conception of invariant morphological computation within an I-language approach which essentially states that the object of research is the "knowledge state", or competence, of a single speaker (Chomsky 1986). This discrepancy has been noted before and a reasonable division between the "competence" and "performance" sides of productivity has been drawn by Corbin (1987). According to her, a morphological process is *disponible* or 'available', if it can be used in the production of new forms, whereas it is *rentable* or 'profitable', if it is actually used in a performance sense (Bauer 2001; Corbin 1987). Availability is an all-or-nothing matter describing whether new forms emerge as outputs of a morphological process, and profitability is a scalar matter, depending on language usage. Another possible conception of the distinction could be in terms of different grammars within one single speaker's mind. Availability describes whether a morphological rule is present within one grammar and profitability whether the outputs of the grammars coincide in certain speech situations. For example, for a morphological process where there is no sociolinguistic variation, the grammars would not differ (or there would not be any different grammars at all), and profitability would equal availability. In cases like the illative where the usage of the allomorphs is tied to certain norms and registers, the outputs of the grammars do not coincide and the availability of a process in each grammar would not match profitability since one grammar is chosen over the other.

The same difference between societal and grammar internal constraints on productivity can also be observed in the distribution of the long and the null illative. Parallel forms of one and the same root within the *sepp* class do in fact exist and therefore the selectional context can not be explained by properties of the root alone but by variation between grammars pertaining to factors outside of narrow morphosyntax like normativity or simply areal or stratal variation (Raag 1998). In other words, the reason why a certain morphological form can or cannot be formed may have its origin in diachrony, language political decision making, or to some extent even individual choices (Bauer 2001), but the narrow computational system in the minimalist implementation is oblivious to such concerns. It is only concerned with narrow morphosyntax. If productivity is understood as exhaustive application of a vocabulary insertion rule over a well-defined class of bases (or as elsewhere) within a grammar, and given the format of rule precedence, any deviation of an expected output of a productive rule must be explained by higher specificity or precedence of another rule, and thus ultimately, a different grammar where this is possible. This means that what is measured in a mixed corpus is a mixture of the outputs of rather different grammars tied to usage in certain speech contexts, with possibly different rules which is in theory not entirely desirable. Nevertheless the corpus studies reviewed above showed that variation of the allomorps, especially for the *sepp* class, is rather small across different corpora.

There is one additional striking difference between these competence and performance side constraints on illative allomorph choice: while there is variation of the short and long allomorphs for some roots, there is never any variation with respect to the combination of illative allomorph and stem allomorph. So for whatever sociolinguistic reason a speaker may choose to use one allomorph over the other (e.g. the long one under the influence of a normatively oriented register), the long illative invariantly occurs with the genitive stem, and the short illative with the partitive stem. That is, while the first condition is "fuzzy" to a certain amount and under the influence of pragmatic and sociolinguistic factors, the latter is invariant and has to be "hard-coded" into the morphology of all Estonian grammars. In other words, while selection of one form over the other can have manifold reasons, some of which do lie outside the explanatory scope of DM,

the regular cooccurrence of the null suffix with the strong stem has a grammar internal footing even across different grammars.

For now, I conjecture that the differences between long and short illative usage for the same roots are bound to different registers and thus different grammars (see the discussions in Walkden 2014, pp. 18–31 and Roberts 2007, §4.2). For the present study of productivity within a DM approach, this means that it is reasonable to try and abstract away from the notion of a process being *rentable* when assessing its productivity, especially in the light of diachronic effects pertaining to the prevalence of different language norms. In the ideal case, the choice of corpus material would be constrained only to a certain register and thus to the grammar that speakers use in this speech situation. For the study at hand, this means trying to focus on written colloquial speech. Before talking about the selection of the corpus for this task, a method for measuring productivity with a corpus will be argumented for.

3.3 Method

Since the productivity of the geminating illative is assessed by conducting a corpus study, this study falls into the field of quantitative approaches to productivity. Very simply stated, under this kind of approach a morphological process (or rule) is more productive the more it is put into action (Fernández-Domínguez 2013). Thus, the most widespread method within this group of approaches is simply counting the number of different forms created by a certain morphological process (for an overview, see Bauer 2001; Fernández-Domínguez 2013). This number is supposed to reflect to what extent a rule is used. As mentioned above, equating only text frequency with productivity does not yet tell anything about whether a process is productive or not. In fact, it can actually measure the opposite since existing irregular morphological forms tend to be frequently used forms. For example, one study found that while the text frequency of regular verbs in English amounts to only 40%, their token frequency is about 96% (Grabowski and Mindt 1995). However, these factors are connected nevertheless: In general, if word forms with certain characteristics form a small class of only few items but occur frequently in texts, the process bringing about the word forms in question is usu-

ally unproductive. Similarly, as has become clear from the discussion above, the geminated illative is at least from the viewpoint of token frequency the preferred form to use in the *elu*-class (Hasselblatt 2000; Kio 2006; A. Metslang 2015). Recall however, that there was some disagreement about usage preference based on the token frequency for the whole class which was attributable to giving frequent forms more weight. Vice versa, if word forms with certain characteristics form a large class but occur only infrequently in texts, the process is productive (Baayen and Lieber 1991; Bauer 2001; Kaalep 2009). Therefore a suitable measurement of productivity has to take other factors into account than only token frequency.

3.3.1 Counting type frequency of short and long illatives of one class of bases

The first approach to productivity taken in the corpus study for this thesis is rather conservative and is based on an early attempt to measuring morphological productivity quantitatively as described in Aronoff (1976). The method is based on the assumption that productivity can be measured by counting all suitable bases for a given morphological process and compare these to the number of attested output types of the rule. This gives a certain relation which can then be compared to other morphological rules, possibly with the same domain of bases. The method can easily be formalized as follows (Baayen and Lieber 1991, p. 803):

$$I = V/S$$

Here I is the index of productivity which shall be calculated, V is the number of actual attested types, and S is the number of types or bases to which the morphological rule could in principle be applied. This means that token frequency is excluded from this measure and every attested instance of a type adds to the count of V. Thus, this approach is very "generous" in measuring productivity since already a single instance of an alleged output form suffices to have an effect on the index of productivity. This approach has been criticized for a number of reasons. The first criticism pertains to the fact that with any productive process, however intricate the restrictions on its bases, the number of bases is potentially infinite, for example due to neologisms or loan words which are adapted into that class. This leads to a possibly infinite S and subsequently a productivity in-

dex converging to zero in the case of productive processes (Baayen and Lieber 1991). Therefore it has been argued that this measure can only be used to measure unproductive processes since their bases form a closed class. Of course, this problem is prevalent in measuring instances of a short illative of CVCV roots as well, but only to some extent as far as I can see. The number of new roots in Q1 is quite small compared to the much larger class of loans in the *sepp* class which pattern accordingly (Pedaja 2006, see section 4.2). The other criticism altogether addresses the rather vague notion of possible bases that the morphological rule could apply to. The only way of obtaining a list of possible bases is a fixed corpus which is likely to miss outcomes of especially productive rules (Baayen and Lieber 1991; Bauer 2001). However, Bauer (2001, pp. 186–187) points out that in a corpus study the problem of providing a measure for S can be circumvented by the following nontrivial assumption: Since a corpus is necessarily a fixed and limited collection of text, V will not represent the total number of actually existing types which are the results of a morphological rule. For example, *blokki* (or possibly *plokki*) is not attested in etTenTen as short illative of *blogi* 'blog', but some marginal examples can be encountered on the internet like the following:

(9) *Viis aastat tagasi kirjutasin siia* ***blokki*** *esimest korda.*
five year.PRT ago write.PST.1SG here blog.ILL first time

Five years ago, I wrote into this blog for the first time.
`https://iibis.wordpress.com/2015/04/28/kevadpaike-ja-jah/`

However, even if a corpus is necessarily only a selection of text it can be assumed that for any affix it is likely that the corpus has sampled potential bases for a morphological rule in the same proportion as it has sampled existing types with that affix (Bauer 2001). That is, because the sample on which a corpus is based is representative in the ideal case, the probability for any token to appear in the corpus – both affixed tokens and bases – is a reflection of the total ratio of usage outside of the corpus. In this case, the number of attested bases in the corpus can be used as factor S (Bauer 2001). Since the starting point of possible bases for the geminating illative in this thesis is a list of CVCV roots gathered from sources independent of the corpus, any root for which not a single illative

singular form (long or short) is attested in the corpus will be excluded from S for the final calculation. Thus, S is in this study the number of all CVCV roots for which there is at least one illative singular in the corpus. Since the basic list for S is established independently of the corpus it avoids the possible problem that some bases which are predominantly used in the short illative in idioms like *eru* 'retired' or *võsa* 'scrub' could happen not to be represented in the corpus in another case form at all. Considering the size of the used corpus etTenTen (270 million words)[15] this is highly unlikely anyways but it makes sure all the forms are represented in the sample. There are indeed some roots of the CVCV class which are not encountered in the nominative in etTenTen. For example, the corpus does not include a single instance of *eru*.NOM 'retired',[16] but 506 illatives.

The biggest problem that Bauer (2001) points out in the discussion about applying the above formula to measure the productivity of nominalizing suffixes in English is that there are many lexicalized nominalizations whose verbal base which they could have been derived from is not included in the corpus, or whose corresponding verbs have actually been formed from the nominalizations through back-formation. In either case, the ratio of nominalizations and possible transparent bases is misleading since the nominalizations have not arisen through a productive word formation process from what looks like its bases (Bauer 2001, pp. 186–189). A similar problem seems to arise in the application on the illatives, namely how to differentiate more lexicalized instances of forms like *marru* 'rage.ILL' or *erru* 'retired.ILL' from other, arguably more regular and semantically more transparent short illatives like *perre* 'family.ILL' or *jõkke* 'river.ILL', since in some grammar and language processing theories these are argued to be stored as whole units and not derived compositionally (e.g. Ullman 2004, see Leminen et al. 2018 for a recent overview). This is a very complex issue beyond the scope of this thesis and for this matter I will simply adhere to the tenet of DM that it is "syntactic hierarchical structure all the way down" (Halle and Marantz 1994, pp. 276–277) and idiomaticized meaning arises only at a later stage in the derivation. As far as I can see, however, neither of Bauer's criticisms

[15] See `http://www.keeleveeb.ee/dict/corpus/ettenten/about.html`.

[16] Among all results for `eru` most instances were names, there were some misspellings of *aru*, and a few partitive forms.

above is a problem for measuring the short illatives. Besides the few instances of roots which do appear in other case forms only rarely, most bases still exist, and if not as stems for synchronic nominal inflection, then as parallel forms. This point can arguably be made for highly frequent forms such as *kätte* 'hand.ILL' which appear in their grammaticalized postpositional function ('into exposure') far more frequently than their transparent nominal counterparts ('into the hand'). In turn, back formation from a lexicalized short illative to a nominal root does not appear to exist.[17]

Now that the main points and criticisms of the first approach have been discussed we can turn to the concrete implementation in this thesis. As discussed, the productivity index in the formula above serves as first indication to compare long and short illative usage of the same class of bases, namely Q1 CVCV roots. The number for S is obtained by manually checking all instances of roots with this structure and counting every root for which there is at least one illative singular in the corpus. Since the indices of the long and short illative shall be compared, the number for V is in one case the number of short illative types in total and in the other the number of long illatives in total. The statement in M. Erelt et al. (1995, §157) that of all bisyllabic Q1 stems ending in a vowel (thus, CVCV) only about 100, or 13% respectively, have a lengthened stem variant that can be "used in reality" ("reaalselt kasutatav") could have been the outcome of a similar approach. All available stems of this class are taken under consideration as S and, presumably, starting at a certain frequency the short illative forms are counted as instances of V which results in an I of 0.13 or 13% while it follows implicitly that if the long illative is a parallel form occurring in reality for all these roots, its I should be 1. In summary, comparing short and long illatives for a whole class of roots serves to give a first assessment of what process of illative formation is more productive with the CVCV class of roots, but it obscures the whole amount of variation found for individual roots in this class. To tackle this issue, another measure is needed.

[17] There is at least one example for a geminated illative as base for further derivation, namely the verb *abi`elluma* 'get married' (< *abi`ellu* 'marriage.ILL').

3.3.2 Comparing token frequences of short and long illatives for each base separately

If gemination does not apply to CVCV roots across the board but is restricted to a list of roots as context, we can expect a small number of frequently used roots showing a high percentage of token frequency for the short illative, a small number of roots showing the total range of intermediate values of long and short illative tokens, and the majority of infrequently used roots showing a near-zero percentage of short illative tokens and instead exhibiting the elsewhere long form. This follows from the character of unproductive morphological cateories. Since irregular morphological forms are usually frequent (Bybee 2007; Haspelmath and Sims 2010), the whole unproductive category (that is, the geminating illatives) will in a corpus be characterized by these few frequent types together with a low number of low-frequency types (Baayen and Renouf 1996). The latter are subject to variation between speakers and are possibly on the verge of transition out of the context for the geminating illative within the language community which results in the application of the elsewhere.

Therefore, in addition to comparing the short and the long illative according to the formula by Aronoff (1976), the ratio between short and long illative usage, i.e. token frequency, will be calculated for each root separately. This has been done on a smaller scale in previous studies, e.g. Kio (2006) and Tšernova (2014), albeit only with a selection of the possible roots. The ratios will then be compared to each other without reference to the total token frequency of single lemmata. This provides a clear picture of how fragmented short illative usage within the CVCV class actually is by checking whether the numbers conform to the generalization in Baayen and Renouf (1996) above and avoids skewing the distribution for the total class of roots by giving highly frequent forms more weight. Simply put, the total token frequency of one illative allomorph of the whole class of bases will not be considered. The method employed here on the one hand avoids cherry picking only roots with a certain usage distribution (for example those with a high token frequency) since all obtainable roots were considered, and on the other that it gives a better picture of the variation for short illative usage for each possible input base instead of only considering the class as a whole. At

the same time, the combination manages to take token frequency and therefore actual language usage into account. The amount of uniformity of short illative usage within the CVCV class can then be measured by comparing the ratios of long and short illatives of different roots within one corpus. For example, if the corpus contains 95% short illatives of a pseudo-root *tiga* but only 10% of another pseudo-root *piga*, in a case where phonological differences can not play any role, it can be stated that the phonological context for the vocabulary insertion rule has been lost and the context must be specified in another way.

3.3.3 Selection of corpus and bases

Since the purpose of the corpus study conducted in this thesis is to give a picture of recent usage of the illative of Q1 CVCV roots, the corpus etTenTen[18] was used for this task. It contains 270 million words from 686000 internet pages downloaded in 2013. There is no indication of the total time frame in which the texts were produced, but since widespread usage of the internet is a fairly recent phenomenon, this corpus can provide a good view of the recent usage of the geminating illative. Although Estonian written in forums has become closer to standard Estonian in the last decades (Haiba 2016), the data as a whole still offers a better view into colloquial written Estonian as, say, texts from newspapers since the latter ones are and have been subject to normative correctives (Raag 1998), or constituted language normativizing institutions themselves, as is typical in contexts of standardized languages (Ammon 2005; but see Kasik 2004 who argues that Estonian newspapers show a tendency to colloquialization).

The method for selecting the bases to measure the productivity of the geminated illative was as follows: First, all suitable non-compound stems with an unequivocal Q1 CVCV structure were extracted from the word list `lemmad2013.txt` from `http://www.eki.ee/tarkvara/wordlist/` which includes about 100000 lemmata from dictionaries. To this the suitable lemmata from `slang.ee`, a free online slang dictionary were added in order to get greater coverage of nonstandard lemmata.[19] In contrast to previous approaches the bases were selected only by reference to their segmental structure and not by reference to certain noun

[18] Accessible at `http://www.keeleveeb.ee/dict/corpus/ettenten`.

[19] Thanks to Albert Zak for helping me with the technical side of this task.

classes as defined in ÕS. The reason for this choice is that short illative formation by gemination does not necessarily depend on other inflectional properties of the stem. For example, the geminated illative *perre* 'family.ILL' is attested in the same way as *uttu* 'fog.ILL' is, but *pere* takes a partitive in *-t* rather than a zero partitive like *udu*. Because of the difference in the partitive, *pere* does not even appear in the possible word list of the *pesa* class (as defined by ÕS) and falls out of the sample in Kio (2006), for example. The geminated illative is thus (to some extent) independent of the other inflectional properties, and so it is reasonable to try and deduce the context only from the phonological and prosodic properties of the stem alone, as has been done in A. Metslang (2015) and Siiman (2018). Since the Estonian nominal system is partly organized with respect to prosodic considerations, especially foot structure (see section 5.3), it is a plausible assumption that this is also the case for the geminating illative. Obvious prima facie evidence for this is that geminated illatives exist only of Q1 CVCV roots.

The form in the word lists is the dictionary form, namely the nominative singular. Since the stems of the nominative and the partitive singular are identical in most word classes and the short illative is always formed on the basis of this same stem, stem changes affecting the stems of other case forms did not have to be considered when compiling the list of roots. The stem vowel which can be omitted in the nominative forms of some stem classes like the *sepp* class is always present in the nominative singular of CVCV roots which made this segmental change unproblematic, too.[20] Therefore, the nominative form was in most cases suitable for getting the right stems.[21] One possible minor drawback of this method for obtaining bases is that the foot structure of a sequence of CV syllables cannot be derived from the orthography of a stem itself without stress marks. For example, *kimono* 'kimono' has initial stress but *pomelo* 'pomelo' is stressed on the second syllable.[22] Since foot structure is computed with respect to main stress, a root of the structure CV'CVCV with main stress on the second syllable would represent

[20]Therefore, cases like *jõgi* : *jõge* 'river' where the theme vowel differs in the nominative from other case forms were also included.

[21]Stem classes where the nominative is in a sense irregular are the classes of words like *habe : habeme* 'beard' or *kole : koleda* 'ugly'. Lemmata of these classes were excluded. In the case where one nominative belonged to two different classes like *maru* : *maru* 'rage' and *maru* : *maruda* 'cool' I only included examples of the CVCV class.

[22]Thanks to Heiki-Jaan Kaalep for providing this example.

the same prosodic structure as a garden variety 'CVCV root with initial stress and could theoretically constitute a suitable base for gemination. Such roots do indeed exist, as *pomelo* above. It is unclear, however, whether the vowel of the stressed syllable is Q1 or Q2 and there do not appear to be any attested short illatives of such roots (Heiki-Jaan Kaalep, p.c.). Therefore, this class of roots was excluded from the present study. The example of *pomelo* also shows that the number of new suitable bases for the geminating short illative is considerably small since they indeed have to conform to a very specific prosodic structure. Another possible source for geminating illatives which was left out were compounds of which the second part was a CVCV root like *läka+köha* 'pertussis' or *aja+lugu* 'history'. Some lemmata like *ime* 'wonder' were at the same time abbreviations or place names. In the cases where I noticed this, I excluded the abbreviations and names from the count. Other lemmata which had to be left out were due to unresolvable homophony with other highly frequent forms and/or wrong tagging in the corpus.[23]

After these considerations and manual filtering the final word list included 655 lemmata. These included words with no stem alternations and a null partitive like *kena* 'beautiful', *prügi* 'trash', *osa* 'part', words with consonant gradation

[23]Lemmata which could not be included were *iga* 'age' and *iga* 'each, every' because a possible short illative of both words, *ikka*, is homophonous with the adverb *ikka* 'still' and the corpus did not differentiate between the forms. The same was the case with *lina* 'linen' whose short illative *linna* is homophonous with the short illative of *linn* 'city' and the results could not be gone through by hand due to their great number. The word *rada* 'track' was excluded from the final count as well since the short illative *ratta* is homophonous with *ratta* 'wheel.GEN' and the annotation could not differentiate between the forms. The tokens for *lukku* were not tagged right in the corpus, so the numbers for the short illative of *lugu* 'story' are somewhat tentatively used. Checking *ajalugu* 'history' separately, the short : long ratio was 2512 : 58. For some lemmata the corpus did show some discrepancies. For example, the query `blogi(sg ill)` delivered 553 results, but `blogisse@word` 936 which is an unexpected result since this token can only be a long illative of *blogi* 'blog'. In this case I used the latter measurement. The same happened with some homophonous forms: *kakku* (*kagu*.S.ILL 'southeast') was falsely annotated as *kakk*.PRT 'loaf' in examples like the following:

(10) *suurim* *laius* *loodest* *kakku* *17.6* *km.*
biggest width northwest.ELA southeast.ILL

the biggest width from the northwest to the southeast 17.6 km.

like *madu* ‘snake’, *sõda* ‘war’, *luba* ‘permit’, words with a partitive in *-t* like *pere* ‘family’, *kõne* ‘speech’, *logo* ‘logo’ and words with irregular *-i* in the nominative which alternates with *-e* like *hani* ‘goose’, *süli* ‘lap’, *lovi* ‘trance’. The vast majority of the words belongs to the first class and thus the *pesa* class which was referred to in Kaalep (2009) and Kio (2006) as well.

