Variation of the Estonian singular long and short illative form: A multivariate analysis[1]

Ann Siiman
University of Tartu

Abstract

This article aims to determine which morphophonological, morphosyntactic and semantic variables are statistically significant when choosing the illative case long or short form in Estonian. The methodological approach taken is multivariate analysis – classification trees. The results are compared to prior research that studied the same topic using univariate analysis. It is found that the important variables for choosing the long or short illative form are the direction of gradation, the quantity degree of the base form, government, stem-final alternation and the stem-final alternation pattern. Compared to the results of univariate analysis, multivariate analysis leads to similar conclusions. However, it appears that the multivariate analysis is more accurate, for example the classification tree method gives hierarchy about factors.

Keywords: morphology, morphophonology, morphosyntax, semantics, corpus linguistics, variation, illative, aditive, Estonian

1 Introduction

Estonian is a language with a rich morphology. For example, declinable and conjugable words can have either gradation (*astmevaheldus*), stem-final alternation (*lõpuvaheldus*) or no stem alternation at all (Erelt et al. 1995: 123). Also, Estonian words must be in one of the first (Q1), second (Q2) or third (Q3) quantity degrees (Erelt et al. 1995: 110–111).

[1] I would like to thank Külli Habicht, Martin Ehala, Aki-Juhani Kyröläinen, and two anonymous reviewers for their highly valuable comments and suggestions. The study was supported by the University of Tartu Raefond Scholarship (funded by the University of Tartu Foundation).

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

There are two types of gradation in Estonian: quantity alternation (*vältevaheldus*) and quality alternation (*laadivaheldus*). Quantity alternation means that a word's stem can be in strong or weak grade. A stem in strong grade is usually in third degree (Q3) and a stem in weak grade is usually in second degree (Q2), for example *hetk* (NOM, Q3): *hetke* (GEN, Q2): *hetke* (PART, Q3) 'a while'; *taevas* (NOM, Q2): *taeva* (GEN, Q3): *taevast* (PART, Q2) 'sky'. Quality alternation means that a stem in strong grade has a stop or *s*, which in weak grade is absent. For example, *vesi* (NOM): *vee* (GEN): *vett* (PART) 'water'; *hammas* (NOM): *hamba* (GEN): *hammast* (PART) 'tooth' (Erelt et al. 2007: 209–210).

The other stem alternation besides gradation is stem-final alternation. If the nominative form of a word ends with a consonant, then the word has stem-final alternation because in genitive form the base vowel is added. If the nominative form of a words ends with a vowel, then the word usually does not have stem-final alternation, except for words belonging to the *nimi, tuli, kole, habe, sai, lagi, käsi, nali* or *pääse* type. Also, *ne-* and *ke-*ending words always have stem-final alternation, e.g. *inimene* (NOM): *inimese* (GEN) 'human' or *tilluke* (NOM): *tillukese* (GEN) 'tiny' (Erelt et al. 2007: 226).

Another feature demonstrating the rich morphology of Estonian is that it has 14 cases (Erelt et al. 2007: 238–239). Furthermore, some Estonian cases show variation. In this article I study the variation of declinable words, more specifically the illative case variation. The illative case has two forms: a long form and a short form. In the academic grammar of Estonian and Handbook of Estonian the long *sse*-ending form (i.e. the long illative) is possible for all word types (Erelt et al. 1995: 56–57; Erelt et al. 2007: 245–247). However, the short illative is a second choice for some word types. Moreover, research about its actual language use shows that for many words the short form is preferred (Hasselblatt 2000; Kio 2006; Kaalep 2009). In some previous studies the short illative is considered an independent case, the so-called aditive case (e.g. Rajandi 1963: 410; Viitso 1976: 152–153; Viks 1992). Currently, the aditive case is not one of the official Estonian cases because it cannot be applied to all declinable words. In this paper the term *illative* is used for the singular long illative case form and the term *aditive* for the singular short illative case form. Both are forms of one case – the illative case.

1.1 Previous studies and purpose of this study

A compendious synopsis of the history of the illative case has been given by Hasselblatt (2000) and Kio (2006). The variation of the illative and aditive forms has been examined in previous studies (e.g. Sõnajalg 1956; Raag 1998; Hasselblatt 2000; Kio 2006; Kaalep 2009). Nevertheless, there is no consensus in the description of Estonian grammar explaining the variation of the illative case for declinable words. The academic grammar of Estonian describes the use of the short illative form as depending on the word phonological-derivational structure (for example, the short form is more common with *ik*-ending words) and frequency of the word use. Some short illative forms are adverbs or part of multi-word expressions (Erelt et al. 1995: 56). In summary, the academic grammar of Estonian is not specific about the use of the short illative form, saying only that it depends on idiolect and is not bound in normative written language (Erelt et al. 1995: 57).

Previous studies have examined how the choice between the illative and aditive is related to morphophonological variables (Metslang 2015), as well as morphosyntactic and semantic variables (Siiman 2016). This paper builds on those prior studies. The method used previously was univariate analysis – chi-square test and standardized Pearson residuals, and the results were controlled with a so-called part-whole method and using the Cramér's V effect size method. It was found that the choice between the illative and aditive may be related to gradation, the type of gradation, stem-final alternation and the stem-final alternation pattern, the final sound of the base form, the number of syllables in the genitive stem, government, multi-word expression, proper or common noun, the proper noun semantic group and the common noun semantic group. The direction of gradation, the quantity degree of the base form, part of speech, syntactic function and meaning of the verb lemma were not statistically significant factors in the choice between the illative and aditive.

The theoretical background of this study is a usage-based approach, which assumes that linguistic structures and usage events are closely related (Barlow & Kemmer 2000: viii). According to Langacker (1987: 494) the usage event is a symbolic expression which a person uses in certain circumstances and for a certain aim. Based on the usage-based approach the corpus material is valuable material for research to describe the structure of the language.

One of the goals of this article is to determine which morphophonological, morphosyntactic and semantic variables the choice between the illative and aditive depends on. The results are then compared to the prior results that used univariate analysis. The illative case variation is shown here as an example of how other morphological variations in Estonian could be examined.

The outline of the paper is as follows. §2 introduces the data and method. §3 provides an overview of the explanatory variables. §4 presents the results – initially with all the variables analysed from previous studies, then with only the variables that were significant in prior studies. Finally, the results are presented separately for all variable groups – morphophonological, morphosyntactic and semantic variables. In §5 a new variable (the number of syllables in the last foot) is added to the analysis and examined whether it changes the results. Uni- and multivariate methods are compared in §6 and a conclusion is given in §7.

2 Data and method

The same data and data collecting principles which were used for examining morphosyntactic and semantic variables in Siiman (2016) are used in this study. This consists of almost all singular long and short illative forms searched from the Keeleveeb corpuses. The Balanced Corpus of Estonian from the Keeleveeb corpuses was not used because it uses data from other corpuses, which is already in the studied material. Also, the Estonian Dialect Corpus was not used because it does not provide enough data to compare the use of the illative case in written Estonian and Estonian dialects.

The search method to collect data involved first queries through Keeleveeb from the Estonian Web Corpus (etTenTen)[2] (270 million words), a corpus collected from the internet. Hence the data is from edited and non-edited Estonian, including internet language. Also, through the Keeleveeb[3] corpus queries from all the corpuses that are in the Estonian Reference Corpus[4] (240 million words) were included. So in total, queries were from about 510 million words. The material is balanced and all word forms are included only once. In the material there are 840 forms: 420

[2] http://www.keeleveeb.ee/dict/corpus/ettenten/ (Accessed 2015-09-27.)
[3] http://www.keeleveeb.ee/ (Accessed 2015-09-10.)
[4] http://www.cl.ut.ee/korpused/segakorpus/index.php (Accessed 2015-09-27.)

illative forms and 420 aditive forms and they all are theoretically variative. Theoretical variativeness is based on the Dictionary of Standard Estonian ÕS 2013 (Erelt et al. 2013). All forms are included only once, so the illative form *majasse* 'house' could be in the initial material multiple times, but in the final 420 illative forms it is included only once. The parallel aditive form *majja* 'house' could be in the final 420 aditive forms only once. In the final data there are 41 words that have the illative and aditive form from the same word in the material.

The classification trees method is applied (see Strobl et al. 2009) because in similar grammatical alternation investigating studies the method was useful and provided dependable results (e.g. Tagliamonte & Baayen 2012; Klavan et al. 2015; Ruutma et al. 2016). Classification trees are easy to interpret: branches of a tree are divided from top to bottom and the first division is usually the one with the most significant explanatory variables. Levels are nodes in the left and right branch, where the tree can be further divided. A tree can be divided as many times as wanted so long as the nodes give meaningful information. In this paper a node had to be at least 25 observations. The software tool used to generate the tree was the statistical program R (version 3.5.1) party package.