3.4 Results

Of all 655 checked lemmata, 332 had at least one illative token. Of these, 78 had at least one short illative, and 327 at least one long illative. Thus, for most lemmata there were no short illatives in the corpus and only for a few lemmata there was not a single long illative token. There were only five lemmata with 100% short illatives: *ladu* ‘storage’, *lagi* ‘ceiling’, *ligu* ‘liquid for soaking’, *magu* ‘stomach’, and *maru* ‘rage’. The variance of short illative usage for one lemma of all suitable stems was spread all from 0% to 100%. Turning to the first measurement by Aronoff (1976) as explained in section 3.3.1, the productivity indices for the short and the long illative are therefore as follows:

$$I_{short} = 0.2394$$
$$I_{long} = 0.9849$$

The long illative is thus far more productive for Q1 CVCV roots than the geminated illative. For the second measurement, comparing token frequencies, only a selection of all bases will be provided in Table 11 for an overview (the whole table can be found in Appendix 1 starting on page 111.). This table includes three small sections of the results which were sorted by total illative token frequency. A few facts can be observed. First, as already mentioned above, the total illative tokens span a large range from the ten thousands to zero. There is widespread variation between the percentages of tokens for single lemmata, between 0% and 100%. The lemmata with a short illative can not be grouped in any meaningful way[24].

The first group of lemmata shows the most frequent lemmata with respect to total illative token frequency in the corpus. It can be observed that for the

[24]But see A. Metslang (2015) and Siiman (2018, 2019) for some statistical tendencies.

Lemma	Gloss	s.ill	l.ill	Total	% s.ill
elu	‘life’	30787	368	31155	98.82
abi	‘help’	16933	68	17001	99.6
tuba	‘room’	9438	1	9439	99.99
maja	‘house’	6733	155	6888	97.75
küla	‘village’	6058	615	6673	90.78
nägu	‘face’	6179	3	6182	99.95
sama	‘same’	82	5418	5500	1.49
...					
kere	‘hull’	76	221	297	25.59
rivi	‘line’	193	101	294	65.65
lugu	‘story’	11	276	287	3.83
pori	‘mud’	217	58	275	78.9
magu	‘stomach’	274	0	274	100
pubi	‘pub’	1	252	253	0.4
käru	‘cart’	39	163	202	19.31
...					
tuju	‘temper’	10	7	17	58.82
kala	‘fish’	0	17	17	0
lõvi	‘lion’	0	17	17	0
hämu	‘dusk’	0	16	16	0
heli	‘sound’	0	16	16	0
lovi	‘trance’	0	16	16	0
õde	‘sister’	0	15	15	0
...					

Table 11: Selection of the results

most part these are the lemmata which show a very clear distribution of the illative allomorphs. Many forms here use a short illative in more than 90% of the cases, but the reverse is true as well for some cases: the illatives of *sama* are long to 98.51%. The general trend is, however, that the frequently used stems tend to not make use of the elsewhere allomorph *-sse* but use the unproductive, irregular geminated illative (irregular in the sense that the vocabulary insertion rule must be specified for an arbitrary list of roots as context). The relationship between frequency and irregularity has already been demonstrated elsewhere: frequently used lemmata tend to be irregular, and irregularly inflected lemmata tend to be high in usage frequency (see e.g. Bybee 2007; Haspelmath and Sims

2010; Kaalep 2009, 2012). The second group is taken from the class of lemmata with intermediate frequency. These show the greatest amount of variation, as can already be observed from the small sample in the table. The percentages of short illatives range from 0.4% to 100% with many values inbetween. Variation is in fact tied to single lemmata and, as it seems, to individual language users. The third rather large group for the most part again behaves very uniformly. There are almost no geminated illatives besides a few outliers like *tuju* 'temper'. The total distribution ordered by percentage of geminated illatives therefore came out as expected: There is a rather small number of lemmata for which the geminated illative is almost exclusively used (25 lemmata over 95%), another class of medium size showing the whole range of variation (53 lemmata between 0.38% and 94.83%) and a long tail of lemmata without geminated illatives.

For some of the outliers or seeming irregularities a reason could be detected, namely possibly prescriptivism, idiosyncratic language usage, and being part of a fixed expression. For some lemmata, the few long illative forms can quite confidently be tied to verbal government and thus to the influence of prescriptivism, as argued in this thesis. For example, of the 6182 instances of an illative of *nägu* 'face' there were only three long illative forms. Of these, one occurred after *suhtuma* 'relate to', and one after *armuma* 'fall in love', and thus exactly after the verbs which are also given in EKK as verbs which require the long illative. Another similar example concerns the lemma *jagu* 'share'. There were 45 geminated illatives and only one long illative which occured after *puutuma* 'concern'. For other lemmata like *kino* 'cinema' there was no such trend, however. A few outliers were somewhat idiosyncratic because they all originated from the same website. This happened for *jõgi* 'river', where of 1613 total illative tokens, four of the six long illatives were from the same source. The same happened for the short illatives of *kagu* 'southwest', where most of the six short illatives of a total of 365 illatives originated from one and the same website. Some geminated forms were almost exclusively tied to fixed expressions. The short illative of *kivi* 'stone' was mostly used in the fixed expression *kivvi raiuma* 'cast in stone', and the one instance of the geminated illative of *liha* 'meat, flesh' occured in the idiom *lihha ja verre* 'into flesh and blood'.

The results of the single lemmata are pretty similar to the numbers in Kio

(2006). Some differences pertained to the lemmata *kava* 'scheme', where this study found 48.62% geminated illatives (versus 67.2% in Kio 2006), less short illatives for *oja* 'creek' (76.14% versus 89.1%), *raba* 'bog' (75.96% versus 87.5%), and slightly more short illatives for *klubi* 'club' (3.72% versus 1.2%). However, even though many of the lemmata could, according to ÕS, be used with both illatives, there was only a preferece in terms of illative frequency in total, but not for all lemmata with possible parallel forms as the results of the sample in Kio (2006) suggest. Thus, in some cases, the findings did not confirm the indications in ÕS. For example, for *lagi* 'ceiling' and *ligu* 'liquid for soaking' were used exclusively in the geminated illative, although the long illative forms *l`aesse* and *l`eosse* are given as parallel forms. On the other hand, ÕS lists *p`illu* as geminated illative of *pilu* 'slit', but this form does not occur in the corpus at all. Besides these single deviations, the findings also show that for many lemmata, parallel forms are marginal at best. Comparing the productivity index of the geminated illative above, 0.2394, to the 13% (or 0.13) of "usage in reality" in M. Erelt et al. (1995, §157), we find that it is a lot higher here. The exact reason for this discrepancy is unknown since EKG does not offer any insight into the methodology used, but some factors could certainly be that the corpus used here simply did not include illatives of all possible stems which led to overrepresentation of the short illatives compared to all lemmata, or that EKG underestimated the amount of short illatives. In the corpus there indeed occurred geminated illatives of lemmata like *sama* 'same' or *kõne* 'speech' which at least according to ÕS could not have a short illative, although these examples were rather rare.

The initial question this corpus study set out to answer was how productive the geminating illative still is today and how fragmented the usage of the geminated illative for the whole prosodically defined Q1 CVCV class of roots is. The results show quite clearly that the geminating illative is part of unproductive nominal morphology. The geminated illative is only applicable for a small number of roots, and within this small class, its usage is rather fragmented and mostly tied to the frequency of illative usage of the lemma and probably to variation of the usage of individual speakers. Therefore, the geminated illative shows many characteristics of unproductive morphology and needs a list of roots as context for vocabulary insertion.

Concluding this whole section, we find that there are two kinds of variation with respect to the different illative allomorphs. In the Q1 CVCV class there was great variation among individual roots of this class which can be explained by their unproductive status. Some roots are included in the context for the consonant geminating -μ allomorphs, others are not. There is no generalization to be made about these roots and this is simply a fact which must be learned by Estonian speakers. For the *sepp* class, however, previous corpus studies have shown that the null illative is productive for the whole class and variation for individual roots in this class is vanishingly small. This is also expected under an account which treats the null allomorph as productive allomorph of the same morpheme which makes reference not to individual roots but is sensitive to the properties of a whole class of roots. The only outliers to this generalization are of course the small numbers of long illatives in the *sepp* class (2.8% according to Kio 2006). These examples can be explained by reference to a different register and thus grammar. Having delineated the question of allomorph variation more clearly, we can proceed to the morphology and phonology of the Estonian nominal system.

4 The Estonian nominal system

This section deals with the specifics of the Estonian nominal system. First, the phenomenon of length alternation will be discussed since it is crucial for the analysis of the null illative in later sections. The analysis of length alternation at the same time predicts possible root structures in Estonian which will be discussed afterward, together with their relation to stems which surface in inflectional paradigms. After that, a morphophonological account for length alternation is devised and argued that all forms which surface as Q2 need to be morphologically derived.

4.1 Stem alternations

Estonian nominal stems undergo different types of alternations like length alternations that modify the quantity of root-internal vowels and consonants, and

consonant gradation which affects root internal consonants qualitatively. In traditional grammars like EKK, these are often treated as *tüvevaheldused* 'stem alternations' together with other phenomena although their exact nature varies considerably. Most stem alternations are arguably regular and there are only few instances of true suppletive allomorphy in Estonian like in the paradigms *mine-* : *lähe-* 'go', *ise* : *enese-* 'self', *see* : *selle-* : *nee-* 'this', *hea* : *pare-* 'good' (M. Erelt, T. Erelt, and Ross 2007, p. M21). In the next two sections we will take a rather cursory look at consonant gradation and length alternation from a phonological point of view before delving into the details of these processes regarding their morphophonology.

4.1.1 Consonant gradation

Consonant gradation is an unproductive phonological process whereby root internal consonants qualitatively change. It concerns nearly exclusively lenis plosives in the second onset position of a bisyllabic stem. By virtue of this process, the plosives either qualitatively change or are completely deleted (M. Erelt, T. Erelt, and Ross 2007, §M29). Consonant gradation is regular to the extent that only certain segments are altered by it in a predictable way, but unpredictable in what roots are affected by this process. It is unproductive since newer loan words are not affected by it. Some contrasting examples of both the CVCV and the *sepp* class are shown in Table 12.

NOM	GEN	Gloss
sõda	sõja	'war'
häda	häda	'emergency'
idu	`eo	'sprout'
udu	udu	'fog'
p`ärg	pärja	'wreath'
`arg	ara	'coward'
j`ärg	järje	'order'
k`irg	kire	'ardour'

Table 12: Types of consonant gradation

Note, for example, the contrast between the lenis consonant and [j] in the

nominative and genitive forms *sõda* and *sõja* but no gradation in *häda*. Therefore, the roots undergoing this process have to be listed. Despite their differences, there are only a few patterns found among all instances of gradation and qualitative gradation only affects lenis consonants in the onset, most of them the lenis plosives. For early generative formalizations of these alternations see Prince (1980).

4.1.2 Length alternation

As shortly discussed in section 2.1.1, Estonian features a prosodic ternary length distinction. Table 3 is repeated here as Table 13:

Q1	Q2	Q3	Gloss		
lina	*linna*	*l`inna*	'linen.NOM'	'city.GEN'	'city.PRT/S.ILL'
kabi	*kapi*	*k`appi*	'hoof.NOM'	'closet.GEN'	'closet.PRT/S.ILL'
lugu	*luku*	*l`ukku*	'story.NOM'	'lock.GEN'	'lock.PRT/S.ILL'

Table 13: The three quantity grades

Despite the purported special typological status of Estonian regarding the three-way contrast, many authors (e.g. Pöchtrager 2006; Prillop 2013, 2018a; Prince 1980) claim that this contrast is only a surface contrast and does not reflect an underlying ternary distinction on the segmental level. One typical explanation which views the quantity contrast as syllabic property traces the ternary contrast back to the combination two binary features, namely length and weight of a syllable (M. Erelt et al. 1995; Hint 1997; Viitso 2003). According to this analysis short light syllables are Q1, long light syllables Q2 and long heavy syllables Q3. Thus, the difference between Q2 and Q3 is found in the kind of accent assigned to a syllable and on the segmental level there is only a contrast between short and long. Q3 indeed has some interesting properties which suggest that the difference between length and overlength can actually not be a property of segments but of prosodic units – syllables or feet (Ehala 2003; Prince 1980):

If the distinction was a matter of segments, one would expect to find all combinations of Q3 and Q2 segments, but this is not the case. Under this conception, it is impossible for a Q2 consonant to be followed by a Q3 vowel and vice versa

(Hint 1997; Prillop 2018a). Furthermore, within these combinations, not even all segments are encountered in all quantities in roots. Thus we find *kiit* 'praise', with a Q2 vowel followed by a Q2 consonant, but we do not find [m,n,l,r] as this Q2 consonant in roots (Pöchtrager 2006; see below). A segmental argumentation of the distinction would have to deal with these regular gaps and the system would seriously overgenerate (Salveste 2012). The Q3:Q2 distinction is only relevant in the first bisyllabic foot counting from the main stressed syllable to the right, but not even found in all possible bisyllabic stem types – only in a (productive) subclass. Q3 is restricted to stressed syllables whereas there is no such constraint on Q1 and Q2. A Q3 syllable can immediately be followed by another Q3 syllable whereas usually stress is assigned every two syllables. The statistically most reliable phonetical perception clue for the distinction between Q3 and Q2 is the quantity ratio between the first and second syllable (Lehiste 1997; Lippus et al. 2013) and speakers can not perceive a difference between Q3 and Q2 without hearing the vowel of the second syllable (Eek and Meister 1997). Q3 is associated with a special intonation contour which speakers need to hear to perceive Q3 (Lehiste 1997; Lippus et al. 2013). Q3 is also a secondary outcome of a minimal word requirement in Estonian: all monosyllables (except short pronouns like *ma* 'I') have to be Q3. Thus, there are words like *saag* 'saw', *saak* 'harvest', and *sakk* 'jag', but no words such as **sag* where a lenis consonant follows a short vowel (Prillop 2018a). Also, all monosyllables which arise (or historically have arisen) through syllable contraction following consonant gradation such as *t̀oa* < *tuba* 'room', *j̀õe* < *jõgi* 'river', *ǹõo* < *nõbu* 'nephew' seem to surface as default Q3, too. These properties suggest that the Q3:Q2 contrast can not be attributed to segments or syllables but higher prosodic domains.

When viewing the Q3:Q2 distinction as property of a certain prosodic domain it can also be noted that what is known as Q3 syllable can be comprised of different combinations of length units which can explain trade-off phenomena happening within a syllable. As mentioned above, there can never be a long vowel followed by an overlong consonant or vice versa in one syllable. Another relevant factor for consonants is the lenis-fortis distinction (see Pöchtrager 2006; Prillop 2013, 2018a,b). There thus are three types of combinations within the first syllable of a word in Q3: A long vowel followed by a lenis, a long vowel followed

by a fortis, and a short vowel followed by a long fortis. All these combinations yield the same surface length of the syllable. If conceived as taking the space of abstract "length points" within a syllable (without implying that there is an underlying ternary quantity contrast on the segmental level), this trade-off phenomenon can be visualized as follows. In this representation, a syllable has a total of four abstract length points which can variably be filled by different segments. The distribution is explainable through the segments, but the ternary contrast arises only on the surface:

		Gloss	nuclear length	non-nuclear length
siid	si::d	'silk'	3	1
kiit	gi:d:	'praise'	2	2
jutt	jud::	'story'	1	3

Table 14: Quantity tradeoff in the nominative (Pöchtrager 2006, p. 193)

What is relevant for this thesis is that certain roots show a Q3:Q2 alternation in certain morphological contexts. Recall that in some paradigms like those of the *sepp* class, Q3 versus Q2 is the only difference between the genitive and partitive forms since the exponents of the cases are null. Therefore, it is precisely this alternation which is relevant from a morphological point of view. This alternation will be called "length alternation" in what follows. The Q1 versus Q2 distinction is never found in a paradigm, and Q1 versus Q3 is only found for Q1 CVCV roots with the geminating illative in Q3. As a consequence of the distribution of Q3 and Q2, length alternation is only present in bisyllabic roots and can also target prosodically heavy bound derivational morphemes like *-lik* or *-kond*.

Length alternation clearly is a phonological phenomenon operating across the board since it is predictable and exceptionless and targets all segments (even non-native ones like [f] or [ʃ]) (Pöchtrager 2006). The segment which changes in a length alternation is always the final segment of the first nucleus, and thus the mechanism is always the same and indeed an automatic phonological process despite the different structures a Q3 syllable can be comprised of. For example, consider the genitives of the same roots as in Table 14, again with the conception of length points:

		Gloss	nuclear length	non-nuclear length
siidi	si:di	‘silk’	2	1
kiidu	gi:du	‘praise’	2	1
jutu	jud:u	‘story’	1	2

Table 15: Quantity tradeoff in the genitive (Pöchtrager 2006, p. 193)

Note that in the genitive the second syllable now includes the stem vowel. In the case of *siid*, the vowel in the first syllable is shortened, in the case of *kiit*, it is the fortis becoming a lenis, and for *jutt*, the long fortis is shortened. While length alternation can indeed target every segment and generate the Q3:Q2 contrast or every segment, not all possible representations are found in roots, that is, in represenations which have not undergone morphologically conditioned length alternation. Concretely, there are Q3 stems like *kell* ‘clock’ and *keel* ‘language’ but no **keell*, or Q3 *linn* ‘city’ and *liin* ‘linen’, but no Q2 **liinn*, or Q3 *kõõm* ‘dandruff’ or *kõmm* ‘bang’, but no Q2 **kõõmm*, and Q3 *vurr* ‘moustache’ and *hiir* ‘mouse’, but no **vuurr* or **hiirr* (Pöchtrager 2006, §4.2.1). That is, Q3 stems such as *paat* ‘boat’ with a long vowel and a fortis exist, but this is not the case for the consonants in the examples above. The only positions these segments ever occur in a Q2 configuration are in contexts where a Q3 stem is shortened in a certain context under length alternation, e.g. *k`onna* ‘toad.PRT’ : *konna* ‘toad.GEN’ or in roots of the *liiga* class like *summa* ‘sum’ (Pöchtrager 2006; Prince 1980). There is thus a clear asymmetry: While the Q3:Q2 distinction can be expressed by all segments in a combinatory manner after application a phonological process, there are certain segmental combinations that are never found with this distinction without the application of length alternation. This again clearly speaks against treatment of the Q3:Q2 contrast as contrast of segmental quantity, and the only difference in segmental quantity is short versus long.

Of course, the surface contrast between Q3 and Q2 still exists and distinguishes morphological forms, so this contrast must be derived morphologically from an underlying binary length distinction. From a phonological viewpoint it is reasonable to assume that it is Q2 which must be derived since some segmental combinations arise only in morphologically complex environments. However,

most previous phonological accounts of overlength (e.g. Bye 1997; Eek and Meister 1997; Prince 1980) posit that it is Q3 which is the "marked" quantity grade or outlier (Ehala 2003). I will argue against this view from a morphophonological side in section 4.3.2 and for now restate some arguments by Ehala (2003) why this is implausible also from the phonological side and why in fact Q2 should be viewed as the odd one out.