3 Explanatory variables

3.1 Morphophonological variables

Gradation (**GRAD**). For every illative and aditive form the variable GRAD indicates if the word has gradation or not. The choice is made based on the Dictionary of Standard Estonian ÕS 2013 (Erelt et al. 2013). GRAD is a binary variable: a word has gradation or it does not have gradation.

Type of gradation (**GRAD_TYPE**). If the word has gradation, the variable GRAD_TYPE indicates whether it has quantity alternation or quality alternation. A word has quantity alternation if the nominative and partitive are in third degree and the genitive form is in second degree. For example *hetk* (NOM, Q3): *hetke* (GEN, Q2): *hetke* (PART, Q3) 'a while'. Or conversely, if nominative and partitive are in second degree and the genitive form is in third degree. For example, *taevas* (NOM, Q2): *taeva* (GEN, Q3): *taevast* (PART, Q2) 'sky'. The word has quality alternation if the nominative and partitive form second syllable first letter is a stop or *s*, which in the genitive is absent. For example *vesi* (NOM): *vee* (GEN): *vett* (PART) 'water'. Or conversely, if genitive form second syllable first letter is

a stop or *s*, which in nominative and partitive is absent. For example *hammas* (NOM): *hamba* (GEN): *hammast* (PART) 'tooth'. Words without gradation are in the level 'no'.

Direction of gradation (**GRAD_DRCT**). If the word has gradation, the variable GRAD_DRCT indicates whether it has strengthening or weakening gradation. The choice is made based on the genitive form of the word. If the genitive form is in the strong grade, then the word has strengthening gradation, e.g. *rooste* (NOM): *rooste* (GEN): *roostet* (PART) 'rust'. If the word is in the weak grade, then the gradation is weakening, e.g. *käsi* (NOM): *käe* (GEN): *kätt* (PART) 'hand'; *mäng* (NOM): *mängu* (GEN): *mängu* (PART) 'game'. Words without gradation are assigned to a level of 'no'.

Quantity degree of the base form (**QN_DGR**). For every form, the variable QN_DGT indicates the quantity degree of the base form. The base form can be in first, second or third degree and the Dictionary of Standard Estonian ÕS 2013 (Erelt et al. 2013) helps to determine it.

Stem-final alternation (**STEM_FINAL_ALT**). Based on the Handbook of Estonian, the variable STEM_FINAL_ALT indicates whether the analysed word has a stem-final alternation or not. If the nominative form of the word ends with a consonant, it always has stem-final alternation because in genitive form the base vowel is added. If the nominative form of the words ends with a vowel, then the word usually does not have stem-final alternation, except for words belonging to the *nimi*, *tuli*, *kole*, *habe*, *sai*, *lagi*, *käsi*, *nali* or *pääse* type. Also, *ne*- and *ke*-ending words always have stem-final alternation, e.g. *inimene* (NOM): *inimese* (GEN) 'human' or *tilluke* (NOM): *tillukese* (GEN) 'tiny'.

Stem-final alternation pattern (**STEM_FINAL_ALT_PTRN**). If the word has stem-final alternation, then the variable STEM_FINAL_ALT_PTRN indicates what the stem-final alternation pattern of the word is based on the Handbook of Estonian (Erelt et al. 2007: 231–233). In the Handbook of Estonian there are nine stem-final alternation patterns. The last five patterns are variants of the first three patterns and the 4th pattern consists of words without aditive forms. Therefore, in this article only the first three patterns are used. The 1st pattern is the most common. In the 1st pattern the initial stem is only used in the singular nominative case and in all other cases, as well as the plural nominative case, the inflectional stem is used. The 1st pattern includes words of the *seminar*, *redel*, *kringel*, *siil*, *sai*, *lagi*, *nali*, *sõber* and *õnnelik* types, e.g. *siil* (NOM): *siili* (GEN): *siili* (PART): *siilide* (pl. GEN): *siile* ~

siili/sid (pl. PART) ‘hedgehog’. The 2nd pattern has the most widespread base form and is used in singular partitive and plural genitive cases. This pattern characterises all *s*-ending words, which have a vowel before *s*, e.g. *panus* (sg. NOM): *panuse* (sg. GEN): *panust* (sg. PART): *panuste* (pl. GEN): *panuseid* (pl. PART) ‘contribution’. The 2nd pattern is common for words belonging to *suur* type, e.g. *suur* (sg. NOM): *suure* (sg. GEN): *suurt* (sg. PART): *suurte* (pl. GEN): *suuri* (pl. PART) ‘big’ and *küünal* type, e.g. *küünal* (sg. NOM): *küünla* (sg. GEN): *küünalt* (sg. PART): *küünalde* (pl. GEN): *küünlaid* (pl. PART) ‘candle’. The 3nd pattern is similar to the 2nd, except that in the singular partitive and plural genitive the short inflectional stem is used instead of the initial stem. This pattern characterises all *ke*- and *ne*-ending words, which have a vowel before *-ne*. The short inflectional stem of these words ends with the string -(V)*s*. This pattern covers words belonging to the *soolane*, *uus–küüs* or *käsi* type, e.g. *uus* (sg. NOM): *uue* (sg. GEN): *uut* (sg. PART): *uute* (pl. GEN): *uusi* (pl. PART) ‘new’; *käsi* (sg. NOM): *käe* (sg. GEN): *kätt* (sg. PART): *käte* (pl. GEN): *käsi* (pl. PART) ‘hand’. (Erelt et al. 2007: 231–233) Words without stem-final alternation are assigned to the level ‘no’.

Final sound of the base form (**FINAL_SOUND**). The variable FINAL_SOUND indicates the final sound of the base form, which is the singular nominative form. It can be a consonant or a vowel.

Number of syllables in the genitive stem (**SYL_GEN**). For every form the variable SYL_GEN indicates the number of syllables in the genitive stem. The forms in the data have one to six syllables in the genitive stems. The variable is divided into four levels: ‘1’, ‘2’, ‘3’ and ‘> 3’, if there are more than 3 syllables in the genitive stem.

In the §5 “New variable: the number of syllables in the last foot” the new morphophonological variable the number of syllables in the last foot of the word is included. It is controlled whether the new variable changes the results of the classification tree analyses.

3.2 Morfosyntactic variables

Part of speech (**P_O_SPCH**). Based on Explanatory Dictionary of the Estonian Language (Langemets et al. 2009) the variable P_O_SPCH indicates the part of speech of every analysed form. Types of declinable words are based on the academic grammar of Estonian (Erelt et al. 1993: 18) ‘numeral’, ‘pronoun’, ‘substantive’ and ‘adjective’.

Syntactic function (**SYN_FUN**). The academic grammar of Estonian categorizes sentences according to the following: predicate, base, object, predicative, adverbial and attribute. Words in the illative case can be adverbials or attributes (Erelt et al. 1993: 9–11). Thus the variable SYN_FUN indicates which of these two levels a word in the illative case has.

Government (**GOV**). The variable GOV indicates the government of a word and is decided based on Mäearu's list "Valik rektsioone" ('Choice of governments'). In that list there are widespread governments like *puutuma millessegi* 'to pertain to something' (lit. 'to concern into something'), *suhtuma millessegi* 'to relate to something; to have an opinion about something' (lit. 'to regard into something'), *uskuma millessegi* 'to believe in something' (lit. 'to believe into something'). For example, if the verb *uskuma* 'believe' is in government structure, substantives *usk* 'belief', *uskuja* 'believer', *uskumine* 'believing' etc. are counted as a government structure parts. Levels are 'yes' if the form is in a government structure and 'no' otherwise.

Multi-word expression (**M_W_E**). The variable M_W_E indicates multi-word expression and is determined by the whole phrase, unlike government structure which is determined by one word. Determinations are based on the database of Estonian verbal multi-word expressions.[5] In description of this database it "contains a subtype of multi-word expressions, namely those consisting of a verb and a particle or a verb and its complements". Expressions with illative or aditive form are in the database, e.g. *jõusse jääma* 'to remain in force', *põhja kõrbema* 'to go out of business, to burn' (lit. 'to burn into the bottom'), *riidesse panema* 'to put on clothes'. Levels are 'yes', if the form is in a multi-word expression phrase, and 'no' otherwise.

3.3 Semantic variables

Proper or common noun (**PN_CN**). If the text is edited, the variable PN_CN determines whether the word is a proper or common noun and is decided by the letter at the beginning of the word. For example, if *Keskus* or *Riigikogu* begins with a capital letter, they are proper nouns. If *keskus* 'centre' or *riigikogu* 'parlament of Estonia' begin with a lower case letter, then they are common nouns. If the text is not edited, then well-known

[5] https://www.cl.ut.ee/ressursid/pysiyhendid/ (Accessed 2015-09-17.)

proper names are decided as proper nouns despite that they begin with lower case letters, e.g. *kopli* 'Kopli, district of Tallinn', *rate* 'Rate.ee, internet site'.

Proper noun semantic group (**PN_SEM**). The variable PN_SEM includes only proper nouns and indicates semantic group: INSTITUTION (*Klubi*, *Gümnaasium*, *Riigikogu* and so with the capital letter in the beginning), PERSON (all person names with capital letters, e.g. *Jeesus Kristus*) and PLACE (all place names with capital letters, also few internet sites and books). The final level is 'common noun'.

Common noun semantic group (**CN_SEM**). The first common noun semantic group is BODY PART, e.g. *käsi* 'hand', also mental *meel* 'mind' and more general *keha* 'body'. The second group is PLACE, in this level belong all substantives that mean real places, e.g. *kodu* 'home', *muuseum* 'museum', *põuetasku* 'breast pocket', also more abstract places when they are destinations, e.g. *keel* 'language/tongue', *valdkond* 'area', *üksus* 'unit'. The third group STATE comprises body states (e.g. *riidesse panema* lit. 'to put into clothes') and mind states, which get their meaning when they are in a phrase (e.g. *jõusse jääma* 'to remain in force', *unustusse jääma* lit. 'to remain into the forgetfulness'). These groups could be a bit subjective because it is hard to define abstract place and abstract state. The fourth group is OTHER, which can include government structures, e.g. *puutuma millessegi* 'to pertain to something' (lit. 'to concern into something'), *suhtuma millessegi* 'to relate to something; to have an opinion about something' (lit. 'to regard into something'), attributes, (e.g. *praegune* 'present'), times/periods (e.g. *periood* 'period'). The last level is 'proper nouns'.

Meaning of the verb lemma (**VERB_LEMMA**). The variable VERB_LEMMA indicates one of four levels: 'movement', 'existence', 'activity' or 'no verb'. The subcategorisation of verbs is quite basic. Verbs in the level 'movement' describe some kind of motion or a way of moving is described, e.g. *jõudma* 'arrive', *lahkuma* 'leave', *sõitma* 'drive'. The 'existence' level is more static, e.g. *jääma* 'stay', *kuuluma* 'belong', *olema* 'be'. Verbs in the 'activity' level describe all active and mental activities, e.g. *lisama* 'add', *suhtuma* 'regard', *töötama* 'work'. Activity verbs have participant, who is aware of his/her action and who controls his/her action. The 'no verb' level means that the analysed form is not related to any verb. For that case there is no verb in string, e.g. *väike kõrvalepõige mõistmisesse* 'little artifice to understanding', *tagasivaade märtsisse 2003* 'flashback to March 2003'. Another option is that based on the corpus string there is no

verb in the clause part where the illative form belongs, e.g. [...] *kuid püsiva järjekindlusega üleväärsusest alaväärsusesse ning* [...] '[...] but constant consistency from superiority to inferiority and [...]'; [...] *Lõuna-Eesti ärikeskkonda tutvustav õppesõit Mooste mõisa kultuuri- ja ettevõtluskeskusesse* '[...] South-Estonian business environment describing learning trip to Mooste manor culture and entrepreneurship centre'.

Table 1 presents all coded variables and levels with frequency division in analysed data.

Table 1: Analysed data by coded variables (N = 840, illative = 420, aditive = 420)

	Variables	**Frequency division**
Morpho-phonological variables	**Gradation (GRAD)**	
	No	551
	Yes	289
	Type of gradation (GRAD_TYPE)	
	No	551
	Quantity alternation	205
	Quality alternation	84
	Direction of gradation (GRAD_DRCT)	
	No	551
	Weakening	277
	Strengthening	12
	Quantity degree of the base form (QN_DGR)	
	Q1	208
	Q2	104
	Q3	528
	Stem-final alternation (STEM_FINAL_ALT)	
	No	242
	Yes	598
	Stem-final alternation pattern (STEM_FINAL_ALT_PTRN)	
	No	242
	1	299
	2	216
	3	83
	Final sound of the base form (FINAL_SOUND)	
	C	480
	V	360
	Number of syllables in the genitive stem (SYL_GEN)	
	1	7
	2	234
	3	171
	>3	428

	Variables	Frequency division
Morpho-syntactic variables	**Part of speech (P_O_SPCH)**	
	Adjective	44
	Numeral	8
	Pronoun	8
	Substantive	780
	Syntactic function (SYN_FUN)	
	Adverbial	795
	Attribute	45
	Government (GOV)	
	No	711
	Yes	129
	Multi-word expression (M_W_E)	
	No	799
	Yes	41
Semantic variables	**Proper or common noun (PN_CN)**	
	Common noun	720
	Proper noun	120
	Proper noun semantic group (PN_SEM)	
	INSTITUTION	28
	PERSON	7
	PLACE	85
	Common noun	720
	Common noun semantic group (CN_SEM)	
	BODY PART	27
	PLACE	460
	STATE	61
	Other	172
	Proper noun	120
	Meaning of the verb lemma (VERB_LEMMA)	
	Movement	481
	Existence	98
	Activity	225
	No verb	36

4 Results

4.1 Classification tree using all 16 variables

Figure 1 shows a classification tree where all 16 variables[6] are included.

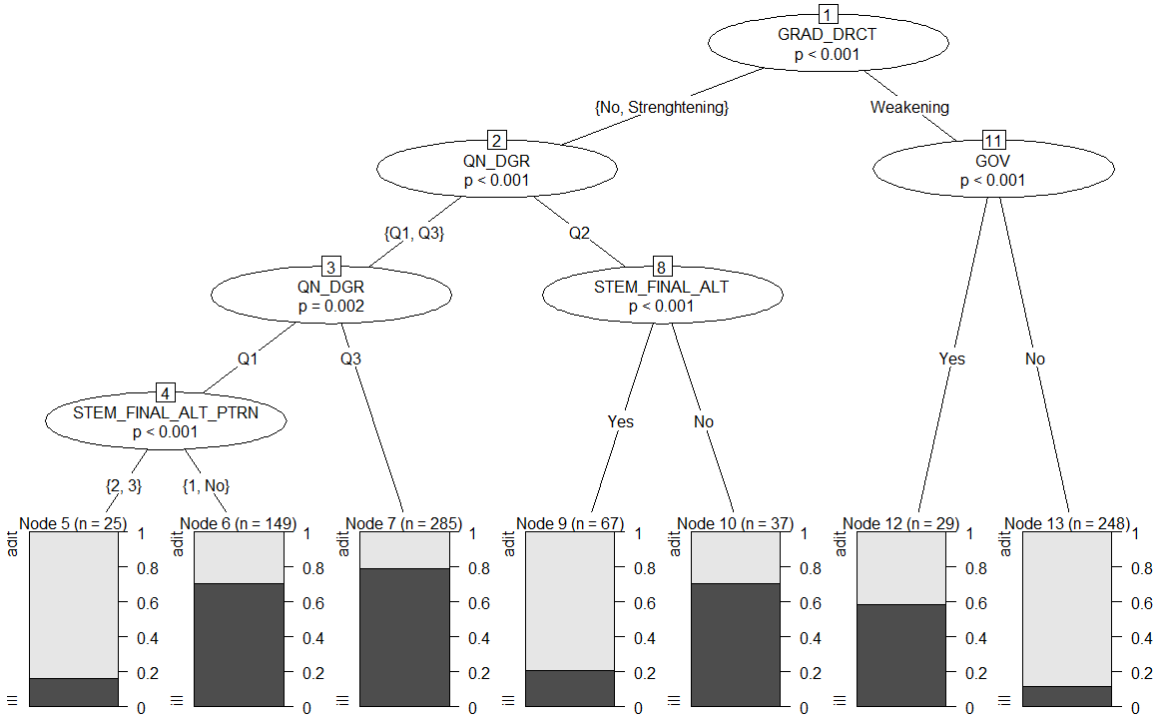

Figure 1: Morphophonological, morphosyntactic and semantic variables to which the choice between the illative and aditive could be related.

Based on Figure 1, the most significant predictor is the direction of gradation. Further splits are made based on the quantity degree of the base form or government variables. In the lower branches there are also stem-final alternation and the stem-final alternation pattern variables.

The direction of gradation variable splits the data into two groups: the first group consists of words without gradation (551) or words with strengthening gradation (12); the second group consists of words with weakening gradation (277).