Diachronically, the most accepted hypothesis for the evolution of the ternary quantity system is that Q3 arose through compensatory lengthening of an initial syllable after loss of segmental material in the second syllable through syncope and apocope. Consequently it is claimed that Q3 arose later while Q2 reflects the original long quantity as in Finnish (see e.g. Collinder 1929; Rätsep 1979; Weske 1879). There are some problems for a correlation between Q3 and compensatory lengthening, however. The main argument in Ehala (2003) goes that there are inflectional types with Q3 forms but which historically did not undergo lengthening due to vowel loss, such as the class of verbs like *ˋõppima* 'study.INF' (compare Finnish *oppima*-). These also undergo length alternation regularly: *ˋõppima* : *õpin* 'study1SG'. Analogy from other inflectional types is implausible since the class of these verbs is very large and analogical levelling usually does not concern all members of a class. The same is true for Q3 forms such as *h ˋambaid* 'tooth.PL.PRT' where there was no loss of segments in the second syllable (compare Finnish *hampaita*). In non-main stressed syllables, Q3:Q2 is never contrastive either (Odden 1997). Thus it seems that Q3 actually reflects the original long state and simply surfaces as default (Ehala 2003).

In what follows, I will argue that it is exactly the configuration which surfaces as Q2 which is the outlier and can arise only through application of length alternation. From a phonological point of view for now note the contrast between nasals and liquids on one side and other consonants on the other, and the above arguments from Ehala (2003). Following Pöchtrager (2006) I assume that Q2 is the outcome of a process which takes place if the vowel of a second nucleus enters the same phonological domain as the first syllable. Thus, the difference between the genitive and partitive forms of *siid* 'silk' (see Table 16) is that they have different bracketings from a phonological point of view which correspond to separate domains of phonological computation. In the genitive, the vowel is part of the

inner domain and shortening obtains, but in the partitive it is not. Here, the inner domain is computed first and the vowel can not influence the phonological realization anymore and the quantity grade of the whole word form equals the phonologically simple nominative (Pöchtrager 2006):

Bracketing	Q	Gloss
(siid)	Q3	'silk.NOM'
(siidi)	Q2	'silk.GEN'
((siid)i)	Q3	'silk.PRT'

Table 16: Phonological bracketings

This assumption and the morphological mechanisms to derive these bracketings will be further motivated by observations about possible prosodic structures of Estonian roots and length alternation in morphology in the following sections. In section 4.3.2 I will argue against the alternative morphophonological approach of the alternation which views Q3 as special.

4.2 Estonian roots and stems

In this section, the possible prosodic structures of Estonian roots and stems and the relation of these two forms to each other will be investigated. These will lend further support to the phonological bracketing in Table 16 which can explain the Q3:Q2 alternation as involving different phonological domains and will at the same time explain other inflectional properties of Estonian nouns. Ultimately this will lay the ground for the analysis of the short illative since this whole form too is characterized by a null exponent and certain prosodic properties of the stem it attaches to. This section draws mainly on Pöchtrager (2006).

There are only a few productive prosodic types of Estonian roots. These can be identified by looking at the adaptation of newer loans into Estonian (Pedaja 2006). The table below summarizes them according to their prosodic structure.

Since the main interests of this thesis are length alternation which is found in bisyllabic feet, and the fleeting vowel which is characteristic of the *sepp* class among others, we will restrict our attention only to Q1 CVCV, Q3 CVC, and Q2 CVCV and zoom into those. The relevant prosodic factors in the adaptation of

NOM.SG	Examples
Q1 CVCV	*bagi* 'buggy', *demo* 'demo', *glögi* (< Finnish *glögi*?) 'glögg'
Q1	*pilates* 'pilates'
Q3 CV	*g`ei* 'gay', *š`õu* 'show', *fond`üü* 'fondue'
Q3 CVC	*spämm* 'spam', *friik* 'freak', *striim* 'stream'
Q2 CVCV	*kreisi* 'crazy', *müsli* 'muesli'
Q2 CVCVC	*bänner* 'banner', *singel* 'single'

Table 17: Prosodic classes of loan nouns 1

loans are the number of syllables, vowel length, and the lenis/fortis distinction (Pöchtrager 2006). Consider the English loans in Table 18:

English	Estonian	NOM.SG
pub	*pubi*	Q1 CVCV
demo	*demo*	Q1 CVCV
code	*kood*	Q3 CVC
beat	*biit*	Q3 CVC
bit	*bitt*	Q3 CVC
coma	*kooma*	Q2 CVCV
hippie	*hipi*	Q2 CVCV

Table 18: Prosodic classes of loan nouns 2

The Q1 classes are characterized by short vowels and lenis consonants in English. If there is no vowel in the second syllable, Estonian adds *i* to avoid subminimal prosodic words. The Q2 CVCV classes either have a long vowel or a fortis consonant in English and are rendered as Q2 in Estonian. These will be discussed later. Consider now the Q3 *sepp* class into which loans like *kood*, *biit*, and *bitt* are adapted. This class is open and very productive: in the sample in Pedaja (2006), 43% of new loan nouns were fitted into this class, and the number of roots in this class is very high in general (Kaalep 2012). This class is characterized by the shapes of the nominative, genitive and partitive. The nominative is in Q3 and lacks the stem vowel (*b`itt*), the genitive is in Q2 with the stem vowel present (*biti*), and the partitive is in Q3 again with the stem vowel present, too (*b`itti*). Considering the regularities within productive stem classes like this one, there should be only one underlying representation for every root from which

the three characteristic stems are derived. What is a viable underlying representation with respect to segmental quantity and the status of the stem vowel?

Consider first quantity which is the only property that differentiates the genitive and the partitive form in this paradigm. As stated before, there is no reason to assume that there is an underlying segmental difference between Q3 and Q2, but only between long and short segments. Now, if this difference is not present underlyingly, one of the forms has to be derived from the other. It was also stated that I assume Q3 to be the default case which arises from underlyingly specified length and does not require explanation. For another implicational argument let's take a look at contexts of morphologically induced neutralization of quantity (Prince 1980). Take for example the Q2 genitives *nätsu* and *notsu*, and *tamiili* and *suahiili* with main stress on the second syllable. Their stress and quantity structure are identical. However, these have different nominatives: Q3 *näts* 'chewing gum' and Q2 *notsu* 'piglet', and Q3 *tamiil* 'tamil' and Q2 *suahiili* 'swahili', respectively (Kaalep 2012). Due to instances of neutralization such as these it is impossible to derive the nominative of a root from the genitive form. However, under the assumption that the difference between Q3 and Q2 is underlyingly not specified, the only difference between the Q3 *sepp* class and the Q2 CVCV class (*liiga* or *hipi* class in what follows) seems to be the presence or absence of the stem vowel in the nominative. Other than in the nominative, the stem vowels are present in all other case forms and before most derivational suffixes. Since quantity does not differ in the representations of roots of both classes, it must be the stem vowel which differentiates them since the *sepp* class lacks it in the nominative but the *liiga* class does not.

Pöchtrager (2006) assumes from a phonological point of view that in the *sepp* class the vowel in the genitive and the partitive actually spells out both the genitive and partitive as different homophonous suffixes. That is, the final *-i* in *biti* is the genitive suffix, while the *-i* in *b`itti* is the partitive suffix. While perfectly fine from the view of phonology, this is undesirable from a morphological point of view since across most declension classes the stem vowel stays invariant across the whole paradigm. So if the vowel in the genitive and partitive actually spelled out specific additional morphosyntactic structure, one would have to assume another identical vowel which surfaces in other case forms like the short illative

and the *-de* plurals and even in contexts of derivational morphology. If these homophonous vowels were in fact exponents of a number of different morphemes, this approach would lead to massive redundancy which is not tenable. The consequence is stating that the vowels are underlyingly linked to the representation of the rest of the root and that the genitive and partitive spell out this vowel in a different way which results in a differing phonological output. This is not really different from the proposal in Pöchtrager (2006), it only devises a morphophonological account for the phonological analysis.

The first hypothesis in this manner could be that the vowel is underlyingly present and is deleted either by a phonological rule in the nominative of these stem classes (Blevins 2008; Odden 1997; Prince 1980). However, a rule of this kind fails to account for trisyllabic roots like *kimono* 'kimono' which retain their vowel in the nominative, contrary to other roots of this class like *mürin* 'rumble' which do not. The second possibility is that the stem vowel is underlyingly present in the root, but not fully specified so that it only surfaces in contexts where it is required by adjacent morphology. One piece of evidence for this claim is found in palatalization. In Estonian, the alveolar consonants /t,n,l,s/ palatalize before /i/, as shown in Table 19.

Example	Gloss
hal^{j}l^{j}i	'grey.GEN'
jut^{j}i	'stroke.GEN'
pal^{j}gi	'beam.GEN'

Table 19: Palatalisation before /i/

However, the same roots show palatalization even in the nominative where the vowel /i/ triggering palatalization is not overtly present. This phenomenon is productive since it affects new loans, too. Consider for example *ligam`en$^{j}$t* 'ligament' or *fant`as^{j}t* 'visionary'. These contain palatalized segments which are not present in the source languages[25] and they are not overtly present in the

[25]Note in general that since palatalization is widespread in new loans with latinate endings such as *-Vnt* or *-Vst* where these are not palatalized in any viable source language (e.g. English *ligament*, German *Ligament*, or Russian *лигамент*), it is not palatalization which cues *i* as stem vowel but the other way around. The default stem vowel *i* triggers palatalization.

nominative either. Crucially, there are minimal pairs of palatalized versus non-palatalized consonants, as in Table 20. The appearance of palatalization is linked to the stem vowel *i* in all examples and the absence of palatalization to any other stem vowel or none, as in Table 21.

NOM	PRT	Gloss	NOM	PRT	Gloss
h`all	h`alla	'frost'	h`alʲlʲ	h`alʲlʲi	'hall'
j`utt	j`uttu	'story'	j`utʲtʲ	j`utʲtʲi	'stroke'
p`alk	p`alka	'wage'	p`alʲk	p`alʲki	'beam'

Table 20: Minimal pairs of palatalized vs. non-palatalized segments 1

	Gloss	NOM	PRT	Gloss
kas	'whether'	k`asʲsʲ	k`asʲsʲi	'cat'
on	'is'	`onʲnʲ	`onʲnʲi	'cabin'

Table 21: Minimal pairs of palatalized vs. non-palatalized segments 2

Now, if the nominatives of these roots differ on the surface only with respect to their stem vowels, and the connection between *i* and palatalization versus other stem vowels and no palatalization is clear, the surface differences should be traced back to some representation of the stem vowel in the underlying representation. Thus we have established that roots of the *sepp* class can not be specified for overlength underlyingly and the stem vowel should be in the representation of the root. Now consider again the neutralized genitives of the *sepp* and the *hipi* classes. Declension class membership should be inferrable from the underlying representation too, since new loans are readily accommodated into both classes with different paradigms. Recall that the root *siid* belongs to a declination class which shows no vowel in the nominative and has a null Q3 partitive, while *liiga* retains its vowel in the nominative and gets a partitive in *-t*. An account which treats Q3 and Q2 as segmentally not different also needs to explain the obvious relationship between Q3 and Q2 forms in the *sepp* class. However, if segments themselves underlyingly can not be specified for overlength, only length, there is no obvious property within the lexical representation of roots of these two classes if one wanted to entertain the hypothesis that the stem vowel

was identically associated with the root, so another account is needed. A parallel representation is not possible:

(11) Vocabulary insertion rules for roots of the *sepp* and *hipi* classes (to be discarded)

a. $\sqrt{siidi} \Leftrightarrow$ /siidi/

b. $\sqrt{liiga} \Leftrightarrow$ /liiga/

Concretely, roots like *siid* could differ from roots like *liiga* in that the vowel is not a fully specified part of the root itself. This is of course a rather straightforward assumption which has been made in many previous descriptions anyways to explain the uniformity of the stem vowel across declination classes and paradigms (e.g. Blevins 2008; M. Erelt, T. Erelt, and Ross 2007; Kaalep 2012). The main difference to previous accounts is here that the vowelless stem is not assumed to be a separate stem form stored in the lexicon but derivable from the single representation that underlies most Estonian nouns. In what follows, the exact status of stem vowels of different roots will be discussed.

First, there indeed are some differences regarding the status of certain segments with respect to being a stem vowel. The number of native Standard Estonian stem vowels is rather small. Of all nine Estonian vowels (/a,e,i,o,u,æ,ø,y,ɤ/), only /a,e,i,u/ (/i/ alternating in some paradigms with /e/) can be native stem vowels. With newer loans like *auto* 'car', *avokaado* 'avocado', or *foto* 'photograph' /o/ has made its way into stem-final nuclei again, but crucially, there are no roots of any class which have /o/ as stem vowel and lack this vowel in the nominative. For example, consider again the contrast between trisyllabic genitives like *redeli* 'ladder' and *kimono* 'kimono' and recall that in the *sepp* class and the *redel* class the stem vowel surfaces in all case forms besides the nominative singular. However, while *redel* lacks the stem vowel in the nominative, *kimono* retains it despite being members of the same trisyllabic declension class. That is, the final *o* in the genitive form *kimono* has to be a fully specified part of the root but the final *i* in the genitive *redeli* must differ from it in some way. Now, under the assumption made before that the stem vowel is present in every root, the difference between *kimono* and *redel* shows that vowels are either full parts of the root

and the roots can not surface without them or that they are defectively specified. Before providing a concrete analysis, let's turn from the *redel* class to the class that this chapter focuses on, namely roots with bisyllabic stems.

When discussing possible phonological structures of Estonian and English roots, Pöchtrager (2006) notes that a crucial difference between the languages is the accommodation of roots with a phonological structures which he calls the *hippie* type. These are bisyllabic structures where the domain head is followed by a fortis consonant and another vowel in the second nucleus. English allows these, like in the root *hippie*. Interestingly, these roots are realized as Q2 in Estonian when loaned, and as usual this is reflected in the orthography as well: *hipi*. Crucially, there are no new loans with a nominative with a prosodic structure like *h`ippi* which consist of a bisyllabic Q3 structure with a vowel in the second nucleus (Pöchtrager 2006).[26] Note that this Q3 structure is found in partitives and short illatives, however, and that new loan words indeed can adopt a Q3 realization, as seen in Table 17 and Table 18. Pedaja (2006, p. 793) mentions at least one unambiguous example where a word of this type in English has been adapted to the Q3 *sepp* class, namely *flopp* < engl. *floppy*, so this also seems to be a possible adaptation. Contrary to the unattested bisyllabic Q3 roots, roots which surface as initial Q2 syllables followed by a vowel are easily found, however. Consider Q2 loans with a stem vowel as part of the root like *liiga* 'league', *teema* 'topic', *floora* 'flora' (Pöchtrager 2006, p. 213) or with a fortis consonant like *tako* 'taco', *feta* 'feta' or *tüpo* 'typo'. These always retain their vowel in the nominative, irrespectively of vowel quality. That is, retainment of the vowel in the nominative is not a matter of vowel quality either, as could have been expected from the contrast between *redel* and *kimono*, but of the difference in underlying representation of the vowel. All new loans with this structure retain their vowel the nominative in Estonian since there is no rule to simply delete root internal segments. The descriptive generalization is therefore: If a bisyllabic root contains a vowel in its second nucleus, it must surface as Q2.

To sum up the discussion about the representation of roots so far, I have

[26]There are some older Estonian exceptions to this generalization like *`aasta* 'year', where the first long syllable is historically a contraction of two syllables, but the stem behaves like a trisyllabic stem also with respect to declension class membership (Blevins 2008; Kager 1996). For Q3 nominatives in *-e* like *j`ulge* see Fn. 29.

argued that there is no underlying contrast between Q3 and Q2 on a segmental level, that stem vowels are either fully part of the root or not, and that the status of the stem vowel is invariably connected to the quantity grade a root surfaces in, and thus also determines declension class membership. That is, the representation of the stem vowel is the deciding factor in putting these observations together. Concretely, I will assume that the special status of the stem vowel in classes like *sepp*, *redel* or *seminar* can be explained by a segment which has melody but no structure. That means, the vowel is specified only for quality but not for any quantity or timing. A similar analysis has been devised for example for French adjectival liaison in Bermúdez-Otero (2018), following Wetzels (1987). A number of French adjectives has alternating root-final consonants which surface only in certain morphological contexts. Thus, *petit* 'small' surfaces in the masculine citation form as [pəti] and in feminine forms as [pətit]. Other adjectives such as *joli* 'pretty' or *lisse* 'smooth' do not alternate. The analysis now states that segments that do not alternate are fixed to a skeletal structure position, while alternating segments are not. Concretely, Bermúdez-Otero (2018, p. 189) represents the adjectives *joli*, *lisse* and *petit* as follows:

(12)
x x x x
| | | |
ʒ ɔ l i

(13)
x x x
| | |
l i s

(14)
x x x x
| | | |
p ə t i t

For *petit*, the alternating final segment is not associated to any structural position and cannot surface without being associated to structure. The association to structure must explicitly be carried out secondarily. The same principle can be applied to Estonian as well. Consider the representations of the roots of *kimono* versus *redel* and *liiga* versus *siidi*. The vowels which are never deleted are fully associated with a structure position and the deleted vowels are not and instead only occupy a floating melody position.

(15)
x x x x x x
| | | | | |
k i m o n o

(16)
x x x x x
| | | | |
r e d e l i

(17)
x x x x x
| | | | |
l i i g a

(18)
x x x x
| | | |
s i i d i

That is, the only difference between the stem vowels is that the vowel is not fully specified for roots where it is lacking in the nominative. As mentioned before, unassociated segments are deleted and irrelevant for prosodic structure because they do not take up any length unit. Now recall that length alternation in morphological contexts was viewed as an automatical phonological process which causes, broadly speaking, shortening, and surfacing of a Q2 structure if the vowel in the second nucleus is part of the same phonological domain like the first syllable. Table 16 is repeated here as Table 22:

Bracketing	Q	Gloss
(siid)	Q3	'silk.NOM'
(siidi)	Q2	'silk.GEN'
((siid)i)	Q3	'silk.PRT'

Table 22: Phonological bracketings

The bracketing was argued to arise due to differences in the phonological computation of the genitive and the partitive where the stem vowel in the genitive was still part of the inner phonological domain. Now this approach can be extended to roots as well. Since roots are the smallest possible morphosyntactically relevant domain where a phonological process could apply, the same mechanism can explain the distribution of length grades in roots. Concretely, the reason why *liiga* or *hipi* can only surface as Q2 is the bracketing: The root must form one phonological domain and cannot be broken up since it already is the smallest possible morphosyntactic and, reasonably, phonological unit of a word. In the case of *liiga*, the vowel with its prosodically relevant structural unit is invariably part of the root and thus shortening to Q2 applies automatically. In the case of *kood* 'code', there is no additional vowel associated to a skeletal position in that domain and Q3 surfaces as default in the nominative. That is, if a stem vowel is a full part of the root, the root must surface as Q2 since it must be one single phonological domain where shortening applies automatically if the right

configuration obtains. The surfacing quantity grades are completely parallel to the bracketings of the nominative and genitive forms in Table 22 above:

Bracketing	Q	Gloss
(kood)	Q3	'code'.NOM
(liiga)	Q2	'league'.NOM

Table 23: Phonological bracketings of roots

With this account of the length distinction which analyzes Q2 as outcome of shortening within one phonological domain, a few facts converge: The analysis predicts possible root structures and how they prosodically surface in the nominative singular without any overt morphology. These structures are invariably connected to the representation of the stem vowels. The analysis also explains why palatalisation obtains even without an overt trigger. Additionally, it captures the link between appearance of the stem vowel in the nominative and other declensional properties like the contrast between the partitive suffixes *-ø* and *-t* since the roots are underlyingly different with regard to their prosodic structure. This confirms the analysis by Kaalep (2012) that the main deciding factors for inflecting a root are indeed the shape of (what surfaces as) the nominative and knowledge of the stem vowel. For case morphology, this analysis captures the generalization about the genitive-partitive distinction of the *sepp* class made in Pöchtrager (2006). In the genitive, it is the same vowel which is part of the first domain of phonological computation, in the partitive it is not.