In the weakening gradation group, government is the most significant predictor. If a word with weakening gradation belongs to government

[6] ctreeilldata = ctree (Adit_ill ~ GRAD + GRAD_TYPE + GRAD_DRCT + QN_DGR + STEM_FINAL_ALT + STEM_FINAL_ALT_PTRN + FINAL_SOUND + SYL_GEN + P_O_SPCH + SYN_FUN + GOV + M_W_E + PN_CN + PN_SEM + CN_SEM + VERB_LEMMA, controls = ctree_control(minbucket=25), data = illdata)
plot(ctreeilldata)

structure, illative is more commonly used, e.g. *asja**sse** puutuma*[7] 'to pertain to something' (lit. 'to concern into a thing'), *loo**sse** suhtuma* 'to relate to a story; to have an opinion about a story' (lit. 'to regard into a story'), *hinnasõja**sse** uskuma* 'to believe in a price war' (lit. 'to believe into a price war'). If a word with weakening gradation does not belong to government structure, the aditive is more frequent, e.g. *garderoobi*[8] 'dressing room', *nimekirja* 'list', *riiki* 'country'.

In the other group, i.e. words without gradation or with strengthening gradation, the most significant predictor is the quantity degree of the base form, which divides into two nodes: first- and third-degree words and second-degree words. If the word has no gradation and is in the third degree of quantity, the illative is preferred, e.g. *alaväärsuse**sse*** 'inferiority'. *Elva**sse*** 'Elva', *tootmise**sse*** 'manufacture'. Third-degree words cannot be with strengthening gradation because they are already in the strongest degree. If the word has first degree of quantity and it has no gradation or strengthening gradation, the significant predictor is the stem-final alternation pattern. The 2nd and the 3rd pattern words (mostly *ne*- and *s*-ending, see §3) seem to have a clear preference for aditive, e.g. *unne* 'sleep', *nimelisse* 'named', *metsasügavusse* 'forest-deepness'. For words without stem-final alternation or the 1st pattern words illative is mostly used, e.g. *lisa**sse*** 'appendix', *mentorklubi**sse*** 'mentor-club', *kuusetüve**sse*** 'bole of fir'.

For second-degree words without gradation or with strengthening gradation the stem-final alternation is the next significant predictor. If these words do not have stem-final alternation, the illative is more commonly used, e.g. *Viljandi**sse*** 'Viljandi', *Poola**sse*** 'Poland', *kütke**sse*** 'fetter'. If there is a stem-final alternation, the aditive is more likely to be chosen, e.g. *teise* 'second/other', *liiklusõnnetusse* 'traffic-accident', *rakendusse* 'application'.

For third-degree words without gradation (285) the illative is more commonly used (226 forms of 285), while third-degree words with weakening gradation (243) make more use of the aditive (204 forms of 243). For third-degree words the significant predictor of the choice between the illative and additive is the direction of gradation, i.e. whether the word is without gradation or with weakening gradation.

---

[7] Illative forms have *sse*-ending in bold because it is an agglutinative ending. Translations are in nominative case.

[8] Aditive forms do not have bold endings, because it is a fusional ending. Translations are in nominative case.

Next it is looked at only *ne*- and *s*-ending words, it means mostly words in the 2nd and the 3rd stem-final alternation pattern. If these kinds of words are first- or second-degree words (89), they seem to have a preference for aditive (74 forms of 89), if they are third-degree words (210), the illative is preferred (159 forms of 210).

It is notable that 8 morphophonological, 4 morphosyntactic and 4 semantic variables were analysed and that the significant predictors were 4 morphophonological variables and 1 morphosyntactic variable (no semantic variables were significant predictors). The result that most of the significant variables were morphophonological variables confirms the claim in the academic grammar of Estonian that the use of the short illative is related to a word's phonological-derivative structure.

Based on the academic grammar of Estonian the use of the short illative is also related to morphosyntactic factors, e.g. whether a word is part of the government structure or part of a multi-word expression. The Figure 1 classification tree confirms that the choice between the illative and aditive is related to whether a word is part of the government structure. The classification tree does not provide information about whether the choice between the illative and additive relates to whether a word is part of a multi-word expression.

The analysis does not claim that the choice is related to words lexical meaning (Erelt et al. 1995: 56), because in the Figure 1 classification tree there are no semantic variables. It is difficult and rather subjective to divide proper or common nouns into a few semantic groups. It could be a reason why semantic variables are not in the classification tree.

4.2 Classification tree using significant variables from prior univariate analyses

It is quite surprising that the most significant predictor when using all variables is the direction of gradation. Only significant variables will be analysed based on prior studies, which are gradation, the type of gradation, stem-final alternation, the stem-final alternation pattern, the final sound of the base form, the number of syllables in the genitive stem, government, multi-word expression, proper or common noun, the proper noun semantic

group and the common noun semantic group.[9] The results are shown in Figure 2.

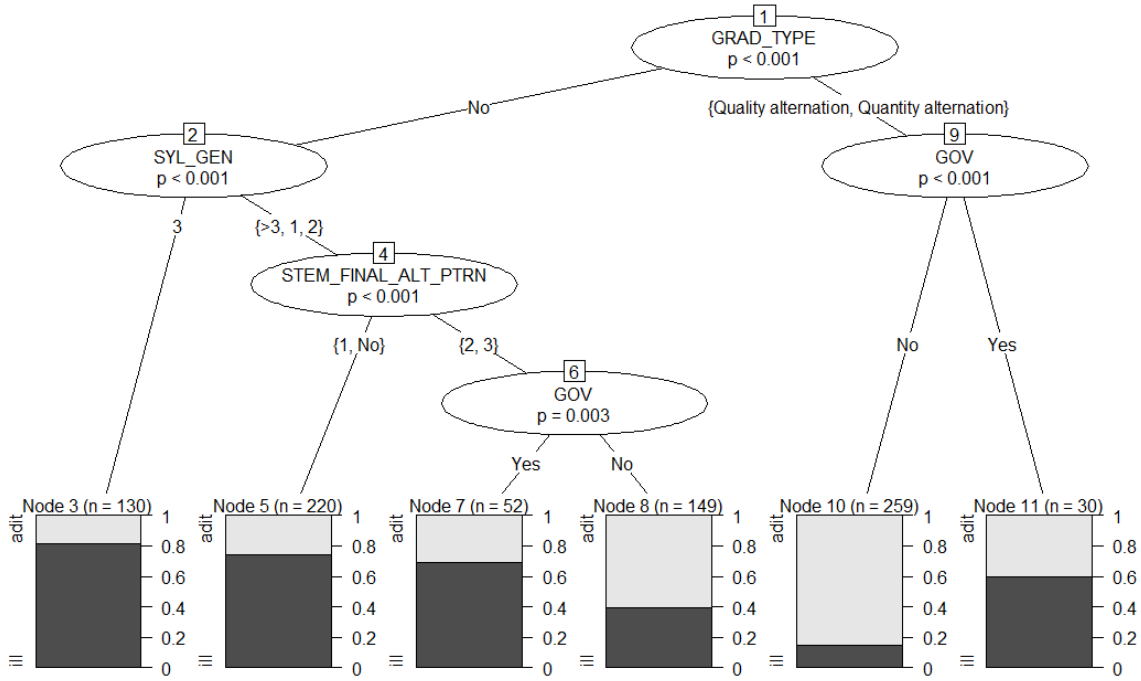

Figure 2: Classification tree using morphophonogical, morphosyntactic and semantic variables for which the choice between the illative and aditive was related based on univariate analysis (Metslang 2015; Siiman 2016).

The strongest predictor is the type of gradation, which splits the tree into two branches: without gradation or with gradation (with quality or quantity alternation). It is interesting that the strongest predictor is the type of gradation, not gradation itself.

Words with gradation have only one split. Words with gradation have a tendency to use illative when they belong to a government structure, e.g. *signaali**sse** suhtuma* ‘to relate to a signal; to have an opinion about a signal’ (lit. ‘to regard into the signal’), *usu**sse** puutuma* ‘to pertain to a belief’ (lit. ‘to concern into belief’), *kõrreroostes**sse** nakatuma* ‘to be infected by blight’ (lit. ‘to infect into blight’). If the word with gradation does not belong to government structure, the aditive is more often used, e.g. *ajalukku* ‘history’, *atmosfääri* ‘atmosphere’, *parki* ‘park’.

[9] ctreeilldata = ctree (Adit_ill ~ GRAD + GRAD_TYPE + STEM_FINAL_ALT + STEM_FINAL_ALT_PTRN + FINAL_SOUND + SYL_GEN + GOV + M_W_E + PN_CN + PN_SEM + CN_SEM, controls = ctree_control(minbucket=25), data = illdata) plot(ctreeilldata)

The number of syllables in the genitive stem is a predictor for words without gradation. A distinction is made between 3-syllable words and other word lengths. 3-syllable words without gradation make more use of illative, e.g. *vestluse**sse*** ‘conversation’, *Hollandi**sse*** ‘Holland’, *kinnise**sse*** ‘closed’. 1-, 2- and more than 3-syllable words without gradation have next predictor the stem-final alternation pattern. For the 1st pattern words or word without stem-final alternation the illative is used more frequently, e.g. *maja**sse*** ‘maja’, *universumi**sse*** ‘universe’, *Soome**sse*** ‘Finland’. The illative is also used more frequently with the 2nd the 3rd pattern words which belong to government structure, e.g. *suhtuma teineteise**sse*** ‘to relate to each other; to have an opinion about each other’ (lit. ‘to regard into each other’), *sisenema administreerimiskeskuse**sse*** ‘to enter to an administration-center’, *puutuma teispoolsuse**sse*** ‘to pertain to the afterlife’ (lit. ‘to concern into the afterlife’). However, if these words do not belong to a government structure then the aditive is more common, e.g. *unne* ‘sleep’, *juhatusse* ‘management’, *üleriigilisse* ‘nationwide’.

There was no predictor like gradation in Figure 1 and 2 classification trees, but it appears that this variable has a central role in choosing between the illative and aditive. In Figure 1, one branch has words without gradation or with strengthening gradation and the other branch has words with weakening gradation. In Figure 2 one branch is words without gradation and the other branch is words with gradation. A significant predictor in both classification trees is also government.

A more specific analysis focusing separately on morphophonological, morphosyntactic and morphosemantic variables is next discussed.

4.3 Classification tree using only morphophonological variables

In Figure 3 a classification tree using 8 morphophonological variables[10] is presented. The 8 variables analysed were gradation, the type of gradation, the direction of gradation, the quantity degree of the base form, stem-final alternation, the stem-final alternation pattern, the final sound of the base form and the number of syllables in the genitive stem.

[10] ctreeilldata = ctree (Adit_ill ~ GRAD + GRAD_TYPE + GRAD_DRCT + QN_DGR + STEM_FINAL_ALT + STEM_FINAL_ALT_PTRN + FINAL_SOUND + SYL_GEN, controls = ctree_control(minbucket=25), data = illdata)
plot(ctreeilldata)

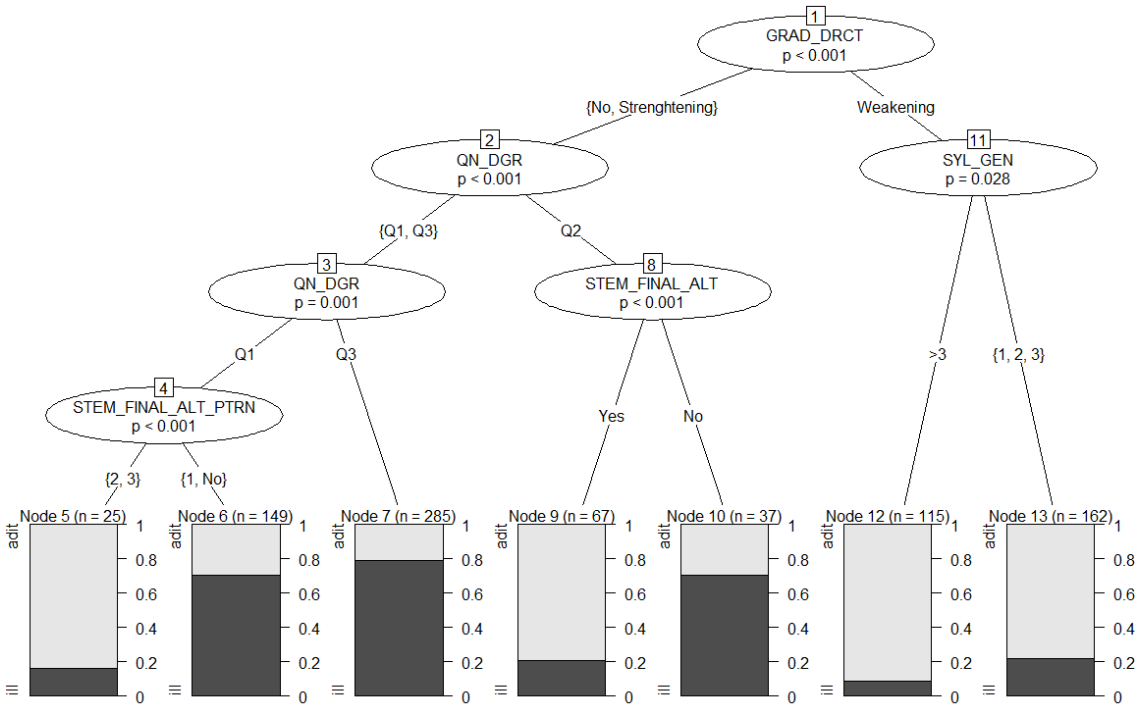

Figure 3: Morphophonological variables to which the choice between the illative and aditive could be related.

As expected, the left branch of the morphophonological variables classification tree is similar to the classification tree shown in Figure 1, in which all the variables were included. The left branch consists of words without gradation and with strengthening gradation. The right branch contains words with weakening gradation, which divides according to the number of syllables a word has in the genitive form. Words with more than three syllables in the genitive stem are more likely in the aditive than words with one, two or three syllables in the genitive stem. It is important to note that all of these words are mostly in the aditive, so the difference is not significant.

4.4 Classification tree using only morphosyntactic variables

Figure 4 shows the results of analysis using only morphosyntactic variables. The choice between the illative and aditive could be related to part of speech, syntactic function, government and multi-word expression.[11]

[11] ctreeilldata = ctree (Adit_ill ~ P_O_SPCH + SYN_FUN + GOV + M_W_E, controls = ctree_control(minbucket = 25), data = illdata)
plot(ctreeilldata)

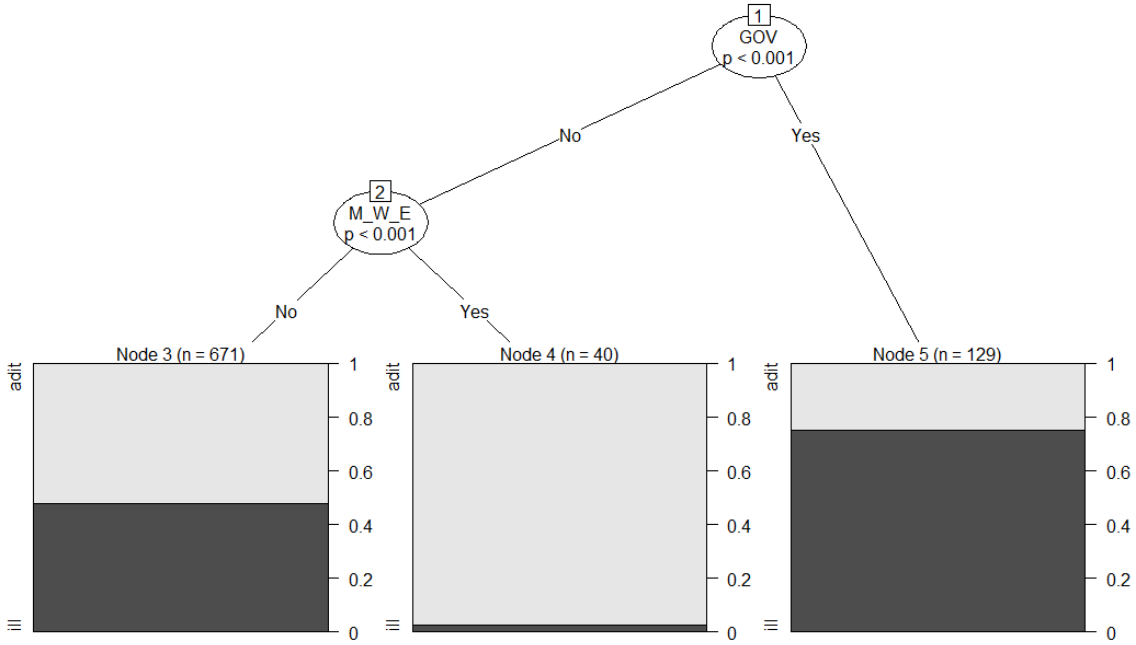

Figure 4: Morphosyntactic variables to which the choice between the illative and aditive could be related.

As shown in Figures 1 and 2, the most significant morphosyntactic predictor is government. If the word belongs to government structure, the illative is used more often, e.g. *asja**sse** puutuma* 'to pertain to something' (lit. 'to concern into a thing'), *loo**sse** suhtuma* 'to relate to a story; to have an opinion about a story' (lit. 'to regard into a story'), *hinnasõja**sse** uskuma* 'to believe in a price war' (lit. 'to believe into a price war'). If the word does not belong to government structure, then the choice between the illative and aditive is related to a multi-word expression. Words that belong to a multi-word expression are mostly in the aditive, e.g. *pähe hakkama* 'to go to your head', *meelde tulema* 'to remind' (lit. 'bring into the mind'), *põhja kõrbema* 'to go out of business; to burn' (lit. 'to burn into the bottom'). In the illative there is only one example – *jõu**sse** jätma* 'to remain in force'. If the word does not belong to a multi-word expression, then the aditive is slightly more common than the illative. A similar tendency has been shown before – that the illative is used more often with words with government structures (Erelt et al. 2007: 247; Siiman 2016: 218) and that the aditive is preferred for words that belong to a multi-word expression (Erelt et al. 1995: 56–57; Kio 2006: 126; Siiman 2016: 219).