The consequence of the floating segment approach is stating that any environment where the stem vowel surfaces needs to supply a structure position.[27]

[27]It is unclear in how far this is a necessary assumption since in some cases the stem vowel is arguably no more than a purely phonological requirement to avoid illegal clusters. In some other cases the stem vowel seems to serve no more purpose than satisfying a requirement on minimal word length in Estonian. As mentioned above, all regular monosyllables are in Q3. In newer originally monosyllabic loans with a short vowel and a final lenis consonant such as *pubi* < engl. *pub* or *blogi* < engl. *blog* the default stem vowel *i* surfaces in the nominative, probably to avoid a subminimal word. In this case, the vowel also looks like a rescue strategy and its representation in the root is unclear. For newer loans Estonian in most cases resorts to the default vowel *i*, or in some phonologically certain environments to *u* (Kaalep 2012, p. 435). Alternatively, the stem vowels could spell out the categorizers *n* or *v* if the root does not end in a vowel and be null if it does, or be a postsyntactically inserted theme vowel altogether. However, this would complicate

Under the analysis presented here, the structural position which makes the vowel surface in the genitive forces its computation in the first domain, and in the partitive it does not. Thus the same generalization has been reached while at the same time disconnecting vowel quality from the allomorphs themselves and allowing the same vowel to surface in many other morphological contexts like the *-de* plural. Two simplified vocabulary insertion rules for the genitive and partitive of the *sepp* class are presented in examples 19 and 20. The function of the index *i* will be discussed in the following sections.

(19)
$$\text{GEN} \Leftrightarrow \; \overset{\text{X}_i}{|}$$

(20)
$$\text{PRT} \Leftrightarrow \; \overset{\text{X}}{|}$$

To recap this section: It was shown how intimately the specifics of length alternation and the representation of different roots with respect to their stem vowel are connected. Length alternation is shortening to Q2 and takes place in a phonological domain which must minimally be the root. The next section deals with the morphophonology of this process.

4.3 The morphophonology of length alternation

4.3.1 Exponents trigger stem alternations

Until now, two complications have arisen with respect to the interplay of the illative allomorphs and stem allomorphy in the *sepp* class: The long illative is built atop the same Q2 stem as the genitive while the short illative is based on the Q3 stem like the partitive, and it furthermore depends on the prosodic structure of the root whether a short illative is available in the first place. As already mentioned in section 3.2.1, the difference between these two constraints is that illative allomorph selection is at least subject to some variability as evident from (some) parallel forms within the *sepp* class whereas stem allomorphy is absolute:

the condititions for locality and the exact identification of the morphosyntactic domain which corresponds to the phonological domain. Therefore, I leave this issue to future investigations and assume for now that any affix which requires a vowel is specified for a structure position. In any case, one necessary assumption for now is that the structural position can only associate to a floating vowel.

for whatever reason a long illative is chosen, it invariably occurs with the Q2 stem, and the short illative invariably with the Q3 stem. Also, recall that these were argued to be genuine allomorphs of one case and not exponents of different case forms. Variation of the stems therefore must depend on the exponents themselves. For example, consider again Table 2, repeated here as Table 24 which shows the interplay of the two kinds of allomorphy:

Allomorph	ILL.SG	Gloss
-sse	*karbisse*	'into the box'
-ø	*k`arpi*	'into the box'

Table 24: Stem and case allomorphy

This section seeks to derive the variation of the stem the illatives are built upon. As argued in section 2.3.1 it is not plausible to assume different featural structures for both illatives since these would not be reflected in any other domain besides allomorph selection and there is no other syntactic reason to actually differentiate between them as has become clear from the data from DP concord and apparent selection. Therefore, the stem alternations must be deduced directly from the exponents.

The consequence of this approach is stating that the exponents must contain a phonologically interpretable unit which is responsible for the stem alternations. The reasoning for this approach is connected to the nature of the phenomenon itself and two principles of modular grammar theories. Concerning the phenomenon, we have to deal with stem alternations that are triggered by exponents which per definition are a phonological instantiation and thus the trigger must reside on an exponent, too. Concerning the first principle, recall that, as shortly outlined in the introduction, DM assumes cyclic vocabulary insertion. The main point here is that the root is phonologically instantiated before number or case (such as the illative) are. Since the exponents must trigger stem alternations, they need to be inserted as well before any alternation can take place. Thus, the trigger has to be included in the phonological representation of the exponent. Second, in a modular grammar theoretical approach like DM, phonology does not have direct access to morphological information and there-

fore should only be able to be influenced indirectly by it. This hypothesis about the morphosyntax-phonology interface is called Indirect Reference Hypothesis (Bermúdez-Otero 2012; Inkelas 1989):

(21) Indirect Reference Hypothesis (Bermúdez-Otero 2012, p. 77)
"A phonological constraint may not refer to syntactic, morphological, or lexical information unless to require alignment between designated prosodic units and the exponents of designated syntactic (word-syntactic or phrase-syntactic) nodes."

The main point of the hypothesis is that phonology may not directly refer to morphosyntactic information for phonological processes. Since length alternation was argued to be the result of a regular phonological process and this process can be observed to happen before two different exponents of the same case, it should also be triggered by the exponents themselves. More generally, if the *-sse* exponent of the illative can trigger length alternation, any exponent which requires a Q2 stem should be able to trigger it. I therefore propose that all triggering exponents contain a phonologically interpretable autosegmental index which causes length alternation. Recall that length alternation involves shortening in bisyllabic stems if the vowel is part of the same phonological domain as the (other segments of the) root. The concrete effect of the autosegmental index on the root is therefore that it forces the extension of the phonological domain onto itself if it is linearly adjacent. To motivate this analyis, let's take a look at the background of length alternation.

Historically, length alternation is a generalized version of a former phonetic alternation called gradation which, broadly speaking, weakened consonants before a closed syllable. In many stems consonant gradation has been preserved, but this phenomenon is unproductive in Estonian today. Remnants of the original environment for gradation are still visible in the nominal paradigms of Estonian, but the condition for gradation is not present anymore. This can most obviously be shown by comparing the allative of *kapp* 'cupboard' to its long illative: *kapi-le* vs. *kapi-sse*. The allative in *-le* does not close the preceding syllable like the long illative in *-sse* but the stem is the same in both cases, and not **kap-pile*, as expected if gradation were synchronically active. In this respect Estonian

contrasts in a few case forms with Finnish which has retained the alternations more transparently (Hakulinen et al. 2004; Kiparsky 2003):

	Estonian	Finnish
ALL	kapi-le	kaapi-lle
ESS	kapi-na	kaappi-na

Table 25: Estonian vs. Finnish

While in Estonian the allative and essive all use the same stem as the genitive, the original triggering environment for gradation is still visible in Finnish: since the allative *-lle* closes the preceding syllable, the weak stem grade *kaapi-* surfaces. The essive in *-na*, however, does not close the syllable and therefore does not cause gradation, and the strong stem grade *kaappi-* 'cupboard' surfaces. For Estonian, these comparative examples show that despite historical shortening of the suffixes, e.g. *-lle* to *-le* and therefore loss of the syllable closing consonant (see Laanest 1975; Rätsep 1979), the stem form does not change since length alternation is opaque today. There is no synchronically active phonetic principle governing the distribution of the stems with respect to the following suffix, and the trigger of length alternation has been morphologized. Estonian speakers need to memorize which suffixes trigger stem alternations. Stem alternations are a defining feature of Estonian inflectional morphology and not restricted to nominal paradigms but are found throughout the language.

Another fact which corroborates that the trigger is located in exponents themselves is the presence of stem alternations in verbal paradigms as well, and differences in which nominal stem different derivational suffixes attach to. Therefore, the trigger for the stem alternations cannot simply be tied to a certain morphosyntactic environment in the nominal paradigm since both occur outside of the nominal domain as well. The same conditions as above apply for the verbal system, too: Length alternation targets all viable loans and all segments, including [ʃ] and [f].

Consonant gradation targets only certain segments in a closed class of stems while length alternation is productive with all segments even in loans. Here, very similar alternations can be observed as in nouns.

INF1	INF2	PRES.1SG	Gloss
kons`ervi-ma	kons`ervi-da	konservi-n	‘conserve’
d`ušši-ma	d`ušši-da	duši-n	‘shower’
kl`ähvi-ma [klæf::ima]	kl`ähvi-da [klæf::ida]	klähvi-n [klæf:in]	‘bark’

Table 26: Length alternation in loan verbs

INF1	INF2	PRES.1SG	Gloss
sidu-ma	sidu-da	s`eo-n	‘tie’
k`arga-ma	kara-ta	k`arga-n	‘hop’
n`ülgi-ma	n`ülgi-da	nüli-n	‘skin’
m`ärga-ma	märja-ta	m`ärga-n	‘wet’
t`õmba-ma	tõmma-ta	t`õmba-n	‘pull’

Table 27: Consonant gradation in verbs

Despite their differences, there is an intriguing link between length alternation and consonant gradation. Some roots of the *sepp* class, for example, show both consonant gradation and length alternation within one paradigm. While the presence or type of consonant gradation again is restricted to certain roots, length alternation is invariant and for example, a genitive is always in Q2 in the *sepp* class. Yet consonant gradation in this class always cooccurs with length alternation. That is, if a root is specified for undergoing consonant gradation, it will undergo gradation always if it undergoes length alternation. In the partitive versus genitive forms of *haab* ‘aspen’, *h`aaba*.PRT and *haava*.GEN, the vowel [a] shortens from Q3 to Q2 in the genitive while at the same time the lenis plosive gradates to [v]. Despite their differences, the processes have to be morphologically related, and making a categorical distinction between these phenomena does not explain why both processes happen at the same time. Therefore, whatever the phonological reality behind these processes is, the trigger has to be morphologically conditioned and the same. This link has of course been discovered a long time ago and this is reflected in the somewhat historicizing terms “strong” and “weak” grades in traditional grammars. Simply put, both shortened and gradated segments conform here to the weak stem forms, irrespectively of which of both processes is found within a paradigm. In fact, all four logically possible

combinations can be found synchronically in Estonian, as seen in Table 28.

PRT	GEN	Gloss	Gradation	Length Alternation
häda	*häda*	'emergency'	✘	✘
sõda	*sõja*	'war'	✓	✘
hidž`aabi	*hidžaabi*	'hijab'	✘	✓
h`aaba	*haava*	'aspen'	✓	✓

Table 28: Combinations of stem alternations

The roots *häda* and *hidž`aab* are not subject to gradation since they are not specified for it, while *sõda* and *haab* are. On the other hand, both *hidž`aab* and *haab* are underlyingly specified as such that they belong to the *sepp* class and consequently show length alternation. Irrespectively of which processes happen for one root, they are always triggered in the same paradigm slots. So the genitive is always in the "weak" grade, whatever processes lead to the formation of this stem, and the partitive is always in the "strong" grade.[28] What then is the connection between the actual phonological processes underlying what we see as length alternation and consonant gradation? Qualitatively they are only similar with respect to their similar diachronic origin and they are not related synchronically. However, both processes go hand in hand if they occur.

If shortening is a regular process operating within a certain phonological domain, this domain must also be responsible for gradation to happen. That is, all instances of gradation are under this formulation very restricted phonological processes which are synchronically active (Odden 1997; Prince 1980). As their context they have certain roots since the processes are unproductive, but crucially they only happen in the same domain as shortening. While the phonological peculiarities of the two stem alternations discussed here are not part of this thesis, they offer a clue that from the viewpoint of morphosyntax, all stem alternations should be treated together. Crucially, the outcome of these operations is always the same, irrespectively of the kind of morpheme which its exponent spells out where the trigger is located on. This striking similarity between the processes can be observed in the fact that the same processes are triggered across

[28]This does not apply to roots with so called "inverse" gradation such as *mõte* 'thought.NOM' : *m`õtte.* 'thought.GEN'.

categories. For example, the factitive-causative derivational suffix *-sta* always selects the "weak" stem of the base when deriving denominal verbs, irrespectively of whether the weak stem involves consonant gradation, length alternation, or both, and irrespectively of the surface quantity grade of the stem. Consider the examples in Table 29 (M. Erelt, T. Erelt, and Ross 2007, SM 10):

Verb	Q	Gloss		NOM	GEN	Gloss
kiibi-sta-	Q2	'chip (e.g. a dog)'	<	*kˋiip*	*kiibi*	'chip'
kurva-sta-	Q2	'sadden'	<	*kˋurb*	*kurva*	'sad'
ˋea-sta-	Q3	'grant adult rights to a minor '	<	*iga*	*ˋea*	'age'

Table 29: Stem alternations before *-sta*

The first verb, *kiibista-* is derived from the weak stem of *kiip* and thus shows length alternation. For *kurvasta-* both length alternation and consonant gradation can be observed in the nominal paradigm, and again, both processes create the weak form which is used as base for the verbal stem which is shortened to Q2 and contains *v*. The last example, *ˋeasta-* shows derivation from a Q1 stem with only consonant gradation. The gradated form arises as Q3 on the surface but irrespectively of the quantity contrast between the first two and the last example, the derivational suffix simply attaches to the stem. Note also that the vowel before *-sta* matches the stem vowel of the noun. Crucially, these changes are always the same as in the nominal paradigm which again confirms that the trigger can neither be category specific nor sensitive to category. As long as the trigger is local to the root, the outcome is the same, irrespectively of whether the trigger is part of a case suffix or a derivational suffix. This result is expected under the account presented here where the trigger is connected to the exponents themselves. The account also makes some predictions concerning the scope of the trigger for the alternations since it demands strict locality of the autosegmental index to the root or stem. Indeed, both environments in which this process is bled and those where it is fed can be encountered.

If length alternation is conceived as automatic process within a certain phonological domain and it is triggered by a linearly adjacent autosegmental index which forces the extension of the bisyllabic domain for shortening onto itself (if the trigger can be part of the second syllable), it is expected that shortening is

bled by intervening phonological material. This is indeed the case, as can be observed from the *-de* plural in the *sepp* class. The allative contains the index since it surfaces with a shortened stem in the singular. Under the reasonable assumption that it is the same exponent which is chosen in the plural as well, we see that shortening does not obtain. This is due to the intervening plural suffix *-de*. Note that this is not a morphosyntactic matter of the plural intervening. If the (archaic) stem vowel plural is used, we can see how shortening still obtains because the phonological structure is the same as in the singular (see Viitso 1998, p. 128 for an overview):

PRT.SG	ALL.SG	ALL.PL-*de*	ALL.PL.STEM	Gloss
r`iiki	riigi-le	r`iiki-de-le	riige-le	'state'
k`ülma	külma-le	k`ülma-de-le	külmi-le	'cold'
p`oissi	poisi-le	p`oissi-de-le	poise-le	'boy'

Table 30: Bleeding stem alternations 1

Similarly, the process is bled by certain overt derivational suffixes which intervene between root and affix. Consider the contrast between nouns and nominals ending in *-us* derived from their roots. In all cases the underived noun undergoes stem alternations in the context of the allative. The affix *-us* is not specified for triggering alternation and thus Q3 is retained in the derived nouns. However, the allative is no longer local to the root and shortening does not obtain.

PRT.SG	ALL.SG	Gloss	NOM.SG	ALL.SG	Gloss
k`urba	kurva-le	'sad'	k`urb-us	k`urb-use-le	'sadness'
m`ärki	märgi-le	'sign'	m`ärk-us	m`ärk-use-le	'comment'
k`ülma	külma-le	'cold'	k`ülm-us	k`ülm-use-le	'cold'

Table 31: Bleeding stem alternations 2

There are also some interesting cases of a previously opaque domain becoming transparent for the process again after derivation which will be discussed in the next section.

Summing up so far, the alternations caused by the derivational suffixes and in verbal paradigms highlight that stem changes are not confined to the nominal

system alone so the trigger for the changes must be formulated independently of the nominal system or of certain slots in the nominal paradigm. The most straight-forward way of capturing this is stating that certain exponents of both inflectional and derivational morphology include the autosegment which causes the changes, and others do not. The morphophonological approach to stem alternations taken in this thesis is analyzing length alternation as phonologically triggered process and locating the trigger for this process on single phonological exponents themselves. The scope of the trigger is locally strictly constrained as seen in the bleeding environments which proves that length alternation is a purely phonological process. Before continuing to the illatives, I will present some morphophonological arguments for why length alternation can not be analyzed as lengthening to Q3 in addition to the phonological arguments in section 4.1.2.

4.3.2 Against morphologically conditioned Q3

Most phonological accounts of the three-way quantity distinction converge in the observation that the Q3:Q2 distinction is not a matter of differences in segmental quantity. Following the further assumption that segmental length only differentiates long and short, the question arises how to derive the three prosodic quantities on the surface from a binary underlying contrast. To capture length alternation before different suffixes it is necessary from the viewpoint of morphophonology to locate a trigger for length alternation either on all the forms which require Q2 or on all which require Q3. Most analyses are not explicitly concerned with the morphological side of these processes and investigate only the phonetic or phonological side of the stem alternations. In the last sections it was argued that length alternation in fact involves shortening of stems to Q2 instead of them surfacing as default Q3. Therefore, all morphemes which require a shortened Q2 stem carry the trigger on their exponents. Some primarily phonological arguments by Ehala (2003) and Pöchtrager (2006) were discussed in section 4.1.2. In this section I will present another morphophonological argument in favor of this view and against length alternation involving lengthening to Q3 under some special condition.

As discussed in section 4.1.2, only few authors (Ehala 2003; Pöchtrager 2006) assume that Q2 is the marked member of the quantity opposition. To quickly recap, the main assumption in Pöchtrager (2006) is that, since the quantity relations between segments in roots and in forms such as the genitive essentially show a trade-off phenomenon of quantity, Q2 can also be explained through regular shortening through a trade-off within a phonological domain. Pöchtrager (2006) traces the difference between genitive and partitive back to a difference in phonological computation of the stem vowels in the genitive and partitive which he assumes to be different overt suffixes and states that the genitive morphologically is specified to belong to the same domain and the partitive not. With minor modifications this is the analysis which has been used in this thesis. On the contrary, in numerous analyses of Estonian ternary length (e.g. Bye 1997; Eek and Meister 1997; Hint 1997; Kiparsky and Pajusalu 2002; Kuznetsova 2018; Prillop 2013, 2018a,b; Prince 1980) overlength arises through lengthening under a certain condition. Under many accounts in prosodic phonology (which most of the previously mentioned authors adhere to), the difference between Q1 and Q2 lies in the number of morae and the difference between Q2 and Q3 in whether both syllables project to a foot or only the first one. This is similar to the approach taken here: Q1 and the longer quantities are distinguished by segmental length and the Q3:Q2 contrast arises later.

All analyses of Estonian quantity which posit that overlength can be distinguished from length by a certain accent or stress would for a feasible morphophonological account need this special stress to be assigned by some morphophonological rule, and such proposals have indeed been made. For example, Kuznetsova (2018) speaks of the heavy versus light foot accent as "a lexicalised (morphologised) prosodic pattern of realisation of the foot nucleus". She analyzes the vocalic partitive plural in the Q3 form *s`uuri* 'big.PL.PRT' as a suprasegmental morpheme which "changes the accent of the whole foot into heavy" (Kuznetsova 2018, p. 236). Also Prince (1980), having identified the representation of Q3 as a single syllable dominated by a foot, speaks of a certain rule assigning this foot in morphological contexts which require the "strong stem", although he explicitly remains agnostic about the exact morphophological formalization of this rule. Caha (2009) tentatively links the foot projection account of Q3 to morphology

and analyzes the genitive-partitive difference of the *sepp* class by saying that the partitive contains a floating mora which causes projection of a foot only of the first syllable and thus lengthening of this syllable to Q3. This analysis restates the markedness of Q3 also in morphological terms: Only exponents which include the extra mora force the first syllable to be projected under a foot and to surface in Q3. A very early morphophonological analysis of the quantity contrasts, Rajandi (1963) argues to analyze all Q3 segments, and concretely, the short illative (both null and geminating), as morphemes consisting only of a Q3 assigning autosegment since the main property of the short illative is precisely the Q3 realization irrespectively of the segmental structure of the stem (the *sepp* class and the CVCV class). In a similar vein, Kiparsky and Pajusalu (2002) handle morphophonological length alternations in Standard Estonian and Seto by positing "prosodic selectional requirements" for affixes. For example, the null partitive has as its selectional requirement that it must be directly preceded by a foot and thus the preceding heavy syllable becomes superheavy and Q3. By contrast, the null genitive can never be directly preceded by a foot and must itself be the weak member of a binary foot. Lengthening to Q3 then happens through optimization at the phonological level.

What connects all these analyses is the fact that they view Q3 as the marked member of the Q3:Q2 opposition which must phonologically and morphologically be derived. When considering that length alternation is distributed among different morphological contexts and not only in nominal paradigms, the necessary consequence of such an approach is that every (morphologically complex) instance of Q3 must be derived, and in turn, any affix which causes lengthening to Q3 under this approach should be able to do this across the board without reference to unmotivated indices on roots and only with reference to their prosody. One difficulty for such an approach has been pointed out by Odden (1997). The main problem are monosyllables which arise through gradation of CVCV roots such as *t`oa* < *tuba* 'room', *j`õe* < *jõgi* 'river', or *n`õo* < *nõbu* 'nephew'. These necessarily occur only in morphologically conditioned contexts such as certain case forms and before derivational affixes like *`easta-* in Table 29 which select or trigger the weak stem. Many analyses assume (as does this thesis) that Q3 is the default which surfaces here. However, this significantly weakens the analy-

sis of Q3 as marked since there is no reason why affixes which normally cooccur with shortened and thus weak stems should assign Q3 only in these cases. Consequently, even these accounts are forced to accept that there are various reasons why Q3 surfaces (see also Bye 1997). If one assumes that Q3 arises as default in certain contexts, this is hard to reconcile with the view that it is the marked quantity grade.

A similar problem arises in the distribution of morphologically conditioned length alternation in general. Recall that there are bisyllabic root classes like the *hipi* class which surface as Q2 throughout their paradigms in contrast to the *sepp* class where length alternation can be observed. Now assume for the sake of the argument that all Q3 case forms like the partitive singular in the *sepp* class are derived by assignment of Q3 in this class since the partitive is null (*s`eppa* 'blacksmith.PRT') but not in the *hipi* class with a *-t* partitive (*hipit* 'hippie.PRT'). An explanation for the difference of the *-ø* partitive to the *-t* partitive needs to make reference to properties of the root anyways, so allomorphy is in principle expected here. However, while it is tenable to state that the possibility for a null partitive with the requirement of Q3 assignment to the root simply makes reference to properties of the root, extending this reasoning to derivational affixes would lead to a very unwelcome result since as a consequence all derivational affixes of all categories would need allomorphs which in some way are sensitive to the root, too. These allomorphs would then in fact even be suppletive allomorphs if hypothesized to contain autosegments since under the view that Q3 must be assigned, there cannot plausibly be any phonological process relating a foot or a mora to the absence of it since this is exactly this contrast which differentiates morphological forms.

However, even if we allowed such allomorphy for derivational suffixes, we would only expect to find either morphemes which have as its allomorphs a Q3 assigning exponent or one which does not manipulate length, and morphemes with no allomorphy which only contain either the Q3 assigning component or nothing. What we do not expect are exponents which always shorten since this is not a logical possibility under these accounts. Under the assumption that length alternation simply takes place whenever the necessary conditions apply, for example before a suffix which contains a trigger for Q3, these analyses thus predict

that there should be affixes which contain the trigger for Q3 – the strong accent or foot or mora causing projection of a foot – and simply apply it across the board, even for stems which do not have Q3 forms in their paradigm, as described also in the problem above. However, this case does not seem to be attested. On the contrary, some derivational suffixes require a Q2 stem even if there is no such stem form within the whole nominal paradigm of the stem (M. Erelt et al. 1995, pp. 421–422). For example, the translative verbal derivational suffix *-ne* always selects a Q2 stem even if the base nouns belong to a paradigm where they do not undergo length alternation (for example in the genitive) as in Table 32.[29] A similar effect is visible before the causative suffix *-da* which also requires a Q2 stem, as seen in Table 33.

NOM	GEN	Gloss		*-ne*	Gloss
h`aige	*h`aige*	'sick'	>	*haigenema*	'get sick'
k`ange	*k`ange*	'austere'	>	*kangenema*	'get austere'
j`ulge	*j`ulge*	'brave'	>	*julgenema*	'dare'

Table 32: Forced Q2 before *-ne*

NOM	GEN	Gloss		*-da*	Gloss
m`ööbel	*m`öövli*	'furniture'	>	*möövbeldama*	'furnish'
g`uugel	*g`uugli*	'Google'	>	*guugeldama*	'google'
bit`uumen	*bit`uumeni*	'bitumen'	>	*bituumendama*	'cover in bitumen'

Table 33: Forced Q2 before *-da*

These examples show that length alternations within a nominal paradigm and in derivation need not to coincide. Therefore it is not possible to simply specify for single roots whether length alternation is possible or not. Shortening simply

[29] All of these examples are not necessarily decisive concerning the distribution of quantity grades among the *sepp* class since the stems before *-ne* already are Q3 in the nominative, and conspicuously all end in *-e*. Pöchtrager (2006, p. 215) provides some other completely parallel Q3 adjectives in the nominatives which end in *-e* like *k`aame* 'pale', *l`eebe* 'mild', or *t`üüne* 'calm' and concludes that these are morphologically complex, too, and not simply Q3 roots. The stems before *-da* neither clearly belong to the *sepp* class since they are trisyllabic stems where in the case of *guugel* and *möövbel* the last *e* (which is pronounced as a schwa, see Kaalep 2012) disappears before addition of a stem vowel.

occurs whenever possible.[30] Crucially, there do not seem to be any derivational suffixes which under a lengthening account would always force Q3 irrespectively of the stem.[31]

This is an explanatory gap in analyses that view Q3 as the marked quantity which requires a morphophonological trigger. If Q3 is assumed to arise only through the trigger and Q3 and Q2 do not differ on the segmental level but only on suprasegmental prosodic structure, together with the facts about presence of length alternation across categories it is unclear why there should not be a derivational affix that forces the projection of a foot or assignment of heavy stress and thus Q3 since the reverse case – forcing Q2 – is in fact attested, as in the examples above. Given the absolutely parallel alternations in the context of derivational affixes, it is surprising why we do not find derivational affixes which would contain this foot and could unequivocally force Q3 of any viable root. I will take this as another argument that length alternation involves shortening to Q2, since from a morphophonological point of view this is the more viable scenario.

5 The Illatives

The previous sections showed that length alternation is phonological shortening, triggered by an autosegmental index in linear adjacence. Now let's finally turn

[30]It is interesting to note, however, that length alternation can access a substantially large phonological domain to the left of the trigger for shortening. In the case of verbal derivations of adjectives in *-e* which presumably already involve some form of analytical morphology given the generalizations about the structure of Q3 stems, it is reasonable to assume that at the point of categorization as a verb and possibly the start of a new derivational phase (see e.g. Embick 2010), the previously complex morphological and thus phonological domain structure becomes opaque for further computation and a new phonological cycle starts. In this new cycle, *j`ulge* is treated as a simple domain since the domain structure has become opaque after computation and shortening obtains, triggered by the verbal derivational suffix (see also Bermúdez-Otero 2012, §2.4). This is a notable difference to the nominal morphology of trisyllabic nouns: Q3 stems like Q3 *tr`eener* 'trainer.NOM' do not undergo alternation if surfacing with the stem vowel in the genitive and the result is Q3 *tr`eeneri* 'trainer.GEN'.

[31]The seemingly adjectival *-e* suffix from before is not a clear example for a possibly Q3 forcing suffix since not all adjectives ending in this segment have transparently arisen through derivation and thus the segmental structure of the bases is unknown. I conjecture here that possible bases such as **leeb-* or **tüün-* still have the same structure like a garden variety root of the *sepp* class and should be specified underlyingly as such. In this case, this *-e* suffix would simply do nothing and constitute a new phonological domain like the partitive singular.

our attention to the illative allomorphs. Recall that the three relevant allomorphs were the productive long illative in *-sse*, the productive null illative *-ø* and the unproductive geminating illative *-μ*. Focusing for now on the long and the null illative, recall that the second instance of allomorphy besides allomorphy of the exponents themselves was stem allomorphy: the long illative takes the weak stem while the null illative takes the strong stem. Table 2 is once again repeated here as Table 34:

Allomorph	ILL.SG	Gloss
-sse	*karbisse*	'into the box'
-ø	*kˋarpi*	'into the box'

Table 34: Stem and case allomorphy

As argued in section 2.3.1, assuming different features for a long and a short illative would be implausible and thus the only way to capture the difference is, again, locating the trigger for the stem change on the exponent itself. I propose that the stem alternations of the short and the long illative can in fact be linked directly to their exponents. Concretely, I propose that only overt PLACE heads within the morphologically complex illative exponent trigger stem alternations and shortening to Q2 while a null head does not. To build up this argument, I will first introduce the decompositional structure I assume for local cases. After that the concrete analyses for the illatives are shown and allomorph choice is connected to other instances of prosodically conditioned allomorph choice in Estonian.

5.1 Exploding directional expressions

Based on observations about the compositional surface form of directional expressions in languages of the world it has been suggested that this compositionality reflects hierarchical structure which directional expressions consist of (Koopman 2010; Pantcheva 2010; Svenonius 2007). For example, in the Daghestanian language Lezgian case suffixes (or suffixed adpositions) come in three varieties of differing phonological complexity: one for expressing location and two different,

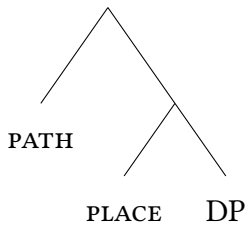
more complex versions for expressing movement to or away from that location. Consider these Lezgian suffixes (Haspelmath 1993; Svenonius 2007):

Gloss	AT	TO	FROM
'in'	*-ø*	*-di*	*-aj*
'on'	*-l*	*-l-di*	*-l-aj*
'at'	*-w*	*-w-di*	*-w-aj*
'behind'	*-q^h*	*-q^h-di*	*-q^h-aj*
'under'	*-k*	*-k-di*	*-k-aj*

Table 35: Lezgian locational suffixes

In this language, the suffixes are clearly built up cumulatively. Concretely, the locational part AT serves as base for additional affixes expressing movement to (*-di*) or away from (*-aj*) the location denoted by the exponent for AT. In that way, the surface representation of these suffixes very transparently mirrors the underlying structure. Note furthermore that in all these cases it is the locational part which is suffixed closest to the root and the directional suffixes build on top of it (Svenonius 2007). Building on previous proposals for semantic decomposition by Jackendoff (1983) and Zwarts (2005), Koopman (2010) and others propose to represent directional expressions (minimally) by PATH, a head denoting motion dominating a head PLACE which expresses stative location (see Asbury et al. 2008; Cinque 2010). Such a structure can be represented as follows:

(22)



The PATH head c-commands the PLACE head which is closer to the DP. When both are linearized in the same direction, PATH will be farther away from the root than PLACE. Turning to Estonian now, this structure can also explain the local case system in a compositional fashion. The six Estonian local cases can be divided into two sets of three cases: the "inner" set beginning in *-s* denoting location at

or movement to or away from an "inner" location, and the "outer" set beginning in *-l* denoting location at or movement to or away from an "outer" location. The short illative does not conform to this generalization since its exponent is null.

Case		Gloss
INE	kapi-s	in the cupboard
ILL	kapi-sse ∼ k`appi-ø	into the cupboard
ELA	kapi-st	out of the cupboard
ADE	kapi-l	on the cupboard
ALL	kapi-le	onto the cupboard
ABL	kapi-lt	from on the cupboard

Table 36: Estonian local cases

To the exclusion of the short illative, the exponents of these cases are clearly built up cumulatively[32] and differ only minimally (see also den Dikken and Dékány 2018). The first segment, and thus the one closest to the root, specifies the set which encodes the type of location, semantically an IN or ON relation between the Figure and the Ground (Talmy 2000; Zwarts 2005).[33] The Figure-Ground relation can be understood as relationship between an entity or substance that is in motion or located with respect to some landmark:

(23) Figure and Ground (Talmy 2000, p. 312)
"The Figure is a moving or conceptually movable entity whose path, site, or orientation is conceived as a variable, the particular value of which is the relevant issue. The Ground is a reference entity, one that has a stationary setting relative to a reference frame, with respect to which the Figure's path, site, or orientation is characterized".

In the local cases discussed here, the Ground is the noun which carries the local case suffix(es) – the DP in the tree diagram above – and the IN or ON relation is expressed by the case suffix. Within both of these sets *e* furthermore marks

[32]This does not imply that the case forms diachronically arose through case stacking. See Ylikoski (2011, 2016) for discussion.

[33]For the proposal that the "inner" set with *-s* is actually unmarked instead of [IN], see (Ylikoski 2016).

movement towards the location, and *t* movement away from the location, both on top of the first segment. Since both denote movement, these can also be assumed to express different features of the same head, namely [TO] and [FROM] carried by the PATH head. The suffixes basically generate minimal pairs between the inner and outer interpretations of nouns irrespectively of whether this expression is stative or directional and it is reasonable to assume that the semantics are indeed fine-grainedly contained in the single segments of the local case exponents, and thus reside on the heads themselves which the case suffixes are comprised of. If the exponents were not built up cumulatively, the similarities would be purely accidental from a synchronic point of view. To the exclusion of the short illatives, the exponents can therefore be summarized as follows:

	PLACE	PATH
[IN]	-s(s)	
[ON]	-l	
[TO]		-e
[FROM]		-t

Table 37: Local case decomposition

The parts *-s* and *-l* are two exponents of the PLACE head spelling out different locational semantics, and the same can be said for *-e* and *-t*. The Estonian local cases thus clearly reflect the underlying structure in terms of being built up cumulatively. Again, the short illatives *-ø* and *-μ* are the outliers to this generalization since they are not segmentable into PLACE and PATH but are portmanteau morphemes expressing both PLACE and PATH since they are directional expressions like the long illative. To account for the different effects the allomorphs have on stems I now propose that only the overt PLACE heads *s*, *ss* and *l* contain the autosegmental index to account for the different stems which the two illatives build on. Before going through the derivations I will make explicit the vocabulary items for the portmanteau short illatives. To account for such morphemes, some work in DM assumes a postsyntactic operation called Fusion which essentially fuses two feature bundles into one (e.g. Acquaviva 2009; Embick 2015; Halle and Marantz 1993).

For the short illative I will assume that PLACE[IN] and PATH[TO] are inserted

as one exponent after application Fusion. Since the short illatives were argued to be allomorphs of one and the same illative case like the long illative it is not possible to formulate a rule for Fusion purely on featural grounds (Embick 2015). On the other hand, since instantiation of linearized structure proceeds cyclically from the inside out it should not be possible to refer to the *exponents* of the short illatives as a condition for an operation which takes place after linearization but *before* vocabulary insertion. However, since the context for the choice of the null allomorphs is very clear, the Fusion rule can be formulated with the same context, too. Recall that the context for the geminating illative allomorph was a list of roots and certain prosodically specified root classes (used with the shorthand PROSODY here). It is only in these contexts that Fusion applies (Acquaviva 2009):

(24) Fusion rule for PLACE[IN] and PATH[TO]
[PLACE[IN]] + [PATH[TO]] ⟶ [PLACE[IN] + PATH[TO]] / $\sqrt{elu}$; PROSODY

This rule specifies that after linearization the heads PLACE[IN] and PATH[TO] can be fused to one head (note the bracketing) if adjacent to either a list of roots or certain prosody, namely prosody of the *sepp* and *seminar* classes. This is necessary to avoid unmotivated null exponents of PLACE[IN] alone, and since both exponents are null, they can't be segmented in any meaningful way either. Now we can devise vocabulary insertion rules for all the Estonian local cases:

(25) Vocabulary insertion rules for the Estonian local cases

a. [PLACE[IN] + PATH[TO]] ⇔ -μ / $\sqrt{elu}$, $\sqrt{toru}$, . . .⌒__

b. [PLACE[IN] + PATH[TO]] ⇔ -ø / PROSODY⌒__

c. [PLACE[IN]] ⇔ -ss / __⌒[PATH[TO]]

d. [PLACE[IN]] ⇔ -s

e. [PLACE[ON]] ⇔ -l

f. [PATH[TO]] ⇔ -e

g. [PATH[FROM]] ⇔ -t

The first two rules match only the complex head which is the outcome of the fusion rule in example 24. Note that while the items in 25c, 25d, and 25f all match

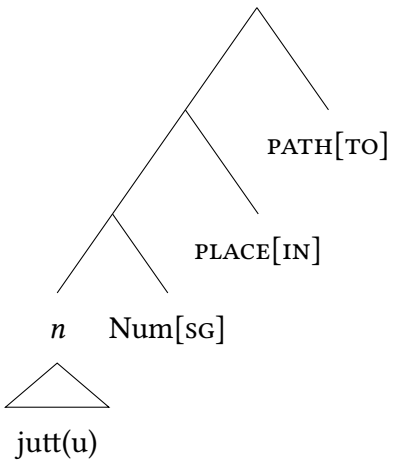
a subset of the features of the complex head and are potential eligible candidates for insertion, they are always ruled out due to the Subset principle (example 3) which dictates that the most specified exponent wins. Note also that to correctly derive the long illative in *-sse* which includes a long [s], the separate rule in 25c is necessary. Other than that, the rules are straightforward.

Turning now to the null illative, recall that it was argued that any exponent which required the stem vowel to surface needed an empty structural position. Therefore, the exponent for the null illative is actually not entirely null, but includes this position. The simplified vocabulary item for the null illative is thus the following (abstracting away from context for now):

(26) [PLACE[IN] + PATH[TO]] ⇔ X
|

The exponent of the null illative is only a structural position which associates to a floating vowel. It does not include any autosegmental index for shortening since Q3 always surfaces as default. This is a rather unexciting result but it follows only from the analysis of the length alternations in Estonian proposed in this thesis. To illustrate this, let's take a look at a few derivations of local cases of the root *jutt* 'story'. First, the null illative. Example 27 shows the structure of the short illative of *jutt* before linearization (the vowel in parentheses stands for the floating vowel):

(27)



After concatenation as in example 2, the linearized structure looks as follows:[34]

(28) $[\sqrt{jutt(u)}$ + *n*]⌢[Num[SG]]⌢[PLACE[IN]]⌢[PATH[TO]]

The derivation proceeds as follows: The root and the nominalizer are inserted. The stem vowel is unassociated in this class of roots. Next, Num[SG] is inserted. It is null and therefore pruned. Now [PLACE[IN]] and [PATH[TO]] are fused since they are local to an exponent of a root of the *sepp* class whose prosodic structure fulfils the right conditions for the Fusion rule described in example 24. Now as exponent for the complex structure [PLACE[IN] + PATH[TO]] the null illative is inserted according to the Subset principle. It contains an empty structural position to later let the stem vowel surface, but it does not contain the autosegmental index. After the derivation, the floating stem vowel associates with the structure position supplied by the null illative and phonological bracketing is established. The root forms one domain to the exclusion of the stem vowel since the domain was not extended and it surfaces as Q3. As in the partitive singular, the stem vowel and all other segments are only concatenated to the output of the phonological computation of the first domain and cannot influence it anymore. The derivation is summarized in Table 38. The segments before the dash constitute the phonological structure of the concatenated string up to the point of the insertion of the morpheme specified in the left column.

	Morpheme	Segments
1.	$[\sqrt{jutt}$ + *n*]	jutt(u)-
2.	Num[SG]	jutt(u)-ø
3.	Fusion of [PLACE[IN]] and [PATH[TO]]	
4.	[PLACE[IN] + PATH[TO]]	jutt(u)-ø
5.	Phonological bracketing	((jutt)u)
6.	surface form	j`uttu

Table 38: Derivation of the null illative singular

[34] I have assumed that the stem vowel is part of the root, but note the remarks in Fn. 27. Since the exact position of the floating stem vowel and thus the content of *n* is unclear, I assume for now that they form a complex head.