4.5 Classification tree using only semantic variables

Lastly, it was analysed which semantic variables – proper or common noun, the proper noun semantic group, the common noun semantic group and meaning of the verb lemma – could be related to the choice between the illative and aditive.[12]

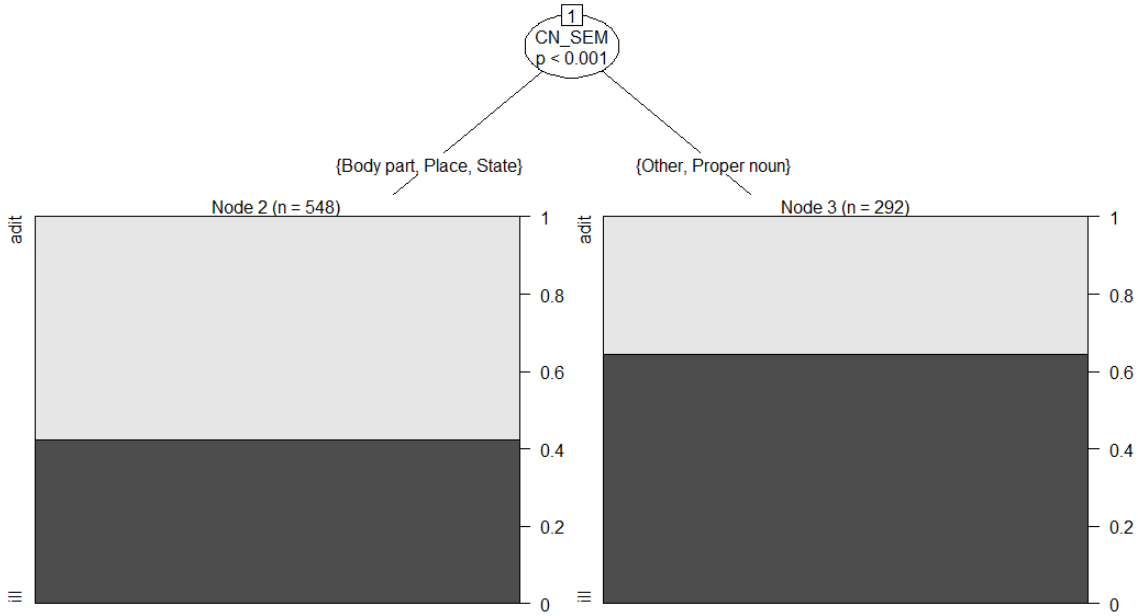

Figure 5: Semantic variables to which the choice between the illative and aditive could be related.

Figure 5 shows that the most significant semantic predictor is the common noun semantic group. For body part, place and state words, the aditive is more commonly used, e.g. *koju* 'home', *kurku* 'throat', *vabadusse* 'freedom'. Words in the group 'other' and proper nouns are more likely to be used in the illative, e.g. *pisikese**sse*** 'tiny', *Pärnu**sse***, *Tartu**sse***.

5 New variable: the number of syllables in the last foot

One morphophonological variable analysed in the previous sections was the number of syllables in the genitive stem: 1, 2, 3 or > 3. However, more than half of the data (428 words out of 840) is in level '> 3'. Thus, to avoid too much data coded in a single level it was decided to consider prosody

[12] ctreeilldata = ctree (Adit_ill ~ PN_CN + PN_SEM + CN_SEM + VERB_LEMMA, controls = ctree_control(minbucket = 25), data = illdata)
plot(ctreeilldata)

and count the number of syllables in the last foot of the word. This means that the syllables are counted from the genitive form last stressed syllable. I am often concerned with secondary stress (not primary stress) when identifying the last stressed syllable. The number of syllables in the last foot can be 1, 2 or 3. It is not always clear, which syllable is the last stressed syllable of a word and how to syllabify a word (e.g. Hint 1980a, 1980b, 1980c). In this article words are syllabified based on Dictionary of Standard Estonian ÕS 2013 (Erelt et al. 2013). There are 4 levels: '1' if there is one syllable in the last foot (e.g. *bakalaureusetöö* 'bachelor thesis', *jõud* 'strength', *tondilugu* 'ghost story'), '2' if there are two syllables in the last foot (e.g. *inimene* 'human', *patsient* 'patient', *tonn* 'ton'), '3' if in the last foot there are three syllables (e.g. *Holland* 'The Netherlands', *Siber* 'Siberia', *Viljandi*) and '2 or 3', if the last foot can be based on the Dictionary of Standard Estonian ÕS 2013 (Erelt et al. 2013) two or three syllables long (e.g. *administreerimiskeskus* 'administration centre', *keskkonnateadlikkus* 'environmentalism', *ministeerium* 'ministry'). In Figure 6 are included all 16 variables and the new variable number of syllables in the last foot (SYL_LF).[13]

[13] ctreeilldata = ctree (Adit_ill ~ GRAD + GRAD_TYPE + GRAD_DRCT + QN_DGR + STEM_FINAL_ALT + STEM_FINAL_ALT_PTRN + FINAL_SOUND + SYL_GEN + SYL_LF + P_O_SPCH + SYN_FUN + GOV + M_W_E + PN_CN + PN_SEM + CN_SEM + VERB_LEMMA, controls = ctree_control(minbucket = 25), data = illdata)
plot(ctreeilldata)

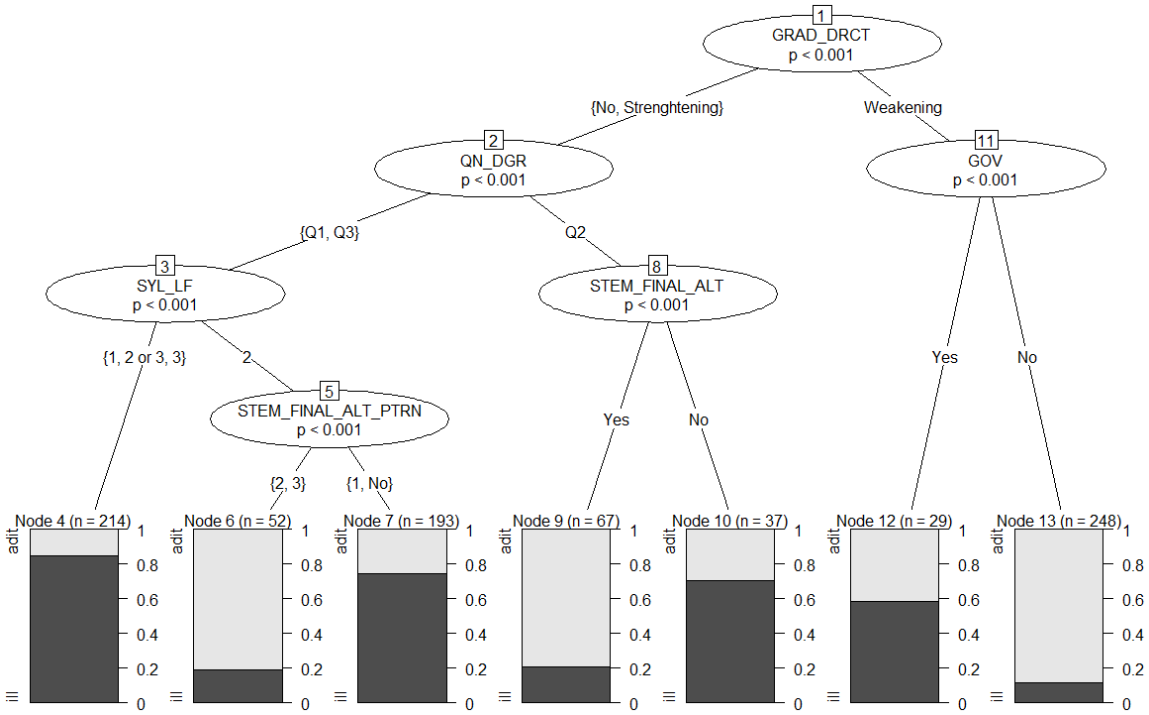

Figure 6: Classification tree using the number of syllables in the last foot variable (SYL_LF)

It turns out that Figure 6 is quite similar to Figure 1 where all 16 variables without new variable were analysed. Again the most significant predictor to choose between the illative and aditive is the direction of gradation, followed by the quantity degree of the base form, government and stem-final alternation. The difference from Figure 1 is that the quantity degree of the base form is not followed by the same variable again, but by the new variable 'the number of syllables in the last foot'. The branches do not split by first- and three-degree words, but by the number of syllables in the last foot. The lowest branch is again the same 'the stem-final alternation pattern'.