Next, consider the allative -*le*. The syntax is the same as in example 27 except for the featural content of PLACE which now carries [ON]. For the purposes of the shortening analysis, the PATH[TO] head with its exponent -*e* is irrelevant since the ablative and adessive also contain -*l* but not -*e* and similarly cause shortening. The relevant vocabulary item containing the autosegmental index is therefore PLACE[ON]. Its only possible exponent and thus elsewhere realization is given in example 29:

(29) PLACE[ON] ⇔ $\begin{array}{cc} \mathrm{X}_i & \mathrm{X} \\ | & | \\ & \mathrm{l} \end{array}$

The exponent first contains the empty structural position which causes the floating stem vowel to associate with it and to surface. Now the structural position carries the autosegmental index which will later force the extension of the phonological domain of the root onto itself and cause surfacing of the bisyllabic foot as Q2. I assume that at least in the syntactic nominal domain for number and case morphology there is only one cycle of phonological computation and that the index is only evaluated in the end of the derivations here (but see Fn. 37). The exponent of course also contains the associated segment -*l* which is characteristic of PLACE[ON]. Now a parallel derivation can be shown for the allative singular.

Until after pruning of Num, the derivation is the same as for the null illative. After that, PLACE[ON] is inserted according to the vocabulary item in example 29. The empty structural position in its exponent is ready to associate with the floating stem vowel and it carries the autosegmental index which is evaluated after the derivational cycle. Next, PATH[TO] is inserted as -*e*, the elsewhere. After the derivation, phonology starts its course. The structural position is associated with the stem vowel of the root. The root plus the associated vowel constitute one domain of phonological computation and the index forces the extension of this domain up to the structure position it is associated to. The domain undergoes computation and shortens to surface Q2. The other segments are left out of this domain and phonologically concatenated to the output and have no effect on the inner computation. The derivation is summarized in 39.

	Morpheme	Segments
1.	$[\sqrt{jutt}$ + *n*]	jutt(u)-
2.	Num[SG]	jutt(u)-ø
3.	PLACE[ON]	jutt(u)-$_i$l
4.	PATH[TO]	jutt(u)$_i$l-e
5.	Phonological bracketing	((juttu)le)
6.	surface form	jutule

Table 39: Derivation of the allative singular

To show one last derivation, consider the allative plural in *-dele* for an instance of alternation bleeding similar to what has been shown in section 4.3.2. This time, Num carries a [PL] feature for plural which contains the empty structural position associating to the stem vowel of the root. PATH and PLACE work identically to the derivation in the singular above. Crucially, although PLACE[ON] carries the autosegmental index, it has no effect in the phonological computation since it is not adjacent to a foot which could be shortened to Q2 and the unassociated structural position is deleted since it has no floating vowel to associate to. The overt plural *-de* blocks shortening of the first foot since it intervenes. This effect is parallel to the lack of alternations in trisyllabic nouns like *tr`eener* : *tr`eeneri*. The derivation is summarized in Table 40.

	Morpheme	Segments
1.	$[\sqrt{jutt}$ + *n*]	jutt(u)-
2.	Num[PL]	jutt(u)-de
3.	PLACE[ON]	jutt(u)de-$_i$l
4.	PATH[TO]	jutt(u)de$_i$l-e
5.	Phonological bracketing	((jutt)udele)
6.	surface form	j`uttudele

Table 40: Derivation of the allative plural

The derivations show that locating the trigger for shortening on the exponents themselves and differentiating the long and short illative precisely with respect to their (supra)segmental structure gives the right results with respect to length alternation.

5.2 The geminating illative

5.2.1 Morphological Length Manipulation

The purpose of this section is to finally make explicit what the function of the illative exponent -μ is and how it can explain the geminated illatives. Both the length alternations of the Estonian nominal stems and the geminating short illatives involve quantitative mutation of stems. As discussed before, the mainly relevant quantity alternation in Estonian is the alternation between Q3 and Q2 since these forms differentiate case forms in single paradigms. These alternations were argued to not involve any quantitative changes on the segmental level. However, the geminating illative does not conform to this generalization since it changes a Q1 root to a Q3 stem. Therefore, the difference here can only be explained by additional quantity which is added in the geminating illative.

Since I assume that the geminating illative exponent is -μ and stands in a paradigmatic relation to other case exponents, the mechanism which combines these exponents with the stem should be identical. When length manipulation is explained via usual affixation of exponents, the positions in a word where length manipulation is expected are exactly the same positions where any other morpheme is expected to surface (Zimmermann 2017). The simplest cases are overtly concatenative prefixation and suffixation where stem and affix follow each other. Since syntax is assumed to be invariant and the other allomorphs of the Estonian illative are garden-variety concatenative suffixes, the non-concatenative quantity manipulating geminating illative should be treated like any other morpheme, namely as exponent which is however, inserted after linearization into the relevant position. Of course, under these assumptions an explanation is required for why the exponent carrying extra quantity is not linearized at one edge of the root but within it.

Non-concatenative morphology such as the geminating illative has been argued to be only an epiphenomenon following principles of independently motivated affixation of phonological primitives (Bermúdez-Otero 2012; Bye and Svenonius 2012; Zimmermann 2017; Zimmermann and Trommer 2013). For length mutation, the affixed primitives are necessarily on the quantity tier. That is, any additive length mutation has to involve additional structural material. In what

follows I will assume that the listed Q1 CVCV roots are manipulated by an operation which has an effect on the quantity structure of the stem. Concretely, the geminating illative allomorph is a length unit which is infixed after the vowel of the first syllable and associates with the following consonantal segment. For ease of exposition and in accordance with most other analyses of this kind this is formalized as infixation of a mora below, but this is by no means the only viable analysis. Any theory separating segmental quality from quantity can similarly account for the phenomenon.

In a similar way to infixation in the languages of the world, morphological length manipulation targets only certain segments of the base and shows a strong edge bias like affixes in general. That is, there are only a rather small number of positions in a root which can be affected by length manipulation and these show a strong preference for the left and right edge in general (Yu 2007; Zimmermann 2017). The approach adhered to in this thesis is Morphological Pivot Affixation (Yu 2007). This theory states that affixes can subcategorize for a restricted number of phonological pivots within a base. That is, the exponent in question carries a phonological subcategorization requirement where it must be placed after vocabulary insertion. For a concrete example, consider in Table 41 a pattern found in Western Shoshone which is basically identical to the Estonian geminating illative. Here the durative aspect on verbs is marked by geminating the medial consonant of the bisyllabic verb stem.

Stem	Gloss	Durative
nɨmi	'travel'	nɨmːi
maka	'feed'	makːa
taikʷa	'speak'	taikːʷa
ɨkʷi	'smell'	ɨkːʷi
hapi	'lie (down)'	hapːi
jɨkʷi	'say'	jɨkːʷi

Table 41: Shoshone durative aspect formation (Crum and Dayley 1993, p. 94), cited in Zimmermann (2017, p. 4)

In line with the formalizations in Zimmermann and Trommer (2013), the sub-

categorization frame of the durative exponent can be defined as follows:[35]

(30) # μ __

This expresses that the exponent which causes lengthening itself subcategorizes for the position after the first mora which is the vowel in the nucleus of the first syllable. The mora (or any other quantity unit) contained in the exponent then associates later with the segment to its right, which is the root-internal consonant.[36]

5.2.2 Analysis

The same principle can be applied to the Estonian geminating illative. The structure which is spelled out is the same as in the null illatives, namely [PLACE[IN] + PATH[TO]], as stated in the vocabulary insertion rule in 25a. After linearization and pruning of the null singular exponent, the complex head becomes linearly adjacent to the root like for the null illative and this relation can trigger allomorphy. As has become clear from the corpus study in this thesis and previous research, the geminating illative is unproductive and requires a list of roots as context for insertion. What are the repercussions of this result in terms of correct formalization of the geminating illative as exponent? First, recall that grammars like EKK treat the geminating illative as involving suppletive stem allomorphy. In a DM-style formalization the rules for vocabulary insertion would then look something like this:

[35] Zimmermann and Trommer (2013, p. 140) give #__μ as subcategorization frame for Shoshone but this must be a mistake since this frame would express that the mora is placed *before* the first mora which is necessarily the first vowel in these bisyllabic stems, and the second onset consonant could not be lengthened under this subcategorization. The same frame is correctly given for the Luganda data in which lengthening patterns accordingly by geminating the *first* onset consonant: kubo : k:ubo; da:la : d:a:la; fumu : f:umu (Zimmermann and Trommer 2013, p. 138).

[36] Alternatively, one could claim that the infix targets the position *after* the second onset and copies the consonant from left to right. It seems that there are no clear empirical generalizations which would speak for one analysis over the other since all these subcategorization patterns are attested in the languages of the world (see Yu 2007; Zimmermann 2017; Zimmermann and Trommer 2013) and both in Estonian and Shoshone the process affects only bisyllabic stems. The direction of association is similarly difficult to resolve. For this reason I will simply adhere to the analysis of Shoshone given above since the pattern is identical to Estonian.

(31)
$$\begin{array}{llll} \sqrt{toru} & \Leftrightarrow & \text{torru} & / \_\_\frown[\textsc{place}[\textsc{in}] + \textsc{path}[\textsc{to}]] \\ \sqrt{toru} & \Leftrightarrow & \text{toru} & \\ \sqrt{elu} & \Leftrightarrow & \text{ellu} & / \_\_\frown[\textsc{place}[\textsc{in}] + \textsc{path}[\textsc{to}]] \\ \sqrt{elu} & \Leftrightarrow & \text{elu} & \\ \sqrt{\ldots} & \Leftrightarrow & \ldots & / \_\_\frown[\textsc{place}[\textsc{in}] + \textsc{path}[\textsc{to}]] \end{array}$$

This list of rules expresses suppletive root allomorphy. The root on the left side of the rule is realized with some phonological form on the right side in the context of the illative singular and in another as the elsewhere. In the case of the illative, however, this rule does not capture that there is a clear non-arbitrary correspondence between the shape of the underlying phonological representation of the root which would surface as elsewhere and the short illative form, namely the geminated root-internal consonant. From a formal point of view, this rule format expresses the same as the relation between different suppletive stems of the verb 'go', for example *mine-* and *lähe-*. Therefore, it is better to formalize the short illative as involving a synchronically active rule altering the segmental structure of the stem to capture the uniformity of the geminated illatives (for discussion see Embick and Halle 2005). As already discussed, I assume that this rule can be explained as synchronic gemination via mora infixation (see Zimmermann and Trommer 2013):

(32)
$$[\textsc{place}[\textsc{in}] + \textsc{path}[\textsc{to}]] \Leftrightarrow \text{-}\mu_{(\#\mu\_\_)} / \left\{ \begin{array}{c} \sqrt{toru} \\ \sqrt{elu} \\ \ldots \end{array} \right\} \frown\_\_$$

This vocabulary insertion rule states that the complex head be instantiated as mora which subcategorizes for the position after the first mora in the root and has as its context a list of roots. Concretely, following cyclic insertion, the root is inserted at some point before the local case exponents and the mora has access to its prosodic structure. It selects the position after the first mora which in CVCV roots always is the vowel in the nucleus of the first syllable. In contrast to the rule in 31 there is only one list, namely the list of roots as context. The phonological shape of the root in the illative singular is in all cases regularly de-

rived through mora infixation.[37] The mora gets inserted into the right segmental position of the stem and associates with the next (i.e. second) onset consonant so this consonant is supplied with an additional length unit and so it lengthens (see Zimmermann 2017; Zimmermann and Trommer 2013).[38] There is no need for root allomorphy or root-specific readjustment rules because the pattern is explained by a phonological mechanism with no additional assumptions for morphosyntax. Interestingly, the very same analysis (although lacking the current theoretical apparatus) was already suggested by Hallap (1964) who analyzes the short illative as morpheme which causes modification of a stem internal segment and stands in an allomorphic relationship to the long illative suffix.

The rule with its list of roots as context does not explain why the roots need to be Q1 CVCV. I conjecture that this is an effect related to the prosodically based organization of the Estonian nominal system (it can't be explained solely through diachronic factors since new roots are added to this list, too). However, it can explain more easily than the suppletion account why (marginal) forms like *kl`uppi* and *v`arru* exist: the roots are simply added to the context of rule application through frequent usage (Kaalep 2009). In purely formal terms, the vocabulary insertion rules expressing suppletion would lack an output form for the input $\sqrt{klubi}$ which would need to arbitrarily get the same phonological shape as the other short illatives. In summary, the geminating illative involves infixation of segmentally empty length units which cause the gemination.

[37] One possible complication which ensues here is how Q3 can be guaranteed after infixation of the length unit. As argued in this thesis, Q2 always arises through phonological computation within a certain domain. This computation must therefore occur quite early on in the derivation. In the case of the geminating illative this must also be the case since otherwise the stem vowel *and* the copied consonant would be part of the domain where shortening to Q2 happens, but this is not the case since the geminated illative is always in Q3. A usual assumption for infix placement by pivot is that the targeted structure already at this point needs a full prosodic structure (see Kalin 2019, and Zimmermann and Trommer 2013, Fn. 4). Under this approach, the autosegmental index would have to operate as soon as it is linearized in the derivations in example 38 and the following. This assumption would not change the outcome of the derivations but require some additional assumptions on phonologic overwriting of previous structure.

[38] The change from lenis to fortis consonants as in *pada* : *p`atta* is not a qualitative change since it occurs also in length alternation, for example in *paadi* : *p`aati*.

5.3 Prosodically conditioned allomorph selection in Estonian

The prosodic structure of roots plays an important role in the choice of allomorphs in Estonian (Blevins 2008; M. Erelt, T. Erelt, and Ross 2007; Kaalep 2009, 2010, 2012; Kager 1996; Siiman 2019). Generally, in the case where there is prosodically conditioned allomorph choice, there is usually an elsewhere allomorph and one that is sensitive to prosody (Kager 1996). I will first introduce some examples of similar effects obtaining in the nominal system of Estonian to turn to the illative singular afterwards.

In general, the clearest examples of real affix allomorphy (in contrast to surface allomorphy) besides the illative allomorphs can be found in the allomorphs of the partitive plural and the partitive singular. Abstracting away from vowel partitive plurals, there are two portmanteau allomorphs which express plural and partitive: *-id* and *-sid*. These are true allomorphs since there is no synchronically active process of *s*-deletion or insertion in Estonian to account for the differences (Kager 1996; Paster 2006). Similarly, the two partitives *-ø* and *-t* are true allomorphs since there is no regular phonological process responsible for such an alternation. As mentioned before, there is usually one or more allomorph(s) sensitive to prosody while another fulfils the role of the elsewhere so it is only necessary to specify the prosodic context for the non-elsewhere allomorphs. This presupposes knowledge of what the elsewhere allomorph is, of course.

The elsewhere of the partitive singular is *-t* (Blevins 2008). Its elsewhere status is confirmed by usage over the zero allomorph in spoken language for some words like *onu* 'uncle' (Keevallik 2003; Siiman 2019), and usage in nonce loans and names (Kaalep 2010). The zero allomorph, on the other hand, is found in the *sepp*, *seminar*, and *pesa* class. For the last class, allomorph choice is not exactly clear since some newer words like *klubi* 'club' and *pubi* 'pub' are inflected with a null partitive, but most other new loans like *midi* 'midi' are not (Siiman 2019).[39] The *sepp* and *seminar* classes are characterized by having a floating vowel in the root and an even number of syllables (in the genitive). This corresponds to the main findings of relevant phonetic and prosodic properties of Estonian roots and the difference between stems with an even and an odd number of syllables

[39]Most other differences can be explained diachronically, see Laanest (1975) and Siiman (2019).

(Kager 1996). For the partitive plural, the elsewhere is *-sid* (Blevins 2008). This is corroborated by its its usage in Q1 CVCV roots with low usage frequency instead of vowel allomorphs (like *küla* 'village' or *nina* 'nose'), and the same is true for roots ending in a long vowel or diphthong: the lower the usage frequency, the higher the prevalence of *-sid* usage over *-id* (*süž`eesid* pro *süž`eid* 'sujet', but *id`eid* 'idea') (Kaalep 2010). Thus we have *-ø* as non-elsewhere for the illative singular and *-id* for the partitive plural. There is a crucial connection between the choice of the partitive singular and the partitive plural: *-id* is in most cases chosen if the partitive singular is *-t*, and the vowel partitive plural is most commonly chosen if the partitive singular ends in a vowel, while *-sid* is the elsewhere. The only exception to this rule is the *kõne* class, a subclass of Q1 CVCV which takes a partitive in *-t* but a partitive in *-sid* (e.g *kõne* : *kõnet* : *kõnesid*) (Kaalep 2010).

The non-elsewhere allomorphs therefore are sensitive to prosodic and phonological properties of the root. According to Kaalep (2012), the properties of a root for declension class membership are the quality of the final segment, position of stress counted from right to left, and length of the final syllable. The sensitivity of case allomorphs for these properties in general can also be observed in the shift from one allomorph to the other for the same root and discrepancies between the norm and actual language usage (Kaalep 2012; Pedaja 2006; Siiman 2019). For example, Kaalep (2012, p. 429) notes that the word *tsunami* 'tsunami' has a normed partitive plural *tsunameid* and the main stress is placed on the first syllable, but in actual usage, main stress is mostly placed onto the second syllable and the partitive plural is *tsunamisid*. Since prosody is counted from the main stress on the second root *tsunami* behaves like a Q2 CVCV root (even if this is not visible in the orthography like in *pomelo* 'pomelo' which is stressed on the second syllable) and takes the *-sid* partitive as elsewhere. Similarly, *mangust* 'mongoose' is normed with main stress and Q3 on the second syllable: *mang`ust* and belongs to the *sepp* type with a null partitive *mang`usti*. In actual usage, the partitive surfaces as Q2 *mangustit* with initial stress (Kaalep 2012).

Another interesting example from the viewpoint of illative allomorph choice below is the *muuseum* word type (M. Erelt, T. Erelt, and Ross 2007; Kaalep 2010). Some rather new loans ending in *-eum*, *-ium*, or *-ion* can be syllabified variably and depending on the syllabification they belong to a certain stem class. For ex-

ample, the root *muuseum* 'museum' can be syllabified as muu.se.um, in which case the short illative is typical, or as muu.seum, where the short illative is impossible (M. Erelt, T. Erelt, and Ross 2007, p. M116). One study found that participants chose the partitive plural *-id* if they preferred a partitive singular in *-t* and a vowel partitive plural if they preferred a null partitive singular. The partitive plural *-sid* was never chosen, however (Peebo 1985). Note that syllabification, correspondent stress placement and allomorph choice are invariably connected. These examples show that most choices of allomorphs can be traced back directly to the prosodic structure of Estonian roots and do not require knowledge of different stem forms. Recall also from section 2.1.2 that most other cases have no allomorphs altogether. With this background, we can now turn to the illatives to provide a near complete picture of prosodically conditioned allomorph choice in the Estonian case system.

5.4 Illative allomorph choice

Turning to the illative, we can first repeat the generalizations about rule precedence of the allomorphs. Taking into account the long, null, and geminating illative, we have found that the geminating illative has as its context a list of roots and is unproductive. The long illative is the elsewhere form, since it is used in the plural and for many different root types. The only allomorph left to explain is the zero allomorph. Kaalep (2009) has devised descriptive generalizations about the choice of a short illative allomorph (both zero and geminating) which formulate the necessary conditions for all allomorphs in dependence of the forms of other exponents:

1. If the partitive singular ends in a vowel:
    (a) If the partitive singular is in Q3, the short illative is homophonous with the partitive singular.
    (b) If the nominative singular ends in a consonant, the short illative is homophonous with the partitive singular.
    (c) If the nominative singular ends in a vowel, the short illative is the same as the partitive singular, but in Q3.

2. If the partitive singular does not end in a vowel:

    (a) Short illative formation is highly unusual.

Under the approach to the representation of stem vowels taken in this thesis these rules can be viewed as epiphenomenal generalization. That is, it is not directly the case that illative formation proceeds on the basis of the partitive singular or has access to its shape, but rather that illative and partitive formation both refer to the prosodic structure of the root. But again, we see that the choice of these prosodically conditioned allomorphs is connected. As discussed in this thesis, rule 1(c) which essentially generates the geminated illative is unproductive. However, separating the rule from the null partitive as in the formulation above and connecting it only to the prosodic structure of the root also helps to explain the CVCV roots which fall under rule 2(a) and have a partitive in *-t*, namely *meri*, *veri*, *tuli*, *lumi*, *pere* and *uni*. These fall under the scope of the unproductive rule to varying extent simply because of their prosodic shape. Similarly, the monosyllabic *pea*, *suu*, *öö* and *töö* are subject to irregular illative formation. Since this rule is only specified for roots again, it is highly specified and rules out the application of the elsewhere.