The strongest predictor is the direction of gradation, which divides the tree into two nodes: words without gradation (551) or words with strengthening gradation (12), which prefer the illative (374 illative forms out of 563). In the other branch are words with weakening gradation (277), which are used in the aditive (231 aditive forms out of 277). Words with weakening gradation split into two groups by government. If the word has weakening gradation and belongs to government structure, it has a tendency to occur in the illative, e.g. *asjasse puutuma* 'to pertain to something' (lit. 'to concern into a thing'), *loosse suhtuma* 'to relate to a story; to have an opinion about a story' (lit. 'to regard into a story'),

*hinnasõja**sse** uskuma* ‘to believe in a price war’ (lit. ‘to believe into a price war’). If a word with weakening gradation does not belong to government structure, then the aditive is more likely to be chosen, e.g. *garderoobi* ‘dressing room’, *nimekirja* ‘list’, *riiki* ‘country.’

Words without gradation or with strengthening gradation are divided by the quantity degree of the base form. In the first group are first- and third-degree words and in the second group are second-degree words. For first- and third-degree words the significant predictor is the number of syllables in the last foot. For second-degree words the significant predictor is stem-final alternation. Second-degree words without gradation or with strengthening gradation with stem-final alternation are mostly in the aditive, e.g. *ajakirjandusse* ‘press’, *liiklusõnnetusse* ‘traffic accident’, *teise* ‘second/other’. Similar words without stem-final alternation are mostly in the illative, e.g. *kütke**sse*** ‘feter’, *loetelu**sse*** ‘list’, *Poola**sse*** ‘Poland’). These same branches were in Figure 1 where the new variable was not taken into account. First- and three-degree words without gradation or with strengthening gradation split by the number of syllables in last foot to 2-syllable words or 1-, 2- or 3- and 3-syllable words. Words in the last branch make more use of the illative, e.g. *peatuse**sse*** ‘halt’, *päevakeskuse**sse*** ‘day-centre’, *Viljandi**sse*** ‘Viljandi’. It is difficult to describe this branch but the conclusion is simple: third-degree *ne*- and *s*-ending words occur more in the illative because in this branch there are mostly third-degree *ne*- and *s*-ending words based on the current data. In the other branch there were words with two syllable foot. If these words had the 2nd or the 3rd stem-final alternation pattern, then the aditive is more frequently used, e.g. *juhatusse* ‘management’, *jäädvustamisse* ‘perpetuate’, *üleriigilisse* ‘nationwide’. The 1st pattern or words without stem-final alternation pattern are more likely in the illative, e.g. *bussitasku**sse*** ‘bus wagon’, *Ruhnu**sse*** ‘Ruhnu’, *voodi**sse*** ‘bed’. Based on the data it is possible to conclude that third-degree *ne*- and *s*-ending words are mostly in the illative and second-degree *ne*- and *s*-ending words are mostly in the aditive. The same conclusion was found in §4.1, where the 2nd and the 3rd pattern first- and second-degree words (89) preferred aditive (74 forms out of 89) and third-degree words (210) were mostly in the illative (159 forms out of 210).

Figure 6 shows that the number of syllables in the last foot is a significant predictor. The number of syllables in the last foot takes into account pronunciation. In further research the number of syllables in the genitive stem could be replaced by the number of syllables in the last foot to be more accurate. The purpose of this article was to analyse previous

variables using multivariate analysis, and therefore it was not possible to not take into account the number of syllables in the genitive stem or to replace this variable.

6 Comparison of univariate and multivariate analysis

In previous studies 8 morphophonological, 4 morphosyntactic and 4 semantic variables were analysed using univariate analysis (Metslang 2015; Siiman 2016). Morphosyntactic and semantic variables were controlled with a so-called part-whole method and the Cramér's V effect size method. It was found that the choice between the illative and aditive could be related to gradation, the type of gradation, stem-final alternation and the stem-final alternation pattern, the final sound of the base form, the number of syllables in the genitive stem, government, multi-word expression, proper or common noun, the proper noun semantic group and the common noun semantic group. From all of the 16 variables the direction of gradation, the quantity degree of the base form, part of speech, syntactic function and meaning of the verb lemma were not statistically significant in the choice between the illative and aditive.

These same variables were analysed in this article using multivariate analysis – classification tree method. Based on the classification tree analyses the most significant predictors in the choice between the illative and aditive are the direction of gradation and the quantity degree of the base form. In a prior study the direction of gradation and the quantity degree of the base form were not statistically significant factors (Metslang 2015). To control for these results the data from Metslang (2015) was analysed using the classification tree method, which resulted in the direction of gradation being the most significant predictor for choosing between the illative and aditive. Words with weakening gradation had only one predictor 'the direction of gradation' and these words made more use of the aditive, e.g. *põhja* 'bottom; north', *selga* 'back', *sõlme* 'knot'. Words with strengthening gradation or without gradation have besides 'the direction of gradation' three more predictors: 'the quantity degree of the base form', 'the number of syllables in the genitive stem' and 'the stem-final alternation pattern'.

Hence, making a new analysis with the classification tree method using data from Metslang (2015) leads to the result that the most significant factor is the direction of gradation and the next most significant factor is the quantity degree of the base form. The direction of gradation was not a

significant factor in Metslang (2015) using univariate analysis because perhaps there were only 12 illative forms with weakening direction of gradation and 12 illative forms with strengthening direction of gradation. The method resulted in the direction of gradation variable being not statistically significant: $X^2(2, N = 1710) = 3.03$, $p = 0.2$. Metslang (2015) and this study results differ because of the different method and data collection principles. Due to balanced data in this study, the data includes more illative case forms and it is possible to get statistically significant results.

Siiman (2016) analysed 4 morphosyntactic and 4 semantic variables. Of the 8 variables, 5 were significant factors. One statistically significant factor was government, which is significant also in this study. Based on uni- and multivariate analysis the words in government structures occur in the illative and words that are not in government structures prefer the aditive. When all 16 variables were analysed none of the semantic variables were significant (see Figure 1). Considering only semantic variables in the classification tree (see Figure 5), then the results of Siiman (2016) and this study are similar: i.e., proper names (people and place names) have a tendency to occur in the illative and common noun place and state phrases are mostly in the aditive. Based on the current analyses the aditive is preferred also with body part words.

Univariate analysis answers the question "With what variables is the illative more often used and with what variables is the aditive more commonly used?" Multivariate analysis answers the question "Which variables are significant in the choice between the illative and aditive?" Thus, univariate analysis gives preliminary results, e.g. words without gradation are mostly in the illative. Multivariate analysis gives more specific results, e.g. third-degree words without gradation are usually in the illative. For first-degree words without gradation, the choice between the illative and aditive may also be related to the stem-final alternation pattern. For second-degree words without gradation the illative and aditive may also be related to stem-final alternation. In summary, the significant factors for the choice between the illative and aditive are the direction of gradation, the quantity degree of the base form, government, stem-final alternation and the stem-final alternation pattern. Based on univariate analysis, there are more significant factors and the direction of gradation and the quantity degree of the base form are not significant factors.

The fewer branches a classification tree has, the easier it is to interpret the tree. If there are many variables, the description of words could be

confusing, e.g. the illative is more common with first-degree words without gradation or with weakening gradation without stem-final alternation or with the 1st stem-final alternation pattern, e.g. *murusse* 'grass', *peresse* 'family', *sõnasse* 'word'.

It appears that the classification tree method is more accurate than univariate analysis because classification tree gives hierarchy about factors, not only *p*-values. In Siiman (2016) factors were hierarchically organised only using the Cramér's V effect size method. Only morphosyntactic and semantic variables were used and the results are similar to the results of the current study.

Based on the Cramér's V effect size method the significant predictors for the choice between the illative and aditive were the common noun semantic group (0.22), government (0.21) and multi-word expressions (0.2). The effect size was smaller with variables the proper noun semantic group (0.15) and proper or common noun (0.12) – variables that were not in this article's classification trees. (Siiman 2016: 227)

Multivariate analysis seems to be well suited for analysing linguistic data since it is less sensitive to sample size – it is possible to determine the minimum number of observations and the results are not missing by the disproportionate distribution of the observations. Univariate analysis is needed to find good preliminary results, but multivariate analysis methods should be used to explore grammatical alternatives.

7 Conclusion

This study examined the variation of the Estonian illative case based on Estonian language material. Using classification trees, it was explained which morphophonological, morphosyntactic and semantic variables most affect the choice between the illative and aditive.