Rules 1(a) and 1(b) express the following generalizations together: A partitive singular ending in a vowel and in Q3 is chosen in the context of all roots which show a bisyllabic stem and length alternation. Descriptively, these include the classes *sepp*, *õnnelik*, *padi : padja*, and *lai : laia*. Roots which fall under rule 1(b) but not 1(a) belong to the *seminar* class. All these roots have a floating stem vowel and a similar prosodic structure, namely an even number of syllables if the stem vowel is counted in. Also to this class belong words like *muuseum* if syllabified as muu.se.um in the nominative. The selection of the null illative therefore has to be sensitive to foot structure and the status of the stem vowel, as are the partitive singular and the partitive plural. The relations between some of the different classes are summarized in Table 42 where the cells with elsewhere exponents are colored gray (see also Siiman 2019, §4.4–4.7):

It is clearly visible that the allomorph choices are connected. The first three classes all have a null partitive singular, a vocalic partitive plural, and a null illative singular. The characteristics of the stems they attach to are an even number

Stem class	PRT.SG	PRT.PL	ILL.SG	
sepp	-ø	-V	-ø	
seminar	-ø	-V	-ø	
muu.se.um	-ø	-V	-ø	
muu.seum	-t	-id	-sse	
redel	-t	-id	-sse	
liiga	-t	-sid	-sse	
kõne	-t	-sid	-sse	∼ -μ
pesa	-ø	-sid	-sse	∼ -μ

Table 42: Connections between case allomorphs

of syllables, and if they are bisyllabic, a heavy first syllable. All these characteristics can be linked to the underlying representation of the root. The next two stems are members of one specific class: The partitive and illative singular are the elsewhere allomorphs and the partitive plural is *-id* only in this class. To the next class belong Q2 bisyllabic stems like *liiga*. These exhibit only elsewhere allomorphs. In the last two Q1 CVCV classes there is only a difference in the partitive singular, otherwise these are identical. Note also that the geminating illative is possible in both classes, irrespectively of the differences in the partitive singular. Here it is still unclear what causes inflection of new loans in one class over the other, and ÕS mentions some variant cases like *logo*, but *-t* is in any case the elsewhere allomorph.

As discussed in section 3.1, variation of the null and the long illative singular allomorphs is small but in these cases the long illative can surface under normative influence and thus a different grammar. Summarizing, the relevant properties for the choice of the null illative are an even number of syllables, and if this number is two, a heavy first syllable. These stems have a floating stem vowel, as in the vocabulary item below. The geminated illative is not primarily sensitive to prosody but has as its context a list of roots. Note that in the case of the short illatives the rule calls for linear adjacence between the complex head and the context for allomorph choice since these allomorphs are found only in the singular whose exponent is always null. The locality conditions in Embick (2010, 2015) then block application in the plural since any intervening overt material would

block the application of the rule, as in the *-de* plural.[40] The vocabulary items for the geminated and null illative with their contexts are shown in examples 33 and 34 below:[41]

(33) $[\textsc{place}[\textsc{in}] + \textsc{path}[\textsc{to}]] \Leftrightarrow \text{-}\mu_{(\#\mu\_)} / \left\{ \begin{array}{l} \sqrt{toru} \\ \sqrt{elu} \\ \dots \end{array} \right\} \frown\!\!\!\!\!\!\_\_$

(34) $[\textsc{place}[\textsc{in}] + \textsc{path}[\textsc{to}]] \Leftrightarrow \begin{array}{c} \text{x} \\ | \end{array} / \left\{ \begin{array}{l} \begin{array}{cccccc} \text{x} & \text{x} & \text{x} & \text{x} & & \\ | & | & | & | & & \\ \text{s} & \text{i} & \text{i} & \text{d} & \text{i} & \end{array} \\ \\ \dots \end{array} \right\} \frown\!\!\!\!\!\!\_\_$

As has been stressed in this thesis, one difference between the vocabulary items of the null and the geminating illative is that the geminating illative is unproductive and the context for its choice is a list of roots while the null illative is productive for a whole class of prosodically defined roots. Previous analyses have not always recognized this fundamental difference and some of the confusion surrounding the short illative allomorphs may in fact be attributable to failing to distinguish between the different contexts of allomorph choice. For

[40] Marginally, there are also null illatives of stem plurals (Viitso 1998) which would also fit this analysis. See also Table 30.

[41] Note that while for the geminated illative the context is a list of roots without reference to their phonological structure, the context for the null illative is the phonological instantiation of the root since this allomorph is sensitive to the prosodic structure of the root. Considering the identity of the morpheme bundle which is instantiated in both cases, this discrepancy is of course not possible since morphosyntactic computation is invariant. Since it is necessary for prosodically conditioned allomorph selection and pruning that the morphemes already be instantiated, I have assumed, following Kalin (2019), that instantiation of roots takes place as first step in the derivation. However, some reference to the identity of roots is necessary to distinguish roots for which a geminated illative is possible, like *maja* 'house' from homophonous names, *Maja*, for which no geminated illative is available. One possible assumption would be stating that roots do not lose their identifier even after being instantiated. See Acquaviva (2009) and Harley (2014) for discussion.

example, (Viks 1982b, p. 576) says that the possibility to use a short illative depends on the phonological structure of a stem, the root's lexical semantics, tradition, frequency, and the speaker's own idiolect. Almost identically, M. Erelt et al. (1995, pp. 56–57) note that the usage of short illatives depends on the phonological structure of a stem, the lexical semantics of a root, usage frequency, being part of a fixed expression, and on the speaker's idiolect. Besides the phonological structure of the stem/root this is certainly not true for the null illative since its usage is almost exceptionless among certain root types. The few outliers can be explained by different grammars. Individual variation of the illative allomorphs pertaining to the idiolect, frequency, being part of a fixed expression (and diachronically, tradition and possibly semantics) is indeed found, but only for the -μ and -*sse* allomorphs, as the corpus data show.

Summing up, this whole section has linked the two different instances of allomorphy surrounding the illatives. Following the account devised for morphologically conditioned length alternation, the exponent of the long illative was argued to include the autosegmental index while both short illatives do not. This followed from considering a fine-grained representation of the allomorphs to parts of which the index could be associated. The last part linked the question about variation between the long and null allomorphs between stem classes to other instances of prosodically conditioned allomorph choice in Estonian.

6 Conclusion

In this thesis I have devised an analysis for the allomorphy of the illative case in Estonian. The main questions for allomorphy of the illative on the one hand pertained to the distribution of the suffix -*sse*, a zero exponent and illative forms which are characterized by root-internal gemination of a consonant, and on the other to length alternations within the stem preceding the allomorph.

To delineate the first problem more clearly I carried out a corpus study about the geminated illative which not only confirmed previous findings that the allomorph is unproductive but that variation between this illative form and the -*sse* exponent for single roots is not widespread. Regarding the second question I extended the phonological account of length alternation by Pöchtrager (2006) by a

morphological implementation to account for length alternations found throughout the language. Contrary to most claims in the literature I argued that of the Q3:Q2 opposition it is the latter which is the marked member, and that consequently it is Q2 which has to be derived morphophonologically. It was argued that every exponent which on the surface stacks onto a Q2 stem form carries an autosegmental index which extends the phonological domain for shortening of the stem unto itself. This approach is superior to a lengthening account since it predicts the existence of forced shortening which can be observed before certain verbal derivational suffixes.

This account of shortening was extended to the illative exponents as well whereby the exponents of the illative could directly be linked to length alternation of the stem forms without resorting to unmotivated different illative features. The geminating illative was argued to be a lengthening infix targeting the root internal consonant for a list of roots. Finally, the distribution of the *-sse* and the zero exponent was linked to other systematic instances of prosodically based allomorph choice in the Estonian case system.

References

Acquaviva, Paolo (2009). "Roots and Lexicality in Distributed Morphology". In: *York-Essex Morphology Meeting 5*. Ed. by Alexandra Galani, Daniel Redinger, and Norman Yeo Yeo. York, pp. 1–21.

Ammon, Ulrich (2005). "Standard und Variation. Norm, Autorität, Legitimation." In: *Standardvariation. Wie viel Variation verträgt die deutsche Sprache?* Ed. by Ludwig Eichinger and Werner Kallmeyer. de Gruyter, pp. 28–40.

Aronoff, Mark (1976). *Word Formation in Generative Grammar*. Cambridge: MIT Press.

Asbury, Anna et al. (2008). "Introduction. Syntax and semantics of spatial P". In: *Syntax and Semantics of spatial P*. Ed. by Anna Asbury. Vol. 120. Linguistik Aktuell/Linguistics Today. Amsterdam: John Benjamins, pp. 1–32.

Baayen, Harald and Rochelle Lieber (1991). "Productivity and English derivation: a corpus-based study". In: *Linguistics* 29, pp. 801–844.

Baayen, Harald and Antoinette Renouf (1996). "Chronicling the *Times*: productive lexical innovations in an English newspaper". In: *Language* 72, pp. 69–96.

Bauer, Laurie (2001). *Morphological Productivity*. Cambridge: Cambridge University Press.

Bermúdez-Otero, Ricardo (2012). "The architecture of grammar and the division of labor in exponence". In: *The Morphology and Phonology of Exponence*. Ed. by Jochen Trommer. Oxford: Oxford University Press, pp. 8–83.

— (2018). "In defence of underlying representations: Latin rhotacism, French liaison, Romanian palatalization". In: *Probus* 30.2, pp. 171–214.

Blevins, James (2008). "Declension Classes in Estonian". In: *Linguistica Uralica* XLIV.4, pp. 241–267.

Bobaljik, Jonathan (2000). "The Ins and Outs of Contextual Allomorphy". In: *University of Maryland Working Papers in Linguistics* 10, pp. 35–71.

— (2015). "Distributed Morphology". Manuscript.

Bonet, Eulàlia and Daniel Harbour (2012). "Contextual Allomorphy". In: *The Morphology and Phonology of Exponence*. Ed. by Jochen Trommer. Oxford: Oxford University Press, pp. 195–235.

Bybee, Joan (2007). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.

Bye, Patrik (1997). "A Generative Perspective on 'Overlength' in Estonian and Saami". In: *Estonian Prosody: Papers from a Symposium*. Ed. by Ilse Lehiste and Jaan Ross. Tallinn: Eesti Keele Instituut, pp. 36–68.

Bye, Patrik and Peter Svenonius (2012). "Non-concatenative morphology as epiphenomenon". In: *The Morphology and Phonology of Exponence*. Ed. by Jochen Trommer. Oxford: Oxford University Press, pp. 427–495.

Caha, Pavel (2009). "The nanosyntax of case". Dissertation. University of Tromsø.

Chomsky, Noam (1986). *Knowledge of language*. New York: Praeger.

Cinque, Guglielmo (2010). "Mapping Spatial PPs. An Introduction". In: *Mapping Spatial PPs*. Ed. by Guglielmo Cinque and Luigi Rizzi. Vol. 6. The Cartography of Syntactic Structures. Oxford University Press, pp. 3–25.

Collinder, Björn (1929). *Über den finnisch-lappischen Quantitätswechsel*. Uppsala.

Corbin, Danielle (1987). *Morphologie dérivationelle et structuration du lexique*. Linguistische Arbeiten 193. Tübingen: Niemeyer.

Crum, Beverly and Jon Dayley (1993). *Western Shoshoni Grammar*. Boise: Boise State University.

Den Dikken, Marcel and Éva Dékány (2018). "Adpositions and Case: Alternative Realisation and Concord". In: *Finno-Ugric Languages and Linguistics* 7.2, pp. 39–75.

Eek, Arvo and Einar Meister (1997). "Simple Perception Experiments on Estonian Word Prosody: Foot Structure vs. Segmental Quality". In: *Estonian Prosody: Papers from a Symposium*. Ed. by Ilse Lehiste and Jaan Ross. Tallinn: Eesti Keele Instituut, pp. 71–99.

Ehala, Martin (2003). "Estonian Quantity: Implications for Moraic Theory". In: *Generative Approaches to Finnic and Saami Linguistics*. Ed. by Diane Nelson and Satu Manninen, pp. 51–80.

Embick, David (2010). *Localism versus Globalism in Morphology and Phonology*. Linguistic Inquiry Monographs 60. Cambridge: MIT Press.

— (2015). *The Morpheme. A Theoretical Introduction*. Interface Explorations 31. Berlin: De Gruyter Mouton.

Embick, David and Morris Halle (2005). "On the status of *stems* in morphological theory". In: *Proceedings of Going Romance 2003*. Ed. by Twan Geerts and Haike Jacobs. Amsterdam/Philadelphia: John Benjamins, pp. 59–88.

Erelt, Mati, Tiiu Erelt, and Kristiina Ross (2007). *Eesti keele käsiraamat 2007*. Eesti Keele Instituut. URL: www.eki.ee/books/ekk09.

Erelt, Mati et al. (1993). *Eesti Keele Grammatika. II, Süntaks*. Ed. by Mati Erelt et al. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

— (1995). *Eesti Keele Grammatika. I, Morfoloogia, Sõnamoodustus*. Ed. by Mati Erelt et al. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Fernández-Domínguez, Jesús (2013). "Morphological Productivity Measurement: Exploring Qualitative versus Quantitative Approaches". In: *English Studies* 94.4, pp. 422–447. DOI: 10.1080/0013838X.2013.780823.

Grabowski, Eva and Dieter Mindt (1995). "A Corpus-Based Learning List of Irregular Verbs in English". In: *ICAME Journal* 19, pp. 5–22.

Haiba, Sabiina (2016). "Kuidas on netikeel muutunud aastatel 2001–2008?" Bachelor's Thesis. Tartu Ülikool.

Hakulinen, Auli et al. (2004). *Iso Suomen Kielioppi*. Webversion. Helsinki: SKS. URL: http://scripta.kotus.fi/visk (visited on 08/31/2017).

Hallap, Valmen (1964). "Morfeemilise positsiooni mõiste". In: *Keel ja Kirjandus* 5, pp. 283–289.

Halle, Morris (1997). "Distributed Morphology: Impoverishment and Fission". In: *MIT Working Papers in Linguistics* 30, pp. 425–449.

Halle, Morris and Alec Marantz (1993). "Distributed morphology and the pieces of inflection". In: *The view from building 20*. Ed. by Kenneth Hale and Samuel Jay Keyser. Cambrigde: MIT Press, pp. 111–176.

— (1994). "Some Key Features of Distributed Morphology". In: *Papers on Phonology and Morphology*. Ed. by Andrew Carnie, Heidi Harley, and Tony Bures. Cambridge: MIT Press, pp. 275–288.

Harley, Heidi (2014). "On the identity of roots". In: *Theoretical Linguistics* 40.3/4, pp. 225–276.

Haspelmath, Martin (1993). *A Grammar of Lezgian*. Vol. 9. Mouton Grammar Library. Berlin: Mouton de Gruyter.

Haspelmath, Martin and Andrea Sims (2010). *Understanding Morphology*. London: Hodder.

Hasselblatt, Cornelius (2000). "Eesti keele ainsuse sisseütlev on lühike". In: *Keel ja Kirjandus* 11, pp. 796–803.

Hiietamm, Katrin (2003). "Definiteness and Grammatical Relations in Estonian". Dissertation. University of Manchester.

Hint, Mati (1997). "The Estonian Quantity Degrees in Prosody and Morphophonology". In: *Estonian Prosody: Papers from a Symposium*. Ed. by Ilse Lehiste and Jaan Ross. Tallinn: Eesti Keele Instituut, pp. 125–135.

Inkelas, Sharon (1989). "Prosodic Constituency in the Lexicon". Ph.D. thesis. Stanford University.

Jackendoff, Ray (1983). *Semantics and Cognition*. Cambridge: MIT Press.

Kaalep, Heiki-Jaan (2009). "Kuidas kirjeldada ainsuse lühikest sisseütlevat kasutamisandmetega kooskõlas?" In: *Keel ja Kirjandus* 6, pp. 411–425.

— (2010). "Mitmuse osastav eesti keele käändesüsteemis". In: *Keel ja Kirjandus* 2, pp. 94–111.

— (2012). "Eesti käänamissüsteemi seaduspärasused". In: *Keel ja Kirjandus* 6, pp. 418–449.

Kager, Rene (1996). "On affix allomorphy and syllable counting". In: *Interfaces in Phonology*. Ed. by Ursula Kleinhenz. Berlin: Akademie Verlag, pp. 155–171.

Kalin, Laura (2019). "When allomorphy meets infixation: Cyclicity and separation". Handout for BCGL 12, December 16, 2019.

Kasik, Reet (2004). "Muutuv meedia – muutuv keel". In: *Oma keel* 2, pp. 19–28.

Keevallik, Leelo (2003). "Colloquial Estonian". In: *Estonian Language*. Ed. by Mati Erelt. Tallinn: Estonian Academy Publishers, pp. 343–378.

Kio, Kati (2006). "Sisseütleva käände kasutus Eesti kirjakeeles". Master's Thesis. Tartu Ülikool.

Kiparsky, Paul (1973). "'Elsewhere' in Phonology". In: *A festschrift for Morris Halle*. Ed. by Stephen Anderson and Paul Kiparsky. New York: Holt, Reinhard and Winston, pp. 93–106.

— (2003). "Finnish Noun Inflection". In: *Generative Approaches to Finnic and Saami Linguistics*. Ed. by Diane Nelson and Satu Manninen, pp. 109–161.

Kiparsky, Paul and Karl Pajusalu (2002). "Seto vowel harmony and the typology of disharmony". Manuscript. URL: www.stanford.edu/~kiparsky/Papers/seto.pdf.

Koopman, Hilda (2010). "Prepositions, Postpositions, Circumpositions, and Particles. The Structure of Dutch PPs". In: *Mapping Spatial PPs*. Ed. by Guglielmo Cinque and Luigi Rizzi. Vol. 6. The Cartography of Syntactic Structures. Oxford University Press, pp. 26–73.

Kuznetsova, Natalia (2018). "Estonian Word Prosody On The Procrustean Bed Of Morae". In: *ESUKA – JEFUL* 9.1, pp. 209–244.

Laakso, Johanna (2017). "Back to the roots? Critical reflections on the 'root' in Finno-Ugristic linguistics". In: *ESUKA – JEFUL* 8.1, pp. 133–148.

Laanest, Arvo (1975). *Sissejuhatus läänemeresoome keeltesse*. Tallinn: Eesti NSV Teaduste Akadeemia.

Lehiste, Ilse (1997). "Search for Phonetic Correlates in Estonian Prosody". In: *Estonian Prosody: Papers from a Symposium*. Ed. by Ilse Lehiste and Jaan Ross. Tallinn: Eesti Keele Instituut, pp. 11–35.

Leminen, Alina et al. (2018). "Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding)". In: *Cortex*. DOI: 10.1016/j.cortex.2018.08.016.

Lippus, Pärtel et al. (2013). "Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian". In: *Journal of Phonetics* 41.1, pp. 17–28.

Marcus, Gary et al. (1995). "German inflection: the exception that proves the rule". In: *Cognitive Psychology* 29, pp. 189–256.

Metslang, Ann (2015). "Ainsuse pika ja lühikese sisseütleva valiku olenemine morfofonoloogilistest tunnustest – korpusanalüüs". In: *Emakeele Seltsi aastaraamat 60 (2014)*, pp. 127–147.

Metslang, Helle (2009). "Estonian grammar between Finnic and SAE: some comparisons". In: *STUF - Language Typology and Universals* 62.1, pp. 49–71.

Mugdan, Joachim (1977). *Flexionsmorphologie und Psycholinguistik*. Tübingen: Narr.

Norris, Mark (2014). "A Theory of Nominal Concord". Dissertation. University of California, Santa Cruz.

Norris, Mark (2018a). “Non-autonomous Accusative Case in Estonian”. In: *Finno-Ugric Languages and Linguistics* 7.2, pp. 7–38.

— (2018b). “Unmarked case in Estonian nominals”. In: *Natural Language & Linguistic Theory* 36, pp. 523–562.

Odden, David (1997). “Some Theoretical Issues in Estonian Prosody”. In: *Estonian Prosody: Papers from a Symposium*. Ed. by Ilse Lehiste and Jaan Ross. Tallinn: Eesti Keele Instituut, pp. 165–195.

Pajusalu, Karl (2015). “Eesti vältesüsteemi olemusest”. In: *Keel ja Kirjandus* 4, pp. 262–267.

Pantcheva, Marina (2010). “The syntactic structure of Locations, Goals, and Sources”. In: *Linguistics* 48.5, pp. 1043–1081.

Paster, Mary (2006). “Phonological Conditions on Affixation”. Dissertation. University of California, Berkeley.

Pedaja, Kati (2006). “Uute laensõnade morfoloogiline kohanemine”. In: *Keel ja Kirjandus* 10, pp. 784–796.

Peebo, Jaak (1985). “*Muuseum*-tüüpi sõnade käänamine”. In: *Kirjakeele teataja 1979–1983. Õigekeelsuskomisjoni otsused*. Ed. by Tiiu Erelt and Henn Saari. Tallinn: Valgus, pp. 10–20.

Pöchtrager, Markus Alexander (2006). “The Structure of Length”. Dissertation. Universität Wien.

Prillop, Külli (2013). “Feet, syllables, moras and the Estonian quantity system”. In: *Linguistica Uralica* XLIX.1, pp. 1–29.

— (2018a). “Mida teeb moora eesti keeles?” In: *Keel ja Kirjandus* LXI.5, pp. 345–364.

— (2018b). “Varjatud muutused eesti keele prosoodilises struktuuris”. In: *Keel ja Kirjandus* LXI.6, pp. 433–452.

Prince, Alan (1980). “A Metrical Theory for Estonian Quantity”. In: *Linguistic Inquiry* 11.3, pp. 511–562.

Raag, Virve (1998). *The Effects of Planned Change on Estonian Morphology*. Studia Uralica Upsaliensia 29. Uppsala.

Rajandi, Henno (1963). “Morfeemide liigitus deskriptiivses lingvistikas”. In: *Keel ja Kirjandus* 6-7, pp. 357–361, 409–414.

Rätsep, Huno (1979). *Eesti keele ajalooline morfoloogia II*. Tartu: Tartu Riiklik Ülikool.

Roberts, Ian (2007). *Diachronic Syntax*. Oxford: Oxford University Press.

Salveste, Nele (2012). "Ikka need välted – no mida neist arvata?" In: *Oma keel* 1, pp. 19–28.

Siiman, Ann (2016). "Ainsuse sisseütleva vormi valiku seos morfosüntaktiliste ja semantiliste tunnustega – materjali ning meetodi sobivus korpusanalüüsiks". In: *Emakeele Seltsi aastaraamat 61 (2015)*, pp. 207–232.

— (2018). "Variation of the Estonian singular long and short illative form: A multivariate analysis". In: *SKY Journal of Linguistics* 31, pp. 139–167.

— (2019). *Vormikasutuse varieerumine ning selle põhjused osastava ja sisseütleva käände näitel*. Dissertationes Philologicae Estonicae Universitatis Tartuensis 45. Tartu: Tartu Ülikool kirjastus.

Svenonius, Peter (2007). "Adpositions, particles and the arguments they introduce". In: *Argument Structure*. Ed. by Eric Reuland, Tanmoy Bhattacharya, and Giorgos Spathas. Linguistik Aktuell/Linguistics Today 108, pp. 63–103.

Talmy, Leonard (2000). *Toward a Cognitive Semantics. Vol I: Concept structuring systems*. Cambridge: The MIT Press.

Tauli, Valter (1968). *Keelekorralduse alused*. Vol. 4. Eesti Teadusliku Seltsi Rootsis. Uppsala.

Tšernova, Nadežda (2014). "Lühikese illatiivi kasutamine sagedus kirjakeeles massimeedia allikate toel". Bachelor's Thesis. Tartu Ülikooli Narva Kolledž.

Türk, Helen (2019). *Consonantal quantity systems in Estonian and Inari Saami*. Dissertationes Linguisticae Universitatis Tartuensis 35. Tartu: Tartu Ülikool kirjastus.

Ullman, Michael T. (2004). "Contributions of memory circuits to language: the declarative/procedural model". In: *Cognition* 92, pp. 231–270.

Viitso, Tiit-Rein (1976). "Eesti muutkondade süsteemist". In: *Keel ja Kirjandus* 3, pp. 148–162.

— (1998). "Estonian". In: *The Uralic Languages*. Ed. by Daniel Abondolo. Routledge, pp. 115–148.

Viitso, Tiit-Rein (2003). “Structure of the Estonian language: phonology, morphology and word formation”. In: *Estonian language*. Linguistica Uralica Supplementary Series 1. Tallinn: Estonian Academy Publishers, pp. 9–92.

Viks, Ülle (1982a). *Klassifikatoorne morfoloogia. Noomen*. Tallinn.

— (1982b). *Ühest morfoloogilisest klassifikatsioonist*. 10-11, pp. 517–525, 575–586.

Walkden, George (2014). *Syntactic Reconstruction and Proto-Germanic*. Oxford Studies in Diachronic and Historical Linguistics 12. Oxford: Oxford University Press.

Weske, Mihkel (1879). *Eesti keele healte õpetus ja kirjutuse wiis*. Tartu: Schnakenburg’i trükk ja kulu.

Wetzels, Leo (1987). “The Timing of Latent Consonants in Modern French”. In: *Studies in Romance Languages*. Ed. by Carol Neidle and Rafael Nuñez Cedeño. Publications in Language Sciences 25. De Gruyter Mouton, pp. 283–318.

Wiese, Richard (1996). *The Phonology of German*. Oxford: Oxford University Press.

Ylikoski, Jussi (2011). “A survey of the origins of directional case suffixes in European Uralic”. In: *Case, animacy and semantic roles*. Ed. by Seppo Kittilä, Katja Västi, and Jussi Ylikoski. Amsterdam: Benjamins, pp. 235–280.

— (2016). “The origins of the western Uralic s-cases revisited: historiographical, functional-typological and Samoyedic perspectives”. In: *Finnisch-Ugrische Forschungen* 63, pp. 6–78.

Yu, Alan (2007). *A Natural History of Infixation*. Oxford Studies in Theoretical Linguistics 15. Oxford: Oxford University Press.

Zimmermann, Eva (2017). *Morphological Length and Prosodically Defective Morphemes*. Vol. 1. Oxford Studies in Phonology and Phonetics. Oxford: Oxford University Press.

Zimmermann, Eva and Jochen Trommer (2013). “The Linearization of Morphological Weight”. In: *Rule Interaction in Grammar*. Ed. by Fabian Heck and Anke Assmann. Linguistische Arbeits Berichte 90. Universität Leipzig, pp. 123–161.

Zwarts, Joost (2005). “Prepositional aspect and the algebra of paths”. In: *Linguistics and Philosophy* 28, pp. 739–779.

Appendix 1: Results of the corpus study

This table contains the results of the corpus study about the geminating illative. Of all 655 checked lemmata, 332 had at least one illative token and only these are included here in the table. The results are sorted according to total illative singular token frequency.

Lemma	S.ILL	L.ILL	Total	% S.ILL
elu	30787	368	31155	98.82
abi	16933	68	17001	99.60
tuba	9438	1	9439	99.99
maja	6733	155	6888	97.75
küla	6058	615	6673	90.78
nägu	6179	3	6182	99.95
sama	82	5418	5500	1.49
häda	3603	56	3659	98.47
süli	3333	6	3339	99.82
pere	2060	730	2790	73.84
rida	2317	2	2319	99.91
lagi	1948	0	1948	100.00
osa	1534	381	1915	80.10
jõgi	1607	6	1613	99.63
klubi	59	1529	1588	3.72
ida	1468	80	1548	94.83
mälu	1447	98	1545	93.66
talu	1259	273	1532	82.18
keha	17	1397	1414	1.20
sõda	1373	9	1382	99.35
tüli	1238	41	1279	96.79
vana	0	1269	1269	0.00
nina	1120	139	1259	88.96
kino	1228	28	1256	97.77
pada	1226	7	1233	99.43
pesa	916	220	1136	80.63
raba	847	268	1115	75.96

Lemma	s.ill	l.ill	Total	% s.ill
võru	1051	32	1083	97.05
kava	492	520	1012	48.62
vale	0	1001	1001	0.00
blogi	0	936	936	0.00
aju	702	229	931	75.40
võsa	808	50	858	94.17
kogu	423	434	857	49.36
taru	434	355	789	55.01
saba	703	54	757	92.87
muda	547	137	684	79.97
ladu	640	0	640	100.00
udu	533	96	629	84.74
maru	600	0	600	100.00
kivi	33	566	599	5.51
uni	572	7	579	98.79
koda	542	2	544	99.63
vaba	0	534	534	0.00
sisu	2	526	528	0.38
eru	505	3	508	99.41
raha	0	479	479	0.00
kagu	6	359	365	1.64
äri	0	354	354	0.00
toru	59	290	349	16.91
kera	302	31	333	90.69
kile	0	323	323	0.00
ala	0	312	312	0.00
ligu	310	0	310	100.00
häbi	0	310	310	0.00
kõri	238	64	302	78.81
kere	76	221	297	25.59
püha	0	296	296	0.00
rivi	193	101	294	65.65
lugu	11	276	287	3.83

Lemma	s.ill	l.ill	Total	% s.ill
pori	217	58	275	78.91
magu	274	0	274	100.00
oma	0	273	273	0.00
nimi	0	268	268	0.00
pubi	1	252	253	0.40
logi	0	224	224	0.00
sõna	0	219	219	0.00
käru	39	163	202	19.31
ihu	12	189	201	5.97
lisa	0	198	198	0.00
edu	0	184	184	0.00
liha	1	178	179	0.56
isa	0	178	178	0.00
oja	134	42	176	76.14
vägi	68	98	166	40.96
segu	0	166	166	0.00
tele	0	138	138	0.00
ema	0	134	134	0.00
tare	98	31	129	75.97
pidu	63	55	118	53.39
suvi	0	117	117	0.00
kena	0	116	116	0.00
janu	0	115	115	0.00
kõne	1	113	114	0.88
tõbi	102	11	113	90.27
vara	0	111	111	0.00
mõju	0	108	108	0.00
nagi	0	97	97	0.00
prügi	9	87	96	9.38
tõde	0	94	94	0.00
jama	0	93	93	0.00
rahu	0	93	93	0.00
ava	0	89	89	0.00

Lemma	s.ill	l.ill	Total	% s.ill
pragu	79	3	82	96.34
õli	0	77	77	0.00
pilu	0	74	74	0.00
tüvi	7	65	72	9.72
ravi	0	71	71	0.00
muru	0	68	68	0.00
vagu	61	3	64	95.31
savi	1	62	63	1.59
mure	0	63	63	0.00
kuju	5	55	60	8.33
haru	15	44	59	25.42
ilu	0	57	57	0.00
muna	0	54	54	0.00
võra	0	53	53	0.00
nohu	0	52	52	0.00
kulu	0	49	49	0.00
vahe	0	47	47	0.00
jagu	45	1	46	97.83
levi	0	46	46	0.00
lõhe	0	45	45	0.00
sugu	0	45	45	0.00
nõgu	35	2	37	94.59
ime	0	37	37	0.00
sada	0	36	36	0.00
kude	0	35	35	0.00
viru	10	23	33	30.30
jada	0	33	33	0.00
jahu	0	29	29	0.00
müra	0	29	29	0.00
vaha	0	28	28	0.00
vibu	0	28	28	0.00
lava	0	27	27	0.00
küna	0	26	26	0.00

Lemma	s.ill	l.ill	Total	% s.ill
vilu	0	25	25	0.00
varu	3	19	22	13.64
hagi	0	22	22	0.00
valu	0	21	21	0.00
lohe	0	20	20	0.00
redu	14	5	19	73.68
hobi	0	19	19	0.00
pusa	0	19	19	0.00
tuju	10	7	17	58.82
kala	0	17	17	0.00
lõvi	0	17	17	0.00
hämu	0	16	16	0.00
heli	0	16	16	0.00
lovi	0	16	16	0.00
naba	0	16	16	0.00
pesu	0	16	16	0.00
õde	0	15	15	0.00
sära	0	15	15	0.00
tava	0	15	15	0.00
ego	0	14	14	0.00
kõva	0	14	14	0.00
sumo	0	14	14	0.00
regi	12	1	13	92.31
melu	0	13	13	0.00
võre	0	13	13	0.00
köha	0	12	12	0.00
olu	0	12	12	0.00
tera	0	12	12	0.00
tulu	0	12	12	0.00
huvi	0	11	11	0.00
kolu	0	11	11	0.00
riba	0	11	11	0.00
tara	0	11	11	0.00

Lemma	s.ill	l.ill	Total	% s.ill
wifi	0	11	11	0.00
salu	0	10	10	0.00
hägu	0	9	9	0.00
logo	0	9	9	0.00
sara	0	9	9	0.00
taju	0	9	9	0.00
tegu	0	9	9	0.00
ludu	1	7	8	12.50
judo	0	8	8	0.00
pala	0	8	8	0.00
tina	0	8	8	0.00
aru	0	7	7	0.00
hüve	0	7	7	0.00
kasu	0	7	7	0.00
kuru	0	7	7	0.00
lüli	0	7	7	0.00
pahe	0	7	7	0.00
pugu	0	7	7	0.00
sodi	0	7	7	0.00
üdi	0	7	7	0.00
viha	0	7	7	0.00
vile	0	7	7	0.00
vine	0	7	7	0.00
buda/buddha	0	6	6	0.00
klade	0	6	6	0.00
kuma	0	6	6	0.00
paha	0	6	6	0.00
soja	0	6	6	0.00
sudu	0	6	6	0.00
uba	0	6	6	0.00
guru	0	5	5	0.00
kadu	0	5	5	0.00
kõla	0	5	5	0.00

Lemma	s.ill	l.ill	Total	% s.ill
nadi	0	5	5	0.00
pagi	0	5	5	0.00
paju	0	5	5	0.00
pära	0	5	5	0.00
pigi	0	5	5	0.00
siga	0	5	5	0.00
tüma	0	5	5	0.00
võlu	0	5	5	0.00
eha	0	4	4	0.00
higi	0	4	4	0.00
kana	0	4	4	0.00
karu	0	4	4	0.00
kiha	0	4	4	0.00
kilo	0	4	4	0.00
liga	0	4	4	0.00
lobi	0	4	4	0.00
möla	0	4	4	0.00
onu	0	4	4	0.00
pinu	0	4	4	0.00
räni	0	4	4	0.00
sula	0	4	4	0.00
tasu	0	4	4	0.00
ufo	0	4	4	0.00
viga	0	4	4	0.00
lodu	1	2	3	33.33
demo	0	3	3	0.00
juha	0	3	3	0.00
jura	0	3	3	0.00
kaja	0	3	3	0.00
kära	0	3	3	0.00
kisa	0	3	3	0.00
kubu	0	3	3	0.00
kula	0	3	3	0.00

Lemma	s.ill	l.ill	Total	% s.ill
kuva	0	3	3	0.00
laga	0	3	3	0.00
lebo	0	3	3	0.00
löga	0	3	3	0.00
mära	0	3	3	0.00
mini	0	3	3	0.00
musi	0	3	3	0.00
nari	0	3	3	0.00
promo	0	3	3	0.00
puru	0	3	3	0.00
rebu	0	3	3	0.00
risu	0	3	3	0.00
silo	0	3	3	0.00
sina	0	3	3	0.00
soga	0	3	3	0.00
vaga	0	3	3	0.00
hala	0	2	2	0.00
halo	0	2	2	0.00
häma	0	2	2	0.00
hame	0	2	2	0.00
iha	0	2	2	0.00
iva	0	2	2	0.00
kibu	0	2	2	0.00
kilu	0	2	2	0.00
kõdu	0	2	2	0.00
läga	0	2	2	0.00
mäda	0	2	2	0.00
mäsu	0	2	2	0.00
memo	0	2	2	0.00
menu	0	2	2	0.00
molu	0	2	2	0.00
mõra	0	2	2	0.00
nibu	0	2	2	0.00

Lemma	s.ill	l.ill	Total	% s.ill
niru	0	2	2	0.00
nisu	0	2	2	0.00
paja	0	2	2	0.00
pila	0	2	2	0.00
proge	0	2	2	0.00
puna	0	2	2	0.00
romu	0	2	2	0.00
rüsa	0	2	2	0.00
tibu	0	2	2	0.00
tihi	0	2	2	0.00
trafo	0	2	2	0.00
vare	0	2	2	0.00
ale	0	1	1	0.00
gala	0	1	1	0.00
homo	0	1	1	0.00
hüva	0	1	1	0
idu	0	1	1	0
ila	0	1	1	0
joga	0	1	1	0
joru	0	1	1	0
juga	0	1	1	0
kabe	0	1	1	0
käbi	0	1	1	0
kahu	0	1	1	0
kalu	0	1	1	0
kama	0	1	1	0
kesa	0	1	1	0
kida	0	1	1	0
kihu	0	1	1	0
kone	0	1	1	0
kribu	0	1	1	0
krusa	0	1	1	0
libu	0	1	1	0

Lemma	s.ill	l.ill	Total	% s.ill
loba	0	1	1	0
lõbu	0	1	1	0
logu	0	1	1	0
luba	0	1	1	0
lüdi	0	1	1	0
madu	0	1	1	0
male	0	1	1	0
mõru	0	1	1	0
mula	0	1	1	0
musa	0	1	1	0
nisa	0	1	1	0
pihu	0	1	1	0
pläga	0	1	1	0
pläru	0	1	1	0
rabi	0	1	1	0
räga	0	1	1	0
rägu	0	1	1	0
rahe	0	1	1	0
raju	0	1	1	0
reha	0	1	1	0
ribi	0	1	1	0
rodu	0	1	1	0
runo	0	1	1	0
sadu	0	1	1	0
säsi	0	1	1	0
soni	0	1	1	0
soru	0	1	1	0
sumu	0	1	1	0
suva	0	1	1	0
taba	0	1	1	0
tabu	0	1	1	0
tädi	0	1	1	0
tala	0	1	1	0

Lemma	s.ill	l.ill	Total	% s.ill
tanu	0	1	1	0
tibi	0	1	1	0
tigu	0	1	1	0
tila	0	1	1	0
tugi	0	1	1	0
vedu	0	1	1	0
visa	0	1	1	0
voli	0	1	1	0

Appendix 2: Abstract

Die vorliegende Arbeit behandelt die Verteilung der Allomorphe des Illativkasus im Estnischen und entwirft eine Analyse im Rahmen der Distributed Morphology. Konkret werden drei Allomorphe untersucht: ein Suffix *-sse*, ein Nullsuffix, und eine Variante des Illativ Singular wobei ein wurzelinterner Konsonant geminiert auftritt. Während das Nullsuffix produktiv ist und kaum mit *-sse* variiert, ist die geminierte Form auf eine kleine Anzahl von Wurzeln beschränkt und daher unproduktiv. Dies wird anhand einer Korpusstudie demonstriert. Zusätzlich zur Allomorphie des Illativmorphems lässt sich im Estnischen eine weitere Form von Allomorphie beobachten: Die Form des gewählten Illativallomorphs hängt systematisch mit prosodischen Alternationen des Nominalstamms zusammen. Diese Veränderungen sind im gesamten Nominalsystem zu finden. Aus diesem Grund wird in der vorliegenden Arbeit argumentiert, dass bestimmte Exponenten autosegmentale Indizes tragen, welche die phonologische Domäne der Wurzel erweitern und so die Stammformen vor Suffixen mit diesem Index gekürzt realisiert werden. Schlussendlich wird argumentiert, dass die Verteilung der Allomorphe *-sse* und *-ø* ähnlich wie die Verteilung von anderen Kasusallomorphen mittels Referenz auf prosodische Eigenschaften des Stamms erklärbar ist.