In the first analysis, all the variables were considered, according to which the significant predictor in choosing the long or short illative case was the direction of gradation followed by the quantity degree of the base form, government, stem-final alternation and the stem-final alternation pattern. It turns out that the choice between the illative and aditive is affected by morphophonological variables, which confirm the claim in the academic grammar of Estonian that the choice between the illative and aditive is related to a word's phonological-derivative structure.

Morphophonological, morphosyntactic and semantic variables were also analysed separately. Considering only morphophonological variables,

the significant predictors for the choice between the illative and aditive were the direction of gradation, the quantity degree of the base form, stem-final alternation, the stem-final alternation pattern and the number of syllables in the genitive stem. Analysis of only morphosyntactic variables indicated that the significant predictors were government and multi-word expression. The same result was obtained in earlier studies, in which government structures prefer the illative (Erelt at al. 2007: 247; Siiman 2016), and in which multi-word expressions are more in the aditive (Erelt et al. 1995: 56–57; Kio 2006: 112–113, 126; Siiman 2016). Considering only semantic variables, the significant predictor for the choice between the illative and aditive was the common noun semantic group. In a previous study, the additive was preferred with the proper noun semantic group (personal names and place names) (Siiman 2016). In this study, the illative was used with proper nouns and with the common noun semantic group 'other'. Furthermore, in both studies the common noun place and state phrases occurred mostly in the aditive. In this study, the aditive also occurred with body part words.

Regarding third-degree words, it turns out that according to this analysis, the choice between the illative and aditive is related to the direction of gradation: words without gradation are more used in the illative and words with weakening gradation prefer the aditive. It was also concluded that in the case of words with a weakening gradation the choice between the illative and aditive is related to government. *ne*- and *s*-ending words (words in the 2nd and the 3rd stem-final alternation pattern) are more likely in the aditive, if they are first- or second-degree words. If these *ne*- and *s*-ending words are third-degree words, then they are more often used in the illative.

One morphophonological variable was added to the 16 variables already analysed – the number of syllables in the last foot. It was found that the analysis would be more accurate if the variable number of last foot could replace the variable the number of syllables in the genitive stem.

Comparing uni- and multivariate analysis, the multivariate method gives more information and is more precise, i.e. it can draw conclusions about the concurrence of several variables. According to the analysis here, the most significant predictors for the choice between the illative and aditive are the direction of gradation and the quantity degree of the base form. However, this result was not obtained in a univariate analysis, and so it can be argued that although a univariate analysis might be suitable for a preliminary analysis, the results should be verified by multivariate analysis.

Then the results can be calculated on the basis of fewer observations and it is possible to set the minimum number of observations.

In the future, the illative variation should also be investigated by other methods. In addition to data from a corpus analysis, surveys could be carried out for studying the illative variation by analogy or experiments could be conducted where Estonian speakers select whether they prefer the singular long or short illative form. The illative variation is a good example of a grammatical alternation, the study of which could be generalised to similar alternation in other languages.

References

Barlow, Michael & Kemmer, Suzanne. 2000. Introduction: a usage-based conception of language. In Barlow, Michael & Kemmer, Suzanne (eds.), *Usage-based models of language*, vii–xxviii. Stanford–California: CSLI Publications.

Erelt, Mati & Kasik, Reet & Metslang, Helle & Rajandi, Henno & Ross, Kristiina & Saari, Henn & Tael, Kaja & Vare, Silvi. 1993. *Eesti keele grammatika II. Süntaks. Lisa: Kiri* [Estonian grammar II: Syntax. Appendix: Orthography]. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

—— 1995. *Eesti keele grammatika I. Morfoloogia. Sõnamoodustus* [Estonian grammar I: Morphology. Word formation]. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.

Erelt, Mati & Erelt, Tiiu & Ross, Kristiina. 2007. *Eesti keele käsiraamat* [Handbook of Estonian]. Tallinn: Eesti Keele Sihtasutus.

Erelt, Tiiu & Leemets, Tiina & Mäearu, Sirje & Raadik, Maire. 2013. *Eesti õigekeelsussõnaraamat ÕS 2013* [The dictionary of standard Estonian ÕS 2013]. Tallinn: Eesti Keele Sihtasutus.

Hasselblatt, Cornelius. 2000. *Eesti keele ainsuse sisseütlev on lühike* [The illative singular in Estonian is short]. *Keel ja Kirjandus* 11. 796–803.

Hint, Mati. 1980a. *Minevikuline ja tulevikuline aines keelesüsteemis. Prosoodiatüübi nihked ja selle tagajärjed* [Past and future subject matter in the language system. Prosody type shift and its consequences]. *Keel ja Kirjandus* 4. 215–223.

—— 1980b. *Minevikuline ja tulevikuline aines keelesüsteemis. Prosoodiatüübi nihked ja selle tagajärjed* [Past and future subject matter in the language system. Prosody type shift and its consequences]. *Keel ja Kirjandus* 5. 270–278.

—— 1980c. *Minevikuline ja tulevikuline aines keelesüsteemis. Prosoodiatüübi nihked ja selle tagajärjed* [Past and future subject matter in the language system. Prosody type shift and its consequences]. *Keel ja Kirjandus* 6. 349–355.

Kaalep, Heiki-Jaan. 2009. *Kuidas kirjeldada ainsuse lühikest sisseütlevat kasutamisandmetega kooskõlas?* [How to describe the short illative singular in harmony with usage data]. *Keel ja Kirjandus* 6. 411–425.

Kio, Kati. 2006. *Sisseütleva käände kasutus eesti kirjakeeles* [The use of the illative case in standard Estonian]. Tartu: University of Tartu. (Master's thesis.)

Klavan, Jane & Pilvik, Maarja-Liisa & Uiboaed, Kristel. 2015. The use of multivariate statistical classification models for predicting constructional choice in spoken, non-standard varieties of Estonian. *SKY Journal of Linguistics* 28. 187–224.

Langacker, Roland W. 1987. *Foundations of cognitive grammar, Vol. 1: Theoretical prerequisites.* Stanford: Stanford University Press.

Langemets, Margit & Tiits, Mai & Valdre, Tiia & Veskis, Leidi & Viks, Ülle & Voll, Piret. 2009. *Eesti keele seletav sõnaraamat* [The explanatory dictionary of the Estonian language]. Eesti Keele Sihtasutus.

Metslang, Ann. 2015. *Ainsuse pika ja lühikese sisseütleva valiku olenemine morfofonoloogilistest tunnustest: Korpusanalüüs* [How the choice between the singular long and short illative case depends on the type of morphophonological variables: A corpus analysis]. *Emakeele Seltsi aastaraamat* 60 (2014). 127–147.

Mäearu, Sirje. *Valik rektsioone* [Choice of governments]. (http://keeleabi.eki.ee/artiklid/rektsioone.html) (Accessed 2015-09-23.)

Raag, Virve. 1998. *The effects of planned change on Estonian morphology*. Uppsala: Studia Uralica Upsaliensia 29. (Doctoral dissertation.)

Rajandi, Henno. 1963. *Morfeemide liigitus deskriptiivses lingvistikas* [Classification of morphemes in descriptive linguistics]. *Keel ja Kirjandus* 7. 409–414.

Ruutma, Mirjam & Kyröläinen, Aki-Juhani & Pilvik, Maarja-Liisa & Uiboaed, Kristel. 2016. *Ambipositsioonide morfosüntaktilise varieerumise kirjeldusi kvantitatiivsete profiilide abil* [Descriptions of the morphosyntactic variation of ambipositions by means of quantitative profiles]. *Keel ja Kirjandus* 2. 92−113.

Siiman, Ann. 2016. Ainsuse sisseütleva vormi valiku seos morfosüntaktiliste ja semantiliste tunnustega: Materjali ning meetodi sobivus korpusanalüüsiks [How the choice between the singular long and short illative case is related to morphosyntactic and semantic variables: Which material and methods are suitable for a corpus analysis]. *Emakeele Seltsi aastaraamat* 61 (2015). 207−232.

Sõnajalg, Ingrid. 1956. *Illatiiv eesti kirjakeeles* [The illative case in standard Estonian]. Tartu: University of Tartu. (Diploma thesis.)

Strobl, Carolin & Malley, James & Tutz, Gerhard. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods* 14(4). 323–348.

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

Viitso, Tiit-Rein. 1976. *Eesti muutkondade süsteemist* [On the inflection system in Estonian]. *Keel ja Kirjandus* 3. 148–162.

Viks, Ülle. 1992. *Väike vormisõnastik I. Sissejuhatus ja grammatika* [A concise morphological dictionary of Estonian I. Introduction and grammar]. Tallinn: Eesti Teaduste Akadeemia, Keele ja Kirjanduse Instituut.

Contact information:

Ann Siiman
Institute of Estonian and General Linguistics
University of Tartu
Jakobi 2
51014 Tartu
Estonia
e-mail: ann(dot)siiman(at)ut(dot)ee